Progress in Probability

Volume 59

Series Editors

Charles Newman
Sidney I. Resnick

Seminar on Stochastic Analysis, Random Fields and Applications V

Centro Stefano Franscini, Ascona, May 2005

Robert C. Dalang
Marco Dozzi
Francesco Russo
Editors

Birkhäuser Verlag
Basel · Boston · Berlin

Editors:

Robert C. Dalang
Institut de Mathématiques
Ecole Polytechnique Fédérale
CH-1005 Lausanne
Switzerland
e-mail: robert.dalang@epfl.ch

Marco Dozzi
Institut Elie Cartan
Université Henri Poincaré
B.P. 239
F-54506 Vandoeuvre-lès-Nancy Cedex
France
e-mail: dozzi@iecn.u-nancy.fr

Francesco Russo
Département de Mathématiques
Institut Galilée
Université Paris 13
F-95430 Villetaneuse
France
e-mail: russo@math.univ-paris13.fr

2000 Mathematics Subject Classification: Primary 60-06; 34A34, 35B35, 35K55, 35K57, 35Q53, 37A25, 37H05, 37L05, 37L40, 39B62, 46B09, 46E30, 47G99, 49N15, 60Gxx, 60Hxx, 60J60, 62M45, 65C30, 65G99, 76F05, 76M35, 82D30, 91B28, 93E20. Secondary 34F05, 35R60, 46N10, 49L20, 49L25, 53D10, 62P05, 76F20, 76M35, 91A23, 91A25, 91A80, 91B16, 91B30, 91B70

Library of Congress Control Number: 2007938508

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

ISBN 978-3-7643-8457-9 Birkhäuser Verlag, Basel – Boston – Berlin

Basel · Boston · Berlin
P.O. Box 133, CH-4010 Basel, Switzerland
Part of Springer Science+Business Media
Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Printed in Germany

ISBN 978-3-7643-8457-9 e-ISBN 978-3-7643-8458-6

9 8 7 6 5 4 3 2 1 www.birkhauser.ch

Table of Contents

Preface

This volume contains the Proceedings of the *Fifth Seminar on Stochastic Analysis, Random Fields and Applications*, which took place at the Centro Stefano Franscini (Monte Verità) in Ascona (Ticino), Switzerland, from May 30 to June 3, 2005. *All papers in this volume have been refereed.*

The previous four editions of this conference occured in 1993, 1996, 1999 and 2002. This Seminar focused on fundamental aspects of stochastic analysis, such as stochastic partial differential equations (spde's) and random fields, but also emphasized applications to fields such as biostochastics, stochastic turbulence and, as in the previous editions, financial mathematics, which was the subject of the *Fifth Minisymposium on Stochastic Methods in Financial Models.*

One of the traditional topics of the Seminar, where a significant part of the organizers' research activity is located, is the area of *stochastic partial differential equations* and more generally infinite-dimensional diffusions. The state of the art of a large part of this subject was presented in several lectures that covered porous media equations, well-posedness for degenerate equations, pathwise integral methods, Navier-Stokes equations, and numerical schemes for spde's. In particular, the study of the equations which are related to fluid mechanics presents many challenging open questions. Within the broad area of random fields, in addition to the study of spde's, there is much activity concerning *random media* in discrete and continuous environments. One typical example of such a model is a system of stochastic differential equations where the drift is a random (and very often an irregular) field.

Among the areas of application we mentioned, research in *biostochastics* is developing in several separate directions. This conference covered: *neuroscience*, in which time series and dynamical systems, but also Gaussian random fields, are important tools; *genomic analysis*, which makes strong use of tools from probability theory such as hidden Markov chains; and *adaptive population evolution*, which naturally involves spde's and measure-valued processes.

As mentioned above, several talks were devoted to *turbulence*. In particular, recent investigations in partial differential equations such as Burgers, Euler and Navier-Stokes equations with stochastic perturbations were presented. As in the volume devoted to the Seminar of 2002, pathwise stochastic methods have also been implemented in several vortex filament models.

Concerning *financial mathematics*, an intense area of activity concerns random volatility models: different types of mean reverting processes, which are Markovian or have long memory, are used to describe the evolution of volatility. Another development makes use of statistical non-parametric estimates of the volatility process to filter market microstructure contaminations. Further research in this broad field is directed towards the valuation of volatility derivatives. Infinite-dimensional stochastic analysis (Dirichlet forms and Malliavin calculus) are used here for sensitivity analysis and for market stability indicators.

Two important topics, which were beginning to emerge during the previous edition of this conference, received special attention:

- *Energy and other commodity markets.* The continuing worldwide process of electricity market deregulation has turned the analysis of the structure of electricity prices into a topic of central importance. Electricity is different from other commodities due to its non storable nature. In this context, infinite-dimensional tools borrowed from the analysis of the term structure of interest rates appear, as well as Lévy processes in order to take into account price peeks.
- *Detection of insider trading.* This remains a challenging subject because of the problems caused by defaults of large companies, which are not anticipated by rating agencies. At the mathematical level, enlargement of filtration techniques and forward stochastic integrals, but also game theory approaches, are used to analyse the asymmetric behaviour of agents.

The *Minisymposium on Stochastic Methods in Financial Models* took place on June 2 and 3. As in other editions of these Ascona conferences, one afternoon was devoted to interactions between practitioners and the academic community. In the first part of the afternoon, talks by Marek Musiela (BNP Paribas), Franco Moriconi (Università di Perugia) and Hélyette Geman (Essec and Paris Dauphine) were featured. The afternoon concluded with two conferences coorganized with the "Dipartimento dell' Economia e delle Finanze" of Ticino, both devoted to energy markets. The session was opened by Paolo Rossi (Director of the "Azienda Elettrica Ticinese") whose presentation was titled *"Energy markets: the increasing need for analyzing data"*: it described the current state of the electricity market in Switzerland. This was followed by the lecture of Prof. René Carmona (Princeton University) whose title was *"Energy trading: new challenges in financial mathematics"*.

Significant financial support for this meeting was provided by the Fonds National Suisse pour la Recherche Scientifique (Berne), the Centro Stefano Franscini (ETH Zürich), and the Ecole Polytechnique Fédérale de Lausanne (EPFL). We take this opportunity to thank these institutions.

Robert C. Dalang, Marco Dozzi and Francesco Russo
May 2007

List of Participants

Andersson, D.	Stockholm, Sweden
Aubin, J.-P.	Paris Dauphine, France
Barndorff-Nielsen, O.E.	Aarhus, Denmark
Becker, C.	Deutsche Bank Frankfurt, Germany
Ben Alaya, M.	Paris 13, France
Benaïm, M.	Neuchâtel, Switzerland
Bernyk, V.	EPF Lausanne, Switzerland
Bessaih, H.	Wyoming, U.S.A.
Biagini, Sara	Perugia, Italy
Bonaccorsi, S.	Trento, Italy
Boufoussi, B.	Marrakech, Morocco
Bouleau, N.	ENPC Paris, France
Carmona, R.	Princeton University, U.S.A.
Castell, F.	Marseille, France
Cohen, S.	Toulouse, France
Cont, R.	Ecole Polytechnique, France
Conus, D.	EPF Lausanne, Switzerland
Corcuera, J.-M.	Barcelona, Spain
Coutin, L.	Toulouse, France
Coviello, R.	Paris 13, France and SNS Pisa, Italy
Cresson, J.	Besançon, France
Cruzeiro, A.B.	IST Lisbon, Portugal
Da Prato, G.	SNS Pisa, Italy
Dalang, R.C.	EPF Lausanne, Switzerland
Darses, S.	Besançon, France
De La Pena, V.	Columbia University, U.S.A.
Debbi, L.	Nancy 1, France
Dozzi, M.	Nancy 2, France
Eberlein, E.	Freiburg, Germany
Elworthy, K.D.	University of Warwick, U.K.
Flandoli, F.	Pisa, Italy
Gradinaru, M.	Nancy 1, France
Grothaus, M.	Kaiserslautern, Germany
Guasoni, P.	Boston University, U.S.A.
Gubinelli, F.	Paris Orsay, France
Haba, Z.	Wroclaw, Poland
Hilbert, A.	Växjö, Sweden
Hu, Y.	Paris 13, France
Huillet, T.	Cergy Pontoise, France
Khoshnevisan, D.	The University of Utah, U.S.A.

Komorowski, T.	Lublin, Poland
Kruk, I.	Paris 13, France
Léandre, R.	Dijon, France
Le Gland, F.	IRISA Rennes, France
Lescot, P.	Saint-Quentin, France
Lisei, H.	Cluj-Napoca, Romania
Madan, D.	Maryland, U.S.A.
Malliavin, P.	Académie des Sciences, France
Marty, W.	Crédit Suisse Zurich, Switzerland
Mazumdar, R.R.	Waterloo, Canada
Méléard, S.	Paris 10, France
Merlini, D.	CERFIM, Switzerland
Mocioalca, O.	University of Kent, U.S.A.
Moriconi, F.	Perugia, Italy
Mulinacci, S.	Milano, Italy
Musiela, M.	BNP Paribas London, U.K.
Nourdin, I.	Paris 6, France
Nualart, E.	Paris 13, France
Øksendal, B.	Oslo, Norway
Pardoux, E.	Marseille, France
Perkins, E.	University of British Columbia, Canada
Piera, F.	Purdue University, U.S.A.
Pratelli, M.	Pisa, Italy
Privault, N.	Poitiers, France
Prum, B.	Evry, France
Röckner, M.	Bielefeld, Germany
Rodriguez, A.	Stockholm, Sweden
Rossi, P.	Azienda Elettrica Ticinese, Switzerland
Roynette, B.	Nancy 1, France
Rüdiger, B.	Koblenz, Germany
Runggaldier, W.J.	Padova, Italy
Russo, F.	Paris 13, France
Sanz-Solé, M.	Barcelona, Spain
Saussereau, B.	Besançon, France
Scandolo, G.	Florence, Italy
Scheutzow, M.	Berlin, Germany
Schmidt, W.	HfB Frankfurt, Germany
Schmiegel, J.	Aarhus, Denmark
Scotti, S.	ENPC-Cermics, France
Simao, I.	Lisbon, Portugal
Stannat, W.	Darmstadt, Germany
Stricker, C.	Besançon, France
Sturm, K.-T.	Bonn, Germany

Sulem, A.	INRIA Rocquencourt, France
Svensson, J.	Stockholm, Sweden
Thieullen, M.	Paris 6, France
Tindel, S.	Nancy 1, France
Truman, A.	Swansea, U.K.
Trutnau, G.	Bielefeld, Germany
Tudor, C.	Paris 1, France
Valkeila, E.	Helsinki, Finland
Vallois, P.	Nancy 1, France
Vargiolu, T.	Padova, Italy
Viens, F.G.	Purdue, U.S.A.
Villa, A.	Grenoble, France
Vizcarra, A.B.	Purdue University, U.S.A.
Walsh, J.B.	University of British Columbia, Canada
Zakaï, M.	Haifa, Israel
Zambrini, J.-C.	Lisbon, Portugal
Zegarlinski, B.	Imperial College London, U.K.

Stochastic Analysis and Random Fields

Progress in Probability, Vol. 59, 3–17

Detection of Dynamical Systems from Noisy Multivariate Time Series

Yoshiyuki Asai and Alessandro E.P. Villa

Abstract. Experimental observations of physical, social, or economical systems may often be reduced to multivariate time series. The observed time series may be investigated as random processes or realizations of stochastic dynamical systems. Studies of natural phenomena should consider that the time series are affected by a random noise such that some realizations of the underlying dynamical system are missed by the observer and some observations correspond to the realizations of a stochastic process associated to the method of measurement. Within this framework we consider discrete time series derived from mappings by the iterations of one observable, typically one of the system's coordinates. The time series were altered by several levels of noise and we show that a pattern detection algorithm was able to detect temporal patterns of events that repeated more frequently than expected by chance. These patterns were related to the generating attractors and were robust with respect to the appearance of spurious points due to the noise. On the basis of this result we propose a filtering procedure aimed at decreasing the amount of noisy events in time series.

Mathematics Subject Classification (2000). 62M45.

Keywords. Multivariate time series, pattern detection algorithm, dynamical systems.

1. Introduction

Discrete time series can represent the occurrences of either a deterministic or a random process. Dynamical system theory provides powerful techniques to assess whether a set of equations (in a suitable embedding space) underlies the dynamics [1, 8, 9, 12, 13, 18, 21, 23, 24]. Beside the characterization of the embedding space, nonlinear time series analysis can determine topological and metric invariants [17]. Physical, social, or economical systems may include deterministic processes, but their observation is limited by the quality of the measurements. The precision of the observation cannot be absolute. It appears that studies of natural phenomena should consider that the observed time series are affected by a random noise such that some realizations are missed by the observer and some observations are not associated to the dynamical system but correspond to the realizations of a stochastic process that depends on the method of measurement. The possibility to filter out the noisy components from the time series observed in nature may be a clue to ascertain the deterministic feature of the underlying dynamical process and to study the topological characteristics of the attractor.

The study of neural dynamics is particularly interesting to this respect [28]. In neurophysiological experiments the discrete time series are obtained from the epochs of action potentials of nervous cells (i.e., *spike trains*). Chaotic determinism in the dynamics of spiking neural networks has been observed in experimental data [7, 10, 19, 22, 23]. This behavior was theoretically predicted and is considered as an important mechanism for representation of learned stimuli in large scale distributed networks [6, 14]. The "synfire chain theory" [2, 3], based on topological assumptions of diverging/converging feed-forward layers of neurons, suggests that whenever the same process repeats in a cell assembly in the brain, the same spatio-temporal firing patterns should appear. Synfire chains may exhibit structures in which a group of neurons excite themselves and maintain elevated firing rates for a long period. Let us note that the synfire chain theory emphasizes the importance of precise timing of spikes (precise temporal coding), while theories of attractor neural networks, generally speaking, do not require it (noisy rate coding).

In the present study we show that a particular pattern detection algorithm developed for the study of temporal activity in electrophysiological recordings [26, 29] is particularly well suited to detect deterministic dynamics in the presence of noise. Starting from mathematically defined mappings such as Hénon, Zaslavskii and Ikeda maps, the algorithm was able to detect temporal patterns of events that repeated more frequently than expected by chance even in presence of an increasing level of observational noise (some points were deleted at random and an equal number of points added at random). The points belonging to all detected patterns were merged together in order to form a reconstructed time series. The reconstructed time series represented a significant fraction of the original points and is related to the generating attractors. On the basis of this result we propose a filtering procedure aimed at decreasing the amount of noisy events in time series.

2. Methods

2.1. Mappings

Below is a list of dissipative mappings that were anlayzed in the present study.

Hénon mapping (2–dimensional)

It is defined by the equations

$$\begin{cases} x_{n+1} = -ax_n^2 + y_n + 1, \\ y_{n+1} = bx_n, \end{cases}$$

$x, y \in \mathbb{R}$, $a, b \in \mathbb{R}$. Let $x_0 = 0.6$, $y_0 = 0.19$ be the initial conditions with parameters $a = 1.6$ and $b = 0.1$.

Zaslavskii map

It is defined by the equations

$$\begin{cases} x_{n+1} = x_n + v(1 + \mu y_n) + \varepsilon v \mu \cos x_n, & (mod.\ 2\pi) \\ y_{n+1} = e^{-\gamma}(y_n + \varepsilon \cos x_n)\ , & \end{cases}$$

where $x, y \in \mathbb{R}$, the parameters are real numbers with $\mu = \frac{1-e^{-\gamma}}{\gamma}$, $v = \frac{4}{3} \cdot 100$. The initial conditions were set to $x_0 = 0.3$ and $y_0 = 0.3$.

Ikeda map

Let

$$z_{n+1} \ = \ p + B \ z_n \ e^{ik - i\alpha/(1+|z_n|^2)}\ , \tag{2.1}$$

where $z \in \mathbb{C}$ and p, B, k, $\alpha \in \mathbb{R}$. We rewrite Eq. (2.1) in its real form as

$$\begin{cases} x_{n+1} = p + B\cos\left(k - \frac{\alpha}{1+x_n^2+y_n^2}\right) x_n - B\sin\left(k - \frac{\alpha}{1+x_n^2+y_n^2}\right) y_n, \\ y_{n+1} = B\cos\left(k - \frac{\alpha}{1+x_n^2+y_n^2}\right) y_n + B\sin\left(k - \frac{\alpha}{1+x_n^2+y_n^2}\right) x_n. \end{cases}$$

We take $p = 1.0$, $B = 0.9$,$k = 0.4$ and $\alpha = 6.0$. The initial conditions were set to $x_0 = 0.3$ and $y_0 = 0.3$.

2.2. Time series

For each mapping described above a new time series $\{W_n\}$ was derived by taking the difference between two consecutive values of the $\{X_n\}$ series and adding a constant K such that $w_n > 0$, $w_n = x_{n+1} - x_n + K$. In order to have data with comparable timing dynamics found in usual neurophysiological experiments, the time series generated from the mapping were scaled in order to have, on average, a base frequency of 3 *events/sec* (i.e., 3 *spikes/s* for neuophysiological data). This means that each point in the time series corresponded to an event in time observed with a 1 *ms* resolution. Ten thousand points ($N = 10,000$) were generated in each series.

The observational noise was simulated by inclusion and deletion of points in the time series and by adding a jitter. Three levels of observational noise were considered: 10%, 20% and 30%. The procedure to generate a noisy time series, given an original $10,000$ points time series was the following. Firstly, the list of points to be deleted was determined by chance assuming an uniform distribution and the given level of noise. In the case of 20% observational noise 20% of the points belonging to the original time series $\{W_n\}$, which was derived according to the mapping equations, were deleted at random. Then, the resulted time series $\{W'_n\}$ contained $0.8 \times N$ points. Secondly, each remaining point w'_i belonging to the series $\{W'_n\}$ was shifted in time by a variable jitter ΔJ distributed uniformly such that $w''_i \in [w' - \Delta J, w' + \Delta J]$. Thirdly, an amount of points, equal to the amount of points that were deleted, was added to the time series $\{W''_n\}$. The added points were generated according to a uniform distribution on the actual interval and were inserted in the time series in such a way that an inserted point could never overlap an existing point. If this overlap occurred, then the position of insertion of the point in the time series was reselected at random. The procedure ended when the number of added points was equal to the number of deleted points. Then, the noisy time series $\{W_n'''\}$ contained exactly N points. Notice that this is a symmetrical case of observational noise. We are currently studying the effect of cases where the two types of noise follow different rates and different distributions.

2.3. Detection of temporal patterns

Temporal patterns of events were detected by applying the Pattern Grouping Algorithm (PGA), designed to identify and evaluate the statistical significance of temporal patterns of spikes formed by three or more different events with slight differences in spike timing [26, 29]. The three adjustable parameters in PGA include the maximal duration of the pattern measured as a delay between the first and the last spike in the sequence of spikes (i.e., the window duration), the level of significance to be used for detection of significant groups, and the upper bound of allowed jitter applied to all the groups. Fig. 1 illustrates the application of PGA to a case study.

The main principles of the PGA algorithm can be outlined as follows. The algorithm can search and cluster individual patterns which differ from each other by a small jitter in spike timing of the order of few ms. The estimation of significance of the detected patterns is done according to three different tests. The first test is an extension of the Pattern Detection Algorithm, PDA [5], which does not rest on the assumption that the spike trains behave like Poisson processes, but just on the assumption that at any time instance t the probability of getting one pre-specified pattern is very low. However, such assumption is not valid for spikes occurring in a burst that can be modeled by non-stationary Poisson processes with high firing rate fluctuation. Two additional tests of significance, FPD, a modified version of Favored Pattern Detection [11], and JTH, Joint Triplet Histogram [20] were applied and only those patterns that passed the three tests were kept for further analysis. The criteria used here for identifying the significant patterns were that they included at least 3 events (triplets), within the same time series, repeated at least 5 times within the time series, lasted less than 800 ms (*window* duration), repeated with an accuracy of ± 5 ms and the significance level be at least 5%.

3. Results

In this section we illustrate in detail the reconstruction procedure for one specific case. From this example it is easy to derive the procedure used for all mappings. The case study is based on the detection of a pattern in the Zaslavskii map with 20% noise. Notice that the PGA algorithm detects a number of repetitive patterns that depends on parameters such as the window duration and the jitter (see [25] for a discussion of this point) and also on the type of mapping.

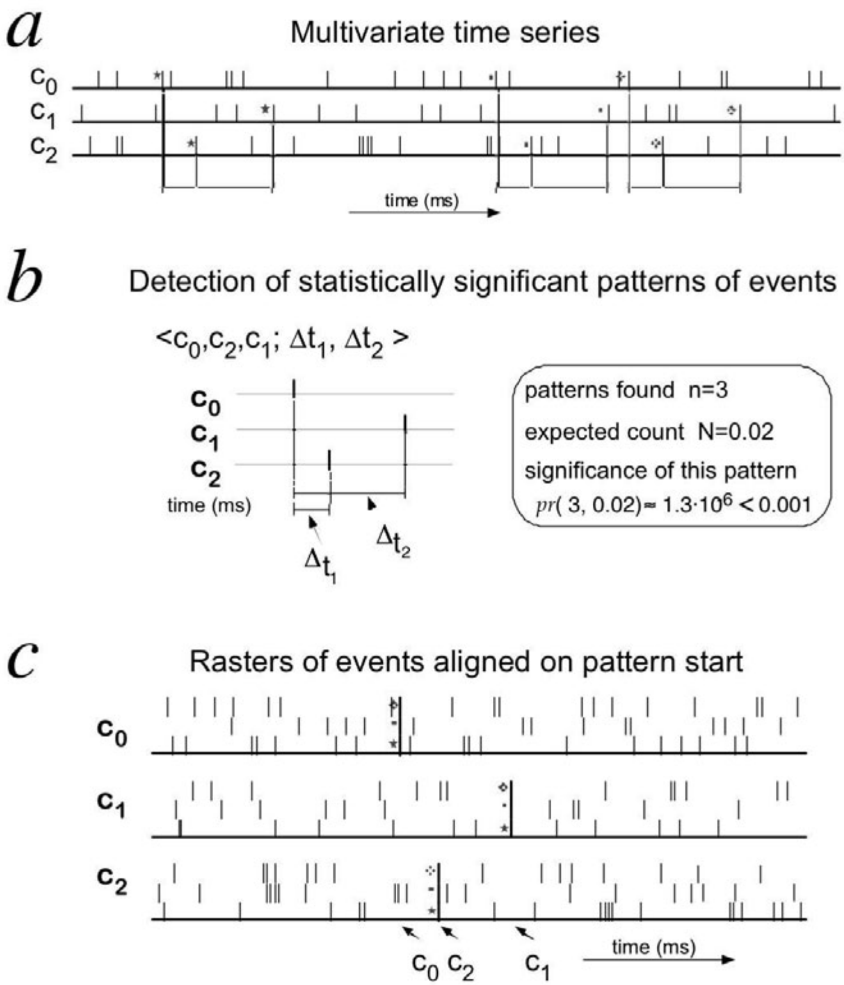


FIGURE 1. Outline of the general procedure followed by pattern detection algorithms. a. Analysis of a set of simultaneously recorded time series. Three variables, labeled c_0, c_1, and c_2, participate to the multivariate time series. In this example three occurrences of a precise pattern are detected. Each occurrence of the pattern has been labeled by a specific marker in order to help the reader to identify the corresponding points. b. Estimation of the statistical significance of the detected pattern. c. Display of pattern occurrences as a raster plot aligned on the pattern start.

The first step always consisted in analyzing the original time series $\{W_n\}$, without noise. In the specific case of the Zaslavskii map the PGA algorithm, with window duration equal to 800 ms and jitter equal to 5 ms, found 153 significant patterns formed by three points (*triplets*) and found 107 significant patterns formed by four points (*quadruplets*). The set formed by these patterns is referred to as the "original" set of patterns. Each pattern of the set appeared several times in the analyzed series and the union of all points belonging to all repeating triplets formed the "reconstructed" time series $\{R_n\}$. A reconstructed time series is formed by two sets of points. The first set, denoted $R \cap W$, is formed by the points that belonged to the original time series and the second set, denoted R_S, is formed by the spurious points, i.e., the points that were introduced by the noise. Among the points of the first set it is important to distinguish the subset, denoted $R \cap R^0$, which is formed by all points that were also observed in the reconstructed series in the absence of noise from the subset, denoted R_B, which is formed by those points that belong to the original time series W but that were not observed in the reconstructed series without noise R^0. From the logical calculus the subset R_B is defined as $R_B = (R \cap W) \cap \neg R^0$. Fig. 2 illustrates these sets of points. In the case of Zaslavskii without noise, the reconstructed series, denoted by $\{R^0_n\}$, included 7148 points ($\approx 71\%$ of the original series). In the case of Zaslavskii with noise level 20%, the PGA algorithm found 56 significant triplets and 17 significant quadruplets. In this case, with a jitter ± 5 ms, we observed that more than one third of the patterns found in the noisy file (triplets: $n = 21/56$, 38%; quadruplets: $n = 6/17$, 35%) belonged also to the original set of patterns. The remaining patterns were found *only* in the noisy time series.

The general notation for a triplet is $< c_0, c_1, c_2 \ ; \ \Delta t_1, \Delta t_2 >$ where c_0, c_1, c_2 refer to the label of the variables of the time series in the case of multivariate time series (Fig. 1). In the case of data from only one series all triplets are formed by points from the same series and by default the notation is $c_0 = c_1 = c_2 = \#1$. The pattern start is set by definition at $t_{start} = 0$, Δt_1 refers to the lag of the second point from pattern start and Δt_2 to the lag of the third point of the triplet.

Let us examine the case of one particular triplet found in both original and noisy time series. The triplet denoted $< 1, 1, 1; 444, 625 >$ means that the second point of the triplet occurs 444 ms after the first event and the third point 625 ms from pattern start. The triplet $< 1, 1, 1; 444, 625 >$ occurred 131 times in the original Zaslavskii time series (Fig. 3.a) and the triplet $< 1, 1, 1; 445, 625 >$ occurred 46 times in the time series with 20% noise level (Fig. 3.b). We assume that these triplets represent the same fundamental event derived from the original dynamical system. It is norteworthy that 33 out of 46 triplets observed in the noisy time series were found both in the original and noisy set. This means that the noise let appear 13 new triplets that were not observed previously. It is interesting to notice the standard deviations of the occurrences of Δt_1 and Δt_2: in the pattern $< 1, 1, 1; 444, 625 >$ (original set) $\sigma_{\Delta t_1} = 1.8$ ms and $\sigma_{\Delta t_2} = 3.8$ ms; in the pattern $< 1, 1, 1; 445, 625 >$ (noisy set) $\sigma_{\Delta t_1} = 2.7$ ms and $\sigma_{\Delta t_2} = 4.3$ ms. Notice that additional patterns characterized by longer lags appear visually in both panels of

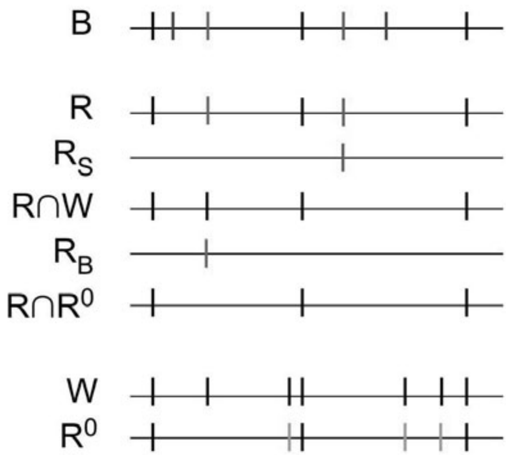


FIGURE 2. Sets of points referred in the time series. W: points in the original time series; B: points in the noisy time series; R: points in the reconstructed time series; R^0: points in the reconstructed time series in the absence of noise. See text for the other definitions.

Fig. 3. Such patterns are missed by PGA either because of their occurrance for window durations larger than the parameter used for pattern search (800 ms for the window duration in this study) or because of their too large jitter (± 5 ms in this study).

The procedure described above was used iteratively and the absolute epochs of the points that belonged to the triplets were recorded in order to form the reconstructed time series. The return map determined by two consecutive inter-events intervals can be used to plot the trajectory of the dynamical system, hence to project the orbits of the attractors. The application of this technique to the mappings of Hénon, Zaslavskii and Ikeda, with various levels of noise, is illustrated at Fig. 4, Fig. 5 and Fig. 6, respectively. These figures show that the procedure is effective in filtering much of the noise and can be considered a filtering procedure. Table 1 gives some quantitative data on the efficiency of the filtering with respect to the original reconstructed time series without noise.

In this table it is interesting to notice that even in presence of 30% noise level the amount of points in the reconstructed time series always included a majority of points that belonged to the original time series (in the range 78% for the Ikeda map to 85% Hénon map). This result indicates that the usage of the PGA algorithm was quiet robust with respect to the appearance of spurious points, which are totally due to the added noise. In Table 1 it is of interest that even in the absence of noise it may be difficult to reconstruct the original time series using the algorithm with the

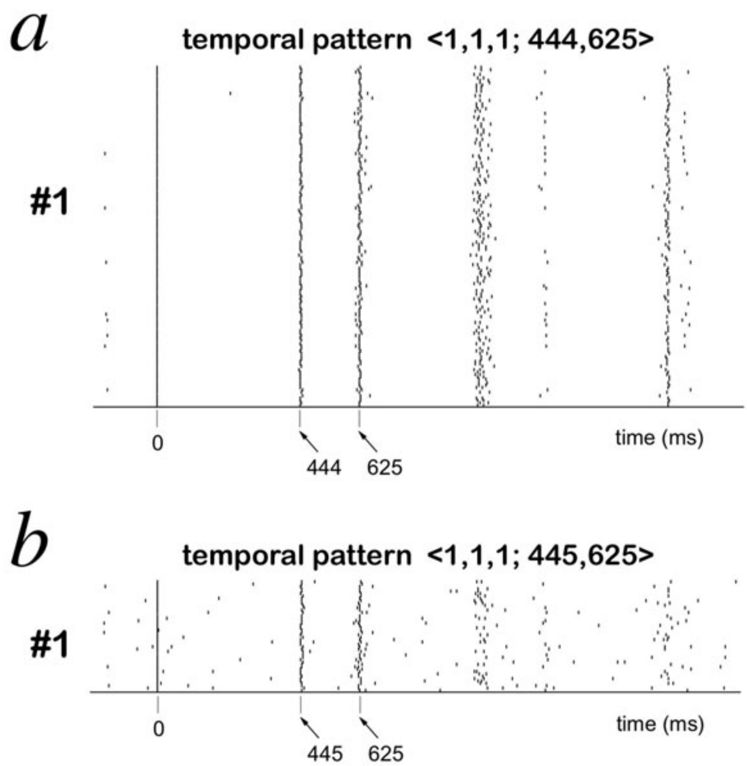


FIGURE 3. Raster display, aligned by displaying the first event in the pattern at time 0, of the activity of a simulated neuron whose dynamics was generated according to the Zaslavskii mapping without noise (see text for the initial conditions and parameters values). a. The pattern repeated 131 times and was composed of 3 events: an event at time 0, the second 444 ms later, and the third event 625 ms later. The abscissa full scale is 2000 ms. The triplet was detected with a fixed accuracy of ± 5 ms. b. The triplet repeated 46 times and was detected within a time series corresponding to the same dynamical system of panel a. with a 20% noise level.

selected parameters. Indeed, in the case of the Ikeda map only 38% of the original points could be found by using PGA vs. 94% for the Hénon mapping. However, the most counterintuitive result is that the noise may help to find patterns that are not belonging to the reconstructed time series R^0. In the case of Ikeda map

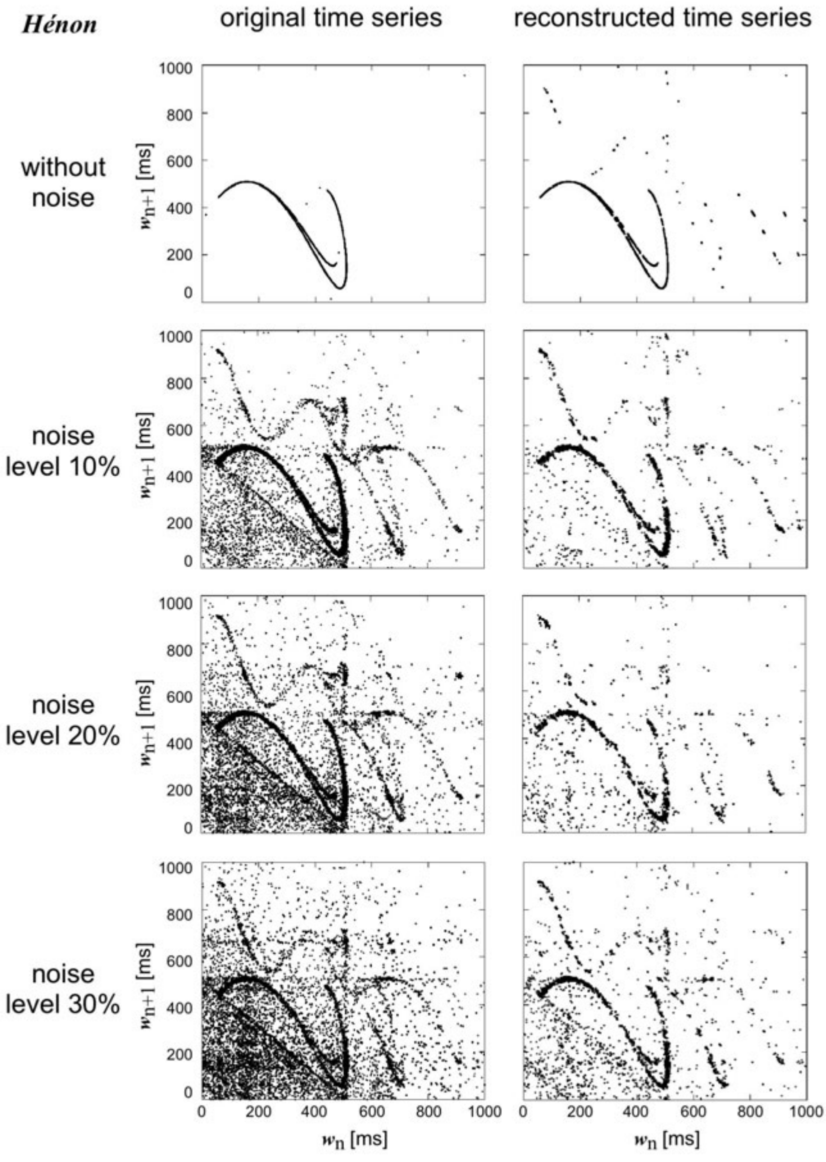


FIGURE 4. PGA based filtering procedure applied to the 2-dimensional Hénon mapping. The left panels show the original return maps with an increasing level of noise (from top to bottom). The right panels show the corresponding return maps obtained from the reconstructed time series.

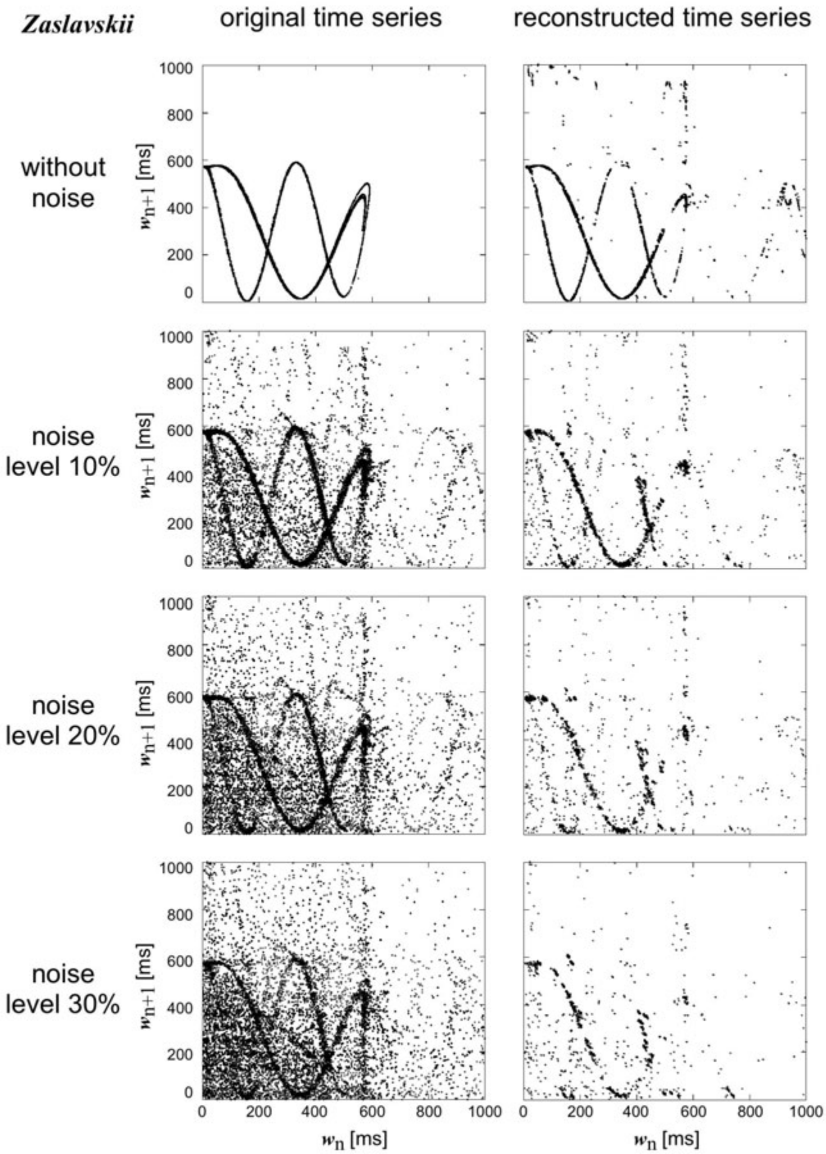


FIGURE 5. PGA based filtering procedure applied to the Zaslavskii mapping. The left panels show the original return maps with an increasing level of noise (from top to bottom). The right panels show the corresponding return maps obtained from the reconstructed time series.

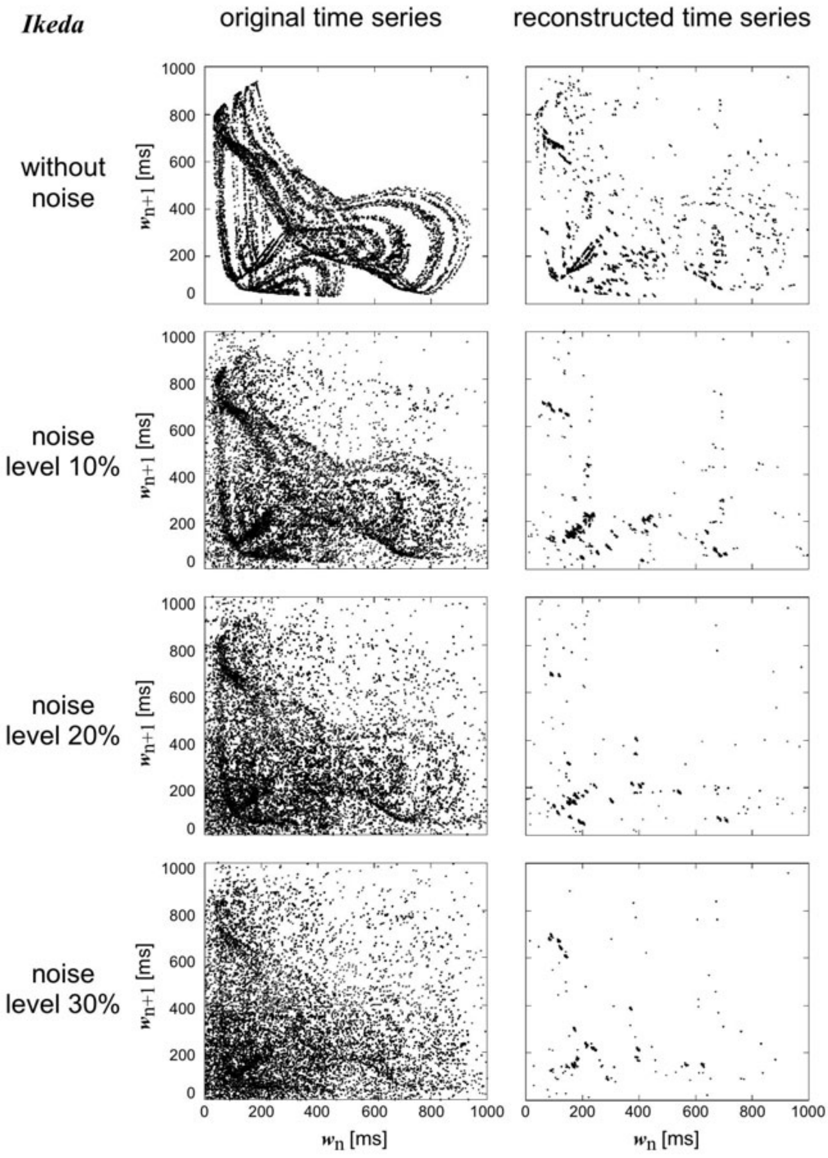


FIGURE 6. PGA based filtering procedure applied to the Ikeda mapping. The left panels show the original return maps with an increasing level of noise (from top to bottom). The right panels show the corresponding return maps obtained from the reconstructed time series.

Mapping	Noise level	Points in the reconstructed time series				
		$R \cap R^0$	R_B	$R \cap W$	R_S	R
Hénon						
	0%	9427	0	9427	0	9427
	10%	4625	168	4793	198	4991
	20%	3228	207	3435	338	3773
	30%	2294	120	2414	436	2850
Zaslavskii						
	0%	7148	0	7148	0	7148
	10%	3460	524	3984	197	4181
	20%	2324	523	2847	299	3146
	30%	1619	329	1948	389	2337
Ikeda						
	0%	3859	0	3859	0	3859
	10%	987	590	1577	108	1685
	20%	477	423	900	179	1079
	30%	309	267	576	164	740

Table 1. Reconstructed time series for various mappings and for several noise levels. W: original time series; R: reconstructed time series by PGA; R^0: reconstructed time series by PGA without noise; $R \cap R^0$: points belonging to the original time series but not included in the reconstructed time series without noise; R_B: all points in the reconstructed time series that were part of the original series but not included in R^0; R_S: Spurious points that belong to the reconstructed time series but did not belong to the original series. See Fig. 2 for an illustration of the series.

with 30% of noise (last line of Table 1) about as many points ($n = 267$) of the original series were not found in the reconstructed series without noise ($n = 309$).

4. Discussion

The current study has presented evidence that time series derived by deterministic dynamics with chaotic attractors are able to produce patterns of events detectable by the PGA algorithm [26, 29]. Another algorithm inspired by very similar ideas has been presented recently [4] and should also be applied in order to validate further our approach. In presence of noise the PGA algorithm was able to reconstruct a time series which is mainly a subset of the original one. With noise levels as high as 30% our approach let produce a time series with only 15% spurious points in the case of the Hénon map and 22% of spurious points in the case of the Ikeda map. In all cases the amount of spurious points was below the rate of noise. It is important to emphasize that our study was not aimed at finding the optimal parameters of the PGA algorithm for the retrieval of the best reconstructed time series. We have been using the algorithm with parameters very much akin to its application in neurophysiological applications [27, 30]. We have demonstrated that this algorithm may retrieve significant points imbedded in a noisy time series and improve the quality of the data for subsequent study, e.g., by classical dynamical system analytical methods, which is of considerable interest for specialists working with practical application of time series analysis. To this aim we can foresee that a search for the optimization of the best choice of the parameters of the algorithm or the application of other algorithm aimed at detecting temporal patterns with variable jitters can provide better results than those presented here. An additional line of study would consist to investigate the effect of different types of noise on the performance of the algorithm. In all cases we suggest that the application of PGA offers as a valid filtering procedure to improve the study of dynamical systems described by noisy data.

A remark of interest concerns the consequences of this study for the interpretation of neural dynamics [28]. The observation of firing patterns in experimental data has been considered as a strong evidence for the existence of "synfire chains" [2, 3]. These structures are formed by diverging/converging feed-forward layers of neurons such that the synchronous activity in one layer can propagate the activity to the next layer with an extremely precise timing. The existence of such structures in the brain has not been demonstrated by any anatomical studies and remains an hypothesis given the experimental difficulty of such an investigation. Insofar the appearance of diverging/converging feed-forward structures has been studied only in computational experiments aimed to simulate critical steps in brain development [15, 16]. Our results show that the variance of the lags of the events belonging to the temporal pattern tended to increase with the lag from pattern start. A similar finding was observed in experimental studies aimed to support the synfire chain hypothesis [20]. In our case the patterns were generated by the dynamical system, in the absence of any synfire activity. This finding raises the question whether significant patterns of spikes detected in neurophysiological experimental data are due to the attractor behavior of neural networks dynamics (a possibility suggested by this study) or if they are produced by synfire chains. No final conclusion can be drawn from the data presented here but the current results support the hypothesis that spatio-temporal patterns of spikes and attractor dynamics may represent two faces of the same coin, as suggested elsewhere [28].

References

[1] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, and L.S. Tsimring, *The analysis of observed chaotic data in physical systems*, Reviews of Modern Physics, **65** (1993), 1331–1392.

[2] M. Abeles, *Local Cortical Circuits*, Springer Verlag, 1982.

[3] M. Abeles, *Corticonics*, Cambridge University Press, 1991.

[4] M. Abeles and I. Gat, *Detecting precise firing sequences in experimental data*, Journal of Neuroscience Methods, **107** (2001), 141–154.

[5] M. Abeles and G. Gerstein, *Detecting spatiotemporal firing patterns among simultaneously recorded single neurons*, J. Neurophysiol., **60** (1988), 909–924.

[6] D.J. Amit and N. Brunel, *Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex*, Cerebral Cortex, **7** (1997), 237–252.

[7] A. Babloyantz and J.M. Salazar, *Evidence of chaotic dynamics of brain activity during the sleep cycle*, Physics Letters A, **111** (1985), 152–155.

[8] G. Boffetta, A. Crisanti, F. Paparella, A. Provenzale, and A. Vulpiani, *Slow and fast dynamics in coupled systems: A time series analysis view*, Physica D, **116** (1998), 301–312.

[9] A. Celletti, C. Froeschlé, I.V. Tetko, and A.E.P. Villa, *Deterministic behaviour of short time series*, Meccanica, **34** (1999), 145–152.

[10] A. Celletti and A.E.P. Villa, *Low dimensional chaotic attractors in the rat brain*, Biological Cybernetics, **74** (1996), 387–394.

[11] J.E. Dayhoff and G.L. Gerstein, *Favored patterns in spike trains. I. Detection*, J. Neurophysiol., **49** (1983), 1334–1348.

[12] J.P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Reviews of Modern Physics, **57** (1985), 617–656.

[13] J. Gao and Z. Zheng, *Local exponential divergence plot and optimal embedding of a chaotic time series*, Physics Letters A, **181** (1993), 153–158.

[14] M. Herrmann, E. Ruppin, and M. Usher, *A neural model of the dynamic activation of memory*, Biological Cybernetics, **68** (1993), 455–463.

[15] J. Iglesias, J. Eriksson, B. Pardo, T. Tomassini, and A.E.P. Villa, *Emergence of oriented cell assemblies associated with spike-timing-dependent plasticity*, Lecture Notes in Computer Science, **3696** (2005), 127–132.

[16] E.M. Izhikevich, J.A. Gally, and G.M. Edelman, *Spike-timing dynamics of neuronal groups*, Cerebral Cortex, **14** (2004), 933–944.

[17] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis.* Cambridge University Press, 2003.

[18] D.T. Kaplan and L. Glass, *Direct test for determinism in a time series*, Physical Review Letters, **68** (1992), 427–430.

[19] G.J. Mpitsos, *Chaos in brain function and the problem of nonstationarity: a commentary*, in E. Basar and T. H. Bullock, editors, Dynamics of Sensory and Cognitive Processing by the Brain, Springer-Verlag, (1989), 521–535.

[20] Y. Prut, E. Vaadia, H. Bergman, I. Haalman, H. Slovin, and M. Abeles, *Spatiotemporal structure of cortical activity - properties and behavioral relevance*, J. Neurophysiol., **79** (1998), 2857–2874.

[21] P.E. Rapp. *Chaos in the neurosciences: cautionary tales from the frontier*, The Biologist, **40** (1993), 89–94.

[22] P.E. Rapp, I.D. Zimmerman, A.M. Albano, G.C. Deguzman, and N.N. Greenbaun, *Dynamics of spontaneous neural activity in the simian motor cortex: the dimension of chaotic neurons*, Physics Letters A, **110** (1985), 335–338.

[23] J.P. Segundo, *Nonlinear dynamics of point process systems and data*, International Journal of Bifurcation and Chaos, **13** (2003), 2035–2116.

[24] G. Sugihara and R.M. May, *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*, Nature, **344**, 734–741.

[25] I.V. Tetko and A.E. Villa, *A comparative study of pattern detection algorithm and dynamical system approach using simulated spike trains*, Lecture Notes in Computer Science, **1327** (1997), 37–42.

[26] I.V. Tetko and A.E.P. Villa, *A pattern grouping algorithm for analysis of spatiotemporal patterns in neuronal spike trains. 1. Detection of repeated patterns*, J. Neurosci. Meth., **105** (2001), 1–14.

[27] I.V. Tetko and A.E.P. Villa, *A pattern grouping algorithm for analysis of spatiotemporal patterns in neuronal spike trains. 2. Application to simultaneous single unit recordings*, J. Neurosci. Meth., **105** (2001), 15–24.

[28] A.E.P. Villa, *Empirical evidence about temporal structure in multi-unit recordings*, in R. Miller, editor, Time and the Brain, chapter 1, Harwood Academic Publishers, (2000), 1–51.

[29] A.E.P. Villa and I.V. Tetko, *Spatiotemporal activity patterns detected from single cell measurements from behaving animals*, Proceedings SPIE, **3728** (1999), 20–34.

[30] A.E.P Villa, I.V. Tetko, B. Hyland, and A. Najem, *Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task*, Proceedings of the National Academy of Sciences of the USA, **96** (1999), 1006–1011.

Yoshiyuki Asai
National Institute of Advanced Industrial Science and Technology (AIST)
Institute for Human Science and Biomedical Engineering
Tsukuba, Japan
and
NeuroHeuristic Research Group, INFORGE
Institute of Computer Science and Organization
University of Lausanne, Switzerland
e-mail: `yoshiyuki.asai@aist.go.jp`

Alessandro E.P. Villa
INSERM, U318, Grenoble, France
Laboratoire de Neurobiophysique, University Joseph Fourier, Grenoble, France
and
NeuroHeuristic Research Group, INFORGE
Institute of Computer Science and Organization
University of Lausanne, Switzerland
`http://www.neuroheuristic.org/`
e-mail: `alessandro.villa@ujf-grenoble.fr`

Progress in Probability, Vol. 59, 19–22

A Bakry-Emery Criterion for Self-Interacting Diffusions

Michel Benaïm and Olivier Raimond

Abstract. We give a Bakry-Emery type criterion for self-interacting diffusions on a compact manifold.

Mathematics Subject Classification (2000). 60J60.

Keywords. Processes with reinforcement, self-interacting diffusion, Bakry-Emery criterion, mass-transportation, Ricci curvature.

Let M be a smooth compact connected Riemannian manifold without boundary and $V : M \times M \to \mathbb{R}$ a smooth function. For every Borel probability measure μ on M let $V\mu : M \to \mathbb{R}$ denote the function defined by $V\mu(x) = \int_M V(x,u)\mu(du)$, and let $\nabla(V\mu)$ denote its gradient.

A *self-interacting diffusion process* associated to V is a continuous-time stochastic process $\{X_t\}$ which is a solution on M to the stochastic differential equation

$$dX_t = dW_t(X_t) - \frac{1}{2}\nabla(V\mu_t)(X_t)dt, \quad X_0 = x \in M,$$

where (W_t) is a Brownian vector field on M, and $\mu_t = \frac{1}{t}\int_0^t \delta_{X_s} ds$ is the *empirical occupation measure* of $\{X_t\}$.

This type of process with reinforcement was introduced in [2] and further studied in [3], [4], with the ultimate goal to:

(a) provide tools allowing us to analyze the long term behavior of $\{\mu_t\}$,
(b) understand the relations connecting this behavior to the nature of V, and,
(c) the geometry of M.

Let $\mathcal{P}(M)$ denote the space of Borel probability measures over M, λ the Riemannian probability on M and $\mathcal{P}_{cd}(M) \subset \mathcal{P}(M)$ the set of measures having a continuous density with respect to λ. Let X_V be the vector field defined on $\mathcal{P}_{cd}(M)$ by

$$X_V(\mu) = -\mu + \Pi_V(\mu)$$

We acknowledge financial support from the Swiss National Science Foundation grant 200021-1036251/1.

where
$$\frac{d\Pi_V(\mu)}{d\lambda} = \frac{e^{-V\mu}}{\int_M e^{-V\mu(y)}\lambda(dy)}.$$
Point **(a)** was mainly addressed in [2] where it was shown that the asymptotic behavior of $\{\mu_t\}$ can be precisely[1] described in terms of the deterministic dynamical system induced by X_V.

Depending on the nature of V, the dynamics of X_V can either be convergent, globally convergent or non-convergent, leading to a similar behavior for $\{\mu_t\}$. A key step toward **(b)** is the next result recently proved in [4].

Theorem 1. *Suppose V is a symmetric function. Then the limit set of $\{\mu_t\}$ (for the topology of weak* convergence) is almost surely a connected subset of $X_V^{-1}(0) = \mathsf{Fix}(\Pi_V)$.*

In (the generic) case where the equilibrium set $X_V^{-1}(0)$ is finite, Theorem 1 implies that $\{\mu_t\}$ converges almost surely. If furthermore, $X_V^{-1}(0)$ reduces to a singleton $\{\mu^*\}$, then $\{\mu_t\}$ converges almost surely to μ^* and we say that $\{\mu_t\}$ is *globally convergent.*

A function $K : M \times M \to \mathbb{R}$ is a *Mercer kernel* provided K is continuous symmetric and defines a positive operator in the sense that
$$\int_{M\times M} K(x,y)f(x)f(y)\lambda(dx)\lambda(dy) \geq 0$$
for all $f \in L^2(\lambda)$. The following result is proved in [4].

Theorem 2. *Assume that (up to an additive constant) V is a Mercer Kernel. Then $\{\mu_t\}$ is* globally convergent.

Example. Suppose $M \subset \mathbb{R}^n$ and $V(x,y) = f(-||x-y||^2)$ where $||\cdot||$ is the Euclidean norm of $\mathbb{R}^n$ and $f : \mathbb{R} \mapsto \mathbb{R}^+$ is a smooth function whose derivatives of all order $f', f'', \ldots$ are nonnegative. Then it was proved by Schoenber [6] that V is a Mercer Kernel.

As observed in [4] the assumption that V is a Mercer Kernel seems well suited to describe *self-repelling diffusions.* On the other hand, it is not clearly related to the geometry of M (see, e.g., the preceding example).

The next theorem has a more geometrical flavor and is robust to smooth perturbations (of M and V). It can be seen as a Bakry-Emery type condition [1] for self-interacting diffusions and is a first step toward **(c)**.

Theorem 3. *Assume that V is symmetric and that for all $x \in M, y \in M, u \in T_xM, v \in T_yM$*
$$\mathsf{Ric}_x(u,u) + \mathsf{Ric}_y(v,v) + \mathsf{Hess}_{x,y}V((u,v),(u,v)) \geq K(||u||^2 + ||v||^2)$$
where K is some positive constant. Then $\{\mu_t\}$ is globally convergent.

[1]We refer the reader to this paper for more details and mathematical statements.

Proof. Let $\mathcal{P}_{ac}(M)$ denote the set of probabilities which are absolutely continuous with respect to λ and let J be the nonlinear free energy function defined on $\mathcal{P}_{ac}(M)$ by

$$J(\mu) = \mathsf{Ent}(\mu) + \frac{1}{2}\int_{M\times M} V(x,y)\mu(dx)\mu(dy)$$

where

$$\mathsf{Ent}(\mu) = \int_M \log\left(\frac{d\mu}{d\lambda}\right) d\mu.$$

The key point is that $X_V^{-1}(0)$ is the critical set of J (restricted to $\mathcal{P}_{cd}(M)$) as shown in [4] (Proposition 2.9). On the other hand, the condition given in the theorem makes J a *displacement K-convex* function in the sense of McCann [5]. Let us briefly explain this latter statement.

Let d_2^W denote the L^2 Wasserstein distance on $\mathcal{P}(M)$ (see, e.g., [7] or [8]). Given $\nu^0, \nu^1 \in \mathcal{P}_{ac}(M)$ McCann [5] proved that there exists a unique geodesic path $t \to \nu^t$ in $(\mathcal{P}_{ac}(M), d_2^W)$ and that ν^t is the image of ν^0 by a map of the form $F_t(x) = \exp_x(t\Phi)$ where Φ is some vector field. Moreover,

$$d_2^W(\nu^0,\nu^t)^2 = \int_M d(x, F_t(x))^2 \nu^0(dx).$$

Set $j(t) = J(\nu^t) = e(t) + \frac{v(t)}{2}$ with $e(t) = \mathsf{Ent}(\nu^t)$ and

$$v(t) = \int_{M\times M} V(x,y)\nu^t(dx)\nu^t(dy) = \int_{M\times M} V(F_t(x), F_t(y))\nu^0(dx)\nu^0(dy).$$

Sturm [7] recently proved the beautiful result that

$$\partial^2 e(t) = \int_M \mathsf{Ric}(\dot{F}_t(x), \dot{F}_t(x))\nu^0(dx)$$

where $\partial^2 e(t) := \liminf_{s\to 0} \frac{1}{s^2}(e(t+s) - 2e(t) + e(t-s))$. Clearly

$$\partial^2 v(t) = \int_{M\times M} \mathsf{Hess}_{F_t(x),F_t(y)} V\left((\dot{F}_t(x), \dot{F}_t(y)), (\dot{F}_t(x), \dot{F}_t(y))\right)\nu^0(dx)\nu^0(dy).$$

Hence, under the assumption of Theorem 3,

$$\partial^2 j(t) \geq \frac{K}{2}\int_{M\times M} (||\dot{F}_t(x)||^2 + ||\dot{F}_t(y)||^2)\nu^0(dx)\nu^0(dy) = K d_2^W(\nu^0,\nu^1)^2.$$

In particular, j is strictly convex. It then follows that J (respectively X_V) has a unique minimum (respectively equilibrium). $\square$

Example. Let $M = S^n \subset \mathbb{R}^{n+1}$ be the unit sphere of dimension $n, f : \mathbb{R} \mapsto \mathbb{R}$ a smooth convex function and

$$V(x,y) = f(-||x-y||^2) = g(\langle x, y\rangle)$$

with $g(t) = f(2t-2)$. By invariance of λ under the orthogonal group $O(n+1)$ it is easily seen (see, e.g., Lemma 4.6 of [2]) that $V\lambda$ is a constant map. Hence $\lambda \in X_V^{-1}(0)$ and here, global convergence means convergence to λ.

For all $(x,y) \in M \times M$, $(u,v) \in T_xM \times T_yM$,

$$\begin{aligned}\mathsf{Hess}_{(x,y)} V((u,v),(u,v)) &= g''(\langle x,y\rangle)\left(\langle x,v\rangle + \langle y,u\rangle\right)^2 \\ &+ g'(\langle x,y\rangle)\left(2\langle u,v\rangle - (||u||^2 + ||v||^2)\langle x,y\rangle\right).\end{aligned}$$

Set $t = \langle x,y\rangle$ and assume (without loss of generality) that $||u||^2 + ||v||^2 = 1$. Then $|2\langle u,v\rangle| \leq 1$ and the last term on the right-hand side of the preceding equality is bounded below by $-tg'(t) - |g'(t)|$. Therefore the condition of Theorem 3 reads

$$tg'(t) + |g'(t)| < 2(n-1) \tag{1}$$

while Theorem 2 would lead to

$$g^{(k)}(t) \geq 0 \,\forall k \in \mathbb{N},\ |t| \leq 1. \tag{2}$$

Remark that condition (1) makes J a displacement-convex function while (2) makes J convex in the usual sense. Of course, none of these conditions is optimal. For instance, suppose that $g(t) = at$. Then (1) reads $|a| < n-1$, and (2) reads $a \geq 0$. On the other hand, this example can be fully analyzed and it was shown in [2] that $\mu_t \to \lambda$ for $a > -(n+1)$ while $\{\mu_t\}$ converges to a "Gaussian" measure with random center, for $a < -(n+1)$.

References

[1] D. Bakry, M. Emery, *Hypercontractivité des semi-groupes de diffusion*, C.R. Acad. Sci. Paris, I, **299** (1984), 775–778.

[2] M. Benaim, M. Ledoux and O. Raimond, *Self-interacting diffusions*, Probab. Theor. Relat. Fields, **122** (2002), 1–41.

[3] M. Benaim and O. Raimond, *Self-interacting diffusions II: Convergence in law*, Annales de l'Institut Henri-Poincaré, **6** (2003), 1043–1055.

[4] M. Benaim and O. Raimond, *Self-interacting diffusions III: Symmetric interactions*, Annals of Probability, **33 (5)** (2005), 1716–1759.

[5] R. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal., **11** (2001), 589–608.

[6] I. J. Schoenberg, *Metric spaces and completely monotone functions*, Ann. of Math., **39** (1938), 811–841.

[7] K. T. Sturm, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, J. Math. Pures. Appl., **84** (2005), 149–168.

[8] C. Villani, *Topics in Mass Transportation*, Graduate studies in Mathematics, AMS, 2003.

Michel Benaïm
Institut de Mathématiques, Université de Neuchâtel, Rue Emile-Argand 11
Case postale 2, CH-2007 Neuchâtel, Switzerland
e-mail: `michel.benaim@unine.ch`

Olivier Raimond
Laboratoire de Modélisation Stochastique et Statistique, Université Paris Sud, France
e-mail: `Olivier.Raimond@math.u-psud.fr`

Progress in Probability, Vol. 59, 23–36

Stationary Solutions for the 2D Stochastic Dissipative Euler Equation

Hakima Bessaih

Abstract. A 2-dimensional dissipative Euler equation, subject to a random perturbation is considered. Using compactness arguments, existence of martingale stationary solutions are proved.

Mathematics Subject Classification (2000). Primary 60H15; Secondary 60G10.

Keywords. Stationary processes, martingale solutions, stochastic Euler equations.

1. Introduction

We are concerned with the dissipative Euler equations for an incompressible fluid perturbed by a multiplicative noise, in an open bounded domain D of $\mathbb{R}^2$ with a smooth boundary ∂D which satisfies the locally Lipschitz condition (see [1]), i.e.,

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u = -\nabla p - \chi u + f + G(u)\zeta, \tag{1.1}$$

where u is the velocity of the fluid, p the pressure, f the external force, ζ is a Gaussian random field white noise in time, subject to the restrictions imposed below, and G is an operator acting on solution. The constant χ will be called the sticky viscosity. u is subject to the incompressibility condition

$$\nabla \cdot u(t,x) = 0, \quad t \in [0,T], \quad x \in D,$$

the boundary condition

$$u \cdot n = 0 \quad \text{on } \partial D,$$

n being the external vector. When $\chi = 0$, (1.1) is the classical Euler equation. For an additive noise, existence of strong solutions (in the probabilistic sense) has been proved in [3] for a bounded domain, in [15] in the whole space and in [8]

I would like to thank the referee for pointing out a mistake in the first version of the paper and for his helpful comments.

on the torus. For a multiplicative noise, existence of martingale solutions can be found in [4] and [7].

2. Notation, hypothesis and main result

Let $\mathcal{V}$ be the space of infinitely differentiable vector fields u on D with compact support strictly contained in D, satisfying $\nabla \cdot u = 0$. We introduce the space H of all measurable vector fields $u : D \longrightarrow \mathbb{R}^2$ which are square integrable, divergence-free, and tangent to the boundary

$$H = \Big\{ u \in \big[L^2(D)\big]^2 ;\ \nabla \cdot u = 0 \text{ in } D,\ u \cdot n = 0 \text{ on } \partial D \Big\} .$$

The space H is a separable Hilbert space with the inner product inherited from $\big[L^2(D)\big]^2$, denoted in the sequel by $< .,. >$ (norm $|.|$). Let V be the following subspace of H:

$$V = \Big\{ u \in \big[H^1(D)\big]^2 ;\ \nabla \cdot u = 0 \text{ in } D,\ u \cdot n = 0 \text{ on } \partial D \Big\} .$$

The space V is a separable Hilbert space with the inner product inherited from $\big[H^1(D)\big]^2$ (norm $\| \ . \ \|$). Identifying H with its dual space H', and H' with the corresponding natural subspace of the dual space V', we have the standard triple $V \subset H \subset V'$ with continuous dense injections. We denote the dual pairing between V and V' by the inner product of H.

Let $b(\cdot,\cdot,\cdot) : V \times V \times V \longrightarrow \mathbb{R}$ be the continuous trilinear form defined as

$$b(u, v, z) = \int_D (u \cdot \nabla v) \cdot z.$$

It is well known that there exists a continuous bilinear operator $B(\cdot,\cdot) : V \times V \longrightarrow V'$ such that

$$< B(u, v), z >= b(u, v, z), \text{ for all } z \in V.$$

By the incompressibility condition, we have

$$< B(u, v), v >= 0 \text{ and } \quad < B(u, v), z >= - < B(u, z), v > .$$

Let K be another separable Hilbert space. Denote by $L_2(K, H)$ the set of Hilbert-Schmidt operators from K to H.

Let $p > 1$ and m be a nonnegative integer; $W^{m,p}$ are the Sobolev spaces. When $p = 2$, then $W^{m,p}$ will be denoted by H^m. Let $0 < \alpha < 1$; then $W^{\alpha,p}(0, T; H)$ is the Sobolev space of all $u \in L^p(0, T; H)$ such that

$$\int_0^T \int_0^T \frac{|u(t) - u(s)|^p}{|t - s|^{1+\alpha p}} dtds < \infty.$$

We impose throughout the paper the following conditions:

1. $W(t)$ is a K-cylindrical Wiener process.
2. $f \in V$.

Let us assume that

$$(\mathbf{G1}) \qquad G : V \longrightarrow L_2(K, V), \ \text{ is globally Lipschitz continuous,}$$

$$(\mathbf{G2}) \qquad \begin{cases} |G(u)|^2_{L_2(K,H)} \le \lambda_0 |u|^2 + \rho_0, \\ |\nabla \wedge G(u)|^2_{L_2(K,H)} \le \lambda_1 |\nabla \wedge u|^2 + \lambda_2 |u|^2 + \rho_1, \ \forall u \in V \end{cases},$$

where $\nabla\wedge u = D_1 u_2 - D_2 u_1$ and $\lambda_0, \lambda_1, \lambda_2, \rho_0, \rho_1$ are positive constants independent of u.

Now let us give the following definition of a stationary martingale solution.

Definition 2.1. A martingale solution of Equation (1.1) consists of a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\ge 0}, \mathbf{P})$, a K-cylindrical Wiener process W and a progressively measurable process $u : [0, \infty) \times \Omega \to H$, with $\mathbf{P}$-a.e. paths

$$u(., \omega) \in C([0, T], D(A^{-\alpha/2})) \cap L^\infty(0, T; V)$$

for all $T > 0$, and $\alpha > 1$ such that $\mathbf{P}$-a.s. the identity

$$\begin{aligned} < u(t), v > + \int_0^t < B(u(s), u(s)), v > ds + \chi \int_0^t < u(s), v > ds \\ =< u(0), v > + \int_0^t < f(s), v > ds + < \int_0^t G(u(s)) dW(s), v > \end{aligned}$$

holds true for all $t \ge 0$ and all $v \in \mathcal{V}$. The space $D(A^{-\alpha/2})$ will be defined in the next section.

Moreover, a stationary martingale solution of Equation (1.1) is a martingale solution such that the process is stationary in H.

Remark 2.2. A function belonging to $C([0, T], D(A^{-\alpha/2})) \cap L^\infty(0, T; V)$ is weakly continuous in H. Hence, for every $t \ge 0$, the mapping $\omega \to u(t, \omega)$ is well defined from Ω to H and it is weakly measurable. Since H is a separable Banach space, it is strongly measurable (see [18, p. 131]). Therefore, it is meaningful to speak about the law of $u(t)$ in H. The stationarity of u in H introduced above has to be understood in this sense.

The existence of martingale solutions has been proved in [4] and in [7]. Here, we are interested in stationary martingale solutions.

Theorem 2.3. *In addition to the assumptions* (**G1**) *and* (**G2**), *assume that*

$$\chi > \frac{3}{2}\lambda_0 \ \text{ and } \ \chi > \frac{\lambda_1}{2}.$$

Then (1.1) *has a stationary martingale solution.*

3. The dissipative Navier-Stokes approximation

For every $\nu > 0$, we consider the equations of Navier-Stokes type

$$\begin{cases} \frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p = \nu\Delta u - \chi u + f + G(u)\frac{\partial W}{\partial t} & \text{in } (0,T)\times D, \\ \nabla \cdot u = 0 & \text{in } (0,T)\times D, \\ \nabla \wedge u = 0 & \text{on } (0,T)\times \partial D, \\ u \cdot n = 0 & \text{on } (0,T)\times \partial D, \\ u|_{t=0} = u_0 & \text{in } D. \end{cases} \tag{3.1}$$

Let $a(\cdot,\cdot) : V \times V \longrightarrow \mathbb{R}$ be the bilinear continuous form defined in [2] as

$$a(u,v) = \int_D \nabla u \cdot \nabla v - \int_{\partial D} k(\sigma)u(\sigma)\cdot v(\sigma)d\sigma,$$

where $k(\sigma)$ is a function defined on the boundary ∂D, and we have the estimates (see [13] for the details)

$$\int_{\partial D} k(\sigma)u(\sigma)\cdot v(\sigma)d\sigma \leq C \parallel u \parallel\parallel v \parallel,$$

and for an arbitrary $\epsilon > 0$,

$$\int_{\partial D} k(\sigma)|u(\sigma)|^2 d\sigma \leq \epsilon \parallel u \parallel^2 + C(\epsilon)|u|^2. \tag{3.2}$$

Moreover, we set

$$D(A) = \left\{u \in V \cap (H^2(D))^2,\ \ \nabla \wedge u = 0\right\},$$

and define the linear operator $A : D(A) \longrightarrow H$, as

$$Au = -\Delta u.$$

We will denote the domain of A^α by $D(A^\alpha)$. Here $D(A^{-\alpha/2})$ denotes the dual of $D(A^{\alpha/2})$, and we perform identification as above to have

$$D(A^{\alpha/2}) \subset V \subset H \subset V' \subset D(A^{-\alpha/2}).$$

In place of Equations (3.1) we will consider the abstract stochastic evolution equation

$$\begin{cases} du(t) + \nu Au(t)dt + B(u(t),u(t))dt = -\chi u(t)dt + f(t)dt + G(u(t))dW(t), \\ u(0) = u_0, \end{cases}$$

for $t \in [0,T]$. Assume that (**G1**) and (**G2**) hold and let $\alpha > 1$ be fixed. We have the continuous embedding (see [1, p. 85, Thm. 4.12 part II])

$$D(A^{\alpha/2}) \subset [H^\alpha(D)]^2 \subset [C(\bar{D})]^2.$$

Let P_n be the operator from $D(A^{-\alpha/2})$ to $D(A^{\alpha/2})$ defined as

$$P_n x = \sum_{i=1}^{n} < x, e_i > e_i, \quad x \in D(A^{-\alpha/2}).$$

Let $B_n(u,u)$ be the Lipschitz operator in P_nH defined as

$$B_n(u,u) = \pi_n B(u,u), \quad u \in P_nH,$$

where $\pi_n : H \to [0,1]$ is a C^∞ function defined as $\pi_n(u) = 1$ for $|u| \leq n$ and $\pi_n(u) = 0$ for $|u| \geq n+1$.

Consider the classical Faedo-Galerkin approximation scheme defined by the processes $u_{n\nu}(t) \in P_n H$, solutions of

$$\begin{cases} du_{n\nu}(t) + \nu A u_{n\nu}(t)dt + P_n B_n(u_{n\nu}(t), u_{n\nu}(t))dt \\ \qquad\qquad = -\chi u_{n\nu} P_n f(t)dt + P_n G(u_{n\nu}(t))dW(t), \\ u_{n\nu}(0) = P_n u_0, \end{cases} \tag{3.3}$$

$t \in [0,T]$.

Lemma 3.1. *There exist positive constants $C_1(p)$ and $\tilde{C}_1$ independent of n and of ν such that for each $p \geq 2$,*

$$\mathbf{E}(\sup_{0\leq s\leq t} |u_{n\nu}(s)|^p) \leq C_1(p), \tag{3.4}$$

and, moreover,

$$\nu \int_0^t \mathbf{E}\|u_{n\nu}(s)\|^2)ds \leq \tilde{C}_1. \tag{3.5}$$

Proof. By Itô's formula, for $p \geq 2$ we have

$$\begin{aligned} d|u_{n\nu}(t)|^p \quad &\leq \quad p|u_{n\nu}(t)|^{p-2} < u_{n\nu}, du_{n\nu} > \\ &\quad + \frac{1}{2}p(p-1)|u_{n\nu}(t)|^{p-2}|G(u_{n\nu})|^2_{L_2(K,V)}dt. \end{aligned}$$

Since $< B(u_{n\nu}, u_{n\nu}), u_{n\nu} >= 0$ and using the hypothesis (**G2**) we get

$$\begin{aligned} &d|u_{n\nu}(t)|^p + \nu p|u_{n\nu}(t)|^{p-2}|\nabla u_{n\nu}|^2 + \chi p|u_{n\nu}(t)|^p \\ &\leq \nu p|u_{n\nu}(t)|^{p-2}\int_{\partial D} k|u_{n\nu}|^2 dt + p|u_{n\nu}(t)|^{p-2} < f, u_{n\nu} > dt \\ &\quad + (1/2)p(p-1)|u_{n\nu}(t)|^{p-2}(\lambda_0|u_{n\nu}(t)|^2 + \rho_0)dt \\ &\quad + p|u_{n\nu}(t)|^{p-2} < G(u_{n\nu})dW, u_{n\nu} > . \end{aligned} \tag{3.6}$$

Using the Hölder inequality and then the Young inequality for the second term on the right-hand side of the above inequality, for a fixed $\epsilon_1 > 0$ we obtain

$$\begin{aligned} |u_{n\nu}(t)|^{p-2} < f, u_{n\nu} > \quad &\leq \quad |u_{n\nu}(t)|^{p-1}|f| \\ &\leq \quad \epsilon_1|u_{n\nu}(t)|^p + C(\epsilon_1, p)|f|^p. \end{aligned}$$

Using Young's inequality for the third term, for a fixed $\epsilon_2 > 0$ we get

$$\frac{1}{2}p(p-1)|u_{n\nu}(t)|^{p-2}\rho_0 \leq \epsilon_2|u_{n\nu}(t)|^p + C(\epsilon_2, p).$$

Thus, by using (3.2) and the previous estimates, we have

$$\begin{aligned} &d|u_{n\nu}(t)|^p + \nu p(1-\epsilon)|u_{n\nu}(t)|^{p-2}|\nabla u_{n\nu}|^2 dt + \chi p|u_{n\nu}|^p dt \\ &\leq C(\epsilon_1, p)|f|^p dt + C(\epsilon_2, p)dt + p|u_{n\nu}(t)|^{p-2} < G(u_{n\nu})dW, u_{n\nu} > \\ &\quad + \left(\frac{\lambda_0}{2}p(p-1) + \epsilon_2 + \epsilon_1 + \nu p C_\epsilon\right)|u_{n\nu}(t)|^p dt. \end{aligned}$$

Now we integrate over $(0,t)$, take the supremum on t and integrate over Ω, we obtain

$$
\begin{aligned}
&\mathbf{E}(\sup_{0\le s\le t}|u_{n\nu}(s)|^p)\\
&\le \mathbf{E}(|u_{n\nu}(0)|^p)+\left(\frac{\lambda_0}{2}p(p-1)+\epsilon_2+\epsilon_1+\nu pC_\epsilon-p\chi\right)\int_0^t \mathbf{E}\left(\sup_{0\le s\le r}|u_{n\nu}(s)|^p\right)dr\\
&\quad +C(\epsilon_2,p)t+C(\epsilon_1,p)\int_0^t \mathbf{E}|f|^p ds\\
&\quad +p\mathbf{E}\left(\sup_{0\le s\le t}\int_0^s |u_{n\nu}(r)|^{p-2}<G(u_{n\nu})dW(r),u_{n\nu}(r)>\right).
\end{aligned}
$$

Let us estimate the last term in the above inequality. By the Burkholder-Davis-Gundy inequality (see [9, p. 82, Thm. 3.14]) we get

$$
\begin{aligned}
&p\mathbf{E}\left(\sup_{0\le s\le t}\int_0^s |u_{n\nu}(r)|^{p-2}<G(u_{n\nu}(r))dw(r),u_{n\nu}(r)>\right)\\
&\le p\mathbf{E}\left(\int_0^t |u_{n\nu}(r)|^{2p-2}|G(u_{n\nu}(r))|^2_{L_2(K,V)}dr\right)^{1/2}.
\end{aligned}
$$

Using (**G2**) in the above inequality and the Cauchy-Schwartz inequality, we get

$$
\begin{aligned}
&p\mathbf{E}\left(\int_0^t |u_{n\nu}(r)|^{2p-2}|G(u_{n\nu}(r))|^2_{L_2(K,V)}dr\right)^{1/2}\\
&\le p\mathbf{E}\left(\int_0^t (\lambda_0|u_{n\nu}(r)|^{2p}+\rho_0|u_{n\nu}(r)|^{2p-2})dr\right)^{1/2}\\
&\le p\mathbf{E}\left(\sup_{0\le s\le t}|u_{n\nu}(s)|^{p/2}\left(\int_0^t (\lambda_0|u_{n\nu}(r)|^p+\rho_0|u_{n\nu}(r)|^{\frac{2p-2}{p}}dr\right)^{1/2}\right)\\
&\le \frac{1}{2}\mathbf{E}\left(\sup_{0\le s\le t}|u_{n\nu}(s)|^p\right)+\frac{p^2}{2}\mathbf{E}\int_0^t\lambda_0\sup_{0\le s\le\sigma}|u_{n\nu}(s)|^p+\frac{p^2}{2}\rho_0\mathbf{E}\int_0^t |u_{n\nu}(s)|^{\frac{2p-2}{p}}ds.
\end{aligned}
$$

Finally, we estimate the last term in the above inequality using Young's inequality. For $\epsilon_3>0$ we obtain

$$
\frac{p^2}{2}\rho_0\mathbf{E}\int_0^t |u_{n\nu}(s)|^{\frac{2p-2}{p}}ds\le \epsilon_3\int_0^t |u_{n\nu}(s)|^p ds+C(\epsilon_3,p).
$$

Collecting all the estimates, we obtain that

$$
\frac{1}{2}\mathbf{E}\left(\sup_{0\le s\le t}|u_{n\nu}(s)|^p\right)\le \mathbf{E}(|u_{n\nu}(0)|^p)+C_2\int_0^t \mathbf{E}\left(\sup_{0\le s\le r}|u_{n\nu}(s)|^p\right)dr+C_3, \quad (3.7)
$$

where

$$
C_2=\frac{\lambda_0}{2}(p(p-1)+p^2)+\epsilon_1+\epsilon_2+\epsilon_3+\nu pC_\epsilon-p\chi,
$$

and

$$C_3 = C(\epsilon_1, p) \int_0^t \mathbf{E}|f|^p + C(\epsilon_2, p) + C(\epsilon_3, p).$$

Using Gronwall's lemma we get (3.4).

Let us go back to (3.6), take $p = 2$ and integrate over $(0, t)$, we get

$$\begin{aligned}
&2\nu \int_0^t |\nabla u_{n\nu}|^2 + 2\chi \int_0^t |u_{n\nu}(t)|^2 \\
&\le |u_{n\nu}(0)|^2 + 2\nu \int_0^t \int_{\partial D} k|u_{n\nu}|^2 + 2 < f, u_{n\nu} > dt \\
&\quad + \int_0^t (\lambda_0 |u_{n\nu}(t)|^2 + \rho_0) + 2 \int_0^t < G(u_{n\nu}) dW, u_{n\nu} > .
\end{aligned}$$

In the above inequality integrate over Ω, then

$$\mathbf{E} \int_0^t < G(u_{n\nu}) dW, u_{n\nu} >= 0.$$

Now use (3.2) to estimate the second term on the left-hand side and the Cauchy-Schwartz inequality to estimate the third term on the left-hand side. Finally, using the estimate (3.4) we get (3.5). ☐

Lemma 3.2. *There exists a positive constant C_4 which does not depend on n and on ν such that*

$$\mathbf{E} \int_0^t \| u_{n\nu}(s) \|^2 \le C_4. \tag{3.8}$$

Proof. Let $\xi_{n\nu} = \nabla \wedge u_{n\nu}$. We apply the curl operator to Equation (3.3) and get for $t \in [0, T]$,

$$d\xi_{n\nu} + \nu A \xi_{n\nu} dt + \nabla \wedge P_n B_n(u_{n\nu}, u_{n\nu}) dt = -\chi \xi_{n\nu} dt + \nabla \wedge P_n f dt + \nabla \wedge (G(u_{n\nu})) dW.$$

By Itô's formula we have

$$\begin{aligned}
d|\xi_{n\nu}|^2 &= 2 < \xi_{n\nu}, d\xi_{n\nu} > + |\nabla \wedge (G(u_{n\nu}))|^2_{L_2(K,V)} \\
&= -2\nu < A\xi_{n\nu}, \xi_{n\nu} > dt - 2 < \nabla \wedge P_n B_n(u_{n\nu}, u_{n\nu}), \xi_{n\nu} > dt \\
&\quad - 2\chi |\xi_{n\nu}|^2 + 2 < \nabla \wedge P_n f, \xi_{n\nu} > dt \\
&\quad + < \nabla \wedge (G(u_{n\nu})) dw, \xi_{n\nu} > + |\nabla \wedge (G(u_{n\nu}))|^2_{L_2(K,V)}.
\end{aligned}$$

Since $\xi_{n\nu}|_{\partial D} = 0$, $< \nabla \wedge P_n B(u_{n\nu}, u_{n\nu}), \xi_{n\nu} >= 0$, and using (**G2**), we get that

$$\begin{aligned}
d|\xi_{n\nu}|^2 + 2\nu |\nabla \xi_{n\nu}|^2 dt &\le -2\chi |\xi_{n\nu}|^2 dt + 2 < \nabla \wedge P_n f, \xi_{n\nu} > dt \\
&\quad + < \nabla \wedge (G(u_{n\nu})) dW, \xi_{n\nu} > + \lambda_1 |\xi_{n\nu}|^2 + \lambda_2 |u_{n\nu}|^2 + \rho_1.
\end{aligned}$$

Now using Young's inequality for the second term on the right-hand of the above inequality and for a fixed $\epsilon_4 > 0$ we obtain

$$\begin{aligned}
d|\xi_{n\nu}|^2 + 2\nu |\nabla \xi_{n\nu}|^2 dt &\le (-2\chi + \lambda_1 + \epsilon_4) |\xi_{n\nu}|^2 dt + C(\epsilon_4, p) |\nabla \wedge P_n f| \\
&\quad + < \nabla \wedge (G(u_{n\nu})) dW, \xi_{n\nu} > + \lambda_2 |u_{n\nu}|^2 + \rho_1.
\end{aligned}$$

We integrate over $(0,t)$ and then over Ω. Since

$$\mathbf{E}\int_0^t < \nabla\wedge(G(u_{n\nu}))dW, \xi_{n\nu} >= 0,$$

we obtain the estimate

$$\mathbf{E}|\xi_{n\nu}(t)|^2 \le \mathbf{E}|\xi_{n\nu}(0)|^2 + (-2\chi+\lambda_1+\epsilon_4)\mathbf{E}\int_0^t |\xi_{n\nu}(s)|^2 ds + C(\epsilon_4)\int_0^t |\nabla\wedge P_n f| + \lambda_2\int_0^t \mathbf{E}|u_{n\nu}|^2 + \rho_1 t, \tag{3.9}$$

Using Gronwall's lemma, we obtain that there exists a positive constant C_5 independent of n and of ν such that

$$\mathbf{E}|\xi_{n\nu}(s)|^2 \le C_5. \tag{3.10}$$

Now let us introduce the elliptic problem

$$\begin{cases} -\Delta u_{n\nu} = \nabla^\perp \xi_{n\nu} & \text{in } D,\\ u_{n\nu}\cdot n = 0 & \text{on } \partial D,\\ \xi_{n\nu} = 0 & \text{on } \partial D, \end{cases} \tag{3.11}$$

where $\nabla^\perp = (D_2, -D_1)$.

We multiply the first equation of (3.11) by $u_{n\nu}$ and integrate over D, we have

$$-< \Delta u_{n\nu}.u_{n\nu} >=< \nabla^\perp\xi_{n\nu}, u_{n\nu} > .$$

Through integration by parts and in virtue of (3.2), we obtain

$$|\nabla u_{n\nu}(t)|^2 \le \epsilon|\nabla u_{n\nu}(t)|^2 + C_\epsilon|u_{n\nu}(t)|^2 + |\xi_{n\nu}(t)|^2$$

for all $t\in(0,T)$ and for an arbitrary $\epsilon > 0$. We integrate the above inequality, respectively, over $(0,t)$ and over Ω, we obtain

$$\begin{aligned} \mathbf{E}\int_0^t |\nabla u_{n\nu}|^2 &\le C\mathbf{E}\left(\int_0^t |u_{n\nu}|^2\right) + \mathbf{E}\left(\int_0^t |\xi_{n\nu}|^2\right)\\ &\le Ct\mathbf{E}\left(\sup_{0\le s\le t}|u_{n\nu}(s)|^2\right) + \mathbf{E}\left(\int_0^t |\xi_{n\nu}|^2\right), \end{aligned}$$

C being a constant independent of n and ν. According to (3.4) and (3.10), this yields the estimate (3.8). □

4. Construction of stationary solutions

Step 1. Take $p=2$ in (3.7) we get that

$$\mathbf{E}|u_{n\nu}(t)|^2 \le \mathbf{E}|u_{n\nu}(0)|^2 + (\lambda_0+\epsilon_2+\epsilon_1+2\nu C_\epsilon - 2\chi)\int_0^t \mathbf{E}|u_{n\nu}(s)|^2 ds + C(\epsilon_2)t + C(\epsilon_1)\int_0^t \mathbf{E}|f|^2 ds.$$

If $\chi > \frac{3}{2}\lambda_0$ and $\chi > \frac{\lambda_1}{2}$ then we can choose ϵ_1, ϵ_2, ϵ_4 and ν_0 in the above inequality and in (3.9) such that using Gronwall lemma we get that

$$\mathbf{E}\|u_{n\nu}(t)\|^2 \leq C \quad \forall t \geq 0 \quad \forall n \geq 1 \tag{4.1}$$

for some constant $C > 0$. This implies that there exists an invariant measure for (3.3) by the classical Krylov-Bogoliubov argument (see [10]). Call $\mu_{n\nu}$ one of such invariant measures. From (4.1) we have

$$\int_{P_n V} |x|^2 \mu_{n\nu}(dx) \leq C \quad \forall n \geq 1. \tag{4.2}$$

There exists a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}\}_t, \mathbf{P})$, possibly larger than the one given at the beginning, that supports a random variable $u_{n\nu}(0)$ which is $\mathcal{F}_0$ measurable, with law $\mu_{n\nu}$, and a cylindrical Wiener process $W(t)$ with values in K. The solution $\tilde{u}_{n\nu}$ with initial condition $u_{n\nu}(0)$ is a stationary process.

Step 2. Now let us prove that the family $\{\mathcal{L}(\tilde{u}_{n\nu})\}_{n\nu}$ is tight in $L^2(0,T;H) \cap C([0,T]; D(A^{-\alpha/2}))$, for all given $\alpha > 1$; in fact we decompose $\tilde{u}_{n\nu}$ as

$$\begin{aligned}
\tilde{u}_{n\nu}(t) &= \tilde{u}_{n\nu}(0) - \nu \int_0^t A\tilde{u}_{n\nu}(s) - \int_0^t P_n B_n(\tilde{u}_{n\nu}(s), \tilde{u}_{n\nu}(s)) \\
&\quad + \int_0^t P_n f(s) + \int_0^t G(\tilde{u}_{n\nu}(s))dW(s) \\
&= J_1 + \cdots + J_5.
\end{aligned}$$

We have from the bound (4.2) on $\mu_{n\nu}$ that

$$\mathbf{E}|J_1|^2 \leq C_6.$$

From (3.7),

$$\mathbf{E} \parallel J_2 \parallel^2_{W^{1,2}(0,T;V')} \leq C_7.$$

Moreover, we have

$$\mathbf{E} \parallel J_4 \parallel^2_{W^{1,2}(0,T;V')} \leq C_8$$

for suitable positive constants C_6, C_7, C_8. Using Lemma 5.1, the uniform assumption (**G1**), and the estimate (3.4) we have

$$\begin{aligned}
\mathbf{E} \parallel J_5 \parallel^2_{W^{\gamma,2}(0,T;H)} &\leq \mathbf{E} \int_0^T \parallel G(\tilde{u}_{n\nu}) \parallel^2_{L_2(K,H)} \\
&\leq \mathbf{E} \int_0^T (\lambda_0 |\tilde{u}_{n\nu}(s)|^2 + \rho_0) ds \\
&\leq C_9(\lambda_0, \rho_0, \gamma)
\end{aligned}$$

for $\gamma \in (0, 1/2)$, C_9 being independent of n and ν.

Since $\alpha > 1$, $D(A^{\alpha/2}) \subset (L^\infty(D))^2$ so that

$$| < B(u,u), v > | \leq C|u| \parallel u \parallel |A^{\alpha/2} v|, \; u \in V, \; v \in D(A^{\alpha/2})$$

for some constant $C > 0$. Hence, we have

$$\| J_3 \|^2_{W^{1,2}(0,T;D(A^{-\alpha/2}))} \leq C_{10} \sup_{0\leq t\leq T} |\tilde{u}_{n\nu}(t)|^2 \int_0^T \| \tilde{u}_{n\nu}(s) \|^2 \, ds$$

for some positive constant C_{10} independent of n and ν. In virtue of (3.4) and (3.8), we obtain that

$$\mathbf{E} \| J_3 \|^2_{W^{1,2}(0,T;D(A^{-\alpha/2}))} \leq C_{11}.$$

Clearly for $\gamma \in (0, 1/2)$, $W^{1,2}(0,T; D(A^{-\alpha/2})) \subset W^{\gamma,2}(0,T; D(A^{-\alpha/2}))$; collecting all the previous inequalities we have

$$\mathbf{E} \| \tilde{u}_{n\nu} \|_{W^{\gamma,2}(0,T;D(A^{-\alpha/2}))} \leq C_{12}, \tag{4.3}$$

for $\gamma \in (0, 1/2)$ and $\alpha > 1$, C_{12} being a positive constant independent of n and ν. By (3.8) and (4.3), we have that the laws $\mathcal{L}(\tilde{u}_{n\nu})$ are bounded in probability in

$$L^2(0,T;V) \cap W^{\gamma,2}(0,T;D(A^{-\alpha/2})).$$

Thus by Theorem 5.2, $\{\mathcal{L}(\tilde{u}_{n\nu})\}$ is tight in $L^2(0,T;H)$. On the other hand, by Theorem 5.3 $\{\mathcal{L}(\tilde{u}_{n\nu})\}$ is tight in $C([0,T]; D(A^{-\beta/2})$, for $\alpha < \beta$.

Step 3. Let us endow $L^2_{loc}(0,\infty;H)$ by the distance

$$d_2(u,v) = \sum_{k=1}^{\infty} 2^{-k} \min\left(|u-v|_{L^2(0,k;H)}, 1 \right),$$

and, similarly, $C(0,\infty; D(A^{-\beta/2}))$ by the distance

$$d_\infty(u,v) = \sum_{k=1}^{\infty} 2^{-k} \min\left(|u-v|_{C[0,k];D(A^{-\beta/2}))}, 1 \right).$$

Hence, we obtain that $\{\mathcal{L}(\tilde{u}_{n\nu})\}_{n\nu}$ is tight in $L^2_{loc}(0,\infty;H) \cap C([0,\infty]; D(A^{-\beta/2}))$, thus $\tilde{u}_{n\nu}$ is a stationary solution in H. Let us choose $\nu = 1/n$. From Prokhorov's theorem (see [9, p. 32]), the set of the laws $\{\mathcal{L}(\tilde{u}_{n\nu})\}$ is relatively compact. By Skorohod's theorem, there exists a basis $(\Omega^1, \mathcal{F}^1, \{\mathcal{F}^1_t\}_{t\geq 0}, \mathbf{P}^1)$ and on this basis, $L^2_{loc}(0,\infty;H) \cap C([0,\infty]; D(A^{-\beta/2}))$-valued random variables u^1, $u^1_{n\nu}$, such that $\mathcal{L}(\tilde{u}_{n\nu}) = \mathcal{L}(u^1_{n\nu})$, on $L^2_{loc}(0,\infty;H) \cap C([0,\infty]; D(A^{-\beta/2}))$, and $u^1_{n\nu} \longrightarrow u^1$ $\mathbf{P^1}$-a.s. in $L^2_{loc}(0,\infty;H) \cap C([0,\infty]; D(A^{-\beta/2}))$. Since $u^1_{n\nu}$ and $\tilde{u}_{n\nu}$ have the same law, $u^1_{n\nu}$ is also a stationary solution. By the a.s. convergence, u^1 is a stationary solution in H.

By (3.4) and (3.8) we have

$$\mathbf{E}\left(\sup_{0\leq s\leq t} |u^1_{n\nu}(s)|^p \right) \leq C_1(p),$$

$$\mathbf{E}\left(\int_0^t \| u^1_{n\nu}(s) \|^2 \right) \leq C_2,$$

for all $n \geq 1$ and $p \geq 2$. Hence, we have that

$$u^1(.,\omega) \in L^2_{loc}(0,\infty;V) \cap L^\infty_{loc}(0,\infty;H) \quad \mathbf{P}\text{-a.s.}$$

and $u^1_{n\nu} \longrightarrow u^1$ weakly in $L^2(\Omega\times(0,\infty);V)$. Let us define the process $M_{n\nu}(t)$ with trajectories in $C([0,\infty];H)$ as

$$M_{n\nu}(t) = u^1_{n\nu}(t) - P_n u^1 + \nu\int_0^t Au^1_{n\nu}(s)ds + \int_0^t P_nB_n(u^1_{n\nu}(s),u^1_{n\nu}(s))ds - \int_0^t P_nf(s)ds.$$

We will prove that $M_{n\nu}(t)$ is a square integrable martingale with respect to the filtration

$$\sigma\left\{u^1_{n\nu}(s), s\le t\right\},$$

with quadratic variation

$$<< M_{n\nu} >>_t = \int_0^t G(u^1_{n\nu})G(u^1_{n\nu})^* ds. \tag{4.4}$$

We shall prove the following lemma.

Lemma 4.1. *Assume that* (3.4) *and* (3.8) *hold. Then*

$$\left\langle\int_0^t P_nB_n(u^1_{n\nu}(s),u^1_{n\nu}(s))ds, v\right\rangle \longrightarrow \left\langle\int_0^t B(u^1(s),u^1(s))ds, v\right\rangle$$

for all $t\in[0,\infty)$ *and* $v\in\mathcal{V}$ $\mathbf{P}$*-a.s.*

Proof.

$$\begin{aligned}\left\langle\int_0^t P_nB_n(u^1_{n\nu}(s),u^1_{n\nu}(s))ds, v\right\rangle &= \left\langle\int_0^t \pi_n(u^1_{n\nu}(s))(u^1_{n\nu}(s))_iD_i(u^1_{n\nu}(s))_j ds, v_j\right\rangle\\ &= -\int_0^t\int_D \pi_n(u^1_{n\nu}(s))(u^1_{n\nu})_i(s)(u^1_{n\nu})_j\frac{\partial(v_n)_j(s)}{\partial x_i}.\end{aligned}$$

That converges $\mathbf{P}$-a.s. to

$$\int_0^t\int_D (u^1)_i(s)(u^1)_j\frac{\partial(v)_j(s)}{\partial x_i} = \left\langle\int_0^t B(u^1(s),u^1(s))ds, v\right\rangle. \qquad\square$$

Since $u_{n\nu}$ and $u^1_{n\nu}$ have the same law, for a real-valued, bounded and continuous function φ on $C([0,s];D(A^{-\beta/2}))$ where $0\le s\le t\le T$, and for all $v,z\in\mathcal{V}$, we have

$$\mathbf{E}(< M_{n\nu}(t) - M_{n\nu}(s), v > \varphi(u_{n\nu})) = 0 \tag{4.5}$$

and

$$\mathbf{E}((< M_{n\nu}(t), v >< M_{n\nu}(t), z > - < M_{n\nu}(s), v >< M_{n\nu}(s), z > - \int_s^t G(u^1_{n\nu}(r))G(u^1_{n\nu}))^*\varphi(u^1_{n\nu})) = 0. \tag{4.6}$$

By (3.4), (3.8) we can take the limit in (4.5) and (4.6) and we obtain

$$\mathbf{E}(< M^1(t) - M^1(s), v > \varphi(u_{n\nu})) = 0 \tag{4.7}$$

and

$$\mathbf{E}((< M^1(t), v >< M^1(t), z > - < M^1(s), v >< M^1(s), z > \\ - \int_s^t G(u^1(r))G(u^1(r)^* \varphi(u^1)) = 0, \quad (4.8)$$

where $M^1(t)$ is defined as

$$M^1(t) = u^1 - u^1(0) + \chi \int_0^t u^1(s)ds + \int_0^t B(u^1(s), u^1(s))ds - \int_0^t f(s)ds$$

P-a.s. in $C([0,T]; D(A^{-\beta/2}))$.

From (4.7) and (4.8), with $v, z \in D(A^{-\beta/2})$, we have that $A^{-\beta/2}M^1(t)$ is a square integrable martingale in H with respect to the filtration

$$\sigma\left\{u^1(s), s \leq t\right\},$$

with quadratic variation

$$<< A^{-\beta/2}M^1 >>_t = \int_0^t A^{-\beta/2}G(u^1)G(u^1)^*A^{-\beta/2}ds.$$

We conclude by a representation theorem (see [9, p. 233]).

5. Appendix

For any progressively measurable process $f \in L^p(\Omega \times [0,T]; L_2(K, H))$ denote by $I(f)$ the Ito integral defined as

$$I(f)(t) = \int_0^t f(s)dW(s), \ t \in [0,T].$$

$I(f)$ is a progressively measurable process in $L^p(\Omega \times [0,T]; H)$.

Lemma 5.1. *Let $p \geq 2$ and $\gamma < 1/2$ be given. Then for any progressively measurable process $f \in L^p(\Omega \times [0,T]; L_2(K, H))$, we have*

$$I(f) \in L^p(\Omega; W^{\gamma,p}(0, T; H)),$$

and there exists a constant $C(p, \gamma) > 0$ independent of f such that

$$\mathbf{E} \parallel I(f) \parallel^p_{W^{\gamma,p}(0,T;H)} \leq C(p, \gamma)\mathbf{E} \int_0^T \parallel f \parallel^p_{L_2(K;H)} dt.$$

Proof. See [11]. □

Theorem 5.2. *Let $B_0 \subset B \subset B_1$ be Banach spaces, B_0 and B_1 reflexive with compact embedding of B_0 in B_1. Let $p \in (1, \infty)$ and $\gamma \in (0, 1)$ be given. Let X be the space*

$$X = L^p(0, T; B_0) \cap W^{\gamma,p}(0, T; B_1)$$

endowed with the natural norm. Then the embedding of X in $L^p(0, T; B)$ is compact.

Theorem 5.3. *Let B_1 and $\tilde{B}$ two Banach spaces such that $B_1 \subset \tilde{B}$ with compact embedding. If the real numbers $\gamma \in (0,1)$ and $p > 1$ satisfy*

$$\gamma p > 1,$$

then the space $W^{\gamma,p}(0,T;B_1)$ is compactly embedded into $C([0,T];\tilde{B})$.

Proof. See [11]. □

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, Pure and Applied Mathematics Series, 2nd edition, Academic Press, 2003.

[2] C. Bardos, *Éxistence et unicité de la solution de l'équation d'Euler en dimensions deux*, Jour. Math. Anal. Appl., **40** (1972), 769–780.

[3] H. Bessaih and F. Flandoli, *2-D Euler equations with non regular force*, Non Diff. Eq. Appl, **6 (1)** (1999), 35–54.

[4] H. Bessaih, *Martingale solutions for stochastic Euler equations*, Stoc. Anal. Appl., **17 (5)** (1999), 713–727.

[5] Z. Brzeźniak, M. Capiński, and F. Flandoli, *Stochastic partial differential equations and turbulence*, Math. Models and Methods in Appl. Sc, **1** (1991), 41–59.

[6] Z. Brzeźniak, M. Capiński, and F. Flandoli, *Stochastic Navier-Stokes equations with multiplicative noise*, Stoc. Anal. Appl, **10 (5)** (1992), 523–532.

[7] Z. Brzeźniak and S. Peszat, *Stochastic Euler equations*, Ann. Probab., **29 (4)** (2001), 1796–1832.

[8] M. Capiński and N. J. Cutland, *Stochastic Euler equations on the torus*, Ann. Appl. Probab., **9 (3)** (1999), 688–705.

[9] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, 1992.

[10] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite-Dimensional Systems*, Cambridge University Press, Cambridge, 1996.

[11] F. Flandoli and D. Gatarek, *Martingale and stationary solutions for stochastic Navier-Stokes equations*, Probab. Theory Relat. Fields, **102** (1995), 367–391.

[12] G. Gallavotti, *Ipotesi per una introduzione alla Meccanica Dei Fluidi*, Quaderni del Consiglio Nazionale delle Ricerche, Gruppo Nazionale di Fisica Matematica, **52** (1996).

[13] J. L. Lions, *Équations Différentielles Opérationelles et Problèmes aux Limites*, Springer-Verlag, Berlin, 1961.

[14] P. L. Lions, *Mathematical Topics in Fluid Mechanics*, vol. 1, Incompressible Models, Oxford Sci. Publ, Oxford, 1996.

[15] R. Mikulevicius and G. Valiukevicius, *On stochastic Euler equation in $\mathbb{R}^d$*, Electron. J. Probab., **5 (6)** (2000).

[16] W. A. Strauss, *On continuity of functions with values in various Banach spaces*, Pacific Journal of Mathematics, **19 (3)** (1966), 543–551.

[17] R. Temam, *Navier-Stokes Equations*, North-Holland, 1984.

[18] K. Yosida, *Functional Analysis*, Classics in Mathematics, Springer-Verlag, Berlin, 1995.

Hakima Bessaih
University of Wyoming
Department of Mathematics
Laramie 82070 WY
USA
e-mail: bessaih@uwyo.edu

Progress in Probability, Vol. 59, 37–55

Volterra Equations Perturbed by a Gaussian Noise

Stefano Bonaccorsi

Abstract. We consider, in a Hilbert space U, a class of Gaussian processes defined by a linear filter with a cylindrical Wiener process as input process. This noise is used as an additive perturbation to a family of fractional order (in time) partial differential equations. We give conditions such that the stochastic convolution process is well defined, both in finite time horizon and in an infinite interval. An important example of noise that is contained in the paper is the fractional Brownian motion.

Mathematics Subject Classification (2000). Primary 60H20; Secondary 60G15.

Keywords. Stochastic integral equations, Gaussian processes.

1. Introduction and motivating example

The purpose of this paper is to study properties of the stochastic convolution process which arises as solution of an infinite-dimensional integral Volterra equation perturbed by a general Gaussian noise,

$$B(t) = \int_0^t K(t,s)\,\mathrm{d}W(s), \qquad t \geq 0. \tag{1.1}$$

We may interpret the integral relationship (1.1) as a filter that takes as input the Wiener process $W(t)$ and outputs the process $B(t)$. Similar types of linear filters are widely used in modeling stochastic systems (see, e.g., Wong and Hajek [26]) and choosing K from suitable families provides interesting examples of distribution processes which generalize ARMA distribution processes. Continuous time, Gaussian fractionally integrated models are extensively used in financial applications: for instance, a general class has been introduced by Comte and Renault [10] for modelling stochastic volatility.

Our interest concerns a noise that is infinite-dimensional. We fix a real separable Hilbert space U, endowed with a complete orthonormal system $\{e_k\}$, and we consider a cylindrical Wiener process W, defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, of the form

$$\langle W(t), h\rangle = \sum_{k=1}^{\infty} \langle h, e_k\rangle \beta^{(k)}(t)$$

for $h \in U$, where $\{\beta^{(k)}(\cdot),\ k \geq 1\}$ is a sequence of real-valued, independent Brownian motions.

In this setting, we look for a mild solution of the equation

$$u(t) = \int_0^t g_\rho(t-s)Au(s)\,\mathrm{d}s + RB(t), \qquad t \in [0,T]. \tag{1.2}$$

Here, A is a self-adjoint, negative defined operator on U; there exists a basis $\{e_k,\ k \in \mathbb{N}\}$ of U such that

$$Ae_k = -\mu_k e_k$$

for an increasing sequence $\{\mu_k,\ k \in \mathbb{N}\}$ of positive real numbers. Moreover, given a sequence $\{\lambda_k,\ k \in \mathbb{N}\}$ of non-negative real numbers, R is a bounded linear operator defined by

$$Re_k = \sqrt{\lambda_k} e_k, \qquad k \geq 1.$$

Remark 1.1. Volterra integro-differential equations with respect to the fractional kernel $g_\rho(t) = \frac{1}{\Gamma(\rho)} t^{\rho-1}$ are widely considered in the literature, as they make a good link between the heat equation ($\rho = 1$) and the wave equation ($\rho = 2$). Such kind of equations naturally appear in different applications in mathematical physics; for some recent developments in the case of a deterministic problem see [14, 17] or the monograph [3]. In the stochastic case, the evolutionary integral equation (1.2) was introduced in [8] in connection to heat equation in materials with memory and in [9] for equations of linear parabolic viscoelasticity. Notice also the recent paper [19] where a different class of noises is considered. In all these papers, an arbitrary completely monotone kernel $a(t)$ is considered; here, on the contrary, we focus on the kernels $g_\rho(t)$, for $\rho \in (0,1)$, as they are related to explicit solutions via Mittag-Leffler's function, see Section 2.2. Our choice also implies that we obtain more precise estimates on the stochastic convolution, as we shall not appeal to the general estimates for the resolvent operator as given, for instance, in [24].

Equation (1.2) is seen as an infinite-dimensional filter of fractional order in time, with input $B(t)$ and output $u(t)$. The above representation clearly shows the causality of the system, i.e., the state of the system is determined by its history and the present perturbation, but does not depend on the future.

It shall be clear that we may consider each component of the (infinite-dimensional) vector $u(t)$ separately; in the language of queueing theory, we may say that the *netput rate* for class-k work is modeled following the law

$$u_k(t) = \sqrt{\lambda_k} B_k(t) - \mu_k (g_\rho * u_k)(t), \qquad t \geq 0,$$

the quantity $\mu_k > 0$ to the output rate of class-k work produced at the station and the quantity λ_k represents the intensity of the input rate. Equation (1.2) can be considered as an input-output system where the properties of the working station (the output rates μ_k and the kernel g_ρ) are fixed, while the large k behavior of the λ_k's defines the spatial behavior of the noise. Our interest is to give conditions on the kernel defining the filter (1.1) and the λ_k's, in order to obtain the existence of a solution in a mild sense (compare Definition 2.5 below) over a finite time interval. The main result of the paper is Theorem 3.1 which concernes with the estimate of the L^2-norm of the stochastic convolution process; several applications are given in Section 3 where, in particular, we shall consider the following classes of kernels.

Remark 1.2.

1. The kernel $K(t,s)$ is of fractional integration type: $K(t,s) = g_\vartheta(t-s)$ for some $\vartheta > \frac{1}{2}$; this case naturally arises in evolution equations of fractional order in time, compare for instance [5]; the main advantage of working with this kernel is the simplicity in the computations; however, this filter does not preserve the stationary increment property of the input and it changes the self-similarity index.
2. The kernel $K_H(t,s)$ defines a fractional Brownian motion $B_H(t)$. This case is of particular interest in the applications because this filter does preserve the stationary increments of the input process, the output process is self-similar (with parameter H) and exhibits long-range dependence whenever $H > \frac{1}{2}$. However, the simplicity of this filter is considered in some cases as "a drawback because a single parameter H determines all the characteristics" [4].
3. There are kernels which cannot be reduced to a (fractional) Brownian behavior; in this sense, we show what happens when the kernel is almost regular (e.g., $K(t,s) = o((t-s)^{\vartheta-1})$ as $t \searrow s$, for $\vartheta \in (\frac{1}{2}, 1)$ no matter how close to 1) or
4. the kernel is less regular than any of the fractional integration kernels (e.g., $K(t,s) \gg (t-s)^{\vartheta-1}$ as $t \searrow s$, for any ϑ no matter how close to $\frac{1}{2}$).

2. Preliminaries

Let us begin our discussion with a description of the class of admissible kernels. Let $K : (0,\infty) \times (0,\infty) \to \mathbb{R}$ be a measurable function such that

$$\begin{cases} \displaystyle\int_0^t K^2(t,s)\,\mathrm{d}s < \infty & \text{for every } t > 0 \\ K(t,s) = 0 & \text{if } s > t. \end{cases} \tag{2.1}$$

We introduce, following [6], the following classes of *singular* and *smooth* kernels.

Definition 2.1.

(i) We say that K is a *singular* (rough) *kernel* if it satisfies condition (2.1) and there exists a measurable function $(t,s) \to \partial_1 K(t,s)$ such that $u \to \partial_1 K(u,s)$

is integrable on every $[t,\tau] \subset (s,\infty)$ and satisfies

$$K(\tau,s) - K(t,s) = \int_t^\tau \partial_1 K(u,s)\,\mathrm{d}u \qquad \text{for } s<t.$$

Notice that $\int_s^t \partial_1 K(u,s)\,\mathrm{d}u$ may be infinite. The main example is the fractional integration kernel $K(t,s) = g_\vartheta(t-s)$ for $\vartheta \in (\frac{1}{2},1)$.

(ii) We say that K is a *smooth kernel* if K is a singular kernel such that

$$K(\tau,s) - K(s,s) = \int_s^\tau \partial_1 K(u,s)\,\mathrm{d}u.$$

This is the case of the fractional Brownian motion of Hurst parameter $H > \frac{1}{2}$, see Section 4.1 below.

In the space of singular kernels we introduce the family of spaces $E_{\gamma,q,p,t}$, for a given set of parameters $\gamma \ge 0$, $p \ge 1$, $q \ge 1$, endowed with the following norm:

$$\|K\|^2_{\gamma,q,p,t} := \left(\int_0^t |(t-s)^\gamma K(t,s)|^{2q}\,\mathrm{d}s\right)^{1/q} + \left(\int_0^t \left(\int_s^t |(r-s)^{\gamma+1}\partial_1 K(r,s)|^p\,\mathrm{d}r\right)\,\mathrm{d}s\right)^{1/p} < \infty. \tag{2.2}$$

We allow q and p to assume the value $+\infty$; in these cases, we suitably modify (2.2):

if $q=+\infty$, the first integral becomes $\sup\limits_{0<s<t} |(t-s)^\gamma K(t,s)|^2$;
if $p=+\infty$, the second integral becomes $\sup\limits_{0<s<r<t} |(r-s)^{\gamma+1}\partial_1 K(r,s)|^2$.

In the sequel we define a set (γ,p,q) of *admissible parameters* if it verifies

$$0<\gamma<\frac{1}{2}, \quad p>\frac{1}{1-\gamma}, \quad q>\frac{1}{1-2\gamma}.$$

We shall remark that, although formally similar, the situation here is quite different from [6], in particular regarding the set of admissible parameters.

Definition 2.2 (Cylindrical Gaussian process). We shall call $B = \{B_t,\ t\in[0,T]\}$ a cylindrical Gaussian process if it has the form

$$\langle B(t),h\rangle = \sum_{k=1}^\infty \langle h,e_k\rangle \int_0^t K(t,s)\,\mathrm{d}\beta^{(k)}(s) := \sum_{k=1}^\infty \sqrt{\lambda_k}\langle h,e_k\rangle B^{(k)}(t). \tag{2.3}$$

2.1. Stochastic calculus with respect to a Gaussian process

Recently, there was an interest to define a stochastic calculus of variations (Malliavin calculus) with respect to a Gaussian process; in this section, we shall introduce the basic ideas provided by Carmona et al. [6], Alòs and Nualart [2]. Let us remark that, for our purposes, it is not necessary to construct a general theory, since all of our integrand functions are deterministic.

We start with a real standard Brownian motion $\beta = \{\beta_t,\ t \in [0,T]\}$ and define a zero mean Gaussian process $B = \{B_t,\ t \in [0,T]\}$ by setting

$$B_t = \int_0^t K(t,s)\,\mathrm{d}\beta(s), \qquad 0 \le t \le T.$$

The covariance function

$$\mathbb{E}(B_t B_s) = R(t,s) \tag{2.4}$$

can be expressed in terms of K as

$$R(t,s) = \int_0^{t\wedge s} K(t,r)K(s,r)\,\mathrm{d}r.$$

Let $\mathcal{E}$ be the set of step functions on $[0,T]$. For any function $f \in \mathcal{E}$ having the representation

$$f(s) = \sum_{k=1}^{N} f_k \mathbb{I}_{(s_k,s_{k+1}]}(s)$$

we define the stochastic integral

$$\mathcal{I}(f) = \int_0^T f(s)\,\mathrm{d}B(s) = \sum_{k=1}^{N} f_k[B(s_{k+1}) - B(s_k)].$$

Then, $\mathcal{I}(f)$ is a Gaussian random variable with zero mean and variance that can be determined via (2.4). The set of Gaussian random variables defined by the elements of $\mathcal{E}$ is a subset of the linear space

$$\mathcal{H}_1 = \{X \in L^2(\Omega) \,|\, \exists\, f_n \in \mathcal{E} \,:\, \mathcal{I}(f_n) \to X \text{ in } L^2\,\}.$$

This space shall be identified with the first Wiener chaos generated by B.

The "reproducing kernel Hilbert space" (RKHS) Λ is defined as the closure of $\mathcal{E}$ with respect to the scalar product

$$\langle \mathbb{I}_{[0,t]}, \mathbb{I}_{[0,s]} \rangle = R(t,s).$$

The mapping $\mathcal{I} : \mathbb{I}_{[0,t]} \to B(t)$ provides an isometry between Λ and $\mathcal{H}_1$.

Remark 2.3. Let us define, for $f \in L^2(0,T)$,

$$\delta(f) = \int_0^T f(s)\,\mathrm{d}\beta(s);$$

δ denotes the Wiener integral with respect to the Brownian motion β.

It is well known that for a Brownian motion β, the space $L^2(\mathbb{R}_+)$ is isometric to the space $\mathcal{H}_1$ of Gaussian random variables. The problem of existence of a similar structure for a general Gaussian process has been addressed by many authors, see [1, 2, 23].

The kernel K defines an operator in $L^2(0,T)$, that will still be denoted by K, given by

$$(Kh)(t) = \int_0^t K(t,s)h(s)\,\mathrm{d}s.$$

We introduce further the "transfer operator" K^* from $\mathcal{E}$ into $L^2(0,T)$: for any $\phi \in \mathcal{E}$ we set $K^*\phi$ to be the unique function such that

$$\int_0^T (K^*\phi)(s)h(s)\,\mathrm{d}s = \int_0^T \phi(s)\partial_s(Kh)(s)\mathrm{d}s$$

for any $h \in L^2(0,T)$. This operator can be naturally extended to the whole Λ, and the following representation holds for any measurable ϕ defined on $(0,t)$:

$$(K^*\phi)(s) = \mathbb{I}_{(0,t)}(s)\phi(s)K(t,s) + \mathbb{I}_{(0,t)}(s)\int_s^t [\phi(r) - \phi(s)]\partial_1 K(r,s)\,\mathrm{d}r. \qquad (2.5)$$

When K is a smooth kernel, we have

$$(K^*\phi)(s) = \mathbb{I}_{(0,t)}(s)\phi(s)K(s,s) + \mathbb{I}_{(0,t)}(s)\int_s^t \phi(r)\partial_1 K(r,s)\,\mathrm{d}r.$$

In particular, for any element $\phi \in \mathcal{E}$ the element $\mathcal{I}(\phi)$ of $\mathcal{H}_1$ can be written as

$$\mathcal{I}(\phi) = \int_0^T (K^*\phi)(s)\,\mathrm{d}\beta(s) = \delta(K^*\phi), \qquad (2.6)$$

where $\delta(u)$ is the integral defined in Remark 2.3.

Let us notice the following diagram:

$$\begin{tikzcd} \Lambda \arrow[rr, "\mathcal{I}"] \arrow[dr, "K^*"'] & & \mathcal{H}_1 \\ & L^2(0,T) \arrow[ur, "\delta"'] & \end{tikzcd}$$

Proposition 2.4. *By the above construction, it follows that*

$$\mathbb{E}|\mathcal{I}(\phi)|^2 = \|\phi\|_\Lambda^2 = \|(K^*\phi)\|_{L^2(0,T)}^2 \qquad (2.7)$$

for any $\phi \in \Lambda$.

2.2. Stochastic Volterra equations

Definition 2.5. We denote by $S(t),\ t \geq 0$ the resolvent operator related to (1.2); following [7], we say that the mild solution to Eq. (1.2) is the process

$$u(t) = \int_0^t S(t-s)\,\mathrm{d}B(s), \quad t \geq 0. \qquad (2.8)$$

We shall write the stochastic convolution in (2.8) componentwise. We introduce the solution $s_\mu(t)$ to the scalar resolvent equation

$$s_\mu(t) + \mu(s_\mu * g_\rho)(t) = 1, \qquad t > 0, \quad \mu > 0; \tag{2.9}$$

the resolvent operator $S(t)$ is defined by

$$S(t)x = \sum_{k=1}^{\infty} s_{\mu_k}(t)\langle x, e_k\rangle e_k$$

for any $x \in U$.

Using the representation (2.3), formula (2.8) can be written as a series

$$\int_0^t S(t-r)\,\mathrm{d}B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \int_0^t s_{\mu_k}(t-r)\,\mathrm{d}B^{(k)}(r)\,e_k.$$

If we denote

$$s^*_\mu(t,r) = (K^* s_\mu(t-\cdot))(r) = s_\mu(t-r)K(t,r) + \int_r^t [s_\mu(t-\tau) - s_\mu(t-r)]\partial_1 K(\tau, r)\,\mathrm{d}\tau,$$

then, with the integration by parts formula (2.6), we can write

$$\int_0^t S(t-r)\,\mathrm{d}B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \int_0^t s^*_{\mu_k}(t,r)\,\mathrm{d}\beta^{(k)}(r)\,e_k.$$

Theorem 2.6. *Assume the convergence of the series*

$$\sum_{k=1}^{\infty} \lambda_k \int_0^t |s^*_{\mu_k}(t,r)|^2\,\mathrm{d}r < \infty, \qquad t \in [0,T]. \tag{2.10}$$

Then the stochastic convolution process $u(t)$ defined in Eq. (2.8) *is a well-defined centered square integrable Gaussian process in U.*

2.3. Mittag-Leffler's function

Let us introduce Mittag-Leffler's function $\mathcal{E}_\rho(t)$ (see [15, Vol. 3, Ch. 18]) defined, for any $\rho > 0$, by the formula

$$\mathcal{E}_\rho(t) = \sum_{k=0}^{\infty} \frac{(-1)^k t^k}{\Gamma(\rho k + 1)}, \qquad t \in \mathbb{R}. \tag{2.11}$$

The asymptotic behavior of $\mathcal{E}_\rho$ can be explicitly controlled. Actually we have, for any $0 < \rho < 2$:

$$\mathcal{E}_\rho(x) = O(\frac{1}{x}), \qquad x \to \infty. \tag{2.12}$$

Notice that the above formula does not hold for $\rho = 2$, since it is $\mathcal{E}_2(x) = O(1)$. Moreover $\mathcal{E}_\rho(x)$ is a completely monotone function for $x \geq 0$ and $0 < \rho \leq 1$.

The interest of this function rests on the following fact, compare [5].

Proposition 2.7. *The scalar resolvent kernel s_μ, solution of Eq.* (2.9), *can be expressed in terms of Mittag-Leffler's function by the formula*

$$s_\mu(t) = \mathcal{E}_\rho(\mu t^\rho). \tag{2.13}$$

Remark 2.8. As we already noticed, in case $\rho \in (0,1)$ the function $s_\mu(t)$ is completely monotonic; recalling Bernstein's theorem, we obtain a representation of $s_\mu(t)$ as a Laplace transform

$$s_\mu(t) = \int_0^\infty e^{-\lambda t} g_\mu(\lambda)\,\mathrm{d}\lambda, \qquad t \ge 0,$$

where

$$g_\mu(\lambda) = \frac{\sin(\rho\pi)}{\pi} \frac{\mu \lambda^{\rho-1}}{\lambda^{2\rho} + \mu^2 + 2\mu\lambda^\rho \cos(\rho\pi)}. \tag{2.14}$$

In analogy with the norm $\|K\|_{\gamma,q,p,t}$ introduced in Eq. (2.2), in order to handle the function $s_\mu(t)$ we introduce the quantity

$$\begin{aligned} \|\varphi\|^2_{-\gamma,q',p',t} = {} & \left(\int_0^t |\varphi(t-s)^{-\gamma}|^{2q'}\,\mathrm{d}s \right)^{1/q'} \\ & + \left(\int_0^t \left| \int_s^t \left| [\varphi(t-u) - \varphi(t-s)](t-s)^{-1-\gamma} \right|^{p'} \mathrm{d}u \right|^2 \mathrm{d}s \right)^{1/p'}. \end{aligned} \tag{2.15}$$

3. Existence of the stochastic convolution

In this section, we give conditions on the kernel $K(t,s)$ and the fractional integration kernel g_ρ in order to control the convergence of the series (2.10).

Again, we employ the fact that we may consider each class of work separately. We consider, for any k, the scalar stochastic convolution process

$$b_k(t) = \int_0^t s_{\mu_k}(t-r)\,\mathrm{d}B^{(k)}(r) = \int_0^t s^*_{\mu_k}(t,r)\,\mathrm{d}\beta^{(k)}(r), \qquad t \in [0,T]. \tag{3.1}$$

Theorem 3.1. *Assume that the kernel $K(t,s)$ belongs to $E_{\gamma,q,p,t}$ for a set of admissible parameters $0 \le \gamma < \frac{1}{2}$, $q > \frac{1}{1-2\gamma}$, $p > \frac{1}{1-\gamma}$. Then it holds that*

$$\|s^*_\mu(t,\cdot)\|^2_{L^2(0,t)} \le O(1) \|s_\mu\|^2_{-\gamma,q',p',t} \|K\|^2_{\gamma,q,p,t}.$$

Proof. Our goal is to evaluate the L^2-norm of the function s^*_μ. Recalling the representation in (2.5),

$$s^*_\mu(t,s) = s_\mu(t-s)K(t,s) + \int_s^t [s_\mu(t-r) - s_\mu(t-s)]\partial_1 K(r,s)\,\mathrm{d}r,$$

we search for an estimate of the L^2-norm of this quantity,

$$\|s^*_\mu(t,\cdot)\|^2_{L^2(0,t)} \le 2\int_0^t |s_\mu(t-s)K(t,s)|^2\,\mathrm{d}s \\ + 2\int_0^t \left|\int_s^t [s_\mu(t-r) - s_\mu(t-s)]\partial_1 K(r,s)\,\mathrm{d}r\right|^2 \mathrm{d}s. \quad (3.2)$$

Let us denote by I_1 and I_2 the first, respectively the second, integral which appears in the right-hand side of formula (3.2).

By Hölder's inequality we have, for any $\gamma \ge 0$,

$$I_1 \le \left(\int_0^t |s_\mu(t-s)(t-s)^{-\gamma}|^{2q'}\,\mathrm{d}s\right)^{1/q'} \left(\int_0^t |K(t,s)(t-s)^{\gamma}|^{2q}\,\mathrm{d}s\right)^{1/q}$$

where q and q' are conjugate exponents: $\frac{1}{q} + \frac{1}{q'} = 1$; also, we have, for a pair of conjugate exponents $\frac{1}{p} + \frac{1}{p'} = 1$,

$$I_2 \le \left\{\int_0^t \left(\int_s^t \left|[s_\mu(t-u) - s_\mu(t-s)](u-s)^{-1-\gamma}\right|^{p'}\mathrm{d}u\right)^2 \mathrm{d}s\right\}^{1/p'} \\ \left\{\int_0^t \left(\int_s^t \left|(u-s)^{\gamma+1}\partial_1 K(u,s)\right|^{p}\mathrm{d}u\right)^2 \mathrm{d}s\right\}^{1/p}.$$

Thanks to definition (2.15) this concludes the proof. □

Using the properties shown in Section 2.3, it is a matter of (tedious) computations to explicitly work out the quantity $\|s_\mu\|^2_{-\gamma,q',p',t}$: it will be the object of next lemma. Hence, to consider the examples proposed in the introduction, it remains to estimate the norm of the kernels in the relevant spaces $E_{\gamma,q,p,t}$.

Lemma 3.2. *Assume that γ, p and q are admissible parameters and let p' and q' denote the conjugate exponents, i.e., $\frac{1}{p} + \frac{1}{p'} = 1$, and similarly for q. Then the quantity $\|s_\mu\|^2_{-\gamma,q',p',t}$ is bounded by*

$$\|s_\mu\|^2_{-\gamma,q',p',t} \le O(1)\Big[\mu^{-(1-2\gamma q')/(\rho q')} + t^{(1-2(\rho+\gamma)q')/q'}\mu^{-2} \\ + t^{(3-2(1+\gamma)p'-2\rho)/p'}\mu^{-2/p'} + \mu^{(2\gamma p'+2p'-3)/\rho p'}\Big]$$

if $2(\gamma+\rho)q' \ne 1$ and $\gamma p' + p' \ne \frac{3}{2} - \rho$.

3.1. Some examples

In this section we show some applications of our results; we consider some specific examples of Gaussian kernels, defined by the filter (1.1) with respect to some kernels of the classes proposed in the introduction.

The first example is a simple perturbation of a fractional convolution operator, provided by

$$K_\vartheta(t,s) = \left(\frac{t}{(t-s)(2t-s)}\right)^\vartheta, \qquad 0 < s < t < T;$$

this kernel belongs to $E_{\vartheta,\infty,\infty,t}$ and its norm remains bounded for $t \to \infty$. In such a case, we obtain from Lemma 3.2 that

$$\|s_\mu\|^2_{-\vartheta,1,1,t} \le O(1)\Big[\mu^{-(1-2\vartheta)/\rho} + \mu^{-2} t^{1-2(\rho+\vartheta)}\Big].$$

Theorem 3.3. *Let $\{-\mu_k,\ k \in \mathbb{N}\}$ be the sequence of eigenvalues of the operator A, g_ρ be the fractional integration kernel, $\rho \in (0,1)$, and $B(t)$ be a Gaussian process in U, defined by Formula* (2.3) *with the kernel $K_\vartheta(t,s)$, $\vartheta \in (0,\frac12)$, defined above. Then the stochastic convolution process $\{u(t),\ t \in [0,T]\}$, defined in Eq.* (2.8), *is well defined in the following cases:*

(i) *if $\rho + \vartheta < \frac12$ and $\displaystyle\sum_{k=1}^\infty \frac{\lambda_k}{\mu_k^2} < +\infty$;*

(ii) *if $\rho + \vartheta = \frac12$ and $\displaystyle\sum_{k=1}^\infty \frac{\lambda_k \log(\mu_k)}{\mu_k^2} < +\infty$;*

(iii) *if $\rho + \vartheta > \frac12$ and $\displaystyle\sum_{k=1}^\infty \frac{\lambda_k}{\mu_k^{(1-2\vartheta)/\rho}} < +\infty$.*

Further, in case (iii), *the stochastic convolution process is well defined for all times:*

$$\sup_{t\ge0} \mathbb{E}|u(t)|^2 < +\infty.$$

3.2. Almost regular kernel

Let us consider the kernel $k(t) = \log(1+\frac1t)$. This kernel is completely monotonic, with a singularity in zero: $k(0+) = +\infty$.

Let us consider the integral I_1 which appears in Eq. (3.2),

$$\begin{aligned} I_1 &= \int_0^t |s_\mu(t-s)K(t,s)|^2\,\mathrm{d}s = \int_0^t |\mathcal{E}_\rho(\mu s^\rho)\log(1+\frac1s)|^2\,\mathrm{d}s \\ &= \mu^{-1/\rho}\int_0^\alpha |\mathcal{E}_\rho(s^\rho)\log(1+\frac{\mu^{1/\rho}}{s})|^2 \mathrm{d}s + \mu^{-1/\rho}\int_\alpha^{\mu^{1/\rho}t} |\mathcal{E}_\rho(s^\rho)\log(1+\frac{\mu^{1/\rho}}{s})|^2 \mathrm{d}s \\ &= I_{1,1} + I_{1,2}. \end{aligned}$$

We may compute explicitly the two terms

$$I_{1,1} \le O(1)\mu^{1/\rho}\log^2(\mu)$$

and

$$I_{1,2} \le O(1)\mu^{1/\rho}\int_\alpha^{\mu^{1/\rho}t} |s^{-\rho}\log(1+\frac{\mu^{1/\rho}}{s})|^2\,\mathrm{d}s = O(1)\begin{cases}\mu^{-2}\log^3(\mu), & \rho = \frac12,\\ \mu^{-1/\rho}\log^2(\mu), & \rho > \frac12.\end{cases}$$

Hence,

$$I_1 \le O(1) \begin{cases} \mu^{-2}\log^3(\mu), & \rho = \frac{1}{2}, \\ \mu^{-1/\rho}\log^2(\mu), & \rho > \frac{1}{2}. \end{cases}$$

Next, we consider I_2; as already mentioned, it holds that $(t-s)\partial_1 K(t,s)$ is bounded on $0 < s < t < T$, and we get

$$I_2 \le O(1) \int_0^t \left(\int_s^t \frac{s_\mu(t-u) - s_\mu(t-s)}{u-s} \, \mathrm{d}u \right)^2 \mathrm{d}s;$$

now – as in the proof of Theorem 3.1 – choosing $\gamma = 0$ and $\rho \ge \frac{1}{2}$ we have

$$I_2 \le O(1) \begin{cases} \mu^{-2}\log(\mu), & \rho = \frac{1}{2}, \\ \mu^{-1/\rho}, & \rho > \frac{1}{2}. \end{cases}$$

Finally, we may summarize our result as follows.

Theorem 3.4. *Let $\{-\mu_k,\ k \in \mathbb{N}\}$ be the sequence of eigenvalues of the operator A, g_ρ be the fractional integration kernel, $\rho \in (0,1)$, and $B(t)$ be a Gaussian process in U, defined by Formula* (2.3) *with the kernel $K(t,s) = \log(1 + \frac{1}{t-s})$. Then the stochastic convolution process $\{u(t),\ t \in [0,T]\}$, defined in Eq.* (2.8)*, is well defined in the following cases:*

(i) *if $\rho < \frac{1}{2}$ and* $\displaystyle\sum_{k=1}^{\infty} \frac{\lambda_k}{\mu_k^2} < +\infty$*;*

(ii) *if $\rho = \frac{1}{2}$ and* $\displaystyle\sum_{k=1}^{\infty} \frac{\lambda_k \log^3(\mu_k)}{\mu_k^2} < +\infty$*;*

(iii) *if $\rho > \frac{1}{2}$ and* $\displaystyle\sum_{k=1}^{\infty} \frac{\lambda_k \log^2(\mu_k)}{\mu_k^{1/\rho}} < +\infty$*.*

3.3. Highly singular kernel

As a model for the case of highly singular kernel, we consider the following example. Let

$$k(t) = \frac{t^{-\frac{1}{2}}}{\log(4 + \frac{1}{t})}, \qquad t > 0,$$

and denote $K(t,s) = k(t-s)$. Notice that $k(t) \in L^2(0,1)$ but $\lim\limits_{t \to 0} t^{\frac{1}{2}-\vartheta} k(t) = +\infty$ for all $\vartheta > 0$.

At first, notice that

$$\int_0^t |(t-s)^\gamma K(t,s)|^{2q} \, \mathrm{d}s < \infty$$

provided $1 \le q \le \frac{1}{1-2\gamma}$, i.e., $q' \ge 2\gamma$; then, the discussion at the end of Section 2.3 implies that, for any $\gamma \in [0, \frac{1}{2})$, we must choose $q = \frac{1}{1-2\gamma}$ and

$$I_1 \le O(1).$$

Next, we evaluate the derivative of $k(t)$:

$$k'(t) = \frac{2 - (1 + 4\,t)\,\log\left[4 + \frac{1}{t}\right]}{2\,t^{3/2}\,(1 + 4\,t)\log\left[4 + \frac{1}{t}\right]^2}.$$

Then it holds $t^{3/2}k'(t) \in L^\infty(0,T)$; hence, we may proceed as in the proof of Theorem 3.1 with $\gamma = \frac{1}{2}$ to get

$$I_2 \le O(1);$$

we conclude that the series in Eq. (2.10) converges if and only if $\sum_{k=1}^{\infty} \lambda_k < \infty$.

4. Fractional Brownian motion

In this section we give an account of the situation concerning the case of a linear stochastic Volterra equation perturbed by a fractional Brownian motion (fBm). Because of its possible interest in the applications, we consider both the case of a singular kernel (case $H < \frac{1}{2}$) and the case of a smooth kernel (case $H > \frac{1}{2}$), although this last example differs in spirit from the remaining of the paper.

Let us start with a review of the main properties of the fBm. Fractional Brownian motion is a natural candidate as a model of noise in several fields of applied mathematics, compare [21]. In the last years, it has been the subject of an intensive study by means of several authors, also from a purely theoretic point of view, compare, e.g., [2, 22] and the references therein.

Definition 4.1. If $0 < H < 1$, the fractional Brownian motion with Hurst parameter H is the Gaussian process $\{B_H(t),\ t \ge 0\}$ satisfying

$$B_H(0) = \mathbb{E}[B_H(t)] = 0 \text{ for all } t \ge 0,$$
$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right) \text{ for all } s,\, t \ge 0.$$

There exist several different representations of the fBm: in the original paper of Mandelbrot and van Ness [20] the "moving average representation" is given,

$$B_H(t) = \frac{1}{c_1(H)} \int_{\mathbb{R}} [(t-s)_+^{H-1/2} - (-s)_+^{H-1/2}]\, \mathrm{d}\beta(s), \qquad t \ge 0,$$

where β is a standard Brownian motion,

$$c_1(H) = \left(\int_0^\infty [(1+s)^{H-1/2} - s^{H-1/2}]^2\, \mathrm{d}s + \frac{1}{2H} \right)^{1/2},$$

and $(x)_+$ denotes the positive part of x.

Other representations may be found, for instance, in [13] and [22]. Here we recall the following result, which is proved in [1].

Proposition 4.2. *Assume that the kernel $K(t,s)$ has the expression*

$$K_H(t,s) = c_H(t-s)^{H-\frac{1}{2}} + c_H(\frac{1}{2}-H)\int_s^t (u-s)^{H-\frac{3}{2}}\left(1-(\frac{s}{u})^{\frac{1}{2}-H}\right)\,\mathrm{d}u, \quad (4.1)$$

where c_H is a normalizing constant, given by

$$c_H = \left(\frac{2H\,\Gamma(\frac{3}{2}-H)}{\Gamma(H+\frac{1}{2})\Gamma(2-2H)}\right)^{1/2}.$$

Then the process

$$B_H(t) = \int_0^t K_H(t,s)\,\mathrm{d}\beta(s)$$

is a fractional Brownian motion with Hurst parameter $H \in (0,1)$.

If $H = \frac{1}{2}$, then $B_H(t)$ coincides with a standard Brownian motion $\beta(t)$. For a discussion on other properties of the fBm, we address to the papers quoted before.

4.1. Hurst parameter $H > \frac{1}{2}$

Let $K(t,s)$ be the kernel defined in Proposition 4.2, with Hurst parameter $H > \frac{1}{2}$. In this case, the elements of the Hilbert space Λ may not be functions but distributions of negative orders, compare [23]. In this case, however, we can find a linear space of functions contained in Λ in the following way. Let $|\mathcal{H}|$ be the space of functions in $L^1([0,T]) \cap L^2([0,T])$ endowed with the scalar product

$$\langle \phi, \psi \rangle_H = \alpha_H \int_0^T \int_0^T |t-r|^{2H-2}\phi(r)\psi(t)\,\mathrm{d}r\,\mathrm{d}t,$$

where $\alpha_H = H(2H-1)$. The isometry formula holds for every function $\phi \in \Lambda$:

$$\mathbb{E}\,|\mathcal{I}(\phi)|^2 = \|\phi\|_H^2.$$

Lemma 4.3. *Assume $H \neq \rho$. The stochastic convolution process $\{b(t),\ t \in [0,T]\}$ defined in* (3.1) *is a zero mean Gaussian random variable, with variance bounded by*

$$\mathbb{E}(b(t)^2) = O(1)\left[\frac{1}{\mu^{2H/\rho}} + \frac{1}{\mu^2}\right] \quad (4.2)$$

for any $t \in [0,T]$.

Proof. In view of the construction of the stochastic integral with respect to a fractional Brownian motion, it only remains to prove (4.2). It suffices to evaluate the $|\mathcal{H}|$-norm of $s_\mu(t-\tau)\mathbb{I}_{(0,t)}(\tau)$, and we have

$$\begin{aligned}
&\alpha_H \int_0^\infty \int_0^\infty s_\mu(t-\tau)\mathbb{I}_{(0,t)}(\tau)s_\mu(t-\sigma)\mathbb{I}_{(0,t)}(\sigma)|\tau-\sigma|^{2H-2}\,\mathrm{d}\tau\,\mathrm{d}\sigma \\
&= \alpha_H \int_0^t \int_0^t s_\mu(t-\tau)s_\mu(t-\sigma)|\tau-\sigma|^{2H-2}\,\mathrm{d}\tau\,\mathrm{d}\sigma \\
&= 2\frac{\alpha_H}{\mu^{2H/\rho}} \int_0^{t\mu^{1/\rho}} \int_0^\sigma \mathcal{E}_\rho(\tau^\rho)|\tau-\sigma|^{2H-2}\,\mathrm{d}\tau\,\mathcal{E}_\rho(\sigma^\rho)\,\mathrm{d}\sigma. \quad (4.3)
\end{aligned}$$

Consider first the inner integral; we have

$$\begin{aligned}\int_0^\sigma \mathcal{E}_\rho(\tau^\rho)|\tau-\sigma|^{2H-2}\,\mathrm{d}\tau &= \sum_{k=0}^\infty (-1)^k \frac{1}{\Gamma(k\rho+1)} \int_0^\sigma \tau^{k\rho}|\sigma-\tau|^{2H-2}\,\mathrm{d}\tau \\ &= \sum_{k=0}^\infty (-1)^k \frac{1}{\Gamma(k\rho+1)} \sigma^{k\rho+2H-1} \frac{\Gamma(k\rho+1)\Gamma(2H-1)}{\Gamma(k\rho+2H)} \\ &= \Gamma(2H-1)\sigma^{2H-1} \sum_{k=0}^\infty (-1)^k \frac{1}{\Gamma(k\rho+2H)} \sigma^{k\rho}\end{aligned}$$

where the last quantity is equal to

$$\Gamma(2H-1)\sigma^{2H-1}\mathcal{E}_{\rho,2H}(\sigma^\rho).$$

We introduce this quantity in (4.3) to get

$$\begin{aligned}\alpha_H \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} & s_\mu(t-\tau)\mathbb{I}_{(0,t)}(\tau)s_\mu(t-\sigma)\mathbb{I}_{(0,t)}(\sigma)|\tau-\sigma|^{2H-2}\,d\tau\,\mathrm{d}\sigma \\ &= O(1)\frac{1}{\mu^{2H/\rho}} \int_0^{t\mu^{1/\rho}} \sigma^{2H-1}\mathcal{E}_{\rho,2H}(\sigma^\rho)\,\mathcal{E}_\rho(\sigma^\rho)\,\mathrm{d}\sigma.\end{aligned}$$

Mittag-Leffler's functions are bounded near the origin, and assuming $H > \frac{1}{2}$ we get also $2H-1>0$, which implies, together with (2.12), that the following decomposition holds:

$$\begin{aligned}\alpha_H \int_0^\infty\int_0^\infty & s_\mu(t-\tau)\mathbb{I}_{(0,t)}(\tau)s_\mu(t-\sigma)\mathbb{I}_{(0,t)}(\sigma)|\tau-\sigma|^{2H-2}\,\mathrm{d}\tau\,\mathrm{d}\sigma \\ &= O(1)\frac{1}{\mu^{2H/\rho}}\left[\int_0^1 \sigma^{2H-1}\,\mathrm{d}\sigma + \int_1^{t\mu^{1/\rho}} \sigma^{2H-1}\sigma^{-\rho}\,\sigma^{-\rho}\,\mathrm{d}\sigma\right] \\ &= O(1)\frac{1}{\mu^{2H/\rho}}\left[1+(t\mu^{1/\rho})^{2H-2\rho}\right] \\ &= O(1)\left[\frac{1}{\mu^{2H/\rho}} + t^{2H-2\rho}\frac{1}{\mu^2}\right]\end{aligned}$$

which proves (4.2). □

Corollary 4.4. *An inspection of the above proof shows that, when $H=\rho$, estimate (4.2) shall be modified by*

$$\mathbb{E}(b(t)^2) = O(1)\frac{\log(\mu)}{\mu^2}.$$

Corollary 4.5. *In the above assumptions, the stochastic convolution $b(t)$ is well defined for $t\in\mathbb{R}_+$ provided*

$$H<\rho,$$

and in such case one has

$$\mathbb{E}(b(t)^2) = O(1)\frac{1}{\mu^{2H/\rho}}$$

for any $t \in \mathbb{R}_+$.

4.2. Fractional Brownian motion with Hurst parameter $H < \frac{1}{2}$

Here we have

$$K_H(t,s) = c_H(t-s)^{H-1/2} + s^{H-1/2}F_1(\frac{t}{s})$$

where

$$F_1(z) = \int_0^{z-1} x^{H-3/2}(1-(x+1)^{H-1/2})\,\mathrm{d}x.$$

From (4.1) we also obtain (compare [2])

$$\frac{\partial K_H}{\partial t}(t,s) = c_H(H-\frac{1}{2})\left(\frac{s}{t}\right)^{1/2-H}(t-s)^{H-3/2}.$$

The following result is proved in [2, Proposition 8]. For a short reference about the main definitions concerning fractional calculus, see the remark at the end of this section.

Proposition 4.6. *The RKHS* Λ *is the space* $I_{T-}^{1/2-H}(L^2)$ *and the operator* K^* *is given by*

$$(K_H^* h)(s) = c_u s^{1/2-H} D_{T-}^{1/2-H}(h_{H-1/2})(s),$$

where h_α *denotes the function* $h_\alpha(x) = x^\alpha h(x)$.

Remark 4.7. The above proposition shows that this case is strictly connected to the case of a fractional derivative kernel $K(t,s) = g_{H+1/2}(t-s)$. Actually, the RKHS Λ coincide in both cases, although the operators K^* differ:

$$(K_1^* h)(s) = D_{T-}^{1/2-H}h(s) \quad \text{while} \quad (K_H^* h)(s) = c_u s^{1/2-H} D_{T-}^{1/2-H}(h_{H-1/2})(s).$$

Using Proposition 4.6, we derive the following expression for s_μ^*:

$$\begin{aligned} s_\mu^*(t,r) \;=\; & c_H(t-r)^{H-1/2}s_\mu(t-r) + (H-\frac{1}{2})c_H s^{1/2-H} \\ & \times \int_r^t [x^{H-1/2}s_\mu(t-x) - r^{H-1/2}s_\mu(t-r)](x-r)^{H-3/2}\,\mathrm{d}x. \end{aligned}$$

which implies, in particular,

$$\begin{aligned} \|s_\mu^*(t,\cdot)\|_{L^2(0,t)}^2 \le{} & 2c_H^2 \int_0^t |s_\mu(t-s)(t-s)^{H-1/2}|^2\,\mathrm{d}s + 2c_H^2(H-\frac{1}{2})^2 \\ \times \int_0^t & \left| s^{1/2-H}\int_s^t [r^{H-1/2}s_\mu(t-r) - s^{H-1/2}s_\mu(t-s)](r-s)^{H-3/2}\,\mathrm{d}r \right|^2 \mathrm{d}s. \end{aligned}$$

Since $H < 1/2$, $r^{H-1/2} \leq s^{H-1/2}$ for any $r \in (s,t)$, and

$$\|s^*_\mu(t,\cdot)\|^2_{L^2(0,t)} \leq 2c_H^2 \int_0^t |s_\mu(t-s)(t-s)^{H-1/2}|^2 \,\mathrm{d}s \qquad (4.4)$$
$$+ 2c_H^2 (H - \frac{1}{2})^2 \int_0^t \left| \int_s^t [s_\mu(t-r) - s_\mu(t-s)](r-s)^{H-3/2} \,\mathrm{d}r \right|^2 \mathrm{d}s.$$

Up to a constant, estimate (4.4) coincides with (3.2) in case $K(t,s) = (t-s)^{H-3/2}$, hence we can appeal to Lemma 3.2 to get the following bound for $\|s^*_\mu(t,\cdot)\|^2_{L^2(0,t)}$:

$$\|s^*_\mu(t,\cdot)\|^2_{L^2(0,t)} = O(1)\left(\mu^{-2H/\rho} + \mu^{-2}\right).$$

However, as before, we shall consider separately the case $\rho = H$, where the above bound becomes

$$\|s^*_\mu(t,\cdot)\|^2_{L^2([0,t])} = O(1)\mu^{-2H/\rho}\log(\mu).$$

Finally, we are in a position to state the main theorem in the case of a fractional Brownian motion of parameter H.

Theorem 4.8. *Let $\{-\mu_k,\ k \in \mathbb{N}\}$ be the sequence of eigenvalues of the operator A, g_ρ be the fractional integration kernel, $\rho \in (0,2)$, and $B(t) = B_H(t)$ be a fractional Brownian motion in U, with Hurst parameter H, defined by* (2.3) *with the kernel $K(t,s) = K_H(t,s)$ defined in Proposition* 4.2. *Then the stochastic convolution process $\{u(t),\ t \in [0,T]\}$, defined in* (2.8), *is well defined in the following cases:*

(i) *if $\rho < H$ and* $\displaystyle\sum_{k=1}^\infty \frac{\lambda_k}{\mu_k^2} < +\infty$;

(ii) *if $\rho = H$ and* $\displaystyle\sum_{k=1}^\infty \frac{\lambda_k \log(\mu_k)}{\mu_k^2} < +\infty$;

(iii) *if $\rho > H$ and* $\displaystyle\sum_{k=1}^\infty \frac{\lambda_k}{\mu_k^{2H/\rho}} < +\infty$.

Remark 4.9. For $\alpha > 0$ we can define the *(right)-fractional Riemann-Liouville integral of order α* of an integrable function f on $[0,T]$ as

$$I^\alpha_{T-} f(x) = \int_x^T f(y) g_\alpha(y-x)\,\mathrm{d}y.$$

We will denote by $I^\alpha_{T-}(L^2)$ the class of functions f in $L^2([0,T])$ which may be represented as an I^α_{T-}-integral of some function $\phi \in L^2([0,T])$.

The fractional derivative can be introduced, in a natural way, as the inverse operation of the fractional integral. The *(right)-fractional Riemann-Liouville derivative of f of order α* is given by

$$D^\alpha_{T-} f(x) = \frac{1}{\Gamma(1-\alpha)}\left(\frac{f(x)}{(T-x)^\alpha} - \alpha \int_x^T \frac{f(r)-f(x)}{(r-x)^{\alpha+1}}\,\mathrm{d}r\right).$$

If $f \in I^{\alpha}_{T-}(L^2)$, the function $\phi = D^{\alpha}_{T-} f$ is the unique element of $L^2([0,T])$ such that $f = I^{\alpha}_{T-}\phi$.

A complete introduction to the subject is given in [25]. Short notes can be found in most of the papers on fractional Brownian motion, see, e.g., [2].

4.3. An example in dimension N

It seems interesting to discuss a specific example, compare [11, Section 5.5] and [9]. Assume that $\mathcal{O}$ is the cube $[0,\pi]^N$ in $\mathbb{R}^N$ with boundary $\partial\mathcal{O}$. Let A be the linear operator

$$\begin{aligned} &D(A) = H^2(\mathcal{O}) \cap H^1_0(\mathcal{O}),\\ &Au = \Delta^m u, \qquad \forall\, u \in D(A), \end{aligned}$$

where Δ represents the Laplace operator. Then

$$e_{n_1,\dots,n_N}(\xi) = \left(\frac{2}{\pi}\right)^{N/2} \sin(n_1\xi_1)\cdot \dots \cdot \sin(n_N\xi_N),$$

while

$$\mu_{n_1,\dots,n_N} = n_1^{2m} + \dots + n_N^{2m}.$$

Let us consider Equation (1.2) where $B(t) = B_H(t)$ is a cylindrical fractional Brownian motion of parameter $H \in (0,1)$. Then, using the results in Theorem 4.8, it turns out that we shall consider the following cases:

Let $\rho < H$. Then, in order to ensure the convergence of the series

$$\sum \frac{1}{(n_1^{2m} + \dots + n_N^{2m})^2},$$

we must assume that $N < 4m$.

Let us consider now the case $\rho > H$. Then we shall consider the series

$$\sum \frac{1}{(n_1^{2m} + \dots + n_N^{2m})^{2H/\rho}}.$$

This series converges if and only if

$$\rho < \frac{4mH}{N}.$$

Notice that the above condition also implies that $4m/N > 1$, hence a solution exists if $N < 4m$.

Corollary 4.10. *In the above framework, a solution to Equation* (1.2) *exists if* $N < 4m$ *and, moreover,* $\rho < \dfrac{4mH}{N}$.

It is also of some interest to consider the special case $\rho = 1$. In this case, Equation (1.2) becomes a linear stochastic differential equation in U:

$$\mathrm{d}u(t) = Au(t)\,\mathrm{d}t + \mathrm{d}B_H(t). \tag{4.5}$$

The above construction defines in which cases such an equation has a mild solution.

Corollary 4.11. *In the above framework, a solution to Equation* (4.5) *exists if* $N < 4mH$. *In particular, if* $N = 1$ *and* $m = 1$, *this implies that a sufficient condition for a solution to* (4.5) *to exist is* $H > \frac{1}{4}$.

References

[1] E. Alós, O. Mazet, and D. Nualart, *Stochastic calculus with respect to Gaussian processes,* Ann. Probab., **29 (2)** (2001), 766–801.

[2] E. Alós and D. Nualart, *Stochastic integration with respect to the fractional Brownian motion,* Stoch. Stoch. Rep., **75 (3)** (2003), 129–152.

[3] E. Bajlekova, *Fractional Evolution Equations in Banach Spaces*, Dissertation, Technische Universiteit Eindhoven, Eindhoven, 2001.

[4] L. Bel, G. Oppenheim, L. Robbiano, and M. C. Viano, *Distribution processes with stationary fractional increments,* In Matignon and Montseny [21], 43–54.

[5] S. Bonaccorsi and L. Tubaro, *Mittag-Leffler's function and stochastic linear Volterra equations of convolution type,* Stochastic Anal. Appl., **21 (1)** (2003), 61–78.

[6] Ph. Carmona, L. Coutin, and G. Montseny, *Stochastic integration with respect to fractional Brownian motion*, Ann. Inst. H. Poincaré Probab. Statist., **39 (1)** (2003), 27–68.

[7] Ph. Clément and G. Da Prato, *Some results on stochastic convolutions arising in Volterra equations perturbed by noise,* Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., **7 (3)** (1996), 147–153.

[8] Ph. Clément and G. Da Prato, *White noise perturbation of the heat equation in materials with memory,* Dynam. Systems Appl., **6 (4)** (1997), 441–460.

[9] Ph. Clément, G. Da Prato, and J. Prüss, *White noise perturbation of the equations of linear parabolic viscoelasticity,* Rend. Istit. Mat. Univ. Trieste, **29 (1-2)** (1998), 207–220.

[10] F. Comte and E. Renault, *Long memory continuous time models,* J. Econometrics, **73 (1)** (1996), 101–149.

[11] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, Cambridge, 1992.

[12] L. Decreusefond and A. S. Üstünel, *Fractional Brownian motion: theory and applications,* In Matignon and Montseny [21], 75–86.

[13] L. Decreusefond and A. S. Üstünel, *Stochastic analysis of the fractional Brownian motion,* Potential Anal., **10 (2)** (1999), 177–214.

[14] M. El-Borai, *Some probability densities and fundamental solutions of fractional evolution equations,* Chaos Solitons Fractals, **14 (3)** (2002), 433–440.

[15] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. III. McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955.

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, Boston, 1994.

[17] G. Gripenberg, Ph. Clément, and S. O. Londen, *Smoothness in fractional evolution equations and conservation laws,* Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), **29 (1)** (2000), 231–251.

[18] K. W. Homan, *An Analytic Semigroup Approach to Convolution Volterra Equations*, Dissertation, Delft University, 2003.

[19] A. Karczewska, *Stochastic Volterra convolution with Lévy process*, Int. J. Pure Appl. Math., **18 (1)** (2005), 109–120.

[20] B. Mandelbrot and J. Van Ness, *Fractional Brownian motions, fractional noises and applications,* SIAM Rev., **10** (1968), 422–437.

[21] D. Matignon and G. Montseny, editors. *Systèmes Différentiels Fractionnaires*, volume 5 of ESAIM Proceedings, Paris, 1998. Société de Mathématiques Appliquées et Industrielles.

[22] I. Norros, E. Valkeila, and J. Virtamo, *An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions,* Bernoulli, **5 (4)** (1999), 571–587.

[23] V. Pipiras and M.S. Taqqu, *Integration questions related to fractional Brownian motion,* Probab. Theory Related Fields, **118 (2)** (2000), 251–291.

[24] J. Prüss, *Evolutionary Integral Equations and Applications.* Birkhäuser Verlag, Basel, 1993. Monographs in Mathematics, 87.

[25] S. Samko, A. Kilbas, and O. Marichev, *Fractional Integrals and Derivatives.* Theory and applications. Gordon and Breach Science Publishers, Yverdon, 1993.

[26] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems.* Springer-Verlag, New York, 1985.

Stefano Bonaccorsi
Dipartimento di Matematica
Università di Trento
via Sommarive 14
38050 Povo (Trento)
Italia
e-mail: `stefano.bonaccorsi@unitn.it`

Progress in Probability, Vol. 59, 57–74

Dirichlet Forms Methods: An Application to the Propagation of the Error Due to the Euler Scheme

Nicolas Bouleau

Abstract. We present recent advances on Dirichlet forms methods either to extend financial models beyond the usual stochastic calculus or to study stochastic models with less classical tools. In this spirit, we interpret the asymptotic error on the solution of an sde due to the Euler scheme (Kurtz and Protter [39]) in terms of a Dirichlet form on the Wiener space, what allows to propagate this error thanks to functional calculus.

Mathematics Subject Classification (2000). 65G99, 65C30, 60H07.

Keywords. Squared field operator, Wiener space, density, Dirichlet process, stochastic differential equation, Dirichlet form, error.

Introduction

Considering a Dirichlet form amounts to consider a strongly continuous symmetric contraction semi-group on an L^2-space which possesses in addition the property of being positive on positive functions (cf. [13, 28, 44]). This is a particular case of Markovian potential theory, with several special features due to the use of Hilbertian techniques and to the fact that positivity and contraction properties extend to infinite dimensional framework thanks to Fatou's lemma in measure theory. Many Dirichlet structures are constructively obtained on the Wiener space and on the fundamental spaces of probability theory (Poisson space, Monte Carlo space) which may be thought as hypotheses in order to study error propagation through stochastic models (cf. [8]).

Since the discovery by M. Fukushima, at the end of the 1970s, that Dirichlet forms allow to extend the stochastic calculus to processes which are not semi-martingales (cf. [27]) a lot of works have been developed in this direction, even beyond the Dirichlet forms framework. To this extend we quote the approach to time-dependent Dirichlet forms developed by Oshima [50] and the more recent approach of Stannat [58] and Trutnau [59] about a new theory of generalized Dirichlet forms. As in finance the heart of the complete market property and more generally of the portfolio management is the stochastic integral, a particular interest has been devoted to methods giving rise to new stochastic integrals.

We shall give, at first, a short outlook on recent results related to Dirichlet forms and connected with financial motivations. We include some Malliavin calculus approaches when they amount to the use of the Ornstein-Uhlenbeck structure on the Wiener space. After recalling, in a second part, the main properties of Dirichlet forms and the interpretation of the functional calculus on the squared field operator in terms of error propagation, we focuse, in a third part, on the question of the asymptotic error due to the resolution of a stochastic differential equation by the Euler scheme. We show that the asymptotic error may be represented by a Dirichlet structure on the Wiener space and we apply this to propagate the error on the example of a level volatility model for pricing and hedging procedures. We put the general question of the validity of such a propagation as an *asymptotic calculus principle*, and we give partial arguments for this principle.

1. Some recent works

First must be mentionned the idea of using Malliavin's integration by parts technique to speed up the computation of the Greeks or other quantities in finance. After the collective papers of Fournié et al. [25, 26], improvements have been brought to complex options [29] and to the more general question of the sensitivity to some parameters with the aim of calibration of a model. As integration by parts formulae are available in more general Dirichlet forms situations than the Ornstein-Uhlenbeck structure on the Wiener space (cf. [8, Chapter V]), the same approach may be performed, for instance, on the Poisson space for studying models with jumps [17].

One of the first successes of Malliavin calculus was about proving existence of densities for solutions of sde's with smooth coefficients and Dirichlet forms methods have been able to extend such results to the case of Lipschitz coefficients [13]. Several authors remarked that these tools give also means of improving the computation of densities and establishing estimates for the laws of random variables with some regularity assumption. Let us quote [5, 11, 14, 38] whose results are not limited to applications in finance. With suitable hypotheses it is possible, to get explicit closed formulae for the density even with some liberty in the choice of a weight function allowing an optimization for Monte Carlo simulation.

After the classical works of M. Fukushima and Y. Le Jan on stochastic calculus for additive functionals of symmetric Markov processes associated with a Dirichlet form [42] the role of past and future σ-algebras have been clarified by Lyons and Zheng (cf. [43, 59]) and the main current of research, in order to leave the semi-martingale context, starts with the abstract definition of a Dirichlet process as sum of a local martingale and a process with zero quadratic variation (see [22]).

Because the quadratic variation, as formal Dirichlet form, does not possess the closedness property, the Dirichlet form framework is replaced here by functional analytic arguments. The integral is generally defined by a discretization procedure (cf. [6, 23, 24]) or by a regularization procedure (see [52, 53]). These ways have been deepened with the center example of the fractional Brownian motion (cf. [1, 19, 20, 30, 31, 54, 62]). The connections of these works with finance are many: attempting to generalize Girsanov's theorem in order to define martingale measures by erasing more general drifts and using generalized stochastic integration (forward, symmetric and backward integrals) in order to deal with exotic models (cf. [21]). About "inside trading" and the use of forward integrals it is worth to quote [41].

At last, let us mention some uses of Dirichlet forms or Malliavin calculus to deal with processes with jumps by equipping the general Poisson space with a differential structure (cf. [15, 47]) and the forthcoming book of P. Malliavin and A. Thalmaier [46] whose last chapter is devoted to calculus of variations for markets with jumps, the other ones being strongly related with the above topics.

2. Dirichlet forms theory seen as error propagation theory

Let us begin with a very simple but crucial remark about the magnitude of errors. If we consider an erroneous quantity with a centered small error and apply to it a nonlinear map, we observe by an easy Taylor expansion argument that

- the error is no more centered in general: a bias appears
- the variance transmits with a first-order calculus.

Now if we go on, applying anew several non-linear applications,

- the variances and the biases keep (except special cases) the same order of magnitude
- the biases follow a second-order differential calculus involving the variances.

With natural notation,

$$\sigma_{n+1}^2 = f_{n+1}'^2(x_n)\sigma_n^2,$$

$$\mathrm{bias}_{n+1} = f_{n+1}'(x_n)\mathrm{bias}_n + \frac{1}{2}f_{n+1}''(x_n)\sigma_n^2.$$

The first relation has been discovered, even in several dimensions, with correlation between the errors, by Gauss at the beginning of the nineteenth century.

From this observation, in order to represent the propagation of small errors we may consider that

1) the variances of errors have to be managed by a quadratic first-order differential operator Γ,
2) the biases of errors have to be represented by a linear second-order differential operator A,

the propagation of errors being the result of the following change of variable formulae:

$$\Gamma[F(X_1,\ldots,X_m),G(Y_1,\ldots,Y_n)]=\sum_{ij}F_i'(X_1,\ldots,X_m)G_j'(Y_1,\ldots,Y_n)\Gamma[X_i,Y_j]$$

$$A[F(X_1,\ldots,X_m)]=\sum_i F_i'(X_1,\ldots,X_m)A[X_i]+\frac{1}{2}\sum_{ij}F_{ij}''(X_1,\ldots,X_m)\Gamma[X_i,X_j].$$

Because of these propagation rules for the variances and the biases, little errors may be thought as *second-order* vectors. This old notion of differential geometry has been revived at the beginning of the 1980s by the study of semi-martingales on manifolds (cf. [18, 48, 56]).

Now, instead of germs of semi-martingales and second-order vectors, we will use Dirichlet forms, carré du champ and generator. There are two important reasons for this, that I shall give just after recalling some definitions and examples.

Definition 2.1. An error structure is a term

$$S=(\Omega,\mathcal{A},\mathbb{P},\mathbb{D},\Gamma)$$

where $(\Omega,\mathcal{A},\mathbb{P})$ is a probability space, and:

(1) $\mathbb{D}$ is a dense subvector space of $L^2(\Omega,\mathcal{A},\mathbb{P})$ (also denoted $L^2(\mathbb{P})$).

(2) Γ is a positive symmetric bilinear application from $\mathbb{D}\times\mathbb{D}$ into $L^1(\mathbb{P})$ satisfying "the functional calculus of class $\mathcal{C}^1\cap\mathrm{Lip}$". This expression means

$$\forall u\in\mathbb{D}^m,\quad \forall v\in\mathbb{D}^n,\quad \forall F\colon\mathbb{R}^m\to\mathbb{R},\quad \forall G\colon\mathbb{R}^n\to\mathbb{R}$$

with F, G being of class $\mathcal{C}^1$ and Lipschitzian, we have $F(u)\in\mathbb{D}$, $G(v)\in\mathbb{D}$ and

$$\Gamma[F(u),G(v)]=\sum_{i,j}\frac{\partial F}{\partial x_i}(u)\frac{\partial G}{\partial x_j}(v)\Gamma\big[u_i,v_j\big]\quad \mathbb{P}\text{-a.s.}$$

(3) The bilinear form $\mathcal{E}[u,v]=\frac{1}{2}\mathbb{E}\big[\Gamma[u,v]\big]$ is "closed". This means that the space $\mathbb{D}$ equipped with the norm

$$\|u\|_{\mathbb{D}}=\Big(\|u\|^2_{L^2(\mathbb{P})}+\mathcal{E}[u,u]\Big)^{1/2}$$

is complete.

If, in addition

(4) the constant function 1 belongs to $\mathbb{D}$ (which implies $\Gamma[1]=0$ by property 2), we say that the error structure is Markovian.

We will always write $\mathcal{E}[u]$ for $\mathcal{E}[u,u]$ and $\Gamma[u]$ for $\Gamma[u,u]$.

With this definition, the form $\mathcal{E}$ is known in the literature as a *local Dirichlet form* on $L^2(\Omega,\mathcal{A},\mathbb{P})$ that possesses a "squared field" operator (or a "carré du champ" operator) Γ. These notions are usually studied on σ-finite measurable spaces. We limit ourselves herein to probability spaces both for the sake of simplicity and because we will use images and products of error structures.

Under very weak additional assumptions, to an error structure (also to a Dirichlet form on a σ-finite measurable space) a strongly-continuous contraction semigroup $\big(P_t\big)_{t\geq 0}$ on $L^2(\mathbb{P})$ can be uniquely associated, which is symmetric with respect to $\mathbb{P}$ and sub-Markov. This semigroup has a generator $(A, \mathcal{D}A)$, a self-adjoint operator that satisfies

$$A\big[F(u)\big] = \sum_i \frac{\partial F}{\partial x_i}(u)A\big[u_i\big] + \frac{1}{2}\sum_{i,j}\frac{\partial^2 F}{\partial x_i \partial x_j}(u)\Gamma\big[u_i, u_j\big] \quad \mathbb{P}\text{-a.s.}$$

for $F\colon \mathbb{R}^m \to \mathbb{R}$ of class $\mathcal{C}^2$ with bounded derivatives and $u \in (\mathcal{D}A)^m$ such that $\Gamma\big[u_i\big] \in L^2(\mathbb{P})$.

Example 2.2 (Ornstein-Uhlenbeck structure in dimension 1). $\Omega = \mathbb{R}$, $\mathcal{A}$ = Borel σ-field $\mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mathcal{N}(0,1)$ reduced normal law, $\mathbb{D} = H^1\big(\mathcal{N}(0,1)\big) = \big\{u \in L^2(\mathbb{P})$, u' in the distribution sense belongs to $L^2(\mathbb{P})\big\}$, $\Gamma[u] = u'^2$, then

$$\big(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(0,1), H^1(\mathcal{N}(0,1)), \Gamma\big)$$

is an error structure with generator

$$\mathcal{D}A = \big\{f \in L^2(\mathbb{P})\colon f'' - xf' \text{ in the distribution sense} \in L^2(\mathbb{P})\big\},$$

$$Af = \frac{1}{2}f'' - \frac{1}{2}I \cdot f'$$

where I is the identity map on $\mathbb{R}$.

Example 2.3 (Monte Carlo structure in dimension 1). $\Omega = [0,1]$, $\mathcal{A}$ = Borel σ-field, $\mathbb{P}$ = Lebesgue measure, $\mathbb{D} = \big\{u \in L^2\big([0,1], dx\big)$; the derivative u' in the distribution sense over $]0,1[$ belongs to $L^2([0,1], dx)\big\}$, $\Gamma[u] = u'^2$.

Example 2.4 (Friedrich's extension of a symmetric operator). Let D be a connected open set in $\mathbb{R}^d$ with unit volume. Let $\mathbb{P} = dx$ be the Lebesgue measure on D. Let Γ be defined on $\mathcal{C}_k^\infty(D)$ via

$$\Gamma[u,v] = \sum_{ij}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j}a_{ij}, \qquad u, v \in \mathcal{C}_k^\infty(D),$$

where the functions a_{ij} satisfy

$$\begin{aligned}
&a_{ij} \in L^2_{loc}(D) \quad \frac{\partial a_{ij}}{\partial x_k} \in L^2_{loc}(D) \quad i,j,k = 1, \ldots, d,\\
&\sum_{ij} a_{ij}(x)\xi_i\xi_j \geq 0 \quad \forall \xi \in D,\\
&a_{ij}(x) = a_{ji}(x) \quad \forall x \in D,
\end{aligned}$$

then the pre-structure $(D, \mathcal{B}(D), \mathbb{P}, \mathcal{C}_k^\infty(D), \Gamma)$ is closable.

Let us now come back to the question of using Dirichlet forms instead of second-order vectors as germs of semi-martingales.

The first reason is the closedness property. That gives all the power to this theory. It is similar to σ-additivity in probability theory. Without the closedness property, we have an apparently more general framework (as additive set functions are more general than σ-additive ones), but it becomes impossible to say anything on objects which are defined by limits, error propagation is limited to explicit closed formulae. Instead, this closedness property allows to extend error calculus to infinite dimensional frameworks and to propagate errors through typically limit objects as stochastic integrals. As David Hilbert argued against intuitionists, more theorems is better. The philosopher Carl Popper made this mistake about axiomatization of probability theory emphasing that his system (without σ-additivity) was more general than that of Kolmogorov (with σ-additivity).

What is particularly satisfying is that this closedness property is preserved by products. Any countable product of error structures is an error structure and the theorem on products (cf. [8]) gives explicitely the domain of the new Γ operator. Starting with the Ornstein-Uhlenbeck structure in dimension 1, the infinite product of this structure by itself gives the Ornstein-Uhlenbeck structure on the Wiener space. Less surprisingly, the image of an error structure, defined in the most natural way, is still an error structure, as an image of a probability space by a measurable map is still a probability space.

The second reason is related to simplicity. Let us come back to the first remark at the beginning of this part. We said that starting with a centered error, centeredness is lost after a non linear map. But what is preserved by image? Which property is an invariant? It is the global property of symmetry with respect to a measure. If the operators describing the error are symmetric with respect to some measure, the image of the error has still this symmetry with respect to the image measure. Centeredness is nothing but symmetry with respect to Lebesgue measure (not a probability measure, a σ-finite measure but this doesn't matter really here).

The gradient and the sharp (#)

In addition to the operators Γ and A we will need the notion of *gradient* which is a linear (Hilbert-valued) version of the standard deviation of the error.

Definition 2.5. Let $\mathcal{H}$ be a Hilbert space. A linear operator D from $\mathbb{D}$ into $L^2(\mathbb{P},\mathcal{H})$ is said to be a gradient (for S) if

$$\forall u \in \mathbb{D} \quad \Gamma[u] = < Du, Du >_{\mathcal{H}} .$$

A gradient always exists as soon the space $\mathbb{D}$ is separable. It satisfies necessarily the chain rule:

Proposition 2.6. *Let D be a gradient for S with values in $\mathcal{H}$. Then $\forall u \in \mathbb{D}^n, \forall F \in \mathcal{C}^1 \cap Lip(\mathbb{R}^n)$,*

$$D[F \circ u] = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i} \circ u D[u_i] \quad a.e.$$

What we denote by the sharp # is a special case of the gradient operator when $\mathcal{H}$ is chosen to be $L^2(\hat{\Omega}, \hat{\mathcal{A}}, \hat{\mathbb{P}})$ where $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{\mathbb{P}})$ is a copy of $(\Omega, \mathcal{A}, \mathbb{P})$. It is particularly usefull for structures on the Wiener space because stochastic calculus and Itô formula are available both on $(\Omega, \mathcal{A}, \mathbb{P})$ and $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{\mathbb{P}})$.

Let us give some definitions and notation we will need later on about the weighted Ornstein-Uhlenbeck structure on the Wiener space: let B be a standard Brownian motion constructed as coordinates of the space $\mathcal{C}([0,1])$ equipped with the Wiener measure and let α be a positive function in $L^1_{loc}[0,1]$; there exists an error structure (cf. [8]) satisfying

$$\Gamma\left[\int_0^1 u(s)dB_s\right] = \int_0^1 \alpha(s)u^2(s)ds$$

for $u \in \mathcal{C}([0,1])$. It is the mathematical expression of the following perturbation of the Brownian path:

$$\omega(s) = \int_0^s dB_u \mapsto \int_0^s e^{-\frac{\alpha(u)}{2}\varepsilon} dB_u + \int_0^s \sqrt{1-e^{-\alpha(u)\varepsilon}}\, d\hat{B}_u,$$

where $\hat{B}$ is an independent standard Brownian motion. This structure possesses the following #-operator:

$$\left(\int_0^1 u(s)dB_s\right)^{\#} = \int_0^1 \sqrt{\alpha(s)}u(s)d\hat{B}_s, \quad \forall u \in L^2([0,1],(1+\alpha)dt),$$

which satisfies for regular adapted processes H

$$\left(\int_0^1 H_s dB_s\right)^{\#} = \int_0^1 \sqrt{\alpha(s)} H_s d\hat{B}_s + \int_0^1 H_s^{\#} dB_s.$$

Let us end this part by a comment on the passage from a random walk to the Brownian motion in the context of erroneous quantities. Donsker's theorem says that if U_n are i.i.d. square integrable centered random variables, the linear interpolation of the random walk $\sum_{k=1}^n U_k$, i.e., the process

$$X_n(t) = \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{[nt]} U_k + (nt-[nt])U_{[nt]+1}\right)$$

for $t \in [0,1]$, where $[x]$ denotes the entire part of x, converges in law on the space $\mathcal{C}([0,1])$ equipped by the uniform norm to a Brownian motion. Invariance principles follow giving a way to approximate properties of the Brownian motion by the corresponding ones of the random walk. A quite natural question is how this may be extended to the case where the U_n's are erroneous. To extend weak convergence of probability measures we use convergence of Dirichlet forms on Lipschitz and $\mathcal{C}^1$ functions. Then supposing the errors on the U_n's are equidistributed and uncorrelated, the error structure of the process X_n converges to the Ornstein-Uhlenbeck structure on the Wiener space (cf. [10]). Invariance principles follow giving approximations of the variance of the error of Brownian functionals, for example for the sup-norm of the paths:

$$\mathbb{E}\Gamma[\|X_n(t)\|_\infty] = \mathbb{E}\Gamma\left[\frac{1}{\sqrt{n}}\max_{1\le k\le n}|S_k|\right] \to \mathbb{E}\left[\int_0^1 (D_s[\|.\|_\infty])^2 ds\right] = \mathbb{E}[\mathcal{T}]$$

where D denotes the Ornstein-Uhlenbeck gradient with values in $L^2([0,1])$ and $\mathcal{T}$ is the random time where the absolute value of the Brownian path reaches its maximum.

3. Propagation of the error due to the Euler scheme

If an asset X is represented by the solution of an sde, prices of options, hedging portfolios and other financial quantities are obtained by stochastic calculus as functionals of X. If we suppose the sde is solved using the Euler scheme, the asymptotic error on X discovered by Kurtz and Protter in the spirit of a functional central limit theorem takes the form of a process solution to an other sde. In order to propagate this assymptotic error through stochastic calculus, we have to take the derivative in a suitable sense of non-differentiable functionals as stochastic integrals. This may be performed by the theory of Dirichlet forms. Let us recall the situation.

The error due to the Euler scheme

In 1991 Thomas Kurtz and Philipp Protter obtained an asymptotic estimate in law for the error due to the Euler scheme.

In the simplest case, considering the sde

$$X_t = x_0 + \int_0^t a(X_s)dB_s + \int_0^t b(X_s)ds,$$

if X_t^n is the Euler approximation of X_t and $U^n = X^n - X$, then $(B, \sqrt{n}U^n)$ converges in law to (B, U) where U is solution to the linear sde

$$dU_t = a'(X_t)U_t dB_t + b'(X_t)U_t dt + \frac{1}{\sqrt{2}}a'(X_t)a(X_t)dW_t, \qquad U_0 = 0,$$

where W is a Brownian motion independent of B.

Such an "extra-Brownian motion" appeared in a work of H. Rootzen [51] who studies limits of integrals of the form $\int_0^t \psi_n(s)dB_s$ where ψ_n is an adapted process. In the case where $\int_0^t f(B_s, s)dB_s$ is computed by the Euler scheme

$$\int_0^t \psi_n(s)dB_s = \sum_{i=0}^{[nt]} f(B_{\frac{i}{n}}, i/n)(B_{\frac{i+1}{n}} - B_{\frac{i}{n}}) + f(B_{\frac{[nt]}{n}}, [nt]/n)(B_t - B_{\frac{[nt]}{n}}),$$

he obtains for regular f,

$$\sqrt{n}\left(\int_0^\cdot \psi_n dB - \int_0^\cdot f(B_s, s)dB_s\right) \overset{d}{\Rightarrow} \frac{1}{\sqrt{2}}\int_0^\cdot f'_x(B_s, s)dW_s.$$

This kind of result is restricted to adapted approximations. As Wong and Zakai have shown (1965) other natural approximations of the Brownian motion give rise to stochastic integrals in the sense of Stratonowitch.

The discovery of the asymptotic error due to the Euler scheme has been followed by a series of works which extend it to the case of an sde with respect to a continuous or discontinuous semi-martingale and which obtain some statements as necessary and sufficient conditions ([34, 35]).

In addition, asymptotic expansions have been recently obtained by the stochastic calculus of variation [45].

In the sequel, we shall consider the result of Kurtz-Protter in dimension 1 under the following form:

Let X_t be the solution starting at x_0 to the sde

$$dX_t = a(X_t, t)dB_t + b(X_t, t)dt,$$

let X_t^n be the approximate solution obtained by the Euler method, which may be written

$$X_0^n = x_0, \qquad dX_t^n = a(X^n_{\frac{[nt]}{n}}, [nt]/n)dB_t + b(X^n_{\frac{[nt]}{n}}, [nt]/n)dt,$$

and let $U_t^n = X_t^n - X_t$ be the approximation error, then if a and b are $\mathcal{C}^1$ with linear growth

$$(B, \sqrt{n}U^n) \overset{d}{\Rightarrow} (B, U) \qquad \text{on} \quad \mathcal{C}([0,1]),$$

where the process U may be represented as

$$U_0 = 0 \qquad dU_t = a'_x(X_t, t)U_t dB_t + b'_x(X_t, t)U_t dt + \frac{1}{\sqrt{2}} a'_x(X_t, t)a(X_t, t)dW_t,$$

which is solved by the usual method of variation of the constant: introducing the process

$$M_t = \exp\left\{\int_0^t a'_x(X_s, s)dB_s - \frac{1}{2}\int_0^t a'^2_x(X_s, s)ds + \int_0^t b'_x(X_s, s)ds\right\}$$

gives

$$U_t = M_t \int_0^t \frac{a(X_s, s)a'_x(X_s, s)}{\sqrt{2}M_s} dW_s.$$

Let us consider the weighted Ornstein-Uhlenbeck error structure on the Wiener space with weight α as explain above. If the coefficients a and b are regular, then $X_t \in \mathbb{D}$ and $X_t^{\#}$ satisfies

$$X_t^{\#} = \int_0^t a'_x(X_s, s)X_s^{\#} dB_s + \int_0^t a(X_s, s)\sqrt{\alpha(s)}d\widehat{B_s} + \int_0^t b'_x(X_s, s)X_s^{\#} ds. \qquad (\star)$$

Comparing with the equation of the asymptotic error due to the Euler scheme,

$$U_t = \int_0^t a'_x(X_s, s)U_s dB_s + \int_0^t a(X_s, s)\frac{a'_x(X_s, s)}{\sqrt{2}} dW_s + \int_0^t b'_x(X_s, s)U_s ds, \qquad (\star\star)$$

shows that

- if we could take a random and adapted weight $\alpha(t) = \frac{1}{2}a_x'^2(X_t, t)$,
- if the obtained structure is closable with carré du champ and if the calculus of the #-operator is still ($\star$),

then $X^\#$ would be the asymptotic error due to the Euler scheme, and we would be able to propagate this error through the stochastic computations obtaining the variance of the error on any r. v. $Y \in \mathbb{D}$ by the equation $\Gamma[Y] = \widehat{\mathbb{E}}[Y^{\#2}]$.

The Ornstein-Uhlenbeck structure with random weight

From now on α is a measurable random process defined on the Wiener space, non-negative, non necessarily adapted. We assume that this process satisfies $\mathbb{E}\int_0^1 \alpha_t dt < +\infty$, and $\alpha(\omega, t) \geq k(t) > 0 \quad \mathbb{P} \times dt$-a.e. where k is deterministic.

Let us denote by $\mathbb{D}_{ou}^k$ the domain of the Ornstein-Uhlenbeck structure with deterministic weight k and by D_{ou}^k its gradient. On the domain

$$\mathbb{D} = \left\{ Y \in \mathbb{D}_{ou}^k : \quad \int_0^1 \mathbb{E}[(D_{ou}^k[Y](t))^2 \frac{\alpha(t)}{k(t)}]dt < +\infty \right\}$$

which is dense, the form

$$\mathcal{E}[Y] = \frac{1}{2}\int_0^1 \mathbb{E}[(D_{ou}^k[Y](t))^2 \frac{\alpha(t)}{k(t)}]dt$$

is Dirichlet and admits

$$\Gamma[Y] = \int_0^1 (D_{ou}^k[Y](t))^2 \frac{\alpha(t)}{k(t)}dt$$

as carré du champ operator.

Indeed, let $\mathcal{V}$ be the space of linear combinations of exponentials of the form $Y = \exp\{i\int_0^1 h_u dB_u\}$ with h deterministic bounded, by $\int_0^1 \mathbb{E}\alpha(t)dt < +\infty$, we have $\mathcal{V} \subset \mathbb{D}$ and $D_{ou}^k[Y] = Y(ih\sqrt{k})$, hence $\mathbb{D}$ is dense.

Let X_n be a Cauchy sequence in L^2 and for $\mathcal{E}$. Let X be the limit of X_n in L^2. Then X_n is Cauchy for $\mathcal{E}_{ou}^k$ which is closed, hence $X \in \mathbb{D}_{ou}^k$ and there exists a sub-sequence $X_{n'}$ such that

$$D_{ou}^k[X_{n'}] \to D_{ou}^k[X] \qquad \mathbb{E} \times dt\text{-p.s.},$$

and by Fatou's lemma

$$\begin{aligned} &\int_0^1 \mathbb{E}\left[(D_{ou}^k[X])^2 \frac{\alpha(t)}{k(t)}\right] dt \\ &= \int_0^1 \mathbb{E}\left[\lim(D_{ou}^k[X_{n'}])^2 \frac{\alpha(t)}{k(t)}\right] dt \leq \liminf \int_0^1 \mathbb{E}\left[(D_{ou}^k[X_{n'}])^2 \frac{\alpha(t)}{k(t)}\right] dt < +\infty \end{aligned}$$

since X_n is Cauchy for $\mathcal{E}$. Hence $X \in \mathbb{D}$. Now again by the Fatou lemma we show as classically that X_n converges to X in $\mathbb{D}$.

Contractions operate on $(\mathcal{E}, \mathbb{D})$ by the functional calculus for D_{ou}^k, hence $(\mathcal{E}, \mathbb{D})$ is a Dirichlet form. The definition of the carré du champ operator ([13, Def. 4.1.2]) is satisfied.

The generator $(A, \mathcal{D}A)$ is given by

$$\mathcal{D}A = \{F \in \mathbb{D}, \exists G \in L^2, \forall H \in \mathbb{D}, \frac{1}{2}\mathbb{E}\int_0^1 D^k_{ou}[F]D^k_{ou}[H]\frac{\alpha(t)}{k(t)}dt = - < G, H >\},$$
$$AF = G,$$

hence if $F \in \mathcal{D}A$, then $\frac{\alpha(t)}{k(t)}D^k_{ou}[F] \in \operatorname{dom}\delta^k_{ou}$ and

$$AF = -\frac{1}{2}\delta^k_{ou}\left[\frac{\alpha}{k}D^k_{ou}F\right],$$

where δ^k_{ou} is the Skorokhod integral with weight k.

Adapted case

Let us now add the hypothesis that α is adapted. If h is in $L^\infty(\mathbb{R}_+)$,

$$\mathbb{E}\Gamma\left[F, \int_0^1 hdB\right] = \mathbb{E}\left[F\int_0^1 h(s)\alpha(s)dB_s\right].$$

If $F, G \in \mathbb{D} \cap L^\infty$,

$$\mathbb{E}[G < DF, h\sqrt{\alpha} >] = -\mathbb{E}[F < DG, h\sqrt{\alpha} >] + \mathbb{E}\left[FG\int h\alpha dB\right].$$

And if v is adapted and in $\operatorname{dom}\delta$,

$$\delta[v] = \int_0^1 v_s\sqrt{\alpha_s}dB_s.$$

At last, for finance, the following properties are important, they use the fact that α is adapted:

$$A[\mathbb{E}[X|\mathcal{F}_s]] = \mathbb{E}[A^s[X]|\mathcal{F}_s]$$

where A^s is constructed as A with the weight $\alpha(t)1_{\{t\le s\}}$,

$$D[\mathbb{E}[X|\mathcal{F}_s]](t) = \mathbb{E}[D[X](t)1_{t\le s}|\mathcal{F}_s],$$

$\mathbb{E}[.|\mathcal{F}_s]$ is an orthogonal projector in $\mathbb{D}$,

$$(\mathbb{E}[X|\mathcal{F}_s])^\# = \mathbb{E}[X^{\#_s}|\mathcal{F}_s],$$

where $\#_s$ is constructed as $\#$ with the weight $\alpha(t)1_{\{t\le s\}}$. If X is $\mathcal{F}_t$-measurable, then AX, $\Gamma[X]$ are $\mathcal{F}_t$-measurables.

Concerning the operator $\#$ we have the formulae

$$\left(\int_0^1 \xi_s\, dB_s\right)^\# = \int_0^1 \xi_s^\#\, dB_s + \int_0^1 \xi_s\sqrt{\alpha_s}d\widehat{B}_s.$$

Hence Formula $(\star)$ is satisfied.

Application to diffusion models

Let us consider the following model of an asset:

$$dX_t = X_t \sigma(X_t, t) dB_t + X_t r(t) dt,$$

and let us put on the Wiener space the Ornstein-Uhlenbeck structure with weight

$$\alpha_t = \frac{a'^2(X_t, t)}{2} = \frac{(\sigma(X_t, t) + X_t \sigma'_x(X_t, t))^2}{2}$$

which represents the asymptotic error due to the Euler scheme. σ is supposed to be strictly positive, $\mathcal{C}^1$and Lipschitz and the preceding hypotheses on α are assumed.

Such a modelling is coherent. The error is attached to the asset X and any functional of X, including the Brownian motion itself, and its error may be computed thanks to the equation

$$dB_t = \frac{dX_t}{X_t \sigma(X_t, t)} - X_t r(t) dt$$

which gives

$$(B_t)^{\#} = \int_0^t \sqrt{\alpha(s)} d\widehat{B}_s \qquad \Gamma[B_t] = \int_0^t \alpha(s) ds.$$

Let us show how financial calculi may be performed before proposing some comments on the use of such an analysis. Putting $M_t = \exp\{\int_0^t \sqrt{\alpha_s} dB_s - \frac{1}{2} \int_0^t \alpha_s ds + \int_0^t r(s) ds\}$ we have

$$\Gamma[X_t] = M_t^2 \int_0^t \frac{X_s^2 \sigma^2(X_s, s)}{M_s^2} \alpha_s ds,$$

$$\Gamma[X_s, X_t] = M_s M_t \int_0^{s \wedge t} \frac{X_u^2 \sigma^2(X_u, u)}{M_u^2} \alpha_u du.$$

The price of a European option with payoff $f(X_T)$ at exercise time T,

$$V_t = \mathbb{E}\left[(\exp - \int_t^T r(s) ds) f(X_T) | \mathcal{F}_t \right],$$

becomes erroneous (in the sense of error structures) with an error obtained thanks to the $\#$:

$$\Gamma[V_t] = \left(\exp -2 \int_t^T r(s) ds) (\mathbb{E}[f'(X_T) M_T | \mathcal{F}_t] \right)^2 \frac{\Gamma[X_t]}{M_t^2},$$

$$\Gamma[V_s, V_t] = \left(\exp(- \int_s^T r(u) du - \int_t^T r(v) dv) \right) \mathbb{E}[f'(X_T) M_T | \mathcal{F}_s]$$
$$\times \mathbb{E}[f'(X_T) M_T | \mathcal{F}_t] \frac{\Gamma[X_s, X_t]}{M_s M_t}.$$

The quantity of asset in the hedging portfolio is

$$H_t = \left(\exp - \int_t^T r(s) ds \right) \mathbb{E}[f'(X_T) M_T | \mathcal{F}_t] \frac{1}{M_t}$$

and we have

$$\Gamma[H_t] = \left(\exp -2\int_t^T r(s)ds\right)\left(\mathbb{E}\left[\frac{M_T}{M_t}(f''(X_T)M_T + f'(X_T)Z_t^T)|\mathcal{F}_t\right]\right)^2 \frac{\Gamma[X_t]}{M_t^2}$$

with

$$Z_t^T = \int_t^T L_s dB_s - \int_t^T \sqrt{\alpha_s} L_s M_s ds,$$
$$L_s = a''_{x^2}(X_s, s) = 2\sigma'_x(X_s, s) + X_s\sigma''_{x^2}(X_s, s).$$

It is still true, as in the case of deterministic weight (cf. [8]), that the proportional error on X_t divided by the volatility:

$$\frac{\sqrt{\Gamma[X_t]}}{X_t} \cdot \frac{1}{\sigma(X_t, t)}$$

is a finite variation process (cf. [4] on the "feed back" effect).

Discussion

Thanks to this construction of an error structure, i.e., a local Dirichlet form with squared field operator, on the Wiener space, hence by image, on $\mathcal{C}([0,1])$ equipped with the law of the process X, we have at our disposal a powerful mean to propagate the error done on X toward sufficiently smooth functionals of X. In order to assess the interest of this tool, the question arises of knowing whether the propagated error is the same as the one we would obtain by a direct computation of the functional thanks to the approximation X^n of X. For instance, in the simplest case, does the convergence in law

$$\sqrt{n}(f(X_t^n) - f(X_t)) \overset{d}{\Rightarrow} f'(X_t)X_t^{\#}$$

hold for $f \in \mathcal{C}^1 \cap Lip$? Can we justify *an asymptotic calculus principle* which says that the Dirichlet form allows effectively to compute the errors on the quantities which are erroneous because of the approximation X^n of X? We will not exhaustively examine this principle here, for it is a too large enterprise. Nevertheless, in the important current of research whose fruitfulness has been confirmed these last twenty years, which may be called the "tightness programm", the authors, among which we must at least quote P.-A. Meyer, W. A. Zheng, J. Jacod, A. N. Shiryaev, A. Jakubowski, J. Mémin, G. Pagès, T. G. Kurtz, P. Protter, L. Słomiński, D. Talay, V. Bally, A. Kohatsu-Higa and many others, have already done a major part of the work by stating their results of convergence in law, of stable convergence, of tightness of processes, under a sufficiently general form for propagating iteratively the properties through stochastic integrals and sde's in the semi-martingale framework.

Let us give some results in the direction of this *asymptotic calculus principle* keeping the hypotheses of the present part III.

Let F be a real function of class $\mathcal{C}^1$ and Lipschitz defined on $\mathcal{C}([0,1])$ equipped with the uniform norm. Such a function satisfies

$$F(x+h) = F(x) + < F'(x), h > + \|h\|\varepsilon_x(h) \qquad \forall x, h \in \mathcal{C}([0,1])$$

where the mapping $x \mapsto F'(x)$ is continuous and bounded with values in the Banach space of Radon measures on $[0,1]$, $\varepsilon_x(h)$ is bounded in x and h, and goes to zero when $h \to 0$ in $\mathcal{C}([0,1])$. Then we have

$$\sqrt{n}(F(X^n) - F(X)) \stackrel{d}{\Rightarrow} (F(X))^{\#} = \int_{[0,1]} X_t^{\#} \, F'(X)(dt).$$

The equality on the right-hand side comes from the functional calculus in error structures (see [10]). Putting $U^n = X^n - X$ as before, the fact that $\sqrt{n}\|U^n\|\varepsilon_X(U^n)$ tends to zero in probability, reduces the proof to the study of the convergence in law of

$$< F'(X), \sqrt{n}U^n >= \sqrt{n} \int (X_t^n - X_t) \, F'(X)(dt)$$

to $\int X_t^{\#} \, F'(X)(dt)$. Considering the measure $F'(X)(dt)$ as the differential of a finite variation process adapted to the constant filtration $\mathcal{G}_t = \mathcal{B}(C(]0,1]))$, the fact that the process to be integrated $\sqrt{n}U^n$ converges stably to $X^{\#}$ implies (cf. [40, Thm. 2.2]) that the stochastic integral $\int \sqrt{n}U^n \, F'(X)(dt)$ converges in law to $\int X_t^{\#} \, F'(X)(dt)$.

We obtain also the convergence in law of the stochastic integrals $H.\sqrt{n}U^n \stackrel{d}{\Rightarrow} H.X^{\#}$ for H deterministic or adapted and that of

$$\sqrt{n}\left(\int_0^1 f(X_s^n, s)dX_s^n - \int_0^1 f(X_s, s)dX_s\right)$$

to

$$\left(\int_0^1 f(X_s, s)dX_s\right)^{\#} = \int_0^1 f'(X_s, s)X_s^{\#}dX_s + \int_0^1 f(X_s, s)dX_s^{\#}$$

for f $\mathcal{C}^1$ and Lipschitz.

More generally, we can make more explicit the research programm of determining *the domain* of the asymptotic calculus.

Let X_n and X be two random variables with values in a measurable set $(E, \mathcal{F})$, and let α_n be a sequence of positive numbers. Let $\mathcal{D}_0$ denote a set of *simple* functions included in $L^2(\mathbb{P}_X)$ and in $L^2(\mathbb{P}_{X_n})$ $\forall n$. Let us suppose that there exists an error structure

$$S = (E, \mathcal{F}, \mathbb{P}_X, \mathbb{D}, \Gamma)$$

such that $\mathcal{D}_0 \subset \mathbb{D}$ and $\forall \varphi \in \mathcal{D}_0$

$$\lim_n \alpha_n \mathbb{E}[(\varphi(X_n) - \varphi(X))^2] = \mathbb{E}[\Gamma[\varphi]]; \tag{3.1}$$

we shall say that the asymptotic calculus principle extends to $\mathcal{D}$ for $\mathcal{D}_0 \subset \mathcal{D} \subset \mathbb{D}$ if the limit (3.1) extends to $\psi \in \mathcal{D}$.

If, as above, a #-operator is available (which occurs as soon as $\mathbb{D}$ is separable), in order to prove (3.1) on $\mathcal{D}$, since # is a closed operator, it suffices for any $\psi \in \mathcal{D}$ to find a sequence $\varphi_p \in \mathcal{D}_0$ such that

1. $\varphi_p \to \psi$ in $L^2(\mathbb{P}_X)$

2. $\varphi_p^{\#}$ converges in $L^2(\mathbb{P}_X \times \widehat{\mathbb{P}_X})$
3. $\alpha_n \mathbb{E}[\psi(X_n) - \psi(X))^2]$ may be approximated uniformly in n by $\alpha_n \mathbb{E}[\varphi_p(X_n) - \varphi_p(X))^2]$.

When $(E, \mathcal{F})$ is a normed vector space, obtaining (3.1) from a convergence in law of $\sqrt{\alpha_n}(\varphi(X_n) - \varphi(X))$ uses generally a uniform integrability of $\alpha_n \|X_n - X\|^2$. We shall go deeper in this problem in a separate work.

Let us end by some remarks from the point of view of finance. The interest of considering a financial asset as erroneous is not evident since it is one of the best known quantities continuously quoted in a financial market. Such an error may be justified (cf. [8]) by the inaccuracy of the instants of transaction, possibly also to represent an infinitesimal bid-ask. But this would rather justify specifically constructed error structures instead of the one induced by the Euler scheme. This error structure is relevant only in order to assess the errors in Monte Carlo simulations performed to calculate financial quantities in a given model.

Several authors ([16, 32]) remarked that the stochastic integral which is the active hedge of a future contingent claim, in a model where the underlying asset is a semi-martingale, is an instance of application of limit theorems on discretization errors. This is different from the Euler scheme error and it would be worth to examine this error from the point of view of an asymptotic Dirichlet form.

A more general and complete study of the bias operators and the Dirichlet form yielded by an approximation, with applications related to the part III of the present paper, is to appear ([12]).

References

[1] E. Alos, O. Mazet, and D. Nualart, *Stochastic calculus with respect to fractional Brownian motion with Hurst parameter lesser than 1/2*, Stochastic Process. Appl., **86** (2000).

[2] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function*, Prob. Th. and Rel. Fields, **2 (2)** (1996), 93–128.

[3] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations: II. Convergence rate of the density*, Monte Carlo Methods and Appl., **104 (1)** (1996), 43–80.

[4] E. Barucci, P. Malliavin, M. E. Mancino, R. Renó, and A. Thalmaier, *The price volatility feedback rate: an implementable indicator of market stability*, Math. Finance, **13** (2003), 17–35.

[5] B. Bouchard, I. Ekeland, and N. Touzi, *On the Malliavin approach to Monte Carlo approximation of conditional expectations*, Finance Stoch., **8** (2004), 45–71.

[6] N. Bouleau, *About stochastic integrals with respect to processes which are not semi-martingales*, Osaka J. Math., **22** (1985), 31–34.

[7] N. Bouleau, *Error calculus and path sensitivity in financial models*, Mathematical Finance, **13 (1)** (2003), 115–134.

[8] N. Bouleau, *Error Calculus for Finance and Physics, the Language of Dirichlet Forms*, De Gruyter, 2003.

[9] N. Bouleau, *Financial Markets and Martingales, Observations on Science and Speculation*, Springer, 2004.

[10] N. Bouleau, *Théorème de Donsker et formes de Dirichlet*, Bull. Sci. Math., **129** (2005), 369–380.

[11] N. Bouleau, *Improving Monte Carlo simulations by Dirichlet forms*, C.R. Acad. Sci. Paris, Ser I, 2005.

[12] N. Bouleau, *When and how an error yields a Dirichlet form*, J. of Functional Analysis, **240 (2)** (2006), 445–494.

[13] N. Bouleau and F. Hirsch, *Dirichlet Forms and Analysis on Wiener Space*, De Gruyter, 1991.

[14] M. E. Caballero, B. Fernandez, and D. Nualard, *Estimation of densities and applications*, J. of Theoretical Prob., **11 (3)** (1998).

[15] L. Denis, A. Grorud, and M. Pontier, *Formes de Dirichlet sur un espace de Wiener-Poisson, application au grossissement de filtration*, Sém. Prob. XXXIV, Lect. N. in Math., **1729** (1999), Springer.

[16] D. Duffie and P. Protter, *From discrete to continuous time finance: weak convergence of the financial gain process*, (1989), unpublished, cited in [40].

[17] Y. El Kahtib and N. Privault, *Computation of Greeks in a market with jumps via Malliavin calculus*, Finance and Stoch., **8** (2004), 161–179.

[18] M. Emery, *Stochastic calculus on Manifolds*, Springer, 1989.

[19] M. Errami and F. Russo, *Covariation de convolution de martingales*, C.R. Acad. Sci. Paris, s1, **326** (1998), 601–609.

[20] D. Feyel and A. de La Pradelle, *On fractional Brownian processes*, Potential Anal., **10 (3)** (1999), 273–288.

[21] F. Flandoli, F. Russo, and J. Wolf, *Some SDEs with distributional drift, I. General calculus*, Osaka J. Math., **40 (2)** (2003), 493–542.

[22] H. Föllmer, *Dirichlet processes*, in Stochastic Integrals, Lect. Notes in Math., **851** (1981), 476–478.

[23] H. Föllmer, *Calcul d'Ito sans probabilité*, in Sém. Prob. XV, Lect. N. in Math., Springer, **850** (1981), 143–150.

[24] H. Föllmer, P. Protter, and A. N. Shiryaev, *Quadratic covariation and an extension of Ito's formula*, Bernoulli, **1** (1995), 149–169.

[25] E. Fournié, J. M. Lasry, J. Lebuchoux, P. L. Lions, and N. Touzi, *Applications of Malliavin calculus to Monte Carlo methods in finance*, Finance and Stoch., **3** (1999), 391–412.

[26] E. Fournié, J. M. Lasry, J. Lebuchoux, and P. L. Lions, *Applications of Malliavin calculus to Monte Carlo methods in finance II*, Finance and Stoch., (2001), 201–236.

[27] M. Fukushima, *Dirichlet Forms and Markov Processes*, North-Holland, 1980.

[28] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet forms and symmetric Markov processes*, De Gruyter, 1994.

[29] E. Gobet and A. Kohatsu-Higa, *Computation of Greeks for barrier and lookback options using Malliavin calculus*, Electron. Com. in Prob., **8** (2003), 51–62.

[30] M. Gradinaru and I. Nourdin, *Approximation at first and second order of the m-variation of the fractional Brownian motion*, Electron. Com. Prob., **8** (2003), 1–26.

[31] M. Gradinaru, F. Russo, and P. Vallois, *Generalized covariation, local time and Stratonowich-Ito's formula for fractional Brownian motion with Hurst index* $\geq 1/4$, Annals of Prob., **31 (4)** (2003), 1772–1820.

[32] T. Hayashi and P. A. Mykland, *Evaluating hedging errors: an asymptotic approach*, Math. Finance, **15 (2)** (2005), 309–343.

[33] J. Jacod, *Théorèmes limites pour les processus*, Lect. Notes Math., Springer, **1117** (1985).

[34] J. Jacod, A. Jakubowski, and J. Mémin, *About asymptotic errors in discretization processes*, Ann. of Prob., **31** (2003), 592–608.

[35] J. Jacod and P. Protter, *Asymptotic error distributions for the Euler method for stochastic differential equations*, Ann. Probab., **26** (1998), 267–307.

[36] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer, 1987.

[37] A. Jakubowski, J. Mémin, and G. Pagès, *Convergence en loi des suites d'intégrales stochastiques sur l'espace de Skorokhod*, Probab. Th. Rel. Fields, **81** (1989), 111–137.

[38] A. Kohatsu-Higa and R. Pettersson, *Variance reduction methods for simulation of densities on Wiener space*, SIAM J. Numer. Anal., **40 (2)** (2002), 431–450.

[39] T. Kurtz and Ph. Protter, *Wong-Zakai corrections, random evolutions and simulation schemes for SDEs*, in: Stochastic Analysis, Acad. Press, (1991), 331-346.

[40] T. Kurtz and Ph. Protter, *Weak limit theorems for stochastic integrals and stochastic differential equations*, Ann. Probab., **19** (1991), 1035–1070.

[41] J. A. Leon, R. Navarro, and D. Nualart, *An anticipating calculus approach to the utility maximization of an insider*, Math. Finance, **13 (1)** (2003), 171–185.

[42] Y. Le Jan, *Mesures associées à une forme de Dirichlet, applications*, Bull. Soc. Math. France, **106** (1978), 61–112.

[43] T. Lyons and W. Zheng, *A crossing estimate for the canonical process on a Dirichlet space and tightness result*, in: Colloque Paul Lévy, Astérisque, **157-158** (1998), 249–271.

[44] Z. M. Ma and M. Röckner, *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*, Springer, 1992.

[45] P. Malliavin and A. Thalmaier, *Numerical error for SDE: Asymptotic expansion and hyperdistributions*, C.R. Acad. Sci. Paris, ser. I, **336** (2003), 851–856.

[46] P. Malliavin and A. Thalmaier, *Stochastic Calculus of Variations in Mathematical Finance*, Springer, 2006.

[47] M. Mensi and N. Privault, *Conditional calculus and enlargement of filtration on Poisson space*, Stoch. Anal. and Appl., **21** (2003), 183–204.

[48] P.-A. Meyer, *Géométrie différentielle stochastique*, in: Sém. Prob. XVI suppl., Lect. N. in Math., Springer, **921** (1982), 165–207.

[49] N. Nualart, *The Malliavin Calculus and Related Topics*, Springer, 1995.

[50] Y. Oshima, *On a construction of Markov processes associated with time dependent Dirichlet spaces*, Forum Math., **4** (1992), 395–415.

[51] H. Rootzén, *Limit distribution for the error in approximation of stochastic integrals*, Ann. Probab., **8** (1980), 241–251.

[52] F. Russo and P. Vallois, *The generalized covariation process and Itô formula*, Stochastic Proc. Appl., **59** (1995), 81–104.

[53] F. Russo and P. Vallois, *Itô formula for C^1-functions of semi-martingales*, Prob. Th. Rel. Fields, **104** (1996), 27–41.

[54] F. Russo and P. Vallois, *Stochastic calculus with respect to a finite quadratic variation process*, Stochastics and Stoch. Rep., **70** (2000), 1–40.

[55] F. Russo, P. Vallois, and J. Wolf, *A generalized class of Lyons-Zheng processes*, Bernoulli, **7 (2)** (2001), 363–379.

[56] L. Schwartz, *Géométrie différentielle du 2ème ordre, semi-martingales et équations différentielles stochastiques sur une variété différentielle*, in: Sém. Prob. XVI suppl., Lect. N. in Math., Springer, **921** (1982), 1–150.

[57] L. Słomiński, *Stability of strong solutions of stochastic differential equations*, Stochastic Process. Appl., **31** (1989), 173–202.

[58] W. Stannat, *The theory of generalized Dirichlet forms and its applications in analysis and stochastics*, Mem. Amer. Math. Soc., **142 (678)** (1999).

[59] G. Trutnau, *Stochastic calculus of generalized Dirichlet forms and applications to stochastic differential equations in infinite dimensions*, Osaka J. Math., **37 (2)** (2000), 315–343.

[60] J. Wolf, *An Ito formula for Dirichlet processes*, Stochastics and Stoch. Rep., **62 (2)** (1997), 103–115.

[61] E. Wong and M. Zakai, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist., **36** (1965), 1560–1564.

[62] M. Zahle, *Integration with respect to fractal functions and stochastic calculus*, Prob. Th. Rel. Fields, **21** (1998), 333–374.

Nicolas Bouleau
Ecole Nationale des Ponts et des Chaussées
6 et 8 avenue Blaise Pascal
Cité Descartes, Champs sur Marne
77455 Marne-la-Vallée Cedex 2
France
e-mail: `bouleau@enpc.fr`

Progress in Probability, Vol. 59, 75–113

Individual-Based Probabilistic Models of Adaptive Evolution and Various Scaling Approximations

Nicolas Champagnat, Régis Ferrière and Sylvie Méléard

Abstract. We are interested in modelling Darwinian evolution, resulting from the interplay of phenotypic variation and natural selection through ecological interactions. Our models are rooted in the microscopic, stochastic description of a population of discrete individuals characterized by one or several adaptive traits. The population is modelled as a stochastic point process whose generator captures the probabilistic dynamics over continuous time of birth, mutation, and death, as influenced by each individual's trait values, and interactions between individuals. An offspring usually inherits the trait values of her progenitor, except when a mutation causes the offspring to take an instantaneous mutation step at birth to new trait values. We look for tractable large population approximations. By combining various scalings on population size, birth and death rates, mutation rate, mutation step, or time, a single microscopic model is shown to lead to contrasting macroscopic limits, of different nature: deterministic, in the form of ordinary, integro-, or partial differential equations, or probabilistic, like stochastic partial differential equations or superprocesses. In the limit of rare mutations, we show that a possible approximation is a jump process, justifying rigorously the so-called trait substitution sequence. We thus unify different points of view concerning mutation-selection evolutionary models.

Keywords. Darwinian evolution, birth-death-mutation-competition point process, mutation-selection dynamics, nonlinear integro-differential equations, nonlinear partial differential equations, nonlinear superprocesses, fitness, adaptive dynamics.

1. Introduction

In this paper, we are interested in modelling the dynamics of populations as driven by the interplay of phenotypic variation and natural selection operating through ecological interactions, i.e., Darwinian evolution. The fundamental property of evolving systems is the propensity of each individual to create and to select the diversity. This feature requires to focus on the stochastic dynamics of each individual in the population. The study of such evolutionary-ecological models is very complicated, and several approximations have been proposed. Firstly, Bolker and Pacala [2] and Dieckmann and Law [11] have introduced the moment equations of the distribution of traits in the population and studied different moment closure heuristics. Secondly, various nonlinear macroscopic models (integro-differential equations, partial differential equations, superprocesses) have been proposed without microscopic justification. Finally, the emerging field of adaptive dynamics has proposed a new class of macroscopic models on the evolutionary time scale, defined as jump processes and ordinary differential equations (trait substitution sequences, Metz et al. [22], canonical equation of adaptive dynamics, Dieckmann and Law [10]). In all these cases and from a biological point of view, the pathway from microscopic to macroscopic models deserves a firm mathematical pavement, at least to clarify the significance of the implicit biological assumptions underlying the choice of a particular model.

In this work, we unify several macroscopic approximations by recovering them from a single microscopic model. In particular, we point out the importance of large population assumptions and that the nature of the approximation strongly depends on the combination of various scalings of the biological parameters (birth and death rates, mutation rate, mutation step and time).

This paper starts (Section 2) with the microscopic description of a population of discrete individuals, whose phenotypes are described by a vector of trait values. The population is modelled as a stochastic Markov point process whose generator captures the probabilistic dynamics over continuous time of birth, mutation and death, as influenced by each individual's trait values and interactions between individuals. The adaptive nature of a trait implies that an offspring usually inherits the trait values of her progenitor, except when a mutation occurs. In this case, the offspring makes an instantaneous mutation step at birth to new trait values. We will refer to the state space parameterized by adaptive traits as the trait space, and will often (slightly abusively) call trait the actual trait value. This process is defined as the solution of a stochastic differential equation driven by point Poisson measures (Section 2.1). In Section 2.2, we give an algorithmic construction of the population point process and propose some simulations, for various parameters, of an asymmetrical example developed in Kisdi [18]. Next, we prove that the point population process is a measure-valued semimartingale and compute its characteristics (Section 2.3). Then we look for tractable approximations, following different mathematical paths. Our first approach (Section 3) aims at deriving deterministic equations to describe the moments of trajectories of the point process, i.e., the statistics of a large number of independent realizations of the process. We explain the difficult hierarchy between these equations coming from competition kernels and preventing, even in the simple mean-field case, decorrelations and tractable moment closure. The alternative approach involves renormalizations of the point process based on a large population limit. The measure-valued martingale properties of the renormalized point process allow us to show that, according to different scalings of birth, death and mutation rates, one obtains qualitatively different limiting partial differential equations and the appearance or not of some demographic stochasticity. We show in Section 4.1 that by itself, the large-population limit leads to a deterministic, nonlinear integro-differential equation. Then, in Section 4.2.1, we combine the large-population limit with an acceleration of birth (hence mutation) and death according to small mutation steps. That yields either a deterministic nonlinear reaction-diffusion model, or a stochastic measure-valued process (depending on the acceleration rate of the birth-and-death process). If now this acceleration of birth and death is combined with a limit of rare mutations, the large-population limit yields a nonlinear integro-differential equation either deterministic or stochastic, depending here again on the speed of the scaling of the birth-and-death process, as described in Section 4.2.2.

In Section 5, we model a time scale separation between ecological events (fast births and deaths) and evolution (rare mutations), for an initially monomorphic population. The competition between individuals takes place on the short time scale. In a large population limit, this leads on the mutation time scale to a jump process over the trait space, where the population stays monomorphic at any time. Thereby we provide a rigorous justification to the notion of trait substitution sequence introduced by Metz et al. [21].

2. Population point process

Even if the evolution manifests itself as a global change in the state of a population, its basic mechanisms, mutation and selection, operate at the level of individuals. Consequently, we model the evolving population as a stochastic interacting individual system, where each individual is characterized by a vector of phenotypic trait values. The trait space $\mathcal{X}$ is assumed to be a closed subset of $\mathbb{R}^d$, for some $d \geq 1$.

We will denote by $M_F(\mathcal{X})$ the set of finite non-negative measures on $\mathcal{X}$. Let also $\mathcal{M}$ be the subset of $M_F(\mathcal{X})$ consisting of all finite point measures:

$$\mathcal{M} = \left\{ \sum_{i=1}^{n} \delta_{x_i},\ n \geq 0, x_1, \dots, x_n \in \mathcal{X} \right\}.$$

Here and below, δ_x denotes the Dirac mass at x. For any $m \in M_F(\mathcal{X})$, any measurable function f on $\mathcal{X}$, we set $\langle m, f \rangle = \int_{\mathcal{X}} f dm$.

We aim to study the stochastic process ν_t, taking its values in $\mathcal{M}$, and describing the distribution of individuals and traits at time t. We define

$$\nu_t = \sum_{i=1}^{I(t)} \delta_{X_t^i},$$

$I(t) \in \mathbb{N}$ standing for the number of individuals alive at time t, and $X_t^1, \ldots, X_t^{I(t)}$ describing the individual's traits (in $\mathcal{X}$).

For a population $\nu = \sum_{i=1}^{I} \delta_{x^i}$, and a trait $x \in \mathcal{X}$, we define the birth rate $b(x, V * \nu(x)) = b(x, \sum_{i=1}^{I} V(x - x^i))$ and the death rate $d(x, U * \nu(x)) = d(x, \sum_{i=1}^{I} U(x - x^i))$ of individuals with trait x; V and U denote the interaction kernels affecting, respectively, reproduction and mortality. Let $\mu(x)$ and $M(x, z)dz$ be, respectively, the probability that an offspring produced by an individual with trait x carries a mutated trait and the law of this mutant trait.

Thus, the population evolution can be roughly summarized as follows. The initial population is characterized by a (possibly random) counting measure $\nu_0 \in \mathcal{M}$ at time 0, and any individual with trait x at time t has two independent random exponentially distributed "clocks": a birth clock with parameter $b(x, V * \nu_t(x))$, and a death clock with parameter $d(x, U * \nu_t(x))$. If the death clock of an individual rings, this individual dies and disappears. If the birth clock of an individual with trait x rings, this individual produces an offspring. With probability $1 - \mu(x)$ the offspring carries the same trait x; with probability $\mu(x)$ the trait is mutated. If a mutation occurs, the mutated offspring instantly acquires a new trait z, picked randomly according to the mutation step measure $M(x, z)dz$.

Thus we are looking for a $\mathcal{M}$-valued Markov process $(\nu_t)_{t \geq 0}$ with infinitesimal generator L, defined for real bounded functions ϕ by

$$\begin{aligned} L\phi(\nu) &= \sum_{i=1}^{I} b(x^i, V * \nu(x^i))(1 - \mu(x^i))(\phi(\nu + \delta_{x^i}) - \phi(\nu)) \\ &\quad + \sum_{i=1}^{I} b(x^i, V * \nu(x^i))\mu(x^i) \int_{\mathcal{X}} (\phi(\nu + \delta_z) - \phi(\nu)) M(x^i, z)dz \\ &\quad + \sum_{i=1}^{I} d(x^i, U * \nu(x^i))(\phi(\nu - \delta_{x^i}) - \phi(\nu)). \end{aligned} \tag{2.1}$$

The first term of (2.1) captures the effect on the population of birth without mutation; the second term that of birth with mutation, and the last term that of death. The density-dependence makes all terms nonlinear.

2.1. Process construction

Let us justify the existence of a Markov process admitting L as infinitesimal generator. The explicit construction of $(\nu_t)_{t \geq 0}$ also yields three side benefits: providing a rigorous and efficient algorithm for numerical simulations (given hereafter), laying the mathematical basis to derive the moment equations of the process (Section 3), and establishing a general method that will be used to derive some large population limits (Sections 4 and 5).

We make the biologically natural assumption that the trait dependency of birth parameters is "bounded", and at most linear for the death rate. Specifically, we assume

Assumptions (H):

There exist constants $\bar{b}$, $\bar{d}$, $\bar{U}$, $\bar{V}$ and C and a probability density function $\bar{M}$ on $\mathbb{R}^d$ such that for each $\nu = \sum_{i=1}^{I} \delta_{x^i}$ and for $x, z \in \mathcal{X}$,

$$\begin{gathered} b(x, V * \nu(x)) \leq \bar{b}, \quad d(x, U * \nu(x)) \leq \bar{d}(1+I), \\ U(x) \leq \bar{U}, \quad V(x) \leq \bar{V}, \\ M(x,z) \leq C\bar{M}(z-x). \end{gathered}$$

These assumptions ensure that there exists a constant $\bar{C}$, such that the total event rate, for a population counting measure $\nu = \sum_{i=1}^{I} \delta_{x^i}$, obtained as the sum of all event rates, is bounded by $\bar{C}I(1+I)$.

Let us now give a pathwise description of the population process $(\nu_t)_{t\geq 0}$. We introduce the following notation.

Notation 2.1. Let $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$. Let $H = (H^1, \ldots, H^k, \ldots) : \mathcal{M} \mapsto (\mathbb{R}^d)^{\mathbb{N}^*}$ be defined by $H\left(\sum_{i=1}^{n} \delta_{x_i}\right) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)}, 0, \ldots, 0, \ldots)$, where $x_{\sigma(1)} \preccurlyeq \cdots \preccurlyeq x_{\sigma(n)}$, for some arbitrary order $\preccurlyeq$ on $\mathbb{R}^d$ (for example, the lexicographic order).

This function H allows us to overcome the following (purely notational) problem. Choosing a trait uniformly among all traits in a population $\nu \in \mathcal{M}$ consists in choosing i uniformly in $\{1, \ldots, \langle \nu, 1\rangle\}$, and then in choosing the individual *number* i (from the arbitrary order point of view). The trait value of such an individual is thus $H^i(\nu)$.

We now introduce the probabilistic objects we will need.

Definition 2.2. Let $(\Omega, \mathcal{F}, P)$ be a (sufficiently large) probability space. On this space, we consider the following four independent random elements:

(i) an $\mathcal{M}$-valued random variable ν_0 (the initial distribution),

(ii) independent Poisson point measures $M_1(ds, di, d\theta)$, and $M_3(ds, di, d\theta)$ on $[0, \infty) \times \mathbb{N}^* \times \mathbb{R}^+$, with the same intensity measure $ds\left(\sum_{k\geq 1} \delta_k(di)\right) d\theta$ (the "clonal" birth and the death Poisson measures),

(iii) a Poisson point measure $M_2(ds, di, dz, d\theta)$ on $[0, \infty) \times \mathbb{N}^* \times \mathcal{X} \times \mathbb{R}^+$, with intensity measure $ds\left(\sum_{k\geq 1} \delta_k(di)\right) dz d\theta$ (the mutation Poisson measure).

Let us denote by $(\mathcal{F}_t)_{t\geq 0}$ the canonical filtration generated by these processes.

We finally define the population process in terms of these stochastic objects.

Definition 2.3. Assume (H). A $(\mathcal{F}_t)_{t\geq 0}$-adapted stochastic process $\nu = (\nu_t)_{t\geq 0}$ is called a population process if a.s., for all $t \geq 0$,

$$\begin{aligned}
\nu_t = \nu_0 &+ \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} \delta_{H^i(\nu_{s-})}\mathbf{1}_{\{i\leq\langle\nu_{s-},1\rangle\}} \\
&\qquad \mathbf{1}_{\{\theta\leq b(H^i(\nu_{s-}),V*\nu_{s-}(H^i(\nu_{s-})))(1-\mu(H^i(\nu_{s-})))\}}M_1(ds,di,d\theta) \\
&+ \int_{[0,t]\times\mathbb{N}^*\times\mathcal{X}\times\mathbb{R}^+} \delta_z\mathbf{1}_{\{i\leq\langle\nu_{s-},1\rangle\}} \\
&\qquad \mathbf{1}_{\{\theta\leq b(H^i(\nu_{s-}),V*\nu_{s-}(H^i(\nu_{s-})))\mu(H^i(\nu_{s-}))M(H^i(\nu_{s-}),z)\}}M_2(ds,di,dz,d\theta) \\
&- \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} \delta_{H^i(\nu_{s-})}\mathbf{1}_{\{i\leq\langle\nu_{s-},1\rangle\}}\mathbf{1}_{\{\theta\leq d(H^i(\nu_{s-}),U*\nu_{s-}(H^i(\nu_{s-})))\}}M_3(ds,di,d\theta).
\end{aligned} \tag{2.2}$$

Let us now show that if ν solves (2.2), then ν follows the Markovian dynamics we are interested in.

Proposition 2.4. *Assume (H) and consider a solution $(\nu_t)_{t\geq 0}$ of Eq. (2.2) such that $E(\sup_{t\geq T}\langle\nu_t,\mathbf{1}\rangle^2) < +\infty$, $\forall T > 0$. Then $(\nu_t)_{t\geq 0}$ is a Markov process. Its infinitesimal generator L is defined for all bounded and measurable maps $\phi : \mathcal{M} \mapsto \mathbb{R}$, all $\nu \in \mathcal{M}$, by (2.1). In particular, the law of $(\nu_t)_{t\geq 0}$ does not depend on the chosen order $\preccurlyeq$.*

Proof. The fact that $(\nu_t)_{t\geq 0}$ is a Markov process is classical. Let us now consider a function ϕ as in the statement. With our notation, $\nu_0 = \sum_{i=1}^{\langle\nu_0,1\rangle}\delta_{H^i(\nu_0)}$. A simple computation, using the fact that a.s.,

$$\phi(\nu_t) = \phi(\nu_0) + \sum_{s\leq t}(\phi(\nu_{s-} + (\nu_s - \nu_{s-})) - \phi(\nu_{s-})),$$

shows that

$$\begin{aligned}
\phi(\nu_t) &= \phi(\nu_0) \\
&+ \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} \left(\phi(\nu_{s-} + \delta_{H^i(\nu_{s-})}) - \phi(\nu_{s-})\right)\mathbf{1}_{\{i\leq\langle\nu_{s-},1\rangle\}} \\
&\qquad \mathbf{1}_{\{\theta\leq b(H^i(\nu_{s-}),V*\nu_{s-}(H^i(\nu_{s-})))(1-\mu(H^i(\nu_{s-})))\}}M_1(ds,di,d\theta) \\
&+ \int_{[0,t]\times\mathbb{N}^*\times\mathcal{X}\times\mathbb{R}^+} (\phi(\nu_{s-} + \delta_z) - \phi(\nu_{s-}))\mathbf{1}_{\{i\leq\langle\nu_{s-},1\rangle\}} \\
&\qquad \mathbf{1}_{\{\theta\leq b(H^i(\nu_{s-}),V*\nu_{s-}(H^i(\nu_{s-})))\mu(H^i(\nu_{s-}))M(H^i(\nu_{s-}),z)\}}M_2(ds,di,dz,d\theta) \\
&+ \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} \left(\phi(\nu_{s-} - \delta_{H^i(\nu_{s-})}) - \phi(\nu_{s-})\right)\mathbf{1}_{\{i\leq\langle\nu_{s-},1\rangle\}} \\
&\qquad \mathbf{1}_{\{\theta\leq d(H^i(\nu_{s-}),U*\nu_{s-}(H^i(\nu_{s-})))\}}M_3(ds,di,d\theta).
\end{aligned}$$

Taking expectations, we obtain

$$\begin{aligned}
E(\phi(\nu_t)) &= E(\phi(\nu_0)) \\
&+ \int_0^t E\Big(\sum_{i=1}^{\langle \nu_s,1\rangle} \Big\{ \big(\phi(\nu_s + \delta_{H^i(\nu_s)}) - \phi(\nu_s)\big) \\
&\qquad\qquad b(H^i(\nu_s), V * \nu_s(H^i(\nu_s)))(1-\mu(H^i(\nu_s))) \\
&+ \int_{\mathcal{X}} (\phi(\nu_s+\delta_z) - \phi(\nu_s))\, b(H^i(\nu_s), V*\nu_s(H^i(\nu_s)))\mu(H^i(\nu_s)) M(H^i(\nu_s), z) dz \\
&+ \big(\phi(\nu_s - \delta_{H^i(\nu_s)}) - \phi(\nu_s)\big)\, d(H^i(\nu_s), U * \nu_s(H^i(\nu_s)))\Big\}\Big) ds
\end{aligned}$$

Differentiating this expression at $t=0$ leads to (2.1). □

Let us show existence and moment properties for the population process.

Theorem 2.5. (i) *Assume (H) and that $E(\langle \nu_0, 1\rangle) < \infty$. Then the process $(\nu_t)_{t\geq 0}$ defined by Definition 2.3 is well defined on $\mathbb{R}_+$.*
(ii) *If furthermore for some $p \geq 1$, $E(\langle \nu_0, 1\rangle^p) < \infty$, then for any $T < \infty$,*

$$E\left(\sup_{t\in[0,T]} \langle \nu_t, 1\rangle^p \right) < \infty. \tag{2.3}$$

Proof. We first prove (ii). Consider the process $(\nu_t)_{t\geq 0}$. We introduce for each n the stopping time $\tau_n = \inf\{t \geq 0,\ \langle \nu_t, 1\rangle \geq n\}$. Then a simple computation using Assumption (H) shows that, neglecting the non-positive death terms,

$$\begin{aligned}
&\sup_{s\in[0,t\wedge\tau_n]} \langle \nu_s, 1\rangle^p \\
&\leq \langle \nu_0, 1\rangle^p + \int_{[0,t\wedge\tau_n]\times\mathbb{N}^*\times\mathbb{R}^+} ((\langle \nu_{s-},1\rangle + 1)^p - \langle \nu_{s-},1\rangle^p)\, \mathbf{1}_{\{i\leq \langle \nu_{s-},1\rangle\}} \\
&\qquad\qquad \mathbf{1}_{\{\theta \leq b(H^i(\nu_{s-}), V*\nu_{s-}(H^i(\nu_{s-})))(1-\mu(H^i(\nu_{s-})))\}} M_1(ds, di, d\theta) \\
&\quad + \int_{[0,t]\times\mathbb{N}^*\times\mathcal{X}\times\mathbb{R}^+} ((\langle \nu_{s-},1\rangle + 1)^p - \langle \nu_{s-},1\rangle^p)\, \mathbf{1}_{\{i\leq \langle \nu_{s-},1\rangle\}} \\
&\qquad\qquad \mathbf{1}_{\{\theta \leq b(H^i(\nu_{s-}), V*\nu_{s-}(H^i(\nu_{s-})))\mu(H^i(\nu_{s-}))M(H^i(\nu_{s-}),z)\}} M_2(ds, di, dz, d\theta).
\end{aligned}$$

Using the inequality $(1+x)^p - x^p \leq C_p(1+x^{p-1})$ and taking expectations, we thus obtain, the value of C_p changing from line to line,

$$\begin{aligned}
E\left(\sup_{s\in[0,t\wedge\tau_n]} \langle \nu_s, 1\rangle^p \right) &\leq C_p \left(1 + E\left(\int_0^{t\wedge\tau_n} \bar{b}\,(\langle \nu_{s-},1\rangle + \langle \nu_{s-},1\rangle^p)\, ds \right)\right) \\
&\leq C_p \left(1 + E\left(\int_0^t (1 + \langle \nu_{s\wedge\tau_n},1\rangle^p)\, ds \right)\right).
\end{aligned}$$

The Gronwall lemma allows us to conclude that for any $T < \infty$, there exists a constant $C_{p,T}$, not depending on n, such that

$$E\left(\sup_{t\in[0,T\wedge\tau_n]} \langle \nu_t, 1\rangle^p\right) \le C_{p,T}. \tag{2.4}$$

First, we deduce that τ_n tends a.s. to infinity. Indeed, if not, one may find a $T_0 < \infty$ such that $\epsilon_{T_0} = P\left(\sup_n \tau_n < T_0\right) > 0$. This would imply that

$$E\left(\sup_{t\in[0,T_0\wedge\tau_n]} \langle \nu_t, 1\rangle^p\right) \ge \epsilon_{T_0} n^p$$

for all n, which contradicts (2.4). We may let n go to infinity in (2.4) thanks to the Fatou lemma. This leads to (2.3).

Point (i) is a consequence of point (ii). Indeed, one builds the solution $(\nu_t)_{t\ge 0}$ step by step. One only has to check that the sequence of jump instants T_n goes a.s. to infinity as n tends to infinity. But this follows from (2.3) with $p = 1$. □

2.2. Examples and simulations

Let us remark that Assumption (H) is satisfied in the case where

$$b(x, V * \nu(x)) = b(x), \quad d(x, U * \nu(x)) = d(x) + \alpha(x) \int_{\mathcal{X}} U(x-y)\nu(dy),$$

where b, d and α are bounded functions.

In the case where moreover, $\mu \equiv 1$, this individual-based model can also be interpreted as a model of "spatially structured population", where the trait is viewed as a spatial location and the mutation at each birth event is viewed as dispersal. This kind of models have been introduced by Bolker and Pacala ([2, 3]) and Law et al. ([19]), and mathematically studied by Fournier and Méléard [15]. The case $U \equiv 1$ corresponds to a density-dependence in the total population size.

We will consider later the particular set of parameters for the logistic interaction model, taken from Kisdi [18] and corresponding to a model of asymmetrical competition:

$$\begin{gathered} \bar{\mathcal{X}} = [0,4], \quad d(x) = 0, \quad \alpha(x) = 1, \quad \mu(x) = \mu, \\ b(x) = 4 - x, \quad U(x-y) = \frac{2}{K}\left(1 - \frac{1}{1 + 1,2\exp(-4(x-y))}\right) \end{gathered} \tag{2.5}$$

and $M(x,z)dz$ is a Gaussian law with mean x and variance σ^2 conditioned to the fact that the mutant stays in $[0,4]$. As we will see in Section 4, the constant K scaling the strength of competition also scales the population size (when the initial population size is proportional to K). In this model, the trait x can be interpreted as body size. Equation (2.5) means that body size influences the birth rate negatively, and creates asymmetrical competition reflected in the sigmoid shape of U (being larger is competitively advantageous).

Let us give an algorithmic construction for the population process (in the general case), simulating the size $I(t)$ of the population, and the trait vector $\mathbf{X}_t$ of all individuals alive at time t.

At time $t = 0$, the initial population ν_0 contains $I(0)$ individuals and the corresponding trait vector is $\mathbf{X}_0 = (X_0^i)_{1 \leq i \leq I(0)}$. We introduce the following sequences of independent random variables, which will drive the algorithm.

- The type of birth or death events will be selected according to the values of a sequence of random variables $(W_k)_{k \in \mathbb{N}^*}$ with uniform law on $[0, 1]$.
- The times at which events may be realized will be described using a sequence of random variables $(\tau_k)_{k \in \mathbb{N}}$ with exponential law with parameter $\bar{C}$.
- The mutation steps will be driven by a sequence of random variables $(Z_k)_{k \in \mathbb{N}}$ with law $\bar{M}(z)dz$.

We set $T_0 = 0$ and construct the process inductively for $k \geq 1$ as follows.

At step $k-1$, the number of individuals is I_{k-1}, and the trait vector of these individuals is $\mathbf{X}_{T_{k-1}}$.

Let $T_k = T_{k-1} + \dfrac{\tau_k}{I_{k-1}(I_{k-1}+1)}$. Notice that $\dfrac{\tau_k}{I_{k-1}(I_{k-1}+1)}$ represents the time between jumps for I_{k-1} individuals, and $\bar{C}(I_{k-1}+1)$ gives an upper bound on the total event rate for each individual.

At time T_k, one chooses an individual $i_k = i$ uniformly at random among the I_{k-1} alive in the time interval $[T_{k-1}, T_k)$; its trait is $X^i_{T_{k-1}}$. (If $I_{k-1} = 0$, then $\nu_t = 0$ for all $t \geq T_{k-1}$.)

- If $0 \leq W_k \leq \dfrac{d(X^i_{T_{k-1}}, \sum_{j=1}^{I_{k-1}} U(X^i_{T_{k-1}} - X^j_{T_{k-1}}))}{\bar{C}(I_{k-1}+1)} = W_1^i(\mathbf{X}_{T_{k-1}})$, the chosen individual dies, and $I_k = I_{k-1} - 1$.
- If $W_1^i(\mathbf{X}_{T_{k-1}}) < W_k \leq W_2^i(\mathbf{X}_{T_{k-1}})$, where

$$W_2^i(\mathbf{X}_{T_{k-1}}) = W_1^i(\mathbf{X}_{T_{k-1}}) + \frac{[1-\mu(X^i_{T_{k-1}})]b(X^i_{T_{k-1}}, \sum_{j=1}^{I_{k-1}} V(X^i_{T_{k-1}} - X^j_{T_{k-1}}))}{\bar{C}(I_{k-1}+1)},$$

  then the chosen individual gives birth to an offspring with trait $X^i_{T_{k-1}}$, and $I_k = I_{k-1} + 1$.
- If $W_2^i(\mathbf{X}_{T_{k-1}}) < W_k \leq W_3^i(\mathbf{X}_{T_{k-1}}, Z_k)$, where

$$\begin{aligned} W_3^i(\mathbf{X}_{T_{k-1}}, Z_k) &= W_2^i(\mathbf{X}_{T_{k-1}}) \\ &+ \frac{\mu(X^i_{T_{k-1}})b(X^i_{T_{k-1}}, \sum_{j=1}^{I_{k-1}} V(X^i_{T_{k-1}} - X^j_{T_{k-1}}))M(X^i_{T_{k-1}}, X^i_{T_{k-1}} + Z_k)}{\bar{C}\bar{M}(Z_k)(I_{k-1}+1)}, \end{aligned}$$

  then the chosen individual gives birth to a mutant offspring with trait $X^i_{T_{k-1}} + Z_k$, and $I_k = I_{k-1} + 1$.
- If $W_k > W_3^i(\mathbf{X}_{T_{k-1}}, Z_k)$, nothing happens, and $I_k = I_{k-1}$.

Then, at any time $t \geq 0$, the number of individuals is defined by $I(t) = \sum_{k\geq 0} 1_{\{T_k \leq t < T_{k+1}\}} I_k$ and the population process is obtained as $\nu_t = \sum_{k\geq 0} 1_{\{T_k \leq t < T_{k+1}\}} \sum_{i=1}^{I_k} \delta_{X^i_{T_k}}$.

The simulation of Kisdi's example (2.5) can be carried out following this algorithm. We can show a very wide variety of qualitative behavior according to the value of the parameters σ, μ and K.

In the following figures, the upper part gives the distribution of the traits in the population at any time, using a grey scale code for the number of individuals holding a given trait. The lower part of the simulation represents the dynamics of the total size $I(t)$ of the population.

These simulations will serve to illustrate the different mathematical scalings described in Sections 4 and 5. Let us observe for the moment the qualitative differences between the cases where K is large (Fig. 1 (c)), in which a wide population density evolves regularly (see Section 4.1) and where μ is small (Fig. 1 (d)), in which the population trait evolves according to a jump process (see Section 5.1).

The simulations of Fig. 2 involve an acceleration of the birth and death processes (see Section 4.2) as

$$b(x,\zeta) = K^\eta + b(x) \quad \text{and} \quad d(x,\zeta) = K^\eta + d(x) + \alpha(x)\zeta.$$

There is a noticeable qualitative difference between Fig. 2 (a) and (b), where $\eta = 1/2$, and Fig. 2 (c) and (d), where $\eta = 1$. In the latter, we observe strong fluctuations in the population size and a finely branched structure of the evolutionnary pattern, revealing a new form of stochasticity in the large population approximation.

More discussions about these simulations are given in [7], especially about the branching pattern of some of them.

2.3. Martingale properties

We finally give some martingale properties of the process $(\nu_t)_{t\geq 0}$, which are the key point of our approach.

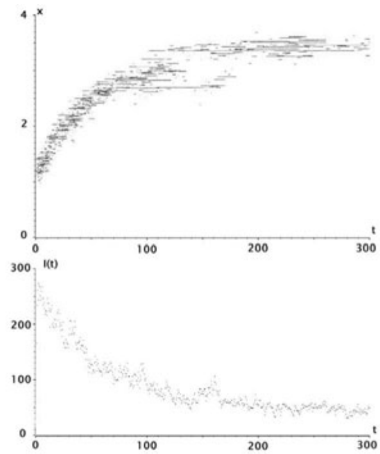
Theorem 2.6. *Assume* (H)*, and that for some* $p \geq 2$*,* $E\left(\langle \nu_0, 1\rangle^p\right) < \infty$*.*

(i) *For all measurable functions* ϕ *from* $\mathcal{M}$ *into* $\mathbb{R}$ *such that for some constant* C*, for all* $\nu \in \mathcal{M}$*,* $|\phi(\nu)| + |L\phi(\nu)| \leq C(1 + \langle \nu, 1\rangle^p)$*, the process*

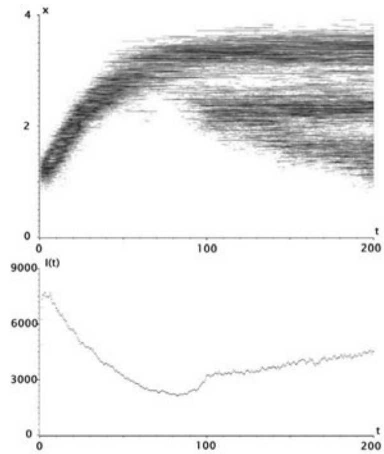
$$\phi(\nu_t) - \phi(\nu_0) - \int_0^t L\phi(\nu_s)ds$$

is a càdlàg $(\mathcal{F}_t)_{t\geq 0}$*-martingale starting from* 0*.*

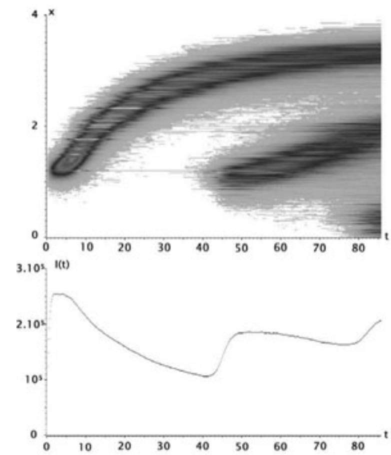
(ii) *Point* (i) *applies to any function* $\phi(\nu) = \langle \nu, f\rangle^q$*, with* $0 \leq q \leq p-1$ *and with* f *bounded and measurable on* $\mathcal{X}$*.*



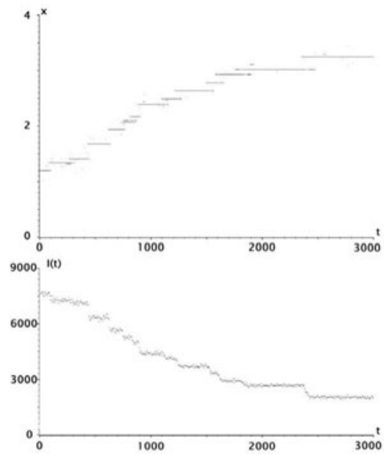
(a) $\mu = 0.03$, $K = 100$, $\sigma = 0.1$.



(b) $\mu = 0.03$, $K = 3000$, $\sigma = 0.1$.

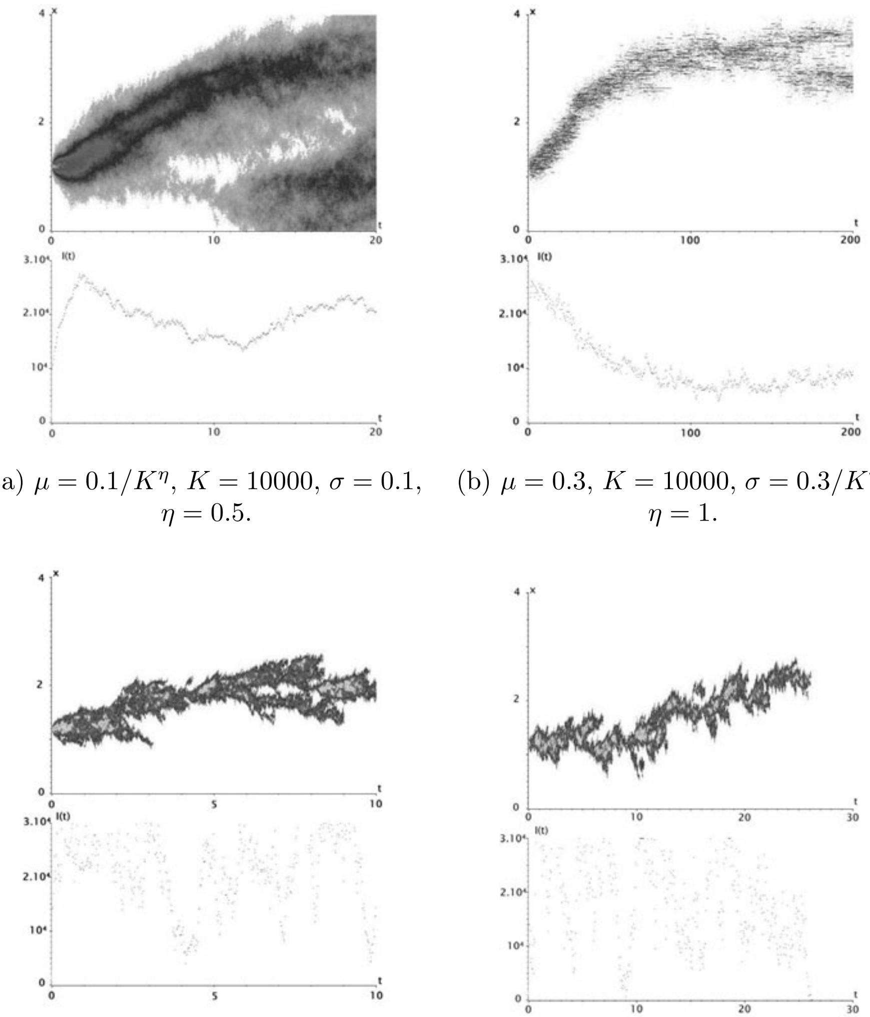


(c) $\mu = 0.03$, $K = 100000$, $\sigma = 0.1$.



(d) $\mu = 0.00001$, $K = 3000$, $\sigma = 0.1$.

Figure 1. Numerical simulations of trait distributions (upper panels, darker is higher frequency) and population size (lower panels). The initial population is monomorphic with trait value 1.2 and contains K individuals. (a–c) Qualitative effect of increasing system size (measured by parameter K). (d) Large system size and very small mutation probability (μ).



(a) $\mu = 0.1/K^{\eta}$, $K = 10000$, $\sigma = 0.1$, $\eta = 0.5$.

(b) $\mu = 0.3$, $K = 10000$, $\sigma = 0.3/K^{\eta/2}$, $\eta = 1$.

(c) $\mu = 0.3$, $K = 10000$, $\sigma = 0.3/K^{\eta/2}$, $\eta = 1$.

(d) $\mu = 0.00001$, $K = 3000$, $\sigma = 0.1$.

FIGURE 2. Numerical simulations of trait distribution (upper panels, darker is higher frequency) and population size (lower panels) for accelerated birth and death and concurrently increased system size. Parameter η (between 0 and 1) relates the acceleration of demographic turnover and the increase of system size. (a) Rescaling mutation step. (b) Rescaling mutation probability. (c–d) Rescaling mutation step in the limit case $\eta = 1$; two samples for the same population. The initial population is monomorphic with trait value 1.2 and contains K individuals.

(iii) *For such a function f, the process*

$$\begin{aligned} M_t^f &= \langle \nu_t, f\rangle - \langle \nu_0, f\rangle \\ &\quad - \int_0^t \int_{\mathcal{X}} \bigg\{ \Big((1-\mu(x)) b(x, V*\nu_s(x)) - d(x, U*\nu_s(x)) \Big) f(x) \\ &\quad + \mu(x) b(x, V*\nu_s(x)) \int_{\mathcal{X}} f(z) M(x,z) dz \bigg\} \nu_s(dx) ds \end{aligned} \tag{2.6}$$

is a càdlàg square integrable martingale starting from 0 *with quadratic variation*

$$\begin{aligned} \langle M^f \rangle_t = \int_0^t \int_{\mathcal{X}} \bigg\{ \Big((1-\mu(x)) b(x, V*\nu_s(x)) - d(x, U*\nu_s(x)) \Big) f^2(x) \\ + \mu(x) b(x, V*\nu_s(x)) \int_{\mathcal{X}} f^2(z) M(x,z) dz \bigg\} \nu_s(dx) ds. \end{aligned} \tag{2.7}$$

Proof. First of all, note that point (i) is immediate thanks to Proposition 2.4 and (2.3). Point (ii) follows from a straightforward computation using (2.1). To prove (iii), we first assume that $E\left(\langle \nu_0, 1\rangle^3\right) < \infty$. We apply (i) with $\phi(\nu) = \langle \nu, f\rangle$. This yields that M^f is a martingale. To compute its bracket, we first apply (i) with $\phi(\nu) = \langle \nu, f\rangle^2$ and obtain that

$$\begin{aligned} \langle \nu_t, f\rangle^2 - \langle \nu_0, f\rangle^2 - \int_0^t \int_{\mathcal{X}} \bigg\{ \Big((1-\mu(x)) b(x, V*\nu_s(x)) (f^2(x) + 2f(x)\langle \nu_s, f\rangle) \\ + d(x, U*\nu_s(x)) (f^2(x) - 2f(x)\langle \nu_s, f\rangle) \Big) \\ + \mu(x) b(x, V*\nu_s(x)) \int_{\mathcal{X}} (f^2(z) + 2f(z)\langle \nu_s, f\rangle) M(x,z) dz \bigg\} \nu_s(dx) ds \end{aligned} \tag{2.8}$$

is a martingale. In another hand, we apply the Itô formula to compute $\langle \nu_t, f\rangle^2$ from (2.6). We deduce that

$$\begin{aligned} &\langle \nu_t, f\rangle^2 - \langle \nu_0, f\rangle^2 \\ &\quad - \int_0^t 2\langle \nu_s, f\rangle \int_{\mathcal{X}} \bigg\{ \Big((1-\mu(x)) b(x, V*\nu_s(x)) - d(x, U*\nu_s(x)) \Big) f(x) \\ &\quad + \mu(x) b(x, V*\nu_s(x)) \int_{\mathcal{X}} f(z) M(x,z) dz \bigg\} \nu_s(dx) ds - \langle M^f \rangle_t \end{aligned} \tag{2.9}$$

is a martingale. Comparing (2.8) and (2.9) leads to (2.7). The extension to the case where only $E\left(\langle \nu_0, 1\rangle^2\right) < \infty$ is straightforward, since even in this case, $E(\langle M^f \rangle_t) < \infty$ thanks to (2.3) with $p = 2$. □

3. Moment equations

Moment equations have been proposed by Bolker and Pacala [2, 3] and Dieckmann and Law [11] as handy analytical models for spatially structured populations.

The philosophy of moment equations is germane to the principle of Monte-Carlo methods: computing the mean path of the point process from a large number of independent realizations. (Another approach, as we shall see in Section 4, is to model the behavior of a single trajectory when it is the initial number of individuals which is made large).

Let us define the deterministic measure $E(\nu)$ associated with a random measure ν by $\displaystyle\int_{\mathcal{X}} \varphi(x)E(\nu)(dx) = E(\int_{\mathcal{X}} \varphi(x)\nu(dx))$. Taking expectations in (2.6), we obtain some formula for $\int_{\mathcal{X}} \varphi(x)E(\nu)(dx)$ involving the expectations of integrals with respect to $\nu(dx)$ or to $\nu(dx)\nu(dy)$. Nevertheless, this equation is very intricate and presents an unresolved hierarchy of nonlinearities. Writing an equation for $E(\nu(dx)\nu(dy))$ could be possible but will involve integrals with respect to $\nu(dx)\nu(dy)\nu(dz)$ and so on. Whether this approach may eventually help describe the population dynamics in the trait space is still unclear.

Let us consider the case of spatially structured population (see Section 2.2) where $d(x,\zeta) = d(x) + \alpha(x)\zeta$, $b(x,\zeta) = b(x)$ and $\mu(x) = 1$. Let $N(t) = E(I(t))$ where $I(t)$ is the number of individuals at time t. Taking expectations on (2.6) with $\varphi \equiv 1$ yields:

$$N(t)=N(0)+\int_0^t E\left(\int_{\mathcal{X}}(b(x)-d(x))\nu_s(dx) - \int_{\mathcal{X}\times\mathcal{X}} \alpha(x)U(x-y)\nu_s(dx)\nu_s(dy)\right)ds. \tag{3.1}$$

In the specific case where b, d and α are independent of (the spatial location) x, (cf. [19]), (3.1) recasts into

$$\dot{N} = (b-d)N - \alpha E\left(\int_{\mathcal{X}\times\mathcal{X}} U(x-y)\nu_t(dx)\nu_t(dy)\right).$$

Even in the specific mean-field case where $U = 1$, we get

$$\dot{N} = (b-d)N - \alpha E\left(\int_{\mathcal{X}\times\mathcal{X}} \nu_t(dx)\nu_t(dy)\right). \tag{3.2}$$

The quadratic term corresponding to spatial correlations can not be simplified and (3.2) allows us to precisely identify the mathematical issues raised by the problem of moment closure. In Section 4.1, we will see that one needs the additional large population hypothesis to decorrelate the quadratic term and to refind the well-known logistic equation.

Nevertheless, even if we are not able to produce a closed equation satisfied by $E(\nu)$, we are able to show, in the general case, the following qualitative important property concerning the absolute continuity of the expectation of ν_t.

Proposition 3.1. *Assume (H), that $E(\langle \nu_0, 1\rangle) < \infty$ and that $E(\nu_0)$ is absolutely continuous with respect to the Lebesgue measure. Then for all $t \geq 0$, $E(\nu_t)$ is absolutely continuous with respect to the Lebesgue measure.*

Remark 3.2. This implies in particular that, when the initial trait distribution $E(\nu_0)$ has no singularity w.r.t. the Lebesgue measure, these singularities, such as Dirac masses, can only appear in the limit of infinite time.

Proof. Consider a Borel set A of $\mathbb{R}^d$ with Lebesgue measure zero. Consider also, for each $n \geq 1$, the stopping time $\tau_n = \inf\{t \geq 0,\ \langle \nu_t, 1\rangle \geq n\}$. A simple computation allows us to obtain, for all $t \geq 0$, all $n \geq 1$,

$$E\left(\langle \nu_{t\wedge\tau_n}, \mathbf{1}_A\rangle\right) \leq E(\langle \nu_0, \mathbf{1}_A\rangle) + \bar{b}\ E\bigg(\int_0^{t\wedge\tau_n} \int_{\mathcal{X}} \mathbf{1}_A(x)\nu_s(dx)ds\bigg)$$
$$+ \bar{b}\ E\bigg(\int_0^{t\wedge\tau_n} \int_{\mathcal{X}} \bigg(\int_{\mathcal{X}} \mathbf{1}_A(z)M(x,z)dz\bigg)\,\nu_s(dx)ds\bigg).$$

By assumption, the first term on the right-hand side is zero. The third term is also zero, since for any $x \in \mathcal{X}$, $\int_{\mathcal{X}} \mathbf{1}_A(z)M(x,z)dz = 0$. By Gronwall's lemma, we conclude that for each n, $E(\langle \nu_{t\wedge\tau_n}, \mathbf{1}_A\rangle)$ is zero. Thanks to (2.3) with $p = 1$, τ_n a.s. grows to infinity with n, which concludes the proof. □

4. Large-population renormalizations of the individual-based process

The moment equation approach outlined above is based on the idea of averaging a large number of independent realizations of the population process initiated with a finite number of individuals. If K scales the initial number of individuals, the alternative approach consists in studying the exact process by letting that system size become very large and making some appropriate renormalizations. Several types of approximations can then be derived, depending on the renormalization of the process.

For any K, let the set of parameters U_K, V_K, b_K, d_K, M_K, μ_K satisfy the Assumption (H). Let ν_t^K be the counting measure of the population at time t. We define the measure-valued Markov process $(X_t^K)_{t\geq 0}$ by

$$X_t^K = \frac{1}{K}\nu_t^K.$$

As the system size K goes to infinity, we need to assume the

Assumption (H1):

The parameters U_K, V_K, b_K, d_K, M_K and μ_K are all continuous, $\zeta \mapsto b(x,\zeta)$ and $\zeta \mapsto d(x,\zeta)$ are Lipschitz for any $x \in \mathcal{X}$, and

$$U_K(x) = U(x)/K, \quad V_K(x) = V(x)/K.$$

A biological interpretation of this renormalization is that larger systems are made up of smaller individuals, which may be a consequence of a fixed amount of available resources to be partitioned among individuals. Thus, the biomass of each interacting individual scales as $1/K$, which may imply that the interaction effect of the global population on a focal individual is of order 1. Parameter K may also be interpreted as scaling the resources available, so that the renormalization of U_K and V_K reflects the decrease of competition for resources.

The generator $\tilde{L}^K$ of $(\nu_t^K)_{t\geq 0}$ is given by (2.1), with parameters U_K, V_K, b_K, d_K, M_K, μ_K. The generator L^K of $(X_t^K)_{t\geq 0}$ is obtained by writing, for any measurable function ϕ from $M_F(\mathcal{X})$ into $\mathbb{R}$ and any $\nu \in M_F(\mathcal{X})$,

$$L^K\phi(\nu) = \partial_t E_\nu(\phi(X_t^K))_{t=0} = \partial_t E_{K\nu}(\phi(\nu_t^K/K))_{t=0} = \tilde{L}^K\phi^K(K\nu)$$

where $\phi^K(\mu) = \phi(\mu/K)$. Then we get

$$\begin{aligned} L^K\phi(\nu) = K\int_{\mathcal{X}} & b_K(x, V*\nu(x))(1-\mu_K(x))(\phi(\nu+\frac{1}{K}\delta_x)-\phi(\nu))\nu(dx) \\ &+ K\int_{\mathcal{X}}\int_{\mathcal{X}} b_K(x,V*\nu(x))\mu_K(x)(\phi(\nu+\frac{1}{K}\delta_z)-\phi(\nu))M_K(x,z)dz\nu(dx) \\ &+ K\int_{\mathcal{X}} d_K(x,U*\nu(x))(\phi(\nu-\frac{1}{K}\delta_x)-\phi(\nu))\nu(dx). \qquad (4.1)\end{aligned}$$

By a similar proof as the one of Section 2.3, we may summarize the moment and martingale properties of X^K.

Proposition 4.1. *Assume that for some $p \geq 2$, $E(\langle X_0^K, 1\rangle^p) < +\infty$.*

(1) *For any $T > 0$, $E(\sup_{t\in[0,T]}\langle X_t^K, 1\rangle^p) < +\infty$.*
(2) *For any bounded and measurable functions ϕ on M_F such that $|\phi(\nu)| + |L^K\phi(\nu)| \leq C(1+ <\nu, 1>^p)$, the process $\phi(X_t^K) - \phi(X_0^K) - \int_0^t L^K\phi(X_s^K)ds$ is a càdlàg martingale.*
(3) *For each measurable bounded function f, the process*

$$\begin{aligned} m_t^{K,f} &= \langle X_t^K, f\rangle - \langle X_0^K, f\rangle \\ &- \int_0^t\int_{\mathcal{X}}(b_K(x,V*X_s^K(x)) - d_K(x,U*X_s^K(x)))f(x)X_s^K(dx)ds \\ &- \int_0^t\int_{\mathcal{X}}\mu_K(x)b_K(x,V*X_s^K(x)\Big(\int_{\mathcal{X}} f(z)M_K(x,z)dz - f(x)\Big)X_s^K(dx)ds\end{aligned}$$

is a square integrable martingale with quadratic variation

$$\begin{aligned}\langle m^{K,f}\rangle_t = \frac{1}{K}\Big\{\int_0^t\int_{\mathcal{X}} & \mu_K(x)b_K(x,V*X_s^K(x)) \\ & \Big(\int_{\mathcal{X}} f^2(z)M_K(x,z)dz - f^2(x)\Big)X_s^K(dx)ds \\ + \int_0^t\int_{\mathcal{X}} & (b_K(x,V*X_s^K(x)) + d_K(x,U*X_s^K(x)))f^2(x)X_s^K(dx)ds\Big\}. \qquad (4.2)\end{aligned}$$

The search of tractable limits for the semimartingales $\langle X^K, f\rangle$ yields the different choices of scalings of the parameters developed in this section. In particular, we obtain the deterministic or stochastic nature of the approximation by studying the quadratic variation of the martingale term, given in (4.2).

4.1. Large-population limit

We assume here that $b_K = b$, $d_K = d$, $\mu_K = \mu$, $M_K = M$.

Theorem 4.2. *Assume Assumptions (H) and (H1). Assume moreover that the initial conditions X_0^K converge in law and for the weak topology on $M_F(\mathcal{X})$ as K increases, to a finite deterministic measure ξ_0, and that $\sup_K E(\langle X_0^K, 1\rangle^3) < +\infty$.*

Then for any $T > 0$, the process $(X_t^K)_{t\geq 0}$ converges in law, in the Skorohod space $\mathbb{D}([0,T], M_F(\mathcal{X}))$, as K goes to infinity, to the unique deterministic continuous function $\xi \in C([0,T], M_F(\mathcal{X}))$ satisfying for any bounded $f : \mathcal{X} \to \mathbb{R}$

$$\langle \xi_t, f\rangle = \langle \xi_0, f\rangle + \int_0^t \int_{\mathcal{X}} f(x)[(1-\mu(x))b(x, V * \xi_s(x)) - d((x, U * \xi_s(x))]\xi_s(dx)ds$$
$$+ \int_0^t \int_{\mathcal{X}} \mu(x)b(x, V * \xi_s(x)) \left(\int_{\mathcal{X}} f(z)M(x,z)dz \right) \xi_s(dx)ds. \tag{4.3}$$

The proof of Theorem 4.2 is let to the reader. It can be adapted from the proofs of Theorem 4.3 and 4.5 below, or obtained as a generalization of Theorem 5.3 in [15]. This result is illustrated by the simulations of Figs. 1 (a)–(c).

Main Examples:

(1) A density case. Following similar arguments as in the proof of Proposition 3.1, one shows that if the initial condition ξ_0 has a density w.r.t. Lebesgue measure, then the same property holds for the finite measure ξ_t, which is then solution of the functional equation:

$$\partial_t \xi_t(x) = [(1-\mu(x))b(x, V * \xi_t(x)) - d(x, U * \xi_t(x))]\, \xi_t(x)$$
$$+ \int_{\mathbb{R}^d} M(y,x)\mu(y)b(y, V * \xi_t(y))\xi_t(y)dy \tag{4.4}$$

for all $x \in \mathcal{X}$ and $t \geq 0$. Desvillettes et al. [9] suggest to refer to ξ_t as the population number density; then the quantity $n_t = \int_{\mathcal{X}} \xi_t(x)dx$ can be interpreted as the total population density over the whole trait space.

(2) The mean field case. As for moment equations (cf. Section 3), the case of spatially structured populations with constant rates b, d, α is meaningful. In this context, (4.4) leads to the following equation on n_t:

$$\partial_t n_t = (b-d)n_t - \alpha \int_{\mathcal{X}\times\mathcal{X}} U(x-y)\xi_t(dx)\xi_t(dy). \tag{4.5}$$

With the assumption $U \equiv 1$, we recover the classical mean-field logistic equation of population growth:

$$\partial_t n_t = (b-d)n_t - \alpha n_t^2.$$

Comparing (4.5) with the first-moment equation (3.2) obtained previously stresses out the "decorrelative" effect of the large system size renormalization (only in case $U \equiv 1$). In (3.2), the correction term capturing the effect of spatial correlations in the population remains, even if one assumes $U \equiv 1$.

(3) Monomorphic and dimorphic cases without mutation. We assume here that the population evolves without mutation (parameter $\mu = 0$); then the population traits are the initial ones.

(a) Monomorphic case: only trait x is present in the population at time $t = 0$. Thus, we can write $X_0^K = n_0^K(x)\delta_x$, and then $X_t^K = n_t^K(x)\delta_x$ for any time t. Theorem 4.2 recasts in this case into $n_t^K(x) \to n_t(x)$ with $\xi_t = n_t(x)\delta_x$, and (4.3) writes

$$\frac{d}{dt}n_t(x) = n_t(x)\big(b(x, V(0)n_t(x)) - d(x, U(0)n_t(x))\big), \tag{4.6}$$

(b) Dimorphic case: when the population contains two traits x and y, i.e., when $X_0^K = n_0^K(x)\delta_x + n_0^K(y)\delta_y$, we can define in a similar way $n_t(x)$ and $n_t(y)$ for any t as before, such that $\xi_t = n_t(x)\delta_x + n_t(y)\delta_y$ satisfies (4.3), which recasts into the following system of coupled ordinary differential equations:

$$\begin{aligned}\frac{d}{dt}n_t(x) &= n_t(x)\big(b(x, V(0)n_t(x) + V(x-y)n_t(y)) - d(x, U(0)n_t(x) + U(x-y)n_t(y))\big)\\ \frac{d}{dt}n_t(y) &= n_t(y)\big(b(y, V(0)n_t(y) + V(y-x)n_t(x)) - d(y, U(0)n_t(y) + U(y-x)n_t(x))\big).\end{aligned} \tag{4.7}$$

4.2. Large-population limit with accelerated births and deaths

We consider here an alternative limit of a large population, combined with accelerated birth and death. This may be useful to investigate the qualitative differences of evolutionary dynamics across populations with allometric demographies (larger populations made up of smaller individuals who reproduce and die faster, see [5, 8]).

Here, we assume for simplicity that $\mathcal{X} = \mathbb{R}^d$. Let us denote by M_F the space $M_F(\mathbb{R}^d)$. We consider the acceleration of birth and death processes at a rate proportional to K^η while preserving the demographic balance. That is, the birth and death rates scale with system size according to

Assumption (H2):

$$b_K(x,\zeta) = K^\eta r(x) + b(x,\zeta), \quad d_K(x,\zeta) = K^\eta r(x) + d(x,\zeta).$$

The allometric effect (smaller individuals reproduce and die faster) is parameterized by the function r, positive and bounded over $\mathbb{R}^d$, and the constant η. A detailed discussion of the biological meaning of these parameters in terms of allometry and life-history scalings can be found in [7]. As in Section 4.1, the interaction kernels V and U are renormalized by K. Using similar arguments as in Section 4.1, the process $X^K = \frac{1}{K}\nu^K$ is now a Markov process with generator

$$L^K\phi(\nu) = K\int_{\mathbb{R}^d}(K^\eta r(x) + b(x, V*\nu(x)))(1-\mu_K(x))(\phi(\nu+\frac{1}{K}\delta_x) - \phi(\nu))\nu(dx)$$
$$+ K\int_{\mathbb{R}^d}(K^\eta r(x) + b(x, V*\nu(x)))\mu_K(x)$$
$$\int_{\mathbb{R}^d}(\phi(\nu+\frac{1}{K}\delta_z) - \phi(\nu))M_K(x,z)dz\nu(dx)$$
$$+ K\int_{\mathbb{R}^d}(K^\eta r(x) + d(x, U*\nu(x)))(\phi(\nu-\frac{1}{K}\delta_x) - \phi(\nu))\nu(dx).$$

As before, for any measurable functions ϕ on M_F such that $|\phi(\nu)| + |L^K\phi(\nu)| \leq C(1+\langle\nu,1\rangle^3)$, the process

$$\phi(X_t^K) - \phi(X_0^K) - \int_0^t L^K\phi(X_s^K)ds \tag{4.8}$$

is a martingale. In particular, for each measurable bounded function f, we obtain

$$\begin{aligned} M_t^{K,f} &= \langle X_t^K, f\rangle - \langle X_0^K, f\rangle \\ &\quad - \int_0^t\int_{\mathbb{R}^d}(b(x, V*X_s^K(x)) - d(x, U*X_s^K(x)))f(x)X_s^K(dx)ds \\ &\quad - \int_0^t\int_{\mathbb{R}^d}\mu_K(x)(K^\eta r(x) + b(x, V*X_s^K(x))) \\ &\qquad\qquad \Big(\int_{\mathbb{R}^d} f(z)M_K(x,z)dz - f(x)\Big)X_s^K(dx)ds, \end{aligned} \tag{4.9}$$

is a square integrable martingale with quadratic variation

$$\begin{aligned} &\langle M^{K,f}\rangle_t \\ &= \frac{1}{K}\Big\{\int_0^t\int_{\mathbb{R}^d}(2K^\eta r(x) + b(x, V*X_s^K(x)) + d(x, U*X_s^K(x)))f^2(x)X_s^K(dx)ds \\ &\quad + \int_0^t\int_{\mathbb{R}^d}\mu_K(x)(K^\eta r(x) + b(x, V*X_s^K(x))) \\ &\qquad\qquad \Big(\int_{\mathbb{R}^d} f^2(z)M_K(x,z)dz - f^2(x)\Big)X_s^K(dx)ds\Big\}. \end{aligned} \tag{4.10}$$

Two interesting cases will be considered hereafter, in which the variance effect $\mu_K M_K$ is of order $1/K^\eta$. That will ensure the deterministic part in (4.9) to converge. In the large-population renormalization (Section 4.1), the quadratic variation of the martingale part was of the order of $1/K$. Here, it is of the order of $K^\eta \times 1/K$. This quadratic variation will thus stay finite provided that $\eta \in (0,1]$, in which case tractable limits will result. Moreover, this limit will be zero if $\eta < 1$ and nonzero if $\eta = 1$, which will lead to deterministic or random limit models.

4.2.1. Accelerated mutation and small mutation steps. We consider here that the mutation rate is fixed, so that mutations are accelerated as a consequence of accelerating birth. We assume

Assumptions (H3):

(1) $\mu_K = \mu$.

(2) The mutation step density $M_K(x,z)$ is the density of a random variable with mean x, variance-covariance matrix $\Sigma(x)/K^\eta$ (where $\Sigma(x) = (\Sigma_{ij}(x))_{1\le i,j\le d}$) and with third moment of order $1/K^{\eta+\varepsilon}$ uniformly in x ($\varepsilon > 0$). (Thus, as K goes to infinity, mutant traits become more concentrated around their 'progenitors').

(3) $\sqrt{\Sigma}$ denoting the symmetrical square root matrix of Σ, the function $\sqrt{\Sigma r\mu}$ is Lipschitz continuous.

The main example is when the mutation step density is taken as the density of a vector of independent Gaussian variables with mean x and variance $\sigma^2(x)/K^\eta$:

$$M_K(x,z) = \left(\frac{K^\eta}{2\pi\sigma^2(x)}\right)^{d/2} \exp[-K^\eta|z-x|^2/2\sigma^2(x)] \tag{4.11}$$

where $\sigma^2(x)$ is positive and bounded over $\mathbb{R}^d$.

Then the convergence results of this section can be stated as follows.

Theorem 4.3. (1) *Assume* (H), (H1), (H2), (H3) *and* $0 < \eta < 1$. *Assume also that the initial conditions* X_0^K *converge in law and for the weak topology on* M_F *as* K *increases, to a finite deterministic measure* ξ_0, *and that*

$$\sup_K E(\langle X_0^K, 1\rangle^3) < +\infty. \tag{4.12}$$

Then, for each $T > 0$, *the sequence of processes* (X^K) *belonging to* $\mathbb{D}([0,T], M_F)$ *converges (in law) to the unique deterministic function* $(\xi_t)_{t\ge 0}$ $\in C([0,T], M_F)$ *satisfying: for each function* $f \in C_b^2(\mathbb{R}^d)$,

$$\begin{aligned}\langle \xi_t, f\rangle = \langle \xi_0, f\rangle &+ \int_0^t \int_{\mathbb{R}^d} (b(x, V*\xi_s(x)) - d(x, U*\xi_s(x)))f(x)\xi_s(dx)ds \\ &+ \int_0^t \int_{\mathbb{R}^d} \frac{1}{2}\mu(x)r(x) \sum_{1\le i,j\le d} \Sigma_{ij}(x)\partial^2_{ij} f(x)\xi_s(dx)ds,\end{aligned} \tag{4.13}$$

where $\partial^2_{ij} f$ *denotes the second-order partial derivative of* f *with respect to* x_i *and* x_j $(x = (x_1, \dots, x_d))$.

(2) *Assume moreover that there exists* $c > 0$ *such that* $r(x)\mu(x)s^*\Sigma(x)s \ge c||s||^2$ *for any* x *and* s *in* $\mathbb{R}^d$. *Then for each* $t > 0$, *the measure* ξ_t *has a density with respect to Lebesgue measure.*

Remark 4.4. In case (2), Eq. (4.13) may be written as

$$\partial_t \xi_t(x) = \Big(b(x, V*\xi_t(x)) - d(x, U*\xi_t(x))\Big)\xi_t(x) + \frac{1}{2}\sum_{1\le i,j\le d} \partial^2_{ij}(r\mu\Sigma_{ij}\xi_t)(x). \tag{4.14}$$

Observe that, for the example (4.11), this equation writes

$$\partial_t \xi_t(x) = \Big(b(x, V * \xi_t(x)) - d(x, U * \xi_t(x)) \Big) \xi_t(x) + \frac{1}{2} \Delta(\sigma^2 r \mu \xi_t)(x). \qquad (4.15)$$

Therefore, Eq. (4.15) generalizes the Fisher reaction-diffusion equation known from classical population genetics (see e.g. [4]).

Theorem 4.5. *Assume* (H), (H1), (H2), (H3) *and* $\eta = 1$. *Assume also that the initial conditions* X_0^K *converge in law and for the weak topology on* $M_F(\mathcal{X})$ *as* K *increases, to a finite (possibly random) measure* X_0, *and that* $\sup_K E(\langle X_0^K, 1\rangle^3) < +\infty$.

Then, for each $T > 0$, *the sequence of processes* (X^K) *converges in law in* $\mathbb{D}([0,T], M_F)$ *to the unique (in law) continuous superprocess* $X \in C([0,T], M_F)$, *defined by the following conditions:*

$$\sup_{t\in[0,T]} E\left(\langle X_t, 1\rangle^3\right) < \infty, \qquad (4.16)$$

and for any $f \in C_b^2(\mathbb{R}^d)$,

$$\begin{aligned}\bar{M}_t^f = \langle X_t, f\rangle - \langle X_0, f\rangle - \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \mu(x) r(x) \sum_{1\le i,j\le d} \Sigma_{ij}(x) \partial_{ij}^2 f(x) X_s(dx) ds \\ - \int_0^t \int_{\mathbb{R}^d} f(x) \left(b(x, V * X_s(x)) - d(x, U * X_s(x))\right) X_s(dx) ds \qquad (4.17)\end{aligned}$$

is a continuous martingale with quadratic variation

$$\langle \bar{M}^f \rangle_t = 2 \int_0^t \int_{\mathbb{R}^d} r(x) f^2(x) X_s(dx) ds. \qquad (4.18)$$

Remark 4.6.
(1) The limiting measure-valued process X appears as a generalization of the one proposed by Etheridge [12] to model spatially structured populations.
(2) The conditions characterizing the process X above can be formally rewritten as equation

$$\begin{aligned}\partial_t X_t(x) = \Big(b(x, V * X_t(x)) - d(x, U * X_t(x)) \Big) X_t(x) \\ + \frac{1}{2} \sum_{1\le i,j\le d} \partial_{ij}^2 (r\mu \Sigma_{ij} X_t)(x) + \dot{M}_t\end{aligned}$$

where $\dot{M}_t$ is a random fluctuation term, which reflects the demographic stochasticity of this fast birth-and-death process, that is, faster than the accelerated birth-and-death process which led to the deterministic reaction-diffusion approximation (4.15).
(3) As developed in Step 1 of the proof of Theorem 4.5 below, a Girsanov's theorem relates the law of X_t and the one of a standard super-Brownian

motion, which leads to conjecture that a density for X_t exists only when $d = 1$, as for the super-Brownian motion.

These two theorems are illustrated by the simulations of Figs. 2 (a), (c) and (d).

Proof of Theorem 4.3. (1) We divide the proof in several steps. Let us fix $T > 0$.

Step 1. Let us first show the uniqueness for a solution of the equation (4.13).

To this aim, we define the evolution equation associated with (4.13). It is easy to prove that if ξ is a solution of (4.13) satisfying $\sup_{t\in[0,T]}\langle \xi_t, 1\rangle < \infty$, then for each test function $\psi_t(x) = \psi(t,x) \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$, one has

$$\langle \xi_t, \psi_t\rangle = \langle \xi_0, \psi_0\rangle + \int_0^t \int_{\mathbb{R}^d} (b(x, V * \xi_s(x)) - d(x, U * \xi_s(x)))\psi(s,x)\xi_s(dx)ds$$
$$+ \int_0^t \int_{\mathbb{R}^d} (\partial_s \psi(s,x) + \frac{1}{2} r(x)\mu(x) \sum_{i,j} \Sigma_{ij}(x)\partial^2_{ij}\psi_s(x))\xi_s(dx)ds.$$

Now, since the function $\sqrt{\Sigma r\mu}$ is Lipschitz continuous, we may define the transition semigroup (P_t) whith infinitesimal generator $f \mapsto \frac{1}{2} r\mu \sum_{i,j}\Sigma_{ij}\partial^2_{ij} f$. Then, for each function $f \in C_b^2(\mathbb{R}^d)$ and fixed $t > 0$, to choose $\psi(s,x) = P_{t-s}f(x)$ yields

$$\langle \xi_t, f\rangle = \langle \xi_0, P_t f\rangle + \int_0^t \int_{\mathbb{R}^d} (b(x, V * \xi_s(x)) - d(x, U * \xi_s(x)))P_{t-s}f(x)\xi_s(dx)ds, \tag{4.19}$$

since $\partial_s \psi(s,x) + \frac{1}{2} r(x)\mu(x)\sum_{i,j}\Sigma_{ij}(x)\partial^2_{ij}\psi_s(x) = 0$ for this choice.

We now prove the uniqueness of a solution of (4.19).

Let us consider two solutions $(\xi_t)_{t\geq 0}$ and $(\bar{\xi}_t)_{t\geq 0}$ of (4.19) satisfying $\sup_{t\in[0,T]}\left\langle \xi_t + \bar{\xi}_t, 1\right\rangle = A_T < +\infty$. We consider the variation norm defined for μ_1 and μ_2 in M_F by

$$||\mu_1 - \mu_2|| = \sup_{f\in L^\infty(\mathbb{R}^d),\, ||f||_\infty \leq 1} |\left\langle \mu_1 - \mu_2, f\right\rangle|.$$

Then, we consider some bounded and measurable function f defined on $\mathcal{X}$ such that $||f||_\infty \leq 1$ and obtain

$$\begin{aligned}
&|\left\langle \xi_t - \bar{\xi}_t, f\right\rangle| \\
&\leq \int_0^t \left| \int_{\mathbb{R}^d} [\xi_s(dx) - \bar{\xi}_s(dx)]\,(b(x, V*\xi_s(x)) - d(x, U*\xi_s(x)))\,P_{t-s}f(x)\right| ds \\
&\quad + \int_0^t \left| \int_{\mathbb{R}^d} \bar{\xi}_s(dx)(b(x, V*\xi_s(x)) - b(x, V*\bar{\xi}_s(x)))P_{t-s}f(x)\right| ds \\
&\quad + \int_0^t \left| \int_{\mathbb{R}^d} \bar{\xi}_s(dx)(d(x, U*\xi_s(x)) - d(x, U*\bar{\xi}_s(x)))P_{t-s}f(x)\right| ds.
\end{aligned} \tag{4.20}$$

Since $||f||_\infty \leq 1$, then $||P_{t-s}f||_\infty \leq 1$ and for all $x \in \mathbb{R}^d$,

$$|(b(x, V*\xi_s(x)) - d(x, U*\xi_s(x)))P_{t-s}f(x)| \leq \bar{b} + \bar{d}(1 + \bar{U}A_T).$$

Moreover, b and d are Lipschitz continuous in their second variable with respective constants K_b and K_d. Thus we obtain from (4.20) that

$$|\langle \xi_t - \bar{\xi}_t, f\rangle| \leq \left[\bar{b} + \bar{d}(1+\bar{U}A_T) + K_b A_T \bar{V} + K_d A_T \bar{U}\right] \int_0^t ||\xi_s - \bar{\xi}_s|| ds.$$

Taking the supremum over all functions f such that $||f||_\infty \leq 1$, and using the Gronwall Lemma, we finally deduce that for all $t \leq T$, $||\xi_t - \bar{\xi}_t|| = 0$. Uniqueness holds.

Step 2. Next, we would like to obtain some moment estimates. First, we check that for all $T < \infty$,

$$\sup_K \sup_{t\in[0,T]} E\big(\langle X_t^K, 1\rangle^3\big) < \infty. \tag{4.21}$$

To this end, we use (4.8) with $\phi(\nu) = \langle \nu, 1\rangle^3$. (To be completely rigorous, one should first use $\phi(\nu) = \langle \nu, 1\rangle^3 \wedge A$, make A tend to infinity). Taking expectation, we obtain that for all $t \geq 0$, all K,

$$\begin{aligned}
&E\left(\langle X_t^K, 1\rangle^3\right) = E\left(\langle X_0^K, 1\rangle^3\right)\\
&+\int_0^t E\bigg(\int_{\mathbb{R}^d}\bigg([K^{\eta+1}r(x) + Kb(x, V * X_s^K(x))]\left\{[\langle X_s^K, 1\rangle + \frac{1}{K}]^3 - \langle X_s^K, 1\rangle^3\right\}\\
&\left\{K^{\eta+1}r(x) + Kd(x, U * X_s^K(x))\right\}\left\{[\langle X_s^K, 1\rangle - \frac{1}{K}]^3 - \langle X_s^K, 1\rangle^3\right\}\bigg)X_s^K(dx)\bigg)ds.
\end{aligned}$$

Dropping the non-positive death term involving d, we get

$$\begin{aligned}
&E\left(\langle X_t^K, 1\rangle^3\right) \leq E\left(\langle X_0^K, 1\rangle^3\right)\\
&\quad+\int_0^t E\bigg(\int_{\mathbb{R}^d}\bigg(K^{\eta+1}r(x)\left\{[\langle X_s^K, 1\rangle + \frac{1}{K}]^3 + [\langle X_s^K, 1\rangle - \frac{1}{K}]^3 - 2\langle X_s^K, 1\rangle^3\right\}\\
&\quad+Kb(x, V * X_s^K(x))\left\{[\langle X_s^K, 1\rangle + \frac{1}{K}]^3 - \langle X_s^K, 1\rangle^3\right\}\bigg)X_s^K(dx)\bigg)ds.
\end{aligned}$$

But for all $x \geq 0$, all $\epsilon \in (0,1]$, $(x+\epsilon)^3 - x^3 \leq 6\epsilon(1+x^2)$ and $|(x+\epsilon)^3 + (x-\epsilon)^3 - 2x^3| = 6\epsilon^2 x$. We finally obtain

$$E\left(\langle X_t^K, 1\rangle^3\right) \leq E\left(\langle X_0^K, 1\rangle^3\right) + C\int_0^t E\left(\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2 + \langle X_s^K, 1\rangle^3\right) ds.$$

Assumption (4.12) and the Gronwall lemma allows us to conclude that (4.21) holds. Next, we wish to check that

$$\sup_K E\big(\sup_{t\in[0,T]} \langle X_t^K, 1\rangle^2\big) < \infty. \tag{4.22}$$

Applying (4.9) with $f \equiv 1$, we obtain

$$\begin{aligned}
\langle X_t^K, 1\rangle =& \langle X_0^K, 1\rangle\\
&+\int_0^t \int_{\mathcal{X}} \left(b(x, V * X_s^K(x)) - d(x, U * X_s^K(x))\right) X_s^K(dx) ds + m_t^{K,1}.
\end{aligned}$$

Hence

$$\sup_{s\in[0,t]} \langle X_s^K, 1\rangle^2 \leq C\bigg(\langle X_0^K, 1\rangle^2 + \bar{b}\int_0^t \langle X_s^K, 1\rangle^2 ds + \sup_{s\in[0,t]} |M_s^{K,1}|^2\bigg).$$

Thanks to (4.12), the Doob inequality and the Gronwall Lemma, there exists a constant C_t not depending on K such that

$$E\big(\sup_{s\in[0,t]} \langle X_s^K, 1\rangle^2\big) \leq C_t \left(1 + E\left(\langle M^{K,1}\rangle_t\right)\right).$$

Using now (4.10), we obtain, for some other constant C_t not depending on K,

$$E\left(\langle M^{K,1}\rangle_t\right) \leq C\int_0^t \left(E\left(\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2\right)\right)ds \leq C_t$$

thanks to (4.21). This concludes the proof of (4.22).

Step 3. We first endow M_F with the vague topology, the extension to the weak topology being handled in Step 6 below. To show the tightness of the sequence of laws $Q^K = \mathcal{L}(X^K)$ in $\mathcal{P}(\mathbb{D}([0,T], M_F))$, it suffices, following Roelly [23], to show that for any continuous bounded function f on $\mathbb{R}^d$, the sequence of laws of the processes $\langle X^K, f\rangle$ is tight in $\mathbb{D}([0,T], \mathbb{R})$. To this end, we use the Aldous criterion [1] and the Rebolledo criterion (see [17]). We have to show that

$$\sup_K E\big(\sup_{t\in[0,T]} |\langle X_s^K, f\rangle|\big) < \infty, \tag{4.23}$$

and the tightness, respectively, of the laws of the predictable quadratic variation of the martingale part and of the drift part of the semimartingales $\langle X^K, f\rangle$.

Since f is bounded, (4.23) is a consequence of (4.22): let us thus consider a couple (S, S') of stopping times satisfying a.s. $0 \leq S \leq S' \leq S+\delta \leq T$. Using (4.10) and (4.22), we get for constants C, C'

$$E\left(\langle M^{K,f}\rangle_{S'} - \langle M^{K,f}\rangle_S\right) \leq CE\left(\int_S^{S+\delta} \left(\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2\right) ds\right) \leq C'\delta.$$

In a similar way, the expectation of the finite variation part of $\langle X_{S'}^K, f\rangle - \langle X_S^K, f\rangle$ is bounded by $C'\delta$.

Hence, the sequence $Q^K = \mathcal{L}(X^K)$ is tight.

Step 4. Let us now denote by Q the limiting law of a subsequence of Q^K. We still denote this subsequence by Q^K. Let $X = (X_t)_{t\geq 0}$ a process with law Q. We remark that by construction, almost surely,

$$\sup_{t\in[0,T]} \sup_{f\in L^\infty(\mathbb{R}^d), ||f||_\infty\leq 1} |\langle X_t^K, f\rangle - \langle X_{t^-}^K, f\rangle| \leq 1/K.$$

This implies that the process X is a.s. strongly continuous.

Step 5. The time $T > 0$ is fixed. Let us now check that almost surely, the process X is the unique solution of (4.13). Thanks to (4.22), it satisfies $\sup_{t\in[0,T]}\langle X_t, 1\rangle < +\infty$ a.s., for each T. We fix now a function $f \in C_b^3(\mathbb{R}^d)$ (the extension of (4.13) to any function f in C_b^2 is not hard) and some $t \leq T$.

For $\nu \in C([0,T], M_F)$, denote

$$\Psi_t^1(\nu) = \langle \nu_t, f\rangle - \langle \nu_0, f\rangle - \int_0^t \int_{\mathbb{R}^d} (b(x, V * \nu_s(x)) - d(x, U * \nu_s(x)))f(x)\nu_s(dx)ds,$$
$$\Psi_t^2(\nu) = -\int_0^t \int_{\mathbb{R}^d} \frac{1}{2}\mu(x)r(x)\sum_{i,j}\Sigma_{ij}(x)\partial_{ij}^2 f(x)\nu_s(dx)ds.$$

We have to show that

$$E_Q\left(|\Psi_t^1(X) + \Psi_t^2(X)|\right) = 0. \tag{4.24}$$

By (4.9), we know that for each K,

$$M_t^{K,f} = \Psi_t^1(X^K) + \Psi_t^{2,K}(X^K),$$

where

$$\Psi_t^{2,K}(X^K) = -\int_0^t \int_{\mathbb{R}^d} \mu(x)(K^\eta r(x) + b(x, V * X_s^K(x)))$$
$$\left(\int_{\mathbb{R}^d} f(z)M_K(x,z)dz - f(x)\right)X_s^K(dx)ds.$$

Moreover, (4.22) implies that for each K,

$$E\left(|M_t^{K,f}|^2\right) = E\left(\langle M^{K,f}\rangle_t\right)$$
$$\leq \frac{C_f K^\eta}{K} E\left(\int_0^t \left\{\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2\right\} ds\right) \leq \frac{C_{f,T}K^\eta}{K},$$

which goes to 0 as K tends to infinity, since $0 < \eta < 1$. Therefore,

$$\lim_K E(|\Psi_t^1(X^K) + \Psi_t^{2,K}(X^K)|) = 0.$$

Since X is a.s. strongly continuous, since $f \in C_b^3(\mathbb{R}^d)$ and thanks to the continuity of the parameters, the functions Ψ_t^1 and Ψ_t^2 are a.s. continuous at X. Furthermore, for any $\nu \in \mathbb{D}([0,T], M_F)$,

$$|\Psi_t^1(\nu) + \Psi_t^2(\nu)| \leq C_{f,T} \sup_{s\in[0,T]} \left(1 + \langle \nu_s, 1\rangle^2\right).$$

Hence using (4.21), we see that the sequence $(\Psi_t^1(X^K) + \Psi_t^2(X^K))_K$ is uniformly integrable, and thus

$$\lim_K E\left(|\Psi_t^1(X^K) + \Psi_t^2(X^K)|\right) = E\left(|\Psi_t^1(X) + \Psi_t^2(X)|\right).$$

We have now to deal with $\Psi_t^{2,K}(X^K) - \Psi_t^2(X^K)$. The convergence of this term is due to the fact that the measure $M_K(x,z)dz$ has mean x, variance $\Sigma(x)/K^\eta$, and third moment bounded by $C/K^{\eta+\varepsilon}$ ($\varepsilon > 0$) uniformly in x. Indeed, if $Hf(x)$ denotes the Hessian matrix of f at x,

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(z)M_K(x,z)dz \\
&= \int_{\mathbb{R}^d}\left(f(x)+(z-x)\cdot\nabla f(x)+\frac{1}{2}(z-x)^*Hf(x)(z-x)+O((z-x)^3)\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad M_K(x,z)dz \\
&= f(x)+\frac{1}{2}\sum_{i,j}\frac{\Sigma_{ij}(x)}{K^\eta}\partial^2_{ij}f(x)+o(\frac{1}{K^\eta}) \qquad (4.25)
\end{aligned}$$

where $K^\eta o(\frac{1}{K^\eta})$ tends to 0 uniformly in x (since f is in C_b^3), as K tends to infinity. Then,

$$\begin{aligned}
\Psi_t^{2,K}(X^K) = -\int_0^t\int_{\mathbb{R}^d}&\mu(x)(K^\eta r(x)+b(x,V*X_s^K(x))) \\
&\times\left(\frac{1}{2}\sum_{i,j}\frac{\Sigma_{ij}(x)}{K^\eta}\partial^2_{ij}f(x)+o(\frac{1}{K^\eta})\right)X_s^K(dx)ds,
\end{aligned}$$

and

$$|\Psi_t^{2,K}(X^K)-\Psi_t^2(X^K)| \leq C_f < X_s^K, 1 > \left(\frac{1}{K^\eta}+K^\eta o(\frac{1}{K^\eta})\right).$$

Using (4.22), we conclude the proof of (4.24).

Step 6. The previous steps imply that $(X^K)_K$ converges to ξ in $\mathbb{D}([0,T],M_F)$, where M_F is endowed with the vague topology. To extend the result to the case where M_F is endowed with the weak topology, we use a criterion proved in Méléard and Roelly [20]: since the limiting process is continuous, it suffices to prove that the sequence $(\langle X^K,1\rangle)$ converges to $\langle\xi,1\rangle$ in law, in $\mathbb{D}([0,T],\mathbb{R})$. One may of course apply Step 5 with $f\equiv 1$, which concludes the proof.

(2) Let us now assume the non-degeneracy property $r(x)\mu(x)s^*\Sigma(x)s \geq c\|s\|^2 > 0$ for each $x\in\mathbb{R}^d, s\in\mathbb{R}^d$. That implies that for each time $t>0$, the transition semigroup $P_t(x,dy)$ introduced in Step 1 of this proof has for each x a density function $p_t(x,y)$ with respect to the Lebesgue measure. Then if we come back to the evolution equation (4.19), we can write

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(x)\xi_t(dx) = \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(y)p_t(x,y)dy\right)\xi_0(dx) \\
&+\int_0^t\int_{\mathbb{R}^d}(b(x,V*\xi_s(x))-d(x,U*\xi_s(x)))\left(\int_{\mathbb{R}^d} f(y)p_{t-s}(x,y)dy\right)\xi_s(dx)ds.
\end{aligned}$$

Using the fact that the parameters are bounded, that $\sup_{t\leq T}\langle\xi_t,1\rangle<+\infty$ and that f is bounded, we can apply Fubini's theorem and deduce that

$$\int_{\mathbb{R}^d} f(x)\xi_t(dx) = \int_{\mathbb{R}^d} H_t(y)f(y)dy$$

with $H \in L^\infty([0,T], L^1(\mathbb{R}^d))$, which implies that ξ_t has a density with respect to the Lebesgue measure for each time $t \leq T$.

Equation (4.14) is then the dual form of (4.13). $\square$

Proof of Theorem 4.5. We will use a similar method as the one of the previous theorem. Steps 2, 3, 4 and 6 of this proof can be achieved exactly in the same way. Therefore, we only have to prove the uniqueness (in law) of the solution to the martingale problem (4.16)–(4.18) (Step 1), and that any accumulation point of the sequence of laws of X^K is solution to (4.16)–(4.18) (Step 5).

Step 1. This uniqueness result is well-known for the super-Brownian process (defined by a similar martingale problem, but with $b = d = 0$, $r = \mu = 1$ and $\Sigma = \mathrm{Id}$, cf. [23]). Following [12], we may use the version of Dawson's Girsanov transform obtained in Evans and Perkins [14, Theorem 2.3], to deduce the uniqueness in our situation, provided the condition

$$E\left(\int_0^t \int_{\mathbb{R}^d} [b(x, V * X_s(x)) - d(x, U * X_s(x))]^2 X_s(dx)ds\right) < +\infty$$

is satisfied. This is easily obtained from the assumption that $\sup_{t\in[0,T]} E[\langle X_t, 1\rangle^3] < \infty$ since the coefficients are bounded.

Step 5. Let us identify the limit. Let us call $Q^K = \mathcal{L}(X^K)$ and denote by Q a limiting value of the tight sequence Q^K, and by $X = (X_t)_{t\geq 0}$ a process with law Q. Because of Step 4, X belongs a.s. to $C([0,T], M_F)$. We have to show that X satisfies the conditions (4.16), (4.17) and (4.18). First note that (4.16) is straightforward from (4.22). Then, we show that for any function f in $C_b^3(\mathbb{R}^d)$, the process $\bar{M}_t^f$ defined by (4.17) is a martingale (the extension to every function in C_b^2 is not hard). We consider $0 \leq s_1 \leq \cdots \leq s_n < s < t$, some continuous bounded maps $\phi_1, \ldots, \phi_n$ on M_F, and our aim is to prove that, if the function Ψ from $\mathbb{D}([0,T], M_F)$ into $\mathbb{R}$ is defined by

$$\Psi(\nu) = \phi_1(\nu_{s_1}) \cdots \phi_n(\nu_{s_n}) \Big\{ \langle \nu_t, f\rangle - \langle \nu_s, f\rangle \\ - \int_s^t \int_{\mathbb{R}^d} \Big(\frac{1}{2}\mu(x)r(x) \sum_{i,j} \Sigma_{ij}\partial_{ij}^2 f(x) + f(x)\,[b(x, V * \nu_u(x)) - d(x, U * \nu_u(x))] \Big) \\ \nu_u(dx)du \Big\},$$

then

$$E\left(\Psi(X)\right) = 0. \tag{4.26}$$

It follows from (4.9) that

$$0 = E\left(\phi_1(X_{s_1}^K) \cdots \phi_n(X_{s_n}^K) \left\{ M_t^{K,f} - M_s^{K,f} \right\}\right) = E\left(\Psi(X^K)\right) - A_K,$$

where A_K is defined by

$$A_K = E\Big(\phi_1(X^K_{s_1})\cdots\phi_n(X^K_{s_n})$$
$$\int_s^t \int_{\mathbb{R}^d} \mu(x)\Big\{b(x, V * X^K_u(x))\Big[\int_{\mathbb{R}^d}(f(z)-f(x))M_K(x,z)dz\Big]$$
$$+ r(x)K\Big[\int_{\mathbb{R}^d}(f(z)-f(x)-\sum_{i,j}\frac{\Sigma_{ij}(x)}{2K}\partial^2_{ij}f(x))M_K(x,z)dz\Big]\Big\}X^K_u(dx)du\Big).$$

It turns out from (4.25) that A_K tends to zero as K grows to infinity, and using (4.22), that the sequence $(|\Psi(X^K)|)_K$ is uniformly integrable, so

$$\lim_K E\left(|\Psi(X^K)|\right) = E_Q\left(|\Psi(X)|\right).$$

Collecting the previous results allows us to conclude that (4.26) holds, and thus $\bar{M}^f$ is a martingale.

We finally have to show that the bracket of $\bar{M}^f$ is given by (4.18). To this end, we first check that

$$\bar{N}^f_t = \langle X_t, f\rangle^2 - \langle X_0, f\rangle^2 - \int_0^t \int_{\mathbb{R}^d} 2r(x)f^2(x)X_s(dx)ds$$
$$- 2\int_0^t \langle X_s, f\rangle \int_{\mathbb{R}^d} f(x)\left[b(x, V*X_s(x)) - d(x, U*X_s(x))\right]X_s(dx)ds$$
$$- \int_0^t \langle X_s, f\rangle \int_{\mathbb{R}^d} \mu(x)r(x)\sum_{i,j}\Sigma_{ij}(x)\partial^2_{ij}f(x)X_s(dx)ds \tag{4.27}$$

is a martingale. This can be done exactly as for $\bar{M}^f_t$, using the semimartingale decomposition of $\langle X^K_t, f\rangle^2$, given by (4.8) with $\phi(\nu) = \langle \nu, f\rangle^2$. On the other hand, Itô's formula implies that

$$\langle X_t, f\rangle^2 - \langle X_0, f\rangle^2 - \langle \bar{M}^f\rangle_t - \int_0^t \langle X_s, f\rangle \int_{\mathbb{R}^d} r(x)\mu(x)\sum_{i,j}\Sigma_{ij}(x)\partial^2_{ij}f(x)X_s(dx)ds$$
$$- 2\int_0^t \langle X_s, f\rangle \int_{\mathbb{R}^d} f(x)\big[b(x, V*X_s(x)) - d(x, U*X_s(x))\big]X_s(dx)ds$$

is a martingale. Comparing this formula with (4.27), we obtain (4.18). □

4.2.2. Rare mutations. In this case, the mutation step density M is fixed and the mutation rate is decelerated proportionally to $1/K^\eta$:

Assumption (H4):

$$M_K = M, \quad \mu_K = \frac{\mu}{K^\eta}.$$

Thus only births without mutation are accelerated.

As in Section 4.2.1, we obtain deterministic or random limits, according to the value of $\eta \in (0,1]$.

Theorem 4.7. (1) *Assume* (H), (H1), (H2), (H4) *and* $0 < \eta < 1$. *Assume also that the initial conditions* X_0^K *converge in law and for the weak topology on* $M_F(\mathcal{X})$ *as* K *increases, to a finite deterministic measure* ξ_0, *and that* $\sup_K E(\langle X_0^K, 1\rangle^3) < +\infty$.

Then, for each $T > 0$, *the sequence of processes* (X^K) *belonging to* $\mathbb{D}([0,T], M_F)$ *converges (in law) to the unique deterministic function* $(\xi_t)_{t\geq 0} \in C([0,T], M_F)$ *weak solution of the deterministic nonlinear integro-differential equation:*

$$\partial_t \xi_t(x) = [b(x, V * \xi_t(x)) - d(x, U * \xi_t(x))]\xi_t(x) + \int_{\mathbb{R}^d} M(y,x)\mu(y)r(y)\xi_t(y)dy \\ - \mu(x)r(x)\xi_t(x). \quad (4.28)$$

(2) *Assume now* $\eta = 1$ *and that* X_0^K *converge in law to* X_0. *Then, for each* $T > 0$, *the sequence of processes* (X^K) *converges in law in* $\mathbb{D}([0,T], M_F)$ *to the unique (in law) continuous superprocess* $X \in C([0,T], M_F)$, *defined by the following conditions:*

$$\sup_{t\in[0,T]} E\left(\langle X_t, 1\rangle^3\right) < \infty,$$

and for any $f \in C_b^2(\mathbb{R}^d)$,

$$\bar{M}_t^f = \langle X_t, f\rangle - \langle X_0, f\rangle - \int_0^t \int_{\mathbb{R}^d} \mu(x)r(x)\int_{\mathbb{R}^d} M(x,z)(f(z)-f(x))dz X_s(dx)ds \\ - \int_0^t \int_{\mathbb{R}^d} f(x)\left(b(x, V*X_s(x)) - d(x, U*X_s(x))\right) X_s(dx)ds$$

is a continuous martingale with quadratic variation

$$\langle \bar{M}^f\rangle_t = 2\int_0^t \int_{\mathbb{R}^d} r(x)f^2(x)X_s(dx)ds.$$

In a SPDE formalism, one can write the last limit as formal solution of the equation

$$\partial_t X_t(x) = [b(x, V*X_t(x)) - d(x, U*X_t(x))]X_t(x) + \int_{\mathbb{R}^d} M(y,x)\mu(y)r(y)X_t(dy) \\ + \dot{M} - \mu(x)r(x)X_t(x), \quad (4.29)$$

where $\dot{M}$ is a random fluctuation term.

The proof of Theorem 4.7 is similar to proofs of Theorems 4.3 and 4.5 and we leave it to the reader. Theorem 4.7 (1) is illustrated in the simulation of Fig. 2 (b).

5. Rare mutation renormalization of the monomorphic process and adaptive dynamics

In the previous section, Eqs. (4.28) and (4.29) have been obtained at the population growth time scale (ecological time scale), under an assumption of rare mutation. Here, we are interested in the behavior of the population process at the evolutionary time scale, when mutations are extremely rare, as illustrated by the simulation of Fig. 1 (d). We hence recover rigorously the stochastic "trait substitution sequence" jump process of adaptive dynamics (Metz et al. [22]) when the initial condition is monomorphic. The biological idea behind such a scaling of the population process is that selection has sufficient time between two mutations to eliminate all disadvantaged traits, so that the population remains monomorphic on the evolutionary timescale. Then the evolution proceeds by successive invasions of mutant traits, replacing the resident trait from which the mutant trait is born, occuring on an infinitesimal timescale with respect to the mutation timescale. Our result emphasizes how the mutation scaling should compare to the system size (K) in order to obtain the correct time scale separation between the "mutant-invasions" (taking place on a short time scale) and the mutations (evolutionary time scale).

5.1. Statement of the result

We consider here a limit of rare mutations combined with the large population limit of Section 4.1 (Assumption (H1) and $b_K = b$, $d_K = d$ and $M_K = M$). We assume

Assumptions (H5):

(i) $\mu_K(x) = u_K \mu(x)$.

(ii) For any constant $C > 0$,

$$e^{-CK} \ll u_K \ll \frac{1}{K \log K}$$

(thus $u_K \to 0$ when $K \to +\infty$), or, equivalently, for any C and $t > 0$,

$$\log K \ll \frac{t}{K u_K} \ll e^{CK}. \tag{5.1}$$

(iii) For any $x \in \mathcal{X}$, $\zeta \mapsto b(x, \zeta)$ and $\zeta \mapsto d(x, \zeta)$ are positive functions, nonincreasing and increasing, respectively, satisfying

$$\begin{aligned} &\forall x \in \mathcal{X},\ b(x, 0) - d(x, 0) > 0, \\ &\lim_{\zeta \to +\infty} \inf_{x \in \mathcal{X}} d(x, \zeta) = +\infty. \end{aligned} \tag{5.2}$$

(iv) There exists a constant $\underline{U} > 0$ such that $U(h) \geq \underline{U}$ for any $h \in \mathbb{R}^d$.

Assumption (H5)-(i) entails the rare mutation asymptotic, and (H5)-(ii) gives the correct scaling between the mutation probability and the system size in order to obtain the correct time scale separation. Observe that (H5)-(ii) implies that $K u_K \to 0$ when $K \to +\infty$, so that the timescale $t/K u_K$, which corresponds to the timescale of mutations (the population size is proportional to K, and each birth event produces a mutant with a probability proportional to u_K, which gives a total mutation rate in the population proportional to Ku_K) is a long timescale. Our result gives the behavior of the population process on this long timescale.

Assumptions (H5)-(iii) and (iv) will allow to bound the population size on the mutation timescale, and to study the behavior of the population when it is monomorphic or dimorphic between two (rare) mutation events. Specifically, the monotonicity properties of b and d in Assumption (H5)-(iii) ensures, for any $x \in \mathcal{X}$, the existence of a unique non-trivial stable equilibrium $\bar{n}(x)$ for the monomorphic logistic equation (4.6) of Example 3 in Section 4.1. Moreover, since $b(x, V(0)u) - d(x, U(0)u) > 0$ for any $u < \bar{n}(x)$ and $b(x, V(0)u) - d(x, U(0)u) < 0$ for any $u > \bar{n}(x)$, any solution to (4.6) with positive initial condition converges to $\bar{n}(x)$.

Concerning the dimorphic logistic equations (4.7), an elementary linear analysis of the equilibrium $(\bar{n}(x), 0)$ gives that it is stable if $f(y, x) < 0$ and unstable if $f(y, x) > 0$, where the function

$$f(y, x) = b(y, V(y - x)\bar{n}(x)) - d(y, U(y - x)\bar{n}(x)) \tag{5.3}$$

is known as the "fitness function" ([21, 22]), which gives a measure of the selective advantage of a mutant individual with trait y in a monomorphic population of trait x at equilibrium. Similarly, the stability of the equilibrium $(0, \bar{n}(y))$ is governed by the sign of $f(x, y)$.

In order to ensure that, when the invasion of a mutant trait is possible, then this invasion will end with the extinction of the resident trait, we will need the following additional assumption:

Assumptions (H6):

Given any $x \in \mathcal{X}$, Lebesgue almost any $y \in \mathcal{X}$ satisfies one of the two following conditions:

(i) either $f(y, x) < 0$ (so that $(\bar{n}(x), 0)$ is stable),
(ii) or $f(y, x) > 0$, $f(x, y) < 0$ and any solution to (4.7) with initial condition with positive coordinates in a given neighborhood of $(\bar{n}(x), 0)$ converges to $(0, \bar{n}(y))$.

In the case of linear logistic density-dependence introduced in Section 2.2 ($b(x, \zeta) = b(x)$ and $d(x, \zeta) = d(x) + \alpha(x)\zeta$), the equilibrium monomorphic density $\bar{n}(x)$ writes $(b(x) - d(x))/\alpha(x)U(0)$ and the condition (H6)-(ii) is actually equivalent to $f(y, x) > 0$ and $f(x, y) < 0$ (see [6]).

Our convergence result writes

Theorem 5.1. *Assume* (H), (H1), (H5) *and* (H6). *Given* $x \in \mathcal{X}$, $\gamma > 0$ *and a sequence of* $\mathbb{N}$*-valued random variables* $(\gamma_K)_{K \in \mathbb{N}}$, *such that* γ_K/K *is bounded in* $\mathbb{L}^1$ *and converges in law to* γ, *consider the process* $(X_t^K, t \geq 0)$ *of Section* 4 *generated by* (4.1) *with initial state* $\frac{\gamma_K}{K}\delta_x$. *Then, for any* $n \geq 1$, $\varepsilon > 0$ *and* $0 < t_1 < t_2 < \ldots < t_n < \infty$, *and for any measurable subsets* $\Gamma_1, \ldots, \Gamma_n$ *of* $\mathcal{X}$,

$$\lim_{K\to+\infty} P\big(\forall i \in \{1,\dots,n\},\ \exists x_i \in \Gamma_i\ :\ \mathrm{Supp}(X^K_{t_i/Ku_K}) = \{x_i\}$$
$$\text{and } |\langle X^K_{t_i/Ku_K}, \mathbf{1}\rangle - \bar{n}(x_i)| < \varepsilon\big) = P(\forall i \in \{1,\dots,n\},\ Y_{t_i} \in \Gamma_i) \quad (5.4)$$

where for any $\nu \in M_F(\mathcal{X})$, $\mathrm{Supp}(\nu)$ *is the support of* ν *and* $(Y_t, t \geq 0)$ *is a Markov jump process with initial state* x *generated by*

$$A\varphi(x) = \int_{\mathbb{R}^d} (\varphi(y) - \varphi(x))g(y,x)M(x,y)dy$$

where

$$g(y,x) = \mu(x)b(x, V(0)\bar{n}(x))\bar{n}(x)\frac{[f(y,x)]_+}{b(y, V(y-x)\bar{n}(x))}, \tag{5.5}$$

and $[\cdot]_+$ *denotes the positive part.*

Corollary 5.2. *With the same notation and assumptions as in Theorem* 5.1*, assuming moreover that* γ_K/K *is bounded in* $\mathbb{L}^q$ *for some* $q > 1$*, the process* $(X^K_{t/Ku_K}, t \geq 0)$ *converges when* $K \to +\infty$*, in the sense of the finite dimensional distributions for the topology on* $M_F(\mathcal{X})$ *induced by the functions* $\nu \mapsto \langle \nu, f\rangle$ *with* f *bounded and measurable on* $\mathcal{X}$*, to the process* $(Z_t, t \geq 0)$ *defined by*

$$Z_t = \begin{cases} \gamma\delta_x & \text{if } t = 0 \\ \bar{n}(Y_t)\delta_{Y_t} & \text{if } t > 0. \end{cases}$$

This corollary follows from the following long time moment estimates.

Lemma 5.3. *Under* (H), (H1), (H5)(iii) (5.2) *and* (iv)*, and if* $\sup_{K\geq 1} E(\langle X^K_0, 1\rangle^q) < +\infty$ *for some* $q \geq 1$*, then*

$$\sup_{K\geq 1}\ \sup_{t\geq 0} E\big(\langle X^K_t, \mathbf{1}\rangle^q\big) < +\infty,$$

and therefore, if $q > 1$*, the family of random variables* $\{\langle X^K_t, \mathbf{1}\rangle\}_{\{K\geq 1,\, t\geq 0\}}$ *is uniformly integrable.*

Proof of Lemma 5.3. Observe that, if we replace $b(x, V*\nu)$ by $\bar{b}$ and $d(x, U*\nu)$ by $g(\underline{U}\langle \nu, \mathbf{1}\rangle)$ where $g(\zeta) := \inf_{x\in\mathcal{X}} d(x,\zeta)$ in the indicator functions of each terms of the construction (2.2) of the process X^K_t, we can stochastically dominate the population size $\langle X^K_t, \mathbf{1}\rangle$ by a birth and death Markov process $(Z^K_t)_{t\geq 0}$ with initial state $Z^K_0 = \langle X^K_0, 1\rangle$ and transition rates

$$\begin{array}{ll} i\bar{b} & \text{from } i/K \text{ to } (i+1)/K, \\ ig(\underline{U}\frac{i}{K}) & \text{from } i/K \text{ to } (i-1)/K. \end{array}$$

Therefore, it suffices to prove that $\sup_{K\geq 0}\sup_{t\geq 0} E((Z^K_t)^q) < +\infty$.

Let us define $p_t^k = P(Z_t^K = k/K)$. Then

$$\begin{aligned}
\frac{d}{dt}E((Z_t^K)^q) &= \sum_{k\geq 1}\left(\frac{k}{K}\right)^q \frac{dp_t^k}{dt} \\
&= \frac{1}{K^q}\sum_{k\geq 1} k^q \Bigg[\bar{b}(k-1)p_t^{k-1} + (k+1)g\left(\underline{U}\frac{k+1}{K}\right)p_t^{k+1} \\
&\qquad\qquad\qquad\qquad - k\left(\bar{b} + g\left(\underline{U}\frac{k}{K}\right)\right)p_t^k\Bigg] \\
&= \frac{1}{K^q}\sum_{k\geq 1}\left[\bar{b}\left(\left(1+\frac{1}{k}\right)^q - 1\right) + g\left(\underline{U}\frac{k}{K}\right)\left(\left(1-\frac{1}{k}\right)^q - 1\right)\right]k^{q+1}p_t^k.
\end{aligned}$$

Now, by (H5) (iii) (5.2), $g(\alpha) \to +\infty$ when $\alpha \to +\infty$, so there exists α_0 such that, for any $\alpha \geq \alpha_0$, $g(\underline{U}\alpha) \geq 2\bar{b}$. Therefore, for $k \geq K\alpha_0$, $\bar{b}((1+1/k)^q - 1) + g(\underline{U}k/K)((1-1/k)^q - 1) \leq -\bar{b}[3 - 2(1-1/k)^q - (1+1/k)^q]$, the term on the right-hand side being equivalent to $-\bar{b}q/k$. Therefore, enlarging α_0 if necessary and using in the first inequality the facts that $(1+\alpha)^q - 1 \leq \alpha(2^q - 1)$ and $(1-\alpha)^q - 1 \leq 0$ for any $\alpha \in [0,1]$, we can write

$$\begin{aligned}
\frac{d}{dt}E((Z_t^K)^q) &\leq \sum_{k=1}^{\lceil K\alpha_0\rceil - 1} \bar{b}(2^q - 1)\left(\frac{k}{K}\right)^q p_t^k - \sum_{k\geq\lceil K\alpha_0\rceil}\frac{\bar{b}q}{2}\left(\frac{k}{K}\right)^q p_t^k \\
&\leq \sum_{k=1}^{\lceil K\alpha_0\rceil - 1}\bar{b}(q/2 + 2^q - 1)\alpha_0^q\, p_t^k - \frac{\bar{b}q}{2}E((Z_t^K)^q) \leq \frac{\bar{b}q}{2}[C - E((Z_t^K)^q)],
\end{aligned}$$

where $C = (1 + 2(2^q-1)/q)\alpha_0^q$. This differential inequality solves as

$$E((Z_t^K)^q) \leq C + [E((Z_0^K)^q) - C]e^{-\bar{b}qt/2},$$

which gives the required uniform bound. □

Proof of Corollary 5.2. Let Γ be a measurable subset of $\mathcal{X}$. Let us prove that

$$\lim_{K\to+\infty} E\big[\langle X_{t/Ku_K}^K, \mathbf{1}_\Gamma\rangle\big] = E\big[\bar{n}(Y_t)\mathbf{1}_{Y_t\in\Gamma}\big]. \tag{5.6}$$

By (H5)-(iii)-(5.2), there exists $\zeta_0 > 0$ such that for any $\zeta > \zeta_0$ and $x \in \mathcal{X}$, $d(x,\zeta) > \bar{b}$. Therefore, by (H5)-(iv), for any $x \in \mathcal{X}$, $\bar{n}(x) \in [0, \zeta_0/\underline{U}]$. Fix $\varepsilon > 0$, and write $[0,\zeta_0/\underline{U}] \subset \cup_{i=1}^p I_i$, where p is the integer part of $\zeta_0/(\underline{U}\varepsilon)$, and $I_i = [(i-1)\varepsilon, i\varepsilon[$. Define $\Gamma_i = \{x \in \mathcal{X} : \bar{n}(x) \in I_i\}$ for $1 \leq i \leq p$, and apply (5.4) to the sets $\Gamma\cap\Gamma_1, \dots, \Gamma\cap\Gamma_p$ with $n = 1$, $t_1 = t$ and the constant ε above. Then, by

Lemma 5.3, for some constant $C>0$ and for sufficiently large K,

$$\begin{aligned}
\limsup_{K\to+\infty} E\big[\langle X^K_{t/Ku_K},\mathbf{1}_\Gamma\rangle\big] &\leq \limsup_{K\to+\infty} E\big[\langle X^K_{t/Ku_K},\mathbf{1}_\Gamma\rangle \mathbf{1}_{\langle X^K_{t/Ku_K},\mathbf{1}\rangle\leq C}\big]+\varepsilon \\
&\leq \sum_{i=1}^p \limsup_{K\to+\infty} E\big[\langle X^K_{t/Ku_K},\mathbf{1}_{\Gamma\cap\Gamma_i}\rangle \mathbf{1}_{\langle X^K_{t/Ku_K},\mathbf{1}\rangle\leq C}\big]+\varepsilon \\
&\leq \sum_{i=1}^p (i+1)\varepsilon P(Y_t\in\Gamma\cap\Gamma_i)+\varepsilon \\
&\leq \sum_{i=1}^p \big(E\big[\bar n(Y_t)\mathbf{1}_{X_t\in\Gamma\cap\Gamma_i}\big]+2\varepsilon P(Y_t\in\Gamma_i)\big)+\varepsilon \\
&\leq E\big[\bar n(Y_t)\mathbf{1}_{Y_t\in\Gamma}\big]+3\varepsilon.
\end{aligned}$$

A similar estimate for the *lim inf* ends the proof of (5.6), which implies the convergence of one-dimensional laws for the required topology.

The same method gives easily the required limit when we consider a finite number of times $t_1,\dots,t_n$. □

Observe that the fact that the limit process is not right-continuous prevents the possibility to obtain a convergence for the Skorohod topology on $\mathbb{D}([0,T], M_F(\mathcal{X}))$.

5.2. Idea of the proof

Theorem 5.1 can be proved in a similar way as in Champagnat [6]. Let us give an idea of the method in order to explain the assumptions, the various parameters appearing in Theorem 5.1 and the tools involved in the proof. It is based on two ingredients: the study of a monomorphic population before the first mutation, and the study of the invasion of a single mutant individual in this population.

1) The first part obtains from large deviation results for the convergence of X^K_t to $n_t(x)\delta_x$ when the initial population is monomorphic with trait x, where $n_t(x)$ satisfies (4.6). Any positive solution to (4.6) converges to $\bar n(x)$ when $t\to+\infty$, and hence reaches a given neighborhood of $\bar n(x)$ in finite time, i.e., on an infinitesimal time scale with respect to the mutation time scale. Large deviations theory allows us to show that the exit time of $\langle X^K_t,\mathbf{1}\rangle$ from this neighborhood behaves as $\exp(KC)$ for some $C>0$ (problem of exit from a domain, Freidlin and Wentzell [16]). Thanks to the right part of Assumption (5.1), we can prove that, with high probability, $\langle X^K_t,\mathbf{1}\rangle$ is close to $\bar n(x)$ when the first mutation occurs. Therefore, the total mutation rate is close to $u_K\mu(x)K\bar n(x)b(x,V(0)\bar n(x))$ and so, on the mutation time scale t/Ku_K, the rate of mutation is close to $\bar n(x)\mu(x)b(x,V(0)\bar n(x))$, which explains the left part of the right-hand side of (5.5). This argument can be made rigorous using stochastic domination results similar to the one used at the beginning of the proof of Lemma 5.3, and leads to the following result:

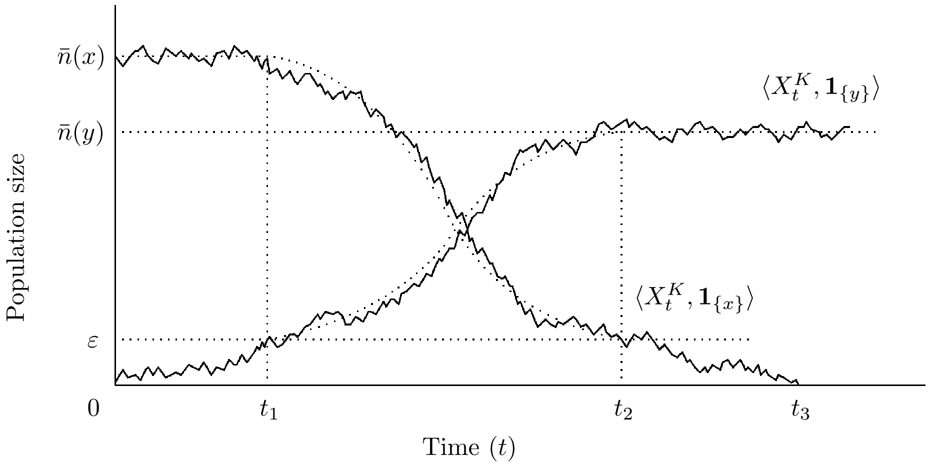


FIGURE 3. The three steps of the invasion and fixation of a mutant trait y in a monomorphic population with trait x. Plain curves represent the resident and mutant densities $\langle X_t^K, \mathbf{1}_{\{x\}}\rangle$ and $\langle X_t^K, \mathbf{1}_{\{y\}}\rangle$, respectively. Dotted curves represent the solution of Eq. (4.7) with initial state $n_0(x) = \bar{n}(x)$ and $n_0(y) = \varepsilon$.

Lemma 5.4. *Let τ_1 denote the first mutation time and $\mathbf{P}^K_{X_0^K}$ the law of X^K with initial state X_0^K. Given $x \in \mathcal{X}$ and a sequence of integers $(z_K)_{K\geq 1}$ such that $z_K/K \to z > 0$,*

(a) *For any $\varepsilon > 0$,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \log K, \sup_{t\in[\log K,\tau_1]} |\langle X_t^K, \mathbf{1}\rangle - \bar{n}(x)| > \varepsilon\right) = 0$$

and

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}(\tau_1 < \log K) = 0.$$

In particular, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$, $X^K_{\log K} \to \bar{n}(x)\delta_x$ and $X^K_{\tau_1-} \to \bar{n}(x)\delta_x$ in probability.

(b) *For any $t > 0$,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) = \exp\big(-\beta(x)t\big),$$

where $\beta(x) = \mu(x)\bar{n}(x)b(x, V(0)\bar{n}(x))$.

2) The study of the invasion of a mutant individual with trait y can be divided in three steps represented in Fig. 3.

Firstly, the invasion of the mutant (between 0 and t_1 in Fig. 3) can be defined as the growth of the mutant density $\langle X_t^K, \mathbf{1}_{\{y\}}\rangle$ from $1/K$ (one individual) to a fixed small level ε (εK individuals). As long as the mutant density is small, the dynamics of the resident density $\langle X_t^K, \mathbf{1}_{\{x\}}\rangle$ is close to the one it followed before the mutation, so it is close to $\bar{n}(x)$ with high probability. Therefore, between 0 and t_1, the birth and death rates of an individual with trait y are close to $b(y, V(y-x)\bar{n}(x))$ and $d(y, U(y-x)\bar{n}(x))$, respectively. Therefore, the number of mutant individuals is close to a binary branching process with the parameters above. When $K \to +\infty$, the probability that such a branching process reaches level εK is close to its survival probability, which writes $[f(y,x)]_+/b(y, V(y-x)\bar{n}(x))$. This gives the second part of the right-hand side of (5.5).

Secondly, once the invasion succeeded (which is possible only if $f(y,x) > 0$), the dynamics of the densities of traits x and y are close to the solution to the dimorphic logistic equation (4.7) with initial state $(\bar{n}(x), \varepsilon)$, represented in dotted curves between t_1 and t_2 in Fig. 3. Because of Assumption (H6), the resident density can be proved to reach level ε with high probability (at time t_2 in Fig. 3).

Finally, a similar argument as in the first step above allows us to prove that the resident population density $\langle X_t^K, \mathbf{1}_{\{x\}}\rangle$ follows approximately a binary branching process with birth rate $b(y, V(x-y)\bar{n}(y))$ and death rate $d(y, U(x-y)\bar{n}(y))$. Since $f(x,y) < 0$ by Assumption (H6), this is a sub-critical branching process, and therefore, the resident trait x disappears in finite time t_3 with high probability.

We can show, using results on branching processes, that t_1 and $t_3 - t_2$ are of order $\log K$, whereas $t_2 - t_1$ depends only on ε. Therefore, the left part of (5.1) ensures that the three steps of the invasion are completed before the next mutation, with high probability. The previous heuristics can be made rigorous using further comparison results, and leads to the following result.

Lemma 5.5. *Assume that the initial population is made of individuals with traits x and y satisfying assumption* (H6) (i) *or* (ii). *Let θ_0 denote the first time when the population gets monomorphic, and V_0 the remaining trait. Let $(z_K)_{K\geq 1}$ be a sequence of integers such that $z_K/K \to \bar{n}(x)$. Then,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}(V_0 = y) = \frac{[f(y,x)]_+}{b(y, V(y-x)\bar{n}(x))},$$

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}(V_0 = x) = 1 - \frac{[f(y,x)]_+}{b(y, V(y-x)\bar{n}(x))},$$

$$\forall \eta > 0, \quad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}\left(\theta_0 > \frac{\eta}{Ku_K} \wedge \tau_1\right) = 0,$$

$$\textit{and} \quad \forall \varepsilon > 0, \quad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}\left(|\langle X^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}(V_0)| < \varepsilon\right) = 1,$$

where $f(y,x)$ has been defined in (5.3).

Once these lemmas are proved, the proof can be completed by observing that the generator A of the process $(Y_t, t \geq 0)$ of Theorem 5.1 can be written as

$$A\varphi(x) = \int_{\mathbb{R}^l} (\varphi(y) - \varphi(x))\beta(x)\kappa(x, dy), \tag{5.7}$$

where $\beta(x)$ has been defined in Lemma 5.4 and the probability measure $\kappa(x, dh)$ is defined by

$$\kappa(x,dy) = \left(1 - \int_{\mathbb{R}^l} \frac{[f(z,x)]_+}{b(z,V(z-x)\bar{n}(x))} M(x,z)dz\right) \delta_x(dy) + \frac{[f(y,x)]_+}{b(y,V(y-x)\bar{n}(x))} M(x,y)dy. \quad (5.8)$$

This means that the process Y with initial state x can be constructed as follows: let $(M(k), k = 0, 1, 2, \ldots)$ be a Markov chain in $\mathcal{X}$ with initial state x and with transition kernel $\kappa(x, dy)$, and let $(N(t), t \geq 0)$ be an independent standard Poisson process. Let also $(T_n)_{n\geq 1}$ denote the sequence of jump times of the Poisson process N. Then, the process $(Y_t, t \geq 0)$ defined by

$$Y_t := M\left(N\left(\int_0^t \beta(Y_s)ds\right)\right)$$

is a Markov process with infinitesimal generator (5.7) (cf. [13, Chapter 6]).

Let P_x denote its law, and define $(S_n)_{n\geq 1}$ by $T_n = \int_0^{S_n} \beta(Y_s)ds$. Observe that any jump of the process Y occurs at some time S_n, but that all S_n may not be effective jump times for Y, because of the Dirac mass at x appearing in (5.8).

Fix $t > 0$, $x \in \mathcal{X}$ and a measurable subset Γ of $\mathcal{X}$. Under P_x, S_1 and Y_{S_1} are independent, S_1 is an exponential random variable with parameter $\beta(x)$, and Y_{S_1} has law $\kappa(x, \cdot)$. Therefore, for any $n \geq 1$, the strong Markov property applied to Y at time S_1 yields

$$P_x(S_n \leq t < S_{n+1},\ Y_t \in \Gamma) = \int_0^t \beta(x)e^{-\beta(x)s} \int_{\mathbb{R}^l} \mathbf{P}_y(S_{n-1} \leq t-s < S_n,\ Y_{t-s} \in \Gamma)\kappa(x,dy)ds \quad (5.9)$$

and

$$P_x(0 \leq t < S_1,\ Y_t \in \Gamma) = \mathbf{1}_{\{x\in\Gamma\}} e^{-\beta(x)t}. \quad (5.10)$$

Using the Markov property at time τ_1 and Lemmas 5.4 and 5.5, we can prove that, when we replace S_n by the n-th mutation time of X^K_{t/Ku_K} and Y_t by the support of X^K_{t/Ku_K} (when it is a singleton) in the LHS of (5.9) and (5.10), the same relations hold in the limit $K \to +\infty$. Therefore, Theorem 5.1 is proved for one-dimensional time marginals. A similar method generalizes to finite dimensional laws.

References

[1] D. Aldous, *Stopping times and tightness*, Ann. Probab., **6** (1978), 335–340.

[2] B. Bolker and S. W. Pacala, *Using moment equations to understand stochastically driven spatial pattern formation in ecological systems*, Theor. Pop. Biol., **52** (1997), 179–197.

[3] B. M. Bolker and S. W. Pacala, *Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal*, Am. Nat., **153** (1999), 575–602.

[4] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation*, John Wiley & Sons, Chichester, 2000.

[5] W. A. Calder III, *Size, Function and Life History*, Harvard University Press, Cambridge, 1984.

[6] N. Champagnat, *A microscopic interpretation for adaptive dynamics trait substitution sequence models*, Stochastic Process. Appl., **116** (2006), 1127–1160.

[7] N. Champagnat, R. Ferrière, and S. Méléard, *Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models*, Theoretical Population Biology, **69** (2006), 297–321.

[8] E. L. Charnov, *Life History Invariants*, Oxford University Press, Oxford, 1993.

[9] L. Desvillettes, C. Prevost, and R. Ferriere, *Infinite dimensional reaction-diffusion for evolutionary population dynamics*, Preprint CMLA, École Normale Superieure de Cachan, 2004.

[10] U. Dieckmann and R. Law, *The dynamical theory of coevolution: A derivation from stochastic ecological processes*, J. Math. Biol., **34** (1996), 579–612.

[11] U. Dieckmann and R. Law, *Relaxation projections and the method of moments*, in: U. Dieckmann, R. Law, and J. A. J. Metz, Editors, The Geometry of Ecological Interactions: Symplifying Spatial Complexity, Cambridge University Press, Cambridge, (2000), 412–455.

[12] A. Etheridge, *Survival and extinction in a locally regulated population*, Ann. Appl. Probab., **14** (2004), 188–214.

[13] S. N. Ethier and T. G. Kurtz, *Markov Processes, Characterization and Convergence*, John Wiley & Sons, New York, 1986.

[14] S. N. Evans and E. A. Perkins, *Measure-valued branching diffusions with singular interactions*, Canad. J. Math., **46** (1994), 120–168.

[15] N. Fournier and S. Méléard, *A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, Ann. Appl. Probab., **14** (2004), 1880–1919.

[16] M. I. Freidlin and A. D. Wentzel, *Random Perturbations of Dynamical Systems*, Springer-Verlag, Berlin, 1984.

[17] A. Joffe and M. Métivier, *Weak convergence of sequences of semimartingales with applications to multitype branching processes*, Adv. Appl. Probab., **18** (1986), 20–65.

[18] E. Kisdi, *Evolutionary branching under asymmetric competition*, J. Theor. Biol., **197** (1999), 149–162.

[19] R. Law, D. J. Murrell, and U. Dieckmann, *Population growth in space and time: spatial logistic equations*, Ecology, **84** (2003), 252–262.

[20] S. Méléard and S. Roelly, *Sur les convergences étroite ou vague de processus à valeurs mesures*, C.R. Acad. Sci. Paris Sér. I Math., **317** (1993), 785–788.

[21] J. A. J. Metz, R. M. Nisbet, and S. A. H. Geritz, *How should we define fitness for general ecological scenarios*, Trends Ecol. Evol., **7** (1992), 198–202.

[22] J. A. J. Metz, S. A. H. Geritz, G. Meszeena, F. A. J. Jacobs, and J. S. van Heerwaarden, *Adaptive Dynamics, a geometrical study of the consequences of nearly faithful reproduction*, in S. J. van Strien and S. M. Verduyn Lunel, Editors, Stochastic and Spatial Structures of Dynamical Systems, North Holland, Amsterdam, (1996), 183–231.

[23] S. Roelly-Coppoletta, *A criterion of convergence of measure-valued processes: application to measure branching processes*, Stoch. Stoch. Rep., **17** (1986), 43–65.

Nicolas Champagnat
Laboratoire d'Écologie, Équipe Éco-Évolution Mathématique
École Normale Supérieure
46 rue d'Ulm, 75230 Paris cedex 05
France
and
Équipe MODALX
Université Paris 10
200 avenue de la République
92001 Nanterre Cedex
France
e-mail: `champagn@clipper.ens.fr`

Régis Ferrière
Laboratoire d'Écologie, Équipe Éco-Évolution Mathématique
École Normale Supérieure
46 rue d'Ulm, 75230 Paris cedex 05
France
and
Department of Ecology and Evolutionary Biology
University of Arizona
Tucson AZ 85721
USA
e-mail: `ferriere@biologie.ens.fr`

Sylvie Méléard
Équipe MODALX
Université Paris 10
200 avenue de la République
92001 Nanterre Cedex
France
e-mail: `sylm@ccr.jussieu.fr`
`sylvie.meleard@u-paris10.fr`

Progress in Probability, Vol. 59, 115–122

A Note on Evolution Systems of Measures for Time-Dependent Stochastic Differential Equations

Giuseppe Da Prato and Michael Röckner

Abstract. We consider a stochastic equation in $\mathbb{R}^n$ with time-dependent coefficients assuming that it has a unique solution and denote by $P_{s,t}, s < t$ the corresponding transition semigroup. Then we consider a family of measures $(\nu_t)_{t\in\mathbb{R}}$ such that $\int_{\mathbb{R}^d} P_{s,t}\varphi(x)\nu_s(dx) = \int_{\mathbb{R}^d} \varphi(x)\nu_t(dx), s \leq t$, for all continuous and bounded functions φ. The family $(\nu_t)_{t\in\mathbb{R}}$ is called an *evolution system of measures indexed by* $\mathbb{R}$. It plays the role of a probability invariant measure for autonomous systems. In this paper we generalize the Krylov–Bogoliubov criterion to prove the existence of an evolution system of measures. Moreover, we study some properties of the corresponding Kolmogorov operator proving in particular that it is dissipative with respect to the measure $\nu(dt, dx) = \nu_t(dx)dt$.

Mathematics Subject Classification (2000). 60H10, 37L40, 37A25.

Keywords. Stochastic differential equations, evolution systems of measures, non-autonomous systems.

1. Notation

We fix $d \in \mathbb{N}$. We shall use the following notation.

The first author would like to thank the University of Bielefeld for its kind hospitality and financial support. This work was also supported by the research program "Equazioni di Kolmogorov" from the Italian "Ministero della Ricerca Scientifica e Tecnologica".
The second-named author would like to thank the Scuola Normale Superiore for a very pleasant stay in Pisa during which most of this work was done. Financial support of the SNS as well as of the DFG-Forschergruppe "Spectral Analysis, Asymptotic Distributions, and Stochastic Dynamics" is gratefully acknowledged.

- $B_b(\mathbb{R}^n)$ is the set of all bounded Borel functions in $\mathbb{R}^n$. It is endowed with the norm
$$\|\varphi\|_0 = \sup_{x\in\mathbb{R}^d} |\varphi(x)|, \quad \varphi \in B_b(\mathbb{R}^n).$$
- $C_b(\mathbb{R}^n)$ is the subspace of $B_b(\mathbb{R}^n)$ of all uniformly continuous functions.
- $\mathcal{B}(\mathbb{R}^n)$ is the σ-algebra of all Borel subsets of $\mathbb{R}^n$.
- $\mathcal{P}(\mathbb{R}^n)$ is the set of all probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
- $C_b^*(\mathbb{R}^n)$ is the topological dual of $C_b(\mathbb{R}^n)$.
  We shall denote by $\langle \cdot,\cdot\rangle$ the duality between $C_b(\mathbb{R}^n)$ and $C_b^*(\mathbb{R}^n)$.

We shall identify $\mathcal{P}(\mathbb{R}^n)$ with a closed convex subset of $C_b^*(\mathbb{R}^n)$ by the mapping
$$\mu \in \mathcal{P}(\mathbb{R}^n) \mapsto F_\mu \in C_b^*(\mathbb{R}^n),$$
where
$$F_\mu(\varphi) = \int_{\mathbb{R}^n} \varphi(y)\mu(dy), \quad \varphi \in C_b(\mathbb{R}^n).$$
We shall write $F_\mu = \mu$ in what follows.

We are concerned with the following stochastic differential equation in $\mathbb{R}^d$,
$$\begin{cases} dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t), \quad t \ge s, \\ X(s) = x, \end{cases} \tag{1.1}$$
where

- $b \in C(\mathbb{R}\times\mathbb{R}^d; \mathbb{R}^d)$,
- $\sigma \in C(\mathbb{R}\times\mathbb{R}^d; L(\mathbb{R}^m, \mathbb{R}^d))$,
- W is a standard Brownian motion in $\mathbb{R}$ taking values in $\mathbb{R}^m$.

We shall assume that problem (1.1) has a unique solution $X(t, s, x)$. We denote by $P_{s,t}$ the transition semigroup,
$$P_{s,t}\varphi(x) = \mathbb{E}[\varphi(X(t, s, x))], \quad s \le t,\ \varphi \in B_b(\mathbb{R}^d),\ x \in \mathbb{R}^d,$$
and by $\pi_{s,t}(x, \cdot)$ the law of $X(t, s, x)$. So,
$$P_{s,t}\varphi(x) = \int_{\mathbb{R}^d} \varphi(y)\pi_{s,t}(x, dy)$$
and
$$P_{s,t}\mathbf{1}_A = \pi_{s,t}(x, A), \quad A \in \mathcal{B}(H).$$
Moreover, the *semigroup law* holds,
$$P_{s,t}P_{t,u} = P_{s,u}, \quad s \le t \le u.$$
We shall denote by $P_{s,t}^*$ the transpose semigroup in $C_b(\mathbb{R}^n)$. It is clear that
$$P_{s,t}^*\mu \in \mathcal{P}(\mathbb{R}^n) \quad \text{for all } \mu \in \mathcal{P}(\mathbb{R}^n),$$
and that
$$P_{t,u}^*P_{s,t}^* = P_{s,u}^*, \quad s \le t \le u.$$

If the coefficients of (1.1) are regular, then by the Itô formula it follows that $u(t,s,x) = P_{s,t}\varphi(x)$ is the solution of the backward *Kolmogorov equation*

$$\begin{cases} D_s u(t,s,x) + K(s)u(t,s,x) = 0, \\ u(t,t,x) = \varphi(x), \end{cases} \tag{1.2}$$

where

$$K(s)\varphi = \frac{1}{2}\,\mathrm{Tr}\,[\sigma(s,x)\sigma^*(s,x)D^2\varphi] + \langle b(s,x), D\varphi\rangle.$$

Let us list some useful properties of $P_{s,t}$ and $P^*_{s,t}$ for $s \le t$, whose proofs are well known,

$$D_s P_{s,t}\varphi = -K(s)P_{s,t}\varphi, \quad \varphi \in C_b^2(\mathbb{R}^d),$$

and

$$D_t P_{s,t}\varphi = P_{s,t}K(t)\varphi, \quad \varphi \in C_b^2(\mathbb{R}^d).$$

Finally,

$$D_s P^*_{s,t} = -P^*_{s,t}K^*(s)$$

and

$$D_t P^*_{s,t} = K^*(t)P^*_{s,t}.$$

2. Evolution system of measures indexed by $\mathbb{R}$

A mapping $\mathbb{R} \to \mathcal{P}(\mathbb{R}^n),\ t \to \nu_t$ is called an *evolution system of measures indexed by* $\mathbb{R}$ if

$$\int_{\mathbb{R}^d} P_{s,t}\varphi(x)\nu_s(dx) = \int_{\mathbb{R}^d} \varphi(x)\nu_t(dx), \quad s \le t,\ \varphi \in C_b(H). \tag{2.1}$$

(2.1) is equivalent to

$$P^*_{s,t}\nu_s = \nu_t, \quad s \le t.$$

So, $\nu_t,\ t \ge 0$, is a measure-valued solution of the Kolmogorov equation (1.2) with values measures in the sense of the paper [1]. The difference with respect to that paper is that ν_t is here defined for t in the whole $\mathbb{R}$.

A solution of (2.1), that is, an evolution system of measures indexed by $\mathbb{R}$, is the natural generalization of the concept of invariant measure for an autonomous system.

If the coefficients of (1.1) are regular, identity (2.1) is equivalent to

$$D_t\nu_t = K^*(t)\nu_t, \quad t \in \mathbb{R}. \tag{2.2}$$

We set

$$Lu(t,x) = D_t u(t,x) + K(t)u(t,x), \quad t \in \mathbb{R},\ x \in \mathbb{R}^d,$$

and consider the formal adjoint

$$L^*\nu(dt,dx) = -D_t\nu(dt,dx) + K^*(t)\nu(dt,dx).$$

So, (2.1) implies (and in regular cases is equivalent to)

$$L^*\nu = 0,$$

where $\nu(dt,dx) = \nu_t dx$.

The evolution system of measures ν_t is called *strongly mixing* if

$$\lim_{s\to-\infty} P_{s,t}\varphi(x) = \int_{\mathbb{R}^d} \varphi(x)\nu_t(dx) \quad \varphi \in C_b(H),\ t \in \mathbb{R}.$$

Example. We take $m = d$, $b(t,x) = A(t)x$ and $\sigma(t,x) = I$, where $A(t) \in L(\mathbb{R}^d)$. We denote by $U(t,s)$ the evolution operator corresponding to $A(t)$. We assume that there are $\omega > 0$ and $M > 0$ such that (see [4] for a sufficient condition based on the Floquet exponents of $A(t)$)

$$\|U(t,s\| \le Me^{-\omega(t-s)}, \quad t \ge s.$$

Then we have

$$X(t,s,x) = U(t,s)x + \int_s^t U(t,\tau)dW(\tau),$$

and so

$$P_{s,t}\varphi(x) = \int_{\mathbb{R}^d} \varphi(U(t,s)x + y)N_{Q(t,s)}(dy) \tag{2.3}$$

where $N_{Q(t,s)}$ is the Gaussian measure in $L(\mathbb{R}^d)$ of mean 0 and covariance operator

$$Q(t,s) = \int_s^t U(t,\tau)U^*(t,\tau)d\tau.$$

By (2.3) it follows that

$$\lim_{s\to-\infty} P_{s,t}\varphi(x) = \int_H \varphi(y)N_{Q(t,-\infty)}(dy).$$

Thus, setting

$$\nu_t = N_{Q(t,-\infty)}, \quad Q(t,-\infty) = \int_{-\infty}^t U(t,\tau)U^*(t,\tau)d\tau,$$

we see that ν_t is a strongly mixing evolution system of measures.

3. Existence of evolution sytems of measures

There are at least (to our knowledge) two methods to prove the existence of evolution sytems of measures. For the first, based on the use of Lyapunov functions, see [2]. The other one is a straightforward generalization of the Krylov–Bogoliubov theorem. We shall briefly present the latter.

We assume here that $P_{s,t}\varphi \in C_b(\mathbb{R}^d)$ for any $\varphi \in C_b(\mathbb{R}^d)$, that is, $P_{s,t}$ is Feller.

Fix $x_0 \in \mathbb{R}^d$. For any $T > 0$, $t \ge -T$, set

$$\mu_{T,t}(E) = \frac{1}{t+T} \int_{-T}^t \pi_{s,t}(x_0,E)ds, \quad E \in \mathcal{B}(\mathbb{R}^d),\ T > 0.$$

Theorem 3.1. *Let $x_0 \in \mathbb{R}^d$ be fixed. Assume that for any $n \in \mathbb{N}$ the set $(\mu_{T,-n})_{T>0}$ is tight. Then there exist evolution sytems of measures for P_t.*

Proof. By the Prokhorov theorem and a diagonal argument there exists a sequence $T_n \uparrow \infty$ such that for all $n \in \mathbb{N}$ the weak limit

$$\mu_{-n} := \lim_{N\to\infty} \mu_{T_N,-n}$$

exists.

Let $t \in \mathbb{R}$ and choose $n \in \mathbb{N}$ such that $t > -n$. Define

$$\nu_t := P^*_{-n,t}\mu_{-n}.$$

Note that this definition is indeed independent of n, since because each $P_{s,t}$ is Feller we have for every $\varphi \in C_b(\mathbb{R}^d)$

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(x)\nu_t(dx) &= \int_{\mathbb{R}^d} P_{-n,t}\varphi(x)\mu_{-n}(dx) \\
&= \lim_{N\to\infty} \int_{\mathbb{R}^d} P_{-n,t}\varphi(x)\mu_{T_n,-n}(dx) \\
&= \lim_{N\to\infty} \frac{1}{-n+T_N}\int_{-T_N}^{-n} P_{s,-n}(P_{-n,t}\varphi)(x_0)ds \\
&= \lim_{N\to\infty} \frac{1}{T_N}\int_{-T_N}^{-n} P_{s,t}\varphi(x_0)ds,
\end{aligned}$$

which is obviously independent of n, $n < t$. Now for $s \le t$ we have for any $n \in \mathbb{N}$, $-n \le s$,

$$\begin{aligned}
P^*_{s,t}\mu_{-n} &= P^*_{s,t}P^*_{-n,s}\mu_{-n} \\
&= (P_{-n,s}P_{s,t})^*\mu_{-n} \\
&= P^*_{-n,t}\mu_{-n} \\
&= \nu_t.
\end{aligned}$$

So, $(\nu_t)_{t\in\mathbb{R}}$ is an evolution system of measures for $(P_{s,t})_{s\le t}$. □

4. The equation $Lu - \lambda u = f$

We are here concerned with the equation

$$D_t u(t,x) + K(t)u(t,x) - \lambda u(t,x) = f(t,x), \quad x \in \mathbb{R}^d,\ t \ge 0, \tag{4.1}$$

where $f \in C_b(\mathbb{R} \times \mathbb{R}^d)$ and $\lambda > 0$. By a *mild solution* of (4.1) we mean a function $u \in C_b(\mathbb{R} \times \mathbb{R}^d)$ such that for any $T \in \mathbb{R}$,

$$u(t,x) = e^{-\lambda(T-t)}P_{t,T}u(T,x) - \int_t^T e^{-\lambda(s-t)}P_{t,s}f(s,x)ds, \quad t \le T,\ x \in \mathbb{R}^d. \tag{4.2}$$

We assume that coefficients of (1.1) are regular. Since $\|P_{t,s}f\|_0 \le \|f\|_0$ for all $f \in C_b(\mathbb{R}^d)$, the following is straightforward. We include the proof for the reader's convenience.

Proposition 4.1. *There exists a unique solution of* (4.1) *given by*

$$u(t,x) = -\int_t^{+\infty} e^{-\lambda(s-t)}P_{t,s}f(s,x)ds, \quad t \in \mathbb{R},\ x \in \mathbb{R}^d. \tag{4.3}$$

Proof. *Existence.* We check that the function u given by (4.3) is a solution of (4.2). Let in fact $T \in \mathbb{R}$. Then we have,

$$u(T,x) = -\int_T^{+\infty} e^{-\lambda(s-T)} P_{T,s} f(s,x) ds, \quad x \in \mathbb{R}^d.$$

Consequently,

$$\begin{aligned} e^{-\lambda(T-t)} P_{t,T} u(T,x) &= -\int_T^{+\infty} e^{-\lambda(s-t)} P_{t,T} P_{T,s} f(s,x) ds \\ &= -\int_T^{+\infty} e^{-\lambda(s-t)} P_{t,s} f(s,x) ds, \end{aligned}$$

and so,

$$\begin{aligned} & e^{-\lambda(T-t)} P_{t,T} u(T,x) - \int_t^T e^{-\lambda(s-t)} P_{t,s} f(s,x) ds \\ &= -\int_T^{+\infty} e^{-\lambda(s-t)} P_{t,s} f(s,x) ds - \int_t^T e^{-\lambda(s-t)} P_{t,s} f(s,x) ds \\ &= -\int_t^{+\infty} e^{-\lambda(s-t)} P_{t,s} f(s,x) ds = u(t,x), \end{aligned}$$

and (4.2) is fulfilled.

Uniqueness. Let u be a solution of (4.2). Since

$$\lim_{T\to+\infty} e^{-\lambda(T-t)} P_{t,T} u(T,x) = 0,$$

(recall that u is bounded), letting $T \to +\infty$ in (4.2) yields (4.3). The proof is complete. □

Assume now in addition that there exists an evolution sytem of measures ν_t. Then, for any $u \in C_0^\infty(\mathbb{R}\times\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} Lu(t,x) \nu_t(dx) dt = 0. \tag{4.4}$$

We have in fact, taking into account (2.2),

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} u(t,x)\nu_t(dx) &= \int_{\mathbb{R}^d} u_t(t,x)\nu_t(dx) + \int_{\mathbb{R}^d} u(t,x) d_t \nu_t(dx) \\ = \int_{\mathbb{R}^d} u_t(t,x)\nu_t(dx) &+ \int_{\mathbb{R}^d} u(t,x) K(t)^* \nu_t(dx) = \int_{\mathbb{R}^d} Lu(t,x)\nu_t(dx). \end{aligned}$$

Integrating with respect to t over $\mathbb{R}$, yields (4.4).

By (4.4) we find the identity

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} Lu(t,x) u(t,x) \nu_t(dx) dt = -\frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}^d} |Du(t,x)|^2 \nu_t(dx) dt.$$

This shows that the operator L is dissipative in the space

$$L^2(\mathbb{R}\times\mathbb{R}^d;\nu_t) \colon = \left\{ u : \int_{\mathbb{R}} \int_{\mathbb{R}^d} |u(t,x)|^2 \nu_t(dx) dt < \infty \right\}.$$

Remark 4.2. If the coefficients of (1.1) are not regular, one can try to take a regularized equation

$$D_t u_n(t,x) + K_n(t)u_n(t,x) = f(t,x),$$

where the K_n are regular, and then to prove the essential m–dissipativity of L proceeding as in the autonomous case, see, e.g., [3].

5. The equation $Lu = f$ with final condition

Here we want to consider the equation

$$D_t u(t,x) + K(t)u(t,x) = f(t,x), \quad t \geq 0, \ x \in \mathbb{R}^d$$

where $f \in C_b((0,T]) \times \mathbb{R}^d)$ with the final condition

$$u(T) = u_0.$$

Lemma 5.1. *Let $v \in C_0^\infty((0,T]) \times \mathbb{R}^d)$. Then we have*

$$\int_0^T \int_{\mathbb{R}^d} Lv(t,x)\nu_t(dx)dt = -\int_{\mathbb{R}^d} v(T,x)\nu_T(dx).$$

Proof. We have

$$D_t \int_{\mathbb{R}^d} v(t,x)\nu_t(dx) = \int_{\mathbb{R}^d} D_t v(t,x)\nu_t(dx) - \int_{\mathbb{R}^d} K(t)v(t,x)\nu_t(dx).$$

Consequently,

$$\int_0^T \int_{\mathbb{R}^d} Lv(t,x)\nu_t(dx)dt = \int_0^T \int_{\mathbb{R}^d} (D_t v(t,x) + K(t)v(t,x))\nu_t(dx)dt$$
$$= \int_0^T D_t \int_{\mathbb{R}^d} v(t,x)\nu_t(dx)dt = -\int_{\mathbb{R}^d} v(T,x)\nu_T(dx).$$

□

Proposition 5.2. *For any $v \in C_0^\infty((0,T]) \times \mathbb{R}^d$ we have*

$$\int_0^T dt \int_H Lv(t,x)\ v(t,x)\ \nu_t(dx)$$
$$= -\frac{1}{2} \int_0^T dt \int_H |D_x u(t,x)|^2 \nu_t(dx) - \int_H v^2(T,x)\nu_T(dx).$$

Proof. The conclusion follows by Lemma 5.1 replacing v with v^2 and using the elementary identity

$$L(v^2) = 2vLv + |Dv(t,x)|^2.$$

□

By Proposition 5.2 it follows that L is dissipative in $L^2((0,T] \times H, \nu)$; consequently it is closable and its closure is dissipative in $L^2([0,T] \times H, \nu)$.

References

[1] V. Bogachev, G. Da Prato, and M. Röckner, *Existence of solutions to weak parabolic equations for measures*, Proc. London Math. Soc., **88 (3)** (2004), 753–774.

[2] V. Bogachev, G. Da Prato, and M. Röckner, *In preparation.*

[3] G. Da Prato, *Kolmogorov Equations for Stochastic PDEs*, Birkhäuser, 2004.

[4] G. Da Prato and L. Tubaro, *Some results on periodic measures for differential stochastic equations with additive noise*, Dynamic Systems and Applications, **1** (1992), 103–120.

Giuseppe Da Prato
Scuola Normale Superiore
Piazza dei Cavalieri 7
56126 Pisa
Italy
e-mail: `daprato@sns.it`

Michael Röckner
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
D–33501 Bielefeld
Germany
e-mail: `roeckner@mathematik.uni-bielefeld.de`

Progress in Probability, Vol. 59, 123–134

Remarks on 3D Stochastic Navier-Stokes Equations

Franco Flandoli

Abstract. Stochastic Navier–Stokes equations could be a suitable model to address questions of statistical fluid mechanics. For stationary measures arising from the Galerkin scheme, energy balance relations are reviewed, a notion of scaling law inspired by Kolmogorov theory is introduced, and a few results and remarks are given in dimensions 2 and 3.

Mathematics Subject Classification (2000). 60H15, 76F05, 76M35.

Keywords. Stochastic Navier–Stokes equations, statistical fluid mechanics, invariant measures, Kolmogorov scaling law.

1. Introduction

Consider the stochastic Navier–Stokes equations on the torus $\mathcal{T} = [0,1]^3$,

$$\frac{\partial u}{\partial t} + (u \cdot \nabla) u + \nabla p = \nu \triangle u + \sum_{i=1}^{\infty} \sigma_i h_i(x) \dot{\beta}_i(t), \tag{1.1}$$
$$\mathsf{div}\ u = 0, \quad \text{periodic b.c.},$$

where $u = u(t,x)$ is the random velocity field, $p = p(t,x)$ is the random pressure field, $\nu > 0$ is the kinematic viscosity, $h_i(x)$ are the eigenfunctions of the Stokes operator, $\beta_i(t)$ are independent Brownian motions, and σ_i are the noise intensities. A certain amount of foundational material is known, like the existence of martingale solutions, the existence of Markov selections and their continuous dependence on initial conditions under special assumptions on the noise, see the review [2] and references therein (although the well-posedness is still open).

One of the most important open problems is concerned with quantitative information on the invariant measures, related, for instance, to statistical fluid dynamics and the laws of turbulence. In conceptual terms, a turbulent fluid is a non-equilibrium system, so no general Gibbs paradigm is expected to hold a priori. Similarly, from the mathematical side it is not gradient-like.

To attack such a problem for the deterministic equation seems extremely difficult, with some attempt by the theory of Ruelle-Sinai-Bowen (not yet applicable even to much simpler models). There is hope that the stochastic model with additive white noise, although artificial, may allow us to capture more easily some feature, see instances of this attempt by [7] and [8] (2D case) and [9] (3D case).

In this note, following Kolmogorov [6] and much of the literature on turbulence (see [4] for a review), we try to understand whether a scaling law may hold for the so-called structure function. We extract and expand some ideas of the work [3] (see also [2]). Equation (1.1) looks like an excellent model for this sort of investigations, as pointed out by [9], since in the limit of vanishing viscosity the mean rate of energy dissipation remains (formally) constant.

Our considerations are limited to the following remarks: in dimension 2 no scaling law is possible, if we adopt a certain rather natural definition of scaling law. In dimension 3 we do not know the answer, but we at least point out the necessity of an intense vortex stretching mechanism in order to hope for a scaling law.

1.1. Notation

We (formally) rewrite equation (1.1) as an abstract stochastic evolution equation in the Hilbert space H,

$$du(t) + \left[\nu Au(t) + B\left(u(t), u(t)\right)\right] dt = \sum_{i=1}^{\infty} \sigma_i h_i d\beta_i\left(t\right),$$

where, denoted by $\mathcal{D}^\infty$ the space of all infinitely differentiable divergence-free, zero mean, periodic fields, H is the closure of $\mathcal{D}^\infty$ in L^2, V is the closure of $\mathcal{D}^\infty$ in H^1, $D(A)$ is the closure of $\mathcal{D}^\infty$ in H^2, $A : D(A) \subset H \to H$ is the operator $Au = -\triangle u$ (componentwise), $\{h_i\}_{i\in\mathbb{N}} \subset H$ is a c.o.s. of eigenvectors of A (with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots$), $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$ (so the noise is the time derivative of an H-valued Brownian motion), in fact often $\sum_{i=1}^{\infty} \lambda_i \sigma_i^2 < \infty$ (to have more regularity of certain measures), $V \subset H \subset V'$ is the usual triple built on these spaces, $\langle \cdot, \cdot \rangle_H$ and $\langle \cdot, \cdot \rangle_{V,V'}$ denote inner product in H and dual pairing between V and V', and $B\left(\cdot, \cdot\right) : V \times V \to V'$ is the bilinear mapping defined as

$$\langle w, B\left(u, v\right)\rangle_{V,V'} = \sum_{i,j=1}^{3} \int_{\mathcal{T}} u_i \frac{\partial v_j}{\partial x_i} w_j dx.$$

Other notation used below will be Q for the operator in H defined as $Qx = \sum_i \sigma_i^2 \langle x, h_i\rangle_H h_i$. Given a monotone diverging sequence of positive integers $\{N_n\}$, we also consider the finite-dimensional Hilbert space H_n spanned by $h_1, \ldots, h_{N_n}$, embed it into H, denote by π_n the orthogonal projection from H to H_n, denote by A_n the restriction of A to H_n and by $B_n\left(.,.\right) : H_n \times H_n \to H_n$ the continuous bilinear operator defined as

$$B_n\left(u, v\right) = \pi_n B\left(u, v\right), \quad u, v \in H_n.$$

2. Preliminaries on Galerkin approximations

2.1. Energy balance for solutions

Consider the stochastic ordinary differential equation in H_n,

$$dX_t^n = \left[-\nu A_n X_t^n - B_n\left(X_t^n, X_t^n\right)\right] dt + \pi_n \sqrt{Q} dW_t, \quad t \geq 0, \tag{2.1}$$

where $(W_t)_{t\geq 0}$ is a Brownian motion in H with covariance operator Q, defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P\right)$. The following result is well known, see [2] for a complete proof.

Lemma 2.1 (L^2 bounds and energy equality). *For every $\mathcal{F}_0$-measurable r.v. $X_0 : \Omega \to H$, there is a unique continuous adapted solution $(X_t^n)_{t\geq 0}$ of equation (2.1) with initial condition $\pi_n X_0$. It satisfies*

$$E\left(\sup_{t\in[0,T]} |X_t^n|_H^2 + \nu \int_0^T \|X_s^n\|_V^2 \, ds\right)$$

$$\leq C\left(E\left|X_0\right|_H^2, TrQ, T\right) |X_t^n|_H^2 + 2\nu \int_0^t \|X_s^n\|_V^2 \, ds$$

$$= |X_0^n|_H^2 + Tr\, \pi_n Q\, t + M_t^n$$

where M_t^n is a square integrable martingale,

$$\frac{1}{2} E\left|X_T^n\right|_H^2 + \nu \int_0^T E\left\|X_s^n\right\|_V^2 ds = \frac{1}{2} E\left|X_0^n\right|_H^2 + \frac{1}{2} Tr\, \pi_n Q\, T.$$

Remark 2.2. This result indicates the right topologies for the solutions of the Navier–Stokes equations (1.1) and gives us the main uniform-in-n bounds to prove the existence of solutions to the martingale problem associated to (1.1), see [2].

Remark 2.3. More important for the purpose of this note, this result is a mean energy balance:

- $\frac{1}{2} Tr\, \pi_n Q = \sum_{i=1}^{N_n} \sigma_i^2$ is the mean rate of energy injected into the system (the finite-dimensional system (2.1))
- $\nu E \int_0^T \|X_s^n\|_V^2 \, ds$ is the mean energy dissipated on $[0, T]$
- $\frac{1}{2} E\left|X_T^n\right|_H^2$ is the mean (kinetic) energy of the system.

For stationary solutions the energy balance is even more interesting (see [2]).

Corollary 2.4. *If $(X_t^n)_{t\geq 0}$ is a stationary, continuous, adapted solution, then*

$$\nu E\left\|X_t^n\right\|_V^2 = \frac{Tr\, \pi_n Q}{2}.$$

Remark 2.5. The mean rate of energy dissipation balances the mean rate of injected energy. In view of Kolmogorov theory, it is very interesting that the mean rate of energy dissipation has a finite limit when the viscosity goes to zero. It is not clear how to realize such a condition without using a white noise and Itô calculus.

2.2. Invariant measures of the Galerkin approximations

It is not difficult to prove that equation (2.1) defines a Markov process with the Feller property (see [2]). Invariant measures are then well defined, for (2.1). By the classical Krylov-Bogoliubov method one has:

Theorem 2.6. *There exists at least one invariant measure μ^n for equation* (2.1).

Remark 2.7. If Q is invertible, then μ^n is unique and ergodic. More interesting, Weinan E and Mattingly [1] in $d = 2$ and Romito [10] in $d = 3$ have proved ergodicity when only very few modes are randomly excited.

The analog of the previous energy balance relations for invariant measures is:

Theorem 2.8. *All invariant measures μ^n satisfy*

$$\mu^n \left[\|\cdot\|_V^2 \right] = \frac{Tr\, \pi_n Q}{2\nu}.$$

Let us remark that, with respect to the problem of obtaining quantitative information on the invariant measures of Navier–Stokes models, this is a first example of a result.

2.3. Galerkin stationary measures for the 3D equation

Let us come to the infinite-dimensional equation. Since Markov property and classical notion of invariant measures are difficult issues, and also in order to take advantage of the finite-dimensional estimates proved above, we do not analyze the limit equation but simply define the following concept, in analogy for instance with infinite volume measures in statistical mechanics (but an analog of DLL conditions is not known).

We say that a Borel probability measure μ on H (shortly $\mu \in \Pr(\mathcal{H}, \mathcal{B})$) is a *Galerkin stationary measure* if there is a subsequence $n_k \uparrow \infty$ and corresponding invariant measures μ_{n_k} of the finite-dimensional equations (2.1) such that $\mu_{n_k} \to \mu$ weakly on $\mathcal{H}$. We denote by $\mathcal{P}_{NS}^{Galerkin}$ the set of all Galerkin stationary measures.

Theorem 2.9. *$P_{NS}^{Galerkin}$ is non empty. Every $\mu \in P_{NS}^{Galerkin}$ satisfies*

$$\nu\mu \left(\|\cdot\|_{\mathcal{V}}^2 \right) \leq \frac{\sum_i \sigma_i^2}{2}.$$

Remark 2.10 (open problem). For the finite-dimensional approximations we have the equality. In the limit, under the assumption $\sum_i \sigma_i^2 \lambda_i$, for the 2D Navier–Stokes equations we still have the equality (from suitable integrability estimates for the vorticity). On the contrary, for the 2D Navier–Stokes equations, the equality is an open problem.

Several other results can be proved (see [2]). If $\sum_i \sigma_i^2 \lambda_i < \infty$, then $\mu(D(A)) = 1$ for every $\mu \in P_{NS}^{Galerkin}$ and $\mu\left[|Ax|_{\mathcal{H}}^{2/3} \right] < \infty$. If the Brownian motion $\sum_{i=1}^{\infty} \sigma_i h_i \beta_i(t)$ is space-homogeneous and partially (in the sense of rotations compatible with the torus) isotropic, then there exists $\mu \in P_{NS}^{Galerkin}$ that is space-homogeneous and partially isotropic.

3. Remarks on K41 theory

3.1. Second-order structure function

Let us work under the assumption $\sum \lambda_i \sigma_i^2 < \infty$, and of a space-homogeneous and partially isotropic noise. Given $\nu > 0$, denote by $P_{NS}^{Galerkin}(\nu)$ the set of all Galerkin stationary measures for equation (1.1) with viscosity ν, that are space-homogeneous and partial isotropic. For all of them we have $\mu(D(A)) = 1$ and

$$\nu\mu \left[\int_{\mathcal{T}} \|Du(x)\|^2 \, dx \right] \leq \frac{\sum \sigma_i^2}{2}.$$

Denote by $\mathcal{M}$ the set of all pairs (μ, ν) such that $\mu \in P_{NS}^{Galerkin}(\nu)$. Let us introduce the *second-order structure function*

$$S_2^{\mu}(r) = \mu \left[\|u(r \cdot e) - u(0)\|^2 \right]$$

for some coordinate unitary vector e, with $r > 0$. The definition is independent of e and the observation point (taken here to be 0).

The following is observed in many experiments on turbulent fluids (see [4] for a review): in log-log coordinates, the plot of the function $r \mapsto S_2^{\mu}(r)$ has, for small viscosity, a plateau with approximate slope 2/3, for a certain range $I(\nu) \subset (0,1)$ of r's:

$$\log S_2^{\mu}(r) \sim \frac{2}{3} \log r + const, \quad r \in I(\nu),$$

and the interval $I(\nu) = [r_-(\nu), r_+(\nu)]$ has (at least) the property

$$\lim_{\nu \to 0} r_-(\nu) = 0.$$

Thus the function $r \mapsto \frac{\log S_2^{\mu}(r)}{\log r}$ is approximately equal to $\frac{2}{3}$ in $I(\nu)$. The approximation is good for small ν and small $r \in I(\nu)$.

It is also customary to write approximate expressions like

$$S_2^{\mu}(r) \sim C r^{2/3}$$

adding some qualification about the range of the r's, as above.

It is important to realize that this is not a limit property in a classical sense. Common concepts would be that $S_2^{\mu}(r)$ has a certain power behavior as $r \to 0$, or as $\nu \to 0$, but this is not the case (we shall see that, for given $\nu > 0$, the expected behavior of $S_2^{\mu}(r)$ as $r \to 0$ is like r^2; and for given $r > 0$, it is like $\frac{1}{\nu}$ as $\nu \to 0$). The previous scaling property has another structure, it specifies a power behavior in an *intermediate* range of the r's, with such range that it extends towards zero when $\nu \to 0$.

Possibly there is not a unique way to capture the previous property in a rigorous manner. We attempt a definition in the next subsections.

3.2. A temptative general definition of scaling law

Let us work with a general function $f(\nu, r)$, having in mind $S_2^\mu(r)$ in the easiest case when there is only one stationary measure μ for a given value of ν (so $S_2^\mu(r)$ in fact depends on (ν, r)).

Definition 3.1. We say that $R \subset (0,1) \times (0,1)$ is an admissible region for a scaling law if it has the form

$$R = \{(\nu, r) \in (0,1) \times (0,1) : r \in I(\nu)\}$$

where $I(\nu) = [r_-(\nu), r_+(\nu)]$, with $r_-, r_+ : (0,1) \to (0,1)$ such that $r_-(\nu) < r_+(\nu)$ and

$$\lim_{\nu \to 0} r_+(\nu) = 0, \quad \lim_{\nu \to 0} \frac{r_-(\nu)}{r_+(\nu)} = 0.$$

Definition 3.2. Let $R \subset (0,1) \times (0,1)$ be an admissible region for a scaling law. Let $\alpha > 0$ and $f : (0,1) \times (0,1) \to (0, \infty)$ be given.

We say that f satisfies a weak scaling law with exponent α on R if

$$\lim_{\substack{\nu \to 0 \\ (\nu, r) \in R}} \frac{\log f(\nu, r)}{\log r} = \alpha.$$

To explain the previous notation let us recall that given a function $g : R \to \mathbb{R}$, one writes

$$\lim_{\substack{\nu \to 0 \\ (\nu, r) \in R}} g(\nu, r) = l$$

if for every $\varepsilon > 0$ there is $\nu_0 > 0$ such that $|g(\nu, r) - l| < \varepsilon$ for every $(\nu, r) \in R \cap ((0, \nu_0) \times (0,1))$.

To understand the definition, let us relate it to another form of scaling law.

Definition 3.3. Let R be a region as above and $\alpha > 0$, $f : (0,1) \times (0,1) \to (0, \infty)$ be given. We say that f satisfies a strong scaling law with exponent α on R if there are $\nu_0 > 0$ and $C > c > 0$ such that

$$c \cdot r^\alpha \le f(\nu, r) \le C \cdot r^\alpha \quad \text{for every } \nu \in (0, \nu_0] \text{ and } r \in I(\nu).$$

Proposition 3.4. *If f satisfies a strong scaling law with exponent α on R, then it satisfies also a weak scaling law with exponent α on R.*

Proof. $\forall\, \nu \in (0, \nu_0]$, $r \in I(\nu)$ we have

$$\frac{\log C}{\log r} + \alpha \le \frac{\log f(\nu, r)}{\log r} \le \frac{\log c}{\log r} + \alpha,$$

hence

$$\frac{\log C}{\log r_+(\nu)} + \alpha \le \frac{\log f(\nu, r)}{\log r} \le \frac{\log c}{\log r_-(\nu)} + \alpha.$$

This implies the result. □

Proposition 3.5. *Let R be a region as above and $\alpha > 0$, $f : (0,1) \times (0,1) \to (0,\infty)$ be given. If f satisfies a weak scaling law with exponent α on R, then for every $\alpha^- < \alpha < \alpha^+$ there is $\nu_0 > 0$ such that*

$$r^{\alpha^+} \leq f(\nu, r) \leq r^{\alpha^-}$$

for every $\nu \in (0, \nu_0]$ and $r \in I(\nu)$.

Proof. For every $\varepsilon > 0$ there is $\nu_0 > 0$ such that $\forall\ \nu \in (0, \nu_0]$, $r \in I(\nu)$ we have

$$\alpha - \varepsilon \leq \frac{\log f(\nu, r)}{\log r} \leq \alpha + \varepsilon,$$

hence

$$r^{\alpha+\varepsilon} \leq f(\nu, r) \leq r^{\alpha-\varepsilon}. \qquad \square$$

Finally, we have to comment on the non trivial requirement $\lim_{\nu\to 0} \frac{r_-(\nu)}{r_+(\nu)} = 0$ imposed in the definition of an admissible region. If we do not impose it, but only the other requirements, functions like $f(\nu, r) = \frac{r^2}{\nu}$ satisfy a strong scaling law with *any* exponent $\alpha \in (0, 2)$, thus such a definition would not correspond to any meaningful concept of scaling law. For a proof, see [2].

3.3. Absence of weak scaling laws in 2D

We are indebted to M. Hairer for some original ideas of this section; see [3] for a different presentation of related results.

Consider equation (1.1) in 2D, namely on the torus $\mathcal{T} = [0,1]^2$. To simplify the exposition, let us work under a set of assumptions on the noise that guarantees the uniqueness of invariant measures, namely that $P_{NS}^{Galerkin}(\nu)$ is a singleton for every $\nu > 0$. See [5] for the most advanced result in this direction and references therein. Under this assumption the structure function depends on (ν, r), so we may write $S_2^\nu(r)$ in place of $S_2^\mu(r)$. In the general case we should just modify some details of the definition of scaling law.

We do not give all the rather classical details, but the essential point in our analysis is that the vorticity field $\xi = \mathsf{curl}\ u$ is orthogonal to the plane of the fluid, or, in other words, the vorticity can be described by the scalar field

$$\xi := \nabla^\perp u, \quad \nabla^\perp u := (-\partial_2 u_1, \partial_1 u_2)$$

that satisfies the equation

$$\frac{\partial \xi}{\partial t} + (u \cdot \nabla)\xi = \nu \triangle \xi + \sum_{i=1}^\infty \sigma_i \nabla^\perp h_i \dot{\beta}_i(t).$$

This equation is a powerful tool to get estimates in stronger topologies than those discussed above for weak solutions. One of the results is:

Lemma 3.6. *Let $\mu \in P_{NS}^{Galerkin}(\nu)$. Then*

$$\nu \cdot \mu \left[\int_{\mathcal{T}} \|Du(x)\|^2 \, dx \right] = \frac{1}{2} \sum_{i=1}^\infty \sigma_i^2,$$

$$\nu \cdot \mu \left[\int_{\mathcal{T}} \left\| D\nabla^{\perp} u(x) \right\|^2 dx \right] = \frac{1}{2} \sum_{i=1}^{\infty} \sigma_i^2 \lambda_i .$$

Since $\int_{\mathcal{T}} \left\| D^2 u \right\|^2 dx = \int_{\mathcal{T}} \left\| D\nabla^{\perp} u \right\|^2 dx$, we readily have

$$\nu \cdot \mu \left[\int_{\mathcal{T}} \left\| D^2 u(x) \right\|^2 dx \right] = \frac{1}{2} \sum_{i=1}^{\infty} \sigma_i^2 \lambda_i .$$

Essentially from Taylor formula we get the following behavior.

Proposition 3.7. *Let σ^2, θ^2 be the constants*

$$\sigma^2 = \frac{1}{2} \sum_{i=1}^{\infty} \sigma_i^2, \qquad \theta^2 = \frac{\sigma^2}{\sum_{i=1}^{\infty} \sigma_i^2 \lambda_i} .$$

Then

$$\frac{\sigma^2}{16} \cdot \frac{r^2}{\nu} \leq S_2^{\nu}(r) \leq \frac{\sigma^2}{2} \cdot \frac{r^2}{\nu} \qquad \forall r \in \left(0, \frac{\theta}{8}\right] . \tag{3.1}$$

Proof. We have to use Taylor's formula, but the measures μ are concentrated a priori only on $W^{2,2}$-vector fields. For sake of brevity, we give the proof under the additional assumption that

$$\mu\left(D(A) \cap C^2(\mathcal{T})\right) = 1$$

for all the measures μ involved. In [3] one may find the proof of a related lemma in the general case, performed by mollification.

By space-homogeneity of μ,

$$\begin{aligned} \mu\left[\|u(re) - u(0)\|^2\right] &\leq r^2 \int_0^1 \mu\left[\|Du(\sigma e)\|^2\right] d\sigma \\ &= r^2 \mu\left[\|Du\|^2\right] = r^2 \frac{\sigma^2}{2\nu} \end{aligned}$$

and thus the right-hand inequality of (3.1) is proved for every $r > 0$.

On the other side, for smooth vector fields we have

$$u(re) - u(0) = Du(0)\, re + r^2 \int_0^1 D^2 u(\sigma e)(e,e)\, d\sigma,$$

and thus

$$\begin{aligned} \mu\left[\|Du \cdot re\|^2\right] &\leq 2\mu\left[\|u(re) - u(0)\|^2\right] \\ &\quad + 2\mu\left[\left\| r^2 \int_0^1 D^2 u(\sigma e)(e,e)\, d\sigma \right\|^2\right] . \end{aligned}$$

Again from space-homogeneity of μ,

$$\mu\left[\left\| r^2 \int_0^1 D^2 u(\sigma e)(e,e)\, d\sigma \right\|^2\right] \leq r^4 \mu\left[\left\|D^2 u\right\|^2\right],$$

and from discrete isotropy we have (see the appendix of [3])

$$\mu\left[\|Du\cdot e\|^2\right]=\frac{1}{2}\mu\left[\|Du\|^2\right].$$

Therefore,

$$\mu\left[\|u(re)-u(0)\|^2\right]\geq\frac{r^2}{4}\mu\left[\|Du\|^2\right]-r^4\mu\left[\left\|D^2u\right\|^2\right].$$

Therefore, by definition of θ,

$$S_2^\nu(r)\geq\left(\frac{1}{4}-\frac{r^2}{\theta^2}\right)\frac{\sigma^2}{2\nu}\cdot r^2.$$

This implies the left-hand inequality of (3.1) for $r\in(0,\frac{\theta}{8}]$. The proof is complete. □

Corollary 3.8. *Let R be an admissible region for a scaling law with the property*

$$\lim_{\nu\to0}\frac{\log\nu}{\log r_\pm(\nu)}=\beta_\pm$$

with $\beta_+>\beta_->0$. Then

$$\lim_{\nu\to0}\frac{\log S_2^\nu(r_\pm(\nu))}{\log r_\pm(\nu)}=2-\beta_\pm$$

and thus there is no exponent $\alpha\in(0,2)$ such that $S_2^\nu(r)$ satisfies a weak scaling law with exponent α on R.

Proof. From the previous lemma, for $r_\pm(\nu)\in(0,\frac{\theta}{8}]$, we have

$$\begin{aligned}\frac{\log\frac{\sigma^2}{2}}{\log r_\pm(\nu)}+2-\frac{\log\nu}{\log r_\pm(\nu)}&\leq\frac{\log S_2^\nu(r_\pm(\nu))}{\log r_\pm(\nu)}\\&\leq\frac{\log\frac{\sigma^2}{16}}{\log r_\pm(\nu)}+2-\frac{\log\nu}{\log r_\pm(\nu)}\end{aligned}$$

and thus we get the result. □

The previous result extends to general admissible regions R for a scaling law, but the proof is less easy; see [3] for a related general result.

3.4. Comments on the 3D case

In dimension three we cannot prove any scaling law but the previous arguments that disprove them in 2D do not work anymore. Let us give a definition of K41 scaling law (in a strong sense) and discuss one of its consequences.

Recall the definition of the *mean energy dissipation rate*:

$$\epsilon=\epsilon(\mu,\nu):=\nu\cdot\mu\left[\int_{\mathcal{T}}\|Du(x)\|^2\,dx\right].$$

To simplify the exposition, assume it is *constant as the viscosity goes to zero*:

$$\epsilon(\mu,\nu)=\epsilon_0.$$

As we remarked above, this is an open problem in 3D. It is true for finite-dimensional models and in 2D. Finally, following Kolmogorov, let us introduce the *dissipation length scale*:

$$\eta(\mu,\nu) = \nu^{3/4}\epsilon^{-1/4}$$

that under the assumption of constant mean dissipation becomes

$$\eta(\mu,\nu) = \nu^{3/4}\eta_0$$

with $\eta_0 = {\epsilon_0}^{-1/4}$. We choose the following definition of K41 scaling law. It is a restricted version of the definition given in [3], for expository purposes.

Definition 3.9. We say that the K41 scaling law holds if $\exists\ \nu_0 > 0$, $C > c > 0$, $r_0 > 0$ such that

$$c \cdot r^{2/3} \leq S_2^{\mu}(r) \leq C \cdot r^{2/3}$$

$\forall\ (\mu,\nu) \in \mathcal{M}$, $\nu \in (0,\nu_0]$

$$\nu^{3/4}\eta_0 < r < r_0.$$

The motivation for the exponents 2/3 and 3/4 comes from dimensional analysis (recalled in [2]). We do not know whether this property is true in 3D, there is only some experimental evidence that it should be approximatively true. Let us recall from [3] one of its consequences on the intensity of vortex stretching.

Given $u \in V$, define the *stress tensor*

$$S_u = \frac{1}{2}\left(Du + Du^T\right),$$

the *vorticity field*

$$\xi = \mathsf{curl}\ u,$$

and the *vortex stretching field*

$$S_u\xi \cdot \xi.$$

We (formally) have

$$\frac{\partial \xi}{\partial t} + (u \cdot \nabla)\xi = \nu \triangle \xi + S_u \xi + \sum_{i=1}^{\infty} \sigma_i (\mathsf{curl}\ h_i)\dot{\beta}_i(t).$$

A formal application of Itô's formula, for $\mu \in \mathcal{P}_{NS}^{Galerkin}(\nu)$, gives us

$$\nu \cdot \mu \int_{\mathcal{T}} \|D\xi\|^2\, dx \leq \mu \int_{\mathcal{T}} S_u\xi \cdot \xi\, dx + \frac{1}{2}\sum_{i=1}^{\infty} \sigma_i^2 \lambda_i.$$

If we assume that K41 is satisfied, arguing on a quantity similar to θ^2 above we can prove (see [3]) that

$$\mu\left[\int_{\mathcal{T}} \left\|D^2 u\right\|^2 dx\right] \geq C\epsilon_0^{3/2} \cdot \nu^{-5/2}$$

and the same is true for $\mu\left[\int_{\mathcal{T}}\|D\xi\|^2\,dx\right]$ that is equal to the left-hand-side. Then the previous formal inequality would give us

$$\mu\left[\int_{\mathcal{T}} S_u\xi\cdot\xi dx\right]\geq C\epsilon_0^{3/2}\nu^{-3/2}.$$

Let us state two rigorous versions of this result, proved in [3].

Theorem 3.10. *If* K41 *holds true, then*

$$\liminf_{k\to\infty}\mu_{n_k}\left[\int_{\mathcal{T}} S_u\mathsf{curl}\ u\cdot\mathsf{curl}\ udx\right]\geq C\epsilon_0^{3/2}\nu^{-3/2}$$

for every $\mu\in\mathcal{P}_{NS}^{Galerkin}(\nu)$ *and every* $\mu_{n_k}\in\mathcal{S}^{k_n}(\nu)$ *such that* $\mu_{k_n}\to\mu$ *in* H*.*

Theorem 3.11. *If* K41 *holds true, then*

$$\mu\left[\int_{\mathcal{T}} S_u\xi\cdot\xi dx\right]\geq C\epsilon_0^{3/2}\nu^{-3/2}.$$

for every $\mu\in\mathcal{P}_{NS}^{Galerkin}(\nu)$ *limit of* $\mu_{n_k}\in\mathcal{S}^{k_n}(\nu)$ *such that*

$$\mu_{n_k}\left[\|\cdot\|_V^{3+\varepsilon}\right]\leq C$$

for some $\varepsilon, C>0$*.*

References

[1] E. Weinan and J. C. Mattingly, *Ergodicity for the Navier–Stokes equation with degenerate random forcing: finite-dimensional approximation*, Comm. Pure Appl. Math., **54 (11)** (2001), 1386–1402.

[2] F. Flandoli, *An Introduction to 3D Stochastic Fluid Dynamics*, CIME Lecture Notes, 2005.

[3] F. Flandoli, M. Gubinelli, M. Hairer, and M. Romito, *Remarks on the K41 scaling law in turbulent fluid*, preprint, 2005.

[4] U. Frisch, *Turbulence*, Cambridge University Press, Cambridge, 1995.

[5] M. Hairer and J. C. Mattingly, *Ergodicity of the 2D Navier–Stokes equations with degenerate forcing term*, *Annals of Mathematics*, **164 (3)** (2006), 993–1032.

[6] A. N. Kolmogorov, *The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers*, reprinted in Proc. Roy. Soc. London, Ser. A, **434** (1991), 9–13.

[7] S. B. Kuksin, *The Eulerian limit for 2D statistical hydrodynamics*, J. Statist. Phys., **115 (1-2)** (2004), 469–492.

[8] S. B. Kuksin and O. Penrose, *A family of balance relations for the two-dimensional Navier–Stokes equations with random forcing*, J. Stat. Phys., **118 (3-4)** (2005), 437–449.

[9] A. Kupiainen, *Statistical theories of turbulence*, in: Advances in Mathematical Sciences and Applications, Gakkotosho, Tokyo, 2003.

[10] M. Romito, *Ergodicity of the finite-dimensional approximations of the 3D Navier–Stokes equations forced by a degenerate noise*, J. Statist. Phys., **114 (1-2)** (2004), 155–177.

Franco Flandoli
Dipartimento Matematica Applicata U. Dini
Università di Pisa
Via Buonarroti 1
C.A.P. 56127, Pisa
Italy
e-mail: flandoli@dma.unipi.it

Progress in Probability, Vol. 59, 135–174

Slices of a Brownian Sheet: New Results and Open Problems

Davar Khoshnevisan

Abstract. We can view a Brownian sheet as a sequence of interacting Brownian motions or *slices*. Here we present a number of results about the slices of the sheet. A common feature of our results is that they exhibit phase transition. In addition, a number of open problems are presented.

Mathematics Subject Classification (2000). 60G60, 60H99.

Keywords. Brownian sheet, capacity, dimension.

1. Introduction

Let $B := \{B(s\,,t)\}_{s,t\geq 0}$ denote a two-parameter Brownian sheet in $\mathbb{R}^d$. That is, B is a centered Gaussian process with covariance matrix,

$$\mathrm{Cov}\left(B_i(s\,,t)\ ,\ B_j(u\,,v)\right) = \min(s\,,u)\min(t\,,v)\delta_{i,j}\,.$$

We can assume without loss of generality that B is continuous. Moreover, it is convenient to think of B as the distribution function of a d-dimensional white noise $\hat{B}$ on $\mathbb{R}^2_+$; i.e., we may think of $B(s\,,t)$ as

$$B(s\,,t) = \hat{B}\left([0\,,s]\times[0\,,t]\right).$$

These properties were discovered first in Čentsov [2].

Choose and fix some number $s > 0$. The *slice of B along s* is the stochastic process $\{B(s\,,t)\}_{t\geq 0}$. It is easy to see that if s is non-random then the slice of B along s is a scaled Brownian motion. More precisely, $t \mapsto s^{-1/2}B(s\,,t)$ is standard d-dimensional Brownian motion. It is not too difficult to see that if s is random, then the slice along s need not be a Brownian motion. For instance, the slice along a non-random s hits points if and only if $d = 1$. But there are random values of s such that the slice along s hits zero up to dimension $d = 3$; see (1.1) below. Nonetheless, one may expect the slice along s to look like Brownian motion in

Research supported in part by a grant from the United States National Science Foundation.

some sense, even for some random values of s. [For example, all slices share the Brownian property that they are continuous paths.]

A common question in infinite-dimensional stochastic analysis is to ask if there are slices that behave differently from d-dimensional Brownian motion in a predescribed manner. There is a large literature on this subject; see the survey paper [12]. In this paper we present some new examples where there is, generally, a "cut-off phenomenon" or "phase transition."

Our first example is related to the zero-set of the Brownian sheet. Orey and Pruitt [23] have proven that $B^{-1}\{0\}$ is non-trivial if and only if the spatial dimension d is three or less. That is,

$$\mathrm{P}\left\{B(s\,,t)=0 \text{ for some } s,t>0\right\}>0 \text{ if and only if } d\leq 3. \tag{1.1}$$

See also Fukushima [9] and Penrose [24]. Khoshnevisan [14] has derived the following refinement: For all non-random, compact sets $E,F\subset(0\,,\infty)$,

$$\mathrm{P}\left\{B^{-1}\{0\}\cap(E\times F)\neq\varnothing\right\}>0 \text{ if and only if } \mathrm{Cap}_{d/2}(E\times F)>0, \tag{1.2}$$

where Cap_β denotes "β-dimensional Riesz capacity." [These capacities are recalled in the appendix.] The Orey–Pruitt theorem (1.1) follows immediately from (1.2) and Taylor's theorem [Appendix A.1].

Now consider the projection $\mathcal{Z}_d$ of $B^{-1}\{0\}$ onto the x-axis. That is,

$$\mathcal{Z}_d := \left\{s\geq 0:\ B(s\,,t)=0 \text{ for some } t>0\right\}.$$

Thus, $s\in\mathcal{Z}_d$ if and only if the slice of B along s hits zero. Of course, zero is always in $\mathcal{Z}_d$, and the latter is a.s. closed. Our first result characterizes the polar sets of $\mathcal{Z}_d$.

Theorem 1.1. *For all non-random, compact sets $F\subset(0\,,\infty)$,*

$$\mathrm{P}\left\{\mathcal{Z}_d\cap F\neq\varnothing\right\}>0 \textit{ if and only if } \mathrm{Cap}_{(d-2)/2}(F)>0.$$

Theorem 1.1 and Taylor's theorem [Appendix A.1] together provide us with a new proof of the Orey–Pruitt theorem (1.1). Furthermore, we can apply a codimension argument [13, Theorem 4.7.1, p. 436] to find that

$$\dim_{\mathrm{H}}\mathcal{Z}_d = 1\wedge\left(2-\frac{d}{2}\right)^+ \qquad \text{a.s.},$$

where $\dim_{\mathrm{H}}$ denotes Hausdorff dimension [Appendix A.3]. Consequently, when $d\in\{2\,,3\}$, the [Hausdorff] dimension of $\mathcal{Z}_d$ is equal to $2-(d/2)$. Oddly enough, this is precisely the dimension of $B^{-1}\{0\}$ as well; see Rosen [27, 28]. But $\mathcal{Z}_d$ is the projection of $B^{-1}\{0\}$ onto the x-axis. Therefore, one might guess that $B^{-1}\{0\}$ and $\mathcal{Z}_d$ have the same dimension because all slices of B have the property that their zero-sets have zero dimension. If B were a generic function of two variables, then such a result would be false, as there are simple counter-examples. Nevertheless, the "homoegenity" of the slices of B guarantees that our intuition is correct in this case.

Theorem 1.2. *If $d \in \{2,3\}$, then the following holds outside a single* P*-null set:*

$$\dim_{\mathrm{H}} \left(B^{-1}\{0\} \cap (\{s\} \times (0,\infty))\right) = 0 \qquad \text{for all } s > 0. \tag{1.3}$$

Remarks 1.3. 1. Equation (1.3) is not valid when $d = 1$. In that case, Penrose [24] proved that $\dim_{\mathrm{H}}(B^{-1}\{0\} \cap (\{s\} \times (0,\infty))) = 1/2$ for all $s > 0$. In particular, Penrose's theorem implies that $\mathcal{Z}_1 = \mathbb{R}_+$ a.s.; the latter follows also from an earlier theorem of Shigekawa [29].

2. Almost surely, $\mathcal{Z}_d = \{0\}$ when $d \geq 4$; see (1.1). This and the previous remark together show that "$d \in \{2,3\}$" covers the only interesting dimensions.
3. The fact that Brownian motion misses singletons in $\mathbb{R}^d$ for $d \geq 2$ implies that the Lebesgue measure of $\mathcal{Z}_d$ is a.s. zero when $d \in \{2,3\}$.
4. It is not hard to see that the probability in Theorem 1.1 is 0 or 1. Used in conjunction with Theorem 1.1, this observation demonstrates that $\mathcal{Z}_d$ is a.s. everywhere-dense when $d \leq 3$.

Next, we consider the random set,

$$\mathcal{D}_d := \{s \geq 0 : \ B(s,t_1) = B(s,t_2) \text{ for some } t_2 > t_1 > 0\}.$$

We can note that $s \in \mathcal{D}_d$ if and only if the slice of B along s has a double point.

Lyons [18] has proven that $\mathcal{D}_d$ is non-trivial if and only if $d \leq 5$. That is,

$$\mathrm{P}\{\mathcal{D}_d \neq \{0\}\} > 0 \text{ if and only if } d \leq 5. \tag{1.4}$$

See also Mountford [21]. Lyons's theorem (1.4) is an improvement to an earlier theorem of Fukushima [9] which asserts the necessity of the condition "$d \leq 6$." Our next result characterizes the polar sets of $\mathcal{D}_d$.

Theorem 1.4. *For all non-random, compact sets $F \subset (0,\infty)$,*

$$\mathrm{P}\{\mathcal{D}_d \cap F \neq \varnothing\} > 0 \textit{ if and only if } \mathrm{Cap}_{(d-4)/2}(F) > 0.$$

Lyons's theorem (1.4) follows at once from this and Taylor's theorem. In addition, a codimension argument reveals that almost surely,

$$\dim_{\mathrm{H}} \mathcal{D}_d = 1 \wedge \left(3 - \frac{d}{2}\right)^+.$$

This was derived earlier by Mountford [21] who used different methods.

Remark 1.5. Penrose [24, 25] has shown that $\mathcal{D}_d = \mathbb{R}^d_+$ a.s. when $d \leq 3$. Also recall Lyons' theorem (1.4). Thus, Theorem 1.4 has content only when $d \in \{4,5\}$.

In summary, our Theorems 1.1 and 1.4 state that certain unusual slices of the sheet can be found in the "target set" F if and only if F is sufficiently large in the sense of capacity. Next we introduce a property which is related to more delicate features of the set F. Before we do so, let us set $d \geq 3$ and define

$$\mathcal{R}(s) := \inf\left\{\alpha > 0 : \ \liminf_{t\to\infty} \frac{(\log t)^{1/\alpha}}{t^{1/2}} |B(s,t)| < \infty\right\} \qquad \text{for all } s > 0.$$

Thus, $\mathcal{R}(s)$ is the critical escape-rate — at the logarithmic level — for the slice of B along s. Because $t \mapsto s^{-1/2}B(s\,,t)$ is a standard Brownian motion for all fixed $s > 0$, the integral test of Dvoretzky and Erdős [7] implies that

$$\mathrm{P}\left\{\mathcal{R}(s) = d-2\right\} = 1 \quad \text{for all } s > 0.$$

That is, the typical slice of B escapes at log-rate $(d-2)$. This leads to the question, "When are all slices of B transient"? Stated succinctly, the answer is: "If and only if $d \geq 5$." See Fukushima [9] for the sufficiency of the condition "$d \geq 5$," and Kôno [16] for the necessity. Further information can be found in Dalang and Khoshnevisan [3]. Next we try to shed further light on the rate of convergence of the transient slices of B. Our characterization is in terms of packing dimension $\dim_{\mathrm{P}}$, which is recalled in Appendix B.2.

Theorem 1.6. *Choose and fix $d \geq 3$, and a non-random compact set $F \subset (0\,,\infty)$. Then with probability* 1*:*

1. $\mathcal{R}(s) \geq d - 2 - 2\dim_{\mathrm{P}} F$ *for all* $s \in F$.
2. *If* $\dim_{\mathrm{P}} F < (d-2)/2$, *then* $\mathcal{R}(s) = d - 2 - 2\dim_{\mathrm{P}} F$ *for some* $s \in F$.

Remark 1.7. The condition that $\dim_{\mathrm{P}} F < (d-2)/2$ is always met when $d \geq 5$.

The organization of this paper is as follows: After introducing some basic real-variable computations in Section 2 we prove Theorem 1.1 in Section 3. Our derivation is entirely harmonic-analytic, and rests on a projection theorem for capacities which may be of independent interest. Theorems 1.4 and 1.2 are respectively proved in Sections 4 and 6. Section 5 contains a variant of Theorem 1.4, and Section 7 contains the proof of Theorem 1.6 and much more. There is also a final Section 8 wherein we record some open problems.

Throughout, any n-vector x is written, coordinatewise, as $x = (x_1, \ldots, x_n)$. Moreover, $|x|$ will always denote the ℓ^1-norm of $x \in \mathbb{R}^n$; i.e.,

$$|x| := |x_1| + \cdots + |x_n|.$$

Generic constants that do not depend on anything interesting are denoted by $c, c_1, c_2, \ldots$; they are *always* assumed to be positive and finite, and their values may change between, as well as within, lines.

Let A denote a Borel set in $\mathbb{R}^n$. The collection of all Borel probability measures on A is always denoted by $\mathcal{P}(A)$.

Acknowledgement. A large portion of this work was motivated by enlightening discussions with Robert Dalang over a period of several years. A great many thanks are due to him. The final version of this paper enjoyed a number of improvements thanks to suggestions made by an anonymous referee to whom I am grateful.

2. Preliminary real-variable estimates

Our analysis depends on the properties of three classes of functions. We develop the requisite estimates here in this section. Aspects of these lemmas overlap with Lemmas 1.2 and 2.5 of Dalang and Khoshnevisan [3].

Here and throughout, we define for all $\epsilon > 0$ and $x \in \mathbb{R}$,

$$\begin{aligned} f_\epsilon(x) &:= \left(\frac{\epsilon}{|x|^{1/2}} \wedge 1\right)^d, \\ F_\epsilon(x) &:= \int_0^1 f_\epsilon(y + |x|)\, dy, \\ G_\epsilon(x) &:= \int_0^1 F_\epsilon(y + |x|)\, dy. \end{aligned} \tag{2.1}$$

Our first technical lemma attaches a "meaning" to f_ϵ.

Lemma 2.1. *Let* $\mathbf{g}$ *denote a* d*-vector of i.i.d. standard-normal variables. Then there exist a constant* c *such that for all* $\sigma, \epsilon > 0$,

$$c f_\epsilon(\sigma^2) \le \mathrm{P}\left\{\sigma|\mathbf{g}| \le \epsilon\right\} \le f_\epsilon(\sigma^2).$$

Proof. This is truly an elementary result. However, we include a proof to acquaint the reader with some of the methods that we use later on.

Let $M := \max_{1\le i\le d} |\mathbf{g}_i|$, and note that $|\mathbf{g}| \ge M$. Therefore,

$$\mathrm{P}\left\{\sigma|\mathbf{g}| \le \epsilon\right\} \le \left(\int_{-\epsilon/\sigma}^{\epsilon/\sigma} \frac{e^{-u^2/2}}{(2\pi)^{1/2}}\, du\right)^d \le \left(\frac{\epsilon}{\sigma}\right)^d,$$

because $(2/\pi)^{1/2}\exp(-u^2/2) \le 1$. The upper bound of the lemma follows because $\mathrm{P}\{\sigma|\mathbf{g}| \le \epsilon\}$ is also at most one. To derive the lower bound we use the inequality $|\mathbf{g}| \le Md$ to find that when $\epsilon \le \sigma$,

$$\begin{aligned} \mathrm{P}\left\{\sigma|\mathbf{g}| \le \epsilon\right\} &\ge \left(\int_{-\epsilon/(\sigma d)}^{\epsilon/(\sigma d)} \frac{e^{-u^2/2}}{(2\pi)^{1/2}}\, du\right)^d \ge \left(\frac{2}{\pi d^2}\right)^{d/2} e^{-1/(2d^2)} \left(\frac{\epsilon}{\sigma}\right)^d \\ &= \left(\frac{2}{\pi d^2}\right)^{d/2} e^{-1/(2d^2)} f_\epsilon(\sigma^2) := c_1 f_\epsilon(\sigma^2). \end{aligned}$$

The same reasoning shows that when $\epsilon > \sigma$,

$$\begin{aligned} \mathrm{P}\left\{\sigma|\mathbf{g}| \le \epsilon\right\} &\ge \left(\int_{-1}^{1} \frac{e^{-u^2/2}}{(2\pi)^{1/2}}\, du\right)^d = \left(\int_{-1}^{1} \frac{e^{-u^2/2}}{(2\pi)^{1/2}}\, du\right)^d f_\epsilon(\sigma^2) \\ &:= c_2 f_\epsilon(\sigma^2). \end{aligned}$$

The lemma follows with $c := \min(c_1\,, c_2)$. □

Next we find bounds for F_ϵ in terms of the function $U_{(d-2)/2}$ that is defined in (A.1).

Lemma 2.2. *There exists $c > 1$ such that such that for all $0 \le y \le 2$ and $\epsilon > 0$,*

$$F_\epsilon(y) \le c\epsilon^d U_{(d-2)/2}(y).$$

In addition, for all $y \ge \epsilon^2$,

$$F_\epsilon(y) \ge \frac{\epsilon^d}{c} U_{(d-2)/2}(y).$$

Proof. Evidently,

$$F_\epsilon(y) = \int_0^1 f_\epsilon(x+y)\,dx \le \epsilon^d \int_0^1 \frac{dx}{(x+y)^{d/2}} = \epsilon^d \int_y^{1+y} \frac{dx}{x^{d/2}},$$

and this is an equality when $y \ge \epsilon^2$. The remainder of the proof is a direct computation. □

As regards the functions G_ϵ, we first note that

$$G_\epsilon(x) = \iint\limits_{[0,1]^2} f_\epsilon(x + |y|)\,dy. \tag{2.2}$$

The following captures a more useful property of G_ϵ.

Lemma 2.3. *There exists $c > 1$ such that for all $0 < x \le 2$ and $\epsilon > 0$,*

$$G_\epsilon(x) \le c\epsilon^d U_{(d-4)/2}(x).$$

If, in addition, $x \ge \epsilon^2$, then

$$G_\epsilon(x) \ge \frac{\epsilon^d}{c} U_{(d-4)/2}(x).$$

Lemma 2.3 follows from Lemma 2.2 and one or two elementary and direct computations.

We conclude this section with a final technical lemma.

Lemma 2.4. *For all $x, \epsilon > 0$,*

$$G_\epsilon(x) \ge \frac{1}{2}\int_0^2 F_\epsilon(x+y)\,dy.$$

Proof. We change variables to find that

$$\int_0^2 F_\epsilon(x+y)\,dy = \frac{1}{2}\int_0^1 F_\epsilon\left(x + \frac{y}{2}\right)\,dy \ge \frac{1}{2}\int_0^1 F_\epsilon(x+y)\,dy,$$

by monotonicity. This proves the lemma. □

3. Proof of Theorem 1.1

In light of (1.2) it suffices to prove that

$$\mathrm{Cap}_{d/2}([0\,,1]\times F)>0 \text{ if and only if } \mathrm{Cap}_{(d/2)-1}(F)>0. \tag{3.1}$$

The following harmonic-analytic fact does the job, and a little more; it must be well known, but we could not find it in a suitable form in the literature.

Recall that a function $f:\mathbb{R}^n\to[0\,,\infty]$ is of *strict positive type* if: (i) f is locally integrable away from $0\in\mathbb{R}^n$; and (ii) the Fourier transform of f is *strictly* positive. Corresponding to such a function f we can define a function $\Pi_m f$ [equivalently, the operator Π_m] as follows:

$$(\Pi_m f)(x) := \int_{[0,1]^m} f(x\otimes y)\,dy \qquad \text{for all } x\in\mathbb{R}^{n-m},$$

where $x\otimes y:=(x_1\,,\ldots,x_{n-m}\,,y_1\,,\ldots,y_m)\in\mathbb{R}^n$ is the tensor product of x and y. It is easy to see that

$$(\Pi_m f)(x) := \iint_{[0,1]^m\times[0,1]^m} f(x\otimes(y-z))\,dy\,dz \qquad \text{for all } x\in\mathbb{R}^{n-m}, \tag{3.2}$$

provided that we identify $[0\,,1]^m$ with the m-dimensional torus endowed with its usual coordinatewise addition (mod 1) group product and the corresponding quotient topology. The preceding is a direct computation when $m=1$; the general case is proved by induction. Then, we have

Theorem 3.1 (Projection theorem for capacities). *Let $n>1$ be an integer, and suppose that $f:\mathbb{R}^n\to[0\,,\infty]$ is of strict positive type and continuous on $\mathbb{R}^n\setminus\{0\}$. Then, for all integers $1\le m<n$ and compact sets $F\subset\mathbb{R}^{n-m}$,*

$$\mathrm{Cap}_f\left([0\,,1]^m\times F\right)=\mathrm{Cap}_{\Pi_m f}(F).$$

The proof is divided into two parts. The first part is easier, and will be dispensed with first.

Proof of Theorem 3.1 *(The Upper Bound).* Let λ_m denote the Lebesgue measure on $[0\,,1]^m$, normalized to have mass one. If $\mu\in\mathcal{P}(F)$, then evidently,

$$I_{\Pi_m f}(\mu)=I_f(\lambda_m\times\mu)\ge\inf_{\nu\in\mathcal{P}([0,1]^m\times F)} I_f(\nu).$$

The equality follows from (3.2) and the theorem of Fubini–Tonelli. But it is clear that $\lambda_m\times\mu\in\mathcal{P}([0\,,1]^m\times F)$, whence $\mathrm{Cap}_{\Pi_m f}(F)\le\mathrm{Cap}_f([0\,,1]^m\times F)$. This completes our proof. □

We need some preliminary developments for the lower bound. For this portion, we identify the hypercube $[0\,,1)^m$ with the m-dimensional torus $\mathbf{T}^m$ in the usual way. In particular, note that $\mathbf{T}^m$ is compact in the resulting quotient topology. Any probability measure μ on $[0\,,1)^m\times F$ can be identified with a probability measure on $\mathbf{T}^m\times F$ in the usual way. We continue to write the latter measure as

μ as well. Throughout the remainder of this section, $f : \mathbb{R}^n \to [0\,,\infty]$ is a fixed function of strict positive type that is also continuous on $\mathbb{R}^n \setminus \{0\}$.

Lemma 3.2. *Suppose $\mathbf{T}^m \times F$ has positive f-capacity. Then, there exists a probability measure $\mathbf{e}_{\mathbf{T}^m\times F}$ — the "equilibrium measure" — on $\mathbf{T}^m \times F$ such that*

$$I_f(\mathbf{e}_{\mathbf{T}^m\times F}) = \left[\mathrm{Cap}_f\left(\mathbf{T}^m \times F\right)\right]^{-1} < \infty.$$

Proof. For all $\epsilon > 0$ we can find $\mu_\epsilon \in \mathcal{P}(\mathbf{T}^m \times F)$ such that

$$I_f(\mu_\epsilon) \le \frac{1+\epsilon}{\mathrm{Cap}_f\left(\mathbf{T}^m \times F\right)}. \tag{3.3}$$

All μ_ϵ's are probability measures on the same compact set $\mathbf{T}^m \times F$. Choose an arbitrary weak limit $\mu_0 \in \mathcal{P}(\mathbf{T}^m \times F)$ of the sequence $\{\mu_\epsilon\}_{\epsilon>0}$, as $\epsilon \to 0$. It follows from Fatou's lemma that

$$\begin{aligned}
\liminf_{\epsilon\to 0} I_f(\mu_\epsilon) &\ge \liminf_{\eta\to 0}\liminf_{\epsilon\to 0} \iint\limits_{\{|x-y|\ge\eta\}} f(x-y)\,\mu_\epsilon(dx)\,\mu_\epsilon(dy)\\
&\ge \liminf_{\eta\to 0} \iint\limits_{\{|x-y|\ge\eta\}} f(x-y)\,\mu_0(dx)\,\mu_0(dy)\\
&= I_f(\mu_0).
\end{aligned}$$

Thanks to (3.3), $I_f(\mu_0)$ is at most equal to the reciprocal of the f-capacity of $\mathbf{T}^m \times F$. On the other hand, the said capacity is bounded above by $I_f(\sigma)$ for all $\sigma \in \mathcal{P}(T^m \times F)$, whence follows the lemma. □

The following establishes the uniqueness of the equilibrium measure.

Lemma 3.3. *Suppose $\mathbf{T}^m \times F$ has positive f-capacity χ. If $I_f(\mu) = I_f(\nu) = 1/\chi$ for some $\mu, \nu \in \mathcal{P}(\mathbf{T}^m \times F)$, then $\mu = \nu = \mathbf{e}_{\mathbf{T}^m\times F}$.*

Proof. We denote by $\mathcal{F}$ the Fourier transform on any and every (locally compact) abelian group G; $\mathcal{F}$ is normalized as follows: For all group characters ξ, and all $h \in L^1(G)$,

$$(\mathcal{F}h)(\xi) = \int_G (x\,,\xi)h(x)\,dx,$$

where $(x\,,\xi)$ is the usual duality relation between $x \in G$ and the character ξ, and "dx" denotes Haar measure (normalized to have mass one if G is compact; counting measure if G is discrete; and mixed in the obvious way, when appropriate). Because f is of positive type and continuous away from the origin,

$$I_f(\mu) = \frac{1}{(2\pi)^n}\int_{\mathbf{T}^m\times\mathbb{R}^{n-m}} (\mathcal{F}f)(\xi)\left|(\mathcal{F}\mu)(\xi)\right|^2\,d\xi; \tag{3.4}$$

see Kahane [10, Eq. (5), p. 134].

Using (3.4) (say) we can extend the definition of $I_f(\kappa)$ to all signed measures κ that have finite absolute mass. We note that $I_f(\kappa)$ is real and non-negative, but could feasibly be infinite; $I_f(\kappa)$ is strictly positive if κ is not identically equal to the zero measure. The latter follows from the strict positivity of f.

Let ρ and σ denote two signed measures that have finite absolute mass. Then, we can define, formally,

$$I_f(\sigma,\rho) := \iint \left[\frac{f(x-y)+f(y-x)}{2}\right] \sigma(dx)\,\rho(dy).$$

This is well defined if $I_f(|\sigma|\,,|\rho|) < \infty$, for instance. Evidently, $I_f(\sigma,\rho) = I_f(\rho\,,\sigma)$ and $I_f(\sigma,\sigma) = I_f(\sigma)$. Finally, by the Cauchy–Schwarz inequality,

$$|I_f(\sigma,\rho)| \le I_f(\sigma) I_f(\rho).$$

Now suppose to the contrary that the μ and ν of the statement of the lemma are distinct. Then, by (3.4),

$$0 < I_f\left(\frac{\mu-\nu}{2}\right) = \frac{I_f(\mu)+I_f(\nu)-2I_f(\mu\,,\nu)}{4} = \frac{\chi^{-1}-I_f(\mu\,,\nu)}{2},$$

where, we recall, $\chi^{-1} = I_f(\mathbf{e}_{\mathbf{T}^m\times F})$ denotes the reciprocal of the f-capacity of $\mathbf{T}^m \times F$. Consequently, $I_f(\mu\,,\nu)$ is strictly less than $I_f(\mathbf{e}_{\mathbf{T}^m\times F})$. From this we can deduce that

$$\begin{aligned} I_f\left(\frac{\mu+\nu}{2}\right) &= \frac{I_f(\mu)+I_f(\nu)+2I_f(\mu\,,\nu)}{4} = \frac{\chi^{-1}+I_f(\mu\,,\nu)}{2} \\ &< I_f(\mathbf{e}_{\mathbf{T}^m\times F}) \le I_f\left(\frac{\mu+\nu}{2}\right). \end{aligned}$$

And this is a contradiction. Therefore, $\mu=\nu$; also μ is equal to $\mathbf{e}_{\mathbf{T}^m\times F}$ because of the already-proved uniqueness together with Lemma 3.2. □

Proof of Theorem 3.1 *(The Lower Bound).* It remains to prove that

$$\mathrm{Cap}_{\Pi_m f}(F) \ge \mathrm{Cap}_f\left([0\,,1]^m \times F\right). \tag{3.5}$$

We will prove the seemingly-weaker statement that

$$\mathrm{Cap}_{\Pi_m f}(F) \ge \mathrm{Cap}_f\left(\mathbf{T}^m \times F\right). \tag{3.6}$$

This is seemingly weaker because $\mathrm{Cap}_f(\mathbf{T}^m \times F) = \mathrm{Cap}_f([0\,,1)^m \times F)$. But, in fact, our proof will reveal that for all $q>1$,

$$\mathrm{Cap}_{\Pi_m f}(F) \ge q^{-m}\mathrm{Cap}_f\left([0\,,q)^m \times F\right).$$

The right-hand side is at least $q^{-m}\mathrm{Cap}_f([0\,,1]^m \times F)$. Therefore, we can let $q \downarrow 1$ to derive (3.5), and therefrom the theorem.

With our ultimate goal (3.6) in mind, we assume without loss of generality that

$$\mathrm{Cap}_f\left(\mathbf{T}^m \times F\right) > 0.$$

Thus, $\mathbf{e}_{\mathbf{T}^m\times F}$ exists and is the unique minimizer in the definition of $\mathrm{Cap}_f(\mathbf{T}^m \times F)$ (Lemmas 3.2 and 3.3).

Let us write any $z \in \mathbf{T}^m \times \mathbb{R}^{n-m}$ as $z=(z',z'')$, where $z' \in \mathbf{T}^m$ and $z'' \in \mathbb{R}^{n-m}$.

For all $a, b \in \mathbf{T}^m \times \mathbb{R}^{n-m}$ define $\tau_a(b) = a + b$. We emphasize that the first m coordinates of $\tau_a(b)$ are formed by addition in $\mathbf{T}^m$ [i.e., component-wise addition mod 1 in $[0,1)^m$], whereas the next $n-m$ coordinates of $\tau_a(b)$ are formed by addition in $\mathbb{R}^{n-m}$. In particular, $\tau_a(\mathbf{T}^m \times F) = \mathbf{T}^m \times (a'' + F)$.

For all $a \in \mathbf{T}^m \times \mathbb{R}^{n-m}$, $\mathbf{e}_{\mathbf{T}^m\times F} \circ \tau_a^{-1}$ is a probability measure on $\tau_a(\mathbf{T}^m \times F)$. Moreover, it is easy to see that $\mathbf{e}_{\mathbf{T}^m\times F}$ and $\mathbf{e}_{\mathbf{T}^m\times F} \circ \tau_a^{-1}$ have the same f-energy. Therefore, whenever $a'' = 0$, $\mathbf{e}_{\mathbf{T}^m\times F} \circ \tau_a^{-1}$ is a probability measure on $\mathbf{T}^m \times F$ that minimizes the f-capacity of $\mathbf{T}^m \times F$. The uniqueness of $\mathbf{e}_{\mathbf{T}^m\times F}$ proves that

$$\mathbf{e}_{\mathbf{T}^m\times F} = \mathbf{e}_{\mathbf{T}^m\times F} \circ \tau_a^{-1} \quad \text{whenever } a'' = 0.$$

See Lemma 3.3. Now let X be a random variable with values in $\mathbf{T}^m \times F$ such that the distribution of X is $\mathbf{e}_{\mathbf{T}^m\times F}$. The preceding display implies that for all $a' \in \mathbf{T}^m$, the distribution of $(X' + a', X'')$ is the same as that of (X', X''). The uniqueness of normalized Haar measure λ_m then implies that X' is distributed as λ_m. In fact, for all Borel sets $A \subset \mathbf{T}^m$ and $B \subset \mathbb{R}^{n-m}$,

$$\begin{aligned}
\mathbf{e}_{\mathbf{T}^m\times F}(A \times B) &= \mathrm{P}\left\{X' \in A\,, X'' \in B\right\} \\
&= \int_{\mathbf{T}^m} \mathrm{P}\left\{X' \in a' + A\,, X'' \in B\right\} da' \\
&= \mathrm{E}\left[\lambda_m(A - X')\ ;\ X'' \in B\right] \\
&= \lambda_m(A)\mathrm{P}\left\{X'' \in B\right\} := \lambda_m(A)\mu(B).
\end{aligned}$$

Now we compute directly to find that

$$\mathrm{Cap}_f\left(\mathbf{T}^m \times F\right) = \frac{1}{I_f\left(\lambda_m \times \mu\right)} = \frac{1}{I_{\Pi_m f}(\mu)} \le \frac{1}{\inf_{\sigma\in\mathcal{P}(F)} I_{\Pi_m f}(\sigma)}.$$

This proves (3.6), and therefore the theorem. □

Finally we are ready to present the following:

Proof of Theorem 1.1. The function U_α is of strict positive type for all $0 < \alpha < d$. The easiest way to see this is to merely recall the following well-known fact from harmonic analysis: In the sense of distributions, $\mathcal{F}U_\alpha = c_{d,\alpha} U_{d-\alpha}$ for a positive and finite constant $c_{d,\alpha}$ [30, Lemma 1, p. 117]. We note also that U_α is continuous away from the origin. Thus, we can combine (1.2) with Theorem 3.1 to find that

$$\mathrm{P}\left\{\mathcal{Z}_d \cap F \neq \varnothing\right\} > 0 \text{ if and only if } \mathrm{Cap}_{\Pi_1 U_{d/2}}(F) > 0. \tag{3.7}$$

But for all $x \ge \epsilon^2 > 0$,

$$\left(\Pi_1 U_{d/2}\right)(x) \asymp \int_0^1 \frac{dy}{|x+y|^{d/2}} = \frac{F_\epsilon(x)}{\epsilon^d}.$$

[By "$f \asymp g$" we mean that f/g is bounded above and below by universal constants.] Therefore, in accord with Lemmas 2.2 and 2.4, $(\Pi_1 U_{d/2})(x) \asymp U_{(d-2)/2}(x)$, simultaneously for all $\epsilon > 0$ and $x \ge 2\epsilon^2$. Because the implies constants in the last inequalities do not depend on ϵ, it follows that $\mathrm{Cap}_{\Pi_1 U_{d/2}}(F) \asymp \mathrm{Cap}_{(d-2)/2}(F)$. This and (3.7) together prove the theorem. □

4. Proof of Theorem 1.4

Let $B^{(1)}$ and $B^{(2)}$ be two independent Brownian sheets in $\mathbb{R}^d$, and define for all $\mu \in \mathcal{P}(\mathbb{R}_+)$,

$$J_\epsilon(\mu) := \frac{1}{\epsilon^d} \iint_{[1,2]^2} \int \mathbf{1}_{\mathbf{A}(\epsilon\,;s,t)}\, \mu(ds)\, dt,$$

where $\mathbf{A}(\epsilon; a, b)$ is the event

$$\mathbf{A}(\epsilon; a, b) := \left\{ |B^{(2)}(a\,, b_2) - B^{(1)}(a\,, b_1)| \leq \epsilon \right\}, \tag{4.1}$$

for all $1 \leq a, b_1, b_2 \leq 2$ and $\epsilon > 0$.

Lemma 4.1. *We have*

$$\inf_{0<\epsilon<1} \inf_{\mu \in \mathcal{P}([1,2])} \mathrm{E}\left[J_\epsilon(\mu)\right] > 0.$$

Proof. The distribution of $B^{(2)}(s\,, t_2) - B^{(1)}(s\,, t_1)$ has a density function that is bounded below, uniformly for all $1 \leq s, t_1, t_2 \leq 2$. □

Next we present a bound for the second moment of $J_\epsilon(\mu)$. For technical reasons, we first alter $J_\epsilon(\mu)$ slightly. Henceforth, we define

$$\hat{J}_\epsilon(\mu) := \frac{1}{\epsilon^d} \iint_{[1,3]^2} \int \mathbf{1}_{\mathbf{A}(\epsilon;s,t)}\, \mu(ds)\, dt.$$

Lemma 4.2. *There exists a constant c such that for all Borel probability measures μ on $\mathbb{R}_+$ and all $0 < \epsilon < 1$,*

$$\mathrm{E}\left[\left(\hat{J}_\epsilon(\mu) \right)^2 \right] \leq \frac{c I_{G_\epsilon}(\mu)}{\epsilon^d} \leq c I_{(d-4)/2}(\mu).$$

Proof. For all $\epsilon > 0$, $1 \leq s, u \leq 2$, and $t, v \in [1\,, 2] \times [3\,, 4]$ define

$$P_\epsilon(s, u; t, v) := \mathrm{P}\left(\mathbf{A}(\epsilon; s, t) \cap \mathbf{A}(\epsilon; u, v) \right).$$

We claim that there exists a constant c_1 — independent of $(s\,, u\,, t\,, v\,, \epsilon)$ — such that

$$P_\epsilon(s, u; t, v) \leq c_1 \epsilon^d f_\epsilon(|s - u| + |t - v|). \tag{4.2}$$

Lemmas 2.3 and 2.4 of Dalang and Khoshnevisan [3] contain closely-related, but non-identical, results.

Let us assume (4.2) for the time being and prove the theorem. We will establish (4.2) subsequently.

Owing to (4.2) and the Fubini–Tonelli theorem,

$$\begin{aligned}
\mathrm{E}\left[\left(\hat{J}_\epsilon(\mu)\right)^2\right] &\le \frac{c_1}{\epsilon^d}\iint \iint_{[1,3]^2\times[1,3]^2} f_\epsilon(|s-u|+|t-v|)\,dt\,dv\,\mu(ds)\,\mu(du)\\
&\le \frac{c}{\epsilon^d}\iint G_\epsilon(s-u)\,\mu(ds)\,\mu(du)\\
&= \frac{cI_{G_\epsilon}(\mu)}{\epsilon^d}.
\end{aligned}$$

See (2.2). This is the first inequality of the lemma. The second follows from the first and Lemma 2.3. Now we proceed to derive (4.2).

By symmetry, it suffices to estimate $P_\epsilon(s,u;t,v)$ in the case that $s \le u$. Now we carry out the estimates in two separate cases.

Case 1. First we consider the case $t_1 \le v_1$ and $t_2 \le v_2$. Define $\hat{B}^{(i)}$ to be the white noise that corresponds to the sheet $B^{(i)}$ $(i=1,2)$. Then, consider

$$\begin{gathered}
H_1^{(1)} := \hat{B}^{(1)}\left([0\,,s]\times[0\,,t_1]\right), \quad H_2^{(1)} := \hat{B}^{(1)}\left([0\,,s]\times[t_1\,,v_1]\right),\\
H_3^{(1)} := \hat{B}^{(1)}\left([s\,,u]\times[0\,,v_1]\right),\\
H_1^{(2)} := \hat{B}^{(2)}\left([0\,,s]\times[0\,,t_2]\right), \quad H_2^{(2)} := \hat{B}^{(2)}\left([0\,,s]\times[t_2\,,v_2]\right),\\
H_3^{(2)} := \hat{B}^{(2)}\left([s\,,u]\times[0\,,v_2]\right).
\end{gathered}$$

Then, the H's are all totally independent Gaussian random vectors. Moreover, we can find independent d-vectors $\{\mathbf{g}_j^{(i)}\}_{1\le i\le 2, 1\le j\le 3}$ of i.i.d. standard-normals such that

$$\begin{gathered}
H_1^{(1)} = (st_1)^{1/2}\mathbf{g}_1^{(1)},\ H_2^{(1)} = (s(v_1-t_1))^{1/2}\mathbf{g}_2^{(1)},\\
H_3^{(1)} = (v_1(u-s))^{1/2}\mathbf{g}_3^{(1)},\\
H_1^{(2)} = (st_2)^{1/2}\mathbf{g}_1^{(2)},\ H_2^{(2)} = (s(v_2-t_2))^{1/2}\mathbf{g}_2^{(2)},\\
H_3^{(2)} = (v_2(u-s))^{1/2}\mathbf{g}_3^{(2)}.
\end{gathered}$$

In addition,

$$\begin{aligned}
P_\epsilon(s,u;t,v) &= \mathrm{P}\left\{\begin{array}{c} \left|H_1^{(2)} - H_1^{(1)}\right| \le \epsilon \\ \left|H_1^{(2)} + H_2^{(2)} + H_3^{(2)} - H_1^{(1)} - H_2^{(1)} - H_3^{(1)}\right| \le \epsilon \end{array}\right\}\\
&\le \mathrm{P}\left\{\left|H_1^{(2)} - H_1^{(1)}\right| \le \epsilon\right\}\\
&\qquad\qquad \times \mathrm{P}\left\{\left|H_2^{(2)} + H_3^{(2)} - H_2^{(1)} - H_3^{(1)}\right| \le 2\epsilon\right\}.
\end{aligned}$$

The first term on the right is equal to the following:

$$\mathrm{P}\left\{(s(t_1+t_2))^{1/2}|\mathbf{g}| \le \epsilon\right\} \le c_2\epsilon^d, \tag{4.3}$$

where $c_2 > 0$ does not depend on (s, t, u, v, ϵ); see Lemma 2.1. Also, the second term is equal to the following:

$$\begin{aligned}\mathrm{P}\left\{(s(v_2 - t_2) + v_2(u - s) + s(v_1 - t_1) + v_1(u - s))^{1/2}\,|\mathbf{g}| \leq 2\epsilon\right\}& \\ \leq \mathrm{P}\left\{(|v - t| + (u - s))^{1/2}\,|\mathbf{g}| \leq 2\epsilon\right\}& \\ \leq c_3 f_\epsilon(|u - s| + |t - v|),&\end{aligned} \tag{4.4}$$

and $c_3 > 0$ does not depend on $(s\,, t\,, u\,, v\,, \epsilon)$. We obtain (4.2) by combining (4.3) and (4.4). This completes the proof of Case 1.

Case 2. Now we consider the case that $t_2 \geq v_2$ and $t_1 \leq v_1$. We can replace the $H_i^{(j)}$'s of Case 1 with the following:

$$\begin{gathered}H_1^{(1)} := \hat{B}^{(1)}\left([0\,, s] \times [0\,, t_1]\right), \quad H_2^{(1)} := \hat{B}^{(1)}\left([0\,, s] \times [t_1\,, v_1]\right), \\ H_3^{(1)} := \hat{B}^{(1)}\left([s\,, u] \times [0\,, v_1]\right), \\ H_1^{(2)} := \hat{B}^{(2)}\left([0\,, s] \times [0\,, v_2]\right), \quad H_2^{(2)} := \hat{B}^{(2)}\left([0\,, s] \times [v_2\,, t_2]\right), \\ H_3^{(2)} := \hat{B}^{(2)}\left([s\,, u] \times [0\,, v_2]\right).\end{gathered}$$

It follows then that

$$P_\epsilon(s, u; t, v) = \mathrm{P}\left\{\begin{array}{c}\left|H_1^{(2)} + H_2^{(2)} - H_1^{(1)}\right| \leq \epsilon \\ \left|H_1^{(2)} + H_3^{(2)} - H_1^{(1)} - H_2^{(1)} - H_3^{(1)}\right| \leq \epsilon\end{array}\right\}.$$

One can check covariances and see that the density function of $H_1^{(2)} - H_1^{(1)}$ is bounded above by a constant $c_1 > 0$ that does not depend on $(s\,, t\,, u\,, v\,, \epsilon)$. Therefore,

$$\begin{aligned}P_\epsilon(s, u; t, v) &\leq c_1 \int_{\mathbb{R}^d} \mathrm{P}\left\{\begin{array}{c}\left|H_1^{(2)} + z\right| \leq \epsilon \\ \left|H_3^{(2)} - H_2^{(1)} - H_3^{(1)} + z\right| \leq \epsilon\end{array}\right\} dz \\ &= c_1 \int_{\{|w| \leq \epsilon\}} \mathrm{P}\left\{\left|H_3^{(2)} - H_1^{(2)} + H_2^{(1)} - H_3^{(1)} + w\right| \leq \epsilon\right\} dw \\ &\leq c_1 (2\epsilon)^d \mathrm{P}\left\{\left|H_3^{(2)} - H_1^{(2)} + H_2^{(1)} - H_3^{(1)}\right| \leq 2\epsilon\right\}.\end{aligned}$$

The component-wise variance of this particular combination of $H_j^{(i)}$'s is equal to $(u - s)(v_1 + v_2) + s(v_1 - t_1 + v_2 - t_2) \geq (u - s) + |t - v|$. Whence follows (4.2) in the present case.

Symmetry considerations, together with Cases 1 and 2, prove that (4.2) holds for all possible configurations of $(s\,, u\,, t\,, v)$. This completes our proof. $\square$

For all $i \in \{1\,, 2\}$ and $s, t \geq 0$, we define $\mathcal{F}_{s,t}^{(i)}$ to be the σ-algebra generated by $\{B^{(i)}(u\,, v)\}_{0 \leq u \leq s,\, 0 \leq v \leq t}$; as usual, we can assume that the $\mathcal{F}^{(i)}$'s are complete and right-continuous in the partial order "$\prec$" described as follows: For all $s, t, u, v \geq 0$,

$(s\,,t) \prec (u\,,v)$ iff $s \le u$ and $t \le v$. [If not, then complete $\mathcal{F}^{(i)}$ and then make it $\prec$-right-continuous.] Based on $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$, we define

$$\mathcal{F}_{s;t,v} := \mathcal{F}^{(1)}_{s,t} \vee \mathcal{F}^{(2)}_{s,v} \qquad \text{for all } s,t,v \ge 0.$$

The following proves that Cairoli's maximal L^2-inequality holds with respect to the family of $\mathcal{F}_{s;t,v}$'s.

Lemma 4.3. *Choose and fix a number $p > 1$. Then for all almost surely non-negative random variables $Y \in \mathcal{L}^p := L^p(\Omega, \vee_{s,t,v\ge 0}\mathcal{F}_{s;t,v}, \mathrm{P})$,*

$$\left\| \sup_{s,t,v\in\mathbf{Q}_+} \mathrm{E}\left[Y \,\middle|\, \mathcal{F}_{s;t,v}\right] \right\|_{\mathcal{L}^p} \le \left(\frac{p}{p-1}\right)^3 \|Y\|_{\mathcal{L}^p}.$$

Proof. We propose to prove that for all $s,s',t,t',v,v' \ge 0$, and all bounded random variables Y that are $\mathcal{F}_{s';t',v'}$-measurable,

$$\mathrm{E}\left[Y \,\middle|\, \mathcal{F}_{s;t,v}\right] = \mathrm{E}\left[Y \,\middle|\, \mathcal{F}_{s\wedge s';t\wedge t',v\wedge v'}\right] \quad \text{a.s.} \tag{4.5}$$

This proves that the three-parameter filtration $\{\mathcal{F}_{s;t,v}\}_{s,t,v\in\mathbf{Q}_+}$ is *commuting* in the sense of Khoshnevisan [13, p. 35]. Corollary 3.5.1 of the same reference [13, p. 37] would then finish our proof.

By a density argument, it suffices to demonstrate (4.5) in the case that $Y = Y_1Y_2$, where Y_1 and Y_2 are bounded, and measurable with respect to $\mathcal{F}^{(1)}_{s',t'}$ and $\mathcal{F}^{(2)}_{s',v'}$, respectively. But in this case, independence implies that almost surely,

$$\mathrm{E}\left[Y \,\middle|\, \mathcal{F}_{s;t,v}\right] = \mathrm{E}\left[Y_1 \,\middle|\, \mathcal{F}^{(1)}_{s,t}\right]\mathrm{E}\left[Y_2 \,\middle|\, \mathcal{F}^{(2)}_{s,v}\right]. \tag{4.6}$$

By the Cairoli–Walsh commutation theorem [13, Theorem 2.4.1, p. 237], $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ are each two-parameter, commuting filtrations. Theorem 3.4.1 of Khoshnevisan [13, p. 36] implies that almost surely,

$$\begin{aligned}
\mathrm{E}\left[Y_1 \,\middle|\, \mathcal{F}^{(1)}_{s,t}\right] &= \mathrm{E}\left[Y_1 \,\middle|\, \mathcal{F}^{(1)}_{s\wedge s',t\wedge t'}\right],\\
\mathrm{E}\left[Y_2 \,\middle|\, \mathcal{F}^{(2)}_{s,v}\right] &= \mathrm{E}\left[Y_2 \,\middle|\, \mathcal{F}^{(2)}_{s\wedge s',v\wedge v'}\right].
\end{aligned}$$

Plug this into (4.6) to obtain (4.5) in the case that Y has the special form Y_1Y_2, as described above. The general form of (4.5) follows from the mentioned special case and density. □

Lemma 4.4. *Choose and fix a number $p > 1$. Then for all almost surely non-negative random variables $Y \in \mathcal{L}^p := L^p(\Omega, \vee_{s,t,v\ge 0}\mathcal{F}_{s;t,v}, \mathrm{P})$, we can find a continuous modification of the three-parameter process $\{\mathrm{E}[Y \,|\, \mathcal{F}_{s;t,v}]\}_{s,t,v\ge 0}$. Consequently,*

$$\left\| \sup_{s,t,v\ge 0} \mathrm{E}\left[Y \,\middle|\, \mathcal{F}_{s;t,v}\right] \right\|_{\mathcal{L}^p} \le \left(\frac{p}{p-1}\right)^3 \|Y\|_{\mathcal{L}^p}.$$

Proof. First suppose $Y = Y_1 Y_2$ where $Y_i \in \mathcal{L}^p(\Omega, \vee_{s,t\geq 0}\mathcal{F}^{(i)}_{s,t}, \mathrm{P})$. In this case, (4.6) holds by independence. Thanks to Wong and Zakai [32], each of the two conditional expectations on the right-hand side of (4.6) has a representation in terms of continuous, two-parameter and one-parameter stochastic integrals. This proves the continuity of $(s\,,t\,,v) \mapsto \mathrm{E}[Y \,|\, \mathcal{F}_{s;t,v}]$ in the case where Y has the mentioned special form. In the general case, we can find $Y^1, Y^2, \ldots$ such that: (i) Each Y^i has the mentioned special form; and (ii) $\|Y^n - Y\|_{\mathcal{L}^p} \leq 2^{-n}$. We can write, for all integers $n \geq 1$,

$$\left|\mathrm{E}[Y^{n+1} \,|\, \mathcal{F}_{s;t,v}] - \mathrm{E}[Y^n \,|\, \mathcal{F}_{s;t,v}]\right| \leq \sum_{k=n}^{\infty} \left|\mathrm{E}[Y^{k+1} - Y^k \,|\, \mathcal{F}_{s;t,v}]\right|.$$

Take supremum over $s, t, v \in \mathbf{Q}_+$ and apply Lemma 4.3 to find that

$$\sum_{n=1}^{\infty} \left\| \sup_{s,t,v\in\mathbf{Q}_+} \left|\mathrm{E}[Y^{n+1} \,|\, \mathcal{F}_{s;t,v}] - \mathrm{E}[Y^n \,|\, \mathcal{F}_{s;t,v}]\right| \right\|_{\mathcal{L}^p} \leq c \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \left\|Y^{k+1} - Y^k\right\|_{\mathcal{L}^p} < \infty.$$

Because each $\mathrm{E}[Y^n \,|\, \mathcal{F}_{s;t,v}]$ is continuous in $(s\,,t\,,v)$, $\mathrm{E}[Y \,|\, \mathcal{F}_{s;t,v}]$ has a continuous modification. The ensuing maximal inequality follows from continuity and Lemma 4.3. □

Lemma 4.5. *There exists a constant c such that the following holds outside a single null set: For all* $0 < \epsilon < 1$, $1 \leq a, b_1, b_2 \leq 2$, *and* $\mu \in \mathcal{P}(\mathbb{R}_+)$,

$$\mathrm{E}\left[\hat{J}_\epsilon(\mu) \,\middle|\, \mathcal{F}_{a;b_1,b_2}\right] \geq \frac{c}{\epsilon^d} \int\limits_{F\cap[a,2]} G_\epsilon(s-a)\,\mu(ds) \cdot \mathbf{1}_{\mathbf{A}(\epsilon/2;a,b)}. \tag{4.7}$$

Remark 4.6. As the proof will show, we may have to redefine the left-hand side of (4.7) on a null-set to make things work seamlessly. The details are standard, elementary probability theory and will go without further mention.

Proof. Throughout this proof we write $\mathcal{E} := \mathcal{E}_{a;b;\epsilon}(\mu) := \mathrm{E}[\hat{J}_\epsilon(\mu) \,|\, \mathcal{F}_{a;b_1,b_2}]$. Evidently,

$$\mathcal{E} \geq \frac{1}{\epsilon^d} \int_{b_1}^{3} \int_{b_2}^{3} \int_{F\cap[a,2]} \mathrm{P}\left(\mathbf{A}(\epsilon; s, t) \,|\, \mathcal{F}_{a;b_1,b_2}\right) \mu(ds)\, dt_2\, dt_1. \tag{4.8}$$

A white-noise decomposition implies the following: For all $s \geq a$, $t_1 \geq b_1$, and $t_2 \geq b_2$,

$$\begin{aligned} B^{(1)}(s\,,t_1) &= B^{(1)}(a\,,b_1) + b_1^{1/2} W_1^1(s-a) + a^{1/2} W_2^1(t_1 - b_1) \\ &\qquad + V^1(s-a\,,t_1-b_1), \\ B^{(2)}(s\,,t_2) &= B^2(a\,,b_2) + b_2^{1/2} W_1^2(s-a) + a^{1/2} W_2^2(t_2 - b_2) \\ &\qquad + V^2(s-a\,,t_2-b_2). \end{aligned}$$

Here: the W_j^i's are standard, linear Brownian motions; the V^i's are Brownian sheets; and the collection $\{W_j^i, V^i, B^i(a\,,b_i)\}_{i,j=1}^2$ is totally independent. By appealing to this decomposition in conjunction with (4.8) we can infer that the following is a lower bound for $\mathcal{E}$, almost surely on the event $\mathbf{A}(\epsilon/2; a, b)$:

$$\begin{aligned}
&\frac{1}{\epsilon^d}\int_{b_1}^3\int_{b_2}^3\int_{F\cap[a,2]}\mu(ds)\,dt_2\,dt_1\\
&\qquad\times \mathrm{P}\left\{\left|\begin{matrix} b_2^{1/2}W_1^2(s-a)+a^{1/2}W_2^2(t_2-b_2)+V^2(s-a\,,t_2-b_2)\\ -b_1^{1/2}W_1^1(s-a)-a^{1/2}W_2^1(t_1-b_1)-V^1(s-a\,,t_1-b_1)\end{matrix}\right|\le\frac{\epsilon}{2}\right\}\\
&=\frac{1}{\epsilon^d}\int_{b_1}^3\int_{b_2}^3\int_{F\cap[a,2]}\mathrm{P}\left\{\sigma\,|\mathbf{g}|\le\frac{\epsilon}{2}\right\}\mu(ds)\,dt_2\,dt_1.
\end{aligned}$$

Here, $\mathbf{g}$ is a d-vector of i.i.d. standard-normals, and σ^2 is equal to the quantity $b_2(s-a)+a(t_2-b_2)+(s-a)(t_2-b_2)+b_1(s-a)+a(t_1-b_1)+(s-a)(t_1-b_1)$. The range of possible values of a and b is respectively $[1\,,2]$ and $[1\,,2]^2$. This means that we can find a constant $c>0$ — independent of $(a\,,b\,,s\,,t)$ — such that $\sigma^2\le c\{|s-a|+|t-b|\}$. Apply this bound to the previous display; then appeal to Lemma 2.1 to find that (4.7) holds a.s., but the null-set could feasibly depend on $(a\,,b\,,\epsilon)$.

To ensure that the null-set can be chosen independently from $(a\,,b\,,\epsilon)$, we first note that the integral on the right-hand side of (4.7) is: (i) continuous in $\epsilon>0$; (ii) independent of $b\in[1\,,2]^2$; and (iii) lower semi-continuous in $a\in[1\,,2]$. Similarly, $(a\,,b\,,\epsilon)\mapsto\mathbf{1}_{\mathbf{A}(\epsilon;a,b)}$ is left-continuous in $\epsilon>0$ and lower semi-continuous in $(a\,,b)\in[1\,,2]^3$. Therefore, it suffices to prove that the left-hand side of (4.7) is a.s. continuous in $(a\,,b)\in[1\,,2]^3$, and left-continuous in $\epsilon>0$. The left-continuity assertion about $\epsilon>0$ is evident; continuity in $(a\,,b)$ follows if we could prove that for all bounded random variables Y, $(a\,,b)\mapsto\mathrm{E}\left[Y\mid\mathcal{F}_{a;b_1,b_2}\right]$ has an a.s.-continuous modification. But this follows from Lemma 4.4. □

Next we state and prove a quantitative capacity estimate.

Proposition 4.7. *Consider the collection of times of double-points:*

$$D(\omega):=\left\{1\le s\le 2:\ \inf_{t\in[1,2]^2}\left|B^{(2)}(s\,,t_2)-B^{(1)}(s\,,t_1)\right|(\omega)=0\right\}.$$

Then there exists a constant $c>1$ such that for all compact, non-random sets $F\subseteq[1\,,2]$,

$$\frac{1}{c}\mathrm{Cap}_{(d-4)/2}(F)\le\mathrm{P}\left\{D\cap F\neq\varnothing\right\}\le c\,\mathrm{Cap}_{(d-4)/2}(F).$$

Proof. Define the closed random sets,

$$D_\epsilon(\omega):=\left\{1\le s\le 2:\ \inf_{t\in[1,2]^2}\left|B^{(2)}(s\,,t_2)-B^{(1)}(s\,,t_1)\right|(\omega)\le\epsilon\right\}.$$

Also, choose and fix a probability measure $\mu \in \mathcal{P}(F)$. It is manifest that D_ϵ intersects F almost surely on the event $\{J_\epsilon(\mu) > 0\}$. Therefore, we can apply the Paley–Zygmund inequality to find that

$$\mathrm{P}\left\{D_\epsilon \cap F \neq \varnothing\right\} \geq \frac{\left(\mathrm{E}[J_\epsilon(\mu)]\right)^2}{\mathrm{E}\left[\left(J_\epsilon(\mu)\right)^2\right]} \geq \frac{\left(\mathrm{E}[J_\epsilon(\mu)]\right)^2}{\mathrm{E}\left[\left(\hat{J}_\epsilon(\mu)\right)^2\right]}.$$

Let $\epsilon \downarrow 0$ and appeal to compactness to find that

$$\mathrm{P}\left\{D \cap F \neq \varnothing\right\} \geq \frac{\liminf_{\epsilon\to 0}\left(\mathrm{E}[J_\epsilon(\mu)]\right)^2}{cI_{(d-4)/2}(\mu)}.$$

[We have used the second bound of Lemma 4.2.] According to Lemma 4.1, the numerator is bounded below by a strictly positive number that does not depend on μ. Therefore, the lower bound of our proposition follows from optimizing over all $\mu \in \mathcal{P}(F)$.

In order to derive the upper bound we can assume, without any loss in generality, that $\mathrm{P}\{D_\epsilon \cap F \neq \varnothing\} > 0$; for otherwise there is nothing to prove.

For all $0 < \epsilon < 1$ define

$$\tau_\epsilon := \inf\left\{s \in F : \inf_{t\in[1,2]^2}\left|B^{(2)}(s\,,t_2) - B^{(1)}(s\,,t_1)\right| \leq \epsilon\right\}.$$

As usual, $\inf\varnothing := \infty$. It is easy to see that τ_ϵ is a stopping time with respect to the one-parameter filtration $\{\mathcal{H}_s\}_{s\geq 0}$, where

$$\mathcal{H}_s := \bigvee_{t,v\geq 0} \mathcal{F}_{s;t,v} \qquad \text{for all } s \geq 0.$$

We note also that there exist $[0\,,\infty]$-valued random variables τ'_ϵ and τ''_ϵ such that: (i) $\tau'_\epsilon \vee \tau''_\epsilon = \infty$ iff $\tau_\epsilon = \infty$; and (ii) almost surely on $\{\tau_\epsilon < \infty\}$,

$$\left|B^{(2)}(\tau_\epsilon\,,\tau'_\epsilon) - B^{(1)}(\tau_\epsilon\,,\tau''_\epsilon)\right| \leq \epsilon.$$

Define

$$p_\epsilon := \mathrm{P}\left\{\tau_\epsilon < \infty\right\}, \quad \text{and} \quad \nu_\epsilon(\bullet) := \mathrm{P}\left(\tau_\epsilon \in \bullet \,\middle|\, \tau_\epsilon < \infty\right).$$

We can note that

$$\inf_{\epsilon>0} p_\epsilon \geq \mathrm{P}\{D \cap F \neq \varnothing\}, \tag{4.9}$$

and this is strictly positive by our earlier assumption. Consequently, ν_ϵ is well defined as a classical conditional probability, and $\nu_\epsilon \in \mathcal{P}(F)$. Now consider the process $\{M^\epsilon\}_{0<\epsilon<1}$ defined as follows:

$$M^\epsilon_{a;b_1,b_2} := \mathrm{E}\left[\hat{J}_\epsilon(\nu_\epsilon) \,\middle|\, \mathcal{F}_{a;b_1,b_2}\right].$$

Thanks to Lemmas 4.4 and 4.5,

$$\begin{aligned}
\mathrm{E}&\left[\sup_{a,b_1,b_2\in\mathbb{R}^3_+}\left(M^\epsilon_{a;b_1,b_2}\right)^2\right]\\
&\qquad\geq \mathrm{E}\left[\left(M^\epsilon_{\tau_\epsilon;\tau''_\epsilon,\tau'_\epsilon}\right)^2\right]\\
&\qquad\geq \frac{cp_\epsilon}{\epsilon^{2d}}\mathrm{E}\left[\left(\int_{F\cap[\tau_\epsilon,2]} G_\epsilon(s-\tau_\epsilon)\,\nu_\epsilon(ds)\right)^2\,\middle|\,\tau_\epsilon<\infty\right]\\
&\qquad\geq \frac{cp_\epsilon}{\epsilon^{2d}}\left(\mathrm{E}\left[\int_{F\cap[\tau_\epsilon,2]} G_\epsilon(s-\tau_\epsilon)\,\nu_\epsilon(ds)\,\middle|\,\tau_\epsilon<\infty\right]\right)^2.
\end{aligned}\tag{4.10}$$

The last line is a consequence of the Cauchy–Schwarz inequality. We can bound the squared term on the right-hand side as follows:

$$\begin{aligned}
\mathrm{E}&\left[\int_{F\cap[\tau_\epsilon,2]} G_\epsilon(s-\tau_\epsilon)\,\nu_\epsilon(ds)\,\middle|\,\tau_\epsilon<\infty\right]\\
&\qquad= \iint_{\{s\in F\cap[u,2]\}} G_\epsilon(s-u)\,\nu_\epsilon(ds)\,\nu_\epsilon(du)\\
&\qquad\geq \frac{1}{2}\iint G_\epsilon(s-u)\,\nu_\epsilon(ds)\,\nu_\epsilon(du) = \frac{1}{2}I_{G_\epsilon}(\nu_\epsilon).
\end{aligned}$$

Plug this in (4.10), and appeal to Lemmas 4.2 and 4.3, to find that

$$\frac{cp_\epsilon}{4\epsilon^{2d}}\left(I_{G_\epsilon}(\nu_\epsilon)\right)^2 \leq \mathrm{E}\left[\sup_{a,b_1,b_2\in\mathbf{Q}_+}\left(M^\epsilon_{a;b_1,b_2}\right)^2\right] \leq 2^6\mathrm{E}\left[\left(\hat{J}_\epsilon(\nu_\epsilon)\right)^2\right] \leq \frac{c}{\epsilon^d}I_{G_\epsilon}(\nu_\epsilon).$$

Solve this, using (4.9), to find that

$$\mathrm{P}\{D\cap F\neq\varnothing\} \leq \frac{c}{I_{G_\epsilon}(\nu_\epsilon)}.\tag{4.11}$$

Choose and fix a number $\eta>0$. In accordance with Lemma 2.3,

$$I_{G_\epsilon}(\nu_\epsilon) \geq \iint_{\{|s-u|\geq\eta\}} U_{(d-4)/2}(s-u)\,\nu_\epsilon(ds)\,\nu_\epsilon(du),$$

for all $0<\epsilon<\eta^{1/2}$. Recall that $\{\nu_\epsilon\}_{\epsilon>0}$ is a net of probability measures on F. Because F is compact, Prohorov's theorem ensures that there exists a subsequential weak limit $\nu_0\in\mathcal{P}(F)$ of $\{\nu_\epsilon\}_{\epsilon>0}$, as $\epsilon\to0$. Therefore, we can apply Fatou's lemma to find that

$$\begin{aligned}
\liminf_{\epsilon\to0} I_{G_\epsilon}(\nu_\epsilon) &\geq \lim_{\eta\to0}\iint_{\{|s-u|\geq\eta\}} U_{(d-4)/2}(s-u)\,\nu_0(ds)\,\nu_0(du)\\
&= I_{(d-4)/2}(\nu_0).
\end{aligned}$$

Together with (4.11), the preceding implies that $\mathrm{P}\{D \cap F \neq \varnothing\}$ is at most some constant divided by $I_{(d-4)/2}(\nu_0)$. This, in turn, in bounded by a constant multiple of $\mathrm{Cap}_{(d-4)/2}(F)$. The proposition follows. $\square$

Proof of Theorem 1.4. Let I and J be disjoint, closed intervals in $(0\,,\infty)$ with the added property that $x < y$ for all $x \in I$ and $y \in J$. Define

$$\mathcal{D}_d(I, J) := \{s > 0 : \; B(s\,,t_1) = B(s\,,t_2) \text{ for some } t_1 \in I \text{ and } t_2 \in J\}\,.$$

We intend to prove that

$$\mathrm{P}\{\mathcal{D}_d(I, J) \cap F \neq \varnothing\} > 0 \text{ if and only if } \mathrm{Cap}_{(d-4)/2}(F) > 0. \tag{4.12}$$

Evidently, this implies Theorem 1.4. Without loss of much generality, we may assume that $I = [\frac{1}{2}\,,\frac{3}{2}]$, $J = [\frac{7}{2}\,,\frac{9}{2}]$, and $F \subseteq [1\,,2]$. Now consider the random fields,

$$\begin{aligned} B^{(2)}(s\,,t) &:= B(s\,,\tfrac{5}{2}+t) - B(s\,,\tfrac{5}{2}) \\ B^{(1)}(s\,,t) &:= B(s\,,\tfrac{5}{2}-t) - B(s\,,\tfrac{5}{2}), \end{aligned}$$

for $0 \leq s,t \leq 5/2$. Then two covariance computations reveal that the random fields $\{B^{(1)}(s\,,\frac{5}{2}-t) - B(s\,,\frac{5}{2})\}_{1\leq s,t\leq 2}$ and $\{B^{(2)}(s\,,\frac{5}{2}+t) - B^{(2)}(s\,,\frac{5}{2})\}_{1\leq s,t\leq 2}$ are independent Brownian sheets. On the other hand, the following are easily seen to be equivalent: (i) there exists $(s\,,t_1\,,t_2) \in [1\,,2]^3$ such that $B^{(1)}(s\,,t_1) = B^{(2)}(s\,,t_2)$; and (ii) there exists $(s\,,t_1\,,t_2) \in [1\,,2] \times I \times J$ such that $B(s\,,t_1) = B(s\,,t_2)$. Therefore, (4.12) follows from Proposition 4.7. This completes our proof. $\square$

5. More on double-points

Consider the random sets

$$\begin{aligned} \hat{\mathcal{D}}_d &:= \left\{(s\,,t_1\,,t_2) \in \mathbb{R}^3_+ : \; B(s\,,t_1) = B(s\,,t_2)\right\}, \\ \bar{\mathcal{D}}_d &:= \left\{(s\,,t_1) \in \mathbb{R}^2_+ : \; B(s\,,t_1) = B(s\,,t_2) \text{ for some } t_2 > 0\right\}. \end{aligned}$$

The methods of this paper are not sufficiently delicate to characterize the polar sets of $\hat{\mathcal{D}}_d$ and $\mathcal{D}_d$. I hasten to add that I believe such a characterization is within reach of the existing technology [14]. Nonetheless it is not too difficult to prove the following by appealing solely to the techniques developed here.

Theorem 5.1. *For all non-random compact sets $E \subset (0\,,\infty)^2$ and $G \subset (0\,,\infty)^3$,*

$$\begin{gathered} \mathrm{Cap}_{d/2}(G) > 0 \implies \mathrm{P}\left\{\hat{\mathcal{D}}_d \cap G \neq \varnothing\right\} > 0 \implies \mathcal{H}_{d/2}(G) > 0, \\ \mathrm{Cap}_{(d-2)/2}(E) > 0 \implies \mathrm{P}\left\{\bar{\mathcal{D}}_d \cap E \neq \varnothing\right\} > 0 \implies \mathcal{H}_{(d-2)/2}(E) > 0. \end{gathered}$$

where $\mathcal{H}_\alpha$ denotes the α-dimensional Hausdorff measure [Appendix A.3*].*

Proof. Let $B^{(1)}$ and $B^{(2)}$ be two independent, two-parameter Brownian sheets on $\mathbb{R}^d$. It suffices to prove that there exists a constant $c > 1$ such that for all non-random compact sets $E \subseteq [1\,,2]^2$ and $G \subseteq [1\,,2]^3$,

$$\begin{aligned} c^{-1}\mathrm{Cap}_{d/2}(G) \le \mathrm{P}\left\{\hat{\mathcal{T}}_d \cap G \neq \varnothing\right\} \le c\mathcal{H}_{d/2}(G), \\ c^{-1}\mathrm{Cap}_{(d-2)/2}(E) \le \mathrm{P}\left\{\bar{\mathcal{T}}_d \cap E \neq \varnothing\right\} \le c\mathcal{H}_{(d-2)/2}(E), \end{aligned} \tag{5.1}$$

where

$$\begin{aligned} \hat{\mathcal{T}}_d &:= \left\{(s\,,t_1,t_2) \in [1\,,2]^3 :\ B^{(2)}(s\,,t_2) = B^{(1)}(s\,,t_1)\right\}, \\ \bar{\mathcal{T}}_d &:= \left\{(s\,,t_1) \in [1\,,2]^2 :\ B^{(2)}(s\,,t_2) = B^{(1)}(s\,,t_1) \text{ for some } t_2 > 0\right\}. \end{aligned}$$

[This sort of reasoning has been employed in the proof of Theorem 1.1 already; we will not repeat the argument here.] We begin by deriving the first bound in (5.1).

Recall (4.1). Choose and fix $\mu \in \mathcal{P}(G)$, and define for all $\epsilon > 0$,

$$\mathcal{J}_\epsilon(\mu) := \frac{1}{\epsilon^d} \iiint \mathbf{1}_{\mathbf{A}(\epsilon;s,t)}\, \mu(ds\, dt_1\, dt_2).$$

The proof of Lemma 4.1 shows that

$$\inf_{0<\epsilon<1}\ \inf_{\mu\in\mathcal{P}([1,2]^3)} \mathrm{E}\left[\mathcal{J}_\epsilon(\mu)\right] > 0.$$

Similarly, we can apply (4.2) to find that

$$\begin{aligned} \mathrm{E}\left[(\mathcal{J}_\epsilon(\mu))^2\right] &\le \frac{c}{\epsilon^d} \iiiint f_\epsilon(|s-u| + |t-v|)\, \mu(ds\, dt_1\, dt_2)\, \mu(du\, dv_1\, dv_2) \\ &\le cI_{d/2}(\mu). \end{aligned}$$

We have used the obvious inequality, $f_\epsilon(x) \le \epsilon^d |x|^{-d/2}$. The lower bound in (5.1) follows from the previous two moment-bounds, and the Paley–Zygmund inequality; we omit the details.

For the proof of the upper bound it is convenient to introduce some notation. Define

$$\begin{aligned} \Delta(s;t) &:= B^{(2)}(s\,,t_2) - B^{(1)}(s\,,t_1) \quad \text{for all } s,t_1,t_2 \ge 0, \\ \mathcal{U}(x;\epsilon) &:= [x_1\,,x_1+\epsilon] \times [x_2\,,x_2+\epsilon] \times [x_3\,,x_3+\epsilon] \quad \text{for all } x \in \mathbb{R}^3,\ \epsilon > 0. \end{aligned}$$

Then,

$$\mathrm{P}\left\{\hat{\mathcal{T}}_d \cap \mathcal{U}(x\,;\epsilon) \neq \varnothing\right\} \le \mathrm{P}\left\{|\Delta(x)| \le \Theta(x\,;\epsilon)\right\},$$

where $\Theta(x\,;\epsilon) := \sup_{y\in\mathcal{U}(x;\epsilon)} |\Delta(y) - \Delta(x)|$. The density function of $\Delta(x)$ is bounded above, uniformly for all $x \in [1\,,2]^3$. Furthermore, $\Delta(x)$ is independent of $\Theta(x\,;\epsilon)$. Therefore, there exists a constant c such that uniformly for all $0 < \epsilon < 1$ and $x \in [1\,,2]^3$,

$$\mathrm{P}\left\{\hat{\mathcal{T}}_d \cap \mathcal{U}(x\,;\epsilon) \neq \varnothing\right\} \le c\mathrm{E}\left[(\Theta(x\,;\epsilon))^d\right] \le c\epsilon^{d/2}. \tag{5.2}$$

The final inequality holds because: (i) Brownian-sheet scaling dictates that $\Theta(x\,;\epsilon)$ has the same law as $\epsilon^{d/2}\Theta(x\,;1)$; and (ii) $\Theta(x\,;1)$ has moments of all order, with bounds that do not depend on $x \in [1\,,2]^3$ [23, Lemma 1.2].

To prove the upper bound we can assume that $\mathcal{H}_{d/2}(G) < \infty$. In this case we can find $x_1, x_2, \ldots \in [1\,,2]^3$ and $r_1, r_2, \ldots \in (0\,,1)$ such that $G \subseteq \cup_{i=1}^{\infty}\mathcal{U}(x_i\,;r_i)$ and $\sum_{i=1}^{\infty} r_i^{d/2} \le 2\mathcal{H}_{d/2}(G)$. Thus, by (5.2),

$$\mathrm{P}\left\{\hat{\mathcal{T}}_d \cap G \neq \varnothing\right\} \le \sum_{i\ge 1} \mathrm{P}\left\{\hat{\mathcal{T}}_d \cap \mathcal{U}(x_i\,;r_i) \neq \varnothing\right\} \le c\sum_{i\ge 1} r_i^{d/2} \le 2c\mathcal{H}_{d/2}(G).$$

This completes our proof of the first bound in (5.1).

In order to prove the lower bound for $\bar{\mathcal{T}}_d$ note that $\bar{\mathcal{T}}_d$ intersects E if and only if $\hat{\mathcal{T}}_d$ intersects $[0\,,1] \times E$. In (3.1) we proved that if E is a one-dimensional, compact set, then $\mathrm{Cap}_{d/2}([0\,,1] \times E) = \mathrm{Cap}_{(d-2)/2}(E)$. A similar proof shows that the same fact holds in any dimension, whence follows the desired lower bound for the probability that $\bar{\mathcal{T}}_d$ intersects E.

To conclude, it suffices to prove that

$$\mathcal{H}_{d/2}([0\,,1] \times E) > 0 \implies \mathcal{H}_{(d-2)/2}(E) > 0.$$

But this follows readily from Frostman's lemma [Appendix A.3]. Indeed, the positivity of $\mathcal{H}_{d/2}([0\,,1] \times E)$ is equivalent to the existence of $\mu \in \mathcal{P}([0\,,1] \times E)$ and a constant c such that the μ-measure of all balls [in $\mathbb{R}^3$] of radius $r > 0$ is at most $cr^{d/2}$. Define $\bar{\mu}(C) := \mu([0\,,1] \times C)$ for all Borel sets $C \subseteq \mathbb{R}^2$. Evidently, $\bar{\mu} \in \mathcal{P}(E)$, and a covering argument, together with the Frostman property of μ, imply that $\bar{\mu}$ of all two-dimensional balls of radius $r > 0$ is at most $cr^{(d/2)-1}$. Another application of the Frostman lemma finishes the proof. □

6. Proof of Theorem 1.2

Define for all $s > 0$, every $\omega \in \Omega$, and all Borel sets $I \subseteq \mathbb{R}_+$,

$$T_d^I(s)(\omega) := \{t \in I :\ B(s\,,t)(\omega) = 0\}\,.$$

Equivalently, $T_d^I(s) = B^{-1}\{0\} \cap (\{s\} \times (0\,,\infty)) \cap I$. It suffices to prove that for all closed intervals $I \subset (0\,,\infty)$,

$$\dim_{\mathrm{H}} T_d^I(s) = 0 \quad \text{for all } s > 0 \quad \text{a.s.} \tag{6.1}$$

[N.B.: The order of the quantifiers!]. This, in turn, proves that

$$\dim_{\mathrm{H}} T_d^{\mathbb{R}_+}(s) = \sup_I \dim_{\mathrm{H}} T_d^I(s) = 0 \quad \text{for all } s > 0,$$

where the supremum is taken over all closed intervals $I \subset (0\,,\infty)$ with rational endpoints. Theorem 1.2 follows suit. Without loss of much generality, we prove (1.3) for $I := [1\,,2]$; the more general case follows from this after a change of notation. To simplify the exposition, we write

$$T_d(s) := T_d^{[1,2]}(s).$$

Consider the following events:

$$\mathbf{G}_k(n) := \left\{ \sup_{\substack{1\le s,t\le 2\\ s\le u\le s+(1/k)\\ t\le v\le t+(1/k)}} |B(u,v)-B(s\,,t)| \le n\left(\frac{\log k}{k}\right)^{1/2} \right\},$$

where $k,n\ge 3$ are integers. We will use the following folklore lemma. A generalization is spelled out explicitly in Lacey [17, Eq. (3.8)].

Lemma 6.1. *For all $\gamma>0$ there exists $n_0=n_0(\gamma)$ such that for all $n,k\ge n_0$,*

$$\mathrm{P}\left(\mathbf{G}_k(n)\right)\ge 1-k^{-\gamma}.$$

Next we mention a second folklore result.

Lemma 6.2. *Let $\{W(t)\}_{t\ge 0}$ denote a standard Brownian motion in $\mathbb{R}^d$. Then, there exists a constant c such that for all integers $m\ge 1$ and $1\le r_1\le r_2\le\ldots\le r_m\le 2$,*

$$\mathrm{P}\left\{\max_{1\le i\le m}|W(r_i)|\le\epsilon\right\}\le c\epsilon^d\prod_{2\le i\le m}\left(\frac{\epsilon}{(r_i-r_{i-1})^{1/2}}\wedge 1\right)^d.$$

Proof. If $|W(r_i)|\le\epsilon$ for all $i\le m$ then $|W(r_1)|\le\epsilon$, and $|W(r_i)-W(r_{i-1})|\le 2\epsilon$ for all $2\le i\le m$. Therefore,

$$\mathrm{P}\left\{\max_{1\le i\le m}|W(r_i)|\le\epsilon\right\}\le \mathrm{P}\left\{|W(r_1)|\le\epsilon\right\}\prod_{2\le i\le m}\mathrm{P}\left\{|W(r_i-r_{i-1})|\le 2\epsilon\right\}.$$

A direct computation yields the lemma from this. □

Now define

$$I_{i,j}(k) := \left[1+\frac{i}{k}\,,1+\frac{(i+1)}{k}\right]\times\left[1+\frac{j}{k}\,,1+\frac{(j+1)}{k}\right],$$

where i and j can each run through $\{0,\ldots,k-1\}$, and $k\ge 1$ is an integer. We say that $I_{i,j}(k)$ is *good* if $I_{i,j}(k)\cap B^{-1}\{0\}\neq\varnothing$. With this in mind, we define

$$N_{i,k} := \sum_{0\le j\le k-1}\mathbf{1}_{\{I_{i,j}(k)\text{ is good}\}}.$$

Lemma 6.3. *Suppose $d\in\{2\,,3\}$. Then, for all $\gamma>0$ there exists $\alpha=\alpha(d\,,\gamma)>1$ large enough that*

$$\max_{0\le i\le k-1}\mathrm{P}\left\{N_{i,k}\ge\alpha(\log k)^{(8-d)/2}\right\}=O\left(k^{-\gamma}\right),$$

as k tends to infinity.

Proof. On $\mathbf{G}_k(n)$ we have the set-wise inclusion,

$$\{I_{i,j}(k)\text{ is good}\}\subseteq\left\{\left|B\left(1+\frac{i}{k}\,,\ 1+\frac{j}{k}\right)\right|\le n\left(\frac{\log k}{k}\right)^{1/2}\right\}.$$

Therefore, for all integers $p \geq 1$,

$$\begin{aligned}
&\mathrm{E}\left[N_{i,k}^p \ ; \ \mathbf{G}_k(n)\right] \\
&\leq \sum_{0\leq j_1, \cdots, j_p\leq k-1} \cdots \sum \mathrm{P}\left\{\max_{1\leq \ell\leq p}\left|B\left(1+\frac{i}{k} \ , \ 1+\frac{j_\ell}{k}\right)\right| \leq n\left(\frac{\log k}{k}\right)^{1/2}\right\} \\
&= \sum_{0\leq j_1, \cdots, j_p\leq k-1} \cdots \sum \mathrm{P}\left\{\max_{1\leq \ell\leq p}\left|\left(1+\frac{i}{k}\right)^{1/2} W\left(1+\frac{j_\ell}{k}\right)\right| \leq n\left(\frac{\log k}{k}\right)^{1/2}\right\} \\
&\leq p! \sum_{0\leq j_1\leq\cdots\leq j_p\leq k-1} \cdots \sum \mathrm{P}\left\{\max_{1\leq \ell\leq p}\left|W\left(1+\frac{j_\ell}{k}\right)\right| \leq n\left(\frac{\log k}{k}\right)^{1/2}\right\},
\end{aligned}$$

where W denotes a standard d-dimensional Brownian motion. Because the latter quantity does not depend on the value of i, Lemma 6.2 shows that

$$\begin{aligned}
&\max_{0\leq i\leq k-1} \mathrm{E}\left[N_{i,k}^p \ ; \ \mathbf{G}_k(n)\right] \\
&\qquad \leq cp!n^{pd}\left(\frac{\log k}{k}\right)^{d/2} \sum_{0\leq j_1\leq\cdots\leq j_p\leq k-1} \cdots \sum \prod_{2\leq \ell\leq p}\left(\frac{\log k}{j_\ell - j_{\ell-1}}\right)^{d/2},
\end{aligned}$$

for all k large, where we are interpreting $1/0$ as one.

Now first consider the case $d = 3$. We recall our (somewhat unusual) convention about $1/0$, and note that

$$\sum_{0\leq j_1\leq\cdots\leq j_p\leq k-1} \cdots \sum \prod_{2\leq \ell\leq p} \frac{1}{(j_\ell - j_{\ell-1})^{3/2}} \leq k\left(\sum_{l\geq 0}\frac{1}{l^{3/2}}\right)^{p-1}. \tag{6.2}$$

Therefore, when $d = 3$ we can find a constant c_1 — independent of $(p\,,k)$ — such that

$$\max_{0\leq i\leq k-1} \mathrm{E}\left[N_{i,k}^p \ ; \ \mathbf{G}_k(n)\right] \leq p!\,\frac{(c_1\log k)^{3p/2}}{k^{1/2}} \leq p!\,(c_1\log k)^{3p/2}. \tag{6.3}$$

By enlarging c_1, if need be, we find that this inequality is valid for all $k \geq 1$. This proves readily that

$$\max_{0\leq i\leq k-1} \mathrm{E}\left[\exp\left(\frac{N_{i,k}}{2(c_1\log k)^{3/2}}\right) \ ; \ \mathbf{G}_k(n)\right] \leq \sum_{p\geq 0} 2^{-p} = 2. \tag{6.4}$$

Therefore, Chebyshev's inequality implies that for all $i, k, p \geq 1$ and $a > 0$,

$$\max_{0\leq i\leq k-1} \mathrm{P}\left\{N_{i,k} \geq 2\gamma c_1^{3/2}(\log k)^{5/2} \ ; \ \mathbf{G}_k(n)\right\} \leq 2k^{-\gamma}. \tag{6.5}$$

Note that c_1 may depend on n. But we can choose n large enough — once and for all — such that the probability of the complement of $\mathbf{G}_k(n)$ is at most $k^{-\gamma}$ (Lemma 6.1). This proves the lemma in the case that $d = 3$.

The case $d = 2$ is proved similarly, except (6.2) is replaced by

$$\sum \cdots \sum_{0 \le j_1 \le \cdots \le j_p \le k-1} \prod_{2 \le \ell \le p} \frac{1}{j_\ell - j_{\ell-1}} \le k \left(\sum_{0 \le l \le k} \frac{1}{l} \right)^{p-1} \le k(c_2 \log k)^{p-1},$$

where c_2 does not depend on $(k\,, p)$, and [as before] $1/0 := 1$. Equation (6.3), when $d = 2$, becomes:

$$\max_{0 \le i \le k-1} \mathrm{E}\left[N_{i,k}^p \ ; \ \mathbf{G}_k(n) \right] \le p!(c_2 \log k)^p.$$

This forms the $d = 2$ version of (6.4):

$$\max_{0 \le i \le k-1} \mathrm{E}\left[\exp\left(\frac{N_{i,k}}{2c_2 \log k} \right) \ ; \ \mathbf{G}_k(n) \right] \le 2.$$

Thus, (6.5), when $d = 2$, becomes

$$\max_{0 \le i \le k-1} \mathrm{P}\left\{ N_{i,k} \ge 2\gamma c_2 (\log k)^2 \ ; \ \mathbf{G}_k(n) \right\} \le 2k^{-\gamma}.$$

The result follows from this and Lemma 6.1 after we choose and fix a sufficiently large n. □

Estimating $N_{i,k}$ is now a simple matter, as the following shows.

Lemma 6.4. *If $d \in \{2\,, 3\}$, then with probability 1,*

$$\max_{0 \le i \le k-1} N_{i,k} = O\left((\log k)^{(8-d)/2} \right) \qquad (k \to \infty).$$

Proof. By Lemma 6.3, there exists $\alpha > 0$ so large that for all $k \ge 1$ and $0 \le i \le k-1$, $\mathrm{P}\{N_{i,k} \ge \alpha (\log k)^{(8-d)/2}\} \le \alpha k^{-3}$. Consequently,

$$\mathrm{P}\left\{ \max_{0 \le i \le k-1} N_{i,k} \ge \alpha (\log k)^{(8-d)/2} \right\} \le \alpha k^{-2}.$$

The lemma follows from this and the Borel–Cantelli lemma. □

We are ready to prove Theorem 1.2. As was mentioned earlier, it suffices to prove (6.1), and this follows from our next result.

Proposition 6.5. *Fix $d \in \{2\,, 3\}$ and define the measure-function*

$$\Phi(x) := \left[\log_+(1/x) \right]^{-(8-d)/2}.$$

Then, $\sup_{1 \le s \le 2} \mathcal{H}_\Phi(T_d(s)) < \infty$ a.s.

The reason is provided by the following elementary lemma whose proof is omitted.

Lemma 6.6. *Suppose φ is a measure function such that $\liminf_{x \downarrow 0} x^{-\alpha} \varphi(x) = \infty$ for some $\alpha > 0$. Then, for all Borel sets $A \subset \mathbb{R}^n$,*

$$\mathcal{H}_\varphi(A) < \infty \implies \mathcal{H}_\alpha(A) < \infty \implies \dim_{\mathrm{H}} A \le \alpha.$$

Now we prove Proposition 6.5.

Proof of Proposition 6.5. We can construct a generous cover of $T_d(s)$ as follows: For all irrational $s \in [i/k\,, (i+1)/k]$, we cover $T_d(s)$ intervals of the form

$$\left[1 + \frac{j}{k}\,, 1 + \frac{(j+1)}{k}\right],$$

where j can be any integer in $\{0, \ldots, k-1\}$ as long as $I_{i,j}(k)$ is good. Therefore, for any measure-function φ,

$$\sup_{\substack{1 \le s \le 2: \\ s \text{ is irrational}}} \mathcal{H}_\varphi^{(1/k)}\left(T_d(s)\right) \le \varphi(1/k) \max_{0 \le i \le k-1} N_{i,k}.$$

Now we choose the measure-function $\varphi(x) := \Phi(x)$ and let $k \to \infty$ to find that $\mathcal{H}_\Phi(T_d(s))$ is finite, uniformly over all irrational $s \in [1\,, 2]$. The case of rational s's is simpler to analyse. Indeed, $T_d(s) = \varnothing$ a.s. for all rational $s \in [1\,, 2]$. This is because d-dimensional Brownian motion ($d \in \{2\,, 3\}$) does not hit zero. □

Remark 6.7. The form of Lemma 6.4 changes dramatically when $d = 1$. Indeed, one can adjust the proof of Lemma 6.4 to find that a.s.,

$$\max_{0 \le i \le k-1} N_{i,k} = O\left(k^{1/2} (\log k)^{3/2}\right) \qquad (k \to \infty).$$

This yields fairly readily that the upper Minkowski dimension [written as $\dim_{\mathrm{M}}$] of $T_1(s)$ is at most $1/2$ simultaneously for all $s > 0$. Let $\dim_{\mathrm{P}}$ denote the packing dimension, and recall (B.3). Then, the preceding and the theorem of Penrose [24] together prove that almost surely,

$$\dim_{\mathrm{H}} T_1(s) = \dim_{\mathrm{P}} T_1(s) = \dim_{\mathrm{M}} T_1(s) = \frac{1}{2} \quad \text{for all } s > 0.$$

7. On rates of escape

Throughout this section, we choose and fix a non-decreasing and measurable function $\psi : (0\,, \infty) \to (0\,, \infty)$ such that $\lim_{t \to \infty} \psi(t) = \infty$. Define, for all Borel-measurable sets $F \subset \mathbb{R}$,

$$\Upsilon_F(\psi) := \int_1^\infty \left[\frac{\mathrm{K}_F\left(1/\psi(x)\right)}{\left(\psi(x)\right)^{(d-2)/2}} \wedge 1\right] \frac{dx}{x},$$

where K_F denotes the Kolmogorov entropy of F; see Appendix B.1 for a definition.

Theorem 7.1. *If $d \ge 3$, then for all non-random, compact sets $F \subset (0\,, \infty)$, the following holds with probability* 1*:*

$$\liminf_{t \to \infty} \inf_{s \in F} \left(\frac{\psi(t)}{t}\right)^{1/2} |B(s\,, t)| = \begin{cases} 0 & \text{if } \Upsilon_F(\psi) = \infty, \\ \infty & \text{otherwise.} \end{cases} \tag{7.1}$$

Remark 7.2. Although the infimum over all $s \in E$ is generally an uncountable one, measurability issues do not arise. Our proof actually shows that the event in (7.1) is a subset of a null set. Thus, we are assuming tacitly that the underlying probability space is complete. This convention applies to the next theorem as well.

Definition 7.3. Let $F \subset (0\,,\infty)$ be non-random and compact, and $\psi : (0\,,\infty) \to (0\,,\infty)$ measurable and non-decreasing. Then we say that $(F, \psi) \in \mathrm{FIN}_{loc}$ if there exists a denumerable decomposition $F = \cup_{n=1}^{\infty} F_n$ of F in terms of closed intervals $F_1, F_2, \ldots$ — all with rational end-points — such that $\Upsilon_{F_n}(\psi) < \infty$ for all $n \geq 1$.

This brings us to the main theorem of this section. Its proof is a little delicate because we have to get three different estimates, each of which is valid only on a certain scale. This proof is motivated by the earlier work of the author with David Levin and Pedro Méndez [15].

Theorem 7.4. *If $d \geq 3$, then for all non-random, compact sets $F \subset (0\,,\infty)$, the following holds with probability 1:*

$$\inf_{s\in F} \liminf_{t\to\infty} \left(\frac{\psi(t)}{t}\right)^{1/2} |B(s\,,t)| = \begin{cases} 0 & \text{if } (F,\psi) \notin \mathrm{FIN}_{loc}, \\ \infty & \text{otherwise.} \end{cases}$$

The key estimate, implicitly referred to earlier, is the following.

Theorem 7.5. *If $d \geq 3$ then there exists a constant c such that for all non-random compact sets $F \subseteq [1\,,2]$ and $0 < \epsilon < 1$,*

$$\frac{1}{c}\left[\epsilon^{d-2}\mathrm{K}_F(\epsilon^2) \wedge 1\right] \leq \mathrm{P}\left\{\inf_{s\in F}\inf_{1\leq t\leq 2} |B(s\,,t)| \leq \epsilon\right\} \leq c\left[\epsilon^{d-2}\mathrm{K}_F(\epsilon^2)\wedge 1\right].$$

Let us mention also the next result without proof; it follows upon combining Theorems 4.1 and 4.2 of our collaborative effort with Robert Dalang [3], together with Brownian scaling:

Lemma 7.6. *If $d \geq 3$, then there exists c such that for all $1 \leq a < b \leq 2$, $0 < \epsilon < 1$, and $n \geq 1$ such that $(b-a) \geq c\epsilon^2$,*

$$\frac{1}{c}(b-a)^{(d-2)/2} \leq \mathrm{P}\left\{\inf_{\substack{a\leq s\leq b\\ 1\leq t\leq 2}} |B(s\,,t)| \leq \epsilon\right\} \leq c(b-a)^{(d-2)/2}.$$

Remark 7.7. Dalang and Khoshnevisan [3] state this explicitly for $d \in \{3\,,4\}$. However, the key estimates are their Lemmas 2.1 and 2.6, and they require only that $d > 2$.

Proof of Theorem 7.5 *(The Upper Bound).* Fix $n \geq 1$. Define $I_j := [j/n\,,(j+1)/n)$, and let $\chi_j = 1$ if $I_j \cap F \neq \varnothing$ and $\chi_j = 0$ otherwise. Then in accordance with Lemma 7.6,

$$\begin{aligned}
&\mathrm{P}\left\{\inf_{s\in F}\inf_{1\leq t\leq 2}|B(s\,,t)| \leq \frac{1}{(cn)^{1/2}}\right\} \\
&\qquad\leq \sum_{n\leq j\leq 2n-1} \mathrm{P}\left\{\inf_{s\in I_j}\inf_{1\leq t\leq 2}|B(s\,,t)| \leq \frac{1}{(cn)^{1/2}}\right\}\chi_j \\
&\qquad\leq cn^{-(d-2)/2}\mathrm{M}_n(F).
\end{aligned}$$

This, in turn, is bounded above by $cn^{-(d-2)/2}\mathrm{K}_F(1/n)$; see (B.1). The lemma follows in the case that $\epsilon = (cn)^{-1/2}$. The general case follows from a *monotonicity argument*, which we rehash (once) for the sake of completeness.

Suppose $(c(n+1))^{-1/2} \le \epsilon \le (cn)^{-1/2}$. Then,

$$\begin{aligned}\mathrm{P}\left\{\inf_{s\in F}\inf_{1\le t\le 2}|B(s\,,t)| \le \epsilon\right\} &\le \mathrm{P}\left\{\inf_{s\in F}\inf_{1\le t\le 2}|B(s\,,t)| \le \frac{1}{(cn)^{1/2}}\right\}\\ &\le cn^{-(d-2)/2}\mathrm{K}_F(1/n)\\ &\le c\epsilon^{d-2}\mathrm{K}_F(c\epsilon^2).\end{aligned}$$

Equation (B.2) implies that $\mathrm{K}_F(c\epsilon^2) = O(\mathrm{K}_F(\epsilon^2))$ as $\epsilon \to 0$, and finishes our proof of the upper bound. $\square$

Before we prove the lower bound we mention a heuristic argument. If, in Lemma 7.6, the condition "$(b-a) \ge c\epsilon^2$" is replaced by $(b-a) \ll \epsilon^2$, then the bounds both change to ϵ^{d-2}. This is the probability that a single Brownian motion hits $\mathcal{B}(0;\epsilon)$ some time during $[1\,,2]$; compare with Lemma C.1. This suggests that the "correlation length" among the slices is of order ϵ^2. That is, slices that are within ϵ^2 of one another behave much the same; those that are further apart than ϵ^2 are nearly independent. We use our next result in order to actually prove the latter heuristic.

Proposition 7.8. *If $d \ge 3$, then there exists a constant c such that for all $1 \le s, u \le 2$ and $0 < \epsilon < 1$, if $|u-s| \ge \epsilon^2$, then*

$$\mathrm{P}\left\{\inf_{1\le t\le 2}|B(s\,,t)| \le \epsilon\ ,\ \inf_{1\le v\le 2}|B(u\,,v)| \le \epsilon\right\} \le c\epsilon^{d-2}|u-s|^{(d-2)/2}.$$

Proof. Without loss of generality we may choose and fix $2 \ge u > s \ge 1$. Now the processes $\{B(s\,,t)\}_{t\ge 0}$ and $\{B(u\,,v)\}_{v\ge 0}$ can be decomposed as follows:

$$B(s\,,t) = s^{1/2}Z(t), \qquad B(u\,,v) = s^{1/2}Z(v) + (u-s)^{1/2}W(v),$$

where W and Z are independent d-dimensional Brownian motions. Thus, we are interested in estimating the quantity p_ϵ, where

$$\begin{aligned}p_\epsilon := \mathrm{P}&\left\{\inf_{1\le t\le 2}|Z(t)| \le \frac{\epsilon}{s^{1/2}}\ ,\ \inf_{1\le v\le 2}\left|Z(v) + \left(\frac{u-s}{s}\right)^{1/2}W(v)\right| \le \frac{\epsilon}{s^{1/2}}\right\}\\ &\le \mathrm{P}\left\{\inf_{1\le t\le 2}|Z(t)| \le \epsilon\ ,\ \inf_{1\le v\le 2}\left|Z(v) + (u-s)^{1/2}W(v)\right| \le \epsilon\right\}.\end{aligned}$$

The proposition follows from Lemma C.2 in Appendix C below. $\square$

Proof of Theorem 7.5 *(The Lower Bound).* We make a discretization argument, once more. Let $n := \mathrm{K}_F(\epsilon^2)$ and find maximal Kolmogorov points $s_1 < \cdots < s_n$ — all in F — such that $s_{i+1} - s_i \ge \epsilon^2$ for all $1 \le i < n$. Define

$$J_\epsilon(n) := \sum_{1\le i\le n}\mathbf{1}_{\{|B(s_i,t)|\le\epsilon \text{ for some } t\in[1,2]\}}.$$

According to Lemma C.1,

$$\frac{1}{c} n\epsilon^{d-2} \le \mathrm{E}\left[J_\epsilon(n)\right] \le cn\epsilon^{d-2}. \tag{7.2}$$

On the other hand, the condition $|s_j - s_i| \ge \epsilon^2$ and Proposition 7.8 together ensure that

$$\mathrm{E}\left[(J_\epsilon(n))^2\right] \le \mathrm{E}[J_\epsilon(n)] + c\left(\mathrm{E}[J_\epsilon(n)]\right)^2.$$

Now to prove the lower bound we first assume that $n\epsilon^{d-2} \le 1$. The previous display implies then that $\mathrm{E}[(J_\epsilon(n))^2] \le c\mathrm{E}[J_\epsilon(n)]$. Combine this inequality with (7.2) and the Paley–Zygmund inequality to find that

$$\mathrm{P}\left\{\inf_{s\in F}\inf_{1\le t\le 2} |B(s\,,t)| \le \epsilon\right\} \ge \mathrm{P}\left\{J_\epsilon(n) > 0\right\} \ge \frac{\left(\mathrm{E}[J_\epsilon(n)]\right)^2}{\mathrm{E}[(J_\epsilon(n))^2]} \ge cn\epsilon^{d-2}.$$

On the other hand, if $n\epsilon^{d-2} \ge 1$, then the left-hand side is bounded away from zero, by a similar bound. This is the desired result. □

Lemma 7.9. *Let $d \ge 3$, and $f : [1\,,2] \to \mathbb{R}^d$ be a fixed, non-random, measurable function. Then there exists a constant c such that for all integers $1 \le k \le n$*

$$\mathrm{P}\left\{\inf_{\substack{1\le s\le k/n\\ 1\le t\le 2}} |B(s\,,t) - f(s)| \le \frac{1}{n^{1/2}}\right\} \le c\left(kn^{-(d-2)/2} + \sum_{n\le i\le n+k-1} \left(\Omega_{i,n}(f)\right)^{d-2}\right),$$

where for all continuous functions h,

$$\Omega_{i,n}(h) := \sup_{i/n\le t\le (i+1)/n} |h(t) - h(i/n)|.$$

Proof. Lemma 7.9 holds for similar reasons as does Proposition 7.8, but is simpler to prove. Indeed, the probability in question is at most

$$\sum_{n\le i\le n+k-1} \mathrm{P}\left\{\inf_{i/n\le s\le (i+1)/n} |B(s\,,t) - f(s)| \le \frac{1}{n^{1/2}}\right\}.$$

This, in turn, is less than or equal to

$$\sum_{n\le i\le n+k-1} \mathrm{P}\left\{\inf_{1\le t\le 2} \left|B(\tfrac{i}{n}\,,t)\right| \le \frac{1}{n^{1/2}} + \sup_{1\le t\le 2} \Omega_{i,n}(B(\bullet\,,t)) + \Omega_{i,n}(f)\right\}.$$

By the Markov property, $B((i/n)\,,\bullet)$ is a d-dimensional Brownian motion that is independent of $\sup_{1\le t\le 2}\Omega_{i,n}(B(\bullet\,,t))$. Standard modulus-of-continuity bounds show that the $L^{d-2}(\mathrm{P})$-norm of $\sup_{1\le t\le 2}\Omega_{i,n}(B(\bullet\,,t))$ is at most a constant times $n^{-(d-2)/2}$; the details will be explained momentarily. Since $(i/n) \ge 1$, these observations, in conjunction with Lemma C.1 [Appendix C] imply the lemma. It remains to prove that there exists a c such that for all $n \ge 1$,

$$\max_{n\le i\le 2n} \mathrm{E}\left[\sup_{1\le t\le 2} \left(\Omega_{i,n}(B(\bullet\,,t))\right)^{d-2}\right] \le cn^{-(d-2)/2}. \tag{7.3}$$

Choose and fix $n \geq 1$, $n \leq i \leq 2n$, and $v \in [i/n\,,(i+1)/n]$. Then the process $t \mapsto B(v\,,t) - B(i/n\,,t)$ is manifestly a martingale with respect to the filtration generated by the infinite-dimensional process $t \mapsto B(\bullet\,,t)$. Consequently, $T \mapsto \sup_{1\leq t\leq T}(\Omega_{i,n}(B(\bullet\,,t)))^{d-2}$ is a sub-martingale, and (7.3) follows from Doob's inequality and Brownian-sheet scaling. This completes our proof. $\square$

Lemma 7.9, together with a monotonicity argument, implies the following.

Lemma 7.10. *Let $d \geq 3$, and $f : [1\,,2] \to \mathbb{R}^d$ be a fixed, non-random, measurable function. Then there exists a constant c such that for all $1 \leq a \leq 2$ and $0 < \epsilon < 1$,*

$$\mathrm{P}\left\{\inf_{a\leq s\leq a+\epsilon^2}\inf_{1\leq t\leq 3}|B(s\,,t) - f(s)| \leq \epsilon\right\} \leq c\left(\epsilon^{d-2} + \sup_{a\leq u\leq a+\epsilon^2}|f(u) - f(a)|^{d-2}\right),$$

Proof of Theorem 7.1. First, assume that $\Upsilon(\psi) < \infty$; this is the *first half.*

Define for all $n = 0, 1, 2, \ldots$,

$$\psi_n := \psi(2^n),$$
$$\mathbf{A}_n := \left\{\inf_{s\in F}\inf_{2^n\leq t\leq 2^{n+1}}|B(s\,,t)| \leq (2^n/\psi_n)^{1/2}\right\}.$$

We combine Theorem 7.5 with the Brownian-sheet scaling to deduce the following:

$$\frac{1}{c}\left[\psi_n^{-(d-2)/2}\mathrm{K}_F(1/\psi_n) \wedge 1\right] \leq \mathrm{P}(\mathbf{A}_n) \leq c\left[\psi_n^{-(d-2)/2}\mathrm{K}_F(1/\psi_n) \wedge 1\right]. \qquad (7.4)$$

After doing some algebra we find that because $\Upsilon_F(\psi)$ is finite, $\sum_{n\geq 1}\mathrm{P}(\mathbf{A}_n) < \infty$. By the Borel–Cantelli lemma,

$$\liminf_{n\to\infty}\left(\frac{\psi_n}{2^n}\right)^{1/2}\inf_{s\in F}\inf_{2^n\leq t\leq 2^{n+1}}|B(s\,,t)| \geq 1 \quad \text{a.s.}$$

If $2^n \leq t \leq 2^{n+1}$, then $(\psi_n/2^n)^{1/2} \leq (2\psi(t)/t)^{1/2}$. It follows that almost surely,

$$\liminf_{t\to\infty}\left(\frac{\psi(t)}{t}\right)^{1/2}\inf_{s\in F}|B(s\,,t)| \geq \frac{1}{2^{1/2}}.$$

But if $\Upsilon_F(\psi)$ is finite then so is $\Upsilon_F(r\psi)$ for any $r > 0$; see (B.2). Therefore, we can apply the preceding to $r\psi$ in place of ψ, and then let $r \to 0$ to find that

$$\Upsilon_F(\psi) < \infty \implies \liminf_{t\to\infty}\left(\frac{\psi(t)}{t}\right)^{1/2}\inf_{s\in F}|B(s\,,t)| = \infty \quad \text{a.s.}$$

This concludes the proof of the first half.

For the *second half* we assume that $\Upsilon_F(\psi) = \infty$. The preceding analysis proves that $\sum_{n\geq 1}\mathrm{P}(\mathbf{A}_n) = \infty$. According to the Borel–Cantelli lemma, it suffices to prove that

$$\limsup_{N\to\infty}\frac{\sum\sum_{1\leq n<m\leq N}\mathrm{P}(\mathbf{A}_n \cap \mathbf{A}_m)}{\left(\sum_{1\leq n\leq N}\mathrm{P}(\mathbf{A}_n)\right)^2} < \infty. \qquad (7.5)$$

Define for all integers $n \geq 1$, and all $s, t \geq 0$,

$$\mathcal{A}_n := \text{ the } \sigma\text{-algebra generated by } \{B(\bullet\,, v)\}_{0 \leq v \leq 2^n},$$
$$\Delta_n(s\,, t) := B(s\,, t + 2^n) - B(s\,, 2^n).$$

The Markov properties of the Brownian sheet imply that whenever $m > n \geq 1$: (i) Δ_m is a Brownian sheet that is independent of $\mathcal{A}_n$; and (ii) $\mathbf{A}_n \in \mathcal{A}_n$. Thus, we apply these properties in conjunction with Brownian-sheet scaling to find that a.s., $\mathrm{P}(\mathbf{A}_m \,|\, \mathcal{A}_n)$ is equal to

$$\begin{aligned}
&\mathrm{P}\left(\inf_{s \in F} \inf_{2^m - 2^n \leq t \leq 2^{m+1} - 2^n} |\Delta_n(s\,, t) - B(s\,, 2^n)| \leq \left(\frac{2^m}{\psi_m}\right)^{1/2} \middle|\, \mathcal{A}_n \right) \\
&\quad = \mathrm{P}\left(\inf_{1 \leq t \leq (2^{m+1} - 2^n)/\alpha} \left| \Delta_n(s\,, t) - \frac{B(s\,, 2^n)}{\alpha^{1/2}} \right| \leq \left(\frac{2^m}{\alpha\psi_m}\right)^{1/2} \,\middle|\, \mathcal{A}_n \right),
\end{aligned}$$

where $\alpha := 2^m - 2^n$. Because $m \geq n + 1$, $(2^{m+1} - 2^n)/\alpha \leq 3$ and $2^m/\alpha \leq 2$. Therefore, almost surely,

$$\mathrm{P}\left(\mathbf{A}_m \,|\, \mathcal{A}_n\right) \leq \mathrm{P}\left(\inf_{s \in F} \inf_{1 \leq t \leq 3} \left| \Delta_n(s\,, t) - \frac{B(s\,, 2^n)}{\alpha^{1/2}} \right| \leq \left(\frac{2}{\psi_m}\right)^{1/2} \,\middle|\, \mathcal{A}_n \right).$$

We can cover E with at most $K := \mathrm{M}_{[2/\psi_m]}(F)$ intervals of the form $I_i := [i/\ell\,, (i+1)/\ell]$, where $\ell := [\psi_m/2]$. Having done this, a simple bound, together with Lemma 7.10 yield the following: With probability one, $\mathrm{P}(\mathbf{A}_m \,|\, \mathcal{A}_n)$ is bounded above by

$$\begin{aligned}
&\sum_{1 \leq i \leq K} \mathrm{P}\left(\inf_{s \in I_i} \inf_{1 \leq t \leq 3} \left| \Delta_n(s\,, t) - \frac{B(s\,, 2^n)}{\alpha^{1/2}} \right| \leq \left(\frac{2}{\psi_m}\right)^{1/2} \,\middle|\, \mathcal{A}_n \right) \\
&\qquad\qquad \leq cK\left(\psi_m^{-(d-2)/2} + \Omega \right),
\end{aligned}$$

where

$$\begin{aligned}
\Omega &:= \alpha^{-(d-2)/2} \max_{1 \leq i \leq K} \mathrm{E}\left[\sup_{s \in I_i} |B(s\,, 2^n) - B(i/\ell\,, 2^n)|^{d-2} \right] \\
&= \alpha^{-(d-2)/2} 2^{n(d-2)/2} \mathrm{E}\left[\sup_{0 \leq s \leq 1/\ell} |B(s\,, 1)|^{d-2} \right] \\
&= c\alpha^{-(d-2)/2} 2^{n(d-2)/2} \ell^{-(d-2)/2}.
\end{aligned}$$

Therefore, the bound $2^n/\alpha \leq 1$ implies that $\Omega \leq c\ell^{-(d-2)/2} \leq c\psi_m^{-(d-2)/2}$. On the other hand, by (B.1) and (B.2), $K \leq \mathrm{K}_F(1/\psi_m)$. Therefore, the preceding paragraph and (7.4) together imply that $\mathrm{P}(\mathbf{A}_m \,|\, \mathcal{A}_n) \leq c\mathrm{P}(\mathbf{A}_m)$ a.s., where c does not depend on $(n\,, m\,, \omega)$. Therefrom, we conclude that $\mathrm{P}(\mathbf{A}_m \,|\, \mathbf{A}_n) \leq c\mathrm{P}(\mathbf{A}_m)$, whence (7.5). □

We are ready to prove Theorem 7.4.

Proof of Theorem 7.4. Suppose, first, that $(F, \psi) \in \mathrm{FIN}_{loc}$. According to Theorem 7.1, we can write $F = \cup_{n\geq 1} F_n$ a.s., where the F_n's are closed intervals with rational end-points, such that

$$\inf_{s\in F_n} \liminf_{t\to\infty} \left(\frac{\psi(t)}{t}\right)^{1/2} |B(s\,,t)| = \infty \quad \text{for all } n \geq 1.$$

This proves that a.s.,

$$\inf_{s\in F} \liminf_{t\to\infty} \left(\frac{\psi(t)}{t}\right)^{1/2} |B(s\,,t)| = \infty,$$

and this is half of the assertion of the theorem.

Conversely, suppose $(F, \psi) \notin \mathrm{FIN}_{loc}$. Then, given any decomposition $F = \cup_{n\geq 1} F_n$ in terms of closed, rational intervals $F_1, F_2, \ldots$,

$$\liminf_{t\to\infty} \inf_{s\in F_n} \left(\frac{\psi(t)}{t}\right)^{1/2} |B(s\,,t)| = 0 \quad \text{for all } n \geq 1. \tag{7.6}$$

Define for all $k, n \geq 1$,

$$O_{k,n} := \left\{ s > 0 : \inf_{t\geq k} \left[\left(\frac{\psi(t)}{t}\right)^{1/2} |B(s\,,t)| \right] < \frac{1}{n} \right\}.$$

Then (7.6) implies that every $O_{k,n}$ is relatively open and everywhere dense in F a.s. By the Baire category theorem, $\cap_{k,n\geq 1} O_{k,n}$ has the same properties, and this proves the theorem. □

With Theorem 7.4 under way, we can finally derive Theorem 1.6 of the Introduction, and conclude this section.

Proof of Theorem 1.6. Throughout, define for all $\alpha > 0$,

$$\psi_\alpha(x) := \left[\log_+(x)\right]^{2/\alpha} \qquad \text{for all } x > 0.$$

Note that for any ψ, as given by Theorem 7.4, and for all $\nu > 0$,

$$\Upsilon_F(\psi) < \infty \quad \text{iff} \quad \int_1^\infty \left[\frac{\mathrm{K}_F\left(1/\psi(x)\right)}{\left(\psi(x)\right)^{(d-2)/2}} \wedge \nu \right] \frac{dx}{x} < \infty.$$

Therefore,

$$\begin{aligned} &\text{if } \mathrm{K}_F(\epsilon) = O\left(\epsilon^{-(d-2)/2}\right) \quad (\epsilon \to 0), \quad \text{then} \\ &\Upsilon_F(\psi) < \infty \text{ if and only if } \int_1^\infty \frac{\mathrm{K}_F\left(1/\psi(x)\right)}{x\left(\psi(x)\right)^{(d-2)/2}}\, dx < \infty. \end{aligned} \tag{7.7}$$

Suppose $d \geq 4$. Then $K_F(\epsilon) \leq c\epsilon^{-1}$, and so by (7.7) and a little calculus,

$$\Upsilon_F(\psi_\alpha) < \infty \text{ if and only if } \int_1^\infty \frac{\mathrm{K}_F(1/s)}{s^{(d-\alpha)/2}}\, ds < \infty.$$

According to this and (B.3), if $\alpha > d - 2 - 2\dim_{\mathrm{M}} F$ is strictly positive, then $\Upsilon_F(\psi_\alpha) < \infty$. Theorem 7.1 then implies that, in this case,

$$\liminf_{t\to\infty} \inf_{s\in F} \frac{(\log t)^{1/\alpha}}{t^{1/2}} |B(s\,,t)| = 0 \quad \text{a.s.}$$

Similarly, if $0 < \alpha < d - 2 - 2\dim_{\mathrm{M}} F$, then

$$\liminf_{t\to\infty} \inf_{s\in F} \frac{(\log t)^{1/\alpha}}{t^{1/2}} |B(s\,,t)| = \infty \quad \text{a.s.}$$

Write $F = \cup_{n\geq 1} F_n$ and "regularize" to find that:

1. If $\alpha > d - 2 - 2\dim_{\mathrm{P}} F$ is strictly positive, then
$$\inf_{s\in F} \liminf_{t\to\infty} \frac{(\log t)^{1/\alpha}}{t^{1/2}} |B(s\,,t)| = 0 \quad \text{a.s.}$$
2. If $0 < \alpha < d - 2 - 2\dim_{\mathrm{P}} F$, then
$$\inf_{s\in F} \liminf_{t\to\infty} \frac{(\log t)^{1/\alpha}}{t^{1/2}} |B(s\,,t)| = \infty \quad \text{a.s.}$$

The theorem follows in the case that $d \geq 4$.

When $d = 3$, the condition $\dim_{\mathrm{M}} F < 1/2$ guarantees that $\mathrm{K}_F(\epsilon) = O(\epsilon^{-1/2})$. Now follow through the proof of the case $d \geq 4$ to finish. □

8. Open problems

8.1. Slices and zeros

Theorem 1.2 is a metric statement. Is there a topological counterpart? The following is one way to state this formally.

Open Problem 1. *Suppose $d \in \{2\,,3\}$. Is it true that outside a single null set, $B^{-1}\{0\} \cap (\{s\} \times (0\,,\infty))$ is a finite set for all $s > 0$?*

I conjecture that the answer is "no." In fact, it is even possible that there exists a non-trivial measure function ϕ such that: (i) $\lim_{r\to 0} \phi(r) = \infty$; and (ii) $\mathcal{H}_\phi$-measure of $B^{-1}\{0\} \cap (\{s\} \times (0\,,\infty))$ is positive for some $s > 0$.

8.2. Smallness of double-points for slices

Theorem 5.1 and a codimension argument together imply that with probability one,

$$\dim_{\mathrm{H}} \hat{\mathcal{D}}_d = \left(3 - \frac{d}{2}\right)_+ \quad\text{and}\quad \dim_{\mathrm{H}} \bar{\mathcal{D}}_d = 2 \wedge \left(3 - \frac{d}{2}\right)_+ .$$

This might suggest that, therefore, none of the slices accrue any of the dimesion.

Open Problem 2. *Define, for all $s \geq 0$,*

$$\mathcal{Y}_d(s) := \left\{(t_1, t_2) \in \mathbb{R}^2_+ :\ B(s\,,t_1) = B(s\,,t_2)\right\}.$$

Then is it the case that if $d \in \{4\,,5\}$, then, outside a single null-set, $\dim_{\mathrm{H}} \mathcal{Y}_d(s) = 0$ for all $s \geq 0$?

I conjecture that the answer is "yes." Answering this might rely on studying closely the methods of the literature on "local non-determinism." See, in particular, Berman [1], Pitt [26], and the recent deep work of Xiao [33]. On the other hand, I believe it should be not too hard to prove that the answer to the corresponding problem for $d \le 3$ is "no," due to the existence of continuous intersection local times [25]. [I have not written out a complete proof in the $d \le 3$ case, mainly because I do not have a proof, or disproof, in the case that $d \in \{4,5\}$. This is the more interesting case because there are no intersection local times.]

Open Problem 1 has the following analogue for double-points.

Open Problem 3. *Let $d \in \{4,5\}$. Then is it true that outside a single null set, $\mathcal{Y}_d(s)$ is a finite set for all $s > 0$?*

The answer to this question is likely to be "no." In fact, as was conjectured for Open Problem 1, here too there might exist slices that have positive $\mathcal{H}_\phi$-measure in some gauge ϕ. If so, then there are in fact values of s for which $\mathcal{Y}_d(s)$ is uncountable.

8.3. Marstrand's Theorem for projections

Marstrand [19] proved that almost every lower-dimensional orthogonal projection of a Borel set A has the same Hausdorff dimension as A; see also Kaufman [11]. Theorem 1.1 proves that a given projection (say, onto the x-axis) of the zero-set of Brownian sheet has the same "Marstrand property." I believe that the proof can be adjusted to show that, in fact, any non-random orthogonal projection of $B^{-1}\{0\}$ has the same Hausdorff dimension as $B^{-1}\{0\}$ itself.

Open Problem 4. *Is there a (random) orthogonal projection such that the said projection of $B^{-1}\{0\}$ has a different Hausdorff dimension than $2-(d/2)$?*

I believe that the answer is "no." However, I have no proof nor counter-proof. Similar questions can be asked about double-points. I will leave them to the interested reader.

8.4. Non-linear SPDEs

Consider d independent, two-dimensional white noises, $\dot{B}_1, \ldots, \dot{B}_d$, together with the following system of d non-interacting stochastic PDEs with additive noise: For a fixed $T > 0$,

$$\begin{aligned} \frac{\partial^2 u^i}{\partial t \partial x}(t,x) &= \hat{B}_i(t,x) + b^i(u(t,x)), \\ u^i(0,x) &= u_0(x) \qquad \text{all } -\infty < x < \infty, \\ \frac{\partial u^i}{\partial t}(0,x) &= u_1(x) \qquad \text{all } -\infty < x < \infty, \end{aligned}$$

where $1 \le i \le N$, and u_0 and u_1 are non-random and smooth, as well as bounded (say). Then, as long as $b := (b^1, \ldots, b^d)$ is bounded and Borel-measurable the law of the process $u := (u^1, \ldots, u^d)$ is mutually absolutely continuous with respect to the law of the two-parameter, d-dimensional Brownian sheet B. See Proposition 1.6 of Nualart and Pardoux [22]. Therefore, the theorems of the preceding sections apply to the process u equally well.

Open Problem 5. *Suppose $\sigma : \mathbb{R}^d \to \mathbb{R}^d$ is a strongly elliptic, bounded, C^∞ function. Is it the case that the results of the previous sections apply to the solution of $(\partial^2 u^i/\partial t \partial x) = b^i(u) + \sigma^i(u) \cdot \hat{B}$ with reasonable boundary conditions?*

There is some evidence that the answer is "yes." See Dalang and Nualart [6] where a closely-related problem is solved.

Finally, we end with an open-ended question about parabolic SPDEs, about which we know far less at this point. We will state things about the additive linear case only. This case seems to be sufficiently difficult to analyse at this point in time.

Open Problem 6. *Consider the following system of linear parabolic SPDE:*

$$\frac{\partial u^i}{\partial t}(t\,, x) = \frac{\partial^2 u^i}{\partial x^2}(t\,, x) + \hat{B}_i(t\,, x),$$

with reasonable boundary conditions. Is there an analysis of the "slices" of u along different values of t that is analogous to the results of the present paper?

Some results along these lines will appear in forthcoming work with Robert Dalang and Eulalia Nualart [4, 5].

Appendix A. Capacity and dimension

For the sake of completeness, we begin with a brief review of Hausdorff measures. Further information can be found in Kahane [10, Chapter 10], Khoshnevisan [13, Appendices C and D], and Mattila [20, Chapter 4].

A.1. Capacity

Recall that $\mathcal{P}(F)$ denotes the collection of all probability measures on the Borel set F, and $|x|$ is the ℓ^1-norm of the vector x. Occasionally we may write $\|x\| := (x_1^2 + \cdots + x_m^2)^{1/2}$ for the ℓ^2-norm of $x \in \mathbb{R}^m$.

Let $f : \mathbb{R}^n \to [0\,,\infty]$ be Borel measurable. Then for all $\mu \in \mathcal{P}(\mathbb{R}^n)$, the *$f$-energy of μ* is defined by

$$I_f(\mu) := \iint f(x-y)\,\mu(dx)\,\mu(dy).$$

If $F \subset \mathbb{R}^n$ is Borel-measurable, then its *f-capacity* can be defined by

$$\mathrm{Cap}_f(F) := \left[\inf_{\mu \in \mathcal{P}(F)} I_f(\mu)\right]^{-1},$$

where $\inf \varnothing := \infty$ and $1/\infty := 0$. If $f : \mathbb{R}_+ \to [0\,,\infty]$ is Borel measurable, then we occasionally abuse notation and write $I_f(\mu) := \iint f(\|x-y\|)\,\mu(dx)\,\mu(dx)$ as well as $I_f(\mu) := \iint f(|x-y|)\,\mu(dx)\,\mu(dy)$. As before, $\mathrm{Cap}_f(F) := [\inf_{\mu\in\mathcal{P}(F)} I_f(\mu)]^{-1}$ in any case.

Let $\beta \in \mathbb{R}$ and $x \in \mathbb{R}\setminus\{0\}$; define

$$U_\beta(x) := \begin{cases} 1, & \text{if } \beta < 0,\\ \log_+(1/|x|), & \text{if } \beta = 0,\\ |x|^{-\beta}, & \text{if } \beta > 0. \end{cases} \tag{A.1}$$

Also, we define U_β at zero by continuously extending U_β to a $[0\,,\infty]$-valued function on all of $\mathbb{R}$. Then we write $I_\beta(\mu)$ in place of $I_{U_\beta}(\mu)$, and $\mathrm{Cap}_\beta(F)$ in place of $\mathrm{Cap}_{U_\beta}(F)$; $I_\beta(\mu)$ is the *Riesz* [or Bessel–Riesz] capacity of μ, and Cap_β is the [Bessel-] *Riesz capacity* of F.

The following is a central property of capacities [13, p. 523].

Taylor's Theorem (Taylor [31]**).** *If $F \subset \mathbb{R}^n$ is compact, then $\mathrm{Cap}_n(F) = 0$. Consequently, for all $\beta \ge n$, $\mathrm{Cap}_\beta(F)$ is zero also.*

A.2. Hausdorff measures

Throughout, we define $\mathcal{B}(x;r) := \{y \in \mathbb{R}^n : |x-y| \le r\}$ to be the closed ℓ^1-ball of radius $r > 0$ about $x \in \mathbb{R}^n$.

A Borel-measurable function $\varphi : \mathbb{R}_+ \to [0\,,\infty]$ is said to be a *measure function* if: (i) φ is non-decreasing near zero; and (ii) $\varphi(2x) = O(\varphi(x))$ as $x \to 0$. Next, we choose and fix a measure function φ and a Borel set A in $\mathbb{R}^n$. For all $r > 0$ we define

$$\mathcal{H}_\varphi^{(r)}(A) := \inf \sum_{j\ge1} \varphi(\delta_j),$$

where the infimum is taken over all $x^{(1)}, x^{(2)}, \ldots \in \mathbb{R}^n$ for which we can find $\delta_1, \delta_2, \ldots \in (0\,,r)$ with $A \subseteq \cup_{j\ge1}\mathcal{B}(x^{(j)};\delta_j)$. The *Hausdorff φ-measure* $\mathcal{H}_\varphi(A)$ of A can then defined as the non-increasing limit,

$$\mathcal{H}_\varphi(A) := \lim_{r\downarrow0} \mathcal{H}_\varphi^{(r)}(A).$$

This defines a Borel [outer-] measure on Borel subsets of $\mathbb{R}^n$.

A.3. Hausdorff dimension

An important special case of $\mathcal{H}_\varphi$ arises when we consider $\varphi(x) = x^\alpha$. In this case we may write $\mathcal{H}_\alpha$ instead; this is the *α-dimensional Hausdorff measure*. The *Hausdorff dimension* of A is

$$\dim_{\mathrm{H}} A := \sup\{\alpha > 0 :\, \mathcal{H}_\alpha(A) > 0\} = \inf\{\alpha > 0 :\, \mathcal{H}_\alpha(A) < \infty\}.$$

The Hausdorff dimension has the following regularity property: If $A_1, A_2, \ldots$ are Borel sets, then

$$\dim_{\mathrm{H}} \bigcup_{i\ge1} A_i = \sup_{i\ge1} \dim_{\mathrm{H}} A_i.$$

In general, this fails if the union is replaced by an uncountable one. For instance, consider the example $\mathbb{R} = \cup_{x\in\mathbb{R}}\{x\}$. The following is a central fact:

Frostman's Lemma (Frostman [8]). *Let A be a compact subset of $\mathbb{R}^n$. Then $\mathcal{H}_\alpha(A) > 0$ if and only if we can find a constant c and a $\mu \in \mathcal{P}(A)$ such that $\mu(\mathcal{B}(x\,;r)) \le cr^\alpha$ for all $r > 0$ and $x \in \mathbb{R}^n$.*

See also Theorem 1 of Kahane [10, p. 130], Theorem 2.1.1 of Khoshnevisan [13, p. 517], and Theorem 8.8 of Mattila [20, p. 112].

Appendix B. Entropy and packing

The material of this appendix can be found, in expanded form and with a detailed bibliography, in Khoshnevisan et al [15]. Throughout, $F \subset \mathbb{R}$ is a Borel-measurable set.

B.1. Minkowksi content and Kolmogorov capacitance

There are various ways to describe the size of the set F. We have seen already the role of capacity, Hausdorff measures, and Hausdorff dimension. Alternatively, we can consider the rate of growth of the *Minkowski content* of F; this is the function $\mathbf{N} \ni n \mapsto \mathrm{M}_n(F)$ defined as follows:

$$\mathrm{M}_n(F) := \#\left\{ i \in \mathbf{Z} : \; F \cap \left[\frac{i}{n}\,, \frac{i+1}{n}\right) \neq \varnothing \right\}.$$

Also, we can consider the *Kolmogorov entropy* (known also as "capacitance" or "packing number") of F; this is the function $(0\,,\infty) \ni \epsilon \mapsto \mathrm{K}_F(\epsilon)$, where $\mathrm{K}_E(\epsilon)$ is equal to the maximum number K for which there exist $x_1, \ldots, x_K \in F$ such that $\min_{i \neq j} |x_i - x_j| \ge \epsilon$. Any such sequence $\{x_i\}_{1 \le i \le \mathrm{K}_F(\epsilon)}$ is referred to as a *Kolmogorov sequence.*

While $\mathrm{M}_n(F)$ is easier to work with, $\mathrm{K}_F(\epsilon)$ has the nice property that $\mathrm{K}_F(\epsilon) \ge \mathrm{K}_F(\delta) \ge 1$ whenever $0 < \epsilon < \delta$. There are two other properties that deserve mention. The first is that [15, Proposition 2.7]

$$\mathrm{K}_F(1/n) \le \mathrm{M}_n(F) \le 3\mathrm{K}_F(1/n) \qquad \text{for all } n \ge 1. \tag{B.1}$$

The second property is the following [15, eq. (2.8)]:

$$\mathrm{K}_E(\epsilon) \le 6\mathrm{K}_F(2\epsilon) \qquad \text{for all } \epsilon > 0. \tag{B.2}$$

B.2. Minkowski and packing dimension

The (upper) *Minkowski dimension* of F is the number

$$\dim_{_\mathrm{M}} F := \limsup_{n \to \infty} \frac{\log \mathrm{M}_n(F)}{\log n}.$$

This is known also as the (upper) "box dimension" of F, and gauges the size of F.

A handicap of the gauge $\dim_{_\mathrm{M}}$ is that it assigns the value 1 to the rationals in $[0\,,1]$; whereas we often wish to think of $\mathbf{Q} \cap [0\,,1]$ as a "zero-dimensional" set. In such cases, a different notion of dimension can be used.

The (upper) *packing dimension* of F is the "regularization" of $\dim_{\mathrm{M}} F$ in the following sense:

$$\dim_{\mathrm{P}} F := \sup \left\{ \dim_{\mathrm{M}} F_k;\ F = \bigcup_{i \geq 1} F_i,\ F_i\text{'s are closed and bounded} \right\}.$$

Then it is not hard to see that $\dim_{\mathrm{P}}(\mathbf{Q} \cap [0\,,1]) = 0$, as desired. Furthermore, we have the relation,

$$\dim_{\mathrm{H}} F \leq \dim_{\mathrm{P}} F \leq \dim_{\mathrm{M}} F. \tag{B.3}$$

See Mattila [20, p. 82]. These are often equalities; e.g., when F is a self-similar fractal. However, there are counter-examples for which either one, or both, of these inequalities can be strict. Furthermore, one has [15, Proposition 2.9] the following integral representations:

$$\dim_{\mathrm{M}} F = \inf \left\{ q \in \mathbb{R} : \int_1^\infty \mathrm{K}_F(1/s)\, \frac{ds}{s^{1+q}} < \infty \right\},$$

$$\dim_{\mathrm{P}} F = \inf \left\{ q \in \mathbb{R} : \begin{array}{c} {}^\exists F_1, F_2, \ldots \text{closed and bounded} \\ \text{such that } F = \bigcup_{i \geq 1} F_i, \text{ and} \\ \int_1^\infty s^{-1-q} \mathrm{K}_{F_n}(1/s)\, ds < \infty \text{ for all } n \geq 1 \end{array} \right\}.$$

Appendix C. Some hitting estimates for Brownian motion

Throughout this section, X and Y denote two independent, standard Brownian motions in $\mathbb{R}^d$, where $d \geq 3$. We will need the following technical lemmas about Brownian motion. The first lemma is contained in Propositions 1.4.1 and 1.4.3 of Khoshnevisan [13, pp. 353 and 355].

Lemma C.1. *For all* $r \in (0\,,1)$,

$$\sup_{a \in \mathbb{R}^d} \mathrm{P}\left\{ \inf_{1 \leq t \leq 3/2} |a + X(t)| \leq r \right\} \leq c r^{d-2} \leq c \mathrm{P}\left\{ \inf_{1 \leq t \leq 2} |X(t)| \leq r \right\}. \tag{C.1}$$

We will also need the following variant.

Lemma C.2. *There exists a constant* c *such that for all* $0 < r < \rho < 1$,

$$\mathrm{P}\left(\inf_{1 \leq t \leq 2} |\rho Y(t) + X(t)| \leq r \;\middle|\; \inf_{1 \leq s \leq 2} |X(s)| \leq r \right) \leq c \rho^{d-2}. \tag{C.2}$$

Remark C.3. The condition "$0 < r < \rho < 1$" can be replaced with "$0 < r \leq \alpha\rho$" for any fixed finite $\alpha > 0$. However, this lemma fails to hold for values of $\rho = o(r)$ as can be seen by first fixing $r > 0$ and then letting ρ tend to 0 on the left-hand side of (C.2): The left-hand side converges to 1 while the right-hand side converges to 0.

Proof. Define $T := \inf\{1 \le t \le 2 : |X(s)| \le r\}$, where $\inf \varnothing := \infty$, as usual. Then,

$$\begin{aligned} \mathrm{P}_1 &:= \mathrm{P}\left(\inf_{T \le t \le 2} |\rho Y(t) + X(t)| \le r \;\middle|\; T < \infty \right) \\ &= \mathrm{P}\left(\inf_{0 \le s \le 2-T} |\rho Y(T+s) + X(T+s)| \le r \;\middle|\; T < \infty \right) \\ &\le \mathrm{P}\left(\inf_{0 \le s \le 2-T} \left|\rho Y(T+s) + \hat{X}(s)\right| \le 2r \;\middle|\; T < \infty \right), \end{aligned}$$

where $\hat{X}(s) := X(T+s) - X(T)$ for all $s \ge 0$. By the strong Markov property of X,

$$\mathrm{P}_1 \le \sup_{1 \le t \le 2} \mathrm{P}\left\{ \inf_{0 \le s \le 1} |\rho Y(t+s) + X(s)| \le 2r \right\}. \tag{C.3}$$

In order to estimate this quantity, let us fix an arbitrary $t \in [1\,, 2]$, and define

$$\begin{aligned} S &:= \inf\{0 \le s \le 1 : \; |\rho Y(t+s) + X(s)| \le 2r\}, \\ Z &:= \int_0^2 \mathbf{1}_{\{|\rho Y(t+s) + X(s)| \le 3r\}}\, ds. \end{aligned}$$

Then,

$$\begin{aligned} \mathrm{E}[Z \,|\, S < \infty] &\ge \mathrm{E}\left[\int_S^2 \mathbf{1}_{\{|\rho Y(t+s) + X(s)| \le 3r\}}\, ds \;\middle|\; S < \infty \right] \\ &\ge \mathrm{E}\left[\int_0^{2-S} \mathbf{1}_{\{|\rho \mathcal{Y}(t+s) + \mathcal{X}(s)| \le r\}}\, ds \;\middle|\; S < \infty \right], \end{aligned}$$

where $\mathcal{Y}(u) := Y(u+S) - Y(S)$ and $\mathcal{X}(u) := X(u+S) - X(S)$ for all $u \ge 0$. The process $u \mapsto \rho Y(t+u) + X(u)$ is a Lévy process, and S is a stopping time with respect to the latter process. Therefore, by the strong Markov property,

$$\begin{aligned} \mathrm{E}[Z \,|\, S < \infty] &\ge \int_0^1 \mathrm{P}\left\{ |\rho \mathcal{Y}(t+s) + \mathcal{X}(s)| \le r \right\} ds \\ &= \int_0^1 \mathrm{P}\left\{ \left(\rho^2(t+s) + s\right)^{1/2} |\mathbf{g}| \le \epsilon \right\} ds \\ &\ge \int_0^1 \mathrm{P}\left\{ \left(\rho^2 t + s\right)^{1/2} |\mathbf{g}| \le \epsilon \right\} ds, \end{aligned}$$

where $\mathbf{g}$ is a d-vector of i.i.d. standard-normal variables. Recall (2.1). Thanks to Lemmas 2.1 and 2.2,

$$\inf_{1 \le t \le 2} \mathrm{E}[Z \,|\, S < \infty] \ge c \int_0^1 f_\epsilon(\rho^2 + s)\, ds = cF_\epsilon(\rho^2) \ge c\epsilon^d \rho^{-(d-2)}.$$

We have appealed to the condition $\rho > \epsilon$ here. Another application of Lemma 2.1 yields the following:

$$\sup_{1 \le t \le 2} \mathrm{E}[Z \,|\, S < \infty] \le \frac{\mathrm{E}[Z]}{\mathrm{P}\{S < \infty\}} \le \frac{c\epsilon^d}{\mathrm{P}\{S < \infty\}}.$$

Recall (C.3) to find that the preceding two displays together imply that $\mathrm{P}_1 \leq c\rho^{d-2}$. Thus, it suffices to prove that

$$\mathrm{P}_2 := \mathrm{P}\left(\inf_{1\leq t\leq T} |\rho Y(t) + X(t)| \leq r \;\middle|\; T < \infty \right) \leq c\rho^{d-2}.$$

The estimate on P_2 is derived by using the method used to bound P_1; but we apply the latter method to the time-inverted Brownian motion $\{tX(1/t)\}_{t>0}$ in place of X. We omit the numerous, messy details. □

References

[1] S. M. Berman, *Local nondeterminism and local times of Gaussian processes*, Indiana Univ. Math. J., **23** (1973/1974), 69–94

[2] N. Čentsov, *Wiener random fields depending on several parameters*, Dokl. Akad. Nauk SSSR (N.S.), **106** (1965), 607–609 (in Russian).

[3] R. C. Dalang and D. Khoshnevisan, *Recurrent lines in two-parameter isotropic stable Lévy sheets*, Stoch. Proc. Appl., **114 (4)** (2004), 81–107.

[4] R. C. Dalang, D. Khoshnevisan, and E. Nualart, *Hitting probabilities for parabolic SPDEs with additive noise*, preprint, 2005.

[5] R. C. Dalang, D. Khoshnevisan, and E. Nualart, *Hitting probabilities for parabolic SPDEs with multiplicative noise*, preprint, 2005.

[6] R. C. Dalang and E. Nualart, *Potential theory for hyperbolic SPDEs*, Ann. Probab., **32 (3a)** (2004), 2099–2148.

[7] A. Dvoretzky and P. Erdős, *Some problems on random walk in space*, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (1950), University of California Press, Berkeley and Los Angeles, (1951), 353–367.

[8] O. Frostman, *Potentiel d'équilibre et capacité des ensembles avec quelques applications à la thérie des fonctions*, Meddel. Lunds. Univ. Mat. Sem., **3** (1935), 1–118.

[9] M. Fukushima, *Basic properties of Brownian motion and a capacity on the Wiener space*, J. Math. Soc. Japan, **36 (1)** (1984), 161–176.

[10] J.-P. Kahane, *Some Random Series of Functions*, Cambridge University Press, Cambridge, 1985.

[11] R. Kaufman, *On Hausdorff dimension of projections*, Mathematika, **15** (1968), 153–155.

[12] D. Khoshnevisan, *Brownian sheet and quasi-sure analysis, Asymptotic methods in stochastics*, Fields Inst. Commun., AMS, **44** (2004), 25–47.

[13] D. Khoshnevisan, *Multiparameter Processes*, Springer, New York, 2002.

[14] D. Khoshnevisan, *Brownian sheet images and Bessel–Riesz capacity*, Trans. Amer. Math. Soc., **351 (7)** (1999), 2607–2622.

[15] D. Khoshnevisan, D. A. Levin, and P. J. Méndez–Hernández, *Exceptional times and invariance for dynamical random walks*, Probab. Th. Rel. Fields, **134 (3)** (2006), 383–416.

[16] N. Kôno, 4-*dimensional Brownian motion is recurrent with positive capacity*, Proc. Japan Acad. Ser. A Math. Sci., **60 (2)** (1984), 57–59.

[17] M. T. Lacey, *Limit laws for local times of the Brownian sheet*, Probab. Th. Rel. Fields, **86 (1)** (1990), 63–85.

[18] T. J. Lyons, *The critical dimension at which quasi-every Brownian path is self-avoiding*, Adv. in Appl. Probab. suppl., (1986), 87–99.

[19] J. M. Marstrand, *The dimension of Cartesian product sets*, Proc. Cambridge Philos. Soc., **50** (1954), 198–202.

[20] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge University Press, Cambridge, 1995.

[21] T. S. Mountford, *Double points and the Ornstein-Uhlenbeck process on Wiener space*, Illinois J. Math., **34 (1)** (1990), 38–48.

[22] D. Nualart and E. Pardoux, *Markov field properties of solutions of white noise driven quasi-linear parabolic PDEs*, Stochastics Stochastics Rep., **48 (1-2)** (1994), 17—44.

[23] S. Orey and W. E. Pruitt, *Sample functions of the N-parameter Wiener process*, Ann. Probab., **1 (1)** (1973), 138–163.

[24] M. D. Penrose, *Quasi-everywhere properties of Brownian level sets and multiple points*, Stoch. Proc. Appl., **36 (1)** (1990), 33–43.

[25] M. D. Penrose, *On the existence of self-intersections for quasi-every Brownian path in space*, Ann. Probab., **17 (2)** (1989), 482–502.

[26] L. D. Pitt, *Local times for Gaussian vector fields*, Indiana Univ. Math. J., **27 (2)** (1978), 309–330.

[27] J. Rosen, *Self-intersections of random fields*, Ann. Probab., **12 (1)** (1984), 108–119.

[28] J. Rosen, *A local time approach to the self-intersections of Brownian paths in space*, Comm. Math. Phys., **88 (3)** (1984), 327–338.

[29] I. Shigekawa, *On a quasi-everywhere existence of the local time of the* 1*-dimensional Brownian motion*, Osaka J. Math., **21 (3)** (1984), 621–627.

[30] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, New Jersey, 1970.

[31] S. J. Taylor, *On the connexion between Hausdorff measures and generalized capacity*, Proc. Cambridge Philos. Soc., **57** (1961), 524–531.

[32] E. Wong and M. Zakai, *Martingales and stochastic integrals for processes with a multi-dimensional parameter*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, **9** (1974), 109–122.

[33] Y. Xiao, *Strong local non-determinism and sample path properties of Gaussian random fields*, preprint, 2005.

Davar Khoshnevisan
Department of Mathematics
The University of Utah
155 S 1400 E Salt Lake City
UT 84112-0090
USA
e-mail: `davar@math.utah.edu`
URL: `http://www.math.utah.edu/~davar`

Progress in Probability, Vol. 59, 175–196

An Estimate of the Convergence Rate in Diffusion Approximation of a Particle Motion under Random Forcing

Tomasz Komorowski

Abstract. Suppose that the trajectory of a particle $\mathbf{x}(t;\mathbf{x},\mathbf{k})$ is a solution of the Newton equation $\ddot{\mathbf{x}}(t;\mathbf{x},\mathbf{k}) = \delta^{1/2}\mathbf{F}(\mathbf{x}(t;\mathbf{x},\mathbf{k}),\dot{\mathbf{x}}(t;\mathbf{x},\mathbf{k}))$, $\mathbf{x}(0) = \mathbf{x}$, $\dot{\mathbf{x}}(0) = \mathbf{k}$, where $\mathbf{F}(\mathbf{x},\mathbf{k})$ is a spatially homogeneous random force field defined over a certain probability space $(\Omega,\Sigma,\mathbb{P})$. It has been proved by Kesten and Papanicolaou in [2] that if $d \geq 3$ and $\mathbf{F}(\mathbf{x},\mathbf{k})$ is sufficiently regular, nondegenerate and mixing in the spatial variable, then the process $(\delta^{1/2}\mathbf{x}(\delta^{-1}t;\mathbf{x},\mathbf{k}),\ddot{\mathbf{x}}(\delta^{-1}t;\mathbf{x},\mathbf{k}))$, $t \geq 0$, converges weakly to a hypoelliptic diffusion. In this paper we prove power-like bounds on the convergence rate for one-dimensional marginals of the process.

Mathematics Subject Classification (2000). 82D30.

Keywords. Stochastic acceleration, random fields, mixing.

1. Introduction

Let $(\mathbf{x},\mathbf{k}) \in \mathbb{R}^{2d}$ and suppose that the trajectory of a particle $\mathbf{x}(t;\mathbf{x},\mathbf{k})$, $t \geq 0$, is a solution of the Newton system of equations

$$\ddot{\mathbf{x}}(t;\mathbf{x},\mathbf{k}) = \delta^{1/2}\mathbf{F}(\mathbf{x}(t;\mathbf{x},\mathbf{k}),\dot{\mathbf{x}}(t;\mathbf{x},\mathbf{k})), \quad \mathbf{x}(0) = \mathbf{x},\, \dot{\mathbf{x}}(0) = \mathbf{k}.$$

Here $\mathbf{F} : \mathbb{R}^d \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ is a random force field defined over a certain probability space $(\Omega,\Sigma,\mathbb{P})$. The parameter $\delta > 0$ corresponds to a magnitude of the field and is supposed to be small. It is obvious that if $t = O(1)$, then the trajectory of the particle is approximately given by $\mathbf{x}(t;\mathbf{x},\mathbf{k}) \approx \mathbf{x} + \mathbf{k}t$. The diffusive behavior of the velocity process is however observed at time scales of order δ^{-1}. It has been shown in [2] for $d \geq 3$ and in [4] for $d = 2$ (in the potential field case) that if the initial velocity does not vanish, the field is sufficiently strongly mixing,

The research of the author was supported by KBN grant nr 2PO3A 031 23.

and satisfies some other regularity assumptions, then the continuous trajectory processes $(\mathbf{x}_\delta(t;\mathbf{x},\mathbf{k}),\mathbf{k}_\delta(t;\mathbf{x},\mathbf{k}))$, where

$$\mathbf{x}_\delta(t;\mathbf{x},\mathbf{k}) := \delta\mathbf{x}(\delta^{-1}t;\delta^{-1}\mathbf{x},\mathbf{k}), \quad \mathbf{k}_\delta(t;\mathbf{x},\mathbf{k}) := \dot{\mathbf{x}}(\delta^{-1}t;\delta^{-1}\mathbf{x},\mathbf{k}), \tag{1.1}$$

converge weakly, as $\delta \to 0+$, to $(\mathbf{x}(t;\mathbf{x},\mathbf{k}),\mathbf{k}(t;\mathbf{k}))$, where $\mathbf{k}(t;\mathbf{k})$ is a diffusion starting at $\mathbf{k}$ (see Section 2.4 below for its definition) and $\mathbf{x}(t;\mathbf{x},\mathbf{k}) := \mathbf{x}+\int_0^t \mathbf{k}(s;\mathbf{k})ds$.

Suppose now that $\phi_0(\mathbf{x},\mathbf{k})$ is a smooth function whose support is contained inside a spherical shell $[(\mathbf{x},\mathbf{k}) \in \mathbb{R}^{2d} : M^{-1} < |\mathbf{k}| < M]$ for some $M > 1$. Then, $\phi_\delta(t,\mathbf{x},\mathbf{k}) := \phi_0(\mathbf{x}_\delta(t;\mathbf{x},\mathbf{k}),\mathbf{k}_\delta(t;\mathbf{x},\mathbf{k}))$ satisfies the Liouville equation

$$\begin{aligned} \frac{\partial\phi_\delta}{\partial t} &= \mathbf{k}\cdot\nabla_{\mathbf{x}}\phi_\delta + \delta^{-1/2}\mathbf{F}\left(\frac{\mathbf{x}}{\delta},\mathbf{k}\right)\cdot\nabla_{\mathbf{k}}\phi_\delta, \\ \phi_\delta(0,\mathbf{x},\mathbf{k}) &= \phi_0(\mathbf{x},\mathbf{k}). \end{aligned} \tag{1.2}$$

The aforementioned weak convergence of stochastic processes implies in particular that $\lim_{\delta\to 0+}\mathbb{E}\phi_\delta(t,\mathbf{x},\mathbf{k}) = \bar{\phi}(t,\mathbf{x},\mathbf{k})$, where $\bar{\phi}(t,\mathbf{x},\mathbf{k})$ is the solution of the Kolmogorov equation corresponding to the limiting diffusion $(\mathbf{x}(t;\mathbf{x},\mathbf{k}),\mathbf{k}(t;\mathbf{k}))$ and $\bar{\phi}(0,\mathbf{x},\mathbf{k}) = \phi_0(\delta\mathbf{x},\mathbf{k})$, see (2.7). In the present paper we set out to find the error estimates in the above convergence. We shall show, see Theorem 2.2 below, that the supremum of $|\mathbb{E}\phi_\delta(t,\mathbf{x},\mathbf{k}) - \bar{\phi}(t,\mathbf{x},\mathbf{k})|$ over a compact subset of $\mathbb{R}_+\times\mathbb{R}^d\times\mathbb{R}^d_*$, where $\mathbb{R}^d_* := \mathbb{R}^d\setminus\{\mathbf{0}\}$, is of order of magnitude δ^α for some $\alpha > 0$. These bounds are useful, e.g., in describing long time asymptotic behavior of the particle trajectory in scales that are longer than δ^{-1}. For example, it can be shown (see [3]) that in the case of the potential force field there exists $\alpha_0 > 0$ such that for each $\alpha \in (0,\alpha_0)$ the processes $\delta^{1+\alpha}\mathbf{x}(t\delta^{-1-2\alpha})$ converge, in an appropriate sense, as $\delta \to 0+$, to a Brownian motion.

The present paper relies to a large extend on the technique developed in [3] in the context of Hamiltonian flows, we shall refer therefore frequently to the respective parts of the aforementioned article. The main tool used to obtain the error estimates is the observation that a suitably modified dynamics of the particle approximately satisfies the martingale problem of Stroock and Varadhan corresponding to the limiting diffusion, see Proposition 4.2 below. The aforementioned modification of the dynamics is made with the help of a certain stopping time τ, see Section 3.2 for its precise definition, in the following way. Before τ, the trajectory remains unchanged and has “almost” the Markov property. After the stopping time, the modified dynamics of the particle motion is the same as that of the limiting diffusion. As a result the law of the modified trajectory process approximately satisfies the martingale problem corresponding to the limiting diffusion. The key observation is that the stopping time in question tends to ∞. An important difference between the argument presented here and that made for a Hamiltonian flow in [3] concerns the fact that the velocity of the particle can become either arbitrarily large, or can degenerate to zero in finite time. Both of these types of behavior are undesirable because they prevent the use of a perturbative argument needed to establish the approximate martingale property stated in Proposition 4.2 below. This forces us to add an additional clause to the stopping rule that deals with such a situation. The probability that the velocity is uncontrollably large can be shown to be small thanks to the fact that this is the case for the limiting diffusion. To deal with the possibility that the particle momentum could vanish we have to assume additionally that the probability of the limiting diffusion reaching a δ-neighborhood of the origin decays as δ^γ, as $\delta \ll 1$, for some $\gamma > 0$, see hypothesis F6) below.

2. Preliminaries and the statement of the main result

2.1. Basic notation

Let $\mathbb{R}_*^{2d} := \mathbb{R}^d \times \mathbb{R}^d$, where as we recall $\mathbb{R}_*^d := \mathbb{R}^d \setminus \{\mathbf{0}\}$. Given a vector $\mathbf{k} \in \mathbb{R}^d$ we denote by $k := |\mathbf{k}|$ its length and $\hat{\mathbf{k}} := \mathbf{k}/k$, provided that $\mathbf{k} \neq \mathbf{0}$. For any $\mathbf{x} \in \mathbb{R}^d$ and $r > 0$ we let $\mathbb{B}_r(\mathbf{x})$ and $\mathbb{S}_r^{d-1}(\mathbf{x})$ be the open ball and sphere of radius $r > 0$ centered at $\mathbf{x}$. In the special case when $\mathbf{x} = \mathbf{0}$, or $r = 1$, we shall omit these parameters in our notation. For a fixed $M > 1$ we define the spherical shell $A(M) := [\mathbf{k} \in \mathbb{R}_*^d : M^{-1} \leq |\mathbf{k}| \leq M^3]$ in the $\mathbf{k}$-space and $\mathcal{A}(M) := \mathbb{R}^d \times A(M)$ in the whole phase space.

For any non-negative integers p, q, r, positive times $T > T_* \geq 0$ and a function $G : [T_*, T] \times \mathbb{R}_*^{2d} \to \mathbb{R}$ that has p, q and r derivatives in the respective variables we define

$$\|G\|_{p,q,r}^{[T_*,T]} := \sum \sup_{(t,\mathbf{x},\mathbf{k}) \in [T_*,T] \times \mathbb{R}^{2d}} |\partial_t^\alpha \partial_{\mathbf{x}}^\beta \partial_{\mathbf{k}}^\gamma G(t, \mathbf{x}, \mathbf{k})|.$$

The summation range covers all integers $0 \leq \alpha \leq p$ and all integer-valued multi-indices with $|\beta| \leq q$ and $|\gamma| \leq r$. In the special case when $T_* = 0$, $T = +\infty$ we write $\|G\|_{p,q,r} = \|G\|_{p,q,r}^{[0,+\infty)}$. We denote by $C_b^{p,q,r}([0, +\infty) \times \mathbb{R}_*^{2d})$ the space of all functions G with $\|G\|_{p,q,r} < +\infty$. We shall also consider spaces of bounded and a suitable number of times continuously differentiable functions $C_b^{p,q}(\mathbb{R}_*^{2d})$ and $C_b^p(\mathbb{R}^d)$ with the respective norms $\|\cdot\|_{p,q}$, $\|\cdot\|_p$.

We shall denote by $C, C_1, \ldots, \alpha_0, \alpha_1, \ldots, \gamma_0, \gamma_1, \ldots$ appearing throughout this article generic positive constants. Unless specified otherwise the constants denoted this way shall not depend on δ.

2.2. Random forcing

Let $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}$ and let $\|X\|_{L^p(\Omega)}$ denote the L^p-norm of a given random variable $X : \Omega \to \mathbb{R}$, $p \in [1, +\infty]$. A random field $\mathbf{F} : \mathbb{R}^d \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ is supposed to satisfy the following conditions:

F1) it is measurable and strictly stationary in the first variable. This means that for any shift $\mathbf{x} \in \mathbb{R}^d$, and a collection of points $(\mathbf{x}_1, \mathbf{k}_1) \ldots, (\mathbf{x}_n, \mathbf{k}_n) \in \mathbb{R}^{2d}$ the laws of $(\mathbf{F}(\mathbf{x}_1 + \mathbf{x}, \mathbf{k}_1), \ldots, \mathbf{F}(\mathbf{x}_n + \mathbf{x}, \mathbf{k}_n))$ and $(\mathbf{F}(\mathbf{x}_1, \mathbf{k}_1), \ldots, \mathbf{F}(\mathbf{x}_n, \mathbf{k}_n))$ are identical.

F2) it is centered, i.e., $\mathbb{E}\,\mathbf{F}(\mathbf{x}, \mathbf{k}) = 0$ for all $(\mathbf{x}, \mathbf{k}) \in \mathbb{R}^{2d}$.

F3) it is smooth, i.e., the realizations of $\mathbf{F}(\mathbf{x},\mathbf{k})$ are $\mathbb{P}$-a.s. C^{n_*}-smooth in $(\mathbf{x},\mathbf{k}) \in \mathbb{R}^{2d}$, with $n_* := [3+d/4]$. We assume also the following control over the field and its derivatives:

$$\tilde{D} := \sum_{0\le i+j\le n_*} \max_{|\alpha|=i} \operatorname*{ess-sup}_{(\mathbf{x},\mathbf{k},\omega)\in\mathbb{R}^{2d}\times\Omega} \left[|\partial_{\mathbf{x}}^{\alpha}\partial_{\mathbf{k}}^{j}\mathbf{F}(\mathbf{x},\mathbf{k};\omega)|\right] < +\infty.$$

F4) the random field is strongly mixing in the uniform sense. More precisely, for any $R>0$ we let $\mathcal{C}_R^i$ and $\mathcal{C}_R^e$ be the σ-algebras generated by random variables $\mathbf{F}(\mathbf{x},\mathbf{k})$, $\mathbf{k}\in\mathbb{R}^d$ with $\mathbf{x}\in\mathbb{B}_R$ and $\mathbf{x}\in\mathbb{B}_R^c$, respectively. The uniform mixing coefficient between the σ-algebras is

$$\phi(\rho) := \sup[\,|\mathbb{P}(B)-\mathbb{P}(B|A)|:\ R>0,\ A\in\mathcal{C}_R^i,\ B\in\mathcal{C}_{R+\rho}^e\,],$$

for all $\rho>0$. We suppose that $\phi(\rho)$ decays faster than any power: for each $p>0$,

$$h_p := \sup_{\rho\ge 0}\rho^p\phi(\rho) < +\infty. \tag{2.1}$$

The two-point spatial covariance matrix of the force field is given by

$$\boldsymbol{R}(\mathbf{y},\mathbf{k}) = [R_{i,j}(\mathbf{y},\mathbf{k})] := \mathbb{E}[F_i(\mathbf{y},\mathbf{k})F_j(\mathbf{0},\mathbf{k})].$$

Note that (2.1) implies that for each $p>0$ there exists a constant $C>0$ such that

$$h_p := \sum_{i=0}^{2n_*}\sum_{|\alpha|=i}\sup_{(\mathbf{y},\mathbf{k})\in\mathbb{R}^{2d}}(1+|\mathbf{y}|^2)^{p/2}|\partial_{\mathbf{y}}^{\alpha}R_{i,j}(\mathbf{y},\mathbf{k})| < +\infty. \tag{2.2}$$

2.3. Certain path-spaces

For fixed integers $d,m\ge 1$ we let $\mathcal{C}^{d,m} := C([0,+\infty);\mathbb{R}^d\times\mathbb{R}_*^m)$: we shall omit the superscripts in the notation of the path space if $m=d$. We shall also write $\mathcal{C}_K$ to denote $C([0,+\infty);\mathbb{R}_*^d)$.

We define $(X(t),K(t)) : \mathcal{C}^{d,m} \to \mathbb{R}^d\times\mathbb{R}_*^m$ as the canonical mapping $(X(t;\pi),K(t;\pi)) := \pi(t)$, $\pi\in\mathcal{C}^{d,m}$ and also let $\theta_s(\pi)(\cdot) := \pi(\cdot+s)$ be the standard shift transformation. For any $u\ge 0$ denote by $\mathcal{M}^u$ the σ-algebra of subsets of $\mathcal{C}$ generated by $(X(t),K(t))$, $t\in[0,u]$. We write $\mathcal{M}$ for the σ algebra of Borel subsets of $\mathcal{C}$. It coincides with the smallest σ-algebra that contains all $\mathcal{M}^t$, $t\ge 0$. We define $\mathcal{C}(M)$ as the set of paths $\pi\in\mathcal{C}$ so that both $K(t)\in A(M)$ and $X(t)=X(0)+\int_0^t K(s)ds,\ t\ge 0$.

2.4. The statement of the main result

Let us define the diffusion matrix $\mathbf{D}(\mathbf{k}) := [D_{mn}(\mathbf{k})]$ for $\mathbf{k}\in\mathbb{R}_*^d$ by

$$D_{mn}(\mathbf{k}) = \frac{1}{2}\int_{-\infty}^{\infty} R_{mn}(s\mathbf{k},\mathbf{k})ds, \qquad m,n=1,\dots,d \tag{2.3}$$

and the drift vector

$$E_m(\mathbf{k}) = \sum_{n=1}^{d} \int_0^{+\infty} s\,[\partial_{x_n} R_{mn}(s\mathbf{k},\mathbf{k}) + \partial_{l_n} R_{mn}(s\mathbf{k},\mathbf{k})]\,ds, \quad m = 1,\ldots,d. \tag{2.4}$$

Let $\mathbf{k}(t;\mathbf{k})$ be a diffusion in $\mathbb{R}_*^d$, starting at $\mathbf{k}$ at $t = 0$, with the generator

$$\mathcal{L}F(\mathbf{k}) = \sum_{m,n=1}^{d} D_{mn}(\mathbf{k})\partial^2_{k_m,k_n} F(\mathbf{k}) + \sum_{m=1}^{d} E_m(\mathbf{k})\partial_{k_m} F(\mathbf{k}). \tag{2.5}$$

Its existence follows from the assumed smoothness of the coefficients of the generator $\mathcal{L}$, see Remark 1, p. 24 of [2]. Let $\mathfrak{Q}_{\mathbf{k}}$ be the corresponding law of the diffusion over $\mathcal{C}_K$.

Note that the substitution $s := sk$ (recall $k = |\mathbf{k}|$) in (2.3) and (2.4) yields $D_{mn}(\mathbf{k}) = k^{-1}d_{mn}(\mathbf{k})$ and $E_m(\mathbf{k}) = k^{-2}\tilde{e}_m(\mathbf{k})$, where

$$d_{mn}(\mathbf{k}) = \frac{1}{2}\int_{-\infty}^{\infty} R_{mn}(s\hat{\mathbf{k}},\mathbf{k})ds, \quad m,n = 1,\ldots,d$$

and

$$\tilde{e}_m(\mathbf{k}) = \sum_{n=1}^{d} \int_0^{+\infty} s\,\left[\partial_{x_n} R_{mn}(s\hat{\mathbf{k}},\mathbf{k}) + \partial_{l_n} R_{mn}(s\hat{\mathbf{k}},\mathbf{k})\right] ds, \quad m = 1,\ldots,d.$$

In light of (2.2) we have

$$\sup_{\mathbf{k}\in\mathbb{R}_*^d} \left(\sum_{m,n=1}^{d} |d_{mn}(\mathbf{k})| + \sum_{m=1}^{d} |\tilde{e}_m(\mathbf{k})| \right) < +\infty.$$

We assume that

F5) there exists a constant $C > 0$ such that

$$\sum_{m,n=1}^{d} d_{mn}(\mathbf{k})\xi_m\xi_n \ge C|\xi|^2, \quad \forall\ \mathbf{k}\in\mathbb{R}_*^d,\ \xi\in\mathbb{R}^d.$$

Define the stopping time $V(\delta) := \min[t \ge 0 : |K(t)| \le \delta]$. We shall assume that:

F6) for each $T > 0$ and a compact set $K \subset \mathbb{R}_*^d$ there exist constants $C, \gamma^* > 0$ such that $\sup_{\mathbf{k}\in K} \mathfrak{Q}_{\mathbf{k}}[V(\delta) \le T] \le C\delta^{\gamma^*}$.

Remark 2.1. Let $\tilde{V}(\delta) := \min[t \ge 0 : |K(t)| \le \delta, \text{ or } |K(t)| \ge \delta^{-3}]$. As a consequence of the above assumption and estimate (2.1) p. 87 of [7] we conclude that the exponent γ^* can be adjusted in such a way that

$$\sup_{\mathbf{k}\in K} \mathfrak{Q}_{\mathbf{k}}[\tilde{V}(\delta) \le T] \le C\delta^{\gamma^*}. \tag{2.6}$$

Let $\mathfrak{Q}_{\mathbf{x},\mathbf{k}}$ be the law of the process $(\mathbf{x}(t;\mathbf{x},\mathbf{k}),\mathbf{k}(t;\mathbf{k}))$ that starts at $t=0$ from $(\mathbf{x},\mathbf{k})$ and is given by $\mathbf{x}(t;\mathbf{x},\mathbf{k})=\mathbf{x}+\int_0^t\mathbf{k}(s;\mathbf{k})ds$, where $\mathbf{k}(t;\mathbf{k})$ is the diffusion described by (2.5). This process is a degenerate diffusion whose generator is given by

$$\tilde{\mathcal{L}}F(\mathbf{x},\mathbf{k})=\mathcal{L}_{\mathbf{k}}F(\mathbf{x},\mathbf{k})+\mathbf{k}\cdot\nabla_{\mathbf{x}}F(\mathbf{x},\mathbf{k}),\quad F\in C_c^\infty(\mathbb{R}_*^{2d}).$$

Here the notation $\mathcal{L}_{\mathbf{k}}$ stresses that the operator $\mathcal{L}$ defined in (2.5) acts on the respective function in the $\mathbf{k}$ variable. We denote by $\mathfrak{M}_{\mathbf{x},\mathbf{k}}$ the expectation corresponding to the path measure $\mathfrak{Q}_{\mathbf{x},\mathbf{k}}$. Then we have the following result.

Theorem 2.2. *Suppose that $T\ge 1$ is given. Assume that $d\ge 3$ and the field $\mathbf{F}$ satisfies the assumptions* F1)–F6)*. Let ϕ_δ be the solution of* (1.2) *and let $\bar\phi\in C_b^{1,1,2}([0,+\infty);\mathbb{R}_*^{2d})$ satisfy*

$$\begin{aligned}\partial_t\bar\phi(t,\mathbf{x},\mathbf{k})&=\tilde{\mathcal{L}}\bar\phi(t,\mathbf{x},\mathbf{k})\\ \bar\phi(0,\mathbf{x},\mathbf{k})&=\phi_0(\mathbf{x},\mathbf{k}),\end{aligned}\tag{2.7}$$

where the initial data $\phi_0\in C^{1,3}(\mathbb{R}_*^{2d})$ is such that $K:=\operatorname{supp}\phi_0(\mathbf{x},\mathbf{k})$ is a compact subset of $\mathbb{R}_*^{2d}$. Then, there exist two constants C, $\alpha_0>0$ such that for all $\delta\in(0,1]$

$$\sup_{(t,\mathbf{x},\mathbf{k})\in[0,T]\times K}\left|\mathbb{E}\phi_\delta\left(t,\mathbf{x},\mathbf{k}\right)-\bar\phi(t,\mathbf{x},\mathbf{k})\right|\le C\delta^{\alpha_0}.\tag{2.8}$$

3. The truncated dynamics

3.1. The random characteristics corresponding to (1.2)

The scaled process $(\mathbf{x}_\delta(t;\mathbf{x},\mathbf{k}),\mathbf{k}_\delta(t;\mathbf{x},\mathbf{k}))$ given by (1.1) satisfies the following system of equations:

$$\begin{cases}\dot{\mathbf{x}}_\delta(t;\mathbf{x},\mathbf{k})=\mathbf{k}_\delta(t;\mathbf{x},\mathbf{k}),\\ \dot{\mathbf{k}}_\delta(t;\mathbf{x},\mathbf{k})=\delta^{-1/2}\,\mathbf{F}\left(\delta^{-1}\mathbf{x}_\delta(t;\mathbf{x},\mathbf{k}),\mathbf{k}_\delta(t;\mathbf{x},\mathbf{k})\right),\\ \mathbf{x}_\delta(0;\mathbf{x},\mathbf{k})=\mathbf{x},\quad \mathbf{k}_\delta(0;\mathbf{x},\mathbf{k})=\mathbf{k}.\end{cases}$$

We denote by $Q^\delta_{\mathbf{x},\mathbf{k}}(\cdot)$ its law over $\mathcal{C}$.

3.2. The definition of stopping times

Let

$$0<\epsilon_0<\epsilon_1<\epsilon_2<1/2,\quad \epsilon_3\in(0,1/2-\epsilon_2),\quad \epsilon_4\in(1/2+\epsilon_0,1-\epsilon_1-\epsilon_2)$$

be some positive constants that will be further determined later on and set

$$M_1=[\delta^{-\epsilon_0}],\quad N=[\delta^{-\epsilon_1}],\quad p=[\delta^{-\epsilon_2}],\quad q=p\,[\delta^{-\epsilon_3}],\quad N_1=Np\,[\delta^{-\epsilon_4}].\tag{3.1}$$

We will specify additional restrictions on the constants ϵ_j as the need for such constraints arises. However, the basic requirement is that ϵ_i, $i\in\{0,1,2,3\}$, should be sufficiently small and ϵ_4 is bigger than $1/2$, less than 1 and can be made as close to 1 as we would need it.

We introduce the following $(\mathcal{M}^t)_{t\geq 0}$-stopping times. Let $\pi \in \mathcal{C}$ be a path. We define the exit time for the $K(\cdot)$ component of the path π from the shell $A(M_1)$, i.e.,

$$T_\delta(\pi) := \inf\Big[\, t \geq 0 : \ |K(t)| \geq M_1^3, \ \text{or } |K(t)| \leq \frac{1}{M_1}\,\Big]. \tag{3.2}$$

Let $t_k^{(p)} := kp^{-1}$ be a mesh of times. We define the "violent turn" stopping time

$$\begin{aligned} S_\delta(\pi) := \inf\Big[\, t \geq 0 : \ &\text{for some } k \geq 0 \text{ we have } t \in \Big[t_k^{(p)}, t_{k+1}^{(p)}\Big) \text{ and} \\ &\hat{K}(t_{k-1}^{(p)}) \cdot \hat{K}(t) \leq 1 - \frac{1}{N}, \ \text{or } \hat{K}\Big(t_k^{(p)} - \frac{1}{N_1}\Big) \cdot \hat{K}(t) \leq 1 - \frac{1}{N}\,\Big], \end{aligned} \tag{3.3}$$

where by convention we set $\hat{K}(-1/p) := \hat{K}(0)$. Note that with the above choice of ϵ_4 we have $Q^\delta_{\mathbf{x},\mathbf{k}}$ a.s. $\hat{K}\left(t_k^{(p)} - 1/N_1\right) \cdot \hat{K}(t_k^{(p)}) > 1 - 1/N$ for all $t_k^{(p)} \leq T_\delta(\pi)$, provided that $\delta \in (0, \delta_0]$ and δ_0 is sufficiently small. Both in (3.3) and in what follows we adopt a customary convention that the infimum of an empty set equals $+\infty$.

For each $t \geq 0$, we denote by $\mathfrak{X}_t(\pi) := \bigcup_{0\leq s\leq t} X(s;\pi)$ the trace of the spatial component of the path π up to time t, and by $\mathfrak{X}_t(q;\pi) := [\mathbf{x} : \text{dist}\,(\mathbf{x}, \mathfrak{X}_t(\pi)) \leq 1/q]$ a tubular region around the path. We introduce the stopping time

$$U_\delta(\pi) := \inf\left[\, t \geq 0 : \exists\, k \geq 1 \ \text{ and } t \in [t_k^{(p)}, t_{k+1}^{(p)}) \text{ for which } X(t) \in \mathfrak{X}_{t_{k-1}^{(p)}}(q)\right]. \tag{3.4}$$

Finally, we set the stopping time

$$\tau_\delta(\pi) := T_\delta(\pi) \wedge S_\delta(\pi) \wedge U_\delta(\pi) \wedge \delta^{-1}. \tag{3.5}$$

The last term appearing on the right-hand side of (3.5) ensures that $\tau_\delta < +\infty$ a.s.

Let p, q, N, N_1, M_1 be the positive integers defined in (3.1). Let $a_1 = 2$ and $a_2 = 3/2$. The functions $\psi_j : \mathbb{R}^d \times \mathbb{S}^{d-1} \to [0,1]$, $j = 1, 2$ are of C^∞ class and satisfy

$$\psi_j(\mathbf{k}, \mathbf{l}) = \begin{cases} 1, & \text{if } \hat{\mathbf{k}} \cdot \mathbf{l} \geq 1 - 1/N \quad \text{and} \quad M_1^{-1} \leq |\mathbf{k}| \leq M_1^3 \\ 0, & \text{if } \hat{\mathbf{k}} \cdot \mathbf{l} \leq 1 - a_j/N, \quad \text{or} \quad |\mathbf{k}| \leq (2M_1)^{-1}, \quad \text{or} \quad |\mathbf{k}| \geq (2M_1)^3. \end{cases}$$

One can construct ψ_j in such a way that for arbitrary nonnegative integers m, n it is possible to find a constant $C_{m,n}$ for which $\|\psi_j\|_{m,n} \leq C_{m,n} N^{m+n} M_1^m$. Let

$$\Psi(t, \mathbf{k}; \pi) := \begin{cases} \psi_1\left(\mathbf{k}, \hat{K}\left(t_{k-1}^{(p)}\right)\right) \psi_2\left(\mathbf{k}, \hat{K}\left(t_k^{(p)} - 1/N_1\right)\right) & \text{for } t \in [t_k^{(p)}, t_{k+1}^{(p)}) \\ & \text{and } k \geq 1 \\ \psi_2(\mathbf{k}, \hat{K}(0)) & \text{for } t \in [0, t_1^{(p)}). \end{cases}$$

Let $\phi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ be a function of the C^∞ class that satisfies $\phi(\mathbf{y}, \mathbf{x}) = 1$, when $|\mathbf{y} - \mathbf{x}| \geq 1/(2q)$ and $\phi(\mathbf{y}, \mathbf{x}) = 0$, when $|\mathbf{y} - \mathbf{x}| \leq 1/(3q)$. Again, in this case we can construct ϕ in such a way that $\|\phi\|_{m,n} \leq Cq^{m+n}$ for arbitrary integers m, n and a suitably chosen constant C. The function $\phi_k : \mathbb{R}^d \times \mathcal{C} \to [0,1]$ for a fixed path π is given by

$$\phi_k(\mathbf{y};\pi) = \prod_{0 \le l/q \le t^{(p)}_{k-1}} \phi\left(\mathbf{y}, X\left(\frac{l}{q}\right)\right).$$

We set

$$\Phi(t,\mathbf{y};\pi) := \begin{cases} 1, & \text{if } 0 \le t < t^{(p)}_1 \\ \phi_k(\mathbf{y};\pi), & \text{if } t^{(p)}_k \le t < t^{(p)}_{k+1}. \end{cases} \tag{3.6}$$

For a given $t \ge 0$, $(\mathbf{y},\mathbf{k}) \in \mathbb{R}^{2d}_*$ and $\pi \in \mathcal{C}$ let us denote $\Theta(t,\mathbf{y},\mathbf{k};\pi) := \Psi(t,\mathbf{k};\pi) \cdot \Phi(t,\mathbf{y};\pi)$. The following lemma can be verified by a direct calculation.

Lemma 3.1. *Let (β_1,β_2) be a multi-index with nonnegative integer components, $m = |\beta_1| + |\beta_2|$. There exists a constant C depending only on m such that $|\partial^{\beta_1}_{\mathbf{y}}\partial^{\beta_2}_{\mathbf{k}}\Theta(t,\mathbf{y},\mathbf{k};\pi)| \le C(T+1)^{|\beta_1|}q^{2|\beta_1|}(NM_1)^{|\beta_2|}$ for all $t \in [0,T]$, $(\mathbf{y},\mathbf{k}) \in \mathcal{A}(2M_1)$, $\pi \in \mathcal{C}$.*

Finally, let us set

$$\mathbf{F}_\delta(t,\mathbf{y},\mathbf{l};\pi,\omega) = \Theta(t,\delta\mathbf{y},\mathbf{l};\pi)\mathbf{F}(\mathbf{y},\mathbf{l};\omega).$$

Note that according to Lemma 3.1 we obtain that

$$|\partial^{\beta_1}_{\mathbf{y}}\partial^{\beta_2}_{\mathbf{k}}\Theta(t,\delta\mathbf{y},\mathbf{l};\pi)| \le C(T+1)^{|\beta_1|}\delta^{|\beta_1|[1-2(\epsilon_2+\epsilon_3)]}(NM_1)^{|\beta_2|} \tag{3.7}$$

for all $t \in [0,T]$, $(\mathbf{y},\mathbf{k}) \in \mathcal{A}(2M_1)$, $\pi \in \mathcal{C}$. For a fixed $(\mathbf{x},\mathbf{k}) \in \mathbb{R}^{2d}_*$, $\delta > 0$ and $\omega \in \Omega$ we consider the modified particle dynamics with the cut-off that is described by the stochastic process $(\boldsymbol{y}^{(\delta)}(t;\mathbf{x},\mathbf{k},\omega), \boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k},\omega))_{t\ge 0}$ whose paths are the solutions of the following equation:

$$\begin{cases} \frac{d\boldsymbol{y}^{(\delta)}(t;\mathbf{x},\mathbf{k})}{dt} = \boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k},), \\ \frac{d\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k})}{dt} = \delta^{-1/2}\,\mathbf{F}_\delta\left(t, \delta^{-1}\boldsymbol{y}^{(\delta)}(t;\mathbf{x},\mathbf{k}), \boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k}); \boldsymbol{y}^{(\delta)}(\cdot;\mathbf{x},\mathbf{k}), \boldsymbol{l}^{(\delta)}(\cdot;\mathbf{x},\mathbf{k})\right), \\ \boldsymbol{y}^{(\delta)}(0;\mathbf{x},\mathbf{k}) = \mathbf{x}, \quad \boldsymbol{l}^{(\delta)}(0;\mathbf{x},\mathbf{k}) = \mathbf{k}. \end{cases} \tag{3.8}$$

We will denote by $\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}$ the law of the modified process $(\boldsymbol{y}^{(\delta)}(\cdot;\mathbf{x},\mathbf{k}), \boldsymbol{l}^{(\delta)}(\cdot;\mathbf{x},\mathbf{k}))$ over $\mathcal{C}$ for a given $\delta > 0$ and by $\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}$ the corresponding expectation. We shall also omit writing the parameters $(\mathbf{x},\mathbf{k})$ if they are obvious from the context. From the construction of the cut-offs we conclude immediately that if $\mathbf{k} \in A(M_1)$, then $(2M_1)^{-1} \le |\boldsymbol{l}^{(\delta)}(t;\mathbf{x},\mathbf{k})| \le (2M_1)^3$ for all $t \ge 0$. We also have the following.

Proposition 3.2. *Assume that the initial velocity $\mathbf{k} \in A(M_1)$. Then, there exists $\delta_0 > 0$ such that for all $\delta \in (0,\delta_0]$,*

$$\hat{\boldsymbol{l}}^{(\delta)}(t) \cdot \hat{\boldsymbol{l}}^{(\delta)}(t^{(p)}_{k-1}) \ge 1 - \frac{2}{N} \tag{3.9}$$

and

$$\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}\left(t_k^{(p)}-\frac{1}{N_1}\right)\geq 1-\frac{3}{2N} \tag{3.10}$$

for $t\in[t_k^{(p)},t_{k+1}^{(p)})$ *and all* $k\geq 0$. *Moreover,*

$$\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}(t_{k-1}^{(p)})\geq 1-\frac{2}{N} \tag{3.11}$$

for $t\in[t_{k-1}^{(p)},t_{k+1}^{(p)})$ *and all* $k\geq 0$.

Proof. We show (3.9) by induction. For $k=0$ the statement reduces to showing that

$$\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}(0)\geq 1-\frac{3}{2N},\quad \forall\, t\in[0,t_1^{(p)}). \tag{3.12}$$

The set $G:=[t\in[0,t_1^{(p)}):\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}(0)<1-3/(2N)]$ is open (in relative topology). We can find therefore a countable family of disjoint open intervals (a_i,b_i) s.t. $G=\bigcup_i(a_i,b_i)\cap[0,t_1^{(i)})$. Since G^c is non-empty (0 belongs to it) we must have $a_i\in G^c$ so $\hat{\boldsymbol{l}}^{(\delta)}(a_i)\cdot\hat{\boldsymbol{l}}^{(\delta)}(0)=1-3/(2N)$. Using the cut-off condition we conclude that $\dot{\boldsymbol{l}}^{(\delta)}(t)=0$ for $t\in(a_i,b_i)$, hence $\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}(0)=1-3/(2N)$ for $t\in(a_i,b_i)$. As a result we conclude that $a_i=b_i$ (or equivalently stating $(a_i,b_i)=\emptyset$) for all i, thus the set G is empty.

Suppose that (3.9) holds for a certain k. Note that

$$\hat{\boldsymbol{l}}^{(\delta)}(t_{k+1}^{(p)})\cdot\hat{\boldsymbol{l}}^{(\delta)}(t_k^{(p)})\geq 1-\frac{2}{N} \tag{3.13}$$

and

$$\hat{\boldsymbol{l}}^{(\delta)}(t_{k+1}^{(p)})\cdot\hat{\boldsymbol{l}}^{(\delta)}\left(t_{k+1}^{(p)}-\frac{1}{N_1}\right)\geq 1-\frac{3}{2N}. \tag{3.14}$$

The estimate (3.13) is a consequence of the inductive assumption (3.10) applied for $t=t_{k+1}^{(p)}$ and the bound

$$\left|\hat{\boldsymbol{l}}^{(\delta)}\left(t_k^{(p)}-\frac{1}{N_1}\right)-\hat{\boldsymbol{l}}^{(\delta)}(t_k^{(p)})\right|\leq\frac{2\tilde{D}M_1}{\delta^{1/2}N_1}\leq\frac{1}{2N}, \tag{3.15}$$

provided that $\epsilon_4-1/2-\epsilon_0-\epsilon_2>0$ and $\delta\in(0,\delta_0]$, where δ_0 is sufficiently small but independent of k. The estimate (3.14) can be obtained in a similar fashion. Now we repeat the argument used for $k=0$ and conclude that (3.13) holds for all $t\in[t_{k+1}^{(p)},t_{k+2}^{(p)})$.

As for the proof of (3.11) it is a conclusion from (3.9) and (3.10). We only need to prove this estimate for $t\in[t_{k-1}^{(p)},t_k^{(p)})$ since for $t\in[t_k^{(p)},t_{k+1}^{(p)})$ it is covered by (3.9). For $k=0$ the proof reduces to showing yet again (3.12) and this has been already done. Suppose therefore that $k\geq 1$. According to (3.9) we have then

$$\hat{\boldsymbol{l}}^{(\delta)}(t)\cdot\hat{\boldsymbol{l}}^{(\delta)}\left(t_{k-1}^{(p)}-\frac{1}{N_1}\right)\geq 1-\frac{3}{2N}$$

for $t \in [t^{(p)}_{k-1}, t^{(p)}_k)$. Using (3.15) with $t^{(p)}_{k-1}$ in place of $t^{(p)}_k$ we obtain from the above estimate that

$$\hat{\boldsymbol{l}}^{(\delta)}(t) \cdot \hat{\boldsymbol{l}}^{(\delta)}\left(t^{(p)}_{k-1}\right) \geq 1 - \frac{2}{N} \quad \text{for } t \in [t^{(p)}_{k-1}, t^{(p)}_k). \qquad \square$$

3.3. Some consequences of the mixing assumption

For any $t \geq 0$ we denote by $\mathcal{F}_t$ the σ-algebra generated by $(\boldsymbol{y}^{(\delta)}(s), \boldsymbol{l}^{(\delta)}(s))$, $s \leq t$. Here we suppress, for the sake of abbreviation, writing the initial data in the notation of the trajectory. In this section we assume that $M > 1$ is fixed, $X_1, X_2 : (\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d^2})^2 \to \mathbb{R}$ are certain continuous functions, Z is a random variable and g_1, g_2 are $\mathbb{R}^{2d}$-valued random vectors. We suppose further that Z, g_1, g_2, are $\mathcal{F}_t$-measurable, while $\tilde{X}_1, \tilde{X}_2$ are random fields of the form

$$\tilde{X}_i(\mathbf{x}, \mathbf{k}) = X_i\left(\left(\nabla^j_{\mathbf{k}}\mathbf{F}(\mathbf{x}, \mathbf{k}), \nabla_{\mathbf{x}}\nabla^j_{\mathbf{k}}\mathbf{F}(\mathbf{x}, \mathbf{k})\right)_{j=0,1,2}\right).$$

For $i = 1, 2$ we denote $g_i := (g^{(1)}_i, g^{(2)}_i) \in \mathbb{R}^{2d}$. We also let

$$U(\theta_1, \theta_2) := \mathbb{E}\left[\tilde{X}_1(\theta_1)\tilde{X}_2(\theta_2)\right], \quad (\theta_1, \theta_2) \in \mathbb{R}^{2d}.$$

The following mixing lemma is a direct consequence of Lemmas 2 and 5 of [2].

Lemma 3.3. (i) *Assume that $r, t \geq 0$ and*

$$\inf_{u \leq t}\left|g^{(1)}_i - \frac{\boldsymbol{y}^{(\delta)}(u)}{\delta}\right| \geq \frac{r}{\delta}, \tag{3.16}$$

for $i = 1, 2$, $\mathbb{P}$-a.s. on the event $[Z \neq 0]$. Then, we have

$$\left|\mathbb{E}\left[\tilde{X}_1(g_1)\tilde{X}_2(g_2)Z\right] - \mathbb{E}\left[U(g_1, g_2)Z\right]\right| \leq 2\phi\left(\frac{r}{2\delta}\right)\|X_1\|_{L^\infty}\|X_2\|_{L^\infty}\|Z\|_{L^1(\Omega)}.$$

(ii) *Let $\mathbb{E}\tilde{X}_1(\mathbf{0}, k) = 0$ for all $k \geq 0$. Furthermore, we assume that g_2 satisfies (3.16), g_1 satisfies*

$$\inf_{u \leq t}\left|g^{(1)}_1 - \frac{\boldsymbol{y}^{(\delta)}(u)}{\delta}\right| \geq \frac{r + r_1}{\delta},$$

and $|g^{(1)}_1 - g^{(1)}_2| \geq r_1\delta^{-1}$ for some $r_1 \geq 0$, $\mathbb{P}$-a.s. on the event $[Z \neq 0]$. Then, we have

$$\begin{aligned}&\left|\mathbb{E}\left[\tilde{X}_1(g_1)\tilde{X}_2(g_2)\,Z\right] - \mathbb{E}\left[U(g_1, g_2)Z\right]\right|\\ &\qquad\leq C\phi^{1/2}\left(\frac{r}{2\delta}\right)\phi^{1/2}\left(\frac{r_1}{2\delta}\right)\|X_1\|_{L^\infty}\|X_2\|_{L^\infty}\|Z\|_{L^1(\Omega)}\end{aligned}$$

for some absolute constant $C > 0$.

4. The approximate martingale problem and the proof of Theorem 2.2

4.1. The augmented path measure

We define functions $D^{(\delta)}_{mn}(\mathbf{k})$, $E^{(\delta)}_m(\mathbf{k})$, $\phi^{(\delta)}(\mathbf{k})$ that are respectively C^{2n_*}, C^{2n_*-1} and C^∞ smooth and satisfy $D^{(\delta)}_{mn}(\mathbf{k}) = D_{mn}(\mathbf{k})$, $E^{(\delta)}_m(\mathbf{k}) = E_m(\mathbf{k})$, $\phi^{(\delta)}(\mathbf{k}) = \mathbf{k}$ for $\mathbf{k} \in A(2M_1)$. We assume that their norms in the respective $C^k_b(\mathbb{R}^d)$ spaces can be bounded by CM_1^I for some $C, I > 0$. Finally, we suppose that the matrix $[D^{(\delta)}_{mn}(\mathbf{k})]$ is symmetric for all $\mathbf{k}$ and uniformly positive definite with the positivity constant CM_1^{-2} for some $C > 0$. Let

$$\begin{aligned}\tilde{\mathcal{L}}^{(\delta)} F(\mathbf{x},\mathbf{k}) := &\sum_{m,n=1}^{d} D^{(\delta)}_{mn}(\mathbf{k})\partial^2_{k_m,k_n} F(\mathbf{x},\mathbf{k}) \\ &+ \sum_{m=1}^{d} E^{(\delta)}_{mn}(\mathbf{k})\partial_{k_m} F(\mathbf{x},\mathbf{k}) + \phi^{(\delta)}(\mathbf{k})\cdot\nabla_{\mathbf{x}} F(\mathbf{x},\mathbf{k}),\end{aligned} \tag{4.1}$$

for any $F \in C^\infty_c(\mathbb{R}^{2d})$. Let $\mathfrak{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}$ be the law of the diffusion corresponding to the generator $\mathcal{L}^{(\delta)}$ on the space $\mathcal{C}$. The following construction of the augmentation of path measures has been carried out in Section 6.1 of [7]. Let $s \ge 0$ and $\pi \in \mathcal{C}$ be fixed. Then, according to Lemma 6.1.1 of ibid. there exists a unique probability measure, that is denoted by $\delta_\pi \otimes_s \mathfrak{Q}^{(\delta)}_{X(s),K(s)}$, such that for any pair of events $A \in \mathcal{M}^s$, $B \in \mathcal{M}$ we have $\delta_\pi \otimes_s \mathfrak{Q}^{(\delta)}_{X(s),K(s)}[A] = \mathbf{1}_A(\pi)$ and $\delta_\pi \otimes_s \mathfrak{Q}^{(\delta)}_{X(s),K(s)}[\theta_s(B)] = \mathfrak{Q}^{(\delta)}_{X(s),K(s)}[B]$. The following result is a direct consequence of Theorem 6.2.1 of [7].

Proposition 4.1. *There exists a unique probability measure $R^{(\delta)}_{\mathbf{x},\mathbf{k}}$ on $\mathcal{C}$ such that $R^{(\delta)}_{\mathbf{x},\mathbf{k}}[A] := Q^{(\delta)}_{\mathbf{x},\mathbf{k}}[A]$ for all $A \in \mathcal{M}^{\tau_\delta}$ and the regular conditional probability distribution of $R^{(\delta)}_{\mathbf{x},\mathbf{k}}[\,\cdot\,|\mathcal{M}^{\tau_\delta}]$ is given by $\delta_\pi \otimes_{\tau_\delta(\pi)} \mathfrak{Q}^{(\delta)}_{X(\tau_\delta(\pi)),K(\tau_\delta(\pi))}$, $\pi \in \mathcal{C}$. This measure shall be also denoted by $Q^{(\delta)}_{\mathbf{x},\mathbf{k}} \otimes_{\tau_\delta} \mathfrak{Q}^{(\delta)}_{X(\tau_\delta),K(\tau_\delta)}$.*

Note that for any $(\mathbf{x},\mathbf{k}) \in \mathbb{R}^{2d}_*$ and $A \in \mathcal{M}^{\tau_\delta}$ we have

$$R^{(\delta)}_{\mathbf{x},\mathbf{k}}[A] = Q^{(\delta)}_{\mathbf{x},\mathbf{k}}[A] = \tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}[A], \tag{4.2}$$

that is, the law of the augmented process coincides with that of the true process, and of the modified process with the cut-offs until the stopping time τ_δ. Hence, according to the uniqueness part of Proposition 4.1, in such a case $Q^{(\delta)}_{\mathbf{x},\mathbf{k}} \otimes_{\tau_\delta} \mathfrak{Q}^{(\delta)}_{X(\tau_\delta),K(\tau_\delta)} = \tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}} \otimes_{\tau_\delta} \mathfrak{Q}^{(\delta)}_{X(\tau_\delta),K(\tau_\delta)}$. We denote by $E^{(\delta)}_{\mathbf{x},\mathbf{k}}$ the expectation with respect to the augmented measure described by the above proposition.

In the following proposition we prove that the augmented measure approximately satisfies the Stroock and Varadhan martingale problem corresponding to the diffusion described by (4.1). To formulate this result we need some more notation. For any $G \in C_b^{1,1,3}([0,+\infty) \times \mathbb{R}_*^{2d})$ we let

$$N_t^{(\delta)}(G) := G(t, X(t), K(t)) - G(0, X(0), K(0)) - \int_0^t (\partial_\varrho + \tilde{\mathcal{L}}^{(\delta)}) G(\varrho, X(\varrho), K(\varrho))\, d\varrho, \quad t \ge 0.$$

When $\pi \in \mathcal{C}$ is given, we let

$$\widehat{N}_t(G) := G(t, X(t), K(t)) - G(0, X(0), K(0)) - \int_0^t (\partial_\varrho + \widehat{\mathcal{L}}_\varrho) G(\varrho, X(\varrho), K(\varrho); \pi)\, d\varrho, \quad t \ge 0,$$

where

$$\widehat{\mathcal{L}}_t G(t, \mathbf{x}, \mathbf{k}; \pi) := \mathbf{k} \cdot \nabla_{\mathbf{x}} G(t, \mathbf{x}, \mathbf{k}) + \Theta^2(t, X(t), K(t); \pi) \mathcal{L}_{\mathbf{k}} G(t, \mathbf{x}, \mathbf{k})$$
$$- \Theta(t, X(t), K(t); \pi) \sum_{m,n=1}^d \partial_{K_m} \Theta(t, X(t), K(t); \pi) D_{m,n}(\mathbf{k}) \partial_{k_n} G(t, \mathbf{x}, \mathbf{k}),$$

and the function $\Theta(\cdot)$ is given by formula (3.6). It follows from the definition of the stopping time $\tau_\delta(\pi)$ and the cut-off function Θ that

$$\nabla_K \Theta(t, X(t), K(t); \pi) = \mathbf{0}, \quad \Theta(t, X(t), K(t); \pi) = 1, \ t \in [0, \tau_\delta(\pi)],$$

hence

$$\widehat{\mathcal{L}}_t G(t, X(t), K(t); \pi) = \tilde{\mathcal{L}} G(t, X(t), K(t); \pi), \ t \in [0, \tau_\delta(\pi)]. \tag{4.3}$$

With this notation we can formulate the following.

Proposition 4.2. *Suppose that* $(\mathbf{x}, \mathbf{k}) \in \mathcal{A}(M_1)$ *and* $\zeta \in C_b((\mathbb{R}_*^{2d})^n)$ *is non-negative,* $0 \le t_1 < \cdots < t_n \le T_* \le t < u \le T$ *and* $G \in C_b^{1,1,3}([T_*, T] \times \mathbb{R}_*^{2d})$. *Denote* $\tilde{\zeta}(\pi) := \zeta(X(t_1), K(t_1), \ldots, X(t_n), K(t_n))$, $\pi \in \mathcal{C}$. *Then, the following are true:*

i) *Suppose that* $\gamma_0' \in (0,1)$ *and* $t - T_* \ge \delta^{\gamma_0'}$. *Then, there exist constants* γ_1, $C > 0$ *such that*

$$\left| \tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)} \left\{ [\widehat{N}_u(G) - \widehat{N}_t(G)] \tilde{\zeta} \right\} \right| \le C \delta^{\gamma_1} (u - t) \|G\|_{1,1,3}^{[T_*,T]} (T+1)^2 \tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)} \tilde{\zeta}.$$

ii) *Suppose that* $\gamma_0 \in (0, 1/2)$ *and that* $v - t \ge \delta^{\gamma_0}$. *Then, there exist constants* γ_1, $C > 0$ *such that*

$$\left| E_{\mathbf{x},\mathbf{k}}^{(\delta)} \left\{ \left[N_v^{(\delta)}(G) - N_t^{(\delta)}(G) \right] \tilde{\zeta} \right\} \right| \le C \delta^{\gamma_1} (v - t) \|G\|_{1,1,3}^{[T_*,T]} (T+1)^2 E_{\mathbf{x},\mathbf{k}}^{(\delta)} \tilde{\zeta}. \tag{4.4}$$

In both statements i) *and* ii) *the choice of the constants* γ_1, C *does not depend on* $(\mathbf{x}, \mathbf{k})$, $\delta \in (0,1]$, ζ, *times* $t_1, \ldots, t_n, T, v, t$, *or the function* G.

The arguments used to demonstrate parts i) and ii) of the above proposition are virtually the same as the ones used in showing Lemma 3.5 and Proposition 3.4 of [3], respectively. The proofs presented ibid. have been carried out for the case of hamiltonian flows, but that fact has not been essential.

4.2. The proof of Theorem 2.2

The crucial tool in proving estimate (2.8) is the following.

Proposition 4.3. *Assume that the dimension $d \geq 3$ and $K \subset \mathbb{R}^{2d}_*$ is compact. Then, one can choose ϵ_i, $i = 0, \ldots, 4$, in such a way that there exist constants $C, \gamma > 0$, independent of δ but possibly dependent on K and T, for which*

$$R^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T] \leq C\delta^\gamma, \quad \forall\, \delta \in (0,1],\ (\mathbf{x},\mathbf{k}) \in K. \tag{4.5}$$

We postpone the demonstration of the above proposition until the following section. In the meantime, taking its assertion for granted, we show how to finish the proof of Theorem 2.2. Let $u \in [\delta^{\gamma_0'}, T]$, where we assume that γ_0' (as in the statement of part i) of Proposition 4.2) belongs to the interval $(1/2, 1)$. Substituting for $G(t, \mathbf{x}, \mathbf{k}) := \bar{\phi}(u - t, \mathbf{x}, \mathbf{k})$, $\zeta \equiv 1$ into (4.4) we obtain (taking $v = u$, $t = \delta^{\gamma_0'}$)

$$\Bigg|\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\Bigg[\phi_0(X(u), K(u)) - \bar{\phi}(u - \delta^{\gamma_0'}, X(\delta^{\gamma_0'}), K(\delta^{\gamma_0'})) - \int\limits_{\delta^{\gamma_0'}}^{u} (\partial_\varrho + \widehat{\mathcal{L}}_\varrho) G(\varrho, X(\varrho), K(\varrho))\, d\varrho\Bigg]\Bigg| \leq C\|G\|^{[0,T]}_{1,1,3}\delta^{\gamma_1}(T+1)^2, \quad \forall\, \delta \in (0,1].$$

Using the fact that $|X(\delta^{\gamma_0'}) - \mathbf{x}| \leq C\delta^{\gamma_0' - \epsilon_0}$, $|K(\delta^{\gamma_0'}) - \mathbf{k}| \leq C\delta^{\gamma_0' - 1/2}$, $\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}$-a.s. for some deterministic constant $C > 0$, cf. (3.8), we obtain that there exist constants $C, \gamma > 0$ such that

$$\Bigg|\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\Bigg[\phi_0(X(u), K(u)) - \bar{\phi}(u, \mathbf{x}, \mathbf{k}) - \int\limits_0^u (\partial_\varrho + \widehat{\mathcal{L}}_\varrho) G(\varrho, X(\varrho), K(\varrho))\, d\varrho\Bigg]\Bigg| \leq C\|G\|^{[0,T]}_{1,1,3}\delta^{\gamma}(T+1)^2, \quad \delta \in (0,1],\ u \in [0,T]. \tag{4.6}$$

On the event $[\tau_\delta \geq T]$ we have however

$$(\partial_\varrho + \widehat{\mathcal{L}}_\varrho) G(\varrho, X(\varrho), K(\varrho)) = (\partial_\varrho + \tilde{\mathcal{L}}_\varrho) G(\varrho, X(\varrho), K(\varrho)) = 0$$

for all $u \in [0, T]$, cf. (4.3).

Since the random variable $[\phi_0(X(u),K(u)) - \bar{\phi}(u,\mathbf{x},\mathbf{k})]\chi_{[\tau_\delta \geq T]}$ is $\mathcal{M}^{\tau_\delta}$ measurable we obtain, using (4.2), that

$$\begin{aligned}
&\left| E^{(\delta)}_{\mathbf{x},\mathbf{k}} \left[\phi_0(X(u),K(u)) - \bar{\phi}(u,\mathbf{x},\mathbf{k}),\, \tau_\delta \geq T\right]\right| \\
&\quad = \left| \tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}} \left[\phi_0(X(u),K(u)) - \bar{\phi}(u,\mathbf{x},\mathbf{k}),\, \tau_\delta \geq T\right]\right| \\
&\quad \overset{(4.6)}{\leq} C\|G\|^{[0,T]}_{1,1,3}\delta^\gamma (T+1)^2 + \left(2\|\phi_0\|_{0,0} + T\|G\|^{[0,T]}_{1,1,2}\right) \tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T].
\end{aligned}$$

Using $\mathcal{M}^{\tau_\delta}$-measurability of the event $[\tau_\delta < T]$ we obtain that $\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T] = R^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T]$, and by virtue of Proposition 4.3 we can estimate the right-hand side of (4.7) by

$$C\|G\|^{[0,T]}_{1,1,3}\delta^\gamma (T+1)^2 + C\delta^\gamma \left(2\|\phi_0\|_{0,0} + T\|G\|^{[0,T]}_{1,1,2}\right). \tag{4.7}$$

Theorem 2.58, p. 53 of [6] allows us to estimate $\|G\|^{[0,T]}_{1,1,3}$ and $\|G\|^{[0,T]}_{1,1,2}$ by $\|\phi_0\|_{1,3}$ and $\|\phi_0\|_{1,2}$, respectively. Summarizing, we proved that the expression in (4.7) can be bounded by $C\delta^\gamma\|\phi_0\|_{1,3}$ for some $C,\gamma > 0$. On the other hand, the expression under the absolute value on the utmost left hand side of (4.7) equals

$$\begin{aligned}
&\int \left[\phi_0(X(u),K(u)) - \bar{\phi}(u,\mathbf{x},\mathbf{k})\,\right] Q^{(\delta)}_{\mathbf{x},\mathbf{k}}(d\pi) \\
&\qquad - \int \left[\phi_0(X(u),K(u)) - \bar{\phi}(u,\mathbf{x},\mathbf{k}),\, \tau_\delta < T\right] Q^{(\delta)}_{\mathbf{x},\mathbf{k}}(d\pi).
\end{aligned}$$

By virtue of Proposition 4.3 the second term can be estimated by

$$2\|\phi_0\|_{0,0} Q^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T] = 2\|\phi_0\|_{0,0} R^{(\delta)}_{\mathbf{x},\mathbf{k}}[\tau_\delta < T] \leq C\delta^\gamma\|\phi_0\|_{0,0}.$$

Since

$$\mathbb{E}\phi_\delta\left(\frac{u}{\delta},\frac{\mathbf{x}}{\delta},\mathbf{k}\right) = \mathbb{E}\phi_0(\mathbf{z}^{(\delta)}(u;\mathbf{x},\mathbf{k}),\mathbf{m}^{(\delta)}(u;\mathbf{x},\mathbf{k})) = \int \phi_0(X(u),K(u)) Q^{(\delta)}_{\mathbf{x},\mathbf{k}}(d\pi)$$

we conclude from the above that the left-hand side of (2.8) can be estimated by $C\delta^\gamma\|\phi_0\|_{1,3}$ for some constants $C,\gamma > 0$ independent of $\delta > 0$.

4.3. The estimate of the stopping time

With no loss of generality we can assume that T, $\delta^{-1} > 1$, since otherwise (4.5) holds with $C = \gamma = 1$. We obviously have then

$$[\,\tau_\delta < T] = [\,U_\delta \leq \tau_\delta,\, U_\delta < T] \cup [\,S_\delta \leq \tau_\delta,\, S_\delta < T] \cup [\,T_\delta \leq \tau_\delta,\, T_\delta < T] \tag{4.8}$$

with the stopping times S_δ, U_δ and T_δ defined in (3.2)–(3.4). Let us denote the events appearing on the right-hand side of (4.8) by $A(\delta)$, $B(\delta)$ and $C(\delta)$, respectively. To show that (4.8) holds we prove that the $R^{(\delta)}_{\mathbf{x},\mathbf{k}}$ probabilities of all these events can be estimated by $C\delta^\gamma$ for some $C,\gamma > 0$: see (4.16), (4.17) and (4.28) below.

4.3.1. An estimate of $R^{(\delta)}_{\mathbf{x},\mathbf{k}}[A(\delta)]$. Note that then

$$A(\delta)\subset\tilde A(\delta):=\left[\left|X\left(\frac{j}{q}\right)-X\left(\frac{i}{q}\right)\right|\le\frac{M_1^3}{q}:1\le i\le j\le[Tq],\quad |i-j|\ge\frac{q}{p}\right]$$

and thus

$$R^{(\delta)}_{\mathbf{x},\mathbf{k}}[A(\delta)]\le[Tq]^2\max\left\{R^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[\left|X\left(\frac{j}{q}\right)-X\left(\frac{i}{q}\right)\right|\le\frac{M_1^3}{q}\right]:\begin{array}{c}1\le i\le j\le[Tq]\\|i-j|\ge\frac{q}{p}\end{array}\right\}.$$

Suppose that $f^{(\delta)}:\mathbb{R}^d\to[0,1]$ is a C^∞-regular function that satisfies $f(\mathbf{x})=1$, if $|\mathbf{x}|\le 3M_1^3/(2q)$ and $f^{(\delta)}(\mathbf{x})=0$, if $|\mathbf{x}|\ge 2M_1^3/q$. We assume furthermore that i,j are positive integers such that $(j-i)/q\in[0,1]$ and $\|f^{(\delta)}\|_3\le q^3/M_1^9$. For any $\mathbf{x}_0\in\mathbb{R}^d$ and $i/q\le t\le j/q$ define

$$G_j(t,\mathbf{x},\mathbf{k};\mathbf{x}_0):=\mathfrak{M}^{(\delta)}_{\mathbf{x},\mathbf{k}}f^{(\delta)}\left(X\left(\frac{j}{q}-t\right)-\mathbf{x}_0\right)$$

for $(\mathbf{x},\mathbf{k})\in\mathcal{A}(M_1)$. Here $\mathfrak{M}^{(\delta)}_{\mathbf{x},\mathbf{k}}$ is the expectation corresponding to $\mathfrak{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}$, cf. Section 4.1. $G_j(t,\mathbf{x},\mathbf{k};\mathbf{x}_0)$ is the unique bounded solution of the Kolmogorov equation

$$\begin{cases}\partial_t G_j(t,\mathbf{x},\mathbf{k};\mathbf{x}_0)+\tilde{\mathcal{L}}^{(\delta)}G_j(t,\mathbf{x},\mathbf{k};\mathbf{x}_0)=0, & i/q\le t\le j/q,\\ G_j(j/q,\mathbf{x},\mathbf{k};\mathbf{x}_0)=f^{(\delta)}(\mathbf{x}-\mathbf{x}_0).\end{cases}$$

It can be then shown using the same argument as in [6], Theorem 2.58, p. 53, that there exist constants $C,I>0$ such that

$$\begin{aligned}\|G_j(\cdot,\cdot,\cdot;\mathbf{x}_0)\|_{1,1,3}^{[i/q,j/q]}&\le CM_1^I\|f^{(\delta)}\|_3\le Cq^3M_1^{I-9}\\&\le C\delta^{(9-I)\epsilon_0-2(\epsilon_2+\epsilon_3)},\quad j\in\{0,\dots,[qT]\}.\end{aligned}\tag{4.9}$$

Hence, using part ii) of Proposition 4.2 with $v=j/q$ and $t=i/q$ (note that $v-t\ge 1/p\ge\delta^{\epsilon_2}$ and $\epsilon_2\in(0,1/2)$), we obtain that there exists $\gamma_1>0$ such that

$$\begin{aligned}&\left|E^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[f^{(\delta)}\left(X\left(\frac{j}{q}\right)-\mathbf{x}_0\right)-G_j\left(\frac{i}{q},X\left(\frac{i}{q}\right),K\left(\frac{i}{q}\right);\mathbf{x}_0\right)\,\Big|\,\mathcal{M}^{i/q}\right]\right|\\&\le C\,\frac{j-i}{q}\,\|G_j(\cdot,\cdot,\cdot;\mathbf{x}_0)\|_{1,1,3}^{[i/q,j/q]}\delta^{\gamma_1},\quad\forall\,\delta\in(0,1].\end{aligned}\tag{4.10}$$

Combining (4.10) and (4.9) we obtain that the left-hand side of (4.10) is less than, or equal to $C\,\delta^{\gamma_1+(9-I)\epsilon_0-3(\epsilon_2+\epsilon_3)}$ for all $\delta\in(0,1]$. Let now $i_0=j-\dfrac{q}{p}$ so that $1\le i\le i_0\le j\le[Tq]$. We have

$$\begin{aligned}R^{(\delta)}_{\mathbf{x},\mathbf{k}}&\left[\left|X\left(\frac{j}{q}\right)-X\left(\frac{i}{q}\right)\right|\le\frac{M_1^3}{q}\right]\le E^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[f^{(\delta)}\left(X\left(\frac{j}{q}\right)-X\left(\frac{i}{q}\right)\right)\right]\\&=E^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[E^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[f^{(\delta)}\left(X\left(\frac{j}{q}\right)-\mathbf{y}\right)\,\Big|\,\mathcal{M}^{i_0/q}\right]_{\big|\mathbf{y}=X(i/q)}\right].\end{aligned}\tag{4.11}$$

According to (4.10) and (4.9) we can estimate the utmost right-hand side of (4.11) by

$$\sup_{\mathbf{x},\mathbf{y},\mathbf{k}} \left\{ \mathfrak{M}^{(\delta)}_{\mathbf{x},\mathbf{k}} f^{(\delta)} \left(X\left(\frac{1}{p}\right) - \mathbf{y} \right) : \mathbf{x},\mathbf{y} \in \mathbb{R}^d,\ \mathbf{k} \in A(2M_1) \right\} + C\,\delta^{\gamma_1 + (9-I)\epsilon_0 - 3(\epsilon_2+\epsilon_3)}. \tag{4.12}$$

Note that obviously

$$\mathfrak{M}^{(\delta)}_{\mathbf{x},\mathbf{k}} f^{(\delta)} \left(X\left(\frac{1}{p}\right) - \mathbf{y} \right) \le \mathfrak{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}} \left[\left| X\left(\frac{1}{p}\right) - \mathbf{y} \right| \le \frac{2M_1^3}{q} \right]. \tag{4.13}$$

We shall focus on estimating the expression appearing on the right-hand side of (4.13).

Let $\partial_m := \partial_{x_m}$ and $\partial_{m+d} := \partial_{k_m}$, $m = 1,\ldots,d$. Suppose that $X_n = \sum_{m=1}^{2d} a_{n,m}(\mathbf{x},\mathbf{k})\partial_m$, $n = 0,\ldots,d$, are certain C^∞ tangent vector fields over R^{2d} whose coefficients satisfy $\|a_{n,m}\|_{p,q} < +\infty$ for all n,m and all non-negative integers p,q. Assume also that $(\mathbf{y}(t;\mathbf{y}_0,\mathbf{l}_0),\mathbf{l}(t;\mathbf{y}_0,\mathbf{l}_0))$ is an $\mathbb{R}^{2d}$-valued diffusion process, which starts at $(\mathbf{y}_0,\mathbf{l}_0)$ and whose generator can be written in the form

$$\mathcal{N}F(\mathbf{y},\mathbf{l}) := \sum_{m=1}^{d} X_m^2 F(\mathbf{y},\mathbf{l}) + X_0 F(\mathbf{y},\mathbf{l}), \quad F \in C_0^\infty(\mathbb{R}^{2d}).$$

Suppose also that the tangent space to $\mathbb{R}^{2d}$ at any point $(\mathbf{y},\mathbf{l})$ is spanned by $X_1,\ldots,X_{2d}$, where $X_{d+1} := [X_0,X_1],\ldots,X_{2d} := [X_0,X_d]$. It is well known from the theory of hypoelliptic diffusions, see e.g. Theorem 5.6, p. 12 [8], that under this condition for each $t > 0$ the random vector $(\mathbf{y}(t;\mathbf{y}_0,\mathbf{l}_0),\mathbf{l}(t;\mathbf{y}_0,\mathbf{l}_0))$ possesses a C^∞-smooth density $q(t,\mathbf{y}_0,\mathbf{l}_0,\mathbf{y},\mathbf{l})$. For $\xi \in \mathbb{S}^{2d-1}$ we let $\mathcal{V}(\mathbf{y},\mathbf{l},\xi) := \sum_{i=1}^{2d}(X_i,\xi)^2_{\mathbb{R}^{2d}}$. Let also $\mathcal{V}(\mathbf{y},\mathbf{l}) := \inf_{\xi\in\mathbb{S}^{2d-1}} \mathcal{V}(\mathbf{y},\mathbf{l},\xi)$ and $C_0 := \sum_{n=0}^{d}\sum_{m=1}^{2d}\|a_{n,m}\|_{0,0}$. Then, according to Corollary 3.25 p. 22 of [5], one can find an integer $I \ge 1$, positive constants λ,ν that are independent of the fields $X_0,\ldots,X_d$ and a constant $K > 0$ depending only on $\|a_{n,m}\|_{p,q}$ for $p+q \le 2$ such that

$$q(t,\mathbf{y}_0,\mathbf{l}_0,\mathbf{y},\mathbf{l}) \le \frac{K}{[\mathcal{V}^2(\mathbf{y}_0)t]^\nu} \exp\left\{ -\lambda\,\frac{|\mathbf{y}-\mathbf{y}_0|^2 + |\mathbf{l}-\mathbf{l}_0|^2}{(1+C_0^2)t} \right\}. \tag{4.14}$$

We suppose first that the coefficients of the generator $\tilde{\mathcal{L}}^{(\delta)}$ (see (4.1)) are of C^∞ class. Let us denote by $[F^{(\delta)}_{mn}(\mathbf{k})]$ the $C^\infty(\mathbb{R}^d_*)$-smooth square root of the matrix $\mathbf{D}^{(\delta)}(\mathbf{k})$. The generator of the canonical process $(X(t),K(t))_{t\ge0}$, considered over the space $(\mathcal{C},\mathcal{M},\mathfrak{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}})$, can be rewritten in the form

$$\tilde{\mathcal{L}}^{(\delta)} F(\mathbf{x},\mathbf{k}) := \sum_{m=1}^{d} X_m^2 F(\mathbf{x},\mathbf{k}) + X_0 F(\mathbf{x},\mathbf{k}), \quad F \in C_0^\infty(\mathbb{R}^{2d}_*),$$

where $X_m(\mathbf{k}) := \sum_{n=1}^{d} F^{(\delta)}_{mn}(\mathbf{k})\partial_{k_n}$, $\quad n = 1,\ldots,d$, and $X_0 := \sum_{m=1}^{d} a_m(\mathbf{k})\partial_{k_m} + \sum_{m=1}^{d} k_m\partial_{x_m}$. Here $a_m(\cdot)$, $m = 1,\ldots,d$, are certain $C^\infty(\mathbb{R}^d_*)$-functions. In fact, as in [2] see pp. 59–60, we can write

$$[X_m, X_0] = \sum_{n=1}^{d} F_{mn}^{(\delta)}(\mathbf{k})\partial_{x_n} + \sum_{n=1}^{d} c_{mn}(\mathbf{k})\partial_{k_n},$$

where we assume further that $\sum(\|F_{mn}^{(\delta)}\|_r + \|c_{mn}\|_r) \le C_r M_1^{2+r}$, $\forall\, r \ge 0$. A straightforward calculation also yields that $\inf_{(\mathbf{x},\mathbf{k})\in\mathbb{R}^{2d}} \mathcal{V}(\mathbf{x},\mathbf{k}) \ge C_1 M_1^{-6}$. In addition the process $K(\cdot)$ is a non-degenerate diffusion whose diffusivity matrix $\mathbf{D}^{(\delta)}$ satisfies the uniform positivity condition with the respective positivity constant CM_1^{-2}.

Suppose that $(\mathbf{y}(t;\mathbf{x},\mathbf{k}), \mathbf{l}(t;\mathbf{x},\mathbf{k}))$ is a diffusion defined over a certain probability space $(\Sigma, \mathcal{W}, \mathbb{W})$ whose generator is given by $\mathcal{N}^{(M_1)} = M_1^{-2I}\left(\sum_{m=1}^{d} X_m^2 + X_0\right)$. From estimate (4.14) we conclude that for all $(\mathbf{x},\mathbf{k}) \in \mathcal{A}(2M_1)$ we have

$$\mathbb{W}\left[\left|\mathbf{y}\left(\frac{M_1^{2I}}{p};\mathbf{x},\mathbf{k}\right) - \mathbf{y}\right| \le \frac{2M_1^3}{q}\right] \le \frac{2^d K\lambda^{d/2}}{(1+C_0^2)^{d/2}} \times \frac{M_1^{d(3-I)-2\nu(I+6)} p^{\nu+d/2}}{q^d}.$$

Since the laws of $(\mathbf{y}(M_1^{2I}t;\mathbf{x},\mathbf{k}), \mathbf{l}(M_1^{2I}t;\mathbf{x},\mathbf{k}))$, $t \ge 0$, and $(X(t), K(t))$, $t \ge 0$, coincide we have

$$\mathfrak{Q}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\left|X\left(\frac{1}{p}\right) - \mathbf{y}\right| \le \frac{2M_1^3}{q}\right] = \mathbb{W}\left[\left|\mathbf{y}\left(\frac{M_1^{2I}}{p};\mathbf{x},\mathbf{k}\right) - \mathbf{y}\right| \le \frac{2M_1^3}{q}\right]$$

and we conclude that

$$\mathfrak{Q}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\left|X\left(\frac{1}{p}\right) - \mathbf{y}\right| \le \frac{2M_1^3}{q}\right] \le \frac{2^d K\lambda^{d/2}}{(1+C_0^2)^{d/2}} \times \frac{M_1^{d(3-I)-2\nu(I+6)} p^{\nu+d/2}}{q^d}. \quad (4.15)$$

Using (4.11), (4.12) and (4.15) we obtain that

$$R_{\mathbf{x},\mathbf{k}}^{(\delta)}[A(\delta)] \le C\frac{M_1^{d(3-I)-2\nu(I+6)} p^{\nu+d/2}}{q^{d-2}} \le C\delta^\gamma, \quad (4.16)$$

for some $C, \gamma > 0$, upon a suitable choice of $\epsilon_0, \epsilon_2, \epsilon_3$. Since the constant $C > 0$ appearing above depends only on the C^2 norms of the coefficients of the diffusion a simple approximation allows us to obtain estimate (4.16) under the assumption that D_{mn} and E_m, $m, n = 1, \ldots, d$, are of C^2 class of regularity.

4.3.2. An estimate of $R_{\mathbf{x},\mathbf{k}}^{(\delta)}[B(\delta)]$. To estimate this term we use essentially the same argument as in Section 3.8.2 of [3]. We conclude then that there exist constants $C, \gamma > 0$, for which

$$R_{\mathbf{x},\mathbf{k}}^{(\delta)}[B(\delta)] \le C\delta^\gamma, \quad \text{for all} \quad (\mathbf{x},\mathbf{k}) \in \mathcal{A}(M_1). \quad (4.17)$$

4.3.3. An estimate of $R^{(\delta)}_{\mathbf{x},\mathbf{k}}[C(\delta)]$. Let K be a compact subset of $\mathbb{R}^{2d}_*$. We show that there exist constants $C,\gamma>0$ such that

$$\sup_{(\mathbf{x},\mathbf{k})\in K} R^{(\delta)}_{\mathbf{x},\mathbf{k}}[T_\delta<T]\le C\delta^\gamma. \tag{4.18}$$

Let $H^{(0)}_\delta(t,\mathbf{k}):=\mathfrak{Q}_{\mathbf{k}}[T_\delta<t]$, $t\ge 0$. Suppose that $H^{(i)}_\delta(t,\mathbf{k})$, $i=1,2$, are the solutions of the following first initial-boundary value problems:

$$\begin{cases} \partial_t H^{(i)}_\delta(t,\mathbf{k})=\tilde{\mathcal{L}}H^{(i)}_\delta(t,\mathbf{k}), & t>0,\,\mathbf{k}\in A(2M_1),\\ H^{(i)}_\delta(0,\mathbf{k})=\phi^{(i)}(\mathbf{k}), & \\ H^{(i)}_\delta(t,\mathbf{k})=1, & t>0,\ \text{and } |\mathbf{k}|=(2^{i-1}M_1)^{-1}, \text{or } |\mathbf{k}|=(2^{i-1}M_1)^3, \end{cases}$$

where $\phi^{(i)}:\mathbb{R}^{2d}\to[0,1]$ is a C^∞ function, equal to 0 on $A(2^{i-2}M_1)$ and 1 outside $A(3\cdot 2^{i-3}M_1)$. Of course (from the maximum principle)

$$H^{(1)}_\delta(t,\mathbf{k})\ge H^{(0)}_\delta(t,\mathbf{k})\ge H^{(2)}_\delta(t,\mathbf{k}),\quad \forall\, t\in[0,T],\,\mathbf{k}\in A(M_1).$$

Suppose also that $\epsilon\in(0,1)$ and $\delta':=\delta^\epsilon$. We have then

$$H^{(1)}_\delta(t,\mathbf{k})\le H^{(0)}_{\delta'}(t,\mathbf{k}),\quad \forall\, t\in[0,T],\,\mathbf{k}\in A(M_1'), \tag{4.19}$$

where $M_1'=[\delta'^{-\epsilon_0}]$.

A crucial estimate of the $\|\cdot\|_{1,3}$ norm of the function $H_\delta(t,\mathbf{k})$ is provided by the following.

Lemma 4.4. *There exist constants C and an integer J such that*

$$\|H^{(i)}_\delta\|^{[0,T]}_{1,3}\le CM_1^J,\quad \forall\,\delta\in(0,1],\,i=1,2. \tag{4.20}$$

Assuming this result, its proof is given in the Appendix, let us show how to establish (4.18). First, note that we can extend function H_δ to the entire $[0,+\infty)\times\mathbb{R}^{2d}$ in such a way that (4.20) holds. We apply part i) of Proposition 4.2 to the function $G_\delta(t,\mathbf{k}):=H_\delta(T-t,\mathbf{k})$. Let $p_1:=[\delta^{-\gamma_0'}]+1$, where $\gamma_0'\in(1/2,1)$. Using (4.4) we conclude that

$$\mathcal{M}^-_n:=\widehat{N}_{n/p_1}(G_\delta)-C'\delta^{\gamma_1}\frac{n}{p_1}\,\|G_\delta\|^{[0,T]}_{1,3},\quad n=0,\dots,[Tp_1],$$

is a $\left(\mathcal{M}_{n/p_1}\right)_{n\ge 0}$ super-martingale under the $\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}$-probability.

Let $\hat{T}_\delta:=$ $\{p_1^{-1}([T_\delta p_1]+1)\}\wedge T$. One can easily verify that $\hat{T}_\delta$ is an $\left(\mathcal{M}_{n/p_1}\right)_{n\ge 0}$ stopping time. As a consequence of the optional stopping theorem we obtain that $\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\mathcal{M}^+_{p_1\hat{T}_\delta}\le 0$, which leads to the estimate

$$\begin{aligned}\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}H^{(1)}_\delta(T-\hat{T}_\delta,K(\hat{T}_\delta))\le H^{(1)}_\delta(T,\mathbf{k})+C'\delta^{\gamma_1}\|H^{(1)}_\delta\|^{[0,T]}_{1,3}\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\hat{T}_\delta\\ +\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[\int_0^{\hat{T}_\delta}\widehat{\mathcal{L}}_sH^{(1)}_\delta(T-s,K(s))ds\right].\end{aligned} \tag{4.21}$$

Using (4.19) we can estimate the first two terms appearing on the right-hand side of (4.21) by

$$H_{\delta'}^{(0)}(T,\mathbf{k}) + C'\delta^{\gamma_1}\|H_\delta^{(1)}\|_{1,3}^{[0,T]}T. \tag{4.22}$$

Using the estimates (4.20) and (2.6) we conclude that, upon the choice of a sufficiently small $\epsilon_0 > 0$, that the expression (4.22) can be estimated by $C\delta^\gamma$ for some $C, \gamma > 0$, provided that $(\mathbf{x},\mathbf{k}) \in K$.

It remains yet to bound the third term on the right hand side of (4.21). The term in question equals

$$\begin{aligned}&\tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H_\delta^{(1)}(T-s,K(s))ds,\ T_\delta < T\right] \\ &+ \tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H_\delta^{(1)}(T-s,K(s))ds,\ T_\delta \ge T\right].\end{aligned} \tag{4.23}$$

Denote the first and the second terms in (4.23) by I and II correspondingly. We can write then

$$\begin{aligned}I =&\tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H_\delta^{(1)}(T-s,K(s))ds,\ T_\delta < T,\ T_\delta \le S(\delta)\wedge U(\delta)\right] \\ &+ \tilde{E}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H_\delta^{(1)}(T-s,K(s))ds,\ S(\delta)\wedge U(\delta) < T_\delta < T\right].\end{aligned} \tag{4.24}$$

The integrand appearing in the first term on the right-hand side of (4.23) equals zero for all $s \in [0,T_\delta]$, because then $\tau_\delta = T_\delta$ and $\widehat{\mathcal{L}}_s H_\delta^{(1)}(T-s,K(s)) = \mathcal{L}H_\delta^{(1)}(T-s,K(s)) = 0$, $s \in [0,T_\delta]$. Note also that the coefficients of the operator $\widehat{\mathcal{L}}_s$ are bounded by CM_1^2. In consequence, the term in question can be estimated by

$$CM_1^3Np_1^{-1}\|H_\delta^{(1)}\|_{1,3} \le C\delta^{1/2},$$

provided that $\varepsilon_0, \varepsilon_1$ are chosen sufficiently small. The factor N in the above expression can be explained by the presence of $\partial_K\Theta$ in the definition of $\widehat{\mathcal{L}}_s$ and estimate (3.7). The second term on the right-hand side of (4.24) can be estimated by

$$CM_1^3Np_1^{-1}\|H_\delta^{(1)}\|_{1,3}\tilde{Q}_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\,S(\delta)\wedge U(\delta) \le \tau_\delta < T\,\right].$$

Since the event $[\,S(\delta)\wedge U(\delta) \le \tau_\delta < T\,]$ is $\mathcal{M}^{\tau_\delta}$-measurable the $\tilde{Q}_{\mathbf{x},\mathbf{k}}^{(\delta)}$-probability of the event equals in fact

$$R_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[\,S(\delta)\wedge U(\delta) \le \tau_\delta < T\,\right] \le R_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[A(\delta)\right] + R_{\mathbf{x},\mathbf{k}}^{(\delta)}\left[B(\delta)\right] \le C\delta^\gamma. \tag{4.25}$$

The last inequality following by virtue of (4.16) and (4.17). Summarizing we have shown that the term I can be estimated $C\delta^\gamma$. As for the term II of (4.23) we can write

$$II = \tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H^{(1)}_\delta(T-s,K(s))ds,\, T_\delta \ge T \ge S(\delta)\wedge U(\delta)\right] \tag{4.26}$$
$$+ \tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H^{(1)}_\delta(T-s,K(s))ds,\, T_\delta \ge T,\, S(\delta)\wedge U(\delta) > T\right].$$

Since on the event under the expectation appearing in the second term on the right-hand side of (4.26) we have $\hat{T}_\delta = T \le \tau_\delta$, we conclude that it vanishes. The first term can be estimated by

$$CM_1^3 N\|H^{(1)}_\delta\|_{1,3} T\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[T \ge \tau_\delta \ge S(\delta)\wedge U(\delta)\right] \overset{(4.25)}{\le} CM_1^3 N\|H^{(1)}_\delta\|_{1,3} T\delta^\gamma.$$

Summarizing, we have shown that there exist $C,\gamma > 0$ such that

$$\left|\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[\int_0^{\hat{T}_\delta} \widehat{\mathcal{L}}_s H^{(1)}_\delta(T-s,K(s))ds\right]\right| \le C\delta^\gamma.$$

Using (4.20) and the definition of the dynamics corresponding to the law $\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}$, cf. (3.8), we obtain that

$$\tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}} H^{(1)}_\delta(T-\hat{T}_\delta, K(\hat{T}_\delta)) \ge \tilde{E}^{(\delta)}_{\mathbf{x},\mathbf{k}} H^{(1)}_\delta(T-T_\delta, K(T_\delta)) - \frac{C}{p_1\delta^{1/2}}\, M_1^J. \tag{4.27}$$

The first term on the right-hand side of (4.27) equals $\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[T_\delta < T\right]$. Choosing γ_0' appropriately we can obtain the estimate

$$\tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[T_\delta < T\right] \le C\delta^\gamma.$$

But, since $C(\delta)$ is $\mathcal{M}^{\tau_\delta}$-measurable we obtain

$$R^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[C(\delta)\right] = \tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[C(\delta)\right] \le \tilde{Q}^{(\delta)}_{\mathbf{x},\mathbf{k}}\left[T_\delta < T\right] \le C\delta^\gamma. \tag{4.28}$$

Appendix A. The proof of Lemma 4.4

In what follows we shall denote by $\partial A(M_1) := [\,|\mathbf{k}| = M_1^{-1}\,] \cup [\,|\mathbf{k}| = M_1^3\,]$ and by $S(M_1)$ *the parabolic boundary* of the region $D(M_1) := (0,T)\times A(M_1)$, i.e., $S(M_1)$ is a union of $S_1(M_1) := \{t=0\}\times A(M_1)$ and $S_2(M_1) := [0,T]\times \partial A(M_1)$. We shall also denote by $C, C_1, \ldots, k, k_1 > 0$ various constants that do not depend on M_1.

Let ψ be an arbitrary C^∞ class function, compactly supported in $D(M_1)$ and such that $\psi_0(\mathbf{k}) := \psi(0,\mathbf{k})$ satisfies $\|\psi_0\|_{H^m(\mathbb{R}^d)} \le C_m M_1^{k_m}$ for some constants $C_m, k_m > 0$. Also the coefficients of the operator $\mathcal{L}$, see (2.5), are of C^{2n_*} class and their respective C^m norms are bounded by $C_m M_1^{k_m}$ for appropriate constants $C_m, k_m > 0$, $m = 0, \ldots, 2n_*$.

Let u be the solution of the first initial-boundary value problem

$$\begin{cases} \partial_t u(t,\mathbf{k}) - \mathcal{L}u(t,\mathbf{k}) = 0, & (t,\mathbf{k}) \in D(M_1), \\ u(t,\mathbf{k}) = \psi(t,\mathbf{k}), & (t,\mathbf{k}) \in S(M_1). \end{cases} \tag{A.1}$$

Thanks to the maximum principle we conclude easily that

$$\max_{t\in[0,T]} \|u(t,\cdot)\|_{L^2(A(M_1))} \le C M_1^k$$

for some constants $C, k > 0$ independent of M_1.

Using the argument from pp. 354-356 of [1] we deduce that

$$\max_{t\in[0,T]} \|u(t,\cdot)\|^2_{L^2(A(M_1))} + \int_0^T \|\nabla u(t,\cdot)\|^2_{L^2(A(M_1))} + \int_0^T \|u(t,\cdot)\|^2_{H^{-1}(A(M_1))} \le C M_1^k$$

for some $C, k > 0$. Let $u_0(t,\mathbf{k}) := u(t,\mathbf{k})$ and $u_i(t,\mathbf{k}) := \partial_t u_{i-1}(t,\mathbf{k})$, $i = 1, \ldots, 2n_*$. These functions satisfy the first boundary value problems

$$\begin{cases} \partial_t u_i(t,\mathbf{k}) - \mathcal{L}u_i(t,\mathbf{k}) = 0, & (t,\mathbf{k}) \in D(M_1), \\ u_i(0,\mathbf{k}) = \mathcal{L}^{i-1}\psi_0(\mathbf{k}), & \mathbf{k} \in A(M_1), \\ u_i(t,\mathbf{k}) = 0, & (t,\mathbf{k}) \in S_2(M_1). \end{cases}$$

Hence,

$$\begin{aligned} &\max_{t\in[0,T]} \|u_i(t,\cdot)\|^2_{L^2(A(M_1))} + \int_0^T \|\nabla u_i(t,\cdot)\|^2_{L^2(A(M_1))} + \int_0^T \|u_i(t,\cdot)\|^2_{H^{-1}(A(M_1))} \\ &\le C_i M_1^{k_i} \end{aligned} \tag{A.2}$$

for some $C_i, k_i > 0$, $i = 0, \ldots, 2n_*$.

From the proof of the L^2 boundary a-priori estimates, see [1] pp. 308–326, we conclude that for appropriate constants $C, k > 0$

$$\begin{aligned} \|u(t,\cdot)\|_{H^2(A_1(M))} &\le C M_1^k \left(\|u(t,\cdot)\|_{L^2(A_1(M_1))} + \|\mathcal{L}u(t,\cdot)\|_{L^2(A_1(M_1))} \right) \\ &= C M_1^k \left(\|u(t,\cdot)\|_{L^2(A_1(M_1))} + \|u_1(t,\cdot)\|_{L^2(A_1(M_1))} \right) \le C' M_1^{k'}. \end{aligned}$$

The last inequality follows from (A.2) applied for $i = 1$. Likewise,

$$\begin{aligned} \|u_1(t,\cdot)\|_{H^2(A_1(M))} &\le C M_1^k \left(\|u_1(t,\cdot)\|_{L^2(A_1(M_1))} + \|\mathcal{L}u_1(t,\cdot)\|_{L^2(A_1(M_1))} \right) \\ &= C M_1^k \left(\|u_1(t,\cdot)\|_{L^2(A_1(M_1))} + \|u_2(t,\cdot)\|_{L^2(A_1(M_1))} \right) \stackrel{(A.2)}{\le} C' M_1^{k'}. \end{aligned} \tag{A.3}$$

From the a-priori estimate concerning higher Sobolev norms, see the proof of Theorem 6.3.5 p. 323 of [1] to deduce the bound on the respective constant, we obtain

$$\begin{aligned} \|u(t,\cdot)\|_{H^4(A_1(M_1))} &\le C M_1^k \left(\|\mathcal{L}u(t,\cdot)\|_{H^2(A_1(M_1))} + \|u(t,\cdot)\|_{L^2(A_1(M_1))} \right) \\ &= C M_1^k \left(\|u_1(t,\cdot)\|_{H^2(A_1(M_1))} + \|u(t,\cdot)\|_{L^2(A_1(M_1))} \right) \stackrel{(A.3)}{\le} C' M_1^{k'}. \end{aligned}$$

We can extend this argument by induction to conclude that

$$\|u(t,\cdot)\|_{H^{2n_*}(A_1(M_1))} \le CM_1^k$$

for some constants $C, k > 0$. Using Sobolev's embedding theorem, see Theorem 5.7.6 p. 270 of [1], we conclude that $\|u\|_{0,3} \le CM_1^k$ for some $C, k > 0$ and since this function satisfies (A.1) we obtain that $\|u\|_{1,3} \le CM_1^k$ for some $C, k > 0$.

Acknowledgement

The author wishes to express his gratitude to Professor Lenya Ryzhik for numerous enlightening discussions on the subject of this article.

References

[1] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Graduate Studies in Mathematics, 19, 1988.

[2] H. Kesten and G. C. Papanicolaou, *A limit theorem for stochastic acceleration*, Comm. Math. Phys., **78** (1980), 19–63.

[3] T. Komorowski and L. Ryzhik, *Diffusion in a weakly random Hamiltonian flow*, Comm. in Math. Physics, **263 (2)** (2006), 277–323.

[4] T. Komorowski and L. Ryzhik, *The stochastic acceleration problem in two dimensions*, Israel Journ. Math., **155** (2006), 157–203.

[5] S. Kusuoka and D. Stroock, Applications of the Malliavin calculus, Part II, J. Fac. Sci. Univ. Tokyo, Sect. IA, Math., **32** (1985), 1–76.

[6] D. Stroock, *An Introduction to the Analysis of Paths on a Riemannian Manifold*, Math. Surv. and Monographs, 74, 2000.

[7] D. Stroock and S. R. S. Varadhan *Multidimensional Diffusion Processes*, Berlin, Heidelberg, New York: Springer-Verlag, 1979.

[8] D. Williams, *To begin at the beginning*, in: Stochastic Integrals, Lecture Notes in Mathematics, Springer-Verlag, **851** (1981), 1–55.

Tomasz Komorowski
Instytut Matematyczny, PAN
Śniadeckich 8
00-950 Warsaw
Poland
and
Institute of Mathematics, UMCS
pl. Marii Curie Skłodowskiej 1
20-031 Lublin
Poland
e-mail: `komorow@hektor.umcs.lublin.pl`

Progress in Probability, Vol. 59, 197–201

Long-Time Behaviour for the Brownian Heat Kernel on a Compact Riemannian Manifold and Bismut's Integration-by-Parts Formula

Rémi Léandre

Abstract. We give a probabilistic proof of the classical long-time behaviour of the heat kernel on a compact manifold by using Bismut's integration-by-parts formula.

Mathematics Subject Classification (2000). 60H07.

Keywords. Heat kernels.

1. Introduction

Let M be a compact connected Riemannian manifold and let Δ be the Laplace-Beltrami operator on M. The heat semi-group associated to Δ has a heat kernel $p_t(x, y)$ associated to the Riemannian probability measure $d\mu$ on M. Since $d\mu$ is the unique invariant probability measure associated to the heat semi-group, it is well known that for all continuous functions f on M,

$$\int_M p_t(x, y)f(y)d\mu \to \int_M f(y)d\mu \tag{1.1}$$

when $t \to \infty$. This classical result is established by analytical techniques, that can be improved in order to get the following theorem:

Main Theorem. *When $t \to \infty$,*

$$\sup_{x,y} |p_t(x, y) - 1| \to 0.$$

Our goal is to provide a new proof of the main theorem by using the Malliavin Calculus.

The Malliavin Calculus until now was only able to give short-time asymptotics of heat kernels. We refer to the surveys of Léandre [10, 13, 14], Kusuoka [9] and Watanabe [17] for the details. For asymptotics of semi-groups by using probabilities, we refer to the book of Kolokoltsov [8]. The main novelty of this paper is that we can get long-time asymptotics of heat kernels by Malliavin Calculus. For that purpose, we use the intrinsic approach to the Malliavin Calculus of Bismut [2], and the intrinsic integration-by-parts on the Brownian motion on the Riemannian manifold of Bismut [2]. We refer to the works of Léandre [11, 12], Driver [3] and Hsu [5] for developments.

The main trick is to choose the auxiliary function $s \to \frac{\exp[Cs]-1}{\exp[Ct]-1}$ for $s \in [0,t]$, $C > 0$, instead of the function $s \to \frac{s}{t}$ that was chosen by Bismut in [2].

2. Proof of the main theorem

Let us consider the Riemannian manifold endowed with the Levi-Civita connection. We consider a Brownian motion issued from 0 in $T_x(M)$ and the Eells-Elworthy-Malliavin equation issued from x:

$$d\gamma_t(x) = \tau_t dB_t$$

where τ_t is the parallel transport on the solution path. Let $e_t : \gamma(.) \to \gamma_t(x)$ be the evaluation map. τ_t appears as a section of $(e_0^* T(M))^* \otimes e_t^* T(M)$ where $T(M)$ is the tangent bundle of M. $e_t^* T(M)$, considered as a bundle on the path space inherits a connection from the Levi-Civita connection on the manifold. Moreover, the parallel transport is an isometry. Let $H(s)$ be a finite energy, deterministic path, issued from 0 in $T_x(M)$, and defined for $0 \le s \le t$. We consider Bismut's tangent vector field $X_s(H(.)) = \tau_s H_s$. Since H is supposed to be deterministic [2, 3, 5, 11], we have

$$\begin{aligned} &E[< df(\gamma_t(x)), X_t(H(.)) >] \\ &= E[f(\gamma_t(x))(\int_0^t < d/dsH(s), \delta B_s > +1/2 \int_0^t < S_{X_s(H(.))}, \delta\gamma_s(x) >)] \end{aligned}$$

where S is the Ricci tensor and δ denotes the Itô integral.

Let us recall that the law of $\gamma_t(x)$ has a smooth density with respect to $d\mu$ which coincides with $p_{t/2}(x,y)$. In the sequel we will forget the factor 1/2 in the previous formula and we will write γ_t instead of $\gamma_t(x)$. We have:

Lemma 2.1. *When* $t \to \infty$, $\operatorname{grad} \log p_t(x,y)$ *remains bounded in* x *and* y.

Proof. According to [2], let us write $q_t(z) = \operatorname{grad}_x \log p_t(z,y)$. Conditionnally to the fact that $\gamma_t = y$, $dB_s - \tau_s^{-1} q_{t-s}(\gamma_s) ds$ is the differential of a Brownian motion (see [2, 2.87]).

Let

$$H_t(s) = \frac{\exp[Cs]-1}{\exp[Ct]-1}$$

for $s \le t$ where $C > 0$. $H_t(0) = 0$ and $H_t(t) = 1$, and $H_t(s)$ is smaller than 1. Let us consider a vector field X on M and let the vector $X_.(X)$ on the path space between $[0, t]$ be given by

$$X_s(X) = \tau_s H_t(s) \tau_t^{-1} X(\gamma_t)$$

Let e_i be the canonical orthonormal basis of $T_x(M)$.

We have:

$$E[< df(\gamma_t), X(\gamma_t) >] = E[f(\gamma_t) \int_0^t < d/ds H_t(s) \tau_t^{-1} X(\gamma_t), \delta B_s >]$$
$$+ E[1/2 f(\gamma_t) \int_0^t < S_{X_s(X)}, \delta\gamma_s >] - \sum_i E[f(\gamma_t) < d(\tau_t^{-1} X(\gamma_t)), X(H_t(.)e_i) >]$$

But by the Arafeva-Bismut formula [1, 2, 12],

$$\nabla_{X_.(H_t(.)e_i)} \tau_t^{-1} = - \int_0^t \tau_s^{-1} R(d\gamma_s, X_s(H_t(.)e_i)) \tau_s \tau_t^{-1}$$

where we consider the curvature tensor R of M and $d\gamma_s$ denotes the Stratonovitch integral. Moreover $\nabla_{X_t(H_t(.)e_i))} X(\gamma_t)$ is bounded.

Let $E_t^{x,y}$ be the expectation when we condition by $\gamma_t = y$. We deduce that:

$$< \mathrm{grad}_y \log p_t(x, y), X(y) >= E_t^{x,y} [\int_0^t < d/ds H_t(s) \tau_t^{-1} X(y), \delta B_s >$$
$$+1/2 \int_0^t < S_{X_s(X)}, \delta\gamma_s > - \sum_i < d(\tau_t^{-1} X(\gamma_t)), X_.(H_t(.)e_i) >] - \mathrm{div}\, X_.$$

We remark that $p_t(x, y) = p_t(y, x)$ and we get an analoguous expression for $\mathrm{grad}_x \log p_t(x, y)$.

Let us consider

$$\alpha_t = \sup_{x,y} (|\mathrm{grad}_x \log p_t(x, y)| + |\mathrm{grad}_y \log p_t(x, y)|).$$

We remark that $\sup_{t>1} \int_0^t |d/ds H_t(s)|^2 ds < \infty$, $\sup_{t>1} \int_0^t H_t(s)^2 ds < \infty$ and that $\sup_{t>1} \int_0^t H_t(s) ds < \infty$ (this fact is false if we replace, as Bismut did, $H_t(s)$ by s/t). We deduce from [2, 2.87] that, for $t > 1$,

$$\begin{aligned} \alpha_t &\le C + C \int_0^{t-1} \alpha_{t-u} d/du H_t(u) du + C \int_0^{t-1} \alpha_{t-u} H_t(u) du \\ &\le C + C \int_0^{t-1} \alpha_{t-u} \exp[C(u-t)] du + \exp[-Ct] \int_0^{t-1} \alpha_{t-u} du. \end{aligned}$$

By remarking that $\exp[-Ct] \le \exp[-Cu]$ if $u \le t$ and after putting $t - u = v$, we deduce that

$$\alpha_t \le C + C \int_1^t \exp[-Cu] \alpha_u du.$$

We deduce the result by using Gronwall's lemma since $\int_1^\infty \exp[-Cu] du < \infty$. $\square$

Proof of the main theorem. $p_t(x,y)d\mu$ is a probability measure on compact M. Therefore, there exists y_t such that $p_t(x,y_t) < 1$. Let y_s be a path joining at time 1 y_t to y: since M is supposed to be compact, we can suppose that its speed is bounded. We have

$$|d/dsp_t(x,y_s)| \leq C\alpha_t p_t(x,y_s) \tag{2.1}$$

Since α_t is bounded when $t \to \infty$, we deduce by Gronwall's lemma from (2.1) that $\sup_{t>1,x,y} p_t(x,y) < \infty$ and therefore that $\mathrm{grad}_x p_t(x,y)$ and $\mathrm{grad}_y p_t(x,y)$ are bounded in x,y when $t \to \infty$. This shows that the family of functions $(x,y) \to p_t(x,y)$ is equicontinuous when $t \to \infty$. The result follows by Ascoli's theorem and (1.1). □

References

[1] I. Y. Araf'eva, *Non-Abelian Stokes formula*, Teo. Mat. Fiz., **43** (1980), 353–356.

[2] J. M. Bismut, *Large Deviations and Malliavin Calculus*, Progress in Math., 45, Birkhäuser, 1984.

[3] B. Driver, *Cameron-Martin type quasi-invariance formula for Brownian motion on compact manifolds*, J. Funct. Anal., **110** (1992), 272–376.

[4] K. D. Elworthy, *Geometric aspects of stochastic analysis*, in: Developments of Mathematics 1950-2000, J. P. Pier, Editor, Birkhäuser, (2000), 437–484.

[5] E. P. Hsu, *Quasi-invariance of the Wiener measure on the path space over a compact manifold*, J. Funct. Anal., **134** (1996), 417–450.

[6] N. Ikeda, I. Shigekawa and S. Shigekawa, *The Malliavin Calculus and long-time asymptotics of certain Wiener integrals*, Proc. Center. Math. Ana. Australian. Ubiv., **9** (1985), 46–113.

[7] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, 1981.

[8] V. N. Kolokoltsov, *Semiclassical Analysis for Diffusions and Stochastic Processes*, Lect. Notes Math., Springer, **1724** (2000).

[9] S. Kusuoka, *More recent theory of Malliavin Calculus*, Sugaku, **5** (1992), 155–173.

[10] R. Léandre, *Applications quantitatives et qualitatives du Calcul de Malliavin*, Paris E.N.S. 1987, M. Métivier et S. Watanabe, Editors, Lect. Notes Math., Springer, **1322** (1988), 109–123. English translation in Geometry of Random Motion (Cornell, 1987), R. Durrett et M. Pinsky, Editors, Contem. Math., **73** (1988), 173–197.

[11] R. Léandre, *Integration by parts and rotationally invariant Sobolev calculus on free loop spaces*, in: XXVII Winter School of Theoretical Physics (Carpacz, 1991), A. Borowiec and R. Gielerak, Editors, J. Geom. Phys., **11** (1993), 517–528.

[12] R. Léandre, *Invariant Sobolev calculus on free loop space*, Acta. Appl. Math., **46** (1997), 267–350.

[13] R. Léandre, *Malliavin Calculus for a general manifold*, in: Séminaire X-EDP 2002/2003, J. M. Bony, J. Y. Chemin, and J. Sjoestrand, Editors, Publ. Ecole Polytechnique, 2003.

[14] R. Léandre, *Stochastic mollifier and Nash inequality*, in: Recent Developments in Stochastic Analysis and Related Topics (Beijing, 2002), S. Albeverio, Z. M. Ma, and M. Roeckner, Editors, World Scientific, (2004), 246–257.

[15] R. Léandre, *A geometrical hypoelliptic diffusion*, Differential Geometry and its Applications, (Praha, 2004), D. Krupka, Editor, (2004), 517–521.

[16] P. Malliavin, *Sur certaines intégrales stochastiques oscillantes*, C.R.A.S., **295** (1982), 295–300.

[17] S. Watanabe, *Stochastic analysis and its application*, Sugaku, **5** (1992), 51–71.

Rémi Léandre
Institut de Mathématiques de Bourgogne
Université de Bourgogne
21078 Dijon
France
e-mail: `remi.leandre@u-bourgogne.fr`

Progress in Probability, Vol. 59, 203–226

Probabilistic Deformation of Contact Geometry, Diffusion Processes and Their Quadratures

Paul Lescot and Jean-Claude Zambrini

Abstract. Classical contact geometry is an odd-dimensional analogue of symplectic geometry. We show that a natural probabilistic deformation of contact geometry, compatible with the very irregular trajectories of diffusion processes, allows one to construct the stochastic version of a number of basic geometrical tools, like, for example, Liouville measure. Moreover, it provides an unified framework to understand the origin of explicit relations (cf. "quadrature") between diffusion processes, useful in many fields. Various applications are given, including one in stochastic finance.

Mathematics Subject Classification (2000). Primary 60J60; Secondary 53D10.

Keywords. Diffusion processes, contact geometry.

1. Introduction

In [12] (afterwards referred to as [Iso]) we have introduced a concept of "stochastic quadrature" for one-dimensional processes that are solutions of stochastic differential equations (SDE):

$$dz(t) = \sqrt{\hbar}\, dw(t) + \widetilde{B}(z(t), t)dt \tag{1.1}$$

with respect to the increasing filtration $\mathcal{P}_t$ of the Brownian process $w(t)$. In Eq. (1.1), $\hbar$ is a positive constant and the drift $\widetilde{B}$ is of the special form

$$\widetilde{B}(q,t) = \hbar \frac{\partial}{\partial q} ln\widetilde{\eta}(q,t) \tag{1.2}$$

The present paper resulted from a visit of the first author at the GFMUL (Lisbon) within the project POCTI/MAT/34924/2000. Both authors are grateful to the Ascona's organizers for the opportunity to present their results.

for $\widetilde{\eta}$ a positive solution of

$$\hbar \frac{\partial \widetilde{\eta}}{\partial t} = -\frac{\hbar^2}{2} \frac{\partial^2 \widetilde{\eta}}{\partial q^2} \equiv H_0 \widetilde{\eta}, \quad t \in \Sigma, \tag{1.3}$$

with Σ a compact or semi-infinite interval of $\mathbb{R}$.

The point of our construction of such stochastic quadratures was to provide an unified framework explaining the origin of explicit relations between some families of diffusion processes, often very useful in computations but hard to guess a priori.

A typical example of stochastic quadratures is Doob's familiar relation between the Brownian and the Ornstein-Uhlenbeck process $z(t)$ starting from x:

$$z(t) = e^{-\beta t} \left[x + w \left(\frac{1}{2\beta} \left(e^{2\beta t} - 1 \right) \right) \right], \tag{1.4}$$

β being the multiplicative constant of the linear drift of $z(t)$.

Another motivation of [Iso] was more geometrical and aimed to answer a question apparently unrelated with our first point: how could we construct a natural stochastic symplectic geometry? Indeed, elementary as it sounds, the basic problem of the construction of a stochastic analogue of the classical Liouville measure, for example, does not seem to have found, as yet, a natural solution. This could be at the origin of conceptual difficulties when trying to construct probability measures for infinite-dimensional symplectic dynamical systems.

Our understanding of the above adjective "natural" stems from the origin of the Brownian $w(t)$ itself: the construction in question should use nothing more than (the geometrical content of) Eq. (1.3) itself.

Our purpose here is to show that the method introduced in [Iso] is considerably more general than we thought initially.

On the one hand, we can add perturbation potentials a and V to H_0 in the parabolic equation (1.3):

$$\begin{aligned} \hbar \frac{\partial \widetilde{\eta}}{\partial t} &= -\frac{\hbar^2}{2} \frac{\partial^2 \widetilde{\eta}}{\partial q^2} + a(q) \frac{\partial \widetilde{\eta}}{\partial q} + V(q,t) \widetilde{\eta} \\ &\equiv H \widetilde{\eta}, \end{aligned} \tag{1.5}$$

and our method will provide for free a maximal class of such potentials allowing to reduce the analysis of stochastic quadratures to the one of Eq. (1.3). However, let us stress at once that the method advocated is in no way restricted to any special class of Hamiltonians.

It will appear that the underlying geometry is, in fact, more general than symplectic. It can be regarded as an $\hbar$-deformation of the classical contact geometry of elementary dynamical systems whose Hamiltonian is the "classical limit" of H in Eq. (1.5).

Although this deformation is quite similar to the one expressing the transition from classical to quantum dynamics where $\hbar$ stands, of course, for Planck's constant (with the crucial difference that Schrödinger's equation is replaced here by its "Euclidean" or "imaginary time" parabolic counterpart (1.5), better suited to a probabilistic analysis) our results are relevant to any field where one needs to seriously compute with diffusion processes.

The organization of this paper is the following:

§2 is a summary of the geometrical study of the classical Hamilton-Jacobi equation and its relation with classical Hamiltonian and Lagrangian dynamics. What we shall need, specifically, is the contact geometrical approach to the Hamilton-Jacobi equation (HJ), which is not very familiar, even in mathematical physics. Instead of presenting immediately its more elegant (but abstract) version in terms of ideal of differential forms, due to E. Cartan, we recall how the more traditional approaches are indeed founded on a special set of differential forms and how the geometry of HJ can be expressed in terms of the Lie dragging of those forms along its symmetries.

The Hamilton-Jacobi-Bellman equation (HJB) will be regarded here as a deformation of its classical counterpart and §3 is devoted to the analysis of its geometrical content, in Cartan's perspective. Theorem 3.1 describes the Lie dragging of the deformed ideal of forms associated with HJB, in terms of the coefficients of the infinitesimal symmetries of this PDE. It is the general result on the geometrical content of HJB valid for any (regular) potential V.

Although the statement of Theorem 3.1 seems, a priori, to have no relations whatsoever with stochastic analysis, one of the symmetries involved, as well as the deformed basic invariance identity for the action functional, strongly suggest what to do to find one, namely, in geometrical terms, a section of the base manifold of independent variables of HJB within the jet space for this equation.

Theorem 3.3 describes the main results of the probabilistic interpretation on the base integral submanifold of the jet space, in the perspective of our stochastic quadratures of diffusions.

§4 and §5 contain a list of explicit examples of quadratures resulting from the general results of §3. Although such quadratures can be computed, in principle, each time the determining equations (3.31) are solved, for any given V, a special (but large!) class of examples is directly accessible using exclusively the geometry of the free HJB equation (i.e., $V=0$). We mention a theorem of Rosencrans explaining how to do this reduction and use it for some explicit class of diffusions. Those readers familiar with the geometrical approach to the quantization problem will recognize, in the special status of the associated class of potentials, the probabilistic counterpart of the special status of the metaplectic representation in quantum physics.

Our last explicit example is inspired by recent results of Patie and Alili in stochastic finance [2, 13] and shows how they can be reinterpreted in our geometrical perspective.

2. Elementary classical contact geometry

Since this theory cannot be regarded as common knowledge, we shall first summarize the part of classical contact geometry relevant to the geometrical approach to first order elementary dynamical systems. We shall limit ourselves to the one-dimensional case $q \in \mathbb{R}$, only because our explicit examples will be in this class. For much more about contact geometry, cf. [3, Chapter 5; 4, Chapter 10]. Let us consider the Hamilton-Jacobi equation associated with an elementary mechanical system of given C^2 class energy (or Hamiltonian) $h(q,p,t)$, where q and p denote, respectively, the configuration (position) and momentum variables. Departing from the tradition we will choose

$$-\frac{\partial S}{\partial t} + h\left(q, -\frac{\partial S}{\partial q}, t\right) = 0 \tag{2.1}$$

for a real-valued function $S(q,t)$. Introducing the function of five variables

$$f(q,t,S,p,E) = -E + h(q,p,t), \tag{2.2}$$

we can regard Eq. (2.1) as the partial differential equation of first order

$$f\left(q, t, S, -\frac{\partial S}{\partial q}, \frac{\partial S}{\partial t}\right) = 0. \tag{2.3}$$

The idea to define partial derivatives of the dependent variable of (2.1) as new variables is called "prolongation".

(Lie's) characteristic equations for Eq. (2.3) are the following:

$$\begin{aligned}
&\dot{q} = \frac{\partial f}{\partial p},\ \dot{t} = -\frac{\partial f}{\partial E} \\
&\dot{p} = -\frac{\partial f}{\partial q} - p\frac{\partial f}{\partial S},\ \dot{E} = \frac{\partial f}{\partial t} - E\frac{\partial f}{\partial S} && (2.4) \\
&\dot{S} = f - p\frac{\partial f}{\partial p} - E\frac{\partial f}{\partial E} && (2.5)
\end{aligned}$$

where $\cdot$ denotes the derivative $\frac{d}{du}$ of the characteristics $\sigma(u) \equiv (q(u), t(u), S(u), p(u), E(u))$, $u \in \Sigma$, an interval of $\mathbb{R}$.

Given the simplicity of Eqs. (1.3) and (1.5), it will be sufficient to consider elementary Hamiltonians h of the form

$$h(q,p,t) = \frac{1}{2}p^2 + V(q,t) \tag{2.6}$$

so, using (2.2) the second equation of (2.4) implies that the parameter u can be identified with t, and therefore $\cdot$ can be regarded as $\frac{d}{dt}$.

The relation (2.2) shows that the four remaining equations split into Hamilton equations for h as in (2.6) and the (generalized) conservation of energy:

$$\dot{q} = \frac{\partial h}{\partial p} = p, \ \dot{p} = -\frac{\partial h}{\partial q} = -\frac{\partial V}{\partial q} \tag{2.7}$$

$$\dot{h} = \frac{\partial h}{\partial t} = \frac{\partial V}{\partial t}. \tag{2.8}$$

Regarding (2.5), we obtain, using again (2.7):

$$\begin{aligned} \dot{S} &= -\left(\frac{1}{2}\,\dot{q}^2 - V\right) \\ &= -L(\dot{q}, q, t), \end{aligned} \tag{2.9}$$

the right-hand side defining (minus) the Lagrangian L of such a mechanical system. This means that, by integration along the characteristics, for $v \geq t$,

$$S(q,t) = \int_{q,t}^{q_v,v} L(\dot{q}(\tau), q(\tau), \tau)d\tau + S_v(q_v) \tag{2.10}$$

where a final condition $S(q(v), v) = S_v(q_v)$ has been introduced. (A final condition and not an initial one as usual because we started, here, from Eq. (2.1) instead of the usual Hamilton-Jacobi equation. Cf. [4] for a probabilistic interpretation.)

Our starting space $\mathbb{R}^5$ of the variables (q, t, S, p, E) is called 1-jet and is often denoted by J^1. According to (2.3), our Hamilton-Jacobi equation corresponds to the hypersurface $\varepsilon \equiv \{f = 0\}$ in J^1.

In symplectic geometry, the basic geometrical object is the Liouville 2-form $\Omega = dp \wedge dq = d\omega$ where $\omega = pdq$, the Poincaré form. In the ("extended") cases where, like here, the conjugate variables $(t, -E)$ are also needed, one starts, instead, from the Poincaré-Cartan form:

$$\omega_{pc} = pdq - Edt. \tag{2.11}$$

In contrast with the symplectic case, the basic geometric object of contact geometry is the contact 1-form:

$$\omega = \omega_{pc} + dS. \tag{2.12}$$

For a given contact Hamiltonian $f \in C^\infty(J^1)$, the associated contact vector field is defined by Eqs. (2.4)–(2.5):

$$\begin{aligned} X_f = {} & \frac{\partial f}{\partial p}\frac{\partial}{\partial q} - \frac{\partial f}{\partial E}\frac{\partial}{\partial t} + \left(f - p\frac{\partial f}{\partial p} - E\frac{\partial f}{\partial E}\right)\frac{\partial}{\partial S} \\ & - \left(\frac{\partial f}{\partial q} + p\frac{\partial f}{\partial S}\right)\frac{\partial}{\partial p} + \left(\frac{\partial f}{\partial t} - E\frac{\partial f}{\partial S}\right)\frac{\partial}{\partial E}. \end{aligned} \tag{2.13}$$

Notice that X_f can be regarded as an (extended) Hamiltonian vector field only when f does not depend on the variable S.

For f as in Eqs. (2.2) and (2.6), in particular, we obtain

$$X_f = p\frac{\partial}{\partial q} + \frac{\partial}{\partial t} + \left(-\frac{1}{2}p^2 + V\right)\frac{\partial}{\partial S} - \frac{\partial V}{\partial q}\frac{\partial}{\partial p} + \frac{\partial V}{\partial t}\frac{\partial}{\partial E}.$$

Using definitions (2.12) and (2.13) it is clear that for any $f \in C^\infty(J^1)$,

$$\omega(X_f) = f \tag{2.14}$$

so that any contact Hamiltonian f can be defined that way.

The Lie algebra of contact vector fields is in bijective correspondence with a Lie algebra on $C^\infty(J^1)$, defined through the Lagrange (or Jacobi) bracket $\{.,.\}_L$:

$$\{f, g\}_L = \omega([X_g, X_f]).$$

As suggested by (2.13), it is only when f and g do not depend on the variable S that this bracket provides a Poisson structure (since, in general, it does not satisfy the Leibniz rule).

A contact vector field X_n for Eq. (2.3) is called an infinitesimal symmetry of this equation if and only if

$$\{f, n\}_L = 0 \quad \text{on } \varepsilon.$$

Let us come back to our elementary system defined by (2.2) and (2.6). If we look for symmetry Hamiltonians n of the simple form

$$n(q, p, t, E) = X(q, t)p - T(t)\,E - \phi(q, t) \tag{2.15}$$

for undetermined real valued coefficients X, T and ϕ, one checks that they are allowed, indeed, every time those coefficients solve the "determining" equations:

$$\begin{aligned} \dot{T} = 2\frac{\partial X}{\partial q}, \quad \frac{\partial X}{\partial t} = \frac{\partial \phi}{\partial q} \\ \frac{\partial \phi}{\partial t} = -\dot{T}V - X\frac{\partial V}{\partial q} - T\frac{\partial V}{\partial t}. \end{aligned} \tag{2.16}$$

Let us summarize the main classical results on Lie dragging along symmetries, relevant to the probabilistic deformation of our elementary systems,

$$\mathcal{L}_{X_n}(\omega_{pc}) = d\phi.$$

We give here the proof of this first relation, as an illustration.

By (2.13), when $f = n$, we have

$$\begin{aligned} X_n &= X\frac{\partial}{\partial q} + T\frac{\partial}{\partial t} - \phi\frac{\partial}{\partial S} + \left(-\frac{\partial X}{\partial q}p + \frac{\partial \phi}{\partial q}\right)\frac{\partial}{\partial p} \\ &\quad + \left(\frac{\partial X}{\partial t}p - \dot{T}E - \frac{\partial \phi}{\partial t}\right)\frac{\partial}{\partial E} \\ &\equiv X_n^q\frac{\partial}{\partial q} + X_n^t\frac{\partial}{\partial t} + X_n^S\frac{\partial}{\partial S} + X_n^p\frac{\partial}{\partial p} + X_n^E\frac{\partial}{\partial E} \end{aligned}$$

where we have introduced a notation for the components of X_n. So by definition (2.11), the properties of Lie derivative and the components of X_n,

$$
\begin{aligned}
\mathcal{L}_{X_n}(pdq - Edt) &= \mathcal{L}_{X_n}(p)dq + p\mathcal{L}_{X_n}(dq) - \mathcal{L}_{X_n}(E)dt - E\mathcal{L}_{X_n}(dt) \\
&= X_n^p dq + pdX_n^q - X_n^E dt - EdT \\
&= \left(-\frac{\partial X}{\partial q}p + \frac{\partial \phi}{\partial q}\right) dq + pdX(q,t) \\
&\quad - \left(\frac{\partial X}{\partial t}p - \dot{T}E - \frac{\partial \phi}{\partial t}\right) dt - E\dot{T}dt \\
&= d\phi(q,t)
\end{aligned}
$$

Since $\Omega = d\omega_{pc}$, it also follows that, formally, $\mathcal{L}_{X_n}(\Omega) = \mathcal{L}_{X_n}(d\omega_{pc}) = d\mathcal{L}_{X_n}(\omega) = dd\phi$, so

$$\mathcal{L}_{X_n}(\Omega) = 0.$$

Regarding the Lagrangian (2.9) of our system, we find, using (2.7) and (2.16),

$$\mathcal{L}_{X_n}(L) + L\dot{T} = \frac{\partial \phi}{\partial t} + p\frac{\partial \phi}{\partial q}. \tag{2.17}$$

As it is well known, the relation (2.17) is the basic invariance identity of the variational calculus for the functional defined on the r.h.s. of (2.10). Indeed, by (2.6), (2.7) and (2.9)

$$
\begin{aligned}
\int \omega_{pc} &= \int pdq - Edt \\
= \int (p\dot{q} - h)dt &= \int Ldt.
\end{aligned}
$$

Now we are going to generalize this whole geometrical picture of classical dynamics to the case where the (smooth) characteristics are replaced by solutions of SDE.

3. Probabilistic deformation of contact geometry

Let us come back to Eq. (1.5). For the moment, we shall consider the case $a = 0$.

We shall start from the same nonlinear change of variables (whose origin dates back to E. Schrödinger, in the context of quantum mechanics, cf. [4]) as in the free case $V = 0$ treated in [Iso]:

$$S = -\hbar \ln \eta, \tag{3.1}$$

for η a positive solution of Eq. (1.5). Then S solves the following (Hamilton-Jacobi-Bellman) equation:

$$-\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial q}\right)^2 - V - \frac{\hbar}{2}\frac{\partial^2 S}{\partial q^2} = 0 \tag{3.2}$$

interpreted here as a deformation of the classical PDE (2.1) for h as in (2.6).

Let us define, in analogy with the classical case,

$$B = -\frac{\partial S}{\partial q}, \quad E = -\frac{\partial S}{\partial t}, \tag{3.3}$$

but with a few changes of signs (w.r.t (2.12)) due to the abovementioned Euclidean counterpart. In order to distinguish the 1-jet of HJB from its classical counterpart of §2, we denote by B the variable playing now the role of the classical momentum p. The geometrical content of Eq. (3.2) for the deformed 1-jet $J^1 = (q, t, S, B, E)$ is contained in the vanishing of differential forms:

$$\omega = Bdq + Edt + dS, \tag{3.4}$$

together with

$$\Omega = d\omega = dBdq + dEdt$$

where we drop the symbol $\wedge$ of exterior multiplication, for simplicity, and the two-form β defining Hamilton-Jacobi-Bellman equation itself, namely

$$\beta = \left(E + \frac{1}{2}B^2 - V\right) dqdt + \frac{\hbar}{2} dBdt.$$

Let us recall that for given S, its "section" mapping lifts up the (q, t) base manifold of independent variables into the 5-dimensional jet space J^1 according to $\{q, t, S(q, t), \frac{\partial S}{\partial q}(q, t), \frac{\partial S}{\partial t}(q, t)\}$. All the "sectioned"forms pull back to zero onto the base 2-submanifold of J^1, which is called an integral submanifold.

Since $d\beta = (-dq + Bdt)d\omega$, it belongs to the ideal I of forms generated by ω, Ω and β. This one is therefore the smallest differential ideal containing ω and β. E. Cartan has shown that, under these conditions, the geometric representation of our PDE (3.2) can be completed (cf. [3, 10]).

Clearly, the ideal I contains the three fundamental ingredients we needed for our contact geometrical approach to classical dynamics. On the basis of our definition of Lagrangian, in the free case [Iso], and of the form of the Lagrangian for such elementary systems in Euclidean approaches to Feynman's ideas (cf. [4]), we infer that the Lagrangian underlying I should be (we shall prove it later on) the function of the independent variables B, q, t:

$$L(B, q, t) = \frac{1}{2}B^2 + V(q, t). \tag{3.5}$$

Let us denote by N a vector field on J^1 playing the role of a classical symmetry contact vector field X_n. In the context of our ideal I one should have:

$$\mathcal{L}_N(I) \subseteq I. \tag{3.6}$$

Such a N has been called, sometimes, isovector [10] and the theory of those in term of differential ideals is due, in essence, to E. Cartan [3].

Because of the linearity of Eq. (1.5), the Lie algebra $\mathcal{G}$ of these isovectors contains an infinite-dimensional abelian ideal $\mathcal{J}$, with canonical supplement $\aleph$.

In the free case $V = a = 0$, $\aleph$ has dimension 6 and possesses a natural basis, each element of which corresponds to a symmetry of our system.

As in this free case, the probabilistic interpretation of our results rests on the fact that, on the integral 2-submanifold, the underlying continuous trajectories will be of the form $t \mapsto z(t)$, solutions of stochastic differential equation (1.1).

It is well known that what plays, then, the role of the strong derivative along smooth (or "classical") paths is the infinitesimal generator of this diffusion ("Bernstein", cf. [4]) process:

$$\widetilde{D} = \frac{\partial}{\partial t} + \widetilde{B}\frac{\partial}{\partial q} + \frac{\hbar}{2}\frac{\partial^2}{\partial q^2} \tag{3.7}$$

where $\widetilde{B}$ was defined in (1.2).

Let us write an isovector N as

$$N = N^q\frac{\partial}{\partial q} + N^t\frac{\partial}{\partial t} + N^S\frac{\partial}{\partial S} + N^B\frac{\partial}{\partial B} + N^E\frac{\partial}{\partial E}.$$

Then we have the

Theorem 3.1. *Along each isovector N of $\aleph = \left\{N \in \mathcal{G} \ s.t \ \frac{\partial N^S}{\partial S} = 0\right\}$ satisfying* (3.6) *for the ideal I generated by (ω, Ω, β) as defined before, and for the Lagrangian L associated with it, given by* (3.5), *we have*

$$\begin{aligned} &(1) \quad \mathcal{L}_N(Bdq + Edt) = -dN^S \\ &(2) \quad \mathcal{L}_N(\Omega) = 0 \\ &(3) \quad \mathcal{L}_N(L) + L\frac{dN^t}{dt} = -DN^S \end{aligned} \tag{3.8}$$

where $D \equiv \frac{\partial}{\partial t} + B\frac{\partial}{\partial q} + \frac{\hbar}{2}\frac{\partial^2}{\partial q^2}$.

A word of caution is needed, before the (tedious, algebrico-geometrical) proof of this theorem.

If it was not for the last term of (3.8), one could suspect that the following proof has nothing to do with stochastic analysis. Note, however, that if we could look at D in (3.8), as an operator whose first-order term in q does not involve the independent variable $B \in J^1$ but the section $\widetilde{B} = \widetilde{B}(q,t) = -\frac{\partial \widetilde{S}}{\partial q}$, like in (3.7), then the relation of Theorem 3.1 with stochastic analysis would be clear. We shall show, in Theorem 3.3, that we are indeed allowed to do this and therefore to find a probabilistic interpretation on the integral 2-submanifold. The traditional notation $\sim$ for the sectioned geometrical objects is an anticipation of Theorem 3.3.

Proof. By definition $\mathcal{L}_N(\omega)$ is sum of multiples (with "Lagrange multipliers") of $\omega, d\omega$ and β. Necessarily, here, there is a function φ such that $\mathcal{L}_N(\omega) = \varphi\omega$. In analogy with (2.14), and following [10], let us define $F_N \equiv \omega(N) = BN^q + EN^t + N^S$ and consider

$$\begin{aligned} \varphi\omega - dF_N &= \mathcal{L}_N(\omega) - d(\omega(N)) \\ &= d\omega(N) \\ &= N^B dq - N^q dB + N^E dt - N^t dE. \end{aligned}$$

After substitution of (3.4), the identification of the coefficients of both sides provides $\varphi = \frac{\partial F_N}{\partial S}$, then

$$\begin{aligned}
&N^q = \frac{\partial F_N}{\partial B}, \quad N^t = \frac{\partial F_N}{\partial E}, \quad N^S = F_N - B\frac{\partial F_N}{\partial B} - E\frac{\partial F_N}{\partial E}\\
&N^B = -\frac{\partial F_N}{\partial q} + B\frac{\partial F_N}{\partial S}, \quad N^E = -\frac{\partial F_N}{\partial t} + E\frac{\partial F_N}{\partial S}
\end{aligned} \tag{3.9}$$

which should be compared with the components of the (real time!) classical contact field (2.13) associated with a contact Hamiltonian F_N.

By hypothesis $\mathcal{L}_N(\omega) \in I$, so $\mathcal{L}_N(d\omega) = d\mathcal{L}_N(\omega) \in dI \subset I$. Therefore the vector field N associated with F_N by (3.9) will be an isovector of I if and only if $\mathcal{L}_N(\beta) \in I$. So there must be two 0-forms α and γ and a 1-form ε such that

$$\mathcal{L}_N(\beta) = \alpha\beta + \varepsilon\omega + \gamma d\omega. \tag{3.10}$$

As before, those Lagrange multipliers should be eliminated. Without loss of generality we can assume that ε has no dS term (cf. [10, p. 658]). So it reduces to

$$\varepsilon = \mu dq + \lambda dt + \rho dB + \nu dE.$$

Explicitly, (3.10) means, after identification of the respective coefficients on both sides,

$$\begin{aligned}
&-N^E - BN^B + N^t\frac{\partial V}{\partial t} + N^q\frac{\partial V}{\partial q} + \left(V - E - \frac{1}{2}B^2\right)\frac{\partial N^q}{\partial q}\\
&+ \left(V - E - \frac{1}{2}B^2\right)\frac{\partial N^t}{\partial t} - \frac{\hbar}{2}\frac{\partial N^B}{\partial q}\\
&= \alpha\left(V - E - \frac{1}{2}B^2\right) + \lambda B - \mu E &(3.11)\\
&\left(V - E - \frac{1}{2}B^2\right)\frac{\partial N^q}{\partial S} - \frac{1}{2}\hbar\frac{\partial N^B}{\partial S} = \lambda &(3.12)\\
&\left(V - E - \frac{1}{2}B^2\right)\frac{\partial N^q}{\partial E} - \frac{1}{2}\hbar\frac{\partial N^B}{\partial E} = -\nu E - \gamma &(3.13)\\
&\left(V - E - \frac{1}{2}B^2\right)\frac{\partial N^q}{\partial B} - \frac{1}{2}\hbar\frac{\partial N^B}{\partial B} - \frac{\hbar}{2}\frac{\partial N^t}{\partial t} = -\alpha\frac{1}{2}\hbar - \rho E &(3.14)\\
&\left(E + \frac{1}{2}B^2 - V\right)\frac{\partial N^t}{\partial S} = \mu &(3.15)\\
&\left(E + \frac{1}{2}B^2 - V\right)\frac{\partial N^t}{\partial E} = -\nu B\\
&\left(E + \frac{1}{2}B^2 - V\right)\frac{\partial N^t}{\partial B} - \frac{\hbar}{2}\frac{\partial N^t}{\partial q} = -\rho B - \gamma &(3.16)\\
&\nu = 0 &(3.17)
\end{aligned}$$

$$\frac{1}{2}\hbar\frac{\partial N^t}{\partial S} = \rho \tag{3.18}$$

$$-\frac{1}{2}\hbar\frac{\partial N^t}{\partial E} = 0. \tag{3.19}$$

So we are left with conditions (3.11), (3.13) and (3.19). From (3.19) we get $\partial N^t/\partial E = 0$, i.e., (see (3.9)) $\frac{\partial^2 F_N}{\partial E^2} = 0$. Therefore, our contact Hamiltonian F_N is affine in E:

$$F_N = a + TE,$$

where a and T may depend only upon (q, t, B, S). After substitution of this F in (3.10), using (3.16), (3.17) and (3.18) we can rewrite (3.13) as

$$\left(V - E - \frac{1}{2}B^2\right)\frac{\partial N^q}{\partial E} - \frac{\hbar}{2}\frac{\partial N^B}{\partial E}$$
$$= -\frac{1}{2}\hbar\frac{\partial N^t}{\partial q} + \left(E + \frac{1}{2}B^2 - V\right)\frac{\partial N^t}{\partial B} + \frac{1}{2}\hbar B\frac{\partial N^t}{\partial S}.$$

Introducing the expressions (3.9) of $N^q, N^t, \ldots$, this reduces to

$$2\left(E + \frac{1}{2}B^2 + V\right)\frac{\partial T}{\partial B} = \hbar\frac{\partial T}{\partial q} - \hbar B\frac{\partial T}{\partial S}.$$

By identification of the coefficients of E, $\frac{\partial T}{\partial B} = 0$ and $\frac{\partial T}{\partial q} = B\frac{\partial T}{\partial S}$. By computing $\frac{\partial}{\partial B}$ of the last relation, we get $\frac{\partial T}{\partial S} = 0$. Finally,

$$T = T_N(t)$$

is true for any $N \in \mathcal{G}$.

Now the relations (3.9) can be rewritten in terms of coefficients $a = a(q, t, S, B)$ and $T = T_N(t)$. Since $N^t = T_N(t)$, (3.15) implies $\mu = 0$ and (3.18) implies $\rho = 0$, and (3.16) implies $\gamma = 0$. Then (3.11) reduces to an identity involving the partial derivatives of a and T_N. By elimination of λ and α using (3.12) and (3.14) in this identity, we can identify the coefficients of E^2 on both sides and get $\frac{\partial^2 a}{\partial B^2} = 0$, i.e.,

$$a = X_N B + h \tag{3.20}$$

where $X_N \equiv N^q$ and h can depend only upon (q, t, S). This means that the contact Hamiltonian F_N is also affine in B:

$$F_N = X_N B + T_N E + h. \tag{3.21}$$

The identification of the coefficients of E provides

$$2\frac{\partial X_N}{\partial q} - 2B\frac{\partial X_N}{\partial S} - \dot{T}_N = 0.$$

Since X_N cannot depend upon B, and $T_N = T_N(t)$ only, $\frac{\partial X_N}{\partial S} = 0$, i.e., $X_N = X_N(q, t)$. Moreover the last relation implies that

$$X_N(q, t) = \frac{1}{2}\dot{T}_N(t)q + l(t).$$

According to the definition of N^S (cf. (3.9)) and of the contact Hamiltonian F_N in (3.21), we have $N^S = h(q,t,S)$. So, for any $N \in \aleph$, N^S depends only on (q,t). It is immediate to check that if $N_1, N_2 \in \aleph$, $[N_1, N_2](S)$ is a function of (q,t), so $\aleph$ is indeed a Lie subalgebra of $\mathcal{G}$.

We are now able to prove the claims of Theorem 3.1:

$$\begin{aligned}\mathcal{L}_N(\omega) &= \varphi\omega \\ &= \frac{\partial F_N}{\partial S}\omega = 0,\end{aligned}$$

whence

$$\mathcal{L}_N(d\omega) = 0.$$

Since, by (3.4), $\omega_{pc} = \omega - dS$, this reduces to

$$\begin{aligned}\mathcal{L}_N(Bdq + Edt) &= \mathcal{L}_N(\omega) - d(\mathcal{L}_N(S)) \\ &= -dN^S.\end{aligned}$$

This proves (1). To prove (3), we need a more explicit version of N^S. Coming back to the reduced version of (3.11) providing (3.20), in the general setting where

$$N^S = h(q,t,S),$$

and substituting there what we deduced from (3.9) for F_N of the form (3.21), i.e.,

$$\begin{aligned}&N^q \equiv X_N = \frac{1}{2}\dot{T}_N(t)q + l(t) \\ &N^t \equiv T_N(t) \\ &N^S = h(q,t,S) \\ &N^B = -\frac{1}{2}\dot{T}_N(t)B - \frac{\partial h}{\partial q} + B\frac{\partial h}{\partial S} \\ &N^E = -\left(\frac{1}{2}\ddot{T}_N q + \dot{l}\right)B - \dot{T}_N E - \frac{\partial h}{\partial t} + E\frac{\partial h}{\partial S},\end{aligned} \tag{3.9$'$}$$

we obtain three equations corresponding to the identification of the coefficients of B^2, B and 1. Respectively,

$$\frac{\partial^2 h}{\partial S^2} = \frac{1}{\hbar}\frac{\partial h}{\partial S} \tag{3.22}$$

$$\frac{1}{2}\ddot{T}_N q + \dot{l} + \frac{\partial h}{\partial q} = \hbar\frac{\partial^2 h}{\partial S \partial q} \tag{3.23}$$

$$\frac{\partial h}{\partial t} + T_N\frac{\partial V}{\partial t} + \left(\frac{1}{2}\dot{T}_N q + l\right)\frac{\partial V}{\partial q} + \frac{\hbar}{2}\frac{\partial^2 h}{\partial q^2} = \left(\frac{\partial h}{\partial S} - \dot{T}_N\right)V. \tag{3.24}$$

We can express the solution of Eq. (3.22) as

$$h(q,t,S) = \hbar\widetilde{\eta}(q,t)e^{\frac{1}{\hbar}S} - \phi(q,t). \tag{3.25}$$

Then (3.23) means

$$\frac{\partial \phi}{\partial q} = \frac{1}{2}\ddot{T}_N q + \ddot{l} \equiv \frac{\partial X_N}{\partial t}, \tag{3.26}$$

therefore,

$$\phi(q,t) = \frac{1}{4}\ddot{T}_N q^2 + \dot{l}q - \sigma(t). \tag{3.27}$$

After substitution of (3.25) in (3.24), we obtain an equation which splits into

$$\hbar\frac{\partial\widetilde{\eta}}{\partial t} = -\frac{\hbar^2}{2}\frac{\partial^2\widetilde{\eta}}{\partial q^2} + V\widetilde{\eta} \tag{3.28}$$

and

$$-\frac{1}{4}\dddot{T}_N q^2 - \ddot{l}q + \dot{\sigma} + T_N\frac{\partial V}{\partial t} + \frac{1}{2}\dot{T}_N q\frac{\partial V}{\partial q} + l\frac{\partial V}{\partial q} + \dot{T}_N V - \frac{\hbar}{4}\ddot{T}_N = 0. \tag{3.29}$$

To prove (3) recall that $N \in \aleph$ implies $N^S = h$ does not depend on S, i.e., $\widetilde{\eta} = 0$ in the relation (3.25) or

$$h = -\phi(q,t). \tag{3.30}$$

Now, for L as in (3.5), since L is a function, i.e., a 0-form,

$$\begin{aligned}
\mathcal{L}_N(L) &= N(L) \\
&= N\left(\frac{1}{2}B^2 + V(q,t)\right) \\
&= BN^B + N^q\frac{\partial V}{\partial q} + N^t\frac{\partial V}{\partial t}.
\end{aligned}$$

Introducing N^B, N^q, N^t given by (3.9′) and $h = -\phi(q,t)$ as in (3.27), using (3.29) one verifies that

$$\mathcal{L}_N(L) + L\frac{dT_N}{dt} = \frac{\partial\phi}{\partial t} + B\frac{\partial\phi}{\partial q} + \frac{\hbar}{2}\frac{\partial^2\phi}{\partial q^2} \equiv D\phi.$$

This is (3.8) for $\phi = -N^S$. □

Remarks 3.2. **1)** In the course of the proof of Theorem 3.1, we have introduced new labels for the coefficients N^q, N^t and N^S (a-priori functions of all the variables of J^1), namely $X_N(q,t)$, $T_N(t)$ and $-\phi_N(q,t)$. Our first reason for this is to turn easier the comparison with the classical expressions (2.15) and (2.16). The second one is to relate directly the formulation of our present results with the ones of the stochastic Noether theorem, proved in [16] without using the jet space J^1 and its 2-integral submanifold of independent variables (q,t), where the probabilistic interpretation $z_t = q$ will be valid. Denoting by F_N the symmetry contact Hamiltonian associated with the isovector $N \in \aleph$, it follows from (3.21), (3.30) and our calculations above that

$$\begin{aligned}
F_N &= F_N(q,B,t,E) \\
&= X_N(q,t)B + T_N(t)E - \phi_N(q,t).
\end{aligned}$$

This is the (Euclidean) deformation of the classical symmetry contact Hamiltonian denoted by $n(q,p,t,E)$ in (2.15). To get a better notion of the deformation in question, let us recall that, in the context of the stochastic Noether theorem for such systems, we had found the following "determining" relations between the coefficients X_N, T_N and ϕ_N: (cf. [16], Lemma 3.5, in one dimension)

$$\begin{aligned}
&\dot{T}_N = 2\frac{\partial X_N}{\partial q}\\
&\frac{\partial X_N}{\partial t} = \frac{\partial \phi_N}{\partial q}\\
&\frac{\partial \phi_N}{\partial t} + \frac{\hbar}{2}\Delta\phi_N = \dot{T}_N V + X_N\frac{\partial V}{\partial q} + T_N\frac{\partial V}{\partial t}.
\end{aligned} \tag{3.31}$$

By (3.9′) here, the first relation is true. The second one was already checked in (3.26). Now the derivative $\frac{\partial}{\partial q}$ of (3.29) coincides with the integrability condition $\frac{\partial^2\phi_N}{\partial t\partial q} = \frac{\partial^2\phi_N}{\partial q\partial t}$ in the relations above:

$$\frac{\partial^2 X_N}{\partial t^2} - 3\frac{\partial X_N}{\partial q}\frac{\partial V}{\partial q} - X_N\frac{\partial^2 V}{\partial q^2} - T_N\frac{\partial^2 V}{\partial q\partial t} = 0, \tag{3.32}$$

after substitution of $X_N = \frac{1}{2}\dot{T}_N q + l(t)$ and $T = T_N(t)$. So the determining relations resulting from our analysis are the deformations of the classical relations (2.16). Except for the change of signs of Euclidean origin, the only deformation involves, in fact, the Laplacian of ϕ_N.

The integrability condition (3.32) is useful computationally. For a given V, it provides easily coefficients X and T allowed for symmetries.

2) If we did not know from the start that, behind the ideal I, there is the parabolic equation (3.28), the above calculation would have proved it.

Let N be an isovector for Eq. (1.5). We consider here again the case $a = 0$. By construction, $e^{\alpha N}$, $\alpha \in \mathbb{R}$, maps (q, t, S, B, E) to $(q_\alpha, t_\alpha, S_\alpha, B_\alpha, E_\alpha)$. Defining $\widetilde{\eta}_\alpha$ by

$$e^{-\frac{1}{\hbar}S_\alpha} = \widetilde{\eta}_\alpha(q_\alpha, t_\alpha)$$

it is known that $\widetilde{\eta}_\alpha$ solves the same PDE as $\widetilde{\eta}(q,t)$ (this is the definition of its symmetry group). If we denote by $e^{\alpha\widetilde{N}} : \widetilde{\eta} \mapsto \widetilde{\eta}_\alpha$ the associated one-parameter group, we obtain the following homomorphism of Lie algebras associated with the section within J^1 mentioned in the introduction:

$$\begin{aligned}
N &= N^q\frac{\partial}{\partial q} + N^t\frac{\partial}{\partial t} + N^S\frac{\partial}{\partial S} + N^B\frac{\partial}{\partial B} + N^E\frac{\partial}{\partial E}\\
&\longrightarrow -\widetilde{N} = +N^t\frac{\partial}{\partial t} + N^q\frac{\partial}{\partial q} - \frac{1}{\hbar}N^S,
\end{aligned} \tag{3.33}$$

where the last formula means that for each regular function $f(q,t)$,

$$\widetilde{N}f(q,t) = -N^t\frac{\partial f}{\partial t} - N^q\frac{\partial f}{\partial q} + \frac{1}{\hbar}N^S f.$$

Let us recall that, by definition, such a mapping preserves all the operations in the less complicated Lie algebra defined on the integral submanifold. Our notation $\sim$ is the same as the one used for sectioning differential forms in an ideal (cf. [3, 10]). We will use it now to give the probabilistic interpretation of Theorem 3.1 on the integral submanifold:

Theorem 3.3. *Let $z_t, t \in \Sigma$, be a solution of the SDE* (1.1) *built in term of positive solutions of Eq.* (1.5) *with $H = -\frac{\hbar^2}{2}\Delta + V(q,t)$ and for any regular V in the Kato class (cf.* [4, 6]*). The probabilistic counterparts of the characteristic equations* (2.7)–(2.9) *for the associated classical system are given, in term of the generator* (3.7), *by*

(1) $\widetilde{D}z = \widetilde{B}$
(2) $\widetilde{D}\widetilde{B} = \frac{\partial V}{\partial q}$
(3) $\widetilde{D}\widetilde{E} = \frac{\partial V}{\partial t}$
(4) $\widetilde{D}\widetilde{S} = -\left(\frac{1}{2}\widetilde{B}^2 + V\right) \equiv -L$ or
(5) $\widetilde{S}(z_t,t) = E_t \int_t^v L(\widetilde{B}_\tau, z_\tau, \tau)d\tau + E_t\widetilde{S}_v(z_v)$
(6) $\qquad\qquad = E_t \int_t^u \widetilde{B} \circ dz_\tau + \widetilde{E}d\tau + E_t\widetilde{S}_v(Z_v)$.

The invariance identity (3) *of Theorem* 3.1 *can rewritten in terms of $N^q = X_N$ and $N^t = T_N$ as*

(7) $X_N\frac{\partial L}{\partial q} + T_N\frac{\partial L}{\partial t} + (\widetilde{D}X_N - \widetilde{B}\dot{T}_N)\frac{\partial L}{\partial \widetilde{B}} + L\dot{T}_N = \widetilde{D}\phi_N.$

Proof. This theorem summarizes properties of diffusions (1.1) which have been proved, along the years, without knowledge of their contact geometrical background (cf. [4, 16]). According to (3.1) and (3.3), we have

$$\widetilde{B}(q,t) = \hbar\nabla\ ln\ \widetilde{\eta}(q,t)$$
$$\widetilde{E}(q,t) = \hbar\partial_t\ ln\ \widetilde{\eta}(q,t)$$

for $\widetilde{\eta}$ a positive solution of the parabolic equation (1.5) (with $a = 0$). The relations (1) to (4) result from direct computation with the generator $\widetilde{D}$ (cf. (3.7)) of $z(t)$.

Eq. (5) follows from (4) by the Itô-Dynkin formula, under integrability conditions.

The geometrical definition (6) of the action $\widetilde{S}$ in term of the Poincaré-Cartan form ω_{pc} of (3.4) needs to be clarified. If E_t denotes the conditional expectation $E[\dots|z(t)]$, we have

$$E_t\int_t^v \widetilde{\omega}_{pc} \equiv E_t\int_t^v \widetilde{B}\circ dz(\tau) + \widetilde{E}d\tau,$$

where $\circ$ denotes the (Fisk) Stratonovich integral in the sense of Itô [11],

$$\begin{aligned}
&= \hbar E_t \int_t^v (\nabla\, ln\, \widetilde{\eta}(z_\tau, \tau) \circ dz(\tau) + \partial_\tau\, ln\, \widetilde{\eta}(z(t), \tau) d\tau) \\
&= \hbar E_t \int_t^v d(ln\, \widetilde{\eta}(z(\tau), \tau)) \\
&= \hbar E_t\, ln\, \widetilde{\eta}(z(v), v) - \hbar\, ln\, \widetilde{\eta}(z(t), t) \\
&= \widetilde{S}(z(t), t) - E_t \widetilde{S}(z(v)).
\end{aligned}$$

Regarding (7), observe that, using (3.7) and the determining equations (3.29),

$$\widetilde{D} X_N - \widetilde{B} \dot{T}_N = \frac{\partial \phi_N}{\partial q} - \widetilde{B} \frac{\partial X_N}{\partial q},$$

and so (7) reduces indeed to (3.8). □

As classically, the invariance identity (7) can be regarded as a basic formula of a (stochastic) calculus of variations. Such a calculus already exists, cf. [6], and allows to obtain directly some of the abovementioned results (but without the geometrical insight).

4. Perturbation of the free case and examples

As indicated in the introduction, Theorem 3.1 and Theorem 3.3 are true for any regular potential V (in the Kato class). As soon as Eq. (3.31) can be solved, an isovector N (and therefore $\widetilde{N}$) is determined and can be used to relate explicitly solutions of heat equations, i.e., for us, to obtain a quadrature of diffusion processes. But there is a special class of potentials V, in Theorem 3.1 and Theorem 3.3, for which all such computations can be done explicitly without using more than the isovectors of the free heat equation ($a = V = 0$). This important class plays the role, in our probabilistic framework, of the quadratic class associated with the metaplectic representation in quantum theory (cf., for instance, [7]). Let us start from a positive solution η_χ of the free equation, such that

$$\eta_\chi(q, 0) = \chi(q) > 0.$$

The above observation can be expressed via the following result of Rosencrans:

Theorem 4.1 ([15]). *If we denote, for a given $\widetilde{N}$ as in (3.33), $(e^{\alpha \widetilde{N}} \eta_\chi)(q, t)$ by $\rho_N(q, t, \alpha)$ and $\rho_N(q, o, \alpha)$ by $\eta^N(q, \alpha)$, then η^N solves*

$$\begin{cases} \hbar \frac{\partial \eta^N}{\partial \alpha} = N^t(q, 0) \frac{\hbar^2}{2} \frac{\partial^2 \eta^N}{\partial q^2} - N^q(q, 0) \hbar \frac{\partial \eta^N}{\partial q} - N^S(q, 0) \eta^N \\ \eta^N(q, 0) = \chi(q). \end{cases}$$

So, choosing an isovector N of the free equation such that $N^q(q, 0) = -\frac{1}{\hbar}(aq + b)$, $N^t(q, 0) = -1$ and $N^S(q, 0) = -(cq^2 + dq + f)$, for a, b, c, d, f real constants, then *(denoting again the parameter by t):*

$$\begin{cases} \hbar\frac{\partial\eta^N}{\partial t} = -\frac{\hbar^2}{2}\frac{\partial^2\eta^N}{\partial q^2} + (aq+b)\frac{\partial\eta^N}{\partial q} + (cq^2+dq+f)\eta^N \\ \eta^N(q,0) = \chi(q). \end{cases} \tag{4.1}$$

Let us consider some examples in this quadratic class:

1) Linear potential $\mathbf{V(q) = \lambda q}$, $\lambda \in \mathbb{R}$

This is the case $a = b = c = f = 0$, $d = \lambda$ of Theorem 4.1. Then

$$\eta_\chi^N(q,t) = e^{-\frac{1}{\hbar}\left(\frac{\lambda^2}{6}t^3 - \lambda t q\right)}\eta_\chi\left(q - \lambda\frac{t^2}{2}, t\right)$$

solves Eq. (4.1) if η_χ solves the free equation. The drift $\widetilde{B}_V$ of the diffusion $z_V(t)$ associated with this perturbation is, by (1.2),

$$\begin{aligned} \widetilde{B}_V(q,t) &= \hbar\frac{\partial}{\partial q} ln\, \eta_\chi^N(q,t) \\ &= \lambda t + \widetilde{B}\left(q - \lambda\frac{t^2}{2}, t\right), \end{aligned}$$

for $\widetilde{B}$ the drift coming from the free equation. The relation between the two families of diffusions is, as expected, the deterministic translation:

$$z_V(t) = z(t) + \lambda\frac{t^2}{2}.$$

2) Quadratic and first order linear perturbations

This is the case $a = \beta\hbar$, $b = 0$, $c = \frac{1}{2}\beta^2$, $d = 0$, $f = -\frac{\hbar}{2}\beta$ in Theorem 4.1, corresponding to (a one-dimensional version of) $A(q) = \beta q$ and $V(q) = \frac{1}{2}\beta^2 q^2 - \frac{\hbar}{2}\beta$ in [11, p. 71]. It was shown there that the relevant parabolic equation is $\hbar\frac{\partial\eta^N}{\partial t} = H\eta^N$ with

$$\begin{aligned} H &= -\frac{\hbar^2}{2}\left(\nabla - \frac{A}{\hbar}\right)^2 + V \\ &= -\frac{\hbar^2}{2}\Delta + \hbar\beta q\nabla \end{aligned}$$

and that the associated drift is of the form

$$\widetilde{B}_V(q,t) = \hbar\frac{\partial}{\partial q}\, ln\, \eta_\chi^N(q,t) - A(q).$$

Using the relation between η_χ^N and the free solution:

$$\eta_\chi^N(q,t) = \eta_\chi\left(e^{\beta t}q, \frac{1}{2\beta}(e^{2\beta t} - 1)\right),$$

the relation between drifts becomes

$$\widetilde{B}_V(q,t) = e^{\beta t}\widetilde{B}\left(e^{\beta t}q, \frac{1}{2\beta}(e^{2\beta t} - 1)\right) - \beta q.$$

After time integration we get

$$z_V(t) = e^{-\beta t}\left[c(\omega) + z\left(\frac{1}{2\beta}(e^{2\beta t}-1)\right)\right],$$

where $c(\omega)$ is an arbitrary random constant.

The simplest illustration is to start from $\chi = 1$. Then $\eta_\chi = 1$ and $z(t)$ is the Wiener $\sqrt{\hbar}w(t)$ itself, whose drift $\widetilde{B} = 0$. Then $\widetilde{B}_V = -\beta q$, so z_V solves

$$dz_V(t) = -\beta z_V(t)dt + \sqrt{\hbar}dw(t).$$

This means that such a perturbation contains Doob's relation (1.4) as a very special case.

Let us stress again that it would be a mistake to understand Theorem 3.1 and Theorem 3.3 as meaning that stochastic quadratures of diffusion processes are available only when the perturbations are quadratic polynomials.

Consider for example:

3) $\mathbf{V(q) = \frac{\alpha^2}{q^2}}$, $\alpha \in \mathbb{R}$
It is easy to verify that the integrability condition (3.32) is satisfied by the following pairs of coefficients (X_N, T_N):

$$\begin{aligned}
X_N &= 0 \qquad T_N = 0\\
X_N &= 0 \qquad T_N = 1\\
X_N &= \frac{1}{2}q, \quad T_N = t\\
X_N &= -qt, \quad T_N = -t^2.
\end{aligned} \tag{4.2}$$

The associated coefficients ϕ_N follow easily from (3.31). Using the notation of [Iso] (cf. [3, p. 190]) the list (4.2) corresponds, respectively, to the isovectors N_3, N_1, N_2 and N_6. There are no other symmetries for this V. But we can still express the diffusion $z_V(t)$ in terms of some (here four) of the isovectors of the free, 6-dimensional Lie algebra $\aleph$ of §3. So, at the expense of reducing the dimension of $\aleph$, many stochastic quadratures are available beyond the quadratic class of potentials involved in Theorem 4.1.

5. A transformation of diffusions relevant to first crossing problems and mathematical finance

As mentioned in the introduction, our approach allows us, for example, to recover and extend explicit results established by Patie [13] and Alili-Patie [2] with a view to the computation of some option prices. Here we set $a = V = 0$ in the above general geometrical structure. Still, this application is not trivial: it requires the use of two of the free isovectors computed in [Iso], namely

$$\begin{aligned} N_4 &= 2t\frac{\partial}{\partial t} + q\frac{\partial}{\partial q} - 2E\frac{\partial}{\partial E} - B\frac{\partial}{\partial B} \\ N_6 &= 2t^2\frac{\partial}{\partial t} + 2qt\frac{\partial}{\partial q} + (\hbar t - q^2)\frac{\partial}{\partial S} - (2qB + 4tE + \hbar)\frac{\partial}{\partial E} + 2(q - tB)\frac{\partial}{\partial B}. \end{aligned}$$

For $\alpha > 0$ and $\beta \in \mathbb{R}$, let us set $\mu = -ln(\alpha)$ and $\lambda = -\alpha\beta$, and define:

$$R_{\alpha,\beta} \underset{\text{def}}{\equiv} e^{\mu N_4} e^{-\frac{\lambda}{2}N_6}.$$

$R_{\alpha,\beta}$ is a differential operator on the space $\mathcal{C}^\infty(M) = \mathcal{C}^\infty(\mathbb{R}^5)$ of smooth functions of (t, q, S, E, B), and it maps the subspace of smooth functions of (t, q) into itself. We extended $R_{\alpha,\beta}$ to $(\mathcal{C}^\infty(\mathbb{R}^5))^2$ by setting:

$$\forall (f, g) \in \mathcal{C}^\infty(\mathbb{R}^5)^2 \quad R_{\alpha,\beta}(f, g) = (R_{\alpha,\beta}(f), R_{\alpha,\beta}(g)).$$

It follows from case 2) of [Iso], p.201, that, for each $f \in \mathcal{C}^\infty(\mathbb{R}^2)$, $e^{-\frac{\lambda}{2}N_6}(f) = g$ is given by

$$g(t, q) = f\left(\frac{t}{1 + \lambda t}, \frac{q}{1 + \lambda t}\right).$$

Also, $e^{\mu N_4}$ maps g to h, where

$$h(t, q) = g\left(e^{2\mu}t, e^{\mu}q\right)$$

(cf. case 1) of [12], p. 201). Therefore $R_{\alpha,\beta} \underset{\text{def}}{\equiv} e^{\mu N_4} e^{-\frac{\lambda}{2}N_6}$ maps f to h, where

$$\begin{aligned} h(t, q) &= g\left(e^{2\mu}t, e^{\mu}q\right) \\ &= f\left(\frac{e^{2\mu}t}{1 + \lambda e^{2\mu}t}, \frac{e^{\mu}q}{1 + \lambda e^{2\mu}t}\right) \\ &= f\left(\frac{\frac{1}{\alpha^2}t}{1 - \alpha\beta\frac{1}{\alpha^2}t}, \frac{\frac{1}{\alpha}q}{1 - \alpha\beta\frac{1}{\alpha^2}t}\right) \\ &= f\left(\frac{t}{\alpha(\alpha - \beta t)}, \frac{q}{\alpha - \beta t}\right), \end{aligned}$$

i.e.,

$$R_{\alpha,\beta}(f)(t, q) = f\left(\varphi_{\alpha,\beta}(t), \psi_{\alpha,\beta}(t, q)\right),$$

where

$$\varphi_{\alpha,\beta}(t) \underset{\text{def}}{\equiv} \frac{t}{\alpha(\alpha - \beta t)}$$

and

$$\psi_{\alpha,\beta}(t, q) \underset{\text{def}}{\equiv} \frac{q}{\alpha - \beta t}.$$

Proposition 5.1. *Let $z(\cdot)$ denote a Bernstein diffusion solving* (1.1) *and let us define $z_{\alpha,\beta}$ by*

$$z_{\alpha,\beta}(\varphi_{\alpha,\beta}(t)) = \psi_{\alpha,\beta}(t, z(t)).$$

Then $z_{\alpha,\beta} = S^{(\alpha,\beta)}(z)$, where $S^{(\alpha,\beta)} : \mathcal{C}(\mathbb{R}^+, \mathbb{R}) \longrightarrow \mathcal{C}([0,T], \mathbb{R})$, with $T = -(\alpha\beta)^{-1}$ for $\beta < 0$ and $+\infty$ otherwise, is Patie's transformation in [13, p. 49].

Proof. Let us set $t_{\alpha,\beta} = \varphi_{\alpha,\beta}(t)$, then $t_{\alpha,\beta} = \frac{t}{\alpha(\alpha-\beta t)}$, whence

$$\begin{aligned} 1 + \alpha\beta t_{\alpha,\beta} &= 1 + \frac{\beta t}{\alpha - \beta t} \\ &= \frac{\alpha}{\alpha - \beta t} \end{aligned}$$

and

$$\begin{aligned} z_{\alpha,\beta}(t_{\alpha,\beta}) &= z_{\alpha,\beta}(\varphi_{\alpha,\beta}(t)) \\ &= \psi_{\alpha,\beta}(t, z(t)) \\ &= (\alpha - \beta t)^{-1} z(t) \\ &= \alpha^{-1}(1 + \alpha\beta t_{\alpha,\beta}) z(t) \\ &= \alpha^{-1}(1 + \alpha\beta t_{\alpha,\beta}) z\left(\frac{\alpha^2 t_{\alpha,\beta}}{1 + \alpha\beta t_{\alpha,\beta}}\right). \end{aligned}$$

Now by definition ([13, p. 49]),

$$S^{(\alpha,\beta)}(\omega)(\tau) = \frac{1 + \alpha\beta\tau}{\alpha}\, \omega\left(\frac{\alpha^2\tau}{1 + \alpha\beta\tau}\right),$$

so

$$z_{\alpha,\beta}(t_{\alpha,\beta}) = S^{(\alpha,\beta)}(z)(t_{\alpha,\beta}), \text{ i.e. } z_{\alpha,\beta} = S^{(\alpha,\beta)}(z). \qquad \square$$

Corollary 5.2. *When $\alpha = 1, z_{1,\beta} = S^{(\beta)}(z)$, where $S^{(\beta)} \underset{\text{def}}{\equiv} S^{(1,\beta)}$ is defined as in* [2], *§2. In particular, for $\beta < 0$ and $\eta_\chi = \chi = 1, z(t) = \sqrt{\hbar}\, w(t)$, as in Ex.* 2) *§4 and $z_{1,\beta}(t) = \sqrt{\hbar}\, w_{0,0}^{0,-\beta^{-1}}(t)$, a Brownian bridge.*

Proof. When $\alpha = 1$, $\mu = 0$ we have $R_{\alpha,\beta} = e^{-\frac{\lambda}{2}N_6}$ and $z_{\alpha,\beta}(\tau) = (1+\beta\tau) z\left(\frac{\tau}{1+\beta\tau}\right)$ $\equiv S^{(\beta)}(z)(\tau)$ by [2], §1 p. 226. Since $R_{\alpha,\beta} = e^{-\frac{\lambda}{2}N_6} = e^{\frac{\beta}{2}N_6}$ the statement about the Brownian Bridge follows from [Iso] p. 201 (setting $\alpha = -\beta$). $\square$

From the structure of the Lie algebra $\mathcal{H}$ it follows:

Theorem 5.3. *For all $\alpha > 0$, $\alpha' > 0$, β and β', one has:*

$$R_{\alpha',\beta'} \circ R_{\alpha,\beta} = R_{\alpha'',\beta''}$$

where

$$\alpha'' \underset{\text{def}}{\equiv} \alpha\alpha'$$

and

$$\beta'' \underset{\text{def}}{\equiv} \alpha\beta' + \frac{\beta}{\alpha'}.$$

Proof. From $[N_4, N_6] = 2N_6$ (see [Iso]) it follows that

$$N_6 N_4 = (N_4 - 2I)N_6,$$

whence, by an easy induction over m,

$$\forall m \in \mathbb{N} \quad N_6 {N_4}^m = (N_4 - 2I)^m N_6,$$

whence

$$N_6\, e^{\mu N_4} = e^{-2\mu}\, e^{\mu N_4}\, N_6.$$

Now another easy induction yields:

$$\forall n \in \mathbb{N} \qquad N_6^n\, e^{\mu N_4} = e^{-2n\mu}\, e^{\mu N_4}\, N_6^n,$$

whence

$$e^{-\frac{\lambda'}{2} N_6}\, e^{\mu N_4} N = e^{\mu N_4}\, e^{-\frac{\lambda'}{2}\, e^{-2\mu}\, N_6}.$$

It follows that

$$\begin{aligned} R_{\alpha',\beta'} \circ R_{\alpha,\beta} &= e^{\mu' N_4}\, e^{-\frac{\lambda'}{2} N_6}\, e^{\mu N_4}\, e^{-\frac{\lambda}{2} N_6} \\ &= e^{\mu' N_4}\, e^{\mu N_4}\, e^{-\frac{\lambda'}{2}\, e^{-2\mu}\, N_6}\, e^{-\frac{\lambda}{2} N_6} \\ &= e^{\mu'' N_4}\, e^{-\frac{\lambda''}{2} N_6} \end{aligned}$$

where

$$\begin{aligned} \mu'' &= \mu + \mu' \\ &= -\ln(\alpha) - \ln(\alpha') \\ &= -\ln(\alpha'') \end{aligned}$$

and

$$\begin{aligned} \lambda'' &= \lambda'\, e^{-2\mu} + \lambda \\ &= (-\alpha'\beta')\, e^{2\ln\alpha} - \alpha\beta \\ &= -\alpha\alpha' \left(\alpha\beta' + \frac{\beta}{\alpha'} \right) \\ &= -\alpha''\beta''. \end{aligned}$$

Therefore

$$R_{\alpha',\beta'} \circ R_{\alpha,\beta} = R_{\alpha'',\beta''}.$$

□

Corollary 5.4. *One has*

$$S^{(\alpha,\beta)} \circ S^{(\alpha',\beta')} = S^{(\alpha'',\beta'')}.$$

Proof. By definition,

$$\begin{aligned} \varphi_{\alpha'',\beta''}(t) &= R_{\alpha'',\beta''}(t) \\ &= R_{\alpha',\beta'}(R_{\alpha,\beta}(t)) \\ &= R_{\alpha',\beta'}(\varphi_{\alpha,\beta}(t)) \\ &= \varphi_{\alpha,\beta}(\varphi_{\alpha',\beta'}(t)) \end{aligned}$$

and

$$\begin{aligned}
\psi_{\alpha'',\beta''}(t,q) &= R_{\alpha'',\beta''}(q)\\
&= R_{\alpha',\beta'}(R_{\alpha,\beta}(q))\\
&= R_{\alpha',\beta'}(\psi_{\alpha,\beta}(t,q))\\
&= \psi_{\alpha,\beta}(\varphi_{\alpha',\beta'}(t),\ \psi_{\alpha',\beta'}(t,q))
\end{aligned}$$

as well as

$$\begin{aligned}
S^{(\alpha'',\beta'')}(z)(\varphi_{\alpha'',\beta''}(t)) &= \psi_{\alpha'',\beta''}(t,z(t))\\
&= \psi_{\alpha,\beta}\left(\varphi_{\alpha',\beta'}(t),\psi_{\alpha',\beta'}(t,z(t))\right)\\
&= \psi_{\alpha,\beta}(\varphi_{\alpha',\beta'}(t),\ S^{(\alpha',\beta')}(z)(\varphi_{\alpha',\beta'}(t)))\\
&= S^{(\alpha,\beta)}(S^{(\alpha',\beta')}(z))[\varphi_{\alpha,\beta}(\varphi_{\alpha',\beta'}(t))]\\
&= (S^{(\alpha,\beta)}\circ S^{(\alpha',\beta')})(z)[\varphi_{\alpha'',\beta''}(t)],
\end{aligned}$$

whence

$$S^{(\alpha'',\beta'')}(z) = (S^{(\alpha,\beta)}\circ S^{(\alpha',\beta')})(z),$$

i.e.,

$$S^{(\alpha,\beta)}\circ S^{(\alpha',\beta')} = S^{\left(\alpha\alpha',\alpha\beta'+\frac{\beta}{\alpha'}\right)},$$

i.e., Patie's formula ([13, p. 60]), modulo correction of a misprint. □

Remark 5.5. The above formula can also be obtained in an elementary fashion:

$$\begin{aligned}
((S^{(\alpha,\beta)}\circ S^{(\alpha',\beta')})(\omega))(\tau) &= (S^{(\alpha,\beta)}(S^{(\alpha'\beta')}(\omega)))(\tau)\\
&= \frac{1+\alpha\beta\tau}{\alpha}(S^{(\alpha',\beta')}(\omega))\left(\frac{\alpha^2\tau}{1+\alpha\beta\tau}\right)\\
&= \frac{1+\alpha\beta\tau}{\alpha}\ \frac{1+\alpha'\beta'\frac{\alpha^2\tau}{1+\alpha\beta\tau}}{\alpha'}\ \omega\left(\frac{\alpha'^2\frac{\alpha^2\tau}{1+\alpha\beta\tau}}{1+\alpha'\beta'\frac{\alpha^2\tau}{1+\alpha\beta\tau}}\right)\\
&= \frac{1+\alpha''\beta''\tau}{\alpha''}\ \omega\left(\frac{\alpha''^2\tau}{1+\alpha''\beta''\tau}\right)\\
&= \left(S^{(\alpha'',\beta'')}(\omega)\right)(\tau).
\end{aligned}$$

Corollary 5.6 ([2]). *$S^{(\beta)}$ satisfies the semigroup property.*

Proof. One applies Corollary 5.4 with $\alpha=\alpha'=1$; in this case, $\alpha''=1$ and $\beta''=\beta+\beta'$ whence:

$$S^{(\beta+\beta')} = S^{(\beta)}\circ S^{(\beta')}.$$

□

6. Conclusion

The method used here is considerably more general than our present results.

Of course, it still holds if we start from the n-dimensional heat equation instead of Eq. (1.3). All the results of Theorem 3.1 and 3.3, §3 are preserved. In fact, the method is the same when the underlying stochastic processes (and therefore operator H_0 of Eq. (1.3)) are more general than diffusions. In particular, the method can be adapted to the class of diffusion processes with jumps introduced in [14].

Finally, let us observe that the analogy alluded to, in the introduction, between our stochastic deformation and the quantization of elementary dynamical systems, is not superficial. For example, the stochastic Noether theorem associated with the isovectors N provides, after coming back from Eq. (3.28) to the Schrödinger equation, quantum first integrals in the L^2 sense. Even in the free one-dimensional case, this list of constants is strictly larger than the one known by traditional means in quantum mechanics (cf. [4] and [1]).

Part of the stochastic deformation strategy illustrated here could also be relevant to a more general one, inspired by ideas of dynamical systems [5].

Acknowledgement

The authors are grateful to the anonymous referee whose careful reading has improved the legibility of this work.

References

[1] S. Albeverio, J. Rezende, and J. C. Zambrini, *Probability and quantum symmetries II. The Theorem of Noether in Quantum Mechanics*, Journal of Mathematical Physics, **47 (6)** (2006), 062107, 61 pp.

[2] L. Alili and P. Patie, *On the first crossing times of a Brownian motion and a family of continuous curves*, C.R. Acad. Sci., Paris, Ser. I, **340** (2005), 225–228.

[3] E. Cartan, *Les systèmes différentiels extérieurs*, Hermann, Paris, 1971.

[4] K. L. Chung and J. C. Zambrini, *Introduction to Random Time and Quantum Randomness*, new edition, World Scientific, 2003.

[5] J. Cresson and S. Darses, *Stochastic embedding of dynamical systems*, in preparation.

[6] A. B. Cruzeiro and J. C. Zambrini, *Malliavin Calculus and Euclidean Quantum Mechanics. I. Functional calculus*, J. Funct. Anal., **96 (1)** (1991), 62–95.

[7] G. B. Folland, *Harmonic Analysis in Phase Space*, Princeton University Press, Princeton N.J., 1989.

[8] R. V. Gamkrelidze, Editor, *Geometry I*, Springer Encyclopedia of Mathematical Sciences, Vol. 28, Chapt. 5., Springer-Verlag, 1991.

[9] M. Giaquinta and S. Hildebrandt, *Calculus of Variations II*, Grundl. der Mathem. Wissench 311, Chapt. 10, Springer-Verlag, 1996.

[10] B. K. Harrison and F. P. Estabrook, *Geometric approach to invariance groups and solution of partial differential systems*, J. Math. Physics, **12 (4)** (1971), 653–666.

[11] K. Itô, *Extension of stochastic integrals*, in: Proceedings of the International Symposium SDE, K. Itô, Editor, Kinokuniya, Tokyo, 1976.

[12] P. Lescot and J. C. Zambrini, *Isovectors for Hamilton-Jacobi-Bellman Equations, Formal Stochastic Differentials and First Integrals in Euclidean Quantum Mechanics*, in: Progress in Probability, Vol. 58, R. Dalang, M. Dozzi, F. Russo, Editors, Birkhäuser, (2004), 187-202.

[13] P. Patie, *On Some First Passage Time Problems Motivated by Financial Applications*, Ph.D., ETH, 2004.

[14] N. Privault and J. C. Zambrini, *Markovian bridges and reversible diffusion processes with jumps*, Ann. Inst. H. Poincaré, **40** (2004), 599–633.

[15] S. I. Rosencrans, *Perturbation algebra of an elliptic operator*, Journal of Mathematical Analysis and Applications, **56** (1976), 317–329.

[16] M. Thieullen and J. C. Zambrini, *Probability and quantum symmetries I. The Theorem of Noether in Schrödinger's Euclidean Quantum Mechanics*, Ann. Inst. H. Poincaré (Phys. Theor.), **67 (3)** (1997), 297–338.

Paul Lescot
LAMFA, CNRS UMR 6140
Sous-équipe "Probabilités et Théorie Ergodique"
Université de Picardie Jules Verne
33 Rue Saint-Leu
F-80039 Amiens Cédex
France
and
INSSET-Université de Picardie
48 Rue Raspail
F-02100 Saint-Quentin
France
e-mail: `paul.lescot@u-picardie.fr`

Jean-Claude Zambrini
Grupo de Física-Matemática da Universidade de Lisboa
Av. Prof. Gama Pinto, 2
1649-003 Lisboa
Portugal
e-mail: `zambrini@cii.fc.ul.pt`

Progress in Probability, Vol. 59, 227–241

Approximation of Stochastic Differential Equations Driven by Fractional Brownian Motion

Hannelore Lisei and Anna Soós

Abstract. The aim of this paper is to approximate the solution of a stochastic differential equation driven by fractional Brownian motion (with Hurst index greater than $\frac{1}{2}$) using a series expansion for the noise. We prove that the solution of the approximating equations converge in probability to the solution of the given equation. We illustrate the approximation through an example from mathematical finance.

Mathematics Subject Classification (2000). Primary 60H10; Secondary 60H40, 60H05.

Keywords. Stochastic differential equations, approximation, fractional Brownian motion.

1. Introduction

We consider the following stochastic differential equation driven by fractional Brownian motion $\Big(B(t)\Big)_{t\in[0,1]}$ with Hurst index $H \in (\frac{1}{2}, 1)$:

$$X(t) = X_0 + \int_{t_0}^{t} F(X(s), s)ds + \int_{t_0}^{t} G(X(s), s)dB(s),\ t \geq t_0, \tag{1.1}$$

where the stochastic integral is defined in terms of fractional integration as investigated by M. Zähle [13], [15]. We assume that a.s. $F \in C(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$, $G \in C^1(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$, and for each $t \in [0,T]$ the functions $F(\cdot, t)$, $\dfrac{\partial G(\cdot, t)}{\partial x}$, $\dfrac{\partial G(\cdot, t)}{\partial t}$ are locally Lipschitz.

We approximate the fractional Brownian motion $\Big(B(t)\Big)_{t\in[0,1]}$ by using the series expansion given in [4]. Let J_ν be the Bessel function of first type of order ν and let $x_1 < x_2 < \ldots$ be the positive, real zeros of J_{-H}, while $y_1 < y_2 < \ldots$ are the positive, real zeros of J_{1-H}. We consider $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ to be two independent sequences of centered Gaussian random variables such that for each $n \in \mathbb{N}$ we have

$$\mathrm{Var}X_n = \frac{2c_H^2}{x_n^{2H} J_{1-H}^2(x_n)}, \quad \mathrm{Var}Y_n = \frac{2c_H^2}{y_n^{2H} J_{-H}^2(y_n)},$$

where $c_H^2 = \dfrac{\sin(\pi H)}{\pi}\Gamma(1+2H)$. We approximate equation (1.1) through

$$X_N(t) = X_0 + \int_0^t F(X_N(s),s)ds + \int_0^t G(X_N(s),s)dB_N(s), \quad \text{for each } N \in \mathbb{N}, \quad (1.2)$$

where

$$B_N(t) = \sum_{n=1}^{N} \frac{\sin(x_n t)}{x_n} X_n + \sum_{n=1}^{N} \frac{1-\cos(y_n t)}{y_n} Y_n, \quad t \in [0,1], N \in \mathbb{N}.$$

We will show that the equation (1.2) has a local solution, which converges in probability to the solution of (1.1) in the interval, where the solutions exist. We illustrate the approximation through the model for the price of risky assets from mathematical finance. The figures are generated by a Matlab program.

Investigations concerning stochastic differential equations driven by a fractional Brownian motion or more general fractional process have been done by L. Coutin and L. Decreusefond [2], F. Klingenhöfer and M. Zähle [7], M. Zähle [14, 15], M. Errami and F. Russo [5], and many others. These studies were motivated by the problems occurring in mathematical finance, telecommunication networks, biology, hydrology etc. The main difficulty raised by the fractional Brownian motion and the processes related to it, is that they are not Markovian, even more, they are not semimartingales. Hence a new approach to stochastic fractional calculus was developed. There exist several ways to define the stochastic integral, pathwise and related techniques, Dirichlet forms, anticipating techniques using Malliavin calculus and Skorohod integration (e.g., [1, 3, 10, 13]). In this paper we use the approach of M. Zähle [13], based on the ideas of Lebesgue-Stieltjes integrals and fractional calculus [11].

2. Series expansion for a fractional Brownian motion

For $\nu \neq -1, -2, \ldots$ **the Bessel function J_ν of the first type of order** ν is defined on the region $\{z \in \mathbb{C} : |\arg z| < \pi\}$ as the absolutely convergent sum

$$J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{\Gamma(k+1)\Gamma(\nu+k+1)} \left(\frac{z}{2}\right)^{\nu+2k}.$$

It is known that for $\nu > -1$ the function J_ν has a countable number of real, positive simple zeros (see [12, Chapter 15]). We fix now $H \in (0,1)$. Let $x_1 < x_2 < \dots$ be the positive, real zeros of J_{-H} and let $y_1 < y_2 < \dots$ be the positive, real zeros of J_{1-H}.

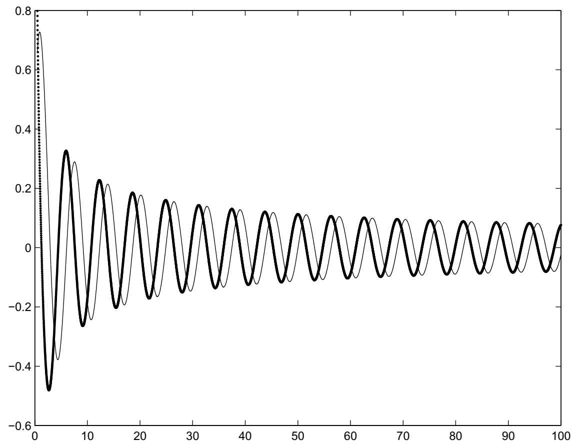


FIGURE 1. Bessel functions: J_{-H} (with '·'), J_{1-H} (with '-'), $H = 0.65$.

Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be two independent sequences of independent Gaussian random variables such that for each $n \in \mathbb{N}$ we have

$$E(X_n) = E(Y_n) = 0$$

and

$$\mathrm{Var}X_n = \frac{2c_H^2}{x_n^{2H} J_{1-H}^2(x_n)}, \quad \mathrm{Var}Y_n = \frac{2c_H^2}{y_n^{2H} J_{-H}^2(y_n)},$$

where

$$c_H^2 = \frac{\sin(\pi H)}{\pi}\Gamma(1+2H).$$

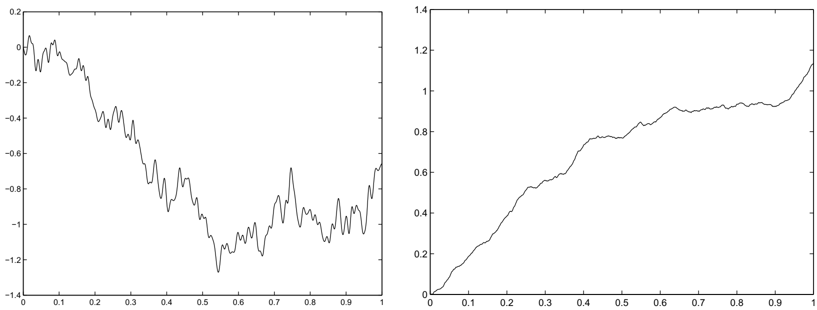
In [4, Theorem 4.5] it is proved that the random process $\Big(B(t)\Big)_{t\in[0,1]}$ given by

$$B(t) = \sum_{n=1}^{\infty} \frac{\sin(x_n t)}{x_n} X_n + \sum_{n=1}^{\infty} \frac{1-\cos(y_n t)}{y_n} Y_n, \quad t \in [0,1],$$

is well defined and both series converge absolutely and uniformly in $t \in [0,1]$. The process B is a fractional Brownian motion with Hurst index H.

For each $N \in \mathbb{N}$ we define the process

$$B_N(t) = \sum_{n=1}^{N} \frac{\sin(x_n t)}{x_n} X_n + \sum_{n=1}^{N} \frac{1-\cos(y_n t)}{y_n} Y_n, \quad t \in [0,1], \tag{2.1}$$



a) $H = 0.65$ b) $H = 0.90$

FIGURE 2. Approximation B_N of fractional Brownian motion.

then using the above-mentioned result from [4] we have

$$P(\lim_{N\to\infty} \sup_{t\in[0,1]} |B(t) - B_N(t)| = 0) = 1. \tag{2.2}$$

Since the functions involved in (2.1) have bounded derivatives, it is easy to prove the following result:

Theorem 2.1. *For all $N \in \mathbb{N}$ the approximating processes $(B_N(t))_{t\in[0,1]}$ are with probability* 1 *Lipschitz continuous.*

3. Fractional integrals and derivatives

Let $a, b \in \mathbb{R}$, $a < b$ and $f, g : \mathbb{R} \to \mathbb{R}$. We use notions and results about fractional calculus, from [11] and [13]:

$$f(a+) := \lim_{\delta\searrow 0} f(a+\delta), \quad f(b-) := \lim_{\delta\searrow 0} f(b-\delta),$$

$$f_{a+}(x) = \mathbb{I}_{(a,b)}(f(x) - f(a+)), \quad g_{b-}(x) = \mathbb{I}_{(a,b)}(g(x) - g(b-)).$$

For $f \in L_1(a,b)$ and $\alpha > 0$ the **left-** and **right-sided fractional Rieman-Liouville integral of f of order** α on (a,b) is given for a.e. x by

$$I^\alpha_{a+} f(x) = \frac{1}{\Gamma(\alpha)} \int_a^x (x-y)^{\alpha-1} f(y)dy,$$

$$I^\alpha_{b-} f(x) = \frac{(-1)^{-\alpha}}{\Gamma(\alpha)} \int_x^b (y-x)^{\alpha-1} f(y)dy.$$

Note that $(-1)^\alpha = e^{i\pi\alpha}$.

For $p > 1$ let $I^\alpha_{a+}(L_p(a,b))$, be the class of functions f which have the representation $f = I^\alpha_{a+}\Phi$, where $\Phi \in L_p(a,b)$, and let $I^\alpha_{b-}(L_p(a,b))$ be the class of functions g which have the representation $g = I_{b-}^{\alpha}\varphi$, where $\varphi \in L_p(a,b)$. If $0 < \alpha < 1$, then the function Φ, respectively φ, in the representations above agree a.s. with the **left-sided** and respectively **right-sided fractional derivative of order** α (in the Weyl representation)

$$\Phi(x) = D_{a+}^{\alpha} f(x) = \frac{1}{\Gamma(1-\alpha)} \left(\frac{f(x)}{(x-a)^{\alpha}} + \alpha \int_a^x \frac{f(x)-f(y)}{(x-y)^{\alpha+1}} dy \right) \mathbb{I}_{(a,b)}(x)$$

and

$$\varphi(x) = D_{b-}^{\alpha} g(x) = \frac{(-1)^{\alpha}}{\Gamma(1-\alpha)} \left(\frac{g(x)}{(b-x)^{\alpha}} + \alpha \int_x^b \frac{g(x)-g(y)}{(y-x)^{\alpha+1}} dy \right) \mathbb{I}_{(a,b)}(x).$$

The convergence at the singularity $y = x$ holds in the L_p-sense. For completeness we denote

$$D_{a+}^{0} f(x) = f(x), D_{b-}^{0} g(x) = g(x), D_{a+}^{1} f(x) = f'(x), D_{b-}^{1} g(x) = g'(x).$$

Let $0 \le \alpha \le 1$. The **fractional integral** of f with respect to g is defined as

$$\int_a^b f(x)dg(x) = (-1)^{\alpha} \int_a^b D_{a+}^{\alpha} f_{a+}(x) D_{b-}^{1-\alpha} g_{b-}(x) dx + f(a+)(g(b-) - g(a+))$$

if $f_{a+} \in I_{a+}^{\alpha}(L_p(a,b)), g_{b-} \in I_{b-}^{1-\alpha}(L_q(a,b))$ for $\frac{1}{p} + \frac{1}{q} \le 1$.

In our investigations we will take $p = q = 2$. If $0 \le \alpha < \frac{1}{2}$, then the above integral can be written as

$$\int_a^b f(x)dg(x) = (-1)^{\alpha} \int_a^b D_{a+}^{\alpha} f(x) D_{b-}^{1-\alpha} g_{b-}(x) dx \tag{3.1}$$

if $f \in I_{a+}^{\alpha}(L_2(a,b))$, $f(a+)$ exists, $g_{b-} \in I_{b-}^{1-\alpha}(L_2(a,b))$ (see [13]).

Without loss of generality we consider $0 < T \le 1$, because for arbitrary $T > 0$ we can rescale the time variable using the H-self similar property of the fractional Brownian motion meaning that $\left(B(ct)\right)_{t \ge 0}$ and $\left(c^H B(t)\right)_{t \ge 0}$ are equal in distribution for every $c > 0$.

We consider $\alpha > 1 - H$. $B_{T-} \in I_{T-}^{1-\alpha}(L_2(0,T))$ is fulfilled, since the fractional Brownian motion B is a.s. Hölder continuous with exponent $\gamma \in (0, H)$ (see [3]). Observe that $B_{N,T-} \in I_{T-}^{1-\alpha}(L_2(0,T))$ (for each $n \in \mathbb{N}$), which follows from the Lipschitz continuity property (see Theorem 2.1).

Let $G \in I^{\alpha}_{0+}(L_2(0,T))$ such that $G(0+)$ exists. It follows by (3.1) that we have the following *stochastic integrals*,

$$\int_0^T G(u)dB(u) = (-1)^{\alpha} \int_0^T D^{\alpha}_{0+}G(u)D^{1-\alpha}_{T-}B_{T-}(u)du$$

and

$$\int_0^T G(u)dB_N(u) = (-1)^{\alpha} \int_0^T D^{\alpha}_{0+}G(u)D^{1-\alpha}_{T-}B_{N,T-}(u)du.$$

4. Stochastic differential equations driven by fractional Brownian motion

Let $\Big(B(t)\Big)_{t\geq 0}$ be a fractional Brownian motion with Hurst parameter H such that $H > \frac{1}{2}$. We investigate stochastic differential equations of the form

$$\begin{aligned} dX(t) &= F(X(t),t)dt + G(X(t),t)dB(t), \\ X(t_0) &= X_0, \end{aligned} \tag{4.1}$$

where $t_0 \in (0,T]$, X_0 is a random vector in $\mathbb{R}^n$ and the random functions F and G satisfy with probability 1 the following conditions:

(C1) $F \in C(\mathbb{R}^n \times [0,T], \mathbb{R}^n), G \in C^1(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$;

(C2) for each $t \in [0,T]$ the functions $F(\cdot,t), \dfrac{\partial G(\cdot,t)}{\partial x^i}, \dfrac{\partial G(\cdot,t)}{\partial t}$ are locally Lipschitz for each $i \in \{1,\dots,n\}$.

We consider the pathwise auxiliary partial differential equation on $\mathbb{R}^n \times \mathbb{R} \times [0,T]$,

$$\begin{aligned} \frac{\partial K}{\partial z}(y,z,t) &= G(K(y,z,t),t), \\ K(Y_0,Z_0,t_0) &= X_0, \end{aligned} \tag{4.2}$$

where Y_0 is an arbitrary random vector in $\mathbb{R}^n$ and Z_0 an arbitrary random variable in $\mathbb{R}$. From the theory of differential equations it follows that with probability 1 there exists a local solution $K \in C^1(\mathbb{R}^n \times \mathbb{R} \times [0,T], \mathbb{R}^n)$ in a neighborhood V of (Y_0, Z_0, t_0) with partial derivatives being Lipschitz in the variable y and

$$\det \left(\frac{\partial K^i}{\partial y^j}(y,z,t) \right)_{1\leq i,j\leq n} \neq 0.$$

We have for $(x,y,t) \in V$

$$\frac{\partial^2 K}{\partial z^2}(y,z,t) = \sum_{j=1}^{n} \frac{\partial G}{\partial x^j}(K(y,z,t),t)G^j(K(y,z,t),t).$$

We also consider the pathwise differential equation (in matrix representation) on $[0,T]$,

$$dY(t) = \left(\frac{\partial K}{\partial y}(Y(t), B(t), t)\right)^{-1} \left[F(K(Y(t), B(t), t), t) - \frac{\partial K}{\partial t}(Y(t), B(t), t)\right] dt,$$
$$Y(t_0) = Y_0, \tag{4.3}$$

which has a unique local solution on a maximal interval $(t_0^1, t_0^2) \subseteq [0,T]$ with $t_0 \in (t_0^1, t_0^2)$ (see Theorem 6.1 from Appendix).

We apply the stochastic Itô formula (see [15, Theorem 5.10]) to the random function $Q(z,t) = K(Y(t), z, t)$ and the fractional Brownian motion B, we obtain

$$\begin{aligned}
&K(Y(t), B(t), t) - K(Y(t_0), B(t_0), t_0) \\
&= \sum_{j=1}^{n} \int_{t_0}^{t} \frac{\partial K}{\partial y^j}(Y(s), B(s), s) dY^j(s) + \int_{t_0}^{t} \frac{\partial K}{\partial z}(Y(s), B(s), s) dB(s) \\
&\quad + \int_{t_0}^{t} \frac{\partial K}{\partial t}(Y(s), B(s), s) ds \\
&= \int_{t_0}^{t} F(K(Y(s), B(s), s), s) ds + \int_{t_0}^{t} G(K(Y(s), B(s), s), s) dB(s).
\end{aligned}$$

Therefore, $X(t) := K(Y(t), B(t), t)$ satisfies

$$X(t) = X_0 + \int_{t_0}^{t} F(X(s), s) ds + \int_{t_0}^{t} G(X(s), s) dB(s).$$

Instead of the process $\Big(B(t)\Big)_{t\in[0,1]}$ we consider its approximations $\Big(B_N(t)\Big)_{t\in[0,1]}$ given in (2.1). For each $N \in \mathbb{N}$ we consider the pathwise differential equation (in matrix representation)

$$\begin{aligned}
dY_N(t) &= \left(\frac{\partial K}{\partial y}(Y_N(t), B_N(t), t)\right)^{-1} \Big[F(K(Y_N(t), B_N(t), t), t) \\
&\qquad - \frac{\partial K}{\partial t}(Y_N(t), B_N(t), t)\Big] dt \\
Y_N(t_0) &= Y_0,
\end{aligned} \tag{4.4}$$

which has a unique local solution Y_N on a maximal interval $(t_1, t_2) \subseteq (t_0^1, t_0^2)$ of existence which contains t_0 (see Theorem 6.2 from Appendix). Applying the stochastic Itô formula, to the random function $Q(z,t) = K(Y_N(t), z, t)$ and the process B_N we obtain

$$\begin{aligned}
&K(Y_N(t),B_N(t),t)-K(Y_N(t_0),B_N(t_0),t_0)\\
&=\sum_{j=1}^{n}\int_{t_0}^{t}\frac{\partial K}{\partial y^j}(Y_N(s),B_N(s),s)dY_N^j(s)+\int_{t_0}^{t}\frac{\partial K}{\partial z}(Y_N(s),B_N(s),s)dB_N(s)\\
&\qquad+\int_{t_0}^{t}\frac{\partial K}{\partial t}(Y_N(s),B_N(s),s)ds\\
&=\int_{t_0}^{t}F(K(Y_N(s),B_N(s),s),s)ds+\int_{t_0}^{t}G(K(Y_N(s),B_N(s),s),s)dB_N(s).
\end{aligned}$$

Therefore, $X_N(t):=K(Y_N(t),B_N(t),t)$ satisfies

$$X_N(t)=X_0+\int_{t_0}^{t}F(X_N(s),s)ds+\int_{t_0}^{t}G(X_N(s),s)dB_N(s),\quad t\in(t_1,t_2).$$

By Theorem 6.2 it follows that we have the following pathwise property:

$$\lim_{N\to\infty}\sup_{t\in(t_1,t_2)}\|Y_N(t)-Y(t)\|=0.$$

Then the continuity properties of K and (2.2) imply that for a.e. $\omega\in\Omega$ it holds

$$\lim_{N\to\infty}\sup_{t\in(t_1,t_2)}\|X_N(t)-X(t)\|=0.$$

By this we have proved that having the solutions and the approximations of the pathwise differential equation, we can obtain the solutions and approximations of the given stochastic differential equation.

Now we prove that having the solution X for the stochastic equation (4.1), we can construct the solution Y of the pathwise differential equation (4.3). This argument assures the uniqueness of the local solution of (4.1), since (4.3) has a unique global solution.

Let K be the solution of (4.2). It is known that in a neighborhood V of (Y_0,Z_0,t_0) the solution is invertible, i.e.

$$(y,z,t)\in V\longmapsto (K(y,z,t),z,t)\text{ has an inverse .}$$

We denote by L the mapping which satisfies

$$L(K(y,z,t),z,t)=y\text{ and }K(L(x,z,t),z,t)=x.$$

On the neighborhood V we have the matrix equality

$$\frac{\partial L}{\partial x}(x,z,t)=\left(\frac{\partial K}{\partial y}(L(x,z,t),z,t)\right)^{-1}\tag{4.5}$$

and by using (4.2) we get

$$\frac{\partial L}{\partial z}(x,z,t) = -\sum_{i=1}^{n} \frac{\partial L}{\partial x^i}(x,z,t) G^i(x,t)$$

and

$$\frac{\partial L}{\partial t}(x,z,t) = -\sum_{i=1}^{n} \frac{\partial L}{\partial x^i}(x,z,t) \frac{\partial K^i}{\partial t}(L(x,z,t),z,t).$$

Applying the stochastic Itô formula to the function $L(x,z,t)$ and the $\mathbb{R}^{n+1}$-valued process $(X(t),B(t))$ we obtain

$$\begin{aligned}
&L(X(t),B(t),t) - L(X(t_0),B(t_0),t_0)\\
&= \sum_{i=1}^{n} \int_{t_0}^{t} \frac{\partial L}{\partial x^i}(X(s),B(s),s)dX^i(s) + \int_{t_0}^{t} \frac{\partial L}{\partial z}(X(s),B(s),s)dB(s)\\
&\quad + \int_{t_0}^{t} \frac{\partial L}{\partial t}(X(s),B(s),s)ds\\
&= \sum_{i=1}^{n} \int_{t_0}^{t} \frac{\partial L}{\partial x^i}(X(s),B(s),s)F^i(X(s),s)ds\\
&\quad - \sum_{i=1}^{n} \int_{t_0}^{t} \frac{\partial L}{\partial x^i}(X(s),B(s),t)\frac{\partial K^i}{\partial t}(L(X(s),B(s),s),B(s),s)ds.
\end{aligned}$$

But $X(s) = K(L(X(s),B(s),s),B(s),s)$ and (4.5) holds, therefore

$$Y(t) := L(X(t),B(t),t)$$

is a local solution of the pathwise equation (4.3). Analogously we can prove that $Y_N(t) := L(X_N(t),B_N(t),t)$ is a local solution of the pathwise equation (4.4). By this, we get the main result of our paper.

Theorem 4.1. *Let B be a fractional Brownian motion approximated through the processes B_N as given in* (2.1) *and* (2.2)*. Let $F,G : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ be random functions satisfying with probability* 1 *the conditions* (C1) *and* (C2)*. Let $t_0 \in (0,T]$ be fixed. Then, each of the stochastic equations*

$$X(t) = X_0 + \int_{t_0}^{t} F(X(s),s)ds + \int_{t_0}^{t} G(X(s),s)dB(s),$$

$$X_N(t) = X_0 + \int_{t_0}^{t} F(X_N(s),s)ds + \int_{t_0}^{t} G(X_N(s),s)dB_N(s), \quad N \in \mathbb{N},$$

admits almost surely a unique local solution on a common interval (t_1, t_2) *(which is independent of* N *and contains* t_0*). Moreover, we have the following approximation result:*

$$P(\lim_{N\to\infty} \sup_{t\in(t_1,t_2)} \|X_N(t) - X(t)\| = 0) = 1.$$

5. Application

We consider the one-dimensional stochastic linear equation from finance mathematics, modeling the price S of a stock,

$$S(t) = S_0 + \int_0^t \mu(s)S(s)ds + \int_0^t \sigma(s)S(s)dB(s),$$

where $\Big(B(t)\Big)_{t\in[0,T]}$ is a fractional Brownian motion with Hurst index $H > \frac{1}{2}$, μ is the interest rate and σ the dispersion function.

It is known (see [7, p. 1022]) that this equation has the following unique solution:

$$S(t) = S_0 \exp\left\{ \int_0^t \mu(u)du + \int_0^t \sigma(u)dB(u) \right\} \text{ for all } t \in [0,T].$$

By the methods of the above section we approximate B through the processes B_N, via (2.1) and (2.2) and consider

$$S_N(t) = S_0 \exp\left\{ \int_0^t \mu(u)du + \int_0^t \sigma(u)dB_N(u) \right\} \text{ for all } t \in [0,T].$$

Using Theorem 4.1 it follows that

$$P(\lim_{N\to\infty} \sup_{t\in[0,T]} |S_N(t) - S(t)| = 0) = 1.$$

In the special case when μ and σ are constants, we have that the price of a stock is

$$S(t) = S_0 e^{\mu t + \sigma B(t)}$$

and we can simulate it by computer using

$$S_N(t) = S_0 e^{\mu t + \sigma B_N(t)}$$

as given in Figure 3.

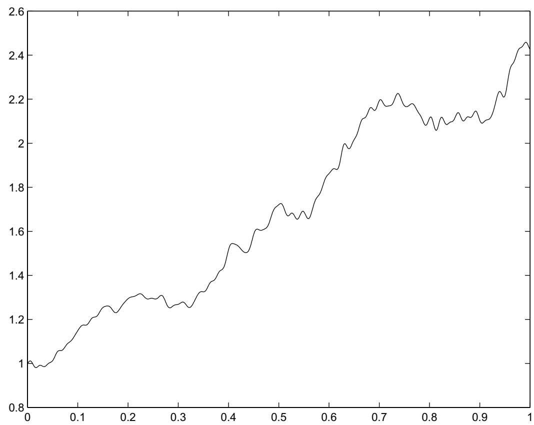


FIGURE 3. Approximated solution S_N (we took $H = 0.9$).

6. Appendix

We prove the existence of the local solution of a deterministic equation with locally Lipschitz function (in the version we need in our paper). We adapt the ideas from the proof of Theorem 1.4 in [9]. We give the proof here in order to make the proof of Theorem 6.2 more understandable.

In what follows $\|\cdot\|$ denotes the norm in $\mathbb{R}^n$.

Theorem 6.1. *Let $A : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n$ be such that for each $u \in \mathbb{R}^n$ the function $A(u,\cdot)$ is continuous and for any $c, T > 0$ we have*

$$\|A(x,t) - A(y,t)\| \le L(c,T)\|x-y\|$$

for all $x, y \in \mathbb{R}^n$ with $\|x\| \le c, \|y\| \le c$ and $t \in [0,T]$, where $L(c,T) > 0$ is the locally Lipschitz constant. We consider the equation

$$U(t) = U_0 + \int_{t_0}^{t} A(U(s),s)ds, \tag{6.1}$$

where $U_0 \in \mathbb{R}^n$ and $t_0 > 0$ fixed. Then equation (6.1) has a local solution, i.e., there exists a maximal interval $(t_1,t_2) \in [0,\infty)$ containing t_0 and a function $U : \mathbb{R}^n \times (t_1,t_2) \to \mathbb{R}^n$ such that (6.1) is satisfied for each $t \in (t_1,t_2)$.

Proof. For any $\tau > 0$ let $M(\tau) = \max\limits_{t\in[0,\tau+1]} \|A(0,t)\|$. We consider

$$\delta = \min\left\{1, \frac{\|U_0\|}{2\|U_0\|L(2\|U_0\|, t_0+1) + M(t_0)}, t_0\right\}.$$

We define the mapping $\mathcal{A} : C([t_0-\delta, t_0+\delta], \mathbb{R}^n) \to C([t_0-\delta, t_0+\delta], \mathbb{R}^n)$,

$$(\mathcal{A}U)(t) := U_0 + \int_{t_0}^{t} A(U(s), s)ds, \quad t \in [t_0-\delta, t_0+\delta].$$

We prove that $\mathcal{A}$ maps the ball $\mathcal{B}(0, R)$ of radius $R = 2\|U_0\|$ centered at 0 of the space $C([t_0-\delta, t_0+\delta], \mathbb{R}^n)$ into itself. For $U \in \mathcal{B}(0, R)$ and for each $t \in [t_0-\delta, t_0+\delta]$ we have the estimates

$$\begin{aligned} \|\mathcal{A}(U)(t)\| &\leq \|U_0\| + \left| \int_{t_0}^{t} \|A(U(s), s) - A(0, s)\| + \|A(0, s)\|\Big) ds \right| \\ &\leq \|U_0\| + (L(R, t_0+1)R + M(t_0))|t - t_0| \leq 2\|U_0\| = R. \end{aligned}$$

Therefore, $\mathcal{A}U \in \mathcal{B}(0, R)$. It is easy to verify that for each $U, V \in \mathcal{B}(0, R)$ and each $t \in [t_0-\delta, t_0+\delta]$ we have

$$\|\mathcal{A}(U)(t) - \mathcal{A}(V)(t)\| \leq L(R, t_0+1)|t - t_0| \sup_{t \in [t_0-\delta, t_0+\delta]} \|U(t) - V(t)\|.$$

For each $N \in \mathbb{N}$ we denote

$$\mathcal{A}^N = \underbrace{\mathcal{A} \circ \cdots \circ \mathcal{A}}_{N \text{ times}}.$$

From the definition of $\mathcal{A}$ it then follows for each $N \in \mathbb{N}$ and each $t \in [t_0-\delta, t_0+\delta]$ that

$$\|\mathcal{A}^N(U)(t) - \mathcal{A}^N(V)(t)\| \leq \frac{\Big(L(R, t_0+1)|t - t_0|\Big)^N}{N!} \sup_{t \in [t_0-\delta, t_0+\delta]} \|U(t) - V(t)\|.$$

Hence

$$\sup_{t \in [t_0-\delta, t_0+\delta]} \|\mathcal{A}^N(U)(t) - \mathcal{A}^N(V)(t)\| \leq \frac{\Big(L(R, t_0+1)\delta\Big)^N}{N!} \sup_{t \in [t_0-\delta, t_0+\delta]} \|U(t) - V(t)\|.$$

For N large enough we have $\dfrac{\Big(L(R, t_0+1)\delta\Big)^N}{N!} < 1$. By a well-known extension of the contraction principle it follows that $\mathcal{A}$ has a unique fixed point in $\mathcal{B}(0, R)$.

We have proved that there exists a solution U defined on the interval $[t_0-\delta, t_0+\delta]$ satisfying (6.1). This solution can be extended to the interval $[t_0-\delta^*, t_0+\delta^*]$ ($\delta^* > \delta$), where on $[t_0-\delta, t_0+\delta]$ we have the above solution U and for $t \geq t_0+\delta$ we use the above method to find a local solution for

$$U(t) = U(t_0+\delta) + \int_{t_0+\delta}^{t} A(U(s), s)ds,$$

and also for $t \le t_0 - \delta$ we use the above method to find a local solution for

$$U(t) = U(t_0 - \delta) + \int_{t_0-\delta}^{t} A(U(s), s)ds.$$

Moreover, δ^* depends only on $\delta, \|U(t_0 + \delta)\|, \|U(t_0 - \delta)\|, M(t_0 + \delta), M(t_0 - \delta)$. Hence, there exists a maximal interval (t_1, t_2) containing t_0 for the existence of the local solution U. $\square$

Theorem 6.2. *Let* $A : \mathbb{R}^{n+1} \times [0, T] \to \mathbb{R}^n$ *be such that for each* $(x, u) \in \mathbb{R}^{n+1}$ *the function* $A(x, u, \cdot)$ *is continuous and we have*

$$\|A(x, u, t) - A(y, v, t)\| \le L(c)(\|x - y\| + |u - v|)$$

for all $x, y \in \mathbb{R}^n$ *with* $\|x\| \le c, \|y\| \le c, |u| \le c, |v| \le c$ *and each* $t \in [0, T]$*, where* $L(c) > 0$ *is the locally Lipschitz constant. Let* $U_0 \in \mathbb{R}^n$ *and* $t_0 \in (0, T]$ *fixed. Assume that* $(v_N)_{N\in\mathbb{N}}$ *is a sequence from* $C[0, T]$ *which converges uniformly to* $v \in C[0, T]$*, i.e.,*

$$\lim_{N\to\infty} \sup_{t\in[0,T]} |v_N(t) - v(t)| = 0.$$

We consider the equations

$$U_N(t) = U_0 + \int_{t_0}^{t} A(U_N(s), v_N(s), s)ds, \quad N \in \mathbb{N} \tag{6.2}$$

and

$$U(t) = U_0 + \int_{t_0}^{t} A(U(s), v(s), s)ds. \tag{6.3}$$

The equations (6.2) *and* (6.3) *have local solutions, i.e., there exists a maximal interval* $(t_1, t_2) \subset [0, T]$ *(which does not depend on* N*) containing* t_0 *and functions* $U_N, U : \mathbb{R}^n \times (t_1, t_2) \to \mathbb{R}^n$ *such that* (6.2) *and* (6.3) *are satisfied for each* $t \in (t_1, t_2)$*. Moreover,*

$$\lim_{N\to\infty} \sup_{t\in(t_1,t_2)} \|U_N(t) - U(t)\| = 0.$$

Proof. For any $\tau > 0$ let $M = \max_{t\in[0,T]} \|A(0, 0, t)\|$. Since $(v_N)_{N\in\mathbb{N}}$ converges uniformly to v in $C[0, T]$, it follows that there exists $m > 0$ such that

$$\sup_{t\in[0,T]} |v_N(t)| + \sup_{t\in[0,T]} |v(t)| \le m \qquad \text{for each } N \in \mathbb{N}.$$

We consider

$$\delta = \min\left\{1, \frac{m}{(\|U_0\| + 2m)L(\|U_0\| + m) + M}, t_0, T - t_0\right\}.$$

We define the mapping $\mathcal{F}_N : C([t_0-\delta,t_0+\delta],\mathbb{R}^n) \to C([t_0-\delta,t_0+\delta],\mathbb{R}^n)$

$$(\mathcal{F}_N Y)(t) := U_0 + \int_{t_0}^{t} A(Y(s), v_N(s), s)ds, \quad t \in [t_0-\delta, t_0+\delta].$$

We prove that $\mathcal{F}_N$ maps the ball $\mathcal{B}(0,R)$ of radius $R = \|U_0\| + m$ centered at 0 of the space $C([t_0-\delta,t_0+\delta],\mathbb{R}^n)$ into itself. For $Y \in \mathcal{B}(0,R)$ and for each $t \in [t_0-\delta,t_0+\delta]$ we have the estimates

$$\begin{aligned}\|\mathcal{F}_N(Y)(t)\| &\le \|U_0\| + \left|\int_{t_0}^{t} \|A(Y(s),v_N(s),s) - A(0,0,s)\| + \|A(0,0,s)\|\Big)ds\right| \\ &\le \|U_0\| + (L(R)(R+m)+M)|t-t_0| \le \|U_0\| + m = R.\end{aligned}$$

Therefore, $\mathcal{F}_N Y \in \mathcal{B}(0,R)$. It is easy to verify that for each $Y, Z \in \mathcal{B}(0,R)$ and each $t \in [t_0-\delta,t_0+\delta]$ we have

$$\|\mathcal{F}_N(Y)(t) - \mathcal{F}_N(Z)(t)\| \le L(R)|t-t_0| \sup_{t\in[t_0-\delta,t_0+\delta]} \|Y(t)-Z(t)\|.$$

Using the contraction principle exactly as in the proof of Theorem 6.1, it follows that $\mathcal{F}_N$ has a unique fixed point in $\mathcal{B}(0,R)$, which is defined on $[t_0-\delta,t_0+\delta]$. This fixed point is the local solution U_N of (6.2). We observe that this interval of existence of the local solution U_N does not depend on N, and $U_N \in \mathcal{B}(0,R)$ for each $N \in \mathbb{N}$. Exactly in the same way we can prove that on the same interval $[t_0-\delta,t_0+\delta]$ there exists a solution $U \in \mathcal{B}(0,R)$ satisfying (6.3). Let $(t_1,t_2) \subset (0,T]$ be the maximal interval (which does not depend on N) containing t_0 such that (6.2) and (6.3) are satisfied for each $t \in (t_1,t_2)$ and there exists $R > 0$ (independent of N) such that $U_N, U \in \mathcal{B}(0,R)$. Then for large N we have

$$\begin{aligned}\|U_N(t) - U(t)\| &\le \left|\int_{t_0}^{t} \|A(U_N(s),v_N(s),s) - A(U(s),v(s),s)\|ds\right| \\ &\le \left|\int_{t_0}^{t} L(R)(\|U_N(s)-U(s)\| + \|v_N(s)-v(s)\|)ds\right|.\end{aligned}$$

By the Gronwall lemma we get

$$\sup_{t\in(t_1,t_2)} \|U_N(t)-U(t)\| \le \sup_{t\in(t_1,t_2)} \|v_N(t)-v(t)\| e^{L(R)(t_2-t_1)}.$$

Therefore,

$$\lim_{N\to\infty} \sup_{t\in(t_1,t_2)} \|U_N(t)-U(t)\| = 0. \qquad \square$$

References

[1] J. Bertoin, *Sur une intégrale pour les processus à α variation bornée*, Annals of Probability, **17** (1989), 1521–1535.

[2] L. Coutin and L. Decreusefond, *Stochastic differential equations driven by a fractional Brownian motion*, Tech. Report 97C004, ENST Paris, 1997.

[3] L. Decreusefond and A. S. Üstünel, *Stochastic analysis of the fractional Brownian motion*, Potential Analysis, **10** (1999), 177–214.

[4] K. Dzhaparidze and H. van Zanten, *A series expansion of fractional Brownian motion*, Probab. Theory Relat. Fields, **130** (2004), 39–55.

[5] M. Errami and F. Russo, *n-Covariation, generalized Dirichlet processes and calculus with respect to finite cubic variation process*, Stochastic Process. Appl., **104** (2003), 259–299.

[6] M. L. Kleptsyna, P. E. Kloeden, and V. V. Anh, *Existence and uniqueness theorems for FBM stochastic differential equations*, Information Transmission Problems, **34** (1998), 41–51.

[7] F. Klingenhöfer and M. Zähle, *Ordinary differential equations with fractal noise*, Proc. Amer. Math. Soc., **127** (1999), 1021–1028.

[8] D. Nualart and A. Rascanu, *Differential equations driven by fractional Brownian motion*, Collect. Math., **53** (2002), 55–81.

[9] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer Verlag, New York, 1983.

[10] F. Russo and P. Vallois, *Forward, backward and symmetric stochastic integration*, Probab. Theory Related Fields, **97** (1993), 403–421.

[11] S. G. Samko, A. A. Kilbas and O. I. Marichev, *Fractional Integrals and Derivatives. Theory and Applications*, Gordon and Breach, 1993.

[12] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1944.

[13] M. Zähle, *Integration with respect to fractal functions and stochastic calculus I*, Probab.Theory Relat. Fields, **111** (1998), 333–374.

[14] M. Zähle, *Ordinary differential equations with fractal noise*, Proc. Am. Math. Soc., **127** (1999), 333–374.

[15] M. Zähle, *Integration with respect to fractal functions and stochastic calculus II*, Math. Nachr., **225** (2001), 145–183.

Hannelore Lisei and Anna Soós
Babeş-Bolyai University
Faculty of Mathematics and Computer Science
Str. Kogălniceanu Nr. 1
400084 Cluj-Napoca
Romania
e-mail: hanne@math.ubbcluj.ro
asoos@math.ubbcluj.ro

Progress in Probability, Vol. 59, 243–259

Critical Exponents for Semilinear PDEs with Bounded Potentials

José Alfredo López-Mimbela and Nicolas Privault

Abstract. Using heat kernel estimates obtained in [18] and the Feynman-Kac formula, we investigate finite-time blow-up and stability of semilinear partial differential equations of the form $\frac{\partial w_t}{\partial t}(x) = \Delta w_t(x) - V(x)w_t(x) + v_t(x)G(w_t(x))$, $w_0(x) \geq 0$, $x \in \mathbb{R}^d$, where v and G are positive measurable functions subject to certain growth conditions, and V is a positive bounded potential. We recover the results of [19] and [14] by probabilistic arguments and in the quadratic decay case $V(x) \sim_{+\infty} a(1+|x|^2)^{-1}$, $a > 0$, we find two critical exponents $\beta_*(a)$, $\beta^*(a)$ with $0 < \beta_*(a) \leq \beta^*(a) < 2/d$, such that any nontrivial positive solution blows up in finite time if $0 < \beta < \beta_*(a)$, whereas if $\beta^*(a) < \beta$, then nontrivial positive global solutions may exist.

Mathematics Subject Classification (2000). 60H30, 35K55, 35K57, 35B35.

Keywords. Semilinear partial differential equations, Feynman-Kac representation, critical exponent, finite time blow-up, global solution.

1. Introduction

Consider a semilinear Cauchy problem of the form

$$\frac{\partial u_t}{\partial t}(x) = Au_t(x) + u_t^{1+\beta}(x), \qquad u_0(x) = \varphi(x), \qquad x \in \mathbb{R}^d, \tag{1.1}$$

where $\beta > 0$ is constant, $\varphi \geq 0$ is bounded and measurable, and A is the generator of a strong Markov process in $\mathbb{R}^d$. It is well known that, for any non-trivial initial value φ, there exists a number $T_\varphi \in (0,\infty]$ such that (1.1) has a unique mild solution u which is bounded on $[0,T] \times \mathbb{R}^d$ for any $0 < T < T_\varphi$, and if $T_\varphi < \infty$, then $\|u_t(\cdot)\|_\infty \to \infty$ as $t \uparrow T_\varphi$. When $T_\varphi = \infty$, the function u is called a global solution of (1.1), and when $T_\varphi < \infty$, one says that u blows up in finite time or that u is nonglobal.

The blow-up behaviors of semilinear equations of the above type have been intensely studied mainly in the analytic literature; see [1, 3, 7, 12, 13] for surveys.

In the case of the fractional power $A = -(-\Delta)^{\alpha/2}$ of the Laplacian, $0 < \alpha \le 2$, it has been proved that, for $d \le \alpha/\beta$, any nontrivial positive solution of (1.1) is nonglobal, whereas if $d > \alpha/\beta$, then the solution of (1.1) is global provided the initial value satisfies $\varphi \le \gamma G_r^\alpha$ for some $r > 0$ and some sufficiently small constant $\gamma > 0$, where G_r^α, $r > 0$, are the transition densities of the stable motion with generator $-(-\Delta)^{\alpha/2}$, see [2, 4, 10, 11, 15].

Critical exponents for blow-up of the semilinear equation

$$\frac{\partial u_t}{\partial t}(x) = \Delta u_t(x) - V(x)u_t(x) + u_t^{1+\beta}(x), \quad u_0(x) = \varphi(x), \quad x \in \mathbb{R}^d, \tag{1.2}$$

where $\varphi \ge 0$ and V is a bounded potential, have been studied in [14, 18, 19], where it is proved that if $d \ge 3$ and

$$0 \le V(x) \le \frac{a}{1 + |x|^b}, \quad x \in \mathbb{R}^d, \tag{1.3}$$

for some $a > 0$ and $b \in [2, \infty)$, then $b > 2$ implies finite time blow-up of (1.2) for all $0 < \beta < 2/d$, whereas if $b = 2$, then there exists $\beta_*(a) < 2/d$ such that blow-up occurs if $0 < \beta < \beta_*(a)$. It is also proved that if

$$V(x) \ge \frac{a}{1 + |x|^b}, \quad x \in \mathbb{R}^d, \tag{1.4}$$

for some $a > 0$ and $0 \le b < 2$, then (1.2) admits a global solution for all $\beta > 0$ and all non-negative initial values satisfying $\varphi(x) \le c/(1 + |x|^\sigma)$ for a sufficiently small constant $c > 0$ and all σ obeying $\sigma \ge b/\beta$.

In this note we give conditions for finite-time blow-up and for existence of nontrivial global solutions of the semilinear problem

$$\frac{\partial u_t}{\partial t}(x) = \Delta u_t(x) - V(x)u_t(x) + v_t(x)G(u_t(x)), \quad u_0(x) = \varphi(x), \quad x \in \mathbb{R}^d, \tag{1.5}$$

where V, φ are as above, and v, G are positive measurable functions subject to certain growth conditions. Using heat kernel estimates obtained in [18] and the Feynman-Kac representation of (1.5) we prove that, for dimensions $d \ge 3$, condition (1.3) with $b > 2$ entails finite time blow-up of any nontrivial positive solution of (1.5) provided

$$G(z) \ge \kappa z^{1+\beta}, \quad z > 0 \quad \text{and} \quad v_t(x) \ge t^\zeta \mathbf{1}_{B_{t^{1/2}}}(x), \quad (x, t) \in \mathbb{R}^d \times \mathbb{R}_+,$$

where $\kappa > 0$ and β, ζ are positive constants satisfying $0 < \beta < 2(1 + \zeta)/d$. (Here and in the sequel, $B_r(x)$ denotes the open ball of radius r centered at x).

We also prove that Eq. (1.5) admits nontrivial global solutions if (1.4) holds with $b < 2$ and $v_t(x)G(z) \le \kappa t^\zeta z^{1+\beta}$, $t \ge 0$, $z \ge 0$, for some positive constants κ, ζ and β.

As to the critical value $b = 2$, we investigate Equation (1.2) with a potential satisfying either (1.3) or (1.4), and with more general nonlinearities. We prove that, in dimensions $d \ge 3$, there exist critical exponents $\beta_*(a)$, $\beta^*(a)$, both decreasing in $a > 0$, given by

$$0 < \beta_*(a) := \frac{2(1+\zeta) - 4ac}{d + 2ac} \leq \beta^*(a) := \frac{2(1+\zeta)}{d + \min(1, a(d+4)^{-2}/64)} < \frac{2(1+\zeta)}{d},$$

where $c > 0$ is independent of a, and such that

a) if $0 \leq V(x) \leq \dfrac{a}{1+|x|^2}$, then (1.2) blows up in finite time provided $0 < \beta < \beta_*(a)$;
b) if $V(x) \geq \dfrac{a}{1+|x|^2}$, then (1.2) admits a global solution for all $\beta > \beta^*(a)$.

We remark that the blow-up behavior of (1.2) with potentials of the class we are considering here remains unknown when $\beta_*(a) \leq \beta \leq \beta^*(a)$, but notice that in the (unbounded) case $V(x) = a|x|^{-2}$, it can be deduced from [1], [8] and [5] that (1.2) admits a unique critical exponent $\beta(a) < 2/d$, given by

$$\beta(a) = \frac{2}{1 + d/2 + \sqrt{a + (d-2)^2/4}}.$$

Namely, if $V(x) = a|x|^{-2}$, then no global nontrivial solution of (1.2) exists if $\beta < \beta(a)$, whereas global solutions exist if $\beta(a) < \beta$. However, the approaches of the papers quoted above are specially suitable for the potential $V(x) = a|x|^{-2}$ and do not apply to our potentials, which are bounded on $\mathbb{R}^d$ in the subcritical case.

In the case of the one-dimensional equation

$$\frac{\partial u_t}{\partial t}(x) = -(-\Delta)^{\alpha/2} u_t(x) - V(x)u_t(x) + \kappa t^\zeta G(u_t(x)), \quad u_0(x) = \varphi(x), \quad x \in \mathbb{R}, \tag{1.6}$$

where $G(z)$ satisfies a suitable growth condition with respect to $z^{1+\beta}$, we show that, for every $\alpha \in (1,2]$ and $\zeta \geq 0$, any nontrivial solution of (1.6) blows up in finite time whenever $0 < \beta < 1 + \alpha\zeta$ and $V : \mathbb{R} \to \mathbb{R}_+$ is integrable. The same happens when $\beta = 1 + \alpha\zeta$ and the L^1-norm of V is sufficiently small. We were not able to investigate here the blow-up properties of (1.6) in the general case $d \geq 1$. From the perspective of our present methods, such investigation requires to derive sharp heat kernel estimates for the operator $\Delta_\alpha - V$, which is a topic of current research.

Let us remark that the heat kernel bounds from [18] play a major role in our arguments. In Section 2 we briefly recall such estimates, and derive some other ones that we will need in the sequel. These estimates are used to obtain semigroup bounds in Section 3. In Section 4 we investigate finite time blow-up of solutions using the Feynman-Kac approach developed in [2] (see also [9]). Section 5 is devoted to proving results on existence of global solutions.

We end this section by introducing some notation and basic facts we shall need.

Let $\Delta_\alpha = -(-\Delta)^{\alpha/2}$ denote the fractional power of the d-dimensional Laplacian, $0 < \alpha \leq 2$. We write $(S_t^\alpha)_{t\geq 0}$ for the semigroup generated by $\Delta_\alpha - V$, i.e.,

$$S_t^\alpha \varphi(y) = \int_{\mathbb{R}^d} \varphi(x) p_t^\alpha(x,y) dx = f_t(y),$$

where f_t denotes the solution of

$$\frac{\partial f_t}{\partial t}(x) = \Delta_\alpha f_t(x) - V(x) f_t(x), \qquad f_0(x) = \varphi(x),$$

and $p_t^\alpha(x,y)$, $t > 0$, are the transition densities of the Markov process in $\mathbb{R}^d$ having $\Delta_\alpha - V$ as its generator. Recall that from the Feynman-Kac formula we have

$$p_t^\alpha(x,y) = G_t^\alpha(x-y) E_x \left[\exp\left(- \int_0^t V(W_s^\alpha)\, ds \right) \Big| W_t^\alpha = y \right], \tag{1.7}$$

where $(W_s^\alpha)_{s\in\mathbb{R}_+}$ is a symmetric α-stable motion, and G_t^α, $t > 0$ are the corresponding α-stable transition densities. In case $\alpha = 2$ we will omit the index α and write

$$G_t(x) = \frac{1}{(4\pi t)^{d/2}} e^{-\frac{|x|^2}{4t}}, \qquad x \in \mathbb{R}^d, \quad t > 0,$$

for the standard Gaussian kernel, and

$$p_t(x,y) = G_t(x-y) E_x \left[\exp\left(- \int_0^t V(W_s)\, ds \right) \Big| W_t = y \right], \qquad t > 0,$$

where $(W_s)_{s\in\mathbb{R}_+}$ is a Brownian motion.

2. Heat kernel bounds of $\Delta - V$

Recall that from Theorem 1.1 in [18] we have:

Theorem 2.1. *Let $d \geq 3$, $b \geq 0$, $a > 0$, and assume that*

$$V(x) \geq \frac{a}{1+|x|^b}, \qquad x \in \mathbb{R}^d.$$

There exist constants $c_1, c_2, c_3 > 0$, and $\alpha_1(a) > 0$, such that for all $x, y \in \mathbb{R}^d$ and $t > 0$ there holds

$$p_t(x,y) \leq \begin{cases} c_2 G_t(c_3(x-y)) \exp\left(-c_1 \left(\frac{t^{1/2}}{1+|x|^{b/2}} \right)^{1-b/2} - c_1 \left(\frac{t^{1/2}}{1+|y|^{b/2}} \right)^{1-b/2} \right) & \text{if } b < 2, \\ c_2 G_t(c_3(x-y)) \max\left(\frac{t^{1/2}}{1+|x|}, 1 \right)^{-\alpha_1(a)} \max\left(\frac{t^{1/2}}{1+|y|}, 1 \right)^{-\alpha_1(a)} & \text{if } b = 2, \\ c_2 G_t(c_3(x-y)) & \text{if } b > 2. \end{cases}$$

We also recall the following estimates, cf. Theorem 1.2 in [18].

Theorem 2.2. *Let $d \geq 3$ and assume that, for some $b \geq 0$ and $a > 0$,*

$$0 \leq V(x) \leq \frac{a}{1+|x|^b}, \qquad x \in \mathbb{R}^d. \tag{2.1}$$

There exist constants $c_4, c_5, c_6 > 0$, and $\alpha_2(a) > 0$, such that for all $t > 0$ and $x, y \in \mathbb{R}^d$ there holds

$$p_t(x,y) \geq \begin{cases} c_6 e^{-2c_5 t} G_t(c_4(x-y)) & \text{if } b < 2, \\ c_6 t^{-\alpha_2(a)} G_t(c_4(x-y)) & \text{if } b = 2, \\ c_6 G_t(c_4(x-y)) & \text{if } b > 2. \end{cases}$$

Remark 2.3. Notice that from Proposition 2.1 of [17] we have

$$\alpha_1(a) = \min(1, a(d+4)^{-2}/64), \qquad a > 0.$$

Moreover, from the arguments in [18], pp. 391–392, it follows that $\alpha_2 = ca$ for some $c > 0$ independent of a.

Let $B_r \subset \mathbb{R}^d$ denote the open ball of radius $r > 0$, centered at the origin. Notice that, under (2.1), Lemma 4.5 and Lemma 5.1 of [18] imply the more precise statement: for $t \geq 1$ and $x, y \in \mathbb{R}^d$,

$$p_t(x,y) \geq \begin{cases} c_6 e^{-2c_5 t} \mathbf{1}_{B_{a_1 t^{1/2}}}(x) \mathbf{1}_{B_{a_1 t^{1/2}}}(y), & \text{if } 0 \leq b < 2, \\ c_6 t^{-\alpha_2(a) - d/2} \mathbf{1}_{B_{a_2 t^{1/2}}}(x) \mathbf{1}_{B_{a_2 t^{1/2}}}(y), & \text{if } b = 2, \end{cases}$$

where c_5, c_6, a_1, a_2 are positive constants and $\alpha_2(a) = ca$ is a linear function of a.

We complete the above results with the following estimate, which yields an extension of Theorem 2.2 to the case $\alpha \in (1,2]$, though only in dimension $d = 1$.

Theorem 2.4. *Let $d = 1$ and $\alpha \in (1,2]$, and assume that $V(x)$ is integrable on $\mathbb{R}$. Then, for all $x, y \in \mathbb{R}$,*

$$p_t^\alpha(x,y) \geq e^{-Ct^{1-1/\alpha}} G_t^\alpha(x-y) \mathbf{1}_{B_{t^{1/\alpha}}}(x) \mathbf{1}_{B_{t^{1/\alpha}}}(y), \quad t > 0,$$

where $C > 0$ is a constant.

Proof. Using (1.7) and Jensen's inequality we have

$$p_t^\alpha(x,y) \geq G_t^\alpha(x-y) \exp\left(-E_x\left[\int_0^t V(W_s^\alpha)\, ds \,\middle|\, W_t^\alpha = y\right]\right).$$

From the scaling property of stable densities we obtain, for $y \in B_{t^{1/\alpha}}$ and $x \in B_{t^{1/\alpha}}$,

$$
\begin{aligned}
&\frac{G_s^\alpha(z-x)G_{t-s}^\alpha(z-y)}{G_t^\alpha(y-x)} \\
&= \frac{s^{-1/\alpha}(t-s)^{-1/\alpha}G_1^\alpha(s^{-1/\alpha}(z-x))G_1^\alpha((t-s)^{-1/\alpha}(z-y))}{t^{-1/\alpha}G_1^\alpha(t^{-1/\alpha}(y-x))} \\
&\leq C_\alpha \frac{s^{-1/\alpha}(t-s)^{-1/\alpha}}{t^{-1/\alpha}}, \qquad 0 < s < t,
\end{aligned}
$$

for some $C_\alpha > 0$. Hence

$$
\begin{aligned}
E_x\left[\int_0^t V(W_s^\alpha)\,ds \Big| W_t^\alpha = y\right] &= \int_{\mathbb{R}} \int_0^t V(z) \frac{G_s^\alpha(z-x)G_{t-s}^\alpha(z-y)}{G_t^\alpha(y-x)}\,dz\,ds \\
&\leq C_\alpha \int_{\mathbb{R}} V(z)\,dz \int_0^t \frac{s^{-1/\alpha}(t-s)^{-1/\alpha}}{t^{-1/\alpha}}\,ds \\
&= C_\alpha t^{1-1/\alpha} \int_{\mathbb{R}} V(z)\,dz \int_0^1 s^{-1/\alpha}(1-s)^{-1/\alpha}\,ds. \qquad (2.2)
\end{aligned}
$$

□

3. Semigroup bounds

In this section we establish some bounds for the semigroup $(S_t)_{t\in\mathbb{R}_+}$ of generator $\Delta - V$. The following proposition will be used in the proof of Theorem 5.2.

Proposition 3.1. *Let $a_1, a_2, \sigma > 0$ and $0 \leq b \leq 2$, and assume that*

$$
V(x) \geq \frac{a_1}{1+|x|^b} \qquad \textit{and} \qquad 0 \leq \varphi(x) \leq \frac{a_2}{1+|x|^\sigma}, \qquad x \in \mathbb{R}^d.
$$

i) *If $b < 2$, then for all $\varepsilon \in (0,1)$ we have*

$$
\|S_t\varphi\|_\infty \leq c_\varepsilon t^{-\sigma(1-\varepsilon)/b}, \qquad t > 0,
$$

for some $c_\varepsilon > 0$.

ii) *If $b = 2$, then for all $\varepsilon \in (0,1)$ there exists $c_\varepsilon > 0$ such that*

$$
\|S_t\varphi\|_\infty \leq c_\varepsilon t^{-(1-\varepsilon)\alpha_1(a_1)-d/2}, \qquad t > 0,
$$

provided $\sigma > d$.

Proof. i) If $b<2$, applying Theorem 2.1 we obtain

$$
\begin{aligned}
S_t\varphi(y) & \\
= & \int_{\mathbb{R}^d}\varphi(x)p_t(x,y)dx \\
\leq & \; c_2\int_{\mathbb{R}^d}\varphi(x)\exp\left(-c_1\left(\frac{t^{1/2}}{1+|x|^{b/2}}\right)^{1-b/2}\right)G_t(c_3(x-y))dx \\
\leq & \; c_2\exp\left(-c_1\left(\frac{t^{1/2}}{1+t^{(1-\varepsilon)/2}}\right)^{1-b/2}\right)\int_{\{|x|\leq t^{(1-\varepsilon)/b}\}}\varphi(x)G_t(c_3(x-y))dx \\
& +c_2\int_{\{|x|>t^{(1-\varepsilon)/b}\}}\varphi(x)G_t(c_3(x-y))dx,
\end{aligned}
$$

hence

$$
S_t\varphi(y)\leq a_2\exp\left(-c_1\left(\frac{t^{1/2}}{1+t^{(1-\varepsilon)/2}}\right)^{1-b/2}\right)+\frac{a_2c_2}{1+t^{(1-\varepsilon)\sigma/b}}.
$$

ii) Let now $b=2$ and $\varepsilon\in(0,1)$. From Theorem 2.1 we know that

$$
\begin{aligned}
S_t\varphi(y) & \\
\leq & \; c_2\int\varphi(x)\max\left(\frac{t^{1/2}}{1+|x|},1\right)^{-\alpha_1(a_1)}\max\left(\frac{t^{1/2}}{1+|y|},1\right)^{-\alpha_1(a_1)}G_t(c_3(x-y))dx \\
\leq & \; c_2\int_{\{|x|<t^{\varepsilon/2}\}}\varphi(x)\max\left(\frac{t^{1/2}}{1+|x|},1\right)^{-\alpha_1(a_1)}G_t(c_3(x-y))dx \\
& +c_2\int_{\{|x|>t^{\varepsilon/2}\}}\varphi(x)\max\left(\frac{t^{1/2}}{1+|x|},1\right)^{-\alpha_1(a_1)}G_t(c_3(x-y))dx \\
\leq & \; c_2\int_{\{|x|<t^{\varepsilon/2}\}}\varphi(x)\left(\frac{t^{1/2}}{1+t^{\varepsilon/2}}\right)^{-\alpha_1(a_1)}G_t(c_3(x-y))dx \\
& +c_2\int_{\{|x|>t^{\varepsilon/2}\}}\varphi(x)G_t(c_3(x-y))dx \\
\leq & \; c_2t^{-(1-\varepsilon)\alpha_1(a_1)/2}\int_{\{|x|<t^{\varepsilon/2}\}}\varphi(x)G_t(c_3(x-y))dx \\
& +\frac{c_2}{(4\pi)^{d/2}}t^{-d/2}\int_{\{|x|>t^{\varepsilon/2}\}}\varphi(x)dx \\
\leq & \; \frac{c_2}{(4\pi)^{d/2}}t^{-(1-\varepsilon)\alpha_1(a_1)/2-d/2}\int_{\{|x|<t^{\varepsilon/2}\}}\varphi(x)dx+c_7t^{-(\sigma-d)\varepsilon/2-d/2}.
\end{aligned}
$$

Hence for some $c_\varepsilon>0$ we have

$$
S_t\varphi(y)\leq c_\varepsilon t^{-(1-\varepsilon)\alpha_1(a_1)/2-d/2},\qquad y\in\mathbb{R}^d,\ t>1,
$$

provided $(1-\varepsilon)\alpha_1(a_1)\leq(\sigma-d)\varepsilon$. □

The following lemma will be used in the proof of Theorem 4.1.

Lemma 3.2. *Let $d \geq 3$, $b \geq 2$, and let $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ be bounded and measurable. Assume that*

$$0 \leq V(x) \leq \frac{a}{1+|x|^b}.$$

Then, for all $t \geq 1$ and $y \in \mathbb{R}^d$ we have

$$S_t\varphi(y) \geq c_0 t^{-\alpha_2 - d/2} \mathbf{1}_{B_{t^{1/2}}}(y) \int_{B_{t^{1/2}}} \varphi(x)\, dx,$$

where $\alpha_2 = 0$ if $b > 2$, and $\alpha_2(a) = ca$ for some $c > 0$ when $b = 2$.

Proof. Let $y \in B_{t^{1/2}}$. Due to Theorem 2.2 and self-similarity of Gaussian densities we have

$$\begin{aligned} S_t\varphi(y) &= \int_{\mathbb{R}^d} \varphi(x) p_t(x,y)\, dx \\ &\geq c_2 t^{-\alpha_2(a)} \int_{B_{t^{1/2}}} \varphi(x) G_t(c_4(x-y))\, dx \\ &\geq c_2 t^{-\alpha_2(a) - d/2} \int_{B_{t^{1/2}}} \varphi(x) G_1(c_4 t^{-1/2}(x-y))\, dx \\ &\geq c_0 t^{-\alpha_2(a) - d/2} \int_{B_{t^{1/2}}} \varphi(x)\, dx. \end{aligned}$$

□

The next lemma, which will be needed in the proof of Theorem 4.1 below, provides lower bounds on certain balls for the distributions of the bridges of the Markov process $(X_t)_{t\in\mathbb{R}_+}$ generated by $\Delta - V$.

Lemma 3.3. *Assume that $d \geq 3$ and let $(X_t)_{t\in\mathbb{R}_+}$ denote the Markov process with generator $\Delta - V$. If for some $b \geq 2$,*

$$0 \leq V(x) \leq \frac{a}{1+|x|^b}, \qquad x \in \mathbb{R}^d,$$

then there exists $c_8 > 0$ such that for all $t \geq 2$, $y \in B_{t^{1/2}}$, $x \in B_1$ and $s \in [1, t/2]$,

$$\mathbb{P}_x(X_s \in B_{s^{1/2}} \mid X_t = y) \geq c_8 t^{-2\alpha_2(a)},$$

where $\alpha_2(a) = 0$ when $b > 2$ and $\alpha_2(a) = ca$ when $b = 2$.

Proof. Since $V(x) \geq 0$, the Feynman-Kac formula (1.7) yields $p_t(x,y) \leq G_t(y-x)$, $t > 0$, $x,y \in \mathbb{R}^d$. An application of Theorem 2.2 and of the Markov property of

$(X_s)_{s\in\mathbb{R}_+}$ gives

$$\begin{aligned}
&\mathbb{P}_x(X_s \in B_{s^{1/2}} \mid X_t = y) \\
&\geq \int_{B_{s^{1/2}}} \frac{p_{t-s}(y,z)p_s(z,x)}{p_t(y,x)}\,dz \\
&= \frac{1}{c_6^2 s^{\alpha_2(a)}(t-s)^{\alpha_2(a)}} \int_{B_{s^{1/2}}} \frac{G_{t-s}(c_4(y-z))G_s(c_4(z-x))}{G_t(c_4(y-x))}\,dz \\
&\geq c_8 t^{-2\alpha_2(a)},
\end{aligned}$$

where we used Lemma 2.2 of [2] to obtain the last inequality. □

We conclude this section with the following lemma, which will be used in the proof of Theorem 5.2.

Lemma 3.4. *Let $d \geq 3$ and $V(x) \geq 0$, $x \in \mathbb{R}^d$. Assume that*

$$V(x) \geq \frac{a}{1+|x|^b}$$

holds for all $|x|$ greater than some $r_0 > 0$, where $a > 0$ and $0 \leq b < 2$. There exists $\gamma > 0$ such that for all bounded measurable $D \subset \mathbb{R}^d$,

$$S_t \mathbf{1}_D(x) \leq c_D t^{-(1+\gamma)}, \quad x \in \mathbb{R}^d, \tag{3.1}$$

for all sufficiently large t, where c_D does not depend on x and t.

Proof. By Theorem 2.1 we have

$$p_t(x,y) \leq c_2 G_t(c_3(x-y)) \exp\left(-c_1\left(\left(\frac{t}{1+|x|^b}\right)^{c_4} + \left(\frac{t}{1+|y|^b}\right)^{c_4}\right)\right)$$

for certain constants $c_1, c_2, c_3, c_4 > 0$. Condition (3.1) is obviously fulfilled for any positive γ if $b = 0$, hence let us assume that $0 < b < 2$. For any bounded measurable $D \subset \mathbb{R}^d$ we have, provided $t > \|D\|^2 := \sup_{y\in D}\|y\|^2$,

$$\begin{aligned}
S_t\mathbf{1}_D(x) &\leq c_2 \int_D G_t(c_3(x-y)) e^{-c\left(\frac{t}{1+|y|^b}\right)^{c_4}}\,dy \\
&\leq \frac{c_2}{(4\pi t)^{d/2}} \int_D dy \\
&\leq c_D t^{-(1+\gamma)},
\end{aligned}$$

with $\gamma = (d-2)/2 > 0$. □

4. Explosion in subcritical dimensions

Recall that if u_t, v_t respectively solve

$$\frac{\partial u_t}{\partial t}(y) = \Delta u_t(y) + \zeta_t(y)u_t(y), \qquad \frac{\partial v_t}{\partial t}(y) = \Delta v_t(y) + \xi_t(y)v_t(y),$$

with $u_0 \geq v_0$ and $\zeta_t \geq \xi_t$ for all $t \geq 0$, then $u_t \geq v_t$, $t \geq 0$. In particular, if $\varphi \geq 0$ is bounded and measurable, and if u_t is a subsolution of

$$\frac{\partial w_t}{\partial t}(y) = \Delta w_t(y) + \kappa w_t^{1+\beta}(y), \qquad w_0 = \varphi, \tag{4.1}$$

where $\kappa, \beta > 0$, then any solution of

$$\frac{\partial v_t}{\partial t}(y) = \Delta v_t(y) + \kappa u_t^{\beta}(y) v_t(y), \qquad v_0 = \varphi,$$

remains a subsolution of (4.1).

Theorem 4.1. *Let $d \geq 3$, $b \geq 2$, $\beta > 0$ and $a > 0$, and assume that*

$$0 \leq V(x) \leq \frac{a}{1+|x|^b}, \qquad x \in \mathbb{R}^d.$$

Let $G : \mathbb{R}_+ \to \mathbb{R}_+$ be such that

$$\frac{G(z)}{z} \geq \kappa z^{\beta}, \qquad z > 0, \tag{4.2}$$

for some $\kappa > 0$. Let $v : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ be a measurable function satisfying

$$v_t(x) \geq t^{\zeta} \mathbf{1}_{B_{t^{1/2}}}(x)$$

for some $\zeta > 0$. Consider the semilinear equation

$$\frac{\partial u_t(x)}{\partial t} = \Delta u_t(x) - V(x) u_t(x) + v_t(x) G(u_t(x)), \quad u_0(x) = \varphi(x), \quad x \in \mathbb{R}^d, \tag{4.3}$$

where $\varphi \geq 0$ is bounded and measurable.

a) *If $b > 2$ and*

$$0 < \beta < \frac{2(1+\zeta)}{d},$$

then any nontrivial positive solution of (4.3) blows up in finite time.

b) *If $b = 2$ and*

$$0 < \beta < \beta_*(a) := \frac{1+\zeta-2ac}{ac+d/2} < \frac{2(1+\zeta)}{d},$$

where $2ac < 1+\zeta$ and $c > 0$ is given in Remark 2.3, then any nontrivial positive solution of (4.3) blows up in finite time.

Proof. Let g_t denote the mild solution of

$$\frac{\partial g_t}{\partial t}(x) = \Delta g_t(x) - V(x) g_t(x) + v_t(x) \frac{G(f_t(x))}{f_t(x)} g_t(x), \quad g_0(x) = \varphi(x),$$

where $f_t = S_t \varphi$ satisfies

$$\frac{\partial f_t}{\partial t}(x) = \Delta f_t(x) - V(x) f_t(x), \quad f_0(x) = \varphi(x).$$

By the Feynman-Kac formula (1.7) we have

$$g_t(y) = \int_{\mathbb{R}^d} \varphi(x) p_t(x,y) E_x \left[\exp \int_0^t v_s(X_s) \frac{G(f_s(X_s))}{f_s(X_s)} \, ds \,\middle|\, X_t = y \right] dx.$$

Let $\alpha_2(a) = 0$ if $b > 2$, and $\alpha_2(a) = ca$ if $b = 2$. Then, for $y \in B_{t^{1/2}}$, and for certain positive constants K_1, K_2, K_3, we have by Lemma 3.2 that

$$
\begin{aligned}
&g_t(y)\\
&\quad \geq \int_{\mathbb{R}^d} \varphi(x) p_t(x,y) E_x \left[\exp K_1 \int_0^t v_s(X_s)(f_s(X_s))^\beta ds \,\middle|\, X_t = y \right] dx\\
&\quad \geq \int_{\mathbb{R}^d} \varphi(x) p_t(x,y) E_x \left[\exp \int_1^{t/2} K_2 s^{\zeta - d\beta/2 - \beta\alpha_2(a)} \mathbf{1}_{B_{s^{1/2}}}(X_s)\, ds \,\middle|\, X_t = y \right] dx\\
&\quad \geq \int_{\mathbb{R}^d} \varphi(x) p_t(x,y) \exp \left(K_2 \int_1^{t/2} s^{\zeta - d\beta/2 - \beta\alpha_2(a)} \mathbb{P}_x \left(X_s \in B_{s^{1/2}} | X_t = y \right) ds \right) dx\\
&\quad \geq \int_{\mathbb{R}^d} \varphi(x) p_t(x,y) \exp \left(K_3 t^{-2\alpha_2(a)} \int_1^{t/2} s^{\zeta - d\beta/2 - \beta\alpha_2(a)}\, ds \right) dx\\
&\quad \geq \int_{\mathbb{R}^d} \varphi(x) p_t(x,y) dx \exp \left(K_4 t^{\zeta - d\beta/2 - (\beta+2)\alpha_2(a) + 1} \right),
\end{aligned}
$$

where we used Lemma 3.3 to obtain the fourth inequality. The above argument shows that g eventually grows to $+\infty$ uniformly on the unit ball B_1 provided

$$\zeta - d\beta/2 - (\beta + 2)\alpha_2(a) > -1.$$

This condition is satisfied for all $0 < \beta < 2(1+\zeta)/d$ if $b > 2$, and for all $0 < \beta < \beta_*(a)$ if $b = 2$. Since g is subsolution of (4.3), the comparison result recalled at the beginning of this section shows that the solution u_t of (4.3) also grows to $+\infty$ uniformly on B_1. A well-known argument [6] involving Condition (4.2) then shows blow-up of (4.3). For the sake of completeness we include this argument here. Given $t_0 \geq 1$, let $\tilde{u}_t = u_{t+t_0}$ and $K(t_0) = \inf_{x \in B_1} u_{t_0}(x)$. The mild solution of (4.3) is given by

$$\tilde{u}_t(x) = \int_{\mathbb{R}^d} p_t(x,y) \tilde{u}_0(y)\, dy + \int_0^t \int_{\mathbb{R}^d} p_{t-s}(x,y) v_{s+t_0}(y) G(\tilde{u}_s(y))\, dy\, ds.$$

Thus, for all $t \in (1,2]$ and $x \in B_1$ we get from Theorem 2.2:

$$
\begin{aligned}
&\tilde{u}_t(x)\\
&\quad \geq \int_{B_1} p_t(x,y) \tilde{u}_0(y)\, dy + \kappa \int_0^t s^\zeta \int_{B_1} p_{t-s}(x,y) \tilde{u}_s^{1+\beta}(y)\, dy\, ds\\
&\quad \geq c_6 K(t_0) \int_{B_1} G_t(c_4(x-y))\, dy + \kappa c_6 \int_0^t s^\zeta \int_{B_1} G_{t-s}(c_4(x-y)) \tilde{u}_s^{1+\beta}(y)\, dy\, ds.
\end{aligned}
$$

Since $\xi := c_4^{-d} \min_{x \in B_1} \min_{s \in [1,2]} \mathbb{P}_x(W_s \in B_{c_4}) > 0$, we have

$$\min_{x \in B_1} \tilde{u}_t(x) \geq \xi c_6 K(t_0) + \kappa \xi c_6 \int_0^t s^\zeta (\min_{x \in B_1} \tilde{u}_s(x))^{1+\beta}\, ds.$$

It remains to choose $t_0 > 0$ sufficiently large so that the blow-up time of the equation

$$v(t) = \xi c_6 K(t_0) + \kappa \xi c_6 \int_0^t s^\zeta v^{1+\beta}(s)\, ds$$

is smaller than 2. $\square$

The following result gives an explosion criterion which is actually valid for any $\alpha \in (1,2]$ and $d = 1$; its proof uses Theorem 2.4 instead of Theorem 2.2 and Lemma 3.3. Here the potential V need not be bounded.

Theorem 4.2. *Let $\alpha \in (1,2]$, $\beta > 0$ and assume that $V : \mathbb{R} \to \mathbb{R}_+$ is integrable. Then the solution of*

$$\frac{\partial u_t}{\partial t}(x) = -(-\Delta)^{\alpha/2} u_t(x) - V(x) u_t(x) + \kappa t^\zeta u_t^{1+\beta}(x), \quad u_0(x) = \varphi(x), \quad x \in \mathbb{R},$$

blows up in finite time whenever $0 < \beta < 1 + \alpha\zeta$. If $\beta = 1 + \alpha\zeta$, the same happens provided $\int_{\mathbb{R}} V(z)\, dz$ is sufficiently small.

Proof. Let g_t denote the mild solution of

$$\frac{\partial g_t}{\partial t}(x) = -(-\Delta)^{\alpha/2} g_t(x) - V(x) g_t(x) + \kappa t^\zeta f_t^\beta(x) g_t(x), \quad g_0(x) = \varphi(x), \quad x \in \mathbb{R},$$

where $f_t = P_t \varphi$ satisfies

$$\frac{\partial f_t}{\partial t}(x) = -(-\Delta)^{\alpha/2} f_t(x), \quad f_0(x) = \varphi(x),$$

and $(P_t)_{t \in \mathbb{R}_+}$ is the α-stable semigroup. The Feynman-Kac formula and Jensen's inequality yield

$$g_t(y) \geq$$
$$\int_{\mathbb{R}} \varphi(x) G_t^\alpha(x-y) \exp\left(E_x \left[\int_0^t \left(-V(W_s^\alpha) + s^\zeta \left(P_s \varphi(W_s^\alpha) \right)^\beta \right) ds \Big| W_t^\alpha = y \right] \right) dx,$$

where, for any $t \geq 1$,

$$\begin{aligned}
& E_x \left[\int_0^t s^\zeta \left(P_s \varphi(W_s^\alpha) \right)^\beta ds \Big| W_t^\alpha = y \right] \\
\geq \;& c_2 E_x \left[\int_1^t s^{-\beta/\alpha + \zeta} \mathbf{1}_{\{B_{s^{1/\alpha}}\}}(W_s^\alpha) \Big| W_t^\alpha = y \right] \\
\geq \;& c_2 \int_1^t \mathbb{P}_x(W_s^\alpha \in B_{s^{1/\alpha}} \mid W_t^\alpha = y) s^{-\beta/\alpha + \zeta} ds \\
\geq \;& c_5 \int_1^t s^{\zeta - \beta/\alpha} ds \\
= \;& \frac{c_5}{1 + \zeta - \beta/\alpha} (t^{1 - \beta/\alpha + \zeta} - 1);
\end{aligned}$$

here we applied Lemma 2.2 of [2]. The last inequality together with (2.2) renders

$$g_t(y) \geq e^{-C_\alpha t^{1-1/\alpha} \int_{\mathbb{R}} V(z)\, dz + \frac{c_5}{1-\beta/\alpha+\zeta}(t^{1+\zeta-\beta/\alpha} - 1)},$$

hence by the same steps as in the proof of Theorem 4.1 (comparison result for PDEs and blow-up argument of [6]), finite time explosion occurs if $\beta < 1+\alpha\zeta$, or if $\beta = 1+\alpha\zeta$ and $\int_{\mathbb{R}} V(z)\,dz$ is sufficiently small. $\square$

Since $0 \le V(x) \le (1+|x|^b)^{-1}$, $x \in \mathbb{R}$, and $1 < b \le 2$ imply integrability of $V(x)$ on $\mathbb{R}$, Theorem 4.2 yields a partial extension of Theorem 4.1 to the case $0 < \alpha \le 2$.

5. Existence of global solutions

We have the following non-explosion result, which is a generalization of Theorem 4.1 in [9].

Theorem 5.1. *Consider the semilinear equation*

$$\frac{\partial w_t}{\partial t}(x) = \Delta w_t(x) - V(x)w_t(x) + t^\zeta G(w_t(x)), \qquad w_0(x) = \varphi(x), \quad x \in \mathbb{R}^d, \tag{5.1}$$

where $\zeta \in \mathbb{R}$, φ is bounded and measurable, and $G : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function satisfying

$$0 \le \frac{G(z)}{z} \le \lambda z^\beta, \quad z \in (0,c), \tag{5.2}$$

for some $\lambda, \beta, c > 0$. Assume that $\varphi \ge 0$ is such that

$$\lambda\beta \int_0^\infty r^\zeta \|S_r\varphi\|_\infty^\beta \, dr < 1$$

and

$$\|\varphi\|_\infty \le c\left(1 - \lambda\beta\int_0^\infty r^\zeta \|S_r\varphi\|_\infty^\beta dr\right)^{1/\beta}. \tag{5.3}$$

Then Equation (5.1) *admits a global solution $u_t(x)$ that satisfies*

$$0 \le u_t(x) \le \frac{S_t\varphi(x)}{\left(1 - \lambda\beta\int_0^t r^\zeta \|S_r\varphi\|_\infty^\beta dr\right)^{1/\beta}}, \qquad x \in \mathbb{R}^d, \quad t \ge 0.$$

Proof. This is an adaptation of the proof of Theorem 3 in [16], see also [9]. Recall that the mild solution of (5.1) is given by

$$u_t(x) = S_t\varphi(x) + \int_0^t r^\zeta S_{t-r} G(u_r(x))\,dr. \tag{5.4}$$

Setting

$$B(t) = \left(1 - \lambda\beta\int_0^t r^\zeta \|S_r\varphi\|_\infty^\beta dr\right)^{-1/\beta}, \qquad t \ge 0,$$

it follows that $B(0) = 1$ and

$$\frac{d}{dt}B(t) = \lambda t^\zeta \|S_t\varphi\|_\infty^\beta \left(1 - \lambda\beta\int_0^t r^\zeta \|S_r\varphi\|_\infty^\beta dr\right)^{-1-1/\beta} = \lambda t^\zeta \|S_t\varphi\|_\infty^\beta B^{1+\beta}(t),$$

hence

$$B(t) = 1 + \lambda \int_0^t r^\zeta \|S_r\varphi\|_\infty^\beta B^{1+\beta}(r)\, dr.$$

Let $(t,x) \mapsto v_t(x)$ be a continuous function such that $v_t(\cdot) \in C_0(\mathbb{R}^d)$, $t \geq 0$, and

$$S_t\varphi(x) \leq v_t(x) \leq B(t)S_t\varphi(x), \qquad t \geq 0,\ x \in \mathbb{R}^d. \tag{5.5}$$

Let now

$$R(v)(t,x) = S_t\varphi(x) + \int_0^t r^\zeta S_{t-r} G(v_r(x))\, dr.$$

Since $v_r(x) \leq B(r)\|S_r\varphi\|_\infty$, $r \geq 0$, we have from (5.5), (5.3) and (5.2) that

$$\begin{aligned} R(v)(t,x) &= S_t\varphi(x) + \int_0^t r^\zeta S_{t-r}\left(\frac{G(v_r)}{v_r} v_r\right)(x)\, dr \\ &\leq S_t\varphi(x) + \lambda \int_0^t r^\zeta (B(r))^\beta \|S_r\varphi\|_\infty^\beta S_{t-r} v_r(x)\, dr \\ &\leq S_t\varphi(x) + \lambda \int_0^t r^\zeta B^{1+\beta}(r) \|S_r\varphi\|_\infty^\beta S_{t-r}\left(S_r\varphi(x)\right)\, dr \\ &= S_t\varphi(x)\left(1 + \lambda \int_0^t r^\zeta \|S_r\varphi\|_\infty^\beta B^{1+\beta}(r)\, dr\right), \end{aligned}$$

where the last inequality follows from (5.5). Hence

$$S_t\varphi(x) \leq R(v)(t,x) \leq B(t)S_t\varphi(x), \qquad t \geq 0,\ x \in \mathbb{R}^d.$$

Let

$$u_t^0(x) = S_t\varphi(x), \quad \text{and} \quad u_t^{n+1}(x) = R(u^n)(t,x), \quad n \in \mathbb{N}.$$

Then $u_t^0(x) \leq u_t^1(x)$, $t \geq 0$, $x \in \mathbb{R}^d$. Since S_t is non-negative, using induction we obtain

$$0 \leq u_t^n(x) \leq u_t^{n+1}(x), \qquad n \geq 0.$$

Letting $n \to \infty$ yields, for $t \geq 0$ and $x \in \mathbb{R}^d$,

$$0 \leq u_t(x) = \lim_{n\to\infty} u_t^n(x) \leq B(t)S_t\varphi(x) \leq \frac{S_t\varphi(x)}{\left(1 - \lambda\beta \int_0^t r^\zeta \|S_r\varphi\|_\infty^\beta dr\right)^{1/\beta}} < \infty.$$

Thus, u_t is a global solution of (5.4) due to the monotone convergence theorem. □

As a consequence of Theorem 5.1, an existence result can be obtained under an integrability condition on φ.

Theorem 5.2. *Let $G : \mathbb{R}_+ \to \mathbb{R}_+$ and $v : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ be measurable functions such that $G(z) \leq \kappa_1 z^{1+\beta}$, $z > 0$, and $v_t(x) \leq \kappa_2 t^\zeta$, $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, where $\beta, \zeta, \kappa_1, \kappa_2 > 0$. Let $0 \leq b \leq 2$, $a > 0$, and assume that*

$$V(x) \geq \frac{a}{1 + |x|^b}, \qquad x \in \mathbb{R}^d.$$

i) *If* $b < 2$, *then the equation*

$$\frac{\partial u_t}{\partial t}(x) = \Delta u_t(x) - V(x)u_t(x) + v_t(x)G(u_t(x)), \qquad w_0 = \varphi, \tag{5.6}$$

admits a global solution for all $\beta > 0$.

ii) *If* $b = 2$ *and*

$$\beta > \beta^*(a) := \frac{2(1+\zeta)}{d + \alpha_1(a)},$$

then (5.6) *admits a global solution.*

Proof. Clearly, it suffices to consider the semilinear equation

$$\frac{\partial u_t}{\partial t}(x) = \Delta u_t(x) - V(x)u_t(x) + \kappa t^{\zeta} u_t^{1+\beta}(x), \qquad u_0(x) = \varphi(x), \tag{5.7}$$

for a suitable constant $\kappa > 0$. Suppose that for some $\sigma > 0$,

$$0 \leq \varphi(x) \leq \frac{C}{1 + |x|^{\sigma}}, \qquad x \in \mathbb{R}^d.$$

i) Assume that $\sigma > b(1+\zeta)/\beta$, and let $\varepsilon \in (0,1)$ be such that $(1-\varepsilon)\beta\sigma/b > 1 + \zeta$. From Proposition 3.1.i) we get

$$\int_1^{\infty} t^{\zeta} \|S_t\varphi\|_{\infty}^{\beta} dt < 1,$$

provided C is sufficiently small.

ii) If $b = 2$ and $\beta > 2(1+\zeta)/(d + \alpha_1(a))$, let $\varepsilon \in (0,1)$ be such that $\beta(d/2 + (1-\varepsilon)\alpha_1(a)) > 1 + \zeta$. From Proposition 3.1.ii), there exists $\sigma > d$ such that

$$\int_1^{\infty} t^{\zeta} \|S_t\varphi\|_{\infty}^{\beta} dt < 1$$

provided C is sufficiently small. $\square$

Remark 5.3. An alternative proof of Theorem 5.2-i) consists in letting the initial value φ in (5.7) be such that

$$\varphi(x) \leq \tau S_1 \mathbf{1}_D(x),$$

for a sufficiently small constant $\tau > 0$, where $D \subset \mathbb{R}^d$ is bounded and Borel measurable. By Lemma 3.4,

$$S_t\varphi(x) \leq \tau S_{t+1}\mathbf{1}_D(x) \leq \tau c_D(1+t)^{-(1+\gamma)},$$

thus showing that $\int_1^{\infty} t^{\zeta}\|S_t\varphi\|^{\beta}\, dt$ can be made arbitrarily close to 0 by choosing τ sufficiently small. By Theorem 5.1 we conclude that (5.7) admits positive global solutions.

Remark 5.4. In the same way as in the above remark we can deal with the semilinear system

$$\begin{cases} \dfrac{\partial u_t}{\partial u}(x) = \Delta u_t(x) - V_1(x)u_t(x) + u_t(x)v_t(x), \quad u_0(x) = \varphi(x), \\ \\ \dfrac{\partial v_t}{\partial t}(x) = \Delta v_t(x) - V_2(x)v_t(x) + u_t(x)v_t(x), \quad v_0(x) = \psi(x), \end{cases} \tag{5.8}$$

where $x \in \mathbb{R}^d$, $d \geq 2$, $\varphi, \psi \geq 0$, and

$$V_1(x) \sim \frac{a_1}{1 + |x|^{b_1}}, \qquad V_2(x) \sim \frac{a_2}{1 + |x|^{b_2}}, \qquad x \in \mathbb{R}^d,$$

with $a_i > 0$ and $b_i \geq 0$, $i = 1, 2$.

Theorem 5.5. *If* $\max(b_1, b_2) < 2$*, then* (5.8) *admits nontrivial positive global solutions.*

Proof. Without loss of generality let us assume that $b := b_1 < 2$. Let $(S_t^1)_{t\geq 0}$ denote the semigroup with generator $L = \Delta - V_1$. By Lemma 3.4, there exists $\gamma > 0$ such that

$$S_t^1 \mathbf{1}_D(x) \leq c_D t^{-(1+\gamma)}, \quad x \in \mathbb{R}^d,$$

for all sufficiently large $t > 0$, where c_D does not depend on x and t. The proof is finished by an application of Theorem 1.1 in [10]. $\square$

References

[1] C. Bandle and H. A. Levine, *Fujita type phenomena for reaction-diffusion equations with convection like terms*, Differential Integral Equations, **7 (5-6)** (1994), 1169–1193.

[2] M. Birkner, J. A. López-Mimbela, and A. Wakolbinger, *Blow-up of semilinear PDE's at the critical dimension. A probabilistic approach*, Proc. Amer. Math. Soc., **130 (8)** (2002), 2431–2442 (electronic).

[3] K. Deng and H. A. Levine, *The role of critical exponents in blow-up theorems: the sequel*, J. Math. Anal. Appl., **243 (1)** (2000), 85–126.

[4] M. Guedda and M. Kirane, *A note on nonexistence of global solutions to a nonlinear integral equation*, Bull. Belg. Math. Soc. Simon Stevin, **6 (4)** (1999), 491–497.

[5] A. El Hamidi and G. Laptev, *Existence and nonexistence results for higher-order semilinear evolution inequalities with critical potential*, J. Math. Anal. Appl., **304** (2005), 451–463.

[6] K. Kobayashi, T. Sirao, and H. Tanaka, *On the growing up problem for semilinear heat equations*, J. Math. Soc. Japan, **29 (3)** (1977), 407–424.

[7] H. A. Levine, *The role of critical exponents in blowup theorems*, SIAM Rev., **32 (2)** (1990), 262–288.

[8] H. A. Levine and P. Meier, *The value of the critical exponent for reaction-diffusion equations in cones*, Arch. Rational Mech. Anal., **109 (1)** (1990), 73–80.

[9] J. A. López-Mimbela and N. Privault, *Blow-up and stability of semilinear PDEs with Gamma generators*, J. Math. Anal. Appl., **370** (2005), 181–205.

[10] J. A. López-Mimbela and A. Wakolbinger, *Length of Galton-Watson trees and blow-up of semilinear systems*, J. Appl. Probab., **35 (4)** (1998), 802–811.

[11] M. Nagasawa and T. Sirao, *Probabilistic treatment of the blowing up of solutions for a nonlinear integral equation*, Trans. Amer. Math. Soc., **139** (1969), 301–310.

[12] C. V. Pao, *Nonlinear Parabolic and Elliptic Equations*, Plenum Press, New York, 1992.

[13] A. A. Samarskii, V. A. Galaktionov, S. P. Kurdyumov, and A. P. Mikhailov, *Blow-up in Quasilinear Parabolic Equations*, volume 19 of de Gruyter Expositions in Mathematics, Walter de Gruyter & Co., Berlin, 1995.

[14] P. Souplet and Q. S. Zhang, *Stability for semilinear parabolic equations with decaying potentials in $\mathbb{R}^n$ and dynamical approach to the existence of ground states*, Ann. Inst. H. Poincaré Anal. Non Linéaire, **19 (5)** (2002), 683–703.

[15] S. Sugitani, *On nonexistence of global solutions for some nonlinear integral equations*, Osaka J. Math., **12** (1975), 45–51.

[16] F. B. Weissler, *Existence and nonexistence of global solutions for a semilinear heat equation*, Israel J. Math., **38 (1-2)** (1981), 29–40.

[17] B. Wong and Q. S. Zhang, *Refined gradient bounds, Poisson equations and some applications to open Kähler manifolds*, Asian J. Math., **7 (3)** (2003), 337–364.

[18] Q. S. Zhang, *Large time behavior of Schrödinger heat kernels and applications*, Comm. Math. Phys., **210 (2)** (2000), 371–398.

[19] Q. S. Zhang, *The quantizing effect of potentials on the critical number of reaction-diffusion equations*, J. Differential Equations, **170 (1)** (2001), 188–214.

José Alfredo López-Mimbela
Centro de Investigación en Matemáticas
Apartado Postal 402
36000 Guanajuato
Mexico
e-mail: `jalfredo@cimat.mx`

Nicolas Privault
Département de Mathématiques
Université de La Rochelle
Avenue Michel Crépeau
17042 La Rochelle Cedex 1
France
e-mail: `nprivaul@univ-lr.fr`

Progress in Probability, Vol. 59, 261–274

Generalized Ornstein–Uhlenbeck Processes on Separable Banach Spaces

V. Mandrekar and B. Rüdiger

Abstract. In this article we reduce the analysis of Banach-valued generalized Ornstein–Uhlenbeck processes to an application of the results in [12, 16], concerning Banach-valued stochastic integrals w.r.t. Lévy processes and compensated Poisson random measures, as well as the results in [11], related to the analysis of Banach-valued stochastic differential equations with Lévy noise, and the corresponding Itô formula studied in [17].

Mathematics Subject Classification (2000). 60H05, 60G51, 60G57, 46B09, 47G99.

Keywords. Stochastic differential equations, stochastic integrals on separable Banach spaces, M-type 2 Banach spaces, martingales measures, compensated Poisson random measures, additive processes, random Banach-valued functions, generalized Ornstein–Uhlenbeck processes, Itô formula.

1. Introduction

O. Barndorff-Nielsen and N. Shepard introduced the generalized Ornstein–Uhlenbeck process as a model for volatility in Finance [5] (see also the article by Z. J. Jurek, W. Vervaat [9] and related comments in Remark 3.4 of Section 3). From the results in [11] it follows that generalized Ornstein–Uhlenbeck processes can be studied also on Banach spaces. In fact, existence and uniqueness of the solutions for the corresponding linear stochastic differential equations (SDEs) (defined in (3.1)–(3.3) below) follow from general results in [11], where SDEs with local Lipschitz drift and noise coefficients have been analyzed. In [12] we give sufficient conditions for the existence of Itô integrals w.r.t Banach-valued Lévy processes and prove that these can also be written as integrals w.r.t. the corresponding compensated Poisson random measure (cPrm). As an example for possible applications of these previous results, together with the Itô formula for Banach-valued jump processes found in [17], we show in this article that these can be used to study pathwise properties of Banach-valued generalized Ornstein–Uhlenbeck processes, also related to properties of the corresponding invariant measures.

In Section 2 we recall some of our previous results on stochastic integrals w.r.t. cPrm and Lévy processes [11, 12, 16, 17]. We recall in particular the Itô formula for Banach-valued jump processes obtained by integration w.r.t. cPrms [17] and the results in [12] where we show that integrals w.r.t. Lévy processes coincide with integrals w.r.t. cPrms [12]. In Section 3 we present the results related to generalized Ornstein–Uhlenbeck process mentioned above.

2. Stochastic Integrals and the Lévy noise on Banach spaces

In [11, 12, 16] we analyzed the stochastic integrals of Banach-valued random functions w.r.t cPrms and additive processes. Here we recall only a small part of these results, in fact only the results which are used in the following section. We restrict in particular our attention to the case of Lévy processes.

In the whole article we assume that

$$q(dsdx)(\omega) := N(dsdx)(\omega) - ds\beta(dx)$$

is the compensated Poisson random measure associated to an E-valued Lévy process $(X_t)_{t\geq 0}$ on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq\infty}, P)$, where E is a separable Banach space with norm $\|\cdot\|_E$ and $(\mathcal{B}(E))$ is the corresponding σ-algebra. When no misunderstanding is possible we write $\|\cdot\|$ instead of $\|\cdot\|_E$. We assume that the filtered probability space satisfies the "usual hypotheses", i.e.:

i) $\mathcal{F}_t$ contain all null sets of $\mathcal{F}$, for all t such that $0 \leq t < +\infty$
ii) $\mathcal{F}_t = \mathcal{F}_t^+$, where $\mathcal{F}_t^+ = \cap_{u>t}\mathcal{F}_u$, for all t such that $0 \leq t < +\infty$, i.e., the filtration is right continuous.

Let us use the notation $E_0 := E \setminus \{0\}$. It is well known that the measures $N(dsdx)(\omega)$ (for ω fixed) and $ds\beta(dx)$ on $\mathcal{B}(E_0)$ are σ-finite (eventually not finite) measures, which are finite on the sets $(0,T] \times \Lambda$, with $\Lambda \in \mathcal{B}(E_0)$ and $0 \notin \overline{\Lambda}$ (where with $\overline{\Lambda}$ we denote the closure of the set Λ).

Let F be a separable Banach space with norm $\|\cdot\|_F$. (When no misunderstanding is possible we write $\|\cdot\|$ instead of $\|\cdot\|_F$.) Let $\mathcal{E}_t := \mathcal{B}(\mathbb{R}_+ \times E_0) \otimes \mathcal{F}_t$ be the product σ-algebra generated by the semi-ring $\mathcal{B}(\mathbb{R}_+ \times E_0) \times \mathcal{F}_t$. Let $T > 0$, and

$$\begin{aligned} M^T(E/F) := \{f : \mathbb{R}_+ \times E_0 \times \Omega \to F, &\text{ such that } f \text{ is } \mathcal{E}_T/\mathcal{B}(F)\text{-measurable,} \\ &f(t,x,\omega) \text{ is } \mathcal{F}_t\text{-adapted} \quad \forall x \in E_0,\ t \in (0,T]\} \end{aligned}$$

There is a "natural definition" of stochastic integral w.r.t. $q(dtdx)(\omega)$ on those sets $(0,T] \times \Lambda$ with $\Lambda \in \mathcal{B}(E_0)$ and such that $0 \notin \overline{\Lambda}$ [16]:

Definition 2.1. Let $t \in (0,T]$, $0 \notin \overline{\Lambda}$, $f \in M^T(E/F)$. Assume that $f(\cdot,\cdot,\omega)$ is Bochner integrable on $(0,T] \times \Lambda$ w.r.t. $ds\beta(dx)$, for all $\omega \in \Omega$ fixed. The *natural integral* of f on $(0,t] \times \Lambda$ w.r.t. the compensated Poisson random measure

$q(dtdx) := N(dtdx)(\omega) - dt\beta(dx)$ is

$$\int_0^t \int_\Lambda f(s,x,\omega)\,(N(dsdx)(\omega) - ds\beta(dx))$$
$$:= \sum_{0<s\leq t} f(s,(\Delta X_s)(\omega),\omega)\mathbf{1}_\Lambda(\Delta X_s(\omega)) - \int_0^t \int_\Lambda f(s,x,\omega)ds\beta(dx)\ \omega \in \Omega,$$

where the last term is understood as a Bochner integral, (for $\omega \in \Omega$ fixed) of $f(s,x,\omega)$ w.r.t. the measure $ds\beta(dx)$.

It is more difficult to define the stochastic integral on those sets $(0,T] \times \Lambda$, $\Lambda \in \mathcal{B}(E_0)$, s.th. $0 \in \overline{\Lambda}$, as it might be that $\beta(\Lambda) = \infty$ (see, e.g., [18] or [10] for the discussion on the properties of Lévy measures $\beta(dx)$).

Definition 2.2. A function f belongs to the set $\Sigma(E/F)$ of *simple functions*, if $f \in M^T(E/F)$, $T > 0$ and there exist $n \in \mathbb{N}$, $m \in \mathbb{N}$, such that

$$f(t,x,\omega) = \sum_{k=1}^{n-1} \sum_{l=1}^{m} \mathbf{1}_{A_{k,l}}(x)\mathbf{1}_{F_{k,l}}(\omega)\mathbf{1}_{(t_k,t_{k+1}]}(t)a_{k,l} \tag{2.1}$$

where $A_{k,l} \in \mathcal{B}(E_0)$ and $0 \notin \overline{A_{k,l}}$, $t_k \in (0,T]$, $t_k < t_{k+1}$, $F_{k,l} \in \mathcal{F}_{t_k}$, $a_{k,l} \in F$. For all $k \in 1,\ldots,n-1$ fixed, $A_{k,l_1} \times F_{k,l_1} \cap A_{k,l_2} \times F_{k,l_2} = \emptyset$ if $l_1 \neq l_2$.

Let $f \in \Sigma(E/F)$ be of the form (2.1), then

$$\int_0^T \int_\Lambda f(t,x,\omega)q(dtdx)(\omega) = \sum_{k=1}^{n-1}\sum_{l=1}^{m} a_{k,l}\mathbf{1}_{F_{k,l}}(\omega)q((t_k,t_{k+1}]\cap(0,T]\times A_{k,l}\cap\Lambda)(\omega) \tag{2.2}$$

for all $\Lambda \in \mathcal{B}(E_0)$, $T > 0$. (Equation (2.2) might either be interpreted as a definition or, like done in [11, 12, 16], as a statement, by first introducing the definition of "natural integral".)

We recall here the definition of a strong-p-integral, $p \geq 1$, (Definition 2.6 below) used in [16] (and in several previous articles, see Remark 2.7.)

First we establish some properties of the functions $f \in M_\beta^{T,p}(E/F)$, where

$$M_\beta^{T,p}(E/F) := \{f \in M^T(E/F) : \int_0^T \int_{E_0} E[\|f(t,x,\omega)\|^p]\, dt\beta(dx) < \infty\}. \tag{2.3}$$

Theorem 2.3 ([16], Theorem 4.2)**.** *Let $p \geq 1$. Let $T > 0$, then for all $f \in M_\beta^{T,p}(E/F)$ and all $\Lambda \in \mathcal{B}(E_0)$, there is a sequence of simple functions $\{f_n\}_{n\in\mathbb{N}}$ satisfying the following property :*

Property P: $f_n \in \Sigma(E/F)\,\forall n \in \mathbb{N}$, *$f_n$ converges $dt \otimes \beta(dx) \otimes P$-a.s. to f on $(0,T] \times \Lambda \times \Omega$, when $n \to \infty$, and*

$$\lim_{n\to\infty} \int_0^T \int_\Lambda E[\|f_n(t,x) - f(t,x)\|^p]\, dt\beta(dx) = 0\,, \tag{2.4}$$

i.e., $\|f_n - f\|$ converges to zero in $L^p((0,T]\times\Lambda\times\Omega, dt\otimes\beta(dx)\otimes P)$, when $n \to \infty$.

Definition 2.4. We say that a sequence of functions f_n is L^p-*approximating* f, if it satisfies property P, i.e., f_n converges $dt \otimes \beta(dx) \otimes P$-a.s. to f on $(0,T] \times \Lambda \times \Omega$, when $n \to \infty$, and satisfies (2.4).

Definition 2.5. Let $p \geq 1$; $L_p^F(\Omega, \mathcal{F}, P)$ is the space of F-valued random variables, such that $E\|Y\|^p = \int \|Y\|^p dP < \infty$. We denote by $\|\cdot\|_p$ the norm given by $\|Y\|_p = (E\|Y\|^p)^{1/p}$. Given $(Y_n)_{n\in\mathbb{N}}, Y \in L_p^F(\Omega, \mathcal{F}, P)$, we write $\lim^p_{n\to\infty} Y_n = Y$ if $\lim_{n\to\infty} \|Y_n - Y\|_p = 0$.

In [16] we introduced the following

Definition 2.6. Let $p \geq 1$, $t > 0$. We say that f is *strong-p-integrable* on $(0,t] \times \Lambda$, $\Lambda \in \mathcal{B}(E_0)$, if there is a sequence $\{f_n\}_{n\in\mathbb{N}} \in \Sigma(E/F)$, which satisfies the property P in Theorem 2.3, and such that the limit of the integrals of f_n w.r.t. $q(dtdx)$ exists in $L_p^F(\Omega, \mathcal{F}, P)$ for $n \to \infty$, i.e.,

$$\int_0^t \int_\Lambda f(t,x,\omega) q(dtdx)(\omega) := \lim_{n\to\infty}^{p} \int_0^t \int_\Lambda f_n(t,x,\omega) q(dtdx)(\omega) \tag{2.5}$$

exists. Moreover, the limit (2.5) does not depend on the sequence $\{f_n\}_{n\in\mathbb{N}} \in \Sigma(E/F)$, for which property P and (2.5) holds.

Remark 2.7. The definition of stochastic integral w.r.t. cPrm defined in Definition 2.6 is also used, e.g., (for the finite-dimensional case) in [2, 6, 19] and recently in [3] (however without calling the stochastic integrals strong-p-integrals). In [16] we compared these integrals with the ones introduced, e.g., in [8]. These are called in [16] simple-p-integrals. In [16] we show in particular that the definition of strong-p-integral is more general than the definition of simple-p-integral. (We refer to [16] for precise statements, i.e., Theorems 6.4–6.7 in [16].)

In [11, 16] we gave sufficient conditions for the existence of the strong-p-integrals, when $p = 1$, or $p = 2$.

Theorem 2.8 ([16], Theorem 4.12)**.** *Let $f \in M_\beta^{T,1}(E/F)$, then f is strong-1-integrable w.r.t. $q(dt, dx)$ on $(0,t] \times \Lambda$, for any $0 < t \leq T$, $\Lambda \in \mathcal{B}(E_0)$. Moreover,*

$$E\left[\| \int_0^t \int_\Lambda f(s,x,\omega) q(dsdx)(\omega)\|\right] \leq 2 \int_0^t \int_\Lambda E[\|f(s,x,\omega)\|] ds \beta(dx)(\omega).$$

Theorem 2.9 ([16], Theorem 4.14)**.** *Suppose $(F, \mathcal{B}(F)) := (H, \mathcal{B}(H))$ is a separable Hilbert space. Let $f \in M_\beta^{T,2}(E/H)$, then f is strong 2-integrable w.r.t. $q(dtdx)$ on $(0,t] \times \Lambda$, for any $0 < t \leq T$, $\Lambda \in \mathcal{B}(E_0)$. Moreover,*

$$E\left[\| \int_0^t \int_\Lambda f(s,x,\omega) q(dsdx)(\omega)\|^2\right] = \int_0^t \int_\Lambda E[\|f(s,x,\omega)\|^2] ds \beta(dx).$$

The following Theorem 2.12 was proven in [16] for the case of deterministic functions on type 2 Banach spaces, and on M-type 2 spaces for functions which do not depend on the random variable x, in [12] for the general case.

We recall here the definition of M-type 2 and type 2 separable Banach space (see, e.g., [13]).

Definition 2.10. A separable Banach space F, with norm $\|\cdot\|$, is of *M-type* 2, if there is a constant K_2, such that for any F-valued martingale $(M_k)_{k\in 1,\ldots,n}$ the following inequality holds:

$$E[\|M_n\|^2] \leq K_2 \sum_{k=1}^{n} E[\|M_k - M_{k-1}\|^2]\,,$$

with the convention that $M_0 = 0$.

We remark that a separable Hilbert space is in particular a separable Banach space of M-type 2.

Definition 2.11. A separable Banach space F is of *type* 2, if there is a constant $\mathcal{K}_2$, such that if $\{X_i\}_{i=1}^n$ is any finite set of centered independent F-valued random variables, such that $E[\|X_i\|^2] < \infty$, then

$$E[\|\sum_{i=1}^{n} X_i\|^2] \leq \mathcal{K}_2 \sum_{i=1}^{n} E[\|X_i\|^2].$$

We remark that any separable Banach space of M-type 2 is a separable Banach space of type 2. Typical examples of separable Banach spaces of M- type 2 are the spaces $L_p(\Omega, P)$, $p \in [2, \infty)$.

Theorem 2.12 ([11], Theorem 3.6)**.** *Suppose that F is a separable Banach space of M-type 2. Let $f \in M_\beta^{T,2}(E/F)$, then f is strong 2-integrable w.r.t. $q(dtdx)$ on $(0,t] \times \Lambda$, for any $0 < t \leq T$, $\Lambda \in \mathcal{B}(E_0)$. Moreover,*

$$E\left[\|\int_0^t \int_\Lambda f(s,x,\omega) q(dsdx)(\omega)\|^2\right] \leq K_2 \int_0^t \int_\Lambda E[\|f(s,x,\omega)\|^2] ds\beta(dx),$$

where K_2 is the constant in the Definition 2.10 of M-type 2 Banach spaces.

Theorem 2.13 ([16], Theorem 4.16)**.** *Suppose that F is a separable Banach space of type 2. Let $f \in M_\beta^{T,2}(E/F)$, and f be a deterministic function, i.e., $f(t,x,\omega)$ $= f(t,x)$, then f is strong 2-integrable w.r.t. $q(dtdx)$ on $(0,t] \times A$, for any $0 <$ $t \leq T$, $A \in \mathcal{B}(E_0)$. Moreover,*

$$E\left[\|\int_0^t \int_\Lambda f(s,x,\omega) q(dsdx)(\omega)\|^2\right] \leq 4\mathcal{K}_2 \int_0^t \int_\Lambda E[\|f(s,x,\omega)\|^2] ds\beta(dx),$$

where $\mathcal{K}_2$ is the constant in the Definition 2.11 of type 2 Banach spaces.

Proposition 2.14 ([16], [11], Proposition 3.12)**.** *Let f satisfy the hypothesis of Theorem 2.8, or 2.12. Then $\int_0^t \int_\Lambda f(s,x,\omega) q(dsdx)(\omega)$, $t \in [0,T]$ is an $\mathcal{F}_t$-martingale with mean zero and is cád-làg.*

In [17], we analyzed the Itô formula for jump processes defined through the strong-p-integrals in Definition 2.6. and proved the following Theorem (in a slightly more general form than here, by adding also the stochastic integrals w.r.t. random functions of bounded variation).

Theorem 2.15 ([17], Theorem 5.1)**.** *Let $p = 1$ or $p = 2$. Let $f \in M_\beta^{T,p}(E/F)$, where F is a separable Banach space of M-type 2, if $p = 2$, or of type 2, if $p = 2$ and f is a deterministic function, i.e., $f(t, x, \omega) = f(t, x)$. Let*

$$Y_t(\omega) := \int_0^t \int_\Lambda f(t,x,\omega) q(dtdx)(\omega) + \int_0^t \int_{\Lambda'} k(s,x,\omega) N(dsdx)(\omega)\,,$$

where $\Lambda \in \mathcal{B}(E_0)$ and $\Lambda' \in \mathcal{B}(E_0)$, with $0 \notin \overline{\Lambda'}$. Moreover let the random function $k(s, x, \omega) \in M^T(E/F)$ be finite P-a.s. for every $s \in [0, T]$, $x \in \Lambda$, and be cádlàg or cáglàd.

Let G be a separable Banach space. Let G be of M-type-2, if $p = 2$. Suppose that the Fréchet derivatives $\partial_s \mathcal{H}(s, y)$ and $\partial_y \mathcal{H}(s, y)$ exists and are uniformly bounded on $[\tau, t] \times F$, and all the second Fréchet derivatives $\partial_s\partial_s \mathcal{H}(s, y)$, $\partial_{s,} \partial_y \mathcal{H}(s, y)$, $\partial_y \partial_s \mathcal{H}(s, y)$ and $\partial_y \partial_y \mathcal{H}(s, y)$ exist and are uniformly bounded on $[\tau, t] \times B(0, R)$, for all $R \geq 0$. Then

$$\begin{aligned}
&\mathcal{H}(t, Y_t(\omega)) - \mathcal{H}(\tau, Y_\tau(\omega)) \\
&\quad= \int_\tau^t \partial_s \mathcal{H}(s, Y_{s_-}(\omega)) ds + \int_\tau^t \int_A \{\mathcal{H}(s, Y_{s_-}(\omega) + f(s,x,\omega)) \\
&\qquad\qquad - \mathcal{H}(s, Y_{s_-}(\omega))\}\, q(dsdx)(\omega) \\
&\qquad + \int_\tau^t \int_A \{\mathcal{H}(s, Y_{s_-}(\omega) + f(s,x,\omega)) - \mathcal{H}(s, Y_{s_-}(\omega)) \\
&\qquad\qquad - \partial_y \mathcal{H}(s, Y_{s_-}(\omega)) f(s,x,\omega)\}\, ds\beta(dx) \\
&\qquad + \int_\tau^t \int_\Lambda \{\mathcal{H}(s, Y_{s_-}(\omega) + k(s,x,\omega)) - \mathcal{H}(s, Y_{s_-}(\omega))\}\, N(dsdx)(\omega) \\
&\qquad\qquad P\text{-a.s.}
\end{aligned} \tag{2.6}$$

In [12] we analyzed the Itô-integral of random functions $H(s, \omega)$ w.r.t. Banach-valued martingales $(M_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq\infty}, P)$, obtained by strong-p integration w.r.t. cPrm. We proved that such integrals can be seen as strong-p-integrals w.r.t. cPrms. We recall here some results, however restricting only to the case where $H(s, \omega)$ is real-valued. For more general cases we refer to [12].

Let $M^T(\mathbb{R}_+/\mathbb{R})$ be the set of progressive measurable processes $(H_t)_{t\in[0,T]}$ with values on $\mathbb{R}$.

Definition 2.16. We denote by $E^T(\mathbb{R}_+/\mathbb{R})$ the set of elementary processes $(H(t, \omega))_{t\in[0,T]}$, i.e., those which are in $M^T(\mathbb{R}_+/\mathbb{R})$, are uniformly bounded and are of the form

$$H(t, \omega) = \sum_{i=1}^{r-1} \mathbf{1}_{(t_i, t_i+1]}(t) H_i(\omega)\,, \tag{2.7}$$

with $H_i(\omega)$ $\mathcal{F}_{t_i}$-adapted, $0 < t_i < t_{i+1} \leq T$.

In the usual way we introduce the stochastic integral of elementary processes w.r.t. martingales on the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq\infty}, P)$.

Definition 2.17. Let $(M_t)_{t\in[0,T]}$ be an $\mathcal{F}_t$-adapted martingale with values on the separable Banach space F. Let $(H(t,\omega))_{t\in[0,T]} \in E^T(\mathbb{R}_+/\mathbb{R})$, $(H(t,\omega))_{t\in[0,T]}$ be of the form (2.7). The stochastic integral $(H\cdot M)_t$, $t\in[0,T]$, of $(H(t,\omega))_{t\in[0,T]}$ w.r.t. $(M_t(\omega))_{t\in[0,T]}$ is defined by

$$(H\cdot M)_t(\omega) := \int_0^t H(s,\omega)dM_s(\omega) := \sum_{i=1}^{r-1} H_i(\omega)[M_{t_{i+1}\wedge t}(\omega) - M_{t_i\wedge t}(\omega)].$$

Let $p=1$ or $p=2$. Let F be a separable Banach space. If $p=2$ we suppose also that F is an M-type 2 Banach space. Let $f\in M_\beta^{T,p}(E/F)$ (defined in (2.3)). We define

$$M_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R}) :=$$
$$\{(H(t,\omega))_{t\in[0,T]} \in M^T(\mathbb{R}_+/\mathbb{R}),\ s.th.\ \textstyle\int_0^T\int_{E_0} E[|H(s)|^p\|f(s,x)\|^p]ds\beta(dx)<\infty\}.$$

Remark 2.18. If $(H(t,\omega))_{t\in[0,T]} \in M_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R}))$ then there exists a sequence of elementary processes $(H_n(t,\omega))_{t\in[0,T]} \in E^T(\mathbb{R}_+/\mathbb{R})$ s.th.

$$\lim_{n\to\infty}\int_0^T\int_{E_0} E[|H_n(s)-H(s)|^p\|f(s,x)\|^p]ds\beta(dx)=0. \tag{2.8}$$

This can be proven e.g. with the analogous techniques used in STEP 1–STEP 4 in the proof of Theorem 2.3 in [16].

We denote with $\mathcal{M}_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R})$ the set of $dt\otimes dP$ equivalence classes in $M_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R})$. $\mathcal{M}_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R})$ is a separable Banach space.

Theorem 2.19 ([12], Theorem 3.6)**.** *Let $(H(t,\omega))_{t\in[0,T]} \in M_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R})$. There is a unique element $(H\cdot M)_t \in \mathcal{M}_{f,\beta}^{T,p}(\mathbb{R}_+/\mathbb{R})$, such that*

$$(H\cdot M)_t = \overset{p}{\lim_{n\to\infty}}(H_n\cdot M)_t = \overset{p}{\lim_{n\to\infty}}\int_0^t H_n dM_s \tag{2.9}$$

for any sequence of elementary processes $(H_n(t,\omega))_{t\in[0,T]} \in E^T(\mathbb{R}_+/\mathbb{R}))$, for which (2.8) *holds.*

Moreover the following properties hold:

1) *The convergence* (2.9) *holds also in the following sense:*

$$P(\sup_{[0,T]}\|(H_n\cdot M)_t-(H\cdot M)_t\|>\epsilon)\to_{n\to\infty}0.$$

It follows that there is a subsequence such that

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\|(H_n\cdot M)_t-(H\cdot M)_t\|=0 \quad P\text{-}a.s. \tag{2.10}$$

2) $(H\cdot M)_t$ *coincides with the strong-p-integral of the function Hf w.r.t. the cPrm q, i.e.,*

$$P\left((H\cdot M)_t = \int_0^t\int_{E_0} H(s,\omega)f(s,x,\omega))q(dsdx)(\omega) \quad \forall t\in[0,T]\right)=1.$$

3) $(H \cdot M)_t$ *is an* $\mathcal{F}_t$*-martingale.*

As already stressed in [12] it follows in particular that the definition of Lévy noise introduced in [4] is well defined on Banach spaces. (See [12] for a more precise statement.)

Definition 2.20. We call $\int_0^t H(s,\omega)dM_s(\omega) := (H \cdot M)_s(\omega)$ the stochastic Itô-integral of $(H(t,\omega))_{t\in[0,T]}$ w.r.t. $(M_t(\omega))_{t\in[0,T]}$, if it is obtained from the limit in (2.10).

Remark 2.21. If f is a deterministic function and $p = 2$, then it is sufficient that F is a separable Banach spaces of type 2.

3. The generalized Ornstein–Uhlenbeck processes on separable Banach spaces

Let $p = 1$, or $p = 2$ and $(F, \mathcal{B}(F))$ be a separable Banach space of type 2. We assume that

$$\int_{0<\|x\|\leq 1} \|x\|^p \beta(dx) < \infty \tag{3.1}$$

We analyze

$$d\eta_t(\omega) = -a\eta_t(\omega)dt + d\xi_t(\omega) \tag{3.2}$$

where

$$\xi_t(\omega) := \int_0^t \int_{0<\|x\|\leq 1} xq(dsdx)(\omega) + \int_0^t \int_{\|x\|>1} xN(dsdx)(\omega)\,, \tag{3.3}$$

$a > 0$ and the initial condition η_0 being independent of the filtration $(\mathcal{F}_t)_{t\geq 0}$ of $(\xi_t)_{t\in[0,T]}$. As a consequence of [12] (or the previous section),

$$d\xi_t(\omega) = \int_{0<\|x\|\leq 1} xq(dsdx)(\omega) + \int_{\|x\|>1} xN(dsdx)(\omega)$$

From the general results in [11, Theorem 4.11], we know that for every $T > 0$ there is a unique path wise solution $(\xi_t)_{t\in[0,T]}$ of (3.2) with initial condition η_0. Moreover if $\eta_0 = x$, $x \in F$, then $(\xi_t)_{t\in[0,T]}$ is Markov ([11, Theorem 5.2]).

Using Itô's formula we show in this section that the solution is

$$\eta_t(\omega) = e^{-at}\left(\eta_0(\omega) + \int_0^t e^{as} d\xi_s(\omega)\right). \tag{3.4}$$

In fact, applying the Itô formula (2.6) in Theorem 2.15 to

$$\mathcal{H}(s,z) := e^{-as} z$$

$$\begin{aligned}
Y_s(\omega) \quad &:= \quad \eta_0(\omega) + \int_0^t e^{as} d\xi_s(\omega) \\
&= \quad \eta_0(\omega) + \int_0^t \int_{0<\|x\|\leq 1} e^{as} xq(dsdx)(\omega) + \int_0^t \int_{\|x\|>1} e^{as} xN(dsdx)(\omega)
\end{aligned}$$

$$f(s,x,\omega) := e^{as}x$$

we obtain

$$\begin{aligned}
& \mathcal{H}(t,Y_t(\omega)) - \mathcal{H}(\tau,Y_\tau(\omega)) \\
= \; & -a\int_\tau^t e^{-as}Y_s(\omega)ds \\
& + \int_\tau^t \int_{0<\|x\|\le 1} \{e^{-as}[Y_{s_-}(\omega)+e^{as}x] - e^{-as}Y_{s_-}(\omega) - x\}ds\beta(dx) \\
& + \int_\tau^t \int_{0<\|x\|\le 1} \{e^{-as}[Y_{s_-}(\omega)+e^{as}x] - e^{-as}Y_{s_-}(\omega)\}q(dsdx)(\omega) \\
& + \int_\tau^t \int_{\|x\|>1} \{e^{-as}[Y_{s_-}(\omega)+e^{as}x] - e^{-as}Y_{s_-}(\omega)\}N(dsdx)(\omega) \\
= \; & -a\int_\tau^t e^{-as}Y_s(\omega)ds \\
& + \int_\tau^t \int_{0<\|x\|\le 1} xq(dsdx)(\omega) + \int_\tau^t \int_{\|x\|>1} xN(dsdx)(\omega)
\end{aligned}$$

and hence (3.2).

Remark that

$$d\eta_t(\omega) = -ae^{-at}\left(\eta_0(\omega) + \int_0^t e^{as}d\xi_s(\omega)\right)dt + e^{-at}d\left(\int_0^t e^{as}d\xi_s(\omega)\right)$$

where

$$d\left(\int_0^t e^{as}d\xi_s(\omega)\right) = e^{as}d\xi_s(\omega).$$

Let us denote for $x \in E_0$, $A \in \mathcal{B}(E_0)$

$$P_t(x,A) = P(\eta_t \in A/\eta_0 = x)$$

the transition probability of η_t. Then

$$P_t(x,A) = \delta_{e^{-at}x} \star Q_t(A) \quad \forall A \in E_0$$

where with $\star$ we denote the convolution, δ is the Dirac measure, and $Q_t(\cdot)$ is the distribution of $\int_0^t e^{-(a(t-s)}d\xi_s$. We remark that this coincides with the distribution of $\int_0^t e^{-as}d\xi_s$, as ξ_s has stationary independent increments. If η_t has an invariant measure, i.e., there exists a measure μ on E s.th.

$$\int_E P_t(x,A)\mu(dx) = \mu(A)\,,$$

then

$$\int_E \delta_{e^{-at}x} \star Q_t(A)\mu(dx) = \mu(A).$$

Theorem 3.1. *Let* η_0 *be independent of the filtration* $(\mathcal{F}_t)_{t\geq 0}$ *of* $(\xi_t)_{t\geq 0}$*, and*

$$\eta_t = e^{-at}\left(\eta_0 + \int_0^t e^{as} d\xi_s\right).$$

Let $(P_t)_{t\geq 0}$ *denote the Markov semigroup associated to* $(\eta_t)_{t\geq 0}$*. Suppose that* μ *is a corresponding invariant measure, and let* $\mathcal{L}(X)$ *denote the law of a random variable* X*, then if* $\int_0^t e^{-as} d\xi_s$ *converges in law, then*

$$\lim_{t\to\infty} \mathcal{L}\left(\int_0^t e^{-as} d\xi_s\right) = \mu. \tag{3.5}$$

Proof of Theorem 3.1.

$$\mu = P_t \star \mu = \mu(e^{-at}\cdot) \star \mathcal{L}\left(\int_0^t e^{-a(t-s)} d\xi_s\right).$$

We note that

$$\mathcal{L}\left(\int_0^t e^{-as} d\xi(s)\right) = \mathcal{L}\left(\int_0^t e^{-a(t-s)} d\xi_s\right).$$

Since $a > 0$, $e^{-at} \to 0$ when $t \to \infty$, gives $\mu(e^{-at}\cdot) \to \delta_0$ when $t \to \infty$, so that there is a measure ν (see e.g. [10]), s.th.

$$\mathcal{L}\left(\int_0^t e^{-as} d\xi_s\right) \to \nu \quad \textit{when} \quad t \to \infty$$

and $\nu = \mu$. □

Let us discuss when Theorem 3.1 can be used to find the invariant measure of the solution (3.4) of (3.2), (3.3). We first prove that $P_t(x,\cdot)$ is infinitely divisible.

Lemma 3.2.

$$\begin{aligned}
&\int_{E_0} e^{ix^\star(y)} P_t(x, dy) \\
= \;& \exp\left[e^{-at} i x^\star(x)\right] \\
& \times \exp\left[\int_0^t \int_{0<\|y\|\leq 1} (e^{ie^{a(t-s)}x^\star(y)} - 1 - ie^{a(t-s)}x^\star(y)) ds\beta(dy)\right] \\
& \times \exp\left[\int_0^t \int_{\|y\|>1} (e^{ie^{a(t-s)}x^\star(y)} - 1) ds\beta(dy)\right] \\
= \;& \exp\left[e^{-at} i x^\star(x) + \int_0^t \psi(e^{a(t-s)}x^\star) ds\right]
\end{aligned}$$

where $\exp(t\psi(x^\star))$ *is the Fourier transform of* ξ_t*.*

Proof. The proof is obtained by applying the Itô formula in Theorem 2.15 to (3.4) and is similar to the computations in the appendix. □

Remark 3.3. Lemma 3.2 implies that $P_t(x,\cdot)$ is infinitely divisible and the corresponding Lévy measure $\beta_t(\cdot)$ is such that

$$\beta_t(\Lambda) = \int_{E_0} \beta(dy) \int_0^t \mathbf{1}_\Lambda(e^{-as}y)ds\,, \quad \Lambda \in \mathcal{B}(\mathbb{R}_+ \times E_0)\,,$$

while the corresponding shift is

$$\gamma := e^{-at}x + \int_{E_0} \beta(dy) \int_0^t e^{-as}y(\mathbf{1}_{0<\|y\|\leq 1}e^{-as}y - \mathbf{1}_{0<\|y\|\leq 1}(y)ds\,.$$

(See e.g. [18] for the definitions and properties related to infinitely divisible laws on $\mathbb{R}^d$, Chapt. 3, Paragraph 17 in particular for such computations, [10] on Banach spaces.)

Remark 3.4. Lemma 3.2 was proven in Lemma 17.1 in Chapt. 3, Paragr. 17, [18], for the real-valued case, however using an approximation by simple functions. The existence of the solution of (3.2) has been proven in Paragr. 17 of [18], only for the case where $(\xi_t)_{t\in[0,T]}$ is of bounded variation, (i.e., has big jumps), while it has been proven on the real line in [9] in an ad hoc way, by defining the stochastic integral of e^{-as} w.r.t. $d\xi_s$ by assuming that an integration by part formula holds, which follows now as a consequence of the Itô formula [17, Theorem 5.1] (or Theorem 2.15 in this article).

Following [18] Theorem 17.5, Chapt. 3, Paragraph 17, it can be shown that if

$$\int_{\|y\|>2} \log\|y\|\beta(dy) < \infty,$$

then the limit distribution μ in (3.5) exists and its Fourier transform $\hat{\mu}$ is

$$\hat{\mu} = \int_0^\infty \psi(e^{-as}x^\star)ds.$$

We note that μ is infinitely divisible as the limit of infinitely divisible Q_t. Similar to [18] one can show that the corresponding Lévy measure $\tilde{\beta}$ and shift γ is

$$\tilde{\beta}(A) := \lim_{t\to\infty} \frac{1}{a} \int_E \beta(dy) \int_0^\infty \mathbf{1}_A(e^{-s}y)ds\,,$$

$$\gamma := \frac{1}{a} \int_{\|y\|>1} \frac{y}{\|y\|}\beta(dy)\,.$$

Appendix A. Lévy processes and Lévy measures

We assume again that $(E, \mathcal{B}(E))$, and $(F, \mathcal{B}(F))$ are separable Banach spaces and use the same notation as in the previous sections. We show how the results in [11, 16] have natural consequences concerning the Lévy processes, which are well known for the real-valued case, e.g., [18]. Let $p=1$, or $p=2$ and $(F, \mathcal{B}(F))$ be a separable Banach space of type 2. We assume that

$$\int_{E_0} \|f(x)\|^p \beta(dx) < \infty, \tag{A.1}$$

then

$$\begin{aligned} \xi_t &:= \int_0^t \int_{\|f(x)\|\le 1} f(x) q(dsdx) + \int_0^t \int_{\|f(x)\|>1} f(x) N(dsdx)(\omega) \\ &= \int_0^t \int_{E_0} f(x) q(dsdx) + \int_0^t \int_{\|f(x)\|>1} f(x)\beta(dx)ds \end{aligned} \tag{A.2}$$

is a Markov process ([11, Theorem 5.2]). In this section we show that any such solution $(\xi_t)_{t\ge 0}$ is a Lévy process. (This implies as a particular case, by taking $f(x)=x$, that any compensated Poisson-Lévy measure is associated to a Lévy process.)

Let us apply the Itô formula ([17, Theorem 5.1] or Theorem 2.15 in this article),

$$\begin{aligned} &\mathcal{H}(\xi_t(\omega)) - \mathcal{H}(\xi_\tau(\omega)) \\ &= \int_\tau^t \int_{0<\|f(x)\|\le 1} \{\mathcal{H}(\xi_{s_-}(\omega)+f(x)) - \mathcal{H}(\xi_{s_-}(\omega))\} \, q(dsdx)(\omega) \\ &\quad + \int_\tau^t \int_{0<\|f(x)\|\le 1} \{\mathcal{H}(\xi_{s_-}(\omega)+f(x)) - \mathcal{H}(\xi_{s_-}(\omega)) - \mathcal{H}'(\xi_{s_-}(\omega))f(x)\} \, ds\beta(dx) \\ &\quad + \int_\tau^t \int_{\{\|f(x)\|>1\}} \{\mathcal{H}(\xi_{s_-}(\omega)+f(x)) - \mathcal{H}(\xi_{s_-}(\omega))\} \, N(dsdx)(\omega) \\ &\quad P-\text{a.s.}, \end{aligned}$$

to $\mathcal{H}(\xi_t(\omega)) = e^{ix^\star(\xi_t)}$, for $x^\star \in E^\star$, $E^\star$ denoting the dual space of E. We get

$$\begin{aligned} &E[e^{ix^\star(\xi(t))} - 1] \\ &= \int_0^t \int_{E_0} E[e^{ix^\star(\xi(s))} \left(e^{ix^\star(f(x))} - 1 - ix^\star(f(x))\mathbf{1}_{\|f(x)\|\le 1}\right) ds\beta(dx) \end{aligned}$$

Let us define $\phi_t(x^\star) = E[e^{ix^\star(\xi(t))}]$, then we get

$$\frac{d}{dt}\phi_t(x^\star) = c\phi_t(x^\star)$$

$$\phi_0(\xi^\star) = 1$$

with

$$c := \int_{E_0} \left(e^{ix^\star(f(x))} - 1 - ix^\star(f(x))\mathbf{1}_{\|f(x)\|\le 1}\right) \beta(dx).$$

Solving the above equation, we get

$$\phi_t(x^\star) = e^{t\int_{E_0}\left(e^{ix^\star(f(x))} - 1 - ix^\star(f(x))\mathbf{1}_{\|f(x)\|\le 1}\beta(dx)\right)}.$$

Thus $(\xi_t)_{t\geq 0}$, solution of (A.2) is a Lévy process with Lévy measure $\beta_f(dx)$, such that $\beta_f(A) = \int_A \beta(f^{-1}(dx))$ for any $A \in \mathcal{B}(E_0)$. Conversely, from the Lévy–Itô decomposition Theorem ([1, Theorem 4.1]) it follows that, given any Lévy process $(\xi_t)_{t\geq 0}$, such that the associated Lévy measure $\beta(dx)$ satisfies (A.1) for $f(x)= x$, then $(\xi_t)_{t\geq 0}$ satisfies (A.2), with $f(x) = x$.

References

[1] S. Albeverio and B. Rüdiger, *The Lévy–Itô decomposition theorem and stochastic integrals, on separable Banach spaces*, Stoch. An. and Appl., **23 (2)** (2005), 217–253.

[2] S. Albeverio, J.-L. Wu, and T.-S. Zhang, *Parabolic SPDEs driven by Poisson white noise*, Stochastic Process. Appl., **74 (1)** (1998), 21–36.

[3] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge Studies in Advanced Mathematics, 93, Cambridge University Press, 2004.

[4] D. Applebaum and J.-L. Wu, *Stochastic partial differential equations driven by Lévy space-time white noise*, Random Operators and Stochastic Equations, **8 (3)** (2000), 245–259.

[5] O. Barndorff-Nielsen and N. Shepard, *Modelling by Lévy processes for financial econometrics*, in: Lévy Processes: Theory and Applications, O. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick, Editors, Birkhäuser Verlag, Basel, (2001), 283–318.

[6] A. Bensoussan and J.-L. Lions, *Contrôle impulsionnel et inéquations quasi variationnelles*, (French) [Impulse control and quasivariational inequalities] Méthodes Mathématiques de l'Informatique [Mathematical Methods of Information Science], 11, Gauthier-Villars, Paris, 1982.

[7] E. Dettweiler, *Banach space-valued processes with independent increments and stochastic integrals*, in: Probability in Banach spaces IV, Proc., Oberwolfach 1982, Springer, Berlin, (1982), 54–83.

[8] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland Mathematical Library, Vol. 24, North Holland Publishing Company, 1989.

[9] Z. J. Jurek and W. Vervaat, *An integral representation for selfdecomposable Banach space-valued random variables*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **62** (1983), 247–262.

[10] W. Linde, *Infinitely Divisible and Stable Measures on Banach spaces*, Teubner-Texte zur Mathematik, Band 58, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1983.

[11] V. Mandrekar and B. Rüdiger, *Existence and uniqueness of pathwise solutions for stochastic differential equations with non Gaussian noise on Banach spaces*, Stochastics, **78 (4)** (2006).

[12] V. Mandrekar and B. Rüdiger, *Lévy noises and stochastic integrals on Banach spaces*, in: Stochastic Partial Differential Equations and Applications - VII, G. Da Prato et al., Editors, Lecture Notes in Pure and Applied Mathematics, CRC Press, (2006), 193–214.

[13] G. Pisier, *Probabilistic methods in the geometry of Banach spaces*, in: Probability and Analysis, Varenna (Como) 1985, G. Letta and M. Pratelli, Editors, LNM 1206, edited by A. Dold, B. Eckmann, Springer Verlag, (1986), 167–241.

[14] G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math., **20** (1975), 326–350.

[15] M. Pratelli, *Intégration stochastique et géométrie des espaces de Banach*, Séminaire de Probabilités XXII, Lecture Notes in Math., 1321, Springer, Berlin, (1988), 129–137.

[16] B. Rüdiger, *Stochastic integration with respect to compensated Poisson random measures on separable Banach spaces*, Stoch. Stoch. Rep., **76 (3)** (2004), 213–242.

[17] B. Rüdiger and G. Ziglio, *The Itô formula for Banach-valued stochastic integrals obtained by integration with respect to compensated Poisson random measures*, Stochastics, **78 (6)** (2006), 377–410.

[18] K. I. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge studies in advanced mathematics, 68, Cambridge University Press, 1999.

[19] A. V. Skorohod, *Studies in the Theory of Random Processes*, Addison-Wesley Publishing Company, Inc, Reading Massachutts. Translated from the Russian by Scripta Technica, 1965.

V. Mandrekar
Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824
USA

B. Rüdiger
Mathematisches Institut
Universität Koblenz-Landau
Campus Koblenz, Universitätsstrasse 1
56070 Koblenz
Germany
and
SFB 611, Institut für Angewandte Mathematik
Abteilung Stochastik
Universität Bonn
Wegelerstr. 6
53115 Bonn
Germany
e-mail: `ruediger@uni-koblenz.de`

Progress in Probability, Vol. 59, 275–303

Approximation of Rough Paths of Fractional Brownian Motion

Annie Millet and Marta Sanz-Solé

Abstract. We consider a geometric rough path associated with a fractional Brownian motion with Hurst parameter $H \in]\frac{1}{4}, \frac{1}{2}[$. We give an approximation result in a modulus type distance, up to the second order, by means of a sequence of rough paths lying above elements of the reproducing kernel Hilbert space.

Mathematics Subject Classification (2000). Primary 60G15; Secondary 60H05, 60H07.

Keywords. Fractional Brownian motion, rough paths.

1. Introduction

Consider a d-dimensional fractional Brownian motion W^H with Hurst parameter $H \in]\frac{1}{4}, \frac{1}{2}[\cup]\frac{1}{2}, 1[$ and integral representation

$$W_t^H = \int_0^1 K^H(t,s)\, dB_s, \tag{1.1}$$

where $K^H(t,s) = 0$, if $s \geq t$ and for $0 < s < t$,

$$K^H(t,s) = c_H\, (t-s)^{H-\frac{1}{2}} + s^{H-\frac{1}{2}} F_1\left(\frac{t}{s}\right) \tag{1.2}$$

with

$$F_1(z) = c_H \left(\frac{1}{2} - H\right) \int_0^{z-1} u^{H-\frac{3}{2}} \left(1 - (u+1)^{H-\frac{1}{2}}\right) du, \tag{1.3}$$

First author was supported by the grants BFM 2003-01345, HF 2003-006, Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain and partially supported by the program SAB 2003-0082, Dirección General de Universidades, Ministerio de Educación y Ciencia, Spain. Second author was supported by the grants BFM 2003-01345, HF 2003-006, Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain.

for $z > 1$ (see, for instance, [1], equation (42)). In (1.1), B denotes a standard d-dimensional Brownian motion and in (1.2), (1.3), c_H denotes a positive real constant depending on H.

Let $p \in]1,4[$ be such that $pH > 1$. In [2], it is proved that the sequence of smooth rough paths based on linear interpolations of W^H converges in the p-variation distance. The limit defines a geometric rough path with roughness p lying above W^H. We will call this object *the enhanced fractional Brownian motion.*

In the recent papers [3, 5], the p-variation distance on rough paths is replaced by a strictly stronger, *modulus type* distance defined as follows:

$$\bar{d}_p(x,y) = \sup_{0\le s<t\le 1}\left(\sum_{i=1}^{[p]} \frac{|x^{(i)}_{s,t} - y^{(i)}_{s,t}|}{(t-s)^{\frac{i}{p}}}\right).$$

In [3], it is proved that the enhanced fractional Brownian motion can actually be obtained by means of the $\bar{d}_p$ distance and also that linear interpolations of W^H define stochastic processes with values in $\mathcal{H}^H$, the reproducing kernel Hilbert space associated with W^H (see Theorem 3.3 in [4] for a description of this space). Then, the authors state a characterization of the topological support of the enhanced fractional Brownian motion among other results.

Our aim in this work is to give a new approximation of the enhanced fractional Brownian motion by means of a sequence of geometric rough paths which, unlike those based on linear interpolations, are not smooth, but also belong to $\mathcal{H}^H$. For the sake of simplicity, we restrict to $[p] = 2$. We are pretty confident that our results extend to $[p] = 3$; however, dealing with higher generality would most likely produce a very technical paper. Our result, as is stated in Theorem 2.1, provides in particular a new approximation of the Lévy area of the fractional Brownian motion.

For any $m \in \mathbb{N}$, we consider the dyadic grid $(t^m_k = k2^{-m}, k = 0,1,\dots,2^m)$ and set $\Delta^m_k =]t^m_{k-1}, t^m_k]$ and $\Delta^m_k B = B_{t^m_k} - B_{t^m_{k-1}}$. Define $B(m)_0 = 0$ and for $t \in \Delta^m_k$, $B(m)_t = B_{t^m_{k-1}} + 2^m(t - t^m_{k-1})\Delta^m_k B$. Our approximation sequence is defined by

$$W(m)^H_t = \int_0^t K^H(t,s)\dot{B}(m)_s\,ds, \tag{1.4}$$

$m \in \mathbb{N}$, where $\dot{B}(m)_s$ denotes the derivative with respect to s of the path $s \mapsto B(m)_s$. Notice that $W(m)^H \in \mathcal{H}^H$.

Let K^H_m be the orthogonal projection of $K^H(t,\cdot)$ on the σ-field generated by $(\Delta^m_k, k = 1,\dots,m)$. That is, for any $0 < s < t \le 1$,

$$K^H_m(t,s) = \sum_{k=1}^{2^m} 2^m \left(\int_{\Delta^m_k \cap]0,t]} K^H(t,u)\,du\right) \mathbb{1}_{\Delta^m_k}(s). \tag{1.5}$$

We clearly have

$$W(m)^H_t = \int_0^1 K^H_m(t,s)\,dB_s. \tag{1.6}$$

For $H \in]\frac{1}{2}, 1[$, we set $\mathbf{W} = (\mathbf{W}_{s,t} = (W^{(1)}_{s,t}, 0 \le s \le t \le 1))$, $\mathbf{W}(\mathbf{m}) = (\mathbf{W}(\mathbf{m})_{s,t} = (W(m)^{(1)}_{s,t}, 0 \le s \le t \le 1))$, while for $H \in]\frac{1}{4}, \frac{1}{2}[$ we set $\mathbf{W} = (\mathbf{W}_{s,t} = (W^{(1)}_{s,t}, W^{(2)}_{s,t}, 0 \le s \le t \le 1))$ and $\mathbf{W}(\mathbf{m}) = (\mathbf{W}(\mathbf{m})_{s,t} = (W(m)^{(1)}_{s,t}, W(m)^{(2)}_{s,t}, 0 \le s \le t \le 1))$, $m \ge 1$.

The main result of the paper states the convergence of $\mathbf{W}(\mathbf{m})$ to $\mathbf{W}$ in the $\bar{d}_p$-metric for $p \in]1, 3[$. For $p \in]1, 2[$, the result is an almost trivial consequence of Lemma 3.2 which establishes Hölder continuity in the $L^2[0,1]$-norm of the kernels K^H, K^H_m, respectively, and a control of the quadratic mean error in the approximation of K^H by K^H_m. For $p \in [2, 3[$, the approximation of the Lévy area relies on representation formulas for the second-order multiple integrals by means of the operator K^* given in (2.3) and introduced in [1] (see also [2]). There are two fundamental ingredients. Firstly, Proposition 2.3, giving the rate of convergence of the approximation at the second-order level in the $L^q(\Omega)$-modulus norm; secondly, Lemma 3.5, an extension of the Garsia-Rademich-Rumsey Lemma for geometric rough paths of any roughness p. Other technical results used in the proofs, mainly on the kernels K^H and K^H_m, are given in the Appendix.

For simplicity, in general we shall not write explicitly the dependence on H; thus W stands for W^H, $K(t,s)$ for $K^H(t,s)$, etc. For any $q \in [1, \infty[$, we denote by $\|\cdot\|_q$ the $L^q(\Omega)$-norm. We make the convention $\sum_{k=a}^b x_k = 0$ if $b < a$ and denote by C positive constants with possibly different values. For additional notions and notation on rough paths, we refer the reader to [6].

2. Approximation result

For $p \in]1, +\infty[$ we set $\tilde{d}_p = \bar{d}_{p\wedge 2}$, that is,

$$\tilde{d}_p(x,y) = \sup_{0 \le s < t \le 1} \left(\sum_{i=1}^{[p]\wedge 2} \frac{|x^{(i)}_{s,t} - y^{(i)}_{s,t}|}{(t-s)^{\frac{i}{p}}} \right).$$

The purpose of this section is to prove the following approximation result.

Theorem 2.1. *Let $H \in]\frac{1}{4}, \frac{1}{2}[$, $p \in]2, 4[$ (resp. $H \in]\frac{1}{2}, 1[$, $p \in]1, 2[$), be such that $pH > 1$ and $q \in [1, +\infty[$. The sequence $\left(\tilde{d}_p(\mathbf{W}(\mathbf{m}), \mathbf{W}), m \ge 1\right)$, converges to 0 in $L^q(\Omega)$ and a.s. Thus for $H \in]\frac{1}{2}, 1[$ and $p \in]1, 2[$, if $\mathcal{G}_p$ denotes the set of dyadic geometric rough paths endowed with the norm $\tilde{d}_p(0, .)$ and P^H denotes the law of the fractional Brownian motion W^H, then the triple $(X, \mathcal{H}^H, P^H)$ is an abstract Wiener space.*

The next proposition provides the auxiliary result to state the approximation of the first component of the enhanced fractional Brownian motion.

Proposition 2.2. *Let* $0 \le s < t \le 1$, $q \in [1,\infty[$.

(i) *For any* $H \in]0, \frac{1}{2}[$, $\lambda \in [0,H[$,

$$\left\| W^{(1)}_{s,t} - W(m)^{(1)}_{s,t} \right\|_q \le C 2^{-m\lambda} |t-s|^{H-\lambda}. \tag{2.1}$$

(ii) *For any* $H \in]\frac{1}{2}, 1[$, $\varepsilon \in [0,H[$, $\mu \in]0, \frac{\varepsilon}{H(2H+1)}[$,

$$\left\| W^{(1)}_{s,t} - W(m)^{(1)}_{s,t} \right\|_q \le C 2^{-m\mu} |t-s|^{H-\varepsilon}. \tag{2.2}$$

Proof. By the hypercontractivity inequality, it suffices to prove the results for $q = 2$. In this case, it is an easy consequence of the identity

$$E\left(\left| W^{(1)}_{s,t} - W(m)^{(1)}_{s,t} \right|^2 \right) = \int_0^1 \left| (K(t,u) - K(s,u)) - (K_m(t,u) - K_m(s,u)) \right|^2 du$$

and of Lemma 3.2. Indeed, by (3.14), we have

$$E\left(\left| W^{(1)}_{s,t} - W(m)^{(1)}_{s,t} \right|^2 \right) \le C|t-s|^{2H}.$$

Hence, if $t - s < 2^{-m}$, we easily obtain (2.1) and (2.2).

Assume now $H \in]0, \frac{1}{2}[$ and $t - s \ge 2^{-m}$. By (3.15), for $\epsilon \in [0,H]$,

$$E\left(\left| W^{(1)}_{s,t} - W(m)^{(1)}_{s,t} \right|^2 \right) \le C 2^{-2mH} \le C 2^{-2m\epsilon} |t-s|^{2(H-\epsilon)}.$$

Hence, (2.1) follows.

Let $H \in]\frac{1}{2}, 1[$ and $t - s \ge 2^{-m}$. Let $\alpha \in]0,1[$; then (3.14) and (3.16) imply

$$\left\| W^{(1)}_{s,t} - W(m)^{(1)}_{s,t} \right\|_q \le C|t-s|^{H(1-\alpha)} 2^{-m\lambda\alpha},$$

with $\lambda \in]0, \frac{1}{2H+1}[$. By taking $\alpha = \frac{\varepsilon}{H}$, we obtain (2.2) with $\mu = \lambda \frac{\varepsilon}{H}$. $\square$

Throughout the rest of this section, $H \in]\frac{1}{4}, \frac{1}{2}[$.

Following [1], let $\mathcal{H}_K$ denote the set of functions $\varphi : [0,1] \to \mathbb{R}$ such that

$$||\varphi||_K^2 = \int_0^1 \varphi(s)^2 K(1,s)^2 ds + \int_0^1 ds \left(\int_s^1 |\varphi(t) - \varphi(s)| \, |K|(dt,s) \right)^2 < +\infty.$$

For any $\varphi \in \mathcal{H}_K$, $0 < s < t$, set

$$\begin{aligned} K^* \left(\mathbb{1}_{]s,t]}(\cdot) \left(\varphi_\cdot - \varphi_s \right) \right)(u) &= \mathbb{1}_{]0,s]}(u) \int_s^t \left(\varphi_r - \varphi_s \right) K(dr,u) \\ &+ \mathbb{1}_{]s,t]}(u) \left(K(t,u) \left(\varphi_u - \varphi_s \right) + \int_u^t \left(\varphi_r - \varphi_u \right) K(dr,u) \right). \end{aligned} \tag{2.3}$$

Following [2],

$$W^{(2)}_{s,t} = \int_0^1 K^* \left(\mathbb{1}_{]s,t]}(\cdot) \left(W_\cdot - W_s \right) \right)(u) dB_u + \frac{1}{2}|t-s|^{2H}.$$

Moreover, by Theorem 9 in [7], for $W(m)$ defined in (1.4) we have

$$W(m)^{(2)}_{s,t} = \int_0^1 K^* \left(\mathbb{1}_{]s,t]}(\cdot)\left(W(m)_\cdot - W(m)_s\right)\right)(u)\dot{B}(m)_u du.$$

Proposition 2.3. *For each $m \in \mathbb{N}$, $0 < s < t \leq 1$, $q \in [1, \infty[$,*

$$\|W^{(2)}_{s,t} - W(m)^{(2)}_{s,t}\|_q \leq C 2^{-m\mu}|t-s|^{2H-\varepsilon}, \tag{2.4}$$

for some positive constants C and any $\varepsilon \in]0, 2H - \frac{1}{2}[$ and $\mu \in]0, \frac{\varepsilon}{2}[$.

Before proving this proposition, we give an equivalent expression for $W(m)^2_{s,t}$, as follows. The integration by parts formula of Malliavin calculus (see, e.g., [8], Equation (1.49)) and (1.6) yield $W(m)^{(2)}_{s,t} = A^1_{s,t}(m) + A^2_{s,t}(m)$, with

$$A^1_{s,t}(m) = \sum_{k=1}^{2^m} \int_0^1 du\, \mathbb{1}_{\Delta^m_k}(u) 2^m K^* \left(\mathbb{1}_{]s,t]}(\cdot)\int_{\Delta^m_k} dB_r \left(W(m)_\cdot - W(m)_s\right)\right)(u), \tag{2.5}$$

$$A^2_{s,t}(m) = \sum_{k=1}^{2^m} \int_0^1 du\, \mathbb{1}_{\Delta^m_k}(u) 2^m K^* \left(\mathbb{1}_{]s,t]}(\cdot)\int_{\Delta^m_k} dr \left(K_m(\cdot, r) - K_m(s,r)\right)\right)(u). \tag{2.6}$$

By definition, for $r \in \Delta^m_k$, $K_m(t,r) = 2^m \int_{\Delta^m_k \cap]0,t]} K(t,u)du = 2^m K(\mathbb{1}_{\Delta^m_k})(t)$. Since $h := K\left(\mathbb{1}_{\Delta^m_k}\right) \in \mathcal{H}_K$, the duality relation given in [7], equation (58) and Lemma 3.3 yield

$$\begin{aligned}
A^2_{s,t}(m) &= \sum_{k=1}^{2^m} \int_0^1 dr\, \mathbb{1}_{\Delta^m_k}(r) 2^{2m} \int_0^1 du\, \mathbb{1}_{\Delta^m_k}(u) K^* \left(\mathbb{1}_{]s,t]}(\cdot) K\left(\mathbb{1}_{\Delta^m_k}\right)_{s,\cdot}\right)(u) \\
&= \sum_{k=1}^{2^m} \int_0^1 dr\, \mathbb{1}_{\Delta^m_k}(r) 2^{2m} \int_s^t K\left(\mathbb{1}_{\Delta^m_k}\right)(du)\left(K\left(\mathbb{1}_{\Delta^m_k}\right)(u) - K\left(\mathbb{1}_{\Delta^m_k}\right)(s)\right) \\
&= \sum_{k=1}^{2^m} \int_0^1 dr \mathbb{1}_{\Delta^m_k}(r) 2^{2m} \frac{\left(K\left(\mathbb{1}_{\Delta^m_k}\right)(t) - K\left(\mathbb{1}_{\Delta^m_k}\right)(s)\right)^2}{2} \\
&= \frac{1}{2}\int_0^1 dr\, |K_m(t,r) - K_m(s,r)|^2 = \frac{1}{2}\|W(m)^{(1)}_{s,t}\|^2_2.
\end{aligned}$$

Thus, since $E|W_t - W_s|^2 = |t-s|^{2H}$, Schwarz's inequality, (3.14), (3.15) imply

$$\left| A^2_{s,t}(m) - \frac{1}{2}|t-s|^{2H} \right| \leq C 2^{-m\varepsilon}|t-s|^{2H-\varepsilon}, \tag{2.7}$$

for some positive constant C and $\varepsilon \in]0, H[$.

Hence, in order to establish (2.4) it suffices to prove that for any small parameter $\varepsilon \in]0, 4H-1[$ and $\mu \in]0, \varepsilon[$,

$$E\left(\left|\int_0^1 K^*\left(1\!\!1_{]s,t]}(\cdot)\left(W_\cdot - W_s\right)\right)(u)dB_u - A^1_{s,t}(m)\right|^2\right) \le C2^{-m\mu}|t-s|^{4H-\epsilon}. \quad (2.8)$$

for all $m \ge 1$. We devote the next lemmas to the proof of this convergence, using the expression of the operator K^* given in (2.3).

Lemma 2.4. *For any $0 \le s < t \le 1$, $m \ge 1$, we set*

$$T_1(s,t) = \int_0^s dB_u \left(\int_s^t (W_r - W_s)\, K(dr,u)\right),$$
$$T_1(s,t,m) = \sum_{k=1}^{2^m} \int_{\Delta_k^m} dB_r\, 2^m \left(\int_{\Delta_k^m \cap]0,s]} du \left(\int_s^t (W(m)_v - W(m)_s)\, K(dv,u)\right)\right).$$

Then for any $\epsilon \in]0, 2H[$ and $\mu \in]0, \epsilon[$, there exists $C > 0$ such that

$$E\left(|T_1(s,t,m) - T_1(s,t)|^2\right) \le C2^{-m\mu}|t-s|^{4H-\epsilon}. \quad (2.9)$$

Proof. Assume $s \in \Delta_I^m$, $I \ge 1$; we consider the decomposition

$$E\left(|T_1(s,t,m) - T_1(s,t)|^2\right) \le C\sum_{j=1}^3 \tau_{1,j}(s,t,m),$$

with

$$\tau_{1,1}(s,t,m) = \sum_{k\in\{1,I-1,I\}} E\Bigg(\Bigg|\int_{\Delta_k^m} dB_r\, 2^m \times \left(\int_{\Delta_k^m\cap]0,s]} du \left(\int_s^t (W(m)_v - W(m)_s)\, K(dv,u)\right)\right)\Bigg|^2\Bigg), \quad (2.10)$$

$$\tau_{1,2}(s,t,m) = \sum_{k\in\{1,I-1,I\}} E\left(\left|\int_{\Delta_k^m\cap]0,s]} dB_r \left(\int_s^t (W_v - W_s)\, K(dv,r)\right)\right|^2\right), \quad (2.11)$$

$$\tau_{1,3}(s,t,m) = E\Bigg(\Bigg|\sum_{k=2}^{I-2}\int_{\Delta_k^m} dB_r\, 2^m \int_{\Delta_k^m} du \Big(\int_s^t (W(m)_v - W(m)_s)\, K(dv,u) - \int_s^t (W_v - W_s)K(dv,r)\Big)\Bigg|^2\Bigg). \quad (2.12)$$

By Lemma 3.4, (3.4), Schwarz's inequality and (3.14), any term on the right-hand side of (2.10) is bounded as follows. Let $\varepsilon \in]0, 2H[$, $\lambda \in]\frac{1-(2H-\varepsilon)}{2}, \frac{1}{2}[$; then $2H - 3 + 2\lambda < -1$, $1 - 2\lambda - (2H - \varepsilon) < 0$ and

$$
\begin{aligned}
&E\left(\left|\int_{\Delta_k^m} dB_r\, 2^m \left(\int_{\Delta_k^m \cap]0,s]} du \left(\int_s^t (W(m)_v - W(m)_s)\, K(dv, u)\right)\right)\right|^2\right) \\
&\le C \int_{\Delta_k^m} dr \int_0^1 d\rho \left|2^m \int_{\Delta_k^m \cap]0,s]} du \int_s^t (K_m(v,\rho) - K_m(s,\rho))\, K(dv, u)\right|^2 \\
&\le C \int_{\Delta_k^m} dr \int_0^1 d\rho\, 2^m \int_{\Delta_k^m \cap]0,s]} du \left(\int_s^t dv\, |v-u|^{2H-3+2\lambda}\right) \\
&\quad \times \left(\int_s^t dv\, |K_m(v,\rho) - K_m(s,\rho)|^2 |v-u|^{-2\lambda}\right) \\
&\le C \int_{\Delta_k^m \cap]0,s]} du\, (s-u)^{2H-2+2\lambda} |t-s|^{2H} \left(|t-u|^{1-2\lambda} - |s-u|^{1-2\lambda}\right) \\
&\le C \int_{\Delta_k^m \cap]0,s]} du\, (s-u)^{2H-2+2\lambda} |t-s|^{2H}\, |t-s|^{2H-\varepsilon}\, |s-u|^{1-2\lambda-(2H-\varepsilon)} \\
&\le C|t-s|^{4H-\varepsilon} \int_{\Delta_k^m \cap]0,s]} du\, |s-u|^{\varepsilon-1} \le C 2^{-m\varepsilon} |t-s|^{4H-\varepsilon}.
\end{aligned}
$$

Each term of the right-hand side of (2.11) can be studied using a similar strategy. Thus we obtain for $\varepsilon \in]0, 2H[$:

$$
\tau_{1,1}(s,t,m) + \tau_{1,2}(s,t,m) \le C 2^{-m\varepsilon} |t-s|^{4H-\varepsilon}. \tag{2.13}
$$

Set for $s \ge 3 \cdot 2^{-m}$, and, hence, $I \ge 4$,

$$
\begin{aligned}
X_r = \sum_{k=2}^{I-2} \mathbb{1}_{\Delta_k^m}(r) 2^m \int_{\Delta_k^m} du \Big(&\int_s^t (W(m)_v - W(m)_s)\, K(dv, u) \\
&- \int_s^t (W_v - W_s) K(dv, r)\Big).
\end{aligned}
$$

Notice that $X_r = \int_0^1 g(r,\rho)\, dB_\rho$, with

$$
\begin{aligned}
g(r,\rho) = \sum_{k=2}^{I-2} \mathbb{1}_{\Delta_k^m}(r) 2^m \int_{\Delta_k^m} du \Big(&\int_s^t K(dv,u)\, (K_m(v,\rho) - K_m(s,\rho)) \\
&- \int_s^t K(dv, r)\, (K(v,\rho) - K(s,\rho))\Big).
\end{aligned}
$$

Hence, by Lemma 3.4 and Schwarz's inequality, $\tau_{1,3}(s,t,m) \le C(\tau_{1,3,1}(s,t,m) + \tau_{1,3,2}(s,t,m))$, with

$$\tau_{1,3,1}(s,t,m) = \sum_{k=2}^{I-2} 2^m \int_{\Delta_k^m} dr \int_{\Delta_k^m} du \int_0^1 d\rho \times \Big| \int_s^t \big(K_m(v,\rho) - K_m(s,\rho)\big)\big(K(dv,u) - K(dv,r)\big)\Big|^2,$$

$$\tau_{1,3,2}(s,t,m) = \sum_{k=2}^{I-2} \int_{\Delta_k^m} dr \int_0^1 d\rho \Big| \int_s^t K(dv,r) \times \big(K_m(v,\rho) - K_m(s,\rho) - K(v,\rho) + K(s,\rho)\big)\Big|^2.$$

Owing to (3.4), (3.7), we have for $\lambda \in]0,1[$, $u, r \in \Delta_k^m$,

$$\begin{aligned}
&\left|\frac{\partial K}{\partial v}(v,u) - \frac{\partial K}{\partial v}(v,r)\right| \\
&\quad \le C \left|\frac{\partial K}{\partial v}(v,u) - \frac{\partial K}{\partial v}(v,r)\right|^{\lambda} \left(\left|\frac{\partial K}{\partial v}(v,u)\right|^{1-\lambda} + \left|\frac{\partial K}{\partial v}(v,r)\right|^{1-\lambda}\right) \\
&\quad \le C 2^{-m\lambda} |v - (u \vee r)|^{H-\frac{3}{2}} \left[(u \wedge r)^{-1} + |v - (u \vee r)|^{-1}\right]^{\lambda}. \qquad (2.14)
\end{aligned}$$

Thus, taking $\lambda := H$ yields $\tau_{1,3,1}(s,t,m) \le C 2^{-2mH} \sum_{j=1}^{2} \tau_{1,3,1,j}(s,t,m)$, with

$$\tau_{1,3,1,1}(s,t,m) = \sum_{k=2}^{I-2} 2^m \int_{\Delta_k^m} dr \int_{\Delta_k^m} du \int_0^1 d\rho \Big(\int_s^t dv |K_m(v,\rho) - K_m(s,\rho)| \times |v - (u \vee r)|^{H-\frac{3}{2}} (u \wedge r)^{-H}\Big)^2,$$

$$\tau_{1,3,1,2}(s,t,m) = \sum_{k=2}^{I-2} 2^m \int_{\Delta_k^m} dr \int_{\Delta_k^m} du \int_0^1 d\rho \Big(\int_s^t dv |K_m(v,\rho) - K_m(s,\rho)| \times |v - (u \vee r)|^{-\frac{3}{2}}\Big)^2.$$

Let $a = 2 - \epsilon$, with $\epsilon \in]0, 2H[$. Schwarz's inequality along with (3.14) yield

$$\begin{aligned}
\tau_{1,3,1,1}(s,t,m) &\le C \sum_{k=2}^{I-2} 2^m \int_{\Delta_k^m} dr \int_{\Delta_k^m} du \left(\int_s^t dv |v - (u \vee r)|^{-a} dv\right) \\
&\quad \times \left(\int_s^t dv |v - s|^{4H-3+a} |u \wedge r|^{-2H}\right) \\
&\le C |t-s|^{4H-\epsilon} \int_{t_1^m}^{t_{I-2}^m} du (s - \overline{u}_m)^{\epsilon-1} (\underline{u}_m)^{-2H} \\
&\le C |t-s|^{4H-\epsilon} s^{\epsilon-2H} \le |t-s|^{4H-\epsilon} 2^{-m(\epsilon-2H)}. \qquad (2.15)
\end{aligned}$$

Indeed, $\int_s^t dv|v-(u\vee r)|^{-2+\epsilon} \le C(s-\overline{u}_m)^{\epsilon-1}$ for $\overline{u}_m$ defined by (3.13). Let $\epsilon\in]0,2H[$ using Schwarz's inequality and (3.14), we obtain

$$\begin{aligned}
\tau_{1,3,1,2}(s,t,m) &\le C\sum_{k=2}^{I-2} 2^m \int_{\Delta_k^m} dr \int_{\Delta_k^m} du\Big(\int_s^t dv|v-(u\vee r)|^{-2-2H+\varepsilon}\Big)\\
&\qquad \times \Big(\int_s^t dv|v-(u\vee r)|^{2H-\varepsilon-1}|v-s|^{2H}\Big)\\
&\le C|t-s|^{4H-\varepsilon}\int_{t_1^m}^{t_{I-2}^m} du\int_s^t dv(v-\overline{u}_m)^{-2-2H+\varepsilon}\\
&\le C|t-s|^{4H-\varepsilon}\int_{t_1^m}^{t_{I-2}^m} du(s-\overline{u}_m)^{-1-2H+\varepsilon}\\
&\le C|t-s|^{4H-\varepsilon}2^{-m(\varepsilon-2H)}. \qquad (2.16)
\end{aligned}$$

From (2.15), (2.16) we deduce that for $\epsilon\in]0,2H[$,

$$\tau_{1,3,1}(s,t,m) \le C|t-s|^{4H-\epsilon}\, 2^{-m\epsilon}. \qquad (2.17)$$

Let $\delta\in]0,2H[$, $\alpha\in]0,2H[$, $\lambda\in]0,1[$ and $\mu\in]\frac{1}{2},1-H[$. Notice that for these choices, $-2\mu+1-2H+\delta<0$. Hölder's inequality together with (3.14) and (3.15) yield for any $\lambda\in]0,1[$,

$$\tau_{1,3,2}(s,t,m) \le C\tau_{1,3,2,1}(s,t,m)^{\lambda}\,\tau_{1,3,2,2}(s,t,m)^{1-\lambda},$$

where

$$\tau_{1,3,2,1}(s,t,m) = \int_{t_1^m}^{t_{I-2}^m} dr\left(\int_s^t dv(v-r)^{2H-3+2\mu}\right)\left(\int_s^t dv(v-r)^{-2\mu}(v-s)^{2H}\right),$$

$$\tau_{1,3,2,2}(s,t,m) = \int_{t_1^m}^{t_{I-2}^m} dr\left(\int_s^t dv(v-r)^{2H-3+2\mu}\right)\left(\int_s^t dv(v-r)^{-2\mu}2^{-2mH}\right).$$

For the first term we have

$$\begin{aligned}
\tau_{1,3,2,1}(s,t,m) &\le C|t-s|^{4H-\delta}\int_{t_1^m}^{t_{I-2}^m} dr(s-r)^{2H-2+2\mu}(s-r)^{-2\mu+1-2H+\delta}\\
&\le C|t-s|^{4H-\delta},
\end{aligned}$$

while for the second one, we obtain

$$\tau_{1,3,2,2}(s,t,m) \le C2^{-2mH}|t-s|^{2H-\alpha}\int_{t_1^m}^{t_{I-2}^m} dr(s-r)^{2H-2+2\mu}(s-r)^{-2\mu+1-2H+\alpha}.$$

Consequently,

$$\tau_{1,3,2}(s,t,m) \le C\,|t-s|^{(4H-\delta)\lambda+(2H-\alpha)(1-\lambda)}\,2^{-2mH(1-\lambda)}.$$

Take α,δ arbitrarily small and $1-\lambda=\frac{\epsilon-H\delta}{2H+\alpha}$. Then for $\beta<\epsilon<2H$, we have proved that

$$\tau_{1,3,2}(s,t,m) \le C|t-s|^{4H-\epsilon}2^{-m\beta}.$$

This inequality, together with (2.13) and (2.17) yields (2.9). □

Lemma 2.5. *For any $0 \le s < t \le 1$, set*

$$T_2(s,t) = \int_s^t dB_u\, K(t,u)(W_u - W_s)\,,$$
$$T_2(s,t,m) = \sum_{k=1}^{2^m} \int_{\Delta_k^m} dB_r\, 2^m \left(\int_{\Delta_k^m \cap]s,t]} du K(t,u)\, (W(m)_u - W(m)_s) \right).$$

Then, for $b \in]0, 2H[$, there exists a constant $C > 0$ such that for each $m \ge 1$

$$E\left(|T_2(s,t,m) - T_2(s,t)|^2 \right) \le C 2^{-mb} |t-s|^{4H-b}. \tag{2.18}$$

Proof. Let $s \in \Delta_I^m$, $t \in \Delta_J^m$. We have

$$E\left(|T_2(s,t,m) - T_2(s,t)|^2 \right) \le C \sum_{j=1}^{3} T_{2,j}(s,t,m),$$

with, for $\mathcal{I} = \{I, I+1, J-2, J-1J\}$,

$$T_{2,1}(s,t,m) = \sum_{k \in \mathcal{I}} E\Big(\Big| \int_{\Delta_k^m} dB_r\, 2^m \int_{\Delta_k^m \cap]s,t]} du K(t,u)\, (W(m)_u - W(m)_s) \Big|^2 \Big),$$
$$T_{2,2}(s,t,m) = \sum_{k \in \mathcal{I}} E\Big(\Big| \int_{\Delta_k^m \cap]s,t]} dB_r\, K(t,r)(W_r - W_s) \Big|^2 \Big),$$
$$T_{2,3}(s,t,m) = E\Big(\Big| \sum_{k=I+2}^{J-3} \int_{\Delta_k^m} dB_r \Big[2^m \int_{\Delta_k^m \cap]s,t]} du K(t,u)\, (W(m)_u - W(m)_s)$$
$$- K(t,r)(W_r - W_s) \Big] \Big|^2 \Big).$$

Owing to Lemma 3.4 applied to the Gaussian process

$$X_r := 1\!\!1_{\Delta_k^m}(r) \int_0^1 dB_\rho \left(2^m \int_{\Delta_k^m \cap]s,t]} du K(t,u) \big(K_m(u,\rho) - K_m(s,\rho) \big) \right)$$

and Schwarz's inequality, we have for any $k = 1, \dots, 2^m$,

$$T(s,t,m,k) := E\left(\left| \int_{\Delta_k^m} dB_r\, 2^m \int_{\Delta_k^m \cap]s,t]} du K(t,u)\, (W(m)_u - W(m)_s) \right|^2 \right)$$
$$\le C 2^{2m} \int_{\Delta_k^m} dr \int_0^1 d\rho \left(\int_{\Delta_k^m \cap]s,t]} du K^2(t,u) \right)$$
$$\times \left(\int_{\Delta_k^m \cap]s,t]} du\, |K_m(u,\rho) - K_m(s,\rho)|^2 \right).$$

Let $k = I, I+1$; since $\int_{\Delta_k^m \cap]s,t]} du K^2(t,u) \le \int_{]s,t]} du K^2(t,u) \le C|t-s|^{2H}$, we have for any $b \in]0, 2H[$,

$$
\begin{aligned}
T(s,t,m,k) &\le C 2^m |t-s|^{2H} \left(\int_{\Delta_k^m \cap]s,t]} du |u-s|^{2H} \right) \\
&\le C 2^m |t-s|^{4H-b} \int_{\Delta_k^m \cap]s,t]} du |u-s|^b \le C 2^{-mb} |t-s|^{4H-b}.
\end{aligned}
$$

Let $k = J-2, J-1, J$ with $J-2 > I+1$, then, for $u \in \Delta_k^m$, (3.5) implies $|K(t,u)|^2 \le C|t-u|^{2H-1}$ and $|t-u| \le C2^{-m}$; we obtain for $b \in]0, 2H[$,

$$
\begin{aligned}
T(s,t,m,k) \le C 2^m &\left(\int_{\Delta_k^m \cap]s,t]} du |t-u|^{2H-1-b} 2^{-mb} du \right) \left(\int_{\Delta_k^m \cap]s,t]} du |u-s|^{2H} \right) \\
&\le C |t-s|^{4H-b} 2^{-mb}.
\end{aligned}
$$

We therefore have proved that for $b \in]0, 2H[$,

$$
T_{2,1}(s,t,m) \le C 2^{-bm} |t-s|^{4H-b}.
$$

The analysis of the term $T_{2,2}(s,t,m)$ is easier. Indeed, the isometry property of the stochastic integral yields for any $k = 1, \ldots, 2^m$,

$$
E\left(\left| \int_{\Delta_k^m \cap]s,t]} dB_r\, K(t,r)(W_r - W_s) \right|^2 \right) = C \int_{\Delta_k^m \cap]s,t]} dr\, K^2(t,r) |r-s|^{2H}. \tag{2.19}
$$

For the particular values of $k \in \mathcal{I}$, the right-hand side of (2.19) can be analyzed following similar ideas as for $T_{2,1}(s,t,m)$, which yields for $b \in]0, 2H[$

$$
T_{2,2}(s,t,m) \le C 2^{-mb} |t-s|^{4H-b}.
$$

We now study $T_{2,3}(s,t,m)$ and note that $T_{2,3}(s,t,m) = 0$ if $|t-s| \le 2^{-m}$. Thus, we may assume that $t - s \ge 2^{-m}$. First, we apply Lemma 3.4 and obtain

$$
T_{2,3}(s,t,m) \le C(T_{2,3,1}(s,t,m) + T_{2,3,2}(s,t,m)),
$$

where

$$
\begin{aligned}
T_{2,3,1}(s,t,m) = \int_{\overline{s}_m}^{\underline{t}_m - 2^{1-m}} dr \int_0^1 d\rho \Big| 2^m \int_{\underline{r}_m}^{\overline{r}_m} du \Big(K(t,u) - K(t,r) \Big) \\
\times \Big(K_m(u,\rho) - K_m(s,\rho) \Big) \Big|^2, \\
T_{2,3,2}(s,t,m) = \int_{\overline{s}_m}^{\underline{t}_m - 2^{1-m}} dr \int_0^1 d\rho \Big| 2^m \int_{\underline{r}_m}^{\overline{r}_m} du K(t,r) \\
\times \Big(\big[K_m(u,\rho) - K_m(s,\rho) \big] - \big[K(r,\rho) - K(s,\rho) \big] \Big) \Big|^2.
\end{aligned}
$$

By Schwarz's inequality and (3.14), for $b \in]0, 2H[$,

$$\begin{aligned}
T_{2,3,1}(s,t,m) &\le \int_{\overline{s}_m}^{\underline{t}_m - 2^{1-m}} dr 2^m \int_{\underline{r}_m}^{\overline{r}_m} du |K(t,u) - K(t,r)|^2 |u-s|^{2H} \\
&\le C|t-s|^{2H} \int_{\overline{s}_m}^{\underline{t}_m - 2^{1-m}} dr 2^m \int_{\underline{r}_m}^{\overline{r}_m} du |K(t,u) - K(t,r)|^2 \\
&\le C 2^{-2mH} |t-s|^{2H} \le C 2^{-mb} |t-s|^{4H-b},
\end{aligned}$$

where the last inequalities follow from (3.19) and $|t-s| \ge 2^{-m}$.

Owing to (3.15), we have for $u \in [\underline{r}_m, \overline{r}_m]$,

$$\begin{aligned}
\int_0^1 d\rho |K_m(s,\rho) - K(s,\rho)|^2 \le & C 2^{-2mH}, \\
\int_0^1 d\rho |K_m(u,\rho) - K(r,\rho)|^2 \le & C \int_0^1 d\rho \Big(|K_m(u,\rho) - K(u,\rho)|^2 \\
& + |K(u,\rho) - K(r,\rho)|^2 \Big) \le C 2^{-2mH}.
\end{aligned}$$

Schwarz's inequality, along with (3.5) and the above estimates yield

$$\begin{aligned}
T_{2,3,2}(s,t,m) &\le C \int_{\overline{s}_m}^{\underline{t}_m - 2^{1-m}} dr 2^{-2mH} \left(|r|^{2H-1} + |t-r|^{2H-1} \right) \\
&\le C 2^{-2mH} \left(t^{2H} - s^{2H} + |t-s|^{2H} + 2^{-2mH} \right) \\
&\le C 2^{-2mH} |t-s|^{2H} \le C 2^{-mb} |t-s|^{4H-b}
\end{aligned}$$

for $b \in]0, 2H[$. Indeed, for each $H \in]0, \frac{1}{2}[$, and $s < t$, $t^{2H} - s^{2H} \le (t-s)^{2H}$ and we are assuming that $2^{-m} < |t-s|$. Thus, (2.18) is proved. □

Lemma 2.6. *For any $0 \le s < t \le 1$, set*

$$\begin{aligned}
T_3(s,t) &= \int_s^t dB_u \int_u^t K(dr,u)(W_r - W_u), \\
T_3(s,t,m) &= \sum_{k=1}^{2^m} 2^m \int_{\Delta_k^m} dB_r \int_{\Delta_k^m \cap]s,t]} du \int_u^t K(dv,u) \left(W(m)_v - W(m)_u \right).
\end{aligned}$$

There exists a positive constant C such that, for any $\epsilon \in]0, 4H-1[$,

$$E \left(|T_3(s,t,m) - T_3(s,t)|^2 \right) \le C 2^{-m\epsilon} |t-s|^{4H-\epsilon}, \tag{2.20}$$

for each $m \ge 1$.

Proof. Assume $s \in \Delta_I^m$, $t \in \Delta_J^m$; we consider the upper bound

$$E \left(|T_3(s,t,m) - T_3(s,t)|^2 \right) \le C \sum_{j=1}^3 T_{3,j}(s,t,m),$$

where for $\mathcal{J} = \{I, I+1, J-1, J\}$,

$$T_{3,1}(s,t,m) = \sum_{k \in \mathcal{J}} E\Big(\big| 2^m \int_{\Delta_k^m} dB_r \int_{\Delta_k^m \cap]s,t]} du \int_u^t K(dv,u) \times (W(m)_v - W(m)_u) \big|^2 \Big), \tag{2.21}$$

$$T_{3,2}(s,t,m) = \sum_{k \in \mathcal{J}} E\Big(\Big| \int_{\Delta_k^m \cap]s,t]} dB_r \int_r^t K(dv,r)(W_v - W_r) \Big|^2 \Big), \tag{2.22}$$

$$T_{3,3}(s,t,m) = E\Big(\Big| \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dB_r \int_{\Delta_k^m} du \times \Big(\int_u^t K(dv,u)(W(m)_v - W(m)_u) - \int_r^t K(dv,r)(W_v - W_r) \Big) \Big|^2 \Big).$$

Lemma 3.4 along with Schwarz's inequality yield for each term of the sum on the right-hand side of (2.21) the upper bound

$$C \int_{\Delta_k^m} dr \int_0^1 d\rho \, 2^m \int_{\Delta_k^m \cap]s,t]} du \left(\int_u^t K(dv,u) \big(K_m(v,\rho) - K_m(u,\rho) \big) \right)^2 .$$

Fix $a \in]2-4H, 1]$. From Schwarz's inequality, (3.4) and (3.14) we deduce the following estimates for this integral:

$$C \int_{\Delta_k^m} dr 2^m \int_{\Delta_k^m \cap]s,t]} du \left(\int_u^t dv |v-u|^{-a} \right) \left(\int_u^t |v-u|^{4H-3+a} \right) \leq C \left(2^{-m} \wedge |t-s| \right) |t-s|^{4H-1}.$$

A similar analysis yields the same result for each term on the right-hand side of (2.22). Consequently,

$$T_{3,1}(s,t,m) + T_{3,2}(s,t,m) \leq C \left(2^{-m} \wedge |t-s| \right) |t-s|^{4H-1}. \tag{2.23}$$

If $|t-s| \leq 2^{-m}$, then $T_{3,3}(s,t,m) = 0$. Hence, let us assume that $t - s \geq 2^{-m}$; in this case $T_{3,3}(s,t,m)$ is equal to $E\left(\int_0^1 dB_r X_r \right)^2$, with $X_r = \int_0^1 dB_\rho g(r,\rho)$, and

$$g(r,\rho) = \sum_{k=I+2}^{J-2} \mathbb{1}_{\Delta_k^m}(r) 2^m \int_{\Delta_k^m} du \Big[\int_u^t K(dv,u)(K_m(v,\rho) - K_m(u,\rho)) - \int_r^t K(dv,r)(K(v,\rho) - K(r,\rho)) \Big].$$

We at first study the contribution to $T_{3,3}(s,t,m)$ of the integrands

$$g_1(r,\rho) = \sum_{k=I+2}^{J-2} \mathbb{1}_{\Delta_k^m}(r) 2^m \int_{\Delta_k^m} du \int_u^{u\vee r} K(dv,u)(K_m(v,\rho) - K_m(u,\rho)),$$
$$g_2(r,\rho) = \sum_{k=I+2}^{J-2} \mathbb{1}_{\Delta_k^m}(r) 2^m \int_{\Delta_k^m} du \int_r^{u\vee r} K(dv,r)(K(v,\rho) - K(r,\rho)),$$

which we denote by $T_{3,3,j}(s,t,m)$, $j=1,2$. Actually, both are similar and therefore we only study the first one. Lemma 3.4, (3.4), (3.14) and Schwarz's inequality imply, for each $a \in]2-4H,1]$,

$$\begin{aligned} T_{3,3,1}(s,t,m) &\le C \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{\Delta_k^m} du \int_u^{u\vee r} dv |v-u|^{-a} \int_u^{u\vee r} dv |v-u|^{4H-3+a} \\ &\le C 2^{-m(4H-1)} |t-s|. \end{aligned} \tag{2.24}$$

We end the analysis of the term $T_{3,3}(s,t,m)$ by studying the contribution of $T_{3,3,3}((s,t,m)$ defined in terms of the integrand

$$\begin{aligned} g_3(r,\rho) = \sum_{k=I+2}^{J-2} \int_{\Delta_k^m} dr 2^m \int_{\Delta_k^m} du \int_{u\vee r}^t \Big[& K(dv,u)(K_m(v,\rho) - K_m(u,\rho)) \\ & - K(dv,r)(K(v,\rho) - K(r,\rho))\Big]. \end{aligned}$$

Notice that $g_3(r,\rho)$ is the sum of two analogous terms where the set Δ_k^m of the integral with respect to the variable u is replaced by $[\underline{r}_m, r[$, $[r, \overline{r}_m[$, respectively. Again, the contribution of both terms is similar, so that we concentrate on the first one. That is, we consider

$$\begin{aligned} T_{3,3,3}^+(s,t,m) := E\Big(\Big| \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dB_r \int_{[\underline{r}_m, r[} du \int_r^t \Big[& K(dv,u) \\ & \times (W(m)_v - W(m)_u) - K(dv,r)(W_v - W_r)\Big]\Big|^2\Big). \end{aligned}$$

As before, all the arguments rely on Lemma 3.4, (3.4), (3.14), a suitable factorization of the integrands along with Schwarz's inequality. In order to deal with the singularity at $v=r$, we first replace the integral with respect to the variable v by $\int_r^{\overline{r}_m + 2^{-m}}$. Given $a \in]2-4H,1[$, the corresponding contribution to $T_{3,3,3}^+(s,t,m)$ is bounded by

$$
\begin{aligned}
& C\sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \int_0^1 d\rho \Big(\Big| \int_r^{\overline{r}_m+2^{-m}} K(dv,u) \\
& \quad \times (K_m(v,\rho)-K_m(u,\rho)) \Big|^2 + \Big| \int_r^{\overline{r}_m+2^{-m}} K(dv,r)(K(v,\rho)-K(r,\rho)) \Big|^2 \Big) \\
& \le C\sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \int_r^{\overline{r}_m+2^{-m}} dv |v-r|^{-a} \int_r^{\overline{r}_m+2^{-m}} dv |v-r|^{4H-3+a} \\
& \le C 2^{-m(4H-1)} |t-s|. \qquad (2.25)
\end{aligned}
$$

Let us finally consider the range $]r_m + 2^{-m}, t[$ for the variable v. We have to study two terms:

$$
\begin{aligned}
M_1(s,t,m) &= \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \int_0^1 d\rho \Big(\int_{\overline{r}_m+2^{-m}}^t dv \\
&\qquad \times |K_m(v,\rho)-K_m(u,\rho)| \left| \frac{\partial K}{\partial v}(v,u) - \frac{\partial K}{\partial v}(v,r) \right| \Big)^2, \\
M_2(s,t,m) &= \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \int_0^1 d\rho \Big(\int_{\overline{r}_m+2^{-m}}^t dv \Big| \frac{\partial K}{\partial v}(v,r) \Big| \\
&\qquad \times \big[(K_m(v,\rho)-K_m(u,\rho)) - (K(v,\rho)-K(r,\rho)) \big] \Big)^2.
\end{aligned}
$$

For $M_1(s,t,m)$, we proceed in a similar way as for the term $\tau_{1,3,1}(s,t,m)$ in Lemma 2.4, as follows. By means of (2.14) we obtain for $\lambda \in]0,1[$ $M_1(s,t,m) \le C2^{-2m\lambda}\left(M_{1,1}(s,t,m) + M_{1.2}(s,t,m)\right)$, with

$$
\begin{aligned}
M_{1,1}(s,t,m) &= \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du\, u^{-2\lambda} \int_0^1 d\rho \Big(\int_{\overline{r}_m+2^{-m}}^t dv \\
&\qquad |K_m(v,\rho)-K_m(u,\rho)| |v-r|^{H-\frac{3}{2}} \Big)^2, \\
M_{1,2}(s,t,m) &= \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \int_0^1 d\rho \Big(\int_{\overline{r}_m+2^{-m}}^t dv \\
&\qquad |K_m(v,\rho)-K_m(u,\rho)| v-r|^{H-\frac{3}{2}-\lambda} \Big)^2.
\end{aligned}
$$

Let $a \in]2-4H, 1[$, $\lambda \in]0, \frac{1}{2}[$. Since $t-s \ge 2^{-m}$, for $u \in [\underline{r}_m, r[$, we have

$$
\int_{\overline{r}_m+2^{-m}}^t dv |v-r|^{2H-3+a} |v-u|^{2H} \le C|t-r|^{4H+a-2}.
$$

Consequently, since $r \geq u \geq \underline{r}_m \geq t_{I+1}$ implies $u \geq \frac{r}{2}$,

$$\begin{aligned} M_{1,1}(s,t,m) &\leq C \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du\, u^{-2\lambda} \Big(\int_{\overline{r}_m+2^{-m}}^t dv |v-r|^{-a} \Big) \\ &\quad \times \Big(\int_{\overline{r}_m+2^{-m}}^t dv |v-r|^{2H-3+a} |v-u|^{2H} \Big) \\ &\leq C \int_s^t r^{-2\lambda} |t-s|^{4H-1} dr \leq C|t-s|^{4H-2\lambda}. \end{aligned} \tag{2.26}$$

Analogously, for $b \in]2+2\lambda-4H, 1[$, $\lambda \in]0, 2H-\frac{1}{2}[$ and $|t-s| \geq 2^{-m}$,

$$\begin{aligned} M_{1,2}(s,t,m) &\leq C \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \Big(\int_{\overline{r}_m+2^{-m}}^t dv |v-r|^{-b} \Big) \\ &\quad \times \Big(\int_{\overline{r}_m+2^{-m}}^t dv |v-r|^{2H-3-2\lambda+b} |v-u|^{2H} \Big) \\ &\leq C \int_s^t |t-r|^{4H-1-2\lambda} dr = C|t-s|^{4H-2\lambda}. \end{aligned} \tag{2.27}$$

Finally, if we additionally use (3.15), we obtain for $a \in]2-4H, 1[$,

$$\begin{aligned} M_2(s,t,m) &\leq C \sum_{k=I+2}^{J-2} 2^m \int_{\Delta_k^m} dr \int_{[\underline{r}_m, r[} du \Big(\int_r^t dv |v-r|^{-a} \Big) \\ &\quad \times \Big(\int_{\underline{r}_m+2^{-m}}^t dv |v-r|^{2H-3+a} 2^{-2mH} \Big) \\ &\leq C \int_s^t |t-r|^{1-a}\, 2^{-m(4H-2+a)} dr \leq C 2^{-mb} |t-s|^{4H-b} \end{aligned} \tag{2.28}$$

for $b \in]0, 4H-1[$. We easily check that (2.20) follows from (2.23)–(2.28). □

Proof of Proposition 2.3. We remark that Lemmas 2.4 to 2.6 yield the upper bound (2.8). Therefore, for $q = 2$, (2.4) follows from (2.7) and (2.8). The hypercontractivity inequality yields the validity of the same inequality for any $q \in]2, \infty[$. □

Proof of Theorem 2.1. Let $H \in]\frac{1}{2}, 1[$ and $p \in]\frac{1}{H}, 2[$. The convergence of $\tilde{d}_p(\mathbf{W}(\mathbf{m}), \mathbf{W})$ to zero in $L^q(\Omega)$ is a consequence of (2.2) and the usual version of the Garsia-Rademich-Rumsey lemma (see, e.g., [9], Theorem 2.1.3).

Consider the metric space $(\mathcal{G}_p, \tilde{d}_p)$. The canonical embedding $\mathcal{H}^H \hookrightarrow \mathcal{G}_p$ is continuous. Indeed, let $\dot{h}_i$, $i = 1, 2$, belong to $L^2([0,1])$. Then for $h_i(.) = \int_0^{\cdot} K(.,r)\dot{h}_i(r)dr$ and $0 \leq s < t \leq 1$,

$$|(h_1)_{s,t}^{(1)} - (h_2)_{s,t}^{(1)}| \leq |t-s|^H \|\dot{h}_1 - \dot{h}_2\|_2 \leq |t-s|^{\frac{1}{p}} \|h_1 - h_2\|_{\mathcal{H}^H}.$$

Consequently, the preceding convergence shows that $(\mathcal{G}_p, \mathcal{H}^H, P^H)$ is an abstract Wiener space.

Let now $H \in]\frac{1}{4}, \frac{1}{2}[$. We follow the outline of the proof of Lemma 3 in [3], but refer to the extension of the Garsia-Rademich-Rumsey lemma stated in Lemma 3.5.

Fix $p \in]2,4[$ such that $pH > 1$. We shall prove that there exists $\theta > 0$ such that for every $q \in [1,\infty[$,

$$E\left(\left|\tilde{d}_p(\mathbf{W}, \mathbf{W(m)})\right|^q\right) \leq C_q 2^{-m\theta q}.$$

Indeed, for a fixed $q \in [1,\infty[$, let $M > q$ and $N = 2M$ satisfy $N > \frac{p}{2(Hp-1)}$. Let $\alpha, \beta > 0$ defined by $\alpha = \frac{2}{p} + \frac{1}{M}$, $\beta = \frac{1}{p} + \frac{1}{N}$.

By virtue of (2.1) and (2.4), we easily check that the random variables

$$\begin{aligned} A_1(m) : &= \int_0^1 \int_0^1 dsdt 1_{\{s<t\}} \frac{|W^{(1)}_{s,t} - W(m)^{(1)}_{s,t}|^{2N}}{|t-s|^{2N\beta}}, \\ A_2(m) : &= \int_0^1 \int_0^1 dsdt 1_{\{s<t\}} \frac{|W^{(2)}_{s,t} - W(m)^{(2)}_{s,t}|^{2M}}{|t-s|^{2M\alpha}}, \end{aligned}$$

satisfy

$$E\big(A_1(m)\big) \leq C 2^{-m\mu 2N}, \; E\big(A_2(m)\big) \leq C 2^{-m\mu 2M}, \tag{2.29}$$

for some $\mu > 0$.

Furthermore, the hypercontractivity property and the inequality (3.14) imply that for $0 \leq s < t \leq 1$ and $q \in [1,\infty[$,

$$\sup_m \left(\|W^{(1)}_{s,t}\|_q + \|W(m)^{(1)}_{s,t}\|_q\right) \leq C\, |t-s|^H.$$

This yields

$$\sup_m E\big(\eta(m)\big) \leq C\,, \tag{2.30}$$

where

$$\eta(m) := \int_0^1 \int_0^1 dsdt 1_{\{s<t\}} \frac{|W^{(1)}_{s,t}|^{2N} + |W(m)^{(1)}_{s,t}|^{2N}}{|t-s|^{2N\beta}}.$$

By Lemma 3.5, we deduce that for any $0 \leq s < t \leq 1$,

$$|W^{(1)}_{s,t} - W(m)^{(1)}_{s,t}| \leq C\, A_1(m)^{\frac{1}{2N}}\, |t-s|^{\frac{1}{p}}, \tag{2.31}$$

$$|W^{(2)}_{s,t} - W(m)^{(2)}_{s,t}| \leq C \left[A_2(m)^{\frac{1}{2M}} + A_1(m)^{\frac{1}{2N}}\, \eta(m)^{\frac{1}{2N}}\right] |t-s|^{\frac{2}{p}}. \tag{2.32}$$

Finally, Schwarz's and Hölder's inequalities together with (2.29)–(2.32) conclude the proof of the theorem. □

3. Appendix

Let $W^H = (W^H_t, t \in [0,1])$ be a d-dimensional fractional Brownian motion with Hurst parameter $H \in]0, \frac{1}{2}[\cup]\frac{1}{2}, 1[$ and integral representation given in (1.1).

Assume $H \in]\frac{1}{2}, 1[$; by computing the integral of the right-hand side of (1.3), we obtain the following expression for the kernel K^H defined in (1.2):

$$K^H(t,s) = c_H \left(H - \frac{1}{2}\right) s^{H-\frac{1}{2}} F_2\left(\frac{t}{s}\right), \tag{3.1}$$

where for $z > 1$,

$$F_2(z) = \int_0^{z-1} u^{H-\frac{3}{2}} (u+1)^{H-\frac{1}{2}}\, du. \tag{3.2}$$

From (1.2), it follows that

$$\frac{\partial K^H}{\partial t}(t,s) = c_H \left(H - \frac{1}{2}\right) \left(\frac{s}{t}\right)^{\frac{1}{2}-H} (t-s)^{H-\frac{3}{2}}. \tag{3.3}$$

holds for any $H \in]0, \frac{1}{2}[\cup]\frac{1}{2}, 1[$ and $0 < s < t < 1$. Consequently, for $H \in]0, \frac{1}{2}[$,

$$\left|\frac{\partial K^H}{\partial t}(t,s)\right| \le C|t-s|^{H-\frac{3}{2}}. \tag{3.4}$$

The next lemma collects some technical estimates on the kernel $K^H(t,s)$.

Lemma 3.1. *Let $0 < s < t < 1$.*

(1) *Assume $H \in]0, \frac{1}{2}[$. Then,*

$$|K^H(t,s)| \le C\left(s^{H-\frac{1}{2}} \mathbb{1}_{]0,\frac{t}{2}[}(s) + (t-s)^{H-\frac{1}{2}} \mathbb{1}_{[\frac{t}{2},t[}(s)\right), \tag{3.5}$$

$$\left|\frac{\partial K^H}{\partial s}(t,s)\right| \le C\left(s^{H-\frac{3}{2}} \mathbb{1}_{]0,\frac{t}{2}[}(s) + (t-s)^{H-\frac{3}{2}} \mathbb{1}_{[\frac{t}{2},t[}(s)\right), \tag{3.6}$$

$$\left|\frac{\partial^2 K^H}{\partial t \partial s}(t,s)\right| \le C(t-s)^{H-\frac{3}{2}}\left(s^{-1} \mathbb{1}_{]0,\frac{t}{2}[}(s) + (t-s)^{-1} \mathbb{1}_{[\frac{t}{2},t[}(s)\right). \tag{3.7}$$

(2) *For $H \in]\frac{1}{2}, 1[$,*

$$|K^H(t,s)| \le C\left((t-s)^{H-\frac{1}{2}} \mathbb{1}_{]0,\frac{t}{2}[}(s) + s^{H-\frac{1}{2}} \mathbb{1}_{[\frac{t}{2},t[}(s)\right), \tag{3.8}$$

$$\left|\frac{\partial K^H}{\partial s}(t,s)\right| \le C(t-s)^{2H-1}\left(s^{-(H+\frac{1}{2})} \mathbb{1}_{]0,\frac{t}{2}[}(s) + (t-s)^{-(H+\frac{1}{2})} \mathbb{1}_{[\frac{t}{2},t[}(s)\right). \tag{3.9}$$

Proof. Assume first $H \in]0, \frac{1}{2}[$. It is easy to check that, for any $u > 0$,

$$0 < 1 - (u+1)^{H-\frac{1}{2}} \le \left(\left(\frac{1}{2} - H\right) u\right) \wedge 1.$$

Hence, for $0 < s < t$, $0 < u < \frac{t}{s} - 1$,

$$\begin{aligned} u^{H-\frac{3}{2}}\left(1 - (u+1)^{H-\frac{1}{2}}\right) \le\ & C u^{H-\frac{1}{2}} \mathbb{1}_{]0,1\wedge(\frac{t}{s}-1)[}(u) \\ & + C u^{H-\frac{3}{2}} \mathbb{1}_{]1\wedge(\frac{t}{s}-1),\frac{t}{s}-1[}(u). \end{aligned} \tag{3.10}$$

Thus, from (1.3), (3.10), it follows that

$$\left|F_1\left(\frac{t}{s}\right)\right| \le C\int_0^{\frac{t}{s}-1} u^{H-\frac{1}{2}}du \le C,$$

for $\frac{t}{2} \le s < t$, while for $0 < s < \frac{t}{2}$,

$$\left|F_1\left(\frac{t}{s}\right)\right| \le C\int_0^1 u^{H-\frac{1}{2}}\, du + C\int_1^\infty u^{H-\frac{3}{2}}\, du \le C.$$

Consequently,

$$\sup_{0\le s<t}\left|F_1\left(\frac{t}{s}\right)\right| \le C, \tag{3.11}$$

and the identity (1.2) yields (3.5).

By differentiating with respect to the variable s in (1.2) and using (3.11), we obtain

$$\left|\frac{\partial K^H}{\partial s}(t,s)\right| \le C\left(|t-s|^{H-\frac{3}{2}} + s^{H-\frac{3}{2}} + s^{-1}t|t-s|^{H-\frac{3}{2}}\right),$$

which yields (3.6). The inequality (3.7) follows by differentiating with respect to the variable s in (3.3).

Suppose now $H \in]\frac{1}{2}, 1[$. Consider the function F_2 given in (3.2). Clearly, if $\frac{t}{s} - 1 \le 1$, that is, if $\frac{t}{2} \le s < t$,

$$\left|F_2\left(\frac{t}{s}\right)\right| \le C.$$

Assume $\frac{t}{s} - 1 > 1$. For any $u \in]1, \frac{t}{s} - 1[$, $(1+u)^{H-\frac{1}{2}} \le Cu^{H-\frac{1}{2}}$. Consequently,

$$\left|F_2\left(\frac{t}{s}\right)\right| \le C\left(\int_0^1 u^{H-\frac{3}{2}}du + \int_1^{\frac{t}{s}-1} u^{2H-2}du\right) \le C\left(\frac{t}{s}\right)^{2H-1}.$$

The previous upper bounds, together with the representation of the kernel K^H given in (3.1), imply

$$\begin{aligned}|K^H(t,s)| &\le C\left(s^{H-\frac{1}{2}}\left(\frac{t}{s}\right)^{2H-1} \mathbb{1}_{]0,\frac{t}{2}[}(s) + s^{H-\frac{1}{2}}\,\mathbb{1}_{[\frac{t}{2},t[}(s)\right)\\ &\le \left(s^{H-\frac{1}{2}}\,\mathbb{1}_{]0,\frac{t}{2}[}(s) + s^{-H+\frac{1}{2}}(t-s)^{2H-1}\,\mathbb{1}_{]0,\frac{t}{2}[}(s) + s^{H-\frac{1}{2}}\,\mathbb{1}_{[\frac{t}{2},t[}(s)\right)\end{aligned}$$

and (3.8) follows.

Differentiating with respect to the variable s in (3.1) yields

$$\begin{aligned}\left|\frac{\partial K^H}{\partial s}(t,s)\right| &\le C\left(s^{H-\frac{3}{2}}F_2\left(\frac{t}{s}\right) + s^{H-\frac{1}{2}}\frac{t}{s^2}\left(\frac{t}{s}-1\right)^{H-\frac{3}{2}}\left(\frac{t}{s}\right)^{H-\frac{1}{2}}\right)\\ &\le C\Big(s^{H-\frac{3}{2}}\left(\frac{t}{s}\right)^{2H-1}\mathbb{1}_{]0,\frac{t}{2}[}(s) + s^{-(H+\frac{1}{2})}t^{H+\frac{1}{2}}(t-s)^{H-\frac{3}{2}}\\ &\qquad + s^{H-\frac{3}{2}}\,\mathbb{1}_{[\frac{t}{2},t[}(s)\Big),\end{aligned}$$

where in the last inequality we have applied the upper bounds for F_2 obtained before. Replacing in the last expression t^{2H-1} by $C(s^{2H-1} + (t-s)^{2H-1})$ and $t^{H+\frac{1}{2}}$ by $C(s^{H+\frac{1}{2}} + (t-s)^{H+\frac{1}{2}})$, respectively, yields

$$\left|\frac{\partial K^H}{\partial s}(t,s)\right| \leq C\left(s^{H-\frac{3}{2}} + (t-s)^{H-\frac{3}{2}} + s^{-(H+\frac{1}{2})}(t-s)^{2H-1}\right). \tag{3.12}$$

If $0 < s < \frac{t}{2}$ then, $s < t-s$ and $(t-s)^{H-\frac{3}{2}} < s^{H-\frac{3}{2}} < s^{-(H+\frac{1}{2})}(t-s)^{2H-1}$, while for $\frac{t}{2} \leq s < t$, the previous inequalities are reversed accordingly. Hence (3.9) clearly follows from (3.12). □

We introduce the notation

$$\underline{t}_m = [2^m t]2^{-m} \quad \text{and} \quad \bar{t}_m = \underline{t}_m + 2^{-m}, \tag{3.13}$$

for any $m \in \mathbb{N}$. Notice that, K^H_m given in (1.5) satisfies $K^H_m(t,s) = 0$ if $s \geq \bar{t}_m$.

In the next result, we give a bound for the approximation in quadratic mean of the kernel K^H by its projection K^H_m.

Lemma 3.2. (1) *Let $H \in]0, \frac{1}{2}[\cup]\frac{1}{2}, 1[$. There exists a positive constant C such that for any $0 < s < t \leq 1$,*

$$\sup_{m\geq 1}\int_0^1 \left(\left|K^H_m(t,u) - K^H_m(s,u)\right|^2 + \left|K^H(t,u) - K^H(s,u)\right|^2\right) du \leq C|t-s|^{2H}. \tag{3.14}$$

(2) *For $H \in]0, \frac{1}{2}[$,*

$$\int_0^1 \left|K^H(t,u) - K^H_m(t,u)\right|^2 du \leq C\left(t \wedge 2^{-m}\right)^{2H}. \tag{3.15}$$

(3) *For $H \in]\frac{1}{2}, 1[$ and any $\lambda \in]0, \frac{1}{2H+1}[$,*

$$\int_0^1 \left|K^H(t,u) - K^H_m(t,u)\right|^2 du \leq C\, 2^{-2m\lambda} t^{2(H-\lambda)}. \tag{3.16}$$

Proof. The operator π_m is a contraction on $L^2[0,1]$. Thus,

$$\begin{aligned}\sup_{m\geq 1}\int_0^1 &\left(\left|K^H_m(t,u) - K^H_m(s,u)\right|^2 + \left|K^H(t,u) - K^H(s,u)\right|^2\right) du \\ &\leq 2\int_0^1 \left|K^H(t,u) - K^H(s,u)\right|^2 du = 2E\left(|W^H_t - W^H_s|^2\right) = 2|t-s|^{2H},\end{aligned}$$

proving (3.14).

By the same argument,

$$\int_0^1 \left|K^H(t,u) - K^H_m(t,u)\right|^2 du \leq 4\int_0^1 |K^H(t,u)|^2\, du = 4\, t^{2H}. \tag{3.17}$$

Therefore (3.15) holds for $t \leq C\, 2^{-m}$.

Fix $t \in \Delta_I^m$ with $I > 7$. We assume first $H \in]0, \frac{1}{2}[$. Consider the decomposition

$$\int_0^1 \left| K^H(t,u) - K_m^H(t,u) \right|^2 \, du \le C \sum_{i=1}^5 T_i(t), \tag{3.18}$$

with

$$T_1(t) = \int_0^{t_2^m} \left| K^H(t,u) - K_m^H(t,u) \right|^2 du,$$

$$T_2(t) = \int_{t_{I-3}^m}^{t_I^m} \left| K^H(t,u) - K_m^H(t,u) \right|^2 du,$$

$$T_3(t) = \sum_{k=3}^{[2^{m-1}t]} \int_{\Delta_k^m} \left| K^H(t,u) - K_m^H(t,u) \right|^2 du,$$

$$T_4(t) = \sum_{k=[2^{m-1}t]+2}^{I-3} \int_{\Delta_k^m} \left| K^H(t,u) - K_m^H(t,u) \right|^2 du,$$

$$T_5(t) = \int_{\Delta^m_{[2^{m-1}t]+1}} \left| K^H(t,u) - K_m^H(t,u) \right|^2 du.$$

Schwarz's inequality and (3.5) imply

$$T_1(t) \le 4 \int_0^{t_2^m} |K^H(t,u)|^2 \, du \le C \int_0^{t_2^m} u^{2H-1} \, du = C \, 2^{-2mH}.$$

Similarly,

$$T_2(t) \le 4 \int_{t_{I-3}^m}^{t_I^m} |K^H(t,u)|^2 \, du \le C \int_{t_{I-3}^m}^{t} |t-u|^{2H-1} \, du = C \, 2^{-2mH}.$$

Let $\lambda \in]H, 1[$ and $k = 3, \ldots, [2^{m-1}t]$, which implies $\Delta_k^m \subset]0, \frac{t}{2}[$. By Schwarz's inequality, the mean value theorem and (3.5), (3.6), we obtain

$$\begin{aligned}
\int_{\Delta_k^m} \left| K^H(t,u) - K_m^H(t,u) \right|^2 \, du &\le 2^m \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \left| K^H(t,u) - K^H(t,v) \right|^2 \\
&\le 2^m \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \left| K^H(t,u) - K^H(t,v) \right|^{2\lambda} \left| |K^H(t,u)| + |K^H(t,v)| \right|^{2(1-\lambda)} \\
&\le C 2^{-m(2\lambda-1)} \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \, \left((u \wedge v)^{2H-1-2\lambda} \right).
\end{aligned}$$

For $u, v \in \Delta_k^m$, $u \wedge v \ge u - 2^{-m}$; thus,

$$T_3(t) \le C 2^{-2m\lambda} \int_{t_2^m}^{t^m_{[2^{m-1}t]}} du \, (u - 2^{-m})^{2H-1-2\lambda} \le C 2^{-2mH}.$$

Fix now $k = [2^{m-1}t] + 2, \dots, I-3$, so that $\Delta_k^m \subset [\frac{t}{2}, t[$. In this case

$$\int_{\Delta_k^m} \left|K^H(t,u) - K_m^H(t,u)\right|^2 du \le C2^{-m(2\lambda-1)} \\ \times \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \, (t-(u \vee v))^{2H-1-2\lambda}.$$

Since for $u, v \in \Delta_k^m$, $t - (u \vee v) \ge t - u - 2^{-m} \ge t_{I-2}^m - u$, the previous estimate implies

$$T_4(t) \le C2^{-2m\lambda} \int_{t_{[2^{m-1}t]}^m}^{t_{I-3}^m} du \, (t_{I-2}^m - u)^{2H-1-2\lambda} \le C2^{-2mH}.$$

We study the term $T_5(t)$ using the same method as for $T_3(t)$, $T_4(t)$, as follows:

$$\begin{aligned} T_5(t) &\le 2^m \int_{\Delta_{[2^{m-1}t]+1}^m} du \int_{\Delta_{[2^{m-1}t]+1}^m} dv \left|K^H(t,u) - K^H(t,v)\right|^2 \\ &\le C2^{-m(2\lambda-1)} \int_{\Delta_{[2^{m-1}t]+1}^m} !du \int_{\Delta_{[2^{m-1}t]+1}^m} dv \left((u \wedge v)^{H-\frac{3}{2}} + (t-(u \vee v)^{H-\frac{3}{2}}\right)^{2\lambda} \\ &\quad \times \left((u \wedge v)^{H-\frac{1}{2}} + (t-(u \vee v)^{H-\frac{1}{2}}\right)^{2(1-\lambda)}. \end{aligned}$$

For $u, v \in \Delta_{[2^{m-1}t]+1}^m$, $u \wedge v > \frac{t}{2} - 2^{-m}$, $u \vee v < \frac{t}{2} + 2^{-m}$ and $t - (u \vee v) > \frac{t}{2} - 2^{-m}$. Thus, the last integral is bounded by

$$\int_{\Delta_{[2^{m-1}t]+1}^m} du \int_{\Delta_{[2^{m-1}t]+1}^m} dv \left(\frac{t}{2} - 2^{-m}\right)^{2H-1-2\lambda}.$$

Moreover, since we are assuming that $t \in \Delta_I^m$, with $I > 7$, $\frac{t}{2} - 2^{-m} \ge 2^{-m+1}$. Thus, we finally obtain for $\lambda = \frac{1}{2}$,

$$T_5(t) \le C2^{-2mH}.$$

Then (3.15) follows from the upper bounds obtained so far for $T_i(t)$, $i = 1, \dots, 5$.

Notice that we have also proved that for $H \in]0, \frac{1}{2}[$,

$$\sum_{k=3}^{I-3} 2^m \int_{\Delta_k^m} du \int_{\Delta_k^m} dv |K^H(t,u) - K^H(t,v)|^2 \le C2^{-2mH}. \tag{3.19}$$

Assume now $H \in]\frac{1}{2}, 1[$ and fix $\lambda \in]0, \frac{1}{2H+1}[$, so that $H - \lambda > 0$. Since the inequality (3.17) holds for any $H \in]0, \frac{1}{2}[\cap]\frac{1}{2}, 1[$, (3.16) holds for any $t \le C2^{-m}$. Let now $t \in \Delta_I^m$, with $I > 7$. We apply a similar method as we used in the case

$H \in]0, \frac{1}{2}[$, using the decomposition (3.18). In fact, owing to (3.8),

$$T_1(t) \le C \int_0^{t_2^m} (t-u)^{2H-1}\, du \le C2^{-m} t^{2H-1},$$
$$T_2(t) \le C \int_{t_{I-3}^m}^{t_I^m} u^{2H-1}\, du \le C2^{-m} t^{2H-1}.$$

Fix $k = 3, \ldots, [2^{m-1}t]$. Schwarz's inequality, along with the mean value theorem and (3.8), (3.9), imply

$$\begin{aligned}\int_{\Delta_k^m} \left|K^H(t,u) - K_m^H(t,u)\right|^2 du &\le 2^m \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \left|K^H(t,u) - K^H(t,v)\right|^{2\lambda} \\ &\quad \times \left||K^H(t,u)| + |K^H(t,v)|\right|^{2(1-\lambda)} \\ &\le C2^{-m(2\lambda-1)} \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \left((t-(u\wedge v)\right)^{(\lambda+1)(2H-1)} (u\wedge v)^{-\lambda(2H+1)} \\ &\le C2^{-2m\lambda}\, t^{(\lambda+1)(2H-1)} \int_{\Delta_k^m} du\, (u - 2^{-m})^{-\lambda(2H+1)}.\end{aligned}$$

Since $\lambda < \frac{1}{2H+1}$, we have

$$T_3(t) \le C2^{-2m\lambda}\, t^{2(H-\lambda)}.$$

Let now $k = [2^{m-1}t] + 2, \ldots, I-3$. With similar arguments as before, we deduce

$$\begin{aligned}\int_{\Delta_k^m} \left|K^H(t,u) - K_m^H(t,u)\right|^2 du &\le 2^m \int_{\Delta_k^m} du \int_{\Delta_k^m} dv \left|K^H(t,u) - K^H(t,v)\right|^{2\lambda} \\ &\quad \times \left||K^H(t,u)| + |K^H(t,v)|\right|^{2(1-\lambda)} \\ &\le C2^{-m(2\lambda-1)} \int_{\Delta_k^m} du \int_{\Delta_k^m} dv (t-(u\vee v))^{\lambda(2H-3)} (u\vee v)^{(1-\lambda)(2H-1)} \\ &\le C2^{-2m\lambda}\, t^{(1-\lambda)(2H-1)} \int_{\Delta_k^m} du\, (t-u-2^{-m})^{\lambda(2H-3)}.\end{aligned}$$

For $\lambda < \frac{1}{2H+1}$, $\lambda(2H-3)+1 > 0$. Hence,

$$T_4(t) \le C2^{-2m\lambda} t^{(1-\lambda)(2H-1)} \int_{\frac{t}{2}}^{t_{I-3}^m} (t-u-2^{-m})^{\lambda(2H-3)} \le C2^{-2m\lambda}\, t^{2(H-\lambda)}.$$

Finally, we study the contribution of $T_5(t)$ as follows.

$$\begin{aligned} T_5(t) &\le 2^m \int_{\Delta^m_{[2^{m-1}t]+1}} du \int_{\Delta^m_{[2^{m-1}t]+1}} dv \left|K^H(t,u) - K^H(t,v)\right|^2 \\ &\le C 2^{-m(2\lambda-1)} \int_{\Delta^m_{[2^{m-1}t]+1}} du \int_{\Delta^m_{[2^{m-1}t]+1}} dv \Big((t-(u\wedge v))^{2H-1} \\ &\quad \times \left((u\wedge v)^{-(H+\frac{1}{2})} + (t-(u\vee v))^{-(H+\frac{1}{2})} \right) \Big)^{2\lambda} \\ &\quad \times \left(\big(t-(u\wedge v)\big)^{H-\frac{1}{2}} + (u\vee v)^{H-\frac{1}{2}} \right)^{2(1-\lambda)}. \end{aligned}$$

For $u, v \in \Delta^m_{[2^{m-1}t]+1}$, $u\wedge v > C_1 t$, $u \vee v < C_2 t$, $t-(u\wedge v) < C_3 t$ and $t-(u\vee v) > C_4 t$. Thus,

$$T_5(t) \le C\, 2^{-m(2\lambda-1)} 2^{-2m} t^{2(H-\lambda)-1} \le C 2^{-2m\lambda}\, t^{2(H-\lambda)}$$

The estimates obtained so far imply (3.16). □

In the next Lemma we prove a simple extension of a well-known integration formula for bounded variation functions.

Lemma 3.3. *For any $h \in \mathcal{H}$, $t \ge 0$,*

$$\int_0^t h(u)h(du) = \frac{h^2(t)}{2}, \tag{3.20}$$

where the integral is understood in the sense of Proposition 5 *in* [7].

Proof. Let $n \ge 1$ and let $h(n)$ be the function obtained by linear interpolation on the n-th dyadic grid of h. We have proved in [7], Theorem 9 that

$$\lim_{n\to\infty} \int_0^t h(n)(u)h(n)(du) = \int_0^t h(u)h(du),$$

for any $t \ge 0$. Since (3.20) is true with h replaced by $h(n)$, the result follows. □

The following result gives an upper bound for the L^2 norm of a Skorohod integral of a Gaussian process.

Lemma 3.4. *Let $X_t = \int_0^1 g(t,s)dB_s$, $t \in [0,1]$, with g a deterministic function belonging to $L^2([0,1]^2)$. Then, the Skorohod integral $\int_0^1 X_s dB_s$ satisfies*

$$E\left(\int_0^1 X_s dB_s\right)^2 \le C \int_0^1 ds \int_0^1 dr |g(s,r)|^2. \tag{3.21}$$

Proof. The isometry property of the Skorohod integral ([8], Equation (1.48)) yields

$$E\left(\int_0^1 X_s dB_s\right)^2 \le C \int_0^1 E(X_s)^2 ds + \int_0^1 ds \int_0^1 dr E(|D_r X_s|^2).$$

Since $E(X_s)^2 = \int_0^1 |g(s,r)|^2 dr$ and the Malliavin derivative $D_r X_s$ is equal to $g(s,r)$, (3.21) follows. □

We conclude this section by proving an extension of the Garsia-Rademich-Rumsey lemma used to estimate $\tilde{d}_p(X,Y)$ when X and Y are geometric rough paths with roughness $p \in [2,\infty[$ (see [6], Definition 3.3.3).

Lemma 3.5. *Let X and Y be geometric rough paths with the same roughness $p \in [2,+\infty[$. Set $k=[p]$. For $i=1,\dots,k$, let $M_i \geq 1$, $\alpha_i = \frac{i}{p}+\frac{1}{M_i}$. Suppose that*

$$\int_0^1\int_0^1 dsdt 1_{\{s\leq t\}} \frac{|X_{s,t}^{(i)}|^{2M_i}+|Y_{s,t}^{(i)}|^{2M_i}}{|t-s|^{2M_i\alpha_i}} \leq A_i, \quad 1\leq i\leq k-1, \tag{3.22}$$

$$\int_0^1\int_0^1 dsdt 1_{\{s\leq t\}} \frac{|X_{s,t}^{(i)}-Y_{s,t}^{(i)}|^{2M_i}}{|t-s|^{2M_i\alpha_i}} \leq B_i, \quad 1\leq i\leq k. \tag{3.23}$$

Then, there exists a constant $C>0$ such that for any $0\leq s<t\leq 1$,

$$|X_{s,t}^{(i)}|+|Y_{s,t}^{(i)}| \leq C\, F_i|t-s|^{\frac{i}{p}}, \quad 1\leq i\leq k-1, \tag{3.24}$$

$$\left|X_{s,t}^{(i)}-Y_{s,t}^{(i)}\right| \leq C\, G_i|t-s|^{\frac{i}{p}}, \quad 1\leq i\leq k. \tag{3.25}$$

where F_i and G_i are defined recursively by

$$F_i = A_i^{\frac{1}{2M_i}} + \sum_{j=1}^{i-1} F_j\, F_{i-j}, \quad 1\leq i\leq k-1, \tag{3.26}$$

$$G_i = B_i^{\frac{1}{2M_i}} + \sum_{j=1}^{i-1} G_j\, F_{i-j}, \quad 1\leq i\leq k. \tag{3.27}$$

Remark 3.6. For rough paths X, Y of roughness $p\in[1,\infty[$, $X_{s,t}^{(1)}-Y_{s,t}^{(1)} = (X-Y)_{s,t}^{(1)}$. The usual version of the Garsia-Rademich-Rumsey lemma yields the following. If

$$\int_0^1\int_0^1 dsdt 1_{\{s\leq t\}} \frac{|X_{s,t}^{(1)}-Y_{s,t}^{(1)}|^{2M_1}}{|t-s|^{2M_1\alpha_1}} \leq B_1,$$

then $|X_{s,t}^{(1)}-Y_{s,t}^{(1)}| \leq C\, B_1^{\frac{1}{2M_1}}|t-s|^{\frac{1}{p}}$. Similarly, if

$$\int_0^1\int_0^1 dsdt 1_{\{s\leq t\}} \frac{|X_{s,t}^{(1)}|^{2M_1}+|Y_{s,t}^{(1)}|^{2M_1}}{|t-s|^{2M_1\alpha_1}} \leq A_1,$$

then $|X_{s,t}^{(1)}|+|Y_{s,t}^{(1)}| \leq C\, A_1^{\frac{1}{2M_1}}|t-s|^{\frac{1}{p}}$.

Proof of Lemma 3.5. Throughout the proof, the constants F_i, $1\leq i\leq k-1$ and G_i, $1\leq i\leq k$ are defined by (3.26), (3.27), respectively. We introduce the following assumption:

$$(H_i) \qquad \int_0^1 \int_0^1 dsdt 1_{\{s\le t\}} \frac{|X^{(i)}_{s,t} - Y^{(i)}_{s,t}|^{2M_i}}{|t-s|^{2M_i\alpha_i}} \le B_i,$$
$$|X^{(j)}_{s,t}| + |Y^{(j)}_{s,t}| \le C\, F_j |t-s|^{\frac{j}{p}}, \quad 1\le j\le i-1,$$
$$|X^{(j)}_{s,t} - Y^{(j)}_{s,t}| \le C\, G_j |t-s|^{\frac{j}{p}}, \quad 1\le j\le i-1,$$

$i\in\{2,\dots,k\}$, and we prove that (H_i) implies

$$|X^{(i)}_{s,t} - Y^{(i)}_{s,t}| \le C\, G_i |t-s|^{\frac{i}{p}}. \tag{3.28}$$

For this, we use an argument similar to the proof of Theorem 2.1.3 in [9].

Indeed, for every $t\in[0,1]$, set

$$I(t) = \int_0^t \frac{|X^{(i)}_{s,t} - Y^{(i)}_{s,t}|^{2M_i}}{|t-s|^{2M_i\alpha_i}}\, ds\,,\ J(t) = \int_t^1 \frac{|X^{(i)}_{t,u} - Y^{(i)}_{t,u}|^{2M_i}}{|u-t|^{2M_i\alpha_i}}\, du\,.$$

Then $\int_0^1 I(t)\, dt = \int_0^1 J(t)\, dt \le B_i$ and there exists $t_0>0$ such that $I(t_0) + J(t_0) \le 2\, A_i$. We construct by induction a decreasing sequence $(t_n,\, n\ge 0)$ such that $\lim_n t_n = 0$ and an increasing sequence $(s_n,\, n\ge 0)$ such that $s_0 = t_0$, $\lim_n s_n = 1$, and such that there exists $C>0$ such that for every $n\ge 1$,

$$\left|X^{(i)}_{t_n,t_0} - Y^{(i)}_{t_n,t_0}\right| \le C \int_0^1 |8\, B_i|^{\frac{1}{2M_i}}\ u^{\frac{i}{p}-1}\, du + C \sum_{j=1}^{i-1} F_j\, G_{i-j}\,, \tag{3.29}$$

$$\left|X^{(i)}_{s_0,s_n} - Y^{(i)}_{s_0,s_n}\right| \le C \int_0^1 |8\, B_i|^{\frac{1}{2M_i}}\ u^{\frac{i}{p}-1}\, du + C \sum_{j=1}^{i-1} F_j\, G_{i-j}\,. \tag{3.30}$$

Then Chen's identity implies as $n\to+\infty$,

$$\begin{aligned}|X^{(i)}_{0,1} - Y^{(i)}_{0,1}| &\le |X^{(i)}_{0,t_0} - Y^{(i)}_{0,t_0}| + |X^{(i)}_{t_0,1} - Y^{(i)}_{t_0,1}| \\ &\quad + \sum_{j=1}^{i-1}\Big(|X^{(j)}_{0,t_0} - Y^{(j)}_{0,t_0}||X^{(i-j)}_{t_0,1}| + |Y^{(j)}_{0,t_0}||X^{(i-j)}_{t_0,1} - Y^{(i-j)}_{t_0,1}|\Big)\,.\end{aligned} \tag{3.31}$$

With the hypothesis (H_i), we obtain (3.28) with $s=0$ and $t=1$.

To construct (t_n), we suppose that t_{n-1} has been chosen. Let d_{n-1} be defined by $d_{n-1}^{\alpha_i} = \frac{1}{2}\, t_{n-1}^{\alpha_i}$. Then there exists $t_n \in]0, d_{n-1}[$ such that

$$I(t_n) \le \frac{4\, B_i}{d_{n-1}} \quad\text{and}\quad \frac{|X^{(i)}_{t_n,t_{n-1}} - Y^{(i)}_{t_n,t_{n-1}}|^{2M_i}}{|t_{n-1}-t_n|^{2M_i\alpha_i}} \le \frac{2I(t_{n-1})}{d_{n-1}}\,.$$

Indeed, the sets where each one of these inequalities may fail has Lebesgue measure less that $\frac{d_{n-1}}{2}$. Furthermore, for every $n\ge 0$, $2\, d_{n+1}^{\alpha_i} = t_{n+1}^{\alpha_i} \le d_n^{\alpha_i} = \frac{1}{2} t_n^{\alpha_i}$ and $|t_n - t_{n+1}|^{\alpha_i} \le t_n^{\alpha_i} = 2\, d_n^{\alpha_i} \le 4\,(d_n^{\alpha_i} - d_{n+1}^{\alpha_i})$. Hence there exists $a\in]0,1[$ such that $t_{n+1}\le a\, t_n$, so that $\lim_n t_n = 0$ and more precisely,

$$t_n \le a^n\, t_0, \tag{3.32}$$

while for any $n \geq 1$,

$$
\begin{aligned}
|X^{(i)}_{t_{n+1},t_n} - Y^{(i)}_{t_{n+1},t_n}| &\leq |2\,I(t_n)|^{\frac{1}{2M_i}}\, d_n^{-\frac{1}{2M_i}}\, |t_n - t_{n+1}|^{\alpha_i} \\
&\leq |8\,B_i|^{\frac{1}{2M_i}}\, |d_n\, d_{n-1}|^{-\frac{1}{2M_i}}\, 4\, |d_n^{\alpha_i} - d_{n+1}^{\alpha_i}| \\
&\leq 4\,\alpha_i \int_{d_{n+1}}^{d_n} |8\,B_i|^{\frac{1}{2M_i}}\, u^{-\frac{1}{M_i}+\alpha_i-1}\, du. \qquad (3.33)
\end{aligned}
$$

Let $b = a^{\frac{1}{p}} < 1$; Chen's identity, (H_i) and (3.33) imply that for any $n \geq 1$,

$$
\begin{aligned}
&\left|X^{(i)}_{t_{n+1},t_0} - Y^{(i)}_{t_{n+1},t_0}\right| \leq \left|X^{(i)}_{t_n,t_0} - Y^{(i)}_{t_n,t_0}\right| + \left|X^{(i)}_{t_{n+1},t_n} - Y^{(i)}_{t_{n+1},t_n}\right| \\
&\quad + \sum_{j=1}^{i-1} \Big(|X^{(j)}_{t_{n+1},t_n} - Y^{(j)}_{t_{n+1},t_n}|\, |X^{(i-j)}_{t_n,t_0}| + |Y^{(j)}_{t_{n+1},t_n}|\, |X^{(i-j)}_{t_n,t_0} - Y^{(i-j)}_{t_n,t_0}| \Big) \\
&\leq \left|X^{(i)}_{t_n,t_0} - Y^{(i)}_{t_n,t_0}\right| + C \int_{d_{n+1}}^{d_n} |8\,B_i|^{\frac{1}{2M_i}}\, u^{\frac{i}{p}-1}\, du \\
&\quad + C \sum_{j=1}^{i-1} (G_j F_{i-j} + F_j G_{i-j})\, |t_n - t_{n+1}|^{\frac{j}{p}} |t_0 - t_n|^{\frac{i-j}{p}}.
\end{aligned}
$$

Since $\sup_{1 \leq j \leq i-1} |t_n - t_{n+1}|^{\frac{j}{p}} \leq t_n^{\frac{1}{p}} \leq C b^n < 1$, an easy induction on n implies that for any $n \geq 1$,

$$
\left|X^{(i)}_{t_n,t_0} - Y^{(i)}_{t_n,t_0}\right| \leq C \int_0^1 |8\,B_i|^{\frac{1}{2M_i}}\, u^{\frac{i}{p}-1}\, du + C \left(\sum_{j=1}^{i-1} G_j F_{i-j} \right) \left(\sum_{l=0}^{n-2} b^l \right),
$$

which implies (3.29). To prove (3.30), we proceed in a similar way, exchanging the endpoints of the interval $[0,1]$. Recall that $s_0 = t_0$; suppose that s_{n-1} has been defined and let δ_{n-1} be such that $|1 - \delta_{n-1}|^{\alpha_i} = \frac{1}{2}|1 - s_{n-1}|^{\alpha_i}$. There exists $s_n \in]\delta_{n-1}, 1[$ such that

$$
J(s_n) \leq \frac{4B_i}{1 - \delta_{n-1}} \quad \text{and} \quad \frac{|X^{(i)}_{s_{n-1},s_n} - Y^{(i)}_{s_{n-1},s_n}|^{2M_i}}{|s_n - s_{n-1}|^{\alpha_i}} \leq \frac{2J(s_{n-1})}{1 - \delta_{n-1}}.
$$

Then for every $n \geq 1$, $2\,|1 - \delta_{n+1}|^{\alpha_i} = |1 - s_{n+1}|^{\alpha_i} \leq |1 - \delta_n|^{\alpha_i} = \frac{1}{2}|1 - t_n|^{\alpha_i}$, so that $s_n \leq \delta_n \leq s_{n+1} \leq \delta_{n+1}$ and for some $\bar{a} \in]0,1[$

$$
1 - s_n \leq \bar{a}^n\, (1 - t_0),
$$

so that $\lim_n s_n = 1$ and computations similar to those proving (3.33) yield

$$
|X^{(i)}_{s_n,s_{n+1}} - Y^{(i)}_{s_n,s_{n+1}}| \leq 4\,\alpha_i \int_{\delta_n}^{\delta_{n+1}} |8\,B_i|^{\frac{1}{2M_i}}\, u^{-\frac{1}{M_i}+\alpha_i-1}\, du.
$$

Thus if $\bar{b} = \bar{a}^{\frac{1}{p}} < 1$, Chen's identity and (H_i) imply

$$\left| X^{(i)}_{t_0,s_n} - Y^{(i)}_{t_0,s_n} \right| \le C \int_{t_0}^{s_n} |8\, B_i|^{\frac{1}{2M_i}}\, u^{\frac{i}{p}-1}\, du + C \left(\sum_{j=1}^{i-1} F_j G_{i-j} \right) \left(\sum_{l=0}^{n-1} \bar{b}^l \right),$$

which completes the proof of (3.30) and hence that of (3.28) for $s = 0$, $t = 1$.

To deduce (3.28), for any $s, t \in [0,1]$ with $s < t$, define $\bar{X}_u = X_{s+(t-s)u}$, $\bar{Y}_u = Y_{s+(t-s)u}$ for $u \in [0,1]$. Then $\bar{X}$ and $\bar{Y}$ are geometric rough paths with the same roughness p. Moreover, for $0 \le u < v \le 1$, $j = 1, \ldots, k$, $\bar{X}^{(j)}_{u,v} = X^{(j)}_{s+(t-s)u,s+(t-s)v}$. In fact, by a change of variables, we see that this identity is obvious for smooth rough paths and therefore it is trivially extended to geometric rough paths.

Furthermore,

$$\begin{aligned}
&\int_0^1 \int_0^1 dudv 1_{\{u<v\}} \frac{|\bar{X}^{(i)}_{u,v} - \bar{Y}^{(i)}_{u,v}|^{2M_i}}{|v-u|^{2M_i \alpha_i}} \\
&\qquad = (t-s)^{-2+2\alpha_i M_i} \int_s^t \int_s^t dudv 1_{\{u<v\}} \frac{|X^{(i)}_{u,v} - Y^{(i)}_{u,v}|^{2M_i}}{|v-u|^{2M_i \alpha_i}} \\
&\qquad \le (t-s)^{-2+2\alpha_i M_i}\, B_i = (t-s)^{2M_i \frac{i}{p}} B_i\,.
\end{aligned}$$

Hence, if the pair (X, Y) satisfies (H_i), then $(\bar{X}, \bar{Y})$ satisfies a similar property with constants $\bar{A}_j = (t-s)^{2M_j \frac{j}{p}} A_j$, $\bar{F}_j = |t-s|^{\frac{j}{p}} F_j$,$1 \le j \le i-1$, $\bar{B}_j = (t-s)^{2M_j \frac{j}{p}} B_j$, $\bar{G}_j = |t-s|^{\frac{j}{p}}$, $1 \le j \le i$. This finishes the proof of (3.28).

Taking in the preceding arguments first $X \equiv 0$ and then $Y \equiv 0$, we see recursively that (3.22) implies (H_i) for any $i = 1, \ldots, k-1$, with $B_i = A_i$. Hence we obtain (3.24). Moreover, we also see that (H_i) holds true for any $i = 1, \ldots, k$, whenever (3.22), (3.23) are satisfied. This concludes the proof. □

Acknowledgments

The first named author whishes to thank the Centre de Recerca Matemàtica in Bellaterra and the Universitat de Barcelona for their support and hospitality in the fall of 2004, when discussions on the content of this paper started.

References

[1] E. Alòs, O. Mazet, and D. Nualart, *Stochastic calculus with respect to Gaussian processes*, The Annals of Probab., **29 (2)** (2001), 766–801.

[2] L. Coutin and Z. Qian, *Stochastic analysis, rough path analysis and fractional Brownian motions*, Probab. Theory Relat. Fields, **122** (2002), 108–140.

[3] L. Coutin, P. Friz, and N. Victoir, *Good rough path sequences and applications to anticipating and fractional stochastic calculus*, The Annals of Probab., to appear.

[4] L. Decreusefond and S. Üstünel, *Stochastic analysis of the fractional Brownian motion*, Potential Analysis, **10** (1999), 177–214.

[5] P. Friz and N. Victoir, *A note on the notion of geometric rough paths*, Prob. Theory Rel. Fields, **136 (3)** (2006), 395–416.

[6] T. Lyons and Z. Qian, *System Control and Rough Paths*, Oxford Mathematical Monographs, Oxford Science Publications, Clarendon Press, Oxford, 2002.

[7] A. Millet and M. Sanz-Solé, *Large deviations for rough paths of the fractional Brownian motion*, Ann. Inst. H. Poincaré, Probab. Stat., **42 (2)** (2006), 245–271.

[8] D. Nualart, *The Malliavin Calculus and Related Topics*, Probability and its Applications, Springer Verlag, 1995.

[9] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Grundlehren des mathematischen Wissenschaften, 233, Springer Verlag, 1979.

Annie Millet
Laboratoire de Probabilités et Modèles Aléatoires
Universités Paris 6 – Paris 7
Boîte courrier 188
4, Place Jussieu
F-75252 Paris Cedex 5
France
and
SAMOS-MATISSE, Université Paris 1
e-mail: `amil@ccr.jussieu.fr`

Marta Sanz-Solé
Facultat de Matemàtiques
Universitat de Barcelona
Gran Via 585
E-08007 Barcelona
Spain
e-mail: `marta.sanz@ub.edu`

Progress in Probability, Vol. 59, 305–330

A One-Dimensional Analysis of Singularities and Turbulence for the Stochastic Burgers Equation in d Dimensions

Andrew D. Neate and Aubrey Truman

Abstract. The inviscid limit of the stochastic Burgers equation, with body forces white noise in time, is discussed in terms of the level surfaces of the minimising Hamilton-Jacobi function, the classical mechanical caustic and the Maxwell set, and their algebraic pre-images under the classical mechanical flow map. The problem is analysed in terms of a reduced (one-dimensional) action function. We give an explicit expression for an algebraic surface containing the Maxwell set and caustic in the polynomial case. Those parts of the caustic and Maxwell set which are singular are characterised. We demonstrate how the geometry of the caustic, level surfaces and Maxwell set can change infinitely rapidly causing turbulent behaviour which is stochastic in nature, and we determine its intermittence in terms of the recurrent behaviour of two processes.

Mathematics Subject Classification (2000). Primary 60H15, 60H30, 35Q53; Secondary 76M35, 35R60, 76F20.

Keywords. Stochastic Burgers equation, turbulence, caustic, Maxwell set.

1. Introduction

Burgers equation has been used in studying turbulence and in modelling the large scale structure of the universe [1, 9, 28], as well as to obtain detailed asymptotics for stochastic Schrödinger and heat equations [10, 11, 29, 30, 31, 32]. It has also played a part in Arnol'd's work on caustics and Maslov's works in semiclassical quantum mechanics [3, 4, 20, 21].

We consider the stochastic viscous Burgers equation for the velocity field $v^\mu(x,t) \in \mathbb{R}^d$, where $x \in \mathbb{R}^d$ and $t > 0$,

$$\frac{\partial v^\mu}{\partial t} + (v^\mu \cdot \nabla)\, v^\mu = \frac{\mu^2}{2}\Delta v^\mu - \nabla V(x) - \epsilon \nabla k_t(x) \dot{W}_t, \qquad v^\mu(x,0) = \nabla S_0(x) + \mathrm{O}(\mu^2).$$

Here $\dot{W}_t$ denotes white noise and μ^2 is the coefficient of viscosity which we assume to be small. We are interested in the advent of discontinuities in the inviscid limit of the Burgers fluid velocity $v^0(x,t)$ where $v^\mu(x,t) \to v^0(x,t)$ as $\mu \to 0$.

Using the Hopf-Cole transformation $v^\mu(x,t) = -\mu^2 \nabla \ln u^\mu(x,t)$, the Burgers equation becomes the Stratonovich heat equation,

$$\frac{\partial u^\mu}{\partial t} = \frac{\mu^2}{2}\Delta u^\mu + \mu^{-2} V(x) u^\mu + \frac{\epsilon}{\mu^2} k_t(x) u^\mu \circ \dot{W}_t, \quad u^\mu(x,0) = \exp\left(-\frac{S_0(x)}{\mu^2}\right) T_0(x),$$

where the convergence factor T_0 is related to the initial Burgers fluid density [14].

Now let

$$A[X] := \frac{1}{2}\int_0^t \dot{X}^2(s)\,\mathrm{d}s - \int_0^t V(X(s))\,\mathrm{d}s - \epsilon \int_0^t k_s(X(s))\,\mathrm{d}W_s, \tag{1.1}$$

and select a path X which minimises $A[X]$. This requires that

$$\mathrm{d}\dot{X}(s) + \nabla V(X(s))\,\mathrm{d}s + \epsilon \nabla k_s(X(s))\,\mathrm{d}W_s = 0.$$

We then define the stochastic action $A(X(0),x,t) := \inf_X \{A[X] : X(t) = x\}$. Setting

$$\mathcal{A}(X(0),x,t) := S_0(X(0)) + A(X(0),x,t)$$

and then minimising $\mathcal{A}$ over $X(0)$ gives $\dot{X}(0) = \nabla S_0(X(0))$. Moreover, it follows that

$$\mathcal{S}_t(x) := \inf_{X(0)} \{\mathcal{A}(X(0),x,t)\}$$

is the minimal solution of the Hamilton-Jacobi equation,

$$\mathrm{d}\mathcal{S}_t + \left(\frac{1}{2}|\nabla \mathcal{S}_t|^2 + V(x)\right)\mathrm{d}t + \epsilon k_t(x)\,\mathrm{d}W_t = 0, \qquad \mathcal{S}_{t=0}(x) = S_0(x).$$

Following the work of Donsker, Freidlin et al [12], $-\mu^2 \ln u^\mu(x,t) \to \mathcal{S}_t(x)$ as $\mu \to 0$. This gives the inviscid limit of the minimal entropy solution of Burgers equation as $v^0(x,t) = \nabla \mathcal{S}_t(x)$ [5].

Definition 1.1. The stochastic wavefront at time t is defined to be the set

$$\mathcal{W}_t = \{x : \quad \mathcal{S}_t(x) = 0\}.$$

For small μ and fixed t, $u^\mu(x,t)$ switches continuously from being exponentially large to small as x crosses the wavefront $\mathcal{W}_t$. However, u^μ and v^μ can also switch discontinuously.

Define the classical flow map $\Phi_s : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\mathrm{d}\dot{\Phi}_s + \nabla V(\Phi_s)\,\mathrm{d}s + \epsilon \nabla k_s(\Phi_s)\,\mathrm{d}W_s = 0, \qquad \Phi_0 = \mathrm{id}, \qquad \dot{\Phi}_0 = \nabla S_0.$$

Since $X(t) = x$ it follows that $X(s) = \Phi_s\left(\Phi_t^{-1}(x)\right)$, where the pre-image $x_0(x,t) = \Phi_t^{-1}(x)$ is not necessarily unique.

Given some regularity and boundedness, the global inverse function theorem gives a caustic time $T(\omega)$ such that for $0 < t < T(\omega)$, Φ_t is a random diffeomorphism; before the caustic time $v^0(x,t) = \dot{\Phi}_t\left(\Phi_t^{-1}(x)\right)$ is the inviscid limit of a classical solution of the Burgers equation with probability 1.

The method of characteristics suggests that discontinuities in $v^0(x,t)$ are associated with the non-uniqueness of the real pre-image $x_0(x,t)$. When this occurs, the classical flow map Φ_t focusses an infinitesimal volume of points $\mathrm{d}x_0$ into a zero volume $\mathrm{d}X(t)$.

Definition 1.2. The caustic at time t is defined to be the set

$$C_t = \left\{ x : \quad \det\left(\frac{\partial X(t)}{\partial x_0}\right) = 0 \right\}.$$

Assume that x has n real pre-images,

$$\Phi_t^{-1}\{x\} = \{x_0(1)(x,t), x_0(2)(x,t), \dots, x_0(n)(x,t)\},$$

where each $x_0(i)(x,t) \in \mathbb{R}^d$. Then the Feynman-Kac formula and Laplace's method in infinite dimensions give for a non-degenerate critical point

$$u^\mu(x,t) = \sum_{i=1}^{n} \theta_i \exp\left(-\frac{S_0^i(x,t)}{\mu^2}\right), \tag{1.2}$$

where $S_0^i(x,t) := S_0(x_0(i)(x,t)) + A(x_0(i)(x,t),x,t)$, and θ_i is an asymptotic series in μ^2. An asymptotic series in μ^2 can also be found for $v^\mu(x,t)$ [33]. Note that $\mathcal{S}_t(x) = \min\{S_0^i(x,t) : i = 1,2,\dots,n\}$.

Definition 1.3. The Hamilton-Jacobi level surface is the set

$$H_t^c = \left\{ x : \quad S_0^i(x,t) = c \text{ for some } i \right\}.$$

The zero level surface H_t^0 includes the wavefront $\mathcal{W}_t$.

As $\mu \to 0$, the dominant term in the expansion (1.2) comes from the minimising $x_0(i)(x,t)$ which we denote $\tilde{x}_0(x,t)$. Assuming $\tilde{x}_0(x,t)$ is unique, we obtain the inviscid limit of the Burgers fluid velocity as $v^0(x,t) = \dot{\Phi}_t(\tilde{x}_0(x,t))$.

If the minimising pre-image $\tilde{x}_0(x,t)$ suddenly changes value between two pre-images $x_0(i)(x,t)$ and $x_0(j)(x,t)$, a jump discontinuity will also occur in the inviscid limit of the Burgers fluid velocity. There are two distinct ways in which the minimiser can change; either two pre-images coalesce and disappear (become complex), or the minimiser switches between two pre-images at the same action value. The first of these occurs as x crosses the caustic and when the minimiser disappears the caustic is said to be cool. The second occurs as x crosses the Maxwell set and again, when the minimiser is involved, the Maxwell set is said to be cool.

Definition 1.4. The Maxwell set is given by

$$\begin{aligned} M_t \quad = \quad & \{x : \exists\, x_0, \check{x}_0 \in \mathbb{R}^d \text{ s.t.} \\ & \quad x = \Phi_t(x_0) = \Phi_t(\check{x}_0),\ x_0 \neq \check{x}_0 \text{ and } \mathcal{A}(x_0,x,t) = \mathcal{A}(\check{x}_0,x,t)\}. \end{aligned}$$

Example (The generic Cusp). Let $V(x,y) = 0$, $k_t(x,y) = 0$ and $S_0(x_0,y_0) = x_0^2 y_0/2$. This initial condition leads to the *generic Cusp*, a semicubical parabolic caustic shown in Figure 1. The caustic C_t (long dash) is given by

$$x_t(x_0) = t^2 x_0^3, \quad y_t(x_0) = \frac{3}{2} t x_0^2 - \frac{1}{t}.$$

The zero level surface H_t^0 (solid line) is

$$x_{(t,0)}(x_0) = \frac{x_0}{2}\left(1 \pm \sqrt{1 - t^2 x_0^2}\right), \quad y_{(t,0)}(x_0) = \frac{1}{2t}\left(t^2 x_0^2 - 1 \pm \sqrt{1 - t^2 x_0^2}\right),$$

and the Maxwell set M_t (short dash) is $x = 0$ for $y > -1/t$.

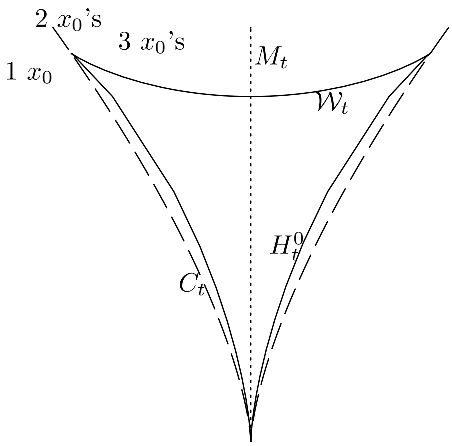


FIGURE 1. Cusp and Tricorn.

Notation. Throughout this paper x, x_0, x_t etc will denote vectors, where normally $x = \Phi_t(x_0)$. Cartesian coordinates of these will be indicated using a sub/superscript where relevant; thus $x = (x_1, x_2, \ldots, x_d)$, $x_0 = (x_0^1, x_0^2, \ldots, x_0^d)$ etc. The only exception will be in discussions of explicit examples in two and three dimensions when we will use (x, y) and (x_0, y_0) etc. to denote the vectors.

2. Some background

We begin by summarising some of the geometrical results established by Davies, Truman and Zhao (DTZ) [6, 7, 8] and presenting some minor generalisations of their results [22, 25]. Following equation (1.1), let the stochastic action be defined,

$$A(x_0, p_0, t) = \frac{1}{2}\int_0^t \dot{X}(s)^2 \,\mathrm{d}s - \int_0^t \left[V(X(s))\,\mathrm{d}s + \epsilon k_s(X(s))\,\mathrm{d}W_s\right],$$

where $X(s) = X(s, x_0, p_0) \in \mathbb{R}^d$ and

$$\mathrm{d}\dot{X}(s) = -\nabla V(X(s))\,\mathrm{d}s - \epsilon\nabla k_s(X(s))\,\mathrm{d}W_s, \quad X(0) = x_0, \quad \dot{X}(0) = p_0,$$

for $s \in [0, t]$ with $x_0, p_0 \in \mathbb{R}^d$. We assume $X(s)$ is $\mathcal{F}_s$ measurable and unique.

Lemma 2.1. *Assume $S_0, V \in C^2$ and $k_t \in C^{2,0}$, $\nabla V, \nabla k_t$ Lipschitz with Hessians $\nabla^2 V, \nabla^2 k_t$ and all second derivatives with respect to space variables of V and k_t bounded. Then for p_0, possibly x_0 dependent,*

$$\frac{\partial A}{\partial x_0^\alpha}(x_0, p_0, t) = \dot{X}(t) \cdot \frac{\partial X(t)}{\partial x_0^\alpha} - \dot{X}_\alpha(0), \qquad \alpha = 1, 2, \ldots, d.$$

Methods of Kolokoltsov et al [18, 19] guarantee that for small t the map $p_0 \mapsto X(t, x_0, p_0)$ is onto for all x_0. Therefore, we can define

$$A(x_0, x, t) = A(x_0, p_0, t)|_{p_0 = p_0(x_0, x, t)},$$

where $p_0 = p_0(x_0, x, t)$ is the random minimiser (assumed unique) of $A(x_0, p_0, t)$ when $X(t, x_0, p_0) = x$. The stochastic action corresponding to the initial momentum $\nabla S_0(x_0)$ is then $\mathcal{A}(x_0, x, t) := A(x_0, x, t) + S_0(x_0)$.

Theorem 2.2. *If Φ_t is the stochastic flow map, then $\Phi_t(x_0) = x$ is equivalent to*

$$\frac{\partial}{\partial x_0^\alpha}\left[\mathcal{A}(x_0, x, t)\right] = 0, \qquad \alpha = 1, 2, \ldots, d.$$

The Hamilton-Jacobi level surface H_t^c is obtained by eliminating x_0 between

$$\mathcal{A}(x_0, x, t) = c \quad \text{and} \quad \frac{\partial \mathcal{A}}{\partial x_0^\alpha}(x_0, x, t) = 0, \quad \alpha = 1, 2, \ldots, d.$$

Alternatively, if we eliminate x to give an expression in x_0, we have the pre-level surface $\Phi_t^{-1} H_t^c$. Similarly the caustic C_t (and pre-caustic $\Phi_t^{-1} C_t$) are obtained by eliminating x_0 (or x) between

$$\det\left(\frac{\partial^2 \mathcal{A}}{\partial x_0^\alpha \partial x_0^\beta}(x_0, x, t)\right)_{\alpha,\beta=1,2,\ldots,d} = 0 \quad \text{and} \quad \frac{\partial \mathcal{A}}{\partial x_0^\alpha}(x_0, x, t) = 0 \quad \alpha = 1, 2, \ldots, d.$$

These pre-images are calculated algebraically which are not necessarily the topological inverse images of the surfaces C_t and H_t^c under Φ_t.

Assume that $A(x_0, x, t)$ is C^4 in space variables with $\det\left(\frac{\partial^2 \mathcal{A}}{\partial x_0^\alpha \partial x^\beta}\right) \neq 0$.

Definition 2.3. A curve $x = x(\gamma)$, $\gamma \in N(\gamma_0, \delta)$, is said to have a generalised cusp at $\gamma = \gamma_0$, γ being an intrinsic variable such as arc length, if $\frac{\mathrm{d}x}{\mathrm{d}\gamma}(\gamma_0) = 0$.

Lemma 2.4. *Let Φ_t denote the flow map and let $\Phi_t^{-1}\Gamma_t$ and Γ_t be some surfaces where if $x_0 \in \Phi_t^{-1}\Gamma_t$, then $x = \Phi_t(x_0) \in \Gamma_t$. Then Φ_t is a differentiable map from $\Phi_t^{-1}\Gamma_t$ to Γ_t with Frechet derivative*

$$D\Phi_t(x_0) = \left(-\frac{\partial^2 \mathcal{A}}{\partial x \partial x_0}(x_0, x, t)\right)^{-1} \left(\frac{\partial^2 \mathcal{A}}{\partial x_0^2}(x_0, x, t)\right).$$

Lemma 2.5. *Let $x_0(s)$ be any two-dimensional intrinsically parameterised curve, and define $x(s) = \Phi_t(x_0(s))$. Let e_0 denote the zero eigenvector of $\left(\frac{\partial^2 \mathcal{A}}{(\partial x_0)^2}\right)$ and assume that* $\ker\left(\frac{\partial^2 \mathcal{A}}{(\partial x_0)^2}\right) = \langle e_0 \rangle$. *Then, there is a generalised cusp on $x(s)$ when $s = \sigma$ if and only if either:*

1. *there is a generalised cusp on* $x_0(s)$ *when* $s = \sigma$*; or,*
2. $x_0(\sigma)$ *is on the pre-caustic and the tangent* $\frac{\mathrm{d}x_0}{\mathrm{d}s}(s)$ *at* $s = \sigma$ *is parallel to* e_0*.*

Proposition 2.6. *The normal to* $\Phi_t^{-1} H_t^c$ *is*

$$n(x_0) = -\left(\frac{\partial^2 \mathcal{A}}{\partial x_0 \partial x_0}\right)\left(\frac{\partial^2 \mathcal{A}}{\partial x_0 \partial x}\right)^{-1} \dot{X}\left(t, x_0, \nabla S_0(x_0)\right).$$

Corollary 2.7. *In two dimensions, let* $\Phi_t^{-1} H_t^c$ *meet* $\Phi_t^{-1} C_t$ *at* x_0 *where* $n(x_0) \neq 0$ *and* $\ker\left(\frac{\partial^2 \mathcal{A}}{(\partial x_0)^2}\right) = \langle e_0 \rangle$*. Then the tangent to* $\Phi_t^{-1} H_t^c$ *at* x_0 *is parallel to* e_0*.*

Proposition 2.8. *In two dimensions, assume that* $n(x_0) \neq 0$ *where* $x_0 \in \Phi_t^{-1} H_t^c$*, so that* $\Phi_t^{-1} H_t^c$ *does not have a generalised cusp at* x_0*. Then* H_t^c *can only have a generalised cusp at* $\Phi_t(x_0)$ *if* $\Phi_t(x_0) \in C_t$*. Moreover, if* $x = \Phi_t(x_0) \in \Phi_t\left\{\Phi_t^{-1} C_t \cap H_t^{-1}\right\}$*, then* H_t^c *will have a generalised cusp.*

Example (The generic Cusp). Figure 2 shows that a point lying on three level surfaces has three distinct real pre-images each on a separate pre-level surface. A cusp only occurs on the corresponding level surface when the pre-level surface intersects the pre-caustic. Thus, a level surface only has a cusp on the caustic, but it does not have to be cusped when it meets the caustic.

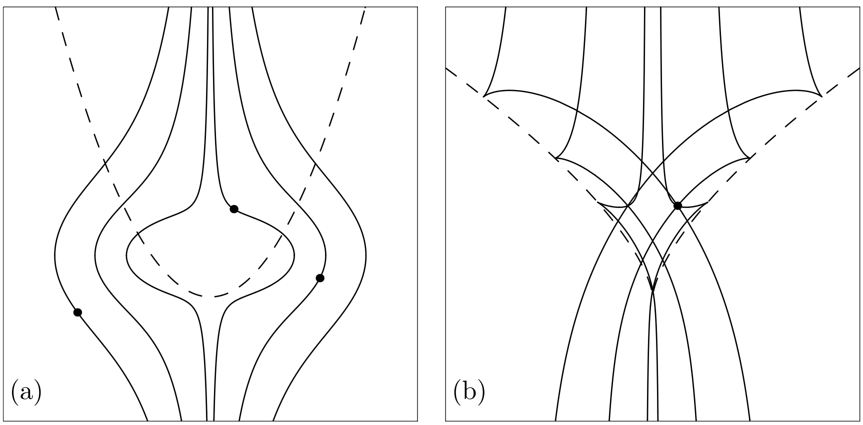


FIGURE 2. (a) The pre-level surface (solid line) and pre-caustic (dashed), (b) the level surface (solid line) and caustic (dashed), both for the generic Cusp with $c > 0$.

Theorem 2.9. *Let*

$$x \in \mathit{Cusp}\left(H_t^c\right) = \left\{x \in \Phi_t\left(\Phi_t^{-1} C_t \cap \Phi_t^{-1} H_t^c\right), x = \Phi_t(x_0), n(x_0) \neq 0\right\}.$$

Then in three dimensions in the stochastic case, with probability 1*,* T_x *the tangent space to the level surface at* x *is at most one-dimensional.*

3. A one-dimensional analysis

In this section we outline a one-dimensional analysis first described by Reynolds, Truman and Williams (RTW) [34].

Definition 3.1. The d-dimensional flow map Φ_t is globally reducible if for any $x = (x_1, x_2, \ldots, x_d)$ and $x_0 = (x_0^1, x_0^2, \ldots, x_0^d)$ where $x = \Phi_t(x_0)$, it is possible to write each coordinate x_0^α as a function of the lower coordinates. That is,

$$x = \Phi_t(x_0) \quad \Rightarrow \quad x_0^\alpha = x_0^\alpha(x, x_0^1, x_0^2, \ldots x_0^{\alpha-1}, t) \text{ for } \alpha = d, d-1, \ldots, 2. \tag{3.1}$$

Therefore, using Theorem 2.2, the flow map is globally reducible if we can find a chain of C^2 functions $x_0^d, x_0^{d-1}, \ldots, x_0^2$ such that

$$\begin{aligned}
x_0^d = x_0^d(x, x_0^1, x_0^2, \ldots x_0^{d-1}, t) \quad &\Leftrightarrow \quad \frac{\partial \mathcal{A}}{\partial x_0^d}(x_0, x, t) = 0,\\
x_0^{d-1} = x_0^{d-1}(x, x_0^1, x_0^2, \ldots x_0^{d-2}, t) \quad &\Leftrightarrow \quad \frac{\partial \mathcal{A}}{\partial x_0^{d-1}}(x_0^1, x_0^2, \ldots, x_0^d(\ldots), x, t) = 0,\\
&\vdots\\
x_0^2 = x_0^2(x, x_0^1, t) \quad &\Leftrightarrow\\
&\frac{\partial \mathcal{A}}{\partial x_0^2}(x_0^1, x_0^2, x_0^3(x, x_0^1, x_0^2, t), \ldots, x_0^d(\ldots), x, t) = 0,
\end{aligned}$$

where $x_0^d(\ldots)$ is the expression only involving x_0^1 and x_0^2 gained by substituting each of the functions $x_0^3, \ldots, x_0^{d-1}$ repeatedly into $x_0^d(x, x_0^1, x_0^2, \ldots, x_0^{d-1}, t)$. This requires that no roots are repeated to ensure that none of the second derivatives of $\mathcal{A}$ vanish. We assume also that there is a favoured ordering of coordinates and a corresponding decomposition of Φ_t which allows the non-uniqueness to be reduced to the level of the x_0^1 coordinate. This assumption appears to be quite restrictive. However, local reducibility at x follows from the implicit function theorem and some mild assumptions on the derivatives of $\mathcal{A}$.

Definition 3.2. If Φ_t is globally reducible, then the reduced action function is the univariate function gained from evaluating the action with equations (3.1),

$$f_{(x,t)}(x_0^1) := f(x_0^1, x, t) = \mathcal{A}(x_0^1, x_0^2(x, x_0^1, t), x_0^3(\ldots), \ldots, x, t).$$

Lemma 3.3. *If Φ_t is globally reducible, modulo the above assumptions,*

$$\left| \det\left(\frac{\partial^2 \mathcal{A}}{(\partial x_0)^2}(x_0, x, t)\right)\Big|_{x_0 = (x_0^1, x_0^2(x, x_0^1, t), \ldots, x_0^d(\ldots))} \right|$$

$$= \prod_{\alpha=1}^{d} \left| \left[\left(\frac{\partial}{\partial x_0^\alpha}\right)^2 \mathcal{A}(x_0^1, \ldots, x_0^\alpha, x_0^{\alpha+1}(\ldots), \ldots, x_0^d(\ldots), x, t) \right]_{\substack{x_0^2 = x_0^2(x, x_0^1, t)\\ \vdots \\ x_0^\alpha = x_0^\alpha(\ldots)}} \right|,$$

where the first term is $f''_{(x,t)}(x_0^1)$ *and the last* $d-1$ *terms are non-zero.*

Theorem 3.4. *Let the classical mechanical flow map* Φ_t *be globally reducible. Then:*

1. $f'_{(x,t)}(x_0^1) = 0$ *and the equations* (3.1) $\Leftrightarrow x = \Phi_t(x_0)$,
2. $f'_{(x,t)}(x_0^1) = f''_{(x,t)}(x_0^1) = 0$ *and the equations* (3.1) $\Leftrightarrow x = \Phi_t(x_0)$ *is such that the number of real solutions* x_0 *changes.*

4. Analysis of the caustic

We begin by parameterising the caustic $0 = \det(D\Phi_t(x_0))$ from Definition 1.2; this equation only involves x_0 and t, and is therefore the pre-caustic. We use this to parameterise the pre-caustic as

$$x_0^1 = \lambda_1, \quad x_0^2 = \lambda_2, \quad \ldots, \quad x_0^{d-1} = \lambda_{d-1} \quad \text{and} \quad x_0^d = x_0^d(\lambda_1, \lambda_2, \ldots, \lambda_{d-1}).$$

The parameters are restricted to be real so that only real pre-images are considered.

Definition 4.1. For any $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{d-1}) \in \mathbb{R}^{d-1}$ the pre-parameterisation of the caustic is given by $x_t(\lambda) := \Phi_t\left(\lambda, x_0^d(\lambda)\right)$.

The pre-parameterisation will be intrinsic if $\ker(D\Phi_t)$ is one-dimensional.

Corollary 4.2. *Let* $x_t(\lambda)$ *denote the pre-parameterisation of the caustic where* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{d-1}) \in \mathbb{R}^{d-1}$. *Then* $f'_{(x_t(\lambda),t)}(\lambda_1) = f''_{(x_t(\lambda),t)}(\lambda_1) = 0$.

Proposition 4.3. *Let* $x_t(\lambda)$ *denote the pre-parameterisation of the caustic where* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{d-1}) \in \mathbb{R}^{d-1}$. *Assume* $f_{(x_t(\lambda),t)}(x_0^1) \in C^{p+1}$, *then, in* d *dimensions, if the tangent to the caustic is at most* $(d-p+1)$*-dimensional at* $x_t(\tilde{\lambda})$,

$$f'_{(x_t(\tilde{\lambda}),t)}(\tilde{\lambda}_1) = f''_{(x_t(\tilde{\lambda}),t)}(\tilde{\lambda}_1) = \ldots = f^{(p)}_{(x_t(\tilde{\lambda}),t)}(\tilde{\lambda}_1) = 0.$$

Proof. Follows by repeatedly differentiating $f'''_{(x_t(\lambda),t)}(\lambda_1) = 0$, which holds if the tangent space at $x_t(\lambda)$ is $(d-2)$-dimensional [22]. □

From Corollary 4.2, there is a critical point of inflexion on $f_{(x,t)}(x_0^1)$ at $x_0^1 = \lambda_1$ when $x = x_t(\lambda)$. Consider an example where for x on one side of the caustic there are four real critical points on $f_{(x,t)}(x_0^1) = 0$. Let them be enumerated $x_0^1(i)(x,t)$ for $i = 1$ to 4 and denote the minimising critical point $\tilde{x}_0^1(x,t)$. Figure 3 illustrates how the minimiser jumps from (a) to (b) as x crosses the caustic. This will cause u^μ and v^μ to jump for small μ and the caustic at such a point is described as being cool.

Definition 4.4. Let $x_t(\lambda)$ be the pre-parameterisation of the caustic. Then $x_t(\lambda)$ is on the cool part of the caustic if $f_{(x_t(\lambda),t)}(\lambda_1) \le f_{(x_t(\lambda),t)}(x_0^1(i)(x_t(\lambda),t))$ for all $i = 1, 2, \ldots, n$, where $x_0^1(i)(x,t)$ denotes an enumeration of all the real roots for x_0^1 to $f'_{(x,t)}(x_0^1) = 0$. If the caustic is not cool, it is hot.

Definition 4.5. The pre-normalised reduced action function evaluated on the caustic is given by $\mathcal{F}_\lambda(x_0^1) := f_{(x_t(\lambda),t)}(x_0^1) - f_{(x_t(\lambda),t)}(\lambda_1)$.

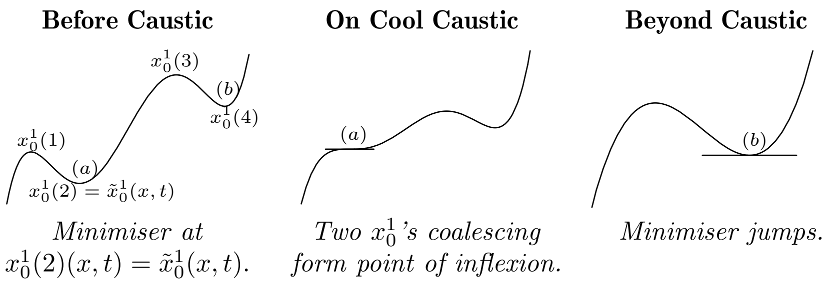


FIGURE 3. The graph of $f_{(x,t)}(x_0^1)$ as x crosses the caustic.

Assume that $\mathcal{F}_\lambda(x_0^1)$ is a real analytic function in a neighbourhood of $\lambda_1 \in \mathbb{R}$. Then,

$$\mathcal{F}_\lambda(x_0^1) = (x_0^1 - \lambda_1)^3 \tilde{F}(x_0^1),$$

where $\tilde{F}$ is real analytic. When the inflexion at $x_0^1 = \lambda_1$ is the minimising critical point of $\mathcal{F}_\lambda$, the caustic will be cool. Therefore, on a hot/cool boundary this inflexion is about to become or cease being the minimiser.

Proposition 4.6. *A* necessary condition *for* $x_t(\lambda) \in C_t$ *to be on a hot/cool boundary is that either* $\tilde{F}(x_0^1)$ *or* $\tilde{G}(x_0^1)$ *has a repeated root at* $x_0^1 = r$ *where*

$$\tilde{G}(x_0^1) = 3\tilde{F}(x_0^1) + (x_0^1 - \lambda_1)\tilde{F}'(x_0^1).$$

Proof. The minimiser could change when either $\tilde{F}$ has a repeated root which is the minimiser, or there is a second inflexion at a lower minimising value [23]. □

The condition is not sufficient as it includes cases where the minimiser is not about to change (see Figure 4).

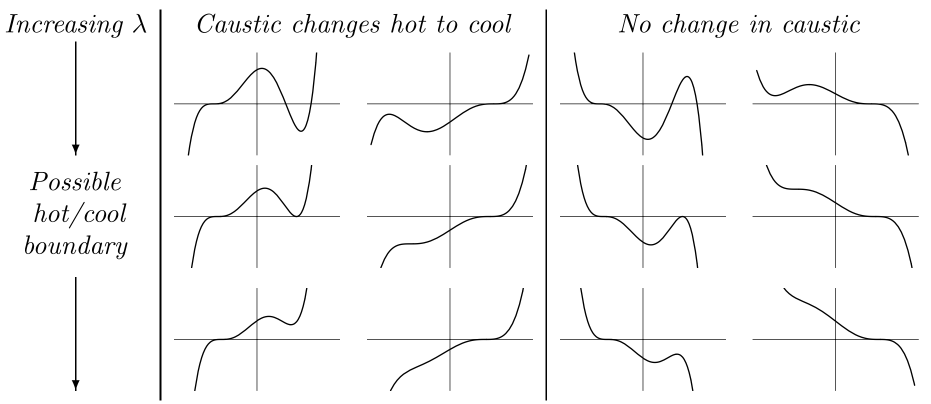


FIGURE 4. Graphs of $\mathcal{F}_\lambda(x_0^1)$ as λ varies.

$$\kappa = \left(-\frac{t^5}{500} - \epsilon \int_0^t W_s \,\mathrm{d}s, \frac{t^3}{50} - \frac{1}{2t}\right)$$

$$\psi = \left(-\frac{t^5(3+8\sqrt{6})}{18000} - \epsilon \int_0^t W_s \,\mathrm{d}s, \frac{t^3(9-\sqrt{6})}{450} - \frac{1}{2t}\right)$$

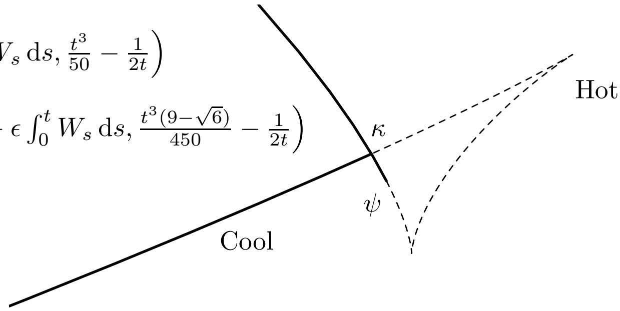
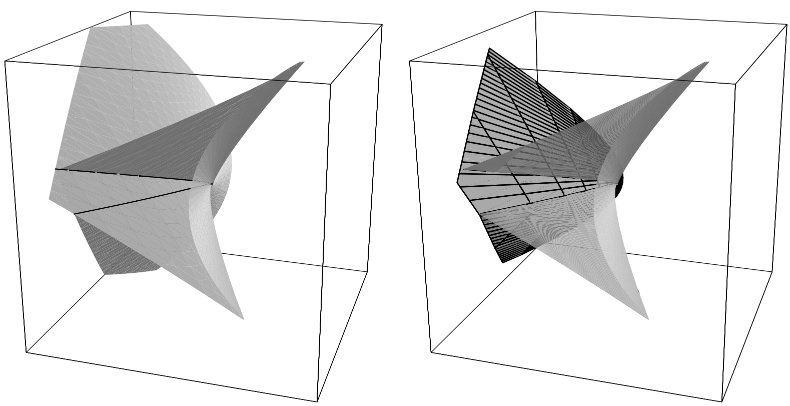


FIGURE 5. Hot and cool parts of the polynomial swallowtail caustic at time $t = 1$.

Example (The polynomial swallowtail). Let $V(x,y) \equiv 0$, $k_t(x,y) \equiv x$, and $S_0(x_0,y_0) = x_0^5 + x_0^2 y_0$. This gives global reducibility, and $k_t(x,y) \equiv x$ means that the effect of the noise is to translate $\epsilon = 0$ picture through $\left(-\epsilon \int_0^t W_s \,\mathrm{d}s, 0\right)$. A simple calculation gives

$$\tilde{F}(x_0) = 12\lambda^2 - 3\lambda t + 6\lambda x_0 - t x_0 + 2x_0^2,$$
$$\tilde{G}(x_0) = 15\lambda^2 - 4\lambda t + 10\lambda x_0 - 2t x_0 + 5x_0^2.$$

Example (The three-dimensional polynomial swallowtail). Let $V(x,y) \equiv 0$, $k_t(x,y) \equiv 0$, and $S_0(x_0,y_0,z_0) = x_0^7 + x_0^3 y_0 + x_0^2 z_0$. The functions $\tilde{F}$ and $\tilde{G}$ can be easily found, and an exact expression for the boundary extracted [22]; this is shown in Figure 6.



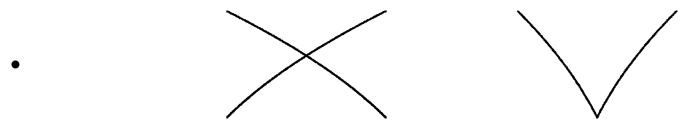
Boundary on the caustic. *Hot and cool parts.*

FIGURE 6. The hot (plain) and cool (mesh) parts of the 3D polynomial swallowtail caustic at time $t = 1$.

5. Swallowtail perestroikas

The geometry of a caustic or wavefront can suddenly change with singularities appearing and disappearing [2]. We consider the formation or collapse of a swallowtail using some earlier works of Cayley and Klein. This section provides a summary of results from [23] where all proofs can be found.

We begin by recalling the classification of double points of a two-dimensional algebraic curve as acnodes, crunodes and cusps (Figure 7).



Acnode. Crunode. Cusp.

FIGURE 7. The classification of double points.

In Cayley's work on plane algebraic curves, he describes the possible triple points of a curve [27] by considering the collapse of systems of double points which would lead to the existence of three tangents at a point. The four possibilities are shown in Figure 8. The systems will collapse to form a triple point with respectively, three real distinct tangents, three real tangents with two coincident, three real tangents all of which are coincident, or one real tangent and two complex tangents. It is the interchange between the last two cases which will lead to the formation of a swallowtail on a curve [15]. This interchange was investigated by Felix Klein [17].

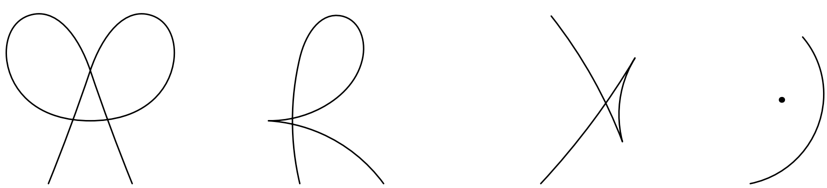


FIGURE 8. Cayley's triple points.

In Section 3, we restricted the pre-parameter to be real to only consider points with real pre-images. This does not allow there to be any isolated double points. We now allow the parameter to vary throughout the complex plane and consider when this maps to real points. We begin by working with a general curve of the form $x(\lambda) = (x_1(\lambda), x_2(\lambda))$ where each $x_\alpha(\lambda)$ is real analytic in $\lambda \in \mathbb{C}$. If $\mathrm{Im}\{x(a + \mathrm{i}\eta)\} = 0$, it follows that $x(a + \mathrm{i}\eta) = x(a - \mathrm{i}\eta)$, so this is a "complex double point" of the curve $x(\lambda)$.

Lemma 5.1. *If $x(\lambda) = (x_1(\lambda), x_2(\lambda))$ is a real analytic parameterisation of a curve and λ is an intrinsic parameter, then there is a generalised cusp at $\lambda = \lambda_0$ if and only if the curves*

$$0 = \frac{1}{\eta} Im\{x_\alpha(a + i\eta)\} \qquad \alpha = 1, 2,$$

intersect at $(\lambda_0, 0)$ in the (a, η) plane.

Now consider a family of parameterised curves $x_t(\lambda) = (x_t^1(\lambda), x_t^2(\lambda))$. As t varies the geometry of the curve can change with swallowtails forming and disappearing.

Proposition 5.2. *If a swallowtail on the curve $x_t(\lambda)$ collapses to a point where $\lambda = \tilde{\lambda}$ when $t = \tilde{t}$, then*

$$\frac{\mathrm{d}x_{\tilde{t}}}{\mathrm{d}\lambda}(\tilde{\lambda}) = \frac{\mathrm{d}^2 x_{\tilde{t}}}{\mathrm{d}\lambda^2}(\tilde{\lambda}) = 0.$$

Proposition 5.3. *Assume that there exists a neighbourhood of $\tilde{\lambda} \in \mathbb{R}$ such that $\frac{\mathrm{d}x_t^\alpha}{\mathrm{d}\lambda}(\lambda) \neq 0$ for $t \in (\tilde{t} - \delta, \tilde{t})$ where $\delta > 0$. If a complex double point joins the curve $x_t(\lambda)$ at $\lambda = \tilde{\lambda}$ when $t = \tilde{t}$ then,*

$$\frac{\mathrm{d}x_{\tilde{t}}}{\mathrm{d}\lambda}(\tilde{\lambda}) = \frac{\mathrm{d}^2 x_{\tilde{t}}}{\mathrm{d}\lambda^2}(\tilde{\lambda}) = 0.$$

These provide a necessary condition for the formation or destruction of a swallowtail, and for complex double points to join or leave the main curve.

Definition 5.4. A family of parameterised curves $x_t(\lambda)$, (where λ is some intrinsic parameter) for which

$$\frac{\mathrm{d}x_{\tilde{t}}}{\mathrm{d}\lambda}(\tilde{\lambda}) = \frac{\mathrm{d}^2 x_{\tilde{t}}}{\mathrm{d}\lambda^2}(\tilde{\lambda}) = 0,$$

is said to have a point of swallowtail perestroika when $\lambda = \tilde{\lambda}$ and $t = \tilde{t}$.

As with generalised cusps, we have not ruled out further degeneracy at these points. Moreover, as Cayley highlighted, these points are not cusped and are barely distinguishable from an ordinary point of the curve [27].

5.1. The complex caustic in two dimensions

The complex caustic is the complete caustic found by allowing the parameter λ in the pre-parameterisation $x_t(\lambda) \in \mathbb{R}^2$ to vary over the complex plane. By considering the complex caustic, we are determining solutions $a = a_t$ and $\eta = \eta_t$ to

$$f'_{(x,t)}(a + i\eta) = f''_{(x,t)}(a + i\eta) = 0,$$

where $x \in \mathbb{R}^2$. We are interested in these points if they join the main caustic at some finite critical time $\tilde{t}$. That is, there exists a finite value $\tilde{t} > 0$ such that $\eta_t \to 0$ as $t \uparrow \tilde{t}$. If this holds, then a swallowtail can develop at the critical time $\tilde{t}$.

Theorem 5.5. *For a two-dimensional caustic, assume that $x_t(\lambda)$ is a real analytic function. If at a time $\tilde{t}$ a swallowtail perestroika occurs on the caustic, then $x = x_{\tilde{t}}(\lambda)$ is a real solution for x to*

$$f'_{(x,\tilde{t})}(\lambda) = f''_{(x,\tilde{t})}(\lambda) = f'''_{(x,\tilde{t})}(\lambda) = f^{(4)}_{(x,\tilde{t})}(\lambda) = 0,$$

where $\lambda = a_{\tilde{t}}$.

Theorem 5.6. *For a two-dimensional caustic, assume that $x_t(\lambda)$ is a real analytic function. If at a time $\tilde{t}$ there is a real solution for x to*

$$f'_{(x,\tilde{t})}(\lambda) = f''_{(x,\tilde{t})}(\lambda) = f'''_{(x,\tilde{t})}(\lambda) = f^{(4)}_{(x,\tilde{t})}(\lambda) = 0,$$

and the vectors $\nabla_x f'_{(x,\tilde{t})}(\lambda)$ and $\nabla_x f''_{(x,\tilde{t})}(\lambda)$ are linearly independent, then x is a point of swallowtail perestroika on the caustic.

Example. Let $V(x,y) = 0, k_t(x,y) \equiv 0$ and $S_0(x_0,y_0) = x_0^5 + x_0^6 y_0$. The caustic has no cusps for times $t < \tilde{t}$ and two cusps for times $t > \tilde{t}$ where $\tilde{t} = 4\sqrt{2} \times 33^{3/4} \times 7^{(-7/4)} = 2.5854\ldots$

At the critical time $\tilde{t}$ the caustic has a point of swallowtail perestroika as shown in Figures 9 and 10. The conjugate pairs of intersections of the curves in Figure 9 are the complex double points. There are five before the critical time and four afterwards. The remaining complex double points do not join the main caustic and so do not influence its behaviour for real times.

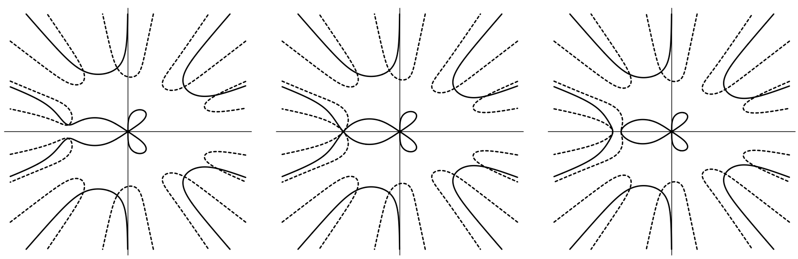


FIGURE 9. $\mathrm{Im}\{x_t(a + \mathrm{i}\eta)\} = 0$ (solid) and $\mathrm{Im}\{y_t(a + \mathrm{i}\eta)\} = 0$ (dashed) in (a, η) plane.

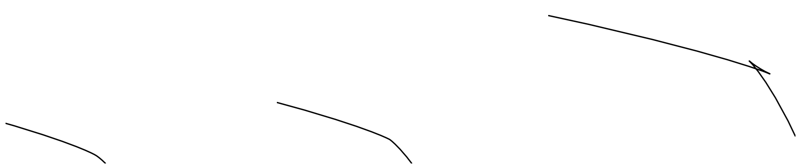


FIGURE 10. Caustic plotted at corresponding times.

5.2. Level surfaces

Unsurprisingly, these phenomena are not restricted to caustics. There is an interplay between the level surfaces and the caustics, characterised by their pre-images.

Proposition 5.7. *Assume that in two dimensions at* $x_0 \in \Phi_t^{-1} H_t^c \cap \Phi_t^{-1} C_t$ *the normal to the pre-level surface* $n(x_0) \neq 0$ *and the normal to the pre-caustic* $\tilde{n}(x_0) \neq 0$ *so that the pre-caustic is not cusped at* x_0*. Then* $\tilde{n}(x_0)$ *is parallel to* $n(x_0)$ *if and only if there is a generalised cusp on the caustic.*

Corollary 5.8. *Assume that in two dimensions at* $x_0 \in \Phi_t^{-1} H_t^c \cap \Phi_t^{-1} C_t$ *the normal to the pre-level surface* $n(x_0) \neq 0$*. Then at* $\Phi_t(x_0)$ *there is a point of swallowtail perestroika on the level surface* H_t^c *if and only if there is a generalised cusp on the caustic* C_t *at* $\Phi_t(x_0)$*.*

Example. Let $V(x, y) = 0$, $k_t(x, y) = 0$, and $S_0(x_0, y_0) = x_0^5 + x_0^6 y_0$. Consider the behaviour of the level surfaces through a point inside the caustic swallowtail at a fixed time as the point is moved through a cusp on the caustic. This is illustrated in Figure 11. Part (a) shows all five of the level surfaces through the point demonstrating how three swallowtail level surfaces collapse together at the cusp to form a single level surface with a point of swallowtail perestroika. Parts (b) and (c) show how one of these swallowtails collapses on its own and how its pre-image behaves.

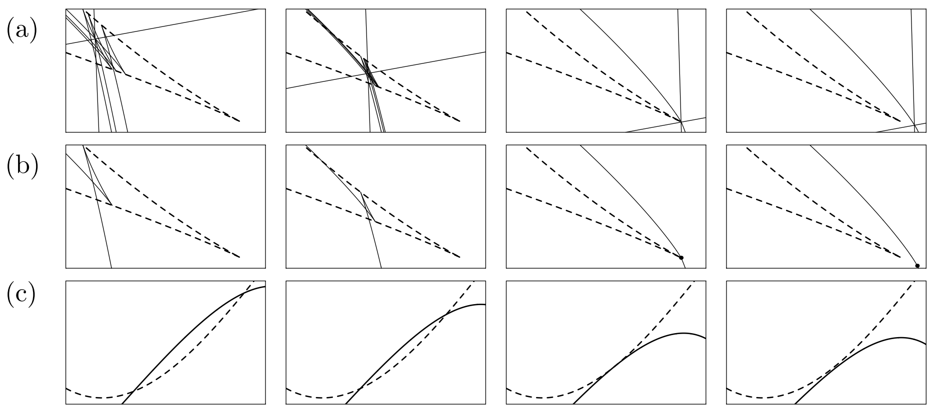


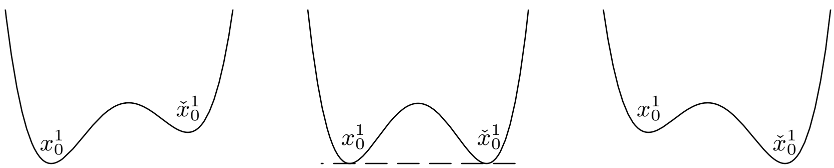
FIGURE 11. (a) All level surfaces (solid line) through a point as it crosses the caustic (dashed line) at a cusp, (b) one of these level surfaces with its complex double point, and (c) its real pre-image.

6. Maxwell sets

A jump will occur in the inviscid limit of the Burgers velocity field if we cross a point at which there are two different global minimisers $x_0(i)(x,t)$ and $x_0(j)(x,t)$ returning the same value of the action.

In terms of the reduced action function, the Maxwell set corresponds to values of x for which $f_{(x,t)}(x_0^1)$ has two critical points at the same height. If this occurs at the minimising value then the Burgers fluid velocity will jump as shown in Figure 12.

Before Maxwell set **On Cool Maxwell set** **Beyond Maxwell set**



Minimiser at x_0^1. *Two x_0's at same level.* *Minimiser jumps.*

FIGURE 12. The graph of the reduced action function as x crosses the Maxwell set.

6.1. The Maxwell-Klein set

We begin with the two-dimensional polynomial case by considering the classification of double points of a curve (Figure 7).

Lemma 6.1. *A point x is in the Maxwell set if and only if there is a Hamilton-Jacobi level surface with a point of self-intersection (crunode) at x.*

Proof. Follows from Definition 1.4. □

Definition 6.2. The Maxwell-Klein set B_t is the set of points which are non-cusp double points of some Hamilton-Jacobi level surface curve.

It follows from this definition that a point is in the Maxwell-Klein set if it is either a complex double point (acnode) or point of self-intersection (crunode) of some Hamilton-Jacobi level surface. Using the geometric results of DTZ outlined in Section 2, it is easy to calculate this set in the polynomial case as the cusps of the level surfaces sweep out the caustic.

Theorem 6.3. *Let D_t be the set of double points of the Hamilton-Jacobi level surfaces, C_t the caustic set, and B_t the Maxwell-Klein set. Then, from Cayley and Klein's classification of double points as crunodes, acnodes, and cusps, by definition, $D_t = C_t \cup B_t$, and the corresponding defining algebraic equations factorise $D_t = C_t^n \cdot B_t^m$, where m, n are positive integers.*

Proof. Follows from Proposition 2.8 and Lemma 6.1. □

Theorem 6.4. *Let $\rho_{(t,c)}(x)$ be the resultant*

$$\rho_{(t,c)}(x) = R\left(f_{(x,t)}(\cdot) - c, f'_{(x,t)}(\cdot)\right),$$

where $x = (x_1, x_2)$. Then $x \in D_t$ if and only if for some c,

$$\rho_{(t,c)}(x) = \frac{\partial \rho_{(t,c)}}{\partial x_1}(x) = \frac{\partial \rho_{(t,c)}}{\partial x_2}(x) = 0.$$

Further,

$$D_t(x) = \gcd\left(\rho_t^1(x), \rho_t^2(x)\right),$$

where $\gcd(\cdot,\cdot)$ *denotes the greatest common divisor and ρ_t^1 and ρ_t^2 are the resultants*

$$\rho_t^1(x) = R\left(\rho_{(t,\cdot)}(x), \frac{\partial \rho_{(t,\cdot)}}{\partial x_1}(x)\right) \quad \textit{and} \quad \rho_t^2(x) = R\left(\frac{\partial \rho_{(t,\cdot)}}{\partial x_1}(x), \frac{\partial \rho_{(t,\cdot)}}{\partial x_2}(x)\right).$$

Proof. Recall that the equation of the level surface of Hamilton-Jacobi functions is merely the result of eliminating x_0^1 between the equations

$$f_{(x,t)}(x_0^1) = c \quad \text{and} \quad f'_{(x,t)}(x_0^1) = 0.$$

We form the resultant $\rho_{(t,c)}(x)$ using Sylvester's formula. The double points of the level surface must satisfy for some $c \in \mathbb{R}$,

$$\rho_{(t,c)}(x) = 0, \quad \frac{\partial \rho_{(t,c)}}{\partial x_1}(x) = 0 \quad \text{and} \quad \frac{\partial \rho_{(t,c)}}{\partial x_2}(x) = 0.$$

Sylvester's formula proves that all three equations are polynomial in c. To proceed we eliminate c between pairs of these equations using resultants giving

$$R\left(\rho_{(t,\cdot)}(x), \frac{\partial \rho_{(t,\cdot)}}{\partial x_1}(x)\right) = \rho_t^1(x) \quad \text{and} \quad R\left(\frac{\partial \rho_{(t,\cdot)}}{\partial x_1}(x), \frac{\partial \rho_{(t,\cdot)}}{\partial x_2}(x)\right) = \rho_t^2(x).$$

Let $D_t = \gcd(\rho_t^1, \rho_t^2)$ be the greatest common divisor of the algebraic ρ_t^1 and ρ_t^2. Then $D_t(x) = 0$ is the equation of double points. □

We now extend this to d dimensions, where the Maxwell-Klein set corresponds to points which satisfy the Maxwell set condition but have both real pre-images (Maxwell) or complex pre-images (Klein).

Theorem 6.5. *Let the reduced action function $f_{(x,t)}(x_0^1)$ be a polynomial in all space variables. Then the set of all possible discontinuities for a d-dimensional Burgers fluid velocity field in the inviscid limit is the double discriminant*

$$D(t) := D_c\left\{D_\lambda\left(f_{(x,t)}(\lambda) - c\right)\right\} = 0,$$

where $D_x(p(x))$ is the discriminant of the polynomial p with respect to x.

Proof. By considering the Sylvester matrix of the first discriminant,

$$D_\lambda\left(f_{(x,t)}(\lambda) - c\right) = K \prod_{i=1}^{m} \left(f_{(x,t)}(x_0^1(i)(x,t)) - c\right),$$

where $x_0^1(i)(x,t)$ is an enumeration of the real and complex roots λ of $f'_{(x,t)}(\lambda) = 0$ and K is some constant. Then the second discriminant is simply

$$D_c\left(D_\lambda\left(f_{(x,t)}(\lambda) - c\right)\right) = K^{2m-2}\prod_{i<j}\left(f_{(x,t)}(x_0^1(i)(x,t)) - f_{(x,t)}(x_0^1(j)(x,t))\right)^2 .$$

□

Theorem 6.6. *The double discriminant $D(t)$ factorises as*

$$D(t) = b_0^{2m-2}\cdot(C_t)^3\cdot(B_t)^2 ,$$

where $B_t = 0$ is the equation of the Maxwell-Klein set and $C_t = 0$ is the equation of the caustic. The expressions B_t and C_t are both algebraic in x and t.

Proof. See [23]. □

Example (The polynomial swallowtail). Let $V(x,y) = 0$, $k_t(x,y) = 0$ and $S_0(x_0,y_0) = x_0^5 + x_0^2 y_0$. The Maxwell-Klein set can be found by factorisation giving

$$\begin{aligned} 0 \;=\; & -675 + 52t^4 - t^8 + 3120t^3x - 224t^7x + 4t^{11}x - 38400t^2x^2 + 1408t^6x^2 \\ & +128000tx^3 - 5400ty + 312t^5y - 4t^9y + 12480t^4xy - 448t^8xy \\ & -76800t^3x^2y - 16200t^2y^2 + 624t^6y^2 - 4t^{10}y^2 + 12480t^5xy^2 \\ & -21600t^3y^3 + 416t^7y^3 - 10800t^4y^4. \end{aligned}$$

Outside of the swallowtail on the caustic there are two real and two complex pre-images whereas inside there are four real and no complex pre-images. Therefore, any part of the Maxwell-Klein set outside of the caustic swallowtail must correspond to Klein double points and any part inside must correspond to the Maxwell set. This is shown in Figure 13.

6.2. The pre-Maxwell set

If the Maxwell set is defined as in Definition 1.4, then the pre-Maxwell set is the set of all the pre-images x_0 and $\check{x}_0$ which give rise to the Maxwell set.

Definition 6.7. The pre-Maxwell set $\Phi_t^{-1}M_t$ is the set of all points $x_0 \in \mathbb{R}^d$ where there exists $x, \check{x}_0 \in \mathbb{R}^d$ such that $x = \Phi_t(x_0)$ and $x = \Phi_t(\check{x}_0)$ with $x_0 \neq \check{x}_0$ and

$$\mathcal{A}(x_0,x,t) = \mathcal{A}(\check{x}_0,x,t).$$

With the caustic and level surfaces, each regular point was linked by Φ_t^{-1} to a single point on the relevant pre-surafce. However, every point on the Maxwell set is linked by Φ_t^{-1} to at least two points on the pre-Maxwell set.

Theorem 6.8. *The pre-Maxwell set is given by the discriminant $D_{\check{x}_0^1}\left(G(\check{x}_0^1)\right) = 0$ where*

$$G(\check{x}_0^1) = \frac{f_{(\Phi_t(x_0),t)}(x_0^1) - f_{(\Phi_t(x_0),t)}(\check{x}_0^1)}{(x_0^1 - \check{x}_0^1)^2}.$$

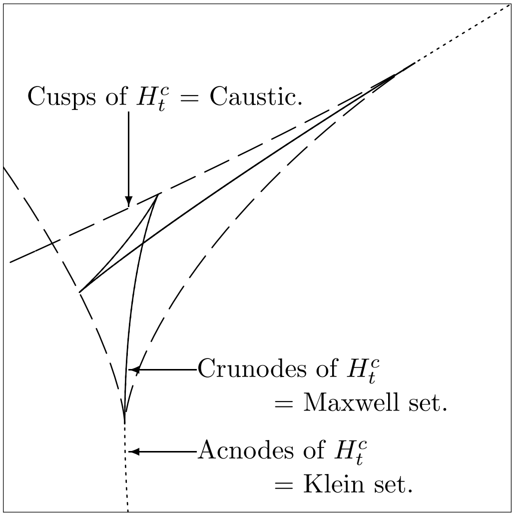


FIGURE 13. The caustic and Maxwell-Klein set.

Proof. From Definition 6.7 and Theorem 3.4 it follows that the pre-Maxwell set is found by eliminating x and $\check{x}_0^1$ between

$$f_{(x,t)}(x_0^1) = f_{(x,t)}(\check{x}_0^1) \quad f'_{(x,t)}(x_0^1) = f'_{(x,t)}(\check{x}_0^1) = 0.$$

This surface would include the pre-caustic where $x_0^1 = \check{x}_0^1$ and so this repeated root must be eliminated. □

We can use this to pre-parameterise the Maxwell set as has been done with the caustic and level surfaces. By restricting the parameter to be real, we only get the Maxwell set as the Klein points have complex pre-images.

We now summarise the results of [25].

Lemma 6.9. *Assume that a point x on the Maxwell set corresponds to exactly two pre-images on the pre-Maxwell set, x_0 and $\tilde{x}_0$. Then the normal to the pre-Maxwell set at x_0 is to within a scalar multiplier given by*

$$\begin{aligned} n(x_0) \quad = \quad & -\left(\frac{\partial^2 \mathcal{A}}{\partial x_0^2}(x_0, x, t)\right)\left(\frac{\partial^2 \mathcal{A}}{\partial x \partial x_0}(x_0, x, t)\right)^{-1} \\ & \left(\dot{X}(t, x_0, \nabla S_0(x_0)) - \dot{X}(t, \tilde{x}_0, \nabla S_0(\tilde{x}_0))\right). \end{aligned}$$

Corollary 6.10. *In two dimensions let the pre-Maxwell set meet the pre-caustic at a point x_0 where $n \neq 0$ and*

$$\ker\left(\frac{\partial^2 \mathcal{A}}{(\partial x_0)^2}(x_0, \Phi_t(x_0), t)\right) = \langle e_0 \rangle,$$

where e_0 is the zero eigenvector. Then the tangent plane to the pre-Maxwell set at x_0, T_{x_0} is spanned by e_0.

Proposition 6.11. *Assume that in two dimensions at $x_0 \in \Phi_t^{-1} M_t$ the normal $n(x_0) \neq 0$ so that the pre-Maxwell set does not have a generalised cusp at x_0. Then the Maxwell set can only have a cusp at $\Phi_t(x_0)$ if $\Phi_t(x_0) \in C_t$. Moreover, if*

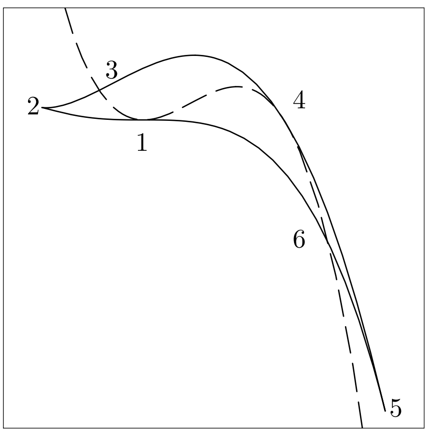
$$x = \Phi_t(x_0) \in \Phi_t \left\{ \Phi_t^{-1} C_t \cap \Phi_t^{-1} M_t \right\},$$

the Maxwell set will have a generalised cusp at x.

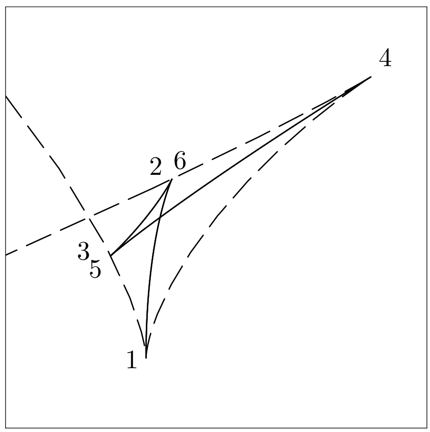
Corollary 6.12. *In two dimensions, if the pre-Maxwell set intersects the pre-caustic at a point x_0, so that there is a cusp on the Maxwell set at the corresponding point where it intersects the caustic, then the pre-Maxwell set touches the pre-level surface $\Phi_t^{-1} H_t^c$ at the point x_0. Moreover, if the cusp on the Maxwell set intersects the caustic at a regular point of the caustic, then there will be a cusp on the pre-Maxwell set which also meets the same pre-level surface $\Phi_t^{-1} H_t^c$ at another point $\check{x}_0$.*

Corollary 6.13. *When the pre-Maxwell set touches the pre-caustic and pre-level surface, the Maxwell set intersects a cusp on the caustic.*

Example (The polynomial swallowtail). Let $V(x,y) = 0$, $k_t(x,y) = 0$ and $S_0(x_0, y_0) = x_0^5 + x_0^2 y_0$.



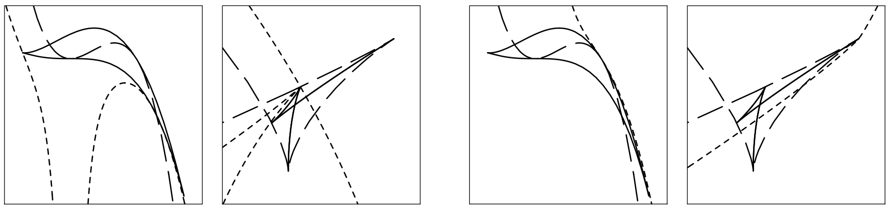
Pre-curves



Curves

FIGURE 14. The caustic (dashed) and Maxwell set (solid line).

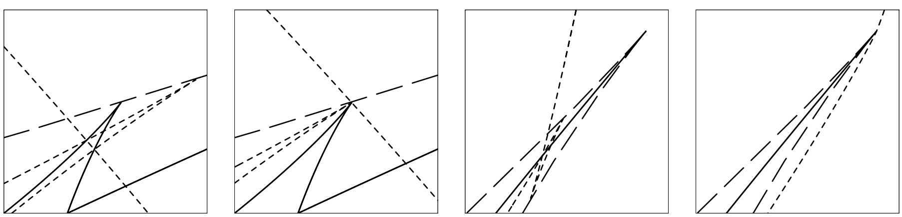
From Proposition 6.11, the cusps on the Maxwell set correspond to the intersections of the pre-curves (points 3 and 6 on Figure 14). But from Corollary 6.12, the cusps on the Maxwell set also correspond to the cusps on the pre-Maxwell set (points 2 and 5 on Figure 14 and also Figure 15). The Maxwell set terminates when it reaches the cusps on the caustic. These points satisfy the condition for a generalised cusp but, instead of appearing cusped, the curve stops and the parameterisation begins again in the sense that it maps back exactly on itself. At such points the pre-surfaces all touch (Figure 15).



Cusp on Maxwell set *Cusp on caustic*

FIGURE 15. The caustic (long dash) and Maxwell set (solid line) with the level surfaces (short dash) through special points.

These two different forms of cusps correspond to very different geometric behaviours of the level surfaces. Where the Maxwell set stops or cusps corresponds to the disappearance of a point of self-intersection on a level surface. There are two distinct ways in which this can happen. Firstly, the level surface will have a point of swallowtail perestroika when it meets a cusp on the caustic. At such a point only one point of self-intersection will disappear, and so there will be only one path of the Maxwell set which will terminate at that point. However, when we approach the caustic at a regular point, the level surface must have a cusp but not a swallowtail perestoika. This corresponds to the collapse of the second system of double points in Figure 8. Thus, two different points of self-intersection coalesce and so two paths of the Maxwell set must approach the point and produce the cusp (see Figure 16).



Approaching the caustic *Approaching a cusp on the caustic*

FIGURE 16. The caustic (long dash) Maxwell set (solid line) and level surface (short dash).

7. Some applications to turbulence in two dimensions

7.1. Real turbulence and the ζ process

Definition 7.1. The turbulent times t are times when the pre-level surface of the minimising Hamilton-Jacobi function *touches* the pre-caustic. Such times t are zeros of a stochastic process $\zeta^c(.)$, i.e., $\zeta^c(t) = 0$.

These turbulent times are times at which the number of cusps on the corresponding level surface will change. We begin with some minor generalisations of results in RTW [34] and also [23, 26].

Proposition 7.2. *Assume Φ_t is globally reducible and that $x_t(\lambda)$ is the pre-parameterisation of a two-dimensional caustic. Then the turbulence process at λ is given by*

$$\zeta^c(t) = f_{(x_t(\lambda_0),t)}(\lambda_0) - c,$$

where $f_{(x,t)}(x_0^1)$ is the reduced action *evaluated at points $x = x_t(\lambda_0)$ where $x_t(\lambda_0) = \Phi_t(\lambda_0, x_0^2(\lambda_0)) \in C_t$, $\lambda = \lambda_0$ satisfying*

$$\dot{X}_t(\lambda) \cdot \frac{\mathrm{d}x_t}{\mathrm{d}\lambda}(\lambda) = 0,$$

where $\dot{X}_t(\lambda) = \dot{\Phi}_t(\lambda, x_{0,\mathrm{C}}^2(\lambda))$ and $x_t(\lambda_0) \in C_t^{\mathrm{c}}$, the cool part of the caustic.

Hence, there are three kinds of real stochastic turbulence:

1. *Cusped*, where there is a cusp on the caustic,
2. *Zero speed*, where the Burgers fluid velocity is zero,
3. *Orthogonal*, where the Burgers fluid velocity is orthogonal to the caustic.

Proof. The number of cusps on the relevant pre-level surface is

$$n_c(t) = \#\left\{\lambda \in \mathbb{R} : f_{(x_t(\lambda),t)}(\lambda) = c\right\},$$

where the roots $\lambda = \lambda_0$ correspond to points in the cool part of the caustic. The pre-surfaces touch when $n_c(t)$ changes, which occurs when

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} f_{(x_t(\lambda),t)}(\lambda) = 0. \qquad \square$$

For stochastic turbulence to be intermittent we require that the process $\zeta^c(t)$ is recurrent.

Proposition 7.3. *Let $V(x,y) = 0$, $k_t(x,y) = x$ and*

$$S_0(x_0, y_0) = f(x_0) + g(x_0)y_0,$$

where f, g, f' and g' are zero at $x_0 = a$ but $g''(a) \neq 0$. Then, for orthogonal turbulence at a,

$$\zeta^c(t) = -a\epsilon W_t + \epsilon^2 W_t \int_0^t W_s \,\mathrm{d}s - \frac{\epsilon^2}{2} \int_0^t W_s^2 \,\mathrm{d}s - c.$$

We note the following result of RTW [34].

Lemma 7.4. *Let W_t be a $BM(\mathbb{R})$ process starting at 0, c any real constant and*

$$Y_t = -a\epsilon W_t + \epsilon^2 W_t \int_0^t W_s \,\mathrm{d}s - \frac{\epsilon^2}{2}\int_0^t W_s^2 \,\mathrm{d}s - c.$$

Then, with probability 1, there exists a sequence of times $t_n \nearrow \infty$ such that

$$Y_{t_n} = 0 \quad \textit{for every } n.$$

We also note that this can be extended to a d-dimensional setting where for a d-dimensional Wiener process $W(t)$ the zeta process can be found explicitly [22].

Theorem 7.5. *In d dimensions, the zeta process is given by*

$$\zeta_t \;=\; f^0_{(x_t^0(\lambda),t)}(\lambda_1) - \epsilon x_t^0(\lambda)\cdot W(t) + \epsilon^2 W(t)\cdot \int_0^t W(s)\,\mathrm{d}s + \frac{\epsilon^2}{2}\int_0^t |W(s)|^2 \,\mathrm{d}s,$$

where $f^0_{(x,t)}(x_0^1)$ denotes the deterministic reduced action function, $x_t^0(\lambda)$ denotes the pre-parameterisation of the deterministic caustic, and λ must satisfy the equation

$$\nabla_\lambda \left(f^0_{(x_t^0(\lambda),t)}(\lambda_1) - \epsilon x_t^0(\lambda)\cdot W(t)\right) = 0.$$

When λ is deterministic, the recurrence of this process can be shown using the same argument as for the two-dimensional case (further results on recurrence can be found in [24]). Here we recapitulate our belief that cusped turbulence will be the most important. As we have shown, when the cusp on the caustic passes through a level surface, it forces a swallowtail to form on the level surface. The points of self-intersection of this swallowtail form the Maxwell set.

7.2. Complex turbulence and the resultant η process

We now consider a completely different approach to turbulence. Let $\left(\lambda, x^2_{0,\mathrm{C}}(\lambda)\right)$ denote the parameterisation of the pre-caustic at time t. When

$$Z_t = \operatorname{Im}\left\{\Phi_t(a+\mathrm{i}\eta, x^2_{0,\mathrm{C}}(a+\mathrm{i}\eta))\right\},$$

is random, the values of $\eta(t)$ for which $Z_t = 0$ will form a stochastic process. The zeros of this new process will correspond to points at which the real pre-caustic touches the complex pre-caustic. The points at which these surfaces touch correspond to swallowtail perestroikas on the caustic. When such a perestroika occurs there is a solution of the equations

$$f'_{(x,t)}(\lambda) = f''_{(x,t)}(\lambda) = f'''_{(x,t)}(\lambda) = f^{(4)}_{(x,t)}(\lambda) = 0.$$

Assuming that $f_{(x,t)}(x_0^1)$ is polynomial in x_0^1 we can use the resultant to state explicit conditions for which this holds [23].

Lemma 7.6. *Let g and h be polynomials of degrees m and n, respectively, with no common roots or zeros. Let $f = gh$ be the product polynomial. Then the resultant is*

$$R(f,f') = (-1)^{mn}\left(\frac{m!n!}{N!}\frac{f^{(N)}(0)}{g^{(m)}(0)h^{(n)}(0)}\right)^{N-1} R(g,g')R(h,h')R(g,h)^2,$$

where $N = m + n$ *and* $R(g, h) \neq 0$.

Since $f'_{(x_t(\lambda),t)}(x_0^1)$ is a polynomial in x_0 with real coefficients, its zeros are real or occur in complex conjugate pairs. Of the real roots, $x_0 = \lambda$ is repeated. So,

$$f'_{(x_t(\lambda),t)}(x_0^1) = (x_0^1 - \lambda)^2 Q_{(\lambda,t)}(x_0^1) H_{(\lambda,t)}(x_0^1),$$

where Q is the product of quadratic factors

$$Q_{(\lambda,t)}(x_0^1) = \prod_{i=1}^{q} \left\{ (x_0^1 - a_t^i)^2 + (\eta_t^i)^2 \right\},$$

and $H_{(\lambda,t)}(x_0^1)$ the product of real factors corresponding to real zeros. This gives

$$f'''_{(x_t(\lambda),t)}(x_0^1)\Big|_{x_0^1=\lambda} = 2 \prod_{i=1}^{q} \left\{ (\lambda - a_t^i)^2 + (\eta_t^i)^2 \right\} H_{(\lambda,t)}(\lambda).$$

We now assume that the real roots of H are distinct as are the complex roots of Q. Denoting $f'''_{(x_t(\lambda),t)}(x_0^1)\Big|_{x_0^1=\lambda}$ by $f_t'''(\lambda)$ etc, a simple calculation gives

$$\begin{aligned} &\left| R_\lambda(f_t'''(\lambda), f_t^{(4)}(\lambda)) \right| \\ &\quad = K_t \prod_{k=1}^{q} (\eta_t^k)^2 \prod_{j \neq k} \left\{ (a_t^k - a_t^j)^4 + 2((\eta_t^k)^2 + (\eta_t^j)^2)(a_t^k - a_t^j)^2 + ((\eta_t^k)^2 - (\eta_t^j)^2)^2 \right\} \\ &\qquad \times |R_\lambda(H, H')| \, |R_\lambda(Q, H)|^2, \end{aligned}$$

K_t being a positive constant. Thus, the condition for a swallowtail perestroika to occur is that

$$\rho_\eta(t) := \left| R_\lambda(f_t'''(\lambda), f_t^{(4)}(\lambda)) \right| = 0,$$

where we call $\rho_\eta(t)$ the *resultant eta process.*

When the zeros of $\rho_\eta(t)$ form a perfect set, swallowtails will spontaneously appear and disappear on the caustic infinitely rapidly. As they do so, the geometry of the cool part of the caustic will rapidly change as the λ shaped sections typical of a swallowtail caustic appear and disappear. Moreover, Maxwell sets will be created and destroyed with each swallowtail that forms and vanishes adding to the turbulent nature of the solution in these regions. We call this 'complex turbulence' occurring at the turbulent times which are the zeros of the resultant eta process.

Complex turbulence can be seen as a special case of real turbulence which occurs at specific generalised cusps of the caustic. Recall that when a swallowtail perestroika occurs on a curve, it also satisfies the conditions for having a generalised cusp. Thus, the zeros of the resultant eta process must coincide with some of the zeros of the zeta process for certain forms of cusped turbulence. At points where the complex and real pre-caustic touch, the real pre-caustic and pre-level surface touch in a particular manner (a double touch) since at such a point two swallowtail perestroikas on the level surface have coalesced.

Thus, our separation of complex turbulence from real turbulence can be seen as an alternative form of categorisation to that outlined in Section 7.1 which could be extended to include other perestroikas.

References

[1] V. I. Arnol'd, S. F. Shandarin, and Y. B. Zeldovich, *The large scale structure of the universe I*, Geophys. Astrophys. Fluid Dynamics, **20** (1982), 111–130.

[2] V. I. Arnol'd, *Catastrophe Theory*, Springer-Verlag, Berlin, 1986.

[3] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, 1989.

[4] V. I. Arnol'd, *Singularities of Caustics and Wave Fronts*, Mathematics and its Applications (Soviet Series), 62, Kluwer Academic Publishers Group, 1990.

[5] C. Dafermos, *Hyperbolic Conservation Laws in Continuum Physics*, Grundlehren der Mathematischen Wissenschaten, 325, Springer-Verlag, Berlin, 2000.

[6] I. M. Davies, A. Truman and H. Zhao, *Stochastic heat and Burgers equations and their singularities – Geometrical and analytic properties (the fish the butterfly, and why.)*, UWS MRRS Preprint 01-03, (2001), available here: `http://www.ma.texas.edu/mp_arc-bin/mpa?yn=01-45`.

[7] I. M. Davies, A. Truman, and H. Zhao, *Stochastic heat and Burgers equations and their singularities I – Geometric properties*, J. Math. Phys., **43** (2002), 3293-3328.

[8] I. M. Davies, A. Truman, and H. Zhao, *Stochastic heat and Burgers equations and their singularities II – Analytical properties and limiting distributions*, J. Math. Phys., **46** (2005), article number 043515.

[9] E. Weinan, K. Khanin, A. Mazel, and Y. Sinai, *Invariant measures for Burgers equations with stochastic forcing*, Ann. Math., **151** (2000), 877–960.

[10] K. D. Elworthy, A. Truman, and H. Zhao, *Stochastic elementary formulae on caustics I: One-dimensional linear heat equations*, UWS MRRS Preprint 01-22, 2001.

[11] K. D. Elworthy, A. Truman, and H. Zhao, *Generalised Ito formulae and space-time Lebesgue-Stieltjes integrals of local times*, Séminaire de Probabilites Strasbourg, **40** (2005), to appear.

[12] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1998.

[13] R. Gilmore, *Catastrophe Theory for Scientists and Engineers*, John Wiley, New York, 1981.

[14] E. Hopf, *The partial differential equation* $u_t + uu_x = \mu u_{xx}$, Comm. Pure and Appl. Math., **3 (3)** (1950), 201–230.

[15] R. C. Hwa and V. L. Teplitz, *Homology and Feynman Integrals*, W. A. Benjamin, New York, 1966.

[16] M. Kac, *Probability and Related Topics in Physical Science*, Interscience Publishers, New York, 1959.

[17] F. Klein, *Über den Verlauf der Abelschen Integrale bei den Kurven vierten Grades*, in: Gesammelte Mathematische Abhandlungen II, R. Fricke and H. Vermeil, Editors, Springer, Berlin, (1922).

[18] V. N. Kolokoltsov, R. L. Schilling, and A. E. Tyukov, *Estimates for multiple stochastic integrals and stochastic Hamilton-Jacobi equations*, Rev. Mat. Iberoamericana, **20** (2004), 333–380.

[19] H. Kunita, *Stochastic differential equations and stochastic flows of homeomorphisms*, in: Stochastic Analysis and Applications, Advances in Probability and Related Topics, Vol. 7, M. A. Pinsky, Editor, Marcel Dekker, New York, 1984.

[20] V. P. Maslov, *Perturbation Theory and Asymptotic Methods*, Dunod, Paris, 1972.

[21] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics*, Mathematical Physics and Applied Mathematics, Vol. 7, Riedel Publishing Company, 1981.

[22] A. D. Neate, *A One-Dimensional Analysis of the Singularities of the d-Dimensional Stochastic Burgers Equation*, PhD thesis, UWS, 2005.

[23] A. D. Neate and A. Truman, *A one-dimensional analysis of real and complex turbulence and the Maxwell set for the stochastic Burgers equation*, J. Phys. A: Math. Gen., **38** (2005), 7093–7127.

[24] A. D. Neate and A. Truman, *A one-dimensional analysis of turbulence and its intermittence for the d-dimensional stochastic Burgers equation*, Markov Processes and Related Fields, 2005, submitted.

[25] A. D. Neate and A. Truman, *Geometrical properties of the Maxwell set for the stochastic Burgers equation*, Lett. Math. Phys., 2006, submitted.

[26] C. Reynolds, *On the polynomial swallowtail and cusp singularities of stochastic Burgers equations*, PhD thesis, UWS, 2002.

[27] G. Salmon, *A Treatise on the Higher Plane Curves* G E Stechert Co., New York, 1934.

[28] S. F. Shandarin and Y. B. Zeldovich, *The large scale structure of the universe 2: turbulence, intermittency, structures in a self gravitating medium*, Rev. Mod. Phys., **6** (1989), 185–220.

[29] A. Truman and H. Zhao, *The stochastic Hamilton Jacobi equations and related topics: a survey*, LMS Lecture Note Ser. 216, Cambridge University Press, (1995),287–303.

[30] A. Truman and H. Zhao, *The stochastic Hamilton Jacobi equation, stochastic heat equations and Schrödinger equations*, Stochastic Analysis and Applications, Proc. of the 5th Gregynog Symp. held in Powys (July 9–14, 1995), I. M. Davies et al., Editors, World Scientific, (1996), 441–464.

[31] A. Truman and H. Zhao, *On Stochastic Diffusion Equations and Stochastic Burgers' Equations*, J. Math. Phys., **37** (1996), 283–307.

[32] A. Truman and H. Zhao, *Quantum Mechanics of Charged Particles in Random Electromagnetic Fields*, J. Math. Phys., **37** (1996), 3180–3197.

[33] A. Truman and H. Zhao, *Stochastic Burgers equations and their semi classical expansions*, Comm. Math. Phys., **194** (1998), 231–248.

[34] A. Truman, C. N. Reynolds, and D. Williams, *Stochastic Burgers equation in d-dimensions – a one-dimensional analysis: hot and cool caustics and intermittence of stochastic turbulence*, in: Probabilistic Methods in Fluids, I. M. Davies et al., Editors, World Scientific, Singapore, (2003), 239–262.

[35] B. L. Van Der Waerden, *Modern Algebra*, Vols. 1 and 2, Frederick Ungar Publishing, New York, 1949.

Andrew D. Neate and Aubrey Truman
Department of Mathematics
University of Wales Swansea
Singleton Park
Swansea, SA2 8PP, Wales
UK
e-mail: a.truman@swan.ac.uk

Progress in Probability, Vol. 59, 331–344

Attractors for Ergodic and Monotone Random Dynamical Systems

Michael Scheutzow

Abstract. We relate ergodicity, monotonicity and attractors of a random dynamical system (rds). Our first result states that an rds which is both monotone and ergodic has a weak random attractor which consists of a single point. Then we show that ergodicity alone is insufficient for the existence of a weak random attractor. In particular we present an rds in $\mathbb{R}^d$, $d \geq 2$ namely an isotropic Brownian flow with drift, whose single-point motion is an ergodic diffusion process and which does not have a weak attractor. It seems that this is the first example of this kind in the literature.

Mathematics Subject Classification (2000). Primary 37H05; Secondary 34F05.

Keywords. Random dynamical system, attractor, diffusion, monotone system, isotropic flow.

1. Introduction

Our motivation for studying the relationship between ergodicity, monotonicity and the existence of a random attractor for a random dynamical system (rds) is to provide a rather general sufficient condition for the existence of a random attractor. Indeed, several authors have proved the existence of random attractors for particular systems using ad-hoc methods and in many cases the systems were in fact monotone and ergodic. When an rds is monotone, then this property is usually very easy to prove. Ergodicity is not always easy to prove – but in most cases still considerably easier than to prove the existence of an attractor. The result in Section 2 provides a stronger conclusion than most other results on attractors in that we show, that the attractor is trivial (i.e., consists of a single point). On the other hand, due to the generality of the set-up, we can only show that the attractor attracts all deterministic compact sets in probability while some authors prove almost sure attraction for a larger class of sets.

The article is organized as follows: we first introduce random dynamical systems and attractors. Section 2 contains the main result about attractors for monotone and ergodic systems. This section is based on the joint paper [7] with Igor Chueshov. Therefore we will omit detailed proofs. In Section 3 we provide a class of examples of ergodic systems which do not have a weak attractor. These examples are random dynamical systems generated by a stochastic differential equation on $\mathbb{R}^d$, $d \geq 2$, with bounded drift and isotropic diffusion part. In these examples, the single-point motion is a nondegenerate diffusion process with additive white noise. To the best of our knowledge, these are the first examples in the literature of this kind. In Section 4, we briefly address the relationship between ergodicity and point attractors and in the last section we discuss the relationship between the existence of a trivial random attractor and *coupling from the past*.

We start by explaining some basic concepts in the theory of random dynamical systems, which has been developed by Ludwig Arnold and his group during the past two decades, see [1]: let (S, d) be a complete, separable metric space and let $\mathbf{T}$ be either $\mathbb{R}$ or $\mathbb{Z}$. Further $\mathbf{T}_+$ denotes the set of nonnegative elements from $\mathbf{T}$. We denote by $\mathcal{B}(S)$ the Borel σ-algebra of subsets of S. By definition, a *random dynamical system* with time $\mathbf{T}_+$ and state space S is a pair (ϑ, φ) consisting of the following two objects:

- A *metric dynamical system* (mds) $\vartheta \equiv (\Omega, \mathcal{F}, \mathbf{P}, \{\vartheta(t), t \in \mathbf{T}\})$, i.e., a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with a family of measure-preserving transformations $\vartheta \equiv \{\vartheta(t) : \Omega \to \Omega,\ t \in \mathbf{T}\}$ such that
  (a) $\vartheta(0) = \mathrm{id}$, $\vartheta(t) \circ \vartheta(s) = \vartheta(t+s)$ for all $t, s \in \mathbf{T}$;
  (b) the map $(t, \omega) \mapsto \vartheta(t)\omega$ is measurable and
  (c) $\vartheta(t)\mathbf{P} = \mathbf{P}$ for all $t \in \mathbf{T}$.
- A (perfect) *cocycle* φ over ϑ of continuous mappings of S with one-sided time $\mathbf{T}_+$, i.e., a measurable mapping
$$\varphi : \mathbf{T}_+ \times \Omega \times S \to S, \quad (t, \omega, x) \mapsto \varphi(t, \omega)x$$
such that the mapping $x \mapsto \varphi(t, \omega)x$ is continuous for every $t \geq 0$ and $\omega \in \Omega$ and it satisfies the cocycle property:
$$\varphi(0, \omega) = \mathrm{id}, \quad \varphi(t+s, \omega) = \varphi(t, \vartheta(s)\omega) \circ \varphi(s, \omega)$$
for all $t, s \geq 0$ and $\omega \in \Omega$.

Note that we do not require φ to be continuous with respect to t in case $\mathbf{T} = \mathbb{R}$.

We call an rds *ergodic* in case there exists a probability measure π on S such that for each $x \in S$, the law of $\varphi(t, \omega)x$ converges weakly to π. Note that we do not assume that the one-point motion is Markovian.

Next, we introduce the concept of a *random attractor*.

Definition 1.1. Let $\mathcal{C}$ be the family of nonempty compact subsets of S. The mapping $\mathcal{A} : \Omega \to \mathcal{C}$ is called an *invariant random compact set* if

(i) $\omega \mapsto d(x, \mathcal{A}(\omega))$ is measurable for each $x \in S$, where $d(x, \mathcal{A}) = \inf_{y \in \mathcal{A}} d(x, y)$,

(ii) there exists a set $\widetilde{\Omega} \in \mathcal{F}$ of full measure which is invariant under $\vartheta(t)$ for each $t \in \mathbf{T}$ such that $\varphi(t,\omega)(\mathcal{A}(\omega)) = \mathcal{A}(\vartheta(t)\omega)$ for all $\omega \in \widetilde{\Omega}$ and $t \in \mathbf{T}_+$.

An invariant random compact set is called a *pullback attractor* if for each compact set $B \subseteq S$

$$\lim_{t\to\infty} \sup_{x\in B} d(\varphi(t,\vartheta(-t)\omega)x, \mathcal{A}(\omega)) = 0 \quad \text{a.s.},$$

and a *weak attractor* if for each compact set $B \subseteq S$

$$\lim_{t\to\infty} \sup_{x\in B} d(\varphi(t,\vartheta(-t)\omega)x, \mathcal{A}(\omega)) = 0 \quad \text{in probability.}$$

Since almost-sure convergence implies convergence in probability, every pullback attractor is also a weak attractor. The converse is not true (see, e.g., [19] for examples). Clearly both notions coincide in case S is compact. The concept of a pullback attractor was proposed independently in [10] and [21]. Weak attractors were introduced in [17]. If an attractor (weak or pullback) exists, then it is unique up to sets of measure zero. Sometimes a random attractor $\mathcal{A}$ is defined a bit differently or more generally by requiring that $\mathcal{A}$ attracts a different class of sets than the compact ones (for example all bounded deterministic sets or a class of random sets).

2. Monotone random dynamical systems

This section is essentially a condensed version of [7]. Proofs and detailed examples can be found there. We alert the reader that in [7] we required an attractor to attract all bounded deterministic sets instead of all compact sets (which necessitates only few changes).

To introduce a *monotone* rds, we need a partially ordered state space S. We will assume that (S,d) is a suitable subset of an ordered Banach space V:

Let V be a real separable Banach space with a cone $V_+ \subset V$. By definition, V_+ is a closed convex set in V such that $\lambda v \in V_+$ for all $\lambda \geq 0$, $v \in V_+$ and $V_+ \cap (-V_+) = \{0\}$. The cone V_+ defines a partial order relation on V via $x \leq y$ iff $y - x \in V_+$ which is compatible with the vector space structure of V. If V_+ has nonempty interior $\operatorname{int} V_+$, we say that the cone V_+ is *solid*. For elements a and b in V such that $a \leq b$ we define the (conic) closed *interval* $[a,b]$ as the set of the form

$$[a,b] = \{x \in V : a \leq x \leq b\}\ .$$

If the cone V_+ is solid, then any bounded set $B \subset V$ is contained in some interval. A cone V_+ is said to be *normal* if every interval $[a,b]$ is bounded.

An rds (ϑ,φ) taking values in a subset S of V is called *monotone* or *order-preserving*, if (possibly up to a universal set of measure zero) $x \leq y$ implies $\varphi(t,\omega)x \leq \varphi(t,\omega)y$ for all $t \in \mathbf{T}_+$ and $\omega \in \Omega$.

Theorem 2.1. *Let (ϑ,φ) be an ergodic and monotone rds taking values in a separable Banach space V with a solid and normal cone V_+. Then the rds has a weak attractor which consists of a single (random) point.*

If we denote the single point in the theorem by $v(\omega)$, then it follows from the invariance of the attractor that $t \mapsto \varphi(t,\omega)v(\omega)$ is stationary since $\varphi(t,\omega)v(\omega) = v(\vartheta(t)\omega)$. The law of v equals π (which appears in the definition of an ergodic rds).

In many applications one has an ergodic and monotone rds on a *subset* of an ordered Banach space. It is easy to construct examples which show that Theorem 2.1 does *not* hold if V is replaced by an arbitrary subset of V (see [7, Example 3]). It does hold however if φ takes values in an *admissible* subset of V, which, by definition, is a subset S of V whose induced topology admits a complete and separable metrization and which has the property that for any compact set K in S, there exist $a, b \in S$, $a \leq b$, such that K is a subset of the relative interior of $[a,b] \cap S$. The latter property implies in particular that each pair $x, y \in S$ is contained in an interval $[a,b]$ with endpoints $a, b \in S$. The condition that V be solid can be dropped in case the state space S is admissible.

We provide the *idea of the proof* of this result in the special case where $S \subseteq [a,b]$ and $a, b \in S$ (note that in this case S is automatically admissible): first one shows that $\|\varphi(t,\omega)a - \varphi(t,\omega)b\|$ converges to zero in probability (this is Proposition 1 in [7] which is rather easy to prove and which does not require the cone V_+ to be solid or normal). Next observe that the sets $\varphi(t,\vartheta(-t)\omega)(S)$ are decreasing in t. Due to the fact that by monotonicity $\varphi(t,\vartheta(-t)\omega)(S)$ is contained in $[\varphi(t,\vartheta(-t)\omega)a, \varphi(t,\vartheta(-t)\omega)b]$ and the norm of the difference of the endpoints of that interval converges to zero in probability (in fact even almost surely!), it is easy to see (using the fact that the cone V_+ is normal) that $\cap_{t\geq 0}\varphi(t,\vartheta(-t)\omega)(S)$ consists of a single point which we call $v(\omega)$. In particular, $\{v(\omega)\}$ attracts the whole space S almost surely. It is not hard to see that $\varphi(t,\omega)v(\omega) = v(\vartheta(t)\omega)$ holds almost surely, but the exceptional set might depend on t. By some appropriate *perfection* (see the Appendix of [7]) one can find a modification of v which is invariant (without exceptional sets depending on t) showing that $\mathcal{A}(\omega) = \{v(\omega)\}$ is a pullback attractor. The basic idea of the proof in the general case is similar but a bit more technical. The reader is referred to [7] for details.

The assumptions of the theorem do not guarantee that the rds has a pullback attractor (see [7, Example 2]). We point out, that another paper dealing with attractors for an ergodic rds (with independent increments) is [13], see also Section 4.

Our result can be applied to one-dimensional rds, to multi-dimensional rds with *cooperative* drift, to certain stochastic delay differential equations, certain parabolic spde's and some (monotone) interacting particle systems like the (dynamical) Ising model. The reader is referred to Section 4 of [7] for details.

3. A counterexample

In this section we will provide an example of an ergodic rds (ϑ,φ) on $\mathbb{R}^d$, $d \geq 2$, which does not have an attractor. In this example, φ is even a cocycle of homeomorphisms rather than just continuous maps and the one-point motion is a diffusion process with additive noise. Roughly speaking, the drift vector points towards the origin and is, on the one hand, strong enough to ensure ergodicity but, on the other hand, weak enough in order to allow the noise to prevent large balls from collapsing, i.e., even though each fixed initial point which is far out, tends towards the origin, there will exist random points (measurable with respect to the future) which will move out to ∞, thus contradicting the existence of a weak attractor. To the best of our knowledge such examples have not appeared in the mathematical literature so far.

In our example, the noise which drives the flow will have to be sufficiently non-degenerate. That is why we assume the noise to be an isotropic Brownian one.

In [20] we proved that there is a positive probability that the diameter of a compact and connected subset of $\mathbb{R}^d$, $d \geq 2$, consisting of more than one point will grow to ∞ as $t \to \infty$ (in fact with linear speed) under an isotropic Brownian flow, the probability being 1 in case the highest Lyapunov exponent of the flow is non-negative (see also [8]). Such an isotropic Brownian flow is not ergodic since its one-point motion is Brownian motion, so we need to add some appropriate drift in order to make it ergodic. If this drift is rather weak, then one may expect that it will not be able to fully compensate the linear expansion of a set under the isotropic flow. We will show below that this is true. We use similar techniques as in the proofs in [20] but some arguments are a bit different. We will rely on two lemmas from [20]: a support lemma and a lemma about the asymptotics of sequences of random variables which are stated below as Lemma 3.5 and Lemma 3.6.

We will first define the concept of an isotropic covariance tensor b, then we will introduce isotropic Brownian fields and finally flows generated by stochastic differential equations which are driven by an isotropic Brownian field.

Definition 3.1. Let $b = (b_{ij}(x))_{i,j=1,\dots,d}$ be a positive semidefinite real matrix for each $x \in \mathbb{R}^d$. We say that b is an *isotropic covariance tensor* or *matrix* if

(i) $x \mapsto b(x)$ is four times continuously differentiable
(ii) $b(0) = E_d$ (the identity matrix)
(iii) $x \mapsto b(x)$ is not constant
(iv) $b(x) = G^* b(Gx) G$ for all $x \in \mathbb{R}^d,\ G \in O(d)$.

(i) is a convenient and not too restrictive smoothness assumption, (ii) a normalization condition, (iii) is assumed to avoid rigid motions, and (iv) ensures that b is invariant under orthogonal transformations – justifying the term *isotropic*.

Definition 3.2. Let $b = (b_{ij}(x))_{i,j=1,\dots,d}$ be an isotropic covariance tensor. An $\mathbb{R}^d$-valued random field $M(t,x)$, $t \geq 0$, $x \in \mathbb{R}^d$, defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, is called an *isotropic Brownian field* with covariance tensor b, if

- $(t,x) \mapsto M(t,x)$ is a zero-mean Gaussian process
- $\mathrm{cov}(M(s,x), M(t,y)) = (s \wedge t)\, b(x-y)$
- $(t,x) \mapsto M(t,x)$ is continuous for almost all $\omega \in \Omega$.

From this definition it is easy to obtain the following properties of M.

Corollary 3.3. *Let M be an $\mathbb{R}^d$-valued isotropic Brownian field. Then the following holds:*

- $t \mapsto M(t,x)$ *is a d-dimensional standard Brownian motion for each $x \in \mathbb{R}^d$.*
- $\langle M(\cdot,x), M(\cdot,y)\rangle_t = b(x-y)\,t$ *for each $x, y \in \mathbb{R}^d$.*

Next, we consider the *Kunita-type* stochastic differential equation (sde)

$$\mathrm{d}X(t) = f(X(t))\mathrm{d}t + M(\mathrm{d}t, X(t)), \tag{3.1}$$

where M is an isotropic Brownian field and $f : \mathbb{R}^d \to \mathbb{R}^d$ satisfies a local Lipschitz and a linear growth condition. It was shown by Kunita ([14, Theorem 4.5.1]), that this equation does not only have a unique solution for every initial condition $X(0) = x \in \mathbb{R}^d$ but that it even generates a stochastic flow of homeomorphisms, i.e., that there exists a family $(\phi_{st})_{0\le s,t<\infty}$ of random homeomorphisms of $\mathbb{R}^d$ such that:

- $\phi_{su} = \phi_{tu} \circ \phi_{st}$ for all $0 \le s, t, u < \infty$ and all $\omega \in \Omega$.
- $\phi_{ss} = \mathrm{id}_{\mathbb{R}^d}$ for all $s \ge 0$ and all $\omega \in \Omega$.
- For each $s \ge 0$, $x \in \mathbb{R}^d$ $(\phi_{st}(x))_{t\ge s}$ solves (3.1) for $t \ge s$ with initial condition $X(s) = x$.
- The map $(s,t,x) \mapsto \phi_{st}(x)$ is continuous for all $\omega \in \Omega$.

In case $f \equiv 0$, this flow is called an *isotropic Brownian flow*. Isotropic Brownian flows have been investigated by a number of authors, e.g., [4] and [15]. Note that the one-point motion starting in x of an isotropic flow is a standard d-dimensional Brownian motion (starting in x).

It is shown in Arnold and Scheutzow ([2]), that the solutions of equation (3.1) even generate an rds (ϑ, φ) (on a suitable mds). The relationship between the cocycle φ and the flow ϕ is given by

$$\varphi(t, \vartheta(s)\omega) = \phi_{s,s+t}(\omega).$$

Note that this identity shows the equivalence of the flow property of ϕ and the cocycle property of φ.

Before we state our result about the non-existence of a weak attractor, we provide more information about the covariance tensor b introduced in Definition 3.1: there exist C^4 functions $B_L, B_N : [0,\infty) \to (-1,1]$, called the *longitudinal*, respectively *transversal* or *normal* correlation functions, satisfying

$$b(x) = (B_L(|x|) - B_N(|x|))\frac{xx^T}{|x|^2} + B_N(|x|)E_d, \qquad x \neq 0,$$

where E_d denotes the $d \times d$ identity matrix (see [4]). Further, $B_N(r), B_L(r) \neq 1$ for $r > 0$. From this representation, it is easy to see that the set of eigenvalues of the matrix

$$a(x) := \begin{pmatrix} E_d & b(x) \\ b(x) & E_d \end{pmatrix}$$

is given by $\{1 \pm B_N(|x|), 1 \pm B_L(|x|)\}$. Note that in particular, the set of eigenvalues of $a(x)$ is bounded and bounded away from zero for all x in a compact subset of $\mathbb{R}^d \setminus \{0\}$. If $(z_1, \ldots, z_d)$ is an orthogonal basis of $\mathbb{R}^d$ with $z_1 = x$ $(x \neq 0)$, then the family

$$\begin{pmatrix} z_i \\ \pm z_i \end{pmatrix}, i = 1, \ldots, d$$

is a basis of eigenvectors of $a(x)$.

We introduce the filtration $\mathcal{F}_t := \sigma(\varphi(s), 0 \leq s \leq t)$ for $t \geq 0$.

Proposition 3.4. *Let M be an isotropic Brownian field on $\mathbb{R}^d$, $d \geq 2$ and $f : \mathbb{R}^d \to \mathbb{R}^d$ a bounded, locally Lipschitz continuous function. If $\beta := \sup_{x \in \mathbb{R}^d} \|f(x)\|$ is sufficiently small, then the rds (ϑ, φ) generated by the sde*

$$\mathrm{d}Y(t) = f(Y(t))\mathrm{d}t + M(\mathrm{d}t, Y(t))$$

does not possess a weak attractor.

Proof. Let $\mathcal{D} \subset \mathbb{R}^d$ be a compact and connected subset of diameter Δ. We show that if $\beta > 0$ is sufficiently small, then for sufficiently large Δ we have

$$\mathbf{P}\left\{\lim_{n \to \infty} \mathrm{diam}(\varphi(n)(\mathcal{D})) = \infty\right\} > 0$$

which implies the non-existence of a weak attractor. Let

$$D_t := \mathrm{diam}(\varphi(t)(\mathcal{D})).$$

We will use the following lemma which is proved in [20], p. 2057.

Lemma 3.5. *Let $(\tilde{D}_n, \mathcal{F}_n)$, $n \geq 0$ be an adapted real-valued process satisfying*

$$\mathbf{E}(\tilde{D}_{n+1} - \tilde{D}_n \,|\, \mathcal{F}_n) \geq c_1 \mathbf{1}_{\{\tilde{D}_n \geq \alpha\}} \tag{3.2}$$

and

$$\mathbf{P}\left\{|\tilde{D}_{n+1} - \tilde{D}_n| \geq \lambda \,|\, \mathcal{F}_n\right\} \leq b_1 \mathrm{e}^{-b_2 \lambda} \quad \text{for all } \lambda > 0 \tag{3.3}$$

for some strictly positive constants α, c_1, b_1, b_2. Then for

$$\gamma = \gamma(c_1, b_1, b_2) = \frac{c_1 b_2^2}{2b_1 + b_2^2} \wedge \frac{b_2}{5}$$

and $\kappa \geq \xi \geq \alpha$,

$$\mathbf{P}\left\{\lim_{n \to \infty} \tilde{D}_n = \infty \text{ and } \tilde{D}_n \geq \xi \ \forall n \geq 0 \,|\, \mathcal{F}_0\right\} \geq 1 - \mathrm{e}^{-\gamma(\kappa - \xi)} \ \text{ on } \{\tilde{D}_0 \geq \kappa\} \tag{3.4}$$

and

$$\mathbf{P}\left\{\liminf_{n \to \infty} \frac{1}{n} \tilde{D}_n \geq c_1 \,|\, \limsup_{n \to \infty} \tilde{D}_n \geq \alpha\right\} = 1. \tag{3.5}$$

We define the sequence Z_n, $n \geq 1$ as follows: for $n \in \mathbb{N}_0$, select $\mathcal{F}_n$-measurable (random) points $x, \tilde{x} \in \varphi(n)(\mathcal{D})$ such that $\|x - \tilde{x}\| = D_n$. Define $Z_{n+1} := 0$ on the set $I_n := \{\omega : \inf\{D_k(\omega) : k \leq n\} < 1\}$. Otherwise, select $\mathcal{F}_n$-measurable points

$y, \tilde{y} \in \varphi(n)(\mathcal{D})$ such that $\|y - x\| = 1$ and $\|\tilde{y} - \tilde{x}\| = 1$ (such points exist since $\varphi(n)(\mathcal{D})$ is connected). For fixed n, define

$$\begin{aligned} A &:= \langle \phi_{n,n+1}(x), \frac{x-\tilde{x}}{\|x-\tilde{x}\|} \rangle, \qquad B := \langle \phi_{n,n+1}(y), \frac{x-\tilde{x}}{\|x-\tilde{x}\|} \rangle, \\ \tilde{A} &:= \langle \phi_{n,n+1}(\tilde{x}), \frac{x-\tilde{x}}{\|x-\tilde{x}\|} \rangle, \qquad \tilde{B} := \langle \phi_{n,n+1}(\tilde{y}), \frac{x-\tilde{x}}{\|x-\tilde{x}\|} \rangle, \\ Z_{n+1} &:= A \vee B - \tilde{A} \wedge \tilde{B} - D_n. \end{aligned}$$

Note that we have $D_{n+1} \geq D_n + Z_{n+1}$ on the complement of the set I_n.

We will show the validity of (3.2) and (3.3) with $\alpha = 1$ for the sequence

$$\tilde{D}_n := D_0 + \sum_{k=1}^{n} Z_k.$$

(3.3) follows easily since $Z_{n+1} = 0$ on I_n, and on the complement of I_n we have

$$-2\beta + N_1 + N_2 + N_3 + N_4 \leq Z_{n+1} \leq 2\beta + N_1 + N_2 + N_3 + N_4,$$

where the N_i are standard normal variables. Once we have also proved (3.2) it follows from (3.4) and the fact that $D_n \geq \tilde{D}_n$ for all n on the set $\{\inf_k \tilde{D}_k \geq 1\}$ that $\lim D_n = \infty$ with positive probability provided that $\Delta > 1$.

Now we verify (3.2) for $\alpha := 1$ provided β is sufficiently small. Clearly, (3.2) is satisfied on the set I_n, so we can assume that $\omega \notin I_n$ (and hence $D_n \geq 1$).

For any $\rho > 0$,

$$\begin{aligned} \mathbf{E}(A \vee B | \mathcal{F}_n) &= \mathbf{E}(A|\mathcal{F}_n) + \mathbf{E}((B-A)^+ | \mathcal{F}_n) \\ &\geq \langle x, \frac{x-\tilde{x}}{\|x-\tilde{x}\|} \rangle - \beta + \mathbf{E}((B-A)^+ | \mathcal{F}_n) \\ &\geq \langle x, \frac{x-\tilde{x}}{\|x-\tilde{x}\|} \rangle - \beta + \rho \mathbf{P}\{B - A \geq \rho | \mathcal{F}_n\}. \end{aligned}$$

We will estimate the last probability from below using the following support lemma which is proved in [20], p. 2053 (using Lemma I.8.3 of [3]). Observe that in order for $\mathbf{P}\{B - A \geq \rho | \mathcal{F}_n\}$ to be strictly positive we need the dimension d of the underlying space to be at least 2.

Lemma 3.6. *Fix a positive integer m. There exists a continuous function $g : (0,\infty)^5 \to (0,\infty)$ which is decreasing in the second, third, and fifth variable and increasing in the first and fourth variable with the following property: let $(S_t)_{t \geq 0}$ be any $\mathbb{R}^m$-valued continuous semimartingale, with Doob decomposition $S_t = N_t + V_t$, N_t being the local martingale part and V_t having locally bounded variation, and $N_0 = V_0 = 0$. Let $\varepsilon > 0$ be given, and define $\tau = \inf\{t : \|S_t\| > \varepsilon\}$. Let $\alpha_1, \alpha_2, \gamma_1$ be positive, such that V has Lipschitz constant no more than γ_1 on $[0, \tau]$ and such that the quadratic variation $a_t = \mathrm{d}\langle N \rangle / \mathrm{d}t$ satisfies*

$$\alpha_2 \|z\|^2 \geq z^T a_t z \geq \alpha_1 \|z\|^2 \tag{3.6}$$

for $0 \le t \le \tau$ *and* $z \in \mathbb{R}^m$. *Then* $\mathbf{P}\{\tau > t_0\} \ge g(\alpha_1, \alpha_2, \gamma_1, \varepsilon, t_0)$ *for every positive* t_0.

To apply the lemma, we pick a function $\psi : \{(x, \tilde{x}, y, t) \in \mathbb{R}^{3d} \times [0,1] : x \neq \tilde{x}, \|y - x\| = 1\} \to \mathbb{R}^d$ with the following properties (in the domain of definition):

- $\psi(x, \tilde{x}, y, 0) = y$
- $\|\psi(x, \tilde{x}, y, t) - \psi(x, \tilde{x}, y, s)\| \le 3|t - s|$
- $\|\psi(x, \tilde{x}, y, t) - x\| \ge 1/2$
- $\langle \psi(x, \tilde{x}, y, 1) - x, \frac{x - \tilde{x}}{\|x - \tilde{x}\|} \rangle \ge 1$.

For fixed n, we apply the lemma with $m = 2d$ to the process

$$S_t := \begin{pmatrix} \int_0^{t\wedge 1} M(\mathrm{d}s, \phi_{n,n+s}(x)) \\ y + \int_0^{t\wedge 1} M(\mathrm{d}s, \phi_{n,n+s}(y)) - \psi(x, \tilde{x}, y, t \wedge 1) \end{pmatrix}, \qquad t \ge 0.$$

S_t, $t \ge 0$ is a continuous $(\mathcal{F}_{n+t})$ −semimartingale with $S_0 = 0$. Fix $\varepsilon = 1/4$ and define τ as in the lemma. By definition of ψ, the bounded variation part of S has Lipschitz constant no more than $\gamma_1 = 3$. It remains to check condition (3.6). The matrix a_t in the lemma is given by

$$a_t = \begin{pmatrix} E_d & b\left(\phi_{n,n+t}(x) - \phi_{n,n+t}(y)\right) \\ b\left(\phi_{n,n+t}(x) - \phi_{n,n+t}(y)\right) & E_d \end{pmatrix}$$

for $0 \le t \le 1$. By definition of τ, the set $\{\|\phi_{n,n+t}(x) - \phi_{n,n+t}(y)\| : t \le \tau \wedge 1\}$ is bounded above by a deterministic constant and is bounded from below by $\|\psi(x, \tilde{x}, y, t \wedge \tau) - x\| - 2\beta - \sqrt{2}\|S_{t\wedge\tau}\| \ge \frac{1}{2} - 2\beta - \sqrt{2}/4 > .04$ provided that $\beta \le .05$.

Using the statement about the eigenvalues of a just before Proposition 3.4, we see that (3.6) is satisfied for some $\alpha_2 \ge \alpha_1 > 0$ which do not depend on the function f as long as $\beta \le .05$.

Therefore, we obtain

$$\mathbf{P}\left\{\sup_{0\le t\le 1} \|S_t\| \ge \frac{1}{4}\right\} \ge \mathbf{P}\{\tau > 1\} \ge g(\alpha_1, \alpha_2, 3, 1/4, 1).$$

Denoting the first d components of S by S^1 and the last d components by S^2 we see that on the set $\left\{\sup_{0\le t\le 1} \|S_t\| \ge \frac{1}{4}\right\}$, we have

$$\begin{aligned} B - A &= \langle \phi_{n,n+1}(y) - \phi_{n,n+1}(x), \frac{x - \tilde{x}}{\|x - \tilde{x}\|} \rangle \\ &= \langle S_1^2 + \psi(x, \tilde{x}, y, 1) + \int_0^1 f(\phi_{n,n+s}(y))\mathrm{d}s \\ &\qquad\qquad -S_1^1 - x - \int_0^1 f(\phi_{n,n+s}(x))\mathrm{d}s, \frac{x - \tilde{x}}{\|x - \tilde{x}\|} \rangle \\ &\ge \langle \psi(x, \tilde{x}, y, 1) - x, \frac{x - \tilde{x}}{\|x - \tilde{x}\|} \rangle - 2\beta - \frac{1}{4} - \frac{1}{4} \\ &\ge 1 - .1 - .5 = .4 \end{aligned}$$

provided $\beta \leq .05$. Hence, defining $\rho := .4$, we get

$$\rho\mathbf{P}\{B - A \geq \rho|\mathcal{F}_n\} \geq \rho g(\alpha_1, \alpha_2, 3, 1/4, 1).$$

An analogous estimate holds for $(\tilde{A}, \tilde{B})$ in place of (A, B). Therefore, we get on the complement of I_n (and hence in particular on the set$\{\tilde{D}_n \geq 1\}$)

$$\begin{aligned}
\mathbf{E}\left(\tilde{D}_{n+1} - \tilde{D}_n|\mathcal{F}_n\right) &= \mathbf{E}\left(Z_{n+1}|\mathcal{F}_n\right) \\
&= \mathbf{E}\left(A \vee B|\mathcal{F}_n\right) - \mathbf{E}\left(\tilde{A} \wedge \tilde{B}|\mathcal{F}_n\right) - D_n \\
&\geq -2\beta + 2\rho g(\alpha_1, \alpha_2, 3, 1/4, 1).
\end{aligned}$$

Therefore (3.2) is satisfied as long as $\beta < (\rho g(\alpha_1, \alpha_2, 3, 1/4, 1)) \wedge .05$ and hence the proof of the proposition is complete. $\square$

Observe that we did not use (3.5) in the proof of the proposition, but that it provides additional information on the (linear) growth of the image of a connected set under the rds. Note further that we did not make any assumptions about the sign of the largest element of the Lyapunov spectrum of the underlying isotropic flow or the Lyapunov spectrum associated to an invariant measure of the rds (ϑ, φ) (which can be positive or negative).

Corollary 3.7. *For each $d \geq 2$ and each isotropic Brownian field M on $\mathbb{R}^d$, there exists a bounded Lipschitz continuous function f such that the rds generated by equation* (3.1) *is ergodic and does not have a weak attractor.*

Proof. By Proposition 3.4, the rds generated by (3.1) has no weak attractor if $\beta = \sup_{x \in \mathbb{R}^d} \|f(x)\|$ is sufficiently small. On the other hand, it is well known that the rds is ergodic in case $\langle f(x), \frac{x}{\|x\|}\rangle < 0$ (see [5]) (to apply Bhattacharya's result, observe that the one-point motion of the rds is the same as that of the equation $\mathrm{d}X(t) = f(X(t))\mathrm{d}t + \mathrm{d}W(t)$, where W is a $d-$dimensional Brownian motion). Obviously, there exist functions f satisfying both properties, so the corollary is proved. $\square$

4. Ergodicity and point attractors

Apart from weak (set) attractors introduced in Definition 1.1, (weak) *point attractors* have been introduced by H. Crauel in [9].

Definition 4.1. Let $\mathcal{A}(\omega)$ be an invariant compact set of an rds (ϑ, φ). $\mathcal{A}$ is called a *(minimal weak) point attractor*, if

(i) for each $x \in S$,

$$\lim_{t\to\infty} d(\varphi(t, \vartheta(-t)\omega)x, \mathcal{A}(\omega)) = 0 \quad \text{in probability and}$$

(ii) for each compact invariant set $\widetilde{\mathcal{A}}(\omega)$ which satisfies (i), we have

$$\widetilde{\mathcal{A}}(\omega) \subseteq \mathcal{A}(\omega) \quad \text{almost surely.}$$

Clearly, a point attractor is unique (if it exists). It is easy to see that any weak (set) attractor contains a point attractor (which may or may not coincide with the set attractor). Kuksin and Shirikyan ([13]) provide sufficient conditions for the existence of a point attractor for an rds with independent increments (a so-called *white noise* rds). These conditions are stronger than ergodicity. They also show that under rather weak conditions, the support of the (unique) invariant measure (i.e., a probability measure on $\Omega \times S$ which is invariant under the *skew-product flow* associated to the rds (ϑ, φ)) coincides with the point attractor almost surely.

In general, neither ergodicity implies the existence of a point attractor nor vice versa. Clearly, the identity $\varphi = \mathrm{id}$ on a compact space S which has more than one element, has a point attractor (namely S) without being ergodic. Below, we sketch an example of a discrete-time rds which is ergodic without having a point attractor. The question whether an ergodic rds consisting of random *homeomorphism* rather that just continuous maps always has a point attractor seems to be open. We do not know if an ergodic rds which is generated by equation (3.1) always has a point attractor.

Example. Take $S = (0,1)$, $\mathbf{T} = \mathbb{Z}$, U, X and Y independent random variables such that U is uniformly distributed on $(0,1/2)$ and X and Y are uniformly distributed on $(0,1)$. Denote $V := X \wedge Y$ and $W := X \vee Y$. Define the random map $g : \Omega \times S \to S$ by

$$g(\omega, x) = \begin{cases} x & \text{if } x \le U \text{ or } x \ge 1-U \\ \frac{x}{V} + U(1 - \frac{1}{V}) & \text{if } U \le x \le U + V(1-2U) \\ -\frac{x}{W-V} + \frac{U(1-V-W)+W}{W-V} & \text{if } U + V(1-2U) \le x \le U + W(1-2U) \\ \frac{x}{1-W} + (U-1)\frac{W}{1-W} & \text{if } U + W(1-2U) \le x \le 1-U. \end{cases}$$

Clearly, the map g is continuous and preserves Lebesgue measure on (0,1) for every ω. Denote the law of g on $C(0,1)$ by Q. The iteration of independent copies of g defines an rds (ϑ, φ) as follows: $\Omega = (C(0,1))^{\mathbb{Z}}$, $(\vartheta\omega)_i = \omega_{i+1}$, $\mathbf{P} = Q^{\mathbb{Z}}$, $\varphi(n, \omega) = \omega_{n-1} \circ \cdots \circ \omega_1 \circ \omega_0$. It is easy to check that the rds is ergodic: since the corresponding one-point motion is a Markov chain with Lebesgue measure on (0,1) as invariant measure, all one needs to check is that the chain is irreducible (which is obvious). Assume that (ϑ, φ) has a point attractor $\mathcal{A}(\omega)$. Then for each $\varepsilon > 0$ there exists some $\delta > 0$ such that $\mathcal{A}(\omega)$ is contained in $[\delta, 1-\delta]$ with probability at least $1 - \varepsilon$.

For any finite subset $B \subset (0,1)$, there exists $n_0 > 0$ such that $\mathbf{P}\{\varphi(n)(B) \subseteq [\delta/2, 1-\delta/2]\} \ge 1 - \varepsilon$ for all $n \ge n_0$. Now we choose a particular finite (random) subset: on a probability space $(\tilde{\Omega}, \tilde{F}, \tilde{\mathbf{P}})$ let $X_1, X_2, \ldots$ be independent and uniformly distributed on (0,1). Let N be so large that $\tilde{\mathbf{P}}(\{X_1, \ldots, X_N\} \nsubseteq [\delta/2, 1 - \delta/2]) \ge 1 - \varepsilon$. On the product space $(\Omega \times \tilde{\Omega}, \mathcal{F} \otimes \tilde{\mathcal{F}}, \mathbf{P} \otimes \tilde{\mathbf{P}})$ define $A_n := \{(\omega, \tilde{\omega}) : \varphi(n, \omega)B(\tilde{\omega}) \nsubseteq [\delta/2, 1-\delta/2]\}$, where $B(\tilde{\omega}) := \{X_1, \ldots, X_N\}$. Since φ preserves Lebesgue measure, we obtain $\mathbf{P} \otimes \tilde{\mathbf{P}}(A_n) \ge 1 - \varepsilon$ for each $n \in \mathbb{N}$. On the other hand, we know that $\limsup_{n\to\infty} \mathbf{P}(A_n(.,\tilde{\omega})) \le \varepsilon$ for almost every $\tilde{\omega}$. Using Fubini's theorem, we get a contradiction in case $\varepsilon < 1/2$, so we found an rds (ϑ, φ) which is ergodic without having a point attractor.

5. Attractors and coupling from the past

Starting with the seminal paper of Propp and Wilson [18], the method of *coupling from the past* became popular to perfectly simulate the invariant probability measure of an ergodic discrete time Markov chain with a large but finite state space. There is nothing new we can add to this topic but we believe that it is worthwhile to highlight the relationship between coupling from the past and the existence of weak random attractors consisting of a single point (we learnt about this relationship from Franco Flandoli, Pisa). Let us consider an ergodic rds (ϑ, φ) with compact metric state space (S, d) and let π be the probability measure on S in the definition of an ergodic rds. If (ϑ, φ) admits a weak attractor which consists of a singleton $\mathcal{A}(\omega) = \{v(\omega)\}$, then the Hausdorff distance between $\varphi(t, \vartheta(-t)\omega)(S)$ and $\mathcal{A}(\omega)$ converges to zero in probability (in fact even almost surely, since the concepts of a weak and a pullback attractor coincide when the state space is compact). If, moreover, S is finite, then there exists some $t_0(\omega)$ such that $\varphi(t, \vartheta(-t)\omega)(S) = \mathcal{A}(\omega)$ for all $t \ge t_0(\omega)$ almost surely. So, if $0 \le t_1 \le t_2 \cdots$ is any sequence (possibly random) such that $\lim_{i\to\infty} t_i = \infty$ almost surely, then $T := \inf\{i \ge 0 : \varphi(t_i, \vartheta(-t_i)\omega)(S)$ consists of a single point$\}$ is finite almost surely and the single point coincides with $v(\omega)$. This means that the output of the *Propp–Wilson algorithm* is the point $v(\omega)$, which has law π. Theorem 2.1 implies that the Propp–Wilson algorithm terminates in particular for any monotone and ergodic rds with a finite state space, no matter whether the single-point motion is Markovian or not.

Observe that for a discrete-time ergodic Markov chain with finite state space S and transition probabilities p_{ij}, $i, j \in S$, it is always possible to find a random map $g : \Omega \times S \to S$ such that $\mathbf{P}\{g(\omega, i) = j\} = p_{ij}$ for all $i, j \in S$ and such that if $g_1, g_2, \dots$ are i.i.d. copies of g, then $\mathbf{P}\{g_n \circ g_{n-1} \circ \cdots \circ g_1(S)$ consists of a single point$\} \to 1$ as $n \to \infty$. In this case one can define the associated canonical rds as follows: $\Omega = (S^S)^{\mathbb{Z}}$, $(\vartheta\omega)_i = \omega_{i+1}$, $Q = \mathcal{L}(g)$, $\mathbf{P} = Q^{\mathbb{Z}}$, $\varphi(n, \omega) = \omega_{n-1} \circ \cdots \circ \omega_1 \circ \omega_0$. (ϑ, φ) is an rds whose one-point motion is the original Markov chain. Due to our construction of g, the rds has a (weak and pullback) attractor which is a singleton and therefore the Propp–Wilson algorithm terminates.

The well-known warning, that coupling from the past does not work if it is replaced by *coupling in the future* (see e.g. [11]) can be formulated in the language of rds and attractors as well: if $0 < t_1 < t_2 < \cdots$ is any sequence of stopping times and $T := \inf\{i \ge 0 : \varphi(t_i, \omega)(S)$ consists of a single point$\}$, then this single point coincides with $\varphi(T, \omega)(v(\omega)) = v(\vartheta(T)\omega)$ which does not always have law π (it does in case T is deterministic).

References

[1] L. Arnold, *Random Dynamical Systems*, Springer, 1998.

[2] L. Arnold and M. Scheutzow, *Perfect cocycles through stochastic differential equations*, Probab. Theory Relat. Fields, **101 (1)** (1995), 65–88.

[3] R. Bass, *Diffusions and Elliptic Operators*, Springer, 1998.

[4] P. Baxendale and T. Harris, *Isotropic stochastic flows*, Ann. Probab., **14 (4)** (1986), 1155–1179.

[5] R. N. Bhattacharya, *Criteria for recurrence and existence of invariant measures for multidimensional diffusions*, Ann. Probab., **6 (2)** (1978), 541–553.

[6] I. Chueshov, *Monotone Random Systems: Theory and Applications*, Springer, 2002.

[7] I. Chueshov and M. Scheutzow, *On the structure of attractors and invariant measures for a class of monotone random systems*, Dynamical Systems: An International Journal, **19 (2)** (2004), 127–144.

[8] M. Cranston, M. Scheutzow, and D. Steinsaltz, *Linear expansion of isotropic Brownian flows*, Electron. Comm. Probab., **4** (1999), 91–101.

[9] H. Crauel, *Random point attractors versus random set attractors*, J. London Math. Soc., **63 (2)** (2001), 413–427.

[10] H. Crauel and F. Flandoli, *Attractors for random dynamical systems*, Probab. Theory Relat. Fields, **100 (3)** (1994), 365–393.

[11] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, 2002.

[12] M. A. Krasnoselskii, *Positive Solutions of Operator Equations*, Noordhoff, 1964.

[13] S. B. Kuksin and A. Shirikyan, *On random attractors for mixing type systems*, Functional Anal. and its Appl., **38 (1)** (2004), 28–37.

[14] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, 1990.

[15] Y. Le Jan, *On isotropic Brownian motions*, Z. Wahrscheinlichkeitstheorie verw. Geb., **70 (4)** (1985), 609–620.

[16] H. Lisei and M. Scheutzow, *On the dispersion of sets under the action of an isotropic Brownian Flow*, in: I. Davies et. al., Editors, Proceedings of the Swansea 2002 Workshop Probabilistic Methods in Fluids, World Scientific, 2003, 224–238.

[17] G. Ochs, *Weak Random Attractors*, Institut für Dynamische Systeme, Universität Bremen, Report 449, 1999.

[18] J. Propp and D. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms, **9 (1-2)** (1996), 232–252.

[19] M. Scheutzow, *Comparison of various concepts of a random attractor: A case study*, Arch. Math., **78 (3)** (2002), 233–240.

[20] M. Scheutzow and D. Steinsaltz, *Chasing balls through martingale fields*, Ann. Probab., **30 (4)** (2002), 2046–2080.

[21] B. Schmalfuß, *Backward cocycles and attractors for stochastic differential equations*, in: V. Reitmann, T. Riedrich, and N. Koksch, Editors, International Seminar on Applied Mathematics – Nonlinear Dynamics: Attractor Approximation and Global Behaviour, Teubner, (1992), 185–192.

Michael Scheutzow
Institut für Mathematik, MA 7-5
Technische Universität Berlin
Str. des 17. Juni 136
D 10623 Berlin
Germany
e-mail: `ms@math.tu-berlin.de`

Progress in Probability, Vol. 59, 345–362

On the Stability of Feynman-Kac Propagators

Wilhelm Stannat

Abstract. The stability of a non-homogeneous measure-valued evolution equation is studied using a variational approach. We apply our results in particular to stochastic optimization algorithms and to the pathwise filter equation. In the latter example the variational approach leads to a new interpretation of the rate of stability.

Mathematics Subject Classification (2000). Primary 35J20; Secondary 60E15, 60G35, 92D15, 93E11.

Keywords. Stability, functional inequalities, ground state transform, stochastic optimization, stochastic filtering.

1. Introduction

This paper is a study of stability properties of measure-valued evolution equations of Feynman-Kac type using a variational approach. In the context of genetic algorithms these equations can be interpreted as equations describing approximately the empirical distribution of types within a population of individuals undergoing time-dependent mutation and selection. This way, the paper is a continuation of the variational approach introduced in [13] to study the long-time behaviour of genetic algorithms. Let us first introduce the algorithm we are interested in. Let $(S, \mathcal{S})$ be an arbitrary measurable space (called the *type space*) and a measurable time-dependent fitness function $\sigma_t : S \to \mathbb{R}$ specifying the fitness $\sigma_t(x)$ at time t of an individual of type x. Furthermore, let $\mathbb{M} = ((X_t), (P_{t,x}))$ be a time-inhomogeneous Markov process on S modelling random mutation of the type of an individual. Let A_t be the generator of $\mathbb{M}$ at time t and denote by $\hat{A}_t$ its dual operator acting on the space of probability measures $\mathcal{M}_1(S)$ over the type space S. We are interested in the long-time behaviour of the (nonlinear) flow (ψ_t) of distribution of types determined by

$$\dot{\psi}_t(\mu) = \hat{A}_t\psi_t(\mu) + \sigma_t\psi_t(\mu) - \langle \sigma_t, \psi_t(\mu)\rangle\psi_t(\mu)\,, \psi_0(\mu) = \mu\,. \tag{1.1}$$

Here we use the notation $\langle f, \eta\rangle := \int f\, d\eta$ for any bounded measurable f and $\eta \in \mathcal{M}_1(S)$. Equation (1.1) governs the empirical distribution of types in a given population with time-dependent mutation A_t and time-dependent selection σ_t in the limit of a large number of individuals (see [4] and [1] for the time-discrete case). In the context of stochastic filtering theory, an equation of type (1.1) describes the conditional distribution of a signal that is modelled as a Markov process in continuous time (see [3]). Using the Feynman-Kac propagator

$$p^{\sigma.}_{s,t} f(x) := E_{s,x}[f(X_t) \exp(\int_s^t \sigma_r(X_r)\, dr)]\,, \quad 0 \le s \le t\,,$$

associated with $\mathbb{M}$ and $\sigma.$, it is easy to see that a solution to (1.1) can be written as

$$\langle f, \psi_t(\mu)\rangle = \frac{\langle p^{\sigma.}_{0,t} f, \mu\rangle}{\langle p^{\sigma.}_{0,t} 1, \mu\rangle}\,.$$

To introduce our variational approach, assume that there exists a "reference" measure $\nu \in \mathcal{M}_1(S)$, that is, a probability measure on the type space S, such that the propagator $(p_{s,t})_{0\le s\le t}$ associated with the time-inhomogeneous Markov process respects ν-classes, that is, $f = 0$ ν-a.e. implies $p_{s,t} f = 0$ ν-a.e. for all $0 \le s \le t$. Then, starting with an absolutely continuous measure μ, it follows that the solution $\psi_t(\mu)$ of (1.1) will be again absolutely continuous w.r.t. ν. Its density $\tilde{h}_t$ satisfies in a weak sense the equation

$$\dot{\tilde{h}}_t = (\hat{A}_t + \sigma_t)\tilde{h}_t - \int \sigma_t \tilde{h}_t\, d\nu \cdot \tilde{h}_t\,. \tag{1.2}$$

In the next step let us introduce a "time-dependent" or parabolic ground state transform to equation (1.2): let $\hat{m}_t$ be a nonnegative ν-integrable solution to the equation

$$\frac{d}{dt}\hat{m}_t = (\hat{A}_t + \sigma_t)\hat{m}_t + \hat{\lambda}_t \hat{m}_t\,, \quad t \ge 0\,, \tag{1.3}$$

for some suitable function $\hat{\lambda}. : [0,T] \to \mathbb{R}$. It then follows that the density h_t of $\psi_t(\mu)$ w.r.t. the new reference measure $\hat{m}_t\, d\nu$ satisfies the equation

$$\dot{h}_t = \hat{A}^*_t h_t - (\hat{\lambda}_t + \int \sigma_t h_t\, \hat{m}_t\, d\nu) h_t\,, \tag{1.4}$$

where now

$$\hat{A}^*_t f := \frac{1}{\hat{m}_t}(\hat{A}_t(f\hat{m}_t) - \hat{A}_t\hat{m}_t\, f)\,.$$

We are interested in the stability of the solution h_t in suitable L^2-spaces. To this end let m_t, $t \in [0,T]$, be a nonnegative solution of the backward equation

$$\frac{d}{dt} m_t = -(A_t + \sigma_t) m_t - \lambda_t m_t\,, \quad t \in [0,T]\,, \tag{1.5}$$

for some suitable function $\lambda. : [0,T] \to \mathbb{R}$. Define the measures $d\nu^*_t := \frac{m_t \hat{m}_t\, d\nu}{\int m_t \hat{m}_t\, d\nu}$. It follows that ν^*_t, $t \in [0,T]$, satisfies the Fokker-Planck equation

$$\frac{d}{dt}\int f\, d\nu^*_t = -\int \hat{A}^*_t f\, d\nu^*_t$$

for suitable test-functions f (see Lemma 2.4). To study stability of equation (1.4) in $L^2(\nu_t^*)$ (instead of $L^1(\hat{m}_t\nu)$), we first consider the linear equation

$$\dot{h}_t = \hat{A}_t^* h_t \,. \tag{1.6}$$

It is easy to see that if $h.$ is a nonnegative solution of (1.6), then

$$\eta_t = \frac{h_t \hat{m}_t \, d\nu}{\int h_t \hat{m}_t \, d\nu}$$

is a solution of (1.1) up to time T (see Proposition 2.2).

We emphasize that equation (1.6) is much easier to study than the original problem. In particular, without further complications, it allows one to consider unbounded selection in contrast to the original approach presented in [5] and [3] which is based on the Feynman-Kac propagator. As the main abstract result of the paper we will show in Theorem 2.3 below that if $\frac{1}{2}\int \Gamma(\hat{A}_t^*)(f,f)\, d\nu_t^*$ satisfies a Poincaré inequality in $L^2(\nu_t^*)$ with constant less than $\kappa_*^{-1}(t)$, it follows that

$$\|h_{1,t} - h_{2,t}\|^2_{L^2(\nu_t^*)} \le e^{-2\int_0^t \kappa_*(s)\, ds} \|h_{1,0} - h_{2,0}\|^2_{L^2(\nu_0^*)} \,, \quad t \in [0,T] \,,$$

for two solutions $h_{i,\cdot}$, $i = 1,2$, of (1.6) with $\int h_{1,0}\, d\nu_0^* = \int h_{2,0}\, d\nu_0^*$ (see Theorem 2.3). Here, $\Gamma(\hat{A}_t^*)$ denotes the carrée du champ operator associated with $\hat{A}_t^*$. As a Corollary, this result implies stability of the solution (ψ_t) to (1.1) for suitable initial conditions in the total variation norm with the same rate (see Corollary 2.5). Consequently, the understanding of the long-time behaviour of (1.1) is reduced with the help of a parabolic ground state transformation to the study of the Poincaré inequality of the quadratic form $\frac{1}{2}\int \Gamma(\hat{A}_t^*)(f,f)\, d\nu_t^*$ (or equivalently to the study of the mass gap of the associated generator).

The real use of Theorem 2.3 and Corollary 2.5 however can only be seen in concrete applications. To this end we will consider in Section 3 two particular examples: The simulated annealing algorithm and the pathwise filter equation in the linear filtering problem. The main feature of the last example will be that the rate of the exponential convergence of ψ_t^Y will be *independent* of the observation Y (see Theorem 3.7). This result complements and partly strengthens Theorem 2.6 in [12] (see Remark 3.8 for a precise comparison). But what is more important is that, due to our approach, the rate of convergence can now be interpreted as the mass gap of the ground state transform associated with the generator of the signal process and the square of the observation function (see Remark 3.4). The analysis relies on the fact that for suitable initial (resp. terminal) conditions, the solution $\hat{m}_t$ of (1.3) and m_t of (1.5) can be calculated explicitely. This is, of course, not possible in the general nonlinear case. Nevertheless, the method can be generalized to the nonlinear case by using the concept of log-concavity to find estimates on $\kappa_*(t)$ (see [14]).

Finally, we would like to mention that the book [3] by P. Del Moral contains many results on stability of Feynman-Kac propagators. In contrast to the present approach, the contraction properties in the book by P. Del Moral are based on the Dobrushin contraction coefficient. It follows that the corresponding stability results cannot be applied in general to random mutation on noncompact type spaces. Typically, a lower bound on the contraction coefficient that is strictly less than 1 can be obtained only locally but not globally.

2. Stability via the variational approach

Let $(S, \mathcal{S})$ be an arbitrary measurable space, $\nu \in \mathcal{M}_1(S)$ a probability measure and A_t, $t \geq 0$, a family of linear operators with dense domain $D(A_t) \subset L^2(\nu)$ and let $\hat{A}_t$ with domain $D(\hat{A}_t)$ be the adjoint operator of A_t in $L^2(\nu)$. We assume that there exists an algebra D of *test functions*, containing the constant functions, such that $D \subset D(A_t) \cap D(\hat{A}_t)$ for all t. We also assume that $A_t 1 = 0$ for all t. We emphasize that $D \subset L^p(\nu)$ for all finite p since D is an algebra. Denote by

$$\Gamma(A_t)(f, g) := A_t(fg) - A_t f\, g - f\, A_t g\,, \quad f, g \in D\,,$$

the carré du champ operator associated with A_t and write $\Gamma(A_t)(f) := \Gamma(A_t)(f, f)$.

Let $\sigma_\cdot : [0, \infty[\to L^4(\nu)$ be measurable and locally bounded in $L^4(\nu)$.

Given $h \in D$, $h > 0$, we define the linear operator $\hat{A}_t^*$ by

$$\hat{A}_t^* f := \frac{1}{h}(\hat{A}_t(fh) - \hat{A}_t h\, f)\,, \quad f \in D\,.$$

Note that if h is a ground state of $\hat{A}_t + \sigma_t$, that is, h is an eigenvector of $-(\hat{A}_t + \sigma_t)$ corresponding to the eigenvalue $-\lambda_*$ given by the infimum of the spectrum of $-(\hat{A}_t + \sigma_t)$, it follows that

$$\hat{A}_t^* f = \frac{1}{h}(\hat{A}_t + \sigma_t - \lambda_*)(hf)\,,$$

so that in this case $\hat{A}_t^*$ coincides with the ground state transform associated with $\hat{A}_t$ and σ_t. We will use the same terminology in the general case too and say that $\hat{A}_t^*$ is the ground state transform associated with $\hat{A}_t$ and h.

For fixed $T > 0$ let $\hat{m}_\cdot : [0, T] \to D_+$ be a strictly positive solution of the forward equation

$$\frac{d}{dt}\hat{m}_t = (\hat{A}_t + \sigma_t)\hat{m}_t + \hat{\lambda}_t m_t \tag{2.1}$$

and $m_\cdot : [0, T] \to D_+$ be a strictly positive solution of the backward equation

$$\frac{d}{dt} m_t = -(A_t + \sigma_t) m_t - \lambda_t m_t \tag{2.2}$$

for locally bounded $\lambda_\cdot$, $\hat{\lambda}_\cdot : [0, T] \to \mathbb{R}$. Here we assume that the derivatives exist in $L^4(\nu)$. Define the measures

$$d\nu_t^* := \frac{m_t \hat{m}_t\, d\nu}{\int m_t \hat{m}_t\, d\nu}\,.$$

In the following let

$$\hat{A}_t^* f := \frac{1}{\hat{m}_t}(\hat{A}_t(\hat{m}_t f) - \hat{A}_t \hat{m}_t\, f)\,, \quad f \in D\,,$$

be the ground state transform associated with $\hat{A}_t$ and $\hat{m}_t$, $t \in [0,T]$. We are then interested in nonnegative solutions of the linear equation

$$\dot{h}_t = \hat{A}_t^* h_t\,, \quad t \in]0,T]\,. \tag{2.3}$$

Definition 2.1. (i) A function $h. \in C([0,T]; L^4(\nu)_+)$ with $h_t \in D$ for all $t \in]0,T]$ is called an *admissible solution (up to time T)* of (2.3)), if $\dot{h}_t$ exists in $L^4(\nu)$ and satisfies (2.3) for all $t \in]0,T]$.

(ii) A function $h \in L^4(\nu)_+$ is called an *admissible initial condition of* (2.3), if there exists an admissible solution h_t, $t \in [0,T]$, of (2.3) with $h_0 = h$.

Proposition 2.2. *Let $h.$ be an admissible solution of* (2.3) *with $h_t \neq 0$ for all $t \in [0,T]$. Then*

$$\eta_t := \frac{h_t \hat{m}_t \, d\nu}{\int h_t \hat{m}_t \, d\nu}\,, \quad t \in [0,T]\,,$$

is a solution of equation (1.1) *up to time T in the sense that*

$$\frac{d}{dt}\int f\, d\eta_t = \int A_t f + \sigma_t f \, d\eta_t - \int \sigma_t \, d\eta_t \cdot \int f \, d\eta_t\,, \quad f \in D\,, \;\; t \in]0,T]\,.$$

Proof. Note that for $f \in D$,

$$\begin{aligned}
\frac{d}{dt}\int f h_t \hat{m}_t \, d\nu &= \int f \dot{h}_t \, \hat{m}_t \, d\nu + \int f h_t \dot{\hat{m}}_t \, d\nu \\
&= \int f(\hat{A}_t^* h_t)\hat{m}_t \, d\nu + \int f h_t(\hat{A}_t \hat{m}_t + \sigma_t \hat{m}_t + \hat{\lambda}_t \hat{m}_t)\, d\nu \\
&= \int f \hat{A}_t(h_t \hat{m}_t) + \sigma_t f h_t \hat{m}_t \, d\nu + \hat{\lambda}_t \int f h_t \hat{m}_t \, d\nu \\
&= \int (A_t f + \sigma_t f) h_t \hat{m}_t \, d\nu + \hat{\lambda}_t \int f h_t \hat{m}_t \, d\nu\,.
\end{aligned}$$

In particular,

$$\frac{d}{dt}\int h_t \hat{m}_t \, d\nu = \int \sigma_t h_t \hat{m}_t \, d\nu + \hat{\lambda}_t \int h_t \hat{m}_t \, d\nu$$

so that

$$\frac{d}{dt}\int f\, d\eta_t = \int A_t f + \sigma_t f \, d\eta_t - \int \sigma_t \, d\eta_t \cdot \int f \, d\eta_t\,, t \in]0,T]\,,$$

which implies the assertion. □

Theorem 2.3. *Let $T > 0$ and m_t, $\hat{m}_t$, $t \in [0,T]$, be as in* (2.2) *and* (2.1). *Let $h_{i,\cdot}$, $i = 1,2$, be admissible solutions to* (2.3) *with $\int h_{1,0}\, d\nu_0^* = \int h_{2,0}\, d\nu_0^*$. If*

$$\mathcal{A}_t^*(f) := \frac{1}{2}\int \Gamma(\hat{A}_t^*)(f)\, d\nu_t^*\,, \quad f \in D\,,$$

satisfies a Poincaré inequality with constant less than $\frac{1}{\kappa_*(t)}$ in $L^2(\nu_t^*)$, then

$$\|h_{1,t} - h_{2,t}\|^2_{L^2(\nu_t^*)} \le e^{-2\int_0^t \kappa_*(s)\, ds}\|h_{1,0} - h_{2,0}\|^2_{L^2(\nu_0^*)}\,, \quad t \in [0,T]\,.$$

For the proof of Theorem 2.3 we need the following

Lemma 2.4. *Let $f \in D$ and $t \in [0,T]$. Then*

$$\frac{d}{dt}\int f\, d\nu_t^* = -\int \hat{A}_t^* f\, d\nu_t^*\,.$$

Proof. Using (2.1) and (2.2) we obtain that

$$\begin{aligned}\frac{d}{dt}\int f\, m_t\hat{m}_t\, d\nu &= -\int f A_t m_t\, \hat{m}_t\, d\nu + \int f m_t \hat{A}_t \hat{m}_t\, d\nu \\ &\quad + (\hat{\lambda}_t - \lambda_t)\int f m_t \hat{m}_t\, d\nu \\ &= -\int \frac{1}{\hat{m}_t}\hat{A}_t(f\hat{m}_t)\, m_t\hat{m}_t\, d\nu + \int \frac{\hat{A}_t\hat{m}_t}{\hat{m}_t} f\, m_t\hat{m}_t\, d\nu \\ &\quad + (\hat{\lambda}_t - \lambda_t)\int f\, m_t\hat{m}_t\, d\nu \\ &= -\int \hat{A}_t^* f\, m_t\hat{m}_t\, d\nu + (\hat{\lambda}_t - \lambda_t)\int f\, m_t\hat{m}_t\, d\nu\,.\end{aligned}$$

In particular,

$$\frac{d}{dt}\int m_t\hat{m}_t\, d\nu = (\hat{\lambda}_t - \lambda_t)\int m_t\hat{m}_t\, d\nu$$

so that

$$\frac{d}{dt}\int f\, d\nu_t^* = -\int \hat{A}_t^* f\, d\nu_t^*\,. \qquad \square$$

Proof of Theorem 2.3. First note that $\frac{d}{dt}\int h_{i,t}\, d\nu_t^* = 0$ by Lemma 2.4, hence $t \mapsto \int h_{i,t}\, d\nu_t^*$ is constant, and thus $\int h_{1,t}\, d\nu_t^* = \int h_{2,t}\, d\nu_t^*$ for all $t \in [0,T]$. Consequently, $\mathrm{Var}_{\,\nu_t^*}(h_{1,t} - h_{2,t}) = \|h_{1,t} - h_{2,t}\|^2_{L^2(\nu_t^*)}$ for all $t \in [0,T]$. It follows that

$$\begin{aligned}\frac{d}{dt}\|h_{1,t} - h_{2,t}\|^2_{L^2(\nu_t^*)} &= -\int \hat{A}_t^*\left((h_{1,t} - h_{2,t})^2\right)\, d\nu_t^* \\ &\quad + 2\int \hat{A}_t^*(h_{1,t} - h_{2,t})(h_{1,t} - h_{2,t})\, d\nu_t^* \\ &= -\int \Gamma(\hat{A}_t^*)(h_{1,t} - h_{2,t})\, d\nu_t^* \\ &\le -2\kappa_*(t)\ \mathrm{Var}_{\,\nu_t^*}(h_{1,t} - h_{2,t}) = -2\kappa_*(t)\|h_{1,t} - h_{2,t}\|^2_{L^2(\nu_t^*)}\,.\end{aligned}$$

Integrating the last inequality we obtain the assertion. $\square$

Theorem 2.3 now implies the following abstract result on the stability of the genetic algorithm (1.1).

Corollary 2.5. *Let $d\mu_i = \tilde{h}_{i,0}\, \hat{m}_0 d\nu \in \mathcal{M}_1(S)$, $i = 1, 2$, be such that $\tilde{h}_{i,0}$ are admissible initial conditions of* (2.3)*. Let $\psi_t(\mu_i)$, $t \in [0,T]$, be the associated solution of* (1.1)*. Then*

$$\|\psi_t(\mu_1) - \psi_t(\mu_2)\|_{var} \le c_{2.5}(t) e^{-\int_0^t \kappa_*(s)\, ds}\|h_{1,0} - h_{2,0}\|_{L^2(\nu_0^*)}$$

for all $t \in [0,T]$. Here

$$c_{2.5}(t) := \frac{\sqrt{\int m_t^{-1} \hat{m}_t \, d\nu \int m_t \hat{m}_t \, d\nu}}{\int \hat{m}_t \, d\nu}.$$

In particular, if $m_t = 1$, then

$$\|\psi_t(\mu_1) - \psi_t(\mu_2)\|_{var} \leq e^{-\int_0^t \kappa_*(s)\,ds} \|h_{1,0} - h_{2,0}\|_{L^2(\nu_0^*)}.$$

Proof. If $f, g \in L^1(\nu)$, $f, g \geq 0$, and $\mu_f := \frac{f\nu}{\int f\,d\nu}$, $\mu_g := \frac{g\nu}{\int g\,d\nu}$, it follows that

$$\|\mu_f - \mu_g\|_{var} \leq \frac{1}{\int g\,d\nu} \|f - g\|_{L^1(\nu)}$$

since

$$\begin{aligned}
\|\mu_f - \mu_g\|_{var} &= \frac{1}{2} \int \left| \frac{f}{\int f\,d\nu} - \frac{g}{\int g\,d\nu} \right| d\nu \\
&\leq \frac{1}{2 \int f\,d\nu \int g\,d\nu} \int \int |f(x)g(y) - f(y)g(x)| \, \nu(dx)\, \nu(dy) \\
&\leq \frac{1}{\int g\,d\nu} \|f - g\|_{L^1(\nu)}.
\end{aligned}$$

Let $h_{i,t}$ be the admissible solution of the linear equation (2.3) with initial condition $h_{i,0}$ and let $\nu_t := \frac{\hat{m}_t \nu}{\int \hat{m}_t\,d\nu}$. Then Theorem 2.3 implies that

$$\begin{aligned}
\|\psi_t(\mu_1) - \psi_t(\mu_2)\|_{var} &\leq \frac{1}{\int \hat{m}_t\,d\nu} \|h_{1,t}\hat{m}_t - h_{2,t}\hat{m}_t\|_{L^1(\nu)} \\
&\leq \frac{\sqrt{\int m_t^{-1} \hat{m}_t \, d\nu \int m_t \hat{m}_t \, d\nu}}{\int \hat{m}_t \, d\nu} \|h_{1,t} - h_{2,t}\|_{L^2(\nu_t^*)} \qquad (2.4) \\
&\leq e^{-\int_0^t \kappa_*(s)\,ds} \|h_{1,0} - h_{2,0}\|_{L^2(\nu_0^*)}.
\end{aligned}$$

□

Remark 2.6. The last corollary reduces the problem of (exponential) stability of the genetic algorithm for suitable initial conditions to the problem of estimating the constant $\kappa_*(t)$. One way to obtain estimates on $\kappa_*(t)$ is to find upper and lower bounds on the solutions m_t and $\hat{m}_t$. Indeed, suppose that the quadratic form $\frac{1}{2}\int \Gamma(\hat{A}_t^*)(f)\,d\nu$ satisfies a Poincaré inequality in $L^2(\nu)$ with constant less than κ^{-1}, then it is easy to see that

$$\kappa_*(t) \geq \frac{\kappa}{\|m_t \hat{m}_t\|_\infty \|m_t^{-1} \hat{m}_t^{-1}\|_\infty}.$$

We will obtain such upper and lower bounds in the particular example of the simulated annealing algorithm in Subsection 3.1 below.

Let us discuss the particular *stationary case* in more detail: Suppose that $A_t = A$ and $\sigma_t = \sigma$ do not depend on t. Let $m \in D(A)$ be a strictly positive solution of $(A + \sigma)m = -\lambda m$ and $\hat{m} \in D(\hat{A})$ be a strictly positive solution of

$(\hat{A}+\sigma)\hat{m} = -\hat{\lambda}\hat{m}$. Assume that $m, \hat{m} \in D$ and that $\int m^2\, d\nu = \int \hat{m}^2\, d\nu = 1$. Theorem 2.3 and Corollary 2.5 then imply the following

Corollary 2.7. *Let $T > 0$ and $d\nu^* := \frac{m\hat{m}\, d\nu}{\int m\hat{m}\, d\nu}$. Assume that*

$$\mathcal{A}^*(f) := \frac{1}{2}\int \Gamma(\hat{A}^*)(f)\, d\nu^*$$

satisfies a Poincaré inequality in $L^2(\nu^*)$ with constant $\frac{1}{\kappa_*}$. If $h.$ is an admissible solution of (2.3) up to time T with $\int h_0\, d\nu^* = 1$, it follows that

$$\|h_t - 1\|^2_{L^2(\nu^*)} \le e^{-2\kappa_* t}\|h_0 - 1\|^2_{L^2(\nu^*)}$$

for all $t \in [0,T]$. Let $d\psi_\infty := \frac{\hat{m}\, d\nu}{\int \hat{m}\, d\nu}$. If $\psi.(\mu)$ is the associated solution of (1.1), it follows that

$$\|\psi_t(\mu) - \psi_\infty\|_{var} \le c_{2.5}e^{-\kappa_* t}\|h_0 - 1\|_{L^2(\nu^*)}\,.$$

Proof. Clearly, $m_t = m$ and $\hat{m}_t = \hat{m}$ are nonnegative solutions of the backward and forward equations (2.2) and (2.1) with $\lambda_t = \lambda$ and $\hat{\lambda}_t = \hat{\lambda}$. Hence Theorem 2.3 gives the first inequality. Note that the measure ψ_∞ is a stationary point of ψ_t, that is $\psi_t(\psi_\infty) = \psi_\infty$. Consequently, inequality (2.4) implies that

$$\|\psi_t(\mu) - \psi_\infty\|_{var} \le c_{2.5}e^{-\kappa_* t}\|h_0 - 1\|_{L^2(\nu^*)}\,,$$

hence the assertion. □

3. Applications

3.1. Stability of the simulated annealing algorithm

We apply the abstract results of Section 2 to the simulated annealing algorithm (cf. [7]). To this end let S be a countable set. Fix a bounded potential $U : S \to \mathbb{R}_+$, a sub-probability kernel q on S (the "proposal matrix") with symmetrizing measure ν having full support and a monotone increasing continuously differentiable function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\beta(0) = 0$ and $\lim_{t\to\infty}\beta(t) = \infty$ (the "cooling schedule"). Without loss of generality we may define $q(x, \{x\}) = 1$ for all $x \in S$. The simulated annealing algorithm with Metropolis sampler associated with U, q and β is then given as the solution of the linear equation

$$\dot{\psi}_t(\mu) = \hat{A}_{\beta(t)}\psi_t(\mu)\,, \psi_0(\mu) = \mu\,, \tag{3.1}$$

where

$$A_\beta f(x) = \int f(y) - f(x)\,\gamma_\beta(x,y)\, q(x,dy)\,, f \in L^2(\nu)\,,$$

and

$$\gamma_\beta(x,y) = \begin{cases} e^{-\beta(U(y)-U(x))^+} & \text{if } y \neq x \\ 1 - \sum_{\xi\neq x}\gamma_\beta(x,\xi)q(x,\{\xi\}) & \text{if } y = x\,. \end{cases}$$

In other words, $\psi_t(\mu)$ is the distribution at time t of the time-inhomogeneous Markov process with generator $A_{\beta(t)}$, $t \ge 0$, and initial distribution μ.

The adjoint operator $\hat{A}_\beta$ of A_β in $L^2(\nu)$ is given by

$$\hat{A}_\beta f(x) := \int f(y) - f(x)\hat{\gamma}_\beta(x,y)q(x,dy) + V_\beta(x)f(x)\,,$$

where $\hat{\gamma}_\beta(x,y) := \gamma_\beta(y,x)$ and $V_\beta(x) := \int \hat{\gamma}_\beta(x,y) - \gamma_\beta(x,y)\,q(x,dy)$. Note that $D := L^\infty(\nu)$ is an algebra contained in $D(A_\beta) \cap D(\hat{A}_\beta)$ for all β. Moreover, since $\hat{A}_\beta : L^\infty(\nu) \to L^\infty(\nu)$ is bounded and the cooling schedule β locally Lipschitz, it follows that for all bounded initial conditions n_0 there exists a unique solution n_t, $t \geq 0$, of the equation

$$\dot{n}_t = \hat{A}_{\beta(t)} n_t \tag{3.2}$$

(see [2], Section III.1). Moreover, $n_0 \geq 0$ implies $n_t \geq 0$ for all $t > 0$. Given a solution $n.$ of (3.2), bounded from above and from below uniformly in x and locally in t, we obtain similarly that every $h \in L^\infty(\nu)$ is an admissible initial condition of equation (2.3). Here, the ground state transform $\hat{A}_t^*$ associated with $\hat{A}_{\beta(t)}$ and n_t is given by

$$\begin{aligned}\hat{A}_t^* f(x) &= \frac{1}{n_t(x)}(\hat{A}_{\beta(t)}(fn_t)(x) - \hat{A}_{\beta(t)} n_t(x) f(x)) \\ &= \frac{1}{n_t(x)}\int (f(y) - f(x)) n_t(y)\hat{\gamma}_{\beta(t)}(x,y)\,q(x,dy)\,, \quad f \in L^\infty(\nu)\,.\end{aligned}$$

In particular,

$$\Gamma(\hat{A}_t^*)(f)(x) = \frac{1}{n_t(x)}\int (f(y) - f(x))^2 n_t(y)\hat{\gamma}_{\beta(t)}(x,y)q(x,dy)\,. \tag{3.3}$$

Proposition 3.1. *Let n_t, $t \geq 0$, be a nonnegative solution of the equation* (3.2)*. Then*

$$\inf_{x\in S} n_0(x) e^{-\beta(t)\mathrm{osc}\,(U)} \leq n_t \leq \sup_{x\in S} n_0(x) e^{\beta(t)\mathrm{osc}\,(U)}\,.$$

Proof. To simplify notations in the following let

$$f^- := \inf_{x\in S} f(x) \qquad \text{and} \qquad f^+ := \sup_{x\in S} f(x)$$

for any $f : S \to \mathbb{R}$. Note that (3.2) is equivalent to

$$\dot{n}_t(x) = \langle n_t \hat{\gamma}_{\beta(t)}(x,\cdot), q(x,\cdot)\rangle - n_t(x)\,. \tag{3.4}$$

Define $g_t(x) := e^{\beta(t)U(x)+t} n_t(x)$, then equation (3.4) implies that

$$\dot{g}_t(x) = \dot{\beta}(t)U(x)g_t(x) + e^{\beta(t)U(x)+t}\langle n_t\hat{\gamma}_{\beta(t)}(x,\cdot), q(x,\cdot)\rangle. \tag{3.5}$$

The detailed balance equation

$$e^{-\beta(t)U(y)}\hat{\gamma}_{\beta(t)}(x,y) = e^{-\beta(t)U(y)}\gamma_{\beta(t)}(y,x) = e^{-\beta(t)U(x)}\gamma_{\beta(t)}(x,y)$$

implies that

$$e^{\beta(t)U(x)+t}\langle n_t\hat{\gamma}_{\beta(t)}(x,\cdot), q(x,\cdot)\rangle = \langle g_t\gamma_{\beta(t)}(x,\cdot), q(x,\cdot)\rangle\,.$$

Hence (3.5) is equivalent to

$$\dot{g}_t(x) = \dot{\beta}(t)U(x)g_t(x) + \langle g_t\gamma_{\beta(t)}(x,\cdot), q(x,\cdot)\rangle\,. \tag{3.6}$$

For the proof of the lower bound now observe that (3.6) implies that $\dot{g}_t(x) \geq (\dot{\beta}(t)U^- + 1)g_t^-$, hence $g_t^- \geq g_0^- + \int_0^t (\dot{\beta}(s)U^- + 1)\, g_s^- ds$. Consequently, $g_t^- \geq g_0^- \exp(\beta(t)U^- + t)$, and thus

$$n_t(x) \geq n_0^- \exp(\beta(t)(U^- - U(x))) \geq n_0^- \exp(-\beta(t)\mathrm{osc}\,(U))\,.$$

Similarly, (3.6) implies that $\dot{g}_t(x) \leq (\dot{\beta}(t)U^+ + 1)g_t^+$, hence $g_t^+ \leq g_0^+ + \int_0^t (\dot{\beta}(s)U^+ + 1)\, g_s^+ ds$. Consequently, $g_t^+ \leq g_0^+ \exp(\beta(t)U^+ + t)$, and thus

$$n_t(x) \leq n_0^+ \exp(\beta(t)(U^+ - U(x))) \leq n_0^+ \exp(\beta(t)\mathrm{osc}\,(U))\,. \qquad \square$$

From Theorem 2.3 and Corollary 2.5 we now obtain the following.

Corollary 3.2. *Assume that the quadratic form*

$$\frac{1}{2}\int\int (f(x) - f(y))^2 q(x,dy)\,\nu(dx)$$

satisfies a Poincaré inequality in $L^2(\nu)$ with constant less than κ^{-1}.

(i) *Let $h_{i,\cdot}$, $i = 1,2$, be admissible solutions of* (2.3) *with $\int h_{1,0}\,d\nu = \int h_{2,0}\,d\nu$. Then*

$$\|h_{1,t} - h_{2,t}\|^2_{L^2(n_t\nu)} \leq e^{-2\kappa\int_0^t e^{-3\beta(s)\mathrm{osc}\,(U)}\,ds}\|h_{1,0} - h_{2,0}\|^2_{L^2(\nu)}\,.$$

Here n_t is a solution of (3.2) *(with initial condition $n_0 = 1$).*

(ii) *Let $\mu_i \ll \nu$ with bounded density $h_{i,0}$, $i = 1,2$. Let $\psi_\cdot(\mu_i)$ be the corresponding solution of the simulated annealing algorithm* (3.1)*. Then*

$$\|\psi_t(\mu_1) - \psi_t(\mu_2)\|_{var} \leq e^{-\kappa\int_0^t e^{-3\beta(s)\mathrm{osc}\,(U)}\,ds}\left(\|h_{1,0}\|_\infty + \|h_{2,0}\|_\infty\right)\,.$$

Proof. Note that $m_t \equiv 1$, $t \geq 0$, is a solution of the backward equation (2.2) with $\lambda_t = 0$ and $\hat{m}_t = n_t$ a solution to the forward equation (2.1) with $\hat{\lambda}_t = 0$, since $\sigma_t = 0$ in this particular case. Note that $\frac{d}{dt}\int n_t\,d\nu = \int \hat{A}_{\beta(t)}n_t\,d\nu = 0$, hence $d\nu_t^* = n_t\,d\nu$, and Theorem 2.3 now implies that

$$\|h_{1,t} - h_{2,t}\|^2_{L^2(n_t\nu)} \leq e^{-2\int_0^t \kappa_*(s)\,ds}\|h_{1,0} - h_{2,0}\|^2_{L^2(\nu)}\,, \tag{3.7}$$

where $\frac{1}{\kappa_*(t)}$ is the constant for the Poincaré inequality of the quadratic form

$$\begin{aligned}
\mathcal{A}_t^*(f) &= \frac{1}{2}\int \Gamma(\hat{A}_t^*)(f)\,d\nu_t^* \\
&= \frac{1}{2}\int\int (f(y) - f(x))^2 n_t(y)\hat{\gamma}_{\beta(t)}(x,y)q(x,dy)\,\nu(dx) \\
&= \frac{1}{2}\int\int (f(x) - f(y))^2 \gamma_{\beta(t)}(x,y)q(x,dy)n_t(x)\,\nu(dx)
\end{aligned}$$

in $L^2(n_t\nu)$ (see (3.3)). Since

$$\gamma_{\beta(t)}(x,y) \geq e^{-\beta(t)\mathrm{osc}\,(U)}\,, \quad x \neq y\,,$$

the upper and lower bound on n_t, obtained in the last Proposition, now imply that

$$\kappa_*(t) \geq \kappa e^{-3\beta(t)\mathrm{osc}\,(U)}\,. \tag{3.8}$$

Inserting inequality (3.8) into (3.7), we obtain the first assertion. The second assertion now is an immediate consequence of the last inequality and Corollary 2.5. □

Remark 3.3. The last corollary shows that polynomial stability holds for the simulated annealing algorithm if $\beta(t) \leq \frac{\log(1+t)}{3\mathrm{osc}\,(U)}$, since in this case

$$\|\psi_t(\mu_1) - \psi_t(\mu_2)\|_{var} \leq \frac{1}{(1+t)^\kappa}\left(\|h_{1,0}\|_\infty + \|h_{2,0}\|_\infty\right)\,.$$

3.2. Stability of the pathwise filter equation

Consider a time-homogeneous Markov process $\mathbb{M} = ((X_t), (P_x))$ on S (the "signal process") with generator A and associated carré du champ operator Γ. Suppose that $\mathbb{M}$ is seen through the observation process

$$dY_t = g(X_t)\,dt + d\tilde{W}_t\,, \quad Y_0 = 0\,,$$

where $g = (g_1, \ldots, g_p)^T : S \to \mathbb{R}^p$ and $(\tilde{W}_t)$ is a p-dimensional Brownian motion independent of $\mathbb{M}$. Given the observation $y = (y_1, \ldots, y_p)^T \in C(\mathbb{R}_+; \mathbb{R}^p)$, the pathwise filter equation is given by

$$\dot{\psi}_t^y(\mu) = \hat{A}_t^y \psi_t^y(\mu) + \sigma_t^y \psi_t^y(\mu) - \int \sigma_t^y\, d\psi_t^y(\mu)\, \psi_t^y(\mu)\,, \psi_0^y(\mu) = \mu\,, \tag{3.9}$$

where

$$A_t^y f = Af - \sum_{i=1}^p y_i(t)\Gamma(g_i, f)$$

and

$$\sigma_t^y(x) = -\langle y(t), Ag(x)\rangle + \frac{1}{2}\langle y(t), \Gamma(g,g)(x)y(t)\rangle - \frac{1}{2}\|g(x)\|^2\,.$$

Here $\Gamma(g,g)(x) = (\Gamma(g_i, g_j)(x))_{1\leq i,j\leq p}$ (cf. [5], p. 19, and [6], p. 75). Up to the density $e^{\langle y(t), g(x)\rangle}$, the solution of equation (3.9) gives a regular conditional distribution of the signal X_t given the observation $Y_\cdot$ up to time t (see also Chapter 11 in [8]).

The linear case

Consider $S = \mathbb{R}^d$ and suppose that the signal process $\mathbb{M}$ is given by the solution of the linear stochastic differential equation

$$dX_t = -BX_t\,dt + \sigma dW_t \tag{3.10}$$

for some d-dimensional Brownian motion and $d \times d$-matrices B and σ, so that the generator of the signal process is given by

$$Af(x) = \frac{1}{2}tr(Qf''(x)) - \langle Bx, f'(x)\rangle\,,$$

where $Q = (q_{ij})_{1\le i,j\le d}$ and $q_{ij} = (\sigma\sigma^T)_{ij}$. We suppose that the signal process is seen through the linear observation

$$dY_t = GX_t\,dt + d\tilde{W}_t\,, Y_0 = 0\,,$$

for some $p \times d$-matrix G. From now on we assume that Q is positive definite.

Remark 3.4. Before we state our main result on the exponential stability of (3.9) let us consider the following heuristic in the particular case where B is symmetric and positive definite, and B, Q and G^TG are simultaneously diagonizable: If $y \equiv 0$ then $A_t^y = A$ and $\sigma_t^y = -\frac{1}{2}\|Gx\|^2$. The ground state m associated with A and $-\frac{1}{2}\|Gx\|^2$ is given by $m(x) = \exp(\frac{1}{2}\langle Q^*x, x\rangle)$, where

$$Q^* = Q^{-1}(B - \sqrt{B^2 + QG^TG}) \tag{3.11}$$

is a symmetric solution of the matrix Riccati equation

$$0 = Q^*QQ^* - 2Q^*B - G^TG\,.$$

In fact, it is easy to check that (3.11) implies that

$$Am - \frac{1}{2}\|Gx\|^2 m = \frac{1}{2}\mathrm{tr}\ (QQ^*)m\,.$$

The ground state transform A^* associated with A and m is given by

$$A^*f = \frac{1}{2}tr(Qf''(x)) - \langle\sqrt{B^2 + QG^TG}x, f'(x)\rangle\,.$$

The symmetrizing measure for A^* is given by

$$\nu_* = N(0, \frac{1}{2}Q\sqrt{B^2 + QG^TG}^{-1}),$$

and A^* has a mass gap of size κ_* (in $L^2(\nu_*)$), where

$$\kappa_* = \min\{\lambda : \lambda \text{ is an eigenvalue of } \sqrt{B^2 + QG^TG}\}\,. \tag{3.12}$$

The case $y = 0$ is somehow the limiting behaviour for a typical observation. Indeed, note that positive definiteness of Q and B implies that

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t GX_s\,ds = 0\ P_\nu\text{-a.s.}\,,$$

where $\nu = N(0, \frac{1}{2}QB^{-1})$. Since in addition $\lim_{t\to\infty}\frac{1}{t}\tilde{W}_t = 0$ a.s. by the strong law of large numbers for Brownian motion, we have that $\lim_{t\to\infty}\frac{1}{t}Y_t = 0$ for a typical observation.

In fact, Theorem 3.7 and Remark 3.8 below imply in this particular case that for *any* (continuous) observation Y the solution $\psi_t^Y(\mu)$ to the pathwise filter equation

$$\dot{\psi}_t^Y(\mu) = \hat{A}_t^Y\psi_t^Y(\mu) + \sigma_t^Y\psi_t^Y(\mu) - \int\sigma_t^Y\,d\psi_t^Y(\mu)\,\psi_t^Y(\mu)\,, \psi_0^Y(\mu) = \mu$$

is exponentially stable in the total variation norm with rate κ_*.

We will need the following result several times:

Lemma 3.5. *Let $T > 0$, $B, G : [0,T] \to \mathbb{R}^{d\times d}$, $C, D : [0,T] \to \mathbb{R}^d$ be continuous. For $t \in [0,T]$ define*

$$L_t f(x) := \frac{1}{2} tr(Qf''(x)) - \langle B(t)x + C(t), f'(x)\rangle$$

and

$$\sigma_t(x) = \langle D(t), x\rangle - \frac{1}{2}\langle G^T(t)G(t)x, x\rangle\,.$$

Let $E : [0,T] \to \mathbb{R}^{d\times d}$ be a symmetric solution of the matrix Riccati equation

$$\dot{E}(t) = E(t)QE(t) - E(t)B(t) - B^T(t)E(t) - G(t)^T G(t)$$

and $F : [0,T] \to \mathbb{R}^d$ be a solution of the linear equation

$$\dot{F}(t) = (E(t)Q - B^T(t))F(t) - (E(t)C(t) - D(t))\,.$$

Then

$$n_t(x) := \exp(\frac{1}{2}\langle E(t)x, x\rangle + \langle F(t), x\rangle)\,, \quad t \in [0,T]\,,$$

is a solution to the forward equation

$$\dot{n}_t = L_t n_t + \sigma_t n_t + \lambda_t n_t$$

with

$$\lambda_t = -\frac{1}{2}\mathrm{tr}(QE(t)) - \frac{1}{2}\langle QF(t), F(t)\rangle + \langle C(t), F(t)\rangle\,.$$

The proof of the lemma is an immediate calculation, so that we omit it.

From now on we make the following two assumptions:

Assumption 1. There exists a symmetric positive definite solution E_∞^+ of the algebraic matrix Riccati equation

$$0 = E_\infty Q E_\infty - E_\infty B - B^T E_\infty - G^T G\,. \tag{3.13}$$

Assumption 2. The symmetric negative semidefinite solution $E(t)$, $t \geq 0$, of the matrix Riccati equation

$$\dot{E}(t) = E(t)QE(t) - E(t)B - B^T E(t) - G^T G\,, E(0) = 0 \tag{3.14}$$

converges to some matrix E_∞^- as $t \to \infty$.

For existence and sufficient conditions on the convergence of the negative semidefinite solution $E(\cdot)$ of (3.14) see Chapter 10 in [9].

It follows from the two assumptions that the matrix $\frac{1}{2}\sqrt{Q}(E_\infty^+ - E_\infty^-)\sqrt{Q}$ is positive definite. Let

$$\kappa_* := \min\{\lambda : \lambda \text{ is an eigenvalue of } \frac{1}{2}\sqrt{Q}(E_\infty^+ - E_\infty^-)\sqrt{Q}\}\,.$$

Remark 3.6. In the particular case, where B is symmetric, Q, B and G^TG are simultaneously diagonizable and $B^2 + QG^TG$ is positive definite, it is easy to see that

$$E_\infty^+ = Q^{-1}(B + \sqrt{B^2 + QG^TG})$$

is a positive definite symmetric solution of (3.13) and

$$E_\infty^- = Q^{-1}(B - \sqrt{B^2 + QG^TG})$$

is the limit for $t \to \infty$ of the negative semidefinite symmetric solution

$$\begin{aligned} E(t) := -Q^{-1}G^TG(E_\infty^+ - E_\infty^- \exp(-2t\sqrt{B^2 + QG^TG}))^{-1} \\ \cdot (I - \exp(-2t\sqrt{B^2 + QG^TG})) \end{aligned}$$

of (3.14). It follows that $\frac{1}{2}\sqrt{Q}(E_\infty^+ - E_\infty^-)\sqrt{Q} = \sqrt{B^2 + QG^TG}$, so that in this case the lowest eigenvalue κ_* coincides with κ_* in (3.12).

To apply the abstract results of Section 2, let $\nu := N(0, (E_\infty^+)^{-1})$ be the reference measure and

$$\begin{aligned} D := \{f \in C^2(\mathbb{R}^d) : \exists M, c \text{ such that } |\partial^\alpha f|(x) \le M \exp(c\|x\|) \\ \text{for all } \alpha \in (\mathbb{N}_0^d)^2\,, |\alpha| \le 2\} \end{aligned}$$

be the test function space.

Theorem 3.7. *Let $y_\cdot \in C([0,\infty); \mathbb{R}^p)$, $y_0 = 0$, and assume that Assumptions* 1 *and* 2 *hold. Let*

$$\kappa_*(s) := \min\{\lambda : \lambda \text{ is an eigenvalue of } \frac{1}{2}\sqrt{Q}(E_\infty^+ - E(s))\sqrt{Q}\}\,, \tag{3.15}$$

and κ_*^+ be the lowest eigenvalue of $\frac{1}{2}\sqrt{Q}E_\infty^+\sqrt{Q}$. Then the solution $\psi_t^y(\mu)$ to the pathwise filter equation (3.9) *is (exponentially) stable with (exponential) rate $\frac{1}{t}\int_0^t \kappa_*(s)\, ds \ge \kappa_*^+$ in the following sense: for initial condition $\mu_i \ll \nu$ with density bounded from below and from above, it follows that*

$$\limsup_{t\to\infty} e^{\int_0^t \kappa_*(s)\, ds} \|\psi_t^y(\mu_1) - \psi_t^y(\mu_2)\|_{var} < \infty\,.$$

Remark 3.8. (i) The main feature of the above theorem is the fact that the (exponential) rate is *independent of the observation* y. Theorem 3.7 complements Theorem 2.6 of [12] under our Assumptions 1 and 2. Note that in the particular case where B is symmetric, Q, B and G^TG are simultaneously diagonizable and $B^2 + QG^TG$ is positive definite, the quantity $\bar{\lambda}$ defined in [12], p. 230, in fact coincides with κ_*. Indeed, in this case, $P_\infty := (E_\infty^+)^{-1}$ is a positive definite solution of the algebraic matrix Riccati equation (9) on p. 229 (note that B in [12] has to be replaced by $-B$), so that $B + P_\infty G^TG = \sqrt{B^2 + QG^TG}$ coincides with $\frac{1}{2}\sqrt{Q}(E_\infty^+ - E_\infty^-)\sqrt{Q}$. We emphasize, that our approach is entirely different from the approach in [12] and that it provides an alternative interpretation of the rate κ_*.

(ii) In the situation of (i) it follows that $\lim_{t\to\infty} E(t) = E_\infty^-$ with exponential rate (see the explicit representation for $E(t)$ in Remark 3.6). It follows that $\lim_{t\to\infty} \kappa_*(t) = \kappa_*$ with an exponential rate too. Consequently,

$$M := \sup_{t\geq 0} \left| \int_0^t \kappa_*(s) - \kappa_* \, ds \right| < \infty$$

which implies for μ_i, $i = 1, 2$, as in the theorem that

$$\limsup_{t\to\infty} e^{\kappa_* t} \|\psi_t^y(\mu_1) - \psi_t^y(\mu_2)\|_{var} < \infty \, ,$$

so that in this case (ψ_t^y) is in fact exponentially stable with an exponential rate κ_*.

(iii) Note that the theorem also gives exponential stability of the pathwise filter equation (3.9) in cases where the signal process is not ergodic. Stability of the Kalman-Bucy filter for possibly nonergodic signals was, of course, known. Extensions to the case of Benes filters have been obtained in [11] and for gradient-type signal processes in [14]. Note that the assumptions on the initial distribution of the signal process made in [14] are rather restrictive. On the other hand, the real advantage is the new and explicit *variational* interpretation of the rate of stability.

Proof of Theorem 3.7. Fix $t > 0$ and an observation $y_\cdot \in C([0,t];\mathbb{R}^p)$, $y_0 = 0$. Let A_s^y and σ_s^y, $s \in [0,t]$, be as above. Note that the adjoint operator $\hat{A}$ of the generator A of the signal process (3.10) in $L^2(\nu)$ is given by

$$\begin{aligned}
\hat{A}f(x) &= \frac{1}{2}\mathrm{tr}(Qf''(x)) - \langle (QE_\infty^+ - B)x, f'(x)\rangle \\
&\quad + \frac{1}{2}\langle (E_\infty^+ Q E_\infty^+ - E_\infty^+ B - B^T E_\infty^+)x, x\rangle f(x) - \frac{1}{2}\mathrm{tr}(\hat{B} - B) f(x) \\
&= \frac{1}{2}\mathrm{tr}(Qf''(x)) - \langle \hat{B}x, f'(x)\rangle \\
&\quad + \frac{1}{2}\langle G^T G x, x\rangle f(x) - \frac{1}{2}\mathrm{tr}(\hat{B} - B) f(x) \, ,
\end{aligned}$$

with $\hat{B} := QE_\infty^+ - B$.

It follows that the adjoint operator $\hat{A}_s^y$ of A_s^y in $L^2(\nu)$ is given by

$$\begin{aligned}
\hat{A}_s^y f(x) &= \frac{1}{2}\mathrm{tr}(Qf''(x)) - \langle \hat{B}x - QG^T y(s), f'(x)\rangle \\
&\quad - (\langle QG^T y(s), E_\infty^+ x\rangle - \frac{1}{2}\langle G^T G x, x\rangle) f(x) - \frac{1}{2}\mathrm{tr}(\hat{B} - B) f(x) \, ,
\end{aligned}$$

so that

$$\begin{aligned}
\hat{A}_s^y f(x) + \sigma_s^y(x) f(x) &= \frac{1}{2}\mathrm{tr}(Qf''(x)) - \langle \hat{B}x - QG^T y(s), f'(x)\rangle \\
&\quad + (-\langle \hat{B}^T G^T y(s), x\rangle + \frac{1}{2}\langle y(s), GQG^T y(s)\rangle - \frac{1}{2}\mathrm{tr}(\hat{B} - B)) f(x) \, .
\end{aligned}$$

Lemma 3.5 applied to $B(s) = \hat{B}$, $C(s) = -QG^Ty(s)$, $D(s) = -\hat{B}^TG^Ty(s)$ and $G(s) = 0$ implies that

$$\hat{m}_s(x) = \exp(\langle \hat{F}(s), x\rangle)\,, \quad s \in [0,t]\,, \tag{3.16}$$

where

$$\hat{F}(s) = -e^{-s\hat{B}^T}\int_0^s e^{r\hat{B}^T}\hat{B}^TG^Ty(r)\,dr$$

is a solution of the forward equation (2.1) with

$$\begin{aligned}\hat{\lambda}_s = &-\frac{1}{2}\langle Q\hat{F}(s), \hat{F}(s)\rangle - \langle QG^Ty(s), \hat{F}(s)\rangle\\ &-\frac{1}{2}\langle y(s), GQG^Ty(s)\rangle + \frac{1}{2}\mathrm{tr}(\hat{B} - B)\end{aligned}$$

and initial condition $\hat{m}_0 \equiv 1$.

Similarly, Lemma 3.5 applied to $B(s) = B$, $C(s) = QG^Ty(s)$, $D(s) = B^TG^Ty(s)$ and $G(s) = G$ and the time-reversal $s \mapsto t - s$, implies that

$$m_s(x) = \exp(\frac{1}{2}\langle E(t-s)x, x\rangle + \langle F(s), x\rangle)\,, s \in [0,t]\,,$$

with $E(s)$, $s \geq 0$, as in (3.14), and

$$\dot{F}(s) = -(E(t-s)Q - B^T)F(s) + (E(t-s)QG^Ty(s) - B^TG^Ty(s))$$

with terminal condition $F(t) = 0$ is a solution of the backward equation (2.2) with

$$\begin{aligned}\lambda_s = &-\frac{1}{2}\mathrm{tr}(QE(t-s)) - \frac{1}{2}\langle QF(s), F(s)\rangle + \langle QG^Ty(s), F(s)\rangle\\ &+\frac{1}{2}\langle y(s), GQG^Ty(s)\rangle\end{aligned}$$

and terminal condition $m_t \equiv 1$.

Consequently,

$$\begin{aligned}\nu_s^*(dx) &= \frac{m_s\hat{m}_s\,d\nu}{\int m_s\hat{m}_s\,d\nu}\\ &= Z_s^{-1}\exp(-\frac{1}{2}\langle (E_\infty^+ - E(t-s))x, x\rangle + \langle F(s) + \hat{F}(s), x\rangle)\,dx\end{aligned}$$

for some normalizing constant Z_s. In particular, the quadratic form generated by the ground state transform $\hat{A}_s^*$ associated with $\hat{A}_s^y$ and $\hat{m}_s$,

$$\mathcal{A}_s^*(f) = \frac{1}{2}\int \langle Qf', f'\rangle\,d\nu_s^*\,, \quad f \in D\,,$$

satisfies a Poincaré inequality with constant less than $\kappa_*(t-s)^{-1}$, where $\kappa_*(t-s)$ is as in (3.15).

We will show in Lemma 3.9 below that for $h_{i,0} \in \mathcal{B}_b(\mathbb{R}^d)_+$ with $\int h_{i,0}\,d\nu_0^* = 1$, there exist admissible solutions $h_{i,s}$, $s \in [0,t]$, of (2.3) with initial condition $h_{i,0}$. Theorem 2.3 now implies that

$$\|h_{1,t} - h_{2,t}\|^2_{L^2(\nu_t^*)} \leq e^{-2\int_0^t \kappa_*(t-s)\,ds}\|h_{1,0} - h_{2,0}\|^2_{L^2(\nu_0^*)}\,.$$

Finally, let $\mu_i \in \mathcal{M}_1(S)$ be absolutely continuous w.r.t. ν with density $\tilde{h}_{i,0}$ bounded from below and from above and let $h_{i,0} := \frac{\tilde{h}_{i,0}}{\int \tilde{h}_{i,0}, d\nu_0^*}$, $i = 1, 2$. Let $\psi_\cdot(\mu_i)$ be the solutions of the pathwise filter equation (3.9) with initial condition μ_i. Then Corollary 2.5 implies that

$$\|\psi_t^y(\mu_1) - \psi_t^y(\mu_2)\|_{\text{var}} \le c_{2.5}(t) e^{-\int_0^t \kappa_*(s)\, ds} \|h_{1,0} - h_{2,0}\|_{L^2(\nu_0^*)} .$$

Note that $m_t \equiv 1$, so that $c_{2.5}(t) = 1$. Since $\tilde{h}_{i,0}$ is bounded from below and from above it follows that $\delta^{-1} \ge h_{i,0} \ge \delta > 0$ for some constant δ. Consequently, $\frac{h_{i,0}}{\int h_{i,0}\, d\nu_0^*} \le \delta^{-2}$ (independent of t). This proves the theorem. $\square$

It remains to show the following:

Lemma 3.9. *Let the notation be as in the proof of the theorem. Let $h_0 \in \mathcal{B}_b(\mathbb{R}^d)_+$, with $\int h_0\, d\nu_0^* = 1$. Then there exists an admissible solution h_s, $s \in [0, t]$, of (2.3) with initial condition h_0.*

Proof. Using (3.16), the ground state transform $\hat{A}_s^*$ associated with $\hat{A}_s^y$ and $\hat{m}_s$ can be written explicitely as

$$\hat{A}_s^* f(x) = \frac{1}{2} tr(Q f''(x)) - \langle \hat{B}x - QG^T y(s) - Q\hat{F}(s), f'(x)\rangle .$$

For any $s \in [0, t]$ and $x \in \mathbb{R}^d$ let $\xi_r(s, x)$, $s \le r \le t$, be the solution of the linear stochastic differential equation

$$d\xi_r(s, x) = -(\hat{B}\xi_r(s, x) - Q(G^T y(t - r) + \hat{F}(t - r)))\, dr + \sigma dW_r\, , \;\; \xi_s(s, x) = x\, .$$

It follows from Theorem V.7.4 in [10] and the time reversal $s \mapsto t - s$ that for any initial condition h_0 that is twice continuously differentiable with polynomially bounded partial derivatives, the function

$$h_s(x) := E[h_0(\xi_t(t - s, x))]\, , \quad s \in [0, t]\, , \;\; x \in \mathbb{R}^d\, ,$$

is a solution of the equation $\dot{h}_t = \hat{A}_t^* h_t$. It is easy to extend the last result to obtain for any nonnegative bounded measurable initial condition h_0 an admissible solution h_s of (2.3). Indeed, the distribution of $\xi_r(s, x)$ is a Gaussian distribution with mean

$$m_r(s, x) = e^{-(r-s)\hat{B}}(x + \int_s^r e^{(u-s)\hat{B}} Q(G^T y(t - u) + \hat{F}(t - u))\, du)$$

and covariance $Q_{s,r} = Q_{r-s}$, where

$$Q_u = \int_0^u e^{-v\hat{B}} Q e^{-v\hat{B}^T}\, dv\, ,$$

which implies that

$$h_s(x) = Z_s^{-1} \int h_0(z) \exp(-\frac{1}{2}\langle Q_s^{-1}(z - m_t(t - s, x)), (z - m_t(t - s, x))\rangle)\, dz\, .$$

Here Z_s is a normalizing constant. If h_0 is bounded, the time-derivative $\dot{h}_s$ exists in $L^4(\nu)$ for all $s \in]0, t]$ which implies the assertion. $\square$

References

[1] P. Del Moral and A. Guionnet, *On the stability of interacting processes with applications to filtering and genetic algorithms*, Ann. Inst. H. Poincaré Probab. Statist., **37** (2001), 155–194.

[2] J. L. Daleckii and M. G. Krein, *Stability of Solutions of Differential Equations in Banach Space*, American Mathematical Society, Providence, RI, 1974.

[3] P. Del Moral, *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*, Springer, Berlin, 2000.

[4] P. Del Moral and L. Miclo, *Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering*, Séminaire de Probabilités, XXXIV, Lecture Notes in Math., Springer, **1729** (2000), 1–145.

[5] P. Del Moral and L. Miclo, *On the stability of non linear Feynman-Kac semigroups*, Preprint, Toulouse, 2001.

[6] W. H. Fleming and S. K. Mitter, *Optimal control and nonlinear filtering for nondegenerate diffusion processes*, Stochastics, **8** (1982), 63–77.

[7] R. Holley and D. Stroock, *Simulated annealing via Sobolev inequalities*, Commun. Math. Phys., **115** (1988), 553–569.

[8] G. Kallianpur, *Stochastic Filtering Theory*, Springer, Berlin, 1980.

[9] H. W. Knobloch and H. Kwakernaak, *Lineare Kontrolltheorie*, Springer, Berlin, 1985.

[10] N. Krylov, *Introduction to the Theory of Diffusion Processes*, American Mathematical Society, Providence, RI, 1995.

[11] D. Ocone, *Asymptotic stability of Benes filters*, Stochastic Anal. Appl., **17** (1999), 1053–1074.

[12] D. Ocone and E. Pardoux, *Asymptotic stability of the optimal filter with respect to its initial condition*, SIAM J. Control And Optimization, **34** (1996), 226–243.

[13] W. Stannat, *On the convergence of genetic algorithms – A variational approach*, Probab. Theor. Rel. Fields, **129** (2004), 113–132.

[14] W. Stannat, *Stability of the filter equation for a time-dependent signal on* $\mathbb{R}^d$, Appl. Math. Optim., **52** (2005), 39–71.

Wilhelm Stannat
Fachbereich Mathematik
Technische Universität Darmstadt
Schlossgartenstr. 7
64289 Darmstadt
Germany
e-mail: `stannat@mathematik.tu-darmstadt.de`

Progress in Probability, Vol. 59, 363–395

Some Applications of the Malliavin Calculus to Sub-Gaussian and Non-Sub-Gaussian Random Fields

Andrew B. Vizcarra and Frederi G. Viens

Abstract. We introduce a boundedness condition on the Malliavin derivative of a random variable to study sub-Gaussian and other non-Gaussian properties of functionals of random fields, with particular attention to the estimation of suprema. We relate the boundedness of the nth Malliavin derivative to a new class of "sub-nth-Gaussian chaos" processes. An expected supremum estimation, extending the Dudley theorem, is proved for such processes. Sub-nth-Gaussian chaos concentration inequalities for the supremum are obtained, using Malliavin derivative conditions; for $n = 1$, this generalizes the Borell-Sudakov inequality to a class of sub-Gaussian processes, with a particularly simple and efficient proof; for $n = 2$ a natural extension to sub-2nd-Gaussian chaos processes is established; for $n \geq 3$ a slightly less efficient Malliavin derivative condition is needed.

Mathematics Subject Classification (2000). Primary 60H07; Secondary 60G15, 60G17.

Keywords. Stochastic analysis, Malliavin derivative, Wiener chaos, concentration, sub-Gaussian process, suprema of processes, Dudley-Fernique theorem, Borell-Sudakov inequality.

1. Introduction

Gaussian analysis, and in particular the Malliavin calculus, are powerful and versatile tools in contemporary probability theory and stochastic analysis. The latter has applications ranging from other areas of probability theory to physics, to finance, to name a few; a very short selection of references might include [2, 5, 6, 7, 13, 14, 15, 16, 17, 18, 23]. We will not attempt to give an overview of such a wide array of areas. Instead, this article presents a new way of using Malliavin derivatives to uncover sub-Gaussian and other non-Gaussian properties of functionals of random fields, with particular attention to the estimation of suprema.

After introducing some standard material on Wiener chaoses and the Malliavin derivative in what we hope is a streamlined and didactic way (Section 2), we introduce the fundamental lemma that serves as a basis and a springboard for non-Gaussian results: it is the observation that if a random variable X has a Malliavin derivative whose norm in $L^2[0,1]$ is almost surely bounded, then X is sub-Gaussian (Lemma 3.3). In Section 3, this lemma is exploited to analyze sub-Gaussian processes. Even though the proofs of the results therein are quite elementary, we believe they may have far-reaching consequences in probability and its applications. For example, even though it is not stated so explicitly, Lemma 3.3 is the key ingredient in the new proofs of existence of Lyapunov exponents for the continuous space stochastic Anderson model and the Brownian directed polymer in a Gaussian environment, obtained, respectively, in [8] and [19]; these existence results had been open problems for many years (see, e.g., [4]). Lemma 3.3, and its application to sub-Gaussian deviations of the supremum of a sub-Gaussian random field (Theorem 3.6, which is a generalization of the so-called Borell-Sudakov inequality, see [1]), are techniques applied in [22] for statistical estimation problems for non-linear fractional Brownian functionals.

Inspired by the power of such applications, we postulate that in order to generalize the concept of sub-Gaussian random variables, one would be well advised to investigate the properties of random fields whose nth Malliavin derivative is bounded. Our study chooses to define the concept of *sub-nth-Gaussian chaos* (or *sub-nth chaos*, for short) random fields slightly differently, in order to facilitate the study of such processes' concentration properties as well as those of their suprema. This is done in Section 4, which also includes an analysis of the relation between the sub-nth chaos property and boundedness of the nth Malliavin derivative. Our proofs in Section 4 are inspired by some of the techniques that worked well in the sub-Gaussian case of Section 3; yet when $n \geq 3$, many technical difficulties arise, and our work opens up as many new problems as it solves in that case.

While we prefer to provide full statements of our results in the main body of this paper, we include here some typical consequences of our work under a simplifying assumption which is nonetheless relevant for some applications, leaving it to the reader to check that the results now given do follow from our theorems.

Assumption. Let n be a positive integer. Let X be a centered separable random field on an index set I. Assume that there exists a non-random metric δ on $I \times I$ such that almost surely, for all $x, y \in I$, for all $0 \leq s_n \leq \cdots \leq s_2 \leq s_1 \leq 1$,

$$|D_{s_n} \cdots D_{s_2} D_{s_1} (X(x) - X(y))| \leq \delta(x, y). \tag{1.1}$$

Conclusions. Let $N(\varepsilon)$ be the smallest number of balls of radius ε in the metric δ needed to cover I. There is a constant C_n depending only on n such that, if the assumption above holds, the following conclusions hold:

Sub-nth Gaussian chaos property: (see Theorem 4.7)

$$\mathbf{E}\left[\exp\left(\frac{1}{C_n}\left|\frac{X(x)-X(y)}{\delta(x,y)}\right|^{2/n}\right)\right] \leq 2;$$

Sub-nth Gaussian chaos extension of the Dudley upper bound: (see Theorem 4.5)

$$\mu := \mathbf{E}\left[\sup_{x\in I} X(x)\right] \leq C_n \int_0^\infty (\log N(\varepsilon))^{n/2}\, d\varepsilon;$$

Sub-nth Gaussian chaos extension of the Borell-Sudakov concentration inequality: (see Corollary 4.14) With

$$\sigma = \operatorname*{ess\,sup}_{\omega\in\Omega} \{\sup |D_{s_n} \cdots D_{s_2} D_{s_1} X(x)| : x \in I;$$
$$0 \leq s_n \leq \cdots \leq s_2 \leq s_1 \leq 1\},$$

for all $\varepsilon > 0$, for u large enough,

$$\mathbf{P}\left[\left|\sup_{x\in I} X(x) - \mu\right| > u\right] \leq 2(1+\varepsilon)\exp\left(-\frac{1}{(1+\varepsilon)}\left(\frac{u}{\sigma}\right)^{2/n}\right).$$

It should be noted that in the sub-2nd-Gaussian chaos case ($n = 2$), we prove (Theorem 4.5, Theorem 4.7 case $n = 2$, Corollary 4.11) the three "Conclusions" above hold under the considerably weaker condition: almost surely,

$$\int\cdots\int_{[0,1]^n} |D_{s_n}\cdots D_{s_2} D_{s_1}(X(x) - X(y))|^2\, ds_1 ds_2 \cdots ds_n \leq \delta^2(x,y). \tag{1.2}$$

When $n \geq 3$, the conditions we need to draw the above conclusions are intermediate between (1.1) and (1.2). However, we conjecture that the conclusions should hold under conditions much closer to (1.2). When $n = 1$, the Dudley-Fernique theorem has been known for many years (see [11]) if one assumes the conclusion of Lemma 3.3; our interpretation of this Lemma appears to be new, although its proof below clearly shows it is a translation of Ustunel's [23, Theorem 9.1.1]; however, our proof of the Borell-Sudakov inequality (Theorem 3.6) under the hypotheses of Lemma 3.3 is new, and the inequality itself might be new for any class of non-Gaussian processes insofar as it does not seem to appear in the literature.

In addition to the obvious practical significance of results such as the "Conclusions" above, we think the reader familiar with classical proofs of such results as the Borell-Sudakov inequality and the Dudley-Fernique theorem, will appreciate the power of Malliavin derivatives: they provide, in Section 3 ($n = 1$), stronger results with elegant, simpler proofs. We hope that beyond the issue of sharpening the results in Section 4 ($n \geq 3$) to come closer to Condition (1.2), this paper will encourage the reader to use our Malliavin-derivative based concentration inequalities in sub-Gaussian and non-sub-Gaussian settings, such as to study the almost-sure moduli of continuity of random fields to extend classical results (see [1] or [21]).

We wish to thank the three organizers of the Fifth Seminar on Stochastic Analysis, Random Fields and Applications (Ascona, Switzerland, 2005) for providing the impetus for the research which led to this paper.

2. Preliminairies

In this didactic section, we present some basic facts about Wiener chaoses and the Malliavin calculus, largely with only sketches of proofs, to be used in the remainder of the article, and as a general quick reference guide. Excellent and complete treatment of these results and many more can be found for instance in the monographs [17] and [23]; both have been a constant source of inspiration for us.

We begin with a Brownian motion $W = \{W(t) : t \in [0,1\}$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and adapted to a filtration $(\mathcal{F}_t)_{t\in[0,1]}$ satisfying the usual conditions (see [9]). With dr representing the Lebesgue measure, the Wiener integral $W(f) = \int_0^1 f(r)\, dW(r)$ of a non-random $f \in \mathcal{H} := L^2([0,1], dr)$ is a centered Gaussian random variable with variance $||f||^2_{\mathcal{H}} = \int_0^1 f^2(r)\, dr$; the set $\mathcal{H}_1$ of all Wiener integrals $W(f)$ when f ranges over all of $\mathcal{H}$ is a set of jointly Gaussian random variables called the first Wiener chaos of W, or Gaussian space of W, whose entire finite-dimensional distributions are thus defined via the formula $\mathbf{E}W(f)W(g) = \langle f; g\rangle_{\mathcal{H}} = \int_0^1 f(r)g(r)\, dr$. The Wiener integral coincides with the Itô integral on $\mathcal{H}_1$, which can be seen via several different procedures, including the fact that both can be approximated in $L^2(\Omega)$ by the same Riemann sums. To construct chaoses of higher order, one may for example use iterated Itô integration. Denote $I_0(f) = f$ for any non-random constant f. Assume by induction that for any $g \in \mathcal{H}^{\otimes n}$, for almost every $(t,\omega) \in L^2([0,1]\times\Omega, dr d\mathbf{P})$,

$$I_n(g) = n! \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} g(s_1, s_2, \dots, s_n)\, dW(s_n) \cdots dW(s_2)\, dW(s_1) \quad (2.1)$$

has been defined. Given a symmetric function $f \in \mathcal{H}^{\otimes n+1}$, let

$$g_t(s_1, s_2, \dots, s_n) = f(t, s_1, s_2, \dots, s_n)\, \mathbf{1}_{s_1 \le t}.$$

We thus see that the function $t \mapsto I_n(g_t)$ is a square-integrable $(\mathcal{F}_t)_{t\in[0,1]}$-martingale. We may then define $I_{n+1}(f)$ to be the Itô integral $(n+1)\int_0^1 I_n(g_t)\, dW(t)$. The set $\mathcal{H}_{n+1}$ spanned by $I_{n+1}(f)$ for all symmetric f in $\mathcal{H}^{\otimes n+1}$ is the $(n+1)$-th Wiener chaos of W.

Remark 2.1. It holds that $L^2(\Omega)$ is the direct sum – with respect to the inner product defined by expectations of products of r.v.'s – of all the Wiener chaoses. Specifically for any $X \in L^2(\Omega)$, there exists a sequence of non-random symmetric functions $f_n \in \mathcal{H}^{\otimes n} = L^2([0,1]^n)$ with $\sum_{n=0}^{\infty} |f_n|^2_{\mathcal{H}^{\otimes n}} < \infty$ such that $X = \sum_{n=0}^{\infty} I_n(f_n)$; moreover $\mathbf{E}[I_n(f_n) I_m(f_m)] = \delta_{m,n} n! |f_n|^2_{\mathcal{H}^{\otimes n}}$ where $\delta_{m,n}$ equals 0 if $m \neq n$ and 1 if $m = n$.

Remark 2.2 (see [17]). The n-th Wiener chaos $\mathcal{H}_n = I_n\left(\mathcal{H}^{\otimes n}\right)$ coincides with the closed linear subspace of $L^2(\Omega)$ generated by all the random variables of the form $H_n(W(h))$ where $h \in \mathcal{H}$, $|h|_{\mathcal{H}} = 1$, and H_n is the n-th Hermite polynomial, defined by $H_0 \equiv 1$, $H_1(x) = x$, and $H_{n+1}(x) = (n+1)^{-1}(xH_n(x) - H_{n-1}(x))$. Moreover, $H_n' = H_{n-1}$.

We believe the easiest way to understand the Malliavin derivative operator is using the following three-step "constructive" presentation; in fact, the essence of the construction of this operator only requires steps 1 and 2(a), as one can arguably see from step 3.

1. We define an operator D from $\mathcal{H}_1$ into $\mathcal{H}$ by the formula
$$D_r W(f) = f(r).$$
Thus the Malliavin derivative finds the integrand which a centered Gaussian r.v. in $\mathcal{H}_1$ is formed from as a Wiener integral. If $X = W(f) + \mu$ where μ is non-random, $D.X = f$, consistent with the fact that the derivative is linear and kills constants.
2. We extend D by a consistency with the chain rule.
   (a) For any m-dimensional Gaussian vector $G = (G_i)_{i=1}^m \in (\mathcal{H}_1)^m$, for any $\Phi \in C^1(\mathbf{R}^m)$ such that $X = \Phi(G) \in L^2(\Omega)$, in order to be consistent with the appellation "derivative", one must set
$$D_r X = \sum_{i=1}^m \frac{\partial \Phi}{\partial g_i}(G)\, D_r G_i = \nabla \Phi(G) \cdot D_r G; \tag{2.2}$$
that is to say, the chain rule must hold. It is a simple matter to check that the above requirement (2.2) can be satisfied for all X of this form, defining D uniquely on them.
   (b) Equivalently, by the chain rule in $C^1(\mathbf{R}^n)$, one can state that formula (2.2) holds for all Y of the form $Y = \Psi(X_1, \ldots, X_n)$ with $\Psi \in C^1(\mathbf{R}^n)$ and all X_i's as in part 2.a, if we replace $D_r G$ by $D_r X$: $D_r Y = \nabla \Psi(X) \cdot D_r X$ holds for any X, Y and Ψ such that the right-hand side is in $L^2(\Omega)$.
3. The following argument can now be used to define D on a much larger set of random variables. For a fixed random variable $Z \in L^2(\Omega)$, we consider the orthogonal chaos decomposition $Z = \sum_{n=0}^\infty I_n(f_n)$ of Remark 2.1. From Remark 2.2, $I_n(f_n)$ can be further approximated in $L^2(\Omega)$: $I_n(f_n) = \sum_{j=1}^\infty X_j$ where $X_j = H_n(W(h_j))$ where H_n is the nth Hermite polynomial and $h_j \in \mathcal{H}$. By step 2.a, $D_r X_j$ is defined for almost all r, as it is trivial to see that $D_r X_j \in L^2(\Omega)$ for any r such that $h_j(r)$ is finite. More to the point, since $h_j \in \mathcal{H}$, we can say that $D.X_j \in L^2(\Omega) \times \mathcal{H}$. We now need to have a criterion that allows us to justify that $D.I_n(f_n)$ exists in the same space $L^2(\Omega) \times \mathcal{H}$ as a limit in that space of the sums of all the Malliavin derivatives $D.X_j$. It turns out that no additional criterion is needed beyond the fact that the symmetric f_n is in $\mathcal{H}^{\otimes n}$. Indeed, using the relation $H_n' = H_{n-1}$, one

proves that the series $\sum_j D_r X_j$ converges to $nI_{n-1}(f_n(\cdot, r))$ in $L^2(\Omega) \times \mathcal{H}$. To complete the program of defining $D.Z$ on as wide a space of Z's as possible, since from Remark 2.1 we have $\int_0^1 \mathbf{E}\left|nI_{n-1}(f_n(\cdot, r))\right|^2 dr = nn! \left|f_n\right|^2_{\mathcal{H}^{\otimes n}}$, we immediately get that $D.Z$ exists in $L^2(\Omega) \times \mathcal{H}$ and has orthogonal decomposition in that space given by

$$D_r Z = \sum_{n=1}^{\infty} nI_{n-1}(f_n(\cdot, r))$$

as soon as

$$\sum_{n=1}^{\infty} nn! \left|f_n\right|^2_{\mathcal{H}^{\otimes n}} < \infty. \tag{2.3}$$

Remark 2.3. The set of all $Z \in L^2(\Omega)$ such that (2.3) holds is called the (Gross-)Sobolev space $\mathbf{D}^{1,2}$ with respect to W and its Malliavin derivative. It is a Hilbert space with respect to the inner product $\langle Z, Z' \rangle = \mathbf{E}[ZZ'] + \int_0^1 \mathbf{E}[D_r Z D_r Z'] \, dr$.

Remark 2.4 (General Chain Rule for Malliavin derivatives). Combining relation (2.2) from Step 2a and Step 3 above, for any $Z \in \left(\mathbf{D}^{1,2}\right)^m$, for any $\Phi \in C^1(\mathbf{R}^m)$ such that $\nabla\Phi(Z) \in L^2(\Omega)$, we get $\Phi(Z) \in \mathbf{D}^{1,2}$ and the general chain rule formula

$$D_r(\Phi(Z)) = \nabla\Phi(Z) \cdot D_r Z. \tag{2.4}$$

3. Sub-Gaussian theory

In this section we develop the concept of sub-Gaussian random variables and processes/fields (a stochastic process defined on an index set that is not a subset of $\mathbf{R}_+$ is normally called a *random field*). We define sufficient Malliavin derivative conditions implying these concepts, and we investigate extensions of the familiar concentration inequalities known as the Dudley-Fernique theorems (on the expected supremum of a process) and the Borell-Sudakov inequalities (on the deviation from this expectation).

Definition 3.1. A centered random variable X is said to be sub-Gaussian relative to the scale σ if for all $\lambda > 0$,

$$\mathbf{E}[\exp \lambda X] \leq \exp \lambda^2 \sigma^2 / 2. \tag{3.1}$$

Remark 3.2. The interpretation of σ^2 above is that of an upper bound on X's variance. More specifically, the following two statements imply (3.1) and are implied by it, with different universal constants c in each implication:

$$\mathbf{E}\left[\exp\left(X^2 / \left(c\sigma^2\right)\right)\right] \leq 2, \tag{3.2}$$

and for all $u > 0$,

$$\mathbf{P}[|X| > u] \leq 2 \exp\left(-\frac{u^2}{2c\sigma^2}\right).$$

For instance, (3.1) implies (3.2) with $c = 5$. Consult Lemma 4.6 for more general results than these implications, and their proofs.

We will use the following fundamental lemma, whose consequences are far-reaching.

Lemma 3.3. *Let X be a centered random variable in $\mathbf{D}^{1,2}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ of the previous section. Assume there exists a non-random constant M such that, $\mathbf{P}$-almost surely,*

$$\int_0^1 |D_r X|^2 \, dr \le M^2. \tag{3.3}$$

Then X is sub-Gaussian relative to $\sigma = M$.

Proof. The following result is due to Üstünel [23, Theorem 9.1.1]: if (3.3) holds, then $\mathbf{P}[|X| > u] \le 2 \exp\left(-u^2/(2M^2)\right)$. The lemma is thus just a translation of this theorem using the definition of sub-Gaussian random variables. □

In the previous section, we saw that in $(\Omega, \mathcal{F}, \mathbf{P})$ a Gaussian random variable is one such that its Malliavin derivative is non-random. The above lemma states that a class of sub-Gaussian centered random variables is obtained by requiring only that their Malliavin derivatives have an almost-surely bounded norm in $\mathcal{H} = L^2[0,1]$. The reader will check that, equivalently, condition (3.3) says that $D.X \in L^\infty(\Omega, \mathcal{H})$, and ess sup$|D.X|^2_{\mathcal{H}}$ is the smallest $M > 0$ satisfying (3.3) almost surely.

Definition 3.4. A pseudo-metric is a symmetric function δ on $I \times I$ such that $\delta(s,u) \le \delta(s,t) + \delta(t,u)$.

The axiom $\delta(s,t) = 0 \Longrightarrow s = t$ need not hold for pseudo-metrics. Examples of pseudo-metrics are the canonical metrics δ_Z of all centered Gaussian fields Z on I: $\delta_Z(s,t) := \sqrt{\mathbf{E}\left[(Z(t) - Z(s))^2\right]}$.

Definition 3.5. A centered process (random field) X on an arbitrary index set I is said to be sub-Gaussian relative to the pseudo-metric δ on I if for any $s, t \in I$, the random variable $X(t) - X(s)$ is sub-Gaussian relative to the scale $\sigma = \delta(s,t)$.

Our first theorem is the extension to the class of sub-Gaussian processes defined via condition (3.3) of the so-called Borell-Sudakov inequality. The classical version of this inequality states that for a centered separable Gaussian field on an index set I, if $\mu := \mathbf{E}\sup_I X < \infty$, then $\mathbf{P}[|\sup_I X - \mu| > u] \le 2\exp\left(-u^2/\left(2\sigma^2\right)\right)$ where $\sigma^2 = \sup_{t \in I} \mathrm{Var}[X(t)]$.

Theorem 3.6. *Let X be a separable random field on I such that all finite-dimensional vectors of X are formed of almost-surely distinct components. Assume $\mu := \mathbf{E}[\sup_I X] < \infty$. Assume for each $t \in I$, $X(t) \in \mathbf{D}^{1,2}$, and there exist a constant $\sigma^2(t)$ such that almost surely*

$$\int_0^1 |D_r X(t)|^2 \, dr \le \sigma^2(t).$$

Then the random variable $\sup_I X - \mu$ *is sub-Gaussian relative to* $\sigma^2 = \sup_{t\in I} \sigma^2(t)$. *In other words*

$$\mathbf{P}\left[\left|\sup_I X - \mu\right| > u\right] \leq 2\exp\left(-\frac{u^2}{2\sigma^2}\right).$$

Proof. Step 1*: Setup.* Separability of X means that its distribution only requires knowledge of X on a countable subset of I, i.e., we can assume I is countable in the expression $\sup_I X$. Hence, by the dominated convergence theorem, the problem reduces to the case of finite I. Thus we assume $I = \{1, 2, \ldots, N\}$ where N is a positive integer and $X = \{X_1, X_2, \ldots, X_N\}$. Now let

$$s_n = \max\{\sigma(1), \sigma(2), \ldots, \sigma(n)\},$$

and

$$S_n = \max\{X_1, X_2, \ldots, X_N\}.$$

Since $\Phi(x, y) = \max(x, y) = x\mathbf{1}_{x\geq y} + y\mathbf{1}_{x<y}$, thus we have $S_{n+1} = \Phi(X_{n+1}, S_n)$ where $\partial\Phi/\partial x(x, y) = \mathbf{1}_{x\geq y}$ and $\partial\Phi/\partial y(x, y) = \mathbf{1}_{x\leq y}$.

Step 2*: Explicit extension of the chain rule.* Unfortunately Φ is not of class C^1, so to keep our proof rigorous, since we will need to use the chain rule formula (2.4) with Φ, we indicate how to extend it for our purposes. We claim the following.

Lemma 3.7. *The chain rule* (2.4) *holds with* Z *any vector of random variables in* $\mathbf{D}^{1,2}$, *for any* Φ *that is of class* C^1 *off of a finite union* T *of hyperplanes, with* $\nabla\Phi$ *bounded, with* $\Phi(Z) \in \mathbf{D}^{1,2}$, *and with* $Z \notin T$ *almost surely.*

See the appendix for a proof of this result which is spelled out for the situation we need.

Step 3*: Induction.* We prove the theorem by induction on n. Our induction hypothesis need only be that $S_n \in \mathbf{D}^{1,2}$ and almost surely,

$$\int_0^1 |D_r S_n|^2 \, dr \leq s_n^2. \tag{3.4}$$

Indeed, this inequality is satisfied with $n = 1$ by hypothesis since $S_1 = X_1$; when $n = N$, Lemma 3.3 applied to $S_N = \sup_I X$ proves that this induction hypothesis implies the statement of the theorem. Therefore, we only need to prove that if $S_n \in \mathbf{D}^{1,2}$ and (3.4) holds for some $n \in \{1, \ldots, N-1\}$, then $S_{n+1} \in \mathbf{D}^{1,2}$ and (3.4) holds for $n+1$. Since $S_{n+1} = \Phi(X_{n+1}, S_n)$, and by hypothesis $X_{n+1} \neq S_n$ almost surely, we can apply the above lemma: for almost every $r \in [0, 1]$,

$$\begin{aligned} D_r S_{n+1} &= \frac{\partial\Phi}{\partial x}(X_{n+1}, S_n) D_r X_{n+1} + \frac{\partial\Phi}{\partial y}(X_{n+1}, S_n) D_r S_n \\ &= \mathbf{1}_{X_{n+1}\geq S_n} D_r X_{n+1} + \mathbf{1}_{X_{n+1}\leq S_n} D_r S_n \\ &= \mathbf{1}_{X_{n+1}> S_n} D_r X_{n+1} + \mathbf{1}_{X_{n+1}< S_n} D_r S_n. \end{aligned}$$

The last equality holds a.s. again because the X_i's are distinct almost surely. Therefore, since the product of the two terms in the last line above is zero, using the induction hypothesis (3.4) and the assumption $\|D.X_{n+1}\|^2_{L^2[0,1]} \le \sigma^2 (n+1)$, we obtain

$$\begin{aligned}\int_0^1 |D_r S_{n+1}|^2 \, dr &= \mathbf{1}_{X_{n+1}\ge S_n} \int_0^1 |D_r X_{n+1}|^2 \, dr + \mathbf{1}_{X_{n+1}<S_n} \int_0^1 |D_r S_n|^2 \, dr \\ &\le \sigma^2 (n+1) \mathbf{1}_{X_{n+1}\ge S_n} + s_n^2 \mathbf{1}_{X_{n+1}<S_n} \\ &\le s_{n+1}^2 \mathbf{1}_{X_{n+1}\ge S_n} + s_{n+1}^2 \mathbf{1}_{X_{n+1}<S_n} \\ &= s_{n+1}^2 .\end{aligned}$$

By induction, the proof of the theorem is complete. □

Remark 3.8. The assumption in the previous theorem that any vector of X's have almost surely distinct components can be easily satisfied using a now classical result on the existence of densities of random vectors. From [17, Theorem 2.1.2] we learn that we only need to check that the matrix of Malliavin derivatives' inner products $\left(\langle D.X_n; D.X_{n'}\rangle\right)_{n,n'=1}^N$ is almost surely invertible, since this implies that the law of X has a density. Thus the theorem's two assumptions can be phrased in terms of Malliavin derivatives, one as a boundedness condition, the other as a non-degeneracy condition. The latter is of course much weaker than the former.

The only assumption on the non-diagonal correlations of X in the above theorem is the finiteness of μ, which has evidently little or nothing to do with the sub-Gaussian property at the process level. The main Malliavin derivative boundedness hypothesis is only a set of one-dimensional distributional hypotheses, which represents a significant improvement over assuming that the entire vector X is jointly Gaussian.

When comparing Lemma 3.3 and Theorem 3.6, one may wonder whether the Borell-Sudakov inequality holds under the weaker hypothesis that each $X(t)$ is sub-Gaussian. This represents a gap which we are not able to fill at this time. It is instructive to note that the main issue here is that the converse of Lemma 3.3 is false: if the r.v. X is sub-Gaussian relative to the scale σ, it does not imply that (3.3) holds for the scale σ, or even for any other scale. To see this, consider a random variable $X = \int_0^1 u(s)\, dW(s)$ where u is adapted to $(\mathcal{F}_s)_{s\in[0,1]}$ and is such that $\int_0^1 u^2(s)\, ds$ is almost surely bounded by σ^2. For any λ, using the exponential martingale $\mathcal{E}(\lambda M)_t$ based on the martingale $M_t = \int_0^t u(s)\, dW(s)$, we immediately get

$$\begin{aligned}\mathbf{E}[\exp \lambda X] &= \mathbf{E}\left[\mathcal{E}(\lambda M)_1 \exp \frac{\lambda^2}{2} \int_0^1 u^2(s)\, ds\right] \\ &\le \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \mathbf{E}\left[\mathcal{E}(\lambda M)_1\right] = \exp\left(\frac{\lambda^2\sigma^2}{2}\right),\end{aligned}$$

which means X is sub-Gaussian relative to the scale σ. But for the specific case of $u(s) = f(W_s)$ where f is bounded and in $C^2(\mathbf{R})$, we can easily find examples of f where (3.3) does not hold for any scale. Using the formula $D_r X = u(r) + \int_r^1 D_r u(s)\, dW(s)$ (see [17]), which in our example yields $D_r X = f(W_r) + \int_r^1 D_r W_s f'(W_s)\, dW_s$, and then using Itô's formula, we get

$$\int_0^1 |D_r X|^2\, dr = \int_0^1 \left| f(W_r) + \int_r^1 f'(W_s)\, dW(s) \right|^2 dr$$

$$= \int_0^1 \left| f(W_1) - \frac{1}{2} \int_r^1 f''(W_s)\, ds \right|^2 dr \tag{3.5}$$

$$= f^2(W_1) - f(W_1) \int_0^1 s f''(W_s)\, ds + \frac{1}{4} \int_0^1 \int_0^1 f''(W_s) f''(W_{s'}) \min(s, s')\, ds ds'. \tag{3.6}$$

Even if f is bounded, it is simple to construct examples where f'' is not bounded: e.g., $f(w) = \sin(w^2)$, with $f''(w) = 2\cos(w^2) - 4w^2 \sin(w^2)$. Whether in line (3.5) or line (3.6), we see that the expression above can take arbitrarily large values with positive probability.

In order to use the Borell-Sudakov inequality efficiently, it is necessary to be able to estimate the expected supremum effectively. We recall here the classical results of Dudley (upper bound) and Fernique (lower bound) for Gaussian processes.

Theorem 3.9. *Let X be a separable Gaussian field on an index set I. Let $\delta_X(s,t) = \left(\mathbf{E}\left[(X(s) - X(t))^2\right]\right)^{1/2}$ be its canonical metric. Let the* metric entropy $N(\varepsilon)$ *be the smallest number of balls of radius ε in the pseudo-metric δ_X needed to cover I. There exist two positive universal constants K and K' such that*

$$\mathbf{E}\left[\sup_I X\right] \le K \int_0^\infty \sqrt{\log N(\varepsilon)} d\varepsilon, \tag{3.7}$$

and, if I is a subset of a group G and the law of X, defined on G, is translation invariant (e.g., $I \subset \mathbf{R}^d$ and $\delta(s,t)$ depends only on $|s-t|$, i.e., X is homogeneous *or* stationary*),*

$$\mathbf{E}\left[\sup_I X\right] \ge K' \int_0^\infty \sqrt{\log N(\varepsilon)} d\varepsilon.$$

For what classes of processes does a result of the same type as the lower bound above hold? This is an open problem which we will not tackle in this paper. Yet the Dudley upper bound (3.7) of this theorem is true, with the same $N(\varepsilon)$, for all processes which are sub-Gaussian relative to the same pseudo-metric δ_Z. This result even extends beyond the sub-Gaussian case, as we are about to see in the next section, which is why we omit the proof that Theorem 3.9 holds for sub-Gaussian processes. Another reason for omiting the proof is that the result is now classical (see [11]). For the sake of completeness, we still record the statement here.

Remark 3.10. If X is sub-Gaussian on I, as in Definition 3.5, relative to the pseudo-metric δ, then with the notation of Theorem 3.9, (3.7) holds.

4. Sub-nth chaos processes

One of the difficulties with Wiener chaos expansions such as $X = \sum_{n=0}^{\infty} I_n(f_n)$ (defined in Remark 2.1) is that they often mask fundamental properties of processes. In particular, a typical sub-Gaussian random variable has components of all orders in its chaos expansion, so that any estimation done term by term using this expansion will miss the sub-Gaussian property, while the entire sum of the expansion, being sub-Gaussian, is thus more akin to its term of order $n = 1$. In this section we introduce a concept which generalizes this idea to higher values of n. We use it to derive a Dudley-type theorem (Subsection 4.1). Then we attempt to relate the concept to iterated Malliavin derivative calculations (Subsection 4.2), and derive an extension of the Borell-Sudakov concentration inequality as a consequence (Subsection 4.3).

Definition 4.1. Let n be a positive integer. A centered random variable X is said to have the sub-nth-Gaussian chaos property (or is a sub-nth chaos r.v., or is a sub-Gaussian chaos r.v. of order n, etc.) relative to the scale M if

$$\mathbf{E}\left[\exp\left(\left(\frac{X}{M}\right)^{2/n}\right)\right] \leq 2.$$

Obviously, when $n = 1$, such an X is sub-Gaussian relative to the scale $\sqrt{5}M$. Our definition is similar to the definition of an Orlicz norm of X, although the only intersection between the concepts appears to occur for $n = 1$ or 2, since Orlicz norms have a requirement of convexity of their Young function, which is not the case here for $n > 2$ (see [11], or [20]).

Remark 4.2. From Definition 3.1 and Remark 3.2, we get the following equivalent definitions of the sub-nth-Gaussian chaos property, up to universal multiplicative scale constants c: for all $\lambda, u > 0$,

$$\mathbf{E}\left[\exp\lambda\left(|X|^{1/n} - \mathbf{E}\,|X|^{1/n}\right)\right] \leq \exp c\lambda^2 M^2/2$$

and

$$\mathbf{P}\left[|X|^{1/n} > u\right] \leq 2\exp\left(-\frac{u^2}{2cM^2}\right).$$

Definition 4.3. Let δ be a pseudo-metric on a set I. A centered random field X on I is said to be a sub-nth-Gaussian chaos field with respect to δ if for any $s, t \in I$, the random variable $X(t) - X(s)$ has the sub-nth-Gaussian chaos property relative to the scale $\delta(s,t)$.

Definition 4.4. Let δ and X be as in the previous definition. We use the notation N_δ, and we say that N_δ is a metric entropy for X, if N_δ is the smallest number of balls of radius ε in the pseudo-metric δ needed to cover I.

4.1. Expected suprema

As announced in the previous section, we now prove a Dudley upper bound for sub-nth chaos processes.

Theorem 4.5. *For each fixed positive integer n, there exists a universal constant C_n depending only on n such that if X defined on I is a separable sub-nth-Gaussian chaos field with respect to the pseudo-metric δ, then with N_δ a metric entropy for X,*

$$\mathbf{E}\sup_{t\in I} X(t) \le C_n \int_0^\infty (\log N_\delta(\varepsilon))^{n/2} d\varepsilon.$$

This theorem is a new result for $n > 2$; it has been established in [24] for $n \le 2$ using convexity of the Orlicz space's Young function. Our proof of this theorem, which works for any integer $n \ge 1$, requires the first two inequalities of the following lemma, which is established in the appendix.

Lemma 4.6. *For every integer n, there exists a universal constant v_n such that, for any sub-nth chaos r.v. X relative to the scale δ, the following inequalities hold: for every $u > 0$,*

$$\mathbf{P}[|X| > u] \le 2\exp\left(-\left(\frac{u}{\delta}\right)^{2/n}\right),$$

$$\mathbf{E}\left[X^2\right] \le (v_n\delta)^2,$$

and for every $\lambda > 0$,

$$\mathbf{E}\left[\exp\left(\lambda\left|\frac{X}{v_n\delta}\right|^{1/n}\right)\right] \le v_n e^{\lambda^2/2}.$$

The converse also holds. Namely, with possibly some other universal constant $v_n' > 1$, each of the three inequalities above implies that X is a sub-nth chaos r.v. relative to the scale $M = v_n'\delta$.

Proof of Theorem 4.5. Our proof is patterned from Michel Ledoux's notes [10] on "Isoperimetry and Gaussian Analysis", although here no Young function convexity is used, and indeed we do not have the restriction $n \le 2$. We may and do assume that I is finite (see Step 1 of proof of Theorem 3.6). If the right-hand side of the conclusion of the theorem is infinite, there is nothing to prove. Therefore we may assume that $\sup_I X$ is integrable.

Step 1*: Chaining argument.* Let $q > 1$ be fixed and let ℓ_0 be the largest integer ℓ in $\mathbb{Z}$ such that $N_\delta(q^{-\ell}) = 1$. For every $\ell \ge \ell_0$, we consider a family of cardinality $N(\ell) := N_\delta(q^{-\ell})$ of balls of radius $q^{-\ell}$ covering I. One may therefore construct a partition $\mathcal{A}_\ell$ of I of cardinality $N(\ell)$ on the basis of this covering with sets of diameter less than $2q^{-\ell}$. In each A of $\mathcal{A}_\ell$, fix a point of I and denote by I_ℓ the collection of these points. For each t in I, denote by $A_\ell(t)$ the element of $\mathcal{A}_\ell$ that contains t. For every t and every ℓ, let then $s_\ell(t)$ be the element of I_ℓ such that $t \in A_\ell(s_\ell(t))$. Note that $\delta(t, s_\ell(t)) \le 2q^{-\ell}$ for every t and $\ell \ge \ell_0$. Also note that

$$\delta(s_\ell(t), s_{\ell-1}(t)) \le 2q^{-\ell} + 2q^{-\ell+1} = 2(q+1)q^{-\ell}.$$

Hence, by the second inequality in the previous lemma, the series $\sum_{\ell>\ell_0}\left(X_{s_\ell(t)}-X_{s_{\ell-1}(t)}\right)$ converges in $L^1(\Omega)$, and also $s_\ell(t)$ converges to t in $L^1(\Omega)$ as $\ell\to\infty$. By the telescoping property of the the above sum, we thus get that almost surely for every t,

$$X_t = X_{s_0} + \sum_{\ell>\ell_0}\left(X_{s_\ell(t)}-X_{s_{\ell-1}(t)}\right) \tag{4.1}$$

where $s_{\ell_0}(t) := s_0$ may be chosen independent of $t\in I$.

Step 2: Applying the lemma. Let c_ℓ be a constant that will be chosen in the next step. It follows from the decomposition (4.1) above, and the identity $\mathbf{E}X_{s_0}=0$, that

$$\begin{aligned}
&\mathbf{E}\left(\sup_{t\in I} X_t\right)\\
&= \mathbf{E}\left[X_{s_0} + \sup_{t\in I}\sum_{\ell>\ell_0}\left(X_{s_\ell(t)}-X_{s_{\ell-1}(t)}\right)\right]\\
&\le \sum_{\ell>\ell_0} c_\ell + \mathbf{E}\left(\sup_{t\in I}\sum_{\ell>\ell_0}\left|X_{s_\ell(t)}-X_{s_{\ell-1}(t)}\right| 1_{\{|X_{s_\ell(t)}-X_{s_{\ell-1}(t)}|>c_\ell\}}\right)\\
&\le \sum_{\ell>\ell_0} c_\ell + \mathbf{E}\left(\sum_{\ell>\ell_0}\sum_{(u,v)\in H_\ell}\left|X_u-X_v\right| 1_{\{|X_u-X_v|>c_\ell\}}\right)\\
&\le \sum_{\ell>\ell_0} c_\ell + \sum_{\ell>\ell_0}\sum_{(u,v)\in H_\ell}\mathbf{E}\left(\left|X_u-X_v\right| 1_{\{|X_u-X_v|>c_\ell\}}\right)
\end{aligned}$$

where $H_\ell=\{(u,v)\in I_\ell\times I_{\ell-1};\delta(u,v)\le 2(q+1)q^{-\ell}\}$. Using Holder's inequality, we get

$$\mathbf{E}\left(\sup_{t\in I} X_t\right) \le \sum_{\ell>\ell_0} c_\ell + \sum_{\ell>\ell_0}\sum_{(u,v)\in H_\ell}\left(\mathbf{E}|X_u-X_v|^2\right)^{1/2}\left(\mathbf{P}(|X_u-X_v|>c_\ell)\right)^{1/2}.$$

Using Lemma 4.6 now, and applying a uniform upper bound for all $(u,v)\in H_\ell$, we get

$$\begin{aligned}
&\mathbf{E}\left(\sup_{t\in I} X_t\right)\\
&\le \sum_{\ell>\ell_0} c_\ell + \sum_{\ell>\ell_0}\sum_{(u,v)\in H_\ell} v_n\delta(u,v)\left(2\exp\left(-\left(\frac{c_\ell}{\delta(u,v)}\right)^{2/n}\right)\right)^{1/2}\\
&\le \sum_{\ell>\ell_0} c_\ell + \sum_{\ell>\ell_0} v_n \mathrm{Card}(H_\ell)2(q+1)q^{-\ell}\left(2\exp\left(-\left(\frac{c_\ell}{2(q+1)q^{-\ell}}\right)^{2/n}\right)\right)^{1/2}.
\end{aligned}$$

Step 3: Choosing c_ℓ. Since $\mathrm{Card}(H_\ell)\le N(\ell)^2$, it is now apparent that a convenient choice for c_ℓ, in order to exploit the summability of $q^{-\ell}$ without having to worry about the size of Card(H_ℓ), is $c_\ell = 2(q+1)q^{-\ell}(4\log N(\ell))^{n/2}$. We thus obtain

$$\begin{aligned}\mathbf{E}\left(\sup_{t\in I} X_t\right) &\le \sum_{\ell>\ell_0} c_\ell + \sum_{\ell>\ell_0} N(\ell)^2\, 2^{3/2}(q+1)q^{-\ell}v_n \exp(-2\log N(\ell)) \\ &\le \sum_{\ell>\ell_0} 2(q+1)q^{-\ell}(4\log N(\ell))^{n/2} + \sum_{\ell>\ell_0} 2^{3/2}(q+1)q^{-\ell}v_n.\end{aligned}$$

Step 4: Conclusion. Now, since for $\ell > \ell_0$, $\log N(\ell) \ge \log 2$, then $(\log N(\ell))^{n/2} \ge (\log 2)^{n/2}$ for $n \ge 1$. It follows that

$$\mathbf{E}\left(\sup_{t\in I} X_t\right) \le k_n(q+1)\sum_{\ell>\ell_0} q^{-\ell}(\log N(\ell))^{n/2}$$

where $k_n = 2\cdot 4^{n/2} + 2^{3/2}v_n \log^{-n/2} 2$. By comparing our series to an integral, since N_δ is decreasing, we get

$$\begin{aligned}\mathbf{E}\left(\sup_{t\in I} X_t\right) &\le k_n\left(\frac{q+1}{1-q^{-1}}\right)(1-q^{-1})\sum_{\ell>\ell_0} q^{-\ell}(\log N(\ell))^{n/2} \\ &\le k_n\left(\frac{q+1}{1-q^{-1}}\right)\sum_{\ell>\ell_0}\int_{q^{-\ell-1}}^{q^{-\ell}} (\log N_\delta(\varepsilon))^{n/2} d\varepsilon \\ &\le C_n \int_0^\infty (\log N_\delta(\varepsilon))^{n/2} d\varepsilon\end{aligned}$$

where $C_n = k_n\left(\frac{q(q+1)}{q-1}\right)$. The theorem is proved with $C_n = \left(2\sqrt{2}+3\right)k_n$. □

4.2. Malliavin derivative conditions

A connection between the above definition of sub-nth chaos r.v.'s and Malliavin derivatives is provided by the following.

Theorem 4.7. *Let X be a random variable in $\mathbf{D}^{n,2}$. That is to say, X has n iterated Malliavin derivatives, and the nth derivative*

$$D^{(n)}_{s_n,\dots,s_2,s_1}X = D_{s_n}\left(D_{s_{n-1}}\left(\cdots D_{s_2}\left(D_{s_1}X\right)\cdots\right)\right)$$

is a member of $L^2\left(\Omega\times\mathcal{H}^{\otimes n}\right)$. With the notation $X = \sum_{m=0}^{n-1} I_m(f_m) + X_n$ where each f_m is a non-random symmetric function in $\mathcal{H}^{\otimes m}$, X_n is a sub-nth-Gaussian chaos random variable in the following two cases.

Case $n = 2$. *Assume*

$$M_2 := \operatorname*{ess\,sup}_{\omega\in\Omega}\left(\int_0^1\int_0^1 \left|D^{(2)}_{s_2,s_1}X\right|^2 ds_2 ds_1\right)^{1/2} < \infty.$$

Then $X_2 = X - \mathbf{E}X - I_1(f_1)$ is a sub-2nd-Gaussian chaos random variable relative to the scale $\pi\sqrt{10}M_2$.

Case $n \geq 3$. *Let*

$$G_n(x) = \sum_{k=n+1}^{\infty} \left(\frac{2k}{n} \sqrt{\frac{e}{2}} \right)^{2k/n} \frac{1}{k!} x^k. \tag{4.2}$$

Assume that almost surely,

$$M_2 := \operatorname*{ess\ sup}_{\omega \in \Omega} \left(\int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} \left| \mathbf{E}\left[D^{(n)}_{s_n,\ldots,s_2,s_1} X | \mathcal{F}_{s_n} \right] \right|^2 ds_n \cdots ds_2 ds_1 \right)^{1/2} < \infty \tag{4.3}$$

and assume there exists M_G non random such that almost surely

$$\int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} G_n \left(|M_G|^{-2} \left| \mathbf{E}\left[D^{(n)}_{s_n,\ldots,s_2,s_1} X | \mathcal{F}_{s_n} \right] \right|^{2/n} \right) ds_n \cdots ds_2 ds_1 \leq 1/2. \tag{4.4}$$

Then X_n is a sub-nth-Gaussian chaos random variable relative to any scale $M \geq \max \left(\log^{-n/2}(3/2) M_2; M_G \right)$.

In particular, with K_u a universal constant, the following choice of M is satisfactory if it is finite:

$$M = \operatorname*{ess\ sup}_{\omega \in \Omega} \frac{\left(\log \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} n! \exp\left(\left| \mathbf{E}\left[D^{(n)}_{s_n,\ldots,s_2,s_1} X | \mathcal{F}_{s_n} \right] \right|^{6/n} \right) ds_n \cdots ds_2 ds_1 \right)^{n/6}}{\log^{n/6}\left(1 + \frac{n!}{K_u} \right)}.$$

Note that f may be taken to be symmetric in the above theorem. Also note that this theorem is presumably inefficient for $n \geq 3$, since the case $n = 2$ has a much more natural conclusion. In fact one may conjecture that up to a universal constant, Condition (4.3) by itself is sufficient to ensure that X_n is a sub-nth-Gaussian chaos random variable relative to M_2; yet we have not found a proof of this fact in general. Our result in the above theorem in the case $n = 2$ matches this conjecture in that case, up to the multiplicative universal constant $\pi\sqrt{10}$ which is presumably not sharp; the proof is self-contained, and of independent interest, but does not seem to allow passage to $n \geq 3$; the proof is also intriguing in that it seems to make rather wasteful use of the hypothesis of boundedness of $\left| D^{(2)} X \right|_{\mathcal{H}^{\otimes 2}}$, and one may wonder whether examples can be found where X is a sub-2nd chaos r.v. without $\left| D^{(2)} X \right|_{\mathcal{H}^{\otimes 2}}$ being bounded. The conjecture does holds for the special case of nth Wiener chaos random variables, i.e., $X = I_n(f_n)$ for some non-random $f \in \mathcal{H}^{\otimes n}$; we have not found an elementary proof of this fact; nevertheless it is a consequence of the proof of a result by C. Borell in [3], where the isoperimetric inequality is used (see Lemma 4.15 below). Lastly, note that the proof of Theorem 4.7 for $n = 2$ does not seem to extend to $n \geq 3$, while it is not possible to adapt the proof for $n \geq 3$ to the case $n = 2$ because, in the latter case, the function G_2 would not have an infinite radius of convergence.

The classical Clark-Ocone representation will be needed to prove Theorem 4.7.

Remark 4.8 (Clark-Ocone Representation). Any random variable X in $\mathbf{D}^{1,2}$ can be written as $X = \mathbf{E}X + \int_0^1 \mathbf{E}\left[D_s X | \mathcal{F}_s\right] dW_s$.

By iterating this proposition, we obtain the following, whose proof is in the appendix.

Lemma 4.9. *Let $X \in \mathbf{D}^{n,2} \subset L^2(\Omega)$ with the Wiener chaos decomposition $X = \sum_{n=0}^{\infty} I_n(f_n)$ where $f_n \in \mathcal{H}^{\otimes n}$ and f_n is symmetric. Then*

$$X = \sum_{m=0}^{n-1} I_m(f_m) + \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} \mathbf{E}(D^{(n)}_{s_n,\dots,s_2,s_1} X | \mathcal{F}_{s_n}) dW_{s_n} \cdots dW_{s_2} dW_{s_1}.$$

Proof of Theorem 4.7*, "Case $n \geq 3$".* From Lemma 4.9, where the functions $(f_m)_{m=0}^{n-1}$ are identified, we have that

$$X_n = X - \sum_{m=0}^{n-1} I_m(f_m) = \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} u(s_1, s_2, \dots, s_n)\, dW_{s_n} \cdots dW_{s_2} dW_{s_1}$$

where the stochastic process $u(s_1, s_2, \dots, \cdot)$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$, and $u \in L^\infty\left(\Omega; L^2\left(\mathcal{H}^{\otimes n}\right)\right)$, that is to say, with the non-random number $\|u\|_{\infty,2} := (1/n!)\, \|u\|_{L^\infty(\Omega; L^2(\mathcal{H}^{\otimes n}))}$, almost surely,

$$\int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} |u(s_1, s_2, \dots, s_n)|^2\, ds_n \cdots ds_2 ds_1 \leq \|u\|_{\infty,2}^2 . \tag{4.5}$$

Let now

$$U = \left| \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} u(s_1, s_2, \dots, s_n)\, dW_{s_n} \cdots dW_{s_2} dW_{s_1} \right|^{1/n} . \tag{4.6}$$

Intuitively, since one way to construct an nth Wiener chaos r.v. is to take a polynomial of degree n and apply it to a Gaussian r.v., the definition of U should presumably give us a sub-Gaussian r.v. In any event, to prove the theorem, we only need to show that

$$\mathbf{E} \exp\left(\frac{U^2}{L^2}\right) \leq 2,$$

where $L = M^{1/n}$.

Step 1*: Taylor expansion.* For simplicity we use the notation $V = U/L$. We simply evaluate the Taylor expansion of the exponential above in the following way, where for the terms with $k = 1, \dots, n$, we used Jensen's inequality:

$$\begin{aligned} \mathbf{E} \exp V^2 &= \mathbf{E} \sum_{k=0}^{\infty} \frac{V^{2k}}{k!} = \mathbf{E}\left[1 + \sum_{k=1}^{n} \frac{V^{2k}}{k!} + \sum_{k=n+1}^{\infty} \frac{V^{2k}}{k!}\right] \\ &\leq 1 + \sum_{k=1}^{n} \frac{\mathbf{E}\left[V^{2n}\right]^{k/n}}{k!} + \sum_{k=n+1}^{\infty} \frac{\mathbf{E}\left[V^{2k}\right]}{k!}. \end{aligned} \tag{4.7}$$

Step 2: Moments evaluations. We have that $V^{2k} = (V^n)^{2k/n} = \left(\left|Y^{(n)}(1)\right|/L^n\right)^{2k/n}$ where $\left(Y^{(n)}(t)\right)_{t\in[0,1]}$ is the $(\mathcal{F}_t)_{t\in[01]}$-martingale defined by

$$Y^{(n)}(t) = \int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-1}} u(s_1, s_2, \ldots, s_n)\, dW_{s_n} \cdots dW_{s_2} dW_{s_1}.$$

The bracket of $Y^{(n)}$ thus satisfies

$$\left\langle Y^{(n)} \right\rangle (t) = \int_0^t ds_1 \left| \int_0^{s_1} \cdots \int_0^{s_{n-1}} u(s_1, s_2, \ldots, s_n)\, dW_{s_n} \cdots dW_{s_2} \right|^2.$$

We begin by evaluating the moments in the tail of the series (4.7). By the Burkholder-Davis-Gundy inequality, for any $k \geq n+1$, since then we have $p := 2k/n > 2$,

$$\mathbf{E}\left[V^{2k}\right] \leq c(2k/n)\, \mathbf{E}\left[\left\langle Y^{(n)} \right\rangle^{k/n} (1)\right] / L^{2k}.$$

where $c(2k/n)$ is the efficient constant defined in Proposition 5.1 in the appendix. Let us evaluate the moments of this bracket by induction. We begin by defining a sequence of martingales: for $t \leq s_1$,

$$Y_{s_1}^{(n-1)}(t) = \int_0^t \cdots \int_0^{s_{n-1}} u(s_1, s_2, \ldots, s_n)\, dW_{s_n} \cdots dW_{s_2};$$

for $t \leq s_2 \leq s_1$,

$$Y_{s_1,s_2}^{(n-2)}(t) = \int_0^t \cdots \int_0^{s_{n-1}} u(s_1, s_2, \ldots, s_n)\, dW_{s_n} \cdots dW_{s_3};$$

more generally for $t \leq s_j \leq s_{j-1} \leq \cdots \leq s_1$,

$$Y_{s_1,s_2,\ldots,s_j}^{(n-j)}(t) = \int_0^t \cdots \int_0^{s_{n-1}} u(s_1, s_2, \ldots, s_n)\, dW_{s_n} \cdots dW_{s_{j+1}};$$

the last iteration is for $t \leq s_{n-1} \leq s_{n-2} \leq \cdots \leq s_1$,

$$Y_{s_1,s_2,\ldots,s_{n-1}}^{(1)}(t) = \int_0^t u(s_1, s_2, \ldots, s_n)\, dW_{s_n}.$$

We now have, iterating the use of the Burkholder-Davis-Gundy inequality, and using Jensen's inequality for the measures ds_{j+1}/s_j on $[0, s_j]$ for each $j = 0, \ldots, n-1$,

$$
\begin{aligned}
& c\,(p)\,\mathbf{E}\left[\left\langle Y^{(n)}\right\rangle^{p/2}(1)\right] \\
& = c\,(p)\,\mathbf{E}\left[\left(\int_0^1 ds_1 \left|Y^{(n-1)}_{s_1}(s_1)\right|^2\right)^{p/2}\right] \\
& \leq c\,(p) \int_0^1 ds_1 \mathbf{E}\left[\left|Y^{(n-1)}_{s_1}(s_1)\right|^p\right] \\
& \leq c\,(p)^2 \int_0^1 ds_1 \mathbf{E}\left[\left(\int_0^{s_1} ds_2 \left|Y^{(n-2)}_{s_1,s_2}(s_2)\right|^2\right)^{p/2}\right] \\
& \leq c\,(p)^2 \int_0^1 ds_1 s_1^{p/2-1} \mathbf{E}\left[\int_0^{s_1} ds_2 \left|Y^{(n-2)}_{s_1,s_2}(s_2)\right|^p\right] \\
& \vdots \\
& \leq c\,(p)^{n-1} \int_0^1 ds_1 s_1^{p/2-1} \int_0^{s_1} ds_2 s_2^{p/2-1} \cdots \\
& \cdots \int_0^{s_{n-3}} ds_{n-2} s_{n-2}^{p/2-1} \int_0^{s_{n-2}} ds_{n-1} \mathbf{E}\left[\left|Y^{(1)}_{s_1,s_2,\ldots,s_{n-1}}(s_{n-1})\right|^p\right] \\
& = c\,(p)^{n-1} \int_0^1 ds_1 s_1^{p/2-1} \int_0^{s_1} ds_2 s_2^{p/2-1} \cdots \\
& \cdots \int_0^{s_{n-3}} ds_{n-2} s_{n-2}^{p/2-1} \int_0^{s_{n-2}} ds_{n-1} \mathbf{E}\left[\left(\int_0^{s_{n-1}} u^2\,(s_1,s_2,\ldots,s_{n-1},s_n)\,ds_n\right)^{p/2}\right].
\end{aligned}
\tag{4.8}
$$

Now for the first terms in the series (4.7), note that an immediate calculation from (4.6) (as in Step 2, with $p = 2$, so that $c\,(2) = 1$) yields

$$
\begin{aligned}
\mathbf{E}\left[U^{2n}\right] &\leq \operatorname*{ess\,sup}_{\omega\in\Omega} \int\cdots\int_{0\leq s_n\leq\cdots\leq s_1\leq 1} \left|u\,(\bar{s})\right|^2 ds_1\cdots ds_n \\
&= \|u\|^2_{\infty,2}\,.
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
\sum_{k=1}^{n} \frac{\mathbf{E}\left[V^{2n}\right]^{k/n}}{k!} &\leq \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{\mathbf{E}\left[U^{2n}\right]}{L^{2n}}\right)^{k/n} \\
&\leq \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{\|u\|^2_{\infty,2}}{L^{2n}}\right)^{k/n}.
\end{aligned}
\tag{4.9}
$$

Step 3. *General conclusion.* We first deal with the terms in (4.9). Let us show that we can find a constant $c\,(n)$ depending only on n such that if $L^{2n} \geq c\,(n)\,\|u\|^2_{\infty,2}$, then the term in (4.9) is bounded by $1/2$. Indeed, since

$$\sum_{k=1}^{n} \frac{1}{k!} \left(\frac{\|u\|_{\infty,2}^{2}}{L^{2n}} \right)^{k/n} \leq \sum_{k=1}^{n} \frac{1}{k!} c(n)^{-k/n}$$
$$= \sum_{k=1}^{n} \frac{1}{k!} \left[c(n)^{-1/n} \right]^{k}$$
$$\leq e^{c(n)^{-1/n}} - 1,$$

it is sufficient to have $e^{c(n)^{-1/n}} = 3/2$, i.e., $c(n) = \log^{-n}(3/2)$, or, in other words,

$$L^{2n} \geq \log^{-n}(3/2) \|u\|_{\infty,2}^{2}. \tag{4.10}$$

Under this constraint, with inequality (4.7), we thus get

$$\mathbf{E}\left[\exp V^2\right] \leq 1 + 1/2 + T \tag{4.11}$$

where the tail term T of the Taylor expansion, is dealt with as follows. We apply line (4.8) above with $p = 2k/n$ and then sum over all $k \geq n+1$. Thus one last use of Jensen's inequality, and the upper bound (4.5) on $|u|$, with the shorthand notation $u(\bar{s}) := u(s_1, s_2, \ldots, s_{n-1}, s_n)$, yield

$$T := \sum_{k=n+1}^{\infty} \frac{\mathbf{E}\left[V^{2k}\right]}{k!} \leq \sum_{k=n+1}^{\infty} \frac{c(2k/n)}{k! L^{2k}} \mathbf{E}\left[\left\langle Y^{(n)} \right\rangle^{k/n} (1) \right]$$
$$\leq \mathbf{E} \sum_{k=n+1}^{\infty} \frac{c(2k/n)^{n-1}}{k! L^{2k}} \int_0^1 ds_1 s_1^{k/n-1} \int_0^{s_1} ds_2 s_2^{k/n-1} \cdots$$
$$\cdots \int_0^{s_{n-3}} ds_{n-2} s_{n-2}^{k/n-1} \int_0^{s_{n-2}} ds_{n-1} \left(\int_0^{s_{n-1}} u^2(\bar{s}) \, ds_n \right)^{k/n}$$
$$\leq \mathbf{E} \sum_{k=n+1}^{\infty} \frac{c(2k/n)^n}{k! L^{2k}} \int \cdots \int_{0 \leq s_n \leq \cdots \leq s_1 \leq 1} ds_1 \cdots ds_n \, |u(\bar{s})|^{2k/n}$$
$$\leq \operatorname*{ess\,sup}_{\omega \in \Omega} \sum_{k=n+1}^{\infty} \frac{c(2k/n)^n}{k! L^{2k}} \int \cdots \int_{0 \leq s_n \leq \cdots \leq s_1 \leq 1} ds_1 \cdots ds_n \, |u(\bar{s})|^{2k/n}. \tag{4.12}$$

Below for any $m \geq 2$, $\|f\|_m$ denotes the L^m norm of any function f on the simplex $0 \leq s_n \leq \cdots \leq s_1 \leq 1$ with respect to Lebesgue measure. Hence we can write from (4.11) and from (4.12):

$$\mathbf{E} \exp V^2 \leq 3/2 + \operatorname{ess\,sup}_{\omega \in \Omega} \sum_{k=n+1}^{\infty} \frac{c(2k/n)^n}{k! L^{2k}} \|u\|_{2k/n}^{2k/n}$$
$$= 3/2 +$$
$$\operatorname*{ess\,sup}_{\omega \in \Omega} \int \cdots \int_{0 \leq s_n \leq \cdots \leq s_1 \leq 1} ds_1 \cdots ds_n \sum_{k=n+1}^{\infty} \frac{c(2k/n)^n}{k!} \left(\frac{|u(s_1, s_2, \ldots, s_{n-1}, s_n)|^{2/n}}{L^2} \right)^k.$$

Now the estimate in Proposition 5.1 in the appendix tells us that

$$\frac{c\left(2k/n\right)^n}{k!} \leq \left(\frac{2k}{n}\sqrt{\frac{e}{2}}\right)^{2k/n} \frac{1}{k!},$$

so that, with $G_n(x)$ as in the statement of the theorem,

$$\mathbf{E}\exp V^2 \leq 3/2 + \operatorname*{ess\,sup}_{\omega\in\Omega} \int\cdots\int_{0\leq s_n\leq\cdots\leq s_1\leq 1} ds_1\cdots ds_n G_n\left(\frac{\left|u\left(\bar{s}\right)\right|^{2/n}}{L^2}\right).$$

Choosing L^2 such that the last term above is less than $1/2$, and with the constraint (4.10), the statement following line (4.4) in the theorem now follows immediately.

Step 4. *Analytic conclusion.* To finish the proof of the theorem in the case $n\geq 3$, we only need to study the function G_n more specifically. Using the Stirling-type formula which is valid for all $k\geq 1$, $k!\geq k^k 3^{-k}$, and using the fact that $n\geq 3$, we get easily

$$G_n(x) \leq \sum_{k=n+1}^{\infty}\left(\frac{4}{k}\right)^{k/3} x^k.$$

For any integer $m\geq 2$, consider the three values $k=3m$, $3m+1$, or $3m+2$. We then obtain $k^{k/3}\geq(3m)^m$. On the other hand, for these same values of k, with $x>1$, we get $x^k\leq\left(x^3\right)^m x^2$. Thus

$$G_n(x) \leq 4^{2/3}x^2\sum_{m=1}^{\infty}\left(\frac{4}{3}\right)^m m^{-m}\left(x^3\right)^m.$$

Using again the Stirling-type formula, valid for all $m\geq 1$, $2^m m^{-m}\leq 1/m!$, we get

$$\begin{aligned} G_n(x) &\leq 4^{2/3}x^2\sum_{m=1}^{\infty}\frac{1}{m!}\left(\frac{2}{3}x^3\right)^m \\ &= 4^{2/3}x^2\left(\exp\left(\frac{2}{3}x^3\right)-1\right). \end{aligned}$$

Thus for $x>1$,

$$G_n(x)\leq 9\cdot 4^{2/3}\left(\exp\left(x^3\right)-1\right),$$

even though the universal constant $9{\cdot}4^{2/3}$ may not be optimal. When $0<x<1$, on the other hand, a similar inequality is found, with a different universal constant; we use the notation K_u for the maximum of the two constants. We may now rewrite the left-hand side of (4.4), which we call Γ, using the last inequality above:

$$\begin{aligned} n!\Gamma &:= \int_{[0,1]^n} G_n\left(L^{-2}\left|u\left(\bar{s}\right)\right|^{2/n}\right) ds_n \cdots ds_2 ds_1 \\ &\leq -K_u + K_u \int_{[0,1]^n} \exp\left(L^{-6}\left|u\left(\bar{s}\right)\right|^{6/n}\right) ds_n \cdots ds_2 ds_1 \\ &= -K_u + K_u \int_{[0,1]^n} \left[\exp\left(\left|u\left(\bar{s}\right)\right|^{6/n}\right)\right]^{1/L^6} ds_n \cdots ds_2 ds_1. \end{aligned}$$

We now make a temporary assumption that $L \geq 1$. This allows us to use Jensen's inequality in the above time integral over the simplex:

$$n!\Gamma \leq -K_u + K_u \left(\int_{[0,1]^n} \exp\left(\left|u\left(\bar{s}\right)\right|^{6/n}\right) ds_n \cdots ds_2 ds_1\right)^{1/L^6}.$$

Hence, since we only need to satisfy the condition (4.4), i.e., $\Gamma \leq 1/2$ almost surely, we only need to have

$$L^6 \geq \frac{\log\left(\int_{[0,1]^n} \exp\left(\left|u\left(\bar{s}\right)\right|^{6/n}\right) ds_n \cdots ds_2 ds_1\right)}{\log\left(1 + \frac{n!}{2K_u}\right)} \tag{4.13}$$

almost surely. Jensen's inequality can then be used to check that this last expression is always larger than the right-hand side of (4.3). The last statement of the theorem is thus proved if the essential supremum $(L^*)^6$ of the right-hand side of (4.13) happens to be greater than 1. If it is not, we leave it to the reader to check that the same conclusion holds by repeating the above calculation (Steps 3 and 4) for the random variable $\tilde{U} = U/L^*$, thereby allowing us not to require $L \geq 1$. $\square$

Proof of Theorem 4.7, *"Case $n = 2$".* The proof is based on Lemma 3.3, applied to the random variable

$$Y = \left(\int_0^1 \left|D_r X_2\right|^2 dr\right)^{1/2} = \left|D.X_2\right|_{\mathcal{H}}.$$

The first step is to prove the following: almost surely,

$$\left|D.Y\right|_{\mathcal{H}}^2 \leq M_2^2 = \operatorname*{ess\,sup}_{\omega\in\Omega} \left(\int_0^1 \int_0^1 \left|D^{(2)}_{s_2,s_1} X_2\right|^2 ds_2 ds_1\right)^{1/2}.$$

Indeed, noting that X and X_2 have the same second malliavin derivative, we have

$$
\begin{aligned}
|D.Y|^2_{\mathcal{H}} &= \int_0^1 \left| D_t \sqrt{\int_0^1 |D_r X_2|^2 \, dr} \right|^2 dt = \int_0^1 \left| \frac{D_t \int_0^1 |D_r X_2|^2 \, dr}{2\sqrt{\int_0^1 |D_r X_2|^2 \, dr}} \right|^2 dt \\
&= \int_0^1 \frac{\left| \int_0^1 (D_r X_2) \left(D^{(2)}_{t,r} X_2 \right) dr \right|^2}{\int_0^1 |D_r X_2|^2 \, dr} dt \\
&\leq \int_0^1 \frac{\int_0^1 |D_r X_2|^2 \, dr \cdot \int_0^1 \left| D^{(2)}_{t,r} X_2 \right|^2 dr}{\int_0^1 |D_r X_2|^2 \, dr} dt \\
&= \int_0^1 \int_0^1 \left| D^{(2)}_{t,r} X_2 \right|^2 drdt = \int_0^1 \int_0^1 \left| D^{(2)}_{t,r} X \right|^2 drdt \leq M_2^2.
\end{aligned}
$$

Thus we can consider that $Z = Y - \mathbf{E}Y$ is a random variable satisfying the hypotheses of Lemma 3.3. We can thus conclude that Z is sub-Gaussian relative to the scale M_2. In particular we get, from Remark 3.2,

$$\mathbf{E}\left[\exp\left(\frac{Z^2}{5M_2^2} \right) \right] \leq 2.$$

Because we will need to find a smaller constant than 2 above, we restate this as

$$\mathbf{E}\left[\exp\left(\frac{Z^2}{10M_2^2} \right) \right] \leq \sqrt{2}. \tag{4.14}$$

We now invoke an exponential Poincaré inequality of Üstünel [23, Theorem 9.2.3(i)]: for any centered random variable V in $\mathbf{D}^{1,2}$,

$$\mathbf{E}\left[\exp V \right] \leq \mathbf{E}\left[\exp\left(\frac{\pi^2}{8} |D.V|^2_{\mathcal{H}} \right) \right].$$

Applying this to $V = X_2/c$ for some constant $c > 0$, we get

$$
\begin{aligned}
\mathbf{E}\left[\exp\left(\frac{X_2}{c} \right) \right] &\leq \mathbf{E}\left[\exp\left(\frac{\pi^2}{8c^2} |D.X_2|^2_{\mathcal{H}} \right) \right] \\
&= \mathbf{E}\left[\exp\left(\frac{\pi^2}{8c^2} (Z + \mathbf{E}Y)^2 \right) \right] \\
&\leq \mathbf{E}\left[\exp\left(\frac{\pi^2}{4c^2} Z^2 \right) \right] \exp\left(\frac{\pi^2}{4c^2} (\mathbf{E}Y)^2 \right). \qquad (4.15)
\end{aligned}
$$

Now if we choose $\pi^2 / \left(4c^2\right) \leq 1 / \left(10M_2^2\right)$, from (4.14), the first term in the last line above is bounded above by $\sqrt{2}$. In order to control the second term, we use the chaos decomposition $X_2 = \sum_{m=2}^{\infty} I_m(f_m)$. We have

$$D_r X_2 = \sum_{m=2}^{\infty} m I_{m-1}(f_m(\cdot, r))$$

and so

$$(\mathbf{E}Y)^2 \le \mathbf{E}\left(|D_\cdot X_2|^2_{\mathcal{H}}\right) = \sum_{m=2}^{\infty} m(m!)|f_m|^2_{\mathcal{H}^{\otimes m}}.$$

We also have that

$$\mathbf{E}\left(|D^{(2)}_\cdot X_2|^2_{\mathcal{H}^{\otimes 2}}\right) = \sum_{m=2}^{\infty} m(m-1)(m!)|f_m|^2_{\mathcal{H}^{\otimes m}} \le M_2^2.$$

Since the second series above is clearly less than the third series, we get that $(\mathbf{E}Y)^2 \le M_2^2$. Certainly, the above choice for c implies $\pi^2/(4c^2) \le (\log\sqrt{2})/M_2^2$. From (4.15) we now get

$$\mathbf{E}\left[\exp\left(\frac{X_2}{\pi\sqrt{5/2}M_2}\right)\right] \le 2. \tag{4.16}$$

The last step in the proof is to allow the use of $|X_2|$ instead of X_2 above. Since we have no information about the symmetry of X_2, we proceed as follows. Since X_2 and $-X_2$ satisfy the same hypotheses, we have that (4.16) holds for X_2 replaced by $-X_2$. Now we can write, with $X' = X_2/\left(\pi\sqrt{10}M_2\right)$, and using the notation $p = \mathbf{P}[X' \ge 0]$,

$$\begin{aligned}
\mathbf{E}\left[\exp\left(|X'|\right)\right] &= \mathbf{E}\left[\exp\left(X'\right)\mathbf{1}_{X'\ge 0}\right] + \mathbf{E}\left[\exp\left(-X'\right)\mathbf{1}_{X'<0}\right] \\
&\le \sqrt{p}\sqrt{\mathbf{E}\left[\exp\left(2X'\right)\right]} + \sqrt{1-p}\sqrt{\mathbf{E}\left[\exp\left(-2X'\right)\right]} \\
&\le \sqrt{2}\left(\sqrt{p} + \sqrt{1-p}\right) \\
&\le 2.
\end{aligned}$$

This finishes the proof of Case $n = 2$ of the theorem. □

4.3. Concentration: the sub-nth chaos property for suprema

We now prove the core of a Borell-Sudakov-type inequality for sub-nth chaos random fields.

Proposition 4.10. *Let X be a separable random field on an index set I such that all finite-dimensional vectors of X are formed of almost-surely distinct components. Assume $\mu := \mathbf{E}[\sup_I X] < \infty$. Assume $X(t) \in \mathbf{D}^{n,2}$ for each $t \in I$. Assume there exist non-random constants $\sigma(t)$ for each $t \in I$ such that almost surely*

$$\left\|D^{(n)}_\cdot X(t)\right\|_2^2 := \int_0^1\int_0^{s_1}\cdots\int_0^{s_{n-1}}\left|D^{(n)}_{s_n,\dots,s_2,s_1}X(t)\right|^2 ds_n\cdots ds_2 ds_1 \le \sigma^2(t).$$

Then $\sup_{t\in I} X(t) \in \mathbf{D}^{1,2}$ and

$$\left\|D^{(n)}_\cdot \sup_{t\in I} X(t)\right\|_2^2 \le \sup_{t\in I}\sigma^2(t).$$

Proof. As in the proof of Theorem 3.6, we can assume without loss of generality that $I = \{1, 2, \ldots, N\}$. Here we have $n \geq 2$. Using the same strategy as in the proof of Theorem 3.6, we denote $X_m = X(m)$ and define $S_m = \max\{X_1, X_2, \ldots, X_m\}$, so that $S_{m+1} = \max\{X_m, S_m\}$. In order to prove that $\max_I X \in \mathbf{D}^{n,2}$, the approximation technique used in the proof of Theorem 3.6 can again be used. We omit the details, only to say that $\mathbf{1}_{X_{m+1}>Sm}$ can be approximated in $\mathbf{D}^{1,2}$ by a smooth function of $X_{m+1} - S$ whose Malliavin derivative tends to 0 for almost every (ω, s) in $L^2(\Omega) \times \mathcal{H}$ because $X_{m+1} - S_m \neq 0$ a.s. In particular, $D.\mathbf{1}_{X_{m+1}>Sm} = 0$ in $L^2(\Omega) \times \mathcal{H}$, and for any $k \leq n$, the kth-order Malliavin derivative of $\mathbf{1}_{X_{m+1}>Sm}$ is 0 in $L^2(\Omega) \times \mathcal{H}^{\otimes k}$ as well.

This justifies the following computation, where equalities hold in $L^2(\Omega) \times \mathcal{H}^{\otimes n}$:

$$\begin{aligned}
D_{s_n,\ldots,s_2,s_1}{}^{(n)} S_{m+1} &= D^{(n-1)}_{s_n,\ldots,s_2}\left(D_{s_1}X_{m+1}\mathbf{1}_{X_{m+1}>Sm} + D_{s_1}S_m\mathbf{1}_{X_{m+1}<Sm}\right)\\
&= D^{(n-2)}_{s_n,\ldots,s_3}\left(\left[D_{s_2}D_{s_1}X_{m+1}\right]\mathbf{1}_{X_{m+1}>Sm} + \left[D_{s_2}D_{s_1}S_m\right]\mathbf{1}_{X_{m+1}<Sm}\right)\\
&\vdots\\
&= \left[D^{(n)}_{s_n,\ldots,s_2,s_1}X_{m+1}\right]\mathbf{1}_{X_{m+1}>Sm} + \left[D^{(n)}_{s_n,\ldots,s_2,s_1}S_m\right]\mathbf{1}_{X_{m+1}<Sm}.
\end{aligned} \tag{4.17}$$

Now, still following the strategy of the proof of Theorem 3.6, we let $\sigma_m^{*2} = \max\{\sigma^2(1), \ldots, \sigma^2(m)\}$, and we assume by induction that $\left\|D_\cdot^{(n)}S_m\right\|_2^2 \leq \sigma_m^{*2}$ almost surely. Our hypothesis and equality (4.17) implies that almost surely

$$\begin{aligned}
\left\|D_\cdot^{(n)}S_{m+1}\right\|_2^2 &= \int_0^1\int_0^{s_1}\cdots\int_0^{s_{n-1}}\left|D^{(n)}_{s_n,\ldots,s_2,s_1}S_{m+1}\right|^2 ds_n\cdots ds_2 ds_1\\
&= \mathbf{1}_{X_{m+1}>Sm}\int_0^1\int_0^{s_1}\cdots\int_0^{s_{n-1}}\left|D^{(n)}_{s_n,\ldots,s_2,s_1}X_{m+1}\right|^2 ds_n\cdots ds_2 ds_1\\
&+ \mathbf{1}_{X_{m+1}<Sm}\int_0^1\int_0^{s_1}\cdots\int_0^{s_{n-1}}\left|D^{(n)}_{s_n,\ldots,s_2,s_1}S_m\right|^2 ds_n\cdots ds_2 ds_1\\
&= \mathbf{1}_{X_{m+1}>Sm}\left\|D_\cdot^{(n)}X_{m+1}\right\|_2^2 + \mathbf{1}_{X_{m+1}<Sm}\left\|D_\cdot^{(n)}S_m\right\|_2^2\\
&\leq \sigma^2(m+1)\,\mathbf{1}_{X_{m+1}>Sm} + \sigma_m^{*2}\mathbf{1}_{X_{m+1}<Sm}\\
&\leq \sigma_{m+1}^{*2}.
\end{aligned}$$

Since $\left\|D_\cdot^{(n)}X_1\right\|_2^2 \leq \sigma_1^{*2} = \sigma^2(1)$ by hypothesis, induction implies the conclusion of the proposition when $m = N$. □

Combining the results of Theorem 4.7 and Proposition 4.10, we state the extension of the Borell-Sudakov inequality in two separate results, depending on whether $n = 2$ or $n \geq 3$.

Corollary 4.11. *Let X and μ be as in Proposition* 4.10 *with $n = 2$. Then $\sup_I X - \mu$ is a sub-2nd chaos random variable. It can be decomposed as*

$$\sup_I X - \mu = \int_0^1 f(s)\, dW_s + X_2$$

where $f \in \mathcal{H}$ and X_2 is a sub-2nd chaos random variable relative to the scale $M = \pi\sqrt{10}\sup_{t\in I}\sigma^2(t)$. In particular we get the following extension of the Borell-Sudakov inequality: for any $u > 0$,

$$\mathbf{P}\left[\left|\sup_I X - \int_0^1 f(s)\, dW_s - \mu\right| > u\right] = \mathbf{P}\left[|X_2| > u\right] \leq 2\exp\left(-\frac{u}{M}\right). \tag{4.18}$$

Proof. The first statement follows immediately from the conclusion of Proposition 4.10 as applied to "Case $n = 2$" in Theorem 4.7. The second statement is an immediate consequence of the tail estimate in Lemma 4.6. □

The presence of the function G_n in Theorem 4.7 case $n \geq 3$ makes it impossible to apply Proposition 4.10 directly. Moreover, the conditional expectation in that same portion of the theorem causes further difficulties, making it necessary to impose slightly stronger conditions on $D^{(n)}X$ than in that theorem, in order to derive a Borell-Sudakov extension.

Proposition 4.12. *Let X and μ be as in Proposition* 4.10 *with $n \geq 3$. Recall the function G_n defined in "Case $n = 3$" of Theorem 4.7. Assume moreover that for any $t \in I$ and for any $s_n \in [0,1]$, there exists a non-random value $M(t)$ not dependent on s_n, such that, almost surely*

$$\int_{s_n}^1 \int_{s_n}^{s_1} \cdots \int_{s_n}^{s_{n-2}} G_n\left(M(t)^{-2}\left|D^{(n)}_{s_n,\ldots,s_2,s_1}X(t)\right|^{2/n}\right) ds_{n-1}\cdots ds_2 ds_1 \leq 1/2$$

and

$$M(t) \geq \sqrt{2e}\left\|D^{(n)}_{\cdot}X(t)\right\|_{\mathcal{H}}.$$

Then the random variable $\sup_I X - \mu$ is a sub-nth chaos r.v. It can be decomposed as $\sup_I X - \mu = \sum_{m=1}^{n-1} I_m(f_m) + X_n$ where each f_m is a non-random symmetric function in $\mathcal{H}^{\otimes m}$, and X_n is a sub-nth-Gaussian chaos random variable relative to the scale

$$M = \sup_{t\in I} M(t).$$

In particular, the extension (4.18) *of the Borell-Sudakov inequality holds for X_n with this M, namely,*

$$\mathbf{P}\left[\left|\sup_I X - \sum_{m=1}^{n-1} I_m(f_m) - \mu\right| > u\right] = \mathbf{P}\left[|X_n| > u\right] \leq 2\exp\left(-\left(\frac{u}{M}\right)^{2/n}\right). \tag{4.19}$$

Remark 4.13. The hypothesis of this proposition is clearly satisfied if there exist constants $\sigma(t)$ such that almost surely, for all $s_1, s_2, \ldots, s_n$, $\left|D^{(n)}_{s_n,\ldots,s_2,s_1} X(t)\right| \leq \sigma(t)$. Then there is a constant k_n depending only on n such that we may take $M = k_n \sup_{t\in I} \sigma(t)$.

Proof of Proposition 4.12. Here, we may not apply Proposition 4.10 directly. Instead, we return to its proof, and use the notation therein. Let $T_{n-1}(s_n) = \left\{(s_i)_{i=1}^{n-1} : s_1 \geq s_2 \geq \cdots \geq s_{n-1} \geq s_n\right\}$, a simplex for any fixed $s_n \in [0,1]$. Let $M_m = M(m)$. Also use the shorthand notation $\bar{s} = (s_i)_{i=1}^{n-1}$. By hypothesis we have

$$\underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left((M_m)^{-2}\left|D^{(n)}_{s_n,\bar{s}} X(m)\right|^{2/n}\right) d\bar{s} \leq \frac{1}{2}. \tag{4.20}$$

We also define

$$M^*_m = \max\{M_1, M_2, \ldots, M_m\}.$$

Then, since G_n is an increasing function, we have, from line (4.17),

$$\begin{aligned}
&\underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left(\left(M^*_{m+1}\right)^{-2}\left|D^{(n)}_{s_n,\bar{s}} S_{m+1}\right|^{2/n}\right) d\bar{s} \\
&= \mathbf{1}_{X(m+1)>S_m} \underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left(\left(M^*_{m+1}\right)^{-2}\left|D^{(n)}_{s_n,\bar{s}} X(m+1)\right|^{2/n}\right) d\bar{s} \\
&\quad + \mathbf{1}_{X(m+1)<S_m} \underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left(\left(M^*_{m+1}\right)^{-2}\left|D^{(n)}_{s_n,\bar{s}} S_m\right|^{2/n}\right) d\bar{s} \\
&\leq \mathbf{1}_{X(m+1)>S_m} \underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left(\left(M_{m+1}\right)^{-2}\left|D^{(n)}_{s_n,\bar{s}} X(m+1)\right|^{2/n}\right) d\bar{s} \\
&\quad + \mathbf{1}_{X(m+1)<S_m} \underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left(\left(M^*_{m}\right)^{-2}\left|D^{(n)}_{s_n,\bar{s}} S_m\right|^{2/n}\right) d\bar{s}.
\end{aligned}$$

Thus, if we assume that

$$\underset{T_{n-1}(s_n)}{\int\cdots\int} G_n\left((M^*_m)^{-2}\left|D^{(n)}_{s_n,\bar{s}} S_m\right|^{2/n}\right) d\bar{s} \leq \frac{1}{2}, \tag{4.21}$$

using (4.20), we obtain that (4.21) holds at rank $m+1$, and thus, by induction, for all $m \leq N$.

The definition (4.2) of G_n shows that the function $x \mapsto G_n\left(|x|^{2/n}\right)$ is convex for all x. Let $M = M^*_N = \max\{M_1, M_2, \ldots, M_N\}$. We may now write, using

Jensen, and (4.21) for $m = N$,

$$\begin{aligned}
&\int_0^1 ds_n \underset{T_{n-1}(s_n)}{\int \cdots \int} G_n\left(M^{-2}\left|\mathbf{E}\left[D^{(n)}_{s_n,\bar{s}}S_N|\mathcal{F}_{s_n}\right]\right|^{2/n}\right)d\bar{s}\\
&\leq \int_0^1 ds_n \underset{T_{n-1}(s_n)}{\int \cdots \int} \mathbf{E}\left[G_n\left(M^{-2}\left|D^{(n)}_{s_n,\bar{s}}S_N\right|^{2/n}\right)|\mathcal{F}_{s_n}\right]d\bar{s}\\
&= \int_0^1 ds_n \mathbf{E}\left[\underset{T_{n-1}(s_n)}{\int \cdots \int} d\bar{s} G_n\left(M^{-2}\left|D^{(n)}_{s_n,\bar{s}}S_N\right|^{2/n}\right)|\mathcal{F}_{s_n}\right]\\
&\leq \frac{1}{2}.
\end{aligned}$$

This establishes Condition (4.4) of Theorem 4.7. We omit the details needed to check the other conditions of this Theorem. Inequality (4.19) is again only a consequence of Lemma 4.6. □

The presence of the Wiener chaos correction terms $\sum_{m=0}^{n-1} I_m(f_m)$ in the statement of the generalizations (4.18) and (4.19) of Borell-Sudakov are somewhat of an annoyance, because these inequalities' proofs present no way of calculating the magnitude of the non-random functions $\{f_m : m = 1, \ldots, n-1\}$. We propose an additional result which shows that asymptotically, these functions are irrelevant.

Corollary 4.14. *With the hypotheses and notation as in Corollary* 4.11 *or Proposition* 4.12*, we have for any* $\varepsilon > 0$*, for* u *large enough,*

$$\mathbf{P}\left[|\sup_I X - \mu| > u\right] \leq 2(1+\varepsilon)\exp\left(-\frac{1}{(1+\varepsilon)}\left(\frac{u}{M}\right)^{2/n}\right).$$

More concisely, we can write

$$\lim_{u\to\infty}\frac{1}{u^{2/n}}\log \mathbf{P}\left[|\sup_I X - \mu| > u\right] \leq -\frac{1}{M^{2/n}}.$$

Proof. First note that, for any $r \in (0,1)$

$$\begin{aligned}
\mathbf{P}\left[|\sup_I X - \mu| > u\right] &= \mathbf{P}\left[|X_n + \sum_{m=1}^{n-1} I_m(f_m)| > u\right]\\
&\leq \mathbf{P}\left[|X_n| > u - \sum_{m=1}^{n-1} |I_m(f_m)|\right]
\end{aligned}$$

$$\leq \mathbf{P}\left[|X_n| > u - \sum_{m=1}^{n-1} |I_m(f_m)|; \sum_{m=1}^{n-1} |I_m(f_m)| \leq ru\right] + \mathbf{P}\left[\sum_{m=1}^{n-1} |I_m(f_m)| > ru\right]$$

$$\leq \mathbf{P}\left[|X_n| > (1-r)u\right] + \sum_{m=1}^{n-1} \mathbf{P}\left[|I_m(f_m)| > \frac{ru}{n-1}\right]. \tag{4.22}$$

The following lemma is a trivial consequence of the results in [10].

Lemma 4.15. *Let $f_m \in \mathcal{H}^{\otimes m}$. Then there exists a constant $M_m(f_m)$ such that*

$$\mathbf{P}\left[|I_m(f_m)| > u\right] \leq \exp\left(-\frac{1}{2}\left(\frac{u}{M_m(f_m)}\right)^{2/m}\right).$$

Armed with this Lemma, and with the inequalities (4.18) or (4.19), and choosing r so that $(1-r) > (1+\varepsilon)^{-n/2}$, we may write from (4.22),

$$\mathbf{P}\left[|\sup_I X - \mu| > u\right] \leq \mathbf{P}\left[|X_n| > u(1+\varepsilon)^{-n/2}\right] + \sum_{m=1}^{n-1} \mathbf{P}\left[|I_m(f_m)| > \frac{ru}{n-1}\right]$$

$$\leq 2\exp\left(-\frac{1}{(1+\varepsilon)}\left(\frac{u}{M}\right)^{2/n}\right) + \sum_{m=1}^{n-1} \exp\left(-\frac{1}{2}\left(\frac{ru}{M_m(f_m)(n-1)}\right)^{2/m}\right)$$

$$\leq 2\exp\left(-\frac{1}{(1+\varepsilon)}\left(\frac{u}{M}\right)^{2/n}\right) + (n-1)\exp\left(-\frac{1}{2}\left(\frac{ru}{K}\right)^{2/(n-1)}\right) \tag{4.23}$$

$$\leq 2(1+\varepsilon)\exp\left(-\frac{1}{(1+\varepsilon)}\left(\frac{u}{M}\right)^{2/n}\right), \tag{4.24}$$

where in line (4.23), the constant K is $(n-1)\max_{m\in\{1,\ldots,n-1\}} M_m(f_m)$ and in line (4.24), u is chosen so large that the second term in (4.23) is less than ε times the first. The first statement of the corollary is proved, and the second follows trivially due to the fact that $\varepsilon > 0$ is arbitrary. □

5. Appendix

5.1. Efficient constant in the Burkholder-Davis-Gundy inequality

Proposition 5.1. *For any square integrable martingale Y, and any $p \geq 2$, we have*

$$\mathbf{E}\left[\sup_{s\in[0,t]} |Y(s)|^p\right] \leq c(p)\,\mathbf{E}\left[|\langle Y\rangle(t)|^{p/2}\right]$$

where the constant $c(p)$ satisfies $c(2) = 1$ and, for any $p > 2$,

$$c(p) = \left(\frac{1}{2}\frac{p^{p+1}}{(p-1)^{p-1}}\right)^{p/2} \leq \left(\sqrt{e/2}\right)^p p^p.$$

Proof. One only needs to keep track of the constants in the classical proof of this inequality: starting with Itô's formula (where the function $f(x) = |x|^p$ is of class C^2),

$$\begin{aligned}
&\mathbf{E}\,|Y(t)|^p \\
&= \mathbf{E}\left[\int_0^t p\,|Y(s)|^{p-1}\operatorname{sgn}(Y(s))\,dY(s) + \frac{1}{2}\int_0^t p(p-1)\,|Y(s)|^{p-2}\,\langle Y\rangle(ds)\right] \\
&= \frac{p(p-1)}{2}\mathbf{E}\left[\int_0^t |Y(s)|^{p-2}\,\langle Y\rangle(ds)\right] \\
&\leq \frac{p(p-1)}{2}\mathbf{E}\left[\left(\sup_{s\in[0,t]}|Y(s)|\right)^{p-2}\langle Y\rangle(t)\right] \\
&\leq \frac{p(p-1)}{2}\mathbf{E}\left[\left(\sup_{s\in[0,t]}|Y(s)|\right)^{p}\right]^{(p-2)/p}\mathbf{E}\left[|\langle Y\rangle(t)|^{p/2}\right]^{2/p}.
\end{aligned}$$

The proposition's constant $c(p)$ follows from some elementary calculations and Doob's inequality

$$\mathbf{E}\left[\left(\sup_{s\in[0,t]}|Y(s)|\right)^{p}\right] \leq (p/(p-1))^p \sup_{s\in[0,t]}\mathbf{E}\left[|Y(s)|^p\right].$$

The second statement in the proposition is equally elementary. □

5.2. Proof of Lemma 3.7

Such a Φ as in the statement of the lemma can be replaced by an approximation Φ_m such that Φ_m is of class C^1, such that $\Phi = \Phi_m$ for all points distant by more than $1/m$ of all hyperplanes, and such that $\Phi - \Phi_m$ and $\nabla\Phi_m$ are both bounded uniformly in m by multiples of $|\nabla\Phi|_\infty$: this can be achieved by interpolating Φ and $\nabla\Phi$ from the boundary of the $1/m$-neighborhood T_m of the union T of the hyperplanes using scaled polynomials. For example, in the case we are interested in, let P be a polynomial of degree 4 on $[-1,1]$, which is increasing and convex, such that $P(-1) = P'(-1) = 0$ and $P(1) = P'(1) = 1$. Define the function $\Phi_m = \Phi$ off the set $T_m = \{|x-y| < 1/m\}$, and on that set define $\Phi_m(x,y) = m^{-1}P(m(x-y)) + y$. This sequence Φ_m has the required property, and in fact $|\nabla\Phi_m|_\infty \leq 1$ and $|\Phi - \Phi_m|_\infty \leq 1$. Now since Φ_m converges to Φ pointwise, the dominated convergence theorem implies that $\Phi_m(Z)$ converges to $\Phi(Z)$ in $L^2(\Omega)$. Moreover, we can write using the chain rule (2.4) for C^1 functions: $\Phi_m(Z) \in \mathbf{D}^{1,2}$ and

$$D_r\Phi_m(Z) = (1 - \mathbf{1}_{T_m}(Z))\,\nabla\Phi(Z)\,D_rZ + \mathbf{1}_{T_m}(Z)\,\nabla\Phi_m(Z)\,D_rZ.$$

Since $\mathbf{1}_{T_m}(Z)$ converges to 0 almost surely, and $D.Z \in L^2(\Omega\times[0,1])$, by the dominated convergence theorem in $L^2(\Omega\times[0,1])$, we have $D.\Phi_m(Z)$ converging to $\nabla\Phi(Z)\,D.Z$ in that space. Now we invoke the fact (see [17]) that the Malliavin derivative operator D is a closed operator from its domain $\mathbf{D}^{1,2}$ into $L^2(\Omega \times [0,1])$, to conclude that $\Phi(Z) \in \mathbf{D}^{1,2}$ and $D.\Phi(Z) = \nabla\Phi(Z)\, D.Z$, as was to be proved.

5.3. Proof of Lemma 4.6

The proofs of this lemma's statements are elementary; we detail some of them. First, we have using Chebyshev's inequality:

$$\begin{aligned}
\mathbf{P}[|X| > u] &= \mathbf{P}\left[\exp(X/\delta)^{2/n} > \exp(u/\delta)^{2/n}\right] \\
&\leq \exp\left(-(u/\delta)^{2/n}\right)\mathbf{E}\left[\exp(X/\delta)^{2/n}\right] \\
&\leq 2\exp\left(-(u/\delta)^{2/n}\right),
\end{aligned}$$

which is the first statement of the lemma. This then implies that

$$\begin{aligned}
\mathbf{E}\left[X^2\right] &= \int_0^\infty \mathbf{P}\left[|X| > \sqrt{u}\right] du \\
&\leq 2\int_0^\infty \exp\left(-(\sqrt{u}/\delta)^{2/n}\right) du = 2\delta^2 \int_0^\infty e^{-v^{1/n}} dv = v_n \delta^2
\end{aligned}$$

where $v_n = 2\int_0^\infty e^{-v^{1/n}} dv$, hence the second statement. The proof of the estimate for $\mathbf{E}\left[\exp\left(\lambda |X|^{1/n}\right)\right]$ is left to the reader.

For the first converse, let $c > 1$ be fixed. Using the estimate $\mathbf{P}[|X| > u] \leq 2\exp\left(-(u/\delta)^{2/n}\right)$, we get

$$\begin{aligned}
\mathbf{E}\left(\exp\left[\left(\frac{X}{c\delta}\right)^{2/n}\right]\right) &= \int_0^\infty \mathbf{P}\left[|X| > c\delta(\log r)^{n/2}\right] dr \\
&\leq 1 + \int_1^\infty 2\exp\left(-c^{2/n}\log r\right) dr \\
&= 1 + 2\int_1^\infty r^{-c^{2/n}} dr \\
&= 1 + \frac{2}{c^{2/n} - 1}.
\end{aligned}$$

Thus we only need to choose $v'_n = c = 3^{n/2}$. The proofs of the other converses are left to the reader.

5.4. Proof of Lemma 4.9

The proof uses three simple facts from the theory of Wiener chaoses. For any symmetric function g in $\mathcal{H}^{\otimes m}$, the first fact is simply the definition of $I_m(g)$ as an iterated Itô integral in (2.1). The second, from Step 2 in Section 2, is the calculation $D_r I_m(g) = m I_{m-1}(g(\cdot, r))$. The last, from Lemma 1.2.4 in [17], says that

$$\mathbf{E}\left[I_m(g_m) | \mathcal{F}_t\right] = I_m\left(g_m \mathbf{1}_{[0,t]}^{\otimes m}\right).$$

For $X \in \mathbf{D}^{n,2}$, we may now calculate, for $s_n \le s_{n-1} \le \ldots \le s_2 \le s_1 \le 1$,

$$\begin{aligned}
D^{(n)}_{s_n,\ldots,s_2,s_1} X &= D^{(n)}_{s_n,\ldots,s_2} \left(\sum_{m=1}^{\infty} m I_{m-1} \left(f_m \left(s_1, \cdot \right) \right) \right) \\
&= \sum_{m=2}^{\infty} m \left(m-1 \right) D^{(n)}_{s_n,\ldots,s_3} \left(I_{m-2} \left(f_m \left(s_1, s_2, \cdot \right) \right) \right) \\
&\vdots \\
&= \sum_{m=n}^{\infty} m \left(m-1 \right) \cdots \left(m-n+1 \right) I_{m-n} \left(f_m \left(s_1, s_2, \ldots, s_n, \cdot \right) \right).
\end{aligned}$$

Thus we obtain

$$\mathbf{E} \left[I_{m-n} \left(f_m \left(s_1, s_2, \ldots, s_n, \cdot \right) \right) | \mathcal{F}_{s_n} \right] = I_{m-n} \left(h_{m,s_1,s_2,\ldots,s_n} \right)$$

where the function h above is defined by

$$h_{m,s_1,s_2,\ldots,s_n} \left(s_{n+1}, \ldots, s_m \right) = f_m \left(s_1, \ldots, s_n, \cdot \right) \prod_{j=n+1}^{m} \mathbf{1}_{s_j \le s_n},$$

which proves that $h_{m,s_1,s_2,\ldots,s_n}$ is symmetric in the variables $s_{n+1}, \ldots, s_m$, and thus we can write

$$\begin{aligned}
&\mathbf{E} \left[I_{m-n} \left(f_m \left(s_1, s_2, \ldots, s_n, \cdot \right) \right) | \mathcal{F}_{s_n} \right] \\
&= (m-n)! \int_0^{s_n} \int_0^{s_{n+1}} \cdots \int_0^{s_{m-1}} f_m \left(s_1, \ldots, s_m \right) dW_{s_m} \cdots dW_{s_{n+1}}.
\end{aligned}$$

The following calculation now finishes the proof of the lemma:

$$\begin{aligned}
&\int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} \mathbf{E} \left[D^{(n)}_{s_n,\ldots,s_2,s_1} X | \mathcal{F}_{s_n} \right] dW_{s_n} \cdots dW_{s_1} \\
&= \sum_{m=n}^{\infty} \frac{m!}{(m-n)!} \int_0^1 \int_0^{s_1} \cdots \\
&\qquad \cdots \int_0^{s_{n-1}} \mathbf{E} \left[I_{m-n} \left(f_m \left(s_1, s_2, \ldots, s_n, \cdot \right) \right) | \, \mathcal{F}_{s_n} \right] dW_{s_n} \cdots dW_{s_1} \\
&= \sum_{m=n}^{\infty} m! \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} \left[\int_0^{s_n} \int_0^{s_{n+1}} \cdots \right. \\
&\qquad \left. \cdots \int_0^{s_{m-1}} f_m \left(s_1, \ldots, s_m \right) dW_{s_m} \cdots dW_{s_{n+1}} \right] dW_{s_n} \cdots dW_{s_1} \\
&= \sum_{m=n}^{\infty} I_m \left(f_m \right).
\end{aligned}$$

References

[1] R. Adler, *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Inst. Math. Stat., Hayward, CA., 1990.

[2] E. Alòs, O. Mazet, and D. Nualart, *Stochastic calculus with respect to Gaussian processes*, Annals of Probability, **29** (2001), 766–801.

[3] C. Borell, *Tail probabilities in Gauss space*, in: Vector Space Measures and Applications (Dublin 1977), Lecture Notes in Math., Springer-Verlag, **644** (1978), 71–82.

[4] R. A. Carmona and S. A. Molchanov, *Parabolic Anderson Model and Intermittency*, Memoirs A.M.S., 418, 1994.

[5] P. Cheridito and D. Nualart, *Stochastic integral of divergence type with respect to fractional Brownian motion with Hurst parameter H in* $(0, 1/2)$, Ann. Institut Henri Poincaré Probab. Stat., **41 (6)** (2005), 1049–1081.

[6] L. Decreusefond and A.-S. Üstünel, *Stochastic analysis of the fractional Brownian motion*, Potential Analysis, **10** (1997), 177–214.

[7] Y.-Z. Hu, B. Oksendal, and A. Sulèm, *Optimal consumption and portfolio in a Black-Scholes market driven by fractional Brownian motion*, Infin. Dimens. Anal. Quantum Probab. Relat. Top., **6 (4)** (2003), 519–536.

[8] I. Florescu and F. Viens, *Sharp estimation of the almost-sure Lyapunov exponent for the Anderson model in continuous space*, Probab. Theory and Rel. Fields, **135 (4)** (2006), 603–644.

[9] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, 1988.

[10] M. Ledoux, *Isoperimetry and Gaussian analysis*, in: Lectures on Probability Theory and Statistics (Saint-Flour, 1994), Lecture Notes in Math., Springer-Verlag, **1648** (1996), 165–294.

[11] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, 1991.

[12] M.-P. Malliavin and P. Malliavin, *Intégrales de Lusin-Calderon pour les fonctions biharmoniques*, Bulletin des Sciences Mathématiques, II. Ser., **101** (1977), 357–384.

[13] P. Malliavin, *Stochastic Analysis*, Springer-Verlag, 2002.

[14] P. Malliavin and A. Thalmaier, *Stochastic Calculus of Variations in Mathematical Finance*, Springer-Verlag, 2005.

[15] B. Maslowski and D. Nualart, *Stochastic evolution equations driven by fBm*, Journal of Functional Analysis, **202** (2003), 277–305.

[16] O. Mocioalca and F. Viens, *Skorohod integration and stochastic calculus beyond the fractional Brownian scale*, Journal of Functional Analysis, **222 (2)** (2004), 385–434.

[17] D. Nualart, *The Malliavin Calculus and Related Topics*, Springer-Verlag, New-York, 1995.

[18] D. Nualart and F. Viens, *Evolution equation of a stochastic semigroup with white-noise drift*, Ann. Probab., **28 (1)** (2000), 36–73.

[19] C. Rovira and S. Tindel, *On the Brownian directed polymer in a Gaussian random environment*, J. Functional Analysis, **222 (1)** (2005), 178–201.

[20] M. Talagrand, *Sample boundedness of stochastic processes under increment conditions*, Ann. Prob., **18 (1)** (1990), 1–49.

[21] S. Tindel, C. A. Tudor, and F. Viens, *Sharp Gaussian regularity on the circle and application to the fractional stochastic heat equation*, Journal of Functional Analysis, **217 (2)** (2004), 280–313.

[22] C. A. Tudor and F. Viens, *Statistical aspects of the fractional stochastic calculus*, Annals of Statistics, 2006, accepted.

[23] A.-S. Üstünel, *An Introduction to Analysis on Wiener Space*, Lecture Notes in Mathematics, Springer-Verlag, 1610, 1995.

[24] M. Weber, *Stochastic processes with values in exponential type Orlicz spaces*, Ann. Prob., **16** (1998), 1365–1371.

Andrew B. Vizcarra
Department of Mathematics
Purdue University
150 N. University St.
West Lafayette, IN 47907-2067
USA
e-mail: avizcarr@purdue.edu

Frederi G. Viens
Department of Statistics and Department of Mathematics
Purdue University
150 N. University St.
West Lafayette, IN 47907-2067
USA
e-mail: viens@purdue.edu

Progress in Probability, Vol. 59, 397–408

Nonlinear Markovian Problems in Large Dimensions

Bogusław Zegarliński

Abstract. We present nonlinear (hypercontractive) Markov semigroups, which are constructed as solutions of infinite-dimensional (semilinear) Cauchy problems, and provide smoothness and ergodicity results. We also discuss a nonlinear path space functional as well as certain nonlinear transition phenomena.

Mathematics Subject Classification (2000). 39B62, 34A34, 37L05.

Keywords. Semilinear Cauchy problems, infinite dimensions, coercive inequalities, nonlinear path space functionals, entropic switching phenomenon.

1. Introduction

In this paper we give a brief account on recent development in the area of nonlinear problems in large-dimensional systems. It is an interesting and rather unexplored domain of analysis which will likely attract considerable attention in the future.

We begin with describing recent results on semilinear Markovian Cauchy problems of [5] and [6]. In particular building up on the considerable achievements in the area of coercive inequalities for a variety of classes of Gibbs measures with nonequivalent tails, we expanded the classical PDE techniques to show existence of nonlinear semigroups on infinite-dimensional spaces which possess strong smoothing (hypercontractivity) and ergodicity properties.

Motivated by this development, we introduce a class of natural nonlinear jump-type semigroups for which a preliminary study suggest certain qualitatively new interesting behaviour.

Finally we discuss a possible nonlinear extension of the path-space functionals which is sensitive to the entropic switching for a family of random variables (while including the classical probability with respect to Wiener measure).

Supported in part by EPSRC GR/R90994/01 & GR/S61690/01.

2. Basic setup

We are interested in analysis on an infinite-dimensional measure space $(\Omega, \Sigma, \mu) = ((\mathbb{M}, \mathcal{B})^{\mathcal{R}}, \mu)$, with $\mathcal{R} \equiv \mathbb{N}, \mathbb{Z}^d, \ldots$, or $\mathcal{G} =$ some more complicated (infinite) graph. It is assumed that in this space we are given a Markov generator associated to a natural Dirichlet form $\mu|\nabla g|_2^2 = -\mu(gLg)$.

In this infinite-dimensional setup one considers the following families of coercive inequalities characterising (natural) Lipschitz random variables on a given measurable space.

- Coercive inequalities for Gibbs measures:

$$\mu v^2 F\left(\frac{v^2}{\mu v^2}\right) \leq c\mu|\nabla v|_2^2 \qquad (\mathbf{FS}_2)$$

  - Gaussian Tails [8], $F(x) \equiv \log(x)$,
    Log-Sobolev Inequality ([7])
    $\iff$ Hypercontractivity in $\mathbb{L}_p(\mu)$,
    $||e^{tL}f||_{L_{p(t)}} \leq ||f||_{\mathbb{L}_2}$
  - Sub-Gaussian Tails [12] ([10, 20]),
    $F(x) \sim (\log(x))^{\beta}$, $\beta \in (0,1)$,
    F-Inequality [1] $\iff$
    Hypercontractivity in $\mathbb{L}_{\Phi_p}(\mu)$, $\Phi_p(x) \equiv x^2 e^{p\cdot F(x^2)}$,
    $||e^{tL}f||_{\Phi_{p(t)}} \leq ||f||_{\mathbb{L}_2}$
  - Super-Gaussian Tails [3, 4, 20].
- Towers of nonequivalent inequalities:

$$\mu v^q \log\left(\frac{|v|^q}{\mu|v|^q}\right) \leq c\mu|\nabla v|_q^q \qquad (\mathbf{LS}_q)$$

  with $q \in [q_0, 2]$, $q_0 > 1$.

While for a long period of time much of the activity in the related area was concentrated on a singular example of the logarithmic Sobolev inequality, the development of recent years providing numerous families of coercive inequalities for variety of qualitatively different Gibbs measures allows us to start thinking of it more like of an extended theory.

3. Nonlinear Markovian Cauchy problems

Coercive inequalities provide a strong technical tool for the PDE theory in finite dimensions. Thus, given coercive inequalities for Gibbs measures on an infinite-dimensional space, it is natural to ask if it is possible to solve certain semilinear Cauchy problems as well as to provide an interesting characterisation of their solutions. In [5] and [6] we made an initial progress in understanding of a family of infinite-dimensional semilinear problems including in particular the following ones:

$$\begin{array}{lcl} \frac{\partial}{\partial t} u(t) & = & Lu(t) + \frac{\lambda}{2} u(t)\ F\left(\frac{u^2(t)}{\mu u^2(t)}\right) \\ u(0) & = & f \end{array} \qquad (\mathbb{C})$$

where L is a Markov generator satisfying $\mu f L g = -\mu \nabla f \cdot \nabla g$.

Firstly, we remark that *the nonlinearity may be neither globally Lipschitz nor monotone*. Secondly, the equation is nonlocal in the sense that the nonlinear term depends on expectation with a given measure. While we are able to treat also some local nonlinearities, the interest of having a normalised quantity inside the function F is in the fact that, due to condition $F(1) = 0$, in this case the nonlinearity vanishes on constants. Thus we have preserved one of the key features of a Markov generator.

The essence of our method lies in the fact that, provided that the coupling constant λ is sufficiently small, because of the coercive inequality the linear Markov generator L dominates the nonlinearity so that the total operator is monotone. The corresponding works are as follows:

- In the first of the cited works we studied measures with Gaussian Tails, [5], we take $F(x) \equiv \log(x)$; that is we work under the assumption that the logarithmic Sobolev inequality holds, $\mu \in \mathbf{LS}_2(c)$.
- In the second work we considered the measures with sub-Gaussian tails, [6]. In this case, $F(x) \sim (\log(x))^\beta \chi(x \geq 1),\ \ \beta \in (0,1)$ and $\mu \in \mathbf{FS}_2$.

4. Examples II

Problems with nonlocal nonlinearity have a long and interesting history. We mention here two important examples:

- Nonlinear Schrödinger Equation [2, 9, 15]:

$$i\hbar \frac{\partial}{\partial t} \Psi = -\frac{\hbar}{2m} \Delta \Psi + V\Psi + \frac{1}{2} kT \Psi \log \frac{\Psi^* \Psi}{\int \Psi^* \Psi}.$$

The normalisation in quantum mechanics is required to provide an interpretation of probability density to the modulus of the wave function. Moreover in case of the logarithmic nonlinearity one preserves a natural notion of probabilistic independence for noninteracting particles.

- Gelfand's Problem:

$$0 = \Delta \varphi + M \frac{e^\varphi}{\int e^\varphi}.$$

This problem appears in many domains of mathematics and theoretical physics, including for example problems of thermomechanics, thermodynamics of selfgravitating gas of charged particles, self-dual gauge theory, and others.

5. The product case

Before we get to the presentation of the general results, we would like to demonstrate that even in the simple product situation when interaction is not present one encounters interesting (and slightly unexpected) phenomena in large-dimensional asymptotics. The configuration space is as before, $\Omega \equiv \mathbf{M}^{\mathbb{N}} \ni \omega \equiv (\omega_j)_{i\in\mathbb{N}}$, where $\mathbf{M}$ is a smooth connected Riemannian, $\dim \mathbf{M} < \infty$, but we assume that the underlying measure is of simple product type, $\mu_0 \equiv \otimes_{i\in\mathbb{N}} \nu_i$, where $\forall i \in \mathbb{N}\ \ \nu_i = \nu_0 \in \mathbf{LS}_2(c_0)$. In such a situation, if the nonlinear interaction involves $F(x) = \log x$, one has a natural factorisation property for initial data of product type. This admits a natural interpretation of probabilistic independence when no many-body interaction between particles is present.

The key mathematical properties are as follows.

- **Existence and uniqueness at weak coupling.** *If* $\lambda \le \frac{1}{c_0}$ *and* $F(x) = \log x$, *then* $\exists!\, u$ *solution of* $(\mathbb{C})$ *for product type sufficiently smooth initial data* $u(t=0) = \prod_{k=1,\dots,n} f_k(\omega_{j_k})$
- **Particle structure for product initial data.**
  - $H_n \equiv \{v = \prod_{k=1,\dots,n} f_k(\omega_{j_k})\}$ is invariant for $\mathfrak{L}v \equiv Lv + \frac{\lambda}{2} v \log \frac{|v|^2}{\mu|v|^2}$.
  - $\exists \varepsilon > 0\, \forall n \in \mathbb{N}\ \ \forall \Psi \in H_n \quad \langle \Psi, \mathfrak{L}\Psi \rangle_{L_2} \le -n \cdot \varepsilon$.
- **Ergodicity** $\lambda < \frac{1}{c_0}$.
  - $\exists\ \lim_{t\to\infty} \mu_0 u_t$.
  - $\mu_0(u_t - \mu_0 u_t)^2 \le Const\, e^{-\varepsilon t} \mu_0(f - \mu_0 f)^2$.
- **Ergodicity breakdown** $\lambda = \frac{1}{c_0}$.

  $\exists\ \infty$ stationary solutions if $(\mathbf{LS}_2)$ holds with equality sign.

  This can be explicitly demonstrated for product Gaussian's (a more general result follows from the works of Rothaus [13, 14]*).*
- **Existence vs nonexistence for large** λ. Let $\nu_0 \in \mathbf{LS}_q$, $q \in (1,2)$, i.e.,

  $$\nu_0(v^q \log \frac{v^q}{\nu_0 v^q}) \le \nu_0 \sum_i |\nabla_i v|^q$$

  (ν_0 must have tails decaying faster than Gaussian's [4]).
  - $\forall \lambda \in \mathbb{R}\ \exists!\, u(t)$ solution of the *free problem* $(\mathbb{C})$ (i.e., with product measure μ_0), for product type time zero data;
  - *not true if interaction is introduced !*

6. Existence and properties of the solutions

The following result providing existence and basic properties of the solution justifies also the name Markovian Cauchy problem.

Theorem 6.1. *Suppose* $\mu \in \mathbf{FS}_2$ *and* $\lambda \in (0, \frac{1}{c})$. *Then a (weak) solution of the semilinear Cauchy problem* $(\mathbb{C})$ *exists and has the following properties.*

- (Constants preservation)
  $f = a \cdot \mathbf{1} \;\; \Rightarrow \forall t \geq 0, \;\; u(t) = a \cdot \mathbf{1}, \quad \mu$-a.e.
- (Positivity)
  $f \geq 0 \;\; \Rightarrow \forall t \geq 0, \;\; u(t) \geq 0, \quad \mu$-a.e.
- (Boundedness)
  $||f||_\infty < \infty \Rightarrow \forall t \geq 0, \;\; ||u(t)||_\infty \leq ||f||_\infty$
- ($\mathbb{L}_2$ - Contractivity)
  $||f||_2 < \infty \Rightarrow \forall t \geq 0, \;\; ||u(t)||_2 \leq ||f||_2$
- (Uniqueness) A weak solution of the semilinear Cauchy problem ($\mathbb{C}$) is unique.

In the infinite-dimensional setting it is interesting that the nonlinear semigroup provided by the solution of the Markovian Cauchy problem possesses the following strong smoothing and ergodicity properties.

- (Hypercontractivity)
  - If $\mu \in \mathbf{LS}_2(c)$, then $\exists C(t) \in (0, \infty)$
    $$||u(t)||_{\mathbb{L}_{q(t)}} \leq C(t)||f||_2$$
    with $q(t) = 1 + \exp(\alpha t)$.
  - If $\mu \in \mathbf{FS}_2$, then $\exists c(t) \in (0, \infty)$
    $$||u(t)||_{\Phi(t)} \leq c(t)||f||_2$$
    with $\Phi(t) \to \Phi_\infty$ as $t \to \infty$.
- (Long time behaviour)
  Suppose $m \in (0, \infty)$ is the best constant s.t.
  $$m\,\mu(v - \mu v)^2 \leq \mu|\nabla v|^2. \qquad (\mathbf{SG})$$
  If $\mu \in \mathbf{FS}_2$ and $\lambda \in (0, (c + 1/m)^{-1})$, then $\exists\, M \in (0, \infty)$, $\forall t \geq 0$,
  $$\mu(u_t - \mu u_t)^2 \leq e^{-Mt}\mu(f - \mu f)^2.$$
- (Gradient estimates: Gaussian tails case)
  Suppose
  $$\Gamma_2(z) \geq \gamma|\nabla|\nabla z||^2 + \rho|\nabla z|^2 \qquad (\mathbf{BE})$$
  with some $\rho, \gamma \in (0, \infty)$, where
  $$\Gamma_2(z) \equiv \frac{1}{2}L|\nabla z|^2 - \nabla z \cdot \nabla Lz.$$
  Then for any $\lambda \in (0, (\gamma \wedge 1)\rho)$, the solution u_t satisfies
  $$\mu|\nabla u_t|^2 \leq e^{-2(\rho-\lambda)t}\mu|\nabla f|^2$$
  provided that $\mu|\nabla f|^2 < \infty$.

Remark 6.2. Unlike as in the linear case, the proof of this result makes an essential use of the first extra term on the r.h.s. of (BE).

Uniform hypercontraction. Consider the following Cauchy problem with a time-dependent normalisation:

$$\begin{array}{rcl} \frac{\partial}{\partial t}u(t) & = & Lu(t) + \frac{\lambda}{2}u(t)\log\left(\frac{u^2(t)}{||u||_q^2(t)}\right) \\ u(0) & = & f \end{array} \qquad (\mathbb{C}_q)$$

where $q \equiv q(t) \equiv 1 + e^{\alpha t}$ with some $\alpha \in (0, \infty)$. Let u_t and v_t be a solution of the Cauchy problem, with initial data f and g, respectively. In this setup, for $w_t \equiv u_t - v_t$, one has the following formal computation (which can be made rigorous by mollification, use of the notion of the weak solution and, after obtaining an integral inequality, removal of the smoothing). We have

$$\frac{d}{dt}\log ||w_t||_q$$
$$= -\frac{\dot{q}}{q^2}\log ||w_t||_q^q + \frac{1}{||w_t||_q^q}\mu\left(|w_t|^q\frac{\dot{q}}{q}\log |w_t|^q + |w_t|^{q-1}\mathrm{sign}(w_t)\frac{\partial}{\partial t}(w_t)\right).$$

Hence one gets

$$\frac{d}{dt}\log ||w_t||_q = \frac{1}{||w_t||_q^q}\mu\left(|w_t|^q\frac{\dot{q}}{q}\log\frac{|w_t|}{||w_t||_q} + |w_t|^{q-1}\mathrm{sign}(w_t)L(w_t)\right)$$
$$+ \frac{1}{||w_t||_q^q}\frac{\lambda}{2}\mu\left(|w_t|^{q-1}\mathrm{sign}(w_t)\left(u_t\log\frac{u_t^2}{||u_t||_q^2} - v_t\log\frac{v_t^2}{||v_t||_q^2}\right)\right).$$

We note that with $u_t(\alpha) \equiv \alpha u_t + (1-\alpha)v_t$, one has

$$\begin{array}{l} \mu\left(|w_t|^{q-1}\mathrm{sign}(w_t)\left(u_t\log\frac{u_t^2}{||u_t||_q^2} - v_t\log\frac{v_t^2}{||v_t||_q^2}\right)\right) \\ = \int_0^1 d\alpha\mu\left(|w_t|^q\log\frac{u_t(\alpha)^2}{||u_t(\alpha)||_q^2}\right) + \int_0^1 d\alpha\mu\left(|w_t|^q\right) \\ \quad - \int_0^1 d\alpha\mu\left(|w_t|^{q-1}\mathrm{sign}(w_t)u_t(\alpha)\right)\cdot\frac{2||u_t(\alpha)||_q\frac{d}{d\alpha}||u_t(\alpha)||_q}{||u_t(\alpha)||_q^2} \\ \leq \mu\left(|w_t|^q\log\frac{|w_t|^2}{||w_t||_q^2}\right) + 4\mu\left(|w_t|^q\right). \end{array}$$

Combining this with the previous computation we arrive at the differential inequality

$$\frac{d}{dt}\log ||w_t||_q \leq \frac{1}{||w_t||_q^q}\mu\left((\lambda + \frac{\dot{q}}{q})|w_t|^q\log\frac{|w_t|}{||w_t||_q} + |w_t|^{q-1}\mathrm{sign}(w_t)L(w_t)\right)$$
$$+ 2\lambda.$$

Assuming that the logarithmic Sobolev inequality is satisfied with a coefficient $c \in (0, \infty)$ and that

$$\lambda + \frac{\dot{q}}{q} = \lambda + \alpha\frac{1}{1 + \exp\{-\alpha t\}} \leq \frac{1}{c}$$

we conclude that

$$||u_t - v_t||_q \le e^{2\lambda t}||f - g||_2.$$

From this inequality one sees that the corresponding semigroup is uniformly hypercontracting (in the sense of the corresponding metric). While in the linear case this comes from the hypercontractivity in the sense of the norm, in nonlinear it may be possible to have different behaviour in different directions. Besides other things the uniform contraction property implies continuity with respect to the initial data as well as the uniqueness of the solution.

7. Nonlinear exponential semigroups

From the point of view of studying Harnack type properties it seems to be natural to consider an associated problem formally obtained by the substitution $v(t) = \log\ u(t)^2$ which transforms the problem ($\mathbb{C}$) to the following one.

$$\begin{array}{lcl} \frac{1}{2}\frac{\partial}{\partial t}v(t) & = & Lv(t) + \frac{1}{4}|\nabla v(t)|^2 - \frac{\lambda}{2}\big\{ \log \mu e^{v(t)} - v(t)\big\} \\ v(0) & = & \log f^2. \end{array} \qquad (\log \mathbb{C})$$

It is interesting to observe that the operator in the curly bracket satisfies a maximum principle and can be interpreted as a nonlinear jump operator. This naturally leads us to study the following nonlinear Markov semigroups (work in progress by [11]).

7.1. Exponentially twisted jump process

Define, with $\psi \equiv \exp$,

$$\mathcal{L}(v) \equiv \psi^{-1} \circ \nu \circ \psi(v) - v.$$

Theorem 7.1. *The Cauchy problem*

$$\begin{array}{lcl} \frac{\partial}{\partial t}v(t) & = & \mathcal{L}(v(t)) \\ v(t=0) & = & f \end{array}$$

has a unique solution $\mathcal{P}_t f$ such that

$$f \ge 0 \Rightarrow \mathcal{P}_t(f) \ge 0,$$
$$\forall a \in \mathbb{R} \quad \mathcal{P}_t(f + a) = \mathcal{P}_t(f) + a,$$
$$\nu(\mathcal{P}_t(f) - \nu\mathcal{P}_t(f))^2 \le e^{-2t}\nu(f - \nu f)^2.$$

Existence of a solution is here a simple matter. The first two properties imply that the semigroup is Markov. The second one seems to be rather a striking property as for a nonlinear semigroup, saying that on a level of linear combinations with constants we have strict linearity.

Remark 7.2. It is interesting to note that in fact $\mathcal{P}_t f$ has an explicit representation similar to the linear Poisson semigroup $P_t f \equiv e^{-t}f + (1 - e^{-t})\nu f$ with generator $Lf \equiv \nu f - f$, from which one can see that

$$\mathcal{P}_t f \ \neq \ \psi^{-1} \circ P_t \circ \psi f,$$

that is, our semigroup is not given by conjugation of the linear Poisson semigroup.

7.2. Nonlinear jump process in dimension ∞

Consider a product measure $\mu \equiv \otimes_{i\in\mathbb{Z}^d}\nu_i$, product probability measure on $\Omega = \mathbb{M}^{\mathbb{Z}^d}$, with $\nu_i \sim \nu$ on $(\mathbb{M}_i, \mathcal{B}_i) \sim (\mathbb{M}, \mathcal{B})$. With this notation we introduce the following nonlinear operator whose domain includes all bounded measurable cylinder functions.

Nonlinear generator. $\mathfrak{L}(v) \equiv \sum_{i\in\mathbb{Z}^d} \mathfrak{L}_i(v) \equiv \sum_{i\in\mathbb{Z}^d} \psi^{-1}\circ\nu_i\circ\psi(v) - v.$

Theorem 7.3. *The Cauchy problem*

$$\begin{aligned} \frac{\partial}{\partial t}v(t) &= \mathfrak{L}(v(t)) \\ v(t=0) &= f \end{aligned}$$

has a unique solution given by a nonlinear Markov semigroup $(\mathfrak{P}_t)_{t\in\mathbb{R}^+}$ *satisfying*

$\star$ *(Locality) For* $\Lambda \subset \mathbb{Z}^d$

$$f(\omega) = f(\omega_\Lambda) \implies \mathfrak{P}_t f(\omega) = \mathfrak{P}_t f(\omega_\Lambda).$$

$\star\star$ *(Super-invariance property)*

$$\mu f \;\le\; \mu\mathfrak{P}_t f.$$

8. Nonlinear path space functionals (NPSF)

In this section we propose a path space description of a process via a family of nonlinear functionals – associated to some Orlicz function – which generalise a notion of the expectation of random variables. While restricted to a characteristic function, they coincide with the probabilities given by the Wiener measure. If the Orlicz function is simply a monomial, the functionals can be understood as the joint (higher) moment of random variables. An interesting phenomenon shows up in case of Orlicz functions with doubling property. In this case the short time correlations may differ dramatically from the long time correlations. *Old news have a different effect as the very recent ones.*

For Φ an Orlicz function, s.t. $\Phi(2\cdot x) \le C\Phi(x)$, with $C \in (0,\infty)$, define

$$E_\rho^\Phi\left(f(X_t)\right) \;\equiv\; \int \Phi\left(f(x)\Phi^{-1}(p_t(x,y)\rho(y))\right) dxdy,$$

$$\begin{aligned} E_\rho^\Phi\left(f_2(X_{t_2}), f_1(X_{t_1})\right) \;\equiv\; &\int dx_2dx_1dy \\ &\Phi\Big(f_2(x_2)\Phi^{-1}\big(p_{t_2-t_1}(x_2,x_1)(\Phi(f_1(x_1)\Phi^{-1}(p_{t_1}(x_1,y)\rho(y)))\cdots\big)\Big), \end{aligned}$$

and for any $n\in\mathbb{N}$,

$$\begin{aligned} E_\rho^\Phi\left(f_n(X_{t_n}),\ldots,f_1(X_{t_1})\right) \;\equiv\; &\int dx_n\cdots dx_1dy \\ &\Phi\Big(f_n(x_n)\Phi^{-1}\big(p_{t_n-t_{n-1}}(x_n,x_{n-1})\Phi\left(\cdots(\Phi(f_1(x_1)\Phi^{-1}(p_{t_1}(x_1,y)\rho(y)))\cdots\right)\big)\Big), \end{aligned}$$

where within the dots in the integrand we apply consecutively the functions $\Phi\Big(f_{j+1}(x_{j+1})\Phi^{-1}\big(p_{t_{j+1}-t_j}(x_{j+1},x_j)\ \circ\ \big)$.

Properties of NPSFs'.

(i) $E^{\Phi}_{\rho}\left(\chi_{A_n}(X_{t_n}),\ldots,\chi_{A_1}(X_{t_1})\right) = \mathbb{P}_{\rho}\Big(X_{t_j}\in A_j, j=1,\ldots,n\Big)$, where $\mathbb{P}_{\rho}$ stands for the Wiener measure with initial distribution ρ.

(ii) If $\Phi(x)=x^{\kappa}$ is a monomial of degree $\kappa>0$, then

$$E^{\Phi}_{\rho}\left(f_n(X_{t_n}),\ldots,f_1(X_{t_1})\right) = \mathbb{E}_{\rho}\left(\prod_{j=1,\ldots,n} f_j^{\kappa}(X_{t_j})\right).$$

(iii) $\forall\ \Phi\in\Delta_2$ (i.e., satisfying $\Phi(2x)\le C\Phi(x)$ with some $C\in(0,\infty)$ for all x), $\exists\ \vartheta:\mathbb{R}\to\mathbb{R}$

$$E^{\Phi}_{\rho}\left(f_n(X_{t_n}),\ldots,f_1(X_{t_1})\right) \le \mathbb{E}_{\rho}\left(\prod_{j=1,\ldots,n} \vartheta(f_j)(X_{t_j})\right).$$

(iv) Different Karamata-Matuszewska indices $\kappa_+\neq\kappa_-$.

$\Big\{\ \Phi(x)\sim x^{\kappa_-}$ for $x\approx 0$ and $\Phi(x)\sim x^{\kappa_+}$ for $x\approx\infty\ \Big\}$.

Entropic switch $\Longrightarrow$ variable long and short time behaviour.

$$E^{\Phi}_{\rho}\left(f_n(X_{t_n}),\ldots,f_1(X_{t_1})\right) \approx \mathbb{E}_{\rho}\left(\prod_{j=1,\ldots,n} f_j^{N_j}(X_{t_j})\right)$$

with $N_j\approx\kappa_-$ if t_j-t_{j-1} large and $N_j\approx\kappa_+$ if t_j-t_{j-1} small.

(v) Mean value inequality. For $x\in\mathbb{R}^d$ and $t>0$, let $\mathcal{O}_t(x)$ be a heat ball. There exists a (nonlinear) function $\theta:\mathbb{R}\to\mathbb{R}$ such that

$$\frac{1}{|\mathcal{O}_t(x)|}\int_{\mathcal{O}_t(x)} E^{\Phi}_{z}\left(f(X_t)\right)\ dz\ \le\ E^{\Phi}_{x}\left(\theta(f)(X_t)\right)$$

where

$$E^{\Phi}_{z}\left(f(X_t)\right) \equiv \int \Phi\left(f(w)\Phi^{-1}\big(p_t(w,z)\big)\right)dw.$$

9. On entropic switching

In this section we discuss briefly an interplay of entropy and nonlinearity. We begin from recalling the following problem in the theory of functional equations.

Multiplicative translation equation (Aczél, Kuczma)**.**

$$G(z,rs) = G\left[G(z,r),s\right].$$

It has the general solution

$$G(z,s) = G_{\Phi}(z,s) \equiv \Phi\left(\Phi^{-1}(z)s\right).$$

Given such a solution we introduce the following object.

Index $\boldsymbol{\iota}$: *For a probability density ρ on $\mathbb{R}^n$ and an Orlicz function satisfying the doubling property $\Phi \in \Delta_2$ define*

$$\boldsymbol{\iota}_\rho(\Phi) = \int_0^\infty \int_{\mathbb{R}^n} G_\Phi(\rho(x), s)\; e^{-s}\; d_n x\; ds.$$

One can quickly see that in case of monomials, the index does not depend on the density ρ and is equal to Euler Gamma functions $\mathbf{\Gamma}_{\text{Euler}}$ of the degree of the monomial. If a reader is keen on nonstandard analysis, we mention that, by the Karamata representation theorem, any element Φ of Δ_2 can be described by a monomial times a slowly varying function and thus it can be regarded as a representation of a nonstandard number differing from its real (standard) part – equal to the power of the corresponding monomial – by an infinitesimal number (represented by a slowly varying function); there are of course plenty of slowly varying functions, that is, there is a large set of infinitesimal numbers (of cardinality larger than $\boldsymbol{c}$). In this way one could think of the index $\boldsymbol{\iota}$ as an extension of the Gamma function.

Properties of the index $\boldsymbol{\iota}$.

$\rho_\sigma(x) \equiv \sigma^{-n}\rho(x/\sigma)$, ($0 \swarrow \sigma \nearrow \infty$),

$$S(\rho_\sigma) \to \pm\infty \implies \boldsymbol{\iota}(\rho_\sigma, \Phi) \to \mathbf{\Gamma}_{\text{Euler}}(\kappa_\pm),$$

where $S(\rho_\sigma)$ denotes the Shannon entropy of the signal ρ_σ and $\kappa_\pm$ are Karamata-Matuszewska indices of Φ.

Particularly interesting is the situation where the asymptotic indices are different. We remark also that the index is well defined for all probability densities even when Shannon's entropy is not finite. (In fact, it is also more sensitive than the entropy with respect to mixtures.)

Two further properties concern the collective behaviour of systems of varying dimension subjected to a common nonlinear action-amplified-reaction mechanism.

Collective switching $\rho_N(\mathbf{x}) \equiv \prod_{j=1,\ldots,N} \rho_j(x_j)$,

$$\mp \inf_N \frac{1}{N} S(\rho_N) > 0 \implies \boldsymbol{\iota}_{\rho_N}(\Phi) \to \mathbf{\Gamma}_{\text{Euler}}(\kappa_\pm)$$

where $\kappa_\pm$ are Karamata-Matuszewska indices of Φ.

Many systems originate from a single activated cell (e.g. a biological cell or perhaps a small business). In the process of expansion a number of cells may be growing (whether by multiplication, setting up subsidiaries or buying up others). If the cells participate in a nonlinear collective mechanism, with the growth the system may be subjected to switching away to a qualitatively different region of existence. To avoid being driven to undesired state, a possible survival strategy involves loosing excessive number of cells. *Somehow a system does know what is its right size.*

The next property is as follows.

Ideal balance principle,

$$\frac{1}{N}S(\rho_N) \approx 0 \implies$$

$$\min\{\mathbf{\Gamma}_{\text{Euler}}(\kappa_\pm)\} < \boldsymbol{\iota}_{\rho_N}(\Phi) < \max\{\mathbf{\Gamma}_{\text{Euler}}(\kappa_\pm)\}.$$

A growing system may perform in a stable way only if positive and negative influences are in a delicate balance. Frequently there is a robust domain of such states allowing for large amplitudes of opposite elements. In some way these seem to be the principle features of stability of large systems (eco-, social, political systems or alive multicell organisms).

Remark 9.1. As one of the consequences, a careful reader may notice that a description of nonlinear electric circuits suggested by N. Wiener (cf. mid of p. 96 in [16]) cannot be adequate in general.

References

[1] F. Barthe, P. Cattiaux and C. Roberto, *Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry*, Revista Math. Iberoamericana, **22** (2006), 993–1067.

[2] I. Bialynicki-Birula and J. Mycielski, *Nonlinear wave mechanics*, Annals of Physics, **100 (1-2)** (1976), 62–93.

[3] S. G. Bobkov and M. Ledoux, *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*, Geom. Funct. Anal., **10 (5)** (2000), 1028–1052.

[4] S. G. Bobkov and B. Zegarliński, *Entropy Bounds and Isoperimetry*, Memoirs AMS, **829** (2005).

[5] P. Fougères and B. Zegarliński, *Semi-linear problems in infinite dimensions*, J. Funct. Analysis, **228** (2005), 39–88.

[6] P. Fougères, C. Roberto, and B. Zegarliński, *Semi-linear problems in infinite dimensions, sub-Gaussian tails*, preprint, Imperial College, 2005.

[7] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math., **97** (1976), 1061–1083.

[8] A. Guionnet and B. Zegarliński, *Lectures on Logarithmic Sobolev Inequalities*, Séminaire de Probabilités 36, Lecture Notes in Math., Springer, **1801** (2003).

[9] T. W. B. Kibble, *Relativistic models of nonlinear quantum mechanics*, Commun. Math. Phys., **64** (1978), 73–82.

[10] R. Latala and K. Oleszkiewicz, *Between Sobolev and Poincaré*, Geometric Aspects of Functional Analysis, Israel Seminar (GAFA) 1996-2000, Lecture Notes in Math., Springer, **1745** (2000), 147–168.

[11] R. Olkiewicz, X. Lihu, and B. Zegarliński, *Nonlinear problems in infinite interacting particle systems*, preprint, Imperial College.

[12] C. Roberto and B. Zegarliński, *Orlicz-Sobolev inequalities for sub-Gaussian measures and ergodicity of Markov semi-groups*, J. Funct. Anal., **243 (1)** (2007), 28–66.

[13] O. S. Rothaus, *Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators*, J. Funct. Anal., **42 (1)** (1981), 110–120.

[14] O. S. Rothaus, *Logarithmic Sobolev inequalities and the spectrum of Sturm-Liouville operators*, J. Funct. Anal., **39 (1)** (1980), 42–56.

[15] S. Weinberg, *Testing quantum mechanics*, Ann. Phys., **194 (2)** (1989), 336–386.

[16] N. Wiener, *Nonlinear Problems in Random Theory*, MIT Press, 1958.

[17] B. Zegarliński, *On log-Sobolev inequalities for infinite lattice systems*, Lett. Math. Phys., **20** (1990), 173–182.

[18] B. Zegarliński, *Log-Sobolev inequalities for infinite one-dimensional lattice systems*, Commun. Math. Phys., **133** (1990), 147–162.

[19] B. Zegarliński, *Dobrushin uniqueness theorem and logarithmic Sobolev inequalities*, J. Funct. Anal., **105** (1992), 77–111.

[20] B. Zegarliński, *Entropy bounds for Gibbs measures with non-Gaussian tails*, J. Funct. Anal., **187 (2)** (2001), 368–395.

[21] B. Zegarliński, *Entropic Switch*, Essays for Science Studies, Vol. I, Matrix Press, 2005.

Bogusław Zegarliński
Department of Mathematics
Imperial College London
UK
e-mail: `b.zegarlinski@imperial.ac.uk`

Stochastic Methods in Financial Models

Progress in Probability, Vol. 59, 411–433

A Tychastic Approach to Guaranteed Pricing and Management of Portfolios under Transaction Constraints

Jean-Pierre Aubin and Patrick Saint-Pierre

Abstract. Dynamic guaranteed pricing and management of a portfolio under transaction constraints is actually a problem straightforwardly set in terms of guaranteed capture basin of a time-dependent target that is viable in a time-dependent environment under (stochastic or tychastic) uncertain systems. The knowledge of the properties of "capture basin" of targets viable in evolving environments under an uncertain evolutionary system can be used for obtaining the corresponding properties for portfolios. They yield at each time *both the evaluation of the capital and the transaction rule*. They can be computed by viability algorithms and software providing the valuation of optimal portfolio and the management of their evolution. The capital function, which is actually the value function of a differential game, is the solution to a free boundary problem for nonlinear partial differential equations with discontinuous coefficients. This survey provides several examples.

Mathematics Subject Classification (2000). Primary 91B28; Secondary 91A23, 91A25, 91A80.

Keywords. Portfolio, pricer, management, options, viability, tychastic, dynamical game, impulse control.

1. Introduction

Dynamic guaranteed pricing and management of a portfolio under transaction constraints is actually a problem straightforwardly set in terms of guaranteed capture basin of a time-dependent target viable in a time-dependent environment under (stochastic or tychastic) uncertain systems.

The authors thank Giuseppe Da Prato, Halim Doss, Georges Haddad, Hélène Frankowska, Dominique Pujal and Jerzy Zabczyk for their hidden contribution, as well as the anonymous reviewer and the editors for their careful reviews and suggestions.

The time-dependent environment can take into account

1. constraints on holding the number of available shares, the cumulated number of shares, bounds on the transactions (up to the interdiction of transactions during certain periods), transaction costs, liquidity constraints, and so on
2. dividend payment for equity, coupon schedules and values, cliquet option (settling periodically and resetting the strike at the spot levels), at-the-money options, etc.
3. bounds on asset prices and their returns
4. redemption for debts, refinancing when the capital reaches threshold values
5. value creation or return on capital.

Therefore, the knowledge of the properties of "capture basin" of targets viable in evolving environments under an uncertain evolutionary system can be used for obtaining the corresponding properties for portfolios. They yield at each time both the evaluation of the capital and the transaction rule. The capital function, which is actually the value function of a differential game, is the solution to a free boundary problem for nonlinear partial differential equations with discontinuous coefficients.

However, we are interested in algorithms and softwares providing the valuation of optimal portfolio and the management of their evolution for all these examples. Analytical closed form formulas are no longer available, as in the standard and familiar Black and Scholes formula. Once discretized in this natural formulation, the Viability Kernel Algorithms designed by [32, P. Saint-Pierre] computes both the guaranteed capture basin of a time-dependent target under (stochastic or tychastic) uncertain systems and the transaction rule. Hence they can be applied and adapted to these problems directly without solving the free boundary problems for nonlinear partial differential equations with discontinuous coefficients and differentiate the valuation function for finding the portfolios. Whenever the number of state and auxiliary variables is small enough (up to 4) to avoid the dimensionality curse, software has been developed for implementing this algorithm.

This paper surveys some results obtained in this direction. Viability issues have already been noticed in some settings by

1. [40, 42, J. Zabczyk], for discrete stochastic systems,
2. [5, 6, 7, J.-P. Aubin & G. Da Prato], [21, 22, 23, G. Da Prato & H. Frankowska], [8, J.-P. Aubin, G. Da Prato & H. Frankowska], [25, H. Doss] and [9, J.-P. Aubin & H. Doss] for stochastic viability,
3. [34, 35, 36, M. Soner & N. Touzi], for capturability under stochastic control systems,
4. [29, D. Pujal & P. Saint-Pierre] and [14, J.-P. Aubin, D. Pujal & P. Saint-Pierre], for tychastic control systems,
5. [11, Aubin & Haddad], for path-dependent evolution of prices,

to quote a few early papers on this topic.

We describe in the first section the additional constraints on the shares and their transactions, beyond a manifold of contracts. The dynamics are introduced in the second section and the conclusions in the third section.

2. Description of the model

2.1. State, regulatory and tychastic variables

We denote by

1. $i = 0, 1, \ldots, n$ assets ($i = 0$ denoting the non-risky asset),
2. T the exercise time, and, at each running date t, $0 \leq t \leq T$, $T - t$ denoting "time to maturity".

The variables of the financial systems considered in this study are

1. the "state variables" of the system made of
   - the prices of the assets $S(t) := (S_0(t), S_1(t), \ldots, S_n(t))$ ($S_0(t)$ being the price of the non-risky asset, and $(S_1(t), \ldots, S_n(t))$ the prices of the risky assets),
   - the number of shares of the assets making up the portfolio $P(t) := (P_0(t), P_1(t), \ldots, P_n(t))$,
   - the value (capital) $W(t) := P_0(t)S_0(t) + \sum_{i=1}^{n} P_i(t)S_i(t)$ of the portfolio, where $P_0(t)S_0(t)$ is the liquid component of the portfolio,
2. the "controls", which are the *transactions* of the risky assets $P'(t) := (P'_0(t), P'_1(t), \ldots, P'_n(t))$, described by the time derivatives or the number of shares,
3. the "tyches" (one of the classical Greek words encapsulating the concept of chance, used here in the sense of un-controlled disturbances, perturbations), which are the returns $R(t) := (R_0(t), R_1(t), \ldots, R_n(t))$, where

$$\forall\, t \geq 0, \;\; R_i(t) \;:=\; \frac{S'_i(t)}{S_i(t)} \;=\; \frac{d\log(S_i(t))}{dt} \text{ if } S_i(t) \;>\; 0,$$

   of the prices of the assets. Here, tyches play the role of random variables in probability and stochastic theories. They provide an alternative mathematical translation of evolution under uncertainty parallel to the usual mathematical translation by a diffusion in the framework of stochastic differential equations. Tyches range over a *tychastic set* (that could be itself a fuzzy set). The size of the tychastic subsets captures mathematically the concept of *"(tychastic) versatility"*, instead of "stochastic volatility": *The larger the tychastic set, the more "versatile" the uncertainty.*

2.2. The viability constraints

Viability theory deals with the problems of evolution under viability constraints bearing on state, regulatory and tychastic variables:

2.2.1. Financial constraints on state variables.

1. Constraints on prices,

$$\forall\, t \in [0,T],\;\; S(t) \in \mathbb{S}(t).$$

2. Constraints on the shares of the portfolio (liquidity constraints),

$$\forall\, t \in [0,T],\;\; P(t) \in \mathbb{P}(t, S(t), W(t)).$$

3. Constraints on the value of the portfolio describing guarantees by a *threshold function* $\mathbf{b}(t,S)$,

$$\forall\, t \in [0,T],\;\; W(t) \geq \mathbf{b}(t, S(t)).$$

4. *Cash-flows* are described by dates T_k payment functions $(S,W) \mapsto \pi(T_k, S, W)$ subtracting to the capital, at dates T_k, amounts $\pi(T_k, S(T_k), W(T_k))$ associated with functions $t \mapsto S(t)$ and $t \mapsto W(t)$.

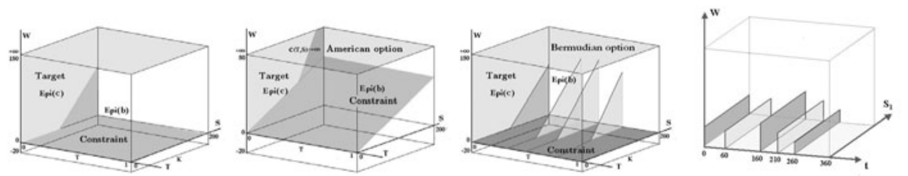


Figure 2.1. *Examples of threshold functions.*
From left to right, threshold functions for European, American, Bermudan options and cash flows. Financial Rules *involve constraints requiring that at each instant, the value of the portfolio must be larger than or equal to a threshold function depending on the time at maturity, the price and the number of shares of the portfolio.*

1. *For portfolios replicating European options, the threshold is equal to zero before the exercise time and to the contingent function at exercise time,*
2. *For portfolios replicating a type of American options, the threshold is equal to a given percentage of the price before the exercise time and to the supremum of this function and the contingent function at exercise time,*
3. *For portfolios replicating Bermudan options, the threshold is equal to zero except at a finite set of dates when it is a contingent function,*
4. *The threshold function can also describe a cash flow that has to be satisfied at each instant.*

No restriction is made in the choice of the threshold function which defines the "financial rules".

2.2.2. Financial constraints on tychastic variables. The returns must obey "tychastic constraints"

$$\forall\, t \in [0,T],\;\; R(t) \in \mathbb{R}(t, S(t), P(t), W(t))$$

where the set-valued map $\mathbb{R}(t, S(t), P(t), W(t))$ is called the *tychastic map*.

We provide below an example of a tychastic map in the case of one risky asset ($n = 1$): The interest rates of the non risky asset $R_0(t)$ are given and the returns $R(t) := R_1(t)$ of the risky asset satisfy

$$\forall\, t \in [0,T], \;\; R(t) \in \mathbb{R}(t, S(t), P(t), W(t)) := [R^\flat(t), R^\sharp(t)],$$

and in particular, when

$$\forall\, t \in [0,T], \;\; R(t) \in \mathbb{R}(t, S(t), P(t), W(t)) := [R - \nu(t), R + \nu(t)],$$

where the function $\nu(\cdot)$ is the tychastic versatility threshold.

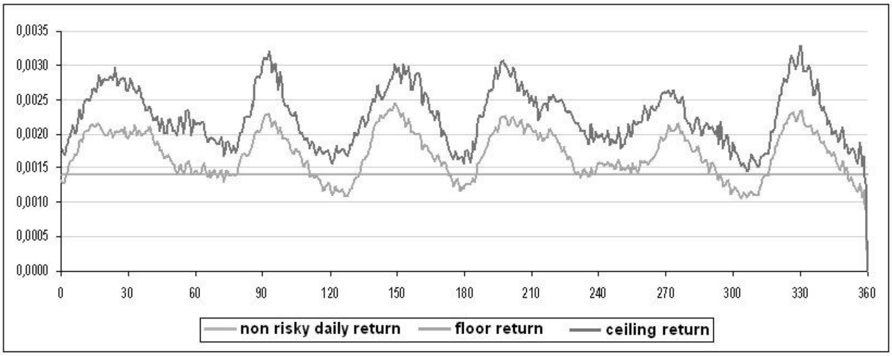


Figure 2.2. *Representation of tychastic uncertainty.*
The picture displays the daily interest rate of the non-risky asset (light gray line), of the daily floor (dark gray) and ceiling (black) returns of the risky asset describing the tychastic scenario.

2.2.3. Financial constraints on regulatory variables. Constraints on transactions are described by subsets $\mathbb{F}(t, S, P, W)$:

$$P'(t) \in \mathbb{F}(t, S(t), P(t), W(t)).$$

The two main examples of constraints on transactions are:

1. *Trading Constraints*, of the form $|P_i'(t)| \leq \gamma_i(t)$, $i = 1, \ldots, n$, $0 \leq \gamma_i(t) \leq +\infty$, the case $\gamma_i(t) = 0$ translating an impossibility of trading at date t, the case $\gamma_i(t) = +\infty$ expressing the absence of trading constraints at this date.
2. *Transaction Costs*,

$$\sum_{i=0}^{n} P_i'(t) S_i(t) = -\delta(P'(t), P(t), S(t), W(t)).$$

"Self-financed portfolios" are the special case when the transaction cost function does not involve transactions, such as

$$\sum_{i=0}^{n} P_i'(t) S_i(t) = 0 \text{ or, more generally } \sum_{i=0}^{n} P_i'(t) S_i(t) = \varphi(t, S(t)) W(t).$$

This is an important case because the shares of the portfolio are no longer state variables, but controls (see Section 3.5).

2.3. The dynamics

The state variables (S_i, P_i, W) must evolve in the time-dependent constrained set $\mathcal{K}(t)$ defined by

$$\mathcal{K}(t) := \{(S,P,W) \mid S \in \mathbb{S}(t),\ P \in \mathbb{P}(t,S,W)\ \&\ W \geq \mathbf{b}(t,S)\}. \qquad (2.1)$$

In order to define option contracts where the option is exercised at an opportune or propitious time $t^\star$, we introduce a time-dependent target $\mathcal{C}(t) \subset \mathcal{K}(t)$ and require that at time $t^\star$,

$$(S(t^\star), P(t^\star), W(t^\star)) \in \mathcal{C}(t^\star)$$

. An example of a target is associated with a "target function" $\mathbf{c}(t,S) \geq \mathbf{b}(t,S)$ in the following way:

$$\mathcal{C}(t) := \{(S,P,W) \mid S \in \mathbb{S}(t),\ P \in \mathbb{P}(t,S,W)\ \&\ W \geq \mathbf{c}(t,S)\}. \qquad (2.2)$$

This means that the option is exercised at the first time $t^\star$ when

$$W(t^\star) \geq \mathbf{b}(t^\star, S(t^\star)).$$

Other option contracts are obtained by taking $\mathbf{b}(t,S) = 0$ and $\mathbf{c}(t,S) = \max(S-K,0)$: the option is exercised as soon as there exists a time $t^\star$ such that $W(t^\star) \geq \max(S(t^\star)-K,0)$. Some contracts may involve as target functions the valuation function of other contracts, as in "barrier options".

The dynamical system governing the evolutions of the state variables: for $i = 0,1,\ldots,n$,

$$\left\{\begin{array}{lrcl} \text{(i)} & S_i'(t) & = & R_i(t)S_i(t),\ i=0,\ldots,n,\ \text{where}\ R(t) \in \mathbb{R}(t,S(t),P(t),W(t)), \\ \text{(ii)} & P_i'(t) & = & u_i(t),\ i=0,\ldots,n,\ \text{where}\ u(t) \in \mathbb{F}(t,S(t),P(t),W(t)), \\ \text{(iii)} & W'(t) & = & R_0(t)W(t) + \displaystyle\sum_{i=1}^{n} P_i(t)S_i(t)(R_i(t)-R_0(t)) + \sum_{i=0}^{n} u_i(t)S_i(t), \end{array}\right. \qquad (2.3)$$

parameterized by the controls $u_i := P_i'$, which are the transactions, and the tyches R_i, which are the rates of the risky assets. This is a "*tychastic control system*" or a differential game against nature.

2.4. Cash-flows

Impulse dynamics are hybrid dynamics introducing discontinuities in the evolutions when the capital hits the threshold function. There is a general theory for dealing with these questions with viability techniques which can be applied to financial models (see [13, Aubin, Lygeros, Quincampoix, Sastry & Seube], [12, Aubin & Haddad], among many articles on this topic).

Cash-Flows are defined by finite sequences of dates $0 =: T_0 < T_1 < T_2 < \ldots < T_{N-1} < T_N =: T$ at which payments $\pi(T_i, S, W, P)$ must be made: We set

$W(T_i^-) := \lim_{t \leq T_i,\, t \to -T_i} W(t)$. At this date, the payment must be done in an impulsive way: The new capital $W(T_i)$ at date T_i becomes:

$$\forall\, i = 1, \ldots, N, \;\; W(T_i) \;=\; W(T_i^-) - \pi(T_i, S(T_i), W(T_i)).$$

A necessary condition is that at date T_i, the capital $W(^-T_i)$ satisfies

$$\forall\, i = 1, \ldots, N, \;\; W(T_i^-) \;\geq \mathbf{b}(T_i, S(T_i)) + \pi(T_i, S(T_i), W(T_i)).$$

3. Guaranteed capture basins and viability kernels

3.1. Definition

Definition 3.1. (Guaranteed Viability Kernel) Given an exercise time T, a time-dependent constrained set $\mathcal{K}(t)$ defined by (2.1) and a time-dependent target $\mathcal{C}(t) \subset \mathcal{K}(t)$ defined by (2.2), its *time-dependent guaranteed capture basin*

$$\mathcal{V}(t) \;:=\; \text{GuarCapt}_{(2.3)}(\mathcal{K}, \mathcal{C})(t)$$

under the tychastic control system (2.3) is the tube $\tau \rightsquigarrow \mathcal{V}(\tau)$, $\tau \in [0, T]$, made of elements $(S, P, W) \in \mathcal{V}(\tau)$ for which there exists a feedback map $\mathbb{G}(t, S, P, W) \in \mathbb{F}(t, S, P, W)$ such that, for any selection of returns $R(t) \in \mathbb{R}(t, S(t), P(t), W(t))$, there exists a time $t^\star \in [0, T]$ such that the evolution of $(S(t), P(t), W(t))$ governed by the system of differential equations

$$\begin{cases} \text{(i)} \quad S_i'(t) \;=\; R_i(t) S_i(t), \;\; i = 0, \ldots, n, \\ \text{(ii)} \quad P'(t) \;=\; \mathbb{G}(t, S(t), P(t), W(t)), \\ \text{(iii)} \quad W'(t) \;=\; R_0(t) W(t) + \displaystyle\sum_{i=1}^{n} P_i(t) S_i(t)(R_i(t) - R_0(t)) \\ \qquad\qquad\qquad + \displaystyle\sum_{i=0}^{n} \mathbb{G}_i(t, S(t), P(t), W(t)) S_i(t), \end{cases}$$

and starting at time τ from (S, P, W) reaches the target at time $t^\star$ in the sense that

$$(S(t^\star), P(t^\star), W(t^\star)) \;\in\; \mathcal{C}(t^\star)$$

and is meanwhile viable in $\mathcal{K}(t)$ in the sense that

$$\forall\, t \in [\tau, t^\star], \;\; (S(t), P(t), W(t)) \;\in\; \mathcal{K}(t).$$

Whenever the time-dependent target $\mathcal{C}(t)$ is equal to

$$\mathcal{C}_{\mathcal{K}}(t) \;:=\; \emptyset \text{ if } 0 \leq t < T \text{ and } \mathcal{C}_{\mathcal{K}}(T) \;:=\; \mathcal{K}(T),$$

then the guaranteed capture basin

$$\text{GuarViab}_{(2.3)}(\mathcal{K})(t) \;:=\; \text{GuarCapt}_{(2.3)}(\mathcal{K}, \mathcal{C}_{\mathcal{K}})(t)$$

is called the *time-dependent guaranteed viability kernel*

$$\mathcal{V}(t) \;:=\; \text{GuarViab}_{(2.3)}(\mathcal{K})(t)$$

of the time-dependent environment $\mathcal{K})(t)$ under the tychastic control system (2.3). In this case, *the time* $t^\star = T$ *is equal to the exercise time* T.

The introduction of non-trivial targets allows us to cover many other option contracts which are exercised as soon as the state $(S(t^\star), P(t^\star), W(t^\star)) = \mathcal{C}(t^\star)$.

The concepts of guaranteed capture basin and viability kernel are among the main topics studied in the viability approach of "robust control" in the theory of differential games against nature (tychastic control problems). We refer to chapter 9 of [2, J-P. Aubin] and the literature on this topic ([1, J-P. Aubin], [10, J-P. Aubin & Frankowska] and the forthcoming [4, J-P. Aubin, A. Bayen, N. Bonneuil & P. Saint-Pierre]) and its bibliography, as well as the survey [19, P. Cardaliaguet, M. Quincampoix & P.Saint-Pierre].

We restrict our attention to the links between the concepts of guaranteed capture basin and viability kernel in the particular case of time-dependent constrained sets $\mathcal{K}(t)$ defined by (2.1), time-dependent target $\mathcal{C}(t) \subset \mathcal{K}(t)$ defined by (2.2) and tychastic control system (2.3).

3.2. Derivation of the valuation function and the transaction rule

Knowing the guaranteed viability kernel, we can deduce easily the answers to the problem of the evaluation of the capital and the management of the shares making up the portfolio in the following way:

Theorem 3.2 (Valuation and Management of the portfolio). *Given an exercise time* T *and the time-dependent constrained sets* $\mathcal{K}(t)$ *defined by* (2.1)*, the* time-dependent guaranteed viability kernel

$$\mathcal{V}(t) \;:=\; \mathrm{GuarViab}_{(2.3)}(\mathcal{K})(t)$$

under the tychastic control system (2.3) *provides*

1. *the* initial capital
$$\mathbb{W}(0,S,P) \;:=\; \inf_{(S,P,W)\in\mathcal{V}(0)} W,$$
2. *the* initial portfolio $\mathbb{Q}(0,S)$*, which minimizes the function* $P \mapsto \mathbb{W}(0,S,P)$ *over the subset* $\mathbb{P}(0,S,\mathbb{W}(0,S,P))$*, i.e., a fixed point of the problem*
$$\mathbb{W}(0,S,\mathbb{Q}(0,S)) \;=\; \mathbb{V}(0,S) \;:=\; \inf_{P\in\mathbb{P}(0,S,\mathbb{W}(0,S,\mathbb{Q}(0,S)))} \mathbb{W}(0,S,P)$$
(whenever the constraints on the shares depend upon W*),*
3. *the* transaction rule
$$P'(t) \;=\; \mathbb{G}(t,S(t),P(t),W(t))$$
defined by the feedback involved in the definition of the time-dependent guaranteed viability kernel.

Consequently, for any evolution of the prices $S(t) \in \mathbb{S}(t)$*, the shares* $P(t)$ *and the capital* $W(t)$ *evolve according to the system of differential equations*

$$\begin{cases} \text{(i)} \quad P'(t) \; = \; \mathbb{G}(t, S(t), P(t), W(t)) \\ \text{(ii)} \quad W'(t) \; = \; R_0(t)W(t) + \displaystyle\sum_{i=1}^{n} P_i(t)S_i(t)(R_i(t) - R_0(t) \\ \qquad\qquad\qquad\qquad + \displaystyle\sum_{i=0}^{n} \mathbb{G}_i(t, S(t), P(t), W(t))S_i(t), \end{cases}$$

starting from the initial portfolio $\mathbb{Q}(0, S)$ *and the initial capital* $\mathbb{V}(0, S) = \mathbb{W}(0, S, \mathbb{Q}(0, S))$*.*

Viability theory studies in depth the properties of the time-dependent viability kernels under tychastic control problems. The key point is that there is an algorithm computing the time-dependent guaranteed viability kernel when time, state, regulatory and tychastic variables are discretized. Difficult convergence theorems guarantee the convergence under adequate assumptions. We could stop our study at this point since our objective is to obtain at each time the capital and the shares. We just summarize few points.

1. The time-dependent guaranteed capture basin and viability kernel can be characterized by tangential conditions, which can be translated by characterizing the valuation function $\mathbb{W}$ as the solution to a free boundary-value problem for a nonlinear first-order partial differential equation, playing the role of a second-order linear partial differential equation of Black and Scholes type. For the sake of simplicity and tractability, we restrict the derivation of this partial differential equation (3.2) to an example presented below, still quite general.
2. Viability and capturability issues for stochastic differential equations are particular cases of the same issues for tychastic systems, thanks to the Stroock and Varadhan Support Theorem ([37, D.W. Stroock & S.R Varadhan]), where the tyches range over the range of the Brownian measure and where the tychastic system involves the Stratonovitch drift. To be more specific, let $\mathbb{X}(x, \omega)$ denote the solution starting at x to the stochastic differential equation

$$dx = \gamma(x)dt + \sigma(x)dW(t)$$

where $W(t)$ ranges over $\mathbb{R}^c$ and the drift $\gamma : \mathbb{R}^d \mapsto \mathbb{R}^d$ and the diffusion $\sigma : \mathbb{R}^d \mapsto \mathcal{L}(\mathbb{R}^c, \mathbb{R}^d)$ are smooth and bounded maps (it is usually denoted by $\mathbb{X}(x, \omega) : t \mapsto \mathbb{X}_\omega^x(t)$ in the stochastic literature). Let us associate with them the Stratonovitch drift $\widehat{\gamma}$ defined by $\widehat{\gamma}(x) := \gamma(x) - \frac{1}{2}\sigma'(x)\sigma(x)$. The associated tychastic system is

$$x'(t) \; = \; \widehat{\gamma}(x(t)) + \sigma(x(t))v(t) \;\; \text{where } v(t) \; \in \; \mathbb{R}^c \tag{3.1}$$

where *the tychastic map is constant and equal to* $\mathbb{R}^c$. Compare with general tychastic systems

$$x'(t) \;=\; \widehat{\gamma}(x(t)) + \sigma(x(t))v(t) \;\text{ where } v(t) \;\in\; Q(x(t))$$

where $Q : \mathbb{R}^d \rightsquigarrow \mathbb{R}^c$ is the *tychastic map* associating with any state x the state-dependent subset $Q(x)$ of tyches.

We denote by $\mathcal{H} \subset \mathcal{C}(0,\infty;X)$ any Borel subset of evolutions satisfying given properties, such as, for instance, the subset

$$\mathcal{V}(K) \;:=\; \{x(\cdot) \in \mathcal{C}(0,\infty;X) \mid \forall\, t \geq 0,\;\; x(t) \;\in\; K\}$$

of evolutions *viable* in K.

The *stochastic core* of $\mathcal{H}$ under the stochastic system is the subset of initial states x from which starts a stochastic process $\omega \mapsto \mathbb{X}(x,\omega)$ such that for almost all $\omega \in \Omega$, $\mathbb{X}(x,\omega) \in \mathcal{H}$:

$$\mathrm{Stoc}_{\mathbb{X}}(\mathcal{H}) \;:=\; \{x \in \mathbb{R}^d \mid \text{for almost all } \omega \in \Omega,\;\; \mathbb{X}(x,\omega) := \mathbb{X}^x_\omega(\cdot) \;\in\; \mathcal{H}\}.$$

We denote by $\mathbb{P}_{\mathbb{X}(x,\cdot)}$ the *law* of the random variable $\mathbb{X}(x,\cdot)$ defined by

$$\mathbb{P}_{\mathbb{X}(x,\cdot)}(\mathcal{H}) \;:=\; \mathbb{P}(\{\omega \mid \mathbb{X}(x,\omega) \in \mathcal{H}\}).$$

Therefore, we can reformulate the definition of the stochastic core of a set $\mathcal{H}$ of evolutions in the form

$$\mathrm{Stoc}_{\mathbb{X}}(\mathcal{H}) \;=\; \{x \in \mathbb{R}^d \mid \mathbb{P}_{\mathbb{X}(x,\cdot)}(\mathcal{H}) = 1\}.$$

In other words, the stochastic core of $\mathcal{H}$ is the set of initial states x such that the subset $\mathcal{H}$ has probability 1 under the law of the stochastic process $\omega \mapsto \mathbb{X}(x,\omega) \in \mathcal{C}(0,\infty;\mathbb{R}^d)$ (if $\mathcal{H}$ is closed, $\mathcal{H}$ is called the *support* of the law $\mathbb{P}_{\mathbb{X}(x,\cdot)}$). Let $\mathcal{S} : \mathbb{R}^c \rightsquigarrow \mathcal{C}(0,+\infty;\mathbb{R}^d)$ denote the solution map associating with any state x the subset $\mathcal{S}(x)$ of all $x(\cdot)$ of tychastic system (3.1). The Stroock-Varadhan support theorem states that under adequate regularity assumptions, this support is equal to the tychastic core

$$\mathrm{Tych}_{\mathbb{S}}(\mathcal{H}) \;:=\; \{x \in X \mid \mathcal{S}(x) \;\subset\; \mathcal{H}\}$$

of initial states $x \in X$ from which **all evolutions** $x(\cdot) \in \mathcal{S}(x)$ of the tychastic system (3.1) satisfy the property $\mathcal{H}$. By taking $\mathcal{H} := \mathcal{V}(K)$, we infer that the stochastic viability kernel coincides with the invariance kernel. For more details on the links between stochasticity and tychasticity, see [9, J.-P. Aubin & H. Doss] and [21, 22, 23, G. Da Prato & H. Frankowska], [8, J.-P. Aubin, G. Da Prato & H. Frankowska]. Many open problems remain to be solved in these directions.

3.3. Options with trading constraints

Consider the case when there exists only one risky asset ($n = 1$). The *constraints* bear on

1. prices of the risky asset:
$$\forall\, t \in [0,T],\;\; S(t) \in [S^\flat(t), S^\sharp(t)]$$
where $S^\flat(t) \geq 0$,
2. the shares of the risky asset (liquidity constraints)
$$\forall\, t \in [0,T],\;\; P(t) \in \left[P^\flat(t), \min\left(P^\sharp(t), \frac{W}{S}\right)\right]$$
(which imply that $P_0(t) \geq 0$ whenever $P^\flat(t) \geq 0$),
3. the values of the portfolio, described by a threshold function $\mathbf{b}(t,S)$
$$\forall\, t \in [0,T],\;\; W(t) \geq \mathbf{b}(t, S(t))$$
where $\mathbf{b}$ may be discontinuous (but at least lower semicontinuous),
4. *trading constraints*:
$$\forall\, t \in [0,T],\;\; |P'(t)| \leq \gamma(t)$$
where γ may be discontinuous (but at least upper semicontinuous); this is the case for treating "*rebalancing*" constraints, when $\gamma(t) = 0$ except at discrete times when transactions are allowed to be made,
5. a "tychastic" translation of uncertainty:
$$\forall\, t \in [0,T],\;\; r(t) - \nu(t) \leq R(t) \leq r(t) + \nu(t)$$
(where the tychastic versatility threshold function $\nu(\cdot)$ is assumed to be Lipschitz).

We denote by $\mathcal{K}(W)$ the subset of triples (t,S,P) such that $0 \leq t \leq T$, $S^\flat(t) \leq S \leq S^\sharp(t)$, $P^\flat(t) \leq P \leq \min(S^\sharp(t), \frac{W}{S})$ and $W \geq \mathbf{b}(t,S)$, and by $\mathcal{C}(t)$ the subset of elements of $\mathcal{K}$ such that $W \geq \mathbf{c}(t,S)$ where $\mathbf{c}(t,S) = +\infty$ if $t < T$ and $\mathbf{b}(T,S) = \mathbf{c}(T,S)$.

One can prove that the function $(t,S,P) \mapsto \mathbb{W}(t,S,P)$ is the unique solution (in an adequate generalized sense) of a free boundary problem for the following (nonlinear) partial differential equation with discontinuous coefficients: for all $(t,S,P) \in \mathcal{K}(W)$,
$$\left\{\begin{array}{l} \dfrac{\partial \mathbb{W}}{\partial t} + \dfrac{\partial \mathbb{W}}{\partial S} r(t) S + \nu(t) S \left|\dfrac{\partial \mathbb{W}}{\partial S} - P\right| - \gamma(t) \left|\dfrac{\partial \mathbb{W}}{\partial P} - S\right| \\ = r_0 \mathbb{W} + PS(r(t) - r_0) \end{array}\right. \tag{3.2}$$
satisfying the final condition $\mathbb{W}(T,S,P) = \mathbf{c}(T,S)$ (see, for instance, [3, J.-P. Aubin]). *This is the tychastic version of the Black and Scholes equation adapted to this problem.*

Observe (informally) that if the versatility $\nu(t) = +\infty$ is infinite and if there is no constraint on the number of shares, then $P = \dfrac{\partial \mathbb{W}}{\partial S}$, which is the famous Δ-hedging rule. If there is no restriction on trading, then we have $S = \dfrac{\partial \mathbb{W}}{\partial P}$.

This is a highly nonlinear problem because not only it involves a first-order nonlinear partial differential equation with discontinuous coefficients (instead of a second linear one as the Black and Scholes) but above all, because the subset $\mathcal{K}(W)$ on which it is defined ... depends upon the solution of this equation.

The transaction rule is given by

$$P'(t) \;=\; -\gamma(t)\frac{\frac{\partial \mathbb{W}}{\partial P} - S}{\left|\frac{\partial \mathbb{W}}{\partial P} - S\right|}.$$

3.4. Example: European options with transaction costs

The tychastic approach allows us to treat transaction costs, whereas the stochastic one raises many difficulties (see [33, Soner H.M., Shreve S.E. & CvitanicJ]) entitled *There is no trivial hedging for option pricing with transaction costs.*

We assume that $S(t) \geq 0$ and that $P(t) \in [0, P^\sharp]$. The threshold function for the European option is defined by

$$\mathbf{b}(t,S) \;=\left\{\begin{array}{l} 0 \text{ if } t < T \\ \max(S-K,0) \text{ if } t = T. \end{array}\right.$$

We consider two types of constraints on the transactions:

- *Trading Constraints*:

$$\forall\, t \geq 0, \;\; |P'(t)| \;\leq\; \gamma(t).$$

- *Transaction costs:*

$$P'(t)S(t) \;=\; -\delta |P'(t)| S(t).$$

The viability kernel algorithm provides the valuation function $\mathbb{W}(0, S, P)$:

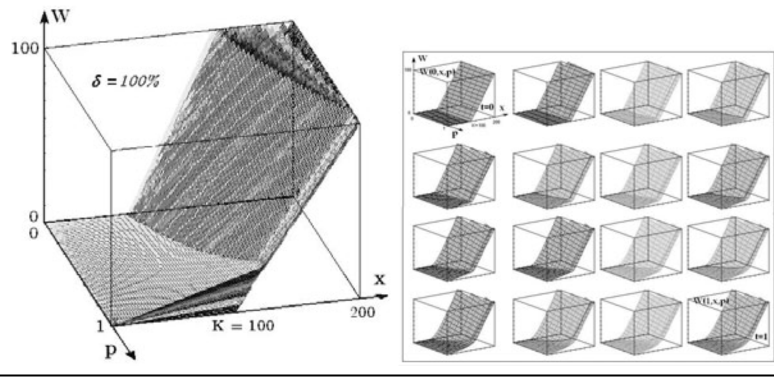


Figure 3.3. *Valuation function.*
The figure displays the valuation function $\mathbb{W}(0,S,P)$) for several values of δ and a fixed exercise time (left) and the value functions for a fixed cost δ and several exercise times.

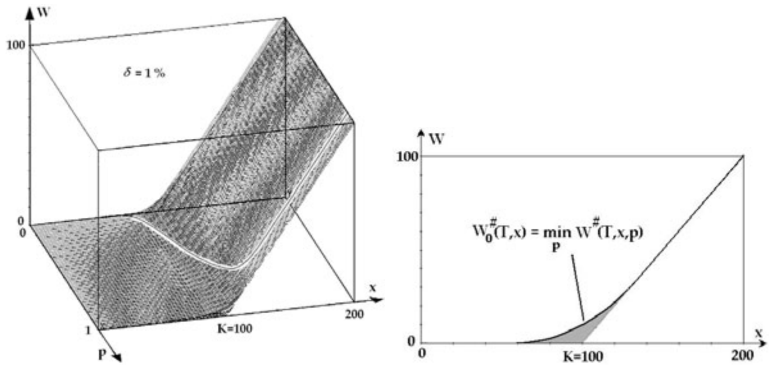


Figure 3.4. *Valuation functions.*
This figure displays the valuation function $\mathbb{V}(0,S) := \inf_{P\in\mathbb{P}(t)} \mathbb{W}(0,S,P)$ for a given exercise time T in the graph of $\mathbb{W}(0,S,P)$ (left), the graph of the function $S \mapsto \mathbb{V}(0,S)$ (right).

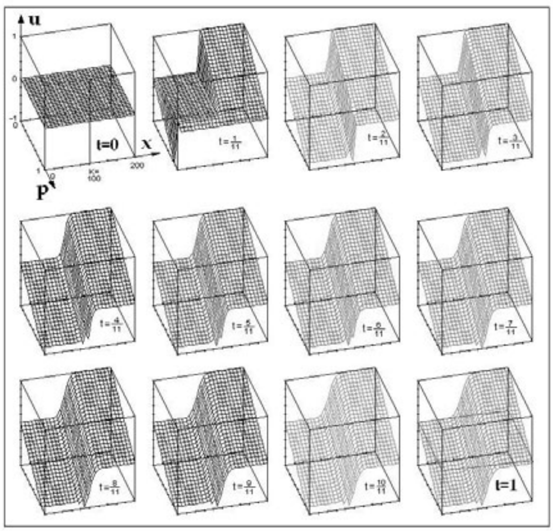


Figure 3.5. *Transaction rules.*
This figure displays the graph of the transaction rule $(S,P) \mapsto \mathbb{G}(t,S,P)$ for several times to maturity. When the time to maturity is equal to 0, $\mathbb{G}(0,S,P) = 0$, because there is no transaction at exercise time. The transactions are negative far below the exercise time and positive far above, a quite intuitive statement.

3.5. Particular case of self-financing portfolios

In the case of self-financing portfolios where

$$\sum_{i=0}^{n} P_i'(t)S_i(t) \;=\; \varphi(t,S(t))W(t)$$

the transactions disappear in the tychastic control system (2.3), which boils down to the simplified tychastic control system

$$\left\{\begin{array}{ll} \text{(i)} & S_i'(t) \;=\; R_i(t)S_i(t) \text{ where } R(t) \;\in\; \mathbb{R}(t,S(t),P(t),W(t)), \\ \text{(ii)} & W'(t) \;=\; (R_0(t)+\varphi(t,S(t)))W(t) + \displaystyle\sum_{i=1}^{n} P_i(t)S_i(t)(R_i(t)-R_0(t)) \\ & \quad \text{where } P(t) \;\in\; \mathbb{P}(t,S(t),W(t)), \end{array}\right. \tag{3.3}$$

where the tyches are still the returns and the controls the numbers of shares instead of their transactions.

The state variables (S,P,W) must evolve in the time dependent constrained set $\mathcal{K}(t)$ defined by

$$\mathcal{K}(t) \;:=\; \{(S,W) \mid S \;\in\; \mathbb{S}(t) \;\&\; W \;\geq\; \mathbf{b}(t,S)\}\,. \tag{3.4}$$

Definition 3.6. (Guaranteed Viability Kernel) Given an exercise time T and the time-dependent constrained set $\mathcal{K}(t)$ defined by (3.4), its *time-dependent guaranteed viability kernel*

$$\mathcal{V}(t) \;:=\; \mathrm{GuarViab}_{(3.3)}(\mathcal{K})(t)$$

under the tychastic control system (3.3) is the tube $\tau \rightsquigarrow \mathcal{V}(\tau)$, $\tau \in [0,T]$, made of elements $(S,W) \in \mathcal{V}(\tau)$ for which there exists a feedback map $\mathbb{G}(t,S,W) \in \mathbb{P}(t,S,W)$ such that, for any selection of returns $R(t) \in \mathbb{R}(t,S(t),W(t))$, the evolution of $(S(t),W(t))$ governed by the system of differential equations

$$\left\{\begin{array}{ll} \text{(i)} & S_i'(t) \;=\; R_i(t)S_i(t), \quad i=0,\ldots,n, \\ \text{(ii)} & W'(t) \;=\; (R_0(t)+\varphi(t,S(t)))W(t) \\ & \qquad\qquad + \displaystyle\sum_{i=1}^{n} \mathbb{G}_i(t,S(t),W(t))S_i(t)(R_i(t)-R_0(t)), \end{array}\right.$$

and starting at time τ from (S,W) is viable in $\mathcal{K}(t)$ in the sense that

$$\forall\, t \in [\tau,T], \;\; (S(t),W(t)) \;\in\; \mathcal{K}(t).$$

Knowing the guaranteed viability kernel, we derive:

Theorem 3.7 (Valuation and Management of the portfolio). *Given an exercise time T and the time-dependent constrained sets $\mathcal{K}(t)$ defined by* (3.4)*, the* time-dependent guaranteed viability kernel

$$\mathcal{V}(t) \;:=\; \mathrm{GuarViab}_{(3.3))}(\mathcal{K})(t)$$

under the tychastic control system (3.3) *provides at each instant t,*

1. *the* capital

$$\forall\, t \in [0,T], \ \ \mathbb{W}(t,S) \ := \inf_{(S,W)\in\mathcal{V}(t)} W,$$

2. *the* management rule

$$\mathbb{P}(t,S) \ = \ \mathbb{G}(t,S,\mathbb{W}(t,S))$$

defined by the feedback involved in the definition of the time-dependent guaranteed viability kernel.

Consequently, for any evolution of the prices $S(t) \in \mathbb{S}(t)$, the shares and the capital are given by $W(t) := \mathbb{W}(t,S(t))$ and $P(t) = \mathbb{P}(t,S(t))$.

The very same viability techniques allow us to treat the "implied versatility" issue. Usually, it is assumed that the portfolio is self-financed. Consider the case of one risky asset. Given the classical contingent function $\max(0, S-K)$ where K is the striking price, an exercise time T and a *constant* tychastic threshold ν, one can associate with any (T,S,K,ν) the initial value $W := \Theta(T,S,K,\nu)$ of the portfolio such that there exists a feedback map $\mathbb{Q}(t,S,W,K,\nu) \in \mathbb{P}(S,W)$ such that, for any selection of returns $v(t) \in [-\nu,+\nu]$, the evolution of $(S(t),W(t))$ governed by the system of differential equations

$$\left\{\begin{array}{rrcl} \text{(i)} & S'(t) & = & r(t)S(t) + v(t)S(t), \\ \text{(ii)} & W'(t) & = & r_0 W(t) + P(t)S(t)(r - r_0 + v(t)) \\ & & & \qquad\qquad \text{where } P(t) := \mathbb{Q}(t,S(t),W(t),K,\nu), \end{array}\right.$$

starting from (S,W) satisfy $W(t) \geq 0$ and

$$W(T) \ := \ \Theta(T,S,K,\nu) \ \geq \ \max(0, S(T)-K).$$

The *implied versatility function* associates with any (T,S,K,W) the largest versatility threshold $\nu := \Lambda(T,S,K,W)$ under which

$$\left\{\begin{array}{rl} \text{(i)} & \forall\, W \geq 0, \ \ \Theta(T,S,K,\Lambda(T,S,K,W)) \leq W, \\ \text{(ii)} & \forall\, \nu \geq 0, \ \ \Lambda(T,S,K,\Theta(T,S,K,\nu)) \ \geq \ \nu. \end{array}\right.$$

These two functions can be characterized in terms of guaranteed viability kernels and computed by the Capture Basin Algorithm instead of inverting the function $\nu \mapsto \Theta(T,S,K,\nu)$ by standard inversion methods which do not take into account its viability property.

3.6. Cash-flow (without transaction costs)

In this example, the constraints are $S(t) \geq 0$, $0 \leq P(t) \leq P^\sharp$, $W(t) \geq 0$ and the cash-flow is made of payments $\pi(T_i,S,W) := \pi_i$. The first figure displays the graph of the function $(t,S) \mapsto \mathbb{W}(t,S)$ and $\mathbb{G}(t,S)$.

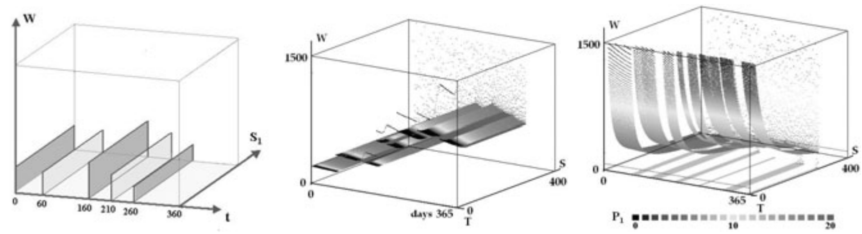


Figure 3.8. *Example of cash flows with constraints on the shares but without transaction constraints: capital and shares in terms of exercise time and prices. Cash flow, capital and shares of the risky asset in terms of exercise time (abscissa) and price of the risky asset (ordinate)*

This portfolio is guaranteed in the sense that whatever the evolution of prices, the capital is sufficient to cover the cash flow:

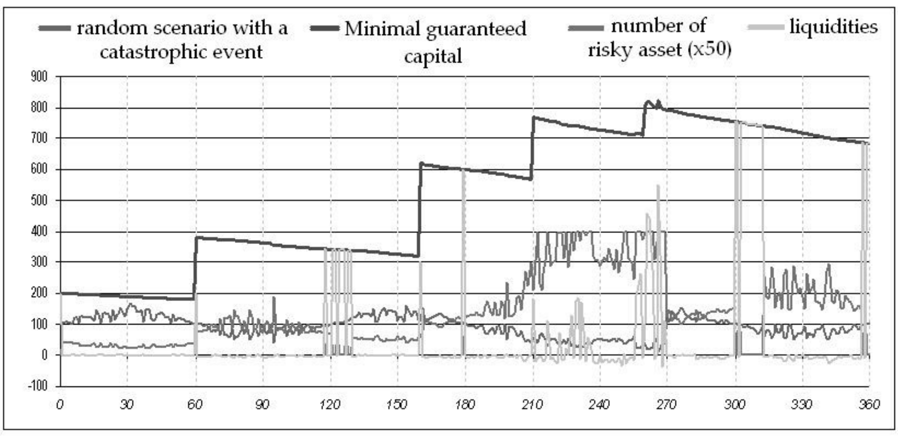


Figure 3.9. *Guaranteed evolution of value and shares.*
The evolution of the price of the risky asset is simulated (dark gray curve). Note the drop of the prices. The picture displays the evolution of the associated value of the portfolio (in black), the number of shares of the risky asset (in gray), the value of the non-risky component of the portfolio in light gray.

4. Options without transaction constraints

We require that at the *exercise time* T, the option is exercised. The threshold function for classical European, American and Bermudan options are

$$\mathbf{b}(t,S) \;=\; \begin{cases} \text{(i)} & 0 \text{ if } t < T \text{ and } \max(S-K,0) \text{ if } t = T \\ & \text{European Options,} \\ \text{(ii)} & \max(S-K,0) \text{ if } t \le T \\ & \text{American Options,} \\ \text{(iii)} & aS \text{ if } t < T \text{ and } \max(S-K,aS) \text{ if } t = T,\;\; 0 < a \le 1 \\ & \text{Quasi-American Options,} \\ \text{(iv)} & 0 \text{ if } t \neq T_i \text{ and } \max(S-K_i,0) \text{ if } t = T_i,\; i = 1,\ldots,n \\ & \text{Bermudan Options.} \end{cases}$$

4.1. European options without transaction costs

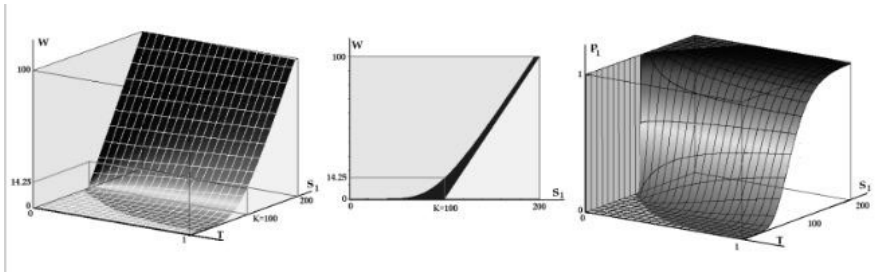


Figure 4.1. *European options without transaction costs.*
This figure displays the valuation function and the price function.
Left and Right: abscissa: Time to Maturity, ordinate: Prices of the Risky Asset.
Left: Price of the European Option, Right: Number of Shares.
Middle: For a fixed exercise time, abscissa: Prices of the Risky Asset, ordinate: Price of the European Option.

The first question which arises is whether the viability kernel algorithm provides the same values as the Black and Scholes formula for the European option (computed with the Cox, Ross and Rubinstein algorithm). The answer is positive and given by Figure 4.2.

Actually, there are two questions: The first one deals with the approximation of the Black and Scholes formula for continuous time by discrete time problems, and the second deals with the computation of the solution to this approximate discrete problem. It is for solving the discretized problem (both with respect to time and space variables) that the Capture Basin Algorithm is used. The other issue deals with the convergence of the solution to the discrete problems to the solution of the continuous time problem. It happens that the discretization of the stochastic problem and of the tychastic problems are quite the same, up to the replacement of the step size Δt in the tychastic discrete system by $\sqrt{\Delta t}$ in some terms of the discrete stochastic system, which provides the Cox, Ross and Rubinstein algorithm in the case of portfolios replicating European options. Hence, by modifying the discretization of the tychastic system by an adequate discretization of the stochastic system, the Viability Kernel Algorithm provides pricers, evaluation of the value of the portfolio and the regulation rule for both mathematical translations of uncertainty, the tychastic one allowing to take into account constraints on the versatility depending upon time, asset prices, and shares of the portfolios.

Maturity	Volatility	Black & Scholes	Capture Basin	Share π_0	Share π_1
1	10%	**6.72**	**6.82**	-64.06	0.7078
0.5	10%	**4.15**	**4.21**	-60.95	0.6507
0.1	10%	**1.52**	**1.53**	-55.47	0.5699
1	20%	**10.45**	**10.46**	-53.23	0.6368
0.5	20%	**6.89**	**6.89**	-52.91	0.5979
0.1	20%	**2.77**	**2.79**	-51.68	0.5446
1	25%	**12.27**	**12.34**	-50.41	0.6274
0.5	25%	**8.22**	**8.26**	-50.84	0.5910
0.1	25%	**3.39**	**3.42**	-50.72	0.5413
1	30%	**14.23**	**14.23**	-48.20	0.6242
0.5	30%	**9.65**	**9.64**	-49.24	0.5887
0.1	30%	**4.03**	**4.04**	-49.97	0.5401
1	50%	**21.73**	**21.76**	-41.69	0.6344
0.5	50%	**15.09**	**15.12**	-44.66	0.5977
0.1	50%	**6.53**	**6.54**	-47.87	0.5441

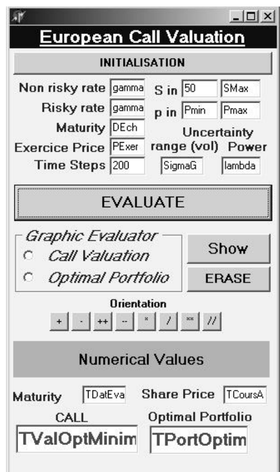


Figure 4.2. *Comparison of algorithms.*

4.2. Other options without transaction costs

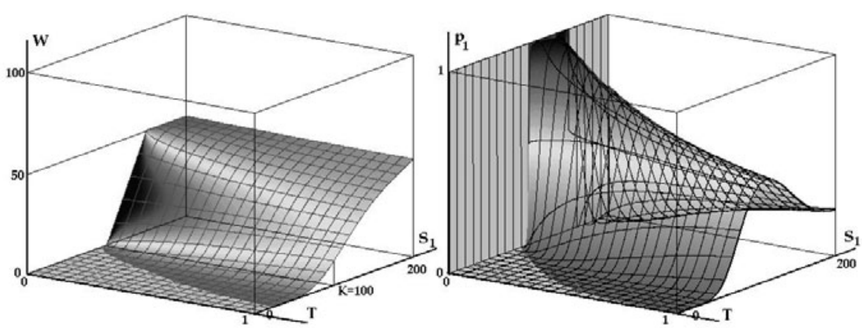


Figure 4.3. *"Capped" options.*
Value and number of shares of risky assets.

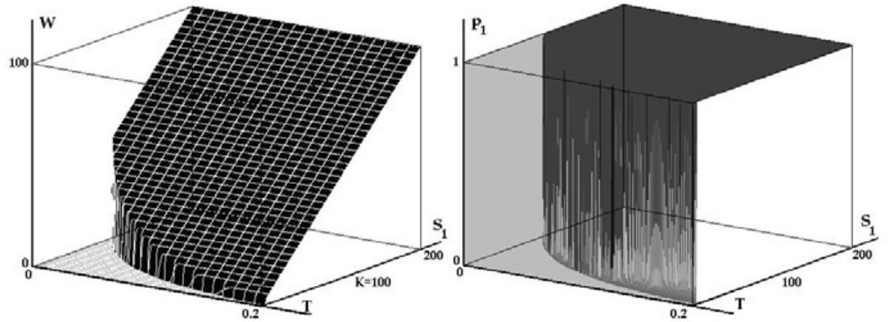


Figure 4.4. *"Asset or nothing" options.*
Value and number of shares of risky assets.

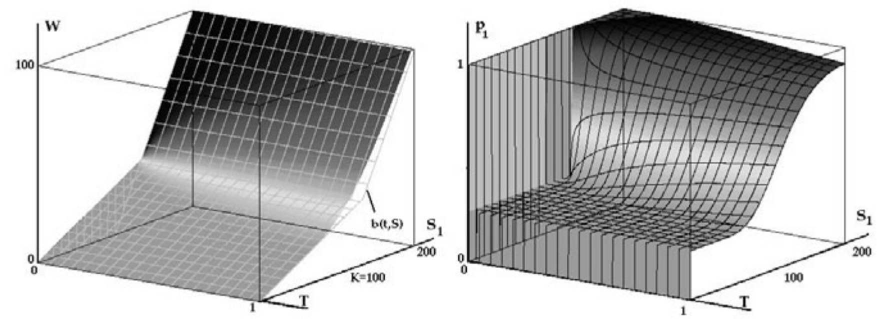


Figure 4.5. *"Non-standard" American options.*
Value and number of shares of risky assets.

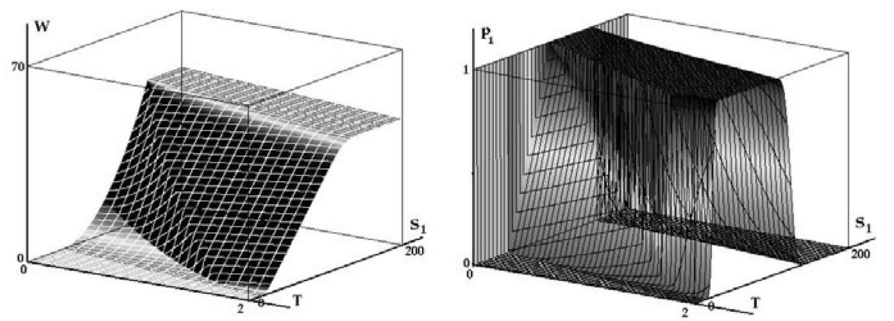


Figure 4.6. *"Non-standard" options.*
Value and number of shares of risky assets under another tychastic dynamics without transactions costs. We take $r(t,S) = \frac{\sqrt{S}}{1000}$, $\vartheta(t) = 0.3\frac{1}{0.01+t^2}$.

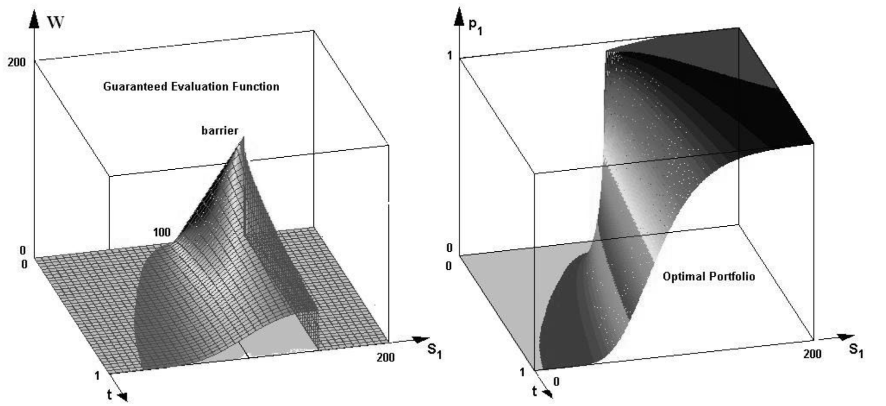


Figure 4.7. *European call with barrier "up in" and "up out". Value and number of shares of risky assets.*

We observe a kind of stability of the shape of the valuation function in all these examples, but the nature of the management rule is very sensitive to the change of contracts.

Many other types of financial products can be characterized as guaranteed viability kernels of suitable constrained environments under adequate more or less "natural" tychastic dynamical systems. The following table mentions some of the existing options or cash flows for which this characterization has been proved and for which some softwares providing both the value of the option and its management rule do exist[1].

Products	Without Transaction Costs	With Transaction Costs	Liquidity Constraints	reba-lancing	CPPI (cushion)
Europan	XXX	XXX	XXX	XXX	X
Bermudian	XXX	XXX	XXX	X	X
Digital	XXX	XX	XXX	XX	X
Lookback	XX		XX		X
Barrier	XXX	X	XXX	X	X
Cash Flows	XXX	XXX	XXX	XXX	X

Lines denote the nature of portfolios replicating options or cash flows, column options indicate the availability of viability characterizations.
XXX: available executables XX: rapidly available executables on demand X: available executables on demand.

[1]Distributed by the company VIMADES (Viability, Markets, Automatics, Decisions).

References

[1] J.-P. Aubin, *Viability Theory*, Birkhäuser, Boston, Basel, 1991.

[2] J.-P. Aubin, *Dynamic Economic Theory: A Viability Approach*, Springer-Verlag, 1997.

[3] J.-P. Aubin, *A concise introduction to viability theory, optimal control and robotics*, cours DEA MVA, École Normale Supérieure de Cachan, 2001.

[4] J.-P. Aubin, A. Bayen, N. Bonneuil, and P. Saint-Pierre, *Viability, Control and Game Theories: Regulation of Complex Evolutionary Systems Under Uncertainty*, Springer-Verlag, 2006.

[5] J.-P. Aubin and G. Da Prato, *Stochastic viability and invariance*, Annali Scuola Normale di Pisa, **27** (1990), 595–694.

[6] J.-P. Aubin and G. Da Prato, *Stochastic Nagumo's viability theorem*, Stochastic Analysis and Applications, **13** (1995), 1–11.

[7] J.-P. Aubin and G. Da Prato, *The viability theorem for stochastic differential inclusions*, Stochastic Analysis and Applications, **16** (1998), 1–15.

[8] J.-P. Aubin, G. Da Prato, H. Frankowska, *Stochastic invariance for differential inclusions*, J. Set-Valued Analysis, **8** (2000), 181–201.

[9] J.-P. Aubin and H. Doss, *Characterization of stochastic viability of any nonsmooth set involving its generalized contingent curvature*, Stochastic Analysis and Applications, **25** (2003), 951–981.

[10] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, 1990.

[11] J.-P. Aubin and G. Haddad, *History (path) dependent optimal control and portfolio valuation and management*, J. Positivity, **6** (2002), 331–358.

[12] J.-P. Aubin and G. Haddad, *Impulse capture basins of sets under impulse control systems*, J. Mathematical Analysis and Applications, **275** (2002), 676–692.

[13] J.-P. Aubin, J. Lygeros, M. Quincampoix, S. Sastry, and N. Seube, *Impulse differential inclusions: a viability approach to hybrid systems*, IEEE Transactions on Automatic Control, **47** (2002), 2–20.

[14] J.-P. Aubin, D. Pujal, and P. Saint-Pierre, *Dynamic management of portfolios with transaction costs under tychastic uncertainty*, in: Numerical Methods in Finance, H. Ben-Ameur and M. Breton, Editors, Kluwer, 2005.

[15] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, 1997.

[16] R. Buckdahn, P. Cardaliaguet, and M. Quincampoix, *A representation formula for the mean curvature motion*, SIAM J. Math.Anal., **33 (4)** (2002), 827–846.

[17] R. Buckdahn, M. Quincampoix, C. Rainer, and A. Rascanu, *Stochastic control with exit time and constraints. Applications to small time attainability of sets*, Appl. Math. Optim., **2** (2004), 99–112.

[18] R. Buckdahn, S. Peng, M. Quincampoix, and C. Rainer, *Existence of stochastic control under state constraints*, Comptes-Rendus de l'Académie des Sciences, **327** (1998), 17–22.

[19] P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre, *Set-valued numerical methods for optimal control and differential games*, in: Stochastic and Differential Games –

Theory and Numerical Methods, Annals of the International Society of Dynamical Games, Birkhäuser, (1999), 177–247.

[20] G. Da Prato and H. Frankowska, *A stochastic Filippov Theorem*, Stochastic Calculus, **12** (1994), 409–426.

[21] G. Da Prato and H. Frankowska, *Invariant measure for a class of parabolic degenerate equations*, Nonlinear Differential Equations Appl., **12 (4)** (2005), 481–501.

[22] G. Da Prato and H. Frankowska, *Stochastic viability for compact sets in terms of the distance function*, Dynamics Systems Appl., **10** (2001), 177–184.

[23] G. Da Prato and H. Frankowska, *Invariance of stochastic control systems with deterministic arguments*, J. Diff. Eqs., **200** (2004), 18–52.

[24] G. Da Prato and J. Zabczyk, *Stochastic Evolution Equations in Infinite Dimensions*, Cambridge University Press, 1992.

[25] H. Doss, *Liens entre équations différentielles stochastiques et ordinaires*, Ann. Inst. Henri Poincaré, Calcul des Probabilités et Statistique, **23** (1977), 99–125.

[26] S. Gautier and L. Thibault, *Viability for constrained stochastic differential equations*, Differential Integral Equations, **6** (1993), 1395–1414.

[27] J. Motyl, *Viability of set-valued Itô equation*, Bull. Pol. Acad, Sci., Math., **47** (1999), 91–103.

[28] D. Pujal, *Valuation et gestion dynamiques de portefeuilles*, Thèse de l'Université de Paris-Dauphine, 2000.

[29] D. Pujal and P. Saint-Pierre, *L'algorithme du bassin de capture appliqué pour évaluer des options européennes, américaines ou exotiques*, Revue de l'Association Française de Finance, **25** (2004), 75–106.

[30] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk, **2** (2000), 21–41.

[31] R. T. Rockafellar and R. Wets, *Variational Analysis*, Springer-Verlag, 1997.

[32] P. Saint-Pierre, *Approximation of the viability kernel*, Applied Mathematics & Optimisation, **29** (1994), 187–209.

[33] H. M. Soner, S. E. Shreve, and J. Cvitanic, *There is no trivial hedging for option pricing with transaction costs*, The Annals of Applied Probability, **5** (1995), 327–355.

[34] H. M. Soner and N. Touzi, *Super-replication under Gamma constraints*, SIAM Journal on Control and Optimization, **39** (2000), 73–96.

[35] H. M. Soner and N. Touzi, *Stochastic target problems, dynamic programming and viscosity solutions*, SIAM Journal on Control and Optimization, **41** (2002), 404–424.

[36] H. M. Soner and N. Touzi, *Dynamic programming for stochastic target problems and geometric flows*, Journal of the European Mathematical Society, **4** (2002), 201–236.

[37] D. W. Stroock and S. R. S. Varadhan, *On the support of diffusion processes with applications to the strong maximum principle*, Proc. of the Sixth Berkeley Symposium on Mathematical Statistics and Probabilty, Univ. of California Press, Vol. III. Probability Theory, (1972), 333–359.

[38] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer-Verlag, 1979.

[39] X. D. H. Truong, *Existence of viable solutions for nonconvex-valued stochastic differential inclusions in Banach spaces*, Portugalae Mathematica, **52** (1995), 241–250.

[40] J. Zabczyk, *Chance and decision: stochastic control in discrete time*, Quaderni, Scuola Normale di Pisa, 1996.

[41] J. Zabczyk, *Pricing options by dynamic programming*, in: Stochastic Processes and Related Topics, H. J. Engelbert, H. Föllmer and J. Zabczyk, Editors, Gordon and Breach, (1996), 153–160

[42] J. Zabczyk, *Stochastic invariance and consistency of financial models*, Atti Accad. Naz. Lincei CI. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., **11 (2)** (2000), 67–80.

Jean-Pierre Aubin and Patrick Saint-Pierre
LASTRE (Laboratoire d'Applications des Systèmes Tychastiques Régulés)
14, rue Domat
75005 Paris
France
e-mail: aubin.jp@gmail.com
Patrick.Saint.Pierre@gmail.com

Progress in Probability, Vol. 59, 435–444

Numerical Aspects of Loan Portfolio Optimization

Claas Becker and Veronika Orlovius

Abstract. The current industry standard is to optimize loan portfolios with respect to variance. In this paper we show that optimization with respect to expected shortfall and expected regret is fairly easy to implement.

Mathematics Subject Classification (2000). Primary 91B28; Secondary 91B30.

Keywords. Loan portfolio optimization, coherent risk measures, expected shortfall.

1. Introduction

It is essential for credit portfolio managers to understand the sources of risk and to have tools at hand to actively manage credit risk. As the CDS market soars, the need for credit risk management tools is even increasing.

The formulation and solution of the portfolio optimization problem dates back to Markovitz and Sharpe[1]. While the assumption of normally distributed returns is a fairly good proxy for market risk, credit risk returns are heavy-tailed and clearly not Gaussian. This is why the development of algorithms and tools for managing portfolio credit risk has lagged behind. The current industry standard is to optimize loan portfolios with respect to variance because this is easy to implement and because it is a common belief that optimization with respect to expected shortfall or expected regret is numerically not tractable. In this paper we show that optimization with respect to expected shortfall and expected regret is fairly easy to implement and that efficient frontiers can be easily computed.

Given n loss variables L_i with returns r_i, we optimize the portfolio loss $L = \sum w_i L_i$ subject to the constraints of a given return $R = \sum w_i r_i$ and constant portfolio volume $\sum w_i = 1$. Unless otherwise stated, we regard L_i as the loss associated to obligor i. On a more strategic level, the L_i's of obligors belonging to

[1]The original formulation was in terms of variance optimization, though they certainly had in mind more general settings.

the same industry sector can be aggregated, thus leading to optimization in terms of industry sectors.

2. Risk measures

In this paper we focus on the optimization of three risk measures: variance, expected shortfall and expected regret.

The variance is given as

$$\mathrm{Var}(L) = \mathrm{cov}(L, L) = \sum_{i,j} w_i w_j \mathrm{cov}(L_i, L_j)$$

where $\mathrm{cov}(L_i, L_j)$ is the covariance matrix of default correlations.

Expected shortfall is defined[2] as

$$\mathrm{ES}_\alpha = E(L|L \geq \mathrm{VaR}_\alpha(L)) = \frac{1}{1-\alpha} \int_{L \geq \mathrm{VaR}_\alpha} L \, dP$$

where VaR_α denotes the value at risk w.r.t. the confidence level $\alpha, 0 < \alpha < 1$.

We assume that the random variable L has a density with respect to Lebesgue measure. In this case the lower quantile

$$q_\alpha(L) = \inf\{x \in \mathbb{R} | P(L \leq x) \geq \alpha\}$$

and the upper quantile

$$q_\alpha(L) = \inf\{x \in \mathbb{R} | P(L \leq x) > \alpha\}$$

coincide.

Expected regret is defined as

$$\mathrm{ER}_K = \int (L - K)^+ \, dP$$

where $K > 0$ is some threshold.

Note that variance and expected regret are no coherent risk measures[3] while expected shortfall is coherent.

One of the early papers on coherent risk measures is [2]. The paper [4] relates coherent risk measures, utility maximization and portfolio optimization.

From a practical point of view, the choice of the confidence level α and of the threshold K is crucial. Usually banks choose α in such a way that $1 - \alpha$ is

[2] More precisely, expected shortfall is defined as $\mathrm{ES}_\alpha = (1-\alpha)^{-1} E(L 1_{\{L > q_\alpha(L)\}}) + q_\alpha(L) \cdot (P(L \leq q_\alpha(L)) - \alpha)$. The second term ensures the coherence if the distribution function has jumps and vanishes if P is continuous w.r.t Lebesgue measure.

[3] A risk measure $\rho : L^1(\Omega, \mathcal{A}, P) \to \mathbb{R}$ is coherent if the following properties hold:

- Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$
- Monotonicity: $\rho(X) \leq \rho(Y)$ if $X \leq Y$
- Positive homogeneity: $\rho(\lambda X) = \lambda \rho(X) \quad \forall \lambda > 0, \forall X \in L^1$
- Translation invariance: $\rho(X + a) = \rho(X) + a \quad \forall a \in \mathbb{R}, \forall X \in L^1$.

the default probability assigned to the bank's rating (or the bank's target rating). However, even losses much smaller than VaR_α may have a significant impact on the bank. Thus, we suggest using a much smaller confidence level, e.g. VaR_α or K equal to the bank's yearly P&L.

3. Formulation of the optimization problem

In the context of portfolios of nontradable loans, the obligor weights w_i will be subject to additional restrictions

$$l_i \leq w_i \leq u_i, \quad i = 1, \ldots, n$$

where l_i denotes the lower bound and u_i denotes the upper bound. For bond portfolios, we may choose $l_i = 0$ or even allow short-selling.

Optimization of the portfolio variance leads to the quadratic optimization problem

$$\min_{w \in R^n} \sum_{i,j=1}^{n} w_i w_j \operatorname{cov}(L_i, L_j)$$

where $w = (w_1, \ldots, w_n)$, subject to constraints

$$\sum_{i=1}^{n} w_i = 1$$

$$l_i \leq w_i \leq u_i, \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} w_i r_i = R, \quad R \in \mathbb{R}^+.$$

This means that for a given return R, we search for the solution with minimal variance. We can thus compute points on the efficient frontier for various returns R.

For expected shortfall and expected regret, we have to know the distribution of L. This is done via a Monte Carlo simulation. Let $y_m = (y_1^m, \ldots, y_n^m) \in \mathbb{R}_+^n$ denote the losses of $(L_1, \ldots, L_n)$ in scenario m, $m = 1, \ldots, M$. If we want to optimize with respect to a reasonably great confidence level α, we have to choose M in the range of, e.g., 10,000 to 100,000.

Expected shortfall optimization leads to the minimization problem (cf. the appendix)

$$\min_{w \in R^n, q} \left(q + \frac{1}{1-\alpha} M^{-1} \sum_{m=1}^{M} z_m \right)$$

subject to constraints

$$z_m \geq \sum_{i=1}^{n} w_i \, y_i^m - q, \quad m = 1, \ldots, M$$

$$z_m \geq 0, \quad m = 1, \ldots, M$$

$$\sum_{i=1}^{n} w_i = 1$$

$$l_i \leq w_i \leq u_i, \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} w_i r_i = R, \quad R \in \mathbb{R}^+.$$

This is an $(M + n + 1)$-dimensional linear programming problem that can be solved using existing LP-solvers. Analogous to the formulation of the variance optimization problem, we keep the return R fixed and search for the solution with minimal expected shortfall.

Optimization of expected regret leads to

$$\min_{w \in R^n} M^{-1} \sum_{m=1}^{M} z_m$$

subject to constraints

$$z_m \geq \sum_{i=1}^{n} w_i \, y_i^m - K, \quad m = 1, \ldots, M$$

$$z_m \geq 0, \quad m = 1, \ldots, M$$

$$\sum_{i=1}^{n} w_i = 1$$

$$l_i \leq w_i \leq u_i, \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} w_i r_i = R, \quad R \in \mathbb{R}^+.$$

As in the preceding two optimization problems, we fix the return R and search for the solution with minimal expected regret.

Details on the derivation of the minimization problems for expected shortfall and expected regret can be found in the appendix.

4. Case study

We used a sample portfolio of 10 obligors intended to represent a German SME portfolio. Alternatively, this sample portfolio can be regarded as the aggregate exposures of obligors belonging to 10 different industry sectors. Table 1 describes the sample portfolio, Table 2 contains the correlation matrix. Based on this correlation information, we generated 20,000 Monte Carlo simulations to obtain the empirical loss distribution. The focus of this computational exercise was to compute efficient frontiers for the three risk measures and to observe whether the optimization algorithm effectively reshapes the loss distribution. It is important to note that the optimization problems are numerically tractable for much bigger portfolios with, e.g., 10,000 different obligors.

Obligor #	Nominal Exposure	LGD	Spread	Initial Weight	Lower Bound	Upper Bound	Annual PD
1	150000	0,26	0,55%	10,0%	5,0%	20,0%	0,30%
2	150000	0,49	0,40%	10,0%	5,0%	20,0%	0,10%
3	150000	0,30	0,80%	10,0%	5,0%	20,0%	0,40%
4	150000	0,19	2,80%	10,0%	5,0%	20,0%	1,00%
5	150000	0,51	0,55%	10,0%	5,0%	20,0%	0,30%
6	150000	0,56	2,00%	10,0%	5,0%	20,0%	0,80%
7	150000	0,42	1,10%	10,0%	5,0%	20,0%	0,50%
8	150000	0,68	2,80%	10,0%	5,0%	20,0%	1,00%
9	150000	0,89	1,40%	10,0%	5,0%	20,0%	0,60%
10	150000	0,37	0,55%	10,0%	5,0%	20,0%	0,30%

Table 1: The sample portfolio.

Obligor #	1	2	3	4	5	6	7	8	9	10
1	1	0,210	0,229	0,216	0,222	0,231	0,188	0,210	0,223	0,136
2	0,210	1	0,215	0,205	0,207	0,214	0,175	0,202	0,215	0,170
3	0,229	0,215	1	0,222	0,227	0,237	0,199	0,212	0,228	0,147
4	0,216	0,205	0,222	1	0,212	0,221	0,182	0,200	0,216	0,143
5	0,222	0,207	0,227	0,212	1	0,230	0,191	0,210	0,219	0,136
6	0,231	0,214	0,237	0,221	0,230	1	0,193	0,214	0,228	0,141
7	0,188	0,175	0,199	0,182	0,191	0,193	1	0,185	0,187	0,098
8	0,210	0,202	0,212	0,200	0,210	0,214	0,185	1	0,208	0,135
9	0,223	0,215	0,228	0,216	0,219	0,228	0,187	0,208	1	0,152
10	0,136	0,170	0,147	0,143	0,136	0,141	0,098	0,135	0,152	1

Table 2: Asset correlations in the sample portfolio.

While variance optimization is quite simple from the computational point of view, expected shortfall and expected regret optimization require optimization with respect to $20,000+10+1$ variables. The computations were done in MATLAB using the linopt function. It is important to note that the optimization problems are still numerically tractable for much bigger portfolios. For example, optimizing a portfolio of $10,000$ obligors w.r.t. expected shortfall or expected regret would require optimization w.r.t. $20,000 + 10,000 + 1$ variables. A reduction of the size of the optimization problem can be achieved by importance sampling, cf. the appendix.

Please note that it does not make any difference whether we optimize with respect to the spreads r_i or the total returns $r_i + r$, where r denotes the funding rate. This is due to the fact that

$$\sum_{i=1}^{n} w_i(r_i + r) = \sum_{i=1}^{n} w_i r_i + r.$$

Note that this argument is no longer true if there is a term structure, i.e., if the funding rate is a function of time.

Figures 1 to 3 depict the efficient frontiers for optimization w.r.t variance, expected shortfall, and expected regret. For these computations, we set the target return R =1.5%.

We observed three effects:

1. Variance optimization does not effectively reshape the loss distribution in the sense of shifting probability mass from very high loss scenarios to lower loss scenarios. In contrast, expected shortfall and expected regret optimization do.

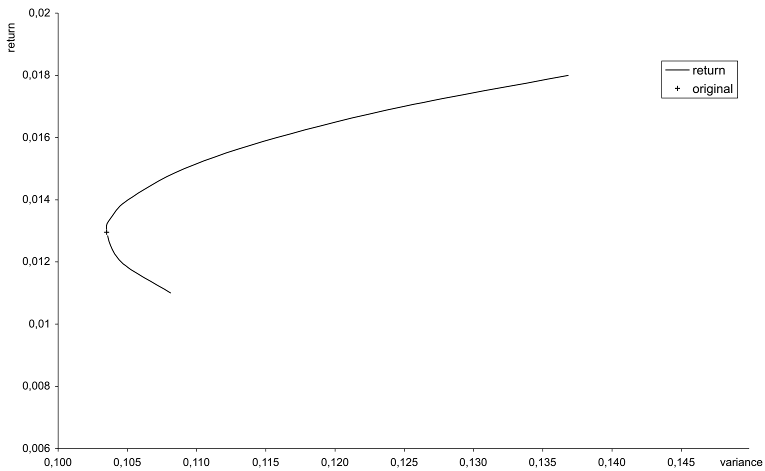


FIGURE 1. Variance efficient frontier.

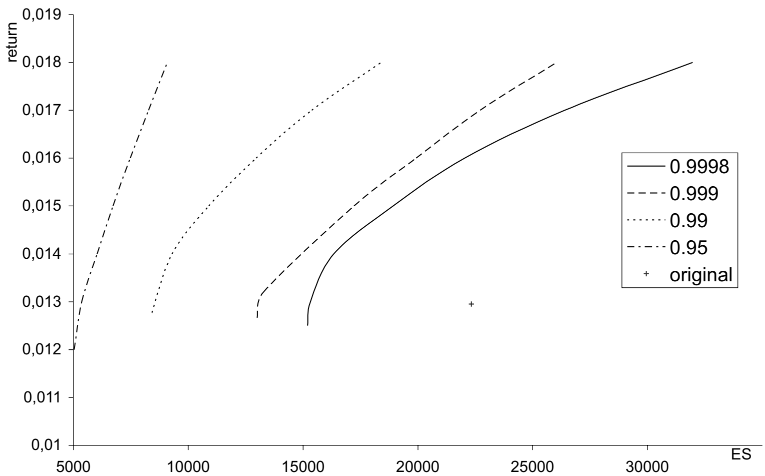


FIGURE 2. Expected shortfall efficient frontier for quantiles corresponding to various confidence levels.

2. The expected shortfall algorithm reacts more flexible to tighter restrictions on the obligor weights than the expected regret algorithm does.
3. Comparing the simplex algorithm to the interior point method, the interior point method, not surprisingly being faster than the simplex algorithm, does not find every minimum. Both algorithms are provided by the linopt function in MATLAB.

Figures 4 to 6 compare the results of the optimization w.r.t. one risk measure with the results w.r.t. the other two risk measures. Not surprisingly, variance-optimal portfolios are not necessarily optimal with respect to expected shortfall or expected regret.

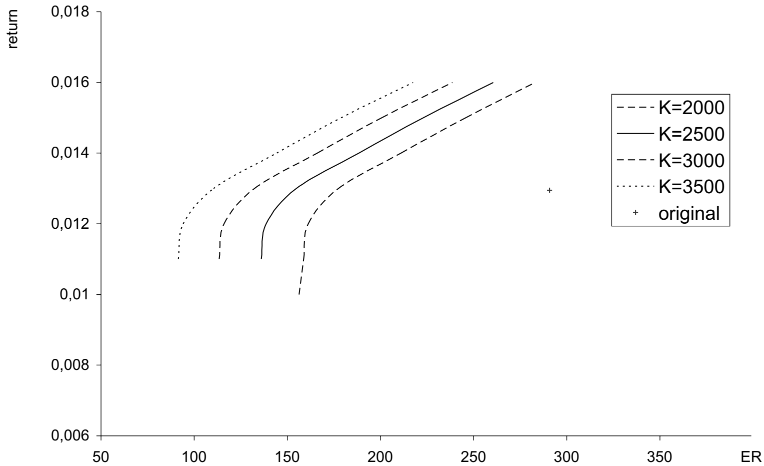


FIGURE 3. Expected regret efficient frontier for various settings of the threshold K.

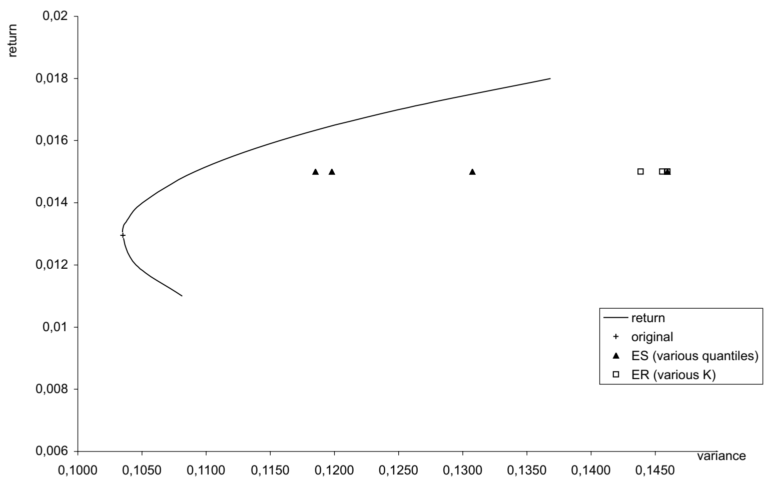


FIGURE 4. Variance efficient frontier vs. ES / ER optimized portfolios.

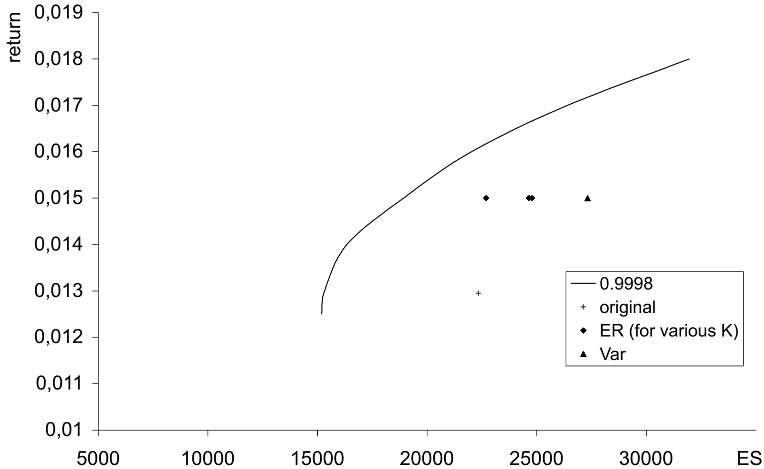


FIGURE 5. ES efficient frontier vs. Variance / ER optimized portfolios.

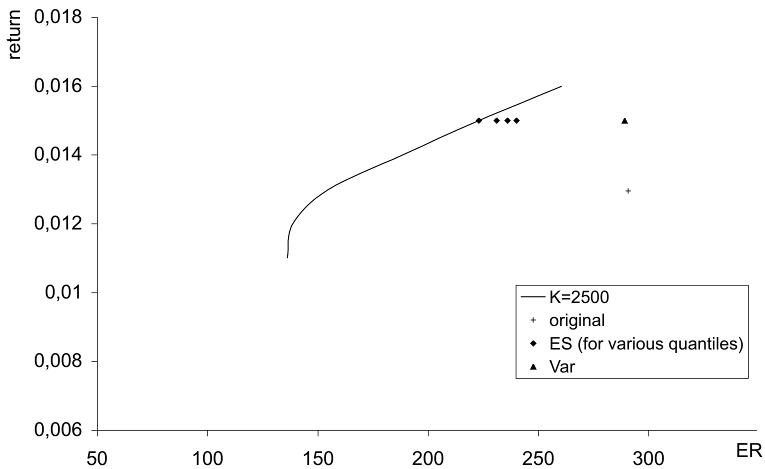


FIGURE 6. ER efficient frontier vs. Variance / ES optimized portfolios.

5. Concluding remarks

Variance optimization is comparatively simple, but leads to portfolios that are not necessarily optimal with respect to expected shortfall or expected regret. Nevertheless, the variance efficient frontier is certainly a useful benchmark for bond portfolio managers: In this case, a decrease in asset value due to a downgrading is much more likely than an actual default. However, if the objective is to manage a portfolio of rather illiquid bank loans, the time horizon is fairly long, corresponding to a buy-and-hold strategy for bond portfolios. Thus, for bank loan portfolios, optimization with respect to expected shortfall or expected regret is certainly more appropriate. Even though optimization with respect to expected shortfall or expected regret looks fairly complicated, it is actually numerically tractable even for portfolios with many obligors and can be done on a PC using standard software.

6. Appendix: The minimization formula

We introduce the abbreviation $f(w,y) = \sum_{i=1}^{n} w_i y_i$ where $y = (y_1, \dots, y_n)$. Assuming that $f(w,(L_1,\dots,L_n))$ has a density g with respect to the n-dimensional Lebesgue measure, we have

$$E(f(w,y)|f(w,y) \geq q_\alpha(w)) = \frac{1}{P(f(w,y) \geq q_\alpha(w))} \int\limits_{f(w,y)\geq q_\alpha(w)} f(w,y)\, dP$$

$$= q_\alpha(w) + \frac{1}{1-\alpha} \int\limits_{f(w,y)\geq q_\alpha(w)} (f(w,y) - q_\alpha(w)) g(y)\, dy\ .$$

Regarding this expression as a function of $q_\alpha(w)$, we arrive at

Theorem 6.1. *Let* $F_\alpha(w,q) = q + \frac{1}{1-\alpha} \int\limits_{f(w,y)\geq q} (f(w,y) - q) g(y)\, dy$. $F_\alpha(w,\cdot) : \mathbb{R}^+ \to \mathbb{R}$ *is finite and convex. Its minimum is at* $q = q_\alpha(w) = \mathrm{VaR}_\alpha(L)$:

$$F_\alpha(w, q_\alpha(w)) = \min_q F_\alpha(w,q).$$

The *proof* is done by setting the derivative w.r.t. q to 0 and can be found in [11].

Rewriting $F_\alpha(w,q)$ as

$$F_\alpha(w,q) = q + \frac{1}{1-\alpha} \int\limits_{y\in R^n} (f(w,y) - q)^+ g(y)\, dy$$

we approximate the integral by the empirical loss distribution

$$q + \frac{1}{1-\alpha} M^{-1} \sum_{m=1}^{M} (f(w,y_m) - q)^+$$

where $y_m = (y_1^m, \dots, y_n^m) \in \mathbb{R}_+^n$ are the empirical losses generated in M Monte Carlo simulations. This idea of replacing the integral expression by the empirical loss distribution which can be quickly calculated is due to Rockafellar and Uryasev. Please note that the existence of a density function g is not essential since, by the Glivenko-Cantelli theorem, the sequence of empirical loss distributions converges weakly to the loss distribution derived by the underlying model.

Replacing $(f(w,y_m) - q)^+$ by dummy variables z_m and imposing additional constraints

$$z_m \geq f(w,y_m) - q, \quad z_m \geq 0, \quad m = 1,\dots,M$$

we get rid of the nonlinearity and arrive at the optimization problem for expected shortfall.

This optimization problem can easily be generalized to the case of importance sampling: In this case the terms in $M^{-1}\sum_{m=1}^{M}(f(w,y_m)-q)^+$ do not have the same weight $\frac{1}{M}$ but individual weights. Details on importance sampling can be found in [5].

The formulation of the optimization problem for expected regret is derived similarly.

The views in this paper express the authors' opinions and do not necessarily represent the views of Deutsche Bank AG or BaFin.

References

[1] C. Acerbi and D. Tasche, *On the coherence of expected shortfall*, Journal of Banking and Finance, **26 (7)**, 1487–1503.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath,*Coherent measures of risk*, Mathematical Finance, **9** (1999), 203–228.

[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley, New York, 1993.

[4] S. Jaschke and U. Küchler, *Coherent risk measures and good-deal bounds*, Finance Stoch., **5 (2)** (2001), 181–200.

[5] M. Kalkbrener, H. Lotter, and L. Overbeck, *Sensible and efficient capital allocation for credit portfolios*, RISK, January 2004, S19-S24.

[6] S. Kealhofer, *Portfolio management of default risk*, Net Exposure, **1 (2)** (1998), 12.

[7] H. Markovitz, *Portfolio selection*, Journal of Finance, **7** (1952), 77–91.

[8] H. Mausser and D. Rosen, *Applying scenario optimization to portfolio credit risk*, Algo Research Quarterly, **2 (2)** (1999), 19–33.

[9] H. Mausser and D. Rosen, *Efficient risk/return frontiers for credit risk*, Algo Research Quarterly, **2 (4)** (1999), 35–48.

[10] M. Padberg, *Linear Optimization and Extensions*, Springer, New York, 2000.

[11] R. T. Rockafellar and St. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk, **2 (3)** (2000), 21–41.

[12] W. Sharpe, *Capital asset prices: A theory of market equilibrium under conditions of risk*, Journal of Finance, **19 (3)** (1964), 425–442.

Claas Becker
Deutsche Bank AG, Taunusanlage 12, 60325 Frankfurt, Germany
e-mail: `claas.becker@db.com`

Veronika Orlovius
Bundesanstalt für Finanzdienstleistungsaufsicht, Graurheindorfer Str. 108, 53117 Bonn, Germany
e-mail: `veronika.orlovius@bafin.de`

Progress in Probability, Vol. 59, 445–455

An Orlicz Spaces Duality for Utility Maximization in Incomplete Markets

Sara Biagini

Abstract. Biagini (2004) and Biagini-Frittelli (2005) faced the utility maximization problem in incomplete markets when the price process of financial assets is described by general semimartingales that are not necessarily locally bounded. They introduced a class of well-controlled admissible strategies in this (very) risky context and then they solved the maximization problem with an (L^∞, ba)-duality technique.

In this note we almost stick to their setup and we show that their dual result can be obtained via an Orlicz spaces duality, naturally associated with the utility function considered. This new formulation gives additional insight into the nature of the loss control in the good trading strategies.

Mathematics Subject Classification (2000). Primary 60G48, 60G44, 46E30, 49N15, 91B28; Secondary 46N10, 91B16.

Keywords. Utility maximization, unbounded semimartingale, Orlicz spaces, duality methods, incomplete markets, σ-martingale measure.

1. Introduction

As in Biagini [3] and Biagini-Frittelli [2], we are interested in the utility maximization problem:

$$\sup_{H \in \mathcal{H}} E[u(x + (H \cdot X)_T)] \tag{1.1}$$

in which $u : \mathbb{R} \to \mathbb{R}$ is the utility function, $x \in \mathbb{R}$ is the constant initial endowment, $T \in (0, +\infty]$ is a fixed time horizon, X is an $\mathbb{R}^d$-valued càd-làg semimartingale defined on a filtered stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ that satisfies the usual assumptions and $\mathcal{H}$ is an appropriate class of admissible integrands. The semimartingale X models the discounted evolution of the prices of d underlyings and it is not assumed to be locally bounded.

The author is very grateful to M. Frittelli and W. Schachermayer.

Assumption 1. *The utility* $u : \mathbb{R} \to \mathbb{R}$ *is a strictly concave increasing differentiable function satisfying the Inada conditions*

$$\lim_{x\to-\infty} u'(x) = +\infty \text{ and } \lim_{x\to+\infty} u'(x) = 0,$$

and having Reasonable Asymptotic Elasticity (RAE(u)), as defined by Schachermayer [9, Definition 1.5]*:*

$$\lim_{x\to-\infty} \frac{xu'(x)}{u(x)} > 1, \tag{1.2}$$

$$\lim_{x\to+\infty} \frac{xu'(x)}{u(x)} < 1. \tag{1.3}$$

Since we exploit a duality technique in solving problem (1.1), we define as usual the convex conjugate of the utility function u, Φ:

$$\Phi : \mathbb{R}_+ \to \mathbb{R}, \quad \Phi(y) \triangleq \sup_{x\in\mathbb{R}} \{u(x) - xy\}.$$

From Assumption 1, Φ is a strictly convex differentiable function satisfying $\Phi(+\infty) = +\infty$, $\Phi(0^+) = u(+\infty)$, $\Phi'(0^+) = -\infty$, $\Phi'(+\infty) = +\infty$ and $(u')^{-1} = -\Phi'$. Moreover, a well-known consequence of the Reasonable Asymptotic Elasticity of u (see Corollary 4.2, Schachermayer [9]) is that Φ satisfies a growth property defined as follows.

Definition 1.1. Let Φ be a convex function with $\mathbb{R}_+$ as proper domain. Then Φ has the G-growth property ($G(\Phi)$ in short) if: for each compact interval $[\lambda_0, \lambda_1]$ contained in $(0, +\infty)$ there exist constants $\alpha > 0$ and $\beta > 0$ such that:

$$\Phi(\lambda y) \leq \alpha\Phi(y) + \beta(y + 1), \text{ for } y > 0 \text{ and } \lambda \in [\lambda_0, \lambda_1].$$

There are two (convex) sets of probabilities that naturally arise in this framework. Firstly,

$$P_\Phi = \left\{Q \ll P \mid E\left[\Phi\left(\frac{dQ}{dP}\right)\right] < +\infty\right\}$$

is the set of P-a.c. probability measures with finite generalized relative entropy. Since Φ has $G(\Phi)$, $Q \in P_\Phi$ iff $Q \in P_{\Phi_\lambda}$ for all $\lambda > 0$, where $\Phi_\lambda(y) \triangleq \Phi(\lambda y)$. Secondly,

$$M_\sigma = \{Q \ll P : X \text{ is a } \sigma\text{-martingale w.r.t. } Q\}$$

consists of all the $P-$absolutely continuous σ-martingale measures for X, i.e., of those $Q \ll P$ such that there exists a process $\eta_Q > 0$ which is predictable, X-integrable w.r.t. the probability P (in short, $\eta_Q \in L(X)(P)$), and with the property that the integral $\eta_Q \cdot X$ is a Q-martingale.

The relevance of the concept of σ-martingale measure in financial mathematics was first shown by Delbaen and Schachermayer [4]. These authors proved that a σ-martingale measure for X is a good pricing instrument when the semimartingale X is non necessarily locally bounded. The set M_σ is thus a generalization of the set of P-absolutely continuous local martingale measures for X. In fact, when X is locally bounded, M_σ boils down to the set of local martingale measures.

Up to this point we haven't said a word about the class $\mathcal{H}$ of integrands in problem (1.1). This is a delicate point. In fact, when X is not locally bounded, the classical set of integrands H that give rise to losses in the trading bounded from below:

$$(H \cdot X)_t \geq -c \ \forall t \in [0, T]$$

may be reduced to the zero integrand, so that the utility maximization problem over this trivial class is meaningless (see the examples in [2, 3] or Remark 2.6 in [9]).

Therefore both in [3] and [2] a more general class of admissible integrands was introduced. In the first reference, an adapted increasing positive process Y is used to control the losses in the trading. But since we maximize utility from terminal wealth (see [3, page 5]), it is harmless to substitute the adapted process with its terminal value $Y_T = W$ in the control.

Since we also work within the general setting of X possibly non locally bounded, we will use the good strategies in the class $\mathcal{H}^W$ as domain in problem (1.1), exactly as done in [2]:

$$\mathcal{H}^W \triangleq \{H \in L(X)(P) \mid \exists c > 0 \text{ s.t. } (H \cdot X)_t \geq -cW, \ t \in [0, T]\}. \tag{1.4}$$

Now the stochastic integrals are no more bounded from below by a constant, but (modulo a scaling factor c) by a "sufficiently integrable" random variable W. We assume here simply that $W > 0$ (and not $W \geq 1$).

The economic significance of this selection of strategies is that in a possibly highly risky market, the agent has to face more risk to improve her maximum expected utility.

Our result is a new proof of the fundamental duality relation proved in [2][1], Theorem 2 and Corollary 1, i.e., (under some extra technical conditions)

$$\sup_{k \in K^W} E[u(x+k)] = \sup_{k \in K^{\mathcal{W}}} E[u(x+k)] = \min_{\lambda > 0, Q \in M_\sigma \cap P_\Phi} \lambda x + E[\Phi(\lambda \frac{dQ}{dP})] \tag{1.5}$$

where we rewrite problem (1.1) in terms of the terminal values from W-admissible strategies $K^W \triangleq \{k \mid k = (H \cdot X)_T, H \in \mathcal{H}^W\}$ and where $K^{\mathcal{W}} = \cup_{W \in \mathcal{W}} K^W$ stands for the union of *all the well-controlled terminal values.* The random control W is in fact allowed to vary in a convex set $\mathcal{W}$, to be defined precisely in Section 2.

The new proof we present here is based on an Orlicz spaces duality $(M^{\widehat{u}}, L^{\widehat{\Phi}})$, in which the Young functions $\widehat{u}, \widehat{\Phi}$ defining the Orlicz spaces are naturally associated to u, Φ (see Sections 3 and 4). This new description gives also an extra insight into the nature of the class of controls $\mathcal{W}$ and henceforth it leads to a better understanding of the W-admissible integrands.

But first we must recollect more of the setup in [3] and [2].

[1]The relation $\sup_{k \in K^W} E[u(x+k)] = \min_{\lambda > 0, Q \in M_\sigma \cap P_\Phi} \lambda x + E[\Phi(\lambda \frac{dQ}{dP})]$ was first proved in [3, Theorem 11] with an (L^∞, ba)-duality technique.

2. More details on the Biagini and Frittelli setup

In order to build a reasonable utility maximization, in [3] as well as in [2] some restrictions were put on the control W: it must satisfy two conditions that are both mathematically useful and economically meaningful. Hereafter we refer to the notation in [2]. The first condition on W guarantees that $\mathcal{H}^W$ is rich enough for trading purposes:

Definition 2.1 ([2]). A random variable $W \in L^0(P)$ is X-**suitable** (or simply suitable) if $W > 0$ P-a.s., and for all $1 \le i \le d$ there exists a process $H^i \in L(X^i)(P)$ such that $P(\{\omega \mid \exists t \ge 0\, H^i_t(\omega) = 0\}) = 0$ and

$$-W \le (H^i \cdot X^i)_t \le W, \text{ for all } t \in [0, T]. \tag{2.1}$$

The second condition implies that the W-admissible trading strategies are compatible with the preferences, i.e. it assures that the expected utility of terminal wealths $x + (H \cdot X)_T$ from all W-admissible trading strategies never equals $-\infty$:

Definition 2.2 ([2]). A random variable $W \in L^0(P)$ is u-**compatible** (or simply compatible) if $W > 0$ P-a.s. and

$$E[u(-cW)] > -\infty \;\; \forall c > 0. \tag{2.2}$$

$\mathcal{W}$ is then defined as the convex set of X-suitable and u-compatible random variables, i.e., it is the set of good loss controls. We recall that when X is locally bounded, $\mathcal{W}$ is always not empty, since $W = 1 \in \mathcal{W}$ ([3, Proposition 5], [2, Proposition 4]). However, in the non-locally bounded case there is no natural selection of the particular $W \in \mathcal{W}$, if there is *any*. This is the reason why we require:

Assumption 2. $\mathcal{W} \neq \emptyset$.

As regards the additional results in [2], the authors showed that the left hand side in the dual relation (1.5) is in general only a supremum, not necessarily a maximum. However, they were able to find the optimal solution f_x in a domain larger than K^W and they showed that f_x can be represented as terminal value of a more general stochastic integral, *not necessarily well-controlled by any* $W \in \mathcal{W}$: this is the main novelty w.r.t. [3]. Finally, they proved that this optimal stochastic integral is a supermartingale w.r.t. every $Q \in M_\sigma \cap P_\Phi$, thus extending to the general case the results that hold true in case X is locally bounded (see Schachermayer [10]).

All this can be recovered also on the basis of our Assumptions 1, 2 and the next Orlicz duality.

3. The Orlicz spaces associated to Φ and to u

Consider the even function $\widehat{\Phi} : \mathbb{R} \to [0, +\infty)$ defined by:

$$\widehat{\Phi}(y) \triangleq \Phi(|y| + \beta) - \Phi(\beta),$$

where $\beta > 0$ is the unique solution of the equation $\Phi'(y) = 0$. It is very easy to see that $\widehat{\Phi}$ is a Young function (see, e.g., the standard reference on Orlicz spaces [8], page 13). In addition, it is a **Nice Young function** because:

1. it is regular and on $\mathbb{R}_+$ it is strictly increasing;
2. $\widehat{\Phi} = 0$ iff $y = 0$;
3. $\widehat{\Phi}'(0) = 0$;
4. $\lim_{y\to+\infty} \widehat{\Phi}'(y) = +\infty$.

As a consequence, the Orlicz space $L^{\widehat{\Phi}} = \{\text{r.v. } f \mid \exists\alpha > 0\, E[\widehat{\Phi}(\alpha f)] < +\infty\}$ is well defined. We recall that $L^{\widehat{\Phi}}$ is a Banach lattice with the pointwise operations and the gauge (or $\widehat{\Phi}$-) norm:

$$\|f\|_{\widehat{\Phi}} = \inf\left\{c > 0 \mid E\left[\widehat{\Phi}\left(\frac{f}{c}\right)\right] \le 1\right\}.$$

The containments

$$L^\infty(P) \subseteq L^{\widehat{\Phi}} \subseteq L^1(P)$$

hold since $\widehat{\Phi}$ is finite, regular on $\mathbb{R}$ and convex.

Remark 3.1. Note that we never require $\Phi(0) < +\infty$ (which is equivalent to requiring $u(+\infty) < +\infty$).

Proposition 3.2. *Since Φ has $G(\Phi)$ (Definition* 1.1*), then $\widehat{\Phi}$ also satisfies $G(\widehat{\Phi})$ on $\mathbb{R}_+$.*

Proof. In fact, fix $y > 0$: $\widehat{\Phi}(\lambda y) \le \widehat{\Phi}(\lambda_1 y) = \Phi(\lambda_1 y + \beta) - \Phi(\beta)$ if $\lambda \in [\lambda_0, \lambda_1]$. Consider first $y \ge y_0 = \max(\frac{\beta}{\lambda_1}, \beta)$: then there exist positive constants K, h, c such that $\Phi(\lambda_1 y + \beta) \le \Phi(2\lambda_1 y) \le K\Phi(y) + h(y+1) \le K\widehat{\Phi}(y) + c(y+1)$. Since $\widehat{\Phi}$ is bounded, say less than C if $y \le \max\{y_0, \lambda_1 y_0\}$, then we immediately derive: $\widehat{\Phi}(\lambda y) \le K\widehat{\Phi}(y) + \max(C, c)(y+1)$. □

Corollary 3.3. *The space $\{f \mid E[\widehat{\Phi}(f)] < +\infty\}$ is linear and coincides with $L^{\widehat{\Phi}}$. As a consequence, given f, there exists* some *$\alpha > 0$ such that $E[\widehat{\Phi}(\alpha f)] < +\infty$ iff for* all *$\alpha > 0$, $E[\widehat{\Phi}(\alpha f)] < +\infty$.*

Proof. $G(\widehat{\Phi})$ is analogous to the Δ_2-condition[2] in Orlicz Spaces theory, so the result follows from [8, Corollaries 3.4.4 and 3.4.5]. □

Remark 3.4. $E[\,\Phi(|f|)\,] < +\infty \Rightarrow f \in L^{\widehat{\Phi}}$. The converse holds only in case $\Phi(0) < +\infty$.

Consider now $\widehat{u}$, the convex conjugate of $\widehat{\Phi}$. Simple calculations show that its expression in terms of u is

$$\widehat{u}(x) = -u(-|x|) - u'(0)|x| + u(0)$$

[2]A Young function Υ is said to satisfy the Δ_2-condition if there exists $K > 0$ such that $\Upsilon(2y) \le K\Upsilon(y)$ for all $y \ge 0$.

and that $\widehat{u}$ is also a Nice Young function. So, we can associate to $\widehat{u}$ the Orlicz space $L^{\widehat{u}}$ as well:

$$L^{\widehat{u}} = \{\text{r.v. } f \mid \exists \alpha > 0\, E[\widehat{u}(\alpha f)] < +\infty\}$$

with the gauge norm:

$$\|f\|_{\widehat{u}} = \inf \left\{ c > 0 \mid E\left[\widehat{u}\left(\frac{f}{c}\right)\right] \leq 1 \right\}.$$

The ratio behind this construction is the following: the concavity of u reflects the risk aversion of the investor, who weights the losses more severely than she considers the gains. Therefore, the Young function $\widehat{u}$ has the same asymptotic behavior of the negative part of u.

As regards growth conditions, being the convex conjugate of a function satisfying $G(\widehat{\Phi})$, $\widehat{u}$ in general does satisfy neither $G(\widehat{u})$ nor the analogous Δ_2-condition. So $L^{\widehat{u}}$ doesn't necessarily have the property stated in Corollary 3.3.

Hence, we introduce a subspace of $L^{\widehat{u}}$ that has exactly the 'homogeneity property' of Corollary 3.3:

$$M^{\widehat{u}} \triangleq \{f \in L^{\widehat{u}} \mid E[\widehat{u}(\alpha f)] < +\infty\, \forall \alpha > 0\}. \tag{3.1}$$

$M^{\widehat{u}}$ is always a linear subspace of the Orlicz space $L^{\widehat{u}}$, but in our context it is also *closed.* The reason is that $\widehat{u}(x) = 0$ iff $x = 0$, as proved in [8, Proposition 3.4.3], where it is also shown that $M^{\widehat{u}}$ coincides with the closure of L^∞, which in general is not dense in $L^{\widehat{u}}$:

$$M^{\widehat{u}} = \overline{L^\infty}^{\widehat{u}}.$$

Hence $M^{\widehat{u}}$ is also a Banach lattice with the inherited $\widehat{u}$-norm. In addition, Theorem 4.1.7 in [8] gives:

$$(M^{\widehat{u}})^* = L^{\widehat{\Phi}}, \tag{3.2}$$

i.e., the topological dual of $M^{\widehat{u}}$ is *exactly* $L^{\widehat{\Phi}}$. To be precise, the dual norm on $L^{\widehat{\Phi}}$ and the $\widehat{\Phi}$-norm are not equal, but equivalent. This subtlety however doesn't affect our application: the triple $(M^{\widehat{u}}, L^{\widehat{\Phi}}, E[\cdot,\cdot])$ is indeed the dual system we will use.

Remark 3.5 (On Asymptotic Elasticity of $\widehat{u}$). In [8, Corollary 2.3.4] it is shown that the Δ_2-growth condition on $\widehat{\Phi}$ is *equivalent* to

$$\lim_{x \to +\infty} \frac{x\widehat{u}'(x)}{\widehat{u}(x)} = \lim_{x \to -\infty} \frac{xu'(x)}{u(x)} > 1.$$

This relation in turn is exactly the Reasonable Asymptotic Elasticity condition on u at $-\infty$, so that we recover part of Assumption 1 (relation (1.2)). However, passing from Φ to $\widehat{\Phi}$ is not harmless: we lose information about the behavior of Φ around 0, which is equivalent to losing information on the behavior of u around $+\infty$. That is one of reasons why it is not clear yet how to relate the Δ_2-condition to condition (1.3) (though an attempt has already been made in [6, Section 6]).

Here is a concrete example of the spaces just introduced.

Example. Let the utility be exponential: $u(x) = -e^{-x}$. Then $\Phi(y) = y \ln y - y$ and:

$$\widehat{u}(x) = e^{|x|} - |x| - 1$$

while

$$\widehat{\Phi}(y) = \Phi(\,|y|+1\,) - \Phi(1) = (\,|y|+1\,)\ln(\,|y|+1\,) - |y|.$$

Therefore:

$$L^{\widehat{u}} = \left\{ f \mid \exists \alpha > 0 \text{ s.t. } E\left[e^{\alpha|f|}\right] < +\infty \right\},$$

$$M^{\widehat{u}} = \left\{ f \mid \forall \alpha > 0 \; E\left[e^{\alpha|f|}\right] < +\infty \right\}$$

and

$$L^{\widehat{\Phi}} = \{ g \mid E\left[(\,|g|+1\,)\ln(\,|g|+1\,)\right] < +\infty \}.$$

We could remove the linear terms of $\widehat{u}, \widehat{\Phi}$ in the above characterizations thanks to convexity. Note that $M^{\widehat{u}}$ consists of those random variables that have all the (absolute) exponential moments finite (e.g., gaussian or bounded variables). On the contrary, an exponentially distributed r.v. f is an example of variable in $L^{\widehat{u}} - M^{\widehat{u}}$ and henceforth it cannot be approximated in $\widehat{u}$-norm with bounded random variables.

4. The Orlicz duality $(M^{\widehat{u}}, L^{\widehat{\Phi}})$ in the utility maximization problem

The definition of $M^{\widehat{u}}$ should remind us of the compatibility condition (2.2) on the loss bound $W \in \mathcal{W}$. In fact, $\mathcal{W} \subset M^{\widehat{u}}_+$, as shown below.

Proposition 4.1. *The set of loss bounds $\mathcal{W}$ is contained in $M^{\widehat{u}}_+$.*

Proof. Fix a generic $W \in \mathcal{W}$: W is positive, so that $u(0) > E[u(-\alpha W)] > -\infty$ for all $\alpha > 0$ by compatibility of W and monotonicity of u. This implies $E[-u(-\alpha W)] < +\infty$. By convexity of $E[-u(-\cdot)]$, we derive: $\alpha W \in L^1(P)$ and finally $E[\widehat{u}(\alpha W)] < +\infty$ for all $\alpha > 0$. □

Proposition 4.2. *If $k \in K^{\mathcal{W}}$ and if $n \in \mathbb{N}$, then $k \wedge n \in M^{\widehat{u}}$.*

Proof. First of all, $k^+ \wedge n \in L^\infty \subset M^{\widehat{u}}$. Also, $0 \le k^- \le cW$ for some $W \in \mathcal{W}$ and some positive c. Since $cW \in M^{\widehat{u}}$ and $\widehat{u}$ is monotone, k^- also belongs to $M^{\widehat{u}}$. The thesis follows then from the identity $k \wedge n = k^+ \wedge n - k^-$. □

The above proposition, together with an application of Fatou's lemma, gives:

Corollary 4.3. *Fix $W \in \mathcal{W}$ and let $C^W \triangleq (K^W - M^{\widehat{u}}_+) \cap M^{\widehat{u}}$. Then*

$$\sup_{k \in K^W} E[u(x+k)] = \sup_{f \in C^W} E[u(x+f)]. \tag{4.1}$$

So, we can formulate the maximization over a Banach lattice $M^{\widehat{u}}$ naturally induced by the problem.

Note that the identity in Corollary 4.3 and the duality result (1.5) from [2] would immediately lead to the dual formula:

$$\sup_{f\in C^W} E[u(x+f)] = \sup_{k\in K^{\mathcal{W}}} E[u(x+k)] = \min_{\lambda>0, Q\in M_\sigma\cap P_\Phi} \lambda x + E[\Phi(\lambda \frac{dQ}{dP})]$$

but we will obtain the same result via an $(M^{\widehat{u}}, L^{\widehat{\Phi}})$-duality, which seems indeed the most natural approach. In what follows, we always refer to this dual system $(M^{\widehat{u}}, L^{\widehat{\Phi}}, E[\cdot,\cdot])$. We indicate with A^0_1 the normalized polar of a set A, i.e., the set of r.v. $g\in A^0$ with $E[g]=1$.

Lemma 4.4. *Set $C^{\widehat{u}} = (K^{\mathcal{W}} - M^{\widehat{u}}_+)\cap M^{\widehat{u}}$. Then:*

$$(C^W)^0_1 = (C^{\widehat{u}})^0_1 = M_\sigma \cap L^{\widehat{\Phi}},$$

that is, the above normalized polars coincide with $M_\sigma\cap L^{\widehat{\Phi}}$, i.e., the σ-martingale measures for X that belong to $L^{\widehat{\Phi}}$ (equivalently, that have finite $\widehat{\Phi}$-entropy).

Proof. We only prove the identity $(C^W)^0_1 = M_\sigma\cap L^{\widehat{\Phi}}$, the one with $C^{\widehat{u}}$ being analogous.

a- $M_\sigma\cap L^{\widehat{\Phi}} \subseteq (C^W)^0_1$. This containment follows from the following considerations: if $W\in\mathcal{W}$, then exactly as in Proposition 6 c) in [2] Fenchel's inequality implies $W\in L^1(Q)$ for all $Q\in M_\sigma\cap L^{\widehat{\Phi}}$. Henceforth, the bound in the definition (1.4) still permits to apply Ansel and Stricker's result [1] and to deduce that

$$H\in\mathcal{H}^W \Rightarrow H\cdot X \text{ is a supermartingale under all } Q\in M_\sigma\cap L^{\widehat{\Phi}}$$

and this obviously implies:

$$E_Q[k]\le 0 \;\forall Q\in M_\sigma\cap L^{\widehat{\Phi}} \quad, \forall k\in C^W \text{ and } \forall W\in\mathcal{W}.$$

b- To prove the opposite inclusion, observe that $(C^W)^0_1$ is made of probability measures $Q\in L^{\widehat{\Phi}}$: therefore, these Qs integrate $W\in M^{\widehat{u}}$. By the suitability assumption on W, the random variables: $\pm(H^i I_A I_{]s,t]}\cdot X)_T$ are in C^W for all $s<t, A\in\mathcal{F}_s$, where the integrands H^i are those in (2.1). Hence for all $i=1,\dots,d$, $H^i\cdot X$ is a Q-martingale for all $Q\in(C^W)^0_1$, which amounts to saying that X is a σ-martingale under Q. □

Lemma 4.5. *The concave functional $I_u : M^{\widehat{u}}\to\mathbb{R}$ is norm-continuous.*

Proof. First we show that I_u is proper and it is norm continuous on the interior of its effective domain.

Thanks to [5, Proposition I.2.5], this is equivalent to showing that there is a non-empty open set O on which I_u is not everywhere equal to $+\infty$ and it is bounded below by a constant $c\in\mathbb{R}$. What we show is that on the open unit ball B of $M^{\widehat{u}}$ the functional I_u is i) finite and ii) uniformly bounded below.

i) If $b\in B$, then by Jensen's inequality $I_u(b)\le u(E[b]) < +\infty$.

ii) If $b \in B$:

$$\|b\|_{\widehat{u}} = \inf\left\{\alpha > 0 \mid E\left[\widehat{u}\left(\frac{b}{\alpha}\right)\right] \leq 1\right\} < 1,$$

and consequently $E[\widehat{u}(b)] \leq 1$, as well as $E[\widehat{u}(b^-)] \leq 1$. By convexity of $\widehat{u}$, this implies $E[b^-] \leq C$ for all b. Hence

$$-I_u(-b^-) = E[-u(-b^-)] = E[\widehat{u}(b^-)] + u'(0)E[b^-] - u(0) \leq 1 + u'(0)C - u(0) = K$$

and so $I_u(b) \geq I_u(-b^-) \geq -K$.

Finally, with a similar technique it is not difficult to see that the effective domain of I_u is the entire $M^{\widehat{u}}$. If $f \in M^{\widehat{u}}$, from the very definition (3.1) we have in particular

$$E[\widehat{u}(f)] < +\infty$$

which implies $E[\widehat{u}(f^-)] < +\infty$ (which is equivalent to $E[u(-f^-)]$ finite) and $f^+ \in L^1$. Hence, $E[u(f)] \in \mathbb{R}$. $\square$

Finally, here is our duality Theorem.

Theorem 4.6. *If there exist $W_0 \in \mathcal{W}$ such that $\sup_{k \in K^{W_0}} E[u(x+k)] < u(+\infty)$, then for all $W \in \mathcal{W}$ we get*

$$\sup_{k \in K^W} E[u(x+k)] = \sup_{k \in K^{\mathcal{W}}} E[u(x+k)] = \sup_{f \in C^{\widehat{u}}} E[u(x+f)] \tag{4.2}$$

$$= \min_{\lambda > 0, Q \in M_\sigma \cap L^{\widehat{\Phi}}} \lambda x + E[\Phi(\lambda \frac{dQ}{dP})] \tag{4.3}$$

$$= \min_{\lambda > 0, Q \in M_\sigma \cap P_\Phi} \lambda x + E[\Phi(\lambda \frac{dQ}{dP})]. \tag{4.4}$$

Proof. We only prove the dual formula

$$\sup_{k \in K^{W_0}} E[u(x+k)] = \sup_{f \in C^{W_0}} E[u(x+f)] = \min_{\lambda > 0, Q \in M_\sigma \cap P_\Phi} \left\{\lambda x + E\left[\Phi\left(\lambda \frac{dQ}{dP}\right)\right]\right\}. \tag{4.5}$$

From this relation it is easy to derive the same result for any other $W \in \mathcal{W}$. In fact, the rhs in formula (4.5) does not depend on the initial W_0.

To start, the equality $\sup_{k \in K^{W_0}} E[u(x+k)] = \sup_{f \in C^{W_0}} E[u(x+f)]$ holds thanks to Corollary 4.3. We then want to show:

$$\sup_{f \in C^{W_0}} E[u(x+f)] = \min_{\lambda > 0, Q \in M_\sigma \cap P_\Phi} \left\{\lambda x + E\left[\Phi\left(\lambda \frac{dQ}{dP}\right)\right]\right\},$$

and the proof will be split into three steps.

Step 1. Set $I_u(f) \triangleq E[u(x+f)], f \in M^{\widehat{u}}$ and let $(I_u)^*$ be the convex conjugate functional of I_u, i.e.,

$$(I_u)^*(g) = \sup_{f \in M^{\widehat{u}}} I_u(f) - E[fg], \quad g \in L^{\widehat{\Phi}}.$$

Thanks to norm-continuity of I_u over $M^{\widehat{u}}$, Fenchel's duality theorem can be applied to get

$$\sup_{f \in C^{W_0}} I_u(x+f) = \min_{g \in (C^{W_0})^0} (I_u)^*(g). \tag{4.6}$$

Step 2. Now we would like to find a concrete expression for $(I_u)^*$. Following the terminology of Rockafellar [7], the space $M^{\widehat{u}}$ is decomposable[3]. Hence Theorem 21 in [7], part a) gives the formula for $(I_u)^*$:

$$(I_u)^*(g) = xE[g] + E[\Phi(g)].$$

So the dual formula (4.6) can be rewritten as

$$\sup_{f \in C^{W_0}} I_u(x+f) = \min_{g \in (C^{W_0})^0} xE[g] + E[\Phi(g)]$$

and after a standard normalization,

$$\sup_{f \in C^{W_0}} I_u(x+f) = \min_{\lambda > 0, Q \in (C^{W_0})^0_1} \lambda x + E\left[\Phi\left(\lambda \frac{dQ}{dP}\right)\right]$$

where the parameter $\lambda > 0$ since the hypothesis $\sup_{k \in K^{W_0}} E[u(x+k)] < u(+\infty)$ implies that the dual optimum cannot be 0.

Step 3. An inspection of the dual function: $\lambda x + E[\Phi(\lambda \frac{dQ}{dP})]$ shows that in fact this expression is finite only on the Q's that belong to P_Φ (here we are also using $G(\Phi)$). Hence we can consider $M_\sigma \cap P_\Phi$, which is a subset of $M_\sigma \cap L^{\widehat{\Phi}}$ (see Remark 3.4), as dual domain, thus proving the identity between (4.3) and (4.4). □

From a (slightly careful) re-reading of the proofs in the last two sections, one could object that no growth conditions on $\widehat{\Phi}$ (or RAE(u)) are actually required, as far as the identity between (4.2) and (4.3) in Theorem 4.6 is concerned. In fact, in [3, Section 2] there are no such requirements on u. Of course, without $G(\Phi)$ (or Δ_2-condition on $\widehat{\Phi}$) Corollary 3.3 wouldn't hold true anymore, but we would continue to have the basic duality (3.2) and we could repeat all the subsequent arguments. However, our goal is to reformulate the setting of [2] up to the point from which we can deduce the same results: and we have already said that in that paper the existence of an optimal primal solution f_x was also shown. To this end, in [2, Assumption 2] it was assumed that $M_\sigma \cap P_\Phi = M_\sigma \cap P_{\Phi_\lambda}$ for all $\lambda > 0$, which is needed to pass from (4.3) to (4.4) in Theorem 4.6. As in all the existing literature, this paves the way to the proof of the existence of the optimal claim f_x. This Assumption 2 is weaker than $G(\Phi)$ but has the disadvantage of being both model and agent dependent. Hence, following [9] we directly assume the RAE(u), which depends only on the utility function.

[3] A space L of random variables on $(\Omega, \mathcal{F}, P)$ is decomposable if, whenever $A \in \mathcal{F}$ and f is a bounded random variable on A, then, for every $x \in L$, $\widetilde{x} = fI_A + xI_{A^c}$ also belongs to L.

Remark 4.7 (Sup vs Max in the primal problem). It is not difficult to see that the domain C^W contains the domain $\mathcal{C} \cdot W$ used in Theorem 2 and Corollary 1, [2]. In spite of this, the supremum of I_u over C^W (or $C^{\widehat{u}}$) is again not a maximum in general, as perhaps should be intuitively clear: if f is the optimal claim in C^W (or in $C^{\widehat{u}}$), then $E[u(-|f|)]$ has to be finite. This is of course more severe than simply requiring $E[u(f)]$ finite, which is equivalent to the two conditions: $E[u(f^+)] < +\infty$ and $E[u(-f^-)] > -\infty$, which imply only $f^- \in L^{\widehat{u}}$ (and not in $M^{\widehat{u}}$!). For an explicit example of this situation we refer to [2], Remark 6.

References

[1] J. P. Ansel and C. Stricker, *Couverture des actifs contingents et prix maximum*, Annales Institute Henri Poincaré, **30** (1994), 303–315.

[2] S. Biagini and M. Frittelli, *Utility maximization in incomplete markets for unbounded processes*, Finance and Stochastics, **9** (2005), 493–517.

[3] S. Biagini, *A new class of strategies with applications to utility maximization*, Università degli Studi di Perugia, Technical Report, 2004.

[4] F. Delbaen and W. Schachermayer, *The fundamental theorem of asset pricing for unbounded stochastic processes*, Mathematische Annalen, **312** (1998), 215–250.

[5] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland, American Elsevier, 1976.

[6] D. Kramkov and W. Schachermayer, *The asymptotic elasticity of utility functions and optimal investment in incomplete markets*, Ann. Appl. Prob., **9 (3)** (1999), 904–950.

[7] R. T. Rockafellar, *Conjugate Duality and Optimization*, Conference Board of Math. Sciences Series, SIAM Publications, No. 16, 1974.

[8] M. M. Rao and Z. D. Ren, *Theory of Orlicz Spaces*, Marcel Dekker Inc., NY, 1991.

[9] W. Schachermayer, *Optimal investment in incomplete markets when wealth may become negative*, The Annals of Applied Probability, **11 (3)** (2001), 694–734.

[10] W. Schachermayer, *A super-martingale property of the optimal portfolio process*, Finance and Stochastics, **4** (2003), 433–457.

Sara Biagini
Dipartimento di Economia, Finanza e Statistica
Università di Perugia, via A. Pascoli 20
06123 Perugia
Italy
e-mail: `s.biagini@unipg.it`

Progress in Probability, Vol. 59, 457–467

No Free Lunch under Transaction Costs for Continuous Processes

Paolo Guasoni

Abstract. We present a version of a No Free Lunch and Hedging Theorem for security markets under transaction costs for continuous processes. We show that the (RNFL) condition, which requires that the absence of free lunches is preserved under a smaller bid-ask spread, is equivalent to the existence of a uniformly strictly consistent price system. We also characterize the super-replication price of bounded contingent claims as the supremum of expected values under all uniformly consistent price systems.

Mathematics Subject Classification (2000). Primary 91B28; Secondary 62P05.

Keywords. Arbitrage, fundamental theorem of asset pricing, transaction costs, hedging.

1. Introduction

We show a version of the No Free Lunch Theorem under transaction costs for continuous processes, which can be proven with a limited set of prerequisites, which include functional analysis, probability and martingale theory, but not stochastic integration.

In the same spirit of Jouini & Kallal [7], we consider the model of a *security market*, where risky assets can only be exchanged with cash and viceversa. In the case of positive prices, this setting is a particular case of the *pure exchange* model introduced by Kabanov [8], and further studied in a series of papers by Kabanov, Rásonyi & Stricker [9, 10, 11], Schachermayer [15] and Campi & Schachermayer [2]. However, in general we allow prices to become negative, which may be useful to model futures and insurance contracts, where the limited liability condition does not necessarily hold.

Partially supported by the National Science Foundation under grant DMS-0532390 at Boston University.

The rest of the paper is organized as follows: in Section 2 we define the model in detail, and state the main theorems. Section 3 contains the proof for the case of continuous processes. With some modifications, this proof can be adapted to the more general case of quasi-left continuous asset prices, but we do not pursue this here.

2. Preliminaries and main results

We consider a market model with a riskless and a risky asset, based on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, satisfying the usual assumptions of right-continuity and saturatedness. The riskless asset is used as numeraire, hence its price is assumed constantly equal to 1. An investor trades in the risky asset, according to the strategy $(\theta_t)_{t\in[0,T]}$, representing the number of shares held at time t. We conventionally set $\theta_0 = \theta_T = 0$, as to deal only with cash payoffs.

The bid (selling) and the ask (buying) prices of the risky asset are denoted by $(S - \kappa)_t$ and $(S + \kappa)_t$. Equivalently, each share traded at price S_t incurs a transaction fee of κ_t. We make the following standing assumption:

Assumption 2.1. (S, κ) is a pair of continuous processes, adapted to the filtration $\mathcal{F}_t$, such that $\kappa_t \geq 0$ a.s. for all $t \in [0, T]$.

The assumption $\kappa_t \geq 0$ a.s. is a minimal requirement to rule out static arbitrage by fictitious trading. Unlike most transaction cost models, here prices can become negative, provided that the spread κ remains positive. The two simplest examples are fixed proportional transaction costs for a positive asset price ($S_t > 0$ and $\kappa_t = kS_t$ for $k > 0$), and constant bid-ask spread (constant κ, arbitrary S).

Definition 2.2. A **simple predictable strategy** is a process $\theta = \sum_{i=1}^{n-1} \theta_{\tau_i} 1_{]\tau_i,\tau_{i+1}]}$, where $0 \leq \tau_1 < \cdots < \tau_n < T$ are stopping times and θ_{τ_i} is $\mathcal{F}_{\tau_i}$-measurable. The **cost process** is defined by

$$\begin{aligned} C_t(\theta) &= \sum_{\tau_i \leq t} (S+\kappa)_{\tau_i} (\theta^i - \theta^{i-1})^+ - \sum_{\tau_i \leq t} (S-\kappa)_{\tau_i} (\theta^i - \theta^{i-1})^- \\ &= \sum_{\tau_i \leq t} S_{\tau_i} (\theta^i - \theta^{i-1}) + \sum_{\tau_i \leq t} \kappa_{\tau_i} |\theta^i - \theta^{i-1}| \end{aligned}$$

and the (final) **liquidation value** is defined as $V(\theta) = -C_T(\theta)$.

The process $-C_t(\theta)$ represents the time evolution in the cash position, which reflects purchases and sales, respectively, at bid and ask prices. The strategy begins and ends with a position in cash only, and the liquidation value coincides with the cash at the final horizon T.

As initially recognized by Harrison & Pliska [6], the development of arbitrage theory in continuous time hinges upon the notion of "admissible strategies", which excludes arbitrage arising from doubling strategies.

In frictionless markets, this is achieved enforcing a fixed credit line at all times, but in a transaction-cost setting this is a delicate issue, which critically depends or the timing allowed for liquidation. As shown by a counterexample of Campi & Schachermayer [1], imposing solvability by immediate liquidation is too restrictive, and leads to a set of strategies which lacks the closedness required in no-arbitrage arguments. By contrast, the following definition allows for future liquidation, while preserving the original idea that, in an arbitrage-free environment, one cannot trade one's way out of losses.

Definition 2.3. A simple strategy is **x-admissible** if for all $t \in [0, T]$ there exists a stopping time $\tau \in [t, T]$ a.s., called **liquidation time**, such that

$$x - C_t(\theta) + \theta_t S_\tau - |\theta_t|\kappa_\tau \geq 0. \tag{2.1}$$

We denote the set of simple x-admissible strategies by $\mathcal{A}_x^s$, and the corresponding set of dominated claims by $\mathcal{C}_x^s = \{V(\theta) : \theta \in \mathcal{A}_x^s\} - L_+^0$ and $\mathcal{C}^s = \cup_{x>0}\mathcal{C}_x^s$.

A **trivial arbitrage** is a strategy $\theta = \alpha 1_{]\sigma,\tau]}$, where $\sigma \leq \tau$ are stopping times and α is $\mathcal{F}_\sigma$-measurable, such that $P(V(\theta) \geq 0) = 1$ and $P(V(\theta) > 0) > 0$. A market satisfies the **(NTA)** if $\theta \in \mathcal{A}^s$ and $P(V(\theta) \geq 0) = 1$ implies that $V(\theta) = 0$.

At an intuitive level, a strategy is x-admissible if it requires a collateral of x. Then, at any time t the broker could freeze the agent's account, and liquidate it to cash at a later date τ without realizing a loss. If $\tau = t$ satisfies (2.1), then immediate liquidation is feasible, but in general one may have to wait for liquidity to improve before closing the position. Also, Definition 2.3 depends on the horizon T, since a later horizon allows more time for liquidation.

In absence of both transaction costs and trivial arbitrage, Definition 2.3 reduces to admissibility in the usual frictionless sense. By contradiction, if $x - C_t(\theta) + \theta_t S_\tau \geq 0$ a.s. but $x - C_t(\theta) + \theta_t S_t < 0$ on some event A of positive probability, then the strategy $\theta_t 1_{]t,\tau]\cap A}$ is a trivial arbitrage opportunity.

Finally, note that a trivial arbitrage is an arbitrage through a buy-and-hold strategy, and is 0-admissible by definition. This natural property does not necessarily hold if solvability by immediate liquidation is required.

The fundamental property of Definition 2.3 is that, when absence of arbitrage holds, x-admissible strategies are characterized by their terminal payoffs:

Proposition 2.4. *If (NTA) holds, then* $\mathcal{A}_x^s = \{\theta \text{ simple} : x + V(\theta) \geq 0 \text{ a.s.}\}$.

This, in turn, implies the convexity of the set of simple admissible strategies:

Corollary 2.5. *If (NTA) holds, then $\mathcal{A}_x^s$ is a convex set.*

Proof. Follows from Proposition 2.4 and the convexity of $\theta \mapsto V(\theta)$. $\square$

Motivated by the Robust No Arbitrage (NA^r) condition proposed by Schachermayer [15] in finite discrete time, we introduce the (RNFL) condition, which requires that the No Free Lunch condition is preserved under a smaller bid-ask spread. This condition will imply that any general admissible strategy has finite variation, whereby the cost process is defined as a classical Stieltjes integral.

Definition 2.6. (S,κ) satisfies **(RNFL)** if there exists a pair $(\tilde{S},\tilde{\kappa})$ whose bid-ask spread is a.s. strictly contained within that of (S,κ), pathwise uniformly:

$$\inf_{t\in[0,T]}(\kappa_t - \tilde{\kappa}_t - |S_t - \tilde{S}_t|) > 0 \quad \text{a.s.} \tag{2.2}$$

and such that

$$C = \overline{(\mathcal{C}^{s(\tilde{S},\tilde{\kappa})} - L^0_+)\cap L^\infty}^{\sigma(L^\infty,L^1)} \cap L^\infty_+ = \{0\}. \tag{NFL}$$

Remark 2.7. It is easily checked that (2.2) is equivalent to

$$\inf_{t\in[0,T]}((S+\kappa)_t - (\tilde{S}+\tilde{\kappa})_t) > 0 \quad \text{and} \quad \inf_{t\in[0,T]}((\tilde{S}-\tilde{\kappa})_t - (S-\kappa)_t) > 0 \quad \text{a.s.}$$

which means that the inner bid and ask prices never touch their outer counterparts. Observe also that (RNFL) implies the *efficient friction* condition

$$\inf_{t\in[0,T]} \kappa_t > 0 \qquad \text{a.s.}$$

so the bid-ask spread is always strictly positive, in pathwise uniform sense.

We now define general admissible strategies as limits of simple admissible strategies.

Definition 2.8. Let (S,κ) satisfy (RNFL). A predictable process $(\theta_t)_{t\in[0,T]}$ is an **x-admissible** strategy if there exists an **approximating sequence** $(\theta^n)_{n\geq 1} \subset \mathcal{A}^s_y$, where $y>0$, such that $\limsup_{n\to\infty} V(\theta^n) \geq -x$ and $\theta^n_t \to \theta_t$ a.s. for t in a dense set of $[0,T]$. We denote the set of x-admissible strategies by $\mathcal{A}_x$ and by $\mathcal{A} = \cup_{x>0}\mathcal{A}_x$.

Then we obtain the following:

Proposition 2.9. *Let* (S,κ) *satisfy* (RNFL) *and* $\theta \in \mathcal{A}$. *Then:*

i) θ *is a predictable finite variation process.*

ii) *For any approximating sequence* $(\theta^n)_{n\geq 1} \subset \mathcal{A}^s$, *we have that*

$$\liminf_{n\to\infty} C_t(\theta^n) \geq \int_0^t S d\theta + \int_0^t \kappa d\|\theta\| \tag{2.3}$$

where the right-hand side is defined in the usual Stieltjes sense.

iii) *There exists an approximating sequence* $(\theta^n)_{n\geq 1}$ *such that in* (2.3) *the limit exists and equality holds.*

The previous proposition leads to the following definition:

Definition 2.10. For $\theta \in \mathcal{A}$, we set

$$C_t(\theta) = \int_{[0,t]} S d\theta + \int_{[0,t]} \kappa d\|\theta\|$$

and similarly $V(\theta) = C_T(\theta)$. We denote by $\mathcal{C}_x = \{V(\theta) : \theta \in \mathcal{A}_x\} - L^0_+$ and $\mathcal{C} = \cup_{x>0}\mathcal{C}_x$ the set of claims dominated by general admissible strategies.

It is immediately seen that Definitions 2.10 and 2.2 coincide for simple strategies. We now turn to the counterparts of martingale measures. Adapting to our setting the definitions of Schachermayer [15] and Campi & Schachermayer [2], we refer to these objects as Consistent Price Systems:

Definition 2.11. Let (S,κ) satisfy Assumption 2.1.

i) A **Consistent Price System** is a pair (M,Q) of a probability Q equivalent to P and a Q-local martingale M lying a.s. within the bid-ask spread, i.e.,

$$|S_t - M_t| \leq \kappa_t \quad \text{a.s. for all } t \in [0,T].$$

If the above inequality is strict, we have a **Strictly Consistent Price System**.

ii) A **Uniformly Strictly Consistent Price System** is a pair (M,Q) as in i) such that M is a.s. strictly contained within the bid-ask spread, pathwise uniformly in $[0,T]$:

$$\inf_{t\in[0,T]} (\kappa_t - |S_t - M_t|) > 0 \quad \text{a.s.} \tag{2.4}$$

iii) $\mathcal{M} \supset \mathcal{M}_s \supset \mathcal{M}_u$ denote respectively the sets of Consistent, Strictly Consistent and Uniformly Strictly Consistent Price Systems.

iv) (S,κ) satisfies the condition **(CPS)**, **(SCPS)**, **(USCPS)** if $\mathcal{M} \neq \emptyset$, $\mathcal{M}_s \neq \emptyset$, $\mathcal{M}_u \neq \emptyset$, respectively.

The main no-arbitrage theorem can then be formulated as:

Theorem 2.12. *Let (S,κ) satisfy Assumption 2.1. Then* (RNFL) $\iff$ (USCPS).

The corresponding hedging theorem becomes:

Theorem 2.13. *Let (S,κ) satisfy Assumption 2.1 and* (RNFL)*. Then for any $X \in L^\infty$ we have that*

$$\inf\{x : x + V(\theta) \geq X \ \textit{a.s. for } \theta \in \mathcal{A}\} = \sup\{E_Q[X] : (M,Q) \in \mathcal{M}_u\}.$$

3. Proofs

We begin with the proof of Proposition 2.4. Here the idea is similar to the frictionless case: if the liquidation value is bounded from below, then this bound cannot be breached before the horizon, otherwise an arbitrage arises. In the presence of transaction costs, this argument needs some refinements, since liquidation values are no longer additive, but only superadditive, with respect to the concatenation of strategies. The key property is that, in absence of trivial arbitrage, one cannot gain "admissibility" by trading. In other words, if a strategy which requires n transactions is x-admissible, then it remains so after removing the last transaction. In the next proof, this is achieved by either early liquidation, or by skipping the last transaction before liquidation.

Lemma 3.1. *Let* $\theta = \sum_{i=1}^{n-1} \theta_{\tau_i} 1_{]\tau_i,\tau_{i+1}]}$ *be a simple strategy such that* $x + V(\theta) \geq 0$ *a.s., and define the "truncated" strategy* $\hat{\theta} = \theta 1_{[0,\tau_{n-2}]} + \theta^{n-2} 1_{]\tau_{n-2},\sigma]}$*, where*

$$\sigma = \begin{cases} \tau_{n-1} & \text{if } x + V(\theta 1_{[0,\tau_{n-1}]}) \geq 0 \\ \tau_n & \text{otherwise.} \end{cases}$$

If (NTA) holds, then $x + V(\hat{\theta}) \geq 0$ *a.s.*

The previous lemma allows to prove Proposition 2.4 by induction on the number of transactions:

Proof of Proposition 2.4. If $\theta = \sum_{i=1}^{n-1} \theta^i 1_{]\tau_i,\tau_{i+1}]}$, we argue by induction on n. The case $n \leq 2$ is trivially satisfied.

We suppose the thesis is true for $n-1$, and we prove it for n. For each time $t \in [0,T]$ we have to find a liquidation time. On the set $\{\tau_{n-1} < t\}$ we simply choose τ_n. On the other hand, θ coincides on the set $\{t \leq \tau_{n-1}\}$ with the strategy $\hat{\theta}$ obtained from Lemma 3.1, which also satisfies $x + V(\hat{\theta}) \geq 0$. Since $\hat{\theta}$ has $n-1$ transactions, by the inductive assumption it admits a liquidation time, which is also valid for θ. □

We now turn to the proof of Lemma 3.1, which requires an auxiliary lemma.

Lemma 3.2. a, b, c, X, Y *be random variables, such that* $b \cdot c > 0$*, and the following conditions hold:*

$$a + bX + c(X - Y) \geq 0 \quad a.s. \tag{3.1}$$

$$P((b+c)(X-Y) < 0 | Y) > 0 \quad a.s. \text{ or } X - Y = 0 \quad a.s., \tag{3.2}$$

then also $a + bX \geq 0$ *a.s.*

Proof. (3.1) implies that

$$(b+c)(X-Y) \geq -(a+bY) \quad \text{a.s.}$$

and therefore

$$a + bY \geq 0 \quad \text{a.s.}, \tag{3.3}$$

otherwise (3.2) is violated. The thesis follows substituting (3.3) in (3.1). □

Proof of Lemma 3.1. By definition of σ and $\hat{\theta}$, we only need to check that $x + V(\hat{\theta}) \geq 0$ on $\{\sigma = \tau_n\}$.

On the set $\{\theta^{n-2}\theta^{n-1} \leq 0\}$ we have that $V(\theta) = V(\theta 1_{[0,\tau_{n-1}]}) + V(\theta 1_{]\tau_{n-1},\tau_n]})$. It follows that $x + V(\theta 1_{[0,\tau_{n-1}]}) \geq 0$ a.s., otherwise $\theta 1_{]\tau_{n-1},\tau_n] \cap \{x + V(\theta 1_{[0,\tau_{n-1}]}) < 0\}}$ is an arbitrage. Hence $\sigma = \tau_{n-1}$ on $\{\theta^{n-2}\theta^{n-1} \leq 0\}$, and on $\{\sigma = \tau_n\}$ we have that $\theta^{n-2}\theta^{n-1} > 0$ and $x + V(\theta 1_{[0,\tau_{n-1}]}) < 0$.

On the set $\{\sigma = \tau_n, |\theta_{\tau_{n-1}}| < |\theta_{\tau_{n-2}}|\}$ we have that $\theta = \alpha\hat{\theta} + (1-\alpha)\theta 1_{[0,\tau_{n-1}]}$ where $\alpha = \theta_{\tau_{n-1}}/\theta_{\tau_{n-2}} \in [0,1]$. Also, $V(\theta) = \alpha V(\hat{\theta}) + (1-\alpha)V(\theta 1_{[0,\tau_{n-1}]})$. It follows that $x + V(\hat{\theta}) \geq 0$ a.s., otherwise the assumption $x + V(\theta) \geq 0$ is violated.

It remains to check that $x + V(\hat{\theta}) \geq 0$ a.s. on $\{\sigma = \tau_n, |\theta^{n-1}| \geq |\theta^{n-2}|\}$. On $\{\theta^{n-2}, \theta^{n-1} > 0\}$ (resp. $\{\theta^{n-2}, \theta^{n-1} < 0\}$), this follows from Lemma 3.2 setting $a = x - C_{\tau_{n-1}}(\theta)$, $b = \theta^{n-2}$, $c = \theta^{n-1} - \theta^{n-2}$, $X = (S-\kappa)_{\tau_n}$ (resp. $X = (S+\kappa)_{\tau_n}$) and $Y = (S+\kappa)_{\tau_{n-1}}$ (resp. $Y = (S-\kappa)_{\tau_{n-1}}$). □

We now prove the intuitively obvious *domination property*, whereby an agent executing a given strategy at better (bid and ask) prices achieves a better payoff than another agent facing worse prices. Indeed, this property is so basic that Lemma 3.3 is formulated pathwise.

Lemma 3.3. *Let* (S, κ) *and* $(\tilde{S}, \tilde{\kappa})$ *be continuous functions such that*

$$\kappa_t - \tilde{\kappa}_t - |S_t - \tilde{S}_t| \geq 0 \quad \text{for all } t \in [0, T].$$

Then for any finite variation function θ *and* $t \in [0, T]$ *we have*

$$C_t^{(S,\kappa)}(\theta) \geq C_t^{(\tilde{S},\tilde{\kappa})}(\theta) + \int_{[0,t]} \left(\kappa - \tilde{\kappa} - |S - \tilde{S}|\right) d\|\theta\|$$

and therefore

$$V^{(S,\kappa)}(\theta) \leq V^{(\tilde{S},\tilde{\kappa})}(\theta) - \int_0^T \left(\kappa - \tilde{\kappa} - |S - \tilde{S}|\right) d\|\theta\|. \tag{3.4}$$

Proof. We have that

$$\begin{aligned} C_t^{(S,\kappa)}(\theta) &= C_t^{(\tilde{S},\tilde{\kappa})}(\theta) - \int_{[0,t]} (S - \tilde{S}) d\theta - \int_{[0,t]} (\kappa - \tilde{\kappa}) d\|\theta\| \\ &= C_t^{(\tilde{S},\tilde{\kappa})}(\theta) - \int_{[0,t]} \left(\kappa - \tilde{\kappa} - (S - \tilde{S}) \frac{d\theta}{d\|\theta\|}\right) d\|\theta\| \\ &\leq C_t^{(\tilde{S},\tilde{\kappa})}(\theta) - \int_{[0,t]} \left(\kappa - \tilde{\kappa} - |S - \tilde{S}|\right) d\|\theta\| \end{aligned}$$

and the proof is complete. □

As an immediate consequence, we obtain the boundedness in L^0 of the set of total variations of simple strategies.

Lemma 3.4. *Let* (S, κ) *satisfy* (RNFL). *Then* $\{\|\theta\|_T : \theta \in \mathcal{A}_x^s\}$ *is bounded in* L^0 *for all* $x > 0$.

Proof. Rearranging (3.4), for any $\theta \in \mathcal{A}_x$ we have that

$$\|\theta\|_T \inf_{t \in [0,T]} (\kappa_t - \tilde{\kappa}_t - |S_t - \tilde{S}_t|) \leq \int_0^T \left(\kappa - \tilde{\kappa} - |S - \tilde{S}|\right) d\|\theta\| \leq x + V^{(\tilde{S},\tilde{\kappa})}(\theta).$$

But (RNFL) implies that the set $\{V^{(\tilde{S},\tilde{\kappa})}(\theta) : \theta \in \mathcal{A}_x\}$ is bounded in L^0, whence the thesis. □

The next lemma is formulated in a pathwise sense:

Lemma 3.5. *Let $(\theta^n)_{n\geq 1}$ and θ be predictable finite variation functions such that $\theta^n_t \to \theta_t$ for all t in a dense set of $[0,T]$. Then $d\theta^n$ converges weakly to $d\theta$, and hence $d\|\theta^n\|$ weakly converges to some $d\eta \geq d\|\theta\|$.*

Proof. For all $\phi \in C^1_c([0,T])$ dominated convergence implies that

$$\lim_{n\to\infty} \int_{[0,t]} \theta^n d\phi = \int_{[0,t]} \theta d\phi.$$

Since $\int_{[0,t]} \phi d\theta \leq \|\theta\|_t \sup_{s\in[0,t]} |\phi_s|$, the map $\phi \mapsto \int_{[0,t]} \phi d\theta$ is continuous in the uniform norm, and the previous equality extends to all $\phi \in C([0,T])$ vanishing at infinity, and weak convergence follows. □

Proof of Proposition 2.9. θ is predictable by definition, and by Lemma 3.4 the set $(\|\theta^n\|_T)_{n\geq 1}$ is bounded in L^0, hence i) follows from Lemma 3.5. ii) also follows from Lemma 3.5, while iii) is obtained from the staircase approximation:

$$\theta^n = \sum_{k=0}^{nT-1} \theta_{k/n} 1_{]k/n,(k+1)/n]}. \qquad \square$$

We prove the no-arbitrage theorem first, and the hedging theorem then follows naturally. We begin with the easy implication.

Lemma 3.6. (USCPS) $\Rightarrow$ (RNFL).

Proof. We check that (S,κ) satisfies (RNFL) with $\tilde{S} = M$ and $\tilde{\kappa} = 0$. In fact, since M is a local martingale under Q, by Lemma 3.3 it follows that $E_Q[V(\theta)] \leq 0$ and hence for its $\sigma(L^\infty, L^1)$-closure. We also have that

$$\kappa_t - \tilde{\kappa}_t - |S_t - \tilde{S}_t| = \kappa_t - |S_t - M_t|$$

which implies (2.2), and the proof is complete. □

The following lemma was first proved by Jouini & Kallal [7], then further investigated by Choulli & Stricker [4] and Cherny [3].

Lemma 3.7. *Let $\mathcal{T} \subset [0,T]$, $(X_t)_{t\in\mathcal{T}}$ a submartingale and $(Y_t)_{t\in\mathcal{T}}$ a supermartingale, such that $X_t \leq Y_t$ a.s for all $t \in \mathcal{T}$. Then there exists a martingale $(M_t)_{t\in\mathcal{T}}$ such that $X_t \leq M_t \leq Y_t$ a.s. for all $t \in \mathcal{T}$.*

Proof. In finite discrete time $\mathcal{T} = \{t_0, \ldots, t_n\}$, we simply set $M_{t_0} = Y_{t_0}$ and recursively define $M_{t_{n+1}} = \alpha_n X_{t_{n+1}} + (1-\alpha_n) Y_{t_{n+1}}$, where α_n satisfies

$$M_{t_n} = \alpha_n E\left[X_{t_{n+1}} \middle| \mathcal{F}_{t_n} \right] + (1-\alpha_n) E\left[Y_{t_{n+1}} \middle| \mathcal{F}_{t_n} \right].$$

Let $\mathcal{T} = [0,T]$. From the discrete case, for each dyadic partition $D_n = \{kT/2^n : 0 \leq k \leq 2^n\}$ we obtain a martingale M^n with respect to the discrete filtration $(\mathcal{F}_t)_{t\in D_n}$, and such that

$$X_t \leq M^n_t \leq Y_t \qquad \text{for all } t \in D_n. \tag{3.5}$$

In particular, $X_T \leq M^n_T \leq Y_T$ for all $n \geq 1$, therefore $(M^n_T)_{n\geq 1}$ is bounded in L^2 under an equivalent probability measure (e.g., $\frac{dQ}{dP} = e^{-(|X_T|+|Y_T|)}/E\left[e^{|X_T|+|Y_T|}\right]$), thus it converges up to a sequence of convex combinations to some random variable M. Then we define the martingale $M_t = E[M_T | \mathcal{F}_t]$, and letting $n \to \infty$ in (3.5) we see that $X_t \le M_t \le Y_t$ a.s. for all $t \in [0, T]$. $\square$

The following is a generalization of Theorem 4.5 in Cherny [3].

Lemma 3.8. *Let $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ be two càdlàg processes. The following conditions are equivalent:*

i) *There exists a martingale $(M_t)_{t\in\mathcal{T}}$ such that*

$$X_t \le M_t \le Y_t \qquad \text{a.s. for all } t \in [0, T]. \tag{3.6}$$

ii) *For all stopping times σ, τ such that $0 \le \sigma \le \tau \le T$ a.s., we have*

$$E[X_\tau | \mathcal{F}_\sigma] \le Y_\sigma \qquad \text{and} \qquad E[Y_\tau | \mathcal{F}_\sigma] \ge X_\sigma. \tag{3.7}$$

Proof. i)⇒ii): from (3.6) and optional sampling, we have that

$$E[X_\tau | \mathcal{F}_\sigma] \le E[M_\tau | \mathcal{F}_\sigma] = M_\sigma \le Y_\sigma,$$

and the second equation in (3.7) follows similarly.

ii)⇒i): Denoting by $\mathcal{O}_t$ the set of stopping times with values in the interval $[t, T]$, we define the auxiliary processes

$$X'_t = \operatorname{ess\,sup}_{\tau\in\mathcal{O}_t} E[X_\tau | \mathcal{F}_t] \qquad \text{and} \qquad Y'_t = \operatorname{ess\,inf}_{\tau\in\mathcal{O}_t} E[Y_\tau | \mathcal{F}_t].$$

Since for $\sigma, \tau \in \mathcal{O}_t$ and $A \in \mathcal{F}_t$ we have that $\sigma 1_A + \tau 1_{\Omega\setminus A} \in \mathcal{O}_t$, we obtain that

$$\operatorname{ess\,sup}_{\tau\in\mathcal{O}_t} E[X_\tau | \mathcal{F}_t] = E\left[\operatorname{ess\,sup}_{\tau\in\mathcal{O}_t} X_\tau \middle| \mathcal{F}_t\right],$$

and therefore X' is a supermartingale. Likewise, Y' is a submartingale, and they both admit càdlàg versions. By ii), for $\sigma, \tau \in \mathcal{O}_t$ we have

$$\begin{aligned} E[X_\tau | \mathcal{F}_t] - E[Y_\sigma | \mathcal{F}_t] &= E[E[X_\tau - Y_\sigma | \mathcal{F}_{\tau\wedge\sigma}] | \mathcal{F}_t] \\ &= E\left[(X_\tau - E[Y_\sigma | \mathcal{F}_\tau])1_{\{\tau\le\sigma\}} + (E[X_\tau | \mathcal{F}_\sigma] - Y_\sigma)1_{\{\sigma<\tau\}} \middle| \mathcal{F}_t\right] \le 0 \quad \text{a.s.} \end{aligned}$$

and hence $X'_t \le Y'_t$ a.s. for all $t \in [0, T]$, and Lemma 3.7 concludes the proof. $\square$

The previous two lemmas are needed for the following result, where the trick is to consider the payoffs with respect to $(\tilde{S}, \tilde{\kappa})$ of strategies admissible for (S, κ).

Lemma 3.9. *If Q is equivalent to P and $E_Q\left[V^{(\tilde{S},\tilde{\kappa})}(\theta)\right] \le 0$ for all $\theta = \alpha 1_{]\sigma,\tau]}$ admissible for (S, κ), then there exists a Q-local martingale M such that $(M, Q) \in \mathcal{M}_u$.*

Proof. As in the frictionless case, we reduce by localization to the case of S bounded. Consider $\theta = \alpha 1_{]\sigma,\tau]}$ admissible for (S, κ), where σ, τ are stopping times. For $\alpha = \pm 1_A$, where $A \in \mathcal{F}_\sigma$, we obtain

$$E_Q\left[\left((\tilde{S} - \tilde{\kappa})_\tau - (\tilde{S} + \tilde{\kappa})_\sigma\right) 1_A\right] \le 0, \quad E_Q\left[\left((\tilde{S} + \tilde{\kappa})_\tau - (\tilde{S} - \tilde{\kappa})_\sigma\right) 1_A\right] \le 0.$$

Since these equations hold for any $A \in \mathcal{F}_\sigma$, it follows that

$$E_Q\left[(\tilde{S} + \tilde{\kappa})_\tau \middle| \mathcal{F}_\sigma\right] \le (\tilde{S} - \tilde{\kappa})_\sigma, \qquad E_Q\left[(\tilde{S} - \tilde{\kappa})_\tau \middle| \mathcal{F}_\sigma\right] \le (\tilde{S} + \tilde{\kappa})_\sigma,$$

and Lemma 3.8 implies the existence of a Q-local martingale M such that

$$|\tilde{S}_t - M_t| \leq \tilde{\kappa}_t \qquad \text{a.s. for all } t \in [0,T],$$

and therefore, by (RNFL) we have

$$\inf_{t\in[0,T]}(\kappa_t - |S_t - M_t|) \geq \inf_{t\in[0,T]}(\kappa_t - \tilde{\kappa}_t - |\tilde{S}_t - S_t|) > 0,$$

and (2.4) follows. □

We finally put all pieces together. We recall the classical Kreps-Yan separation theorem. Recent proofs can be found in Schachermayer [14, Theorem 3.1] and Kabanov & Stricker [12, Lemma 3].

Theorem 3.10 (Kreps [13]**, Yan** [16]**).** *Let $-L^\infty_+ \subset C \subset L^\infty$ be a convex cone, closed in the $\sigma(L^1, L^\infty)$ topology, such that $C \cap L^\infty_+ = \{0\}$. Then there exists a probability Q equivalent to P such that $\frac{dQ}{dP} \in L^\infty$ and $E_Q[C] \leq 0$.*

Proof of Theorem 2.12. By the (RNFL) condition we can apply the Kreps-Yan Theorem (3.10), so that there exists a probability Q, equivalent to P such that $E_Q[C] \leq 0$, and Lemma 3.9 concludes the proof. □

Now we prove the hedging theorem by the usual separation arguments, as in Delbaen & Schachermayer [5, Theorem 5.7]:

Proof of Theorem 2.13. Observe that $x + V^{(S,\kappa)}(\theta) \geq X$ for some $\theta \in \mathcal{A}$ if and only if $X - x \in \mathcal{C}$. Hence it is sufficient to prove that

$$\begin{aligned} X - x \in \mathcal{C} &\quad \Rightarrow \quad E_Q[X] \leq x \quad \text{for all } Q \in \mathcal{M}_u \\ X - x \notin \mathcal{C} &\quad \Rightarrow \quad E_Q[X] > x \quad \text{for some } Q \in \mathcal{M}_u. \end{aligned} \tag{3.8}$$

Let $x + V^{(S,\kappa)}(\theta) \geq X$ for some $\theta \in \mathcal{A}$, and $(M,Q) \in \mathcal{M}_u$. Since by assumption $V^{(S,\kappa)}(\theta)$ can be approximated by a sequence $V^{(S,\kappa)}(\theta^n)$ random variables bounded from below, by Lemma 3.3 we have

$$E_Q\left[x + V^{(S,\kappa)}(\theta)\right] \leq E_Q\left[x + V^{(M,0)}(\theta)\right] \leq x.$$

To see (3.8), suppose that $X - x \notin \mathcal{C}$. Since C is $\sigma(L^\infty, L^1)$-closed, the Hahn-Banach theorem yields a continuous linear functional l such that $\sup_{\xi\in C} l(\xi) < l(X - x)$. Since $-L^\infty_+ \subset C$, $l(L^\infty_+) \geq 0$, and $\sup_{\xi\in C} l(\xi) \leq 0$ because C is a cone. Normalizing by $l(1)$, we can then identify l with a probability measure Q', absolutely continuous with respect to P, and since $X - x \notin \mathcal{C}$, it follows that $E_{Q'}[X - x] > 0$.

By Theorem 2.12, there exists Q, equivalent to P, such that $\sup_{\xi\in C} E[\xi] \leq 0$. If we define $Q_\varepsilon = \varepsilon Q + (1-\varepsilon)Q'$, we obtain that Q_ε is equivalent to P for any $\varepsilon \in (0,1)$, $\sup_{\xi\in C} E_{Q_\varepsilon}[\xi] \leq 0$ and for small ε, $E_{Q_\varepsilon}[X - x] > 0$ by continuity. Then Lemma 3.9 implies that $(M, Q_\varepsilon) \in \mathcal{M}_u$ for some Q_ε-local martingale M, and the proof is complete. □

References

[1] L. Campi and W. Schachermayer, private communication, 2005.

[2] L. Campi and W. Schachermayer, *A super-replication theorem in Kabanov's model of transaction costs*, Finance Stoch., **10 (4)** (2006), 579–596.

[3] A. Cherny, *General arbitrage pricing model: probability and possibility approaches*, preprint, 2005.

[4] T. Choulli and C. Stricker, *Séparation d'une sur- et d'une sousmartingale par une martingale*, in: Séminaire de Probabilités, XXXII, Lecture Notes in Math., Springer, **1686** (1998), 67–72.

[5] F. Delbaen and W. Schachermayer, *A general version of the fundamental theorem of asset pricing*, Math. Ann., **300 (3)** (1994), 463–520.

[6] J. M. Harrison and S. R. Pliska, *Martingales and stochastic integrals in the theory of continuous trading*, Stochastic Process. Appl., **11 (3)** (1981), 215–260.

[7] E. Jouini and H. Kallal, *Martingales and arbitrage in securities markets with transaction costs*, J. Econom. Theory, **66 (1)** (1995), 178–197.

[8] Y. Kabanov, *Hedging and liquidation under transaction costs in currency markets*, Finance Stoch., **3 (2)** (1999), 237–248.

[9] Y. M. Kabanov and C. Stricker, *The Harrison-Pliska arbitrage pricing theorem under transaction costs*, J. Math. Econom., **35 (2)** (2001), 185–196. Arbitrage and control problems in finance.

[10] Y. Kabanov, M. Rásonyi, and C. Stricker, *No-arbitrage criteria for financial markets with efficient friction*, Finance Stoch., **6 (3)** (2002), 371–382.

[11] Y. Kabanov, M. Rásonyi, and C. Stricker, *On the closedness of sums of convex cones in L^0 and the robust no-arbitrage property*, Finance Stoch., **7 (3)** (2003), 403–411.

[12] Y. Kabanov and C. Stricker, *A teachers' note on no-arbitrage criteria*, in: Séminaire de Probabilités, XXXV, Lecture Notes in Math., Springer, **1755** (2001), 149–152.

[13] D. M. Kreps, *Arbitrage and equilibrium in economies with infinitely many commodities*, J. Math. Econom., **8 (1)** (1981), 15–35.

[14] W. Schachermayer, *A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time*, Insurance Math. Econom., **11 (4)** (1992), 249–257.

[15] W. Schachermayer, *The fundamental theorem of asset pricing under proportional transaction costs in finite discrete time*, Math. Finance, **14 (1)** (2004), 19–48.

[16] J. A. Yan, *Caractérisation d'une classe d'ensembles convexes de L^1 ou H^1*, in: Seminar on Probability, XIV (Paris, 1978/1979), Lecture Notes in Math., Springer, **784** (1980), 220–222.

Paolo Guasoni
Boston University
Dept. of Mathematics and Statistics
111 Cummington st.
Boston MA 02215
USA
e-mail: guasoni@bu.edu

Progress in Probability, Vol. 59, 469–492

Robustness of the Hobson–Rogers Model with Respect to the Offset Function

Vera Blaka Hallulli and Tiziano Vargiolu

Abstract. In this paper we analyse the robustness of the Hobson–Rogers model with respect to the offset function, which depends on the whole past of the risky asset and is thus not fully observable. We prove that, if the offset function is the realisation of a stationary process, then the error in pricing a derivative asset decreases exponentially with respect to the observation window. We present sufficient conditions on the volatility in order to characterise the invariant density and three examples.

Mathematics Subject Classification (2000). Primary 91B28; Secondary 60G10.

Keywords. Hobson–Rogers model, robustness, invariant measures.

1. Introduction

The year 1973 is a milestone in the modeling of financial markets: in fact, in that year the papers of Black and Scholes [2] and Merton [15], where an explicit formula for the price of call and put options was present, saw the light. The formula now known universally as "the Black and Scholes formula" links the price of a call option to quantities which are observed in the market (current price, strike price, time to maturity) and a parameter, the volatility, which gives an idea of how rapidly the asset prices can change. The two papers cited above influenced financial markets so deeply that every investment bank today has to deal with "the Black and Scholes approach": this is also witnessed by the Nobel prize in 1997.

The so-called "Black and Scholes model" is however valid only as a first approximation: in fact, it was soon realised that the assumption of a constant volatility was in contrast with the empirical observations of derivative prices in real markets, which suggest that the volatility is not constant, but rather depends both on time to maturity and on the strike price.

Both authors gratefully acknowledge financial support from the CNR Strategic Project *Modellizzazione matematica di fenomeni economici.*

In the last years a growing interest has been raised for models where the asset prices' dynamics do not depend only on their current values, but also on past values: these models can be usually seen as generalisations of the so-called level-dependent volatility models, where the volatility is usually a function of time and current price level, and the market is complete. By making the volatility depend also on the past prices of the risky assets, it is also possible to enrich the model by reproducing correlations and dependencies which are observed in practice. Among these models, the one proposed by Hobson–Rogers [12] is the only case (to the authors' knowledge) where the model is equivalent to a 2-dimensional Markov model, thus the problem of pricing and hedging a derivative asset is led to the solution of a linear PDE. In particular, one component of this Markov process represents the price and the other one represents the so-called *offset function of order* 1, which is an integral depending on all the past history of the asset price, and is thus not fully observable.

There are two ways of using the Hobson–Rogers model in practice. One is to consider a finite horizon approximation, where the offset function is defined only on a finite observation interval of the past price. Unfortunately, the authors proved in a previous paper [10] that it is impossible to obtain a Markov system in this way. The other way is to use the pricing PDE with a misspecified initial offset function, thus making a mistake both on the path of the process as on the calculation of the price of the derivative assets. This approach is studied in detail in this paper. One can then search for the initial offset value which minimises this error. We find out that, for all the contingent claims which are Lipschitz continuous functions of the log-price of the asset, this error is proportional to the variance of the offset function at time 0. By assuming that we can observe the past prices of the risky asset on an interval of length R, this variance decreases exponentially with respect to R, and is proportional to the variance of the offset function at time $-R$. If we also assume that the offset function is a stationary process, we can calculate this variance, which does not depend on R: in this way, if one wants an error less than a given ε in pricing a derivative asset, one only has to observe the past price for a sufficient time R.

The paper is organised as follows. In Section 2 we present the Hobson–Rogers model. In Section 3 we make a survey, based on [10], on the reasons why a version of the Hobson–Rogers model with finite observation horizon loses Markovianity. In Section 4 we study the robustness of the Hobson–Rogers model with respect to the misspecification of the offset function, and in Section 5 we provide an estimate of the minimum observation horizon required for having an error less than a given threshold. In Section 6 we provide a way to calculate the variance of the offset function at the beginning of the observation window in terms of the invariant measure of the offset function, and provide sufficient conditions on the volatility in order to have a characterisation of the invariant density. Section 7 presents three examples.

2. The Hobson–Rogers model

We define the discounted log-price process $Z(t)$ at time t as

$$Z(t) = \log(S(t)e^{-rt})$$

where r is the (constant) risk-free interest rate, and the *offset function* of order m, denoted by $P^{(m)}(t)$, by

$$P^{(m)}(t) = \int_0^\infty \lambda e^{-\lambda u}(Z(t) - Z(t-u))^m du \quad \text{for} \quad m = 0, \dots, n, \tag{2.1}$$

the constant λ being a parameter of the model which describes the rate at which past information is discounted. Then, for some value n, we assume the following.

Assumption 2.1. $Z(t)$ satisfies the SDE

$$dZ(t) = -\frac{1}{2}\sigma^2(P^{(1)}(t), \dots, P^{(n)}(t))dt + \sigma(P^{(1)}(t), \dots, P^{(n)}(t))\; dW(t)$$

where $\sigma(\cdot)$ and $\sigma^2(\cdot)$ are globally Liptschitz, $\sigma(\cdot)$ is strictly positive and $(W_t)_{t\in\mathbb{R}}$ is a so-called two-sided Brownian motion [3] under a probability measure $\mathbb{P}$, which is chosen such that $(S(t)e^{-rt})_t$ is a $\mathbb{P}$-martingale.

This probability measure $\mathbb{P}$ is in fact known as *risk-neutral* probability or *martingale measure*), and the existence of such a $\mathbb{P}$ is equivalent to the non-existence of arbitrage opportunities in the market (see [10, 12] and the references therein for details).

This model can be seen as a "good" model because no new Brownian motions (or other source of uncertainty) have been introduced in the specification of the price process. This means that the market is complete and any contingent claim is hedgeable (see [10] for details). On the other hand, it is possible to allow $\sigma(\cdot)$ to be a function of the price level $S(t)$ also. So, this model can be extended to include the class of level-dependent volatility processes as a special case. The reason for the definition of the processes $P^{(m)}(t)$, $m = 0, \dots, n$, is seen in the following lemma.

Lemma 2.2. *$(Z, P^{(1)}, \dots, P^{(n)})$ is an $(n+1)$-dimensional Markov process, and the offset processes $P^{(m)}(t)$ satisfy the coupled SDEs*

$$\begin{aligned} dP^{(m)}(t) &= mP^{(m-1)}(t)\; dZ(t) + \frac{m(m-1)}{2}P^{(m-2)}(t)\; d\langle Z\rangle(t) - \lambda P^{(m)}(t)\; dt, \\ &\quad m > 1 \\ dP^{(1)}(t) &= dZ(t) - \lambda P^{(1)}(t)\; dt. \end{aligned} \tag{2.2}$$

Proof. See [12]. □

Being $(Z, P^{(1)}, \dots, P^{(n)})$ an $(n+1)$-dimensional Markov process, we can easily employ the Kolmogorov equation when pricing a contingent claim with final payoff $h(S(T))$. In fact, (for sake of simplicity consider from now on the case $n = 1$ and denote $P(t) \equiv P^{(1)}(t)$) its price $V(t) = \mathbb{E}[h(S(T))|\mathcal{F}_t]$ is of the form $V(t) = F(t, S(t), P(t))$, where F is the solution of the Kolmogorov equation

$$F_t + rsF_s - \lambda p F_p + \left(\frac{1}{2}s^2 F_{ss} + sF_{ps} + \frac{1}{2}F_{pp} - \frac{1}{2}F_p\right)\sigma^2(p) = rF \tag{2.3}$$

subject to the boundary condition

$$F(p, s, T) = h(s).$$

Besides, the solution of the hedging problem is a closed formula: it is enough to use the Itô formula on F and to make some calculations to obtain that the hedging strategy at time t is given by

$$\Delta(t) = F_s(t, S(t), P(t)) + \frac{F_p(t, S(t), P(t))}{S(t)}.$$

In conclusion this model allows us to construct a process for the price, but we can see that some difficulties arise. In fact, for the computation of $P(0)$ (or in general $P(t)$), we need to know the path of S on all its past $(-\infty, 0)$ (or $(-\infty, t)$). This requirement is unusual in the modelisation of financial markets, where one usually meets models that start from a certain moment in time (usually 0). In fact, the requirement of an infinite horizon in the past raises mathematical and "practical" (or better economical) complications. From the mathematical side, we would have to define a stochastic calculus with time ranging on all the real line. Once that this is done, we would have to establish that P is well defined: in fact, remember that P is the integral of a process on $(-\infty, 0)$, so one must also prove that this integral is well defined. From the economical side, assets that "existed forever" do not exist in the real market. Thus, one has to establish what can be used instead of the price path of S when the asset still did not exist.

While these problems seem less worrying than stated, mainly due to the exponential weight in (2.1), still theoretical (and practical) solutions to these issues are not present in literature, at least to the authors' knowledge. For this reason, we will explore two different approaches to avoid these problems.

The first one consists in specifying a model with finite horizon and to make the volatility depend on integrals of the price path. Unfortunately up to now all the models of this kind present in literature [1, 8] do not give a Markovian structure as the Hobson-Rogers model does, unless one uses from the beginning a level-dependent volatility model: in the next Section 3 we present a survey, based on [10], of these results.

The second one is the following. The problem of pricing a contingent claim with the Hobson-Rogers model is equivalent to solve the PDE (2.3), once the initial conditions $S(0) = s$, $P(0) = p$ are specified. While the price $S(0)$ is observed in the market, in order to calculate the true value $P(0)$ one would have to observe the asset in all its past. Since this is impossible, one has to use the model with a misspecification $\tilde{P}(0)$. Our aim will be then to search for the initial condition $\tilde{P}(0)$ which minimizes the error of pricing the contingent claim $h(S(T))$. This will be done from Section 4 on.

3. A finite delay model

Now we analyse a modification of the Hobson-Rogers model where we consider a finite time horizon and we make the risky asset's dynamics depend on integrals of the price path. Inspired by a model in [8], the model that we study is

$$dS(t) = S(t)\sigma(Y(t), Z(t))\ dW(t)$$

where the processes Y and Z are defined as

$$Y(t) = \int_0^\tau e^{-\lambda v} f(S(t-v))\ dv = \int_{t-\tau}^t e^{\lambda(u-t)} f(S(u))\ du, \qquad Z(t) = S(t-\tau)$$

where f is a strictly monotone function and τ is a given finite delay. Notice that for $f(x) = \log x$ and $\tau = +\infty$ one has that $\lambda Y(t) = \log S(t) - P^{(1)}(t)$, $P^{(1)}$ being the first offset function of the Hobson-Rogers model. Our scope is now to find a self-financing portfolio V which replicates the option with payoff $h(S(T))$ (or, more generally, $h(S(T), Y(T))$). Unlike in the Hobson-Rogers model, here the process (S, Y) is not Markov, and this is more due to the finite horizon nature of Y rather than to the specification of the volatility, more general than the Hobson-Rogers' one.

One can immediately think of using the state variables $(S(t), Y(t), Z(t))$, but this entails the usage of anticipative stochastic calculus. In fact, by making use of the Itô formula on a deterministic function of $(S(t), Y(t), S(t-\tau))$, we end up with stochastic differentials of the kind $G(t, S(t), Y(t), S(t-\tau))dS(t-\tau)$, where $G(t, S(t), Y(t), S(t-\tau))$ is not adapted to the filtration of the differential $dS(t-\tau)$ but "anticipates" (see [14] and the references therein). Conversely, we would have to prove that the portfolio dynamics could be written in the form $dV(t) = \Delta(t)\ dS(t)$, with Δ adapted to the filtration of S. In doing this, we will surely lose the Markovianity of the original Hobson-Rogers model.

One can be tempted to explore the following shortcut: though (S, Y) is in general not a Markov process, we make the strong assumption that for every final payoff of the form $h(S(T), Y(T))$ there exists a deterministic function F such that

$$V(t) = \mathbb{E}[h(S(T), Y(T))|\mathcal{F}_t] = F(t, S(t), Y(t)). \tag{3.1}$$

If this assumption is true, then the self-financing portfolio depends in a deterministic way only on the current values of S and Y. Unfortunately, the next result states that the assumption (3.1) is equivalent to σ not depending on y, z, that is, to S to be Markov; moreover, in this case, (3.1) is only true for h not depending on Y and the function F depending on t, s only.

Theorem 3.1. *If assumption* (3.1) *is true, then* $\sigma_z = \sigma_y = 0$.

The interested reader can find the proof in [10].

Remark 3.2. In this failed try, we were inspired by the positive results in [8]. We however have to say that in that paper the authors analyse a controlled system (which gives more degrees of freedom in reaching Markovianity), and also in that situation the authors succeed in reducing the system to the current values of S

and Y only when the dynamics of S is linear and with some restriction on the coefficients.

4. Robustness of the Hobson-Rogers model

As already announced, now we focus ourselves in establishing what happens if our Markov process (P, Z) starts from a misspecified initial condition $(\tilde{P}(0), Z(0))$ instead of the true initial condition $(P(0), Z(0))$.

From now on, denote with $\Sigma := (P, Z)$ the process with the correct (but not known) initial conditions and by $\tilde{\Sigma} = (\tilde{P}, \tilde{Z})$ the process starting from the misspecified initial conditions $(\tilde{P}(0), Z(0))$. Then the evolution of (both Σ and) $\tilde{\Sigma}$ is given by

$$\begin{cases} d\tilde{P}(t) &= -\left(\frac{1}{2}\sigma^2(\tilde{P}(t), \tilde{Z}(t)) + \lambda \tilde{P}(t)\right) dt + \sigma(\tilde{P}(t), \tilde{Z}(t))\, dW(t), \\ \tilde{P}(0) &\neq P(0), \\ d\tilde{Z}(t) &= -\frac{1}{2}\sigma^2(\tilde{P}(t), \tilde{Z}(t))\, dt + \sigma(\tilde{P}(t), \tilde{Z}(t))\, dW(t), \\ \tilde{Z}(0) &= Z(0), \end{cases}$$

the dynamics of Σ being driven by the same differential equation with the "right" initial conditions.

Now we present two estimates on the dependence of the process Σ (or $\tilde{\Sigma}$) on the initial condition: the first one is an L^2-estimate on $\sup_{0 \le u \le T} |\Sigma(u) - \tilde{\Sigma}(u)|$, and the second one is an L^2-estimate on $|\Sigma(T) - \tilde{\Sigma}(T)|$. Assume that the functions $\sigma(p, z)$ and $\sigma^2(p, z)$ are globally Lipschitz in (p, z) with respect to the Euclidean norm, in the sense that for $f = \sigma, \sigma^2$ there exists $K \ge 0$ (called *Lipschitz constant* of f) such that

$$|f(p, z) - f(\tilde{p}, \tilde{z})| \le K|(p, z) - (\tilde{p}, \tilde{z})| = K\sqrt{(p - \tilde{p})^2 + (z - \tilde{z})^2} \quad \forall (p, z), (\tilde{p}, \tilde{z}).$$

Theorem 4.1. *If σ, σ^2 are globally Lipschitz with Lipschitz constants, respectively, L, M, then for $t \in [0, T]$ we have*

$$\mathbb{E}\left[\sup_{0 \le u \le t} |\Sigma(u) - \tilde{\Sigma}(u)|^2\right] \le 3\mathbb{E}[|P(0) - \tilde{P}(0)|^2] e^{c(L,M,T)t}$$

where $c(L, M, T) = 2M^2 T + 6\lambda^2 T + 20L^2$, and

$$\mathbb{E}[|\Sigma(t) - \tilde{\Sigma}(t)|^2] \le 3\mathbb{E}[|P(0) - \tilde{P}(0)|^2] e^{C(L,M,T)t}$$

where $C(L, M, T) = 2M^2 T + 6\lambda^2 T + 5L^2$.

Results of this kind are classical in the theory of SDEs: we present the proof in order to show that the constants $C(L, M, T)$ and $c(L, M, T)$ are the best possible for our equations.

Proof. We have that

$$\sup_{0 \le u \le t} |\Sigma(u) - \tilde{\Sigma}(u)|^2 \le \sup_{0 \le u \le t} |Z(u) - \tilde{Z}(u)|^2 + \sup_{0 \le u \le t} |P(u) - \tilde{P}(u)|^2$$

which yields

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le u\le t}|\Sigma(u)-\tilde{\Sigma}(u)|^2\right] \\
&\le \mathbb{E}\left[\sup_{0\le u\le t}|Z(u)-\tilde{Z}(u)|^2\right]+\mathbb{E}\left[\sup_{0\le u\le t}|P(u)-\tilde{P}(u)|^2\right]=(1)+(2).
\end{aligned}$$

For the first term on the right-hand side, applying Doob's inequality and the Lipschitz property of σ and σ^2, we have

$$\begin{aligned}
(1) &= \mathbb{E}\left[\sup_{0\le u\le t}\left|Z(0)-\tilde{Z}(0))+\int_0^u \frac{1}{2}(\sigma^2-\tilde{\sigma}^2)\ ds+\int_0^u(\sigma-\tilde{\sigma})\ dW(s)\right|^2\right] \\
&\le 2\mathbb{E}\left[\sup_{0\le u\le t}\left|\int_0^t \frac{1}{2}(\sigma^2-\tilde{\sigma}^2)ds\right|^2\right]+2\mathbb{E}\left[\sup_{0\le u\le t}\left|\int_0^u(\sigma-\tilde{\sigma})dW(s)\right|^2\right] \\
&\le \frac{1}{2}T\int_0^t \mathbb{E}|\sigma^2-\tilde{\sigma}^2|^2 ds+8\int_0^t \mathbb{E}|\sigma-\tilde{\sigma}|^2 ds \\
&\le \left(\frac{1}{2}M^2T+8L^2\right)\int_0^t \mathbb{E}\left[|Z-\tilde{Z}|^2+|P-\tilde{P}|^2\right] ds
\end{aligned}$$

where σ, $\tilde{\sigma}$ is a shorthand notation for $\sigma(P(s),Z(s))$, $\tilde{\sigma}(s)=\sigma(\tilde{P}(s),\tilde{Z}(s))$. For the second term we have

$$\begin{aligned}
(2) &\le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2+3T\int_0^t \mathbb{E}\left[\left|\frac{1}{2}(\sigma^2-\tilde{\sigma}^2)+\lambda(P-\tilde{P})\right|^2\right] ds \\
&\quad +3\mathbb{E}\left[\sup_{0\le u\le t}\left|\int_0^u(\sigma-\tilde{\sigma})dW(s)\right|^2\right] \\
&\le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2+3T\int_0^t \mathbb{E}\left[\frac{1}{2}|\sigma^2-\tilde{\sigma}^2|^2+2\lambda^2|P-\tilde{P}|^2\right] ds \\
&\quad +12\int_0^t \mathbb{E}|\sigma-\tilde{\sigma}|^2 ds \\
&\le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2+\left(\frac{3}{2}M^2T+12L^2\right)\int_0^t \mathbb{E}|Z-\tilde{Z}|^2 ds \\
&\quad +(3M^2T+6\lambda^2T+12L^2)\int_0^t \mathbb{E}|P-\tilde{P}|^2 ds \\
&\le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2 \\
&\quad +\left(\frac{3}{2}M^2T+6\lambda^2T+12L^2\right)\int_0^t \mathbb{E}\left[|P-\tilde{P}|^2+|Z-\tilde{Z}|^2\right] ds,
\end{aligned}$$

then

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le u\le t}|\Sigma(u)-\tilde{\Sigma}(u)|^2\right] \\
&\quad\le\ 3\mathbb{E}|P(0)-\tilde{P}(0)|^2+(2M^2+6\lambda^2T+20L^2)\int_0^t\mathbb{E}|\Sigma(s)-\tilde{\Sigma}(s)|^2ds,
\end{aligned}
$$

and the theorem follows from the Gronwall lemma applied to

$$v(t)=\mathbb{E}[\sup_{0\le u\le t}|\Sigma(u)-\tilde{\Sigma}(u)|^2].$$

For the estimate on $\mathbb{E}[|\Sigma(t)-\tilde{\Sigma}(t)|^2]$, the proof proceeds in a similar way, without applying Doob's inequality to the term $\mathbb{E}[\sup_{0\le u\le t}|\int_0^u(\sigma-\tilde{\sigma})dW(s)|^2]$. □

Corollary 4.2. *If $h : C^0[0,T]\to\mathbb{R}$ is the payoff of a path-dependent claim such that the function $z(\cdot)\to h(e^{z(\cdot)})$ is globally Lipschitz, then*

$$\left|\mathbb{E}[h(S_T)]-\mathbb{E}[h(\tilde{S}_T)]\right|^2\le 3J^2\mathbb{E}|P(0)-\tilde{P}(0)|^2e^{c(L,M,T)T} \tag{4.1}$$

where J is the Lipschitz constant of $z(\cdot)\to h(e^{z(\cdot)})$. If h is a simple European claim, then an analogous estimate holds, with $C(L,M,T)$ instead of $c(L,M,T)$ and J the Lipschitz constant of $z\to h(e^z)$.

Proof. We have that

$$
\begin{aligned}
\mathbb{E}|h(S(T))-h(\tilde{S}(T))|^2 &\le J^2\mathbb{E}\|Z(\cdot)-\tilde{Z}(\cdot)\|^2_{C^0}\\
&\le J^2\mathbb{E}\left[\sup_{0\le t\le T}(|Z(t)-\tilde{Z}(t)|^2+|P(t)-\tilde{P}(t)|^2)\right],
\end{aligned}
$$

and from Theorem 4.1 we obtain Equation (4.1). □

We can see that the difference between the processes Σ and $\tilde{\Sigma}$ depends on the difference between the initial conditions $P(0)$ and $\tilde{P}(0)$. Unfortunately, we cannot obtain any improvement on the coefficients $c(L,M,T)$ or $C(L,M,T)$ in the case $\sigma=\sigma(P)$.

Remark 4.3. Notice that in Corollary 4.2 the function $z\to h(e^z)$ is required to be globally Lipschitz, so a little caution must be used. For example, if the function $h:\mathbb{R}\to\mathbb{R}$ is globally Lipschitz and piecewise C^1, then

$$\frac{\partial h(e^z)}{dz}=e^zh'(e^z)$$

is bounded (thus $z\to h(e^z)$ is globally Lipschitz) if and only if h' decreases faster than e^z.

Consider now some examples.

Example (European put). The payoff is $h(s) = (K - s)^+$. We have

$$\frac{\partial h(e^z)}{\partial z} = -e^z \mathbf{I}_{z<\log K},$$

then the Lipschitz constant in this case is less or equal than

$$I = \sup_z \left|\frac{\partial h(s)}{\partial s}\right| = e^{\log K} = K.$$

Example (European call). The payoff is now given by $h(s) = (s - K)^+$. We can write this expression as $h(s) = s - K - (K - s)^+$, so the error is the same as when pricing the put.

Example (Asian put). The payoff is now given by $h(s(\cdot)) = (K - \int_0^T s(t)dt)^+$. For two generic paths $z, \bar{z} \in C^0([0,T])$, if both $\int_0^T e^{z(u)}\, du, \int_0^T e^{\bar{z}(u)}\, du$ are less than K, then

$$\begin{aligned}
\left|h(e^{z(\cdot)}) - h(e^{\bar{z}(\cdot)})\right| &= \left|\left(K - \int_0^T e^{z(u)}\, du\right)^+ - \left(K - \int_0^T e^{\bar{z}(u)}\, du\right)^+\right| \\
&\le \left|\int_0^T e^{z(u)}\, du - \int_0^T e^{\bar{z}(u)}\, du\right| \\
&\le \left|\int_0^T |z(u) - \bar{z}(u)| e^{\sup(z(u),\bar{z}(u))}\, du\right| \\
&\le \|z - \bar{z}\|_{C^0} \left|\int_0^T e^{\sup(z(u),\bar{z}(u))}\, du\right| \le 2K\|z - \bar{z}\|_{C^0}, \quad (4.2)
\end{aligned}$$

where in the last line we applied the inequality

$$\int_0^T e^{\sup(z(u),\bar{z}(u))}\, du = \int_{\{u:z(u)>\bar{z}(u)\}} e^{z(u)}\, du + \int_{\{u:\bar{z}(u)>z(u)\}} e^{\bar{z}(u)}\, du \le 2K.$$

If (say) $\int_0^T e^{\bar{z}(u)}\, du > K$ and $\int_0^T e^{z(u)}\, du \le K$, then we can choose $\tilde{z} \in C^0$ such that $\int_0^T e^{\tilde{z}(u)}\, du = K$ and $\|\tilde{z} - z\|_{C^0} \le \|\bar{z} - z\|_{C^0}$ (for example, $\tilde{z} := tz + (1-t)\bar{z}$ for a suitable $t \in (0,1)$). Then

$$\begin{aligned}
\left|h(e^{z(\cdot)}) - h(e^{\bar{z}(\cdot)})\right| &= K - \int_0^T e^{z(u)}\, du = \left|\int_0^T e^{\tilde{z}(u)}\, du - \int_0^T e^{z(u)}\, du\right| \\
&\le 2K\|\tilde{z} - z\|_{C^0} \le 2K\|\bar{z} - z\|_{C^0}
\end{aligned}$$

by Equation (4.2). If both $\int_0^T e^{z(u)}\, du, \int_0^T e^{\bar{z}(u)}\, du$ are greater than K, there is nothing to prove. Then the Lipschitz constant in this case is equal to $2K$.

Example (Lookback put). The payoff is now given by

$$h(s(\cdot)) = \left(K - \max_{0 \le t \le T} s(t)\right)^+.$$

As above, if both $\max e^{z(\cdot)}$, $\max e^{\bar{z}(\cdot)}$ are less than K, then we calculate

$$\begin{aligned}\left|h(e^{z(\cdot)}) - h(e^{\bar{z}(\cdot)})\right| &\leq \left|\max_{0\leq u\leq T} e^{z(u)} - \max_{0\leq u\leq T} e^{\bar{z}(u)}\right| \leq \|e^{z(\cdot)} - e^{\bar{z}(\cdot)}\|_{C^0} \\ &\leq \|z-\bar{z}\|_{C^0}\|e^{\max(z,\bar{z})}\|_{C^0} \leq K\|z-\bar{z}\|_{C^0}.\end{aligned}$$

If at least one of the quantities $\max e^{z(\cdot)}$, $\max e^{\bar{z}(\cdot)}$ is greater than K, an argument similar to the one of the previous example applies. Thus, in this case the Lipschitz constant is equal to K.

5. Using past information

We have seen in Section 4 that the error in pricing derivative assets depends on the difference between the true offset function $P(0)$ and the misspecified value $\tilde{P}(0)$, which we can choose. Of course, our aim will be to choose it in order to minimise the final error. In doing this, we are entitled to use not only the current value of $S(0)$, but also past values.

More in detail, we assume (as it is reasonable) that we know all the past values of the price $S(t)$ (thus, of $Z(t)$) for $t \in [-R, 0]$, where $R > 0$ is a given real number which represents the width of an observation window in the past. As before, the process $P(t)$ remains unobserved also in the past. However, it turns out that we can make the uncertainty on P decay exponentially with respect to the width R of the observation window. Again, we represent this uncertainty by defining the process $\tilde{P}$, starting from the misspecified condition $\tilde{P}(-R)$ and following the dynamics

$$d\tilde{P}(t) = -\lambda\tilde{P}(t)\,dt + dZ(t), \quad t \in (-R, 0] \tag{5.1}$$

$$\tilde{P}(-R) \neq P(-R) \tag{5.2}$$

while the process P always follows the dynamics given by Equation (2.2). Notice that this time, as we can observe Z in the interval $[-R, 0]$, we have no uncertainty on this process.

The following lemma shows that, as both the dynamics of $\tilde{P}$ and P depend on the known values of Z, the difference between $P(0)$ and $\tilde{P}(0)$ decays exponentially with respect to the width R, as announced.

Lemma 5.1. *For every choice of $\tilde{P}(-R)$, we have*

$$|P(0) - \tilde{P}(0)| = e^{-\lambda R}|P(-R) - \tilde{P}(-R)|. \tag{5.3}$$

Proof. By calculating the Itô differential of the process $(e^{\lambda t}P(t))_t$, we have

$$\begin{aligned}d(e^{\lambda t}P(t)) &= e^{\lambda t}\,dP(t) + \lambda e^{\lambda t}P(t)\,dt \\ &= e^{\lambda t}(dZ(t) - \lambda P(t)\,dt) + \lambda e^{\lambda t}P(t)\,dt = e^{\lambda t}\,dZ(t)\end{aligned}$$

and, analogously,

$$de^{\lambda t}\tilde{P}(t) = e^{\lambda t}\,dZ(t).$$

This means that, calculating the two processes in the two points $t = -R, 0$, we have

$$
\begin{aligned}
P(0) &= e^{-\lambda R} P(-R) + \int_{-R}^{0} e^{\lambda t}\, dZ(t), \\
\tilde{P}(0) &= e^{-\lambda R} \tilde{P}(-R) + \int_{-R}^{0} e^{\lambda t}\, dZ(t).
\end{aligned}
$$

The lemma follows by calculating the difference. □

Remark 5.2. Notice that Equation (5.1) entails

$$
\begin{aligned}
\tilde{P}(0) &= e^{-\lambda R} \tilde{P}(-R) + Z(0) - e^{-\lambda R} Z(-R) - \int_{-R}^{0} \lambda e^{\lambda t} Z(t)\, dt \\
&= \int_{0}^{R} \lambda e^{-\lambda u} (Z(0) - Z(-u))\, du + e^{-\lambda R}(Z(0) - Z(-R) + \tilde{P}(-R)).
\end{aligned}
$$

This can be seen by the properties of stochastic integrals of deterministic functions, or directly from Equation (2.1) (which obviously extends to $\tilde{P}$).

Now we are in the position of solving the following problem: for a given $\varepsilon > 0$ we want to find a minimum observation time R such that the error when pricing a contingent claim h is less than ε.

Corollary 5.3. *If h is a general path-dependent claim as in Corollary* 4.2 *and*

$$
R > \frac{\log(\frac{3J^2 E|P(-R) - \tilde{P}(-R)|^2}{\varepsilon^2}) + c(L, M, T)T}{2\lambda}, \tag{5.4}
$$

then

$$
|E[h(S_T)] - E[h(\tilde{S}_T)]| < \varepsilon. \tag{5.5}
$$

Moreover, if $h(S(T))$ is the payoff of a simple European claim, then to obtain the same estimate it is sufficient that

$$
R > \frac{\log(\frac{3I^2 E|P(-R) - \tilde{P}(-R)|^2}{\varepsilon^2}) + C(L, M, T)T}{2\lambda}.
$$

Proof. From (5.4) we have

$$
2\lambda R > \log\Big(\frac{3J^2 \mathbb{E}|P(-R) - \tilde{P}(-R)|^2)}{\varepsilon^2}\Big) + c(L, M, T)T,
$$

that yields

$$
[c(L, M, T)T - 2\lambda R] + \log(3J^2 \mathbb{E}|P(-R) - \tilde{P}(-R)|^2) < \log \varepsilon^2.
$$

By taking the exponential of both the members we obtain

$$
3J^2 \mathbb{E}|P(-R) - \tilde{P}(-R)|^2 e^{c(L,M,T)T - 2\lambda R} < \varepsilon^2.
$$

From (4.1) and (5.3) we have

$$
|\mathbb{E}[h(Z_T)] - \mathbb{E}[h(\tilde{Z}_T)]|^2 \leq 3J^2 \mathbb{E}|P(-R) - \tilde{P}(-R)|^2 e^{C(L,M,T)T - 2\lambda R}; \tag{5.6}
$$

this implies that (5.5) is verified. For the case of a European claim, the proof is the same with $c(L, M, T)$ instead of $C(L, M, T)$. □

6. Stationarity

So far, we have seen that the problem of estimating the pricing error when we misspecify the offset function $\tilde{P}$ is led to the knowledge of $\mathbb{E}[|P(-R) - \tilde{P}(-R)|^2]$, which is in general not allowed as we do not know the initial distribution of $P(-R)$, even if we can decide the value $\tilde{P}(-R)$.

The situation can be much simplified if we make the crucial assumption that the 2-dimensional process (P, Z) is stationary, or that the process P itself is stationary. In this case, if we want the error to be (for example) less than a given $\varepsilon > 0$, it is sufficient to fix $\tilde{P}(-R)$ as being equal to the mean of the invariant measure of P (this minimises the quantity $\mathbb{E}[|P(-R) - \tilde{P}(-R)|^2]$, which is thus equal to the variance of $P(-R)$) and to observe the risky asset in the past for a sufficiently long time R. In fact, if the process P is stationary and admits a unique invariant measure, under suitable assumptions the marginal distribution of $P(t)$ converges, for $t \to +\infty$, to the invariant measure, regardless of the initial condition of P. This means that, if we assume that the process P started in the past at a time $T << -R$ from an arbitrary initial condition, the distribution of $P(-R)$ can be approximated very well by the invariant measure. Thus, the situation of finding $\mathbb{E}[|P(-R) - \tilde{P}(-R)|^2]$ boils down to finding the variance of the invariant measure for P, provided we let $\tilde{P}(-R)$ be equal to the mean of the invariant measure.

While the general case when the volatility σ depends on both P and Z seems more difficult to analyse, much can be said in the case when σ depends only on P. In this case the process P is a Markov process with the evolution

$$dP(t) = m(P(t))\ dt + \sigma(P(t))\ dW(t) \tag{6.1}$$

where $m(x) = -\frac{1}{2}\sigma^2(x) - \lambda x$. We now give sufficient conditions for the existence and uniqueness for the invariant distribution. For this purpose we use the following theorem from [11], that gives a condition for the existence of the invariant measure.

Theorem 6.1. *Assume that there exists a function $V \in C^2(\mathbb{R})$ such that*

$$V(x) \geq 0, \qquad \sup_{|x|>R} LV(x) := -A_R \to -\infty \quad as \quad R \to \infty$$

where $LV(x) := m(x)V'(x) + \frac{1}{2}\sigma^2(x)V''(x)$ and R is arbitrary. Then there exists a solution of Equation (6.1) which is a stationary Markov process.

Take $V(x) = x^2$, then

$$LV(x) = \left(-\frac{1}{2}\sigma^2(x) - \lambda x\right) x + \frac{1}{2}\sigma^2(x) = \frac{1}{2}(1-x)\sigma^2(x) - \lambda x^2.$$

Now if we assume

$$\sigma^2(x) \leq a|x| + b, \tag{6.2}$$

it follows that

$$\begin{aligned} LV(x) &\geq -\frac{1}{2}(x-1)(a|x|+b) - \lambda x^2 = \\ &= -\frac{1}{2}ax|x| - \frac{1}{2}bx + \frac{1}{2}(a|x|+b) - \lambda x^2. \end{aligned} \tag{6.3}$$

If $x > 0$, then $LV(x) \to -\infty$ when $R \to \infty$. If $x < 0$, then

$$LV(x) \geq \left(\frac{1}{2}a - \lambda\right) x^2 - \frac{1}{2}bx + \frac{1}{2}(a|x|+b) \to -\infty$$

if $a < 2\lambda$. We can thus conclude with the following result.

Theorem 6.2. *If Assumption* (6.2) *holds with $a < 2\lambda$, there exists an invariant measure for the process* (6.1).

In order to obtain also uniqueness results, we will need additional assumptions. If the process P has an invariant probability with density $\mu(x)$, from the backward Kolmogorov equation, we have

$$\begin{aligned} 0 &= -\frac{d[m(x)\mu(x)]}{dx} + \frac{1}{2}\frac{d^2[\sigma^2(x)\mu(x)]}{dx^2} \\ 0 &= \frac{d}{dx}\left[-m(x)\mu(x) + \frac{1}{2}\frac{d\sigma^2(x)\mu(x)}{dx}\right]; \end{aligned} \tag{6.4}$$

this implies that

$$\frac{1}{2}\frac{d\sigma^2(x)\mu(x)}{dx} = m(x)\mu(x) + c.$$

Assume that $c = 0$ and $y(x) = \sigma^2(x)\mu(x)$: then we have

$$\begin{aligned} \int \frac{dy}{y} &= \int \frac{2m(x)}{\sigma^2(x)}dx \\ \ln y &= \int_{x_0}^{x} \frac{2m(u)}{\sigma^2(u)}du + \ln C \\ y(x) &= Ce^{\int_{x_0}^{x} \frac{2m(u)}{\sigma^2(u)}du} \end{aligned} \tag{6.5}$$

where C is an arbitrary constant and x_0 is an arbitrary point. If the relation

$$\mu(x) = C\frac{e^{G(x)}}{\sigma^2(x)} \tag{6.6}$$

where $G(x) = \int_{x_0}^{x} \frac{2m(u)}{\sigma^2(u)}du$, gives a density, this is the invariant density for our process P.

Now we study the conditions for existence and uniqueness of the invariant measure for the process P when σ satisfies the following assumption:

Assumption 6.3. There exist $a \in [0, 2\lambda)$, $b, \varepsilon > 0$ such that

$$\varepsilon \leq \sigma^2(x) \leq a|x| + b.$$

Theorem 6.4. *If σ satisfies Assumption* (6.3)*, then there exists an unique invariant measure for P, with density given by* (6.6)*. Moreover, if $P^{-T,\eta}$ follows the dynamics* (6.1) *with initial condition $P^{-T,\eta}(-T) = \eta$ with $-T < -R$, then for every initial distribution η and $E \in \mathbb{R}$, we have*

$$\lim_{T\to\infty} \mathbb{E}[(P^{-T,\eta}(-R) - E)^2] = \int_{\mathbb{R}} (x - E)^2 \mu(x)\, dx.$$

Proof. By results contained in [11], it is sufficient to prove that

$$\int_{-\infty}^{\infty} \frac{e^{G(x)}}{\sigma^2(x)}\, dx < \infty$$

and that

$$\int_{-\infty}^{0} \frac{e^{-G(x)}}{\sigma^2(x)}\, dx = \int_{0}^{\infty} \frac{e^{-G(x)}}{\sigma^2(x)}\, dx = +\infty$$

where

$$G(x) = \int_0^x \left(-1 - \frac{2\lambda u}{\sigma^2(u)}\right)\, du = -x - \int_0^x \frac{2\lambda u}{\sigma^2(u)}\, du.$$

If $x \geq 0$,

$$\begin{aligned} G(x) &\leq -x - \frac{2\lambda}{a}\int_0^x \frac{au+b-b}{au+b} du + C = -x - \frac{2\lambda}{a}x + \frac{2\lambda}{a^2}\int_0^x \frac{-b}{au+b} du + C \\ &= -x - \frac{2\lambda}{a}x + \frac{2\lambda b}{a^2}\ln(ax+b) + C_1 =: n_1(x). \end{aligned}$$

If $x < 0$,

$$G(x) \leq -x - 2\lambda\int_0^x \frac{u}{\varepsilon}\, du = -x - \frac{\lambda}{\varepsilon}x^2 =: n_2(x)$$

where as usual C, C_1, are some constants. Then $e^{G(x)} \leq e^{n_1(x)}$ if $x \geq 0$ and $e^{G(x)} \leq e^{n_2(x)}$ if $x < 0$. So, we can write

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{e^{G(x)}}{\sigma^2(x)} dx \leq \int_{-\infty}^{\infty} \frac{e^{G(x)}}{\varepsilon} dx \\ &\leq K_1 \int_{-\infty}^{0} \frac{e^{-x-\frac{\lambda}{\varepsilon}x^2}}{\varepsilon}\, dx + K_2 \int_0^{+\infty} \frac{e^{-x(1+\frac{2\lambda}{a})}(ax+b)^{\frac{2\lambda b}{a^2}}}{\varepsilon} dx < +\infty \end{aligned}$$

where K_1 and K_2 are constants. Besides,

$$\begin{aligned} \int_{-\infty}^{0} e^{-G(x)}\, dx &\geq \int_{-\infty}^{0} e^{-n_2(x)}\, dx = K_1 \int_{-\infty}^{0} e^{x+\frac{\lambda}{\varepsilon}x^2}\, dx = +\infty, \\ \int_0^{\infty} e^{-G(x)}\, dx &\geq \int_0^{+\infty} e^{-n_1(x)}\, dx = K_2 \int_0^{+\infty} e^{x(1+\frac{2\lambda}{a})}(ax+b)^{-\frac{2\lambda b}{a^2}}\, dx = +\infty. \end{aligned}$$

□

7. Some examples

Now we analyse some particular specifications for σ. The first two are present in the original Hobson-Rogers paper and in other related papers (see [12]), while the third is suggested by the fact that affine processes are very often used in mathematical finance, and they have a well-established theory.

7.1. The case $\sigma(P) = \min\{\sqrt{a+bP^2}, N\}$

This example comes from the original Hobson-Rogers paper [12]:

$$\sigma(P) = \min\{\sqrt{a+bP^2}, N\} \tag{7.1}$$

where $a > 0$, $b > 0$ and $N > 0$ are some constants. As σ satisfies Assumption 6.3 for each possible value of $a, b, N > 0$, we can calculate the function $G(x)$:

$$G(x) = -(x - x_0) - \int_{x_0}^{x} \frac{2\lambda u}{\sigma^2(u)} du.$$

When $x < \sqrt{\frac{N^2-a}{b}}$, the function $G(x)$ becomes

$$G(x) = -(x - x_0) - \int_{x_0}^{x} \frac{2\lambda u}{N^2}\, du = (x_0 - x) - \frac{2\lambda}{N^2}\left(\frac{x^2}{2} - \frac{x_0^2}{2}\right) = -x - \frac{\lambda}{N^2}x^2 + L_1$$

where L_1 is a constant. In this case the function $\mu(x)$ is equal to

$$\mu(x) = C\frac{e^{G(x)}}{N^2} = K\frac{e^{-\frac{\lambda}{N^2}x^2 - x}}{N^2} = K_1\frac{e^{-\frac{\lambda}{N^2}(x+\frac{N^2}{2\lambda})^2}}{N^2}$$

where K and K_1 are constants. When $x \in \left[-\sqrt{\frac{N^2-a}{b}}, +\sqrt{\frac{N^2-a}{b}}\right]$, the function $G(x)$ is

$$G(x) = -(x - x_0) - \int_{-\sqrt{\frac{N^2-a}{b}}}^{x} \frac{2\lambda u}{a + bu^2} du = -x - \frac{\lambda}{b}\ln(a + bx^2) + L_2$$

where L_2 is a constant. The function $\mu(x)$ is equal to

$$\mu(x) = K_2\frac{e^{-x}(a + bx^2)^{-\frac{\lambda}{b}}}{a + bx^2} = K_2 e^{-x}(a + bx^2)^{-\frac{\lambda}{b}-1}$$

where K_2 is a constant. Now we see the case when $x > \sqrt{\frac{N^2-a}{b}}$. In this case the function $G(x)$ is

$$G(x) = -(x - x_0) - \int_{\sqrt{\frac{N^2-a}{b}}}^{x} \frac{2\lambda u}{N^2} du = -x - \frac{\lambda}{N^2}x^2 + L_3$$

for some constant L_3. Then

$$\mu(x) = K_3\frac{e^{-\frac{\lambda}{N^2}(x+\frac{N^2}{2\lambda})^2}}{N^2}$$

where K_3 is a constant. The function $\mu(x)$ must be continuous at the points $x_1 = -\sqrt{\frac{N^2-a}{b}}$ and $x_2 = \sqrt{\frac{N^2-a}{b}}$, so that at this point we have

$$\lim_{x \to x_1^-} \mu(x) = \lim_{x \to x_1^+} \mu(x), \qquad \text{and} \lim_{x \to x_2^-} \mu(x) = \lim_{x \to x_2^+} \mu(x),$$

that implies

$$K_2 = K_1 e^{-\frac{\lambda(N^2-a)}{bN^2} - \frac{N^2}{4\lambda}} N^{\frac{2\lambda}{b}}, \quad K_3 = K_1.$$

In conclusion, the invariant density is

$$\mu(x) = \begin{cases} K_1 e^{-\frac{\lambda(N^2-a)}{bN^2} - \frac{N^2}{4\lambda}} N^{\frac{2\lambda}{b}} e^{-x} (a + bx^2)^{-\frac{\lambda}{b}-1} & |x| \le \sqrt{\frac{N^2-a}{b}} \\ K_1 \frac{e^{-\frac{\lambda}{N^2}(x + \frac{N^2}{2\lambda})^2}}{N^2} & |x| \ge \sqrt{\frac{N^2-a}{b}}. \end{cases}$$

For the mean and the covariance of the process P under the invariant measure, there is not an explicit form. For this reason, a numerical calculation is required.

Example. As in [9], we take

$$a = 0.04, \quad b = 0.2, \quad \lambda = 1, \quad N = 1,$$

so we have

$$L = \sup_{x \in \mathbb{R}} \left| \frac{\partial \sigma}{\partial x} \right| = \frac{\sqrt{b(N^2 - a)}}{N} = 0.438178$$

and

$$M = \sup_{x \in \mathbb{R}} \left| \frac{\partial \sigma^2}{\partial x} \right| = 2\sqrt{b(N^2 - a)} = 0.876356,$$

then we have

$$\mathbb{E}[P] = -0.022293, \qquad \mathrm{Var}[P] = 0.022437.$$

We want to find R such that (5.5) is verified for $\varepsilon = 10^{-2}$. If $J = 1$ (as is often the case), by taking different maturities, we find these results both for a general path-dependent claim as for a European one:

	path-dependent claim		European claim	
T	$c(L, M, T)$	R	$c(L, M, T)$	R
0.25	5.724000	3.971457	2.844000	3.611457
0.5	7.608000	5.157957	4.728000	4.437957
1.0	11.376000	8.943957	8.496000	7.503957
2.0	18.912000	22.167957	16.032000	19.287957
3.0	26.448000	42.927957	23.568000	38.607957
4.0	33.984000	71.223957	31.104000	65.463957
5.0	41.520000	107.055957	38.640000	99.855957

In this case, if we want to make an error of less than $\varepsilon = 10^{-2}$ in pricing (for example) a 6-months contingent claim, we have to observe the underlying asset for at least 5.15 years in the case of a path-dependent contingent claim and at least 4.43 years in the case of a European contingent claim.

Of course the situation can change, depending on the parameters. Take for example (always from [9])

$$a = 0.49, \quad b = 2.45, \quad \lambda = 1, \quad N = 2.236068.$$

Now we have

$$L = \sup_{x\in\mathbb{R}} \left|\frac{\partial\sigma}{\partial x}\right| = 1.486573, \qquad M = \sup_{x\in\mathbb{R}} \left|\frac{\partial\sigma^2}{\partial x}\right| = 6.648158$$

and

$$\mathbb{E}[P] = 1.281530, \qquad \mathrm{Var}[P] = 2.674600.$$

If again we want to find R such that (5.5) is verified for $\varepsilon = 10^{-2}$ and $J = 1$, this time we find these results both for a general path-dependent claim as for a European one:

	path-dependent claim		European claim	
T	$c(L, M, T)$	R	$c(L, M, T)$	R
0.25	67.797000	14.121001	34.648500	9.977439
0.5	91.396000	28.495376	58.247500	20.208251
1.0	138.594000	74.943376	105.445500	58.369126
2.0	232.990000	238.636376	199.841500	205.487876
3.0	327.386000	496.725376	294.237500	447.002626
4.0	421.782000	849.210376	388.633500	782.913376
5.0	516.178000	1296.091376	483.029500	1213.220126

In this case, if we want to make an error of less than $\varepsilon = 10^{-2}$ in pricing (for example) a 6-months contingent claim, we have to observe the underlying asset for at least 28.49 years in the case of a path-dependent contingent claim and at least 20.20 years in the case of a European contingent claim.

7.2. The case $\sigma^2(P) = \frac{a+bP^2}{c+d'P^2}$

Consider σ of the form

$$\sigma^2(P) = \frac{a + bP^2}{c + d'P^2}$$

where a, b, c, d' are some positive numbers. As σ satisfies Assumption 6.3 for each possible value of $a, b, c, d' > 0$, as in the previous section we calculate the function G:

$$\begin{aligned}
G(x) &= -(x - x_0) - 2\lambda \int_{x_0}^{x} \frac{(c + d'u^2)u}{a + bu^2} du = -x - \lambda \int_{x_0^2}^{x^2} \frac{c + d'u^2}{a + bu^2} du^2 + c_0 \\
&= -x - \frac{\lambda c}{b} \int_{bx_0^2+a}^{bx^2+a} \frac{1}{y} dy - \frac{\lambda d'}{b^2} \int_{bx_0^2+a}^{bx^2+a} \frac{a + bu^2 - a}{a + bu^2} d(bu^2 + a) + c_1 \\
&= -x - \frac{\lambda c}{b} \ln(bx^2) - \frac{\lambda d'}{b^2}(bx^2 + a) + \frac{\lambda d' a}{b^2} \ln(bx^2 + a) + c_2 \\
&= -x - \frac{\lambda(bc - ad')}{b^2} \ln(bx^2 + a) - \frac{\lambda d'}{b^2}(bx^2 + a) + c_2. \qquad (7.2)
\end{aligned}$$

The function μ is

$$\mu(x) = C\frac{e^{G(x)}}{\sigma^2(x)} = C\frac{e^{-x}(bx^2+a)^{-\frac{\lambda}{b^2}(bc-ad')}e^{-\frac{\lambda d'}{b^2}(bx^2+a)+c_1}}{\frac{a+bx^2}{c+d'x^2}}$$
$$= K\frac{e^{-\frac{\lambda d'}{b}(x+\frac{b}{2\lambda d'})^2}(bx^2+a)^{-\frac{\lambda}{b^2}(bc-ad')-1}}{c+d'x^2} \tag{7.3}$$

and it is the density of the unique invariant measure of the process P. Also in this case, we cannot calculate explicitly the mean and the variance of the process P, so a numerical integration is again required.

Example. We take

$$a = 0.452, \quad b = 3.012, \quad c = 1.0, \quad d' = 0.261, \quad \lambda = 1.02.$$

We calculate the Lipschitz constants L and M for the functions σ and σ^2. We have

$$L = \sup_{x\in\mathbb{R}}\left|\frac{\partial\sigma}{\partial x}\right| = 1.22302 \quad \text{and} \quad M = \sup_{x\in\mathbb{R}}\left|\frac{\partial\sigma^2}{\partial x}\right| = 3.67938.$$

In fact, denote

$$k(x) := \frac{\partial\sigma(x)}{\partial x} = \frac{(bc-ad')x}{(c+d'x^2)^{\frac{3}{2}}\sqrt{a+bx^2}}$$

which reaches its maximum for $x = \pm\sqrt{\frac{-d'a+\sqrt{d'^2a^2+4abcd'}}{4d'b}}$. Then

$$L = \left|k\Big(\pm\sqrt{\frac{-d'a+\sqrt{d'^2a^2+4abcd'}}{4d'b}}\Big)\right| = 1.22302.$$

Similarly, let us denote

$$g(x) := \frac{\partial\sigma^2}{\partial x} = \frac{2(bc-ad')x}{(c+d'x^2)^2}$$

which reaches its maximum for $x = \pm\sqrt{\frac{c}{3d'}}$. Then

$$M = \left|g\Big(\pm\sqrt{\frac{c}{3d'}}\Big)\right| = \frac{2|bc-ad'|\sqrt{\frac{c}{3d'}}}{(c+d'\frac{c}{3d'})^2} = 3.67938.$$

We obtain

$$\mathbb{E}[P] = -0.324053, \qquad \mathrm{Var}[P] = 0.612203$$

and we have these results respectively for a path-dependent and for a European contingent claim:

	path-dependent claim		European claim	
T	$c(L,M,T)$	R	$c(L,M,T)$	R
0.25	38.245077	9.499770	15.808408	6.750178
0.5	46.574596	16.228215	24.137927	10.729032
1.0	63.233633	35.809752	40.796964	24.811385
2.0	96.551707	99.471410	74.115038	77.474675
3.0	129.869782	195.797846	107.433113	162.802745
4.0	163.187856	324.789061	140.751187	280.795593
5.0	196.505930	486.445055	174.069261	431.453220

7.3. The case $\sigma^2(P) = a + bP$

Suppose that the process P is a so-called affine process [5], i.e., σ is given by

$$\sigma^2(P) = a + bP \tag{7.4}$$

where a and b are two arbitrary constants. So, Equation (6.1) becomes

$$dP(t) = \left(-\left(\frac{b}{2} + \lambda\right) P(t) - \frac{a}{2}\right)\, dt + \sqrt{a + bP(t)}dB(t). \tag{7.5}$$

Clearly, there is a solution to (7.5) when the process $a + bP(t)$ is non-negative for all t. So, the domain D implied by the non-negativity is

$$D = \{x \in R : a + bx > 0\}.$$

We will therefore need to assume, in effect, that the process $a + bP(t)$ has a sufficiently strong positive drift on the boundary point $x = -\frac{a}{b}$. Under the following assumption, we have a unique (strong) solution for the stochastic equation (7.5).

Assumption 7.1. We assume that $2\lambda a > b^2$.

In fact, for x such that $a + bx = 0$, $b[-(\frac{1}{2}b + \lambda)x - \frac{1}{2}a] > \frac{b^2}{2}$, i.e., equivalently $(1 + \frac{2\lambda}{b})x + (1 + \frac{a}{b}) < 0$, this implies $2\lambda a > b^2$. See [5].

Theorem 7.2. *Under Assumption* (7.1)*, there is a unique (strong) solution P of the stochastic differential equation* (7.5) *in the domain D. Moreover, for this solution P, we have $a + bP(t) > 0$ for all t almost surely.*

Since σ is not Lipschitz, we cannot apply Theorem 4.1, but we have to formulate an analogous result here.

Theorem 7.3. *If the coefficient σ satisfies* (7.4)*, then for $t \in [0, T]$ we have*

$$\mathbb{E}\left[\sup_{0\le u\le t} |\Sigma(u) - \tilde{\Sigma}(u)|^2\right] \le \left[3\mathbb{E}|P(0) - \tilde{P}(0)|^2 + \frac{10b^2}{\theta} t\right] e^{c(\theta,T)t} \tag{7.6}$$

where θ is an arbitrary parameter and $c(\theta,T)=\left[3(\frac{b}{2}+\lambda)^2+\frac{b^2}{2}\right]T+10b^2\theta$, *and*

$$\mathbb{E}|\Sigma(t)-\tilde{\Sigma}(t)|^2\leq\left[3\mathbb{E}|P(0)-\tilde{P}(0)|^2+\frac{5b^2}{2\theta'}t\right]e^{C(\theta',T)t} \tag{7.7}$$

where θ' is an arbitrary parameter and $C(\theta',T)=\frac{b^2}{2}(T+5\theta')+3(\frac{b}{2}+\lambda)^2T$.

Proof. We have that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq u\leq t}|\Sigma(u)-\tilde{\Sigma}(u)|^2\right]\\
&\leq\ \mathbb{E}\left[\sup_{0\leq u\leq t}|Z(u)-\tilde{Z}(u)|^2\right]+\mathbb{E}\left[\sup_{0\leq u\leq t}|P(u)-\tilde{P}(u)|^2\right]=(1)+(2).
\end{aligned}$$

For the first term we have

$$\begin{aligned}
(1)\ &=\ \mathbb{E}\left[\sup_{0\leq u\leq t}\left|-\frac{1}{2}\int_0^u b(P(s)-\tilde{P}(s))ds\right.\right.\\
&\qquad\left.\left.+\int_0^u\left(\sqrt{a+bP(s)(s)}-\sqrt{a+b\tilde{P}(s)}\right)dW(s)\right|^2\right]\\
&\leq\ 2\mathbb{E}\left[\sup_{0\leq u\leq t}\left|\frac{1}{2}\int_0^u b(P(s)-\tilde{P}(s))\ ds\right|^2\right]\\
&\qquad+2\mathbb{E}\left[\sup_{0\leq u\leq t}\left|\int_0^u\left(\sqrt{a+bP(s)}-\sqrt{a+b\tilde{P}(s)}\right)dW(s)\right|^2\right]\\
&\leq\ \frac{b^2}{2}T\int_0^t\mathbb{E}|P(s)-\tilde{P}(s)|^2\ ds+8b^2\int_0^t\mathbb{E}|P(s)-\tilde{P}(s)|\ ds\\
&\leq\ \left(\frac{b^2}{2}T+4b^2\theta\right)\int_0^t\mathbb{E}|P(s)-\tilde{P}(s)|^2\ ds+\frac{4b^2}{\theta}t,
\end{aligned}$$

where in the third line we apply the inequality

$$\left|\sqrt{a+bP}-\sqrt{a+b\tilde{P}}\right|\leq b\sqrt{|P-\tilde{P}|}$$

and in the last line the inequality

$$|P-\tilde{P}|\leq\frac{\theta}{2}|P-\tilde{P}|^2+\frac{1}{2\theta}$$

which holds for any real number $\theta > 0$. Then,

$$
\begin{aligned}
(2) \quad &= \mathbb{E}\left[\sup_{0\le u\le t} |P(0)-\tilde{P}(0) - \left(\frac{b}{2}+\lambda\right)\int_0^u (P-\tilde{P})ds \right. \\
&\quad \left. + \int_0^u \left(\sqrt{a+bP}-\sqrt{a+b\tilde{P}}\right)dW(s)|^2\right] \\
&\le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2 + 3\left(\frac{b}{2}+\lambda\right)^2 T\int_0^t \mathbb{E}|P-\tilde{P}|^2 ds \\
&\quad +12\int_0^t \mathbb{E}|\sqrt{a+bP}-\sqrt{a+b\tilde{P}}|^2 ds \\
&\le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2 + 3\left(\frac{b}{2}+\lambda\right)^2 T\int_0^t \mathbb{E}|P-\tilde{P}|^2\ ds \\
&\quad +6b^2\theta\int_0^t \mathbb{E}|P-\tilde{P}|^2 ds + \frac{6b^2}{\theta}t \\
&= 3\mathbb{E}|P(0)-\tilde{P}(0)|^2 + \left[3\left(\frac{b}{2}+\lambda\right)^2 T + 6b^2\theta\right]\int_0^t \mathbb{E}|P-\tilde{P}|^2 ds + \frac{6b^2}{\theta}t.
\end{aligned}
$$

Then

$$
\mathbb{E}\left[\sup_{0\le u\le t}|\Sigma(u)-\tilde{\Sigma}(u)|^2\right] \le 3\mathbb{E}|P(0)-\tilde{P}(0)|^2 + \frac{10b^2}{\theta}t + c(\theta,T)\int_0^t \mathbb{E}|\Sigma(s)-\tilde{\Sigma}(s)|^2 ds.
$$

Similarly as in Theorem 4.1, the result follows from Gronwall's lemma applied to $v(t) = \mathbb{E}\left[\sup_{0\le u\le t}|\Sigma(u)-\tilde{\Sigma}(u)|^2\right]$. □

The parameters θ and θ' which minimize the right-hand side of Equation (7.6) and Equation (7.7) are

$$
\theta = \frac{-5b^2t + \sqrt{25b^4t^2 + 3\mathbb{E}|P(0)-\tilde{P}(0)|^2}}{3\mathbb{E}|P(0)-\tilde{P}(0)|^2}, \qquad \theta' = \frac{1}{4}\theta.
$$

Now we calculate the function $G(x)$. The inequality $a+bx \ge 0$ is equivalent to $x \ge -\frac{a}{b}$ if $b>0$ and to $x \le -\frac{a}{b}$ if $b<0$. Consider the case $b>0$.

$$
\begin{aligned}
G(x) &= -\left(x+\frac{a}{b}\right) - \int_{-\frac{a}{b}}^x \frac{2\lambda u}{a+bu}du \\
&= -\left(1+\frac{2\lambda}{b}\right)x + \frac{2\lambda a}{b^2}\ln(a+bx) - \frac{4\lambda a}{b^2} - \frac{a}{b}.
\end{aligned}
$$

So, the function $\mu(x)$ is

$$
\mu(x) = C\frac{e^{G(x)}}{\sigma^2(x)} = K\frac{e^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}}}{a+bx}
$$

where $K = Ce^{-\frac{4\lambda a}{b^2} - \frac{a}{b}}$ is constant. For $\mu(x)$ to be a density, the quantity

$$\int_{-\frac{a}{b}}^{\infty} e^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}-1}dx$$

must be finite. This is true if $(1+\frac{2\lambda}{b}) > 0$ which is always true, and $\frac{2\lambda a}{b^2} - 1 > -1$, i.e., $a > 0$.

Now we analyze the case $b < 0$. In this case $x \le -\frac{a}{b}$, then,

$$\begin{aligned} G(x) &= -(x-x_0) - \int_{x_0}^{x} \frac{2\lambda u}{a+bu}\, du \\ &= \Big(\frac{2\lambda}{b} - 1\Big)x - \frac{2\lambda a}{b^2}\ln(bx+a) + C_1 \end{aligned}$$

where in the first line we change the variable of integration to $y = a + bu$ and C_1 is some constant. Similarly, as in the case $b > 0$, the function $\mu(x)$ is a density if

$$\int_{-\infty}^{-\frac{a}{b}} e^{-(1-\frac{2\lambda}{b})x}(a+bx)^{-\frac{2\lambda a}{b^2}-1}dx$$

is finite. This is true when $(1-\frac{2\lambda}{b}) < 0$ (equivalently, when $b > 2\lambda > 0$), and $-\frac{2\lambda a}{b^2} - 1 > -1$. But this is absurd because we supposed that $b < 0$. In conclusion,

$$\mu(x) = K' e^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}-1}$$

is an invariant density for our process P in $(-\frac{a}{b}, +\infty)$ if and only if $a > 0$ and $b > 0$. In this case we can calculate the marginal mean and variance for the process P under the invariant measure. For the mean we have that for all $t \in \mathbb{R}$,

$$\begin{aligned} E[P(t)] &= \int_{-\frac{a}{b}}^{\infty} xKe^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}-1}dx \\ &= \frac{1}{b}\int_{-\frac{a}{b}}^{\infty} K(bx+a-a)e^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}-1}dx \\ &= \frac{1}{b}\int_{-\frac{a}{b}}^{\infty} Ke^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}}dx - \frac{a}{b} \\ &= \frac{2\lambda a}{b(b+2\lambda)}\int_{-\frac{a}{b}}^{\infty} Ke^{-(1+\frac{2\lambda}{b})x}(a+bx)^{\frac{2\lambda a}{b^2}-1}dx - \frac{a}{b} \\ &= \frac{2\lambda a}{b(b+2\lambda)} - \frac{a}{b} = -\frac{a}{b+2\lambda}. \end{aligned}$$

Now we calculate $E[P^2(t)]$:

$$\begin{aligned}
E[P^2(t)] &= \int_{-\frac{a}{b}}^{\infty} Kx^2 e^{-(1+\frac{2\lambda}{b^2})x}(a+bx)^{\frac{2\lambda a}{b}-1}dx \\
&= \frac{1}{b^2}\int_{-\frac{a}{b}}^{\infty} K(bx+a-a)^2 e^{-(1+\frac{2\lambda}{b^2})x}(a+bx)^{\frac{2\lambda a}{b}-1}dx \\
&= \frac{1}{b^2}\int_{-\frac{a}{b}}^{\infty} K e^{-(1+\frac{2\lambda}{b^2})x}(a+bx)^{\frac{2\lambda a}{b}+1}dx \\
&\quad -\frac{2a}{b^2}\int_{-\frac{a}{b}}^{\infty} K e^{-(1+\frac{2\lambda}{b^2})x}(a+bx)^{\frac{2\lambda a}{b}}dx + \frac{a^2}{b^2} \\
&= \frac{2\lambda a + b^2}{b^2(b+2\lambda)(1+\frac{2\lambda}{b})}\frac{2\lambda a}{b^2}b - \frac{4\lambda a^2}{b^2(b+2\lambda)} + \frac{a^2}{b^2} \\
&= \frac{a^2+2\lambda a}{(b+2\lambda)^2} = \frac{a(a+2\lambda)}{(b+2\lambda)^2}.
\end{aligned}$$

So that, the variance of the invariant measure of the process P is equal to

$$\mathrm{Var}[P(t)] = E[P^2(t)] - E[P(t)]^2 = \frac{2\lambda a}{(b+2\lambda)^2}.$$

Acknowledgment

Both the authors gratefully acknowledge financial support from the CNR Strategic Project *Modellizzazione matematica di fenomeni economici*. The authors also wish to thank Paolo Dai Pra, Gianni Di Masi, Franco Flandoli, Alessandro Platania, Chris Rogers and Giovanna Villani for useful discussions on this topic, and an anonymous referee for having pointed out some ambiguities on the first draft of the manuscript.

References

[1] M. Arriojas, Y. Hu, S.-E. Mohammed, and G. Pap, *A delayed Black and Scholes formula*, Stoch. Anal. Appl., **25 (2)** (2007), 471–492.

[2] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, **81** (1973), 637–659.

[3] K. Burdzy, *Some path properties of iterated Brownian motion*, Seminar on Stochastic Processes, E. Cinlar et al., Editors, Birkhäuser, Boston, (1993), 67–87

[4] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge University Press, 1996.

[5] D. Duffie and R. Kan, *A yield-factor model of interest rates*, Mathematical Finance, **6** (1996), 379–406.

[6] B. Dupire, *Pricing with a smile*, Risk Magazine, **7 (1)** (1994), 18–20.

[7] B. Dupire, *Pricing and hedging with smiles*, Mathematics of Derivative Securities, M. A. H. Dempster et al., Editors, Cambridge Univ. Press., Publ. Newton Inst., (1997), 103–111.

[8] I. Elsanosi, B. J. Øksendal, and A. Sulem, *Some solvable stochastic control problems with delay*, Stochastics and Stochastics Reports, **71** (2000), 69–89.

[9] P. Foschi and A. Pascucci, *Calibration of the Hobson & Rogers model: empirical tests*, preprint AMS Acta, University of Bologna, 2005.

[10] V. Hallulli and T. Vargiolu, *Financial models with dependence on the past: a survey*, in: Applied and Industrial Mathematics in Italy, M. Primicerio et al., Editors, Series on Advances in Mathematics for Applied Sciences, World Scientific, **69** (2005), 348–359.

[11] R. Z. Has'minskii, *Stochastic Stability of Differential Equations*, Sifthoff Noorthoff, 1980.

[12] D. G. Hobson and L. C. G. Rogers, *Complete models with stochastic volatility*, Mathematical Finance, **8 (1)** (1998), 27–48.

[13] X. Mao, *Stochastic Differential Delay Equations*, Berlin, March 2003, `http://sfde.math.siu.edu/berlinschool1.html`.

[14] S.-E. A. Mohammed, *Stochastic Differential Systems with Memory*, Berlin, March 2003, `http://sfde.math.siu.edu/berlinschool1.html`.

[15] R. C. Merton, *Theory of rational option pricing*, Bell Journal of Economic and Management Science, **4** (1973), 141–183.

Vera Blaka Hallulli and Tiziano Vargiolu
Department of Pure and Applied Mathematics
University of Padova
via Belzoni 7
I-35131 Padova
Italy
e-mail: `hallulli@math.unipd.it`
`vargiolu@math.unipd.it`

Progress in Probability, Vol. 59, 493–506

PDE Approach to Utility Maximization for Market Models with Hidden Markov Factors

Hideo Nagai and Wolfgang J. Runggaldier

Abstract. We consider the problem of maximizing expected utility from terminal wealth for a power utility of the risk-averse type assuming that the dynamics of the risky assets are affected by hidden "economic factors" that evolve as a finite-state Markov process. For this partially observable stochastic control problem we determine a corresponding complete observation problem that turns out to be of the risk sensitive type and for which the Dynamic programming approach leads to a nonlinear PDE that, via a suitable transformation, can be made linear. By means of a probabilistic representation we obtain a unique viscosity solution to the latter PDE that induces a unique viscosity solution to the former. This probabilistic representation allows us to obtain, on the one hand, regularity results, on the other hand, a computational approach based on Monte Carlo simulation.

Mathematics Subject Classification (2000). Primary 93E20; Secondary 91B28, 49L20, 49L25.

Keywords. Portfolio optimization, hidden Markov factors, stochastic control under partial information, risk-sensitive control, HJB-equations, viscosity solutions.

1. Introduction

We consider a market model with one locally riskless security and a certain number of risky securities. The goal is to find an admissible self-financing investment strategy that maximizes the expected utility from terminal wealth at a given maturity and with a power utility function of the risk-averse type.

We assume that the dynamics of the risky assets are affected by exogenous "economic factors" that evolve as a finite-state Markov process. We allow these economic factors to be hidden, i.e., they may not be observed directly. Information about these factors can therefore be obtained only by observing the prices of the risky assets.

Our problem is thus of the type of a partially observed stochastic control problem and we shall determine its solution by solving a corresponding complete observation control problem. After discussing some problems that arise for a complete observation problem based on unnormalized filter values, we construct an equivalent complete observation control problem, where the new state is given by the pair (p_t, Y_t) consisting of the conditional state probability vector (normalized filter) p_t for the hidden factor process and of the log-asset prices Y_t. This pair forms a Markov process also in our more general setup where the coefficients in the security price dynamics are nonlinearly dependent upon the factors. The equivalent complete observation control problem turns out to be of the type of a risk sensitive stochastic control problem. It is approached by the method of Dynamic Programming (DP) that leads to a nonlinear HJB equation. Applying a transformation that is by now rather classical, this nonlinear HJB equation is transformed into a linear one. By means of a probabilistic representation as expectation of a suitable function of the underlying Markov process (p_t, Y_t), we obtain a unique viscosity solution to the latter PDE that induces a unique viscosity solution to the former. This probabilistic representation allows us to obtain, on the one hand, regularity results on the basis of classical results on expectations of functions of diffusion processes; on the other hand, it allows us to obtain a computational approach based on Monte Carlo simulation. This latter computational approach is important since, as we shall show, an explicit analytic solution is very difficult to obtain in the given setup.

Portfolio optimization problems under partial information are becoming more and more popular, also because of their practical interest. They have been studied using both major portfolio optimization methodologies, namely Dynamic Programming (DP) and the "Martingale Method"(MM). While DP has a longer tradition in general, also MM has been applied already since some time for the cases when the drift/appreciation rate in a diffusion-type market model is supposed to be an unknown constant, a hidden finite-state Markov process, or a linear-Gaussian factor process. Along this line are the papers [8, 9, 10, 22] and, more recently, [5, 20]. The case when the volatility is driven by a hidden process is studied in [16]. After the early paper [3], a DP-approach for a finite-horizon linear-Gaussian model with one unobserved factor that is independent of the risky asset has been used in [18]. In this latter paper the author also ends up with a nonlinear PDE. However, instead of using a transformation to reduce the equation to a linear one, the author introduces an auxiliary problem of the linear-quadratic type and obtains from the latter the solution of the former problem. When investment decisions are modelled to take place in discrete time, the entire portfolio optimization problem reduces to one in discrete time and here a DP-approach under partial information can be found in [19]. A risk-sensitive finite horizon control problem under partial information for a general linear-Gaussian model has been considered in [13] where, by solving two kinds of Riccati differential equations, it was possible to construct an optimal strategy. The results are extended to the case of infinite time horizon in [15] by studying the asymptotics of the solutions of inhomogeneous (time dependent) Riccati differential equations as the time horizon goes to infinity.

In relation to the literature as described above, in the present paper we consider the portfolio maximization problem under a hidden Markov setting, where the coefficients of the security prices are nonlinearly dependent on economic factors that evolve as a k-state Markov chain (Section 2). The problem is reformulated in Section 3 as a risk-sensitive stochastic control problem under complete observation, and in Section 4 an optimal strategy is constructed from the solution of the corresponding HJB-equation.

2. Problem setup

Let us consider a market model with $N+1$ securities $(S_t^0, S_t) := (S_t^0, S_t^1, \ldots, S_t^N)^*$, were S^* stands for the transpose of the matrix S, and an economic factor process X_t, which is supposed to be a finite state Markov chain taking its values in the set of the unit vectors $E = \{e_1, e_2, \ldots, e_k\}$ in R^k. The bond price S_t^0 is assumed to satisfy the ordinary differential equation:

$$dS_t^0 = r(t, S_t) S_t^0 dt, \quad S_0^0 = s^0,$$

where $r(t, S)$ is a nonnegative, bounded and locally Lipschitz continuous function in $S \in R_+^N = \{(x^1, \ldots, x^N); x^i \geq 0, \ i = 1, 2, \ldots, N\}$. The other security prices S_t^i, $i = 1, 2, \ldots, N$, are assumed to be governed by the following stochastic differential equations:

$$\begin{aligned} &dS_t^i = S_t^i \{a^i(t, X_t, S_t)dt + \textstyle\sum_{j=1}^N b_j^i(t, S_t) dW_t^j\}, \\ &S_0^i = s^i, \quad i = 1, \ldots, N, \end{aligned} \tag{2.1}$$

where the $a^i(t, X, S)$ and $b_j^i(t, S)$ are bounded and, for each t and X, locally Lipschitz continuous functions in S, b is uniformly non degenerate, i.e., $z^* bb^* z \geq c|z|^2$, $\forall z \in R^N$, $\exists c > 0$ and $W_t = (W_t^j)_{j=1,\ldots,N}$ is an N-dimensional standard Brownian motion process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and is independent of X_t. The Markov chain X_t can be expressed in terms of a martingale M_t of the pure jump type, namely

$$\begin{aligned} &dX_t = \Lambda(t) X_t dt + dM_t, \\ &X_0 = \xi, \end{aligned}$$

where $\Lambda(t)$ is the Q matrix (transition intensity matrix) of the Markov chain and ξ is a random variable taking its values in E. Set

$$\mathcal{G}_t = \sigma(S_u; u \leq t)$$

and let us denote by h_t^i, $(i = 0, 1, \ldots, N)$ the portfolio proportion of the amount invested in the i-th security relative to the total wealth V_t that the investor possesses. It is defined as follows:

Definition 2.1. $(h_t^0, h_t) \equiv (h_t^0, h_t^1, h_t^2, \ldots, h_t^N)^*$ is said to be an investment strategy if the following conditions are satisfied.

i) h_t is an R^N-valued $\mathcal{G}_t$-progressively measurable stochastic process such that

$$\sum_{i=1}^{N} h_t^i + h_t^0 = 1.$$

ii) $P(\int_0^T |h_s|^2 ds < \infty) = 1.$

The set of all investment strategies will be denoted by $\mathcal{H}(T)$. When $(h_t^0, h_t^*)_{0 \le t \le T} \in \mathcal{H}(T)$, we shall often write $h \in \mathcal{H}(T)$ for simplicity.

For given $h \in \mathcal{H}(T)$, and under the assumption of self-financing, the wealth process $V_t = V_t(h)$ satisfies

$$\begin{cases} \dfrac{dV_t}{V_t} &= \sum_{i=0}^{N} h_t^i \dfrac{dS_t^i}{S_t^i} \\ &= h_t^0 r(t, S_t)dt + \sum_{i=1}^{m} h_t^i \{a^i(t, X_t, S_t)dt + \sum_{j=1}^{N} b_j^i(t, S_t)dW_t^j\} \\ V_0 &= v. \end{cases}$$

Taking into account i) above, V_t turns out to be the solution of

$$\begin{cases} \dfrac{dV_t}{V_t} &= r(t, S_t)dt + h_t^*(a(t, X_t, S_t) - r(t, S_t)\mathbf{1})dt + h_t^* b(t, S_t)dW_t, \\ V_0 &= v, \end{cases}$$

where $\mathbf{1} = (1, 1, \ldots, 1)^*$. Our problem is the following. For a given constant $\mu < 1$, $\mu \neq 0$, maximize the expected (power) utility of terminal wealth up to the time horizon T, namely,

$$J(v; h; T) = \frac{1}{\mu} E[V_T(h)^\mu] = \frac{1}{\mu} E[e^{\mu \log V_T(h)}], \tag{2.2}$$

where h ranges over the set $\mathcal{A}(0, T)$ of all admissible strategies that will be defined below in (3.13).

We consider here the maximization problem with partial information, since the economic factors X_t are, in general, not directly observable and so one has to select the strategies only on the basis of past information of the security prices.

3. Reduction to risk-sensitive stochastic control under complete information

There are a priori more possible approaches to determine an equivalent complete observation control problem. One may base it on a Zakai-type equation for an unnormalized filter. One may however also base it on normalized filters. Each approach has its advantages and disadvantages, the major advantage for the Zakai-type approach being that the dynamics are linear. In Subsection 3.1 we first discuss such an approach in a form related to [13] and show that, in our setting, an explicit solution is difficult to obtain despite the linearity of the dynamics for the unnormalized filter. Although we therefore abandon this approach in favour of one based on normalized filter values, we still wanted to discuss it here because it forms a basis for the other approach that will be derived in Subsection 3.2 and that is related to [13] and [15]. We want to point out that, in the given setup, the standard approach leading to the so-called "separated problem" fails because of questions of measurability with respect to the full and the observation filtrations and the fact that in a crucial expectation there appears the product of the function of interest with a Radon-Nikodym derivative (see (3.5) and the comment preceding (3.6)).

Before discussing the individual approaches, let us introduce some notation and expressions that will be used in the sequel.

Let us set

$$Y_t^i = \log S_t^i, \quad i = 0, 1, 2, \ldots, N,$$

with $Y_t = (Y_t^1, Y_t^2, \ldots, Y_t^N)^*$ and $\mathbf{e}^Y = (e^{Y^1}, \ldots, e^{Y^N})^*$. Then

$$dY_t^0 = R(t, Y_t)dt$$

and

$$dY_t = \bar{A}(t, X_t, Y_t)dt + B(t, Y_t)dW_t, \tag{3.1}$$

where

$$\begin{array}{lll} \bar{A}^i(t,x,y) & = & a^i(t,x,\mathbf{e}^y) - \frac{1}{2}(bb^*)^{ii}(t,\mathbf{e}^y), \\ B_j^i(t,y) & = & b_j^i(t,\mathbf{e}^y), \;\; R(t,y) = r(t,\mathbf{e}^y). \end{array}$$

Putting

$$\eta(t,x,y,h) := \frac{1-\mu}{2}h^*BB^*(t,y)h - R(t,y) - h^*(A(t,x,y) - R(t,y)\mathbf{1}), \tag{3.2}$$

with

$$A^i(t,x,y) = a^i(t,x,\mathbf{e}^y),$$

by Itô's formula we see that

$$dV_t^\mu = V_t^\mu\{-\mu\eta(t,X_t,Y_t,h_t)dt + \mu h_t^* B(t,Y_t)dW_t\}, \;\; V_0 = v^\mu, \tag{3.3}$$

and so

$$\begin{aligned} V_t^\mu = v^\mu \exp\{-\mu \textstyle\int_0^t \eta(s,X_s,Y_s,h_s)ds \\ +\mu \textstyle\int_0^t h_s^* B(s,Y_s)dW_s - \frac{\mu^2}{2}\int_0^t h_s^* BB^*(s,Y_s)h_s ds\}. \end{aligned}$$

3.1. Approach via a Zakai-type equation

Given our assumptions on the boundedness of the coefficients, let us introduce a new probability measure $\hat{P}$ on $(\Omega, \mathcal{F})$ defined by

$$\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{F}_T} = \rho_T,$$

where

$$\rho_T = e^{-\int_0^T \bar{A}^*(t,X_t,Y_t)(BB^*)^{-1}B(t,Y_t)dW_t - \frac{1}{2}\int_0^T \bar{A}^*(BB^*)^{-1}\bar{A}(t,X_t,Y_t)dt}.$$

Under the probability measure $\hat{P}$,

$$\hat{W}_t = W_t + \int_0^t B^*(BB^*)^{-1}(s, Y_s)\bar{A}(s, X_s, Y_s)ds$$

is a Brownian motion process and Y_t satisfies

$$dY_t = B(t, Y_t)d\hat{W}_t. \tag{3.4}$$

The criterion (2.2) can be rewritten under the new probability measure as

$$\begin{aligned}
&\tfrac{1}{\mu}E[V_T^{\mu}] \\
&= \tfrac{1}{\mu}v^{\mu}\hat{E}[e^{-\mu\int_0^T \eta(s,X_s,Y_s,h_s)ds+\mu\int_0^T h_s^* B(s,Y_s)dW_s - \frac{\mu^2}{2}\int_0^T h_s^* BB^*(s,Y_s)h_s ds}\rho_T^{-1}] \\
&= \tfrac{1}{\mu}v^{\mu}\hat{E}[e^{-\mu\int_0^T \eta(s,X_s,Y_s,h_s)ds+\int_0^T Q^*(s,X_s,Y_s,h_s)dY_s - \frac{1}{2}\int_0^T Q^*BB^*Q(s,X_s,Y_s,h_s)ds}]
\end{aligned} \tag{3.5}$$

where

$$Q(t, X_t, Y_t, h_t) = (BB^*(t, Y_t))^{-1}\bar{A}(t, X_t, Y_t) + \mu h_t.$$

Since the argument of the expectation in (3.5) is of the form of a Radon-Nikodym derivative multiplied with the function of interest, we shall treat it as a whole considering the process

$$\begin{aligned}
H_t = \exp\{-\mu\textstyle\int_0^t \eta(s, X_s, Y_s, h_s)ds + \int_0^t Q^*(s, X_s, Y_s, h_s)dY_s \\
-\tfrac{1}{2}\textstyle\int_0^t Q^*BB^*(s, Y_s)Q(s, X_s, Y_s, h_s)ds\}
\end{aligned} \tag{3.6}$$

and

$$q_t^i = \hat{E}[H_t X_t^i | \mathcal{G}_t],$$

where $X_t^i = \mathbf{1}_{\{e_i\}}(X_t)$. Then

$$E\{V_T^{\mu}\} = v^{\mu}\hat{E}\{\hat{E}[H_T|\mathcal{G}_T]\} = v^{\mu}\sum_{i=1}^{k}\hat{E}\{\hat{E}[H_T X_T^i|\mathcal{G}_T]\} = v^{\mu}\hat{E}\{\sum_{i=1}^{k} q_T^i\} \tag{3.7}$$

where (see Corollary 3.3 in [1]; see also Section 7.3 in [4]) the q_t^i satisfy

$$\begin{aligned}
dq_t^i &= (\Lambda(t)q_t)^i dt - \mu\eta(t, e_i, Y_t, h_t)q_t^i dt + q_t^i Q^*(t, e_i, Y_t, h_t)dY_t, \\
q_0^i &= p_0^i \equiv P(\xi = e_i), \quad i = 1, 2, \ldots, k.
\end{aligned} \tag{3.8}$$

Next we give some arguments to show that, as mentioned in the introduction, an explicit solution to the problem (3.7) and (3.8) is difficult to obtain.

Set $q_t = (q_t^i)$. Then (q_t, Y_t) can be regarded as the controlled process for the stochastic control problem of maximizing the criterion

$$J = v^{\mu}\hat{E}\{\sum_{i=1}^{k} q_T^i\}.$$

Let us introduce the value function

$$w(t, q, y) = \sup_{h\in\mathcal{A}(t,T)} \hat{E}\{\sum_{i=1}^{k} q_T^i(t)\}$$

where, analogously to $\mathcal{A}(0,T)$, $\mathcal{A}(t,T)$ denotes the admissible strategies over the interval $[t,T]$, $q^i_s(t)$, $t \le s \le T$ is a solution of (3.8) with the initial condition $q^i_t(t) = q^i$ and Y_s, $t \le s \le T$, is a solution of (3.4) with initial condition $Y_t = y$.

The Bellman equation for w then becomes

$$\begin{cases} \frac{\partial w}{\partial s} + \sup_h L_s(h)w = 0\,, \qquad t \le s \le T, \quad (q,y) \in [0,\infty)^k \times R^N, \\ w(T,q,y) = \sum_{i=1}^k q^i, \end{cases}$$

where

$$\begin{cases} L_s(h) = \frac{1}{2}\sum_{i,j}[BB^*(s,y)]^{ij}\frac{\partial^2}{\partial y^i \partial y^j} + \sum_{i,j} q^i[Q^*(s,e_i,y,h)B(s,y)]^j \frac{\partial^2}{\partial q^i \partial y^j} \\ \qquad\qquad + \frac{1}{2}\sum_{i,j} q^i Q^*(s,e_i,y,h)BB^*Q(s,e_j,y,h)q^j \frac{\partial^2}{\partial q^i \partial q^j} \\ \qquad\qquad + \sum_i\{[q^*\Lambda(s)^*]^i - \mu\eta(s,e_i,y,h)q^i\}\frac{\partial}{\partial q^i}. \end{cases}$$

As can now be easily seen, an explicit solution of this Bellman equation is rather difficult to obtain and so we abandon this approach in favour of one based on the normalized filter that will however continue the main line of the arguments of the present section.

3.2. Approach based on the normalized filter

In order to derive the corresponding full information control problem we put

$$p^i_t = P(X_t = e_i|\mathcal{G}_t), \quad i = 1,\dots,k, \tag{3.9}$$

and use the notation

$$f(s,p_s,y,h) = \sum_{i=1}^k f(s,e_i,y,h)p^i_s, \tag{3.10}$$

for a given function $f(s,x,y,h)$ on $[0,T]\times E\times R^N\times R^N$, while the defined function is on $[0,T]\times\Delta_{k-1}\times R^N\times R^N$ with Δ_{k-1} the $(k-1)$-dimensional simplex

$$\Delta_{k-1} = \{(d_1,d_2,\dots,d_k); d_1+d_2+\cdots+d_k = 1,\ 0\le d_i\le 1,\ i=1,\dots,k\}.$$

It is known that these (normalized) conditional probabilities p^i_t, $i = 1,2,\dots,k$, satisfy the equation ("Wonham filter", see [11, 21])

$$\begin{aligned} dp^i_t = (\Lambda(t)p_t)^i dt + p^i_t[\bar{A}^*(t,e_i,Y_t) - \bar{A}^*(t,p_t,Y_t)] \\ \cdot[BB^*(t,Y_t)]^{-1}[dY_t - \bar{A}(t,p_t,Y_t)dt], \end{aligned}$$

namely,

$$\begin{aligned} dp_t = \Lambda(t)p_t dt + D(p_t)[\bar{A}^*(t,Y_t) - \mathbf{1}\bar{A}^*(t,p_t,Y_t)] \\ \cdot[BB^*(t,Y_t)]^{-1}[dY_t - \bar{A}(t,p_t,Y_t)dt], \end{aligned} \tag{3.11}$$

where $\bar{A}(t,Y)$ is an $(N\times k)$-matrix defined by $\bar{A}(t,Y) = (\bar{A}^i(t,e_j,Y))$ and $D(p)$ is a diagonal matrix of which the component in position ii is p^i.

In full analogy with (3.6) we now define

$$\begin{aligned} \hat{H}_t = \exp\{-\mu\int_0^t \eta(s,p_s,Y_s,h_s)ds + \int_0^t Q^*(s,p_s,Y_s,h_s)dY_s \\ -\frac{1}{2}\int_0^t Q^*BB^*(s,Y_s)Q(s,p_s,Y_s,h_s)ds\}. \end{aligned} \tag{3.12}$$

We then have

$$\begin{aligned}
d(\hat{H}_t p_t^i) &= \hat{H}_t dp_t^i + p_t^i d\hat{H}_t + d\langle \hat{H}, p^i \rangle_t \\
&= \hat{H}_t (\Lambda(t) p_t)^i dt \\
&\quad + \hat{H}_t p_t^i [\bar{A}^*(t, e_i, Y_t) - \bar{A}^*(t, p_t, Y_t)][BB^*(t, Y_t)]^{-1}[dY_t - \bar{A}(t, p_t, Y_t)dt] \\
&\quad - \mu \hat{H}_t p_t^i \eta(t, p_t, Y_t, h_t)dt + \hat{H}_t p_t^i Q^*(t, p_t, Y_t, h_t) dY_t + d\langle \hat{H}, p^i \rangle_t \\
&= (\Lambda(t) \hat{H}_t p_t)^i dt - \mu \eta(t, e_i, Y_t, h_t) \hat{H}_t p_t^i dt + \hat{H}_t p_t^i Q^*(t, e_i, Y_t, h_t) dY_t,
\end{aligned}$$

where the last equality is obtained from noticing that, given the previous definitions, the following three equalities hold:

$$\begin{aligned}
& d\langle \hat{H}, p^i \rangle_t = \hat{H}_t p_t^i [\bar{A}(t, e_i, Y_t)^* - \bar{A}(t, p_t, Y_t)^*][BB^*]^{-1} \bar{A}(t, p_t, Y_t) dt \\
& \qquad\qquad\qquad + \hat{H}_t p_t^i \mu h_t^* [\bar{A}(t, e_i, Y_t)^* - \bar{A}(t, p_t, Y_t)^*] dt; \\
& -\mu \eta(t, e_i, Y_t, h_t) \hat{H}_t p_t^i dt + \mu \eta(t, p_t, Y_t, h_t) \hat{H}_t p_t^i dt \\
& \qquad\qquad\qquad = \hat{H}_t p_t^i \mu h_t^* [\bar{A}(t, e_i, Y_t)^* - \bar{A}(t, p_t, Y_t)^*] dt; \\
& \hat{H}_t p_t^i [\bar{A}(t, e_i, Y_t)^* - \bar{A}(t, p_t, Y_t)^*][BB^*]^{-1}[dY_t - \bar{A}(t, p_t, Y_t)dt] \\
& \qquad\qquad\qquad + \hat{H}_t p_t^i Q(t, p_t, Y_t, h_t)^* dY_t \\
& = \hat{H}_t p_t^i Q^*(t, e_i, Y_t, h_t) dY_t - \hat{H}_t p_t^i [\bar{A}(t, e_i, Y_t)^* - \bar{A}(t, p_t, Y_t)^*][BB^*]^{-1} \bar{A}(t, p_t, Y_t).
\end{aligned}$$

Therefore, we see that $q_t^i = \hat{H}_t p_t^i$, thus showing that q_t^i are indeed un-normalized conditional probabilities and

$$\hat{E}[H_T | \mathcal{G}_T] = \sum_{i=1}^{k} q_T^i = \hat{H}_T.$$

We have thus proved the following proposition, which establishes the equivalence of the original incomplete information control problem with the present corresponding complete one. The latter has as state variable process the (finite-dimensional) Markovian pair (p_t, Y_t) satisfying (3.11) and (3.4) respectively, and as objective function $\frac{1}{\mu} v^\mu \hat{E}[\hat{H}_T]$, where $\hat{H}_T$ depends, see (3.12), on the chosen strategy h_t.

Proposition 3.1. *The criterion* (2.2) *can be expressed equivalently as*

$$J(v; h; T) \equiv \frac{1}{\mu} E[V_T^\mu] = \frac{1}{\mu} v^\mu \hat{E}[H_T] = \frac{1}{\mu} v^\mu \hat{E}[\hat{H}_T].$$

Notice that, for Markovianity, we have to consider as state variables in the complete observation problem the pair (p_t, Y_t) and not just p_t alone, because in our original problem the coefficients depend on S_t and therefore on Y_t. Notice also that the state-variable pair (p_t, Y_t) is finite-dimensional.

The criterion expressed in the rightmost equivalent form above can be shown to be of the form of a risk-sensitive stochastic control problem in finite dimension. To this effect let us introduce another change of measure with the Girsanov density defined by

$$\left.\frac{d\tilde{P}}{d\hat{P}}\right|_{\mathcal{G}_T} = \zeta_T \quad = \quad e^{\int_0^T Q^*(s,p_s,Y_s,h_s)dY_s - \frac{1}{2}\int_0^T Q^*BB^*Q(s,p_s,Y_s,h_s)ds}$$
$$= \quad e^{\int_0^T Q^*(s,p_s,Y_s,h_s)B(s,Y_s)d\hat{W}_s - \frac{1}{2}\int_0^T Q^*BB^*Q(s,p_s,Y_s,h_s)ds}.$$

Notice that the new probability measure $\tilde{P}$ depends, through ζ_T, on the chosen strategy h_t. In order that $\tilde{P}$ is a probability measure we have to require that the set $\mathcal{A}(0,T)$ of admissible strategies is given by

$$\mathcal{A}(0,T) = \left\{ h \in \mathcal{H}(T) \mid \hat{E}\{\zeta_T\} = E\{\rho_T \zeta_T\} = 1 \right\}. \tag{3.13}$$

Under the probability measure $\tilde{P}$ we now have that

$$\tilde{W}_t = \int_0^t B^{-1}(s,Y_s)dY_s - \int_0^t B^*(s,Y_s)Q(s,p_s,Y_s,h_s)ds$$

is a standard $\mathcal{G}_t$-Brownian motion process and we have

$$\begin{aligned} dY_t &= B(t,Y_t)d\tilde{W}_t + BB^*(t,Y_t)Q(t,p_t,Y_t,h_t)dt \\ &= B(t,Y_t)d\tilde{W}_t + \{\bar{A}(t,p_t,Y_t) + \mu BB^*(t,Y_t)h_t\}dt \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} dp_t = \;& D(p_t)[\bar{A}^*(t,Y_t) - \mathbf{1}\bar{A}^*(t,p_t,Y_t)][BB^*(t,Y_t)]^{-1}B(t,Y_t)d\tilde{W}_t \\ & + \{\Lambda(t)p_t + \mu D(p_t)[\bar{A}^*(t,Y_t) - \mathbf{1}\bar{A}^*(t,p_t,Y_t)]h_t\}dt. \end{aligned} \tag{3.15}$$

Since

$$\frac{1}{\mu}v^\mu \hat{E}[\hat{H}_T] = \frac{1}{\mu}v^\mu \tilde{E}[\exp\{-\mu\int_0^T \eta(s,p_s,Y_s,h_s)ds\}],$$

we are reduced to considering the risk-sensitive stochastic control problem that consists in maximizing

$$\frac{1}{\mu}v^\mu \tilde{E}[\exp\{-\mu\int_0^T \eta(s,p_s,Y_s,h_s)ds\}] \tag{3.16}$$

subject to the controlled process (p_t, Y_t) on $\Delta_{k-1} \times R^N$ being governed by the controlled stochastic differential equations (3.15) and (3.14) defined on the filtered probability space $(\Omega, \mathcal{F}, \mathcal{G}_t, \tilde{P})$.

The solution to this latter complete observation problem forms the subject of the next Section 4.

4. HJB-equation

For ease of notation, given $t \in [0,T]$, let us now introduce for $s \in [t,T]$ the vector process

$$Z_s := [p_s, Y_s]^*, \quad p_s \in \Delta_{k-1},\; Y_t \in R^N,$$

so that, putting

$$\beta(s, Z_s) := \begin{bmatrix} \Lambda(s)p_s \\ \bar{A}(s, p_s, Y_s) \end{bmatrix}, \quad \text{a } (k+N)\text{-vector}$$

$$\alpha(s, Z_s) := \begin{bmatrix} D(p_s)[\bar{A}^*(s, Y_s) - \mathbf{1}\bar{A}^*(s, p_s, Y_s)](BB^*)^{-1}B(s, Y_s) \\ B(s, Y_s) \end{bmatrix},$$

which is a $(k+N) \times N)$-matrix and

$$\beta_\mu(s, Z_s; h_s) := \beta(s, Z_s) + \mu\alpha(s, Z_s)B^*(s, Y_s)h_s, \quad \text{a } (k+N)\text{-vector}, \tag{4.1}$$

from (3.15) and (3.14) the dynamics of Z_s on $(\Omega, \mathcal{F}, \mathcal{G}_s, \tilde{P})$ and for $s \in [t, T]$ become

$$\begin{cases} dZ_s &= \beta_\mu(s, Z_s; h_s)ds + \alpha(s, Z_s)d\tilde{W}_s \\ Z_t &= z, \end{cases} \tag{4.2}$$

where the strategy h_s affects the evolution of Z_s directly through the drift β_μ and, recalling the comment before (3.13), indirectly also through the measure $\tilde{P}$, i.e., through $\tilde{W}_s$.

Recall now the objective function (2.2) and its representation in Proposition 3.1 and in (3.16) that are all defined for the initial time $t = 0$. For a generic t with $0 \le t \le T$ and for $V_t = v$, $Z_t = z$, put

$$J(t; v; z, h; T) = \frac{1}{\mu} v^\mu G(t, z, h)$$

where, letting with some abuse of notation $\eta(s, Z_s, h_s) := \eta(s, p_s, Y_s, h_s)$ with $\eta(s, p_s, Y_s, h_s)$ as in (3.2) and with the notation as in (3.10), we define

$$G(t, z, h) = \tilde{E}_{t,z}\left\{\exp\left[-\mu\int_t^T \eta(s, Z_s, h_s)\, ds\right]\right\}.$$

In view of the HJB equation put now

$$w(t, z) := \sup_{h \in \mathcal{A}(t,T)} \log G(t, z, h) \tag{4.3}$$

so that

$$\sup_{h \in \mathcal{A}(0,T)} J(v; h; T) = \frac{1}{\mu} v^\mu e^{w(0, Z_0)}.$$

Based on the definition of $\eta(t, z, h)$ and the dynamics of Z in (4.2) with drift β_μ as in (4.1), we may now formally write for $w(t, z)$ in (4.3) the following Bellman equation of the dynamic programming approach,

$$\begin{cases} \frac{\partial w}{\partial t} + \frac{1}{2}tr[\alpha\alpha^* D^2 w] + \frac{1}{2}(\nabla w)^*\alpha\alpha^*\nabla w \\ \quad + \sup_h \left[\beta_\mu(t, z, h)^*\nabla w + \mu\gamma^*(t, z)h - \frac{1}{2}\mu(1-\mu)h^*BB^*h\right] \\ \quad + \mu R(t, z) = 0, \\ w(T, z) = 0, \end{cases} \tag{4.4}$$

where

$$\gamma(t, z) = A(t, p, Y) - R(t, z)\mathbf{1}.$$

Given our assumptions that b is uniformly non-degenerate, the maximizing $\hat{h}$ in (4.4) is

$$\hat{h} = \hat{h}(t,z) = \frac{1}{1-\mu}\,(BB^*)^{-1}(t,z)\;[B(t,z)\alpha^*(t,z)\nabla w(t,z) + \gamma(t,z)] \tag{4.5}$$

and (4.4) itself becomes

$$\begin{cases} \frac{\partial w}{\partial t} + \frac{1}{2}tr[\alpha\alpha^* D^2 w] + \frac{1}{2(1-\mu)}(\nabla w)^*\alpha\alpha^*\nabla w + \Phi^*\nabla w + \Psi = 0 \\ w(T,z) = 0 \end{cases} \tag{4.6}$$

where, for simplicity of notation, we have put

$$\begin{array}{rcl} \Phi(t,z) & := & \beta(t,z) + \frac{\mu}{1-\mu}\alpha(t,z)B^{-1}(t,z)\gamma(t,z) \\ \Psi(t,z) & := & \mu R(t,z) + \frac{\mu}{2(1-\mu)}\gamma^*(t,z)(BB^*)^{-1}(t,z)\gamma(t,z), \end{array}$$

which is a nonlinear second order PDE. We shall now transform (4.6) into a linear PDE by following a by now classical procedure (see, e.g., [6, 7]) and according to which we put

$$v(t,z) = e^{\frac{1}{1-\mu}\,w(t,z)}.$$

With this transformation (4.6) becomes now

$$\begin{cases} \frac{\partial v}{\partial t} + \frac{1}{2}tr[\alpha\alpha^* D^2 v] + \Phi^*(t,z)\nabla v + \frac{\Psi(t,z)}{1-\mu}v = 0 \\ v(T,z) = 1. \end{cases} \tag{4.7}$$

It can now be easily seen that $v(t,z)$ is a viscosity solution for (4.7) if and only if $w = (1-\mu)\log v$ is a viscosity solution for (4.6).

Notice that, in spite of the linearity of the PDE in (4.7), an explicit analytic solution is very difficult to obtain in our setting (to this effect see also the Remark 4.2 at the end of this section). However, the linearity of the PDE leads to a Feynman-Kac representation of the solution, which makes it then possible to compute it numerically by simulation as we shall mention also below. Set then

$$\bar{v}(t,z) = E_{t,z}\left\{\exp\left[\frac{1}{1-\mu}\int_t^T \Psi(s,Z_s)\,ds\right]\right\} \tag{4.8}$$

where Z_s now satisfies, instead of (4.2), the following:

$$\begin{cases} dZ_s & = & \Phi(s,Z_s)dt + \alpha(s,Z_s)dW_s \\ Z_t & = & z \end{cases} \tag{4.9}$$

where W_s is a Wiener process and which, given our assumptions of bounded and locally Lipschitz continuous coefficients with b uniformly non degenerate, admits a unique strong/pathwise solution. A solution to this equation can rather easily be simulated for the purpose of calculating then numerically the value of $\bar{v}(t,z)$.

Finally, using also the boundedness of $\bar{v}$, from Theorem 4.4.3 and Appendix 7.7.2 in [14] it follows that $\bar{v}(t,z)$ is the unique viscosity solution for (4.7) and, consequently, $\bar{w} = (1-\mu)\log\bar{v}$ is the unique viscosity solution for (4.6). Thus we have the following proposition.

Proposition 4.1. *Under the assumptions in Section* 2, *Equation* (4.6) *has a unique viscosity solution* w *and it is expressed as* $w(t,z) = (1-\mu)\log\bar{v}$, *where* $\bar{v}$ *is the function defined by* (4.8).

Under stronger assumptions on r, a^i, b^i_j such that they are $\mathcal{C}^2$ functions with derivatives of polynomial growth we have by Theorem 5.5 in [2] that $\bar{v}(t,z)$, and therefore also $\bar{w}(t,z)$, are of class $\mathcal{C}^2$ and with derivatives of polynomial growth. The formal Bellman equation (4.4) becomes thus an equation having a classical solution and the function $\hat{h}$ in (4.5) exists and $\hat{h}(t,Z_t)$ is thus an optimal control.

We close this section with the following remark that is intended to better explain why an explicit analytic solution to (4.7) is difficult to obtain.

Remark 4.2. We show here the expressions for the coefficients of the HJB equation (4.7) in the simplest case when the coefficients in the asset price dynamics (2.1) are autonomous and do not depend on the asset price itself and the factor process X_t is a two-state homogeneous Markov process with $Q-$matrix

$$\Lambda^* = \begin{pmatrix} \lambda_1 & -\lambda_2 \\ -\lambda_1 & \lambda_2 \end{pmatrix}.$$

Denote by p_t the conditional state probability for state 1 in the generic period t, i.e.,

$$p_t = P\{X_t = e_1 | \mathcal{G}_t\}.$$

We have now

$$\alpha\alpha^*(p)$$
$$= \begin{pmatrix} p^2(1-p^2)(a(e_1)-a(e_2))^2B^{-2} & -p^2(1-p^2)(a(e_1)-a(e_2))^2B^{-2} & p(1-p)(a(e_1)-a(e_2)) \\ -p^2(1-p^2)(a(e_1)-a(e_2))^2B^{-2} & p^2(1-p^2)(a(e_1)-a(e_2))^2B^{-2} & p(1-p)(a(e_2)-a(e_1)) \\ (a(e_1)-a(e_2))p(1-p) & (a(e_2)-a(e_1))p(1-p) & B^2 \end{pmatrix}$$

$$\Phi(p) = \begin{pmatrix} \lambda_1 p - \lambda_2(1-p) \\ -\lambda_1 p + \lambda_2(1-p) \\ a(e_1)p + a(e_2)(1-p) - \frac{1}{2}B^2 \end{pmatrix}$$
$$+\frac{\mu}{1-\mu}\begin{pmatrix} p(1-p)(a(e_1)-a(e_2))B^{-2}(a(e_1)p+a(e_2)(1-p)-R) \\ p(1-p)(a(e_2)-a(e_1))B^{-2}(a(e_1)p+a(e_2)(1-p)-R) \\ a(e_1)p+a(e_2)(1-p)-R \end{pmatrix}$$

$$\Psi(p) = \mu R + \frac{\mu}{1-\mu}\left[a(e_1)p + a(e_2)(1-p) - R\right]^2 B^{-2},$$

and from here it can be seen that, even in this simple case, an explicit solution of the HJB equation (4.7) is difficult to obtain.

5. Conclusions and computational remarks

Given our expected utility maximization problem for a power utility of the risk averse type, where the coefficients in the asset price dynamics are driven by a hidden finite state Markov process representing "economic factors", we have first discussed a corresponding complete observation control problem based on unnormalized conditional probabilities (unnormalized filter) satisfying a linear Zakai-type equation and shown that for this problem it is difficult to obtain an explicit solution. We have then studied an equivalent complete observation problem based on normalized filter values. For this problem we have studied the corresponding HJB equation that has been shown to admit a unique viscosity solution that can be computed as an expectation according to (4.8) and (4.9). Under sufficient regularity assumptions this solution has enough regularity so that an optimal investment strategy exists and can be computed from the solution of the HJB equation according to (4.5). This strategy is a function of the process $Z_s = [p_s, Y_s]^*$ formed by the pair consisting of the filter p_s in (3.9) for the unobserved factor process X_s and the log-prices Y_s, all of which are accessible to the economic agent.

Since a solution can be obtained in the form of an expectation according to (4.8) and (4.9), it can in general be computed by Monte Carlo simulation. This is important since, as discussed in Section 4, also for the complete observation problem based on normalized filter values an analytic solution is very difficult to obtain.

References

[1] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models*, Springer-Verlag, New York, 1995.

[2] A. Friedman, *Stochastic Differential Equations and Applications*, Academic Press, 1975.

[3] G. Gennotte, *Optimal portfolio choice under incomplete information*, J. of Finance, **41** (1986), 733–746.

[4] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer Verlag, 1985.

[5] U.G. Haussmann and J. Sass, *Optimal terminal wealth under partial information for HMM stock returns*, Contemporary Mathematics, AMS, **351** (2004), 171–185.

[6] H. Ishii, H. Nagai, and F. Teramoto, *A singular limit on risk sensitive control and semiclassical analysis*, in: Proceedings of the 7th Japan-Russia Symp. on Prob. and Math. Stat., World Scientific, (1996), 164–173.

[7] H. Kaise and H. Nagai, *Bellman-Isaacs equations of ergodic type related to risk-sensitive control and their singular limits*, Asymptotic Analysis, **16** (1998), 347–362.

[8] I. Karatzas and X. Zhao, *Bayesian adaptive portfolio optimization*, in: Handbk. Math. Finance: Option Pricing, Interest Rates and Risk Management, E. Jouini, J. Cvitanić, and M.Musiela, Editors, Cambridge Univ. Press, (2001), 632–669.

[9] P. Lakner, *Utility maximization with partial information*, Stochastic Processes and their Applications, **56 (2)** (1995), 247–273.

[10] P. Lakner, *Optimal trading strategy for an investor: the case of partial information*, Stochastic Processes and their Applications, **76** (1998), 77–97.

[11] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes: I. General Theory*, Springer-Verlag, Berlin, 1977.

[12] R. C. Merton, *An intertemporal capital asset pricing model*, Econometrica, **41** (1973), 867–887.

[13] H. Nagai, *Risk-senstive dynamic asset management with partial information*, Stochastics in Finite and Infinite Dimension, a volume in honor of G. Kallianpur, Rajput et al., Editors, Birkhäuser, (2000), 321–340.

[14] H. Nagai, *Stochastic Differential Equations*, Kyoritsu Shuppan, 1999.

[15] H. Nagai and S. Peng, *Risk-sensitive dynamic portfolio optimization with partial information on infinite time horizon*, Annals of Applied Probability, **12 (1)** (2002), 173–195.

[16] H. Pham and M.-C. Quenez, *Optimal portfolio in partially observed stochastic volatility models*, The Annals of Applied Probability, **11 (1)** (2001), 210–238.

[17] E. Platen and W. J. Runggaldier, *A benchmark approach to filtering in finance*, Asia Pacific Financial Markets, **11 (1)** (2005), 79–105.

[18] R. Rishel, *Optimal portfolio management with partial observations and power utility function*, in: Stochastic Analysis, Control, Optimization and Applications: Volume in Honour of W. H. Fleming, W. McEneany, G. Yin and Q. Zhang, Editors, Birkhäuser, (1999), 605–620.

[19] W. J. Runggaldier and A. Zaccaria, *A stochastic control approach to risk management under restricted information*, Math. Finance, **10** (2000), 277–288.

[20] J. Sass and U. G. Haussmann, *Optimizing the terminal wealth under partial information: The drift process as a continuous time Markov chain*, Finance and Stochastics, **8** (2004), 553–577.

[21] W. M. Wonham, *Some applications of stochastic differential equations to optimal nonlinear filtering*, SIAM J. Control Opt., **2** (1965), 347–369.

[22] G. Zohar, *A generalized Cameron-Martin formula with applications to partially observed dynamic portfolio optimization*, Math. Finance, **11** (2001), 475–494.

Hideo Nagai
Department of Mathematical Science for Social Systems
Graduate School of Engineering Science
Osaka University
Toyonaka, 560-8531
Japan
e-mail: `nagai@sigmath.es.osaka-u.ac.jp`

Wolfgang J. Runggaldier
Dipartimento di Matematica Pura ed Applicata
Universita degli Studi di Padova
Via Trieste, 63
35121 Padova
Italy
e-mail: `runggal@math.unipd.it`

Progress in Probability, Vol. 59, 507–519

Generalizations of Merton's Mutual Fund Theorem in Infinite-Dimensional Financial Models

Maurizio Pratelli

Abstract. This is a review paper, concerning some extensions of the celebrated *Merton's mutual fund theorem* in infinite-dimensional financial models, in particular, the so-called *Large Financial Markets* (where a sequence of assets is taken into account) and *Bond Markets Models* (where there is a continuum of assets).

In order to obtain these results, an infinite-dimensional stochastic integration theory is essential: the paper illustrates briefly a new theory introduced to this extent by M. De Donno and the author.

Mathematics Subject Classification (2000). Primary 91B28; Secondary 60H05, 91B70.

Keywords. Mutual fund theorem, utility maximization, admissible strategies, stochastic models for financial markets, infinite-dimensional stochastic integration.

1. Introduction

The *Mutual Fund Theorem* (also called the *separation theorem*) is a central result in the problem of maximizing the investor's expected utility of the terminal wealth of a portfolio of risky and riskless assets. It states that (under suitable assumptions) the investor's allocation decision can be separated in two steps.

In the first step, an efficient portfolio of risky assets is determined (the mutual fund); and in the second step the investor decides the allocation between this efficient portfolio and the riskless asset. The efficient portfolio is identical for all investors regardless their attitude towards risk, as reflected by their utility functions.

Before introducing the results, let us fix some notation.

This work is the result of discussions and collaboration with Marzia De Donno.

We indicate by $\mathbf{S}_t = \big(S_t^0, \ldots, S_t^n\big)_{0\le t\le T}$ the available assets on the market. We suppose that the riskless asset S_t^0 is always equal to 1: this simplifies the exposition, since it avoids the introduction of the riskless interest rate, and is not restrictive (this simply means that we consider discounted prices).

The risky assets $(S_t^i)_{0\le t\le T}$ are supposed to be *semimartingales* adapted to some filtration $\big(\mathcal{F}_t\big)_{0\le t\le T}$ on a probability space $\big(\Omega, \mathcal{F}, \mathbf{P}\big)$.

The portfolio's strategy $\mathbf{H}_t = (H_t^0, H_t^1, \ldots, H_t^n)$ is an $(n+1)$-dimensional predictable stochastic process such that the vector stochastic integral $\int_0^t \mathbf{H}_s \mathrm{d}\mathbf{S}_s$ is defined: H_t^i represents the number of assets S^i held at time t, and the stochastic integral is the mathematical representation for the gain from trade.

The (discounted) value of the portfolio at time t is the random variable $X_t = \sum_{i=0}^n H_t^i S_t^i$, and the portfolio is said to be *self-financing* if $X_t = X_0 + \int_0^t \mathbf{H}_s \mathrm{d}\mathbf{S}_s$.

An alternative representation of the portfolio's strategy is to consider the $(n+1)$-dimensional stochastic process $\mathbf{u}_t$ where u_t^i is the *proportion* of the capital invested in the asset i. The process $\mathbf{u}_t$ is also called the *relative portfolio.*

One has evidently

$$u_t^i \quad = \quad \frac{H_t^i S_t^i}{X_t} \quad = \quad \frac{H_t^i S_t^i}{\sum_{j=0}^n H_t^j S_t^j}.$$

This representation of the strategy is suitable when using control techniques: since $\sum_{i=0}^n u_t^i = 1$, it is convenient to consider $(u_t^1, \ldots, u_t^n)$ as a *free* control and consequently $u_t^0 = 1 - \sum_{i=1}^n u_t^i$.

In order to keep the exposition as simple as possible, we restrict ourselves to the problem of maximizing the expected utility from terminal wealth (more generally, one can consider the problem of maximizing the utility from consumption and terminal wealth, take into account restrictions on the allowed strategies...).

More precisely, we consider an utility function $U : \mathbb{R} \to [-\infty, +\infty[$, and, given an initial endowment x, the problem is to maximize $\mathbb{E}\big[U(X_T)\big]$ over all possible random variables X_T , where X_T is the value at time T of a self-financing portfolio with $X_0 = x$.

We consider the case where $U(x) = -\infty$ for $x < 0$ (negative wealth is not allowed), and for positive x, the function U satisfies the so-called Inada conditions: it is strictly increasing, strictly concave, continuously differentiable and $U'(0) = \lim_{x\to 0+} U'(x) = +\infty$, $U'(+\infty) = \lim_{x\to+\infty} U'(x) = 0$.

After previous results by Markowitz in the context of a single period model (see [26]), the continuous time version was proved by Merton ([28, 29]) in the case where asset prices are diffusion processes with constant drift and volatility coefficients: many extensions were subsequently given in terms of various incomplete markets (and with constraints on the strategies) by several authors. See, for instance, [6, 20, 21, 22].

Section 2 of the present review paper gives an outline of Merton's original method (without a complete proof) and Section 3 gives (almost as an exercise) an alternative proof based on stochastic integral representation of martingales in a Brownian filtration.

Section 4 introduces infinite-dimensional models for financial markets and a theory of stochastic integration explicitely developed for the investigation of these models, while Section 5 exposes some extensions of the Mutual Fund theorem.

2. An outline of the classical proof

In this Section, we give an outline of Merton's classical proof (based on stochastic control methods), closely following the presentation given by Bjork (see [2, Chapter 19]).

The level of this section is heuristic: besides the original papers [28] and [29], the interested reader can find an accurate presentation of Merton's results (together with a concise introduction to stochastic optimal control) in the quoted book by Biork.

According to the model of Samuelson–Merton–Black–Scholes, the risky assets are supposed to satisfy the equation

$$\mathrm{d}S_t^i \;=\; S_t^i\Big(\mu^i\,\mathrm{d}t \,+\, \sum_{j=1}^{n}\sigma_{ij}\,\mathrm{d}W_t^j\Big) \tag{2.1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ is a vector of $\mathbb{R}^n$, $\mathbf{W} = (W^1, \ldots, W^n)$ is an n-dimensional Wiener process and $\boldsymbol{\sigma} \;=\; \big[\sigma_{i,j}\big]_{i,j=1,\ldots,n}$ is a $n \times n$ invertible matrix: under these assumptions, the model is *arbitrage free* and *complete.*

By using (as in the previous section) the *relative portfolio* $\mathbf{u}_t = (u_t^1, \ldots, u_t^n)$ as a *control*, the equation of the corresponding porfolio value is

$$\mathrm{d}X_t^{\mathbf{u}} \;=\; X^{\mathbf{u}}\left(\mathbf{u}_t.\boldsymbol{\mu}\,\mathrm{d}t \,+\, \boldsymbol{\sigma}^*\mathbf{u}_t.\mathrm{d}\mathbf{W}_t\right). \tag{2.2}$$

Therefore, $X_t^{\mathbf{u}}$ is a *diffusion process* with *infinitesimal generator*

$$\mathcal{A}_t^{\mathbf{u}} \quad=\quad x\,\mathbf{u}.\boldsymbol{\mu}\,\frac{\partial}{\partial x} \,+\, \frac{x^2}{2}\,\|\boldsymbol{\sigma}^*\mathbf{u}\|^2\,\frac{\partial^2}{\partial x^2}.$$

As it is usual in stochastic optimal control, one considers the *optimal value function*

$$V(t,x) \;=\; \sup_{\mathbf{u}\in\mathcal{U}}\mathbb{E}\Big[U\Big(X_T^{t,x,\mathbf{u}}\Big)\Big]$$

where $\mathcal{U}$ is the class of admissible controls (in this case, *all* controls) and $X^{t,x,\mathbf{u}}$ is the process which starts from x at time t and follows the dynamics given by (2.2).

Under suitable assumptions (obviously satisfied in this simple model with constant coefficients) the function V is the solution of the Hamilton–Jacobi–Bellman equation

$$\begin{cases} \frac{\partial V}{\partial t} + \sup_{\mathbf{u}\in\mathbb{R}^d}\Big[\mathcal{A}^{\mathbf{u}}V(t,x)\Big] \;=\; 0 \\ V(T,x) \;=\; U(x). \end{cases}$$

Handling the HJB equation in practice, is given in two steps:

- given (t,x) and the function V, find $\hat{\mathbf{u}}(t,x,V)$ solution of
$$\mathcal{A}^{\hat{\mathbf{u}}}V(t,x) \;=\; \max_{\mathbf{u}\in\mathbb{R}^n}\left[\mathcal{A}^{\mathbf{u}}V(t,x)\right],$$
- solve the equation
$$\begin{cases} \frac{\partial V}{\partial t} + \mathcal{A}^{\hat{\mathbf{u}}(t,x,V)}V(t,x) \;=\; 0 \\ V(T,x) \;=\; U(x). \end{cases}$$

The solution of $\arg\max_{\mathbf{u}\in\mathbb{R}^n}\Big(x\,\boldsymbol{\mu}.\mathbf{u}\,V_x + \frac{x^2}{2}\|\boldsymbol{\sigma}^*\mathbf{u}\|^2\,V_{xx}\Big)$ is given by $\hat{\mathbf{u}} = \frac{-V_x}{xV_{xx}}\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^*\right)^{-1}\boldsymbol{\mu}$. Before summarizing these results in a complete statement, denote $a = \sum_{i=1}^n\left((\boldsymbol{\sigma}\boldsymbol{\sigma}^*)^{-1}\boldsymbol{\mu}\right)_i$ and $\mathbf{f} = \frac{(\boldsymbol{\sigma}\boldsymbol{\sigma}^*)^{-1}\boldsymbol{\mu}}{a}$. We have the following theorem (see [2, Theorem 19.10]):

Theorem 2.1 (Mutual Fund Theorem). *The optimal portfolio is an allocation between the riskless asset and a* fund *(more precisely a* portfolio*) which consists only of risky assets and corresponds to the control* $\mathbf{f}$.

At each time t, the relative allocation of wealth between the fund and the riskless asset is given by $m^f(t) = -\frac{a\,V_x(t,X_t)}{X_t\,V_{xx}(t,X_t)}$ *and* $m^0(t) = 1 - m^f(t)$.

In this simple situation with constant deterministic coefficients (the model investigated by Merton) the solution of the H.J.B. equation is classical, but in more general situations the solution has to be understood in the viscosity sense. For a comprehensive presentation of recent advanced results in this direction the reader can be addressed to the two interesting courses at *"Scuola Normale Superiore"* given by N. Touzi and M. Soner (see [33] and [31]).

3. A proof based on stochastic analysis

From now on, we prefer to use the process $\mathbf{H}_t$ (as defined in Section 1) for the representation of the strategy, rather than the relative portfolio.

The starting point of this approach is that, if we indicate by $\hat{X}(x)$ the optimal solution of the utility maximization problem, then $U'\big(\hat{X}(x)\big)$ is proportional to the density of the *equivalent martingale probability* $\big(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\big)$.

The *intuition* for this statement can be given as follows: if $\mathbf{K}_s$ is another n-dimensional predictable process and we consider the strategy $\big(\mathbf{H}_s + t\,\mathbf{K}_s\big)$, we have
$$\mathbb{E}\Big[U\Big(\hat{X}(x) + t\int_0^T \mathbf{K}_s\mathrm{d}\mathbf{S}_s\Big)\Big] \;\le\; \mathbb{E}\Big[U\Big(\hat{X}(x)\Big)\Big],$$
and hence the derivative with respect to t, for $t=0$, has to be 0. More precisely,
$$0 \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\mathbb{E}\Big[U\Big(\hat{X}(x) + t\int_0^T \mathbf{K}_s\mathrm{d}\mathbf{S}_s\Big)\Big] \;=\; \mathbb{E}\Big[U'\Big(\hat{X}(x)\Big)\,.\,\int_0^T \mathbf{K}_s\mathrm{d}\mathbf{S}_s\Big]$$
whatever is the strategy $\mathbf{K}$ (provided that suitable integrability conditions are satisfied): necessarily $U'\big(\hat{X}(x)\big)$ (which is a positive r.v.) is proportional to $\big(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\big)$.

Obviously this intuition needs a rigorous proof: the most general formulation (in the framework of incomplete markets) is given in [24].

Let us write the equation (2.1) in a vector form: given $\mathbf{x} \in \mathbb{R}^n$, we indicate by $D\big[\mathbf{x}\big]$ the diagonal matrix $D\big[\mathbf{x}\big] = \mathrm{diag}\big[x^1, \dots, x^n\big]$.

The equation (2.1) can be rewritten as

$$d\mathbf{S}_t = D\big[\mathbf{S}_t\big]\big(\boldsymbol{\mu}\mathrm{d}t + \boldsymbol{\sigma}\mathrm{d}\mathbf{W}_t\big) = D\big[\mathbf{S}_t\big]\boldsymbol{\sigma}\mathrm{d}\big(\mathbf{W}_t + \boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\, t\big) = D\big[\mathbf{S}_t\big]\boldsymbol{\sigma}\mathrm{d}\mathbf{W}^*_t. \quad (3.1)$$

The process $\mathbf{W}^*_t = \mathbf{W}_t + \boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\, t$ is a n-dimensional Wiener process under the probability $\mathbf{Q}$ given by the formula

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} = \exp\Big(- \int_0^T \boldsymbol{\sigma}^{-1}\boldsymbol{\mu}.\mathrm{d}\mathbf{W}_s - \frac{1}{2}\int_0^T \|\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\|^2 \mathrm{d}s\Big).$$

Consider the scalar process $Z_t = \frac{\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\|}.\mathbf{W}^*_t$: Z is a one-dimensional $\mathbf{Q}$-Wiener process and $\hat{X}(x) = \big(U'\big)^{-1}\big(y\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\big)$ is measurable with respect to the filtration generated by $\big(Z_t\big)_{0\le t\le T}$. Therefore we have the equality $\hat{X}(x) = x + \int_0^T \gamma_s \mathrm{d}Z_s$, where γ_s is a suitable scalar predictable process.

The equation (3.1) can be rewritten in the form

$$\mathrm{d}\mathbf{W}^*_t = \boldsymbol{\sigma}^{-1} D\big[\frac{1}{\mathbf{S}_t}\big].\mathrm{d}\mathbf{S}_t.$$

We have therefore

$$\begin{aligned}\hat{X}(x) &= x + \int_0^T \frac{\gamma_s}{\|\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\|}\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}.\mathrm{d}\mathbf{S}_s = x + \int_0^T \frac{\gamma_s}{\|\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\|}\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\sigma}^{-1} D\big[\frac{1}{\mathbf{S}_s}\big].\mathrm{d}\mathbf{S}_s \\ &= x + \int_0^T \frac{\gamma_s}{\|\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}\|}\big(\boldsymbol{\sigma}\boldsymbol{\sigma}^*\big)^{-1}\boldsymbol{\mu}\, D\big[\frac{1}{\mathbf{S}_s}\big].\mathrm{d}\mathbf{S}_s.\end{aligned}$$

The result of Theorem 2.1 can be rewritten in this form: for every (ω, t), the optimal *relative portfolio* $\mathbf{u}_t(\omega)$ is proportional to the vector $\big(\boldsymbol{\sigma}\boldsymbol{\sigma}^*\big)^{-1}\boldsymbol{\mu}$ and this is equivalent to saying that the *optimal strategy* $\mathbf{H}_t(\omega)$ is proportional to $\big(\boldsymbol{\sigma}\boldsymbol{\sigma}^*\big)^{-1}\boldsymbol{\mu}\, D\big[\frac{1}{\mathbf{S}_t(\omega)}\big]$. So we have obtained the *mutual fund theorem.*

In order to extend this method of proof to more general situations, it is worth pointing out the essential steps:

- the value of the optimal portfolio $\hat{X}(x)$ exists and is equal to $\big(U'\big)^{-1}\big(y\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\big)$ with a suitable positive constant y;
- the density of the equivalent martingale probability is measurable with respect to a *smaller filtration* $\big(\mathcal{G}_t\big) \subseteq \big(\mathcal{F}_t\big)$ and on this filtration there is a *stochastic integral representation property* with respect to a (k-dimensional) $\mathbf{P}$-martingale $(N_t)_{0\le t\le T}$;
- the martingale (N_t) can be written as the value of a portfolio (and identifies the *mutual fund*).

Concerning the first statement, we have a general result given by Kramkov–Schachermayer (see [24] Thm. 2.0 for details): let us first define the set of the so-called *equivalent martingale measures.*

Definition 3.1. We indicate by $\mathcal{M}$ the set of all equivalent probabilities $\mathbf{Q}$ with the property that, for every strategy $\mathbf{H}$, if the process $Y_t = \int_0^t \mathbf{H}_s \mathrm{d}\mathbf{S}_s$ is uniformly bounded from below, then it is a $\mathbf{Q}$-supermartingale.

It is usually assumed that the set $\mathcal{M}$ is non-empty: this is in some sense equivalent to an *Absence of Arbitrage* condition (see [12] and [13] for a more precise formulation).

The result stated in [24] is the following: if the market is *complete* (more precisely, if the set $\mathcal{M}$ is a singleton) then $\hat{X}(x)$ exists and is equal to $\big(U'\big)^{-1}\big(y\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\big)$ (with a suitable positive constant y), for every positive x if the utility function U satisfies an additional property (reasonable asymptotic elasticity), and given a general utility function U if x is not too big.

4. Infinite-dimensional financial models

There are two situations, in stochastic models for finance, where infinite-dimensional models are used: Large Financial Markets and Bond Markets.

Large Financial Markets were modeled in [4] as markets containing an infinite, countable, set of traded assets, represented by a sequence of semimartingales $\big(S_t^n\big)_{0\le t\le T}$, $n = 0, 1, \ldots$, on a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbf{P}\big)$.

In the Bond Market models, it is conventional to assume that at every time $t \ge 0$ there exists a bond $P(t,T)$ that matures at time T for $t \le T \le T^*$: we have in this case a *continuum* of stochastic processes $\big(P(t,T)\big)_{0\le t\le T\le T^*}$.

From the point of view of *infinite-dimensional stochastic integration*, much attention has been devoted to Bond Market models: see for instance [3, 5, 15].

The usual approach is to model $P(t,.)$ as a stochastic process with values in a suitable (Hilbert) space $\mathcal{H}$ of continuous functions defined on $[0,T^*]$: for instance, in the papers [5] or [15], $\mathcal{H}$ is an appropriate weighted Sobolev space. The natural space where the integrands should take values is the dual space $\mathcal{H}'$, and the quoted papers contain an adaptation of results of infinite-dimensional stochastic integration.

A different approach was investigated by Bjork et al. (see [3]): they consider the Bond price process as a stochastic process with values in the space of continuous functions on $[0,T^*]$, and develop a theory of stochastic integration where the integrand ϕ_t takes values in the space of signed Radon measures on $[0,T^*]$.

A different method was introduced by M. De Donno and the author in the papers [10] for the case of a sequence of semimartingales and [11] for the case of Bond Markets: we shall expose this approach in greater detail.

Let I be a set and consider an indexed family $\big(S_t^x\big)_{x\in I}$ of semimartingales defined on a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbf{P}\big)$: in our applications, I will be $\mathbb{N}$ or $[0,T^*]$ (and in the second case we impose that the application $x \to S^x$ is continuous with respect to the topology of semimartingales introduced by Émery in [17]).

We consider $\mathbf{S} = (S^x)_{x \in I}$ as a stochastic process with values in the product space $\mathbb{R}^I$: when the latter is endowed with the product topology, its dual space is formed by the finite linear combinations of Dirac's deltas (δ_x).

We call *simple integrand* a process $\mathbf{H}$ of the form $\mathbf{H}(\omega, t) = \sum_{i \le n} H^i(\omega, t)\, \delta_{x_i}$, where $x_1, \ldots, x_n \in I$ and every H^i is a scalar-bounded predictable process: given a simple integrand $\mathbf{H}$, it is natural to define the *stochastic integral*

$$\int_{]0,t]} \mathbf{H}_s \, \mathrm{d}\mathbf{S}_s = \int_{]0,t]} \sum_{i \le n} H^i_s \, \mathrm{d}S^i_s. \tag{4.1}$$

Note that a simple integrand is the mathematical counterpart of a real world portfolio, which is based on a finite number of assets.

In order to obtain a larger class of integrands, it is convenient to introduce processes with values in the set of non-continuous (unbounded) linear functionals on $\mathbb{R}^I$. Denoting by $\mathcal{U}$ the set of these unbounded functionals, we give the following definition:

Definition 4.1. Let $\mathbf{H}$ be a $\mathcal{U}$-valued process. We say that $\mathbf{H}$ is *integrable* with respect to $\mathbf{S}$ if there exists a sequence $(\mathbf{H}^n)$ of simple integrands such that

(i) $\mathbf{H}^n$ converges to $\mathbf{H}$ a.s.;
(ii) $\left(\int \mathbf{H}^n_s \mathrm{d}\mathbf{S}_s\right)$ converges to a semimartingale Y for the semimartingale topology.

We call $\mathbf{H}$ a *generalized integral* and define $\int \mathbf{H}\, \mathrm{d}\mathbf{S} = Y$.

The above definition needs some explanations: the statement (i) means that, for a.e. (ω, t), if $x \in \operatorname{Dom} \mathbf{H}(\omega, t)$, then $\mathbf{H}^n(\omega, t)(x)$ converges to $\mathbf{H}(\omega, t)(x)$. Almost surely means *outside of a set negligible for every semimartingale* S^x: a more precise and formal definition can be found in [10] and [11].

It is clear that Definition 4.1 makes sense only provided that the limit semimartingale Y does not depend on the approximating sequence: this was proved in [10] (Proposition 5.1) for the case of a sequence of semimartingales and [11] (Proposition 2.3) for the case of Bond Market models.

We wish also to point out that the Definition 4.1 of an integrable process is suggested by the notion of an integrable function with respect to a vector-valued measure (see [12], Section IV.10.7).

In order to compare this approach of infinite-dimensional stochastic integration with the previously cited approaches, let us point out that in the finite-dimensional case a fundamental result is the following:

Proposition 4.2. *Let f be a positive function: f satisfies an inequality of the form $f \le x + \int_0^T \mathbf{H}_s \mathrm{d}\mathbf{S}_s$ (with a suitable admissible strategy $\mathbf{H}$ and a positive constant x) if and only if, for every $\mathbf{Q} \in \mathcal{M}$, one has $\mathbb{E}^{\mathbf{Q}}\big[f\big] \le x$.*

The result of Proposition 4.2 was proved by El Karoui and Quenez (see [16]) in the case of diffusion processes, and by Delbaen-Schachermayer in the general semimartingale framework (see [12], and also [13] for a comprehensive presentation). It is worth pointing out that this result is strictly linked to the so-called *optional* *decomposition* (proved, in the general semimartingale case, by D. Kramkov [23]): in fact the optional decomposition is a more general result (the paper [30] by H. Pham gives an infinite-dimensional version of this decomposition, in the framework of jump-diffusion processes).

Proposition 4.2 is an essential step in the convex duality approach to the *utility maximization problem*, along the lines of the general papers by Kramkov and Schachermayer ([24] and [25]). The very technical proof is based on two properties of the (finite-dimensional) stochastic integrals:

(a) the so-called *Memin theorem*, which states that limit of stochastic integrals (for the semimartingale topology) is still a stochastic integral;
(b) the *Ansel–Stricker lemma*, which states that, if $\mathbf{M}$ is a local martingale, $\mathbf{H}$ is $\mathbf{M}$-integrable and the stochastic process $\int_0^t \mathbf{H}_s \mathrm{d}\mathbf{M}_s$ is uniformly bounded from below, then it is a *supermartingale*.

The extension of (a) is not satisfied by the approaches given, e.g., by Carmona–Tehranchi or Ekeland–Taflin, while is satisfied with Definition 4.1. More precisely, we have the following result (see [10] and [11]):

Theorem 4.3. *Let* $\mathbf{H}^n$ *be a sequence of generalized integrands such that* $\left(\int \mathbf{H}^n \mathrm{d}\mathbf{S}\right)$ *is a Cauchy sequence in the space of semimartingales: then there exists a generalized integrand* $\mathbf{H}$ *such that* $\lim_{n\to\infty} \int \mathbf{H}^n \mathrm{d}\mathbf{S} = \int \mathbf{H} \mathrm{d}\mathbf{S}$.

Unfortunately, the Ansel–Stricker lemma is false for generalized integrands (see [10] and [11] for counterexamples). Therefore the definition of *admissible strategy* has to be modified in the following way:

Definition 4.4. A generalized integrand $\mathbf{H}$ is called an admissible strategy if there exist a constant x and a sequence of approximating elementary integrands $\mathbf{H}^n$ such that:

(i) $\int_0^t \mathbf{H}^n_s \mathrm{d}\mathbf{S}_s \geq x$ a.s. for every t;
(ii) the sequence $\int \mathbf{H}^n \mathrm{d}\mathbf{S}$ converges to $\int \mathbf{H} \mathrm{d}\mathbf{S}$ for the semimartingale topology.

With this definition of admissible strategy, the results of Proposition 4.2 and the convex duality approach of [24] and [25] can be extended to infinite-dimensional models: see [9] for the case of Large Financial Markets and [11] for Bond Market models.

It is worth pointing out that there are different papers which investigate, by different methods, the problem of *utility maximization* within a Bond Market model: these are, for instance, the papers by Ekeland–Taflin (see [15]) or Ringer–Tehranchi ([32]). The latter paper, in particular, obtains a *mutual fund theorem*.

5. Generalizations of the Mutual Fund theorem

Let us first insist more on the *No Arbitrage* conditions for an infinite-dimensional model. When we have an infinite family of semimartingales $\left(S^x\right)_{x\in I}$, we indicate by $\mathcal{M}$ the set of all equivalent probabilities $\mathbf{Q}$ such that, for every finite subset

$(x_1, \dots, x_n) \subset I$, the property described in Definition 3.1 is satisfied by the n-dimensional semimartingale $\left(S^{x_1}, \dots, S^{x_n}\right)$: we suppose that the set $\mathcal{M}$ is non-empty and we say that the market is *complete* when $\mathcal{M}$ is a singleton.

The integral defined in Section 4 (Definition 4.1), which satisfies a sort of *Memin's theorem* (Thm. 4.3) is a good mathematical tool in order to face the *utility maximization problem* in an infinite-dimensional market; and when the model satisfies the properties listed at the end of Section 3, it is natural to expect that a *mutual fund theorem* can be obtained.

For instance, the paper [9] contains such a theorem for the case of Large Financial Markets, and [8] a similar result for the case of Bond Markets. Rather than to enumerate such results, we prefer to develop an example in the case of Large Financial Markets.

Let us first mention that these models were introduced by Kabanov and Kramkov (see [18] and [19]) in order to study the existence (or non-existence) of *Asymptotic Arbitrage* possibilities: to this aim, they model a Large Financial Market as a sequence of finite-dimensional financial models.

But problems such as *completeness* or *pricing of derivatives* are hard to study in this framework: to this extent, Bjork and Näslund (see [4]) choose to model a Large Financial Market as a sequence of semimartingales defined on a fixed filtered probability space and investigate the consequences of *diversification* of risk sources.

Let us examine in greater detail a *Factor Model* as introduced in [4]. We assume that every asset price depends on a systematic source of randomness which affects all the assets and on an idiosyncratic source of randomness which is typical for that asset. In particular, we assume that the price processes evolve according to the following dynamics:

$$dS_t^i = S_{t-}^i \left(\alpha_i dt + \beta_i d\hat{N}_t + \sigma_i dW_t^i \right)$$

where $(W^i)_{i \geq 1}$ is a sequence of independent Wiener processes and $\hat{N}_t = N_t - \lambda t$ is a compensated Poisson process with intensity λ (independent of W^i for all i). The Poisson process models some shocks which may occur in the market and may affect all the assets. As in [4], the coefficients $\alpha_i, \beta_i, \sigma_i$ are constants: in particular we assume that $\beta_i, \sigma_i \geq \epsilon > 0$ for all i and that there exists M such that $\sup_i(|\alpha_i|, \beta_i, \sigma_i) \leq M$.

Björk and Näslund studied the questions of No Arbitrage and completeness and showed that an asymptotic *well diversified* portfolio can be defined (as limit of a sequence of portfolios based on the first n assets), in order to complete the market. The intuitive notion of well-diversified portfolio can be translated in a more formal way into the definition of *generalized integrand* given in Section 4: a thorough investigation of completeness (via the integral defined in the previous section) was given by M. De Donno in [7]. Here, we want to analyze the problem of utility maximization in order to obtain a *mutual fund theorem.*

We take as filtration $(\mathcal{F}_t)_{t \leq T}$ the (completed) filtration generated by the price processes, hence by $\{(W^i)_{i \geq 1}, N\}$. It is well known that every local martingale L

has necessarily the form

$$L_t = L_0 + \int_0^t K_s d\hat{N}_s + \sum_{i\geq 1} \int_0^t H_s^i dW_s^i, \tag{5.1}$$

where $K, (H^i)_{i\geq 1}$ are predictable processes and

$$\int_0^T |K_s| ds + \sum_{i\geq 1} \int_0^T (H_s^i)^2 ds < \infty \qquad \text{a.s.} \tag{5.2}$$

Let $\mathbf{Q}$ be a probability measure equivalent to $\mathbf{P}$. Then its density has the form $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P} = \mathcal{E}(L_T)$ (we recall that $\mathcal{E}$ denotes the stochastic exponential), where L has the form (5.1), with $L_0 = 0$. Furthermore, we have that $K_s > -1$ in order to ensure that $\mathcal{E}(L_1) > 0$ and L is such that $\mathcal{E}(L_t)$ is a uniformly integrable martingale.

By Girsanov's theorem, it follows that the process $\tilde{W}_t^i = W_t^i - \int_0^t H_s^i \,\mathrm{d}s$ is a $\mathbf{Q}$-Wiener process, while the process $\tilde{N}_t = \hat{N}_t - \int_0^t K_s \,\mathrm{d}s = N_t - \int_0^t (1 + K_s)\,\mathrm{d}s$ is a $\mathbf{Q}$-martingale (namely $\int_0^t (1 + K_s)\,\mathrm{d}s$ is the $\mathbf{Q}$-compensator of the point process N).

Since every $(S^i)_{i\geq 1}$ is locally bounded, we have that $\mathbf{Q} \in \mathcal{M}$ if and only if S^i is a $\mathbf{Q}$-local martingale and this occurs if and only if

$$H_t^i = -\frac{\alpha_i + \beta_i K_t}{\sigma_i}$$

for all $i \geq 1$. Then, by condition (5.2), it must be $\int_0^T \sum_i (\alpha_i + \beta_i K_t)^2 \sigma_i^{-2} dt < \infty$: it is easy to check that this implies that the sequence (α_i/β_i) converges to some real number h_0. This implies that $K_t = \frac{-h_0}{\lambda} = k$, $H_t^i = \frac{-(\alpha_i + \beta_i h_0)}{\sigma_i} = h^i$, and that there exists a unique equivalent martingale measure $\mathbf{Q}$, provided that $h_0 < \lambda$ (the uniform integrability of the martingale $\mathcal{E}(L_t)$ is a consequence of the Novikov condition).

Conversely, on the n-dimensional market, there are infinitely many equivalent martingale measures. In particular, the point process N may have any intensity, and, possibly, even a stochastic compensator. We can see immediately the difference among every finite (n-dimensional) market and the large (infinite-dimensional) market:

- every n-dimensional market is incomplete, while the large market is complete;
- in every n-dimensional market the utility maximization problem is difficult to solve and there is not a mutual fund theorem, while in the large market the problem becomes easy and we have a mutual fund theorem.

Let us see in greater detail the proof of the last sentence. As in Section 3, the value $\hat{X}(x)$ of the optimal portfolio can be written in the form $\left(U'\right)^{-1}\left(y\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\right)$ with a suitable positive constant y. Note that

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} = \mathcal{E}(L_T) = \mathcal{E}\left(\sum_{j\geq 1} h_j W_T^j - h_0 \hat{N}_T\right)$$
$$= \exp\left(T \sum_{i\geq 0} h_i^2\right) \mathcal{E}\left(\sum_{j\geq 1} h_j \tilde{W}_T^j - h_0 \tilde{N}_T\right).$$

Denote by $\tilde{W}_h$ the process $\sum_{j\geq 1} h_j \tilde{W}^j$. This is a Brownian motion with respect to the probability $\mathbf{Q}$ as well as the process $\tilde{N}$ is a $\mathbf{Q}$-compensated Poisson process (with compensator $\lambda(1 - h_0/\lambda)\, t = (\lambda - h_0)\, t)$. Furthermore, both $\tilde{W}_h$ and $\tilde{N}$ coincide with the values of two self-financing portfolios: more precisely, there exists a pair of generalized strategies $\mathbf{H}^1$ and $\mathbf{H}^2$ such that

$$\tilde{W}_h = \int \mathbf{H}^1 \mathrm{d}\mathbf{S}\,, \qquad \tilde{N} = \int \mathbf{H}^2 \mathrm{d}\mathbf{S}\,. \tag{5.3}$$

This is a consequence of market completeness, for more details one can consult [7].

Observe that $\tilde{W}_h$ and $\tilde{N}$ can be interpreted as mutual funds, each composed of a small part of each asset. In particular $\tilde{W}_h$ does not depend on the systematic risk and contain a small part of all the idiosyncratic risks, while $\tilde{N}$ is based only on the systematic risk.

$\hat{X}(x)$ is measurable with respect to the filtration generated by $(\tilde{W}_h, \tilde{N})$, hence it admits a representation as

$$\hat{X}(x) = x + \int_0^T \phi_s(x) d(\tilde{W}_h)_s + \int_0^T \psi_s(x) d\tilde{N}_s.$$

This, combined with (5.3), allows us to find the optimal strategy $\hat{\mathbf{H}}(x) = \phi(x)\mathbf{H}^1 + \psi(x)\mathbf{H}^2$. Note that $\mathbf{H}^1$ and $\mathbf{H}^2$ depend only on the density of the equivalent martingale measure, while $\phi(x)$ and $\psi(x)$ are the sole processes affected by the choice of the utility function. So, we can claim a mutual fund theorem:

Theorem 5.1. *For any utility function U, the optimal portfolio consists of an allocation between the risk free asset, the mutual fund $\tilde{W}_h$ and the mutual fund $\tilde{N}$.*

References

[1] J. Ansel and C. Stricker, *Couverture des actifs contingents and prix maximum*, Ann. Inst. H. Poincaré, **30** (1994), 303–315.

[2] T. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, Second Edition, 2004.

[3] T. Björk, G. Di Masi, Y. Kabanov, and W. Runggaldier, *Towards a general theory of bond markets*, Finance Stoch., **1** (1997), 141–174.

[4] T. Björk and B. Näslund, *Diversified porfolios in continuous time*, European Finance Review, **1** (1999), 361–378.

[5] R. Carmona and M. Tehranchi, *A characterization of hedging portfolios for interest rates contingent claims*, Ann. Appl. Prob., **14** (2004), 1267–1294.

[6] J. C. Cox and C. F. Huang, *Optimum consumption and portfolio policies when asset prices follow a diffusion process*, J. Econ. Theory, **49** (1989), 33–83.

[7] M. De Donno, *A note on completeness in large financial markets*, Math. Finance, **14** (2004), 295–315.

[8] M. De Donno, *The term structure of interest rates as a random field: a stochastic integration approach*, Stoch. Processes and Applications to Math. Finance, Proceedings of the Ritsumeikan University Int. Symposium, (2004), 27–51.

[9] M. De Donno, P. Guasoni, and M. Pratelli, *Super-replication and utility maximization in large financial markets*, Stoch. Proc. Appl., **115 (12)** (2005), 2006–2022.

[10] M. De Donno and M. Pratelli, *Stochastic integration with respect to a sequence of semimartingales*, Séminaire de Probabilités 39, Lect. Notes Math., Springer, **1874** (2006), 119–135.

[11] M. De Donno and M. Pratelli, *A theory of stochastic integration for bond markets*, Ann. Appl. Probab., **15 (4)** (2005), 2773–2791.

[12] F. Delbaen and W. Schachermayer, *A general version of the fundamental theorem of asset pricing*, Math. Ann., **300** (1994), 463–520.

[13] F. Delbaen and W. Schachermayer, *The Mathematics of Arbitrage*, Springer, 2006.

[14] N. Dunford and J.T. Schwartz, *Linear Operators I*, Interscience Publishers, Wiley, 1988.

[15] I. Ekeland and E. Taflin, *A theory of bond portfolios*, Ann. Appl. Prob., **15** (2005), 1260–1305.

[16] N. El Karoui and M. C. Quenez, *Dynamic programming and pricing of contingent claims in an incomplete market*, SIAM J. Control Optimiz., **33** (1995), 29–66.

[17] M. Émery, *Une topologie sur l'espace des semimartingales*, Séminaire de Probabilités XIII, Springer Lecture Notes in Math., **784** (1980), 152–160.

[18] Y. Kabanov and D. Kramkov, *Large financial markets: asymptotic arbitrage and contiguity*, Prob. Theory and its Applications, **39** (1996), 222–229.

[19] Y. Kabanov and D. Kramkov, *Asymptotic arbitrage in large financial markets*, Finance Stoch., **2** (1998), 143–172.

[20] I. Karatzas, J. P. Lehoczky, and S. E. Shreve, *Optimal portfolio and consumption decisions for a small investor on a finite horizon*, SIAM J. Control Optimiz., **25** (1987), 1557–1586.

[21] I. Karatzas, J. P. Lehoczky, S. E. Shreve, and G. L. Xu, *Martingale and duality methods for utility maximization in an incomplete market*, SIAM J. Control Optimiz., **29** (1991), 702–730.

[22] A. Khanna and M. Kulldorf, *A generalization of the mutual fund theorem*, Finance Stochast., **3** (1999), 167–185.

[23] D. Kramkov, *Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets*, Prob. Theory and Related Fields, **105** (1996), 459–479.

[24] D. Kramkov and W. Schachermayer, *The asymptotic elasticity of utility functions and optimal investment in incomplete markets*, Ann. Appl. Prob., **9** (1999), 904–950.

[25] D. Kramkov and W. Schachermayer, *Necessary and sufficient conditions in the problem of optimal investment in incomplete markets*, Ann. Appl. Prob., **13** (2003), 1504–1516.

[26] H. M. Markowitz, *Portfolio Selection*, Wiley, New York, 1959.

[27] J. Mémin, *Espace de semi-martingales et changement de probabilité*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **52** (1980), 9–39.

[28] R. C. Merton, *Lifetime portfolio selection under uncertainty: the continuous-time case*, Rev. Econ. Statist., **51** (1969), 247–257.

[29] R. C. Merton, *Optimum consumption and portfolio rules in a continuous-time model*, J. Econ. Theory, **3** (1971), 373–413.

[30] H. Pham, *A predictable decomposition in an infinite assets model with jumps. Application to hedging and optimal investment in infinite assets models with jumps*, Stoch. and Stochastic Reports, **5** (2003), 343–368.

[31] H. M. Soner, *Stochastic Optimal Control in Finance*, Pubblicazioni della Scuola Normale Superiore, Pisa, 2004.

[32] N. Ringer and M. Tehranchi, *Optimal portfolio choice in the bond market*, 2005, preprint.

[33] N. Touzi, *Stochastic Control Problems, Viscosity Solutions ans Application to Finance*, Pubblicazioni della Scuola Normale Superiore, Pisa, 2005.

Maurizio Pratelli
Dipartimento di Matematica
Largo Bruno Pontecorvo, 5
I-56127 Pisa
Italy
e-mail: `pratelli@dm.unipi.it`