Annals of the International Society of Dynamic Games

Pierre Cardaliaguet
Ross Cressman
Editors

Advances in Dynamic Games

Theory, Applications, and Numerical Methods for Differential and Stochastic Games

Birkhäuser

Birkhäuser

Annals of the International Society of Dynamic Games
Volume 12

For further volumes:
http://www.springer.com/series/4919

Pierre Cardaliaguet • Ross Cressman
Editors

Advances in Dynamic Games

Theory, Applications, and Numerical Methods for Differential and Stochastic Games

Birkhäuser

Editors
Pierre Cardaliaguet
CEREMADE
Université Paris-Dauphine
Paris, France

Ross Cressman
Department of Mathematics
Wilfrid Laurier University
Waterloo, ON, Canada

ISBN 978-0-8176-8354-2 ISBN 978-0-8176-8355-9 (eBook)
DOI 10.1007/978-0-8176-8355-9
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012945410

Mathematics Subject Classification (2010): 91A15, 91A22, 91A23, 91A24, 91A25, 91A80

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.birkhauser-science.com)

We dedicate this volume of the Annals to the memory of

Thomas L. Vincent

Tom was a dedicated and valued member of the International Society of Dynamic Games who hosted the Eleventh International Symposium on Dynamic Games and Applications in Tucson, Arizona. He was sorely missed by friends and colleagues at the Fourteenth Symposium held in Banff, Alberta.

Preface

The Annals of the International Society of Dynamic Games have a strong track record of reporting recent advances in dynamic games by selecting articles primarily from papers based on presentations at an international symposium of the Society of Dynamic Games. This edition, Volume 12, continues the tradition, with most contributions connected to the 14th International Symposium on Dynamic Games and Applications held in Banff, Alberta, Canada in June 2010. The symposium was cosponsored by St. Francis Xavier University, Antigonish, Nova Scotia, Canada; by the Group for Research in Decision Analysis (GERARD); and by the Chair in Game Theory and Management, HEC Montréal, Canada.

The volume contains 20 chapters that have been peer-reviewed according to the standards of the international journals in game theory and its applications. The chapters are organized into four parts: evolutionary game theory (Part I), theoretical developments in dynamic and differential games (Part II), pursuit-evasion games and search games (Part III), and applications of dynamic games (Part IV). Beginning with its first volume in 2011, the journal *Dynamic Games and Applications* has provided another important venue for the dissemination of related research. Combined with the Annals, this development points to a bright future for the theory of dynamic games as it continues to evolve.

Part I is devoted to *evolutionary game theory and applications*. It contains four chapters.

David Ramsey examines age-structured game-theoretic models of mating behavior in biological species, a topic of early interest when evolutionary game theory began in the 1970s. Ramsey extends recent progress in this area first by allowing the individual mating rate to depend on the proportion searching for mates and then by incorporating asymmetries in the newborn sex ratio or the time for which males and females are fertile. An iterative best-response procedure is used to generate the equilibrium age distribution of fertile individuals.

Mike Mesterton-Gibbons and Tom Sherratt consider the evolutionary consequences of signaling and of dominance in conflicts between two individuals. In particular, it is shown that, when dominance over the opponent is sufficiently advantageous, the evolutionarily stable strategy (ESS) is for only winners of the conflict to signal in long contests and for neither winners nor losers to signal in short contests.

Quanyan Zhu, Hamidou Tembine, and Tamer Başar formulate a multiple-access control game and show that there is a convex set of pure strategy Nash equilibria. The paper also addresses how to select one equilibrium from this set through game-theoretic solutions such as ESS as well as through the long-run behavior of standard evolutionary dynamics applied to this game that has a continuum of pure strategies.

Andrei Akhmetzhanov, Frédéric Grognard, Ludovic Mailleret, and Pierre Bernhard study the evolution of a consumer–resource system assuming that the reproduction rate of the resource population is constant. The consumers' optimal behavior is found over one season when they all act for the common good. The authors then show that selfish mutants can successfully invade this system but are eventually as vulnerable to invasion as the initially cooperative residents.

Part II contains eight chapters on *theoretical developments of dynamic and differential games*.

Sergey Chistyakov and Leon Petrosyan analyze coalition issues in *m*-person differential games with prescribed duration and integral payoffs. They show that components of the Shapley value are absolutely continuous and thus differentiable functions along any admissible trajectory.

Yurii Averboukh studies two-player, non-zero-sum differential games and characterizes the set of Nash equilibrium payoffs in terms of nonsmooth analysis. He also obtains sufficient conditions for a pair of continuous payoff functions to generate a Nash equilibrium.

Anne Souquière studies two-player, non-zero-sum differential games played in mixed strategies and characterizes the set of Nash equilibrium payoffs in this framework. She shows in particular that the set of publicly correlated equilibrium payoffs is the same as the set of Nash equilibrium payoffs using mixed strategies.

Dean Carlson and George Leitmann explain how to solve non-zero-sum differential games with equality constraints by using a penalty method approach. Under the assumption that the penalized problem has an open-loop Nash equilibrium, they show that this open-loop Nash equilibrium converges to an open-loop Nash equilibrium for the constrained problem.

Paul Frihauf, Miroslav Krstic, and Tamer Başar investigate how to approximate the stable Nash equilibria of a game by solving a differential equation in which the players only need to measure their own payoff values. The approximation method is based on the so-called extremum-seeking approach.

Miquel Oliu-Barton and Guillaume Vigeral obtain Tauberian-type results in (continuous-time) optimal control problems: they show an equivalence between the long-time average and the convergence of the discounted problem as the discount rate tends to 0.

Lucia Pusillo and Stef Tijs propose a new type of equilibrium for multicriteria noncooperative games. This "E-equilibrium" is based on improvement sets and captures the idea of approximate and exact solutions.

Olivier Guéant studies a particular class of mean field games, with linear-quadratic payoffs (mean field games are obtained as the limit of stochastic differential games when the number of interacting agents tends to infinity). The author shows that the system of equations associated with these games can be transformed into a simple system of coupled partial differential equations, for which he provides a monotonic scheme to build solutions.

Part III is devoted to *pursuit-evasion games and search games* and contains six contributions.

Sourabh Bhattacharya and Tamer Başar investigate the effect of an aerial jamming attack on the communication network of a team of unmanned aerial vehicles (UAVs) flying in a formation. They analyze the problem in the framework of differential game theory and provide analytical and approximate techniques to compute nonsingular motion strategies of UAVs.

Serguei A. Ganebny, Serguei S. Kumkov, Stéphane Le Menec, and Valerii S. Patsko study a pursuit-evasion game with two pursuers and one evader having linear dynamics. They perform a numerical construction of the level sets of the value function and explain how to produce feedback-optimal control.

Stéphane Le Menec presents a centralized algorithm to design cooperative allocation strategies and guidance laws for air defense applications. One of its main features is a capability to generate and counter alternative target assumptions based on concurrent beliefs of future target behaviors, i.e., a Salvo Enhanced No Escape Zone (SENEZ) algorithm.

Alexander Belousov, Alexander Chentsov, and Arkadii Chikrii study pursuit-evasion games with integral constraints on the controls. They derive sufficient conditions for the game to terminate in finite time.

Anna Karpowicz and Krzysztof Szajowski study the angler's fishing problem, in which an angler has at most two fishing rods. Using dynamic programming methods, the authors explain how to find the optimal times to start fishing with only one rod and then to stop fishing altogether to maximize the angler's satisfaction.

Ryusuke Hohzaki deals with a non-zero-sum three-person noncooperative search game, where two searchers compete for the detection of a target and the target tries to evade the searchers. He shows that, in some cases, there is cooperation between two searchers against the target and that the game can then be reduced to a zero-sum one.

Part IV contains two papers dedicated to the *applications of dynamic games to economics and management science.*

Alessandra Buratto formalizes a fashion licensing agreement where the licensee produces and sells a product in a complementary business. Solving a Stackelberg differential game, she analyzes the different strategies the licensor can adopt to sustain his brand.

Pietro De Giovanni and Georges Zaccour consider a closed-loop supply chain with a single manufacturer and a single retailer. They characterize and compare the feedback equilibrium results in two scenarios. In the first scenario, the manufacturer invests in green activities to increase the product-return rate while the retailer controls the price. In the second scenario, the players implement a cost revenue sharing contract in which the manufacturer transfers part of its sales revenues and the retailer pays part of the cost of the manufacturer's green activities program that aims at increasing the return rate of used products.

Acknowledgements

The selection of contributions to this volume started during the 14th International Symposium on Dynamic Games and Applications held in Banff. Our warmest thanks go to all the referees of the papers. Without their invaluable efforts this volume would not have been possible. Finally, our thanks go to the editorial staff at Birkhäuser, and especially Tom Grasso, for their assistance throughout the editing process. It has been an honor to serve as editors.

Paris, France — Pierre Cardaliaguet
Waterloo, Ontario, Canada — Ross Cressman
March 2012

Contents

Contributors

Andrei R. Akhmetzhanov Department of Biology (Theoretical Biology Laboratory), McMaster University, Life Sciences Building, 1280 Main St. West, Hamilton, ON L8S 4K1, Canada

INRIA Sophia Antipolis – Méditerranée (Project BIOCORE), 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France

Yurii Averboukh Institute of Mathematics and Mechanics UrB RAS, S. Kovalevskaya Street 16, GSP-384, Ekaterinburg 620990, Russia

Tamer Başar Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801-2307, USA

Aleksandr Andreevich Belousov Glushkov Institute of Cybernetics, NAS of Ukraine, Kiev, Ukraine

Sourabh Bhattacharya Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Pierre Bernhard INRIA Sophia Antipolis – Méditerranée (Project BIOCORE), 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France

Alessandra Buratto Department of Mathematics, Via Trieste 63, I-35131 Padova, Italy

Dean A. Carlson American Mathematical Society, Mathematical Reviews, 416 Fourth Street, Ann Arbor, MI 48103, USA

Aleksander Georgievich Chentsov Ural Division of RAS, Institute of Mathematics and Mechanics, Ekaterinburg, Russia

Arkadii Alekseevich Chikrii Glushkov Institute of Cybernetics, NAS of Ukraine, Kiev, Ukraine

Sergey Chistyakov Faculty of Applied Mathematics and Control Processes, St. Petersburg University, Universitetskiy pr. 35, St. Petersburg 198504, Russia

Pietro De Giovanni Department of Information, Logistics and Innovation, VU University Amsterdam, de Boelelaan 1105, 3A-31, 1081 HV Amsterdam, The Netherlands

Paul Frihauf Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA

Sergey Alexandrovich Ganebny Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia,

Frédéric Grognard INRIA Sophia Antipolis – Méditerranée (Project BIOCORE), 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France

Olivier Guéant UFR de Mathématiques, Laboratoire Jacques-Louis Lions, Université Paris-Diderot. 175, rue du Chevaleret, 75013 Paris, France

Ryusuke Hohzaki Department of Computer Science, National Defense Academy, 1-10-20 Hashirimizu, Yokosuka 239-8686, Japan

Anna Karpowicz Bank Zachodni WBK, Rynek 9/11, 50-950 Wrocław, Poland

Miroslav Krstic Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA,

Sergey Sergeevich Kumkov Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia

Stéphane Le Ménec EADS/MBDA, Paris, France

George Leitmann Graduate School, University of California at Berkeley, Berkeley, CA 94720, USA

Ludovic Mailleret INRA Sophia Antipolis (UR880), 400 route des Chappes, BP 167, 06903 Sophia Antipolis, France

Mike Mesterton-Gibbons Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, FL 32306-4510, USA

Miquel Oliu-Barton Institut Mathématique de Jussieu, UFR 929, Université Paris 6, Paris, France

Valerii Semenovich Patsko Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia

Leon Petrosyan Faculty of Applied Mathematics and Control Processes, St. Petersburg University, Universitetskiy pr. 35, St. Petersburg 198504, Russia

Lucia Pusillo Dima – Department of Mathematics, University of Genoa, Via Dodecaneso 35, 16146 Genoa, Italy

David M. Ramsey Department of Mathematics and Statistics, University of Limerick, Ireland

Tom N. Sherratt Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

Anne Souquière Université de Bretagne Occidentale, Laboratoire de Mathématiques, UMR 6205, 6 Avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex, France

Krzysztof Szajowski Institute of Mathematics and Computer Science, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

Hamidou Tembine Department of Telecommunications, École Supérieure d'Electricité (SUPELEC), 3 rue Joliot-Curie, 91192 Gif-Sur-Yvette Cedex, France

Stef Tijs CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Guillaume Vigeral CEREMADE, Université Paris-Dauphine, Paris, France

Georges Zaccour GERAD, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada

Quanyan Zhu Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1308 W. Main St. Urbana, IL 61801, USA

Chapter 1
Some Generalizations of a Mutual Mate Choice Problem with Age Preferences

David M. Ramsey

Abstract This paper considers some generalizations of the large population game theoretic model of mate choice based on age preferences introduced by Alpern et al. [Alpern et al., Partnership formation with age-dependent preferences. Eur. J. Oper. Res. (2012)]. They presented a symmetric (with respect to sex) model with continuous time in which the only difference between members of the same sex is their age. The rate at which young males enter the adult population (at age 0) is equal to the rate at which young females enter the population. All adults are fertile for one period of time and mate only once. Mutual acceptance is required for mating to occur. On mating or becoming infertile, individuals leave the pool of searchers. It follows that the proportion of fertile adults searching and the distribution of their ages (age profile) depend on the strategies that are used in the population as a whole (called the strategy profile). They look for a symmetric equilibrium strategy profile and corresponding age profile satisfying the following condition: any individual accepts a prospective mate if and only if the reward obtained from such a pairing is greater than the individual's expected reward from future search. It is assumed that individuals find prospective mates at a fixed rate. The following three generalizations of this model are considered: (1) the introduction of a uniform mortality rate, (2) allowing the rate at which prospective mates are found to depend on the proportion of individuals who are searching, (3) asymmetric models in which the rate at which males and females enter the population and/or the time for which they are fertile differ.

Keywords Dynamic game • Mate choice problem • Policy iteration • Equilibrium profile

---

D.M. Ramsey (✉)
Department of Mathematics and Statistics, University of Limerick, Ireland,
e-mail: david.ramsey@ul.ie

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_1,

1.1 Introduction

Many models of mate choice have been based on common preferences. According to such preferences, individuals prefer attractive partners and each individual of a given sex agrees on the attractiveness of a member of the opposite sex. Some work has been carried out on models in which preferences are homotypic, i.e. individuals prefer partners who are similar (e.g. in character) to themselves in some way. In such models the attractiveness and character of an individual are assumed to be fixed. One obvious characteristic upon which mate choice might be based is the age of a prospective partner (and the searcher himself/herself). By definition, the age of an individual must change over time. Very little theoretical work has been carried out on such problems. This article extends a model considered by Alpern et al. [4].

Janetos [8] was the first to present a model of mate choice with common preferences. He assumed that only females are choosy and the value of a male to a female comes from a distribution known to the females. There is a fixed cost for observing each prospective mate, but there is no limit on the number of males a female can observe. Real [19] developed these ideas.

In many species both sexes are choosy and such problems are game theoretic. Parker [17] presents a model in which both sexes prefer mates of high value. He concludes that assortative mating should occur with individuals being divided into classes. Class i males are paired with class i females and there may be one class of males or females who do not mate. Unlike the models of Janetos [8] and Real [19], Parker's model did not assume that individuals observe a sequence of prospective mates.

In the mathematics and economics literature such problems are often formulated as marriage problems or job search problems. McNamara and Collins [12] consider a job search game in which job seekers observe a sequence of job offers and, correspondingly, employers observe a sequence of candidates. Both groups have a fixed cost of observing a candidate or employer, as appropriate. Their conclusions are similar to those of Parker [17]. Real [20] developed these ideas within the framework of mate choice problems. For similar problems in the economics literature see e.g. Shimer and Smith [21] and Smith [22].

In the above models it is assumed that the distribution of the value of prospective partners has reached a steady state. There may be a mating season and as it progresses the distribution of the value of available partners changes. Collins and McNamara [6] were the first to formulate such a model as a one-sided job search problem with continuous time. Ramsey [18] considers a similar problem with discrete time. Johnstone [9] presents numerical results for a discrete time, two-sided mate choice problem with a finite horizon. Alpern and Reyniers [3] use a more analytic approach to similar mate choice problems. These models are further developed and analyzed in Alpern and Kantrantzi [1] and Mazalov and Falko [11]. Burdett and Coles [5] consider a dynamic model in which the outflow resulting from partnership formation is balanced by job seekers and employers coming into the employment market. Alpern and Reyniers [2] consider a similar model in which individuals have homotypic preferences.

This paper is an extension of the work by Alpern et al. [4]. They consider a problem of mutual mate choice in which all individuals of a sex are identical except for their age. They first consider a problem with discrete time in which males are fertile for m periods and females are fertile for n periods. Without loss of generality, we may assume that $m \geq n$. At each moment of time a number a_1 (b_1) of young males (females) of age 1 enter the adult population. All the other adult individuals age by 1 unit. If a male (female) reaches the age $m+1$ $(n+1)$ without having mated, then he (she) is removed from the population. The ratio $R = a_1/b_1$ is called the *incoming sex ratio* (ISR). The ratio of the number of adult males searching for a mate to the number of females searching for a mate is called the *operational sex ratio* (OSR) and is denoted by r.

At each moment an individual of the least common sex in the mating pool is matched with a member of the opposite sex with probability ε. The age of the prospective partner is chosen at random from the distribution of the age of members of the appropriate sex. Suppose males are at least as common as females (i.e. $r \geq 1$). It follows that a male is matched with a female with probability ε/r. Given a male is matched with a female, her age is chosen at random from the age of females. Similarly, if females are at least as common as males (i.e. $r \leq 1$), then in each period a searching female is matched with a male with probability εr. The age of such a male is chosen at random from the distribution of male age. When two individuals are matched, they must decide whether to accept or reject their prospective partner. Mating only occurs by mutual consent. On mating two individuals are removed from the population of searchers. It follows that the steady state distributions of the ages of males and females in the population of searchers depend on the strategies used within the population as a whole (the strategy profile).

The reward obtained by a pair on mating is taken to be the expected number of offspring produced over the period of time for which both individuals are fertile. It is assumed that offspring are produced at a rate depending on the ages of the partners in such a way that the reward obtained by an individual on mating is non-increasing in the age of the prospective partner. In this case, the equilibrium strategies of males and females are threshold strategies in which each individual defines the maximum acceptable age of a prospective partner as a function of the individual's age. This maximum acceptable age is non-decreasing in the age of the individual. One example of such a reward function is the *simple fertility model*, according to which the payoff of a pair on mating is simply the number of periods for which both partners remain fertile. Equilibrium strategy profiles and the corresponding age profiles are derived for a selection of problems of this form.

In addition, they define a continuous time model of a symmetric mate choice problem in which both males and females enter the adult population at the same rate and are fertile for one unit of time. It is assumed that when both males and females use the same strategy (and thus $r = 1$), individuals meet prospective partners as a Poisson process of rate λ (called the interaction rate). Hence, an individual expects to meet λ prospective partners during their fertile period. The payoff obtained by a pair on mating is equal to the length of time for which both remain fertile. A policy iteration algorithm is defined to approximate a symmetric equilibrium of the game.

It should be noted that when the strategy used by females differs from the strategy used by males, then the OSR may well differ from one. Since a matching of a male with a female must correspond exactly to one matching of a female with a male, it follows that the interaction rate depends on the OSR. Given the assumption regarding the interaction rate at a symmetric equilibrium, females should meet males at rate $\frac{2\lambda r}{r+1}$, while males should meet females at rate $\frac{2\lambda}{r+1}$. On the other hand, the policy iteration algorithm assumes that this interaction rate is always λ. This assumption affects the dynamics of the evolution of the threshold rule. However, since at a symmetric equilibrium the OSR will be equal to 1, a fixed point of such a regime will also be a fixed point of a suitably adapted algorithm in which the interaction rate varies according to the OSR.

The paper presented here considers three extensions of this continuous time model. Section 1.2 outlines the original model. Section 1.3 adapts this model to include a fixed mortality rate for fertile individuals. Section 1.4 considers a model in which the interaction rate depends on the proportion of fertile members of the opposite sex who are searching for a mate. Section 1.5 considers a model of an asymmetric game in which males are fertile for longer than females and/or the ISR differs from 1. For convenience and ease of exposition, these adaptations are considered separately, but they can be combined relatively easily. Section 1.6 gives some numerical results, while Sect. 1.7 gives a brief conclusion and some directions for future research.

1.2 The Original Model with Continuous Time

1.2.1 Outline of the Model

We consider a symmetric (with respect to sex) model in which the rate at which new males enter the adult population equals the rate at which females enter, i.e. $R = 1$. Suppose individuals are fertile for one unit of time, there is no mortality over this period and the rate at which they meet prospective partners is λ.

Each prospective partner is chosen at random from the set of members of the opposite sex that are searching for a mate. When two prospective partners meet, they decide whether to accept or reject the other on the basis of his/her age. Acceptance must be mutual, in order to form a breeding pair. If acceptance is not mutual, both individuals continue searching. No recall of previously encountered prospective partners is possible.

The strategy of an individual defines the set of ages of acceptable prospective mates at each age. We look for a symmetric equilibrium of such a game in which males and females use the same strategy. It is clear that at such an equilibrium the OSR is also equal to 1.

Suppose the rate at which a male of age x and a female of age y produce offspring is $\gamma(x,y)$, where $\gamma(x,y) \geq 0$. The reward of both partners when a male of age x pairs with a female of age y is given by $u(x,y)$, where

$$u(x,y) = \int_0^{1-\max\{x,y\}} \gamma(x+t,y+t)\,\mathrm{d}t.$$

Suppose the rate at which fertile partners produce offspring is independent of their ages. We may assume $\gamma(x,y) = 1$. In this case $u(x,y) = 1 - \max\{x,y\}$. This is simply the period of time for which both of the partners remain fertile. In the following analysis, we assume the reward is of this form.

The equilibrium condition is as follows: each individual should accept a prospective mate if and only if the reward gained from such a mating is greater than the expected reward from future search. An equilibrium can be described by a strategy pair. It is assumed that all males follow the first strategy in this pair and females follow the second. At a symmetric equilibrium males and females use the same strategy.

Note that the reward of an individual of age x from mating with a prospective partner of age y, $1-\max\{x,y\}$, is non-increasing in y. Hence, if an individual of age x should accept a prospective partner of age y, then he/she should accept a prospective partner of age $\leq y$. Thus at a symmetric equilibrium each individual uses a threshold rule such that an individual of age x accepts any prospective partner of age $\leq f(x)$. The function f will be referred to as the threshold profile.

The future expected reward at age 1 is 0. Hence, an individual of age 1 will accept any prospective mate, i.e. $f(1) = 1$. Suppose an individual of age x meets a prospective mate of age $\leq x$. By mating with such a prospective mate, the individual obtains a payoff of $1-x$, which for $x < 1$ is greater than the payoff obtained from continued search. Hence, $f(x) \geq x$ with equality if and only if $x = 1$. In addition, $f'(x) \geq 0$, since at equilibrium an individual of age x can ensure himself/herself the same reward as an individual of age $x+\delta$ by rejecting all prospective partners until age $x+\delta$ and then following the threshold profile $f(x)$. It should be noted that an individual of age $\leq f(0)$ will be acceptable to any member of the opposite sex.

Define $a(x)$ to be the *steady state* proportion of individuals of age x that are still searching for a mate. It should be noted that this proportion depends on the acceptance rule being used in the population [i.e. on $f(x)$]. The proportion of fertile individuals that have not mated is $\overline{a}$, where $\overline{a} = \int_0^1 a(x)\,\mathrm{d}x$. It follows that the density function of the age of available, fertile individuals is given by $\hat{a}(x) = a(x)/\overline{a}$. The function a will be referred to as the age profile.

We now derive a differential equation which the equilibrium threshold profile must satisfy. Consider a male of age x. The probability of encountering a unmated female in a small interval of time of length δ is $\lambda\delta$. We consider two cases:

1. $x < f(0)$. In this case the male is acceptable to females of any age y. The female is acceptable if $y \leq f(x)$. The probability that the female is acceptable is given by

$$\frac{\int_0^{f(x)} a(u)\,\mathrm{d}u}{\overline{a}}.$$

Given a male is still searching at age x, the probability he mates between age x and age $x+\delta$ is given by

$$\frac{\lambda\delta}{\overline{a}}\int_0^{f(x)} a(u)\,\mathrm{d}u + O(\delta^2).$$

Hence,

$$a(x+\delta) = a(x)\left[1-\frac{\lambda\delta}{\overline{a}}\int_0^{f(x)} a(u)\,\mathrm{d}u\right] + O(\delta^2)$$

$$\frac{a(x+\delta)-a(x)}{\delta} = -\frac{\lambda a(x)}{\overline{a}}\int_0^{f(x)} a(u)\,\mathrm{d}u + O(\delta).$$

Letting $\delta \to 0$, we obtain

$$a'(x) = -\frac{\lambda a(x)}{\overline{a}}\int_0^{f(x)} a(u)\,\mathrm{d}u. \tag{1.1}$$

2. $x \geq f(0)$. In this case, the male must also be acceptable to the female, i.e. $x \leq f(y)$. Since f is an increasing function, it follows that $f^{-1}(x) \leq y$. Hence, acceptance is mutual if $f^{-1}(x) \leq y \leq f(x)$. Given a male is still searching at age x, the probability he mates between age x and age $x+\delta$ is given by

$$\frac{\lambda\delta}{\overline{a}}\int_{f^{-1}(x)}^{f(x)} a(u)\,\mathrm{d}u + O(\delta^2).$$

Calculations analogous to the ones made in Point 1 lead to

$$a'(x) = -\frac{\lambda a(x)}{\overline{a}}\int_{f^{-1}(x)}^{f(x)} a(u)\,\mathrm{d}u. \tag{1.2}$$

The decisions of the players at equilibrium are illustrated in Fig. 1.1.

For $x < f(0)$, dividing both sides of the equation by $a(x)$ and differentiating with respect to x, we obtain the following second order differential equation:

$$a(x)a''(x) - [a'(x)]^2 = -\frac{\lambda[a(x)]^2 a(f(x))f'(x)}{\overline{a}}. \tag{1.3}$$

Equation (1.3) is very difficult to solve directly, even numerically, due to the presence of the composite function $a \circ f$. We have the boundary condition $a(0) = 1$, but in order to use a difference equation to estimate $a(\delta)$ for small δ, we need to know $a(f(0))$, where $f(0) > 0$. One additional problem is that there is no boundary condition for $f(0)$. For this reason, we will define a policy iteration algorithm to numerically compute the equilibrium threshold rule and age profile.

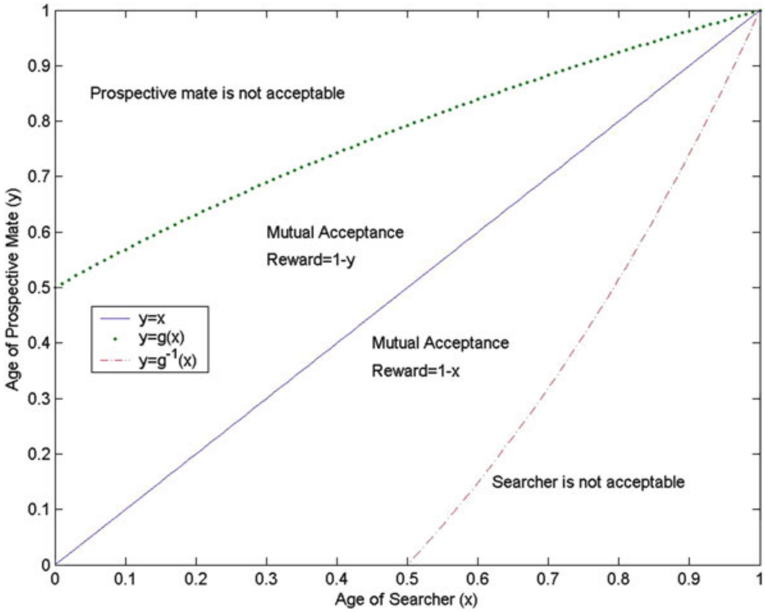


Fig. 1.1 Illustration of the decisions made at equilibrium

1.2.2 Numerical Computation of a Symmetric Equilibrium

At first glance, it appears that the following procedure might work: choose an arbitrary male strategy f_1; determine an optimal female response strategy f_2; determine an optimal male response f_3 to f_2, and so on, hoping that the sequence converges to a limit. If there is a limit, this defines a symmetric equilibrium of the game considered. In a true two person game, such a procedure is at least feasible in principle. However, it will not work in the present setting. To see this, suppose the females know the male strategy f_1 and consider the problem faced by a female of age y. In order to know which males to accept, she needs to know her optimal future return from searching. For this she needs to know (a) the rate at which she will be matched (this is assumed to be known), (b) which males will accept her at her future ages (this is known from f_1), and (c) the age profile, call it a_1, of the males she will be matched with. However, in the scheme we proposed above, she will *not* know this, as it is not determined solely by the male strategy f_1, but also depends on what the females are doing. In theory, we could determine f_2 based on f_1 and a previous female strategy, say f_0, and determine a_1 as the age profile corresponding to f_1 and f_0. However, as we showed in Sect. 1.2.1, the determination of a_1 is difficult. So we use a different iterative procedure, described below.

Suppose we begin by positing an initial male strategy, denoted f_1 (where f_1 is a non-decreasing function), *and* any non-increasing initial male age profile, denoted a_1. Given f_1 and a_1, we can indeed determine a female's optimal response (which is a threshold rule), denoted f_2. This calculation will be derived in Sect. 1.2.3. We write this calculation as

$$f_2 = H_1(f_1, a_1).$$

To continue the process, we then need to compute the function a_2 defining the probability that an individual female using f_2 is still searching for a mate at age x when the male age profile is a_1 and the males adopt strategy f_1. Note that this is not by definition the age profile of females when all males use f_1 and all females use f_2. We denote this computation (derived in Sect. 1.2.3) as

$$a_2 = H_2(f_1, a_1).$$

Define

$$(f_2, a_2) = H(f_1, a_1) = (H_1(f_1, a_1), H_2(f_1, a_1)).$$

Since the game is symmetric with respect to sex, we may define the optimal response of an individual, f_{i+1} and the probability that when such an individual uses this strategy he/she is still searching at age x, $a_{i+1}(x)$, when the members of the opposite sex use the threshold profile f_i and have age profile a_i as follows:

$$(f_{i+1}, a_{i+1}) = H(f_i, a_i) = (H_1(f_i, a_i), H_2(f_i, a_i)).$$

Theorem 1.1 (From Alpern et al. [4]). *Suppose that for some initial strategy-age profile pair (f_1, a_1), the iterates $(f_{i+1}, a_{i+1}) = H(f_i, a_i)$ converge to a limit (f, a). Then*

- *The strategy pair (f, f) is a symmetric equilibrium and*
- *Both sexes have the invariant age profile $a = a(x)$.*

Proof. In the limit we will have

$$(f, a) = H(f, a).$$

It follows from the definition of H_1 that f is the best response function of females when males adopt f and their age profile is a. Similarly, f is the best response function of males when females adopt f and have age profile a. Hence, it suffices to show that when all individuals use the threshold strategy f, then the age profile in both sexes is a. The second part of the iteration indicates that the probability that an individual using f is still searching at age x is given by $a(x)$. This individual is using the same strategy as the rest of the population, thus $a(x)$ is simply the proportion of individuals of age x who are still in the mating pool, as required. □

1.2.3 Definition of the Mapping H

In order to simplify the notation used, we define $f^{-1}(x) = 0$ for $x \leq f(0)$, otherwise $f^{-1}(x)$ is the standard inverse function. First, we consider the best response f_{i+1} to the pair of profiles (f_i, a_i). Denote $\bar{a}_i = \int_0^1 a_i(x)\,\mathrm{d}x$. Suppose an individual is still searching at age x. The optimal expected future reward from search is equal to the reward obtained by accepting the oldest acceptable prospective partner, i.e. $1 - f_{i+1}(x)$. Suppose the next encounter with an available mate occurs at age W. The probability density function of W is given by $p(w|W > x) = \lambda\,\mathrm{e}^{-\lambda(w-x)}$, for $x \leq w < 1$. It should be noted that there is an atom of probability at $w = 1$ of mass equal to the probability that an individual does not meet another prospective partner given that he/she is still searching at age x. In this case, the reward from search is defined to be 0. Suppose the age of the prospective mate is y. The pairing is mutually acceptable if $y \in [f_i^{-1}(w), f_{i+1}(w)]$. If $y \in [f_i^{-1}(w), w]$, then the searcher obtains a reward of $1 - w$. If $y \in (w, f_{i+1}(w)]$, then the searcher obtains a reward of $1 - y$. In all other cases, the future expected reward from search is $1 - f_{i+1}(w)$. Conditioning on the age of the prospective mate and taking the expected value, it follows that

$$1 - f_{i+1}(x) = \frac{\lambda\,\mathrm{e}^{\lambda x}}{\bar{a}_i}\int_x^1 \mathrm{e}^{-\lambda w}\Bigg[[1 - f_{i+1}(w)]\int_0^{f_i^{-1}(w)} a_i(y)\,\mathrm{d}y + \int_{f_i^{-1}(w)}^{w}(1-w)a_i(y)\,\mathrm{d}y + \\ + \int_w^{f_{i+1}(w)}(1-y)a_i(y)\,\mathrm{d}y + \int_{f_{i+1}(w)}^{1}[1 - f_{i+1}(w)]a_i(y)\,\mathrm{d}y\Bigg]\,\mathrm{d}w.$$

Dividing by $\mathrm{e}^{\lambda x}$ and differentiating with respect to x, after some simplification we obtain

$$f'_{i+1}(x) = \frac{\lambda}{\bar{a}_i}\left\{[f_{i+1}(x) - x]\int_{f_i^{-1}(x)}^{x} a_i(y)\,\mathrm{d}y + \int_x^{f_{i+1}(x)}[f_{i+1}(x) - y]a_i(y)\,\mathrm{d}y\right\}. \quad (1.4)$$

Equation (1.4) can be solved numerically, using the boundary condition $f(1) = 1$ and estimating $f_{i+1}(x)$ sequentially at $x = 1 - h, 1 - 2h, \ldots, 0$.

Once f_{i+1} has been estimated, we can estimate the corresponding age profile. The calculations are analogous to the calculations carried out in the previous section. We have

$$a'_{i+1}(x) = -\frac{\lambda a_{i+1}(x)}{\bar{a}_i}\int_{f_i^{-1}(x)}^{f_{i+1}(x)} a_i(y)\,\mathrm{d}y. \quad (1.5)$$

Equation (1.5) can be solved numerically, using the boundary condition $a(0) = 1$ and estimating $a(x)$ sequentially at $x = h, 2h, \ldots, 1$.

A proof that the iteration procedure is well defined can be found in Alpern et al. [4]. This proof may be adapted to show that the procedures proposed for the three extensions considered below are also well defined.

1.3 A Symmetric Game with a Fixed Mortality Rate

We now adapt the model presented above by assuming that mortality affects fertile individuals at a constant rate of μ (independently of sex and status, single or mated). We first derive the expected reward obtained by a pair composed of a male of age x and a female of age y, denoted $u(x,y)$. This is given by the expected time for which both partners survive and are fertile. Note that the death of one of the partners occurs at rate 2μ. Suppose $x \leq y$. If both partners survive a period of $1-y$ (i.e. until the female becomes infertile), then they both receive a payoff of $1-y$. Otherwise, the reward obtained is the time until the death of the first of the partners, denoted by Z. It follows that

$$u(x,y) = (1-y)\int_{1-y}^{\infty} 2\mu\,\mathrm{e}^{-2\mu z}\,\mathrm{d}z + \int_{0}^{1-y} 2\mu z \mathrm{e}^{-2\mu z}\,\mathrm{d}z = \frac{1-\mathrm{e}^{-2\mu(1-y)}}{2\mu}. \quad (1.6)$$

Analogously, when $x > y$ it can be shown that

$$u(x,y) = \frac{1-\mathrm{e}^{-2\mu(1-x)}}{2\mu}.$$

Suppose that the threshold profile and age profile of females is given by the profile pair (f_i, a_i). As before, we define $\overline{a}_i = \int_0^1 a_i(x)\,\mathrm{d}x$. We now consider the optimality criterion for an individual male of age x. As in the original model, an individual should always accept a prospective partner who is younger. Such a male should accept an older female if and only if the expected reward obtained from such a pairing is greater than the male's expected reward from future search, $r_{i+1}(x)$. It follows from Eq. (1.6) that the optimal response (the threshold profile f_{i+1}) must satisfy

$$r_{i+1}(x) = \frac{1-\mathrm{e}^{-2\mu[1-f_{i+1}(x)]}}{2\mu} \Rightarrow f_{i+1}(x) = 1 + \frac{\ln[1-2\mu r_{i+1}(x)]}{2\mu}. \quad (1.7)$$

We now derive a differential equation for $r_{i+1}(x)$ by conditioning on the time of the next event, where the death of a male and meeting a prospective partner are defined to be events. Events occur at rate $\lambda + \mu$, thus given the male is still searching at age x, the time at which the next event occurs, W, has density function $p(w|W > x) = (\lambda + \mu)\mathrm{e}^{-(\lambda+\mu)(w-x)}$ for $x \leq w < 1$. Note that W has an atom of probability at $w = 1$ of mass equal to the probability that no event occurs before the male becomes infertile. Given that an event occurs before the male becomes infertile, this event is his death with probability $\frac{\mu}{\lambda+\mu}$. If no event occurs before he reaches age 1 or the first event is his death, the reward of the male is 0. Considering the time of the next event, the type of this event and the age of the prospective partner, we obtain

$$
\begin{aligned}
r_{i+1}(x) = \int_x^1 \frac{\lambda}{\bar{a}_i} \exp[-(\lambda+\mu)(w-x)] \\
\times \left[\int_0^{f_i^{-1}(w)} r_{i+1}(w) a_i(y)\,\mathrm{d}y + \int_{f_i^{-1}(w)}^{w} \frac{(1-\mathrm{e}^{-2\mu(1-w)}) a_i(y)\,\mathrm{d}y}{2\mu} \right. \\
\left. + \int_w^{f_{i+1}(w)} \frac{(1-\mathrm{e}^{-2\mu(1-y)}) a_i(y)\,\mathrm{d}y}{2\mu} + \int_{f_{i+1}(w)}^1 r_{i+1}(w) a_i(y)\,\mathrm{d}y \right] \mathrm{d}w.
\end{aligned}
$$

Dividing by $\mathrm{e}^{(\lambda+\mu)x}$ and differentiating with respect to x, after some simplification we obtain

$$
\begin{aligned}
r'_{i+1}(x) = \mu r_{i+1}(x) \\
- \frac{\lambda}{\bar{a}_i} \left(\int_{f_i^{-1}(x)}^{x} \left[\frac{1-\mathrm{e}^{-2\mu(1-x)}}{2\mu} - r_{i+1}(x) \right] a_i(y)\,\mathrm{d}y \right. \\
\left. + \int_x^{f_{i+1}(x)} \left[\frac{1-\mathrm{e}^{-2\mu(1-y)}}{2\mu} - r_{i+1}(x) \right] a_i(y)\,\mathrm{d}y \right). \qquad (1.8)
\end{aligned}
$$

The functions f_{i+1} and r_{i+1} can be calculated numerically from Eqs. (1.7) and (1.8) using the boundary conditions $f_{i+1}(1) = 1$ and $r_{i+1}(1) = 0$. Given $r_{i+1}(x)$ and $f_{i+1}(x)$ for a sequence of values $x \in \{x_0, x_0+h, x_0+2h, \ldots, 1\}$, we can evaluate $r_{i+1}(x_0 - h)$ using a numerical procedure to solve Eq. (1.8). We can then evaluate $f_{i+1}(x_0 - h)$ directly from Eq. (1.7).

Having calculated f_{i+1}, we can then estimate $a_{i+1}(x)$, the probability that a male using the threshold profile f_{i+1} is still searching at age x given the threshold profile and age profile of females, (f_i, a_i). A male of age x will leave the population of searchers in the time interval $[x, x+\delta]$ if he either finds a mate or dies in that time interval. Analogous calculations to the ones used to obtain Eq. (1.1) lead to the differential equation

$$
a'_{i+1}(x) = -a_{i+1}(x) \left[\mu + \frac{\lambda}{\bar{a}_i} \int_{f_i^{-1}(x)}^{f_{i+1}(x)} a_i(y)\,\mathrm{d}y \right]. \qquad (1.9)
$$

Equation (1.9) can be solved using the boundary condition $a_{i+1}(0) = 1$ and using a numerical procedure to evaluate $a_{i+1}(x)$ at $x = 0, h, 2h, \ldots, 1$.

It should be noted that this model can be relatively easily modified to allow the mortality rate of individuals to depend on their status (either single or paired), but not on sex. In order to generalize this model to one in which the mortality rate can depend on sex, we have to generalize the model considered above to allow asymmetries between the sexes. Asymmetric problems will be considered in Sect. 1.5.

1.4 A Model in Which the Interaction Rate Depends on the Proportion of Adults Searching

The model presented in Sect. 1.2 assumes that as long as the OSR is equal to 1 individuals meet prospective mates at a constant rate regardless of the strategy profile used (i.e. independently of the proportion of adult individuals who are searching for a partner). One might think of this model as describing a population in which all the singles are concentrated in a particular area (i.e. a type of "singles bar" model). We might consider a model under which the adult population mixes randomly. In this case, we assume that when a male meets a female the probability of her being single is equal to the proportion of females searching for a mate. In reality, it seems likely that the rate of finding prospective mates would be found at an increasing rate as the proportion of adults searching increases. However, it would be realistic to assume that singles can concentrate their search in such a way that the probability of an encounter being with another single is greater than the proportion of the opposite sex who are single. Hence, the two models described above define the two extremes of a spectrum for modelling encounters between searchers.

For ease of presentation, we only consider the "randomly mixing" model under which the rate of meeting prospective mates is proportional to the fraction of individuals of the opposite sex searching for a mate. As in Sect. 1.2, we only consider symmetric equilibria of symmetric games of this form. We define an iterative procedure $(f_{i+1}, a_{i+1}) = H(f_i, a_i)$, where f_{i+1} defines the best response of an individual (without loss of generality we may assume a male) when the threshold and age profiles of females are given by f_i and a_i, respectively. As before, define $\overline{a}_i = \int_0^1 a_i(x)\,\mathrm{d}x$. It is assumed that the rate at which individuals meet prospective partners is $\lambda \overline{a}_i$.

Firstly, we define the best response, f_{i+1}. Suppose a male is still searching at age x. The optimal expected future reward from search is equal to the reward obtained by accepting the presently oldest acceptable female, i.e. $1 - f_{i+1}(x)$. Suppose the next encounter with a single female occurs at age W. The probability density function of W is given by $p(w|W > x) = \lambda \overline{a}_i \mathrm{e}^{-\lambda \overline{a}_i (w-x)}$, for $x \le w < 1$. As before, there is an atom of probability at $w = 1$ of mass equal to the probability that the male does not meet another available female given that he/she is still searching at age x. In this case, the male's reward from search is defined to be 0. Suppose the age of the female is y. The pairing is mutually acceptable if $y \in [f_i^{-1}(w), f_{i+1}(w)]$. If $y \in [f_i^{-1}(w), w]$, then they obtain a reward of $1 - w$. If $y \in (w, f_{i+1}(w)]$, then they obtain a reward of $1 - y$. In all other cases, the future expected reward of the male from search is $1 - f_{i+1}(w)$. Conditioning on the age of the female and taking the expected value, it follows that

$$1 - f_{i+1}(x) = \lambda \mathrm{e}^{\lambda \overline{a}_i x} \int_x^1 \mathrm{e}^{-\lambda \overline{a}_i w} \Bigg[[1 - f_{i+1}(w)] \int_0^{f_i^{-1}(w)} a_i(y)\,\mathrm{d}y + \int_{f_i^{-1}(w)}^{w} (1-w) a_i(y)\,\mathrm{d}y$$
$$+ \int_w^{f_{i+1}(w)} (1-y) a_i(y)\,\mathrm{d}y + \int_{f_{i+1}(w)}^1 [1 - f_{i+1}(w)] a_i(y)\,\mathrm{d}y \Bigg] \mathrm{d}w.$$

Dividing by $e^{\lambda x}$ and differentiating with respect to x, after some simplification we obtain

$$f'_{i+1}(x) = \lambda \left[[f_{i+1}(x) - w] \int_{f_i^{-1}(w)}^{w} a_i(y)\,dy + \int_{w}^{f_{i+1}(w)} [f_{i+1}(x) - y] a_i(y)\,dy \right]. \quad (1.10)$$

Using the boundary condition $f_{i+1}(1) = 1$, we can estimate $f_{i+1}(x)$ for $x = 1-h, 1-2h, \ldots, 0$ by solving Eq. (1.10) numerically.

Having calculated f_{i+1}, we now estimate a_{i+1}, where $a_{i+1}(x)$ is the probability that a male using the optimal response is still searching at age x. This male finds prospective mates at rate $\lambda \overline{a}_i$. Given an optimally responding male of age x meets a female of age y, such a pairing is mutually acceptable if and only if $f_i^{-1}(x) \leq y \leq f_{i+1}(x)$. Analogous calculations to the ones used to obtain Eq. (1.1) lead to

$$a'_{i+1}(x) = -\lambda a_{i+1}(x) \int_{f_i^{-1}(x)}^{f_{i+1}(x)} a_i(y)\,dy. \quad (1.11)$$

We can estimate $a_{i+1}(x)$ for $x = h, 2h, \ldots, 1$ using the boundary condition $a_{i+1}(0)=1$ and solving Eq. (1.11) numerically.

1.5 An Asymmetric Game

In this section we assume that males are fertile for a period of t units, while females are fertile for 1 unit of time. Also, young males enter the adult population at a rate R times the rate at which young females enter the adult population. Without loss of generality, it may be assumed that $t \geq 1$. The OSR r depends on the strategy profile used. It is assumed that there is no mortality.

As stated earlier, there is an intrinsic problem with the formulation of the original model. Although the ISR is one, when the strategy used depends on sex, the OSR may differ from 1 (see Alpern et al. [4]). Suppose $r \neq 1$ and individual males meet females at the same rate as which individual females meet males. It follows that the ratio of the number of times a male meets a female to the number of times a female meets a male must differ from 1. This is clearly a contradiction.

In order to generalize the model, we assume that the rate at which singles meet other singles (of either sex) is λ_0. It follows that the rate at which single females meet prospective mates is λ_f, where $\lambda_f = \frac{\lambda_0 r}{1+r}$. Similarly, the rate at which single males meet prospective mates is λ_m, where $\lambda_m = \frac{\lambda_0}{1+r}$. This satisfies the Fisher condition (see Houston and McNamara [7]) that the ratio of the number of times a male meets a female to the number of times a female meets a male must be equal to 1.

In the case of the symmetric problem, this problem is sidestepped by the assumption that the equilibrium is symmetric with respect to sex. This means that at equilibrium $\lambda_f = \lambda_m = \lambda = \frac{\lambda_0}{2}$. It follows from this argument that a fixed point of the policy iteration algorithm defined in Sect. 1.2 is also a fixed point in an analogous algorithm where the encounter rates depend on the strategy profiles used as described directly above.

As in the original model, it is assumed that the payoff obtained on forming a pair is the length of time for which both members of the pair remain fertile. From the form of this payoff function and the optimality criterion, it follows that at equilibrium both males and females should use sex specific threshold strategies. Also, using an argument similar to the one given in Sect. 1.2, it is easy to show that these thresholds are non-decreasing in age and individuals should always accept a prospective partner who will remain fertile for a longer period than themselves.

Suppose males use the threshold profile $f(x)$, for $0 \leq x \leq t$, and females use the threshold profile $g(y)$, for $0 \leq y \leq 1$. We use the definition of the inverse function adapted to this problem, i.e. $f^{-1}(x) = 0$, for $x \leq f(0)$, otherwise f^{-1} is the standard standard inverse function. Define $a(x)$ to be the number of searching males of age x relative to the number of females of age 0. Hence, $a(0) = R$. Define $b(y)$ to be the proportion of females still searching at age y. Thus, $b(0) = 1$. Let $\overline{a} = \int_0^t a(x)\,\mathrm{d}x$ and $\overline{b} = \int_0^1 b(y)\,\mathrm{d}y$. It follows that $r = \overline{a}/\overline{b}$. The density function of the age of searching males is given by $\hat{a}$, where $\hat{a}(x) = a(x)/\overline{a}$. Similarly, the density function of the age of searching females is given by $\hat{b}(y) = b(y)/\overline{b}$.

We now define a policy iteration algorithm to define an equilibrium profile (f,g). In order to do this, we must define initial threshold and age profiles for both sexes. Let f_1, a_1, g_1 and b_1 be the initial male threshold profile, male age profile, female threshold profile and female age profile, respectively. We define an iterative procedure $(f_{i+1}, a_{i+1}, g_{i+1}, b_{i+1}) = H(f_i, a_i, g_i, b_i)$ as described below. Assume that initially females will not mate and males are ready to mate with any female. Thus the initial threshold and age profiles of males are given by $f_1(x) = 1$ and $a_1(x) = R$, for all $0 \leq x \leq t$. The initial threshold and age profiles of females are given by $g_1(y) = 0$ and $b_1(y) = 1$, for all $0 \leq y \leq 1$. The initial OSR is $r_1^f = Rt$ (the superfix here indicates that this is the OSR used to calculate the following female threshold and age profiles in the procedure). In general, $r_i^f = \overline{a}_i/\overline{b}_i$. Given the present pair of profiles for males (f_i, a_i) and OSR r_i^f, we calculate the best response of a female g_{i+1} and the probability an optimally behaving female is still searching at age y, $b_{i+1}(y)$ as follows:

The rate at which females find prospective mates is $\lambda_i^f = \frac{\lambda_0 r_i^f}{1+r_i^f}$. Suppose a female is still searching at age y. Her optimal reward from future search is given by the length of time for which the presently oldest acceptable male remains fertile, i.e. $t - g_{i+1}(y)$. Suppose the next prospective mate is of age x and appears when the female is of age w. The male will be fertile for at least as long as the female if $t - x \geq 1 - w \Rightarrow x \leq t - 1 + w$. Hence, when $f_i^{-1}(w) \leq x \leq t - 1 + w$, then a pair is formed and both individuals obtain a reward of $1 - w$. If $t - 1 + w \leq x \leq g_{i+1}(w)$, then a pair is formed and both individuals obtain a reward of $t - x$. Otherwise, both individuals continue searching. By conditioning on the age of the female at the next encounter with a male and his age, we obtain

$$t-g_{i+1}(y) = \int_y^1 \frac{\lambda_i^f \mathrm{e}^{-\lambda_i^f(w-y)}}{\overline{a}_i} \left[\int_0^{f_i^{-1}(w)} [t-g_{i+1}(w)]a_i(x)\,\mathrm{d}x + \int_{f_i^{-1}(w)}^{t-1+w} [1-w]a_i(x)\,\mathrm{d}x \right.$$
$$\left. + \int_{t-1+w}^{g_{i+1}(w)} [t-x]a_i(x)\,\mathrm{d}x + \int_{g_{i+1}(w)}^{1} [t-g_{i+1}(w)]a_i(x)\,\mathrm{d}x \right] \mathrm{d}w.$$

Dividing by $\mathrm{e}^{\lambda_i^f y}$ and differentiating with respect to y, after some simplification we obtain

$$g'_{i+1}(y) = \frac{\lambda_i^f}{\overline{a}_i} \left[[g_{i+1}(y)+1-y-t] \int_{f_i^{-1}(y)}^{t-1+y} a_i(x)\,\mathrm{d}x + \int_{t-1+y}^{g_{i+1}(y)} [g_{i+1}(y)-x]a_i(x)\,\mathrm{d}x \right]. \tag{1.12}$$

Using Eq. (1.12) and the boundary condition $g_{i+1}(1) = t$, we can numerically calculate $g_{i+1}(y)$ for $y \in \{1-h, 1-2h, \ldots, 0\}$.

Now we consider the probability that a female using this optimal response will still be searching at age y. This is denoted by $b_{i+1}(y)$. Such a female meets prospective partners at rate λ_i^f. If she meets a male of age x when she is y years old, mating occurs if and only if $f_i^{-1}(y) \le x \le g_{i+1}(y)$. Using an argument analogous to the one to derive Eq. (1.1), it follows that

$$b_{i+1}(y) = -\frac{b_{i+1}(y)\lambda_i^f}{\overline{a}_i} \int_{f_i^{-1}(y)}^{g_{i+1}(y)} a_i(x)\,\mathrm{d}x. \tag{1.13}$$

Using Eq. (1.13) and the boundary condition $b_{i+1}(0) = 1$, we can numerically calculate $b_{i+1}(y)$ for $y \in \{h, 2h, \ldots, 1\}$.

We then calculate the optimal response of a male given that the female threshold and age profiles are g_{i+1} and b_{i+1}, respectively and the OSR is assumed to be given by $r_i^m = \overline{a}_i/\overline{b}_{i+1}$. The rate at which males find females is thus $\lambda_i^m = \frac{\lambda_0}{1+r_i^m}$. It should be noted that this is not by definition the OSR when males use the threshold profile f_i and females use the profile g_{i+1}.

Suppose a male is still searching at age x. His optimal reward from future search is given by the length of time for which the presently oldest acceptable female remains fertile, i.e. $1-f_{i+1}(x)$. Suppose the next prospective mate is of age y and appears when the male is of age w. From the definition of the inverse of the threshold function used here, the youngest female who will accept such a male, is of age $g_{i+1}^{-1}(w)$. A female should accept a male who will be fertile for a longer period than her. It follows that for $w \le t-1$, then $g_{i+1}^{-1}(w) = 0$. Also, for $w > t-1$, we have $1-g_{i+1}^{-1}(w) \ge t-w$. Hence, $g_{i+1}^{-1}(w) \le 1+w-t$. It follows that $g_{i+1}^{-1}(w) \le \max\{0, 1+w-t\}$. When $g_{i+1}^{-1}(w) \le y \le 1+w-t$, then a pair is formed and both individuals obtain a reward of $t-w$. If $1+w-t \le y \le f_{i+1}(w)$, then a pair is formed and both individuals obtain a reward of $1-y$. Otherwise, both individuals continue searching. By conditioning on the age of the female at the next encounter with a male and his age, we obtain

$$1-f_{i+1}(x) = \int_x^1 \frac{\lambda_i^m \mathrm{e}^{-\lambda_i^m(w-x)}}{\bar{b}_i} \left[\int_0^{g_{i+1}^{-1}(w)} [1-f_{i+1}(w)] b_i(y)\,\mathrm{d}y \right. + \int_{g_{i+1}^{-1}(w)}^{\max\{0,1+w-t\}} [t-w] b_i(y)\,\mathrm{d}y + \int_{\max\{0,1+w-t\}}^{f_{i+1}(w)} [1-y] b_i(y)\,\mathrm{d}y \left. + \int_{f_{i+1}(w)}^{1} [1-f_{i+1}(w)] b_i(y)\,\mathrm{d}y \right] \mathrm{d}w.$$

Dividing by $\mathrm{e}^{\lambda_i^m x}$ and differentiating with respect to x, after some simplification we obtain that for $x < t-1$

$$f'_{i+1}(x) = \frac{\lambda_i^m}{\bar{b}_i} \int_0^{f_{i+1}(x)} [f_{i+1}(x) - y] b_i(y)\,\mathrm{d}y \tag{1.14}$$

and for $x > t-1$

$$f'_{i+1}(x) = \frac{\lambda_i^m}{\bar{b}_i} \left[[t-x-1+f_{i+1}(x)] \int_{g_{i+1}^{-1}(x)}^{1+x-t} b_i(y)\,\mathrm{d}y + \int_{1+x-t}^{f_{i+1}(x)} [f_{i+1}(x)-y] b_i(y)\,\mathrm{d}y \right]. \tag{1.15}$$

Using Eqs. (1.14) and (1.15), together with the boundary condition $f_{i+1}(t) = 1$ and the continuity of f_{i+1}, we can numerically estimate $f_{i+1}(x)$ for $x \in \{t-h, t-2h, \ldots, 0\}$.

We define $a_{i+1}(x)$ to be the probability that a male using this best response is still searching at age x. Using an argument analogous to the one used in deriving Eq. (1.1), we obtain

$$a'_{i+1}(x) = -\frac{a_{i+1}(x)\lambda_i^m}{\bar{b}_i} \int_{g_{i+1}^{-1}(y)}^{f_{i+1}(y)} b_i(y)\,\mathrm{d}y. \tag{1.16}$$

Using Eq. (1.16) and the boundary condition $a_{i+1}(0) = R$, we can numerically calculate $a_{i+1}(x)$ for $x \in \{h, 2h, \ldots, t\}$.

We have thus updated each of the four profiles and defined the mapping H. Suppose the mapping H has a fixed point (f,a,g,b). In this case, the best response of females to the male threshold and age profiles (f,a) is to use the threshold profile g. The probability that such an optimally behaving female is still searching at age y is given by $b(y)$. Since this female is using the same strategy as the other females, b gives the age profile of the females. Using a similar argument, f is the optimal response of a male to (g,b) and a gives the age profile of the males. It follows that the OSR defined by the iterative procedure is equal to the actual OSR given the quartet of profiles (f,a,g,b). Hence, any fixed point of the mapping H is an equilibrium of this asymmetric game.

It should be noted that the algorithm described above must be used when looking for a asymmetric equilibrium of a symmetric problem. The algorithm described in Sect. 1.2 generally does not work in this case, since the OSR at such an equilibrium may well differ from 1 and so the rate at which prospective mates are found is sex dependent.

Also, the model presented above can easily be modified to introduce constant mortality rates and encounter rates which are dependent on the proportion of adult individuals who are searching for a mate (as described in Sects. 1.3 and 1.4, respectively).

1.6 Numerical Results

A MATLAB programme was written to estimate the equilibrium threshold rule and age profiles at points $0, h, 2h, \ldots, 1$ based on the appropriate difference equations, using the trapezium rule to calculate the required integrals and double precision arithmetic. The inverse to a threshold rule was estimated at the same points using linear interpolation. Comparison of different step sizes suggested that using a step size of $h = 10^{-4}$ allowed estimation of the threshold and age profile to at least three decimal places for $\lambda \leq 50$. The maximum value of the second derivative of the threshold profile is increasing in λ and for larger values of λ a more accurate procedure would be necessary to achieve the same accuracy.

1.6.1 Model with Mortality

Table 1.1 gives the expected reward and initial threshold (in brackets) at equilibrium for various mortality rates and interaction rates. The case $\mu = 0$ corresponds to the original model. Figure 1.2 illustrates the threshold rules evolved for $\lambda = 20$ and various mortality rates. Figure 1.3 illustrates the corresponding age profiles. When the mortality rate increases, we expect that individuals become less choosy. It would thus seem that the threshold used would be increasing in the mortality rate. This seems to be the case when the mortality rate is relatively low. However, the threshold profile on its own does not tell us how choosy individuals are. Figure 1.2 shows that

Table 1.1 Expected rewards and initial threshold (in round brackets) at equilibrium for various mortality rates, μ, and interaction rates, λ

	$\mu = 2$	$\mu = 1$	$\mu = 0.5$	$\mu = 0.2$	$\mu = 0.1$	$\mu = 0$
$\lambda = 2$	0.109 (0.858)	0.205 (0.737)	0.292 (0.655)	0.365 (0.605)	0.395 (0.589)	0.426 (0.574)
$\lambda = 5$	0.168 (0.721)	0.311 (0.514)	0.445 (0.411)	0.563 (0.362)	0.611 (0.348)	0.665 (0.335)
$\lambda = 10$	0.201 (0.592)	0.366 (0.343)	0.527 (0.250)	0.675 (0.213)	0.736 (0.204)	0.805 (0.195)
$\lambda = 20$	0.221 (0.462)	0.395 (0.218)	0.574 (0.148)	0.740 (0.123)	0.810 (0.116)	0.895 (0.105)

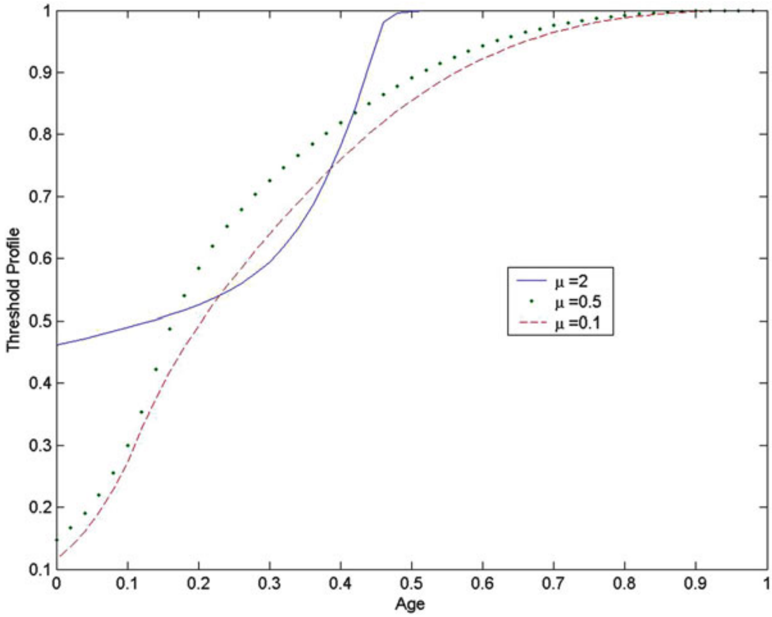


Fig. 1.2 Effect of mortality rate on the equilibrium threshold profile ($\lambda = 20$)

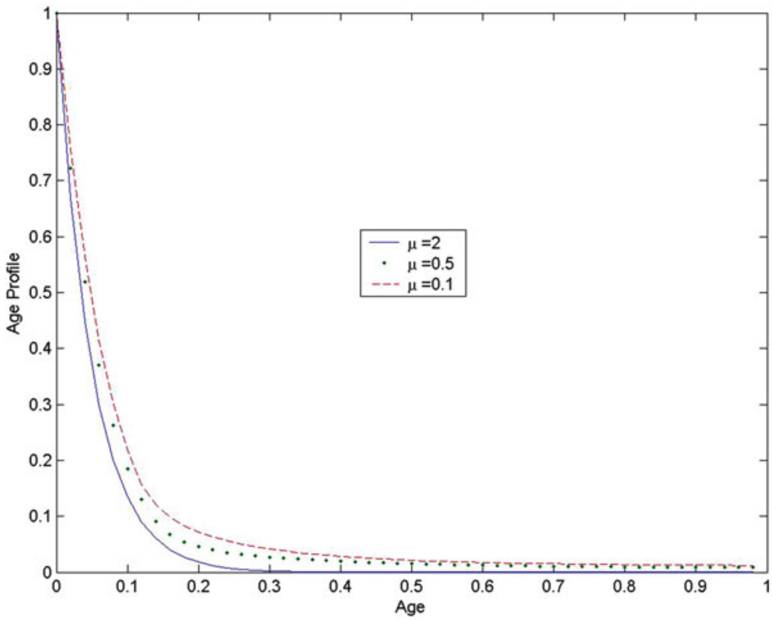


Fig. 1.3 Effect of mortality rate on the equilibrium age profile ($\lambda = 20$)

Table 1.2 Expected rewards at equilibrium under the singles bar model and model of a randomly mixing population

λ	Original (singles bar) model	Randomly mixing
2	0.4264	0.3082
5	0.6645	0.4636
10	0.8054	0.5775
20	0.8954	0.6774

for ages between 0.2 and 0.4 the threshold used at equilibrium when $\mu = 2, \lambda = 20$ is lower than the threshold used in the case where $\mu = 0.5, \lambda = 20$ (i.e. it seems that increasing mortality increases the choosiness of individuals at some ages). However, in the case $\mu = 2$, $\lambda = 20$ (i.e. relatively high interaction and mortality rates), there are virtually no individuals of age greater than 0.3 in the mating pool. Individuals always accept a prospective mate of age below 0.462 and hence at equilibrium the probability of an individual rejecting a prospective partner is virtually zero.

When the mortality rate is high, the age profile of the mating pool is very highly concentrated on young ages. As long as a young individual survives, he/she will almost certainly mate with the next perspective partner and the expected payoff obtained is much more dependent on the mortality rate than the maximum length of time for which an individual can remain fertile. Thus young individuals will increase their threshold only very slowly. This remains true until an individual attains the age at which the youngest individuals begin rejecting him/her. At this point the probability of rejection increases very rapidly due to the shape of the age profile. It follows that for high mortality rates the equilibrium threshold profile is similar to a step function. However, at equilibrium the probability of meeting an unacceptable partner is virtually zero. Thus an individual who always mates with the first prospective partner would have virtually the same reward at equilibrium as an individual using the equilibrium threshold. Hence, at equilibrium there would be very low selection pressure on the threshold used.

1.6.2 Model with Interaction Rates Dependent on the Proportion of Searchers

Table 1.2 gives the expected reward from search at equilibrium for the original (singles bar) model and the model of a randomly mixing population for various interaction rates. Since it is assumed that there is no mortality of fertile individuals, the initial threshold is simply one minus the expected reward.

It should also be noted that if a proportion $\overline{a}$ of the adult population are searching for a mate at equilibrium, such an equilibrium is also stable in a game where the interaction rate is fixed to be $\lambda\overline{a}$ when the sex ratio is one. However, the dynamics of the policy iteration procedure corresponding to these two problems are different.

Table 1.3 Expected reward of females, males (in round brackets) and OSR (in square brackets) when $\lambda_0 = 4$

	$R = 0.5$	$R = 1$	$R = 2$
$T = 1$	0.253 (0.506) [0.293]	0.427 (0.427) [1.000]	0.506 (0.253) [3.413]
$T = 2$	0.356 (0.714) [0.602]	0.585 (0.586) [2.412]	0.653 (0.327) [8.656]
$T = 5$	0.412 (0.826) [1.590]	0.658 (0.659) [7.148]	0.717 (0.359) [26.004]

1.6.3 Asymmetric Model

Table 1.3 gives the expected reward of females, males (in round brackets) and OSR [in square brackets] when $\lambda_0 = 4$. It should be noted that the case $T = R = 1$ corresponds to the original symmetric model with $\lambda = 2$. Various values of λ_0 were used for this parameter set and the equilibrium found was always symmetric. Also, the problem with $T = 1$ and $R = 0.5$ is equivalent to the problem with $T = 1$ and $R = 2$ with the roles of the sexes reversed.

The sum of the rewards of females is by definition equal to the sum of the rewards of males. It follows that the ratio of the expected reward of a female to the expected reward of a male is equal to the ratio of the number of males entering the adult population to the number of females entering the adult population (i.e. R). The minor deviations from this rule are due to numerical errors in the iterative procedure.

1.7 Conclusion

This paper has generalized a model of mate choice with age based preferences introduced by Alpern et al. [4] by (a) introducing a uniform mortality rate, (b) allowing the rate at which prospective mates are found to depend on the proportion of individuals searching, (c) considering models which are asymmetric with respect to sex.

It may well be interesting to generalize the model to allow variable mortality rates. It seems reasonable to assume that the mortality rate increases with age and in this case it is expected that the equilibrium strategy will be of the same form, i.e. a threshold strategy according to which younger mates are always preferred. However, it is possible that the mortality rate is higher for young adults than for middle-aged adults. In this case, the equilibrium strategy may well be of a more complex form, since a middle-age mate may be preferable to a young mate.

It would also be interesting to look at the interplay between resource holding potential (RHP) (see Parker [16]) and age. For example, as a human ages his/her RHP (i.e. qualifications, earnings, wealth, social position) increases. Hence, from this point of view the attractiveness of an individual may well increase over time. However, from the point of view of the model considered here, older individuals are less attractive as mates as they will be fertile for a shorter period.

Mauck et al. [10] note that the average number of surviving offspring per brood in a population of storm petrels is increasing in the age of the partners. There are various explanations for this (e.g. fitter individuals may live longer, individuals (or pairs) may become increasingly efficient at rearing offspring, or simply that older pairs invest more in reproduction than in survival). However, this may mean that age preferences may be to some degree homotypic. This is due to the fact that an old individual may well prefer an old partner, since it is more important for them to maximize their present reproduction rate. Younger partners may well prefer young partners, since they have time to adapt to each other and perfect their method of rearing. It would be interesting to see how the form of the equilibrium function depends on these factors (by considering a wider range of payoff functions). Also, our model assumes that the ages of prospective partners are independent. It may well be that individuals concentrate their search on prospective partners of a "suitable age".

According to our model, individuals only mate once during their life, whereas in reality pairs can divorce or an individual can remate after the death of a partner. Since individuals may be of different qualities, it might be optimal for an individual to divorce a partner who turns out not to be good as expected (see McNamara and Forslund [13], McNamara et al. [14]). It is intended that future work will extend the model to allow individuals to remate after a partner dies or becomes infertile.

Of course, mate choice may depend on other factors, such as attractiveness (common preferences) and compatibility (homotypic preferences). It would be interesting to see how these factors might interact with age. Due to the necessarily complex nature of such models, it would seem that simulations based on replicator dynamics would be a sensible approach to such problems (see Nowak [15]).

Acknowledgements I would like to thank Prof. Steve Alpern for the conversations, advice and encouragement that have aided my work on this paper.

References

1. Alpern, S., Katrantzi, I.: Equilibria of two-sided matching games with common preferences. Eur. J. Oper. Res. **196**(3), 1214–1222 (2009)
2. Alpern, S., Reyniers, D.: Strategic mating with homotypic preferences. J. Theor. Biol. **198**, 71–88 (1999)
3. Alpern, S., Reyniers, D.: Strategic mating with common preferences. J. Theor. Biol. **237**, 337–354 (2005)
4. Alpern, S., Katrantzi, I., Ramsey, D.: Partnership formation with age-dependent preferences. Eur. J. Oper. Res. (2012)
5. Burdett, K., Coles, M.G.: Long-term partnership formation: marriage and employment. Econ. J. **109**, 307–334 (1999)
6. Collins, E.J., McNamara, J.M.: The job-search problem with competition: an evolutionarily stable strategy. Adv. Appl. Prob. **25**, 314–333 (1993)
7. Houston, A.I., McNamara, J.M.A.: Self-consistent approach to paternity and parental effort. Phil. Trans. R. Soc. Lond. B **357**, 351–362 (2002)

8. Janetos, A.C.: Strategies of female mate choice: a theoretical analysis. Behav. Ecol. Sociobiol. **7**, 107–112 (1980)
9. Johnstone, R.A.: The tactics of mutual mate choice and competitive search. Behav. Ecol. Sociobiol. **40**, 51–59 (1997)
10. Mauck, R.A.: Huntington, C.E., Grubb, T.C.: Age-specific reproductive success: evidence for the selection hypothesis. Evolution **58**(4), 880–885 (2004)
11. Mazalov, V., Falko, A.: Nash equilibrium in two-sided mate choice problem. Int. Game Theory Rev. **10**(4), 421–435 (2008)
12. McNamara, J.M., Collins, E.J.: The job search problem as an employer-candidate game. J. Appl. Prob. **28**, 815–827 (1990)
13. McNamara, J.M., Forslund, P.: Divorce rates in birds: prediction from an optimization model. Am. Nat. **147**, 609–640 (1996)
14. McNamara, J.M., Forslund, P., Lang, A.: An ESS model for divorce strategies in birds. Philos. Trans. R. Soc. Lond. B Biol. Sci. **354**, 223–236 (1999)
15. Nowak, M.: Evolutionary Dynamics: Exploring the Equations of Life. Belknap Press, Cambridge, Massachusetts (2006)
16. Parker, G.A.: Assessment strategy and the evolution of animal conflicts. J. Theor. Biol. **47**, 223–243 (1974)
17. Parker, G.A.: Mate quality and mating decisions. In: Bateson, P. (ed.) Mate Choice, pp. 227–256. Cambridge University Press, Cambridge, UK (1983)
18. Ramsey, D.M.: A large population job search game with discrete time. Eur. J. Oper. Res. **188**, 586–602 (2008)
19. Real, L.A.: Search theory and mate choice. I. Models of single-sex discriminination. Am. Nat. **136**, 376–404 (1990)
20. Real, L.A.: Search theory and mate choice. II. Mutual interaction, assortative mating, and equilibrium variation in male and female fitness. Am. Nat. **138**, 901–917 (1991)
21. Shimer, R., Smith, L.: Assortative matching and search. Econometrica **68**, 343–370 (2000)
22. Smith, L.: The marriage model with search frictions. J. Pol. Econ. **114**, 1124–1144 (2006)

Chapter 2
Signalling Victory to Ensure Dominance: A Continuous Model

Mike Mesterton-Gibbons and Tom N. Sherratt

Abstract A possible rationale for victory displays—which are performed by the winners of contests but not by the losers—is that the displays are attempts to decrease the probability that the loser of a contest will initiate a future contest with the same individual. We explore the logic of this "browbeating" rationale with a game-theoretic model, which extends previous work by incorporating the effects of contest length and the loser's strategic response. The model predicts that if the reproductive advantage of dominance over an opponent is sufficiently high, then, in a population adopting the evolutionarily stable strategy or ESS, neither winners nor losers signal in contests that are sufficiently short; and only winners signal in longer contests, but with an intensity that increases with contest length. These predictions are consistent with the outcomes of recent laboratory studies, especially among crickets, where there is now mounting evidence that eventual winners signal far more frequently than losers after fighting, and that post-conflict displays are more likely to be observed after long contests.

Keywords Contest behavior • Evolutionarily stable strategies • Post-conflict displays

M. Mesterton-Gibbons (✉)
Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, FL 32306-4510, USA
e-mail: mesterto@math.fsu.edu

T.N. Sherratt
Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada
e-mail: sherratt@connect.carleton.ca

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_2,

2.1 Introduction

Bower [4] defines a victory display as a display performed by the winner of a contest but not by the loser. He offers in essence two possible adaptive explanations of their function: that they are an attempt to advertise victory to other members of a social group that do not pay attention to contests, or cannot otherwise identify the winner, and thus alter their behavior ("function within the network"), or that they are an attempt to decrease the probability that the loser of a contest will initiate a new contest with the same individual ("function within the dyad"). In an earlier paper [20], we called the first rationale advertising, and the second one browbeating; and we used game-theoretic models to explore the logic of both rationales. These models showed that both rationales are logically sound; moreover, all other things being equal, the intensity of victory displays will be highest through advertising in groups where the reproductive advantage of dominating an opponent is low, and highest through browbeating in groups where the reproductive advantage of dominance is high.

Here we further consider the browbeating rationale, leaving the case of an advertising rationale for future work. By the browbeating rationale, a victory display is an attempt to decrease the probability that the loser of a contest will initiate a new contest with the same individual. As long as there is a chance that the loser will challenge the winner to another fight in the future, the winner has won a battle for dominance, but not the war. If, on the other hand, the victory ensures that the loser will never challenge, then victory is tantamount to dominance. Thus browbeating is an attempt to ensure that victory equals dominance, and the essence of modelling this phenomenon is to observe a distinction between losing and subordination.

Although we have previously demonstrated that browbeating is a plausible mechanism for victory displays, our earlier model assumed—as opposed to predicted—that a loser does not display, and hence dodged the question of why victory displays should be respected. Moreover, our original model assumed that all contests were of equal length, which leaves open the question as to whether individuals should be more or less likely to signal their dominance after long (close) fights than after short (one-sided) fights.

Accordingly, our purpose here is twofold. First, it is to relax the assumption that the loser does not display. Second, our purpose is also to address the context-dependent nature of the display, which lies outside the scope of our original browbeating model. Specifically, in a recent study investigating fighting behavior in the spring field cricket, *Gryllus veletis*, Bertram et al. [3] found that the intensity of post-conflict signals (aggressive song rate and body jerk rate) were dependent on whether the individual was a winner or loser (with winners signalling more intensely than losers) and on the duration of the contest (with short fights producing less intense signals). Ting et al. (unpublished) came to similar conclusions after analysing the outcomes of fights in the fall field cricket, *Gryllus pennsylvanicus*. Likewise, post-conflict displays in the black-capped chickadee, *Poecile atricapillus*—albeit more common among losers than among winners—were more likely to occur after highly aggressive contests [15]. Collectively, these recent studies suggest that context dependency might be a general feature of post-conflict displays. Clearly, if mathematical models are to be of value in understanding victory displays, then they should help explain not only the display, but also who displays, and with what intensity. Here we present a simple model that addresses both phenomena.

2.2 Mathematical Model

The dominance status of a victor relative to the vanquished is determined by a combination of dominance and what we refer to as "non-subordination." A contest outcome is one of dominance if one individual subordinates to the other but the second does not, and of non-subordination if neither defers to the other. To capture the idea that dominance over an opponent contributes more to (long-term) fitness than non-subordination, which contributes more than being dominated, let there be a fitness benefit of 1 for dominating the other individual, of b for non-subordination, and of 0 for being dominated. Thus b, a dimensionless parameter assumed to satisfy

$$0 < b < 1 \tag{2.1}$$

throughout, is an inverse measure of the reproductive advantage of dominance, which is greatest in the limit as $b \to 0$ and least in the limit as $b \to 1$.

We assume that a fight is inevitable; and that its cost can be ignored, by virtue of being the same for both individuals. Neither animal has information about its own or its opponent's strength. So we assume that each animal is equally likely to win, since each has probability $\frac{1}{2}$ of being stronger. However, the length T of the contest is experienced, and the shorter it is, the more likely it is that the winner is the stronger animal.

These assumptions are most nearly satisfied when a fight is provoked between two new neighbors with no prior knowledge of one another's strength and hence no established dominance relation. Of course, they are idealizations. Yet not only do they yield a tractable analytical model, but also they enable us to avoid confounding the evolution of post-conflict displays with the evolution of basic aggression thresholds per se.

Let strategy $u = (u_1, u_2)$ mean that a u-strategist displays with intensity u_1 as winner but u_2 as loser. Let $q(w,l)$ be the probability that a display of intensity w by the winner against a display of intensity l by the loser elicits submission on the part of the loser. (Thus a u-strategist wins with probability $\frac{1}{2}$ against a v-strategist, but wins and dominates only with the smaller probability $\frac{1}{2}q(u_1, v_2)$.) If a u-strategist wins, then with probability $q(u_1, v_2)$ it also dominates and its payoff is 1; but with probability $1 - q(u_1, v_2)$ it fails to dominate and its payoff is b. Let $c_w(s)$ denote the cost to a winner of displaying with intensity s. Then, conditional upon winning, a u-strategist's payoff is $q(u_1, v_2) \cdot 1 + \{1 - q(u_1, v_2)\} \cdot b - c_w(u_1)$. Likewise, conditional upon losing, a u-strategist's payoff is $q(v_1, u_2) \cdot 0 + \{1 - q(v_1, u_2)\}b - c_l(u_2)$, where $c_l(s)$ denotes the cost to a loser of displaying with intensity s. Multiplying each of the above payoffs by $\frac{1}{2}$ and adding, we find that the reward to a u-strategist in a population of v-strategists is

$$f(u,v) \;=\; \frac{1}{2}\big\{(1-b)q(u_1,v_2) - bq(v_1,u_2) - c_w(u_1) - c_l(u_2)\big\} + b. \tag{2.2}$$

We need to place conditions on the functions c_w, c_l and q. First, for c_w and c_l, it seems reasonable to suppose that $c_w(0) = 0$, $c'_w(s) > 0$, $c''_w(s) \geq 0$ (as in [20]) and $c_l(0) = 0$, $c'_l(s) > 0$, $c''_l(s) \geq 0$. For the sake of simplicity, we satisfy these conditions by taking

$$c_w(s) \;=\; \gamma_w \theta s, \quad c_l(s) \;=\; \gamma_l \theta s \tag{2.3}$$

with

$$\gamma_w \;<\; \gamma_l \tag{2.4}$$

throughout, where $\theta(> 0)$ has the dimensions of INTENSITY^{-1}, so that $\gamma_w(> 0)$ and $\gamma_l(> 0)$ are dimensionless measures of the marginal cost of displaying for a winner and a loser, respectively.

Second, for q, the following seem reasonable: $q(\infty, l) = 1$ for any finite l, and $q(w,l) = \delta$ for all $w \leq l$ where δ is the base probability that winning will lead to dominance—a winner cannot increase its chance of converting its win into dominance unless it is displaying with at least as strong an intensity as the loser. The shorter the contest, the more likely it is that the loser will feel heavily outgunned and concede dominance; hence δ is a decreasing function of contest length T. For the sake of simplicity, we take

$$\delta \;=\; \mathrm{e}^{-T/\mu}, \tag{2.5}$$

where μ is a scaling factor (the length of a contest that would reduce the probability of achieving dominance without a display from 1 to approximately 37 %). We also require $\partial q/\partial w > 0$ and $\partial q/\partial l < 0$ for all $w > l$. Again for the sake of simplicity, we satisfy all conditions on q by taking

$$q(w,l) \;=\; \begin{cases} \delta + (1-\delta)\{1 - \mathrm{e}^{-\theta(w-l)}\} & \text{if} \quad w \geq l \\ \delta & \text{if} \quad w < l \end{cases} \tag{2.6}$$

throughout. Note the asymmetry here: a display by the loser is not a second chance to win the fight. On the contrary, it is merely an attempt to reduce the probability that losing implies subordination.

2.3 ESS Analysis

A strategy v is an evolutionarily stable strategy or ESS in the sense of Maynard Smith [17] if it is uniquely the best reply to itself; that is, in present circumstances, if v_1 is a winner's best reply to a loser's v_2 and v_2 is a loser's best reply to a winner's v_1.

From Appendix A, if the marginal cost of displaying is so high for a winner that $\gamma_w \geq 1-b$, then $v_1 = 0$ (i.e., not displaying) is a winner's best reply to any v_2; and likewise, if the marginal cost of displaying is so high for a loser that $\gamma_l \geq b$, then $v_2 = 0$ is a loser's best reply to any v_1. These are not interesting cases. Accordingly, we assume henceforward that $\gamma_w < 1-b$ and $\gamma_l < b$ invariably hold. That is, we assume $\min(\rho, \zeta) > 1$, where

$$\rho \;=\; \frac{1-b}{\gamma_w}, \quad \zeta \;=\; \frac{b}{\gamma_l}. \tag{2.7}$$

Then, from Appendix A, and in particular from the discussion following (A8), the game defined by (2.2)–(2.6) has a unique ESS if

$$\frac{T}{\mu} \;<\; \max\left\{\ln\left(\frac{\rho}{\rho-1}\right), \ln\left(\frac{\zeta}{\zeta-1}\right)\right\}, \tag{2.8}$$

although it has no ESS if the above inequality is reversed. Subject to (2.8), if also

$$\frac{T}{\mu} \;<\; \min\left\{\ln\left(\frac{\rho}{\rho-1}\right), \ln\left(\frac{\zeta}{\zeta-1}\right)\right\}, \tag{2.9}$$

then from (2.5) and (A5) the ESS is $v = (0,0)$: neither a winner nor a loser displays. If, on the other hand, (2.8) holds with (2.9) reversed, then one of two cases arises. If

$$\ln\left(\frac{\zeta}{\zeta-1}\right) \;<\; \frac{T}{\mu} \;<\; \ln\left(\frac{\rho}{\rho-1}\right), \tag{2.10}$$

then the ESS is still $v = (0,0)$ by the remark following (A6). If

$$\ln\left(\frac{\rho}{\rho-1}\right) \;<\; \frac{T}{\mu} \;<\; \ln\left(\frac{\zeta}{\zeta-1}\right), \tag{2.11}$$

however, then it follows from (A6) that the ESS is given by $\theta v = (\lambda, 0)$, where

$$\lambda \;=\; \ln(\rho\{1-\mathrm{e}^{-T/\mu}\}). \tag{2.12}$$

Thus the relative magnitudes of ρ and ζ determine the ESS. For $\rho > \zeta$ or

$$\frac{\gamma_l}{\gamma_w} \;>\; \frac{b}{1-b}, \tag{2.13}$$

there is no ESS if $T > \mu \ln(\frac{\zeta}{\zeta-1})$; but if $T < \mu \ln(\frac{\zeta}{\zeta-1})$, then the unique ESS is given by $v = (0,0)$ for $T < \mu \ln(\frac{\rho}{\rho-1})$ and by $\theta v = (\lambda, 0)$ for $T > \mu \ln(\frac{\rho}{\rho-1})$, with λ defined by (2.12). If (2.13) is reversed, or $\rho < \zeta$, then the unique ESS for $T < \mu \ln(\frac{\rho}{\rho-1})$ is $v = (0,0)$; and for $T > \mu \ln(\frac{\rho}{\rho-1})$ there is no ESS.

2.4 Discussion

Fighting behavior and their associated signals have been the subject of extensive empirical and theoretical study (see [10, 12]). However, much of this work has focused on the behaviors that occur before and during aggressive interactions, and relatively little is known about behaviors that occur after the outcomes have been decided [4]. Here we have developed and explored a game-theoretic model of post-conflict signalling, seeking to identify who should tend to signal following termination of conflict, with what intensity, and the factors that shape this intensity. We have focused on the hypothesis that post-conflict signalling by the victor serves to reinforce dominance, reducing the chances that the loser will try it on again, although there may be other complementary adaptive explanations for such displays, including advertising of victory to bystanders, and non-adaptive explanations such as emotional release [4].

Post-conflict victory displays [4] have been reported in a range of organisms, including humans [22] and birds [9], but they have been most intensively researched in crickets (Orthoptera). Crickets often perform aggressive songs and body jerks both during and after an agonistic conflict [1, 3, 13]. In a study of the field cricket, *Teleogryllus oceanicus*, Bailey and Stoddart [2] proposed that if the display of a victorious male is sufficiently intense, then it may indicate to the loser that the fight is unlikely to be reversed by further combat, enabling the victor to divert its time and energy to other activities such as mating. Conversely, low signalling intensity of the winner may suggest to the loser that re-engagement could potentially produce a reversal, hence some future reward to the loser. This is precisely the situation we have attempted to model here. Indeed, Bailey and Stoddart [2] went further and argued that the winner's post-conflict display could be used as an indication of the winner's position in a broader dominance hierarchy, showing that hierarchies constructed using an index based on post-conflict signalling correlated well with those produced by more classical methods. Intriguingly, Logue et al. [16] recently reported that contests between male field crickets *Teleogryllus oceanicus* that were unable to sing were more aggressive than interactions between males that were free to signal, supporting the view that signalling can serve to mitigate the costs of fighting in these species.

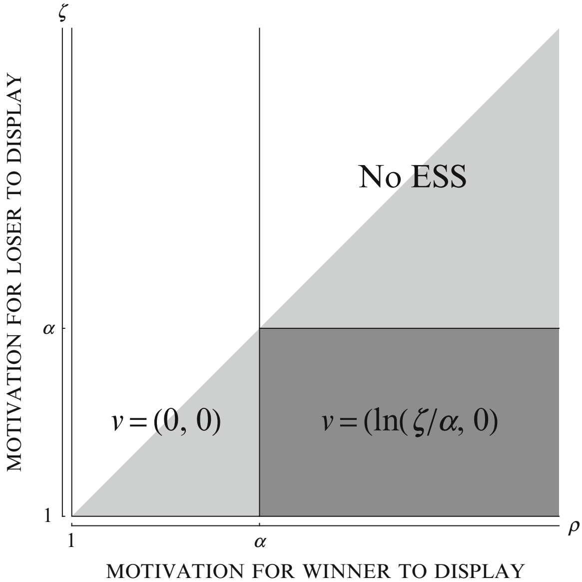
As predicted by our current model, there is now a considerable amount of evidence from the cricket literature that eventual winners tend to signal far more frequently than losers after fighting [1, 3]. One factor driving this basic result in our model (and most likely in the experiments) is our assumption that the marginal cost of signalling is lower for the winner than the loser, i.e., $\gamma_w < \gamma_l$; see (2.4). We consider this an entirely realistic condition given that the victor is likely to have "more left in the tank" than the vanquished (see [4], pp. 121–122 for a similar argument). Indeed, in these cases costly signals may serve as an honest indicator of how much the victor has in reserve, and thereby intimidate the opponent into submission. The "Ali shuffle" [8] is potentially one such example of an honest demonstration of a fighter's superiority. Analogous behaviors, which may have evolved at least in part to provide an honest signal of an individual's ability (in this case to escape predators), include stotting by Thomson's gazelles [5], push-up displays by anolis lizards [14] and aerial singing by skylarks [7].

Another factor favoring the result that eventual winners tend to signal far more frequently than losers after fighting is a sufficiently low value of the parameter b, i.e., a sufficiently high reproductive advantage to dominance. Indeed it follows from the analysis in Sect. 2.3 that there can be no post-conflict signalling at the ESS in our model if $\gamma_w/\gamma_l > (1-b)/b$. On the contrary, the winner victory-displays at the ESS only if (2.13) holds, i.e., if $\gamma_l/\gamma_w > b/(1-b)$. We have already noted that a high value of γ_l/γ_w is one way to favor this inequality, and a low value of b is clearly another.

Our model not only predicts that winners are more likely to signal than losers, but also that signalling should be more intense the longer and more intense the contest. This prediction arises simply because the ease of victory is itself a signal of dominance, a fact unlikely to be reversed through signalling either by the winner or by the loser. This phenomenon is captured in our model by making the baseline probability that winning will lead to outright dominance, i.e., δ, a decreasing function of the contest duration T; see (2.5). Recent studies on crickets ([3], Ting et al. unpublished) support the prediction that post-conflict displays are more likely to be observed after long contests. Further evidence for this general property comes from a recent comparative study by Jang et al. [13], who examined post-conflict "dominance" displays by winners in pairwise contests of males of four different species (*Gryllus pennsylvanicus, G. rubens, G. vernalis* and *G. fultoni*, respectively). The latter two field cricket species do not fight as intensively as the former two species and, as might be anticipated, do not display as frequently following conflict, or with such vigor, as the former two species. This pattern once again suggests that only close or costly fighting selects for victory displays, although more comparative data are clearly needed.

For the model to be compared with observations in a quantitative sense, an obvious question is, how big is the reproductive value of dominance in crickets? That is, how small is b? The observation of multiple fights in a laboratory setting (Janice Ting, personal communication) and the very presence of clear dominance hierarchies [2] suggest that there is a significant reproductive advantage to dominance. Suppose, in the first instance, that the reproductive advantage of dominance is high enough to ensure $b < \frac{1}{2}$. Then (2.13) must hold, corresponding to the shaded triangle in Fig. 2.1, and there are two critical values of T/μ, a smaller value $\ln(\frac{\rho}{\rho-1})$, below which $v = 0$ is the ESS, and a larger value $\ln(\frac{\zeta}{\zeta-1})$, above which no ESS exists. Between these two critical values, the ESS is determined by $\theta v = (\lambda, 0)$, with λ defined by (2.12). We see that the intensity of the winner's display is zero until contest length reaches the first critical value, after which intensity increases with contest length until the second critical value is reached; and the corresponding intensity of the loser's display is zero. This behavior is illustrated in Fig. 2.2 in the limit as $\zeta \to 1$ from above, so that the second critical value recedes towards infinity.

Fig. 2.1 ESS regions in the $\rho-\zeta$ plane for fixed values of T/μ, hence fixed values of $\alpha = 1/(1-\mathrm{e}^{-T/\mu})$. The *shaded triangle* is where $\rho = (1-b)/\gamma_w$ exceeds $\zeta = b/\gamma_l$, i.e., where (2.13) holds. High reproductive advantage of dominance corresponds to low b and hence low ζ, i.e., just above the horizontal axis



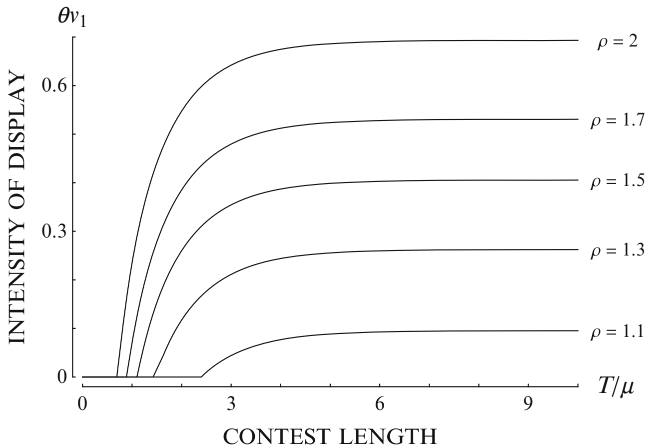


Fig. 2.2 Scaled intensity of winner's victory display as a function of scaled contest length for various values of the parameter $\rho = (1-b)/\gamma_w$ (assumed to exceed 1) in the limit as $\zeta \to 1$ from above, where $\zeta = b/\gamma_l$

In general, however, both critical values are finite. For contest lengths below the first critical value, neither the winner nor the loser displays at the ESS. For contest lengths between the two critical values, only the winner displays, with intensity that increases with T. For contest lengths greater than the second critical value, the ESS breaks down as described at the end of the appendix, and in such a way that a loser's optimal response will sometimes be to match the winner's display. Thus, according to our model, a loser should be expected to display only if the contest is so long that its length exceeds the second critical value. Those unusual biological examples in which only the loser displays (e.g., [15]) may potentially be explained by some sort of subservient signal to assure dominance to the victor, thereby reducing future conflict [4].

There is an intriguing parallel between one of our results on victory displays and a result concerning winner effects that Mesterton-Gibbons [19] found, several years before victory displays were first reviewed by Bower [4]. A winner effect is an increased probability of victory in a later contest following victory in an earlier contest [21], which in Mesterton-Gibbons [19] is mediated through increased self-perception of strength. The greater the likelihood of a later victory, the more likely it is that the earlier victory will eventually lead to dominance over the opponent. Thus a winner effect may also be regarded as an attempt to convert victory into dominance, even though there is no display. The result discovered by Mesterton-Gibbons [19] is that there can be no winner effect unless $b < \frac{1}{2}$, where b has exactly the same interpretation as in our current model, i.e., an inverse measure of the reproductive advantage of dominance. Thus, to the extent that victory displays and winner effects can both be regarded as factors favoring dominance, such factors are most operant when $b < \frac{1}{2}$.

Finally, for the sake of tractability, we did not explicitly model the variation of strength that supports any variation of contest length observed in nature. On the contrary, we assumed that T is fixed for a theoretical population; and we obtained an evolutionarily stable response to that T, which is likewise fixed for the theoretical population. Over many such theoretical populations, each with a different T, however, there will be many different ESS responses; and in effect we have implicitly assumed that the variation of ESS with T thus engendered will reasonably approximate the variation of signal intensity with contest length observed within a single real population. Essentially this assumption—phrased more generally, that ESS variation over many theoretical populations each characterized by a different parameter value will reasonably approximate variation of behavior with respect to that parameter within a single real population—is widely adopted in the literature, although rarely made explicit, as here. Indeed essentially this assumption is made whenever a game-theoretic model predicts the dependence of an ESS on a parameter that varies within a real population, but whose variance is not accounted for by the model.

Acknowledgements We are grateful to Lauren Fitzsimmons and two anonymous reviewers for constructive feedback on earlier versions of the manuscript.

Appendix A. ESS Conditions

Strategy v is a strong, global evolutionarily stable strategy or ESS in the sense of [17] if (and only if) it is uniquely the best reply to itself, in the sense that $f(v,v) > f(u,v)$

for all $u \neq v$; or, equivalently for our model, if v_1 is a winner's best reply to a loser's v_2 and v_2 is a loser's best reply to a winner's v_1.[1]

From (2.2) and (2.6) we have

$$\frac{\partial f}{\partial u_1} = \begin{cases} -\frac{1}{2}\gamma_w\theta & \text{if} \quad u_1 < v_2 \\ \frac{1}{2}\{(1-b)(1-\delta)\,\mathrm{e}^{-\theta(u_1-v_2)} - \gamma_w\}\theta & \text{if} \quad u_1 > v_2 \end{cases} \tag{A1}$$

with $\partial^2 f/\partial u_1^2 = -\frac{1}{2}\theta^2(1-b)(1-\delta)\,\mathrm{e}^{-\theta(u_1-v_2)} < 0$ for $u_1 > v_2$ but $\partial^2 f/\partial u_1^2 = 0$ for $u_1 < v_2$. So, with respect to u_1, f decreases from $u_1 = 0$ to $u_1 = v_2$. What happens next depends on the limit of $\partial f/\partial u_1$ as $u_1 \to v_2$ from above, which is $\frac{1}{2}\{(1-b)(1-\delta) - \gamma_w\}\theta$. If this quantity is not positive, then f continues to decrease, and so the maximum of f with respect to u_1 occurs at $u_1 = 0$. So a winner's best reply is $u_1 = 0$ whenever $\delta > 1 - \gamma_w/(1-b)$ (which is true in particular if $\gamma_w > 1-b$). If, on the other hand, $\delta < 1 - \gamma_w/(1-b)$, then there is a local maximum for $u_1 > v_2$ where $\partial f/\partial u_1 = 0$ or

$$\theta u_1 = \theta v_2 + \ln\left(\frac{(1-b)(1-\delta)}{\gamma_w}\right). \tag{A2}$$

The value of f at this local maximum exceeds the value at $u_1 = 0$ only if

$$\theta v_2 < \frac{(1-b)(1-\delta)}{\gamma_w} - 1 - \ln\left(\frac{(1-b)(1-\delta)}{\gamma_w}\right). \tag{A3}$$

Note that the right-hand side of (A3) is always positive (because $x - 1 - ln(x) > 0$ for all $x > 1$). In sum, a winner's best reply is $u_1 = 0$ unless $\delta < 1 - \gamma_w/(1-b)$ and (A3) holds, in which case, the best reply is given by (A2). In particular, zero is always a winner's best reply if $\gamma_w > 1-b$.

Similarly,

$$\frac{\partial f}{\partial u_2} = \begin{cases} \frac{1}{2}\{b(1-\delta)\,\mathrm{e}^{\theta(u_2-v_1)} - \gamma_l\}\theta & \text{if} \quad u_2 < v_1 \\ -\frac{1}{2}\gamma_l\theta & \text{if} \quad u_2 > v_1 \end{cases} \tag{A4}$$

with $\partial^2 f/\partial u_2^2 = \frac{1}{2}\theta^2 b(1-\delta)\,\mathrm{e}^{\theta(u_2-v_1)} > 0$ for $u_2 < v_1$ but $\partial^2 f/\partial u_2^2 = 0$ for $u_2 > v_1$. Note that the limit of $\partial f/\partial u_2$ as $u_2 \to v_1$ from below is $\frac{1}{2}\{b(1-\delta) - \gamma_l\}\theta$. Because

[1]In general, strategy v is an ESS if it does not pay a potential mutant to switch from v to any other strategy, and v need not satisfy the strong condition $f(v,v) > f(u,v)$ for all $u \neq v$. If there is at least one alternative best reply u such that $f(u,v) = f(v,v)$ but v is a better reply than u to all such u ($f(v,u) > f(u,u)$), then v is called a weak ESS. For our model, however, any ESS is a strong ESS, as is typical of continuous games ([18], p. 408).

$\partial^2 f/\partial u_2^2 > 0$, if the limit is negative, i.e., if $\delta > 1 - \gamma_l/b$, then f decreases with respect to u_2 and has its maximum where $u_2 = 0$, so that a loser should not display. If, on the other hand, the limit is positive, i.e., $\delta < 1 - \gamma_l/b$, then f at least partly increases with respect to u_2 for $u_2 < v_1$; and so the maximum of f with respect to u_2 occurs either at $u_2 = 0$ or $u_2 = v_1$, depending on which has the higher value of f. Let x_c denote the unique positive root of the equation $(1-\delta)(1-\mathrm{e}^{-x}) = \gamma_l x/b$. Then straightforward algebra reveals that the maximum is at 0 if $\theta v_1 > x_c$ but at v_1 if $\theta v_1 < x_c$. In sum, a loser's best reply is $u_2 = 0$ unless $\delta < 1 - \gamma_l/b$ and $\theta v_1 < x_c$, in which case, the best reply is v_1. Clearly, $\theta v_1 < x_c$ holds for $v_1 = 0$, so that $u_2 = 0$ is in particular the best reply to $v_1 = 0$; however, this result follows more readily directly from (A4). Also, note that zero is always a loser's best reply if $\gamma_l > b$.

For $v = (v_1, v_2)$ to be an ESS it must be a best reply to itself, i.e., we require v_1 to be a winner's best reply to the loser's v_2 at the same time as v_2 is a loser's best reply to the winner's v_1. If

$$\delta \;>\; \max\left(1 - \frac{\gamma_w}{1-b}, 1 - \frac{\gamma_l}{b}\right) \tag{A5}$$

then the unique ESS is $v = (0,0)$, because it follows from the discussion after (A1) that $v_1 = 0$ is the best reply to any v_2, and hence to $v_2 = 0$; and from the discussion after (A4) that $v_2 = 0$ is the best reply to any v_1, and hence to $v_1 = 0$. If

$$1 - \frac{\gamma_w}{1-b} \;<\; \delta \;<\; 1 - \frac{\gamma_l}{b} \tag{A6}$$

then $v = (0,0)$ is still the ESS by the above discussion and the remark at the end of the preceding paragraph. If instead

$$1 - \frac{\gamma_l}{b} \;<\; \delta \;<\; 1 - \frac{\gamma_w}{1-b} \tag{A7}$$

then the ESS is given by $\theta v = (\lambda, 0)$ where λ is defined by (2.12), because it follows from (A2) that $v_1 = \lambda/\theta$ is the best reply to $v_2 = 0$; and $v_2 = 0$ is still the best reply to any v_1, and hence to $v_1 = 0$.

If, on the other hand,

$$\delta \;<\; \min\left(1 - \frac{\gamma_w}{1-b}, 1 - \frac{\gamma_l}{b}\right) \tag{A8}$$

then an ESS does not exist. Consider a population in which a winner displays with small positive intensity v_1. Then $\theta v_1 < x_c$; and, from the discussion following (A4), a loser's best reply is to match the display. From (A2), a winner's best reply is now to increase the intensity of its display, because (A3) invariably holds; and a loser's best reply in turn is again to match the display. Continuing in this manner, we observe an "arms race" of increasing display intensity, until either (A3) or $\theta v_1 < x_c$ is violated. If the former, then a winner's best reply is not to display, which a loser matches, so that it pays for a winner to display at higher intensity; if the latter, then a loser's best reply becomes no display, but now a winner's best reply is to display with intensity λ/θ. Either way, the unstable cycle continues ad infinitum. The study of victory displays is still in its infancy, and researchers are still trying to characterize when it occurs and with what frequency. Therefore, there is no study into its temporal dynamics (within or between generations). Furthermore, full analysis of the evolutionary dynamics when no ESS exists is beyond the scope of this paper. Nevertheless, we broach this issue in Appendix B.

Appendix B. What Happens When There Is No ESS?

In this appendix we remark on why, when no ESS exists, the evolutionary dynamics require a more sophisticated approach than the one we have taken in this paper and cannot readily be addressed by the standard framework of discrete evolutionary games with replicator dynamics (e.g., [6, 11]). To make our point as expeditiously as possible, we explore circumstances in which $b < \frac{1}{2}$ and hence $\rho > \zeta$ but T/μ exceeds the second critical value $\ln(\frac{\zeta}{\zeta-1})$ of Sect. 2.3 (corresponding to the shaded triangle within the no-ESS region of Fig. 2.1).

Accordingly, consider a mixture of three strategies that appear to evoke the discussion towards the end of Appendix A, namely, a non-signalling strategy, denoted by N or Strategy 1; the ESS signalling strategy for the dark shaded rectangle of Fig. 2.1, denoted by S or Strategy 2; and a matching strategy, denoted by M or strategy 3, which displays with the ESS intensity corresponding to $\ln(\frac{\rho}{\rho-1}) < T/\mu < \ln(\frac{\zeta}{\zeta-1})$ after winning, but matches the winner's display after losing. From the viewpoint of a focal u-strategist against a v-strategist, these three strategies are defined, respectively, by $u = (0,0)$ for N; $u = (\lambda/\theta, 0)$ for S; and $u = (\lambda/\theta, v_1)$ for M. Let the proportions of N, S and M be x_1, x_2 and x_3, respectively (so that $x_1 + x_2 + x_3 = 1$); and let a_{ij} be the reward to strategy i against strategy j (for $1 \le i, j \le 3$). Then from (2.2), (2.6) and (2.12) we have $a_{11} = \frac{1}{2}\{(1-2b)q(0,0) - c_w(0) - c_l(0)\} + b = \frac{1}{2}\delta + (1-\delta)b$, $a_{12} = \frac{1}{2}\{(1-b)q(0,0) - bq(\lambda/\theta, 0) - c_w(0) - c_l(0)\} + b = \frac{1}{2}\{(1-b)\delta + b(1+1/\rho)\}$, and so on, yielding the reward matrix

$$A = \begin{bmatrix} \frac{1}{2}\delta + (1-\delta)b & \frac{\rho(1-b)\delta+(\rho+1)b}{2\rho} & a_{12} \\ \frac{\rho-1+b+(1-\delta)\rho b}{2\rho} - \frac{1}{2}\gamma_w \ln(\rho/\alpha) & \frac{\rho-1+2b}{2\rho} - \frac{1}{2}\gamma_w \ln(\rho/\alpha) & a_{12} - \frac{1}{2}\gamma_w \ln(\rho/\alpha) \\ a_{21} & a_{21} - \frac{1}{2}\gamma_l \ln(\rho/\alpha) & a_{11} - \frac{1}{2}(\gamma_w + \gamma_l)\ln(\rho/\alpha) \end{bmatrix} \tag{B1}$$

where $\alpha = 1/(1-\delta)$, as in Fig. 2.2.

Because $\gamma_w \rho = 1 - b$ and $\rho/\alpha = \rho(1-\delta) > 1$ by assumption, $a_{11} - a_{21} = \frac{1}{2}\gamma_w\{1 - \rho/\alpha + \ln(\rho/\alpha)\}$ must be negative (because $1 - x + \ln(x) < 0$ for all $x > 1$). Thus $a_{11} < a_{21}$, and Strategy 1 is not an ESS. However, because $a_{11} - a_{21} + a_{22} - a_{12} = 0$, it also follows that $a_{22} > a_{12}$. So if Strategy 2 is also not an ESS, then it must be Strategy 3 that invades. But $a_{22}-a_{32}=\frac{1}{2}\gamma_l\{(1/\rho-1/\alpha)\zeta+\ln(\rho/\alpha)\}$ may have either sign, and in particular will always be positive for sufficiently small ζ, that is, for ζ sufficiently close to α ($\zeta > \alpha$ having been assumed). Furthermore, if we suppose that the point (ρ,ζ) in Fig. 2.1 has migrated from the signalling ESS region (dark shaded rectangle) into the no-ESS region (shaded triangle immediately above) because environmental pressures have increased the value the value of ζ (by decreasing γ_l), then it is precisely such sufficiently small values of ζ that are relevant. Thus S will often be an ESS of the discrete game defined by the matrix A even though it is no longer an ESS of the continuous game described in the main body of our paper.

The upshot is that replicator dynamics cannot readily be used to describe what happens when S is not an ESS of our continuous game; the dynamics described verbally towards the end of Appendix A are not adequately reflected by a mix of M, N and S. They require a much more sophisticated approach, and we leave the matter open for future work.

Nevertheless, let us suppose that ζ is indeed large enough for M to invade S, that is,

$$\zeta \ > \ \frac{\rho\alpha}{\rho-\alpha}\ln(\rho/\alpha). \tag{B2}$$

Then $a_{22}-a_{32}<0$, and because $a_{33}-a_{23}+a_{22}-a_{32}=0$, we must have $a_{33}-a_{23}>0$. That is, from (B1), $a_{33}>a_{13}-\frac{1}{2}\gamma_w\ln(\rho/\alpha)$. Thus $a_{33}>\max(a_{13},a_{23})$ will hold for sufficiently small γ_w, making M the unique ESS of the discrete game defined by A. Otherwise (i.e., for larger values of γ_w), $a_{32}>a_{22}$, $a_{21}=a_{31}>a_{11}$ and $a_{13}>a_{33}$ will hold simultaneously, so that M can invade S, S or M can invade N and N can invade M. In these circumstances, the population will eventually reach a polymorphism of N and M at which $x_1=(a_{13}-a_{33})/(a_{13}-a_{33}+a_{31}-a_{11})$, $x_2=0$ and $x_3=1-x_1$. All of these results have been verified by numerical integration, for relevant parameter values, of the replicator equations $\dot{x}_i=x_i\{(Ax)_i-x\cdot Ax\}, i=1,2,3$, where $x=(x_1,x_2,x_3)$ and an overdot denotes differentiation with respect to time (see, e.g., [11], p. 68).

References

1. Alexander, R.D.: Evolutionary change in cricket acoustical communication. Evolution **16**, 443–467 (1962)
2. Bailey, W.J., Stoddart, J.A.: A method for the construction of dominance hierarchies in the field cricket *Teleogryllus oceanicus* (Le Guillou). Anim. Behav. **30**, 216–220 (1982)
3. Bertram, S.M., Rook, V.L.M., Fitzsimmons, L.P.: Strutting their stuff: victory displays in the spring field cricket, *Gryllus veletis*. Behaviour **147**, 1249–1266 (2010)
4. Bower, J.L.: The occurrence and function of victory displays within communication networks. In: McGregor, P. (ed.) Animal Communication Networks, pp 114–126. Cambridge University Press, Cambridge (2005)

5. Caro, T.M.: The functions of stotting in Thomson's gazelles: some tests of the predictions. Anim. Behav. **34**, 663–684 (1986)
6. Cressman, R.: Evolutionary Dynamics and Extensive Form Games. MIT Press, Cambridge (2003)
7. Cresswell, W.: Song as a pursuit-deterrent signal, and its occurrence relative to other anti-predation behaviours of skylark (*Alauda arvensis*) on attack by merlins (*Falco columbarius*). Behav. Ecol. Sociobiol. **34**, 217–223 (1994)
8. Golus, C.: Muhammad Ali. Twenty-First Century Books, Minneapolis (2006)
9. Grafe, T., Bitz, J.H.: An acoustic postconflict display in the duetting tropical boubou (*Laniarius aethiopicus*): a signal of victory? BMC Ecol. **4**(1) (2004). http://www.biomedcentral.com/1472-6785/4/1
10. Hardy, I.C.W., Briffa, M. (eds.): Animal Contests. Cambridge University Press, Cambridge (2013)
11. Hofbauer, J., Sigmund, K.: Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge (1998)
12. Huntingford, F., Turner, A.K.: Animal Conflict. Chapman & Hall, London (1987)
13. Jang, Y., Gerhardt, H.C., Choe, J.C.: A comparative study of aggressiveness in eastern North American field cricket species (genus *Gryllus*). Behav. Ecol. Sociobiol. **62**, 1397–1407 (2008)
14. Leal, M., Rodríguez-Robles, J.A.: Signalling displays during predator-prey interactions in a puerto rican anole, *Anolis cristatellus*. Anim. Behav. **54**, 1147–1154 (1997)
15. Lippold, S., Fitzsimmons, L.P., Foote, J.R., Ratcliffe, L.M., Mennill, D.J.: Post-contest behaviour in black-capped chickadees (*Poecile atricapillus*): loser displays, not victory displays, follow asymmetrical countersinging exchanges. Acta Ethol. **11**, 67–72 (2008)
16. Logue, D.M., Abiola, I.O., Rains, D., Bailey, N.W., Zuk, M., Cade, W.H.: Does signalling mitigate the cost of agonistic interactions? a test in a cricket that has lost its song. Proc. R. Soc. Lond. B **277**, 2571–2575 (2010)
17. Maynard, S.J.: Evolution and the Theory of Games. Cambridge University Press, Cambridge (1982)
18. McGill, B.J., Brown, J.S.: Evolutionary game theory and adaptive dynamics of continuous traits. Annu. Rev. Ecol. Syst. **38**, 403–435 (2007)
19. Mesterton-Gibbons, M.: On the evolution of pure winner and loser effects: a game-theoretic model. Bull. Math. Biol. **61**, 1151–1186 (1999)
20. Mesterton-Gibbons, M., Sherratt, T.N.: Victory displays: a game-theoretic analysis. Behav. Ecol. **17**, 597–605 (2006)
21. Rutte, C., Taborsky, M., Brinkhof, M.W.G.: What sets the odds of winning and losing? Trends Ecol. Evol. **21**, 16–21 (2006)
22. Tracy, J.L., Matsumoto, D.: The spontaneous expression of pride and shame: evidence for biologically innate nonverbal displays. Proc. Natl. Acad. Sci. U.S.A. **105**, 11655–11660 (2008)

Chapter 3
Evolutionary Games for Multiple Access Control

Quanyan Zhu, Hamidou Tembine, and Tamer Başar

Abstract In this paper, we formulate an evolutionary multiple access control game with continuous-variable actions and coupled constraints. We characterize equilibria of the game and show that the pure equilibria are Pareto optimal and also resilient to deviations by coalitions of any size, i.e., they are strong equilibria. We use the concepts of price of anarchy and strong price of anarchy to study the performance of the system. The paper also addresses how to select one specific equilibrium solution using the concepts of normalized equilibrium and evolutionarily stable strategies. We examine the long-run behavior of these strategies under several classes of evolutionary game dynamics, such as Brown–von Neumann–Nash dynamics, Smith dynamics, and replicator dynamics. In addition, we examine correlated equilibrium for the single-receiver model. Correlated strategies are based on signaling structures before making decisions on rates. We then focus on evolutionary games for hybrid additive white Gaussian noise multiple-access channel with multiple users and multiple receivers, where each user chooses a rate and splits it over the receivers. Users have coupled constraints determined by the capacity regions. Building upon the static game, we formulate a system of hybrid evolutionary game dynamics using

†This material is based upon work supported in part by the U.S. Air Force Office of Scientific Research (AFOSR) under grant number FA9550-09-1-0249, and in part by the AFOSR MURI Grant FA9550-10-1-0573.

The material in this paper was partially presented in [9] and [10].

Q. Zhu (✉) • T. Başar
Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1308 W. Main Street, Urbana, IL 61801, USA
e-mail: zhu31@illinois.edu; basar1@illinois.edu

H. Tembine
Department of Telecommunications, École Supérieure d'Electricité (SUPELEC), 3 rue Joliot-Curie, 91192 Gif-Sur-Yvette Cedex, France
e-mail: tembine@ieee.org

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_3,

G-function dynamics and Smith dynamics on rate control and channel selection, respectively. We show that the evolving game has an equilibrium and illustrate these dynamics with numerical examples.

Keywords Game theory • Evolutionary game dynamics • Capacity region • Access control • Hybrid dynamics • G-functions

3.1 Introduction

Recently much interest has been devoted to understanding the behavior of multiple access controls under constraints. A considerable amount of work has been carried out on the problem of how users can obtain an acceptable throughput by choosing rates independently. Motivated by an interest in studying a large population of users playing a game over time, evolutionary game theory was found to be an appropriate framework for communication networks. It has been applied to problems such as power control in wireless networks and mobile interference control [1, 5, 6, 11].

The game-theoretical models considered in previous studies on user behavior in code division multiple access (CDMA) [4, 33] are static one-shot noncooperative games in which users are assumed to be rational and optimize their payoffs independently. Evolutionary game theory, on the other hand, studies games that are played repeatedly and focuses on the strategies that persist over time, yielding the best fitness of a user in a noncooperative environment on a large time scale.

In [19], an additive white Gaussian noise (AWGN) multiple-access-channel problem was modeled as a noncooperative game with pairwise interactions, in which users were modeled as rational entities whose only interests were to maximize their own communication rates. The authors obtained the Nash equilibrium NE/NEs of Nash equilibrium/equilibria of the two-user game and introduced a two-player evolutionary game model with *pairwise interactions* based on replicator dynamics. However, the case where interactions are not pairwise arises frequently in communication networks, such as the CDMA or the orthogonal frequency-division multiple access (OFDMA) in a Worldwide Interoperability for Microwave Access (WiMAX) environment [11].

In this work, we extend the study of [19] to wireless communication systems with an arbitrary number of users corresponding to each receiver. We formulate a static noncooperative game with m users subject to rate capacity constraints and extend the constrained game to a dynamic evolutionary game with a large number of users whose strategies evolve over time. Unlike evolutionary games with discrete and finite numbers of actions, our model is based on a class of continuous games, known as *continuous-trait games*. Evolutionary games with continuum action spaces are encountered in a wide variety of applications in evolutionary ecology, such as evolution of phenology, germination, nutrient foraging in plants, and predator–prey foraging [7, 23].

In addition to the single-receiver model, we investigate the case with multiple users and receivers. We first formulate a static hybrid noncooperative game with N users who rationally make decisions on the rates as well as the channel selection subject to rate the capacity constraints of each receiver. We extend the static game to a dynamic evolutionary game by viewing rate selections governed by a fitness function parameterized by the channel selections. Such a concept of a hybrid model appeared in [32, 36] in the context of hybrid power control in CDMA systems. The strategies of channel selections determine the long-term fitness of the rates chosen by each user. We formulate such dynamics based on generalized Smith dynamics and generating fitness function (G-function) dynamics.

3.1.1 Contribution

The main contributions of this work can be summarized as follows. We first introduce a game-theoretic framework for local interactions between many users and a single receiver. We show that the static continuous-kernel rate allocation game with coupled rate constraints has a convex set of pure NEs, coinciding with the maximal face of the polyhedral capacity region. All the pure equilibria are Pareto optimal and are also strong equilibria, resilient to simultaneous deviation by coalition of any size. We show that the pure NEs in the rate allocation problem are 100 % efficient in terms of price of anarchy (PoA) and constrained strong price of anarchy (CSPoA). We study the stability of strong equilibria, normalized equilibria, and evolutionarily stable strategies (ESSs) using evolutionary game dynamics such as Brown–von Neumann–Nash dynamics, generalized Smith dynamics, and replicator dynamics. We further investigate the correlated equilibrium of the multiple-access game where the receiver can send signals to the users to mediate the behaviors of the transmitters.

Based on the single-receiver model, we then propose an evolutionary game-theoretic framework for the hybrid additive white Gaussian noise multiple-access channel. We consider a communication system of multiple users and multiple receivers, where each user chooses a rate and splits it over the receivers. Users have coupled constraints determined by the capacity regions. We characterize the NE of the static game and show the existence of the equilibrium under general conditions. Building upon the static game, we formulate a system of hybrid evolutionary game dynamics using G-function dynamics and Smith dynamics on rate control and channel selection, respectively. We show that the evolving game has an equilibrium and illustrate these dynamics with numerical examples.

3.1.2 Organization of the Paper

The rest of the paper is structured as follows. We present in Sect. 3.2.1 the evolutionary game model of rate allocation in additive white Gaussian multiple-access wireless networks and analyze its equilibria and Pareto optimality in Sect. 3.2.2. In

Table 3.1 List of notations

Symbol	Meaning
$\mathcal{N}$	Set of N users
Ω	Subset of N users
$\mathcal{J}$	Set of J receivers
$\mathcal{A}_i$	Action set of user i
P_i	Maximum power of user i
h_i	Channel gain of user i
α_i	Rate of user i
p_{ij}	Probability that user i will select receiver j
u_i	Payoff of user i
$\overline{U}_i$	Expected payoff of user i
C	Capacity region of a set $\mathcal{N}$ of users in a single-receiver case
$C(j)$	Capacity region of a set $\mathcal{N}$ of users at receiver j
λ_i	Distribution over action set $\mathcal{A}_i$
μ	Population state

Table 3.2 List of acronyms

Abbreviation	Meaning
AWGN	Additive white Gaussian noise
MAC	Multiple-access control
MISO	Multi-input and single-output
CCE	Constrained correlated equilibrium
ESS	Evolutionarily stable state (or strategy)
NE	Nash equilibrium
PoA	Price of anarchy
SPoA	Strong price of anarchy

Sect. 3.2.3, we present strong equilibria and the PoA of the game. In Sect. 3.2.4, we discuss how to select one specific equilibrium such as normalized equilibrium and ESSs. Section 3.2.5 studies the stability of equilibria and the evolution of strategies using game dynamics. Section 3.2.6 analyzes the correlated equilibrium of the multiple-acess game.

In Sect. 3.3.1, we present the hybrid rate control model where users can choose the rates and the probability of the channels to use. In Sect. 3.3.2, we characterize the NE of the constrained hybrid rate control game model, pointing out the existence of the NE of the hybrid model and methods to find it. In Sect. 3.3.3, we apply evolutionary dynamics to both rates and channel selection probabilities. We use simulations to demonstrate the validity of these proposed dynamics and illustrate the evolution of the overall evolutionary dynamics of the hybrid model. Section 3.4 concludes the paper. For the reader's convenience, we summarize the notations in Table 3.1 and the acronyms in Table 3.2.

3.2 AWGN Mutiple-Access Model: Single-Receiver Case

We consider a communication system consisting of several receivers and several senders (Fig. 3.1). At each time, there are several simultaneous local interactions (typically, at each receiver there is a local interaction). Each local interaction corresponds to a noncooperative one-shot game with common constraints. The opponents do not necessarily stay the same from one given time slot to the next. Users revise their rates in view of their payoffs and the coupled constraints (for example, by using an evolutionary process, a learning process, or a trial-and-error updating process). The game evolves over time. Users are interested in maximizing a fitness function based on their own communication rates at each time, and they are aware of the fact that the other users have the same goal. The coupled power and rate constraints are also common knowledge. Users must choose independently their own coding rates at the beginning of the communication, where the rates selected by a user may be either deterministic or chosen from some distribution. If the rate profile arrived at as a result of these independent decisions lies in the capacity region, users will communicate at that operating point. Otherwise, either the receiver is unable to decode any signal and the observed rates are zero or only one of the signals can be decoded. The latter occurs when all the other users are transmitting at or below a safe rate. With these assumptions, we can define a constrained noncooperative game. The set of allowed strategies for user i is the set of all probability distributions over $[0, +\infty[$, and the payoff is a function of the rates. In addition, the rational action (rate) sets must lie in the capacity regions (the payoff is zero if the constraint is violated). To study the interactions between the selfish or partially cooperative users and their stationary rates in the long run, we propose to model the problem of rate allocation in Gaussian multiple-access channels as an evolutionary game with a continuous action space and coupled constraints. The development of evolutionary game theory is a major contribution of biology to competitive decision making and the evolution of cooperation. The key concepts

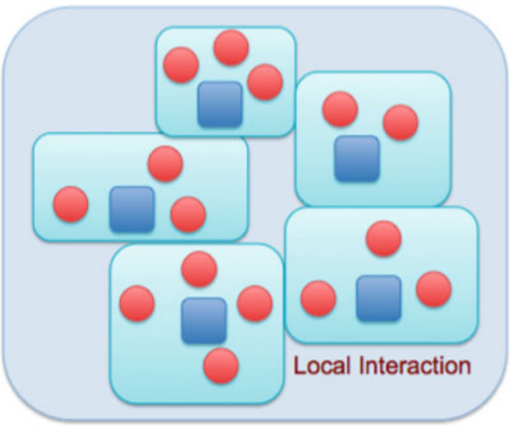


Fig. 3.1 A population: distributed receivers (*blue rectangles*) and senders (*red circles*)

of evolutionary game theory are (a) *evolutionarily stable state* [25], which is a refinement of equilibrium, and (b) *evolutionary game dynamics*, such as replicator dynamics [29], which describes the evolution of strategies or frequencies of use of strategies in time [7, 21].

The single population evolutionary rate allocation game is described as follows: there is one population of senders (users) and several receivers. The number of senders is large. At each time, there are many one-shot finite games called *local interactions*, which models the interactions among a finite number of users in the population. Each sender of the population chooses from his or her set of strategies $\mathcal{A}_i$, which is a nonempty, convex, and compact subset of $\mathbb{R}$. Without loss of generality, we can suppose that user i chooses his or her rate in the interval $\mathcal{A}_i = [0, C_{\{i\}}]$, where $C_{\{i\}}$ is the rate upper bound for user i (to be made precise subsequently) as outside of the capacity region the payoff (to be defined later) will be zero. Let $\Delta(\mathcal{A}_i)$ be the set of probability distributions over the pure strategy set $\mathcal{A}_i$. The set $\Delta(\mathcal{A}_i)$ can be interpreted as the set of mixed strategies for the N-person game at the local interaction. In the case where the N-person local interaction is identical at all local interactions in the population, the set $\Delta(\mathcal{A}_i)$ can also be interpreted as the set of distributions of strategies among the population. Let $\lambda_i \in \Delta(\mathcal{A}_i)$ and E be a λ_i- measurable subset of $\mathbb{R}$; then $\lambda_i(E)$ represents the fraction of users choosing a strategy from E at time t. A distribution $\lambda_i \in \Delta(\mathcal{A}_i)$ is sometimes called the "state" of the population. We denote by $\mathbb{B}(\mathcal{A}_i)$ the Borel σ-algebra on $\mathcal{A}_i$ and by $d(\lambda, \lambda')$ the distance between two states measured with respect to the weak topology. An example of such a distance could be the classical Wasserstein distance or the Monge–Kantorovich distance between two measures.

Each user's payoff depends on the opponents' behavior through the distribution of the opponents' choices and of their strategies. The payoff of user i in a local interaction with $(N-1)$ other users is given as a function $u_i: \mathbb{R}^N \longrightarrow \mathbb{R}$. The rate profile $\alpha \in \mathbb{R}^N$ must belong to a common capacity region $C \subset \mathbb{R}^N$ defined by $2^N - 1$ linear inequalities. The expected payoff of a sender i transmitting at a rate a when the state of the population is $\mu \in \Delta(\mathcal{A}_i)$ is given by $F_i(a, \mu)$. The expected payoff for user i is

$$F_i(\lambda_i, \mu) := \int_{\alpha \in C} u_i(\alpha)\ \lambda_i(\mathrm{d}\alpha_i) \bigotimes_{j \neq i} \mu(\mathrm{d}\alpha_j).$$

The population state is subject to the "mixed extension" of capacity constraints $\mathcal{M}(C)$. This will be discussed in Sect. 3.2.5 and made more precise later.

3.2.1 Local Interactions

Local interaction refers to the problem setting of one receiver and its uplink additive white Gaussian noise (AWGN) multiple-access channel with N senders with coupled constraints (or actions). The signal at the receiver is given by $Y = \xi + \sum_{i=1}^{N} X_i$, where X_i is a transmitted signal of user i and ξ is a zero-mean Gaussian noise with variance σ_0^2. Each user has an individual power constraint $\mathbb{E}(X_i^2) \leq P_i$ and channel gain h_i. The optimal power allocation scheme is to transmit at the maximum power available, i.e., P_i, for each user. Hence, we consider the case in which the maximum power is attained. The decisions of the users, then, consist of choosing their communication rates, and the receiver's role is to decode, if possible. The capacity region is the set of all vectors $\alpha \in \mathbb{R}_+^N$ such that users $i \in \mathcal{N} := \{1,2,\ldots,N\}$ can reliably communicate at rate $\alpha_i,\ i \in \mathcal{N}$. The capacity region C for this channel is the set

$$C = \left\{ \alpha \in \mathbb{R}_+^N \;\middle|\; \sum_{i\in\Omega} \alpha_i \leq \log\left(1 + \sum_{j\in\Omega} \frac{P_j h_j}{\sigma_0^2}\right), \forall\, \emptyset \subsetneq \Omega \subseteq \mathcal{N} \right\}.$$

Example 3.1 (Example of capacity region with three users). In this example, we illustrate the capacity region with three users. Let $\alpha_1, \alpha_2, \alpha_3$ be the rates of the users and $P_i = P$, $h_i = h, \forall i \in \{1,2,3\}$. Based on (3.1), we obtain a set of inequalities

$$\begin{cases} \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_3 \geq 0, \\ \alpha_i \leq \log\left(1 + \frac{Ph}{\sigma_0^2}\right), i = 1,2,3, \\ \alpha_i + \alpha_j \leq \log\left(1 + 2\frac{Ph}{\sigma_0^2}\right), i \neq j, i, j = 1,2,3 \\ \alpha_1 + \alpha_2 + \alpha_3 \leq \log\left(1 + 3\frac{Ph}{\sigma_0^2}\right), \end{cases}$$

or, in compact notation, $M_3\gamma_3 \leq \zeta_3$, where

$$\gamma_3 := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \in \mathbb{R}_+^3, \qquad \zeta_3 := \begin{bmatrix} C_{\{1\}} \\ C_{\{2\}} \\ C_{\{3\}} \\ C_{\{1,2\}} \\ C_{\{1,3\}} \\ C_{\{2,3\}} \\ C_{\{1,2,3\}} \end{bmatrix}, \qquad M_3 := \begin{bmatrix} 1\;0\;0 \\ 0\;1\;0 \\ 0\;0\;1 \\ 1\;1\;0 \\ 1\;0\;1 \\ 0\;1\;1 \\ 1\;1\;1 \end{bmatrix} \in \mathbb{Z}^{7\times 3}.$$

Note that M_3 is a totally unimodular matrix. By letting $Ph = 25,\ \sigma_0^2 = 0.1$, we sketch in Fig. 3.2 the capacity region with three users.

The capacity region reveals the competitive nature of interactions among senders: if a user i wants to communicate at a higher rate, then one of the other users must lower his or her rate; otherwise, the capacity constraint is violated. We let

$$r_{i,\Omega} := \log\left(1 + \frac{P_i h_i}{\sigma_0^2 + \sum_{i'\in\Omega, i'\neq i} P_{i'} h_{i'}}\right), \qquad i \in \mathcal{N}, \Omega \subseteq \mathcal{N}$$

denote the bound on the rate of a user when the signals of the $|\Omega| - 1$ other users are treated as noise.

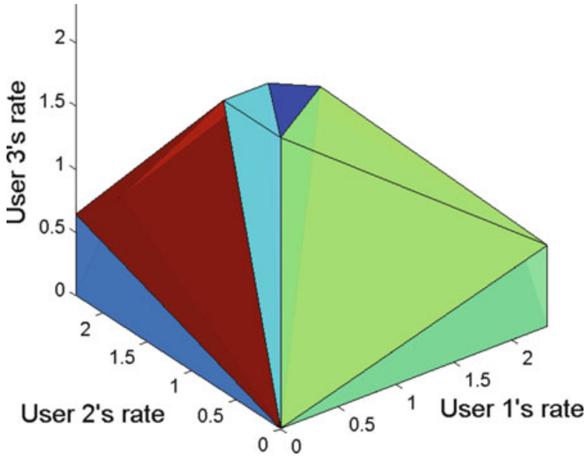


Fig. 3.2 Capacity region with three users

Due to the noncooperative nature of the rate allocation, we can formulate the one-shot game

$$\Xi = \langle \mathcal{N}, (\mathcal{A}_i)_{i \in \mathcal{N}}, (u_i)_{i \in \mathcal{N}} \rangle,$$

where the set of users $\mathcal{N}$ is the set of players, $\mathcal{A}_i$, $i \in \mathcal{N}$, is the set of actions, and u_i, $i \in \mathcal{N}$, are the payoff functions.

3.2.2 Payoffs

We define $u_i : \prod_{i=1}^{N} \mathcal{A}_i \to \mathbb{R}_+$ as follows:

$$u_i(\alpha_i, \alpha_{-i}) = \mathbb{1}_C(\alpha) g_i(\alpha_i) \tag{3.1}$$

$$= \begin{cases} g_i(\alpha_i) & \text{if } (\alpha_i, \alpha_{-i}) \in C, \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

where $\mathbb{1}_C$ is the indicator function, α_{-i} is a vector consisting of other players' rates, i.e., $\alpha_{-i} = [\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_N]$, and g_i is a positive and strictly increasing function for each fixed α_{-i}. Since the game is subject to coupled constraints, the action set $\mathcal{A}_i$ is coupled and dependent on other players' actions. Given the strategy profile α_{-i} of other players, the constrained action set $\mathcal{A}_i$ is given by

$$\mathcal{A}_i(\alpha_{-i}) := \{\alpha_i \in [0, C_{\{i\}}],\ (\alpha_i, \alpha_{-i}) \in C\}. \tag{3.3}$$

We then have an asymmetric game. The minimum rate that user i can guarantee in the feasible regions is $r_{i,\mathcal{N}}$, which is different than $r_{j,\mathcal{N}}$.

Each user i maximizes $u_i(\alpha_i, \alpha_{-i})$ over the coupled constraint set. Owing to the monotonicity of the function g_i and the inequalities that define the capacity region, we obtain the following lemma.

Lemma 3.1. *Let $\overline{BR}_i(\alpha_{-i})$ be the best reply to the strategy α_{-i}, defined by*

$$\overline{BR}_i(\alpha_{-i}) = \arg \max_{y \in \mathcal{A}_i(\alpha^{-i})} u_i(y, \alpha_{-i}).$$

$\overline{BR}_i$ is a nonempty single-valued correspondence (i.e., a standard function) and is given by

$$\max\left(r_{i,\mathcal{N}}, \min_{\Omega \in \Gamma_i}\left\{C_\Omega - \sum_{k \in \Omega \setminus \{i\}} \alpha_k\right\}\right), \tag{3.4}$$

where $\Gamma_i = \{\Omega \in 2^{\mathcal{N}}, i \in \Omega\}$.

Proposition 3.1. *The set of NEs is*

$$\left\{(\alpha_i, \alpha_{-i}) \mid \alpha_i \geq r_{i,\mathcal{N}}, \sum_{i \in \mathcal{N}} \alpha_i = C_{\mathcal{N}}\right\}.$$

All these equilibria are optimal in the Pareto sense.

Proof. Let β be a feasible solution, i.e., $\beta \in C$. If $\sum_{i=1}^N \beta_i < C_{\mathcal{N}} = \log(1 + \sum_{i \in \mathcal{N}} \frac{P_i h_i}{\sigma_0^2})$, then at least one of the users can improve his or her rate (and, hence, payoff) to reach one of the faces of the capacity region. We now check the strategy profile on the face

$$\left\{(\alpha_i, \alpha_{-i}) \,\middle|\, \alpha_i \geq r_{i,\mathcal{N}}, \sum_{i=1}^N \alpha_i = C_{\mathcal{N}}\right\}.$$

If $\beta \in \{(\alpha_i, \alpha_{-i}) \mid \alpha_i \geq r_{i,\mathcal{N}}, \sum_{i=1}^N \alpha_i = C_\Omega\}$, then, from Lemma 3.1, $\overline{BR}_i(\beta_{-i}) = \{\beta_i\}$. Hence, β is a strict equilibrium. Moreover, this strategy β is Pareto optimal because the rate of each user is maximized under the capacity constraint. These strategies are social welfare optimal if the total utility $\sum_{i=1}^N u_i(\alpha_i, \alpha_{-i}) = \sum_{i=1}^N g_i(\alpha_i)$ is maximized subject to constraints. □

Note that the set of pure NEs is a convex subset of the capacity region.

3.2.3 Robust Equilibria and Efficiency Measures

3.2.3.1 Constrained Strong Equilibria and Coalition Proofness

An action profile in a local interaction between N senders is a constrained k-strong equilibrium if it is feasible and no coalition of size k can improve the rate transmissions of each of its members with respect to the capacity constraints. An action is a constrained strong equilibrium [18] if it is a constrained k-strong equilibrium for any size k. A strong equilibrium is then a policy from which no coalition (of any size) can deviate and improve the transmission rate of every member of the coalition (group of the simultaneous moves) while possibly lowering the transmission rate of users outside the coalition group. This notion of constrained strong equilibria[1] is very attractive because it is resilient against coalitions of users. Most games do not admit any strong equilibrium, but in our case we will show that the multiple-access channel game has several strong equilibria.

Theorem 3.1. *Any rate profile on the maximal face of the capacity region* C:

$$\text{Face}_{\max}(C) := \left\{ (\alpha_i, \alpha_{-i}) \in \mathbb{R}^N \;\middle|\; \alpha_i \geq r_N, \sum_{i=1}^{N} \alpha_i = C_{\mathcal{N}} \right\}$$

is a constrained strong equilibrium.

Proof. If the rate profile α is not on the maximal face of the capacity region, then α is not resilient to deviation by a single user. Hence, α cannot be a constrained strong equilibrium. This shows that a necessary condition for a rate profile to be a strong equilibrium is to be in the subset $\text{Face}_{\max}(C)$. We now prove that the condition $\alpha \in \text{Face}_{\max}(C)$ is sufficient. Let $\alpha \in \text{Face}_{\max}(C)$. Suppose that k users deviate simultaneously from the rate profile α. Denote by Dev the set of users that deviate simultaneously (eventually by forming a coalition). The rate constraints of the deviants are

1. $\alpha'_i \geq 0, \ \forall i \in \text{Dev}$;
2. $\sum_{i \in \bar{\Omega}} \alpha'_i \leq C_{\bar{\Omega}}, \ \forall \bar{\Omega} \subseteq \text{Dev}$;
3. $\sum_{i \in \Omega \cap \text{Dev}} \alpha'_i \leq C_\Omega - \sum_{i \in \Omega, i \notin \text{Dev}} \alpha_i, \ \ \forall \Omega \subseteq \mathcal{N}, \ \Omega \cap \text{Dev} \neq \emptyset$.

In particular, for $\Omega = \mathcal{N}$ we have $\sum_{i \in \text{Dev}} \alpha'_i \leq C_{\mathcal{N}} - \sum_{i \notin \text{Dev}} \alpha_i$. The total rate of the deviants is bounded by $C_{\mathcal{N}} - \sum_{i \notin \text{Dev}} \alpha_i$, which is not controlled by the deviants. The deviants move to $(\alpha'_i)_{i \in \text{Dev}}$ with

$$\sum_{i \in \text{Dev}} \alpha'_i < C_{\mathcal{N}} - \sum_{i \notin \text{Dev}} \alpha_i \, .$$

[1] Note that the set of constrained strong equilibria is a subset of the set of NEs (by taking coalitions of size one) and any constrained strong equilibrium is Pareto optimal (by taking coalition of full size).

Then there exists i such that $\alpha_i > \alpha_i'$. Since g_i is nondecreasing, this implies that $g_i(\alpha_i) > g_i(\alpha_i')$. A user i who is a member of coalition Dev does not improve his or her payoff. If the rates of some of the deviants are increased, then the rates of some other users from the coalition must decrease. If $(\alpha_i')_{i\in \mathrm{Dev}}$ satisfies

$$\sum_{i\in \mathrm{Dev}} \alpha_i' = C_{\mathcal{N}} - \sum_{i\notin \mathrm{Dev}} \alpha_i\,,$$

then some users in coalition Dev have increased their rates compared with $(\alpha_i)_{i\in \mathrm{Dev}}$ while others in Dev have decreased their rates of transmission (because the total rate is the constant $C_{\mathcal{N}} - \sum_{i\notin \mathrm{Dev}} \alpha_i$). The users in Dev with a lower rate $\alpha_i' \leq \alpha_i$ do not benefit by being a member of the coalition (the Shapley criterion of membership of coalition does not hold). And this holds for any $\emptyset \subsetneqq \mathrm{Dev} \subseteqq \mathcal{N}$. This completes the proof. □

Corollary 3.1. *In the constrained rate allocation game, NEs and strong equilibria in pure strategies coincide.*

3.2.3.2 Constrained Potential Function for Local Interaction

We introduce the following function:

$$W(\alpha) = 1\!\!1_C(\alpha) \sum_{i=1}^{N} g_i(\alpha_i)\,,$$

where $1\!\!1_C$ is the indicator function of C, i.e., $1\!\!1_C(\alpha) = 1$ if $\alpha \in C$ and 0 otherwise. Function W satisfies

$$W(\alpha) - W(\beta_i, \alpha_{-i}) = g_i(\alpha_i) - g_i(\beta_i),\ \forall \alpha, (\beta_i, \alpha_{-i}) \in C.$$

If g_i is differentiable, then one has

$$\frac{\partial}{\partial \alpha_i} W(\alpha) = g_i'(\alpha_i) = \frac{\partial}{\partial \alpha_i} u_i$$

in the interior of the capacity region C, and W is a constrained potential function [3] in pure strategies.

Corollary 3.2. *The local maximizers of W in C are pure NEs. Global maximizers of W in C are both constrained strong equilibria and social optima for the local interaction.*

3.2.3.3 Strong Price of Anarchy

Throughout this subsection, we assume that the functions g_i are the identity function, i.e., $g_i(x) = id(x) := x$. One metric used to measure how much the performance of decentralized systems is affected by the selfish behavior of its components is the *price of anarchy* (PoA). We present a similar price for strong equilibria under the coupled rate constraints. This notion of PoA can be seen as an *efficiency metric* that measures the *price of selfishness* or decentralization and has been extensively used in the context of congestion games or routing games where typically users have to minimize a cost function [37, 38]. In the context of rate allocation in the multiple-access channel, we define an equivalent measure of PoA for rate maximization problems. One of the advantages of a strong equilibrium is that it has the potential to reduce the distance between the optimal solution and the solution obtained as an outcome of selfish behavior, typically in cases where the capacity constraint is violated at each time. Since the constrained rate allocation game has strong equilibria, we can define the strong price of anarchy (SPoA), introduced in [12], as the ratio between the payoff of the worst constrained strong equilibrium and the social optimum value which $C_{\mathcal{N}}$.

Theorem 3.2. *The SPoA of the constrained rate allocation game is 1 for* $g_i(x) = x$.

Note that for $g_i \neq id$, the constrained SPoA (CSPoA) can be less than one. However, the optimistic PoA of the *best constrained equilibrium*, also called the *price of stability* [13], is one for any function g_i, i.e., the efficiency of the "best" equilibria is 100 %.

3.2.4 Selection of Pure Equilibria

We showed in the previous sections that our rate allocation game has a continuum of pure NEs and strong equilibria. We address now the problem of selecting one equilibrium that has a certain desirable property: the normalized pure NE, introduced in [26]; see also [20, 22, 28]. We introduce the problem of constrained maximization faced by each user when the other rates are at the maximal face of polytope C:

$$\max_{\alpha} \quad u_i(\alpha) \tag{3.5}$$

$$\text{s.t.} \alpha_1 + \cdots + \alpha_N = C_{\mathcal{N}}, \tag{3.6}$$

for which the corresponding Lagrangian for user i is given by

$$L_i(\alpha, \zeta) = u_i(\alpha) - \zeta_i \left(\sum_{i=1}^{N} \alpha_i - C_{\mathcal{N}} \right).$$

From Karush–Kuhn–Tucker optimality conditions it follows that there exists $\zeta \in \mathbb{R}^N$ such that

$$g_i'(\alpha_i) = \zeta_i, \ \sum_{i=1}^{N} \alpha_i = C_{\mathcal{N}}.$$

For a fixed vector ζ with identical entries, define the normal form game $\Gamma(\zeta)$ with N users, where actions are taken as rates and the payoffs given by $L(\alpha, \zeta)$. A normalized equilibrium is an equilibrium of the game $\Gamma(\zeta^*)$, where ζ^* is normalized into the form $\zeta_i^* = \frac{c}{\tau_i}$ for some $c > 0, \tau_i > 0$. We now have the following result due to Goodman [20], which implies Rosen's condition on uniqueness for strict concave games.

Theorem 3.3. *Let u_i be a smooth and strictly concave function in α_i, with each u_i convex in α_{-i}, and let there exist some ζ such that the weighted nonnegative sum of the payoffs $\sum_{i=1}^{N} \zeta_i u_i(\alpha)$ is concave in α. Then, the matrix $G(\alpha,\zeta) + G^T(\alpha,\zeta)$ is negative definite (which implies uniqueness), where $G(\alpha,\zeta)$ is the Jacobian with respect to α of*

$$h(\alpha,\zeta) := [\zeta_1 \nabla_1 u_1(\alpha), \zeta_2 \nabla_2 u_2(\alpha), \ldots, \zeta_N \nabla_N u_N(\alpha)]^T$$

and G^T is the transpose of matrix G.

This now leads to the following corollary for our problem.

Corollary 3.3. *If g_i are nondecreasing strictly concave functions, then the rate allocation game has a unique normalized equilibrium that corresponds to an equilibrium of the normal form game with payoff $L(\alpha,\zeta^*)$ for some ζ^*.*

3.2.5 Stability and Dynamics

In this subsection, we study the stability of equilibria and several classes of evolutionary game dynamics under a symmetric case, i.e., $P_i = P, h_i = h, g_i = g, \mathcal{A}_i = \mathcal{A}$, $\forall i \in \mathcal{N}$. We will drop subscript index i where appropriate. We show that the associated evolutionary game has a unique pure constrained ESS.

Proposition 3.2. *The collection of rates $\alpha = \left(\frac{C_{\mathcal{N}}}{N}, \ldots, \frac{C_{\mathcal{N}}}{N}\right)$, i.e., the Dirac distribution concentrated on the rate $\frac{C_{\mathcal{N}}}{N}$, is the unique symmetric pure NE.*

Proof. Since the constrained rate allocation game is symmetric, there exists a symmetric (pure or mixed) NE. If such an equilibrium exists in pure strategies, each user transmits with the same rate r^*. It follows from Proposition 3.1 of Sect. 3.2.2 and the bound $r_N \leq \frac{C_{\mathcal{N}}}{N}$ that r^* satisfies $Nr^* = C_{\mathcal{N}}$ and r^* is feasible. □

Since the set of feasible actions is convex, we can define a convex combination of rates in the set of feasible rates. For example, $\varepsilon\alpha' + (1-\varepsilon)\alpha$ is a feasible rate if

α' and α are feasible. The symmetric rate profile $(r, r, \ldots, r)$ is feasible if and only if $0 \leq r \leq r^* = \frac{C_{\mathcal{N}}}{N}$. We say that rate r is a constrained ESS if it is feasible and for every *mutant strategy* mut $\neq \alpha$ there exists $\varepsilon_{\text{mut}} > 0$ such that

$$\begin{cases} r_\varepsilon := \varepsilon \text{ mut} + (1-\varepsilon) r \in C & \forall \varepsilon \in (0, \varepsilon_{\text{mut}}) \\ u(r, r_\varepsilon, \ldots, r_\varepsilon) > u(\text{mut}, r_\varepsilon, \ldots, r_\varepsilon) & \forall \varepsilon \in (0, \varepsilon_{\text{mut}}) \end{cases}$$

Theorem 3.4. *The pure strategy $r^* = \frac{C_{\mathcal{N}}}{N}$ is a constrained ESS.*

Proof. Let mut $\leq r^*$ The rate ε mut $+ (1-\varepsilon) r^*$ is feasible, which implies that mut $\leq r^*$ (because r^* is the maximum symmetric rate achievable). Since mut $\neq r^*$, mut is strictly lower than r^*. By monotonicity of the function g, one has $u(r^*, \varepsilon \text{ mut} + (1-\varepsilon) r^*) > u(\text{mut}, \varepsilon \text{ mut} + (1-\varepsilon) r^*)$, $\forall \varepsilon$. This completes the proof. □

3.2.5.1 Symmetric Mixed Strategies

Define the mixed capacity region $\mathcal{M}(C)$ as the set of measure profile $(\mu_1, \mu_2, \ldots, \mu_N)$ such that

$$\int_{\mathbb{R}_+^{|\Omega|}} \left(\sum_{i \in \Omega} \alpha_i \right) \bigotimes_{i \in \Omega} \mu_i(\mathrm{d}\alpha_i) \leq C_\Omega, \ \forall \Omega \subseteq 2^{\mathcal{N}}.$$

Then, the payoff of the action $a \in \mathbb{R}_+$ satisfying $(a, \lambda, \ldots, \lambda) \in \mathcal{M}(C)$ can be defined as

$$F(a, \mu) = \int_{[0,\infty[^{N-1}} u(a, b_2, \ldots, b_N) \ \nu_{N-1}(\mathrm{d}b), \tag{3.7}$$

where $\nu_k = \bigotimes_1^k \mu$ is the product measure on $[0, \infty[^k$. The constraint set becomes the set of probability measures on $\mathbb{R}_+$ such that

$$0 \leq \mathbb{E}(\mu) := \int_{\mathbb{R}_+} \alpha_i \ \mu(\mathrm{d}\alpha_i) \leq \frac{C_{\mathcal{N}}}{N} < C_{\{1\}}.$$

Lemma 3.2. *The payoff can be obtained as follows:*

$$\begin{aligned} F(a, \mu) &= \mathbb{1}_{[0, C_{\mathcal{N}} - (N-1)\mathbb{E}(\mu)]} \times g(a) \times \int_{b \in \mathcal{D}_a} \nu_{N-1}(\mathrm{d}b) \\ &= \mathbb{1}_{[0, C_{\mathcal{N}} - (m-1)\mathbb{E}(\mu)]} \times g(a) \nu_{N-1}(\mathcal{D}_a), \end{aligned}$$

where $\mathcal{D}_a = \{(b_2, \ldots, b_N) \mid (a, b_2, \ldots, b_N) \in C\}$.

Proof. If the rate does not satisfy the capacity constraints, then the payoff is 0. Hence the *rational* rate for user i is lower than $C_{\{i\}}$. Fix a rate $a \in [0, C_{\{i\}}]$. Let $D_\Omega^a := C_\Omega - a\delta_{\{1 \in \Omega\}}$. Then, a necessary condition to have a nonzero payoff is

$(b_2,\ldots,b_N) \in \mathcal{D}_a$, where $\mathcal{D}_a = \{(b_2,\ldots,b_N) \in \mathbb{R}_+^{N-1},\ \sum_{i\in\Omega, i\neq 1} b_i \leq D_\Omega^a,\ \Omega \subseteq 2^{\mathcal{N}}\}$. Thus, we have

$$\begin{aligned} F(a,\mu) &= \int_{\mathbb{R}_+^{N-1}} u(a,b_2,\ldots,b_N)\ \nu_{N-1}(\mathrm{d}b) \\ &= \int_{b\in\mathbb{R}_+^{N-1},\ (a,b)\in C} g(a)\ \nu_{N-1}(\mathrm{d}b) \\ &= \mathbb{1}_{[0,C_{\mathcal{N}}-(N-1)\mathbb{E}(\mu)]}\ g(a) \times \int_{b\in\mathcal{D}_a} \nu_{N-1}(\mathrm{d}b). \end{aligned}$$ □

3.2.5.2 Constrained Evolutionary Game Dynamics

The class of evolutionary games in a large population provides a simple framework for describing strategic interactions among large numbers of users. In this subsection, we turn to modeling the behavior of the users who play such games. Traditionally, predictions of behavior in game theory are based on some notion of equilibrium, typically Cournot equilibrium, Bertrand equilibrium, NE, Stackelberg solution, Wardrop equilibrium, or some refinement thereof. These notions require the assumption of equilibrium knowledge, which posits that each user correctly anticipates how his opponents will act. The equilibrium knowledge assumption is too strong and is difficult to justify, in particular in contexts with large numbers of users. As an alternative to the equilibrium approach, we propose an explicitly dynamic updating choice, a procedure in which users myopically update their behavior in response to their current strategic environment. This dynamic procedure does not assume the automatic coordination of users' actions and beliefs, and it can derive many specifications of users' choice procedures. These procedures are specified formally by defining a revision of rates called *revision protocol* [27]. A revision protocol takes current payoffs and the current mean rate and maps to conditional switch rates, which describe how frequently users in some class playing rate α who are considering switching rates switch to strategy α'. Revision protocols are flexible enough to incorporate a wide variety of paradigms, including ones based on imitation, adaptation, learning, optimization, etc.

We use here a class of continuous evolutionary dynamics. We refer to [14, 30, 31] for evolutionary game dynamics with or without time delays. The normalized continuous-time evolutionary game dynamics on the measure space $(\mathcal{A},\mathcal{B}(\mathcal{A}),\mu)$ is given by

$$\dot{\lambda}_t(E) = \int_{x\in E} V(x,\lambda_t)\mu(\mathrm{d}x), \tag{3.8}$$

where

$$V(x,\lambda_t) = K\left[\int_{a\in\mathcal{A}} \beta_x^a(\lambda_t)\mu(\mathrm{d}a) - \frac{\lambda_t(E)}{\mu(E)}\int_{a\in\mathcal{A}} \beta_x^a(\lambda_t)\mu(\mathrm{d}a)\right],$$

β_x^a represents the rate of mutation from a to x, and K is a growth parameter. $\beta_a^x(\lambda_t) = 0$ if (x, λ_t) or (a, λ_t) is not in the (mixed) capacity region, and E is a μ-measurable subset of $\mathcal{A}$. At each time t, probability measure λ_t satisfies $\frac{\mathrm{d}}{\mathrm{d}t}\lambda_t(\mathcal{A}) = 0$.

We examine the Brown–von Neumann–Nash dynamics, Smith dynamics, and replicator dynamics, where $F(a, \lambda_t)$ is the payoff, as given in (3.7).

[RD-1] Constrained Brown–von Neumann–Nash dynamics.

$$\beta_a^x(\lambda_t) = \begin{cases} \max(F(a,\lambda_t) - \int_x F(x,\lambda_t)\,\mathrm{d}x, 0) & \text{if } (a,\lambda_t),\ (x,\lambda_t) \in \mathcal{M}(C), \\ 0 & \text{otherwise.} \end{cases}$$

As we can see, β_x^a is independent of a. Thus, in the unconstrained case, the first double integral becomes

$$\int_{a\in\mathcal{A}} \int_{x\in E} \beta_x^a(\lambda_t)\mu(\mathrm{d}x)\mu(\mathrm{d}a) = \int_{x\in E} \beta_x^a(\lambda_t)\mu(\mathrm{d}x),$$

and the second term

$$\frac{\lambda_t(E)}{\mu(E)} \int_{a\in\mathcal{A}} \int_{x\in E} \beta_a^x(\lambda_t)\mu(\mathrm{d}x)\mu(\mathrm{d}a) = \lambda_t(E) \int_{a\in\mathcal{A}} \beta_a^x(\lambda_t)\mu(\mathrm{d}a).$$

The difference between the two terms gives rise to

$$\dot{\lambda}_t(E) = \int_{x\in E} \beta_x^a(\lambda_t)\mu(\mathrm{d}x) - \lambda_t(E) \int_{a\in\mathcal{A}} \beta_a^x(\lambda_t)\mu(\mathrm{d}a).$$

The unnormalized continuous-time evolutionary game dynamics on the measure space $(\mathcal{A}, \mathcal{B}(\mathcal{A}), \mu)$ is given by

$$\dot{\lambda}_t(E) = \int_{x\in E} V(x,\lambda_t)\lambda_t(\mathrm{d}x), \tag{3.9}$$

where

$$V(x,\lambda_t) = K\left[\int_{a\in\mathcal{A}} \beta_x^a(\lambda_t)\lambda_t(\mathrm{d}a) - \int_{a\in\mathcal{A}} \beta_a^x(\lambda_t)\lambda_t(\mathrm{d}a)\right].$$

[RD-2] Constrained replicator dynamics. Let

$$\beta_a^x(\lambda_t) = \begin{cases} \max(F(a,\lambda_t) - F(x,\lambda_t), 0) & \text{if } (a,\lambda_t),\ (x,\lambda_t) \in \mathcal{M}(C), \\ 0 & \text{otherwise.} \end{cases}$$

In the unconstrained case, we can easily check that

$$V(x,\lambda_t) = K\int_{a\in\mathcal{A}}[\max(F(x,\lambda_t)-F(a,\lambda_t),0)-\max(F(a,\lambda_t)-F(x,\lambda_t),0)]\lambda_t(\mathrm{d}a),$$

which has the replicator form. Using the fact that

$$[\max(F(x,\lambda_t)-F(a,\lambda_t),0)-\max(F(a,\lambda_t)-F(x,\lambda_t),0)] = F(x,\lambda_t)-F(a,\lambda_t),$$

we obtain

$$V(x,\lambda_t) = K\int_{a\in\mathcal{A}}[F(x,\lambda_t)-F(a,\lambda_t)]\lambda_t(\mathrm{d}a) = K\left[F(x,\lambda_t)-\int_{a\in\mathcal{A}}F(a,\lambda_t)\lambda_t(\mathrm{d}a)\right],$$

where the difference between the payoff at x and the average payoff.

[RD-3] Constrained θ-Smith dynamics.

$$\beta_a^x(\lambda_t) = \begin{cases} \max(F(a,\lambda_t)-F(x,\lambda_t),0)^\theta & \text{if } (a,\lambda_t),\ (x,\lambda_t)\in\mathcal{M}(C) \\ 0 & \text{otherwise} \end{cases},\ \theta\geq 1.$$

A common property that applies to all these dynamics is that the set of NEs is a subset of rest points (stationary points) of the evolutionary game dynamics. Here we extend the concepts of these dynamics to evolutionary games with a continuum action space and coupled constraints, and interactions with more than two users. The counterparts of these results in discrete action space can be found in [21, 27].

Theorem 3.5. *Any NE of a game is a rest point of the following evolutionary game dynamics: constrained Brown–von Neumann–Nash, generalized Smith dynamics, and replicator dynamics. Futhermore, the ESS set is a subset of the rest points of these constrained evolutionary game dynamics.*

Proof. It is clear for pure equilibria using the revision protocols β of these dynamics. Let λ be an equilibrium. For any rate a in the support of λ, $\beta_x^a = 0$ if $F(x,\lambda)\leq F(a,\lambda)$. Thus, if λ is an equilibrium, then the difference between the microscopic inflow and outflow is $V(a,\lambda)=0$, given that a is the support of measure λ. The last assertion follows from the fact that the ESS (if it exists) is an equilibrium that is a rest point of these dynamics. □

Let λ be a finite Borel measure on $[0,C_{\{i\}}]$ with full support. Suppose g is continuous on $[0,C_{\{i\}}]$. Then λ is a rest point of the Brown–von Neumann–Nash dynamics if and only if λ is a symmetric NE. Note that the choice of topology is an important issue when defining the convergence of dynamics and the stability of the dynamics. The most used topology in this area is the topology of the weak convergence to measure the closeness of two states of the system. Different

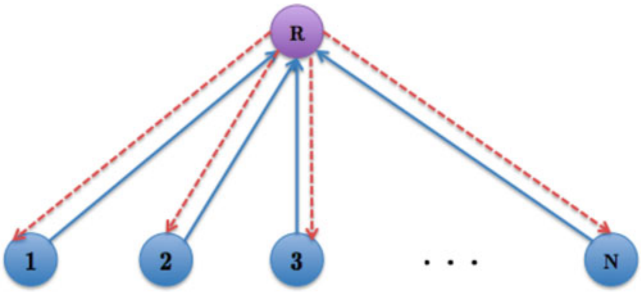


Fig. 3.3 Signaling between multiple senders and a receiver

distances (Prohorov metric, metric on bounded and Lipschitz continuous functions on $\mathcal{A}$) have been proposed. We refer the reader to [24] and references therein for more details on *evolutionarily robust strategy* and stability notions.

3.2.6 Correlated Equilibrium

In this subsection, we analyze constrained correlated equilibria of multiple-access control (MAC) games. Building on the signaling in the one-shot game, we formulate a system of evolutionary MAC games with evolutionary game dynamics that describe the evolution of signaling, beliefs, rate control, and channel selection.

We focus on correlated equilibrium in the single-receiver case. Correlated strategies are based on signaling structures before decisions are made on rates. Different scenarios (with or without a mediator, virtual mediator, or cryptographic multistage signaling structure) have been proposed in the literature [15–17, 35].

Figure 3.3 illustrates the signaling between multiple transmitters and one receiver. The receiver can act as a signaling device to mediate the behaviors of the transmitters. The correlated equilibrium has a strong connection with cryptography in that the private signal sent to the users can be realized by the coding and decoding in the network [35].

Let $\mathcal{B}$ be the set of signals $\beta = [\beta_i, \beta_{-i}] \in \mathbb{R}^N$. The values β from the set of signals need to be in the feasible set $C \subset \mathbb{R}^N$. Let $\mu \in \Delta\mathcal{B}$ be a probability measure over the set $\mathcal{B}$. A constrained correlated equilibrium (CCE) μ^* needs to satisfy the following set of inequalities:

$$\int \mathrm{d}\mu^*(\beta_i, \beta_{-i}) \left[u_i(\alpha_i, \alpha_{-i}, \mid \beta_i) - u_i(\alpha_i', \alpha_{-i} \mid \beta_i)\right] \geq 0, \forall i \in \mathcal{N}, \alpha_i' \in \mathcal{A}_i(\alpha_{-i}).$$

Define a rule of assignment of user i as a map from his signals to his action set $\bar{r}_i : \ \beta_i \longmapsto \alpha_i$. A CCE is then characterized by

$$\int d\mu^*(\beta)\,[u_i(\alpha_i,\alpha_{-i}\mid\beta_i)-u_i(r_i(\beta_i),\alpha_{-i})]$$
$$\geq 0, \forall i\in\mathcal{N}, \forall r_i \text{ such that } \bar{r}_i(.)\in\mathcal{A}_i(\alpha_{-i}). \quad (3.10)$$

Theorem 3.6. *The set of constrained pure NEs of the MISO game is given by*

$$\textit{max-face}(C)=\left\{(\alpha_1,\ldots,\alpha_N)\mid \alpha_i\geq 0,\ \sum_{k\in\mathcal{N}}\alpha_k=C_{\mathcal{N}}\right\}.$$

We can characterize the CCE using the preceding results as follows.

Lemma 3.3. *Any mixture of constrained pure NEs of the MISO game is a CCE.*

Note that the set of CCEs is bigger than the set of constrained NEs. For example, in a two-user case, the distribution $\frac{1}{2}\delta_{(r_1,C_{\{1,2\}}-r_1)}+\frac{1}{2}\delta_{(C_{\{1,2\}}-r_2,r_2)}$ is different than the Dirac distribution $\delta_{(\frac{r_1+C_{\{1,2\}}-r_2}{2},\frac{r_2+C_{\{1,2\}}-r_1}{2})}$.

Proof. Let μ be a probability distribution over some constrained pure equilibria. Then $\mu\in\Delta(\text{max-face}(C))$. For any β such that $\mu(\beta)\neq 0$ one has

$$[u_i(\alpha_i,\alpha_{-i}\mid\beta_i)-u_i(\bar{r}_i(\beta_i),\alpha_{-i})]\geq 0$$

for any measurable function $\bar{r}_i:\ [0,C_{\{i\}}]\longrightarrow[0,C_{\{i\}}]$. Thus, μ is a CCE. □

Corollary 3.4. *Any convex combination of extreme points of the convex compact set*

$$\textit{max-face}(C)=\left\{\alpha=(\alpha_1,\ldots,\alpha_N)\;\middle|\;\alpha_i\geq r_i,\ \sum_{k\in\mathcal{N}}\alpha_k=C_{\mathcal{N}}\right\}$$

is a CCE. Moreover, any probability distribution over the maximal face of the capacity region max-face(C) is a CCE distribution.

3.3 Hybrid AWGN Multiple-Access Control

In this section, we extend the single-receiver case to one with multiple receivers. The multi-input and multioutput (MIMO) channel access game has been studied in the context of power allocation and control. For instance, the authors in [6] formulate a two-player zero-sum game where the first player is the group of transmitters and the second one is the set of MIMO subchannels. In [5], the authors formulate an N-person noncooperative power allocation game and study its equilibrium under two different decoding schemes.

Here, we formulate a hybrid multiple-access game where users are allowed to select their rates and channels under capacity constraints. We first obtain general results on the existence of the equilibrium and methods to characterize it. In addition, we investigate the long-term behavior of the strategies and apply evolutionary game dynamics to both rates and channel selection probabilities. We show that G-function-based dynamics is appropriate for our hybrid model by viewing the channel selection probabilities as strategies that determine the fitness of rate selection. Using the generalized Smith dynamics for channel selection, we are able to build an overall hybrid evolutionary dynamics for the static model. Based on simulations, we confirm the validity of the proposed dynamics and the correspondence between the rest point of the dynamics and the NE.

3.3.1 Hybrid Model with Rate Control and Channel Selection

In this subsection, we establish a model for multiple users and multiple receivers. Each user needs to decide the rate at which to transmit and the channel to pick. We formulate a game $\overline{\Xi} = \langle \mathcal{N}, (\mathcal{A}_i)_{i\in\mathcal{N}}, (\overline{U}_i)_{i\in\mathcal{N}}\rangle$, in which the decision variable is $(\alpha_i, \mathbf{p}_i)$, and $\mathbf{p}_i = [p_{ij}]_{j\in\mathcal{J}}$ is a J-dimensional vector, where p_{ij} is the probability that user $i \in \mathcal{N}$ will choose channel $j \in \mathcal{J}$ and p_{ij} needs to satisfy the probability measure constraints

$$\sum_{j\in\mathcal{J}} p_{ij} = 1, p_{ij} \geq 0, \forall i \in \mathcal{N}. \tag{3.11}$$

The game $\overline{\Xi}$ is *asymmetric* in the sense that the sets of strategies of the users are different and the payoffs are not symmetric.

Let $C_{j,\Omega} := \log(1 + \sum_{i\in\Omega} \frac{P_{ij}h_{ij}}{\sigma_0^2})$ be the capacity for a subset $\Omega \subseteq \mathcal{N}$ of users at receiver $j \in \mathcal{J}$ and $r_{ij,\Omega} := \log(1 + \frac{P_{ij}h_{ij}}{\sigma_0^2 + \sum_{i'\in\Omega, i'\neq i} P_{i'j}h_{i'j}})$ the bound rate of a user i when the signals of the $|\Omega| - 1$ other users are treated as noise at receiver j. Each receiver j has a capacity region $C(j)$ given by

$$C(j) = \left\{ (\alpha, \mathbf{p}_j) \in [0,1]^N \times \mathbb{R}_+^N \;\middle|\; \sum_{i\in\mathcal{N}} \alpha_i p_{ij} \leq C_{j,\Omega}, \forall\, \emptyset \subset \Omega_j \subseteq \mathcal{N},\ j \in \mathcal{J} \right\}. \tag{3.12}$$

The expected payoff function $\overline{U}_i : \prod_{i=1}^N \mathcal{A}_i \longrightarrow \mathbb{R}_+$ of the game is given by

$$\overline{U}_i(\alpha_i, \mathbf{p}_i, \alpha_{-i}, \mathbf{p}_{-i}) = \mathbb{E}_{\mathbf{p}_i}[u_{ij}(\alpha, \mathbf{P})] = \sum_{j\in\mathcal{J}} p_{ij} u_{ij}(\alpha, \mathbf{P}), \tag{3.13}$$

where $\alpha = (\alpha_i, \alpha_{-i}) \in \mathbb{R}_+^N$ and $\mathbf{P} = (\mathbf{p}_i, \mathbf{p}_{-i}) \in [0,1]^{N\times J}$, with $\mathbf{p}_i \in [0,1]^J, \mathbf{p}_{-i} \in [0,1]^{(N-1)\times J}$. Assume that the utility u_{ij} of a user i transmitting to receiver j is only dependent on the user himself and is described by a positive and strictly increasing function $g_i : \mathbb{R}_+ \to \mathbb{R}_+$, i.e., $u_{ij} = g_i, \forall j \in \mathcal{J}$, when capacity constraints are satisfied.

With the presence of coupled constraints (3.12) from each receiver and probability measure constraint (3.11), each sender has his own individual optimization problem (IOP) given as follows:

$$\begin{aligned}\max_{\mathbf{p}_i,\alpha_i} \quad & \textstyle\sum_{j\in\mathcal{J}} p_{ij} g_i(\alpha_i p_{ij});\\ \text{s.t.} \quad & \textstyle\sum_{j\in\mathcal{J}} p_{ij} = 1, \forall i \in \mathcal{N};\\ & p_{ij} \geq 0, \forall i \in \mathcal{N}, j \in \mathcal{J};\\ & (\alpha, \mathbf{p}_j) \in C(j), \forall j \in \mathcal{J}.\end{aligned}$$

Denote the feasible set of IOP by $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, where

$$\mathcal{F}_1 = \left\{ \alpha \in \mathbb{R}_+^N \,\middle|\, (\alpha, \mathbf{P}) \in \cap_{j\in\mathcal{J}} C(j), \mathbf{P} \in \mathcal{F}_2 \right\}, \tag{3.14}$$

$$\mathcal{F}_2 = \left\{ \mathbf{P} \in \mathbb{R}^{N\times J} \,\middle|\, \sum_{j\in\mathcal{J}} p_{ij} = 1, p_{ij} \geq 0, \forall i \in \mathcal{N}, j \in \mathcal{J} \right\}. \tag{3.15}$$

The action set of each user can thus be described by

$$\mathcal{A}_i(\alpha_{-i}, \mathbf{p}_{-i}) = \{(\alpha_i, \alpha_{-i}) \in \mathcal{F}_1, (\mathbf{p}_i, \mathbf{p}_{-i}) \in \mathcal{F}_2\}. \tag{3.16}$$

3.3.1.1 An Example

Suppose we have three users and three receivers, that is, $\mathcal{N} = \{1,2,3\}$ and $\mathcal{J} = \{1,2,3\}$. The capacity region at receiver 1 is given by

$$C(1) = \left\{ \begin{array}{c} \alpha_i \geq 0,\ i \in \{1,2,3\} \\ p_{11}\alpha_1 \leq \log(1 + \frac{P_1 h_1}{\sigma_0^2}) \\ p_{21}\alpha_2 \leq \log(1 + \frac{P_2 h_2}{\sigma_0^2}) \\ p_{31}\alpha_3 \leq \log(1 + \frac{P_3 h_3}{\sigma_0^2}) \\ p_{11}\alpha_1 + p_{21}\alpha_2 \leq \log(1 + \frac{P_1 h_1 + P_2 h_2}{\sigma_0^2}) \\ p_{11}\alpha_1 + p_{31}\alpha_3 \leq \log(1 + \frac{P_1 h_1 + P_2 h_2}{\sigma_0^2}) \\ p_{21}\alpha_2 + p_{31}\alpha_3 \leq \log(1 + \frac{P_1 h_1 + P_2 h_2}{\sigma_0^2}) \\ p_{11}\alpha_1 + p_{21}\alpha_2 + p_{31}\alpha_3 \leq \log(1 + \frac{P_1 h_1 + P_2 h_2 + P_3 h_3}{\sigma_0^2}) \\ 0 \leq p_{i1} \leq 1,\ i \in \{1,2,3\} \end{array} \right\}.$$

This can be rewritten as

$$C(1) = \left\{ \mathbf{p}_1 = \begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \end{bmatrix} \in [0,1]^3, \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \in \mathbb{R}^3_+ \;\middle|\; M_3 \begin{bmatrix} p_{11}\alpha_1 \\ p_{21}\alpha_2 \\ p_{31}\alpha_3 \end{bmatrix} \leq \begin{bmatrix} C_{1,\{1\}} \\ C_{1,\{2\}} \\ C_{1,\{3\}} \\ C_{1,\{1,2\}} \\ C_{1,\{1,3\}} \\ C_{1,\{2,3\}} \\ C_{1,\{1,2,3\}} \end{bmatrix} \right\},$$

where $C_{1,\Omega} = \log\left(1 + \sum_{i\in\Omega} \frac{P_{i1}h_{i1}}{\sigma_0^2}\right)$ and M_3 is a totally unimodular matrix: $M_3 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Capacity regions at receivers 2 and 3 can be obtained in a similar way.

3.3.2 Characterization of Constrained Nash Equilibria

In this subsection, we characterize the NEs of the defined game $\overline{\Xi}$ under the given capacity constraint. We use the following theorem to prove the existence of an NE for the case where the rates are predetermined; this result is then used to solve the game for the case where both the rates and the connection probabilities are (joint) decision variables.

Theorem 3.7 (Başar and Olsder [34]). *Let $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \cdots \times \mathcal{A}_N$ be a closed, bounded, and convex subset of $\mathbb{R}^N$, and for each $i \in \mathcal{N}$ let the payoff functional $\overline{U}_i : \mathcal{A} \to \mathbb{R}$ be jointly continuous in $\mathcal{A}$ and concave in a_i for every $a_j \in \mathcal{A}_j, j \in \mathcal{N}, j \neq i$. Then, the associated N-person non-zero-sum game admits an NE in pure strategies.*

Applying Theorem 3.7, we have the following results immediately.

Proposition 3.3. *Suppose $\alpha_i, i \in \mathcal{N}$, are predetermined feasible rates. Let feasible set $\mathcal{F}$ be closed, bounded, and convex. If g_i in the IOP are continuous on $\mathcal{F}$ and concave in $\mathbf{p}_i$ (without the assumption of their being positive and strictly increasing) and the expected payoff functions $\overline{U}_i : \mathbb{R}^N_+ \times [0,1]^{N\times J} \to \mathbb{R}$ are concave in $\mathbf{p}_i$ and continuous on $\mathcal{F}$, then the static game admits an NE.*

The existence result in Proposition 3.3 only captures the case where the rates α_i are predetermined, and it relies on the convexity requirement of the utility functions g_i. We can actually obtain a stronger existence result by observing that the formulated game $\overline{\Xi}$ is a potential game with a potential function given by

$$\Psi(\alpha,\mathbf{P}) = \sum_i f_i(\alpha_i,\mathbf{p}_i) = \sum_i \sum_j p_{ij} g_i(\alpha_i p_{ij}), \tag{3.17}$$

where $f_i = \sum_j p_{ij} g_i(\alpha_i p_{ij})$, the expected payoff to user i. This is captured by the proposition below.

Proposition 3.4. *The hybrid rate control game $\overline{\Xi}$ admits an NE.*

Proof. Let us formulate a centralized optimization problem (COP) as follows:

$$\begin{aligned} \max_{\alpha,\mathbf{P}} \quad & \Psi(\alpha,\mathbf{P}), \\ \text{s.t.} \quad & (\alpha,\mathbf{P}) \in \mathcal{F}. \end{aligned}$$

Using the result in [3], we can conclude that if there exists a solution to the COP, then there exists an NE to the game $\overline{\Xi}$. Since $\mathcal{F}$ is compact and nonempty, and the objective function is continuous, then there exists a solution to the COP and thus an NE to the game. □

The foregoing problem is generally not convex, and the uniqueness of the NE may not be guaranteed. However, we can still further characterize the NE through the following propositions.

Proposition 3.5. *Let $\beta_{ij} := \alpha_i p_{ij}$. Without predetermining α, suppose that $(\mathbf{p}_{-i}, \alpha_{-i})$ is feasible. A best response strategy at receiver $j \in \mathcal{J}$ for user i must satisfy*

$$0 \leq p_{ij}\alpha_i \leq C_{j,\Omega_j} - \sum_{k \neq i} \alpha_k p_{kj}, \forall \Omega_j, \tag{3.18}$$

where $\Omega_j := \{\Omega \in 2^{\mathcal{N}} \mid i' \in \Omega, p_{i'j} > 0\}$ is the set of users transmitting to receiver j. Since g_i is a nondecreasing function, the best correspondence at j is

$$\beta_{ij}^* = \alpha_i^* p_{ij}^* = \max\left(r_{ij,\mathcal{N}}, \min_{\Omega_j} \left(C_{j,\Omega_j} - \sum_{i' \neq i} \alpha_{i'} p_{i'j} \right) \right), \tag{3.19}$$

where $r_{ij,\mathcal{N}}$ is the bound on the rate of user i when signals of $|\mathcal{N}|-1$ other users are treated as noise.

Proof. The proof is immediate by observing that the rate of user i at receiver j must satisfy (3.18) due to the coupled constraints. Thus, the maximum rate that user i can use to transmit to receiver j without violating the constraints is clearly the minimum of $C_{j,\Omega_j} - \sum_{i' \neq i} \alpha_{i'} p_{i'j}$ over all Ω_j. Since the payoff is a nondecreasing function, the best response for i at receiver j is given by (3.19). □

Proposition 3.6. *Let* $K_i^* = \arg\max_{j\in\mathcal{J}} g_{ij}(\beta_{ij})$. *If* $K_i^* = \{k^*\}$ *is a singleton, then the best response for user i is to choose*

$$\begin{cases} p_{ij} = 1 & \text{if } j = k^*, \\ p_{ij} = 0 & \text{otherwise,} \end{cases}$$

and we can determine α_i *by* $\alpha_i = \frac{\beta_{ik^*}}{p_{ik^*}}$.

If $|K_i^*| \geq 2$, *then the best response correspondence is*

$$\begin{cases} \mathbf{p}_i \in \Delta(K_i^*) & \text{if } j \in K_i^*, \\ 0 & \text{otherwise.} \end{cases}$$

We can determine α_i *from* β_{ij} *by* $\alpha_i = \sum_{j\in K_i^*} \beta_{ij}$.

Proof. Since the expected utility is given in the form of

$$U_i(\alpha_i, \mathbf{p}_i, \alpha_{-i}, \mathbf{p}_{-i}) = \mathbb{E}_{\mathbf{p}_i}[u_{ij}(\alpha_i p_{ij})],$$

the expected utility under the best response is $U_i = \mathbb{E}_{\mathbf{p}_i}[u_{ij}(\beta_{ij})]$. If for a singleton k^* such that $k^* = \arg\max_{i\in\mathcal{N}} g_{ij}(\beta_{ij})$, we can assign all the weight $p_{ik^*} = 1$ to maximize the expected utility. Since $\beta_{ik^*} = \alpha_i p_{ik^*}$, then $\alpha = \beta_{ik^*}/p_{ik^*} = \beta_{ik^*}$. If the set K^* is not a singleton, then without loss of generality, we can pick two indices $j, j' \in K^*$ such that $\beta_{ij} = \alpha_i p_{ij}$ and $\beta_{ij'} = \alpha_i p_{ij'}$, leading to $u_{ij}(\beta_{ij}) = u_{ij'}(\beta_{ij'})$. Since the utilities to transmit using j and j' are the same, we can assign an arbitrary (two-point) distribution, p_{ij} and $p_{ij'}$, over them, with $p_{ij} + p_{ij'} = 1$. Therefore, $\beta_{ij} + \beta_{ij'} = \alpha_i(p_{ij} + p_{ij'}) = \alpha_i$. □

3.3.3 Multiple-Access Evolutionary Games

Interactions among users are dynamic, and users can update their rates and channel selection with respect to their payoffs and the known coupled constraints. Such a dynamic process can generally be modeled by an evolutionary process, a learning process, or a trial-and-error updating process. In classical game theory, the focus is on strategies that optimize payoffs to the players, whereas in evolutionary game theory, the focus is on strategies that will persist through time. In this subsection, we formulate evolutionary game dynamics based on the static game discussed in Sect. 3.3.1. We use generalized Smith dynamics for channel selection and G-function-based dynamics for rates. Combining them, we set up a framework of hybrid dynamics for the overall system.

The action of each user has two components $(\alpha_i, \mathbf{p}_i) \in \mathbb{R}_+ \times [0,1]^J$. We use $\mathbf{p}_i$ as strategies that determine the fitness of user i's rate α_i to receiver j. The rate selection evolves according to the channel selection strategy $\mathbf{P}$. We may view channel selection as an inner game that involves a process on a short time scale, but rate selection is an outer game that represents the dynamical link via fitness on a longer time scale [7, 8].

3.3.3.1 Learning the Weight Placed on Receiver

Let α be a fixed rate on the capacity region. We assume that user i occasionally tests the weights p_{ij} with alternative receivers, keeping the new strategy if and only if it leads to a strict increase in payoff. If the choice of receivers' weights of some users decreases the payoff or violates the constraints due to a strategy change by another user, then the user starts a random search for a new strategy, eventually settling on one with a probability that increases monotonically with its realized payoff. For the foregoing generating-function-based dynamics, the weight of switching from receiver j to receiver j' is given by

$$\eta^i_{jj'}(\alpha,\mathbf{P}) = \max(0, u_{ij'}(\alpha,\mathbf{P}) - u_{ij}(\alpha,\mathbf{P}))^{\theta},\ \theta \geq 1,$$

if the payoff obtained at receiver j' is greater than the payoff obtained at receiver j and the constraints are satisfied; otherwise, $\eta^i_{jj'}(p,\alpha) = 0$. The frequency of uses of each receiver is then seen as the selection strategy of receivers.

The expected change at each receiver is the difference between the incoming and the outgoing flows. The dynamics is also called *generalized Smith dynamics* [2] and is given by

$$\dot{p}_{ij}(t) = \sum_{j'\in\mathcal{J}} p_{ij'}(t)\eta^i_{j'j}(\alpha,\mathbf{P}(t)) - p_{ij}(t)\sum_{j'\in\mathcal{J}} \eta^i_{jj'}(\alpha,\mathbf{P}(t)). \tag{3.20}$$

Let $\chi_{ij}(\alpha,\mathbf{P}(t)) := \sum_{j'\in\mathcal{J}} p_{ij'}(t)\eta^i_{j'j}(\alpha,\mathbf{P}(t)) - p_{ij}(t)\sum_{j'\in\mathcal{J}} \eta^i_{jj'}(\alpha,\mathbf{P}(t))$. Hence, the dynamics can be rewritten as $\dot{p}_{ij} = \chi_{ij}(\alpha,\mathbf{P}(t))$. For $\theta = 1$ the dynamics is known as *Smith dynamics* and has been used to describe the evolution of road traffic congestion in which the fitness is determined by the strategies chosen by all drivers. It has also been studied in the context of resource selection in hybrid systems and a migration constraint problem in wireless networks in [2].

Proposition 3.7. *Any equilibrium of the game $\overline{\Xi}$ with predetermined rates is a rest point of the generalized Smith dynamics (3.20).*

Proof. The transition rate between receivers preserves the sign in the sense that, for every user, the incoming flow from receiver j' to j is positive if and only if the constraints are satisfied and the payoff to j exceeds the payoff to j'. Let α be a feasible point. If the right-hand side of (3.20) is nonzero for some splitting strategy $\mathbf{P}$, then

$$\begin{aligned} d &:= \sum_{j\in\mathcal{J}} \dot{p}_{ij}u_{ij}(\alpha,\mathbf{P}) = \sum_{j\in\mathcal{J}} \chi_{ij}u_{ij}(\alpha,\mathbf{P}) \\ &= \sum_{j,j'\in\mathcal{J}} p_{ij'}\left(u_{ij}(\alpha,\mathbf{P}) - u_{ij'}(\alpha,\mathbf{P})\right)\eta^i_{j'j} \\ &= \sum_{j,j'\in\mathcal{J}} p_{ij'}\max\left[0, \left(u_{ij}(\alpha,\mathbf{P}) - u_{ij'}(\alpha,\mathbf{P})\right)\right]\eta^i_{j'j}, \end{aligned}$$

which is strictly positive. Thus, if $(\alpha,\mathbf{P})$ is an NE, then $(\alpha,\mathbf{P})$ satisfy the constraints, and $p_{ij}=0$, or $\eta^{i}_{jj'}(\alpha,\mathbf{P})=0$. This implies that $(\alpha,\mathbf{P})$ satisfies also $\chi(\alpha,\mathbf{P}))=0$. □

The following proposition says that the equilibria are exactly the rest points of (3.20).

Proposition 3.8. *Any rest point of the dynamics (3.20) is an NE of the game* $\overline{\Xi}$.

The proof of Proposition 3.8 can be obtained using Theorem III in [2]. Since the probability of switching from receiver j to j' is proportional to $\eta^{i}_{jj'}$, which preserves the sign of payoff difference, we can use Theorems III in [2]. It follows that the dynamics generated by η satisfy the Nash stationarity property.

3.3.3.2 G-Function-Based Dynamics

We introduce here the generating fitness function (*G-function*)-based dynamics with projection onto the capacity region. The G-function approach has been successfully applied to nonlinear continuous games by Vincent and Brown [7, 8]. It is appropriate for our hybrid model because we can regard the channel selection as the variables in a fitness function. Users choose channel selection probabilities to try to increase the fitness of their rate choice. In our rate allocation game, to deal with constraints, we use projection onto capacity regions to maintain the feasibility of the trajectories. Starting from a point in the polytope C, users revise and update their strategies according to a rate proportional to the gradient and its payoff subject to the capacity constraints. Let G_{ij} be the fitness-generating function of user i at receiver j defined on $\mathbb{R}^N\times\mathbb{R}^{N\times J}$ satisfying

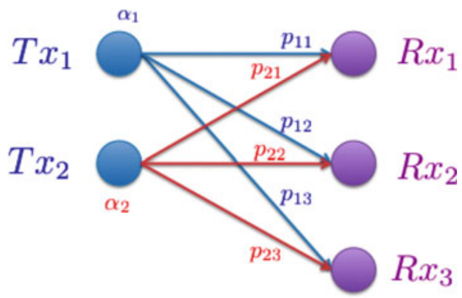
$$G_{ij}(v,\alpha,\mathbf{P})\Big|_{v=\mathbf{p}_i}=\left(C_{j,\mathcal{N}}-p_{ij}\beta_{ij}(t)-\sum_{i'\in\mathcal{N}\backslash\{i\}}p_{i'j}\beta_{i'j}(t)\right)$$

if the pair $(\alpha,\mathbf{P})$ is in the hybrid capacity region. Notice that the term $C_{j,\mathcal{N}}-\sum_{i'\neq i}p_{i'j}\beta_{i'j}(t)$ is the maximum rate of i using channel j at time t. Hence, the G-function-based dynamics is given by

$$\dot{\beta}_{ij}=-\bar{\mu}\left[p_{ij}\beta_{ij}-C_{j,\mathcal{N}}+\sum_{i'\neq i}p_{i'j}\beta_{i'j}\right]p_{ij}\beta_{ij}, \qquad (3.21)$$

with initial conditions $\beta_{ij}(0)\leq C_{j,\{i\}}$, where $\beta=[\beta_{ij}]$ is defined in Proposition 3.5, which is of the same dimension as α, and $\alpha_i(t)=\sum_{j\in\mathcal{J}}\beta_{ij}(t)$; $\bar{\mu}$ is an appropriate parameter chosen for the rate of convergence.

Fig. 3.4 Two users and three receivers



3.3.3.3 Hybrid Dynamics

We now combine the two evolutionary game dynamics described in the previous subsections. Variables α and $\mathbf{P}$ are both evolving in time. The overall dynamics are given by

$$\begin{cases} \dot{p}_{ij}(t) = \sum_{j'\in\mathcal{J}} p_{ij'}(t)\eta^{i}_{j'j}(\alpha(t),\mathbf{P}(t)) - p_{ij}(t)\sum_{j'\in\mathcal{J}} \eta^{i}_{jj'}(\alpha(t),\mathbf{P}(t)), \\ \dot{\beta}_{ij}(t) = -\bar{\mu}\left[p_{ij}(t)\beta_{ij}(t) - C_{j,\mathcal{N}} + \sum_{i'\neq i} p_{i'j}(t)\beta_{i'j}(t)\right] p_{ij}(t)\beta_{ij}(t), \\ \alpha_i(t) = \sum_{j\in\mathcal{J}} \beta_{ij}(t),\ \beta_{ij}(0) \leq C_{j,\{i\}}, \forall j \in \mathcal{J},\ i \in \mathcal{N}. \end{cases} \tag{3.22}$$

All the equilibria of the hybrid evolutionary rate control and channel selection are rest points of the preceding hybrid dynamics. The following result can be obtained directly from Proposition 3.7 and (3.21).

Proposition 3.9. *Let $(\beta^*,\mathbf{P}^*)$ be interior rest points of the hybrid dynamics, i.e., $\beta^*_{ij} > 0,\ p^*_{ij} > 0$ and $\chi(\alpha^*,\mathbf{P}) = 0$. Then for all j,*

$$\sum_{i=1}^{N} p^*_{ij}\beta^*_{ij} = C_{j,\mathcal{N}};\ \ \chi\left(\sum_{j=1}^{N}\beta^*_{ij},\mathbf{P}^*\right) = 0.$$

3.3.4 Numerical Examples

In this subsection, we illustrate the evolutionary dynamics in (3.21) and (3.22) by examining a two-user and three-receiver communication system as depicted in Fig. 3.4. Let $h_{i1} = 0.1, h_{i2} = 0.2, h_{i3} = 0.3$, for $i \in \{1,2\}$. Each transmission power P_i is set to 1 mW for all $i = 1,2$ and the noise level is set to $\sigma^2 = -20$ dBm.

In the first experiment, we assume that the rates of the users are predetermined to be $\alpha = [10,20]^{\mathrm{T}}$, the Smith dynamics in (3.21) yield in Figs. 3.5 and 3.6 the response of $\mathbf{p}_1$ and $\mathbf{p}_2$. It can be seen that the dynamics converge very fast within less than half a second.

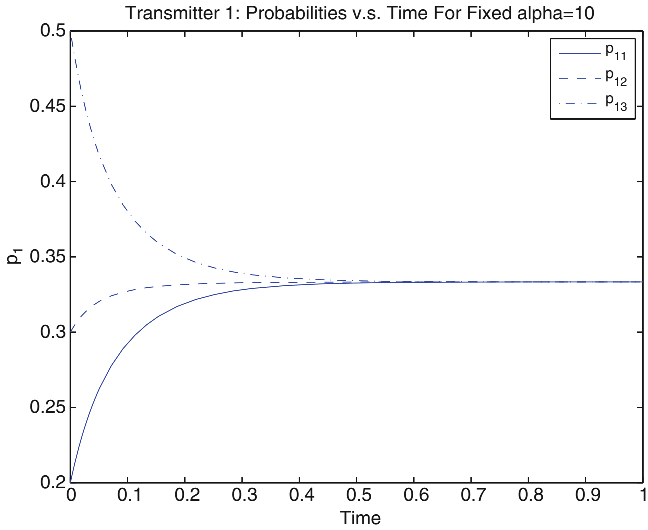


Fig. 3.5 Transmitter 1: probabilities vs. time for fixed $\alpha_1 = 10$

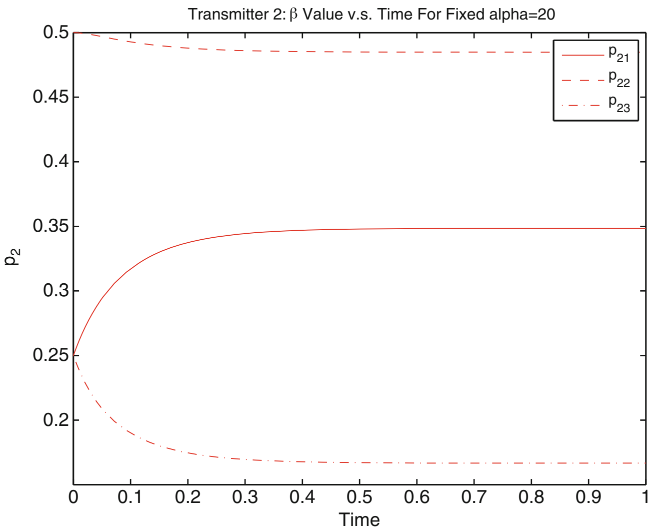


Fig. 3.6 Transmitter 2: probabilities vs. time for fixed $\alpha_2 = 20$

In the second experiment, we assume that the probability matrix $\mathbf{P}$ was optimally found by the users using (3.20). Figures 3.7 and 3.8 show that the β values converge to an equilibrium from which we can find the optimal value for α. Since these dynamics are much slower compared to Smith dynamics on $\mathbf{P}$, our assumption of knowledge of optimal $\mathbf{P}$ for a slowly varying α becomes valid.

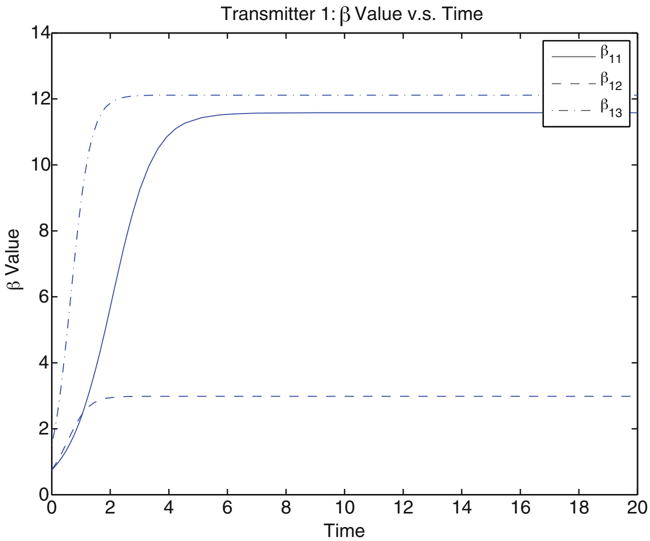


Fig. 3.7 Transmitter 1: β value vs. time

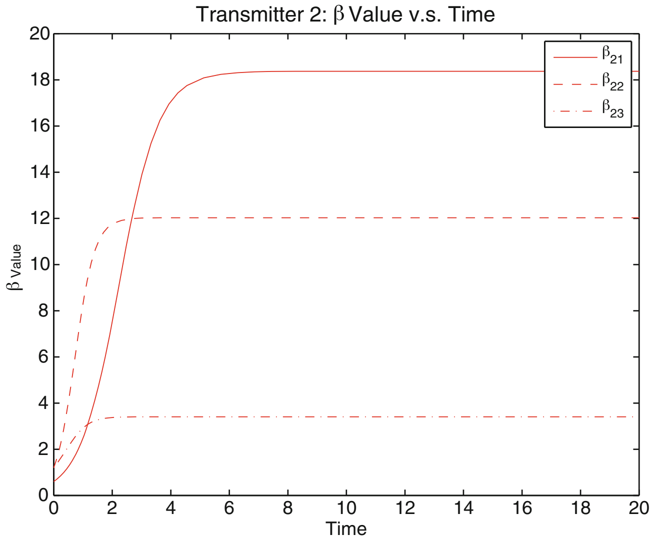


Fig. 3.8 Transmitter 1: β value vs. time

In the next experiment, we simulate the hybrid dynamics in (3.22). Let the probability p_{ij} of user i choosing transmitter j and the transmission rates be initialized as follows:

$$\mathbf{P}(0) = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}, \ \alpha(0) = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}.$$

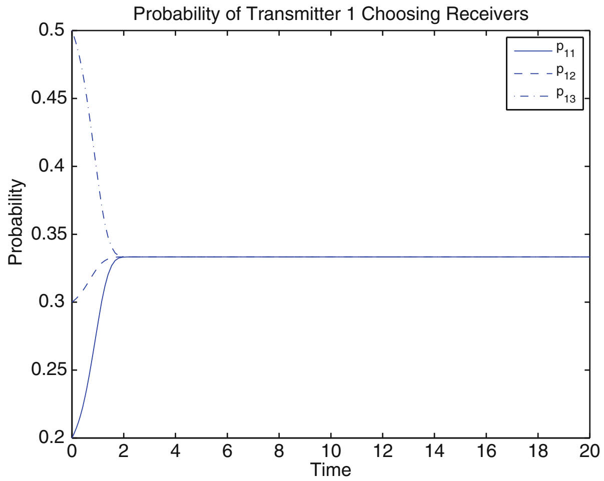


Fig. 3.9 Probability of Transmitter 1 choosing receivers

We let the parameter $\bar{\mu} = 0.9$. Figure 3.9 shows the evolution of the weights of user 1 on each of the receivers. The weights converge to $p_{1j} = 1/3$ for all j within 2 s, leading to an unbiased choice among receivers. In Fig. 3.10, we show the evolution of the weights of the second user on each receiver. At equilibrium, $\mathbf{p}_2 = [0.3484, 0.4847, 0.1669]^{\mathrm{T}}$. It appears that user two favors the second transmitter over the other ones. Since the utility u_{ij} is of the same form, the optimal response set K_i^* is naturally nonempty and contains all the receivers. As shown in Proposition 3.6, the probability of choosing a receiver at equilibrium is randomized among the three receivers and can be determined by the rates α chosen by the users.

The β-dynamics determines the evolution of α in (3.22). In Fig. 3.11, we see that the evolutionary dynamics yield $\alpha = [15.87, 23.19]^T$ at equilibrium. It is easy to verify that they satisfy the capacity constraints outlined in Sect. 3.2. It converges within 5 s and appears to be much slower than in Figs. 3.9 and 3.10. Hence, it can be seen that $\mathbf{P}$-dynamics may be seen as the inner-loop dynamics, whereas β-dynamics can be seen as an outer-loop evolutionary dynamics. They evolve on two different time scales. In addition, thanks to Proposition 3.8, finding the rest points for the preceding dynamics ensures that we will find the equilibrium.

3.4 Concluding Remarks

In this paper, we have studied an evolutionary multiple-access channel game with a continuum action space and coupled rate constraints. We showed that the game has a continuum of strong equilibria that are 100 % efficient in the rate optimization problem. We proposed the constrained Brown–von Neumann–Nash dynamics,

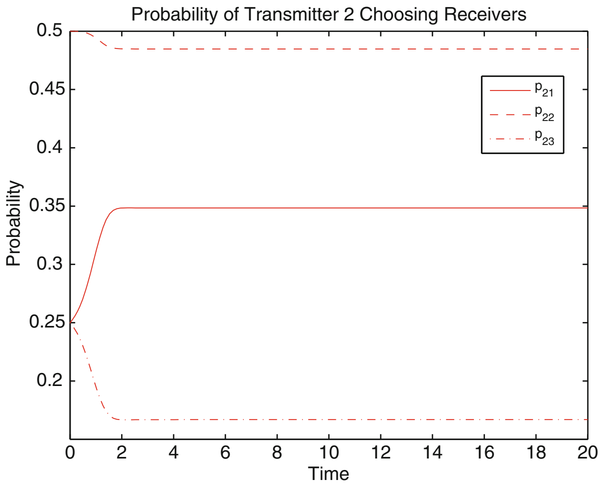


Fig. 3.10 Probability of Transmitter 2 choosing receivers

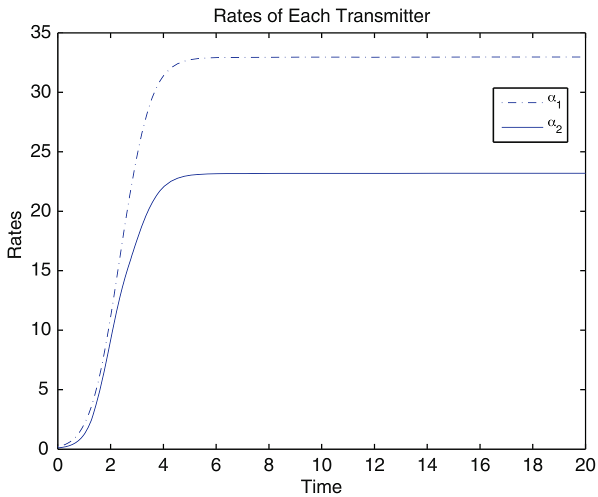


Fig. 3.11 Rates of each transmitter

Smith dynamics, and the replicator dynamics to study the stability of equilibria in the long run. In addition, we introduced a hybrid multiple-access game model and its corresponding evolutionary game-theoretic framework. We analyzed the NE for the static game and suggested a system of evolutionary-game-dynamics-based method to find it. It was found that the Smith dynamics for channel selections are a lot faster than the β-dynamics, and the combined dynamics yield a rest point that corresponds to the NE. An interesting extension that we leave for future research is to introduce a dynamic channel characteristic: the gains $h_{ij}(t)$ are *time-dependent random variables*. Another interesting question is to find an equilibrium structure in the case of multiple-access games with *nonconvex capacity regions*.

References

1. Tembine, H., Altman, E., ElAzouzi, R., Hayel, Y.: Evolutionary games in wireless networks. IEEE Trans. Syst. Man Cybern. B Cybern. **40**(3), 634–646 (2010)
2. Tembine, H., Altman, E., El-Azouzi, R., Sandholm, W.H.: Evolutionary game dynamics with migration for hybrid power control game in wireless communications. In: Proceedings of 47th IEEE Conference on Decision and Control (CDC), Cancun, Mexico, pp. 4479–4484 (2008)
3. Zhu, Q.: A Lagrangian approach to constrained potential games: theory and example. In: Proceedings of 47th IEEE Conference on Decision and Control (CDC), Cancun, Mexico, pp. 2420–2425 (2008)
4. Saraydar, C.U., Mandayam, N.B., Goodman, D.J.: Efficient power control via pricing in wireless data networks. IEEE Trans. Commun. **50**(2), 291–303 (2002)
5. Belmega, E.V., Lasaulce, S., Debbah, M., Jungers, M., Dumont, J.: Power allocation games in wireless networks of multi-antenna terminals. Springer Telecomm. Syst. J. **44**(5–6), 1018–4864 (2010)
6. Palomar, D.P., Cioffi, J.M., Lagunas, M.A.: Uniform power allocation in MIMO channels: a game theoretic approach. IEEE Trans. Inform. Theory **49**(7), 1707–1727 (2003)
7. Vincent, T.L., Vincent, T.L.S.: Evolution and control system design. IEEE Control Syst. Mag. **20**(5), 20–35 (2000)
8. Vincent, T.L., Brown, J.S.: Evolutionary Game Theory, Natural Selection, and Darwinian Dynamics. Cambridge University Press, Cambridge (2005)
9. Zhu, Q., Tembine, H., Başar, T.: A constrained evolutionary Gaussian multiple access channel game. In: Proceedings of the International Conference on Game Theory for Networks (GameNets), Istanbul, Turkey, 13–15 May 2009
10. Zhu, Q., Tembine, H., Başar, T.: Evolutionary games for hybrid additive white Gaussian noise multiple access control. In: Proceedings of GLOBECOM (2009)
11. Altman, E., El-Azouzi, R., Hayel, Y., Tembine, H.: Evolutionary power control games in wireless networks. In: NETWORKING 2008 Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet, pp. 930–942. Springer, New York (2008)
12. Andelman, N., Feldman, M., Mansour, Y.: Strong price of anarchy, Games and Economic Behavior, **65**(2), 289–317 (2009)
13. Anshelevich, E., Dasgupta, A., Kleinberg, J., Tardos, E., Wexler, T., Roughgarden, T.: The price of stability for network design with fair cost allocation. In: Proceedings of the FOCS, pp. 59–73 (2004)
14. Tembine, H., Altman, E., El-Azouzi, R., Hayel, Y.: Multiple access game in ad-hoc networks. In: Proceedings of 1st International Workshop on Game Theory in Communication Networks (GameComm) (2007)
15. Forges, F.: Can sunspots replace a mediator? J. Math. Econ. **17**, 347–368 (1988)
16. Forges, F.: Sunspot equilibrium as a game-theoretical solution concept. In: Barnett, W.A., Cornet, B., Aspremont, C., Gabszewicz, J.J., Mas-Colell, A. (eds.) Equilibrium Theory and Applications: Proceedings of the 6th International Symposium in Economic Theory and Econometrics, pp. 135–159. Cambridge University Press, Cambrdige (1991)
17. Forges, F., Peck, J.: Correlated equilibrium and sunspot equilibrium. Econ. Theory **5**, 33–50 (1995)

18. Aumann R. J.: Acceptable points in general cooperative n-person games, Contributions to the Theory of Games IV, Annals of Mathematics Study 40, Princeton University Press, 287–324 (1959)
19. Gajic, V., Rimoldi, B.: Game theoretic considerations for the Gaussian multiple access channel. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT) (2008)
20. Goodman, J.C.: A note on existence and uniqueness of equilibrium points for concave N-person games. Econometrica **48**(1), 251 (1980)
21. Hofbauer, J., Sigmund, K.: Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge (1998)
22. Ponstein, J.: Existence of equilibrium points in non-product spaces. SIAM J. Appl. Math. **14**(1), 181–190 (1966)
23. McGill, B.J., Brown, J.S.: Evolutionary game theory and adaptive dynamics of continuous traits. Annu. Rev. Ecol. Evol. Syst. **38**, 403–435 (2007)
24. Shaiju, A.J., Bernhard, P.: Evolutionarily robust strategies: two nontrivial examples and a theorem. In: Proceedings of 13-th International Symposium on Dynamic Games and Applications (ISDG) (2006)
25. Maynard Smith, J., Price, G.M.: The logic of animal conflict. Nature **246**, 15–18 (1973)
26. Rosen, J.B.: Existence and uniqueness of equilibrium points for concave N-person games. Econometrica **33**, 520–534 (1965)
27. Sandholm, W.H.: Population Games and Evolutionary Dynamics. MIT Press, Cambridge (2010)
28. Takashi, U.: Correlated equilibrium and concave games. Int. J. Game Theory **37**(1), 1–13 (2008)
29. Taylor, P.D., Jonker, L.: Evolutionarily stable strategies and game dynamics. Math. Biosci. **40**, 145–156 (1978)
30. Tembine, H., Altman, E., El-Azouzi, R., Hayel, Y.: Evolutionary games with random number of interacting players applied to access control. In: Proceedings of IEEE/ACM International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), March 2008.
31. Tembine, H., Altman, E., El-Azouzi, R.: Delayed evolutionary game dynamics applied to the medium access control. In: Proceedings of the 4th IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS) (2007)
32. Alpcan, T., Başar, T.: A hybrid noncooperative game model for wireless communications. In: Proceedings of 11th International Symposium on Dynamic Games and Applications, Tucson, AZ, December 2004
33. Alpcan, T., Başar, T., Srikant, R., Altman, E.: CDMA uplink power control as a noncooperative game. Wireless Networks **8**, 659–670 (2002)
34. Başar, T., Olsder, G.J.: Dynamic Noncooperative Game Theory, SIAM Series in Classics in Applied Mathematics, 2nd ed. SIAM, Philadelphia, (1999)
35. Dodis, Y., Halevi, S., Rabin, T.: A cryptographic solution to a game theoretic problem. In: Annual International Cryptology Conference, Santa Barbara CA, USA. Lecture Notes in Computer Science, vol. 1880, pp. 112–130 Springer-Verlag London, UK (2000)
36. Alpcan, T., Başar, T.: A hybrid system model for power control in multicell wireless data networks. Perform. Eval. **57**, 477–495 (2004)
37. Başar, T., Zhu, Q.: Prices of anarchy, information, and cooperation in differential games. J. Dynamic Games Appl. **1**(1), 50–73 (2011)
38. Roughgarden, T.: Selfish Routing and the Price of Anarchy. MIT Press, Cambridge (2005)

Chapter 4
Join Forces or Cheat: Evolutionary Analysis of a Consumer–Resource System

Andrei R. Akhmetzhanov, Frédéric Grognard, Ludovic Mailleret, and Pierre Bernhard

Abstract In this contribution we consider a seasonal consumer–resource system and focus on the evolution of consumer behavior. It is assumed that consumer and resource individuals live and interact during seasons of fixed lengths separated by winter periods. All individuals die at the end of the season and the size of the next generation is determined by the the consumer–resource interaction which took place during the season. Resource individuals are assumed to reproduce at a constant rate, while consumers have to trade-off between foraging for resources, which increases their reproductive abilities, or reproducing. Firstly, we assume that consumers cooperate in such a way that they maximize each consumer's individual fitness. Secondly, we consider the case where such a population is challenged by selfish mutants who do not cooperate. Finally we study the system dynamics over many seasons and show that mutants eventually replace the original cooperating population, but are finally as vulnerable as the initial cooperating consumers.

A.R. Akhmetzhanov (✉)
Department of Biology (Theoretical Biology Laboratory), McMaster University, Life Sciences Building, 1280 Main St. West, Hamilton, ON L8S 4K1, Canada

INRIA Sophia Antipolis – Méditerranée (Project BIOCORE), 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France
e-mail: akhmetzhanov@gmail.com

F. Grognard • P. Bernhard
INRIA Sophia Antipolis – Méditerranée (Project BIOCORE), 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France
e-mail: Frederic.Grognard@inria.fr; Pierre.Bernhard@inria.fr

L. Mailleret
UMR ISA, INRA (1355), University of Nice Sophia Antipolis, CNRS (7255), 400 route des Chappes, BP 167, 06903, Sophia Antipolis, France

INRIA Sophia Antipolis - Méditerrane (Project BIOCORE), 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis, France
e-mail: Ludovic.Mailleret@sophia.inra.fr

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_4,

Keywords Consumer–resource system • Mutant invasion • Hierarchical differential games

4.1 Introduction

Among the many ecosystems found on Earth, one can easily identify many examples of resource–consumer systems like e.g. plant–grazer, prey–predator or host–parasitoid systems known in biology [13]. Usually, individuals involved in such systems (bacteria, plants, insects, animals, *etc.*) have conflicting interests and models describing such interactions are based on principles of game theory [2, 7, 8, 16]. Hence, the investigation of such models is of interest to both game theoreticians and behavioral and evolutionary biologists.

One of the main topics of evolutionary theory is addressing whether individuals should behave rationally throughout their lifetime. Darwin's statement of the survival of the fittest indicates that evolution selects the best reproducers, so that the evolutionary process should result in selecting organisms which appear to behave rationally, even though they may know little about rationality. Evolutionary processes may thus result in organisms which actually maximize their number of descendants [19]; this is true in systems in which density dependence can be neglected, or in which the relation between the organisms and their environment is fairly simple [14]. Otherwise, such a rule may not apply and evolution is expected to yield a population which employs an evolutionarily stable strategy; such a strategy will not allow them to get the maximum possible number of descendants, but cannot be beaten by any strategy a deviant organism may choose to follow [10, 21]. In the following, since we will be concerned with populations in which some organisms may deviate from the others, we will use the terminology from Adaptive Dynamics [6] and designate by "mutants" the organisms adopting a strategy different from the one of the main population, which will be referred to as the resident population.

In this work we study the fate of mutants based on an example of a seasonal consumer–resource system with optimal consumers as introduced by [1] using a semi-discrete approach [9]. In such a system, consumer and resource individuals are active during seasons of fixed length T separated by winter periods. To give an idea of what such a system could represent, the resource population could be annual plants and the consumer population some univoltine phytophagous insect species. All consumers and resources die at the end of the season and the size of the next generation is determined by the number of offspring produced during the previous season (i.e. offspring are made of seeds or eggs which mature into active stages at the beginning of the season). We assume that consumers have to share their time between foraging for resources, which increases their reproductive abilities, or reproducing. The reproduction of the resource population is assumed to occur at a constant rate.

In nature several patterns of life-history can be singled out, but they frequently contain two main phases: *growth phase* and *reproduction phase*. The transition between these two phases is said to be strict when the consumers only feed at the beginning of their life and only reproduce at the end, or there could exist an *intermediate phase* between them where growth and reproduction occur simultaneously. Such types of behaviors are called *determinate* and *indeterminate growth patterns* respectively [17]. Time-sharing between laying eggs and feeding for the consumers will be modeled by the variable u: $u = 1$ means feeding, $u = 0$ on the other hand means reproducing. Intermediate values $u \in (0,1)$ describe a situation where, for some part of the time, the individual is feeding and, for the other part of the time, it is reproducing.

Firstly, we consider a population of consumers maximizing their common fitness, all consumers being individuals having the same goal function and *acting for the common good*; these will be the residents. We then suppose that a small fraction of the consumer population starts to behave differently from the main population, and accordingly will call them mutants. The aim of this paper is to investigate how mutants will behave in the environment shaped by the residents, and what consequences can be expected for multi-season consumer–resource systems.

4.2 Main Model

4.2.1 Previous Work

Let us first consider a system of two populations: resources and consumers without any mutant. The consumer population is modeled with two state variables: the average energy of one individual p and the number of consumers c present in the system, while the resource population is described solely by its density n. We suppose that both populations are structured in *mature* (adult insects/plants) and *immature* stages (eggs/seeds). During the season, mature consumers and resources interact and reproduce themselves. Between seasons (during winter periods) all mature individuals die and immature individuals become mature in the next season.

We suppose that no consumers have any energy ($p = 0$) at the beginning of the season. The efficiency of reproduction is assumed to be proportional to the value of p; it is thus intuitive that consumers should feed on the resource at the beginning and reproduce at the end once they have gathered enough energy. The consumers thus face a trade-off between investing their time in feeding ($u = 1$) or laying eggs ($u = 0$). According to [1], the within season dynamics are given by

$$\dot{p} = -\kappa p + \eta n u, \quad \dot{n} = -\delta c n u, \tag{4.1}$$

where we assume that neither population suffers from intrinsic mortality; κ, η and δ are constants. After rescaling the time and state variables, the constants κ and η can be eliminated and the system of Eq. (4.1) can be rewritten in the simpler form:

$$\dot{p} = -p + n u, \quad \dot{n} = -c n u, \tag{4.2}$$

where c is a rescaled parameter which is proportional to the number of consumers present in the system.

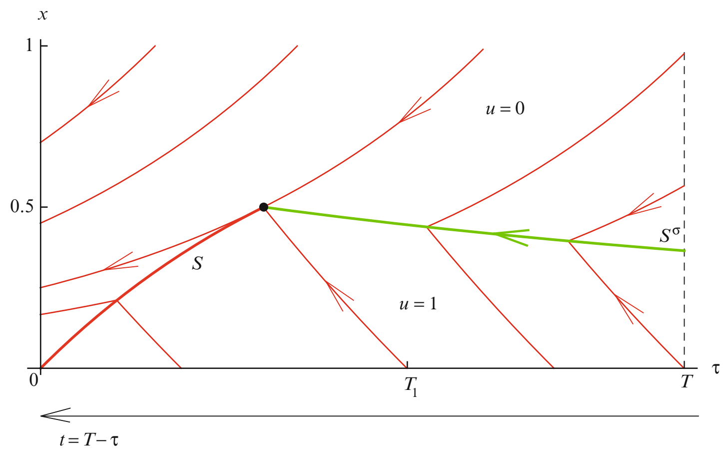


Fig. 4.1 Optimal collective behavior of the residents illustrated in the (τ, x) plane [see Eq. (4.4)] where τ is reverse time. On the figure, solutions are then initiated at $(T, p(0)/n(0))$ where T is the length of the season

The amount of immature offspring produced during the season depends on the sizes of the populations

$$J = \int_0^T \theta c(1-u(t))p(t)\,\mathrm{d}t, \quad J_n = \int_0^T \gamma n(t)\,\mathrm{d}t, \tag{4.3}$$

where θ and γ are constants.

We assume that consumers maximize the value J, their number of descendants, which is a classical measure of fitness. We see that this is an optimal control problem which can be solved using dynamic programming [3] or the Pontryagin maximum principle [18]. To compute the solution of this problem, the constants c, θ and γ can be omitted from (4.3), without loss of generality.

All the equations describing the problem are homogeneous of degree one in the state variables, which can be only positive. This is a particular case of Noether's theorem [15] in the calculus of variations about problems which are invariant under a group of transformations [4]. Hence, the dimension of the phase space of the optimal control problem (4.2)–(4.3) can be lowered by one unit by the introduction of a new variable $x = p/n$. In this case its dynamics can be written in the form:

$$\dot{x} = -x(1-cu) + u, \tag{4.4}$$

and the Bellman function $\tilde{U}(p,n,t) = \int_{T-t}^{\mathrm{T}} (1-u(s))p(s)\mathrm{d}s$ with the starting point at $(p(t), n(t)) = (p, n)$ can be expressed as $\tilde{U}(p,n,t) = nU(x,t)$.

The solution of the optimal control problem (4.2)–(4.3) has been obtained in [1] and the optimal behavioral pattern for $c = 1.5$ and $T = 2$ is shown in Fig. 4.1.

These solutions are not restricted to the case where consumers have no energy at the initial time. The region with $u = 1$ is separated from the region with $u = 0$ by a switching curve S and a singular arc S^σ such that

$$S\colon \quad x = 1 - \mathrm{e}^{-\tau} \tag{4.5}$$

$$S^\sigma\colon \quad \tau = -\log x + \frac{2}{xc} - \frac{4}{c}, \tag{4.6}$$

where $\tau = T - t$. They are shown in Fig. 4.1 by thick curves. Along the singular arc S^σ the consumer uses intermediate control $u = \hat{u}$:

$$\hat{u} = \frac{2x}{2 + xc}. \tag{4.7}$$

When $p(0) = 0$, one might identify a bang-bang control pattern for *short seasons* $T \leq T_1$ and a bang-singular-bang pattern for *long seasons* $T > T_1$. The value T_1 is computed as

$$T_1 = \frac{\log(c+1) + (c-2)\log 2}{c-1}, \tag{4.8}$$

so that it depends on the number of consumers present in the system.

The optimal value of the amount of offspring produced by an individual can be computed using this solution. In the following, we focus on the behavior of mutants appearing in a population of consumers adopting the type of behavior given in Fig. 4.1.

4.2.2 *Consumer–Mutant–Resource System*

Suppose that there is a subpopulation of consumers that deviate from the residents' behavior. Let us assume that these are selfish and maximize their own fitness, and not the fitness of the whole population, taking into account that the main resident population acts as if the mutants were kin (i.e. residents do not understand that mutants are selfish). This means that the residents adjust their strategy by changing the control whenever its level is intermediate. Such adjustment is possible only when some certain conditions are satisfied and mutant subpopulation is small enough (see Sect. 4.3.2).

Denote the proportion of mutants in the whole population of consumers by ε and the variables describing the state of the mutant and resident populations by symbols with subscripts "m" and "r" respectively. Then the number of mutants and residents will be $c_\mathrm{m} = \varepsilon c$ and $c_\mathrm{r} = (1-\varepsilon)c$ and the dynamics of the system can written as

$$\dot{p}_\mathrm{r} = -p_\mathrm{r} + nu_\mathrm{r}, \quad \dot{p}_\mathrm{m} = -p_\mathrm{m} + nu_\mathrm{m}, \quad \dot{n} = -nc\left[(1-\varepsilon)u_\mathrm{r} + \varepsilon u_\mathrm{m}\right], \tag{4.9}$$

similarly to (4.2). The variable $u_{\mathrm{m}} \in [0,1]$ defines the decision pattern of the mutants. The control $u_{\mathrm{r}} \in [0,1]$ is the decision pattern of the residents and defined by the solution of the optimal control problem (4.2)–(4.3).

The number of offspring in the next season is defined similarly to (4.3):

$$J_{\mathrm{r}} = \int_0^T \theta(1-u_{\mathrm{r}}(t))c_{\mathrm{r}}p_{\mathrm{r}}(t)\,\mathrm{d}t,\; J_{\mathrm{m}} = \int_0^T \theta(1-u_{\mathrm{m}}(t))c_{\mathrm{m}}p_{\mathrm{m}}(t)\,\mathrm{d}t,\; J_n = \int_0^T \gamma n(t)\,\mathrm{d}t, \tag{4.10}$$

where the mutant chooses its control u_{m} striving to maximize its fitness J_{m}.

We can see that the problem under consideration is described in terms of a two-step optimal control problem (or a hierarchical differential game): in the first step we define the optimal behavior of the residents (see Sect. 4.2.1), in the second step we identify the optimal response of the mutants to this strategy.

4.3 Optimal Free-Riding

Since θ and γ are constants, they can be omitted from the description of the optimization problem $J_{\mathrm{m}} \to \max\limits_{u_{\mathrm{m}}}$. In this case the functional $J_{\mathrm{m}}/(\theta c_{\mathrm{m}})$ can be taken instead of the functional J_{m}.

Let one introduce the Bellman function $\tilde{U}_{\mathrm{m}}$ for the mutant population. It satisfies the Hamilton–Jacobi–Bellman (HJB) equation

$$\frac{\partial \tilde{U}_{\mathrm{m}}}{\partial t} + \max_{u_{\mathrm{m}}}\Bigg[\frac{\partial \tilde{U}_{\mathrm{m}}}{\partial p_{\mathrm{r}}}(-p_{\mathrm{r}}+nu_{\mathrm{r}}) + \frac{\partial \tilde{U}_{\mathrm{m}}}{\partial p_{\mathrm{m}}}(-p_{\mathrm{m}}+nu_{\mathrm{m}}) - \frac{\partial \tilde{U}_{\mathrm{m}}}{\partial n} nc((1-\varepsilon)u_{\mathrm{r}}+\varepsilon u_{\mathrm{m}}) + p_{\mathrm{m}}(1-u_{\mathrm{m}})\Bigg] = 0. \tag{4.11}$$

Introducing new variables $x_{\mathrm{r}} = p_{\mathrm{r}}/n$ and $x_{\mathrm{m}} = p_{\mathrm{m}}/n$ and using a transformation of the Bellman function of the form $\tilde{U}_{\mathrm{m}}(p_{\mathrm{r}}, p_{\mathrm{m}}, n, t) = nU_{\mathrm{m}}(x_{\mathrm{r}}, x_{\mathrm{m}}, t)$, we can reduce the dimension of the problem by one using Noether's theorem [15]. The modified HJB-equation (4.11) takes the following form

$$\begin{aligned}\mathcal{H} \doteq -\nu + \max_{u_{\mathrm{m}}}\Big\{ &\lambda_{\mathrm{r}}[-x_{\mathrm{r}}(1-c((1-\varepsilon)u_{\mathrm{r}}+\varepsilon u_{\mathrm{m}}))+u_{\mathrm{r}}] \\ &+\lambda_{\mathrm{m}}[-x_{\mathrm{m}}(1-c((1-\varepsilon)u_{\mathrm{r}}+\varepsilon u_{\mathrm{m}}))+u_{\mathrm{m}}] \\ &-U_{\mathrm{m}}c((1-\varepsilon)u_{\mathrm{r}}+\varepsilon u_{\mathrm{m}}) + x_{\mathrm{m}}(1-u_{\mathrm{m}})\Big\} = 0,\end{aligned} \tag{4.12}$$

where the components of the gradient of the Bellman function are denoted by $\partial U_{\mathrm{m}}/\partial x_{\mathrm{r}} = \lambda_{\mathrm{r}}$, $\partial U_{\mathrm{m}}/\partial x_{\mathrm{m}} = \lambda_{\mathrm{m}}$ and $\partial U_{\mathrm{m}}/\partial \tau = \nu$, variable τ denotes backward time, $\tau = T - t$. The optimal control can be defined as $u_{\mathrm{m}} = \mathrm{Heav}(\mathcal{A}_{\mathrm{m}})$, where

$\mathcal{A}_\mathrm{m} = \partial\mathcal{H}/\partial u_\mathrm{m} = \lambda_\mathrm{r} x_\mathrm{r} \varepsilon c + \lambda_\mathrm{m}(1 + x_\mathrm{m}\varepsilon c) - U_\mathrm{m}\varepsilon c - x_\mathrm{m}$ and $\mathrm{Heav}(\cdot)$ is *a unit step function* whose value is *zero* for negative argument and *one* for positive argument.

One of the efficient ways to solve the HJB-equation is to use the method of characteristics (see e.g. [12]). The system of characteristics for Eq. (4.12) is

$$x_\mathrm{r}' = x_\mathrm{r}(1 - c((1-\varepsilon)u_\mathrm{r} + \varepsilon u_\mathrm{m})) - u_\mathrm{r},\ x_\mathrm{m}' = x_\mathrm{m}(1 - c((1-\varepsilon)u_\mathrm{r} + \varepsilon u_\mathrm{m})) - u_\mathrm{m},$$
$$\lambda_\mathrm{r}' = -\lambda_\mathrm{r},\ \lambda_\mathrm{m}' = -\lambda_\mathrm{m} + 1 - u_\mathrm{m},\ U_\mathrm{m}' = -U_\mathrm{m} c((1-\varepsilon)u_\mathrm{r} + \varepsilon u_\mathrm{m}) + x_\mathrm{m}(1 - u_\mathrm{m}), \tag{4.13}$$

where the prime denotes differentiation with respect to backward time τ. The terminal condition $U_\mathrm{m}(x_\mathrm{r}, x_\mathrm{m}, T) = 0$ gives $\lambda_\mathrm{r}(T) = \lambda_\mathrm{m}(T) = 0$. Thus $\mathcal{A}_\mathrm{m}(T) < 0$ and $u_\mathrm{m}(T) = 0$ (mutants should reproduce at the very end of their life).

4.3.1 First Steps

If we emit the characteristic field from the terminal surface $t = T$ with $u_\mathrm{r} = u_\mathrm{m} = 0$, then

$$x_\mathrm{r}' = x_\mathrm{r}, \quad x_\mathrm{m}' = x_\mathrm{m}, \quad \lambda_\mathrm{r}' = -\lambda_\mathrm{r}, \quad \lambda_\mathrm{m}' = -\lambda_\mathrm{m} + 1, \quad U_\mathrm{m}' = x_\mathrm{m},$$
$$\lambda_\mathrm{r}(T) = \lambda_\mathrm{m}(T) = 0, \quad U_\mathrm{m}(T) = 0.$$

We get the following equations for state and conjugate variables and for the Bellman function: $x_\mathrm{r} = x_\mathrm{r}(T)\,\mathrm{e}^{\tau}$, $x_\mathrm{m} = x_\mathrm{m}(T)\,\mathrm{e}^{\tau}$, $\lambda_\mathrm{r} = 0$, $\lambda_\mathrm{m} = 1 - \mathrm{e}^{-\tau}$, $U_\mathrm{m} = x_\mathrm{m}(1 - \mathrm{e}^{-\tau})$.

From this solution we can see that there could exist a switching surface S_m:

$$S_\mathrm{m}: \quad x_\mathrm{m} = 1 - \mathrm{e}^{-(T-t)}, \tag{4.14}$$

such that $\mathcal{A}_\mathrm{m} = 0$ on it, where the mutant changes its control. Equation (4.14) is similar to (4.5). However, we should take into account the fact that there is also a hypersurface S_r, where the resident changes its control from $u_\mathrm{r} = 0$ to $u_\mathrm{r} = 1$ independently of the decision of the mutant. Hence it is important to define which surface, S_r or S_m the characteristic intersects first, see Fig. 4.2. Suppose that this is the surface S_r. Since the control u_r changes its value on S_r, the HJB-equation (4.12) also changes and, as a consequence, the conjugate variables ν, λ_r and λ_m could possibly be discontinuous. Let us denote the incoming characteristic field (in backward time) by "$-$" and the outcoming field by "$+$". Consider a point of intersection of the characteristic and the surface S_r with coordinates $(x_{r_1}, x_{m_1}, \tau_1)$. Thus $x_{r_1} = 1 - \mathrm{e}^{-\tau_1}$ and the normal vector ϑ to the switching surface is written in the form

$$\vartheta = \nabla S_\mathrm{r} = (\partial S_\mathrm{r}/\partial x_\mathrm{r}, \partial S_\mathrm{r}/\partial x_\mathrm{m}, \partial S_\mathrm{r}/\partial \tau)^\mathrm{T} = (-1, 0, 1 - x_{r_1})^\mathrm{T}.$$

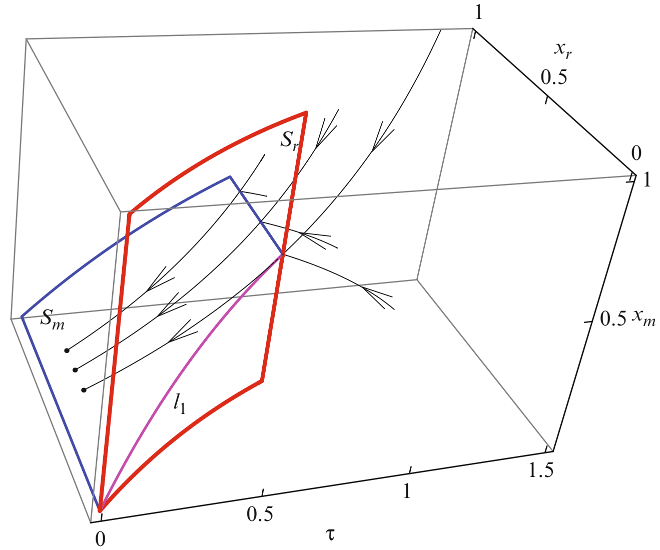


Fig. 4.2 Some family of optimal trajectories emanating from the terminal surface

From the incoming field we have the following information about the co-state $\lambda_{\rm r}^- = 0$, $\lambda_{\rm m}^- = x_{r_1}$, $\nu^- = x_{m_1}(1-x_{r_1})$. Since the Bellman function is continuous on the surface $S_{\rm r}$, we have: $U_{\rm m}^+ = U_{\rm m}^- = U_{\rm m} = x_{m_1}x_{r_1}$. The gradient $\nabla U_{\rm m}$ has a jump in the direction of the normal vector ϑ: $\nabla U_{\rm m}^+ = \nabla U_{\rm m}^- + k\vartheta$. Here k is an unknown scalar. Thus

$$\lambda_{\rm r}^+ = -k, \quad \lambda_{\rm m}^+ = x_{r_1}, \quad \nu^+ = x_{m_1}(1-x_{r_1}) + k(1-x_{r_1}). \tag{4.15}$$

If we suppose that the control of the mutant is the same, $u_{\rm m}^+ = 0$ (in this case $\mathcal{A}_{\rm m}^+$ should be negative), the HJB-equation (4.12) has the form

$$-\nu^+ + \lambda^+[-x_{r_1}(1-(1-\varepsilon)c)+1] - \lambda_{\rm m}^+ x_{m_1}(1-(1-\varepsilon)c) - (1-\varepsilon)cU_{\rm m} + x_{m_1} = 0. \tag{4.16}$$

Substituting the values from (4.15) into (4.16) we get: $k[-2(1-x_{r_1}) - x_{r_1}(1-\varepsilon)c] = 0$, which leads to $k = 0$ and, actually, there is no jump in the conjugate variables. They keep the same values as in (4.15) and $\mathcal{A}_{\rm m}^+ = \mathcal{A}_{\rm m}^-$.

Conversely, the mutant may react to the decision of the resident and also change its control on $S_{\rm r}$ from $u_{\rm m}^- = 0$ to $u_{\rm m}^+ = 1$. This is fulfilled if $\mathcal{A}_{\rm m}^+ > 0$. Substitution of the values ν^+, $\lambda_{\rm r}^+$ and $\lambda_{\rm m}^+$ from (4.15) to the HJB-equation (4.12) gives $k = (x_{r_1} - x_{m_1})/(x_{r_1}c + (1-x_{r_1}))$ and

$$\mathcal{A}_{\rm m}^+ = \lambda_{\rm r}^+ x_{r_1}\varepsilon c + \lambda_{\rm m}^+(x_{m_1}\varepsilon c + 1) - \varepsilon c U_{\rm m} - x_{m_1} = (x_{r_1} - x_{m_1})\frac{(1-\varepsilon)x_{r_1}c + (1-x_{r_1})}{x_{r_1}c + (1-x_{r_1})},$$

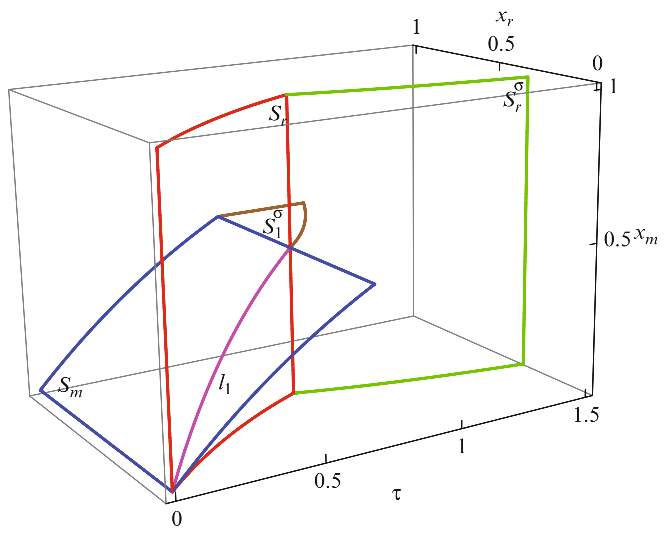


Fig. 4.3 Construction of the singular arc S_1^σ

which is positive when $x_{r_1} > x_{m_1}$. In Fig. 4.2 this corresponds to the points of the surface S_r which are below the line l_1: $x_\mathrm{r} = x_\mathrm{m} = 1 - \mathrm{e}^{-\tau}$. For the optimal trajectories which go through such points: $u_\mathrm{r}(\tau_1 + \delta) = u_\mathrm{m}(\tau_1 + \delta) = 1$, where δ is arbitrarily small. One can show that there will be no more switches of the control. However, if we consider a trajectory going from a point above l_1, then $u_\mathrm{r}(\tau_1 + \delta) = 1$ and $u_\mathrm{m}(\tau_1 + \delta) = 0$; a switch of the control u_m from zero to one then takes place later (in backward time). After that, there will be no more switches.

Now consider a trajectory emitted from the terminal surface which first intersects the surface S_m rather than the surface S_r. In this case the situation depicted in Fig. 4.3 takes place: one might expect the appearance of a singular arc S_1^σ there. The following are necessary conditions for its existence

$$\mathcal{H} = 0 = \mathcal{H}_0 + \mathcal{A}_\mathrm{m} u_\mathrm{m}, \quad \mathcal{H}_0 = -\nu - \lambda x_\mathrm{r} - \lambda_\mathrm{m} x_\mathrm{m} + x_\mathrm{m}, \tag{4.17}$$

$$\mathcal{A}_\mathrm{m} = 0 = \lambda_\mathrm{r} x_\mathrm{r} \varepsilon c + \lambda_\mathrm{m}(x_\mathrm{m} \varepsilon c + 1) - \varepsilon c U_\mathrm{m} - x_\mathrm{m}, \tag{4.18}$$

$$\mathcal{A}'_\mathrm{m} = \{\mathcal{A}_\mathrm{m} \mathcal{H}_0\} = 0 \doteq \mathcal{A}_{m1}, \tag{4.19}$$

where the curly brackets denote the Poisson (Jacobi) brackets. If ξ is a vector of state variables and ψ is a vector of conjugate ones (in our case $\xi = (x_\mathrm{r}, x_\mathrm{m}, \tau)$ and $\psi = (\lambda_\mathrm{r}, \lambda_\mathrm{m}, \nu)$), then the Poisson brackets of two functions $F = F(\xi, \psi, U_\mathrm{m})$ and $G = G(\xi, \psi, U_\mathrm{m})$ are given by the formula: $\{F\,G\} = \langle F_\xi + \psi F_{U_\mathrm{m}}, G_\psi \rangle - \langle F_\psi, G_\xi + \psi G_{U_\mathrm{m}} \rangle$. Here $\langle \cdot, \cdot \rangle$ denotes the scalar product and e.g. $F_\psi = \partial F / \partial \psi$.

After some algebra, (4.19) takes the form

$$\mathcal{A}_{m1} = \nu \varepsilon c + x_\mathrm{m} + \lambda_\mathrm{r} x_\mathrm{r} \varepsilon c - (x_\mathrm{m} + 1)(1 - \lambda_\mathrm{r}) = 0 \tag{4.20}$$

We can derive the variable ν from (4.17) and substitute it into (4.20). We get $\mathcal{A}_{m1} = x_{\rm m} - 1 + \lambda_{\rm m} = 0$. This leads to $\lambda_{\rm m} = 1 - x_{\rm m}$ and

$$\lambda_{\rm r} = \frac{x_{\rm m} + \varepsilon U_{\rm m} + (1 - x_{\rm m})(x_{\rm m}\varepsilon c + 1)}{x_{\rm r}\varepsilon c},$$

which is obtained from (4.18).

To derive the singular control $u_{\rm m} = \tilde{u}_{\rm m} \in (0,1)$ along the singular arc, one should write the second derivative: $\mathcal{A}''_{\rm m} = 0 = \{\{\mathcal{A}_{\rm m}\mathcal{H}\}\mathcal{H}\} = \{\{\mathcal{A}_{\rm m}\mathcal{H}_0\}\mathcal{H}_0\} + \{\{\mathcal{A}_{\rm m}\mathcal{H}_0\}\mathcal{A}_{\rm m}\}\tilde{u}_{\rm m}$. Thus

$$\tilde{u}_{\rm m} = \frac{\{\{\mathcal{A}_{\rm m}\mathcal{H}_0\}\mathcal{H}_0\}}{\{\mathcal{A}_{\rm m}\{\mathcal{A}_{\rm m}\mathcal{H}_0\}\}} = \frac{2x_{\rm m}}{2 + x_{\rm m}\varepsilon c}, \tag{4.21}$$

which has the same form as (4.7).

The equation for the singular arc S_1^σ can be obtained from the system of dynamic equations (4.13) by substituting $u_{\rm r} = 0$ and $u_{\rm m} = \tilde{u}_{\rm m}$ from (4.21):

$$x'_{\rm m} = -\frac{x_{\rm m}^2\varepsilon c}{2 + x_{\rm m}\varepsilon c}, \quad x_{\rm m}(\tau = \log 2) = 1/2\,.$$

Finally, we have the analogous expression to (4.6):

$$S_1^\sigma: \quad T - t = -\log x_{\rm m} + \frac{2}{x_{\rm m}\varepsilon c} - \frac{4}{\varepsilon c} \tag{4.22}$$

for $\varepsilon \neq 0$. If $\varepsilon = 0$, the surface $S_{\rm m}$ is a hyperplane $x_{\rm m} = 1/2$.

After these steps we have the structure of the solution shown in Fig. 4.3.

4.3.2 Optimal Motion Along the Surface $S_{\rm r}^\sigma$

According to the computations done in Sect. 4.2.1, resident consumers must adopt a behavior $u_{\rm r}$ which keeps the surface $S_{\rm r}^\sigma$ invariant (see Fig. 4.3). In a mutant-free population, this is done by playing the singular control (4.7), but if mutants are present in the population, the dynamics of the system are modified and the mutant-free singular control (4.7) does not make $S_{\rm r}^\sigma$ invariant any more. However, residents may still make $S_{\rm r}^\sigma$ invariant by adopting a different behavior, denoted $\hat{u}_{\rm r}$, as long as the mutants' influence, i.e. ε, is not too large. To compute $\hat{u}_{\rm r}$, we notice that it should make $x_{\rm r}$ follow the dynamics depicted in Fig. 4.1, i.e. $\dot{x}_{\rm r} = -x_{\rm r}(1 - cu_{\rm r}) + u_{\rm r}$ with $u_{\rm r} = \hat{u}$ defined in Eq. (4.7). We get that $\hat{u}_{\rm r}$ should be computed from:

$$x'_{\rm r} = -\frac{x_{\rm r}^2 c}{2 + x_{\rm r}c} = x_{\rm r}(1 - c((1-\varepsilon)\hat{u}_{\rm r} + \varepsilon u_{\rm m})) - \hat{u}_{\rm r}\,,$$

so that

$$\hat{u}_{\mathrm{r}} = \frac{2x_{\mathrm{r}}(1+x_{\mathrm{r}}c)}{(1+(1-\varepsilon)x_{\mathrm{r}}c)(2+x_{\mathrm{r}}c)} - \frac{x_{\mathrm{r}}\varepsilon c u_{\mathrm{m}}}{1+(1-\varepsilon)x_{\mathrm{r}}c}. \tag{4.23}$$

Thus, the residents will be able to keep S_{r}^{σ} invariant provided $\hat{u}_{\mathrm{r}} \in [0,1]$ for all points belonging to S_{r}^{σ} and for all possible values of $u_{\mathrm{m}} \in [0,1]$.

To identify for which parameters of the model this is possible, we may notice that $\hat{u}_{\mathrm{r}}$ is a linear function of u_{m} and decreasing. Moreover,

$$\hat{u}_{\mathrm{r}}(u_{\mathrm{m}}=0) = \frac{2x_{\mathrm{r}}(1+x_{\mathrm{r}}c)}{(1+(1-\varepsilon)x_{\mathrm{r}}c)(2+x_{\mathrm{r}}c)} \leq 2x_{\mathrm{r}}\frac{1+x_{\mathrm{r}}c}{2+x_{\mathrm{r}}c} \leq 1\,,$$

since $x_{\mathrm{r}} \leq 1/2$. Conversely, when $\hat{u}_{\mathrm{m}} = 1$, $\hat{u}_{\mathrm{r}} = \frac{2x_{\mathrm{r}}}{2+x_{\mathrm{r}}c}\frac{1+x_{\mathrm{r}}c-\varepsilon c-\varepsilon c x_{\mathrm{r}}c/2}{1+(1-\varepsilon)x_{\mathrm{r}}c}$. If this value is larger than 0 for any x_{r} belonging to S_{r}^{σ}, invariance of S_{r}^{σ} is ensured. A condition for this to occur is

$$\varepsilon < 1/c. \tag{4.24}$$

It is interesting to notice that $\hat{u}_{\mathrm{r}}(u_{\mathrm{m}}=0)$ is larger than the original $\hat{u}$ in (4.7), since the residents must compensate for the non-eating mutants. Conversely, when $u_{\mathrm{m}}=1$, $\hat{u}_{\mathrm{r}} < \hat{u}$. The tipping point takes place when $u_{\mathrm{m}} = \hat{u}$, which ensures $\hat{u}_{\mathrm{r}} = u_{\mathrm{m}}$; mutants behaving like the original residents allow the residents to behave as such.

In this paper we consider only the values of ε satisfying (4.24), i.e. such that the residents are able to adopt their optimal behavior, in spite of the presence of mutants. Otherwise, the influence of the mutants on the system may be too large, and the residents would not have the possibility to stick to their fitness maximization program.

The control $\hat{u}_{\mathrm{r}} = \hat{u}_{\mathrm{r}}(x_{\mathrm{r}}, x_{\mathrm{m}}, \tau, u_{\mathrm{m}})$ is defined in feedback form, i.e. it depends on the time and on the state of the system. The corresponding Hamiltonian (4.12) needs to be modified to

$$\hat{\mathcal{H}} = \mathcal{H}(x_{\mathrm{r}}, x_{\mathrm{m}}, U_{\mathrm{m}}, \lambda_{\mathrm{r}}, \lambda_{\mathrm{m}}, \nu, \hat{u}_{\mathrm{r}}(x_{\mathrm{r}}, x_{\mathrm{m}}, \tau, u_{\mathrm{m}}), u_{\mathrm{m}}), \tag{4.25}$$

so that the coefficient multiplying the control u_{m} becomes

$$\hat{\mathcal{A}}_{\mathrm{m}} = \frac{\partial \hat{\mathcal{H}}}{\partial u_{\mathrm{m}}} = \frac{\lambda_{\mathrm{m}}(1+x_{\mathrm{r}}(1-\varepsilon)c + x_{\mathrm{m}}\varepsilon c) - \varepsilon c U_{\mathrm{m}}}{1+(1-\varepsilon)x_{\mathrm{r}}c} - x_{\mathrm{m}}\,. \tag{4.26}$$

This expression allows us to compute the optimal behavior of the mutants on the surface S_{r}^{σ}, but the calculations are quite complicated. To make things simpler, let us first consider the particular case of vanishingly small values of ε and study the optimal behavioral pattern.

4.3.3 Particular Case of a Vanishingly Small Population of Mutants

4.3.3.1 On the Singular Surface S_r^σ

If $\varepsilon \cong 0$, the mutants' influence on the system is negligible and, to make S_r^σ invariant, the resident should apply the mutant-free singular behavior computed in (4.7): $\hat{u}_\mathrm{r} = 2x_\mathrm{r}/(2+x_\mathrm{r}c)$. In addition, Eqs. (4.25) and (4.26) take the following form

$$\hat{\mathcal{H}} = -\nu + \frac{\lambda_\mathrm{r} x_\mathrm{r}^2 c}{2+x_\mathrm{r}c} + \lambda_\mathrm{m}\left(-x_\mathrm{m}\frac{2-x_\mathrm{r}c}{2+x_\mathrm{r}c} + u_\mathrm{m}\right) - U_\mathrm{m}\frac{2x_\mathrm{r}c}{2+x_\mathrm{r}c} + x_\mathrm{m}(1-u_\mathrm{m}) \quad (4.27)$$

$$\hat{\mathcal{A}}_\mathrm{m} = \lambda_\mathrm{m} - x_\mathrm{m}\,. \quad (4.28)$$

If the trajectory originates (in backward time) from some point belonging to S_r^σ such that $x_\mathrm{m}^\sigma \doteq x_\mathrm{m}(\tau = \log 2) > 1/2$, then $u_\mathrm{m}(\tau = \log 2) = 0$ and the system of characteristics for the Hamiltonian (4.27) is

$$x_\mathrm{r}' = -\frac{x_\mathrm{r}^2 c}{2+x_\mathrm{r}c},\ x_\mathrm{m}' = x_\mathrm{m}\frac{2-x_\mathrm{r}c}{2+x_\mathrm{r}c},\ \lambda_\mathrm{m}' = -\lambda_\mathrm{m} + 1\,,\ U_\mathrm{m}' = -U_\mathrm{m}\frac{2x_\mathrm{r}c}{2+x_\mathrm{r}c} + x_\mathrm{m} \quad (4.29)$$

with boundary conditions: $\tau = \log 2$, $x_\mathrm{r} = 1/2$, $x_\mathrm{m} = x_\mathrm{m}^\sigma$, $\lambda_\mathrm{m} = 1/2$, $U_\mathrm{m} = x_\mathrm{m}^2/2$. Thus $\lambda_\mathrm{m} = 1 - \mathrm{e}^{-\tau}$ and there exists a switching curve $\hat{S}$, which is defined as: $x_\mathrm{m} = 1 - \mathrm{e}^{-\tau}$ in addition to $\tau = -\log x_\mathrm{r} + 2/(x_\mathrm{r}c) - 4/c$. Thus $\hat{S} = S_\mathrm{m} \cap S_\mathrm{r}^\sigma$.

The switching curve $\hat{S}$ ends at the point with coordinates $(x_{r_2}, x_{m_2}, \tau_2)$ where the characteristics become tangent to it and the singular arc $\hat{S}^\sigma$ appears (see Fig. 4.4). Before determining the coordinates of this point, let us define the singular arc, denoted $\hat{S}^\sigma$. From (4.27)–(4.28) we get

$$\nu = \frac{\lambda_\mathrm{r} x_\mathrm{r}^2 c}{2+x_\mathrm{r}c} - \lambda_\mathrm{m} x_\mathrm{m}\frac{2-x_\mathrm{r}c}{2+x_\mathrm{r}c} - U_\mathrm{m}\frac{2x_\mathrm{r}c}{2+x_\mathrm{r}c} + x_\mathrm{m}, \quad \lambda_\mathrm{m} = x_\mathrm{m} \quad (4.30)$$

along the singular arc. Substitution of (4.30) into equation $\hat{\mathcal{A}}_\mathrm{m}' = 0$ gives $x_\mathrm{m} = (2+x_\mathrm{r}c)/4$.

In addition, the intermediate control $\hat{u}_\mathrm{m}$ can be derived from $\hat{\mathcal{A}}_\mathrm{m}'' = 0$ and is equal to

$$\hat{u}_\mathrm{m} = \frac{1}{2+x_\mathrm{r}c}\,,$$

which is positive and belongs to the interval between zero and one.

We see that the coordinates x_{r_2}, x_{m_2} and τ_2 can be defined by the following equations

$$x_{m_2} = \frac{2+x_{r_2}c}{4} = 1 - \mathrm{e}^{-\tau_2}, \quad \tau_2 = -\log x_{r_2} + \frac{2}{x_{r_2}c} - \frac{4}{c}\,,$$

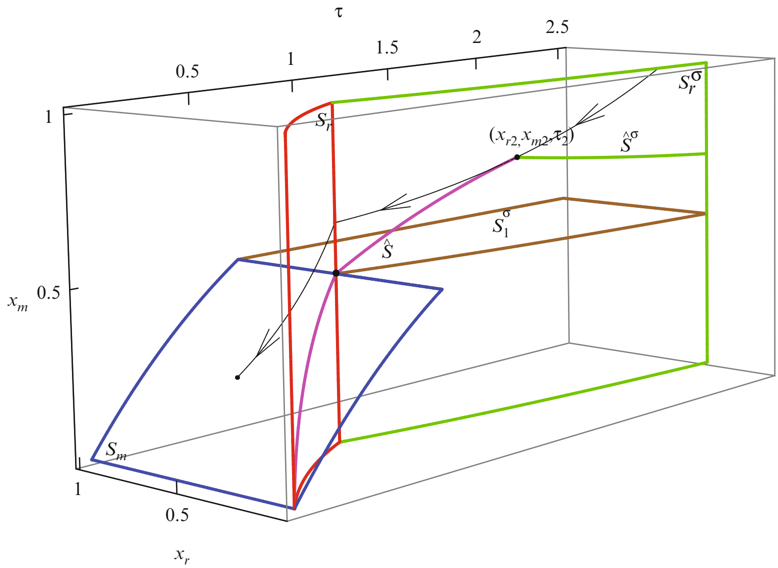


Fig. 4.4 Optimal behavior on the surface S_r^σ

which comes from the fact that the point $(x_{r_2}, x_{m_2}, \tau_2)$ belongs to $\hat{S}^\sigma$ and is located on the intersection of the curves $\hat{S}^\sigma$ and $\hat{S}$. This result is illustrated in Fig. 4.4.

4.3.3.2 Outside the Singular Surface S_r^σ

If the state is outside the surface S_r^σ, things are a little easier since at least the behavior of the residents, u_r, is constant and equal to 0 or 1, depending on the respective value of τ and x_r.

We can actually show that the surface S_1^σ (where $u_\mathrm{r} = 0$) can be extended further by considering the situation in Fig. 4.3. Indeed, the following conditions are fulfilled for this region:

$$\mathcal{H}\Big|_{u_\mathrm{r}=0} = -\nu - \lambda_\mathrm{r} x_\mathrm{r} - \lambda_\mathrm{m} x_\mathrm{m} + x_\mathrm{m} = 0, \quad \mathcal{A}_\mathrm{m} = \lambda_\mathrm{m} - x_\mathrm{m} = 0, \quad \mathcal{A}'_\mathrm{m} = 0\,.$$

Therefore, $\nu = -\lambda_\mathrm{r} x_\mathrm{r} - \lambda_\mathrm{m} x_\mathrm{m} + x_\mathrm{m}$, $\lambda_\mathrm{m} = x_\mathrm{m}$ and the condition $\mathcal{A}'_\mathrm{m} = 0$: $-1 + 2x_\mathrm{m} = 0$ gives $x_\mathrm{m} = 1/2$, which is precisely the definition of S_1^σ when $\varepsilon = 0$ [see Eq. (4.22)].

Consider now the region where x_r is smaller than on the surface S_r^σ (see Fig. 4.4), where $u_\mathrm{r} = 1$. There is a switching surface which extends the surface S_m and is defined by the same Eq. (4.14). However, there could also exist a singular arc S_2^σ starting from some points of S_m. Such an arc must satisfy the following conditions

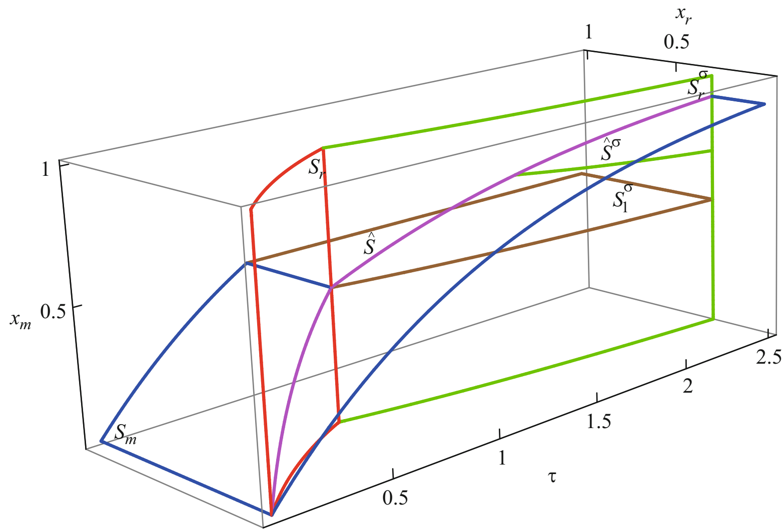


Fig. 4.5 Optimal behavioral pattern for $c = 3$

$$\mathcal{H}\Big|_{u_r=1} = -\nu - \lambda_r(x_r(1-c)-1) - \lambda_m x_m(1-c) - zU_m + x_m = 0 \quad (4.31)$$

$$\mathcal{A}_m = \lambda_m - x_m = 0, \quad \mathcal{A}'_m = 0, \quad (4.32)$$

which give a possible candidate for a singular arc S_2^σ: $x_m = 1/(2-c)$. We see that its appearance is possible only for $c < 1$, since x_m must belong to S_m. For $c > 1$ the structure of the solution in the domain below the surface S_r^σ is actually simpler and consists only of the switching surface S_m, see Fig. 4.5. Notice that in the case $x_r(0) = x_m(0) = 0$ investigated below, the existence of the singular arc S_2^σ is not relevant, since it cannot be reached from such initial conditions.

4.3.4 Computation of the Value Functions in the Case $\varepsilon = 0$

Following [1], we assume that at the beginning of the season the energy of consumers is zero: $x_r(0) = x_m(0) = 0$. Therefore, we should take into account only the trajectories coming from these initial conditions. The phase space is reduced in this case to the one shown in Fig. 4.6. One can see that there are three different regions depending on the length of the season T. If it is short enough, i.e. $T \leq T_1$ (see Eq. (4.8)), then the behavior of the mutant coincides with the behavior of the resident and the main population cannot be invaded: actually, the behavior of the mutant coincides with the behavior of the residents. If the length of the season is between T_1 and T_2, there is a period in the life-time of a resident when it applies

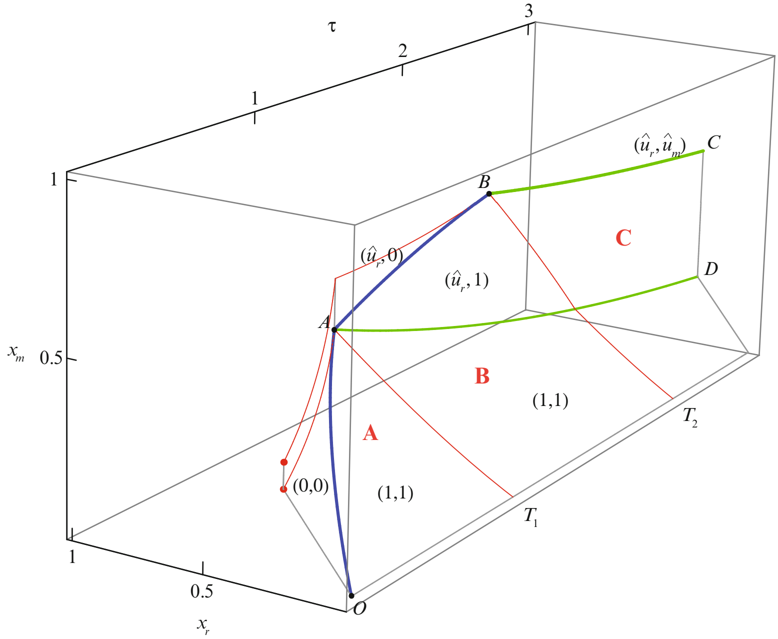


Fig. 4.6 The reduced optimal pattern for trajectories satisfying the initial conditions $x_r(0) = x_m(0) = 0$ with $c = 3$

an intermediate strategy and spares some amount of the resource for its future use. Mutants are able to use this fact and there exists a strategy that guarantees them better results.

Let us introduce the analogue of the value function $\tilde{U}_m$ for the resident and denote it by $\tilde{U}_r$:

$$\tilde{U}_r(p_r, p_m, n, t) = \int_{T-t}^{T} p_r(s)(1 - u_r(s))\,ds\,.$$

The value $\tilde{U}_r(0, 0, n(0), T)$ represents the amount of eggs laid by the resident during a season of length T. Its value depends on the state of the system and the following transformation can be done $\tilde{U}_r(p_r, p_m, n, t) = nU_r(x_r, x_m, t)$. In the following, we omit some parameters and write the value function in the simplified form $U_r(T) \doteq U_r(0, 0, T)$ where the initial conditions $x_r(0) = x_m(0) = 0$ have been taken into account.

In the region **A** (see Fig. 4.6) the value functions for both populations (of mutants and residents) are equal to each other $U_m(T) = U_r(T) = x_1\,e^{-c(T-\tau_1)}$. Here the value τ_1 can be defined from the intersection of the trajectory and the switching curve $S_r \cap S_m$:

$$1 - e^{-\tau_1} = \frac{e^{(c-1)(T-\tau_1)} - 1}{c - 1}\,.$$

To obtain the value functions in the regions **B** and **C**, one must solve the system of characteristics (4.29) in the case when the characteristics move along the surface S_{r}^{σ} and $u_{\mathrm{m}} = 1$. This leads to the following characteristic equations for the Hamiltonian (4.27):

$$x_{\mathrm{r}}' = -\frac{x_{\mathrm{r}}^2 c}{2 + x_{\mathrm{r}} c}, \quad x_{\mathrm{m}}' = x_{\mathrm{m}} \frac{2 - x_{\mathrm{r}} c}{2 + x_{\mathrm{r}} c} - 1, \quad U_{\mathrm{m}}' = -U_{\mathrm{m}} \frac{x_{\mathrm{r}}^2 c}{2 + x_{\mathrm{r}} c},$$

and consequently

$$x_{\mathrm{m}} = C_1 x_{\mathrm{r}}^2 \mathrm{e}^{\tau} + x_{\mathrm{r}} z + 1, \quad U_{\mathrm{m}} = C_2 x_{\mathrm{r}}^2, \quad C_1, C_2 = \mathrm{const}, \tag{4.33}$$

where C_1 and C_2 are defined from the boundary conditions, while Eq. (4.6) is also fulfilled.

Along the singular arc $\hat{S}^{\sigma}$ the mutant uses the intermediate strategy (4.21). In this case,

$$U_{\mathrm{m}}' = -U_{\mathrm{m}} c \hat{u}_{\mathrm{r}} + x_{\mathrm{m}} (1 - \tilde{u}_{\mathrm{m}}) = -U_{\mathrm{m}} \frac{2 x_{\mathrm{r}} c}{2 + x_{\mathrm{r}} c} + \frac{1 + x_{\mathrm{r}} c}{4}.$$

Since $x_{\mathrm{r}}' = -\frac{x_{\mathrm{r}}^2 c}{2 + x_{\mathrm{r}} c}$, we have $\frac{\mathrm{d}U_{\mathrm{m}}}{\mathrm{d}_{\mathrm{r}}} = \frac{2U_{\mathrm{m}}}{x_{\mathrm{r}}} - \frac{(1 + x_{\mathrm{r}} c)(2 + x_{\mathrm{r}} c)}{4 x_{\mathrm{r}}^2 c}$. Thus

$$U_{\mathrm{m}} = C_3 x_{\mathrm{r}}^2 + \frac{4 + 3 x_{\mathrm{r}} c (3 + 2 x_{\mathrm{r}} c)}{24 x_{\mathrm{r}} c}, \quad C_3 = \mathrm{const}. \tag{4.34}$$

We now undertake to compute the limiting season length T_2 that separates the region **B** from the region **C**. The coordinates of the point B were obtained in the previous section. To define the coordinates of the point $(x_{r_2}^{\sigma}, x_{m_2}^{\sigma}, \tau_2^{\sigma})$ of intersection of the optimal trajectory with the curve AD, we use the dynamics of motion along the surface S_{r}^{σ} with $u_{\mathrm{r}} = \hat{u}_{\mathrm{r}}$ and $u_{\mathrm{m}} = 1$ (4.33): $x_{\mathrm{m}} = C_1 x_{\mathrm{r}}^2 \mathrm{e}^{\tau} + x_{\mathrm{r}} z + 1$, where the constant C_1 should be chosen such that: $x_{m_2} = C_1 x_{r_2}^2 \mathrm{e}^{\tau_2} + x_{r_2} c + 1$, $x_{m_2} = \frac{2 + x_{r_2} c}{4} = 1 - \mathrm{e}^{-\tau_2}$. Therefore $C_1 = \frac{(x_{r_2} c - 2)(3 x_{r_2} c + 2)}{16 x_{r_2}^2}$. After that the coordinates: $x_{r_2}^{\sigma}$, $x_{m_2}^{\sigma}$ and τ_2^{σ} can be defined from the following conditions

$$x_{m_2}^{\sigma} = x_2^{\sigma} = C_1 (x_{r_2}^{\sigma})^2 \mathrm{e}^{\tau_2^{\sigma}} + x_{r_2}^{\sigma} c + 1, \quad \tau_2^{\sigma} = -\log x_{r_2}^{\sigma} + \frac{2}{x_{r_2}^{\sigma} c} - \frac{4}{c}. \tag{4.35}$$

The boundary value T_2 can be obtained from $T_2 = \tau_2^{\sigma} + \log(x_{r_2}^{\sigma}(c-1)+1)/(c-1)$.

Now we compute the value functions $U_{\mathrm{r}}(T)$ and $U_{\mathrm{m}}(T)$ for the region **B** ($T_1 < T \leq T_2$), where only the mutant uses bang-bang control. For the resident population we have

$$U_{\mathrm{r}}(T) = U_{r_2} \mathrm{e}^{-c(T - \tau_2)}, \quad U_{r_2} = x_{r_2}(1 - x_{r_2}) + \frac{1 - 2 x_{r_2}}{c}, \tag{4.36}$$

where the point with coordinates $(x_{r_2}, x_{r_2}, \tau_2)$ defines the intersection of the trajectory and surface $S_{\rm r}^{\sigma}$:

$$\tau_2 = -\log x_{r_2} + \frac{2}{x_{r_2} c} - \frac{4}{c}, \quad x_{r_2} = \frac{\mathrm{e}^{(c-1)(T-\tau_2)} - 1}{c-1}. \tag{4.37}$$

For the mutant population the value function $U_{\rm m}$ in the region with $u = \hat{u}$ and $u_{\rm m} = 1$ satisfies the equation resulting from (4.33):

$$U_{\rm m}^{(\hat{u},1)} = x_{m_1}^2 \left(x_{\rm r}/x_{r_1}\right)^2, \tag{4.38}$$

where $(x_{r_1}, x_{m_1}, \tau_1)$ is the point of intersection of the trajectory with the curve AB (see Fig. 4.6). Using (4.38) and notation from (4.37), we can write $U_{\rm m}(T) = U_{m_2}\,\mathrm{e}^{-c(T-\tau_2)}$, $U_{m_2} = x_{m_1}^2 \left(x_{r_2}/x_{r_1}\right)^2$, which is analogous to (4.36).

In the region **C** the value function for the resident has the same form as in (4.36), but it has a different form for the mutant. Suppose that the optimal trajectory intersects the surface S^{σ} at the point with coordinates $(\tilde{x}_{r_2}, \tilde{x}_{m_2}, \tilde{\tau}_2)$. Then the Bellman function at this point is given by

$$\tilde{U}_{m_2} = \tilde{x}_{r_2}^2 \left(\frac{c^2}{16} - \frac{4 + 3\tilde{x}_{r_2} c}{24\tilde{x}_{r_2}^3 c}\right) + \frac{3\tilde{x}_{r_2} c(2\tilde{x}_{r_2} c + 3) + 4}{24\tilde{x}_{r_2} c},$$

which is written using (4.34) with definition of the constant C_3 from the given boundary conditions.

When the optimal trajectory moving along the surface S^{σ} intersects the curve AD at some point with coordinates $(\tilde{x}_{r_2}^{\sigma}, \tilde{x}_{m_2}^{\sigma}, \tilde{\tau}_2^{\sigma})$ (see Fig. 4.6), the Bellman function can be expressed as follows: $\tilde{U}_{m_2}^{\sigma} = \tilde{U}_{m_2}\,\tilde{x}_{r_2}^{\sigma}/\tilde{x}_{r_2}$. Thus $U_{\rm m}(T) = \tilde{U}_{m_2}^{\sigma}\,\mathrm{e}^{-c(T-\tilde{\tau}_2^{\sigma})}$.

The difference in the value functions (number of offspring per mature individual) of themutant and optimally behaving resident is presented in Fig. 4.7. It is shown that as soon as the season length is longer than T_1, residents may be out-competed by selfish "free riding" mutants (see also Fig. 4.8). Otherwise the pay-off functions of the mutants and residents are the same. Therefore, if the season length is shorter than T_1, the optimal strategy of the resident is evolutionary stable in the sense that it cannot be beaten by any other strategy [11]. Thus, in the present example, collective optimal strategies of the bang-bang type are also evolutionary stable, while those of the bang-singular-bang type may always be outcompeted by alternative strategies. Whether such properties also hold in more general settings is an important topic of future research.

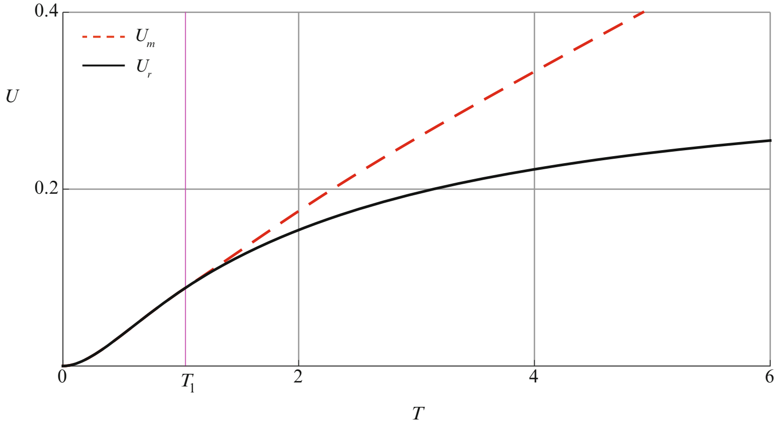


Fig. 4.7 Difference in the value functions of the resident and the mutant ($c = 3$)

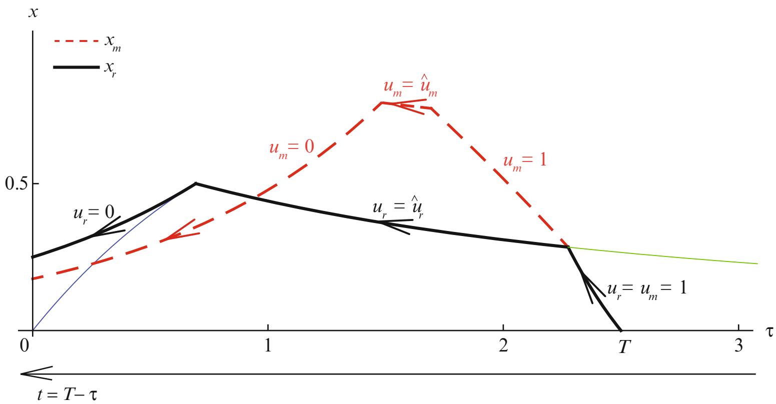


Fig. 4.8 Optimal free-riding by the mutant

4.3.5 Generalization to Sufficiently Small But Non-zero Values of ε

In this section we consider the case of non-zero ε such that the condition (4.24) remains fulfilled. This means that the trajectory intersecting the singular surface S_{r}^{σ} does not cross it, but moves along it due to the residents who make it invariant through the behavior $\hat{u}_{\mathrm{r}}$ (4.23).

In this case, the phase space can also be divided into two regions: according to whether x_{r} is smaller or larger than on S_{r}^{σ}. In both of these regions the structure of the solution has similar properties to the case considered above when ε is arbitrarily small. On the surface S_{r}^{σ} the optimal behavior is also similar to that of the previous case.

In the region with larger x_r values than the ones on the surface S_r^σ, there is a part of the switching surface S_m and a singular arc S_1^σ where the mutant uses an intermediate strategy. The surface S_1^σ can be defined using the expression (4.22). In the other region, we also have a part of S_m and a singular arc S_2^σ which is different from S_1^σ and may not exist for some values of the parameters c and ε.

To identify the values for which the surface S_2^σ is a part of the solution, let us write the necessary conditions as in (4.31)–(4.32): $\mathcal{H}\big|_{u_\mathrm{r}=1} = 0$, $\mathcal{A}_\mathrm{m} = 0$, $\mathcal{A}'_\mathrm{m} = \{\mathcal{A}_\mathrm{m}\mathcal{H}\} = 0$. Using these equations, we are able to obtain the values of λ_r, λ_m and ν on the surface S_2^σ and substitute them into the second derivative $\mathcal{A}''_\mathrm{m} = \{\{\mathcal{A}_\mathrm{m}\mathcal{H}\}\mathcal{H}\} = 0$ to derive the expression for the singular control applied by the mutant on this surface:

$$u_\mathrm{m} = \frac{2x_\mathrm{m} - (1-\varepsilon)c(1+x_\mathrm{m})}{2-(1-\varepsilon)c + x_\mathrm{m}\varepsilon c}\,. \tag{4.39}$$

There are several conditions which must be satisfied. First of all, the control (4.39) should be between zero and one

$$0 \le \frac{2x_\mathrm{m} - (1-\varepsilon)c(1+x_\mathrm{m})}{2-(1-\varepsilon)c + x_\mathrm{m}\varepsilon c} \le 1\,. \tag{4.40}$$

Second, the Kelley condition should also be fulfilled [12, p. 200]:

$$\frac{\partial}{\partial u_\mathrm{m}}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{\partial \mathcal{H}}{\partial u_\mathrm{m}} = \{\mathcal{A}_\mathrm{m}\{\mathcal{A}_\mathrm{m}\mathcal{H}\}\} \le 0\,.$$

This leads to the inequality

$$2-(1-\varepsilon) + x_\mathrm{m}\varepsilon c \ge 0\,. \tag{4.41}$$

In particular, conditions (4.40) and (4.41) together give $x_\mathrm{m} \le 2/(2-c)$.

To construct the singular arc S_2^σ, we should substitute the singular control u_m from (4.40) and $u_\mathrm{r} = 1$ into the equation describing the dynamics (4.13): $x'_\mathrm{m} = x_\mathrm{m}(1 - c((1-\varepsilon)u_\mathrm{r} + \varepsilon u_\mathrm{m})) - u_\mathrm{m}$, with the boundary conditions obtained from the tangency condition for the optimal trajectory from the domain $u_\mathrm{m} = u_\mathrm{r} = 1$ intersecting the switching surface S_m:

$$x_m\left(-\log\left(1 - \frac{1}{2-c(1-\varepsilon)}\right)\right) = \frac{1}{2-c(1-\varepsilon)}\,.$$

Such tangency occurs only if $0 \le \frac{1}{2-c(1-\varepsilon)} \le 1$, which comes from the condition that a singular surface S_m exists only for $0 \le x_\mathrm{m} < 1$. This gives the following inequality: $1 - c(1-\varepsilon) \ge 0$ for the existence of the surface S_2^σ. One can check that the inequalities (4.40)–(4.41) are fulfilled as well. The optimal behavioral pattern for a particular case is shown in Fig. 4.9.

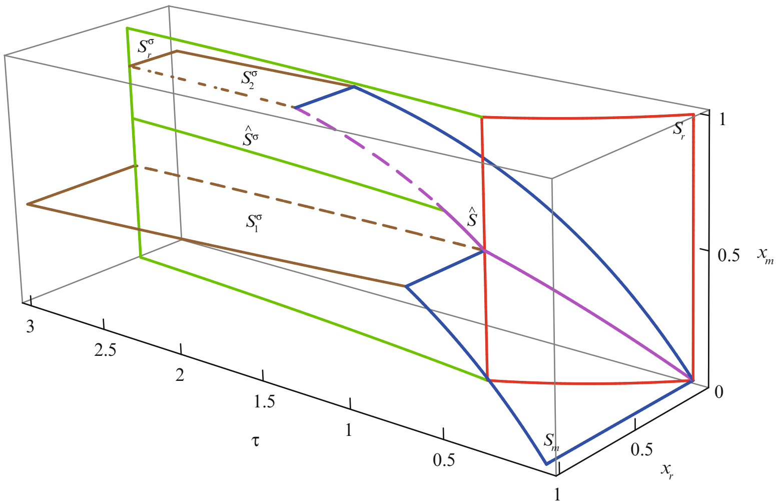


Fig. 4.9 Structure of the optimal behavioral pattern for $c = 1.25$ and $\varepsilon = 0.35$

4.4 Long-Term Evolution of the System

Model (4.2) was introduced in [1] as the intra-seasonal part of a more complex multi-seasonal model of population dynamics in which consumers and resources live for one season only. It was assumed that the (immature) offspring produced by the consumers and resources in season i and defined by the system of Eq. (4.3), mature during the inter-season to form the initial consumer and resource populations of season $(i+1)$, up to some overwintering mortality. The consumer and resource population densities at the beginning of season $i+1$ is thus $c_{i+1} = \mu_c J_i$, $n_{i+1}(t=0) = \mu_n J_{n,i}$, with J_i and $J_{n,i}$ defined in (4.3) ($\mu_n, \mu_c < 1$ allow for overwintering mortality).

In the presence of a mutant invasion, things differ slightly as the total consumer population is structured into $c_{r_i} = (1-\varepsilon_i)c_i$ residents and $c_{m_i} = \varepsilon_i c_i$ mutants that have different reproduction strategies. Assuming that reproduction is asexual and an offspring simply inherits the strategy of their parent, the inter-seasonal dynamics are as follows: $c_{r_{i+1}} = \alpha\tilde{U}_\mathrm{r}(c_i, \varepsilon_i, n_i, T) = (1-\epsilon_{i+1})c_{i+1}$, $c_{m_{i+1}} = \alpha\tilde{U}_\mathrm{m}(c_i, \varepsilon_i, n_i, T) = \epsilon_{i+1}c_{i+1}$ and $n_{i+1} = \beta\tilde{V}(c_i, \varepsilon_i, n_i, T)$, where $\alpha = \mu_c\theta$, $\beta = \mu_n\gamma$, and the functions $\tilde{U}_\mathrm{r} = (1-\varepsilon_i)c_i \int_0^T (1-u_\mathrm{r}(t))p_\mathrm{r}(t)\,\mathrm{d}t$, $\tilde{U}_\mathrm{m} = \varepsilon_i c_i \int_0^T (1-u_\mathrm{m}(t))p_\mathrm{m}(t)\,\mathrm{d}t$, $\tilde{V} = \int_0^T n(t)\,\mathrm{d}t$ can be computed from the solution of the optimal control problem (4.10) with the dynamics given by (4.9). As stated earlier, the energies of both the mutants and residents are zero at the beginning of each season ($p_\mathrm{r}(0) = p_\mathrm{m}(0) = 0$). For the particular case $\varepsilon = 0$, the values $\tilde{U}_\mathrm{r}$ and $\tilde{U}_\mathrm{m}$ were derived analytically in Sect. 4.3.4, but these are not useful in a multi-season study where the frequency of mutants is bound to evolve. In the following, we therefore resorted to a numerical investigation, in order to decipher the long-term fate of the mutants' invasion.

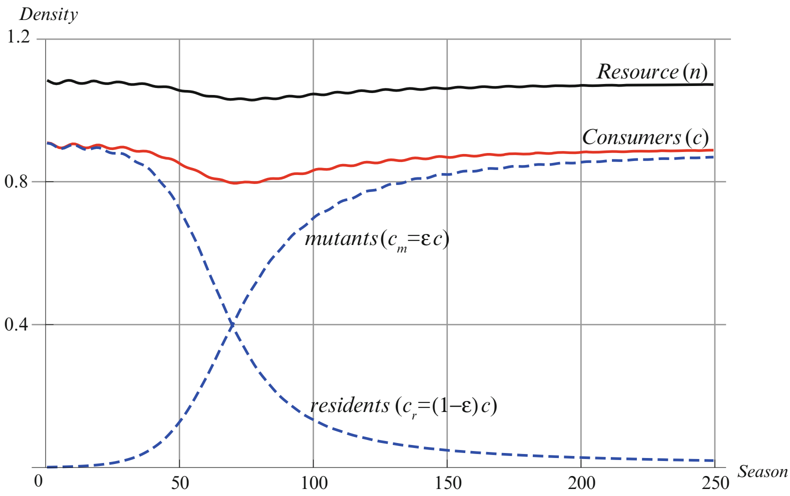


Fig. 4.10 Effect of an invasion by mutants on the system

Here, we follow an adaptive dynamics type approach and assume that, among all possible behaviors [1], the resident consumer and the resource population are at a (globally stable) equilibrium. We investigate what happens when a small fraction of mutants appear in the resident consumer population. We actually assume that resident consumers are "naive" in the sense that even if the mutant population becomes large through the season-to-season reproduction process, the resident consumers keep their collective optimal strategy and treat mutants as cooperators, even if they do not cooperate.

The case that we investigated is characterized by $\alpha = 2$, $\beta = 0.5$ and $T = 4$. Initially, the system is near the all-residents long-term stable equilibrium point $c = 0.9055$ and $n = 1.0848$. At the beginning of some season, a mutant population of small size $c_{\mathrm{m}} = 0.001$ then appears ($\varepsilon \approx 1.1\ 10^{-3} < 1/c$). We see in Fig. 4.10 that the mutant population increases its frequency within the consumer population and modifies the dynamics of the system. Despite this drastic increase, it is however noteworthy to underline that $c_i < 1$ in all seasons, so that $\varepsilon < 1/c_i$ is true and the analysis presented in this paper is valid for all seasons.

The naive behavior of the consumers is detrimental to their progeny: as the seasons pass, mutant consumers progressively take the place of the collectively optimal residents and even replace them in the long run (Fig. 4.10), making the mutation successful. We should however point out that the mutants' strategy, as described in (4.10), is also a kind of "collective" optimum: in some sense, it is assumed that mutants cooperate with other mutants. If the course of evolution drives the resident population to 0 and only mutants survive in the long run, this means that the former mutants become the new residents, with exactly the same strategy as the one of the former residents they replaced. Hence, they are also prone to being invaded by non-cooperating mutants. The evolutionary dynamics of this naive resident-selfish mutant-resource thus appears to be a never-ending process: selfish mutants can invade and replace collectively optimal consumers, but at the end transform into collectively optimal consumers as well, and a new selfish mutant invasion can start again. We are actually not in a "Red Queen Dynamics" context, since we focused on the evolution of one species only, and not co-evolution [20]. Yet, what the Red Queen said to Alice seems to fit the situation we have just described very well: "here, you see, it takes all the running you can do to keep in the same place" [5].

Acknowledgements This research has been supported by grants from the Agropolis Foundation and RNSC (project ModPEA, covenant support number 0902-013), and from INRA (call for proposal "Gestion durable des résistances aux bio-agresseurs", project Metacarpe, contract number 394576). A.R.A. was supported by a Post-Doctoral grant from INRIA Sophia Antipolis – Méditerranée and by the grant of the Russian Federal Agency on Education, program 1.2.1, contract P938.

References

1. Akhmetzhanov, A.R., Grognard, F., Mailleret, L.: Optimal life-history strategies in seasonal consumer–resource dynamics. Evolution **65**(11), 3113–3125 (2011). doi:10.1111/j.1558-5646.2011.01381.x
2. Auger, P., Kooi, B.W., de la Parra, R.B., Poggiale, J.-C.: Bifurcation analysis of a predator-prey model with predators using hawk and dove tactics. J. Theor. Biol. **238**(3), 597–607 (2011). doi:10.1016/j.jtbi.2005.06.012
3. Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton (1957)
4. Carathéodory, C.: Calculus of variations and partial equations of the first order. Holden-Day, San Francisco (1965)
5. Carroll, L.: Through the Looking-Glass, and What Alice Found There. MacMillan and Co., London (1871)
6. Dercole, F., Rinaldi, S.: Analysis of Evolutionary Processes: The Adaptive Dynamics Approach and its Applications. Princeton University Press, Princeton (2008)
7. Hamelin, F., Bernhard, P., Wajnberg, É.: Superparasitism as a differential game. Theor. Popul. Biol. **72**(3), 366–378 (2007). doi:10.1016/j.tpb.2007.07.005
8. Houston, A., Székely, T., McNamara, J.: Conflict between parents over care. Trends Ecol. Evol. **20**(1), 33–38 (2005). doi:10.1016/j.tree.2004.10.008
9. Mailleret, L., Lemesle, V.: A note on semi-discrete modelling in life sciences. Phil. Trans. R. Soc. A **367**(1908), 4779–4799 (2009). doi:10.1098/rsta.2009.0153
10. Maynard Smith, J.: Evolution and the Theory of Games. Cambridge University Press, Cambridge (1982)
11. Maynard Smith, J., Price, G.R.: The logic of animal conflict. Nature **246**(5427), 15–18 (1973). doi:10.1038/246015a0
12. Melikyan, A.A.: Generalized Characteristics of First Order PDEs: Applications in Optimal Control and Differential Games. Birkhäuser, Boston (1998)
13. Murray, J.D.: Mathematical Biology. Springer, Berlin (1989)
14. Mylius, S.D., Diekmann, O.: On evolutionarily stable life histories, optimization and the need to be specific about density dependence. Oikos **74**(2), 218–224 (1995)
15. Noether, E.: Invariante Variationsprobleme. Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse 235–257 (1918)

16. Perrin, N., Mazalov, V.: Local competition, inbreeding, and the evolution of sex-biased dispersal. Am. Nat. **155**(1), 116–127 (2000). doi:10.1086/303296
17. Perrin, N., Sibly, R.M.: Dynamic-models of energy allocation and investment. Annu. Rev. Ecol. Syst. **24**, 379–410 (1993)
18. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The Mathematical Theory of Optimal Processes. Wiley, New York (1962)
19. Schaffer, W.M.: The application of optimal control theory to the general life history problem. Am. Nat. **121**, 418–431 (1983)
20. Van Valen, L.: A New evolutionary law. Evol. Theory **1**, 1–30 (1973)
21. Vincent, T.L., Brown, J.S.: Evolutionary Game Theory, Natural Selection and Darwinian Dynamics. Cambridge University Press, Cambridge (2005)

Chapter 5
Strong Strategic Support of Cooperative Solutions in Differential Games

Sergey Chistyakov and Leon Petrosyan

Abstract The problem of strategically provided cooperation in m-person differential games for a prescribed duration and integral payoffs is considered. The Shapley value operator is chosen as the cooperative optimality principle. It is shown that components of Shapley value are absolutely continuous and, thus, differentiable functions along any admissible trajectory. The main result consists in the fact that if in any subgame along the cooperative trajectory the Shapley value belongs to the core of this subgame, then the payoffs as components of the Shapley value can be realized in a specially constructed strong Nash equilibrium, i.e., an equilibrium that is stable against the deviation of coalitions.

Keywords Strong Nash equilibrium • Time consistency • Shapley value • Cooperative trajectory

5.1 Introduction

Like the analysis in [9], in this paper the problem of strategic support of cooperation in a differential m-person game with prescribed duration T and independent motions is considered. Based on the initial differential game, a new associated differential game (CD game) is designed. In addition to the initial game, it models players' actions in connection with the transition from a strategic form of the game to a cooperative one with the principle of optimality chosen in advance. The model makes it possible to refuse cooperation at any time instant t for any coalition of players. As the cooperative principle of optimality, the Shapley value operator is considered. Under certain assumptions, it is shown that the components of the

S. Chistyakov (✉) • L. Petrosyan
Faculty of Applied Mathematics and Control Processes, St. Petersburg University, Universitetskiy pr. 35, St. Petersburg 198504, Russia
e-mail: svch50@mail.ru; spbuoasis7@peterlink.ru

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_5,

Shapley value along any admissible trajectory are absolutely continuous functions of time. In the foundation of the CD-game construction lies the so-called imputation distribution procedure described in [9] (see also [1]). The theorem established by the authors states that if at each time instant along the conditionally optimal (cooperative) trajectory future payments to each coalition of players according to the imputation distribution procedure exceed the maximal guaranteed value that this coalition can achieve in the CD game, then there exists a strong Nash equilibrium in the class of recursive strategies first introduced in [2]. In other words, the aforementioned equilibrium exists if in any subgame along the conditionally optimal trajectory the Shapley value belongs to its core. The proof of this theorem uses results and methods published in [2, 3]. The proved theorem is also true for other value operators possessing the property of absolute continuity along admissible trajectories of the differential game under consideration. The motions of players in the game are independent. Thus the motion equations have the form

$$\frac{\mathrm{d}x^{(i)}}{\mathrm{d}t} = f^{(i)}(t,x^{(i)},u^{(i)}),\ i \in I = [1;m], \tag{5.1}$$

$$x^{(i)} \in R^{n(i)}, u^{(i)} \in P^{(i)} \in CompR^{k(i)}$$

$$x^{(i)}(t_0) = x_0^{(i)}, \quad i \in I. \tag{5.2}$$

The payoffs of players $i \in I = [1:m]$ have the integral form

$$H^{(i)}_{t_0,x_0}(u(\cdot)) = \int_{t_0}^{T} h^{(i)}(t,x(t,t_0,x_0,u(\cdot)))\,\mathrm{d}t. \tag{5.3}$$

Here $u(\cdot) = (u^{(1)}(\cdot),\dots,u^{(m)}(\cdot))$ is a given m-vector of open-loop controls:

$$x(t,t_0,x_0,u(\cdot)) = \Big(x^{(1)}(t,t_0,x_0,u^{(1)}(\cdot)),\dots,x^{(m)}(t,t_0,x_0,u^{(m)}(\cdot))\Big),$$

where $x^{(i)}(\cdot) = x(\cdot,t_0,x_0^{(i)},u^{(i)}(\cdot))$ is the solution of the Cauchy problem for the ith subsystem of (5.1) with corresponding initial conditions (5.2) and admissible open-loop control $u^{(i)}(\cdot)$ of player i.

Admissible open-loop controls of players $i \in I$ are Lebesgue measurable open-loop controls

$$u^{(i)}(\cdot) : t \mapsto u^{(i)}(t) \in R^{k(i)}$$

such that

$$u^{(i)}(t) \in P^{(i)} \text{ for all } t \in [t_0,T].$$

It is supposed that all of the functions

$$f^{(i)} : R \times R^{k(i)} \times P^{(i)} \to R^{k(i)}, \quad i \in I,$$

are continuous, locally Lipschitz with respect to $x^{(i)}$, and satisfy the condition $\exists \lambda^{(i)} > 0$ such that

$$||f^{(i)}(t, x^{(i)}, u^{(i)})|| \leq \lambda^{(i)}(1 + ||x^{(i)}||) \quad \forall x^{(i)} \in R^{k(i)}, \quad \forall u^{(i)} \in P^{(i)}.$$

Each of the functions

$$h^{(i)} : R \times R^{k(i)} \times P^{(i)} \to R, \quad i \in I$$

is also continuous.

It is supposed that at each time instant $t \in [t_0, T]$, the players have information about the trajectory (solution) $x^{(i)}(\tau) = x(\tau, t_0, x_0, u^{(i)}(\cdot))$ of the system (5.1), (5.2) on the time interval $[t_0, t]$ and use recursive strategies [1, 2].

5.2 Recursive Strategies

Recursive strategies were first introduced in [1] to justify the dynamic programming approach in zero-sum differential games, known as the method of open-loop iterations in nonregular differential games with a nonsmooth value function. The ε-optimal strategies constructed with the use of this method are universal in the sense that they remain ε-optimal in any subgame of the previously defined differential game (for every $\varepsilon > 0$). Exploiting this property it became possible to prove the existence of ε-equilibrium (Nash equilibrium) in non-zero-sum differential games (for every $\varepsilon > 0$) using the so-called "punishment strategies" [4].

The basic idea is that when one of the players deviates from the conditionally optimal trajectory, other players after some small time delay start to play against the deviating player. As a result, the deviating player is not able to obtain much more than he could have gotten using the conditionally optimal trajectory. Punishment of the deviating player at each time instant using the same strategy is possible because of the universal character of ε-optimal strategies in zero-sum differential games.

In this paper the same approach is used to verify the stability of cooperative agreements in the game $\Gamma(t_0, x_0)$ and, as in the aforementioned case, the principal argument is the universal character of ε-optimal recursive strategies in specially defined zero-sum games $\Gamma_S(t_0, x_0)$, $S \subset I$, associated with the non-zero-sum game $\Gamma(t_0, x_0)$.

The recursive strategies lie somewhere in between piecewise open-loop strategies [6] and ε-strategies introduced by Pshenichny [10]. The difference from piecewise open-loop strategies consists in the fact that, as in the case of Pshenichny's ε-strategies, the moments of correction of open-loop controls are not prescribed from the beginning of the game but are defined during the course of the game. At the same time, they differ from Pshenichny's ε-strategies in the fact that the formation of open-loop controls occurs in a finite number of steps.

The recursive strategies $U_i^{(n)}$ of player i with the maximal number of control corrections n is a procedure for an admissible open-loop formation by player i in the game $\Gamma(t_0,x_0),(t_0,x_0) \in D$.

At the beginning of the game $\Gamma(t_0,x_0)$, player i, using the recursive strategy $U_i^{(n)}$, defines the first correction instant $t_1^{(i)} \in (t_0,T]$ and his admissible open-loop control $u^{(i)} = u^{(i)}(t)$ on the time interval $[t_0,t_1^{(i)}]$. Then, if $t_1^{(i)} < T$, possessing information about the state of the game at time instant $t_1^{(i)}$, he chooses the next moment of correction $t_2^{(i)}$ and his admissible open-loop control $u^{(i)} = u^{(i)}(t)$ on the time interval $(t_1^{(i)},t_2^{(i)}]$ and so on. Then whether the admissible control on the time interval $[t_0,T]$ is formed at the kth step ($k \leq n-1$) or at step n, player i will end up with the process by choosing at time instant $t_{n-1}^{(i)}$ his admissible control on the remaining time interval $(t_{n-1}^{(i)},T]$.

5.3 Associated Zero-Sum Games and Corresponding Solutions

For each given state $(t_*,x_*) \in D$ and nonvoid coalition $S \subset I$ consider the zero-sum differential game $\Gamma_S(t_*,x_*)$ between coalition S and $I \backslash S$ with the same dynamics as in $\Gamma(t_*,x_*)$ and the payoff of coalition S equal to the sum of payoffs of the players $i \in S$ in the game $\Gamma(t_*,x_*)$:

$$\sum_{i \in S} H_{t_*x_*}^{(i)}(u^{(S)}(\cdot),u^{(I\backslash S)}(\cdot)) = \sum_{i \in S} H_{t_*x_*}^{(i)}(u(\cdot)) = \sum_{i \in S} \int_{t_0}^{T} h^{(i)}(t,x(t),u(t))\,\mathrm{d}t,$$

where

$$u^{(S)}(\cdot) = \{u^{(i)}(\cdot)\}_{i \in S},$$
$$u^{(I\backslash S)}(\cdot) = \{u^{(j)}(\cdot)\}_{j \in I\backslash S},$$
$$u(\cdot) = (u^{(S)}(\cdot),u^{(I\backslash S)}(\cdot)) = (u^{(1)}(\cdot),\ldots,u^{(m)}(\cdot)).$$

The game $\Gamma_S(t_*,x_*)$, $S \subset I$, $(t_*,x_*) \in D$, as $\Gamma(t_*,x_*),(t_*,x_*) \in D$ we consider in the class of recursive strategies. Under the conditions formulated previously, each of the games $\Gamma_S(t_*,x_*)$, $S \subset I$, $(t_*,x_*) \in D$ has a value

$$val\Gamma_S(t_*,x_*).$$

If $S = I$, then the game $\Gamma_S(t_*,x_*)$ becomes a one-player optimization problem. We suppose that in this game there exists an optimal open-loop solution. We denote the corresponding trajectory-solution of (5.1), (5.2) on the time interval $[t_0,T]$ by

$$x_0(\cdot) = (x_0^{(1)}(\cdot),\ldots,x_0^{(m)}(\cdot))$$

and call it the *conditionally optimal cooperative trajectory*. This trajectory may not necessarily be unique. Then on the set D the mapping

$$v(\cdot) : D \to R^{2^I}$$

is defined with coordinate functions

$$v_S(\cdot) : D \to R, \;\; S \subset I,$$
$$v_S(t_*, x_*) = val\Gamma_S(t_*, x_*).$$

This mapping attributes to each state $(t_*, x_*) \in D$ a characteristic function $v(t_*, x_*) : 2^I \to R$ of a non-zero-sum game $\Gamma(t_*, x_*)$ and thus an m-person classical cooperative game $(I, v(t_*, x_*))$.

Let $E(t_*, x_*)$ be the set of all imputations in the game $(I, v(t_*, x_*))$. The multivalue mapping

$$M : (t_*, x_*) \mapsto M(t_*, x_*) \subset E(t_*, x_*) \subset R^m,$$
$$M(t_*, x_*) \neq \Lambda \quad \forall (t_*, x_*) \in D,$$

is called an *optimality principle* (defined over the family of games $\Gamma(t_*, x_*)$, $(t_*, x_*) \in D$) and the set $M(t_*, x_*)$ a *cooperative solution of the game* $\Gamma(t_*, x_*)$ *corresponding to this principle*.

In what follows we shall consider only single-valued mappings of the form

$$M : (t_*, x_*) \mapsto M(t_*, x_*) \in E(t_*, x_*),$$
$$(t_*, x_*) \in D.$$

Concretely speaking we shall consider the Shapley value as the optimality principle, i.e., a mapping $Sh(\cdot) : D \to R^m$ in which to each state $(t_*, x_*) \in D$ corresponds the Shapley value $Sh(t_*, x_*)$ in the game $(I, v(t_*, x_*))$.

As follows from [8], under the preceding conditions, the following lemma holds.

Lemma 5.1 (Fridman [5]). *The functions* $v_S(\cdot) : D \to R, S \in I$, *are locally Lipschitz.*

Since the solution of the Cauchy problem (5.1), (5.2) in the sense of Caratheodory is absolutely continuous, the next theorem follows from Lemma 5.1.

Theorem 5.1. *For every solution of the Cauchy problem (5.1), (5.2) in the sense of Caratheodory*

$$x(\cdot) = (x^{(1)}(\cdot), \ldots, x^{(m)}(\cdot)),$$

corresponding to the m-system of open-loop controls

$$u(\cdot) = (u^{(1)}(\cdot), \ldots, u^{(m)}(\cdot))$$
$$(x^{(i)}(\cdot) = x(\cdot, t_0, x_0^{(i)}, u^{(i)}(\cdot)), \quad i \in I),$$

the functions

$$\varphi_S : [t_0, T] \to R, \quad S \subset I, \quad \varphi_S(t) = v_S(t, x(t))$$

are absolutely continuous functions on the time interval $[t_0, T]$.

Since each of the coordinate functions of the mapping $Sh(\cdot)$ is a linear combination of $v_S(\cdot), S \subset I$,

$$Sh_i(t_*, x_*) = \sum_{S \subset I:\, i \in S} \frac{(|S|-1)!(m-|S|)!}{m!} [v_S(t_*, x_*) - v_{S \setminus \{i\}}(t_*, x_*)],$$

from Theorem 5.1 we obtain the following corollary.

Corollary 5.1. *For each solution of the Cauchy problem (5.1), (5.2) in the sense of Caratheodory*

$$x(\cdot) = (x^{(1)}(\cdot), \ldots, x^{(m)}(\cdot)),$$

the functions

$$\alpha_i^{Sh} : [t_0, T] \to R, \quad \alpha_i^{Sh}(t) = Sh_i(t, x(t)), \quad i \in I,$$

are absolutely continuous.

5.4 Realization of Cooperative Solutions

We shall connect the realization of the single-valued solution of the game $\Gamma(t_0, x_0)$ with the known *imputation distribution procedure* (IDP) [7, 8].

By the IDP of the solution $M(t_0, x_0)$ of the game $\Gamma(t_0, x_0)$ along the conditionally optimal trajectory $x_0(\cdot)$ we understand function

$$\beta(t) = (\beta_1(t), \ldots, \beta_m(t)), \quad t \in [t_0, T], \tag{5.4}$$

satisfying

$$M(t_0, x_0) = \int_{t_0}^{T} \beta(t)\, \mathrm{d}t \tag{5.5}$$

and

$$\int_{t}^{T} \beta(t)\, \mathrm{d}t \in E(t, x_0(t)) \quad \forall t \in [t_0, T], \tag{5.6}$$

where $E(t, x_0(t))$ is the set of imputations in the game $(I, v(t, x_0(t)))$.

The IDP $\beta(t),\ t \in [t_0, T]$ of the solution $M(t_0, x_0)$ of the game $\Gamma(t_0, x_0)$ is called *dynamically stable (time-consistent)* along the conditionally optimal trajectory $x_0(\cdot)$ if

$$\int_t^T \beta(t)\,\mathrm{d}t = M(t, x_0(t)), \quad \forall t \in [t_0, T]. \tag{5.7}$$

The solution $M(t_0,x_0)$ of the game $\Gamma(t_0,x_0)$ is *dynamically stable (time-consistent)* if along at least one conditionally optimal trajectory the dynamically stable IDP exists.

Using the corollary from Theorem 5.1 we have the following result.

Theorem 5.2. *For any conditionally optimal trajectory $x_0(\cdot)$ the following IDP of the solution $Sh(t_0,x_0)$ of the game $\Gamma(t_0,x_0)$*

$$\beta(t) = -\frac{\mathrm{d}}{\mathrm{d}t} Sh(t, x_0(t)), \quad t \in [t_0, T], \tag{5.8}$$

is the dynamically stable IDP along this trajectory. Therefore, the solution $Sh(t_0,x_0)$ of the game $\Gamma(t_0,x_0)$ is dynamically stable.

5.5 Strategic Support of the Shapley Value

If in the game a cooperative agreement is reached and each player receives his payoff according to the IDP (5.8), then it is natural to suppose that those who violate this agreement are to be punished. The effectiveness of the punishment (sanctions) comes to question of the existence of the strong Nash equilibrium in the differential game $\Gamma^{Sh}(t_0,x_0)$, which differs from $\Gamma(t_0,x_0)$ only by player payoffs.

The payoff of player i in $\Gamma^{Sh}(t_0,x_0)$ is equal to

$$H_{t_0,x_0}^{(Sh,i)}(u(\cdot)) = -\int_{t_0}^{t(u(\cdot))} \frac{\mathrm{d}}{\mathrm{d}t} Sh_i(t, x_0(t))\,\mathrm{d}t + \int_{t(u(\cdot))}^{T} h^{(i)}(t, x(t, t_0, x_0, u(\cdot)))\,\mathrm{d}t,$$

where $t(u(\cdot))$ is the last time instant $t \in [t_0, T]$ for which

$$x_0(\tau) = x(\tau, t_0, x, u(\cdot)) \quad \forall \tau \in [t_0, t].$$

In this paper we use the following definition of the strong Nash equilibrium.

Definition 5.1. Let $\gamma = \langle I, \{X_i\}_{i\in I}, \{K_i\}_{i\in I}\rangle$ be an m-person game in normal form; here $I = [1:m]$ is the set of players, X_i the set of strategies of player i, and

$$K_i : X = X_1 \times X_2 \times \cdots \times X_m \to R$$

the payoff function of player i. We shall say that in the game γ there exists a *strong Nash equilibrium* if

$$\forall \varepsilon > 0 \quad \exists x^{\varepsilon} = (x_1^{\varepsilon}, x_2^{\varepsilon}, \ldots, x_m^{\varepsilon}) \in X$$

such that

$$\forall S \subset I, \forall x_S \in X_S = \prod_{i \in S} X_i,$$

$$\sum_{i \in S} K_i(x_S, x^{\varepsilon}_{I \setminus S}) - \varepsilon \leq \sum_{i \in S} K_i(x^{\varepsilon}),$$

where

$$x^{\varepsilon}_{I \setminus S} = \{x^{\varepsilon}_j\}_{j \in I \setminus S} \quad (x^{\varepsilon}_{I \setminus S} \in X_{I \setminus S}).$$

Let $C(t_*, x_*)$ be the core of the game $(I, v(t_*, x_*))$.

Theorem 5.3. *If for at least one conditionally optimal trajectory $x_0(\cdot)$ in the game $\Gamma(t_0, x_0)$ the condition*

$$Sh(t, x_0(t)) \in C(t, x_0(t)) \quad \forall t \in [t_0, T], \tag{5.9}$$

holds, then in the game $\Gamma^{Sh}(t_0, x_0)$ there exists a strong Nash equilibrium.

The idea of the proof is as follows. Condition (5.9) can be rewritten in the form

$$\sum_{i \in S} Sh_i(t, x_0(t)) \geq v_S(t, x_0(t)), \quad \forall S \subset I, \quad \forall t \in [t_0, T]. \tag{5.10}$$

This means that at each time instant $t \in [t_0, T]$, moving along the conditionally optimal trajectory $x_0(\cdot)$, no coalition can guarantee itself a payoff $[t, T]$ more than according to IDP (5.8), i.e., more than

$$\sum_{i \in S} \int_t^T \beta(\tau)\, \mathrm{d}\tau = -\sum_{i \in S} \int_t^T \frac{\mathrm{d}}{\mathrm{d}t} Sh(\tau, x_0(\tau))\, \mathrm{d}\tau = \sum_{i \in S} Sh_i(t, x_0(t));$$

at the same time, on the time interval $[t_0, t]$, according to the IDP, the coalition already received a payoff equal to

$$\sum_{i \in S} \int_{t_0}^t \beta_i(\tau)\, \mathrm{d}\tau = -\sum_{i \in S} \int_{t_0}^t \frac{\mathrm{d}}{\mathrm{d}t} Sh_i(\tau, x_0(\tau))\, \mathrm{d}\tau$$
$$= \sum_{i \in S} Sh_i(t_0, x_0) - \sum_{i \in S} Sh_i(t, x_0(t)).$$

Consequently, in the game $\Gamma^{Sh}(t_0, x_0)$, no coalition can guarantee a payoff of more than

$$\sum_{i \in S} Sh_i(t_0, x_0)$$

i.e., more than $Sh(t_0, x_0)$. According to the cooperative solution $x_0(\cdot)$ but moving always in the game $\Gamma^{Sh}(t_0, x_0)$ along the conditionally optimal trajectory, each coalition will receive its payoff according to the Shapley value. Thus no coalition can benefit from the deviation from the conditionally optimal trajectory, which in this case is natural to call a "strongly equilibrium trajectory."

5.6 Conclusion

Let us conclude with some remarks about the limits of our approach. The main condition that guarantees a strong strategic support of the Shapley value in the m-person differential game under consideration is the fact that the Shapley value belongs to the core of any subgame along a cooperative trajectory. This can be guaranteed only if the cores are not void and the characteristic functions in the subgames are convex. At the same time, one can easily verify that if instead of the Shapley value any fixed imputation from the core is taken as the optimality principle, then for strong strategic support of this imputation the principle condition is that the cores in subgames along the cooperative trajectory will be nonempty.

In addition, strategic support of the cooperation proposed here based on the notion of a strong Nash equilibrium is coalition proof in the sense that no coalition can force its members to deviate from the cooperative trajectory because in any deviating coalition there will be at least one player who is not interested in the deviation.

References

1. Chistyakov, S.V.: To the solution of game problem of pursuit. Prikl. Math. i Mech. **41**(5), 825–832 (1977) (in Russian)
2. Chistyakov, S.V.: Operatory znacheniya antagonisticheskikx igr (Value Operators in Two-Person Zero-Sum Differential Games). St. Petersburg Univ., St. Petersburg (1999)
3. Chentsov, A.G.: On a game problem of converging at a given instant of time. Math. USSR Sbornic **28**(3), 353–376 (1976)
4. Chistyakov, S.V.: O beskoalizionnikx differenzial'nikx igrakx (On coalition-free differential games). Dokl. Akad. Nauk. **259**(5), 1052–1055 (1981). English transl. in Soviet Math. Dokl. **24**(1), 166–169 (1981)
5. Fridman, A.: Differential Games. Wiley, New York (1971)
6. Petrosjan, L.A.: Differential Games of Pursuit. World Scientific, Singapore (1993)
7. Petrosjan, L.A.: The shapley value for differential games. In: Olsder, G.J. (ed.) Annals of the International Society of Dynamic Games, vol. 3, pp. 409–417, Birkhauser (1995)
8. Petrosjan, L.A., Danilov, N.N.: Stability of solutions in nonzero-sum differential games with integral payoffs, pp. 52–59. Viestnik Leningrad University, no. 1 (1979)
9. Petrosjan, L.A., Zenkevich, N.A.: Principles of stable cooperation. Math. Games Theory Appl. **1**(1), 102–117 (2009) (in Russian)
10. Pschenichny, B.N.: E-strategies in differential games. In: Topics in Differential Games, pp. 45–56. North-Holland Pub. Co. New York (1973)

Chapter 6
Characterization of Feedback Nash Equilibrium for Differential Games

Yurii Averboukh

Abstract We investigate the set of Nash equilibrium payoffs for two-person differential games. The main result of the paper is the characterization of the set of Nash equilibrium payoffs in terms of nonsmooth analysis. In addition, we obtain the sufficient conditions for a couple of continuous functions to provide a Nash equilibrium. This result generalizes the method of the system of Hamilton–Jacobi equations.

Keywords Nash equilibrium • Differential games • Nonsmooth analysis

6.1 Introduction

In this paper, we characterize Nash equilibrium payoffs for two-person differential games. We consider non-zero-sum differential games in the framework of discontinuous feedback strategies. This approach was first proposed by Krasovskii for zero-sum differential games [12]. The existence of the Nash equilibrium was established in the works of Kononenko [11] and Kleimenov [10]. The proof is based on the punishment strategy technique. This technique makes it possible to characterize a set of Nash equilibrium payoffs [8, 10]. Further, the technique was applied for the Nash equilibrium of stochastic differential games [5]. Another approach is based on the system of Hamilton–Jacobi equations. This approach was developed in [1] in the case of differentiable value functions. The Nash equilibrium strategies for some class of the game in the one-dimensional case was constructed on the basis of generalized solutions of the system of Hamilton–Jacobi equations by Cardaliaguet and Plaskacz [6]. In addition, Cardaliaguet investigated the stability of

Y. Averboukh (✉)
Institute of Mathematics and Mechanics UrB RAS, S. Kovalevskaya Street 16, GSP-384, Ekaterinburg 620990, Russia
e-mail: ayv@imm.uran.ru

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_6,

constructed solutions [4]. Bressan and Shen investigated the Nash equilibrium using a hyperbolic system of conservation laws [2, 3]. An approach based on singular surfaces was considered by Olsder [13].

In this paper, we develop the approach of Kononenko [11], Kleimenov [10], and Chistyakov [8]. The main result is the characterization of the set of Nash equilibrium payoffs in terms of nonsmooth analysis. In addition we obtain the sufficient conditions for a pair of continuous functions to provide a Nash equilibrium. This result generalizes the method of the systems of Hamilton–Jacobi equations.

6.2 Main Result

We consider the doubly controlled system

$$\dot{x} = f(t,x,u,v),\ \ t \in [t_0, \vartheta_0],\ \ x \in \mathbb{R}^n,\ \ u \in P,\ \ v \in Q. \tag{6.1}$$

Here u and v are the controls of Players I and II, respectively. Payoffs are terminal. Player I wants to maximize $\sigma_1(x(\vartheta_0))$, whereas Player II wants to maximize $\sigma_2(x(\vartheta_0))$. We assume that sets P and Q are compacts, function f, σ_1, and σ_2 are continuous, and f is Lipschitz continuous with respect to the phase variable and satisfies the sublinear growth condition with respect to x.

We use the control design suggested in [10]. This control design follows the Krasovskii discontinuous feedback formalization. A feedback strategy of Player I is a pair of functions $U = (u(t,x,\varepsilon), \beta_1(\varepsilon))$. Here $u(t,x,\varepsilon)$ is a function of position $(t,x) \in [t_0, \vartheta_0] \times \mathbb{R}^n$ and the precision parameter ε, $\beta_1(\varepsilon)$ is a continuous function of the precision parameter. We suppose that $\beta_1(\varepsilon) \to 0$, $\varepsilon \to 0$. Analogously, a feedback strategy of Player II is a pair $V = (v(t,x,\varepsilon), \beta_2(\varepsilon))$.

Let a position (t_*, x_*) be chosen. The step-by-step motion is defined in the following way. We suppose that the ith player chooses his own precision parameter ε_i. Let Player I choose the partition of the interval $[t_*, \vartheta_0]$ $\Delta_1 = \{\tau_j\}_{j=0}^r$. Assume that the mesh of the partition Δ is less than ε_1. Suppose that Player II chooses the partition $\Delta_2 = \{\xi_k\}_{k=1}^{\nu}$ of the mesh to be less than ε_2. The solution $x[\cdot]$ of Eq. (6.1) with initial date $x[t_*] = x_*$ such that the control of Player I is equal to $u(\tau_j, x[\tau_j], \varepsilon_1)$ on $[\tau_j, \tau_{j+1})$, and the control of Player II is equal to $v(\xi_k, x[\xi_k], \varepsilon_2)$ on $[\xi_k, \xi_{k+1})$ is called a step-by-step motion. Denote it by $x[\cdot, t_*, x_*; U, \varepsilon_1, \Delta_1; V, \varepsilon_2; \Delta_2]$. The set of all step-by-step motions from the position (t_*, x_*) under strategies U and V and precision parameters ε_1 and ε_2 is denoted by $X(t_*, x_*; U, \varepsilon_1; V, \varepsilon_2)$. The step-by-step motion is called consistent if $\varepsilon_1 = \varepsilon_2$.

A limit of step-by-motions $x[\cdot, t^k, x^k; U, \varepsilon_1^k, \Delta_1^k; V\varepsilon_2^k, \Delta_2^k]$ is called a constructive motion if $t^k \to t_*, x^k \to x_*, \varepsilon_1^k \to 0, \varepsilon_2^k \to 0$, as $k \to \infty$. Denote by $X(t_*, x_*; U, V)$ the set of constructive motions. By the Arzela–Ascoli theorem, the set of constructive motions in nonempty. A consistent constructive motion is a limit of step-by-step motions $x[\cdot, t^k, x^k; U, \varepsilon^k, \Delta_1^k; V, \varepsilon^k, \Delta_2^k]$ such that $t^k \to t_*$, $x^k \to x_*$, $\varepsilon_1^k \to 0$, $\varepsilon_2^k \to 0$, as $k \to \infty$. Denote the set of all consistent constructive motions by $X^c(t_*, x_*; U, V)$. This set is also nonempty.

The following definition of the Nash equilibrium is used.

Definition 6.1. Let $(t_*,x_*) \in [t_0,\vartheta_0] \times \mathbb{R}^n$. The pair of strategies U^N and V^N is said to be a Nash equilibrium solution at the position (t_*,x_*) if, for all strategies U and V, the following inequalities hold:

$$\max\left\{\sigma_1(x[\vartheta_0]) : x[\cdot] \in X(t_*,x_*,U,V^N)\right\}$$
$$\leq \min\left\{\sigma_1(x^c[\vartheta_0]) : x^c[\cdot] \in X^c(t_*,x_*,U^N,V^N)\right\}.$$
$$\max\left\{\sigma_2(x[\vartheta_0]) : x[\cdot] \in X(t_*,x_*,U^N,V)\right\}$$
$$\leq \min\left\{\sigma_2(x^c[\vartheta_0]) : x^c[\cdot] \in X^c(t_*,x_*,U^N,V^N)\right\}.$$

The payoff $(\sigma_1(x[\vartheta_0]),\sigma_2(x[\vartheta_0]))$ determined by a Nash equilibrium solution is called a Nash equilibrium payoff of a game. In the typical case, there are many Nash equilibria with different payoffs. The set of all Nash equilibrium payoffs is called a Nash value of a game and is denoted by $\mathcal{N}(t_*,x_*)$. One can consider a multivalued map taking (t_*,x_*) to $\mathcal{N}(t_*,x_*)$.

The set $\mathcal{N}(t_*,x_*)$ is nonempty under the Isaacs condition [10, 11]. The proof is based on the punishment strategy technique. If the Isaacs condition is not fulfilled, then the Nash equilibrium solution exists in the class of mixed strategies or in the class of the pair counterstrategy/strategy [10].

Below we suppose that the Isaacs condition holds: for all $t \in [t_0,\vartheta_0], x,s \in \mathbb{R}^n$

$$\min_{u\in P}\max_{v\in Q}\langle s, f(t,x,u,v)\rangle = \max_{v\in Q}\min_{u\in P}\langle s, f(t,x,u,v)\rangle.$$

Remark 6.1. If the Isaacs condition does not hold, then one can consider the solution in the class of mixed strategies. To this end, we consider the doubly controlled system

$$\dot{x} = \int_P\int_Q f(t,x,u,v)\nu(\mathrm{d}v)\mu(\mathrm{d}u), \quad t \in [t_0,\vartheta], \quad x \in \mathbb{R}^n, \quad \mu \in \mathrm{rpm}(P), \quad \nu \in \mathrm{rpm}(Q). \tag{6.2}$$

Here μ is a generalized control of Player I, ν is a generalized control of Player II, and $\mathrm{rpm}(P)$ and $\mathrm{rpm}(Q)$ are sets of regular probabilistic measures on P and Q, respectively. We endow the sets $\mathrm{rpm}(P)$ and $\mathrm{rpm}(Q)$ with $*$-weak topology. The obtained topology spaces are compacts. It is easy to show that the Isaacs condition is fulfilled for system (6.2). Henceforth we will not mention the change from system (6.1) to system (6.2).

Consider the zero-sum differential game Γ_1 with its dynamic determined by (6.1) and the payoff determined by $\sigma_1(x(\vartheta_0))$. We assume that Player I wants to maximize $\sigma_1(x(\vartheta_0))$, while Player II is interested in minimizing it. There exists a value of the game Γ_1. Denote it by ω_1. Analogously, consider the zero-sum differential game with dynamics (6.1) and the payoff σ_2. We assume that Player II wants to maximize

$\sigma_2(x(\vartheta_0))$, whereas Player I want to minimize it. Denote the value function of this game by ω_2.

Consider the initial value problem

$$\dot{x} \in \mathcal{F}(t,x) \triangleq \mathrm{co}\{f(t,x,u,v) : u \in P, v \in Q\}, \ \ x(t_*) = x_*.$$

By $\mathrm{Sol}(t_*,x_*)$ denote the set of its solutions.

Proposition 6.1. *Let the multivalued map* $\mathcal{T} : [t_0,\vartheta_0] \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^2)$ *satisfy the following conditions:*

(N1) $\mathcal{T}(\vartheta_0,x) = \{(\sigma_1(x),\sigma_2(x))\}$ *for all* $x \in \mathbb{R}^n$.
(N2) $\mathcal{T}(t,x) \subset [\omega_1(t,x),\infty) \times [\omega_2(t,x),\infty)$ *for all* $(t,x) \in [t_0,\vartheta_0] \times \mathbb{R}^n$.
(N3) *for all* $(t_*,x_*) \in [t_0,\vartheta_0] \times \mathbb{R}^n$, $(J_1,J_2) \in \mathcal{T}(t_*,x_*)$, *there exists a motion* $y(\cdot) \in \mathrm{Sol}(t_*,x_*)$ *such that*

$$(J_1,J_2) \in \mathcal{T}(t,y(t)), \ \ t \in [t_*,\vartheta_0].$$

Then $\mathcal{T}(t,x) \subset \mathcal{N}(t,x)$ *for all* $(t,x) \in [t_0,\vartheta_0] \times \mathbb{R}^n$.

This proposition follows from [10, Theorem 1.4].

Henceforth, we limit our attention to closed multivalued maps. The map $\mathcal{T} : [t_0,\vartheta_0] \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^2)$ is called closed if its graph is closed, i.e., $\mathrm{Cl}[\mathcal{T}] = \mathcal{T}$. Here Cl denotes the closure of the graph:

$$[\mathrm{Cl}\mathcal{T}](t,x) \triangleq \Big\{(J_1,J_2) : \exists \Big\{\Big(t^k,x^k\Big)\Big\}_{k=1}^{\infty} \subset [t_0,\vartheta_0] \times \mathbb{R}^n \ \exists \Big\{\Big(z_1^k,z_2^k\Big)\Big\} \subset \mathbb{R}^2 : \\ \Big(z_1^k,z_2^k\Big) \in \mathcal{T}(t^k,x^k), \ \Big(t^k,x^k\Big) \to (t,x), \ \Big(z_1^k,z_2^k\Big) \to (J_1,J_2), \ \text{as } k \to \infty\Big\}.$$

Let I be an indexing set. Let multivalued maps $\mathcal{T}^\alpha : [t_0,\vartheta_0] \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^2)$, $\alpha \in I$, satisfy conditions (N1)–(N3). Define the map $\mathcal{T}^* : [t_0,\vartheta_0] \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^2)$ by the rule $\mathcal{T}^* = \mathrm{Cl}\mathcal{T}$, where

$$\mathcal{T}(t,x) \triangleq \left[\bigcup_{\alpha \in I} \mathcal{T}^\alpha(t,x)\right].$$

The multivalued map $\mathcal{T}^*$ is closed, has compact images, and satisfies conditions (N1)–(N3). By $\mathcal{T}^+$ denote the closure of the pointwise union of all upper semicontinuous multivalued maps from $[t_0,\vartheta_0] \times \mathbb{R}^n$ to $\mathbb{R}^2$ satisfying conditions (N1)–(N3). It follows [10] that $\mathcal{T}^+(t,x) = \mathcal{N}(t,x)$ for all $(t,x) \in [t_0,\vartheta_0] \times \mathbb{R}^n$.

Condition (N1) is a boundary condition, and condition (N2) is connected with the theory of zero-sum differential games. Further, we formulate condition (N3) in terms of viability theory and obtain the infinitesimal form of this condition.

Theorem 6.1. *Let the map* $\mathcal{T} : [t_0,\vartheta_0] \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^2)$ *be closed. Then condition* (N3) *is equivalent to the following one: for all* $(t_*,x_*) \in [t_0,\vartheta_0] \times \mathbb{R}^n$, $(J_1,J_2) \in \mathcal{T}(t_*,x_*)$ *there exist* $\theta > t_*$ *and* $y(\cdot) \in \mathrm{Sol}(t_*,x_*)$ *such that*

$$(J_1,J_2) \in \mathcal{T}(t,y(t)), \quad t \in [t_*,\theta].$$

Theorem 6.1 is proved in Sect. 6.4.

To obtain the infinitesimal form of condition (N3), we define a derivative of a multivalued map. By dist denote the following planar distance between the point $(J_1,J_2) \in \mathbb{R}^2$ and the set $A \subset \mathbb{R}^2$:

$$\mathrm{dist}[(J_1,J_2),A] \triangleq \inf\{|\zeta_1 - J_1| + |\zeta_2 - J_2| : (\zeta_1,\zeta_2) \in A\}.$$

Define the directional derivative of the multivalued map by the rule

$$\mathrm{D}_H\mathcal{T}(t,x;(J_1,J_2),w) \triangleq \liminf_{\delta\downarrow 0, w'\to w} \frac{\mathrm{dist}[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta w')]}{\delta}.$$

Theorem 6.2. *Let* $\mathcal{T}:[t_0,\vartheta_0]\times\mathbb{R}^n \to \mathcal{P}(\mathbb{R}^2)$ *be closed. Then condition* (N3) *at the position* $(t_*,x_*) \in [t_0,\vartheta_0]\times\mathbb{R}^n$ *is equivalent to the following one:*

$$\sup_{(J_1,J_2)\in\mathcal{T}(t_*,x_*)} \inf_{w\in\mathcal{F}(t_*,x_*)} \mathrm{D}_H\mathcal{T}(t_*,x_*;(J_1,J_2),w) = 0. \tag{6.3}$$

Theorem 6.2 is proved in Sect. 6.4.

Introduce the set

$$\widehat{\partial}\mathcal{T}(t_*,x_*;(J_1;J_2)) \triangleq \{w : \mathrm{D}_H\mathcal{T}(t_*,x_*;(J_1,J_2),w) = 0\}.$$

Condition (6.3) can be formulated in the following way:

$$\widehat{\partial}\mathcal{T}(t_*,x_*;(J_1;J_2)) \cap \mathcal{F}(t_*,x_*) = 0, \quad \forall (J_1,J_2) \in \mathcal{T}(t_*,x_*).$$

This statement follows from the proof of Theorem 6.2.

Let us introduce a sufficient condition for the function $(c_1,c_2):[t_0,\vartheta_0]\times\mathbb{R}^n \to \mathbb{R}^2$ to provide a Nash equilibrium. Denote

$$H_1(t,x,s) \triangleq \max_{u\in P}\min_{v\in Q}\langle s, f(t,x,u,v)\rangle,$$

$$H_2(t,x,s) \triangleq \max_{v\in Q}\min_{u\in P}\langle s, f(t,x,u,v)\rangle.$$

Let $(c_1,c_2):[t_0,\vartheta_0]\times\mathbb{R}^n \to \mathbb{R}^2$, $(t,x) \in [t_0,\vartheta_0]\times\mathbb{R}^n$, $w \in \mathbb{R}^n$. Define a modulus derivative at the position (t,x) in the direction $w \in \mathbb{R}^n$ by the rule

$$\mathrm{d}_{abs}(c_1,c_2)(t,x;w) \triangleq \liminf_{\delta\downarrow 0, w'\to w} \frac{|c_1(t+\delta,x+\delta w') - c_1(t,x)| + |c_2(t+\delta,x+\delta w') - c_2(t,x)|}{\delta}.$$

Corollary 6.1. *Suppose that the function* $(c_1,c_2):[t_0,\vartheta_0]\times\mathbb{R}^n\to\mathbb{R}^2$ *is continuous,* $(c_1(\vartheta_0,\cdot),c_2(\vartheta_0,\cdot)) = (\sigma_1(\cdot),\sigma_2(\cdot))$*, for each* i *the function* c_i *is a viscosity supersolution of the equation*

$$\frac{\partial c_i}{\partial t}+H_i(t,x,\nabla c_i)=0, \tag{6.4}$$

and for all $(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n$

$$\inf_{w\in\mathcal{F}(t,x)} \mathrm{d}_{abs}(c_1,c_2)(t,x;w)=0.$$

Then for all $(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n$ *the couple* $(c_1(t,x),c_2(t,x))$ *is a Nash equilibrium payoff of the game.*

Corollary 6.1 follows from the definition of modulus derivative and the property of the upper solution of equation (6.4) [14]: $\omega_i(t,x)\le c_i(t,x)$ for all $(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n$.

Let us show that the proposed method is a generalization of the method based on the system of Hamilton–Jacobi equations. This method provides a Nash solution in the class of continuous strategies [1].

Proposition 6.2. *Let the function* $(\varphi_1,\varphi_2):[t_0,\vartheta_0]\times\mathbb{R}^n\to\mathbb{R}^2$ *be differentiable, and* $(\varphi_1(\vartheta_0,\cdot),\varphi_2(\vartheta_0,\cdot))=(\sigma_1(\cdot),\sigma_2(\cdot))$*. Suppose that the function* (φ_1,φ_2) *satisfies the following condition: for all positions* $(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n$ *there exist* $u^n\in P$*,* $v^n\in Q$ *such that*

$$\max_{u\in P}\left\langle\nabla\varphi_1(t,x),f(t,x,u,v^n)\right\rangle=\left\langle\nabla\varphi_1(t,x),f(t,x,u^n,v^n)\right\rangle, \tag{6.5}$$

$$\max_{v\in Q}\left\langle\nabla\varphi_2(t,x),f(t,x,u^n,v)\right\rangle=\left\langle\nabla\varphi_2(t,x),f(t,x,u^n,v^n)\right\rangle \tag{6.6}$$

$$\frac{\partial\varphi_i(t,x)}{\partial t}+\left\langle\nabla\varphi_i(t,x),f(t,x,u^n,v^n)\right\rangle=0,\ \ i=1,2. \tag{6.7}$$

Then the function (φ_1,φ_2) *satisfies the conditions of Corollary 6.1.*

This proposition is proved in Sect. 6.4.

If for each position $(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n$ and the pair of directions $s_1,s_2\in\mathbb{R}^n$ there exists the pair of controls (u^n,v^n) such that

$$\max_{u\in P}\left\langle s_1,f(t,x,u,v^n)\right\rangle=\left\langle s_1,f(t,x,u^n,v^n)\right\rangle,$$

$$\max_{v\in Q}\left\langle s_2,f(t,x,u^n,v)\right\rangle=\left\langle s_2,f(t,x,u^n,v^n)\right\rangle,$$

then the Hamiltonians $\mathcal{H}_i$ are well defined by the rule

$$\mathcal{H}_i(t,x,s_1,s_2)\triangleq\left\langle s_i,f(t,x,u^n,v^n)\right\rangle,\ \ i=1,2.$$

In this case, condition (6.7) is equal to the following one: (φ_1, φ_2) is a solution of the system

$$\frac{\partial \varphi_i}{\partial t} + \mathcal{H}_i(t, x, \nabla\varphi_1, \nabla\varphi_2) = 0, \;\; i = 1, 2.$$

6.3 Example

Consider the non-zero-sum differential game with the dynamic

$$\begin{cases} \dot{x} = u, \\ \dot{y} = v, \end{cases} \tag{6.8}$$

$t \in [0,1]$, $u, v \in [-1,1]$. Payoffs are determined by the formulas $\sigma_1(x,y) \triangleq -|x-y|$, $\sigma_2(x,y) \triangleq y$. We recall that each player wants to maximize his payoff.

To determine the multivalued map $\mathcal{N} : [0,1] \times \mathbb{R}^2 \to \mathcal{P}(\mathbb{R}^2)$, we use auxiliary multivalued maps $\mathcal{S}_i : [0,1] \times \mathbb{R}^2 \to \mathcal{P}(\mathbb{R})$ such that

$$\mathcal{S}_i(t, x_*, y_*) \triangleq \big\{ z \in \mathbb{R} : \omega_i(t, x_*, y_*) \leq z \leq c_i^+(t, x_*, y_*) \big\}.$$

Here

$$c_i^+(t, x, y) \triangleq \sup_{u \in \mathcal{U}, v \in \mathcal{V}} \sigma_i \left(x + \int_t^{\vartheta_0} u(\xi)\, \mathrm{d}\xi, y + \int_t^{\vartheta_0} v(\xi)\, \mathrm{d}\xi \right).$$

Obviously,

$$\mathcal{N}(t, x_*, y_*) \subset \mathcal{S}_1(t, x_*, y_*) \times \mathcal{S}_2(t, x_*, y_*). \tag{6.9}$$

First we determine the map $\mathcal{S}_2$. The value function of the game Γ_2 is equal to $\omega_2(t, x_*, y_*) = y_* + (1-t)$. In addition, $c_2^+(t, x_*, y_*) = y_* + (1-t)$. Consequently,

$$\mathcal{S}_2(t, x_*, y_*) = \{ y_* + (1-t) \}. \tag{6.10}$$

Let us determine the set $\mathcal{S}_1$. The programmed iteration method [7] yields that

$$\omega_1(t, x_*, y_*) = -|x_* - y_*|.$$

Moreover,

$$c_1^+(t, x_*, y_*) = \min\big\{ -|x_* - y_*| + 2(1-t), 0 \big\}.$$

We obtain that

$$\mathcal{S}_1(t, x_*, y_*) = \big[\omega_1(t, x_*, y_*), c_1^+(t, x_*, y_*) \big]. \tag{6.11}$$

Now we determine the map $\mathcal{N}(t,x_*,y_*)$. The linearity of the right-hand side of (6.8) and the convexity of control spaces yield that any measurable control functions $u(\cdot)\in\mathcal{U}$ and $v(\cdot)\in\mathcal{V}$ can be substituted by the constant controls $u\in P$, $v\in Q$. We have that for all $(J_1,J_2)\in\mathcal{S}_1(t,x_*,y_*)\times\mathcal{S}_2(t,x_*,y_*)$

$$\begin{aligned}&\mathrm{D}_H\mathcal{N}(t,x_*,y_*;(J_1,J_2),(u,v))\\&\quad\geq\liminf_{\delta\downarrow 0,v'\to v}\frac{|y_*+\delta v'+(1-t-\delta)-y_*-(1-t)|}{\delta}=|v-1|.\end{aligned}$$

Therefore, if $\mathrm{D}_H\mathcal{N}(t,x_*,y_*;(J_1,J_2),w)=0$ for a couple $w=(u,v)$, then $v=1$.

First we consider the case $y_*\geq x_*$. Let $(J_1,J_2)\in\mathcal{N}(t,x_*,y_*)$. There exists a motion $(x(\cdot),y(\cdot))\in\mathrm{Sol}(t,x_*,y_*)$ such that $(J_1,J_2)\in\mathcal{N}(\theta,x(\theta),y(\theta))$, $\theta\in[t,1]$. Since $\mathrm{D}_H\mathcal{N}(t,x_*,y_*;(J_1,J_2),w)=0$ only if $v=1$, there exists $u\in[-1,1]$ such that $x(1)=x_*+u(1-t)$, $y(1)=y_*+(1-t)$. Consequently, $y(1)\geq x(1)$. From condition (N2) we obtain that

$$\begin{aligned}J_1=\sigma_1(x(1),y(1))&=-y(1)+x(1)=-y_*-(1-t)+x_*+u(1-t)\\&=-y_*+x_*+(1-t)(u-1)\leq -y_*+x_*=-|x_*-y_*|.\end{aligned}$$

The equality is achieved only if $u=1$. Condition (N1) yields that the inclusion

$$\mathcal{N}(t,x_*,y_*)\subset\{(-|x_*-y_*|,y_*+(1-t))\}$$

is fulfilled. Substituting the value $(1,1)$ for w in the formula for $\mathrm{D}_H\mathcal{N}(t,x_*,y_*;(-|x_*-y_*|,y_*+(1-t)),w)$ we claim that for $y_*\geq x_*$

$$\mathcal{N}(t,x_*,y_*)=\big\{(-|x_*-y_*|,y_*+(1-t))\big\}.$$

Now let $y_*<x_*$. We shall show that

$$\begin{aligned}\mathcal{N}(t,x_*,y_*)&=\mathcal{S}_1(t,x_*,y_*)\times\mathcal{S}_2(t,x_*,y_*)\\&=\Big[-|x_*-y_*|,\min\big\{-|x_*-y_*|+2(1-t),0\big\}\Big]\times\big\{y_*+(1-t)\big\}.\end{aligned}$$

Clearly, conditions (N1) and (N2) hold for this map. Let γ_0 be a maximal number of segment $[0,2]$ such that $-|x_*-y_*|+\gamma_0(1-t)\leq 0$. If $(J_1,J_2)\in\mathcal{N}(t,x_*,y_*)$, then $J_2=y_*+(1-t)$, $J_1=-|x_*-y_*|+d(1-t)$ for some $d\in[0,\gamma_0]$. Let us prove that there exists a number $\delta>0$ with the property

$$(J_1,J_2)\in\mathcal{N}(t+\delta,x_*+\delta u,y_*+\delta)\tag{6.12}$$

for $u = 1-d$. It is sufficient to prove that

$$J_1 \in \Big[y_* - x_* + \delta d, \min\big\{y_* - x_* + \delta d + 2(1-t-\delta), 0\big\}\Big].$$

Indeed, $y_* - x_* + d(1-t) \geq y_* - x_* + \delta d$ for $\delta < (1-t)$. Since $d \leq \gamma_0$, we obtain that

$$J_1 = y_* - x_* + d(1-t) \leq y_* - x_* + \delta d + \gamma_0(1-t-\delta).$$

Also,

$$y_* - x_* + \delta d + \gamma_0(1-t-\delta) \leq \min\big\{y_* - x_* + d\delta + 2(1-t-\delta); 0\big\}.$$

Actually, since $\gamma_0 \leq 2$, the following inequality is fulfilled:

$$y_* - x_* + \delta d + \gamma_0(1-t-\delta) \leq y_* - x_* + d\delta + 2(1-t-\delta).$$

Moreover, $y_* - x_* + \delta d + \gamma_0(1-t-\delta) \leq \delta d - \gamma_0\delta \leq 0$. Thus the condition

$$J_1 \leq \min\big\{y_* - x_* + \delta d + 2(1-t-\delta), 0\big\} = y_* - x_* + \delta d + \gamma_1(1-t-\delta)$$

is valid also. It follows from (6.12) that

$$\mathrm{D}_H \mathcal{N}(t, x_*, y_*; (J_1, J_2), (1-d, 1)) = 0.$$

Since $\mathcal{N}(t, x_*, y_*)$ coincide with the set $\mathcal{S}_1(t, x_*, y_*) \times \mathcal{S}_2(t, x_*, y_*)$ in this case, we claim that the set $\mathcal{N}(t, x_*, y_*)$ is a Nash value of the game at the position (t, x_*, y_*).

Let us compare the obtained result with the method based on the system of Hamilton–Jacobi equations [1]. In the considered case the system of equations is given by

$$\begin{cases} \dfrac{\partial \varphi_1}{\partial t} + \dfrac{\partial \varphi_1}{\partial x} u^n(t,x,y) + \dfrac{\partial \varphi_1}{\partial y} v^n(t,x,y) = 0, \\[2ex] \dfrac{\partial \varphi_2}{\partial t} + \dfrac{\partial \varphi_2}{\partial x} u^n(t,x,y) + \dfrac{\partial \varphi_2}{\partial y} v^n(t,x,y) = 0. \end{cases} \tag{6.13}$$

Here the values $u^n(t,x,y)$ and $v^n(t,x,y)$ are determined by the following conditions:

$$\frac{\partial \varphi_1(t,x,y)}{\partial x} u^n(t,x,y) = \max_{u \in P}\left[\frac{\partial \varphi_1(t,x,y)}{\partial x} u\right],$$

$$\frac{\partial \varphi_2(t,x,y)}{\partial y} v^n(t,x,y) = \max_{v \in Q}\left[\frac{\partial \varphi_2(t,x,y)}{\partial y} v\right].$$

It follows from Proposition 6.2 that if a pair of functions (φ_1, φ_2) is a solution of system (6.13), then $\varphi_2(t,x,y) = y + (1-t)$. Thus, $v^n(t,x,y) = 1$. Consequently, system (6.13) reduces to the equation

$$\frac{\partial \varphi_1}{\partial t} + \left| \frac{\partial \varphi_1}{\partial x} \right| + \frac{\partial \varphi_1}{\partial y} = 0. \tag{6.14}$$

By [14, Theorem 5.6] we obtain that the function

$$\varphi_1(t,x,y) = \begin{cases} x - y, & x \leq y, \\ -x + y + 2(1-t), & x > y, -x + y + 2(1-t) < 0, \\ 0, & x > y, -x + y + 2(1-t) \geq 0 \end{cases}$$

is a minimax (viscosity) solution of Eq. (6.14). Indeed if φ_1 is smooth at (t,x,y), then Eq. (6.14) is fulfilled in the classical sense. On the planes $\{(t,x,y) : x = y\}$ and $\{(t,x,y) : -x + y + 2(1-t) = 0\}$ we have that the Clarke subdifferential is the convex hull of two limits of partial derivatives of the function φ_1. By the well-known properties of subdifferentials and superdifferentials, the continuity, and the positive homogeneity of Eq. (6.14), we obtain that φ_1 satisfies conditions U4 and L4 of [14].

The function φ_1 is nonsmooth. Since the minimax solution is unique, and any classical solution is minimax, we claim that system (6.13) has no classical solution. One may obtain from the formulae for $\mathcal{N}(t,x,y)$ that $(\varphi_1(t,x,y), \varphi_2(t,x,y)) \in \mathcal{N}(t,x,y)$. Moreover,

$$\varphi_1(t,x,y) = \max\Big\{ J_1 \in \mathbb{R} : \exists J_2 \in \mathbb{R} \ \ (J_1, J_2) \in \mathcal{N}(t,x,y) \Big\},$$

$$\{\varphi_2(t,x,y)\} = \Big\{ J_1 \in \mathbb{R} : \exists J_2 \ \ (J_1, J_2) \in \mathcal{N}(t,x,y) \Big\}.$$

In other words, the value $(\varphi_1(t,x,y), \varphi_2(t,x,y))$ is the maximal Nash equilibrium payoff of the game at the position (t,x,y).

One can check that the pair of functions (φ_1, φ_2) satisfies the conditions of Corollary 6.1. Simultaneously, there exists a family of functions satisfying the conditions of Corollary 6.1. Actually, for $\gamma \in [0,2]$ define

$$c_1^\gamma(t, x_*, y_*) = \begin{cases} -|x_* - y_*|, & y_* \geq x_*, \\ \min\big\{ -|x_* - y_*| + \gamma(1-t); 0 \big\}, & y_* < x_*, \end{cases}$$

$$c_2^\gamma(t, x_*, y_*) = y_* + (1-t).$$

Let us show that the pair of functions (c_1^γ, c_2^γ) satisfies the conditions of Corollary 6.1. We have that in our case,

$$H_1(t,x,y,s_x,s_y) = |s_x| - |s_y|, \ \ H_2(t,x,y,s_x,s_y) = |s_y| - |s_x|.$$

First we prove that the functions c_i^γ are the supersolutions of equations (6.4). By [14, condition U4] it suffices to show that for all $(t,x,y) \in [t_0,\vartheta_0] \times \mathbb{R}^2$ $(a,s_x,s_y) \in D^- c_i^\gamma(t,x,y)$ the following inequality holds:

$$a + H_i(s_x,s_y) \leq 0, \quad i = 1,2. \tag{6.15}$$

Here D^- denotes the subdifferential [14, (6.10)]. The computing of subdifferentials gives that

$$D^- c_1^\gamma(t,x,y) = \begin{cases} \{(0,-1,1)\}, & y > x, \\ \{(0,0,0)\}, & y < x < y + \gamma(1-t), \\ \{(-\gamma,-1,1)\}, & x > y + \gamma(1-t), \\ \{(0,\lambda,-\lambda) : \lambda \in [0,1]\}, & x = y, \\ \{(-\lambda\gamma,-\lambda,\lambda) : \lambda \in [0,1]\}, & x = y + \gamma(1-t), \end{cases}$$

$$D^- c_2^\gamma(t,x,y) = \{(-1,0,1)\}.$$

Substituting the values of subdifferentials, we obtain that (6.15) is valid for $i = 1,2$.

Also $c_i(1,x_*,y_*) = \sigma_i(x_*,y_*)$. Moreover, $\mathrm{d}_{abs}(c_1^\gamma,c_2^\gamma)(t,x_*,y_*;1-d,1) = 0$ for

$$d = \begin{cases} 0, & y_* \geq x_*; \\ \max\Big\{r \in [0,\gamma] : -|x_* - y_*| + r(1-t) \leq 0\Big\}, & y_* < x_*. \end{cases}$$

Note that $(\varphi_1,\varphi_2) = (c_1^2,c_2^2)$.

6.4 Weak Invariance of the Set of Values

In this section, the statements formulated in Sect. 6.2 are proved.

Proof of Theorem 6.1. If condition (N3) holds, then one can set $\theta = \vartheta_0$.

Now suppose that for all (t_*,x_*), $(J_1,J_2) \in \mathcal{T}(t_*,x_*)$ there exist $\theta \in [t_*,\vartheta_0]$ and a motion $y(\cdot) \in \mathrm{Sol}(t_*,x_*)$ such that the following condition is fulfilled:

$$(J_1,J_2) \in \mathcal{T}(t,y(t)), \quad t \in [t_*,\theta]. \tag{6.16}$$

Let Θ be a set of moments θ satisfying condition (6.16) for some $y(\cdot) \in \mathrm{Sol}(t_*,x_*)$. Denote $\tau \triangleq \sup\Theta$. We have that $\tau \in \Theta$. Indeed, let a sequence $\{\theta_k\}_{k=1}^\infty \subset \Theta$ tend to τ. One can assume that $\theta_k < \theta_{k+1} \leq \tau$. For every k condition (6.16) is valid under $\theta = \theta_k$, $y(\cdot) = y_k(\cdot) \in \mathrm{Sol}(t_*,x_*)$. The compactness of $\mathrm{Sol}(t_*,x_*)$ yields that $y_k(\cdot) \to y^*(\cdot)$, as $k \to \infty$; here $y^*(\cdot)$ is an element of $\mathrm{Sol}(t_*,x_*)$.

The closeness of the map $\mathcal{T}$ gives that for all k $(J_1,J_2)\in\mathcal{T}(t,y^*(t))$, $t\in[t_*,\theta_k]$. By the same argument we claim that $(J_1,J_2)\in\mathcal{T}(\tau,y^*(\tau))$. Denote $x^*=y^*(\tau)$.

Let us show that $\tau=\vartheta_0$. If $\tau<\vartheta_0$, then there exist a motion $\hat{y}(\cdot)\in\mathrm{Sol}(\tau,x^*)$ and a moment $\theta'>\tau$ such that $(J_1,J_2)\in\mathcal{T}(t,\hat{y}(t))$, $t\in[\tau,\theta']$. Consider a motion

$$\tilde{y}(t)\triangleq\begin{cases} y^*(t), & t\in[t_*,\tau],\\ \hat{y}(t), & t\in[\tau,\theta'].\end{cases}$$

By the definition of θ' it follows that (6.16) is valid under $\theta=\theta'$, $y(\cdot)=\tilde{y}(\cdot)$. Thus $\theta'\in\Theta$, but this contradicts with the choice of τ. Consequently, $\tau=\vartheta_0$, and condition (N3) holds. □

Proof of Theorem 6.2. Let us introduce a graph of the map $\mathcal{T}$

$$\mathrm{gr}\mathcal{T}\triangleq\{(t,x,J_1,J_2):(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n,(J_1,J_2)\in\mathcal{T}(t,x)\}.$$

One can reformulate the condition of Theorem 6.1 in the following way: the graph of $\mathcal{T}$ is weakly invariant under the differential inclusion

$$\begin{pmatrix}\dot{x}\\ \dot{J}_1\\ \dot{J}_2\end{pmatrix}\in\widehat{\mathcal{F}}(t,x)\triangleq\mathrm{co}\left\{\begin{pmatrix} f(t,x,u,v)\\ 0\\ 0\end{pmatrix}:u\in P,v\in Q\right\}.$$

The condition of weak invariance of the multivalued map $\mathcal{T}$ under differential inclusion $\widehat{\mathcal{F}}$ is equivalent [9, 14] to the condition

$$D_t(\mathrm{gr}\mathcal{T})(t,x,J_1,J_2)\cap\widehat{\mathcal{F}}(t,x)\neq\varnothing \tag{6.17}$$

for all $(t,x)\in[t_0,\vartheta_0]\times\mathbb{R}^n$, $(J_1,J_2)\in\mathcal{T}(t,x)$. Here D_t denotes the right-hand derivative in t. It is defined in the following way. Let $\mathcal{G}\subset[t_0,\vartheta_0]\times\mathbb{R}^m$, $\mathcal{G}[t]$ denote a section of G by t:

$$\mathcal{G}[t]\triangleq\{w\in\mathbb{R}^m:(t,x)\in\mathcal{G}\},$$

and the symbol d denote the Euclidian distance between a point and a set. Following [9, 14] set

$$(D_t\mathcal{G})(t,y)\triangleq\left\{h\in\mathbb{R}^m:\liminf_{\delta\to 0}\frac{\mathrm{d}(y+\delta h;\mathcal{G}[t+\delta])}{\delta}=0\right\}.$$

Let us show that conditions (6.3) and (6.17) are equivalent.

Condition (6.3) means that for every couple $(J_1,J_2)\in\mathcal{T}(t,x)$ the following condition holds:

$$\inf_{w\in\mathcal{F}(t,x)}\ \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0}\frac{\mathrm{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w+\gamma))\big]}{\delta}=0.$$

The lower boundary by w in the formula

$$\inf_{w\in\mathcal{F}(t,x)} \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w+\gamma))\big]}{\delta}$$

is attained for all $(J_1,J_2)\in\mathcal{T}(t,x)$. Indeed, let $\{w^r\}_{r=1}^{\infty}$ be a minimizing sequence. By the compactness of $\mathcal{F}(t,x)$ one can assume that $w^r\to w^*$, $r\to\infty$, $w^*\in\mathcal{F}(t,x)$. Let us show that

$$\begin{aligned}\tilde{b} &\triangleq \inf_{w\in\mathcal{F}(t,x)} \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w+\gamma))\big]}{\delta}\\ &= \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w^*+\gamma))\big]}{\delta}. \end{aligned} \tag{6.18}$$

Indeed, for every $r\in\mathbb{N}$ there exist sequences $\{\delta^{r,k}\}_{k=1}^{\infty}$, $\{\gamma^{r,k}\}_{k=1}^{\infty}$ such that $\delta^{r,k}, \|\gamma^{r,k}\|\to 0$, as $k\to\infty$, and

$$\begin{aligned} b^r &\triangleq \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,t+\delta(w^r+\gamma))\big]}{\delta}\\ &= \lim_{k\to\infty} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta^{r,k},t+\delta^{r,k}(w^r+\gamma^{r,k}))\big]}{\delta^{r,k}}. \end{aligned}$$

Let $\hat{k}(r)$ be a number such that

$$\delta^{r,\hat{k}(r)}, \|\gamma^{r,\hat{k}(r)}\|, \left| \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta^{r,\hat{k}(r)},t+\delta^{r,\hat{k}(r)}(w^r+\gamma^{r,\hat{k}(r)}))\big]}{\delta^{r,\hat{k}(r)}} - b^r \right| < 2^{-r}.$$

Set $\hat{\delta}^r \triangleq \delta^{r,\hat{k}(r)}$, $\hat{\gamma}^r \triangleq \gamma^{r,\hat{k}(r)} + w^r - w^*$. Note that $\hat{\delta}^r, \|\hat{\gamma}^r\|\to 0$, $r\to\infty$.

We have that

$$\begin{aligned} &\inf_{w\in\mathcal{F}(t,x)} \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w+\gamma))\big]}{\delta}\\ &\le \liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w^*+\gamma))\big]}{\delta}\\ &\le \lim_{r\to\infty} \frac{\operatorname{dist}\big[(J_1,J_2),\mathcal{T}(t+\hat{\delta}^r,x+\hat{\delta}^r(w^*+\hat{\gamma}^r))\big]}{\hat{\delta}^r}. \end{aligned} \tag{6.19}$$

Further,

$$\begin{aligned}
&\frac{\mathrm{dist}\big[(J_1,J_2),\mathcal{T}(t+\hat\delta^r,x+\hat\delta^r(w^*+\hat\gamma^r))\big]}{\hat\delta^r}\\
&\quad=\frac{\mathrm{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta^{r,\hat k(r)},x+\delta^{r,\hat k(r)}(w^*+\gamma^{r,\hat k(r)}+w^r-w^*))\big]}{\delta^{r,\hat k(r)}}\\
&\quad=\frac{\mathrm{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta^{r,\hat k(r)},x+\delta^{r,\hat k(r)}(w^k+\gamma^{r,\hat k(r)}))\big]}{\delta^{r,\hat k(r)}}\\
&\quad\leq b^r+2^{-r}\to\tilde b,\quad r\to\infty.
\end{aligned}$$

We have that in (6.19) the right- and left-hand sides are equal. This means that condition (6.18) is valid.

Thus, condition (6.3) is equivalent to the following one: for all $(J_1,J_2)\in\mathcal{T}(t,x)$ there exists $w\in\mathcal{F}(t,x)$ such that

$$\begin{aligned}
&\liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0}\frac{\mathrm{dist}\big[(J_1,J_2),\mathcal{T}(t+\delta,x+\delta(w+\gamma))\big]}{\delta}\\
&=\liminf_{\delta\downarrow 0,\gamma\in\mathbb{R}^n,\|\gamma\|\downarrow 0}\inf\left\{\frac{|\zeta_1-J_1|+|\zeta_2-J_2|}{\delta}:(\zeta_1,\zeta_2)\in\mathcal{T}(t+\delta,x+\delta(w+\gamma))\right\}=0.
\end{aligned}\tag{6.20}$$

Now let us prove that this condition is equivalent to condition (6.17).

First we assume that condition (6.17) is valid. This means that there exist sequences $\{\delta^k\}_{k=1}^\infty\subset\mathbb{R}$, $\{\gamma^k\}_{k=1}^\infty\subset\mathbb{R}^n$, $\{\varepsilon_1^k\}_{k=1}^\infty,\{\varepsilon_2^k\}_{k=1}^\infty\subset\mathbb{R}$ such that

- $\delta^k,\|\gamma^k\|,\varepsilon_1^k,\varepsilon_2^k\to 0$, as $k\to\infty$;
- $\left(t+\delta^k,x+\delta^k(w+\gamma^k),J_1+\delta^k\varepsilon_1^k,J_2+\delta^k\varepsilon_2^k\right)\in\mathrm{gr}\mathcal{T}$.

One can reformulate the second condition as

$$(J_1+\delta^k\varepsilon_1^k,J_2+\delta^k\varepsilon_2^k)\in\mathcal{T}(t+\delta^k,t+\delta^k(w+\gamma^k)).$$

Thus,

$$\inf\left\{\frac{|\zeta_1-J_1|+|\zeta_2-J_2|}{\delta^k}:(\zeta_1,\zeta_2)\in\mathcal{T}(t+\delta^k,x+\delta(w+\gamma^k))\right\}=\varepsilon_1^k+\varepsilon_2^k.$$

By the choice $\{\varepsilon_1^k\}$, $\{\varepsilon_2^k\}$ we obtain that condition (6.20) holds.

Now let condition (6.20) be fulfilled, and prove that (6.17) is valid. Indeed, let $\{\delta^k\}_{k=1}^\infty$, $\{\gamma\}_{k=1}^\infty$ be a minimizing sequence. By the compactness of the sets $\mathcal{T}\big(t+\delta^k,x+\delta^k(w+\gamma^k)\big)$ for each k there exist ε_1^k and ε_2^k such that

$$\big(J_1+\delta^k\varepsilon_1^k,J_2+\delta^k\varepsilon_2^k\big)\in\mathcal{T}(t,x+\delta^k(w+\gamma^k)).$$

It follows from (6.20) that $\varepsilon_1^k, \varepsilon_2^k \to 0$, $k \to \infty$. Let us estimate $\mathrm{d}((w+\delta^k w, J_1, J_2), \mathrm{gr}\mathcal{T}[t+\delta^k])$. We have that

$$\left(t+\delta^k, x+\delta^k(w+\gamma^k), J_1+\delta^k\varepsilon_1^k, J_2+\delta^k\varepsilon_2^k\right) \in \mathrm{gr}\mathcal{T}.$$

Consequently,

$$\mathrm{d}\Big((w+\delta^k w, J_1, J_2), \mathrm{gr}\mathcal{T}[t+\delta^k]\Big) \leq \delta^k\sqrt{\|\gamma^k\|^2+(\varepsilon_1^k)^2+(\varepsilon_2^k)^2}.$$

The convergence $\delta^k, \|\gamma^k\|, \varepsilon_1^k, \varepsilon_2^k \to 0$ as $k \to \infty$ yields the equality

$$\begin{pmatrix} w \\ 0 \\ 0 \end{pmatrix} \in D_t(\mathrm{gr}\mathcal{T})(t,x,J_1,J_2).$$

Since $\widehat{\mathcal{F}}(t,x) = \mathcal{F} \times \{(0,0)\}$, we claim that (6.17) is fulfilled. □

Proof of Proposition 6.2. It follows from (6.5) and the Isaacs condition that

$$\left\langle \nabla\varphi_1(t,x), f(t,x,u^n,v^n)\right\rangle \geq \max_{u\in P}\min_{v\in Q}\left\langle \nabla\varphi_1(t,x), f(t,x,u,v)\right\rangle = H_1(t,x,\nabla\varphi_1(t,x)).$$

Analogously, it follows from (6.6) and the Isaacs condition that

$$\left\langle \nabla\varphi_2(t,x), f(t,x,u^n,v^n)\right\rangle \geq \max_{v\in Q}\min_{u\in P}\left\langle \nabla\varphi_2(t,x), f(t,x,u,v)\right\rangle = H_2(t,x,\nabla\varphi_2(t,x)).$$

Therefore, using (6.7) we claim that

$$\frac{\partial\varphi_i(t,x)}{\partial t} + H_i(t,x,\nabla\varphi_i(t,x)) \leq 0, \ \ i=1,2.$$

Since the function φ_i is differentiable, its subdifferential at the position (t,x) is equal to $\{\partial\varphi_1(t,x)/\partial t, \nabla\varphi_1(t,x)\}$. Consequently, the function φ_1 is the upper solution of Eq. (6.4) for $i=1$ [14, Condition (U4)]. Analogously, the function φ_2 is the upper solution of Eq. (6.4) for $i=2$.

Now let us show that $\mathrm{d}_{abs}(\varphi_1,\varphi_2)(t,x;w) = 0$ for $w \in \mathcal{F}(t,x)$. Put $w = f(t,x,u^n,v^n)$. Indeed,

$$\mathrm{d}_{abs}(\varphi_1,\varphi_2)(t,x;w) = \liminf_{\delta\downarrow 0, \|\gamma\|\to 0} \frac{|\varphi_1(t+\delta,x+\delta(w+\gamma))-\varphi_1(t,x)|+|\varphi_2(t+\delta,x+\delta(w+\gamma))-\varphi_2(t,x)|}{\delta}.$$

Let $\{\delta^k\}_{k=1}^{\infty} \subset \mathbb{R}$, $\{\gamma^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ be a minimizing sequence. Then

$$\mathrm{d}_{abs}(\varphi_1,\varphi_2)(t,x;w) = \lim_{k\to\infty}$$
$$\frac{|\varphi_1(t+\delta^k, x+\delta^k(w+\gamma^k)) - \varphi_1(t,x)| + |\varphi_2(t+\delta^k, x+\delta(w+\gamma^k)) - \varphi_2(t,x)|}{\delta^k}$$
$$= \lim_{k\to\infty}\frac{1}{\delta^k}\left[\left|\frac{\partial\varphi_1(t,x)}{\partial t}\delta^k + \langle\nabla\varphi_1(t,x), \delta^k(w+\gamma^k)\rangle + o(\delta^k)\right|\right.$$
$$\left. + \left|\frac{\partial\varphi_2(t,x)}{\partial t}\delta^k + \langle\nabla\varphi_2(t,x), \delta^k(w+\gamma^k)\rangle + o(\delta^k)\right|\right]$$
$$= \left|\frac{\partial\varphi_1(t,x)}{\partial t} + \langle\nabla\varphi_1(t,x), w\rangle\right| + \left|\frac{\partial\varphi_2(t,x)}{\partial t} + \langle\nabla\varphi_2(t,x), w\rangle\right|.$$

By choice of $w = f(t,x,u^n,v^n)$ and condition (6.7) we have that

$$\frac{\partial\varphi_1(t,x)}{\partial t} + \langle\nabla\varphi_1(t,x), w\rangle = \frac{\partial\varphi_2(t,x)}{\partial t} + \langle\nabla\varphi_2(t,x), w\rangle = 0.$$

Thus $\mathrm{d}_{abs}(\varphi_1,\varphi_2)(t,x;w) = 0$. □

Acknowledgements This work was supported by the Russian Foundation for Basic Research (Grant No. 09-01-00436-a), a grant of the president of the Russian Federation (Project MK-7320.2010.1), and the Russian Academy of Sciences Presidium Programs of Fundamental Research, Mathematical Theory of Control.

References

1. Basar, T., Olsder G.J.: Dynamic Noncooperative Game Theory. SIAM, Philadelphia (1999)
2. Bressan, A., Shen, W.: Semi-cooperative strategies for differential games. Int. J. Game Theory **32**, 561–59 (2004)
3. Bressan, A., Shen, W.: Small BV solutions of hyperbolic noncooperative differential games. SIAM J. Control Optim. **43**, 194–215 (2004)
4. Cardaliaguet, P.: On the instability of the feedback equilibrium payoff in a nonzero-sum differential game on the line. Ann. Int. Soc. Dyn. Games **9**, 57–67 (2007)
5. Cardaliaguet, P., Buckdahn, R., Rainer, C.: Nash equilibrium payoffs for nonzero-sum stochastic differential games. SIAM J. Control Optim. **43**, 624–642 (2004)
6. Cardaliaguet, P., Plaskacz, S.: Existence and uniqueness of a Nash equilibrium feedback for a simple nonzero-sum differential game. Int. J. Game Theory **32**, 33–71 (2003)
7. Chentsov, A.G.: On a game problem of converging at a given instant time. Mat. USSR Sb. **28**, 353–376 (1976)
8. Chistyakov, S.V.: On noncooperative differential games. Dokl. AN USSR **259**, 1052–1055 (1981) (in Russian)
9. Guseinov, H.G., Subbotin, A.I., Ushakov, V.N.: Derivatives for multivalued mappings with applications to game-theoretical problems of control. Problems Control Inform. Theory **14**, 155–167 (1985)

10. Kleimenov, A.F.: Non zero-sum differential games. Ekaterinburg, Nauka (1993) (in Russian)
11. Kononenko, A.F.: On equilibrium positional strategies in nonantagonistic differential games. Dokl. AN USSR **231**, 285–288 (1976) (in Russian)
12. Krasovskii, N.N., Subbotin, A.I.: Game-Theoretical Control Problems. Springer, New York (1988)
13. Olsder, G.J.: On open- and closed-loop bang-bang control in nonzero-sum differential games. SIAM J. Control Optim. **49**(4), 1087–1106 (2001)
14. Subbotin, A.I.: Generalized Solutions of First-order PDEs. The Dynamical Perspective. Birkhauser, Boston, Ins., Boston (1995)

Chapter 7
Nash Equilibrium Payoffs in Mixed Strategies

Anne Souquière

Abstract We consider non zero sum two players differential games. We study Nash equilibrium payoffs and publicly correlated equilibrium payoffs. If players use deterministic strategies, it has been proved that the Nash equilibrium payoffs are precisely the reachable and consistent payoffs. Referring to repeated games, we introduce mixed strategies which are probability distributions over pure strategies. We give a characterization of the set of Nash equilibrium payoffs in mixed strategies. Unexpectedly, this set is larger than the closed convex hull of the set of Nash equilibrium payoffs in pure strategies. Finally, we study the set of publicly correlated equilibrium payoffs for differential games and show that it is the same as the set of Nash equilibrium payoffs using mixed strategies.

Keywords Non cooperative differential games • Nash equilibrium payoff • Publicly correlated equilibrium payoff • Mixed strategy

7.1 Introduction

We study equilibria for non zero sum differential games. In general, for a given equilibrium concept, existence and characterization of the equilibria highly depend on the strategies used by the players. There are mainly three types of strategies:

- Non-anticipative strategies or memory-strategies where the control depends on the entire past history of the game (trajectory and controls played so far).

A. Souquière (✉)
Institut TELECOM, TELECOM Bretagne, UMR CNRS 3192 Lab-STICC,
Technopole Brest Iroise, CS 83818, 29238 Brest Cedex, France

Laboratoire de Mathématiques, Université de Bretagne Occidentale,
UMR 6205, 6 Avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex, France
e-mail: anne.souquiere@telecom-bretagne.eu

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_7,

- Feed-back strategies where the current control depends only on the actual state of the system.
- Open-loop controls where the control depends only on time.

Looking for Nash equilibrium payoffs in feedback strategies, one usually computes Nash equilibrium payoffs as functions of time and space. This leads to a system of non linear partial differential equations for which there is no general result for existence nor uniqueness of a solution. If the system admits regular enough solutions, they allow to compute the optimal feedbacks [3, 12]. There are few examples for this approach, the results essentially deal with linear quadratic differential games where solutions are sought amongst quadratic functions. For linear quadratic games, there are conditions for existence of Nash equilibria in feedback strategies and for existence and uniqueness of Nash equilibria in open-loops. Some numerical methods can be applied to compute equilibria [11]. The drawback is that feedback equilibria are highly unstable [5], except in some particular cases of one dimensional games [6].

In the case of deterministic differential games where players use non-anticipative strategies, there are existence and characterization results for Nash equilibrium payoffs in [15, 16, 20]. Our aim is to extend this characterization to the case where players use mixed non-anticipative strategies, namely random combination of memory-strategies. The disadvantage of using non-anticipative strategies is that the associated equilibria lack weak consistency compared to feedback strategies. Their main interest is that they allow to characterize some kind of upper hull of all Nash equilibrium payoffs using reasonable strategies.

The notion of mixed strategies is strongly inspired by repeated games. The folk theorem for repeated games characterizes Nash equilibrium payoffs as feasible and individually rational [2, 18]. As in repeated games, the difficulty is that mixed strategies are unobservable [13].

Deterministic nonzero sum differential games are close to stochastic games, where there is a characterization of the set of correlated equilibria in case the punishment levels do not depend on the past history. This characterization, relying on “rational payoffs” [19] is close to ours and to the characterization of Nash equilibrium payoffs for stochastic games [10]. Our point is to give the link between these two sets. However, in our case, the punishment level varies with time and the specific conditions on the game comparable to the ones in [10] do not hold.

The notion of publicly correlated strategies has strong links with non zero sum stochastic differential games. As for the deterministic case, there is a general result of existence and characterization [7] in case players use non-anticipative strategies which is quite close to ours. For non degenerate stochastic differential games, there is a general result for existence of a Nash equilibrium in feedback strategies [4] based on the existence of smooth enough solutions for the system of partial differential equations defining the equilibrium. Another approach [14] uses BSDEs to check the existence of solutions, prove the existence of a Nash equilibrium and optimal feedbacks. Note that the equilibria defined through this last approach are in fact equilibria in non-anticipative strategies [17] when they both exist.

Here we deal with deterministic non zero sum differential games in mixed strategies and we study Nash equilibria and publicly correlated equilibria. We now expose the framework of our game.

We consider a two players non zero sum differential game in $\mathbb{R}^N$ that runs for $t \in [t_0, T]$. The dynamics of the game is given by:

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), v(t)) \ t \in [t_0, T], \ u(t) \in U \text{ and } v(t) \in V \\ x(t_0) = x_0 \end{cases} \tag{7.1}$$

We first define open-loop controls: we denote by $\mathcal{U}(t_0)$ (respectively $\mathcal{V}(t_0)$) the set of measurable controls of Player I (respectively Player II):

$$\mathcal{U}(t_0) := \{u(\cdot) : [t_0, T] \to U, \ u \text{ measurable}\}$$
$$\mathcal{V}(t_0) := \{v(\cdot) : [t_0, T] \to V, \ v \text{ measurable}\}$$

Under suitable regularity assumptions on the dynamics given below, if controls $u \in \mathcal{U}(t_0)$ and $v \in \mathcal{V}(t_0)$ are played, they define a unique solution of the dynamics (7.1) denoted by $t \mapsto X_t^{t_0, x_0, u, v}$ defined on $[t_0, T]$.

W.l.o.g., the payoffs only depend on the final position of the system. If the controls (u, v) are played, for $i = 1, 2$, Player i's payoff is $\mathfrak{J}_i(t_0, x_0, u, v) = g_i(X_T^{t_0, x_0, u, v})$ In this nonzero sum game, each player tries to maximize his final payoff.

In order to play the game, we need to define strategies. We first consider deterministic or pure strategies:

Definition 7.1 (Pure Strategy). A pure strategy for Player I at time t_0 is a map $\alpha : \mathcal{V}(t_0) \to \mathcal{U}(t_0)$ which satisfies the following conditions:

1. α is a measurable map from $\mathcal{V}(t_0)$ to $\mathcal{U}(t_0)$ where $\mathcal{V}(t_0)$ and $\mathcal{U}(t_0)$ are endowed with the Borel σ-field associated with the L^1 distance,
2. α is non-anticipative with delay, i.e. there exists some delay $\tau > 0$ such that for any $v_1, v_2 \in \mathcal{V}(t_0)$, if $v_1 \equiv v_2$ a.e. on $[t_0, t]$ for some $t \in [t_0, T]$, then $\alpha(v_1) \equiv \alpha(v_2)$ a.e. on $[t_0, (t + \tau) \wedge T]$

We denote by $\mathcal{A}(t_0)$ (respectively $\mathcal{B}(t_0)$) the set of pure strategies for Player I (respectively Player II) and by $\tau(\alpha)$ the delay of the strategy $\alpha \in \mathcal{A}(t_0)$.

The main interest of this definition is that we can associate to any $(\alpha, \beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$ a unique trajectory $t \mapsto X_t^{t_0, x_0, \alpha, \beta}$ defined on $[t_0, T]$ as stated in Lemma 7.1. This allows to define the payoffs $\mathfrak{J}_i(t_0, x_0, \alpha, \beta)$ for all $(\alpha, \beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$ and $i = 1, 2$ by $\mathfrak{J}_i(t_0, x_0, \alpha, \beta) = g_i(X_T^{t_0, x_0, \alpha, \beta})$. The point of the paper is to study the impact of introducing mixed strategies and correlated strategies on the equilibria.

We define mixed strategies as probability distributions over pure strategies:

Definition 7.2 (Mixed Strategy). A mixed strategy $((\Omega_\alpha, \mathcal{P}(\Omega_\alpha), \mathbf{P}_\alpha), \alpha)$ is a probability space $(\Omega_\alpha, \mathcal{F}_\alpha, \mathbf{P}_\alpha)$ and an application $\alpha : \Omega_\alpha \times \mathcal{V}(t_0) \to \mathcal{U}(t_0)$ that satisfies:

- α is a measurable map from $\Omega_\alpha \times \mathcal{V}(t_0)$ to $\mathcal{U}(t_0)$ where Ω_α is endowed with the σ-field $\mathcal{F}_\alpha$ and $\mathcal{V}(t_0)$ and $\mathcal{U}(t_0)$ with the Borel σ-field associated with the L^1 distance,
- α is non anticipative with delay, i.e. there is some delay $\tau > 0$ such that for any $\omega_\alpha \in \Omega_\alpha$, the pure strategy $\alpha(\omega_\alpha, \cdot)$ is non anticipative with delay τ.

We denote by $\mathcal{A}_r(t_0)$ (respectively $\mathcal{B}_r(t_0)$) the set of mixed strategies for Player I (respectively Player II).

The payoff associated to some mixed strategies $(\alpha, \beta) \in \mathcal{A}_r(t_0) \times \mathcal{B}_r(t_0)$ is defined thanks to Lemma 7.1 through $\mathfrak{J}_i(t_0, x_0, \alpha, \beta) = \int_{\Omega_\alpha \times \Omega_\beta} \mathfrak{J}_i(t_0, x_0, \alpha(\omega_\alpha), \beta(\omega_\beta))$ $d\mathbf{P}_\alpha \otimes d\mathbf{P}_\beta(\omega_\alpha, \omega_\beta)$. We also have to define random controls:

Definition 7.3 (Random Control). A random control $((\Omega, \mathcal{F}, \mathbf{P}), (u, v))$ on $[t_0, T]$ is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a collection of pairs of open-loop controls $((u, v)(\omega))_{\omega \in \Omega}$ such that for all $\omega \in \Omega$, $u(\omega) \in \mathcal{U}(t_0)$ and $v(\omega) \in \mathcal{V}(t_0)$ and $\omega \mapsto (u(\omega), v(\omega))$ is a measurable map from $(\Omega, \mathcal{F})$ to $\mathcal{U}(t_0) \times \mathcal{V}(t_0)$ endowed with the σ-field generated by the product of the σ-fields associated with the L^1-distance in $\mathcal{U}(t_0)$ and $\mathcal{V}(t_0)$.

We will call finite random control any random control with finite associated probability space. Note that we can naturally define the payoff associated to a random control.

In order to introduce publicly correlated strategies, we first need some correlation device.

Definition 7.4 (Correlation Device). A correlation device $((\Omega, \mathcal{F}, \mathbf{P}), \mathbf{C})$ is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a stochastic process $\mathbf{C} : [t_0, T] \times \Omega \to \mathbb{R}^N$ generating the natural filtration $(\mathcal{F}_t)$. We will often denote the correlation device with $\mathbf{C}$.

In order to correlate the strategies, we do not assume that Ω is finite, but we assume that both players observe $\mathbf{C}(t)$ at time t. The following definitions are adapted from [7]. We first introduce admissible controls:

Definition 7.5 (C-Admissible Control). For any correlation device $((\Omega, \mathcal{F}, \mathbf{P}), \mathbf{C})$ with associated natural filtration $(\mathcal{F}_t)$, a $\mathbf{C}$-admissible control $\tilde{u}$ for Player I is a $(\mathcal{F}_t)$-measurable process $\tilde{u} : [t_0, T] \times \Omega \to U$ progressively measurable with respect to $(\mathcal{F}_t)$, and symmetrically for Player II. The set of $\mathbf{C}$-admissible controls on $[t_0, T]$ is denoted by $\tilde{\mathcal{U}}_\mathbf{C}(t_0)$ for Player I and $\tilde{\mathcal{V}}_\mathbf{C}((t_0)$ for Player II. We will omit to mention the correlation device as far as no confusion is possible.

We will identify admissible controls and denote it by $\tilde{u}_1 \equiv \tilde{u}_2$ on $[t_0, t]$ as soon as $\mathbf{P}(\tilde{u}_1 = \tilde{u}_2 \text{ a.e. on } [t_0, t]) = 1$.

We define correlated strategies the following way:

Definition 7.6 (Publicly Correlated Strategies). Correlated strategies are a triplet $(\mathbf{C}, \alpha, \beta)$:

- $((\Omega, \mathcal{F}, \mathbf{P}), \mathbf{C})$ is a correlation device generating the natural filtration $(\mathcal{F}_t)$,

- A map $\alpha : \tilde{\mathcal{V}}_{\mathbf{C}}(t_0) \to \tilde{\mathcal{U}}_{\mathbf{C}}(t_0)$ which is strongly non-anticipative with delay [7]: there exists $\tau(\alpha) > 0$ such that $\forall (\mathcal{F}_t)$-stopping time S and for all $\tilde{v}_1, \tilde{v}_2 \in \tilde{\mathcal{V}}_{\mathbf{C}}(t_0)$, if $\tilde{v}_1 \equiv \tilde{v}_2$ on $[\![t_0, S]\!]$, then $\alpha(\tilde{v}_1) \equiv \alpha(\tilde{v}_2)$ on $[\![t_0, (S + \tau(\alpha)) \wedge T]\!]$,
- A map $\beta : \tilde{\mathcal{U}}(t_0)_{\mathbf{C}} \to \tilde{\mathcal{V}}_{\mathbf{C}}(t_0)$ which is a strongly non-anticipative strategy with delay.

We denote by $\mathcal{A}_c(t_0)$ the set of publicly correlated strategies.

Note that our definition is somehow broader than the usual definition of correlated strategies in repeated games where the correlation signal is given only at the beginning of the game. Our correlation device is closer to the autonomous correlation device described in [19]. We will call **C**-correlated strategies any publicly correlated strategies using the correlation device **C**. Note that we can associate a unique pair of **C**-admissible controls to any **C**-correlated strategies and therefore define a unique payoff associated to any publicly correlated strategies as recalled in Lemma 7.1.

We assume that the payoff functions g_1 and g_2 are Lipschitz continuous and bounded, and assume Isaacs'condition: for all $(x, \xi) \in \mathbb{R}^N \times \mathbb{R}^N$

$$\inf_{u \in U} \sup_{v \in V} \langle f(x,u,v), \xi \rangle = \sup_{v \in V} \inf_{u \in U} \langle f(x,u,v), \xi \rangle.$$

In this case the two-players zero sum game whose payoff function is g_1 (respectively g_2) has a value. We denote by

$$\mathbf{V}_1(t,x) := \sup_{\alpha \in \mathcal{A}(t)} \inf_{\beta \in \mathcal{B}(t)} \mathfrak{J}_1(t,x,\alpha,\beta) = \inf_{\beta \in \mathcal{B}(t)} \sup_{\alpha \in \mathcal{A}(t)} \mathfrak{J}_1(t,x,\alpha,\beta) \tag{7.2}$$

the value of the zero sum game with payoff function g_1 where Player I aims at maximizing his payoff and

$$\mathbf{V}_2(t,x) := \inf_{\alpha \in \mathcal{A}(t)} \sup_{\beta \in \mathcal{B}(t)} \mathfrak{J}_2(t,x,\alpha,\beta) = \sup_{\beta \in \mathcal{B}(t)} \inf_{\alpha \in \mathcal{A}(t)} \mathfrak{J}_2(t,x,\alpha,\beta) \tag{7.3}$$

the value of the zero sum game with payoff function g_2 where Player II is the maximizer. We recall that these definitions remain unchanged whether $\alpha \in \mathcal{A}(t)$ or $\mathcal{A}_r(t)$ and $\beta \in \mathcal{B}(t)$ or $\mathcal{B}_r(t)$ [8]. Our assumptions also guarantee that these value functions are Lipschitz continuous.

As we are interested in nonzero sum games, we need equilibrium concepts:

Definition 7.7 (Nash Equilibrium Payoff in Pure Strategies). The pair $(e_1, e_2) \in \mathbb{R}^2$ is a Nash equilibrium payoff in pure strategies (PNEP) for the initial conditions (t_0, x_0) if for all $\epsilon > 0$, there exists $(\alpha, \beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$ such that

1. For $i = 1, 2$, $|\mathfrak{J}_i(t_0, x_0, \alpha, \beta) - e_i| \leq \epsilon$
2. For all $\alpha' \in \mathcal{A}(t_0)$: $\mathfrak{J}_1(t_0, x_0, \alpha', \beta) \leq \mathfrak{J}_1(t_0, x_0, \alpha, \beta) + \epsilon$
3. For all $\beta' \in \mathcal{B}(t_0)$: $\mathfrak{J}_2(t_0, x_0, \alpha, \beta') \leq \mathfrak{J}_2(t_0, x_0, \alpha, \beta) + \epsilon$

We denote by $\mathcal{E}_p(t_0, x_0)$ the set of all PNEPs for the initial conditions (t_0, x_0).

Definition 7.8 (Nash Equilibrium Payoff in Mixed Strategies). The definition is the same as above replacing $\mathcal{A}(t_0)$ with $\mathcal{A}_r(t_0)$ and $\mathcal{B}(t_0)$ with $\mathcal{B}_r(t_0)$.

We denote by $\mathcal{E}_m(t_0,x_0)$ the set of all Nash equilibrium payoffs in mixed strategies (MNEPs) for the initial conditions (t_0,x_0).

Definition 7.9 (Publicly Correlated Equilibrium Payoff). The payoff $(e_1,e_2) \in \mathbb{R}^2$ is a publicly correlated equilibrium payoff (PCEP) for the initial conditions (t_0,x_0) if for all $\epsilon > 0$, there exists some correlated strategies $(\mathbf{C},\alpha,\beta) \in \mathcal{A}_c(t_0)$ such that:

1. For $i = 1,2$, $|\mathfrak{J}_i(t_0,x_0,\alpha,\beta) - e_i| \le \epsilon$
2. For all $(\mathbf{C},\alpha',\beta) \in \mathcal{A}_c(t_0)$: $\mathfrak{J}_1(t_0,x_0,\alpha',\beta) \le \mathfrak{J}_1(t_0,x_0,\alpha,\beta) + \epsilon$
3. For all $(\mathbf{C},\alpha,\beta') \in \mathcal{A}_c(t_0)$: $\mathfrak{J}_2(t_0,x_0,\alpha,\beta') \le \mathfrak{J}_2(t_0,x_0,\alpha,\beta) + \epsilon$

We denote by $\mathcal{E}_c(t_0,x_0)$ the set of all PCEPs for the initial conditions (t_0,x_0).

According to [15, 16, 20], the PNEPs for the initial conditions (t_0,x_0) are exactly the "reachable and consistent payoffs" $(e_1,e_2) \in \mathbb{R}^2$ satisfying: $\forall \epsilon > 0$, $\exists (u_\epsilon, v_\epsilon) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$ such that:

- $\forall i,\ |e_i - \mathfrak{J}_i(t_0,x_0,u_\epsilon,v_\epsilon)| \le \epsilon$
- $\forall i,\ \forall t \in [t_0,T],\ g_i(X_T^{t_0,x_0,u_\epsilon,v_\epsilon}) \ge \mathbf{V}_i(t,X_t^{t_0,x_0,u_\epsilon,v_\epsilon}) - \epsilon$

where $\mathbf{V}_i$ refers to (7.2) or (7.3). Furthermore, the set of PNEPs is non empty.

In this paper, we study MNEPs. First of all, noticing that any pure strategy can be considered as a trivial mixed strategy, the set $\mathcal{E}_m(t_0,x_0)$ is a non empty superset of $\mathcal{E}_p(t_0,x_0)$. It appears that the set $\mathcal{E}_m(t_0,x_0)$ is in fact compact, convex and generally strictly larger than the closed convex hull of the set $\mathcal{E}_p(t_0,x_0)$. Our main result (Theorem 7.1 below) states that:

The payoff $e = (e_1,e_2) \in \mathbb{R}^2$ is a MNEP iff for all $\epsilon > 0$, there exists a random control $((\Omega,\mathcal{F},\mathbf{P}),(u_\epsilon,v_\epsilon))$ such that $\forall i = 1,\ 2$:

- e is ϵ-reachable: $|\mathfrak{J}_i(t_0,x_0,u_\epsilon,v_\epsilon)] - e_i| \le \epsilon$
- (u_ϵ,v_ϵ) are ϵ-consistent:
  $\forall t \in [t_0,T]$, denoting by $\mathcal{F}_t = \sigma\left((u_\epsilon,v_\epsilon)(s),\ s \in [t_0,t]\right)$:

$$\mathbf{P}\left\{\mathbf{V}_i(t,X_t^{t_0,x_0,u_\epsilon,v_\epsilon}) \le \mathbf{E}\left[g_i(X_T^{t_0,x_0,u_\epsilon,v_\epsilon})\middle|\mathcal{F}_t\right] + \epsilon\right\} \ge 1 - \epsilon$$

The proof heavily relies on techniques introduced for repeated games in [1] known as "jointly controlled lotteries" and on the fact that we work with non-anticipative strategies with delay.

Finally, studying publicly correlated equilibria, we show that the set of PCEPs is equal to the set of MNEPs. The idea of the proof uses the similarity between correlated equilibrium payoffs and equilibrium payoffs of stochastic non zero sum differential games.

We complete this introduction by describing the outline of the paper. In Sect. 7.2, we recall the assumptions on the differential game we study. In Sect. 7.3, we give the main properties of the set of MNEPs and present an example where the set of

MNEPs is strictly larger than the convex hull of the set of PNEPs. In Sect. 7.4, we prove the equivalence between the sets of MNEPs and of PCEPs. We postpone to the last section the proof of the characterization of the set of MNEPs.

7.2 Definitions

7.2.1 Assumptions and Notations

Throughout the paper, for any $x, y \in \mathbb{R}^N$, we will denote by $\langle x, y \rangle$ the scalar product, by $\|x\|$ the euclidian norm and by $\|x\|_1$ the L^1 norm in $\mathbb{R}^N$: $\|x\|_1 = \max_{i=1\dots N} |x_i|$. The ball with center x and radius r will be denoted by $B(x,r)$. For any set S, $\mathbf{1}_S$ denotes the indicator function of S: for all $s \in S$, $\mathbf{1}_S(s) = 1$ and for all $s \notin S$, $\mathbf{1}_S(s) = 0$.

We first define the assumptions on the differential game we are dealing with. The dynamics of the game is given by (7.1):

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), v(t)) \ t \in [t_0, T], \ u(t) \in U \text{ and } v(t) \in V \\ x(t_0) = x_0 \end{cases}$$

where

$$\begin{cases} U \text{ and } V \text{ are compact subsets of some finite dimensional spaces} \\ U \text{ and } V \text{ have infinite cardinality,} \\ f : \mathbb{R}^N \times U \times V \to \mathbb{R}^N \text{ is bounded, continuous and uniformly} \\ \qquad \text{Lipschitz continuous with respect to } x \end{cases}$$

These assumptions guarantee existence and uniqueness of the trajectory $t \mapsto X_t^{t_0,x_0,u,v}$ for $t \in [t_0, T]$ generated by any pair of controls $(u,v) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$. The second assumption is a technical assumption simplifying the proof of the main theorem.

We will always assume that players observe the controls played so far.

7.2.2 Payoffs Associated to a Pair of Strategies

In order to study equilibrium payoffs of this game we have introduced pure and mixed strategies. The major interest of working with non-anticipative strategies with delay is the following useful result:

Lemma 7.1 (Controls Associated to a Pair of Strategies). *1. For any pair of pure strategies $(\alpha, \beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$ there is a unique pair of controls $(u_{\alpha\beta}, v_{\alpha\beta}) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$ such that $\alpha(v_{\alpha\beta}) = u_{\alpha\beta}$ and $\beta(u_{\alpha\beta}) = v_{\alpha\beta}$.*

2. *For any pair of mixed strategies* $(\alpha,\beta) \in \mathcal{A}_r(t_0) \times \mathcal{B}_r(t_0)$*, and any* $\omega = (\omega_\alpha, \omega_\beta) \in \Omega_\alpha \times \Omega_\beta$*, there is a unique pair of controls* $(u_\omega, v_\omega) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$ *such that* $\alpha(\omega_\alpha)(v_\omega) = u_\omega$ *and* $\beta(\omega_\beta)(u_\omega) = v_\omega$*. Furthermore, the map* $\omega \mapsto (u_\omega, v_\omega)$ *is measurable from* $\Omega_\alpha \times \Omega_\beta$ *endowed with* $\mathcal{F}_\alpha \otimes \mathcal{F}_\beta$ *into* $\mathcal{U}(t_0) \times \mathcal{V}(t_0)$ *endowed with the Borel* σ*-field associated with the* L^1 *distance.*
3. *For any correlated strategies* $(\mathbf{C},\alpha,\beta) \in \mathcal{A}_c(t_0)$*, there is a unique pair of* $\mathbf{C}$*-admissible controls* $(\tilde{u}_{\alpha\beta}, \tilde{v}_{\alpha\beta}) \in \mathcal{U}_{\mathbf{C}}(t_0) \times \mathcal{V}_{\mathbf{C}}(t_0)$ *such that* $\alpha(\tilde{v}_{\alpha\beta}) = \tilde{u}_{\alpha\beta}$ *and* $\beta(\tilde{u}_{\alpha\beta}) = \tilde{v}_{\alpha\beta}$*.*

Proof. The first two results are in [9], whereas the third is a straightforward extension of the result established for admissible strategies in [7]. Given any pair of pure strategies $(\alpha,\beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$, we denote by $(X_\cdot^{t_0,x_0,\alpha,\beta})$ the map $t \mapsto X_t^{t_0,x_0,u_{\alpha\beta},v_{\alpha\beta}}$ defined on $[t_0,T]$ where $X_\cdot^{t_0,x_0,u_{\alpha\beta},v_{\alpha\beta}}$ is the unique solution of the dynamics (7.1). This allows us to define the payoff associated to any pair of strategies.

Given correlated strategies $(\mathbf{C},\alpha,\beta)$, the final payoff of Player i is:

$$\mathfrak{J}_i(t,x,\alpha,\beta) = \mathbf{E}\left[g_i\left(X_T^{t,x,\tilde{u}_{\alpha\beta},\tilde{v}_{\alpha\beta}}\right)\right] := \mathbf{E}\left[g_i\left(X_T^{t,x,\alpha,\beta}\right)\right]$$

Notice that pure strategies are degenerated correlated strategies using some trivial correlation device. Finally, note that in a zero sum game, using correlated strategies with a fixed correlation device leads to the same value as using pure strategies. Indeed, fix the device $((\Omega,\mathcal{F},\mathbf{P}),\mathbf{C})$ and denote by $(\mathbf{C},\tilde{\alpha},\tilde{\beta})$ any $\mathbf{C}$-correlated strategies and (α,β) any pair of pure strategies. For $i=1,2$:

$$\begin{aligned}
\sup_{\tilde{\beta}}\inf_{\tilde{\alpha}}\mathbf{E}\left[g_i(X_T^{t,x,\tilde{\alpha},\tilde{\beta}})\right] &\geq \sup_{\beta}\inf_{\tilde{\alpha}}\mathbf{E}\left[g_i(X_T^{t,x,\tilde{\alpha},\beta})\right] \\
&= \sup_{\beta}\inf_{\alpha} g_i(X_T^{t,x,\alpha,\beta}) = \mathbf{V}_i(t,x) \\
&= \inf_{\alpha}\sup_{\beta} g_i\left(X_T^{t,x,\alpha,\beta}\right) \\
&= \inf_{\alpha}\sup_{\tilde{\beta}}\mathbf{E}\left[g_i\left(X_T^{t,x,\alpha,\tilde{\beta}}\right)\right] \\
&\geq \inf_{\tilde{\alpha}}\sup_{\tilde{\beta}}\mathbf{E}\left[g_i\left(X_T^{t,x,\tilde{\alpha},\tilde{\beta}}\right)\right]
\end{aligned}$$

On the other hand we have:

$$\sup_{\tilde{\beta}}\inf_{\tilde{\alpha}}\mathbf{E}\left[g_i\left(X_T^{t,x,\tilde{\alpha},\tilde{\beta}}\right)\right] \leq \inf_{\tilde{\alpha}}\sup_{\tilde{\beta}}\mathbf{E}\left[g_i\left(X_T^{t,x,\tilde{\alpha},\tilde{\beta}}\right)\right]$$

and in the end for any **C**-correlated strategies $\tilde{\alpha}, \tilde{\beta}$, for $i = 1,2$:

$$\sup_{\tilde{\beta}} \inf_{\tilde{\alpha}} \mathfrak{J}_i \left(t, x, \tilde{\alpha}, \tilde{\beta} \right) = \mathbf{V}_i(t,x) = \inf_{\tilde{\alpha}} \sup_{\tilde{\beta}} \mathfrak{J}_i \left(t, x, \tilde{\alpha}, \tilde{\beta} \right) \qquad \square$$

7.2.3 Definitions

We will call reachable in mixed strategies a payoff $(e_1, e_2) \in \mathbb{R}^2$ which completes only the first part of the definition 7.8: $\forall \epsilon > 0$, $\exists (\alpha_\epsilon, \beta_\epsilon) \in \mathcal{A}_r(t_0) \times \mathcal{B}_r(t_0)$ such that:

$$\forall i = 1,2, \ |e_i - \mathfrak{J}_i(t_0, x_0, \alpha_\epsilon, \beta_\epsilon)| \leq \epsilon$$

A pair of strategies $(\alpha_\epsilon, \beta_\epsilon) \in \mathcal{A}_r(t_0) \times \mathcal{B}_r(t_0)$ satisfying:

$$\begin{cases} \forall \alpha \in \mathcal{A}_r(t_0), \ \mathfrak{J}_1(t_0, x_0, \alpha_\epsilon, \beta_\epsilon) \geq \mathfrak{J}_1(t_0, x_0, \alpha, \beta_\epsilon) - \epsilon \\ \forall \beta \in \mathcal{B}_r(t_0), \ \mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta_\epsilon) \geq \mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta) - \epsilon \end{cases}$$

will be called ϵ-optimal. Note that we just have to check the ϵ-optimality of α_ϵ (respectively β_ϵ) against pure strategies $\beta \in \mathcal{B}(t_0)$ (respectively $\alpha \in \mathcal{A}(t_0)$), if α_ϵ and β_ϵ are defined on a finite probability space.

7.3 Nash Equilibrium Payoffs Using Mixed Strategies

7.3.1 Characterization

Theorem 7.1 (Characterization of Nash Equilibrium Payoffs Using Mixed Strategies). *The payoff $e = (e_1, e_2) \in \mathbb{R}^2$ is a MNEP iff for all $\epsilon > 0$, there exists a finite random control $((\Omega, \mathcal{P}(\Omega), \mathbf{P}), (u_\epsilon, v_\epsilon))$ such that $\forall i = 1,\ 2$:*

- *e is ϵ-reachable: $|\mathfrak{J}_i(t_0, x_0, u_\epsilon, v_\epsilon)] - e_i| \leq \epsilon$*
- *(u_ϵ, v_ϵ) are ϵ-consistent:*
  $\forall t \in [t_0, T]$, denoting by $\mathcal{F}_t = \sigma\left((u_\epsilon, v_\epsilon)(s),\ s \in [t_0, t]\right)$:

$$\mathbf{P}\left\{ \mathbf{V}_i(t, X_t^{t_0, x_0, u_\epsilon, v_\epsilon}) \leq \mathbf{E}\left[g_i(X_T^{t_0, x_0, u_\epsilon, v_\epsilon}) \middle| \mathcal{F}_t \right] + \epsilon \right\} \geq 1 - \epsilon$$

Note that the characterization could be given using trajectories following [20] rather than controls, provided the trajectory stems from the dynamics (7.1).

We just give the idea of the proof which is postponed to Sect. 7.5.

The fact that any MNEP satisfies such a characterization is in fact quite natural if we extend the definition to any random control. Otherwise, there would exist profitable deviations for one of the players. The way to restrict the definition only to finite random controls is given through appropriate projection as shown in Sect. 7.4.

The sufficient condition is not intuitive. We have to build non anticipative strategies with delay such that no unilateral deviation is profitable. The idea is to build a trigger strategy: follow the same trajectory as the one defined through the consistent controls (u_ϵ, v_ϵ) as long as no deviation occurs and punish any deviation in such a way that if a deviation occurred at the point $(t, x(t))$ the deviating player, say i, will be rewarded with his guaranteed payoff $\mathbf{V}_i(t, x(t))$. The unique difficulty is to coordinate the choice of the trajectory to be followed each time there is some node in the trajectories generated by (u_ϵ, v_ϵ). To this end, players will use some small delay at each node in order to communicate through jointly controlled lottery. Assume for example that the trajectory is splitting in two, one generated by ω_1 with probability $1/2$ and another generated by ω_2 with probability $1/2$. During the small communication delay, Player I chooses either the control u_1 or u_2 and Player II selects v_1 or v_2. If (u_1, v_1) or (u_2, v_2) are played, players will follow the trajectory generated by ω_1 and the one generated by ω_2 otherwise. Note that if each player selects each communication control with probability $1/2$ no unilateral cheating in the use of the control may change the probability of the outcome: each trajectory will be followed with probability $1/2$. This jointly controlled lottery procedure is easily extended to any finite probability over the trajectories. Of course, if one player does not use the communication control, he will be punished and get his guaranteed payoff which, by assumption, is not profitable.

7.3.2 Convexity of the Set of Nash Equilibrium Payoffs Using Mixed Strategies

Proposition 7.1. *The set $\mathcal{E}_m(t_0, x_0)$ of all MNEPs for the initial conditions (t_0, x_0) is convex and compact in $\mathbb{R}^2$.*

Proof. Compactness comes from the fact that the payoff functions are bounded.

Let $(\mathrm{e}^1, \mathrm{e}^2) \in \mathbb{R}^4$ be a pair of Nash equilibrium payoffs in mixed strategies. We will prove that $(\lambda\,\mathrm{e}^1 + (1-\lambda)\,\mathrm{e}^2)$ is a Nash equilibrium payoff in mixed strategies for all $\lambda \in (0,1)$. We will simply build a finite random control satisfying the characterization property of Theorem 7.1. As for $j = 1, 2$, e^j is a Nash equilibrium payoff, we may choose random controls $((\Omega^j, \mathcal{P}(\Omega^j), \mathbf{P}^j), (u^j, v^j))$ such that $\forall i, j = 1,\ 2$:

- $|\mathbf{E}^j(g_i(X_T^{t_0,x_0,u^j,v^j})) - e_i^j| \leq \frac{\epsilon}{3}$
- $\forall t \in [t_0, T]$, denoting by $\mathcal{F}_t^j = \sigma\left((u^j, v^j)(s),\ s \in [t_0, t]\right)$:

$$\mathbf{P}^j \left\{ \mathbf{V}_i(t, X_t^{t_0,x_0,u^j,v^j}) \leq \mathbf{E}^j \left(g_i(X_T^{t_0,x_0,u^j,v^j}) | \mathcal{F}_t^j \right) + \frac{\epsilon}{3} \right\} \geq 1 - \epsilon$$

We need to build controls close to the initial pairs (u^j, v^j), $j = 1, 2$, but with some tag in order to distinguish them. Set some small delay $\delta > 0$ such that for all $x \in B(x_0, \delta \|f\|_\infty)$, for all $(u, v) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$, for all $i = 1, 2$, for all $t \geq t_0 + \delta$:

$$\begin{cases} \left| \mathbf{V}_i(t, X_t^{t_0,x_0,u,v}) - \mathbf{V}_i(t-\delta, X_{t-\delta}^{t_0,x,u,v}) \right| \leq \frac{\epsilon}{3} \\ \left| g_i(X_T^{t_0,x_0,u,v}) - g_i(X_{T-\delta}^{t_0,x,u,v}) \right| \leq \frac{\epsilon}{3} \end{cases} \tag{7.4}$$

We now choose some $u_1 \neq u_2 \in U$ and $v_1 \neq v_2 \in V$ and set for $j = 1, 2$:

$$\begin{cases} \bar{u}^j(s) = u_j & \text{for } s \in [t_0, t_0 + \delta) \\ \bar{u}^j(s) = u^j(s-\delta) & \text{for } s \in [t_0 + \delta, T] \\ \bar{v}^j(s) = v_j & \text{for } s \in [t_0, t_0 + \delta) \\ \bar{v}^j(s) = v^j(s-\delta) & \text{for } s \in [t_0 + \delta, T] \end{cases}$$

We will denote by $\bar{X}^j_\cdot = X_\cdot^{t_0,x_0,\bar{u}^j,\bar{v}^j}$ for $j = 1, 2$. We immediately get thanks to (7.4) $\forall i, j = 1,\ 2$:

$$|\mathbf{E}^j(g_i(\bar{X}_T^j)) - e_i^j| \leq 2\frac{\epsilon}{3} \leq \epsilon \tag{7.5}$$

and for all $t \in [t_0, T]$, denoting by $\bar{\mathcal{F}}_t^j = \sigma\left((\bar{u}^j, \bar{v}^j)(s),\ s \in [t_0, t]\right)$:

$$\mathbf{P}^j \left\{ \mathbf{V}_i(t, \bar{X}_t^j) \leq \mathbf{E}^j \left(g_i(\bar{X}_T^j) | \bar{\mathcal{F}}_t^j \right) + \epsilon \right\} \geq 1 - \epsilon \tag{7.6}$$

For $i, j = 1, 2$, denote by

$$\Sigma_t^{ij} = \left\{ \mathbf{V}_i(t, \bar{X}_t^j) \leq \mathbf{E}^j \left(g_i(\bar{X}_T^j) | \bar{\mathcal{F}}_t^j \right) + \epsilon \right\}$$

We now define a new finite random space $\Omega = \{1, 2\} \times \Omega^1 \times \Omega^2$ endowed with the probability $\mathbf{P}$ defined for all $\omega = (j, \omega^1, \omega^2)$ by:

$$\begin{cases} \mathbf{P}(j, \omega^1, \omega^2) = \lambda \mathbf{P}^1(\omega^1)\mathbf{P}^2(\omega^2) & \text{if } j = 1 \\ \mathbf{P}(j, \omega^1, \omega^2) = (1-\lambda)\mathbf{P}^1(\omega^1)\mathbf{P}^2(\omega^2) & \text{if } j = 2 \end{cases}$$

and define on Ω the random control (u, v) defined by:

$$\begin{cases} (u, v)(j, \omega^1, \omega^2) = (\bar{u}^1, \bar{v}^1)(\omega^1) & \text{if } j = 1 \\ (u, v)(j, \omega^1, \omega^2) = (\bar{u}^2, \bar{v}^2)(\omega^2) & \text{if } j = 2 \end{cases}$$

We will denote by $X_\cdot = X_\cdot^{t_0,x_0,u,v}$. It remains to prove that for $i = 1, 2$:

- $|\mathbf{E}[g_i(X_T)] - \lambda e_i^1 - (1-\lambda) e_i^2| \leq \epsilon$

- $\forall t \in [t_0, T]$, denoting by $\mathcal{F}_t = \sigma\left((u,v)(s),\ s \in [t_0,t]\right)$:

$$\mathbf{P}\left\{\mathbf{V}_i(t,X_t) \leq \mathbf{E}\left[g_i(X_T)\middle|\mathcal{F}_t\right] + \epsilon\right\} \geq 1-\epsilon$$

The first relation is easy to prove. For $i = 1,2$, we have:

$$\begin{aligned}\left|\mathbf{E}[g_i(X_T)] - \lambda e_i^1 - (1-\lambda)e_i^2\right| &= \left|\lambda \mathbf{E}^1[g_i(\bar{X}_T^1)] + (1-\lambda)\mathbf{E}^2[g_i(\bar{X}_T^2)] - \lambda e_i^1 - (1-\lambda)e_i^2\right| \\ &\leq \lambda\left|\mathbf{E}^1[g_i(\bar{X}_T^1)] - e_i^1\right| + (1-\lambda)\left|\mathbf{E}^2[g_i(\bar{X}_T^2)] - \mathrm{e}_i^2\right| \\ &\leq \epsilon \text{ thanks to (7.5)}\end{aligned}$$

In order to prove the second inequality, for $i = 1,2$, we denote by

$$\Sigma_t^i = \left\{\mathbf{V}_i(t,X_t) \leq \mathbf{E}\left(g_i(X_T)\middle|\mathcal{F}_t\right) + 3\epsilon\right\}$$

We have for $i = 1,2$ and $t \in [t_0+\delta, T]$:

$$\begin{aligned}\mathbf{E}\left(g_i(X_T)\middle|\mathcal{F}_t\right) &= \mathbf{E}\left(g_i(X_T)(\mathbf{1}_{\{1\}\times\Omega^1\times\Omega^2} + \mathbf{1}_{\{2\}\times\Omega^1\times\Omega^2})\middle|\mathcal{F}_t\right) \\ &= \mathbf{E}\left(g_i(\bar{X}_T^1)\middle|\bar{\mathcal{F}}_t^1\right)\mathbf{1}_{\{1\}\times\Omega^1\times\Omega^2} + \mathbf{E}\left(g_i(\bar{X}_T^2)\middle|\bar{\mathcal{F}}_t^2\right)\mathbf{1}_{\{2\}\times\Omega^1\times\Omega^2}\end{aligned}$$

Therefore, assuming w.l.o.g. that the functions g_i are non negative and using (7.6):

$$\begin{aligned}\mathbf{E}\left(g_i(X_T)\middle|\mathcal{F}_t\right) &\geq [\mathbf{V}_i(t,\bar{X}_t^1) - \epsilon]\mathbf{1}_{\{1\}\times\Sigma_t^{i1}\times\Omega^2} + [\mathbf{V}_i(t,\bar{X}_t^2) - \epsilon]\mathbf{1}_{\{2\}\times\Omega^1\times\Sigma_t^{i2}} \\ &\geq \mathbf{V}_i(t,X_t)\mathbf{1}_{\{1\}\times\Sigma_t^{i1}\times\Omega^2} + \mathbf{V}_i(t,X_t)\mathbf{1}_{\{2\}\times\Omega^1\times\Sigma_t^{i2}} - \epsilon\end{aligned}$$

And finally:

$$\begin{aligned}\mathbf{P}\left(\Sigma_t^i\right) &\geq \mathbf{P}\left[\{1\}\times\Sigma_t^{i1}\times\Omega^2 \cup \{2\}\times\Omega^1\times\Sigma_t^{i2}\right] \\ &\geq \lambda(1-\epsilon) + (1-\lambda)(1-\epsilon) \\ &\geq 1-\epsilon\end{aligned}$$

Note that for $t \in [t_0, t_0+\delta]$, the preceding relation is straightforward. □

7.3.3 *Comparison Between the Sets of Nash Equilibrium Payoffs in Pure and Mixed Strategies*

We have just proven that the set of MNEPs is convex; therefore it contains the closed convex hull of the set of PNEPs. When trying to compare these two sets, it appears that in general, they are not equal. This result is not intuitive because the guaranteed

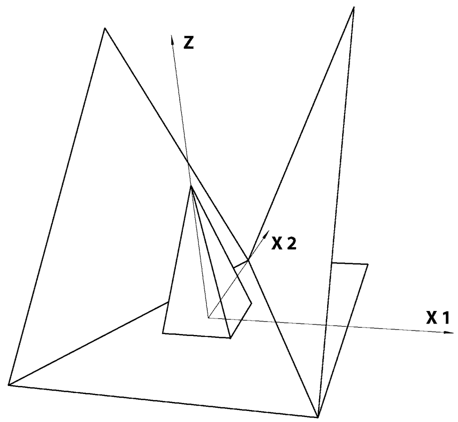


Fig. 7.1 Function g_1

payoffs are exactly the same whether players use pure or mixed strategies. It appears because players may correlate their strategies throughout the whole game and not only at the beginning of it.

Proposition 7.2. *There exist nonzero sum differential games in which the set of MNEPs is larger than the convex hull of the set of PNEPs.*

Proof. We will build a counter-example where an MNEP does not belong to the closed convex hull of the PNEPs.

Consider the simple game in finite time in $\mathbb{R}^2$ with dynamics:

$$\dot{x} = u + v \qquad u,\ v \in [-1/2, 1/2]^2$$

starting from the origin $O = (0,0)$ at time $t = 0$ and ending at time $t = T = 1$. The set of all reachable points in this game is the unit ball in $\mathbb{R}^2$ for the L^1 norm.

The payoff functions are the Lipschitz continuous functions defined as follows:

$$g_1 : \begin{cases} g_1(x) = 1 - 4|x_2| & \text{for } |x_2| \leq 1/4 \text{ and } |x_2| \geq |x_1| \\ g_1(x) = 1 - 4|x_1| & \text{for } |x_1| \leq 1/4 \text{ and } |x_1| \geq |x_2| \\ g_1(x) = x_2 + 2|x_1| - 1 & \text{for } x_2 \geq -2|x_1| + 1 \\ g_1(x) = 0 & \text{elsewhere} \end{cases}$$

In fact, g_1 is the nonnegative function defined on the unit square shown in Fig. 7.1

$$g_2 : \begin{cases} g_2(x) = 0 & \text{for } x_2 \geq 0 \\ g_2(x) = -x_2 & \text{for } x_2 \leq 0 \end{cases}$$

The game clearly fulfill the regularity assumptions listed in the introduction. We will denote by L_g the greater of the Lipschitz-constants of g_1 and g_2 for the L^1-norm.

The set of all reachable payoffs is $[0,2]\times\{0\}\cup\bigcup_{y\in(0,1]}([0,1-y],y)$. It is also clear that

$$\begin{cases}\mathbf{V}_1(t,x)=g_1(x)\\ \mathbf{V}_2(t,x)=g_2(x)\end{cases}$$

The initial values are $\mathbf{V}_1(0,O)=1$ and $\mathbf{V}_2(0,O)=0$, implying any Nash equilibrium payoff has to reward Player I with at least 1 and Player II with a non-negative payoff. In pure strategies, no trajectory can end up at time T at some x such that $x_2<0$ because this would cause Player I to earn strictly less than 1. We then have $e_2=0$ corresponding to $x_2\geq 0$ for every PNEP. We can easily compute

$$\mathcal{E}_p(0,O)=[1,2]\times\{0\}=\overline{\mathrm{Conv}}\mathcal{E}_p(0,O).$$

It is the set of all reachable payoffs such that $e_1\geq 1$ and $e_2\geq 0$.

We now will compute some finite random control (u,v) leading to a final payoff of 1 for Player I and positive for Player II. The controls are as follows:

- For $t\in[0,3/4]$, play $u=v=(1/2,0)$ and join $(3/4,0)$ at $t=3/4$.
- From $t=3/4$ on, with probability one half: play $u=v=(1/2,1/2)$ and join $(1,1/4)$ at $t=1$ to get the payoff $(5/4,0)$.
- From $t=3/4$ on, with probability one half: play $u=v=(1/2,-1/2)$ and join $(1,-1/4)$ at $t=1$ to get the payoff $(3/4,1/4)$.

The final payoff will be $(e_1,e_2)=(1,1/8)\notin\overline{\mathrm{Conv}}\mathcal{E}_p(0,O)$. It remains to prove that this payoff is a MNEP.

We will denote by $X_\cdot=X_\cdot^{0,O,u,v}$. We use the characterization of the MNEPs of Theorem 7.1 and prove that along the trajectories the condition

$$\forall i=1,2\ \forall t\in[0,1]\ \mathbf{E}[g_i(X_T)|\mathcal{F}_t]\geq\mathbf{V}_i(t,X_t)$$

is satisfied. Indeed, along the trajectories:

$$\begin{array}{lll}
\text{for } t\in[0,1/4] & \mathbf{V}_1(t,X_t)=1-4t\in[0,1] & \text{and } \mathbf{E}\left(g_1(X_T)\middle|\mathcal{F}_t\right)=1\\
\text{for } t\in[1/4,1/2] & \mathbf{V}_1(t,X_t)=0 & \text{and } \mathbf{E}\left(g_1(X_T)\middle|\mathcal{F}_t\right)=1\\
\text{for } t\in[1/2,3/4] & \mathbf{V}_1(t,X_t)=2t-1\in[0,1/2] & \text{and } \mathbf{E}\left(g_1(X_T)\middle|\mathcal{F}_t\right)=1\\
\text{for } t\in(3/4,1]: & \text{either } \mathbf{V}_1(t,X_t)=3t-7/4\in[1/2,5/4] & \text{and } \mathbf{E}\left(g_1(X_T)\middle|\mathcal{F}_t\right)=5/4\\
 & \text{or } \quad \mathbf{V}_1(t,X_t)=t-1/4\in[1/2,3/4] & \text{and } \mathbf{E}\left(g_1(X_T)\middle|\mathcal{F}_t\right)=3/4
\end{array}$$

and

$$\begin{cases}
\text{for } t\in[0,3/4] & \mathbf{V}_2(t,X_t)=0 & \text{and } \mathbf{E}\left(g_2(X_T)\middle|\mathcal{F}_t\right)=1/8\\
\text{for } t\in(3/4,1]: \text{either} & \mathbf{V}_2(t,X_t)=0 & \text{and } \mathbf{E}\left(g_2(X_T)\middle|\mathcal{F}_t\right)=0\\
\qquad\qquad\qquad \text{or} & \mathbf{V}_2(t,X_t)=t-3/4\in[0,1/4] & \text{and } \mathbf{E}\left(g_2(X_T)\middle|\mathcal{F}_t\right)=1/4
\end{cases}$$

This proves that the final payoff $(e_1,e_2)=(1,1/8)$ is a MNEP. □

7.4 Publicly Correlated Equilibrium Payoffs

We recall that $\mathcal{E}_c(t_0,x_0) \supset \mathcal{E}_p(t_0,x_0)$. We are going to state some characterization of publicly correlated equilibrium payoffs (PCEPs) and compare the set of MNEPs and the set of PCEPs.

Theorem 7.2. *The set of publicly correlated equilibrium payoffs is equal to the set of Nash equilibrium payoffs using mixed strategies.*

Proof. To begin with, we will show that $\mathcal{E}_m(t_0,x_0) \subseteq \mathcal{E}_c(t_0,x_0)$. First note that a simple adaptation of the proof in [7] states that:

Proposition 7.3 (Characterization of publicly correlated equilibrium payoffs). *The payoff $e=(e_1,e_2)\in\mathbb{R}^2$ is a PCEP for the initial conditions (t_0,x_0) iff for all $\epsilon>0$, there exists a random control $((\Omega,\mathcal{F},\mathbf{P}),(u_\epsilon,v_\epsilon))$, such that $\forall i=1,\ 2$:*

1. $\left|\mathbf{E}[g_i(X_T^{t_0,x_0,u_\epsilon,v_\epsilon})]-e_i\right|\le\epsilon$
2. $\forall t\in[t_0,T]$, *if we denote by* $(\mathcal{F}_t^\epsilon)=\sigma\{(u_\epsilon,v_\epsilon)(s),\ s\in[t_0,t]\}$
   $\mathbf{P}\left\{\mathbf{E}\left[g_i(X_T^{t_0,x_0,u_\epsilon,v_\epsilon})\middle|\mathcal{F}_t^\epsilon\right]\ge\mathbf{V}_i(t,X_t^{t_0,x_0,u_\epsilon,v_\epsilon})-\epsilon\right\}\ge 1-\epsilon$

This characterization and Theorem 7.1 ensure that any MNEP is in fact a PCEP.

We now will prove that $\mathcal{E}_m(t_0,x_0)\supseteq\mathcal{E}_c(t_0,x_0)$. Note that the only difference between the characterizations of MNEPs and PCEPs is that the latest relies on a random control possibly defined on an infinite underlying probability space, whereas MNEPs are characterized through finite random controls. We will consider some PCEP satisfying the characterization of Proposition 7.3 and we will prove that we are able to build a finite random control satisfying the characterization of Theorem 7.1, implying it will be a MNEP.

Consider some PCEP e. Fix ϵ and consider the ϵ^2-optimal random control $((\Omega,\mathcal{F},\mathbf{P}),(u_\epsilon,v_\epsilon))$. Denote by $X_\cdot^\epsilon=X_\cdot^{t_0,x_0,u_\epsilon,v_\epsilon}$ and set for all $\omega\in\Omega$: $X_\cdot^\epsilon(\omega)=X_\cdot^{t_0,x_0,(u_\epsilon,v_\epsilon)(\omega)}$. Note that this random control satisfies:

$$\begin{cases}\left|\mathbf{E}[g_i(X_T^\epsilon)]-e_i\right|\le\epsilon^2\\ \forall t\in[t_0,T],\text{ if we denote by }\mathcal{F}_t^\epsilon=\sigma\left\{(u_\epsilon,v_\epsilon)(s),\ s\in[t_0,t]\right\}:\\ \quad\mathbf{P}\left\{\mathbf{E}\left[g_i(X_T^\epsilon)\middle|\mathcal{F}_t^\epsilon\right]\ge\mathbf{V}_i(t,X_t^\epsilon)-\epsilon^2\right\}\ge 1-\epsilon^2\end{cases}\tag{7.7}$$

If Ω is finite, there is nothing left to prove. Else, we will build a finite random control rewarding a payoff close to e and consistent.

We set $h>0$ and $\bar h>0$ to be defined later such that there exist $N_h,N_{\bar h}\in\mathbb{N}^*$ such that $T-t_0=N_h h$ and $(T-t_0)\|f\|_\infty=N_{\bar h}\bar h$. We build the following time partition $G_h=\{t_k=t_0+kh\}_{k=0,\dots,N_h}$ and the grid in $\mathbb{R}^N$: $G_{\bar h}=\{x_0+\sum_{i=1}^n k_i\bar h e_i\}_{(k_i)\in\{-N_{\bar h},\dots,0,\dots,N_{\bar h}\}^n}$ where $(e_i)_{i=1\dots n}$ is a basis of $\mathbb{R}^N$. We now introduce a projection on the grid:

$$\begin{aligned}&\mathbb{R}^N\to G_{\bar h}\\ \Pi:\ &x\ \mapsto\min\{x_i\in G_{\bar h}/\ d_1(x,x_i)=\inf_{x_j\in G_{\bar h}}d_1(x,x_j)\}\end{aligned}$$

where the minimum is taken with respect to the lexicographic order and d_1 is the distance associated to the norm $\|x\|_1$.

To any $(t_k, x_i, x_j) \in G_h \times G_{\bar{h}} \times G_{\bar{h}}$ we associate, if it exists some $\varphi(t_k, x_i, x_j) = (x, u, v) \in \mathbb{R}^N \times \mathcal{U}(t_k) \times \mathcal{V}(t_k)$ such that $\Pi(x) = x_i$ and $\Pi(X_{t_{k+1}}^{t_k, x, u, v}) = x_j$. We will set $\varphi_x(t_k, x_i, x_j) = x$ and $\varphi_c(t_k, x_i, x_j) = (u, v)$.

We now are able to build a finite random control on $(\Omega, \mathcal{F}, \mathbf{P})$. To any $\omega \in \Omega$ we associate $(u_\eta, v_\eta)(\omega)$ in the following way:

- Fix $(u_0, v_0) \in U \times V$
- $(u_\eta, v_\eta)(\omega)|_{[t_0, t_1)} = (u_0, v_0)$
- For all $k = 1 \ldots N_h - 1$, for all $s \in [t_k, t_{k+1})$:

$$(u_\eta, v_\eta)(\omega)(s) = \varphi_c\left(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega))\right)(s - h)$$

Note that the definition of (u_η, v_η) is non anticipative. From now on, we will denote by $X_\cdot^\eta = X_\cdot^{t_0, x_0, u_\eta, v_\eta}$ and set for all $\omega \in \Omega$: $X_\cdot^\eta(\omega) = X_\cdot^{t_0, x_0, (u_\eta, v_\eta)(\omega)}$.

We now would like to prove that the set of finitely many random control (u_η, v_η) defined on $(\Omega, \mathcal{F}, \mathbf{P})$ satisfies for $i = 1, 2$, for some constants C_1, C_2, C_3:

- $|\mathbf{E}[g_i(X_T^\eta)] - e_i| \leq C_1 \epsilon$
- $\forall t \in [t_0, T]$, if we denote by $\mathcal{F}_t^\eta = \sigma\{(u_\eta, v_\eta)(s),\ s \in [t_0, t]\}$
  $\mathbf{P}\left\{\mathbf{E}\left(g_i(X_T^\eta) \middle| \mathcal{F}_t^\eta\right) \geq \mathbf{V}_i(t, X_t^\eta) - C_2 \epsilon\right\} \geq 1 - C_3 \epsilon$

First of all, we shall prove that the trajectories generated by (u_η, v_η) and (u_ϵ, v_ϵ) are close for sufficiently small values of h and $\bar{h}$.

For all $k = 0 \ldots N_h - 1$, we have

$$\begin{aligned}
&\|X_{t_{k+1}}^\eta(\omega) - X_{t_k}^\epsilon(\omega)\|_1 \\
&\quad \leq \left\| X_{t_k}^{t_{k-1}, X_{t_k}^\eta(\omega), \varphi_c(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega)))} - X_{t_k}^\epsilon(\omega) \right\|_1 \\
&\quad \leq \left\| X_{t_k}^{t_{k-1}, \varphi(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega)))} - X_{t_k}^\epsilon(\omega) \right\|_1 \\
&\qquad + \left\| X_{t_k}^{t_{k-1}, X_{t_k}^\eta(\omega), \varphi_c(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega)))} - X_{t_k}^{t_{k-1}, \varphi(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega)))} \right\|_1 \\
&\quad \leq \bar{h} + \left\| \varphi_x(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon)(\omega)) - X_{t_k}^\eta(\omega) \right\|_1 \mathrm{e}^{L_f h} \\
&\quad \leq \bar{h} + \left(\left\| X_{t_k}^\eta(\omega) - X_{t_{k-1}}^\epsilon(\omega) \right\|_1 + \bar{h} \right) \mathrm{e}^{L_f h}
\end{aligned}$$

because by definition, $\Pi(X_{t_k}^\epsilon(\omega)) = \Pi(X_{t_k}^{t_{k-1}, \varphi(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega)))})$ and $\Pi(X_{t_{k-1}}^\epsilon(\omega)) = \Pi(\varphi_x(t_{k-1}, \Pi(X_{t_{k-1}}^\epsilon(\omega)), \Pi(X_{t_k}^\epsilon(\omega))))$ and points in $B(x_0, (T - t_0) \|f\|_\infty)$ having the same projection on $G_{\bar{h}}$ are at most $\bar{h}$ distant. Using backward induction, and noticing that $\|X^{\eta}_{t_1}(\omega) - X^{\epsilon}_{t_0}(\omega)\|_1 \leq \|f\|_\infty h$, we have that for all $k = 0 \dots N_h - 1$:

$$\left\| X^{\eta}_{t_{k+1}}(\omega) - X^{\epsilon}_{t_k}(\omega) \right\|_1 \leq \bar{h}(1 + \mathrm{e}^{L_f h}) \sum_{i=0}^{k-1} \mathrm{e}^{\mathrm{i} L_f h} + h \mathrm{e}^{k L_f h} \|f\|_\infty$$
$$\leq 2\bar{h} \frac{T - t_0}{h} \mathrm{e}^{L_f (T - t_0)} + h \mathrm{e}^{L_f (T - t_0)} \|f\|_\infty$$

In order to minimize the distance between $X^{\epsilon}_{\cdot}(\omega)$ and $X^{\eta}_{\cdot}(\omega)$, we set for example $\bar{h} = h^2$ in order to get for all $k = 0 \dots N_h$:

$$\|X^{\epsilon}_{t_k}(\omega) - X^{\eta}_{t_k}(\omega)\|_1 \leq h \left[\mathrm{e}^{L_f (T - t_0)} (2(T - t_0) + \|f\|_\infty) + \|f\|_\infty \right]$$

and for all $t \in [t_k, t_{k+1})$:

$$\|X^{\epsilon}_{t}(\omega) - X^{\eta}_{t}(\omega)\|_1 \leq h \left[\mathrm{e}^{L_f (T - t_0)} (2(T - t_0) + \|f\|_\infty) + 3\|f\|_\infty \right]$$

Finally choosing h small enough:

$$\sup_{t \in [t_0, T]} \|X^{\epsilon}_{t}(\omega) - X^{\eta}_{t}(\omega)\|_1 \leq \epsilon \tag{7.8}$$

It is now easy to check that the final payoff using (u_η, v_η) is close to the payoff generated by (u_ϵ, v_ϵ). Indeed for all $i = 1, 2$:

$$\left| \mathfrak{J}_i(t_0, x_0, u_\epsilon, v_\epsilon) - \mathfrak{J}_i(t_0, x_0, u_\eta, v_\eta) \right| \leq \int_\Omega \left| g_i(X^{\epsilon}_T(\omega)) - g_i(X^{\eta}_T(\omega)) \right| \mathrm{d}\mathbf{P}(\omega)$$
$$\leq \int_\Omega L_g \left\| X^{\epsilon}_T(\omega) - X^{\eta}_T(\omega) \right\|_1 \mathrm{d}\mathbf{P}(\omega)$$
$$\leq L_g \epsilon$$

where L_g is maximum of the Lipschitz constant of the payoff functions g_1 and g_2. Using the Assumption (7.7) on (u_ϵ, v_ϵ), we get for all $i = 1, 2$ and $\epsilon < 1$:

$$\left| \mathbf{E}\left(g_i(X^{\eta}_T) \right) - e_i \right| \leq L_g \epsilon + \epsilon^2 \leq (L_g + 1)\epsilon$$

It remains to prove that the trajectories generated by (u_η, v_η) are consistent. For all $t \in [t_0, T]$, for all $i = 1, 2$, using (7.8) we get:

$$\mathbf{V}_i(t, X^{\eta}_t) \leq \mathbf{E}\left(\mathbf{V}_i(t, X^{\epsilon}_t) \middle| \mathcal{F}^{\eta}_t \right) + L_V \epsilon \tag{7.9}$$

where L_V is maximum of the Lipschitz constant of the value functions $\mathbf{V}_1$ and $\mathbf{V}_2$, and

$$\mathbf{E}\left(g_i(X_T^{\epsilon})\big|\mathcal{F}_t^{\eta}\right) \le \mathbf{E}\left(g_i(X_T^{\eta})\big|\mathcal{F}_t^{\eta}\right) + L_g\epsilon \tag{7.10}$$

We now have to use the Assumption (7.7) on (u_ϵ, v_ϵ): if we denote by

$$\Sigma_{it}^{\epsilon} := \left\{\omega / \ \mathbf{V}_i(t, X_t^{\epsilon}) \le \mathbf{E}\left(g_i(X_T^{\epsilon})\big|\mathcal{F}_t^{\epsilon}\right) + \epsilon^2\right\}$$

we know that $\mathbf{P}(\Sigma_{it}^{\epsilon}) \ge 1-\epsilon^2$. Then, denoting by K an upper bound of the payoff functions, for all $t \in [t_0, T]$, for all $i = 1, 2$, we get:

$$\begin{aligned}\mathbf{V}_i(t, X_t^{\epsilon}) &\le \mathbf{E}\left(g_i(X_T^{\epsilon})\big|\mathcal{F}_t^{\epsilon}\right)\mathbf{1}_{\Sigma_{it}^{\epsilon}} + K\mathbf{1}_{(\Sigma_{it}^{\epsilon})^c} + \epsilon^2 \\ &\le \mathbf{E}\left(g_i(X_T^{\epsilon})\big|\mathcal{F}_t^{\epsilon}\right) + K\mathbf{1}_{(\Sigma_{it}^{\epsilon})^c} + \epsilon\end{aligned}$$

assuming w.l.o.g. that the functions g_i are non negative.

Going back to our estimate of $\mathbf{V}_i(t, X_t^{\eta})$ as computed in (7.9) and noticing that the filtration $(\mathcal{F}_t^{\eta})$ is a subfiltration of $(\mathcal{F}_t^{\epsilon})$, we have:

$$\begin{aligned}\mathbf{V}_i(t, X_t^{\eta}) &\le \mathbf{E}\left[\mathbf{E}\left(g_i(X_T^{\epsilon})\big|\mathcal{F}_t^{\epsilon}\right)\big|\mathcal{F}_t^{\eta}\right] + \mathbf{E}\left[K\mathbf{1}_{(\Sigma_{it}^{\epsilon})^c}\big|\mathcal{F}_t^{\eta}\right] + \epsilon + L_V\epsilon \\ &\le \mathbf{E}\left[g_i(X_T^{\epsilon})\big|\mathcal{F}_t^{\eta}\right] + K\mathbf{P}\left[(\Sigma_{it}^{\epsilon})^c\big|\mathcal{F}_t^{\eta}\right] + (L_V+1)\epsilon \\ &\le \mathbf{E}\left[g_i(X_T^{\eta})\big|\mathcal{F}_t^{\eta}\right] + K\mathbf{P}\left[(\Sigma_{it}^{\epsilon})^c\big|\mathcal{F}_t^{\eta}\right] + (L_V+L_g+1)\epsilon \text{ due to (7.10)}\end{aligned}$$

We rewrite this last inequality introducing the constant $C^* = \max(L_V, L_g, 1, K)$:

$$\mathbf{V}_i(t, X_t^{\eta}) \le \mathbf{E}\left[g_i(X_T^{\eta})\big|\mathcal{F}_t^{\eta}\right] + C^*\mathbf{P}\left[(\Sigma_{it}^{\epsilon})^c\big|\mathcal{F}_t^{\eta}\right] + 3C^*\epsilon \tag{7.11}$$

Using the assumption $\mathbf{P}((\Sigma_{it}^{\epsilon})^c) \le \epsilon^2$, we get that $\mathbf{P}\left\{\mathbf{P}\left((\Sigma_{it}^{\epsilon})^c\big|\mathcal{F}_t^{\eta}\right) \ge \epsilon\right\} \le \epsilon$. This implies for all $t \in [t_0, T]$, for all $i = 1, 2$:

$$\begin{aligned}\mathbf{P}\left\{\mathbf{V}_i(t, X_t^{\eta}) \le \mathbf{E}\left[g_i(X_T^{\eta})\big|\mathcal{F}_t^{\eta}\right] + 4C^*\epsilon\right\} &\ge \mathbf{P}\left\{\mathbf{P}((\Sigma_{it}^{\epsilon})^c|\mathcal{F}_t^{\eta}) \le \epsilon\right\} \\ &\ge 1-\epsilon\end{aligned}$$

Finally, for all $\epsilon > 0$, we have built finitely many controls (u_η, v_η) defining a finite random control satisfying for $\epsilon < 1$ for $i = 1, 2$:

$$\left|\mathbf{E}[g_i(X_T^{\eta})] - e_i\right| \le 2C^*\epsilon$$

and for all $t \in [t_0, T]$ for $i = 1, 2$:

$$\mathbf{P}\left\{\mathbf{V}_i(t, X_t^{\eta}) \le \mathbf{E}\left[g_i(X_T^{t_0, x_0, u_\eta, v_\eta})\big|\mathcal{F}_t^{\eta}\right] + 4C^*\epsilon\right\} \ge 1-\epsilon$$

This proves that e is a MNEP. □

Note that we have in fact proven that any MNEP can be approximated through a PCEP and vice versa.

7.5 Proof of the Main Theorem

Proof. We start with the proof of the necessary condition.

Consider a Nash equilibrium payoff $e = (e_1, e_2)$ and a pair of associated $\frac{\epsilon^2}{2}$-optimal mixed strategies $(\alpha_\epsilon, \beta_\epsilon)$. We will consider the random control defined on $\Omega = \Omega_{\alpha_\epsilon} \times \Omega_{\beta_\epsilon}$ using the probability $\mathbf{P} = \mathbf{P}_{\alpha_\epsilon} \otimes \mathbf{P}_{\beta_\epsilon}$ by $(u_\epsilon, v_\epsilon)(\omega_\alpha, \omega_\beta) = (u_{\omega_\alpha \omega_\beta}, v_{\omega_\alpha \omega_\beta})$. We will denote the associated trajectories by $X^\epsilon_\cdot = X^{t_0, x_0, u_\epsilon, v_\epsilon}_\cdot$. We have for small ϵ, for all $i = 1, 2$:

$$\left|\mathbf{E}[g_i(X^\epsilon_T)] - e_i\right| \leq \frac{\epsilon^2}{2} \leq \epsilon.$$

We will prove that these controls are ϵ-consistent. Suppose on the contrary that there exists $\bar{t} \in [t_0, T]$ such that for example:

$$\mathbf{P}\left\{\mathbf{E}\left(g_1(X^\epsilon_T)\middle|\mathcal{F}_{\bar{t}}\right) \geq \mathbf{V}_1(\bar{t}, X^\epsilon_{\bar{t}}) - \epsilon\right\} < 1 - \epsilon.$$

Denote by

$$\Sigma_\epsilon := \left\{(\omega_\alpha, \omega_\beta) / \ \mathbf{E}\left(g_1(X^\epsilon_T)\middle|\mathcal{F}_{\bar{t}}\right) \geq \mathbf{V}_1(\bar{t}, X^\epsilon_{\bar{t}}) - \epsilon\right\}.$$

As we want to build trigger strategies, we have to introduce Maximin strategies:

Lemma 7.2 (Maximin Strategy). *For all $\epsilon > 0$, for all $t \in (t_0, T)$, there exists $\tau > 0$ such that if we denote by $\mathcal{A}_\tau(t) = \{\alpha \in \mathcal{A}(t) / \ \tau(\alpha) \geq \tau\}$ there exists $\alpha_g^{\epsilon,t} : B(x_0, (t - t_0)\|f\|_\infty) \to \mathcal{A}_\tau(t)$ such that:*

$$\forall x \in B(x_0, (t - t_0)\|f\|_\infty), \ \inf_{v \in \mathcal{V}(t)} g_1\left(X_T^{t,x,\alpha_g^{\epsilon,t}(x)(v),v}\right) \geq \mathbf{V}_1(t, x) - \epsilon$$

Proof of Lemma 7.2. We will build the Maximin strategy $\alpha_g^{\epsilon,t}(\cdot)$ as a collection of finitely many pure strategies with delay. For all $x \in B(x_0, (t - t_0)\|f\|_\infty)$, there exists some pure strategy $\alpha_x \in \mathcal{A}(t)$ such that:

$$\inf_{v \in \mathcal{V}(t)} g_1(X_T^{t,x,\alpha_x(v),v}) \geq \mathbf{V}_1(t, x) - \epsilon/2$$

For continuity reasons, there exists a Borelian partition $(O_i)_{i=1,\ldots I}$ of the ball $B(x_0,(t - t_0)\|f\|_\infty)$ such that for any i there exists some $x_i \in O_i$ such that

$$\forall z \in O_i, \ \inf_{v \in \mathcal{V}(t)} g_1\left(X_T^{t,z,\alpha_{x_i}(v),v}\right) \geq \mathbf{V}_1(t, z) - \epsilon$$

and for all $x \in B(x_0, (t-t_0)\|f\|_\infty)$, we define the Maximin strategy $\alpha_g^{\epsilon,t}(x)$ as the strategy that associates to any $v \in \mathcal{V}(t)$ the control:

$$\alpha_g^{\epsilon,t}(x)(v) = \sum_i \alpha_{x_i}(v)\mathbf{1}_{x \in O_i}$$

Note that we have by construction:

$$\forall x \in B(x_0, (t-t_0)\|f\|_\infty), \; \inf_{v \in \mathcal{V}(t)} g_1\left(X_T^{t,x,\alpha_g^{\epsilon,t}(x)(v),v}\right) \geq \mathbf{V}_1(t,x) - \epsilon$$

As the definition of the Maximin strategy relies on a finite collection of pure strategies with delay, there exists some strictly positive delay τ such that $\forall x \in B(x_0, (t-t_0)\|f\|_\infty)$, $\alpha_g^{\epsilon,t}(x)$ is a pure strategy with delay greater than or equal to τ. □

Choose some delay $\delta > 0$ small enough such that if we denote the Lipschitz constant of the value function $\mathbf{V}_1$ by L, we have $L(1+\|f\|_\infty)\delta \leq \epsilon^2/4$. We now build a mixed strategy α defined on Ω_{α_ϵ} using $\mathbf{P}_{\alpha_\epsilon}$ in the following way: for all $v \in \mathcal{V}(t_0)$

- $\alpha(\omega_{\alpha_\epsilon})(v)(s) \equiv \alpha_\epsilon(\omega_{\alpha_\epsilon})(v)(s)$ for $s \in [t_0, \bar{t}+\delta)$
- If there exists $\omega \in \Omega$ such that $(u_\epsilon, v_\epsilon)(\omega) \equiv (\alpha_\epsilon(\omega_{\alpha_\epsilon})(v), v)$ on $[t_0, \bar{t})$ and $\omega \in \Sigma_\epsilon$, then go on playing $\alpha(\omega_{\alpha_\epsilon})(v)(s) \equiv \alpha_\epsilon(\omega_{\alpha_\epsilon})(v)(s)$ for $s \in [\bar{t}+\delta, T]$
- Else, play $\alpha(\omega_{\alpha_\epsilon})(v) = \alpha_g^{\frac{\epsilon}{4},\bar{t}+\delta}(X_{\bar{t}+\delta}^{t_0,x_0,\alpha(\omega_{\alpha_\epsilon})(v),v})(v|_{[\bar{t}+\delta,T]})$ for all $t \in [\bar{t}+\delta, T]$

Note that $(\alpha(\omega_{\alpha_\epsilon}), \beta_\epsilon(\omega_{\beta_\epsilon}))$ generates the same controls as $(\alpha_\epsilon(\omega_{\alpha_\epsilon}), \beta_\epsilon(\omega_{\beta_\epsilon}))$ for all $(\omega_{\alpha_\epsilon}, \omega_{\beta_\epsilon}) \in \Sigma_\epsilon$ and the same controls as $(\alpha_\epsilon(\omega_{\alpha_\epsilon}), \beta_\epsilon(\omega_{\beta_\epsilon}))$ on $[t_0, \bar{t})$ if $(\omega_{\alpha_\epsilon}, \omega_{\beta_\epsilon}) \notin \Sigma_\epsilon$. Computing the payoff of (α, β_ϵ) and using the fact that Σ_ϵ is $(\mathcal{F}_{\bar{t}})$-measurable:

$$\begin{aligned}
\mathfrak{J}_1(t_0,x_0,\alpha,\beta_\epsilon) &= \mathbf{E}\left(g_1\left(X_T^{\bar{t}+\delta, X^\epsilon_{\bar{t}+\delta}, \alpha_g^{\frac{\epsilon}{4},\bar{t}+\delta}(X_{\bar{t}+\delta}),\beta_\epsilon}\right)\mathbf{1}_{\Sigma_\epsilon^c}\right) + \mathbf{E}(g_1(X_T^\epsilon)\mathbf{1}_{\Sigma_\epsilon}) \\
&\geq \mathbf{E}(\mathbf{V}_1(\bar{t}+\delta, X^\epsilon_{\bar{t}+\delta})\mathbf{1}_{\Sigma_\epsilon^c}) - \frac{\epsilon}{4}(1-\mathbf{P}(\Sigma_\epsilon)) + \mathbf{E}(g_1(X_T^\epsilon)\mathbf{1}_{\Sigma_\epsilon}) \\
&\geq \mathbf{E}(\mathbf{V}_1(\bar{t}, X^\epsilon_{\bar{t}})\mathbf{1}_{\Sigma_\epsilon^c}) - L(1+\|f\|_\infty)\delta - \frac{\epsilon}{4}(1-\mathbf{P}(\Sigma_\epsilon)) \\
&\quad + \mathbf{E}(g_1(X_T^\epsilon)\mathbf{1}_{\Sigma_\epsilon}) \\
&\geq \mathbf{E}(\mathbf{E}(g_1(X_T^\epsilon)|\mathcal{F}_{\bar{t}})\mathbf{1}_{\Sigma_\epsilon^c}) + \mathbf{E}(g_1(X_T^\epsilon)\mathbf{1}_{\Sigma_\epsilon}) + \frac{3\epsilon}{4}(1-\mathbf{P}(\Sigma_\epsilon)) - \epsilon^2/4 \\
&\geq \mathbf{E}(g_1(X_T^\epsilon)\mathbf{1}_{\Sigma_\epsilon^c}) + \mathbf{E}(g_1(X_T^\epsilon)\mathbf{1}_{\Sigma_\epsilon}) + \frac{3\epsilon}{4}(1-\mathbf{P}(\Sigma_\epsilon)) - \epsilon^2/4 \\
&> \mathfrak{J}_1(t_0,x_0,\alpha_\epsilon,\beta_\epsilon) + \frac{\epsilon^2}{2}
\end{aligned}$$

This is in contradiction with the $\frac{\epsilon^2}{2}$-optimality of $(\alpha_\epsilon, \beta_\epsilon)$.

We now will prove the sufficient condition.

Consider some payoff $e = (e_1, e_2)$ reachable and consistent as in Proposition 7.1. For all $\epsilon > 0$, we will build ϵ-optimal strategies rewarding a payoff ϵ close to e.

Fix $\epsilon > 0$. Set δ small enough such that:

1. $\forall t \in [t_0, T]$, $\forall x \in \mathbb{R}^N$, $\forall y \in B(x, \delta \|f\|_\infty)$, for all $i = 1, 2$:

$$|\mathbf{V}_i(t,x) - \mathbf{V}_i(t+\delta, y)| \leq \epsilon \tag{7.12}$$

2. $\forall t \in [t_0, T]$, $\forall x \in \mathbb{R}^N$, $\forall y \in B(x, \delta \|f\|_\infty)$, $\forall (u,v) \in \mathcal{U}(t) \times \mathcal{V}(t)$, for all $i = 1, 2$:

$$|g_i(X_T^{t,x,u,v}) - g_i(X_T^{t,y,u,v})| \leq \epsilon \tag{7.13}$$

3. $\exists N_\delta \in \mathbb{N}^*$ such that $N_\delta \delta = T - t_0$

We introduce the time partition $(\theta_0 = t_0, \ldots, \theta_k = t_0 + k\delta, \ldots, \theta_{N_\delta} = T)$. Set $\eta = \frac{\epsilon}{N_\delta}$. Using the assumption, choose a random control $((\Omega, \mathcal{P}(\Omega), \mathbf{P}), (u_\eta, v_\eta))$ rewarding a payoff η-close to e and η-consistent, denoting by $(\mathcal{F}_t)$ the filtration $(\mathcal{F}_t) = (\sigma\{(u_\eta, v_\eta)(s),\ s \in [t_0, t]\})$:

$$\mathbf{P}\left\{\mathbf{V}_i(t, X_t^{t_0,x_0,u_\eta,v_\eta}) \leq \mathbf{E}[g_i(X_T^{t_0,x_0,u_\eta,v_\eta})|\mathcal{F}_t] + \eta\right\} \geq 1 - \eta \tag{7.14}$$

We will set $X_.^\eta = X_.^{t_0,x_0,u_\eta,v_\eta}$ and for any $\omega \in \Omega$: $X_.^\eta(\omega) = X_.^{t_0,x_0,(u_\eta,v_\eta)(\omega)}$.

If the random control is in fact deterministic, we already know a way to build some pure strategies $(\alpha_\epsilon, \beta_\epsilon)$ that are ϵ-optimal and reward a payoff ϵ-close to e (cf. the construction of Proposition 6.1 in [20] for example). If the controls (u_η, v_η) are real random controls, we have to build ϵ-optimal mixed strategies rewarding a payoff ϵ-close to e. The idea of the optimal strategies $(\alpha_\epsilon, \beta_\epsilon)$ is to build "trigger" mixed strategies that are correlated in order to generate controls close to (u_η, v_η). We will use some jointly correlated lottery at each "node" of the trajectories generated by (u_η, v_η) and, if the opponent does not play the expected control, the player who detected the deviation swaps to the "punitive strategy". The proof proceeds in several steps. First of all, we have to build jointly controlled lotteries for each "node". Then we build the optimal strategies, and check that they reward a payoff close to e and that they are optimal.

To begin with, we introduce the explosions that are kind of "nodes" in the trajectories generated by (u_η, v_η):

Definition 7.10 (Explosion). Consider a finite random control $((\Omega, \mathcal{P}(\Omega), \mathbf{P}), (u_\epsilon, v_\epsilon))$ associated to its natural filtration $(\mathcal{F}_t^\epsilon)$. We set $\mathcal{F}_{t_0^-}^\epsilon = \{\emptyset, \Omega\}$. An explosion is any $t \in [t_0, T)$ such that $\mathcal{F}_{t^-}^\epsilon \neq \mathcal{F}_{t^+}^\epsilon$.

Assume that (u_η, v_η) generates $\bar{M}$ distinct pairs of deterministic controls with $\bar{M} \geq 2$ and M explosions with $1 \leq M \leq \bar{M} - 1$ denoted by $\{\tau_i\}$. We introduce an auxiliary time step τ to be defined later such that $\tau < \min_{j\neq k}|\tau_j - \tau_k|/2$, $\tau < T - \max_j \tau_j$ and $\exists \bar{N} \in \mathbb{N}\backslash\{0,1\}$ such that $\bar{N}\tau = \delta$. This ensures that there is no explosion on $[T-\tau, T]$. We introduce another time partition $(t_0, \dots, t_k = t_0 + k\tau, \dots, t_{N_\delta \bar{N}} = T)$.

We now will explain how to correlate the strategies at each explosion using jointly controlled lotteries.

First note that we can approximate the real probability $\mathbf{P}$ through a probability $\mathbf{Q}$ taking rational values in such a way that the random control $((\Omega, \mathcal{F}_T, \mathbf{Q}), (u_\eta, v_\eta))$ rewards a payoff 2η close to e and 2η consistent: for all $t \in [t_0, T]$:

$$\mathbf{Q}\left\{\mathbf{V}_i(t, X_t^\eta) \leq \mathbf{E}_{\mathbf{Q}}\left(g_i(X_T^\eta)|\mathcal{F}_t\right) + 2\eta\right\} \geq 1 - 2\eta \tag{7.15}$$

7.5.1 *Explosion Procedure*

Suppose $\bar{\tau}$ is an explosion with $\bar{\tau} \in [t_k, t_{k+1})$ and consider $\omega_1, \omega_2 \in \Omega$ such that $\bar{t} = \sup\{t/\ (u_\epsilon, v_\epsilon)(\omega_1)(s) \equiv (u_\epsilon, v_\epsilon)(\omega_2)(s)$ on $[t_0, t]\}$. We assume that the filtration $\mathcal{F}_{t_k}$ is generated by the atoms $\{\Omega_l\}_{l\in L}$. We have for some l: $\omega_1, \omega_2 \in \Omega_l$. By definition of the delay τ, there is no other explosion on (t_k, t_{k+1}). The definition of an explosion allows us to set $\Omega_l := \bigsqcup_{i=1}^{I} \Omega_i^l$ with $\Omega_i^l \in \mathcal{F}_{t_{k+1}}$, $2 \leq I \leq \bar{M}$. We consider the rational conditional probabilities $\mathbf{Q}(\Omega_i^l|\Omega_l) = \frac{q_i(t_k^l)}{q(t_k^l)}$. We build a jointly controlled lottery as in [1]: consider the auxiliary two players process with outcome matrix G, $q(t_k^l)$ distinct actions $u_a : [t_k, t_{k+1}] \to U$ for Player I and $q(t_k^l)$ distinct actions $v_b : [t_k, t_{k+1}] \to V$ for Player II. Note that as we assumed that U and V have infinite cardinality, we can define correlation controls (u_a, v_b) as distinct constant controls and use distinct controls for each explosion. The matrix G is build in such a way that the only possible outcomes are $G(a,b) \in \{1 \dots I\}$ and each row and each column of G contains exactly $q_i(t_k^l)$ times the outcome i for all $i \in \{1 \dots I\}$. Note that if Player II plays some fixed v_b and Player I plays each u_a with equiprobability $\frac{1}{q(t_k^l)}$, then the outcome will be i with probability $\frac{q_i(t_k^l)}{q(t_k^l)} = \mathbf{Q}(\Omega_i^l|\Omega_l)$ and symmetrically, if Player I plays some fixed u_a and Player II plays each v_b with probability $\frac{1}{q(t_k^l)}$, then the outcome will be i with probability $\mathbf{Q}(\Omega_i^l|\Omega_l)$. Note that this explosion procedure allows the players to correlate their controls on any Ω_i^l with probability $\mathbf{Q}(\Omega_i^l|\Omega_l)$ in such a way that no unilateral cheating in the use of the correlation controls may change the outcome of the correlation matrix G. We introduce a way to punish the opponent if he is not playing the expected control through punitive strategies:

Lemma 7.3 (Punitive Strategy). *For all $\epsilon > 0$, for all $t \in (t_0, T)$, there exists $\tau > 0$ such that if we denote by $\mathcal{A}_\tau(t) = \{\alpha \in \mathcal{A}(t)/\ \tau(\alpha) \geq \tau\}$ there exists $\alpha_p^{\epsilon,t} : B(x_0, (t - t_0)\|f\|_\infty) \to \mathcal{A}_\tau(t)$ such that:*

$$\forall x \in B(x_0, (t-t_0)\|f\|_\infty),\ \sup_{v\in\mathcal{V}(t)} g_2\left(X_T^{t,x,\alpha_p^{\epsilon,t}(x)(v),v}\right) \leq \mathbf{V}_2(t,x) + \epsilon$$

Proof of Lemma 7.3. The proof is similar to the proof of Lemma 7.2. □

We now have everything needed to define the ϵ-optimal strategies.

7.5.2 Definition of the Strategies $(\alpha_\epsilon, \beta_\epsilon)$

We recall that the idea of the strategy for Player I is to play the same control as $u_\eta(\omega)$, $\omega \in \Omega$ as long as there is no explosion and as long as Player II plays $v_\eta(\omega)$. If an explosion takes place on $[t_k, t_{k+1})$ meaning $\mathcal{F}_{t_{k+1}}$ is generated by the atoms $(\Omega_i)_{i\in I}$, play on this interval some correlation control as defined by the corresponding explosion procedure. Then observe at t_{k+1} the control played by the opponent on $[t_k, t_{k+1})$ and deduce from the explosion procedure on which Ω_i the game is now correlated and play $u_\eta(\omega_i)$, $\omega_i \in \Omega_i$ from t_{k+1} on until the next explosion as long as Player II plays $v_\eta(\omega_i)$. Player I repeats the same procedure at each explosion. As soon as Player I detects that Player II played some unexpected control, he swaps to the punitive strategy.

In order to define the strategy in a more convenient way, we have to introduce some auxiliary random processes depending only on the past, namely $\bar{\Omega}$ keeping the information on which trajectory generated by (u_η, v_η) is currently being followed and $\mathbf{S}$ such that $\mathbf{S} = \emptyset$ if no deviation was observed in the past and $\mathbf{S} = (t_k, x)$ where $t_k \in \{t_0 \dots t_{N_\delta \bar{N}}\}$ means that some deviation occurred on $[t_k, t_{k+1})$ and the punitive strategy is to be played from the state (t_{k+2}, x) because there is a delay between the time at which deviation is detected and the time from which punitive strategy is played.

First of all, in order to build the strategy α_ϵ for example, we will define the associated underlying finite probability space. We will define it by induction on the number of explosions. We will always assume that an explosion procedure is defined using constant correlation controls that are not used in any other explosion procedure. This allows to build the set Ω_{α_ϵ} by backward induction adding new correlation controls for each explosion.

Any $\omega_{\alpha_\epsilon} \in \Omega_{\alpha_\epsilon}$ prescribes one correlation control for any of the possible explosion procedures. Fix any sequence of correlation controls (u_i) possibly leading to the explosion $\bar{\tau} \in [t_k, t_{k+1})$ associated to the atom Ω_l of $\mathcal{F}_{t_k}$. Consider the set of correlation controls $\{u_a\}$ associated to this explosion. Then, the conditional probability of each u_a given (u_i) is by definition $\frac{1}{q(t_k^l)}$:

$$\mathbf{P}_{\alpha_\epsilon}[\omega_{\alpha_\epsilon} \ni u_a | \omega_{\alpha_\epsilon} \ni (u_i)] = \frac{1}{q(t_k^l)} \tag{7.16}$$

We now define the strategy α_ϵ using auxiliary random processes:

$$\mathbf{S}^{\alpha_\epsilon} : \Omega_{\alpha_\epsilon} \times \mathcal{V}(t_0) \times \{t_k\}_{k=0\dots N_\delta \bar{N}} \to \emptyset \cup \left([t_0, T] \times \mathbb{R}^N\right)$$

and

$$\bar{\Omega}^{\alpha\epsilon} : \Omega_{\alpha_\epsilon} \times \mathcal{V}(t_0) \times \{t_k\}_{k=0\ldots N_\delta \bar{N}} \to \mathcal{F}_T.$$

At time t_0, for any ω_{α_ϵ}, for any control $v \in \mathcal{V}(t_0)$, we set $\mathbf{S}^{\alpha\epsilon}_{t_0}(\omega_{\alpha_\epsilon}, v) = \emptyset$ and $\bar{\Omega}^{\alpha\epsilon}_{t_0}(\omega_{\alpha_\epsilon}, v) = \Omega$ and fix $u_0 \in U$. For all $k \in \{0, \ldots N_\delta \bar{N} - 1\}$, if $\alpha_\epsilon(\omega_{\alpha_\epsilon})(v)$ is built on $[t_0, t_k)$, we define $\alpha_\epsilon(\omega_{\alpha_\epsilon})(v)$ further by:

1. If $\mathbf{S}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v) \neq \emptyset$, for example $\mathbf{S}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v) = (t_i, x)$, this means that Player II did not play the expected control from $t_s \in [t_i, t_{i+1})$ on, then play the punitive strategy $\alpha_\epsilon(v)|_{[t_k, t_{k+1})} = \alpha_p^{\eta, t_{i+2}}(x)(v|_{[t_{i+2}, T]})|_{[t_k, t_{k+1})}$ as defined in Lemma 7.3 and set $\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \emptyset$ and $\mathbf{S}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \mathbf{S}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$.
2. If $\mathbf{S}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v) = \emptyset$, then
   - If there is no explosion on $[t_k, t_{k+1})$ for $(u_\eta, v_\eta)(\omega)$, $\omega \in \bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$, then play $\alpha_\epsilon(\omega_{\alpha_\epsilon})(v)|_{[t_k, t_{k+1})} = u_\eta(\omega)|_{[t_k, t_{k+1})}$ for some $\omega \in \bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$ and set $\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$. If $k \geq 1$ and if $v|_{[t_{k-1}, t_k]} \not\equiv v_\eta(\omega)|_{[t_{k-1}, t_k]}$ for all $\omega \in \bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$ then set $\mathbf{S}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = (t_{k-1}, X^{t_0, x_0, \alpha_\epsilon, v}_{t_{k+1}})$, else set $\mathbf{S}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \emptyset$
   - If there is an explosion on $[t_k, t_{k+1})$ for $(u_\eta, v_\eta)(\omega)$, $\omega \in \bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$, play on $[t_k, t_{k+1})$ the control u_a in ω_{α_ϵ} corresponding to the current explosion procedure then consider the control v played by Player II on $[t_{k-1} \vee t_0, t_k + \frac{\tau}{2}]$:
     - If $k \geq 1$ and if $v|_{[t_{k-1}, t_k]} \not\equiv v_\eta(\omega)|_{[t_{k-1}, t_k]}$ for all $\omega \in \bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v)$ then set $\mathbf{S}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = (t_{k-1}, X^{t_0, x_0, \alpha_\epsilon, v}_{t_{k+1}})$ and $\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \emptyset$ and define $\alpha_\epsilon(\omega_{\alpha_\epsilon})$ further using the procedure at step 1, else
     - If $v|_{[t_k, t_k + \frac{\tau}{2}]} \not\equiv v_b$ for any of the v_b prescribed by the explosion procedure, then set $\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \emptyset$ and $\mathbf{S}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \emptyset$. If $k < N_\delta \bar{N} - 1$, play $\alpha_\epsilon(\omega_{\alpha_\epsilon})(v)|_{[t_{k+1}, t_{k+2})} = u_0$ and set $\mathbf{S}^{\alpha\epsilon}_{t_{k+2}}(\omega_{\alpha_\epsilon}, v) = (t_k, X^{t_0, x_0, \alpha_\epsilon, v}_{t_{k+2}})$ and $\bar{\Omega}^{\alpha\epsilon}_{t_{k+2}}(\omega_{\alpha_\epsilon}, v) = \emptyset$.
     - Else, Player II played one of the expected controls for example v_b. Assume $\bar{\Omega}^{\alpha\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, v) = \Omega_l$. Consider $G(a, b) = \kappa$ and set $\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \Omega^l_\kappa$ and $\mathbf{S}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v) = \emptyset$. If $k < N_\delta \bar{N} - 1$, play $\alpha_\epsilon(\omega_{\alpha_\epsilon})(v)|_{[t_{k+1}, t_{k+2})} = u_\eta(\omega)|_{[t_{k+1}, t_{k+2})}$ for some $\omega \in \bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v)$ and set $\bar{\Omega}^{\alpha\epsilon}_{t_{k+2}}(\omega_{\alpha_\epsilon}, v) = \bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\omega_{\alpha_\epsilon}, v)$ and $\mathbf{S}^{\alpha\epsilon}_{t_{k+2}}(\omega_{\alpha_\epsilon}, v) = \emptyset$.

Note that this procedure indeed defines a mixed strategy. It also ensures that for all $k = 0, \ldots, N_\delta \bar{N}$, at time t_k, $\bar{\Omega}^{\alpha\epsilon}_{t_k}(v)$ is either the empty set or one of the atoms of $\mathcal{F}_{t_k}$ and $\{\mathbf{S}^{\alpha\epsilon}_T(v) \in \{t_k\} \times \mathbb{R}^N\} \in \mathcal{F}^{\alpha_\epsilon, v}_{t_k}$ where $\mathcal{F}^{\alpha_\epsilon, v}_{t_k} = \sigma((\alpha_\epsilon(v), v)(s),\ s \in [t_0, t_k])$.

The strategy β_ϵ is built symmetrically using the auxiliary random processes $\bar{\Omega}^{\beta\epsilon}$ and $\mathbf{S}^{\beta\epsilon}$.

7.5.3 *Payoff of the Strategies* $(\alpha_\epsilon, \beta_\epsilon)$

We will first study the controls generated if Player I plays α_ϵ and Player II plays some pure strategy β with delay $\tau(\beta)$ such that β generates no deviation. We will say that β generates no deviation as soon as for all $k \in \{0,\ldots,N_\delta \bar{N}\}$, $\mathbf{S}^{\alpha_\epsilon}_{t_k}(\beta) = \emptyset$ (equivalently $\mathbf{S}^{\alpha_\epsilon}_T(\beta) = \emptyset$), even if $\mathbf{S}^{\alpha_\epsilon}_T(\beta) = \emptyset$ does not imply that the control generated by β on $[T-\tau, T]$ is one of the v_η.

We will first consider the values taken by the process $\bar{\Omega}^{\alpha_\epsilon}(\beta)$.

Lemma 7.4. *If the strategies* (α_ϵ, β) *are played where* β *is some pure strategy with delay such that for all* $k \in \{0,\ldots,N_\delta \bar{N}\}$, $\mathbf{S}^{\alpha_\epsilon}_{t_k}(\beta) = \emptyset$, *then for all* $k \in \{0,\ldots,N_\delta \bar{N}\}$, *for all* $F \in \mathcal{F}_{t_k}$*:*

$$\mathbf{P}_{\alpha_\epsilon}\left[\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) \subset F\right] = \mathbf{Q}(F)$$

Proof. We will prove the Lemma by induction on k for all F such that F is an atom of the filtration $\mathcal{F}_{t_k}$.

For $k = 0$, this is obviously true for the filtration $\mathcal{F}_{t_0}$ is trivial and $\bar{\Omega}^{\alpha_\epsilon}_{t_0}(\beta) = \Omega$.

Assume that the property of the Lemma is true at stage k, $k < N_\delta \bar{N} - 1$ and that $\mathcal{F}_{t_k}$ is generated by the atoms $\{\Omega^k_i\}_{i\in I}$. We know that for all k, $\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) \in \{\Omega^k_i\}_{i\in I} \cup \emptyset$. Assume now that $\mathcal{F}_{t_{k+1}} = \sigma\left(\{\Omega^{k+1}_j\}_{j\in J}\right)$ where the Ω^{k+1}_j are the atoms of $\mathcal{F}_{t_{k+1}}$. Assume that $\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) = \Omega^k_i$.

- If there exists $j \in J$ such that $\Omega^k_i = \Omega^{k+1}_j$, this means that no explosion takes place on $[t_k, t_{k+1})$ for the controls $(u_\eta, v_\eta)(\omega)$, $\omega \in \Omega^k_i$. As $\mathbf{S}^{\alpha_\epsilon}_{t_k}(\beta) = \emptyset$, the strategy α_ϵ will generate on $[t_k, t_{k+1})$ the control $u_\eta(\omega)$ for any $\omega \in \Omega^k_i$ and we will get $\bar{\Omega}^{\alpha_\epsilon}_{t_{k+1}}(\beta) = \bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) = \Omega^k_i$. This implies $\mathbf{P}_{\alpha_\epsilon}\left[\bar{\Omega}^{\alpha_\epsilon}_{t_{k+1}}(\beta) = \Omega^k_i\right] \geq \mathbf{P}_{\alpha_\epsilon}\left[\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) = \Omega^k_i\right]$. On the other hand, the definition of the process $\bar{\Omega}^{\alpha_\epsilon}(\beta)$ ensures that $\bar{\Omega}^{\alpha_\epsilon}_{t_{k+1}}(\beta) \subseteq \bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta)$ leading to

$$\mathbf{P}_{\alpha_\epsilon}\left[\bar{\Omega}^{\alpha_\epsilon}_{t_{k+1}}(\beta) = \Omega^k_i\right] = \mathbf{P}_{\alpha_\epsilon}\left[\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) = \Omega^k_i\right] = \mathbf{Q}(\Omega^k_i)$$

- Assume now that $\Omega^k_i \neq \Omega^{k+1}_j$ for all $j \in J$. This means there is an explosion on $[t_k, t_{k+1})$ for the controls $(u_\eta, v_\eta)(\omega)$, $\omega \in \Omega^k_i$ and $\Omega^k_i = \bigsqcup_{j=j_0}^{j_i} \Omega^{k+1}_j$. Assume that we have for some ω_{α_ϵ} $\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, \beta) = \Omega^k_i$. Recall that $\mathcal{F}^{\alpha_\epsilon, v}_{t_k} = \sigma((\alpha_\epsilon(v), v)(s),\ s \in [t_0, t_k])$. Note that $\mathbf{S}^{\alpha_\epsilon}_{t_k}(\omega_{\alpha_\epsilon}, \beta) = \emptyset$, implying on $[t_k, t_{k+1})$ the strategy α_ϵ will generate one of the correlation control $u_a \in \Omega_{\alpha_\epsilon}$ prescribed by the explosion procedure for Ω^k_i. The conditional probability that the control generated by α_ϵ at time t_k is u_a given all correlation controls played so far is

$$\mathbf{P}_{\alpha_\epsilon}\left[\alpha_\epsilon(\beta)|_{[t_k, t_{k+1})} = u_a \middle| \mathcal{F}^{\alpha_\epsilon, \beta}_{t_k}\right] = \frac{1}{q(t^i_k)} \times \mathbf{P}_{\alpha_\epsilon}\left[\bar{\Omega}^{\alpha_\epsilon}_{t_k}(\beta) = \Omega^k_i \middle| \mathcal{F}^{\alpha_\epsilon, \beta}_{t_k}\right]$$

due to (7.16) because every correlation control being unique, the only way to play u_a is when $\bar{\Omega}^{\alpha\epsilon}_{t_k}(\beta) = \Omega^k_i$. Given the controls played on $[t_0, t_k)$ for any trajectory such that $\bar{\Omega}^{\alpha\epsilon}_{t_k}(\beta) = \Omega^k_i$, at time t_k, the pure strategy β being a strategy with delay will generate on $[t_k, t_k + \tau(\beta)]$ the same control for example v_b whatever the control u_a chosen by Player I on $[t_k, t_{k+1})$. Note that we must have $v|_{[t_k, t_k+\tau/2)}$ is equivalent to one of the constant correlation controls, else, Player I would detect some deviation at time t_{k+1} and set $\mathbf{S}^{\alpha\epsilon}_{t_{k+2}}(\beta) \neq \emptyset$. In the end, Player II has to play on $[t_k, t_k + \tau/2)$ one of the correlation control v_b, and always plays the same control whatever the control u_a played by Player I. Finally, we will get for all $j = j_0 \dots j_i$:

$$\mathbf{P}_{\alpha\epsilon}\left[\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\beta) = \Omega^{k+1}_j \middle| \mathcal{F}^{\alpha\epsilon,\beta}_{t_k}\right] = \mathbf{Q}(\Omega^{k+1}_j | \Omega^k_i) \times \mathbf{P}_{\alpha\epsilon}\left[\bar{\Omega}^{\alpha\epsilon}_{t_k}(\beta) = \Omega^k_i \middle| \mathcal{F}^{\alpha\epsilon,\beta}_{t_k}\right]$$

and finally taking the expectation w.r.t. $\mathbf{P}_{\alpha\epsilon}$:

$$\begin{aligned}\mathbf{P}_{\alpha\epsilon}\left[\bar{\Omega}^{\alpha\epsilon}_{t_{k+1}}(\beta) = \Omega^{k+1}_j\right] &= \mathbf{Q}(\Omega^{k+1}_j | \Omega^k_i)\mathbf{P}_{\alpha\epsilon}\left[\bar{\Omega}^{\alpha\epsilon}_{t_k}(\beta) = \Omega^k_i\right] \\ &= \mathbf{Q}(\Omega^{k+1}_j | \Omega^k_i)\mathbf{Q}(\Omega^k_i) = \mathbf{Q}(\Omega^{k+1}_j)\end{aligned}$$

We have proven that for all $k \in \{0, \dots, N_\delta \bar{N} - 1\}$, for all atom Ω^k_i of the filtration $\mathcal{F}_{t_k}$,

$$\mathbf{P}_{\alpha\epsilon}\left[\bar{\Omega}^{\alpha\epsilon}_{t_k}(\beta) = \Omega^k_i\right] = \mathbf{Q}(\Omega^k_i).$$

Noticing that there is no explosion on $[T - \tau, T]$, we get $\mathcal{F}_T = \mathcal{F}_{t_{N_\delta \bar{N}-1}}$ and due to the definition of the strategy and the fact that $\mathbf{S}^{\alpha\epsilon}_T(\beta) = \emptyset$, we get $\bar{\Omega}^{\alpha\epsilon}_T(\beta) = \bar{\Omega}^{\alpha\epsilon}_{t_{N_\delta \bar{N}-1}}(\beta)$, hence the result. □

We still assume that Player I plays α_ϵ and Player II plays some pure strategy β such that β generates no deviation and we will compute the payoff $\mathfrak{J}_i(t_0, x_0, \alpha_\epsilon, \beta)$ for $i = 1, 2$.

Lemma 7.5. *If the strategies (α_ϵ, β) are played where β is some pure strategy with delay such that $\mathbf{S}^{\alpha\epsilon}_T(\beta) = \emptyset$, then for all $i = 1, 2$:*

$$|\mathfrak{J}_i(t_0, x_0, \alpha_\epsilon, \beta) - e_i| \leq \frac{3\epsilon}{N_\delta}$$

Corollary. *The strategies $(\alpha_\epsilon, \beta_\epsilon)$ reward a payoff $\frac{3\epsilon}{N_\delta}$ close to e.*

Proof of the Corollary. The proof of the Corollary is straightforward. Indeed, as β_ϵ is a mixed strategy, namely a finite probability distribution on finitely many pure strategies $\beta_\epsilon(\omega_{\beta_\epsilon})$ generating $\mathbf{S}^{\alpha\epsilon}_T(\beta_\epsilon(\omega_{\beta_\epsilon})) = \emptyset$ against α_ϵ, we get for $i = 1, 2$:

$$|\mathfrak{J}_i(t_0, x_0, \alpha_\epsilon, \beta_\epsilon) - e_i| \leq \int_{\Omega_{\beta_\epsilon}} |\mathfrak{J}_i(t_0, x_0, \alpha_\epsilon, \beta_\epsilon(\omega_{\beta_\epsilon})) - e_i| \mathrm{d}\mathbf{P}_{\beta_\epsilon}(\omega_{\beta_\epsilon}) \leq \frac{3\epsilon}{N_\delta} \qquad \square$$

Proof of Lemma 7.5. We recall that the explosions are denoted by τ_i, $i \in I$. For all $i \in I$, there exists $k(\tau_i) \in \{0,\ldots,N_\delta \bar{N}-2\}$ such that $\tau_i \in [t_{k(\tau_i)}, t_{k(\tau_i)+1})$. We denote by

$$\Delta := [t_0, T-\tau) \setminus \left(\cup_{i\in I} [t_{k(\tau_i)}, t_{k(\tau_i)+1}) \right)$$

Assume that $\mathcal{F}_T = \sigma(\{\Omega_j\}_{j=1\ldots\bar{M}})$ where the Ω_j are the atoms of $\mathcal{F}_T$ and players are using (α_ϵ, β) as in the assumptions of the Lemma. Notice that $\forall \omega_{\alpha_\epsilon} \in \{\bar{\Omega}_T^{\alpha_\epsilon}(\beta) = \Omega_j\}$, $\forall \omega_j \in \Omega_j$, the control of Player I generated by $(\alpha_\epsilon(\omega_{\alpha_\epsilon}), \beta)$ satisfies $u_{\alpha_\epsilon(\omega_{\alpha_\epsilon})\beta}(s) \equiv u_\eta(\omega_j)(s)$ $\forall s \in \Delta$. Consequently, $\forall \omega_{\alpha_\epsilon} \in \{\bar{\Omega}_T^{\alpha_\epsilon}(\beta) = \Omega_j\}$, the control of Player II generated by $(\alpha_\epsilon(\omega_{\alpha_\epsilon}), \beta)$ satisfies $v_{\alpha_\epsilon(\omega_{\alpha_\epsilon})\beta}(s) \equiv v_\eta(\omega_j)(s)$ $\forall s \in \Delta$, else we would get $\mathbf{S}_T^{\alpha_\epsilon}(\omega_{\alpha_\epsilon}, \beta) \neq \emptyset$. Therefore, for any ω_{α_ϵ} satisfying $\bar{\Omega}_T^{\alpha_\epsilon}(\omega_{\alpha_\epsilon}, \beta) = \Omega_j$ and any $\omega_j \in \Omega_j$, for all $t \in [t_0, T]$:

$$\left\| X_t^{t_0,x_0,(\alpha_\epsilon,\beta)(\omega_{\alpha_\epsilon})} - X_t^\eta(\omega_j) \right\| \le M\tau(1+\|f\|_\infty)\, \mathrm{e}^{L_f(T-t_0)}$$

where L_f denotes the Lipschitz constant of f and M the number of explosions.

We can choose τ small enough in order that for $i = 1,2$, for all $t \in [t_0, T]$:

$$\begin{cases} \left| g_i(X_T^{t_0,x_0,(\alpha_\epsilon,\beta)(\omega_{\alpha_\epsilon})}) - g_i\left(X_T^\eta(\omega_j)\right) \right| \le \eta \\ \left| \mathbf{V}_i(t, X_t^{t_0,x_0,(\alpha_\epsilon,\beta)(\omega_{\alpha_\epsilon})}) - \mathbf{V}_i(t, X_t^\eta(\omega_j)) \right| \le \eta \end{cases} \tag{7.17}$$

leading, for any $j = 1\ldots\bar{M}$ and any $\omega_j \in \Omega_j$, to

$$\begin{aligned} &\left| \mathbf{E}_{\alpha_\epsilon}\left[g_i(X_T^{t_0,x_0,\alpha_\epsilon,\beta}) \mathbf{1}_{\bar{\Omega}_T^{\alpha_\epsilon}(\beta)=\Omega_j} \right] - \mathbf{E}\left[g_i(X_T^\eta)\mathbf{1}_{\Omega_j} \right] \right| \\ &\quad \le \left| g_i(X_T^\eta(\omega_j))\mathbf{P}_{\alpha_\epsilon}(\bar{\Omega}_T^{\alpha_\epsilon}(\beta) = \Omega_j) - g_i(X_T^\eta(\omega_j))\mathbf{Q}(\Omega_j) \right| + \eta \mathbf{P}_{\alpha_\epsilon}(\bar{\Omega}_T^{\alpha_\epsilon}(\beta) = \Omega_j) \\ &\quad \le \eta \mathbf{Q}(\Omega_j) \text{ due to Lemma 7.4} \end{aligned}$$

Finally, for all $i = 1,2$:

$$\begin{aligned} &|\mathfrak{J}_i(t_0,x_0,\alpha_\epsilon,\beta) - \mathfrak{J}_i(t_0,x_0,u_\eta,v_\eta)| \\ &\quad \le \sum_{j=1}^{\bar{M}} \left| \mathbf{E}_{\alpha_\epsilon}\left[g_i(X_T^{t_0,x_0,\alpha_\epsilon,\beta}) \mathbf{1}_{\bar{\Omega}_T^{\alpha_\epsilon}(\beta)=\Omega_j} \right] - \mathbf{E}\left[g_i(X_T^\eta)\mathbf{1}_{\Omega_j} \right] \right| \\ &\quad \le \sum_{j=1}^{\bar{M}} \eta \mathbf{Q}(\Omega_j) = \eta \end{aligned}$$

Using now (7.15), we have for all $i = 1,2$:

$$\begin{aligned} |\mathfrak{J}_i(t_0,x_0,\alpha_\epsilon,\beta) - e_i| &\le |\mathfrak{J}_i(t_0,x_0,\alpha_\epsilon,\beta) - \mathfrak{J}_i(t_0,x_0,u_\eta,v_\eta)| \\ &\quad + |\mathfrak{J}_i(t_0,x_0,u_\eta,v_\eta) - e_i| \le 3\eta \end{aligned}$$

and the strategies (α_ϵ, β) reward a payoff $\frac{3\epsilon}{N_\delta}$ close to e. □

7.5.4 *Optimality of the Strategies* $(\alpha_\epsilon, \beta_\epsilon)$

It remains to prove that the strategies $(\alpha_\epsilon, \beta_\epsilon)$ are optimal. We will prove it for β_ϵ: there exists some constant C_α satisfying

$$\forall \beta \in \mathcal{B}(t_0),\ \mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta) \leq \mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta_\epsilon) + C_\alpha \epsilon. \tag{7.18}$$

Consider some pure strategy with delay β. If β generates no deviation ($\mathbf{S}_T^{\alpha_\epsilon}(\beta) = \emptyset$), then we have just proven that:

$$\mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta) \leq e_2 + \frac{3\epsilon}{N_\delta} \leq \mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta_\epsilon) + \frac{6\epsilon}{N_\delta}. \tag{7.19}$$

It remains to prove the same kind of result as (7.18) for any pure strategy β generating some unexpected controls (leading for some ω_{α_ϵ} to $\mathbf{S}_T^{\alpha_\epsilon}(\omega_{\alpha_\epsilon}, \beta) \neq \emptyset$). The idea of the proof is first to build some pure strategy $\tilde{\beta}$ generating the same controls as β against α_ϵ as long as no deviation occurs and generating no deviation against α_ϵ, that is some non-deviating extension of β. We then will compare the payoffs induced by β and $\tilde{\beta}$.

Lemma 7.6 (Non Deviating Extension $\tilde{\beta}$ of Some Pure Strategy β of Player II). *To any pure strategy with delay β, one can associate a pure strategy with delay $\tilde{\beta}$ satisfying:*

1. $\mathbf{S}^{\alpha_\epsilon}(\tilde{\beta}) = \emptyset$
2. *The pairs of strategies $(\alpha_\epsilon(\omega_{\alpha_\epsilon}), \beta)$ and $(\alpha_\epsilon(\omega_{\alpha_\epsilon}), \tilde{\beta})$ generate the same pairs of controls on* $[t_0, T-\tau] \times \{\mathbf{S}_T^{\alpha_\epsilon}(\beta) = \emptyset\} \cup_{k \in \{0,\dots,N_\delta \bar{N}\}} [t_0, t_k] \times \{\mathbf{S}_T^{\alpha_\epsilon}(\beta) \in \{t_k\} \times \mathbb{R}^N\}$.

Proof. We juste give the sketch of the proof. The strategy $\tilde{\beta}$ is built the following way. We need auxiliary random processes in order to keep in mind

- Which trajectory generated by (u_η, v_η) is followed.
- Wether β deviated in the past: there exists $t \in (t_0, t_k)$ such that $\mathbf{S}^{\alpha_\epsilon}(\tilde{\beta})_t \neq \emptyset$.
- If the strategy played by Player I is α_ϵ.

For all time interval $[t_k, t_{k+1}]$,

- If Player I deviated from α_ϵ, play any control.
- If β deviated, then play the expected control (either $v_\eta(\omega)$ for the ω corresponding to the followed trajectory or any expected correlation control in case there is some explosion). Then check whether Player I played the expected controls corresponding to α_ϵ.
- If β did not deviate, then if there is no explosion, first play $v_\eta(\omega)$ for the ω to be followed then check if β deviated and if Player I played the expected strategy, if there is some explosion, if β is going to play some expected correlation control on $[t_k, t_k + \tau(\beta)]$ then play this control on the first half of the time interval, then

check if β deviates on $[t_k, t_k + \tau/2]$, if it does not deviate, play β for the remaining time interval and otherwise, play any correct correlation control, then check if Player I deviated.

In this way we are able to build a pure strategy with delay. Indeed, $\tilde{\beta}$ is anticipative with respect to β but non anticipative with respect to the control u of the opponent. Furthermore $\tilde{\beta}$ satisfies $\mathbf{S}_T^{\alpha_\epsilon}(\tilde{\beta}) = \emptyset$ and $\bar{\Omega}_T^{\alpha_\epsilon}(\tilde{\beta}) \neq \emptyset$. As long as β generates no deviation, the controls generated by (α_ϵ, β) and $(\alpha_\epsilon, \tilde{\beta})$ are the same. □

We have for any deviating pure strategy β:

$$\begin{aligned}\mathfrak{J}_2(t_0, x_0, \alpha_\epsilon, \beta) = &\sum_{i=0}^{N_\delta \bar{N}-1} \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\beta})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}\right) \\ &+ \mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\beta}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=\emptyset}\right)\end{aligned} \tag{7.20}$$

Assume that for example $\mathbf{S}_T^{\alpha_\epsilon}(\beta) = (t_i, x)$. This means that some deviation occurred on $[t_i, t_{i+1})$. There exists $k \in \{1 \dots N_\delta\}$ such that $[t_i, t_{i+1}) \subset [\theta_{k-1}, \theta_k)$. Using the definition of the strategy α_ϵ and introducing the non deviating extension $\tilde{\beta}$ of β:

$$\begin{aligned} g_2(X_T^{t_0,x_0,\alpha_\epsilon,\beta})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=(t_i,x)} &= g_2(X_T^{t_{i+2},x,\alpha_p^{\eta,t_{i+2}}(x),\beta})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=(t_i,x)} \\ &\leq (\mathbf{V}_2(t_{i+2},x)+\eta)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=(t_i,x)} \\ &\leq \left[\mathbf{V}_2(t_i, X_{t_i}^{t_0,x_0,\alpha_\epsilon,\beta}) + \eta + \epsilon\right]\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=(t_i,x)} \text{ due to (7.12)} \\ &\leq \left[\mathbf{V}_2(t_i, X_{t_i}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}) + 2\epsilon\right]\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=(t_i,x)} \\ &\leq \left[\mathbf{E}_{\alpha_\epsilon}(\mathbf{V}_2(\theta_k, X_{\theta_k}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})|\mathcal{F}_{t_i}^{\alpha_\epsilon\tilde{\beta}}) + 3\epsilon\right]\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=(t_i,x)} \\ &\quad \text{due to (7.12) because } (\theta_k - t_i) \leq \delta \end{aligned}$$

We introduce this last inequality because our estimate of $\mathbf{V}_2(t_i, X_{t_i}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})$ induces some error term of length η, therefore we need to sum up at most N_δ such error terms in order to bound the global error to some ϵ.

In the end we have for all $t_i \in [\theta_{k-1}, \theta_k)$:

$$g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\beta}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \leq \left[\mathbf{E}_{\alpha_\epsilon}\left(\mathbf{V}_2\left(\theta_k, X_{\theta_k}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)|\mathcal{F}_{t_i}^{\alpha_\epsilon\tilde{\beta}}\right) + 3\epsilon\right]\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \tag{7.21}$$

The point now is to get an estimate of $\mathbf{V}_2(\theta_k, X_{\theta_k}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})$. We will prove the following Lemma:

Lemma 7.7. *For all* $t \in \{t_k\}_{k=0\ldots N_\delta \bar{N}}$*, for all pure strategy* $\tilde{\beta}$ *generating no deviation against* α_ϵ*, we have:*

$$\mathbf{P}_{\alpha_\epsilon}\left\{\mathbf{V}_2(t, X_t^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}) \leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\middle|\mathcal{F}_t^{\alpha_\epsilon\tilde{\beta}}\right) + 4\eta\right\} \geq 1 - 2\eta$$

where $(\mathcal{F}_t^{\alpha_\epsilon\tilde{\beta}}) = \sigma((\alpha_\epsilon,\tilde{\beta})(s), s \in [t_0,t])$

Proof. The proof is straightforward for the trajectories generated by (u_η, v_η) and $(\alpha_\epsilon, \tilde{\beta})$ differs on small time intervals cf. (7.17) and (u_η, v_η) are consistent cf. (7.15). □

We will denote by:

$$\Sigma_t^{\alpha_\epsilon\tilde{\beta}} = \left\{\mathbf{V}_2(t, X_t^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}) \leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\middle|\mathcal{F}_t^{\alpha_\epsilon\tilde{\beta}}\right) + 4\eta\right\}$$

We now will compute a more precise estimate of $\mathbf{V}_2(\theta_k, X_{\theta_k}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})$ denoting by $\|g\|_\infty$ some bound of the payoff functions g_1 and g_2:

$$\begin{aligned}
\mathbf{V}_2(\theta_k, X_{\theta_k}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}) &\leq \left[\mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\middle|\mathcal{F}_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}\right) + 4\eta\right]\mathbf{1}_{\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}} + \|g\|_\infty \mathbf{1}_{(\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}})^c} \\
&\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\middle|\mathcal{F}_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}\right)\mathbf{1}_{\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}} + \|g\|_\infty \mathbf{1}_{(\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}})^c} + 4\eta \\
&\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\middle|\mathcal{F}_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}\right) + \|g\|_\infty \mathbf{1}_{(\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}})^c} + 4\eta \qquad (7.22)
\end{aligned}$$

assuming g_2 is non negative, which is possible without lack of generality because this function is bounded.

It remains to introduce this estimate (7.22) in inequality (7.21) to get for all $i \in \{0,\ldots,N_\delta\bar{N}-1\}$, if $t_i \in [\theta_{(k-1)}, \theta_k)$:

$$\begin{aligned}
& g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\beta}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \\
&\quad \leq \left[\mathbf{E}_{\alpha_\epsilon}\left(\mathbf{V}_2\left(\theta_k, X_{\theta_k}^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\middle|\mathcal{F}_{t_i}^{\alpha_\epsilon\tilde{\beta}}\right) + 3\epsilon\right]\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \\
&\quad \leq \mathbf{E}_{\alpha_\epsilon}\left(\mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\middle|\mathcal{F}_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}\right) + \|g\|_\infty \mathbf{1}_{\left(\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}}\right)^c} + 4\eta\middle|\mathcal{F}_{t_i}^{\alpha_\epsilon\tilde{\beta}}\right) \\
&\qquad \times \mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \\
&\qquad + 3\epsilon\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \text{ thanks to (7.22)} \\
&\quad \leq \mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\middle|\mathcal{F}_{t_i}^{\alpha_\epsilon\tilde{\beta}}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \\
&\qquad + \|g\|_\infty \mathbf{P}_{\alpha_\epsilon}\left((\Sigma_{\theta_k}^{\alpha_\epsilon\tilde{\beta}})^c\middle|\mathcal{F}_{t_i}^{\alpha_\epsilon\tilde{\beta}}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} + 7\epsilon\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}
\end{aligned}$$

Note that at time t_i there is no deviation, implying $\mathcal{F}_{t_i}^{\alpha\epsilon\tilde{\beta}} = \mathcal{F}_{t_i}^{\alpha\epsilon\beta}$ and

$$g_2(X_T^{t_0,x_0,\alpha_\epsilon,\beta})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\big|\mathcal{F}_{t_i}^{\alpha\epsilon\beta}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}$$
$$+\|g\|_\infty \mathbf{P}_{\alpha_\epsilon}\left((\Sigma_{\theta_k}^{\alpha\epsilon\tilde{\beta}})^c\big|\mathcal{F}_{t_i}^{\alpha\epsilon\beta}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} + 7\epsilon\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}$$

Using the fact that $\{\mathbf{S}^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N\}$ is $(\mathcal{F}_{t_i}^{\alpha\epsilon\beta})$-measurable due to the definition of the strategy α_ϵ, we get for $i = 0\ldots N_\delta\bar{N}-1$, if $t_i \in [\theta_{k-1},\theta_k)$:

$$g_2(X_T^{t_0,x_0,\alpha_\epsilon,\beta})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}\big|\mathcal{F}_{t_i}^{\alpha\epsilon\beta}\right)$$
$$+\|g\|_\infty \mathbf{E}_{\alpha_\epsilon}\left(\mathbf{1}_{(\Sigma_{\theta_k}^{\alpha\epsilon\tilde{\beta}})^c}\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}\big|\mathcal{F}_{t_i}^{\alpha\epsilon\beta}\right) + 7\epsilon\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N} \tag{7.23}$$

We now use this estimate to compute the expectation of the payoff in case there is some deviation:

$$\sum_{i=0}^{N_\delta\bar{N}-1}\mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\beta}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}\right)$$
$$\leq \sum_{i=0}^{N_\delta\bar{N}-1}\mathbf{E}_{\alpha_\epsilon}\left(\mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}|\mathcal{F}_{t_i}^{\alpha\epsilon\beta}\right)\right)$$
$$+\sum_{k=1}^{N_\delta}\sum_{i=(k-1)\bar{N}}^{k\bar{N}-1}\mathbf{E}_{\alpha_\epsilon}\left(\|g\|_\infty\mathbf{E}_{\alpha_\epsilon}\left(\mathbf{1}_{(\Sigma_{\theta_k}^{\alpha\epsilon\tilde{\beta}})^c}\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}|\mathcal{F}_{t_i}^{\alpha\epsilon\beta}\right)\right)$$
$$+7\epsilon \text{ due to (7.23)}$$
$$\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\neq\emptyset}\right)$$
$$+\|g\|_\infty\sum_{k=1}^{N_\delta}\mathbf{E}_{\alpha_\epsilon}\left(\mathbf{1}_{(\Sigma_{\theta_k}^{\alpha\epsilon\tilde{\beta}})^c}\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in[\theta_{k-1},\theta_k)\times\mathbb{R}^N}\right)+7\epsilon$$
$$\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\neq\emptyset}\right)+\|g\|_\infty\sum_{k=1}^{N_\delta}\mathbf{P}_{\alpha_\epsilon}\left(\left(\Sigma_{\theta_k}^{\alpha\epsilon\tilde{\beta}}\right)^c\right)+7\epsilon$$
$$\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\neq\emptyset}\right)+\|g\|_\infty\sum_{k=1}^{N_\delta}\frac{2\epsilon}{N_\delta}+7\epsilon \text{ thanks to Lemma 7.7}$$
$$\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2\left(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}}\right)\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\neq\emptyset}\right)+2\|g\|_\infty\epsilon+7\epsilon \tag{7.24}$$

Going back to our estimate of $\mathfrak{J}_2(t_0,x_0,\alpha_\epsilon,\beta)$ as in (7.20) we have:

$$\begin{aligned}
&\mathfrak{J}_2(t_0,x_0,\alpha_\epsilon,\beta)\\
&= \sum_{i=0}^{N_\delta \bar{N}-1} \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\beta})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\in\{t_i\}\times\mathbb{R}^N}\right) + \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=\emptyset}\right)\\
&\leq \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)\neq\emptyset}\right) + (2\|g\|_\infty + 7)\epsilon\\
&\quad + \mathbf{E}_{\alpha_\epsilon}\left(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})\mathbf{1}_{\mathbf{S}_T^{\alpha_\epsilon}(\beta)=\emptyset}\right) \text{ due to (7.24)}\\
&\leq \mathbf{E}_{\alpha_\epsilon}(g_2(X_T^{t_0,x_0,\alpha_\epsilon,\tilde{\beta}})) + (2\|g\|_\infty + 7)\epsilon\\
&\leq \mathfrak{J}_2(t_0,x_0,\alpha_\epsilon,\beta_\epsilon) + \frac{6\epsilon}{N_\delta} + (2\|g\|_\infty + 7)\epsilon \text{ thanks to (7.19)}
\end{aligned}$$

This proves that β_ϵ is $(13+2\|g\|_\infty)\epsilon$ optimal. The proof is symmetric to state that α_ϵ is $(13+2\|g\|_\infty)\epsilon$ optimal.

Finally, we have build mixed strategies $(\alpha_\epsilon, \beta_\epsilon)$ rewarding a payoff 3ϵ close to e and $(13+2\|g\|_\infty)\epsilon$ optimal. This proves e is a Nash equilibrium payoff.

□

References

1. Aumann, R.J., Maschler, M.B.: Repeated Games with Incomplete Information. MIT Press, Cambridge (1995)
2. Aumann, R.J., Shapley, L.S.: Long-Term Competition – A Game Theoretic Analysis Mimeo, Hebrew University (1976), reprinted in Essays in Game Theory in Honor of Michael Maschler, (N. Megiddo, ed.), Springer-Verlag, 1–15 (1994)
3. Basar, T., Olsder, G.J.: Dynamic Noncooperative Game Theory, 2nd ed. Academic, London (1995)
4. Bensoussan, A., Frehse, J.: Stochastic games for N players. J. Optim. Theory Appl. **105**(3), 543–565 (2000)
5. Bressan, A., Shen, W.: Semi-cooperative strategies for differential games. Int. J. Game Theory **32**, 561–593 (2004)
6. Bressan, A., Shen, W.: Small BV solutions of hyperbolic noncooperative differential games. SIAM J. Control Optim. **43**(1), 194–215 (2004)
7. Buckdahn, R., Cardaliaguet, P., Rainer, C.: Nash equilibrium payoffs for nonzero sum stochastic differential games. SIAM J. Control Optim. **43**(2), 624–642 (2004)
8. Cardaliaguet, P.: Representations formulae for some differential games with asymetric information. J. Optim. Theory Appl. **138**(1), 1–16 (2008)
9. Cardaliaguet, P., Quincampoix, M.: Deterministic differential games under probability knowledge of initial condition. Int. Game Theory Rev. **10**(1), 1–16 (2008)
10. Dutta, P.K.: A folk theorem for stochastic games. J. Econ. Theory **66**(1), 1–32 (1995)
11. Engwerda, J.C.: LQ Dynamic Optimization and Differential Games. Wiley, New York (2005)
12. Friedman, A.: Differential Games, Wiley-Interscience, New York (1971)

13. Gossner, O.: The folk theorem for finitely repeated games with mixed strategies. Int. J. Game Theory **24**, 95–107 (1995)
14. Hamadène, S., Lepeltier, J.-P., Peng, S.: BSDEs with continuous coefficients and stochastic differential games. In: El Karoui et al. (eds.) Backward Stochastic Differential Equations, Pitman Res. Notes in Math. Series, vol. 364, pp.161–175. Longman, Harlow (1997)
15. Kleimenov, A.F.: Nonantagonist Differential Games. "Nauka" Uralprime skoje Otdelenie Ekaterinburg (1993) (in russian)
16. Kononenko, A.F.: Equilibrium positional strategies in non-antagonistic differential games Dokl. Akad. Nauk SSSR **231**(2), 285–288 (1976). English translation: Soviet Math. Dokl. **17**(6), 1557–1560 (1977) (in Russian)
17. Rainer, C.: On two different approaches to nonzero sum stochastic differential games. Appl. Math. Optim. **56**, 131–144 (2007)
18. Rubinstein, A.: Equilibrium in supergames with the overtaking criterion. J. Econ. Theory **31**, 227–250 (1979)
19. Solan, E.: Characterization of correlated equilibria in stochastic games. Int. J. Game Theory **30**, 259–277 (2001)
20. Tolwinski, B., Haurie, A., Leitmann, G.: Cooperative equilibria in differential games. J. Math. Anal. Appl. **119**, 182–202 (1986)

Chapter 8
A Penalty Method Approach for Open-Loop Variational Games with Equality Constraints

Dean A. Carlson and George Leitmann

Abstract In this paper we consider the possibility of solving a variational game with equality constraints by using a penalty method approach. Under the assumption that the unconstrained penalized games have open loop Nash equilibria we give conditions on our model to ensure that there exists a subsequence of penalty parameters converging to infinity for which the corresponding sequence of solutions to the penalized games converges to an open loop Nash equilibrium of the constrained game. Our conditions are based on classical growth and convexity conditions found in the calculus of variations. We conclude our paper with some remarks on obtaining the solutions of the penalized games via Leitmann's direct method.

Keywords Differential games • Nash equilibria • Equality constraints • Penalty methods

8.1 Introduction

For the past several years, we have been studying a class of variational games which may be viewed as an extension of the calculus of variations. In particular, our focus has been on exploiting a direct solution method, originally due to G. Leitmann in [4], to investigate sufficient conditions for open-loop Nash equilibria. The study of such problems pre-dates J. Nash's work in non-cooperative games, and their study can be

D.A. Carlson (✉)
American Mathematical Society, Mathematical Reviews, 416 Fourth Street, Ann Arbor, MI 48103, USA
e-mail: dac@ams.org

G. Leitmann
Graduate School, University of California at Berkeley, Berkeley, CA 94720, USA
e-mail: gleit@berkeley.edu

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_8,

found in the 1920s with a series of mathematical papers by Roos [5–10] exploring the dynamics of competition in economics. The last of Roo's papers, provides an extensive investigation into general variational games and provides analogues of the standard first-order necessary conditions, such as the Euler–Lagrange equations, the Weierstrass necessary condition, transversality conditions, Legendre's necessary condition and the Jacobi necessary condition. To date, most of these papers dealt only with unconstrained problems (i.e., free problems of Lagrange type). In this paper we investigate problems with equality constraints. Our approach is to consider the feasibility of a penalty method for these problems which extends our recent paper Carlson and Leitmann [1] from the case of a single-player game to an N-player game. Penalty methods, of course, are not new and they have been used in a variety of settings. However, in the study of games a quick search of MathSciNet produced only 22 papers pertaining to penalty methods and games.

The remainder of the paper is organized as follows. In Sect. 8.2, we define the class of games we consider and introduce the penalized game. In the next section we digress to discuss some relevant results concerning growth conditions and sequentially weak relative compactness. We prove our main result in Sect. 8.4. In Sect. 8.5 we present an example illustrating our results and we conclude with some brief remarks indicating how other known techniques might be useful.

8.2 The Class of Games Considered

We consider an N-person game in which the state of player $j = 1, 2, \ldots, N$ is a real-valued function $x_j(\cdot) : [a,b] \to \mathbb{R}^{n_j}$ with fixed initial value $x_j(a) = x_{aj}$. The objective of each player is to minimize a Lagrange type functional

$$I_j(\mathbf{x}(\cdot)) = \int_a^b L_j(t, \mathbf{x}(t), \dot{x}_j(t))\,\mathrm{d}t, \tag{8.1}$$

over all of his/her possible admissible trajectories (see below), $\dot{x}_j(\cdot)$ satisfying the fixed end condition $\mathbf{x}(a) = \mathbf{x}_a$ and the equality constraint

$$g_j(t, \mathbf{x}(t), \dot{x}_j(t)) = 0, \quad \text{a.e. } a < t < b. \tag{8.2}$$

The notation used here is that $\mathbf{x} = (x_1, x_2, \ldots x_N) \in \prod_{j=1}^{N} \mathbb{R}^{n_j} \doteq \mathbb{R}^{\mathbf{n}}$, in which $\mathbf{n} = n_1 + n_2 + \cdots + n_N$. We assume, for each $j = 1, 2, \ldots N$, that $(L_j, g_j)(\cdot,\cdot,\cdot) : A_j \to \mathbb{R}^2$ is a continuous function defined on the open set $A_j \subset \mathbb{R} \times \mathbb{R}^{\mathbf{n}} \times \mathbb{R}^{n_j}$ with the additional property that each $g_j(t, \mathbf{x}, p_j) \geq 0$ for all $(t, \mathbf{x}, p_j) \in A_j$.

Remark 8.1. The above framework does allow us to include more general equality constraints. Indeed if $g_j(t, \mathbf{x}, p_j)$ is not nonnegative we can replace it by its square. Additionally, if there is more than one constraint we can simply add them together.

Clearly, the trajectories of the other players influences the decision of the jth player and so each player is unable to minimize independently of the other players. As a consequence, the players seek to play a (open-loop) Nash equilibrium instead. To introduce this concept we first introduce the following notation. For each fixed $j = 1, 2, \ldots, N$, $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^{\mathbf{n}}$, and $y_j \in \mathbb{R}^{n_j}$ we use the notation

$$[\mathbf{x}^j, y_j] \doteq (x_1, x_2, \ldots, x_{j-1}, y_j, x_{j+1}, \ldots x_N)$$

and if $\mathbf{x}(\cdot) : [a,b] \to \mathbb{R}^{\mathbf{n}}$ and $y(\cdot) : [a,b] \to \mathbb{R}^{n_j}$ we use the notation

$$[\mathbf{x}^j, y](\cdot) \doteq (x_1(\cdot), x_2(\cdot), \ldots, x_{j-1}(\cdot), y(\cdot), x_{j+1}(\cdot), \ldots x_N(\cdot))$$

With this notation we have the following definitions.

Definition 8.1. We say a function $\mathbf{x}(\cdot) = (x_1(\cdot), x_2(\cdot), \ldots, x_N(\cdot)) : [a,b] \to \mathbb{R}^{\mathbf{n}}$ is an admissible trajectory for the constrained variational game (8.1), (8.2) if and only if it is absolutely continuous, satisfies the fixed end conditions

$$x_j(a) = x_{aj}, \quad j = 1, 2, \ldots, N, \tag{8.3}$$

satisfies the equality constraints (8.2), satisfies $(t, x_j(t), \dot{x}_j(t)) \in A_j$ for almost all $t \in [a,b]$ and such that $I_j(\mathbf{x}(\cdot))$ exists for all $j = 1, 2, \ldots, N$.

Definition 8.2. Given an admissible trajectory $\mathbf{x}(\cdot)$ for the constrained variational game (8.1), (8.2) we say a function $y_j(\cdot) : [a,b] \to \mathbb{R}^{n_j}$ is an admissible trajectory for player j relative to $\mathbf{x}(\cdot)$ if and only if the function $[\mathbf{x}^j, y_j](\cdot)$ is an admissible trajectory for the constrained variational game.

With these definitions we can now give the definition of a Nash equilibrium.

Definition 8.3. An admissible trajectory for the constrained variational game (8.1), (8.2) $\mathbf{x}^*(\cdot) : [a,b] \to \mathbb{R}^{\mathbf{n}}$ is called a Nash equilibrium if and only if for each player $j = 1, 2, \ldots, N$ and each function $y_j(\cdot) : [a,b] \to \mathbb{R}^{n_j}$ that is an admissible trajectory for player j relative to $\mathbf{x}^*(\cdot)$ one has

$$\begin{aligned} I_j(\mathbf{x}^*(\cdot)) &= \int_a^b L_j(t, \mathbf{x}^*(t), \dot{x}_j^*(t))\,\mathrm{d}t \\ &\leq \int_a^b L_j(t, [\mathbf{x}^*(t)^j, y_j(t)], \dot{y}_j(t))\,\mathrm{d}t \\ &= I_j([\mathbf{x}^{*j}, y_j](\cdot)). \end{aligned} \tag{8.4}$$

Remark 8.2. From the above definitions it is clear that when all of the players "play" a Nash equilibrium, then each player's strategy is his best response to that of the other players. In other words, if player j applies any other admissible trajectory relative to the Nash equilibrium, than his equilibrium trajectory, his cost functional will not decrease.

Remark 8.3. The above dynamic game clearly is not the most general structure one can imagine, even in a variational framework. In particular, the cost functionals are coupled only through their state variables and not through their strategies (i.e., their time derivatives). While not the most general, one can argue that this form is general enough to cover many cases of interest since in a "real-world setting," an individual player will not know the strategies of the other players (see e.g., Dockner and Leitmann [3]).

To solve games of the type described above, one usually tries to solve the first-order necessary conditions to obtain a candidate for the Nash equilibrium and then apply a sufficient condition to verify that it is one. For the above constrained problem, the first-order necessary conditions are complicated as a result of the equality constraints. For such problems one must find a multiplier for each of the constraints which in its most general form is a measure. As a consequence of this fact we choose to consider a family of unconstrained games in which the objective of each player incorporates the constraint multiplied by a positive constant. We now describe this family of games.

8.2.1 *The Penalized Games*

To define the penalized games for each $\lambda > 0$ define the function $L_{\lambda,j} : A_j \to \mathbb{R}$ by the formula

$$L_{\lambda,j}(t,\mathbf{x},p_j) = L(t,\mathbf{x},p_j) + \lambda g_j(t,\mathbf{x},p_j), \tag{8.5}$$

for each $(t,\mathbf{x},p_j) \in A_j$. With this integrand we consider the unconstrained game in which each player tries to minimize the integral functional

$$I_{\lambda,j}(\mathbf{x}(\cdot)) = \int_a^b L_{\lambda,j}(t,\mathbf{x}(t),\dot{x}_j(t))\,\mathrm{d}t, \quad j = 1,2,\ldots N, \tag{8.6}$$

over all of his admissible trajectories $x_j(\cdot)$ satisfying the fixed end condition $x_j(a) = x_{aj}$. Of course, the set of admissible trajectories for this family of unconstrained games is larger than the set of admissible trajectories for the original constrained game. For completeness we give the following definitions.

Definition 8.4. For a given $\lambda > 0$, a function $\mathbf{x}(\cdot) : [a,b] \to \mathbb{R}^{\mathbf{n}}$ is an admissible trajectory for the unconstrained game (8.6) if it is absolutely continuous, satisfies the fixed end condition (8.3), satisfies $(t,\mathbf{x}(t),\dot{x}_j(t)) \in A_j$ for almost all $t \in [a,b]$ and $I_{\lambda,j}(\mathbf{x}(\cdot))$ exists for all $i = 1,2,\ldots,N$.

Definition 8.5. Given an admissible trajectory $\mathbf{x}(\cdot)$ for the unconstrained game (8.6), we say a function $y_j(\cdot) : [a,b] \to \mathbb{R}^{n_j}$ is admissible for player j relative to $\mathbf{x}(\cdot)$ if the trajectory $[\mathbf{x}(\cdot)^j, y_j(\cdot)]$ is an admissible trajectory for the unconstrained game (8.6).

Definition 8.6. Given a fixed $\lambda > 0$, we say an admissible trajectory $\mathbf{x}^*_\lambda(\cdot) : [a,b] \to \mathbb{R}^{\mathbf{n}}$ for the unconstrained variational (8.6) is a Nash equilibrium of for each $j = 1,2,\ldots N$ and any function $y_j(\cdot) : [a,b] \to \mathbb{R}^{n_j}$ that is admissible for player j relative to $\mathbf{x}^*_\lambda(\cdot)$ one has

$$\begin{aligned} I_{\lambda,j}(\mathbf{x}^*_\lambda(\cdot)) &= \int_a^b L_{\lambda,j}(t,\mathbf{x}^*_\lambda(t),\dot{x}^*_{\lambda j}(t))\,\mathrm{d}t \\ &\leq \int_a^b L_{\lambda,j}(t,[\mathbf{x}^*_\lambda(t)^j, y_j(t)],\dot{y}_j(t))\,\mathrm{d}t \\ &= I_{\lambda,j}([\mathbf{x}^{*j}_\lambda, y_j](\cdot)). \end{aligned} \tag{8.7}$$

We notice that if $\mathbf{y}(\cdot)$ is an admissible trajectory for the constrained game (8.1), (8.2) then it is an admissible trajectory for the unconstrained game (8.6) for any value of $\lambda \geq 0$ and $I_j(\mathbf{y}(\cdot)) = I_{\lambda,j}(\mathbf{y}(\cdot))$. Thus, if it is the case that $\mathbf{x}^*_\lambda(\cdot)$ is both a Nash equilibrium for the unconstrained game (8.6) and if it is also an admissible trajectory for the constrained game (8.1), (8.2), then it is a Nash equilibrium for the constrained game. Indeed if $y(\cdot)$ is admissible for player j for the constrained game relative to $\mathbf{x}^*_\lambda(\cdot)$ (which implies that $g_j(t,[\mathbf{x}^*_\lambda(t)^j, y(t)_j],\dot{y}(t)) = 0$) then we have

$$I_j(\mathbf{x}^*_\lambda(\cdot)) = I_{\lambda,j}(\mathbf{x}^*_\lambda(\cdot)) \leq I_{\lambda,j}([\mathbf{x}^{*j}_\lambda, y](\cdot)) = I_j([\mathbf{x}^{*j}_\lambda, y](\cdot)).$$

The above observation is useful only if we find that a Nash equilibrium for one of the penalized games is an admissible trajectory for the constrained game. The idea of a penalty method is that as the penalty parameter λ grows to infinity the penalized term tends to zero. We now give conditions for when this occurs.

Lemma 8.1. *Assume for each $j = 1,2,\ldots,N$ that there exists constants A^*_j and B^*_j such that for each admissible trajectory for the unconstrained games, $\mathbf{x}(\cdot)$ one has $A^*_j \leq I_j(\mathbf{x}(\cdot)) \leq B^*_j$. Further assume that there exists a $\lambda_0 > 0$ such that for all $\lambda > \lambda_0$ the unconstrained penalized games have Nash equilibria $\mathbf{x}^*_\lambda(\cdot)$ and that corresponding to each there exists an absolutely continuous function $\mathbf{y}_\lambda(\cdot)$ such that for each $j = 1,2,\ldots,N$ the trajectories $[\mathbf{x}^{*j}, y_{\lambda,j}](\cdot)$ are admissible for the constrained game (i.e., $g(t,[\mathbf{x}^{*j}, y_{\lambda,j}](t),\dot{y}_{\lambda j}(t)) = 0$ a.e. $t \in [a,b]$). Then one has,*

$$\lim_{\lambda \to +\infty} \int_a^b g_j(t,\mathbf{x}^*_\lambda(t),\dot{x}^*_{\lambda,j}(t))\,\mathrm{d}t = 0, \quad j = 1,2,\ldots,N.$$

Proof. To prove this result we proceed by contradiction and assume that for some $j = 1,2\ldots,N$ there exists a sequence $\{\lambda_k\}$ and an $\epsilon_0 > 0$ such that

$$\int_a^b g_j\left(t,\mathbf{x}^*_{\lambda_k}(t),\dot{x}^*_{\lambda_k,j}(t)\right)\,\mathrm{d}t > \epsilon_0.$$

From our assumptions we now have the following inequalities

$$\begin{aligned} A_j^* + \epsilon_0 \lambda_k &\leq I_{\lambda_k,j}\left(\mathbf{x}^*_{\lambda_k}(\cdot)\right) \\ &\leq I_{\lambda_k,j}\left(\left[\mathbf{x}^{*j}_{\lambda_k}, y_{\lambda_k,j}\right](\cdot)\right) \\ &= I_j\left(\left[\mathbf{x}^{*j}_{\lambda_k}, y_{\lambda_k,j}\right](\cdot)\right) \\ &\leq B_j^* \end{aligned}$$

for each $k = 1, 2, \ldots$. Letting $k \to +\infty$ produces the obvious contradiction. □

Remark 8.4. In the above, we gave conditions under which the penalty term vanishes as the penalty parameter $\lambda \to +\infty$. This however does not imply that

$$\lim_{\lambda \to \infty} g_j(t, \mathbf{x}^*_\lambda(t), \dot{x}^*_{\lambda,j}(t)) = 0$$

for almost all $t \in [a,b]$. Furthermore, we also know nothing about the convergence of the trajectories $\{\mathbf{x}^*_\lambda(\cdot)\}$ as $\lambda \to \infty$.

Remark 8.5. The existence of the A_j^*'s can be realized by assuming that the integrands $L_j(\cdot,\cdot,\cdot)$ are bounded below, which is not an unusual assumption for minimization problems. The existence of the admissible trajectories $\mathbf{y}_\lambda(\cdot)$ is much more difficult to satisfy, but it is easy to see that such a trajectory exists if the equality constraints are not dependent on the other players and if there exists feasible trajectories for the original constrained game. That is, $g_j(\cdot,\cdot,\cdot) : [a,b] \times \mathbb{R}^{n_j} \times \mathbb{R}^{n_j} \to \mathbb{R}$ and there exists at least one trajectory $\mathbf{y}(\cdot)$ satisfying the fixed end condition (8.3) such that

$$g_j(t, y_j(t), \dot{y}_j(t)) = 0, \quad \text{a.e. } t \in [a,b],\ j = 1, 2, \ldots, N.$$

In this case one only needs to take $\mathbf{y}_\lambda(\cdot) = \mathbf{y}(\cdot)$ for all $\lambda > \lambda_0$. Finally, the existence of the constants B_j^* is perhaps the most difficult to verify, unless one assumes that the integrands $L_j(\cdot,\cdot,\cdot)$ are also bounded above. However, we note that in our proof, this condition can be weakend slightly by assuming that for each $j = 1, 2, \ldots, N$ one has $I_j(\mathbf{x}(\cdot)) \leq B_j^*$ for all feasible trajectories for the original constrained game (8.1), (8.2).

We now begin to investigate the convergence properties of the family of Nash equilibria $\left\{\mathbf{x}^*_\lambda(\cdot)\right\}$.

8.3 Growth Conditions and the Compactness of the Set of Nash Equilibria

In this section we begin by reviewing some classical notions concerning the weak topology of absolutely continuous functions and criteria for compactness of a sequence of absolutely continuous functions. Following this discussion we apply these ideas to our game model and the compactness of the set of Nash equilibria $\{\mathbf{x}_\lambda(\cdot)\}_{\lambda>0}$. Following this result we discuss the lower semicontinuity properties of the integral functionals $I_j(\cdot)$ and $I_{\lambda,j}(\cdot)$ with respect to the weak topology of absolutely continuous funcntions. This will allow us to present our main result in the next section. These questions have their roots in the classical existence results of the calculus of variations.

The existence theory of the calculus of variations is a delicate balance between the compactness properties of sets of admissible trajectories and the conditions imposed on the integral functional to insure lower semicontinuity. Fortunately, this is a well studied problem for the cases we consider here and indeed the results are now classical. We begin first by discussing growth conditions and the weak topology in the class of absolutely continuous functions.

The space of absolutely continuous functions, denoted as $AC([a,b];\mathbb{R}^m)$, is a subspace of the set of continuous functions $z(\cdot):[a,b]\to\mathbb{R}^m$ with the property that their first derivatives $\dot{z}(\cdot)$ are Lebesgue integrable. Clearly they include the class of piecewise smooth trajectories. Further, we also know that the fundamental theorem of calculus holds, i.e.,

$$z(t) = z(a) + \int_a^t \dot{z}(s)\,\mathrm{d}s$$

for every $t\in[a,b]$ and moreover, whenever

$$z(t) = z(a) + \int_a^t \xi(s)\,\mathrm{d}s, \quad a\le t\le b,$$

holds for some Lebesgue integrable function $\xi(\cdot):[a,b]\to\mathbb{R}^m$, then necessarily we have $\dot{z}(t)=\xi(t)$ for almost all $t\in[a,b]$. The convergence structure imposed on this space of functions is the usual weak topology which we define as follows.

Definition 8.7. A sequence $\{z_k(\cdot)\}_{k=1}^{+\infty}$ in $AC([a,b];\mathbb{R}^m)$ converges weakly to a function $y(\cdot)\in AC([a,b];\mathbb{R}^m)$ if there exists a sequence $\{t_k\}_{k=1}^{+\infty}\subset[a,b]$ that converges to $\hat{t}\in[a,b]$ such that $z_k(t_k)\to y(\hat{t})$ as $k\to\infty$ and for every bounded measurable function $\psi(\cdot):[a,b]\to\mathbb{R}^m$ one has (in which $\langle\cdot,\cdot\rangle$ denotes the usual scalar product in $\mathbb{R}^m$)

$$\lim_{k\to\infty}\int_a^b \langle\psi(s),\dot{z}_k(s)\rangle\,,\mathrm{d}s = \int_a^b \langle\psi(s),\dot{y}(s)\rangle\,\mathrm{d}s.$$

We make the following observations concerning the above definition. First, since we are only interested in absolutely continuous functions satisfying the fixed endpoint conditions (8.3), for any sequence of interest for us here we can take $t_k = a$ so that the first condition in the above definition is automatically satisfied. Secondly, the convergence property of the sequence of derivatives is referred to as weak convergence in $L^1([a,b];\mathbb{R}^n)$ (the space of Lebesgue integrable functions) of the derivatives. As a consequence of these two observations, we need to consider the weak compactness of a set of integrable functions. To this end we have the following well known theorem.

Theorem 8.1. *Let $\{h(\cdot) : [a,b] \to \mathbb{R}^m\}$ be a family of Lebesgue integrable functions. The following two statements are equivalent.*

1. *The family $\{h(\cdot)\}$ is sequentially weakly relatively compact in $L^1([a,b];\mathbb{R}^m)$.*
2. *There is a constant $M \in \mathbb{R}$ and a function $\Phi(\cdot) : [0,\infty) \to \mathbb{R}$ such that*

$$\lim_{\zeta \to \infty} \frac{\Phi(\zeta)}{\zeta} = +\infty \quad \text{and} \quad \int_a^b \Phi(h(s))\,\mathrm{d}s \leq M$$

for all $h(\cdot)$ in the family.

Remark 8.6. The above theorem is a consequence of many people and we refer the interested reader to the discussion in Cesari [2, Sect. 10.3]. The weak relative compactness of the family $\{h(\cdot)\}$ means that there exists a subsequence $\{h_k(\cdot)\}_{k=1}^{+\infty}$ which converges weakly in $L^1([a,b];\mathbb{R}^m)$ to some integrable function.

In light of the above theorem, we impose the following growth condition ϕ on the integrands $L_j(\cdot,\cdot,\cdot)$.

Growth condition ϕ. For each $j = 1,2,\ldots,N$ there exists a function $\Phi_j(\cdot)$: $[0,\infty) \to \mathbb{R}$ that is bounded below satisfying $\lim_{\zeta\to\infty} \Phi_j(\zeta)/\zeta = +\infty$ such that $L_j(t,\mathbf{z},p_j) \geq \Phi_j(|p_j|)$ for almost every $t \in [a,b]$ and all $(\mathbf{z},p_j) \in \times\mathbb{R}^{\mathbf{n}} \times \mathbb{R}^{n_j}$.

Remark 8.7. Observe that when the growth condition ϕ holds, the fact that each $\Phi_j(\cdot)$ is bounded below ensures the existence of the constants A_j^* required in Lemma 8.1.

Theorem 8.2. *Assume that the integrands $L_j(\cdot,\cdot,\cdot)$ satisfy the growth condition ϕ and that there exists constants B_j^*, $j = 1,2,\ldots,N$, such that one has $I_j(\mathbf{x}(\cdot)) \leq B_j^*$ for all feasible trajectories of the original constrained game (8.1), (8.2). Further assume that for each λ sufficiently large there exists a Nash equilibrium $\mathbf{x}_\lambda^*(\cdot)$ for the penalized unconstrained variational game (8.6) and that there exists an admissible trajectory $\mathbf{y}_\lambda(\cdot)$ such that the trajectories $[\mathbf{x}_\lambda^{*j}, y_{\lambda,j}](\cdot)$ are feasible trajectories for the constrained variational game (8.1), (8.2). Then there exists a sequence $\lambda_k \to +\infty$ as $k \to \infty$ and a function $\mathbf{x}^*(\cdot) \in AC([0,1];\mathbb{R}^{\mathbf{n}})$ satisfying the endpoint conditions (8.3) and such that $\{\mathbf{x}_{\lambda_k}^*(\cdot)\}_{k=1}^{+\infty}$ converges weakly in $AC([a,b];\mathbb{R}^{\mathbf{n}})$ to $\mathbf{x}^*(\cdot)$ as $k \to \infty$.*

Proof. As a consequence of the growth condition ϕ and the other assumed conditions, for each $j = 1, 2, \ldots, N$ and all λ sufficiently large we have

$$\begin{aligned}\int_a^b \Phi_j(|\dot{x}^*_{\lambda,j}(t)|)\,\mathrm{d}t &\le \int_a^b L_j(t, \mathbf{x}^*(t), \dot{x}^*_{\lambda,j}(t))\,\mathrm{d}t \\ &\le \int_a^b L_j(t, \mathbf{x}^*(t), \dot{x}^*_{\lambda,j}(t)) + \lambda g(t, \mathbf{x}^*(t), \dot{x}^*_{\lambda,j}(t))\,\mathrm{d}t \\ &\le \int_a^b L_j(t, [\mathbf{x}^{*j}_{\lambda}, y_{\lambda,j}](t), \dot{y}_{\lambda,j}(t))\,\mathrm{d}t \\ &\le B^*_j.\end{aligned}$$

This implies, in view of Theorem 8.1, that the set of functions $\{\dot{x}^*_{\lambda,j}(\cdot)\}_{\lambda>\lambda_0}$ are relatively weakly sequentially compact in $L^1([a,b];\mathbb{R}^{n_j})$ for each $j = 1, 2, \ldots, N$ and morover since $x^*_{\lambda,j}(a) = x_{aj}$ for each $\lambda > \lambda_0$ and each j, it follows that $\{x^*_{\lambda,j}(\cdot)\}_{\lambda>\lambda_0}$ is relatively sequentially weakly compact. This means, by a diagonalization argument, that we can find a sequence of $\{\lambda_k\}_{k=1}^{+\infty}$ and a function $\mathbf{x}^*(\cdot) : [a,b] \to \mathbb{R}^{\mathbf{n}}$ such that $\lambda_k \to +\infty$ as $k \to +\infty$ and the corresponding sequence $\{\mathbf{x}^*_{\lambda_k}(\cdot)\}_{k=1}^{+\infty}$ converges weakly to $\mathbf{x}^*(\cdot)$ in $AC([a,b];\mathbb{R}^{\mathbf{n}})$. □

Remark 8.8. We note that when the $L_j(\cdot,\cdot,\cdot)$ satisfy the growth condition ϕ and the assumptions of Lemma 8.1 hold we can conclude that, when the Nash equilibria $\mathbf{x}^*_{\lambda}(\cdot)$ exist for all $\lambda > 0$ sufficiently large, there exists a sequence $\{\lambda_k\}_{k=1}^{+\infty}$, with $\lambda_k \to +\infty$ as $k \to +\infty$, and an absolutely continuous trajectory $\mathbf{x}^*(\cdot)$ such the $\mathbf{x}^*_{\lambda_k}(\cdot) \to \mathbf{x}^*(\cdot)$ weakly in $AC([a,b];\mathbb{R}^{\mathbf{n}})$ and that

$$\lim_{k\to\infty} \int_a^b g_j(t, \mathbf{x}^*_{\lambda_k}(t), \dot{x}^*_{\lambda_k,j}(t))\,\mathrm{d}t = 0, \quad j = 1, 2, \ldots, N.$$

It remains to investigate whether the limit function $\mathbf{x}^*(\cdot)$ is a Nash equilibrium for the constrained variational game. To see this we need to investigate the lower semicontinuity properties of the integral functionals $I_j(\cdot)$ and $I_{\lambda,j}(\cdot)$. To do this we view these functionals as being defined on the product spaces of functions $\mathcal{S}_j \doteq C([a,b];\mathbb{R}^{\mathbf{n}}) \times L([a,b];\mathbb{R}^{n_j})$. For the topology on $\mathcal{S}_j$ we use the topology of pointwise convergence in $C([a,b];\mathbb{R}^{\mathbf{n}})$ and the weak topology of $L([a,b];\mathbb{R}^{n_j})$. To be precise we have the following definition.

Definition 8.8. We say a sequence of functions $\{(\mathbf{z}_k(\cdot), p_k(\cdot))\}_{k=1}^{+\infty} \subset \mathcal{S}$, where $\mathcal{S} \doteq C([a,b];\mathbb{R}^{\mathbf{n}}) \times L([a,b];\mathbb{R}^m)$ converges to $(\hat{\mathbf{z}}(\cdot), \hat{p}(\cdot)) \in \mathcal{S}$ if $\mathbf{z}_k(t) \to \hat{\mathbf{z}}(t)$ pointwise almost everywhere as $k \to +\infty$ and $p_k(\cdot) \to \hat{p}(\cdot)$ weakly in $L([a,b];\mathbb{R}^m)$ as $k \to +\infty$.

Remark 8.9. We notice that if we have a sequence of admissible trajectories for the unconstrained variational game, say $\{\mathbf{x}_k(\cdot)\}_{k=1}^{+\infty}$ which converges weakly in $AC([a,b];\mathbb{R}^{\mathbf{n}})$ to an admissible trajectory $\hat{\mathbf{x}}(\cdot) \in AC([a,b];\mathbb{R}^{\mathbf{n}})$, then the sequence $\{(\mathbf{x}_k(\cdot), \dot{x}_{k,j}(\cdot))\}_{k=1}^{+\infty}$ automatically converges to $(\hat{\mathbf{x}}(\cdot), \dot{\hat{x}}_j(\cdot))$ in $\mathcal{S}_j$ for each $j = 1, 2, \ldots, N$.

The lower semicontinuity property for functionals defined on $\mathcal{S}$ is now given in the next definition.

Definition 8.9. A functional $\mathcal{K}(\cdot) : \mathcal{S} \to \mathbb{R}$ is said to be lower semicontinuous on $\mathcal{S}$ if $\liminf_{k\to+\infty} \mathcal{K}((\mathbf{z}_k(\cdot), p_k(\cdot))) \geq \mathcal{K}((\hat{\mathbf{z}}(\cdot), \hat{p}(\cdot)))$ for every $(\hat{\mathbf{z}}(\cdot), \hat{p}(\cdot)) \in \mathcal{S}$ and every sequence $\{(\mathbf{z}_k(\cdot), p_k(\cdot))\}_{k=1}^{+\infty} \subset \mathcal{S}$ that converges weakly to $(\hat{\mathbf{z}}(\cdot), \hat{p}(\cdot))$.

With this definition we can state the following theorem.

Theorem 8.3. *Let $f(\cdot,\cdot,\cdot) : [a,b] \times \mathbb{R}^{\mathbf{n}} \times \mathbb{R}^m \to \mathbb{R}$ be a continuous function such that $p \mapsto f(t,x,p)$ is convex for all $(t,x) \in [a,b] \times \mathbb{R}^{\mathbf{n}}$ and such that there exists an integrable function $\eta(\cdot) : [a,b] \to [0,\infty)$ such that $f(t,x,p) \geq -\eta(t)$ for all $(t,x,p) \in [a,b] \times \mathbb{R}^n \times \mathbb{R}^n$. Then the functional $\mathcal{J}(\cdot) : \mathcal{S} \to \mathbb{R}$ defined by*

$$\mathcal{J}(\mathbf{z}(\cdot), p(\cdot)) = \int_a^b f(t, \mathbf{x}(t), p(t))\,\mathrm{d}t$$

is lower semicontinuous with respect to the topology of pointwise almost everywhere convergence in $\mathbf{x}(\cdot)$ and weak L^1-convergence in $p(\cdot)$ (as defined in Definition 8.8).

Proof. See Cesari [2, Theorem 10.8i]. □

Remark 8.10. This theorem has a long history dating to the beginning of the twentieth century with the work of L. Tonelli. For a discussion of these matters we refer the reader to the comprehensive monograph of Cesari [2, Chap. 10].

Our next result provides gives us what we desire, namely that if there exists a family of Nash equilibria $\{\mathbf{x}_\lambda^*(\cdot)\}_{\lambda>0}$ for the unconstrained variational games (8.6), then there exists a subsequence which converges to a Nash equilibrium of the constrained variational game. Unfortunately, we have to restrict our model to the case when the equality constraints are uncoupled. That is we now suppose that $g_j(\cdot,\cdot,\cdot) : [a,b] \times \mathbb{R}^{n_j} \times \mathbb{R}^{n_j} \to \mathbb{R}$.

Theorem 8.4. *Assume that for each $j = 1,2,\ldots,N$ the integrands $L_j(\cdot,\cdot,\cdot)$ satisfy the growth condition ϕ and that $p \mapsto L_j(t,\mathbf{x},p)$ is convex for each $(t,\mathbf{x}) \in [a,b] \times \mathbb{R}^{\mathbf{n}}$. Further suppose that each $g_j(\cdot,\cdot,\cdot) : [a,b] \times \mathbb{R}^{n_j} \times \mathbb{R}^{n_j} \to [0,+\infty)$ is continuous and that the maps $p \mapsto g_j(t,x,p)$ are convex on $\mathbb{R}^{n_j}$ for each $(t,x) \in [a,b] \times \mathbb{R}^{n_j}$. Additionally, assume that there exists constants B_j^* such that $I_j(\mathbf{y}(\cdot)) \leq B_j^*$ for each trajectory $\mathbf{y}(\cdot)$ which is admissible for the constrained game (8.1), (8.2). Then, if for each sufficiently large $\lambda > 0$ there exists a Nash equilibrium, $\mathbf{x}_\lambda^*(\cdot)$, of the unconstrained penalized game (8.6) and if there exists at least one admissible trajectory $\mathbf{y}(\cdot)$ for the constrained game, then there exists a trajectory $\mathbf{x}^*(\cdot)$ which is admissible for the constrained variational game such that $\mathbf{x}_{\lambda_k}^*(\cdot) \to \mathbf{x}^*(\cdot)$ weakly in $AC([a,b];\mathbb{R}^{\mathbf{n}})$ as $k \to \infty$ and moreover is a Nash equilibrium for the constrained variational game.*

Proof. We first notice that the existence of the sequence $\lambda_k \to +\infty$ as $k \to +\infty$ and the trajectory $\mathbf{x}^*(\cdot)$ such that $\mathbf{x}_{\lambda_k}^*(\cdot) \to \mathbf{x}^*(\cdot)$ weakly in $AC([a,b];\mathbb{R}^{\mathbf{n}})$ as $k \to +\infty$

has been established in Theorem 2. Moreover, as a result of Lemma 1 (see also the remarks following its proof) we also have that

$$\lim_{k\to\infty}\int_a^b g_j(t,x^*_{\lambda_k,j}(t),\dot{x}^*_{\lambda_k,j}(t))\,\mathrm{d}t = 0, \quad j=1,2,\ldots N.$$

Now, since the functions $g_j(\cdot,\cdot,\cdot)$ are nonnegative and convex in their last n_j arguments we have that the integral functionals

$$\mathcal{G}_j((z(\cdot),p(\cdot))) = \int_a^b g_j(t,z(t),p(\cdot))\,\mathrm{d}t, \quad j=1,2,\ldots N,$$

are lower semicontinuous on $C([a,b];\mathbb{R}^{n_j})\times L([a,b];\mathbb{R}^{n_j})$. In particular this means that

$$0 = \lim_{k\to+\infty}\int_a^b g_j(t,x^*_{\lambda_k,j}(t),\dot{x}^*_{\lambda_k,j}(t))\,\mathrm{d}t \geq \int_a^b g_j(t,x^*_j(t),\dot{x}^*_j(t))\,\mathrm{d}t \geq 0,$$

which implies that $g_j(t,x^*_j(t),x^*_j(t)) = 0$ for almost all $t\in[a,b]$. This of course says that $\mathbf{x}^*(\cdot)$ is an admissible trajectory for the constrained variational game. It remains to show that it is a Nash equilibrium. To see this fix $j=1,2,\ldots N$ and let $y_j(\cdot):[a,b]\to\mathbb{R}^{n_j}$ be an admissible trajectory for player j relative to $\mathbf{x}^*(\cdot)$ and consider the following inequalities for λ_k

$$\begin{aligned}
I_j(\mathbf{x}^*_{\lambda_k}(\cdot)) &= \int_a^b L_j(t,\mathbf{x}^*_{\lambda}(t),\dot{x}^*_{\lambda_k,j}(t))\,\mathrm{d}t\\
&\leq I_{\lambda_k,j}(\mathbf{x}^*_{\lambda_k}(\cdot))\\
&= \int_a^b L_j(t,\mathbf{x}^*_{\lambda_k}(t),\dot{x}^*\lambda_k,j(t)) + \lambda_k g_j(t,x^*_{\lambda_k,j}(t),\dot{x}^*_{\lambda_k,j}(t))\,\mathrm{d}t\\
&\leq I_{\lambda_k,j}([\mathbf{x}^{*j}_{\lambda_k},y_j(t)](\cdot))\\
&= \int_a^b L_j(t,[\mathbf{x}^{*j}_{\lambda_k},y_j](t),\dot{y}_j(t)) + \lambda_k g_j(t,y_j(t),\dot{y}_j(t))\,\mathrm{d}t,\\
&= \int_a^b L_j(t,[\mathbf{x}^{*j}_{\lambda_k},y_j](t),\dot{y}_j(t))\,\mathrm{d}t
\end{aligned}$$

where the first inequality is a result of the nonnegativeness of $g_j(\cdot,\cdot,\cdot)$, the second inequality is a consequence of the fact that $\mathbf{x}^*_{\lambda_k}(\cdot)$ is a Nash equilibrium for the unconstrained variational game with $\lambda=\lambda_k$ and the last equality follows because $g_j(t,y_j(t),\dot{y}_j(t))=0$ for almost all $t\in[a,b]$ by the definition of $y_j(\cdot)$. Letting $k\to\infty$ in the above gives

$$\begin{aligned} I_j(\mathbf{x}^*(\cdot)) &\leq \liminf_{k\to\infty} I_j(\mathbf{x}^*_{\lambda_k}(\cdot)) \\ &= \liminf_{k\to\infty} I_{\lambda_k,j}(\mathbf{x}^*_{\lambda_k}(\cdot)) \\ &\leq \liminf_{k\to\infty} I_{\lambda_k,j}([\mathbf{x}^{*j}_{\lambda_k}, y_j](\cdot)) \\ &= I_j([\mathbf{x}^{*j}, y_j](t)). \end{aligned}$$

In the above, the first inequality is a consequence of the lower semicontinuity properties of the functionals $I_j(\cdot)$, the second equality follows from properties of limit inferiors and the nonnegativity of $g_j(\cdot,\cdot,\cdot)$, the next inequality follows from the Nash equilibria properties of the $\mathbf{x}^*_{\lambda_k}(\cdot)$'s and the last equality follows from the fact that $g_j(t, y_j(t), \dot{y}_j(t)) = 0$ a.e. on $[a,b]$ and Lebesgue's dominated convergence theorem. Since the trajectory $y_j(\cdot)$ is any trajectory for player j which is admissible relative to $\mathbf{x}^*(\cdot)$ the desired result is proved. □

Remark 8.11. The above theorem provides us with what we desire. Unfortunately, we must restrict ourselves to equality constraints that are uncoupled with respect to each of the players. The above proof breaks down in the general case at the very end since we cannot ensure the equality $g_j(t, [\mathbf{x}^{*j}_{\lambda_k}, y_j](t), \dot{y}_j(t)) = 0$ for almost all $t \in [a,b]$.

8.4 Example: The Single Player Case

A firm produces a good k whose quantity at time $t \in [a,b]$ is given by a production process

$$\dot{k}(t) = f(k(t)) + c(t), \quad t \in (0,b),$$

in which $f(k(t))$ is a production rate and $c(t)$ denotes a rate of external investments required for the production (i.e., amount of raw materials). The goal of the firm is to maximize its profit. The price per unit of each unit is given by a demand $p = p(k(t))$, a function that depends on the available inventory of the firm, and the cost of production $C(c(t))$ depends on the external investment rate at time t. Thus the objective of the firm is to maximize a functional of the form

$$\begin{aligned} I(k(\cdot)) &= \int_0^b p(k(t))k(t) - C(c(t))\,\mathrm{d}t \\ &= \int_a^b p(k(t))k(t) - C(\dot{k}(t) - f(k(t)))\,\mathrm{d}t. \end{aligned} \tag{8.8}$$

over all inventory streams $t \mapsto k(t)$ satisfying a given initial level $k(0) = k_0$. This gives a simple unconstrained variational game. The production processes of the firm generates a pollutant $s(t)$ at each time t which the government has mandated must be reduced to a fraction of its initial level (i.e. to $\alpha s(0)$, $\alpha \in (0,1)$) over the time interval $[0,b]$. That is, $s(b) = \alpha s(0)$. Each firm generates pollution according to the following process:

$$\begin{aligned} \dot{s}(t) &= g(k(t)) - \mu s(t), \quad t \in (0,b), \\ s(0) &= s_0, \quad s(b) = \alpha s_0, \end{aligned} \tag{8.9}$$

in which $g(k(t))$ denotes the rate of production of the pollutant by the firm and $\mu > 0$ is a constant representing the "natural" abatement of the pollutant. This gives our differential constraint. Thus the problem for the firm is to maximize its profit given by (8.8) while satisfying the pollution constraint (8.9) and the end conditions $(k(0), s(0)) = (k_0, s_0)$ and $s(b) = s_b$ (here of course we interpret $s_b = \alpha s_0$ but this is not necessary for the formulation of the problem).

If we choose for specificity $p(k) = \pi k$, $f(k) = \alpha k + \beta$, $g(k) = \gamma k$ and $C(c) = \frac{1}{2}c^2$ with all of the coefficients positive constants, the above calculus of variations problem becomes one in which the objective functional is quadratic and the differential side constraint becomes linear.

To apply our theory we consider the family of unconstrained variational problems (P_λ) of minimizing

$$\int_0^b \frac{1}{2}[\dot{k}(t) - \alpha k(t)) - \beta]^2 - \pi k(t)^2 + \lambda[\dot{s}(t) + \mu s(t) - \gamma k(t)]^2 \, \mathrm{d}t, \tag{8.10}$$

over all piecewise continuous $(k(\cdot), s(\cdot)) : [0,b] \to \mathbb{R}^2$ satisfying the end conditions

$$(k(0), s(0)) = (k_0, s_0), \quad s(b) = s_b. \tag{8.11}$$

We define $L_\lambda(\cdot,\cdot)$ by the integrand of the objective. That is,

$$L_\lambda((k,s),(p,q)) = \frac{1}{2}[p - \alpha k - \beta]^2 - \pi k^2 + \lambda[q + \mu s - \gamma k]^2.$$

Further we note that since $k(b)$ is unspecified the solution $(k^\lambda(\cdot), s^\lambda(\cdot))$ must satisfy the transversality condition

$$\left.\frac{\partial L_\lambda}{\partial p}\right|_{((k^\lambda(b), s^\lambda(b)), (\dot{k}^\lambda(b), \dot{s}^\lambda(b)))} = \dot{k}^\lambda(b) - \alpha k^\lambda(b) - \beta = 0.$$

This supplies us with the terminal condition for the state $k^\lambda(t)$ at $t = b$.

The Euler–Lagrange equations for (P_λ) is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}[\dot{k}(t)-\alpha k(t)-\beta] &= -\alpha(\dot{k}(t)-\alpha k(t)-\beta)-2\pi k(t)\\ &\quad -2\lambda\gamma(\dot{s}(t)+\mu s(t)-\gamma k(t))\\ 2\lambda\frac{\mathrm{d}}{\mathrm{d}t}[\dot{s}(t)+\mu s(t)-\gamma k(t)] &= 2\mu\lambda(\dot{s}(t)+\mu s(t)-\gamma k(t))).\end{aligned}$$

For a solution $(k^\lambda(\cdot),s^\lambda(\cdot))$ of the above system define $\Lambda(\cdot)=\dot{s}(\cdot)+\mu s(\cdot)-\gamma k(\cdot)$ and observe that the second equation becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(t)=\mu\Lambda(t),\quad t\in(0,b),$$

which has the general solution $\Lambda(t)=\Lambda_{0\lambda}\,\mathrm{e}^{\mu t}$, where $\Lambda_{0\lambda}$ is a constant to be determined. Observe that the constant $\Lambda_{0\lambda}$ does depend on λ since in general $k^\lambda(\cdot)$ will. Substituting $\Lambda(\cdot)$ for $(\dot{s}^\lambda(t)+\mu s^\lambda(t)-\gamma k^\lambda(t))$ into the first equation gives us the uncoupled equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}[\dot{k}^\lambda(t)-\alpha k^\lambda(t)-\beta]=-\alpha(\dot{k}^\lambda(t)-\alpha k^\lambda(t)-\beta)-2\pi k^\lambda(t)-2\lambda\gamma\Lambda_{0\lambda}\,\mathrm{e}^{\mu t},$$

or after simplifying becomes

$$\ddot{k}^\lambda(t)+(2\pi-\alpha^2)k^\lambda(t)=\alpha\beta-2\lambda\gamma\Lambda_{0\lambda}\,\mathrm{e}^{\mu t}.$$

The general solution of this equation has the form $k^\lambda(t)=k_c(t)+k_1(t)+k_2(t)$ where $k_c(\cdot)$ is the general solution of the homogeneous equation $\ddot{k}(t)+(2\pi-\alpha^2)k(t)=0$, $k_1(\cdot)$ solves the nonhomogeneous equation $\ddot{k}(t)+(2\pi-\alpha^2)k(t)=\alpha\beta$ and $k_2(\cdot)$ solves the nonhomogeneous equation $\ddot{k}(t)+(2\pi-\alpha^2)k(t)=-2\lambda\gamma\Lambda_{0\lambda}\,\mathrm{e}^{\mu t}$. Each of these equations are easy to solve. For simplicity we assume $\alpha^2-2\pi>0$ (to insure real roots of the characteristic equation). Using elementary techniques we have

$$k_c(t)=A^*\mathrm{e}^{rt}+B^*\mathrm{e}^{-rt},\quad k_1(t)=\frac{\alpha\beta}{\alpha^2-2\pi},\quad \text{and}\quad k_2(t)=\frac{2\lambda\gamma\Lambda_{0\lambda}}{\alpha^2-2\pi-\mu^2}\,\mathrm{e}^{\mu t}.$$

giving us

$$k^\lambda(t)=-\frac{\alpha\beta}{\alpha^2-2\pi}+\frac{2\lambda\gamma\Lambda_{0\lambda}}{\alpha^2-2\pi-\mu^2}\,\mathrm{e}^{\mu t}+A^*\,\mathrm{e}^{rt}+B^*\,\mathrm{e}^{-rt},$$

in which $r=\sqrt{\alpha^2-2\pi}$. Using the initial value for $k(0)=k_0$ and the transversality condition we obtain the following two equations for A^* and B^*.

$$A^* + B^* = k_0 + \frac{\alpha\beta}{\alpha^2 - 2\pi} - \frac{2\lambda\gamma\Lambda_{0\lambda}}{\alpha^2 - 2\pi - \mu^2}$$

$$(r-\alpha)\,\mathrm{e}^{rb}A^* + (-r-\alpha)\,\mathrm{e}^{-rb}B^* = \beta - \frac{\alpha^2\beta}{\alpha^2 - 2\pi} - \frac{2\lambda\gamma\Lambda_{0\lambda}(\mu-\alpha)}{\alpha^2 - 2\pi - \mu^2}\,\mathrm{e}^{\mu b}.$$

Using Cramer's rule we get the following expressions for A^* and B^*:

$$\begin{aligned} A^* &= \frac{1}{\Delta}\Bigg[k_0(-r-\alpha)\,\mathrm{e}^{-rb} + \frac{\alpha\beta}{\alpha^2-2\pi}\left[(-r-\alpha)\,\mathrm{e}^{-rb} + \alpha\right] \\ &\qquad + \frac{2\lambda\gamma\Lambda_{0\lambda}}{\alpha^2-2\pi-\mu^2}\left[(\mu-\alpha)\,\mathrm{e}^{\mu b} - (-r-\alpha)\,\mathrm{e}^{-rb}\right]\Bigg] \\ B^* &= \frac{1}{\Delta}\Bigg[\beta - k_0(r-\alpha)\,\mathrm{e}^{rb} - \frac{\alpha\beta}{\alpha^2-2\pi}\left[\alpha + (r-\alpha)\,\mathrm{e}^{rb}\right] \\ &\qquad \frac{2\lambda\gamma\Lambda_{0\lambda}}{\alpha^2-2\pi-\mu^2}\left[(r-\alpha)\,\mathrm{e}^{rb} - (\mu-\alpha)\,\mathrm{e}^{\mu b}\right]\Bigg], \end{aligned}$$

in which $\Delta = 1/[(-r-\alpha)\,\mathrm{e}^{-rb} - (r-\alpha)\,\mathrm{e}^{rb}]$ (i.e., the reciprocal of the determinant of the coefficient matrix of the 2×2 linear system for A^* and B^*). To proceed further, and at the same time trying to keep things managable, observe that we can write $A^* = \mathcal{A} + \mathcal{B}\Lambda_{0\lambda}\lambda$ and $B^* = \mathcal{C} + \mathcal{D}\Lambda_{0\lambda}\lambda$ in which $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ are constants independent of $\Lambda_{0\lambda}$ and λ. We can continue this observation to obtain that

$$k^\lambda(t) = -\frac{\alpha\beta}{\alpha^2-2\pi} + \mathcal{E}\Lambda_{0\lambda}\lambda\,\mathrm{e}^{\mu t} + (\mathcal{A} + \mathcal{B}\Lambda_{0\lambda}\lambda)\,\mathrm{e}^{rt} + (\mathcal{C} + \mathcal{D}\Lambda_{0\lambda}\lambda)\,\mathrm{e}^{-rt},$$

in which $\mathcal{E}$ is also a constant that is independent of $\Lambda_{0\lambda}$ and λ.

We now determine $s^\lambda(\cdot)$. To this end we use the definition of $\Lambda(t) = \Lambda_{0\lambda}\,\mathrm{e}^{\mu t}$ to obtain the differential equation

$$\begin{aligned} \dot{s}^\lambda(t) + \mu s^\lambda(t) &= \gamma k^\lambda(t) + \Lambda_{0\lambda}\,\mathrm{e}^{\mu t} \\ &= -\frac{\alpha\beta}{\alpha^2-2\pi} + (\mathcal{E}\Lambda_{0\lambda}\lambda + \Lambda_{0\lambda})\,\mathrm{e}^{\mu t} \\ &\qquad + (\mathcal{A} + \mathcal{B}\Lambda_{0\lambda}\lambda)\,\mathrm{e}^{rt} + (\mathcal{C} + \mathcal{D}\Lambda_{0\lambda}\lambda)\,\mathrm{e}^{-rt}. \end{aligned}$$

This is a first-order linear differential equation which is easily solved by multiplying both sides by $\mathrm{e}^{\mu t}$. The unique solution satisfying the fixed initial condition $s^\lambda(0) = s_0$ is given by

$$\begin{aligned} s^\lambda(t) &= s_0\,\mathrm{e}^{-\mu t} - \frac{\alpha\beta}{\mu(\alpha^2-2\pi)}(1-\mathrm{e}^{-\mu t}) + \frac{1}{2\mu}(\mathcal{E}\Lambda_{0\lambda}\lambda + \Lambda_{0\lambda})(\mathrm{e}^{\mu t} - \mathrm{e}^{-\mu t}) \\ &\quad \frac{1}{r+\mu}(\mathcal{A} + \mathcal{B}\Lambda_{0\lambda}\lambda)(\mathrm{e}^{rt} - \mathrm{e}^{-\mu t}) + \frac{1}{-r+\mu}(\mathcal{C} + \mathcal{D}\Lambda_{0\lambda}\lambda)(\mathrm{e}^{-rt} - \mathrm{e}^{-\mu t}). \end{aligned}$$

Using the end condition $s^\lambda(b) = s_b$ we find $\Lambda_{0\lambda}$ to be

$$\Lambda_{0\lambda} = \left[\frac{e^{\mu b} - e^{-\mu b}}{2\mu} + \left[\frac{e^{\mu b} - e^{-\mu b}}{2\mu}\mathcal{E} + \frac{e^{rb} - e^{-\mu b}}{r+\mu}\mathcal{B} + \frac{e^{-rb} - e^{-\mu b}}{-r+\mu}\mathcal{D}\right]\lambda\right]^{-1}$$
$$\times \left[s_b - s_0\, e^{-\mu b} - \frac{1 - e^{-\mu b}}{\mu(\alpha^2 - 2\pi)}\alpha\beta + \frac{e^{rb} - e^{-\mu b}}{r+\mu}\mathcal{A} + \frac{e^{-rb} - e^{-\mu b}}{-r+\mu}\mathcal{C}\right],$$

which we observe has the form

$$\Lambda_{0\lambda} = \frac{\mathcal{F}}{\mathcal{G} + \mathcal{H}\lambda},$$

where $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{H}$ are constants depending only on the data.

In the above we have successfully solved the first-order necessary conditions (i.e., the Euler–Lagrange equations and the relevant transversality condition) for each positive λ. Further, it is clear that we have obtained the unique solution in this case. Moreover, since the objective functional is quadratic in the derivative arguments it is clear that there exists an optimal solution for each λ which must be $(k^\lambda(\cdot), s^\lambda(\cdot))$.

We now explore what happens when $\lambda \to +\infty$. First we observe that $\Lambda_{0\lambda} \to 0$ as $\lambda \to +\infty$. This means that we have

$$\lim_{\lambda\to+\infty} \dot{s}^\lambda(t) + \mu s^\lambda(t) - \gamma k^\lambda(t) = \lim_{\lambda\to+\infty} \Lambda_{0\lambda}\, e^{\mu t} = 0, \quad \text{for all} \quad t \in [a,b],$$

which implies that in the limit the constraint is satisfied. Further, we notice that $\Lambda_{0\lambda}\lambda \to \mathcal{F}/\mathcal{H}$ as $\lambda \to +\infty$ from which one easily sees that $(k^\lambda(\cdot), s^\lambda(\cdot)) \to (k^*(\cdot), s^*(\cdot))$ as $\lambda \to +\infty$ where

$$k^*(t) = -\frac{\alpha\beta}{\alpha^2 - 2\pi} + \frac{\mathcal{E}\mathcal{F}}{\mathcal{H}}\, e^{\mu t} + \left(\mathcal{A} + \frac{\mathcal{B}\mathcal{F}}{\mathcal{H}}\right) e^{rt} + \left(\mathcal{C} + \frac{\mathcal{D}\mathcal{F}}{\mathcal{H}}\right) e^{-rt}$$

$$s^*(t) = s_0\, e^{-\mu t} - \frac{\alpha\beta}{\mu(\alpha^2 - 2\pi)}(1 - e^{-\mu t}) + \frac{\mathcal{F}\mathcal{E}}{2\mu\mathcal{H}}(e^{\mu t} - e^{-\mu t})$$
$$\frac{1}{r+\mu}\left(\mathcal{A} + \frac{\mathcal{B}\mathcal{F}}{\mathcal{H}}\right)(e^{rt} - e^{-\mu t}) + \frac{1}{-r+\mu}\left(\mathcal{C} + \frac{\mathcal{D}\mathcal{F}}{\mathcal{H}}\right)(e^{-rt} - e^{-\mu t}).$$

Therefore, as a consequence of our theoretical results we now know that $t \mapsto (k^*(t), s^*(t))$ is a solution of our original problem.

8.5 Example: The Two-Player Case

If we extend the above example to two players each firm produces an equivalent good whose quantity at time $t \in [a,b]$ is given by a production process

$$\dot{k}_j(t) = f_j(k_j(t)) + c_j(t), \quad t \in (a,b), \quad j = 1,2,$$

in which $f_j(k_j(t))$ is a production rate and $c_j(t)$ denotes a rate of external investments required for the production (i.e., amt of raw materials). The goal of each firm is to maximize its profit. The price per unit of each unit is given by a demand function $p = p(k_1(t), k_2(t))$, a function that depends on the total inventory of each firm, and the cost of production $C_j(c_j(t))$ depends on the external investment rate at time t. Thus the objective of each firm is to maximize a functional of the form

$$\begin{aligned} I_j(k_1(\cdot), k_2(\cdot)) &= \int_a^b p(k_1(t), k_2(t))k_j(t) - C_j(c_j(t))\,\mathrm{d}t \\ &= \int_a^b p(k_1(t), k_2(t))k_j(t) - C_j(\dot{k}_j(t) - f_j(k_j(t)))\,\mathrm{d}t. \quad (8.12) \end{aligned}$$

over all inventory streams $t \mapsto k_j(t)$ satisfying a given initial level $k_j(a) = k_{aj}$. This gives a simple unconstrained variational game. The production processes of each firm generate the same pollutant $s(t) = s_1(t) + s_2(t)$ at each time t (here $s_i(t)$ denotes the pollutant level due to the ith player) which the government has mandated must be reduced to a fraction of its initial level (i.e. to $\alpha s(a)$, $\alpha \in (0,1)$) over the time interval $[a,b]$. That is, $s(b) = \alpha s(a)$. Each firm generates pollution according to the following process:

$$\begin{aligned} \dot{s}_j(t) &= g_j(k_j(t)) - \mu s_j(t), \quad t \in (a,b), \quad (8.13) \\ s_j(a) &= s_{aj}, \end{aligned}$$

in which $g_j(k_j(t))$ denotes the rate of production of the pollutant by firm $j = 1,2$ and $\mu > 0$ is a constant representing the “natural” abatement of the pollutant. This gives our differential constraint. Thus the problem for each firm is to maximize its profit given by (8.12) while satisfying the pollution constraint (8.13) and the end conditions $(k_j(a), s_j(a)) = (k_{aj}, s_{aj})$ and $s_j(b)) = s_{bj}$ (here of course we interpret $s_{bj} = \alpha s_{aj}$.

Making the same simplifying assumptions on the form of the relevant functions as in the single player case we obtain a linear-quadratic game. In this case the technique used above can be modified to solve the Euler–Lagrange equation. Unfortunately, in this case it is well known the necessary conditions are not sufficient and so one can not guarantee that the solutions obtained for each λ are Nash equilibria. One technique which might be applicable would be a variation of Leitmann’s direct method which provides sufficient conditions for open-loop Nash equilibria. To do this requires the investigation of variational games with unspecified right endpoint conditions. In the single player case such a modification has been investigated briefly by Wagener [11].

8.6 Conclusions

In this paper we explored the use of a penalty method to find open-loop Nash equilibria for a class of variational games. We showed that using classical assumptions, with roots in the calculus of variations, it was possible to establish our results. We presented an example of a single-player game in detail and gave some indication of the difficulties encountered when treating the multi-player case. Our analysis suggests that a new extension of Leitmann's direct method to problems with unspecified right endpoint conditions could prove useful in using this penalty method to determine Nash equilibria.

Acknowledgements The authors would like to dedicate this paper to the memory of our friend and colleague Thomas L. Vincent. Additionally, the first author would like to offer his best wishes to his friend and co-author George Leitmann on the occasion of his eighty-fifth birthday.

References

1. Carlson, D.A., Leitmann, G.: An equivalent problem approach to absolute extrema for calculus of variations problems with differential constraints. Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms **18**(1), 1–15 (2011)
2. Cesari, L.: Optimization-theory and applications: problems with ordinary differential equations. In: Applications of Applied Mathematics, vol. 17. Springer, New York (1983)
3. Dockner, E.J., Leitmann, G.: Coordinate transformation and derivation of open-loop Nash equilibrium. J. Optim. Theory Appl. **110**(1), 1–16 (2001)
4. Leitmann, G.: A note on absolute extrema of certain integrals. Int. J. Non-Linear Mech. **2**, 55–59 (1967)
5. Roos, C.F.: A mathematical theory of competition. Am. J. Math. **46**, 163–175 (1925)
6. Roos, C.F.: Dynamical economics. Proc. Natl. Acad. Sci. **13**, 145–150 (1927)
7. Roos, C.F.: A dynamical theory of economic equilibrium. Proceedings of the National Academy of Sciences **13**, 280–285 (1927)
8. Roos, C.F.: A dynamical theory of economics. J. Polit. Econ. **35**(5), 632–656 (1927)
9. Roos, C.F.: Generalized Lagrange problems in the calculus of variations. Trans. Am. Math. Soc. **30**(2), 360–384 (1928)
10. Roos, C.F.: Generalized Lagrange problems in the calculus of variations. Trans. Am. Math. Soc. **31**(1), 58–70 (1929)
11. Wagener, F.O.O.: On the Leitmann equivalent problem approach. J. Optim. Theory Appl. **142**(1), 229–242 (2009)

Chapter 9
Nash Equilibrium Seeking for Dynamic Systems with Non-quadratic Payoffs

Paul Frihauf, Miroslav Krstic, and Tamer Başar

Abstract We consider general, stable nonlinear differential equations with N inputs and N outputs, where in the steady state, the output signals represent the payoff functions of a noncooperative game played by the steady-state values of the input signals. To achieve locally stable convergence to the resulting steady-state Nash equilibria, we introduce a non-model-based approach, where the players determine their actions based only on their own payoff values. This strategy is based on the extremum seeking approach, which has previously been developed for standard optimization problems and employs sinusoidal perturbations to estimate the gradient. Since non-quadratic payoffs create the possibility of multiple, isolated Nash equilibria, our convergence results are local. Specifically, the attainment of any particular Nash equilibrium is not assured for all initial conditions, but only for initial conditions in a set around that specific stable Nash equilibrium. For non-quadratic costs, the convergence to a Nash equilibrium is not perfect, but is biased in proportion to the perturbation amplitudes and the higher derivatives of the payoff functions. We quantify the size of these residual biases.

Keywords Extremum seeking • Learning • Nash equilibria • Noncooperative games

P. Frihauf (✉) • M. Krstic
Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA
e-mail: pfrihauf@ucsd.edu; krstic@ucsd.edu

T. Başar
Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801-2307, USA
e-mail: basar1@beckman.illinois.edu

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_9,

9.1 Introduction

We study the problem of solving noncooperative games with N players in real time by employing a non-model-based approach, where the players' actions are the inputs to a general, stable nonlinear differential equation, whose outputs are the players' payoff values. By utilizing deterministic extremum seeking with sinusoidal perturbations, the players achieve local attainment of their Nash strategies without the need for any model information. We analyze games where the dynamic system acts on a faster time scale compared to the time scale of the players' strategies and where the players have non-quadratic payoff functions, which may result in multiple, isolated Nash equilibria. We study the effect of these non-quadratic functions on the players' convergence to a Nash equilibrium.

Most algorithms designed to achieve convergence to a Nash equilibrium require modeling information for the game and assume that the players can observe the actions of the other players. Players update their actions based on the gradient of their payoff functions in [34]. In [25], the stability of general player adjustments is analyzed under the assumption that a player's response mapping is a contraction, which, for games with quadratic payoff functions, is ensured by a diagonal dominance condition. Distributed iterative algorithms are designed in [24] for the computation of equilibria in a general class of non-quadratic convex Nash games with conditions for the contraction of general nonlinear operators to achieve convergence specified. A strategy known as fictitious play (employed in finite games) depends on the actions of the other players so that a player can devise a best response. A dynamic version of fictitious play and gradient response, which also includes an entropy term, is developed in [37] and is shown to converge to a mixed-strategy Nash equilibrium in cases that previously did not converge. In [44], a synchronous distributed learning algorithm, where players remember their own actions and utility values from the previous two time steps, is shown to converge in probability to the set of restricted Nash equilibria.

Other works have focused on learning strategies where players determine their actions based on their own payoff functions. No-regret learning for players with finite strategy spaces is shown in [20] to have similar convergence properties as fictitious play. In [18], impossibility results are established, which show that uncoupled dynamics do not converge to a Nash equilibrium in general. These results are extended in [19] to show that for finite games, uncoupled strategies that incorporate random experimentation and a finite memory lead to almost sure convergence to a Nash equilibrium. In [13], players in finite games employ regret testing, which utilizes only a player's payoff values and randomly evaluates alternate strategies, to converge in probability to the set of stage-game Nash equilibria. Another completely uncoupled learning strategy, where players experiment with alternative strategies, is presented in [42] and leads to behaviors that come close to a pure Nash equilibrium "a high proportion of the time." Lower bounds on the number of steps required to reach a Nash equilibrium, which are exponential in the number of players, are derived in [17] for finite games when players use uncoupled strategies. An approach, which is similar to our Nash seeking method (found in [23] and in this paper), is studied in [39] to solve coordination problems in mobile sensor networks. Additional results on learning in games can be found in [9, 14, 38]. Other diverse engineering applications of game theory include the design of communication networks in [1, 4, 27, 35], integrated structures and controls in [33], and distributed consensus protocols in [6,28,36]. A comprehensive treatment of static and dynamic noncooperative game theory can be found in [5].

The results of this work extend the methods of extremum seeking [3,26,29,31,40, 41], originally developed for standard optimization problems. The extremum seeking method, which performs non-model based gradient estimation, has been used in a variety of applications, such as steering vehicles toward a source in GPS-denied environments [10, 11, 43], controlling flow separation [7] and Tokamak plasmas [8], reducing the impact velocity of an electromechanical valve actuator [32], and optimizing nonisothermal continuously stirred tank reactors [15] and HCCI engine control [22].

In this work, N players in a noncooperative game that has a dynamic mapping from the players' actions to their payoff values employ an extremum seeking strategy to stably attain a pure Nash equilibrium. The key feature of our approach is that the players are not required to know the mathematical model of their payoff function or the underlying model of the game. They need to measure only their own payoff values to determine their respective real-valued actions. Consequently, this learning strategy is *radically uncoupled* according to the terminology of [13]. We consider payoff functions that satisfy a diagonal dominance condition and are non-quadratic, which allows for the possibility of multiple, isolated Nash equilibria. Our convergence result is local in the sense that convergence to any particular Nash equilibrium is assured only for initial conditions in a set around that specific stable Nash equilibrium. Moreover, this convergence is biased in proportion to the perturbation amplitudes and the higher derivatives of the payoff functions.

The paper is organized as follows: we provide the general problem statement in Sect. 9.2 and introduce the Nash equilibrium seeking strategy in Sect. 9.3. In Sects. 9.4 and 9.5, we use general averaging theory and singular perturbation theory to analyze the convergence properties of the game. Finally, we provide a numerical example for a two-player game in Sect. 9.6 and conclude with Sect. 9.7.

9.2 Problem Statement

Consider a noncooperative game with N players and a dynamic mapping from the players' actions u_i to their payoff values J_i, which the players wish to maximize. Specifically, we consider a general nonlinear model,

$$\dot{x} = f(x,u), \tag{9.1}$$

$$J_i = h_i(x), \quad i = 1,\ldots,N, \tag{9.2}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^N$ is a vector of the players' actions, $J_i \in \mathbb{R}$ is the payoff value of player i, $f : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}^n$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ are smooth, and h_i is a possibly non-quadratic function. The inclusion of the dynamic system (9.1) in the game structure and the consideration of non-quadratic payoff functions is motivated by oligopoly games with nonlinear demand and cost functions [30].

We make the following assumptions about the N-player game:

Assumption 9.1. *There exists a smooth function $l : \mathbb{R}^N \to \mathbb{R}^n$ such that*

$$f(x,u) = 0 \quad \textit{if and only if} \quad x = l(u). \tag{9.3}$$

Assumption 9.2. *For each $u \in \mathbb{R}^N$, the equilibrium $x = l(u)$ of the system (9.1) is locally exponentially stable.*

Hence, we assume that for any action by the players, the plant is able to stabilize the equilibrium. We can relax the requirement for each $u \in \mathbb{R}^N$ as we need to only be concerned with the action sets of the players, namely, $u \in U = U_1 \times \cdots \times U_N \subset \mathbb{R}^N$.

The following assumptions are central to our Nash seeking scheme as they ensure that at least one stable Nash equilibrium exists.

Assumption 9.3. *There exists at least one, possibly multiple, isolated stable Nash equilibria $u^* = [u_1^*, \ldots, u_N^*]$ such that*

$$\frac{\partial (h_i \circ l)}{\partial u_i}(u^*) = 0, \tag{9.4}$$

$$\frac{\partial^2 (h_i \circ l)}{\partial u_i^2}(u^*) < 0, \tag{9.5}$$

for all $i \in \{1, \ldots, N\}$.

Assumption 9.4. *The matrix,*

$$\Lambda = \begin{bmatrix} \frac{\partial^2 (h_1 \circ l)(u^*)}{\partial u_1^2} & \frac{\partial^2 (h_1 \circ l)(u^*)}{\partial u_1 \partial u_2} & \cdots & \frac{\partial^2 (h_1 \circ l)(u^*)}{\partial u_1 \partial u_N} \\ \frac{\partial^2 (h_2 \circ l)(u^*)}{\partial u_1 \partial u_2} & \frac{\partial^2 (h_2 \circ l)(u^*)}{\partial u_2^2} & & \\ \vdots & & \ddots & \\ \frac{\partial^2 (h_N \circ l)(u^*)}{\partial u_1 \partial u_N} & & & \frac{\partial^2 (h_N \circ l)(u^*)}{\partial u_N^2} \end{bmatrix}, \tag{9.6}$$

is diagonally dominant and hence, nonsingular.

By Assumptions 9.3 and 9.4, Λ is Hurwitz.

For this game, we seek to attain a Nash equilibrium u^* stably by employing an algorithm that does not require the players to know the actions of the other players, the mathematical form of the payoff functions h_i, or the dynamical system f.

9.3 Nash Equilibrium Seeking

Deterministic extremum seeking, which is a non-model-based real-time optimization strategy, enables player i to attain a Nash equilibrium by evolving its action u_i according to the measured value of its payoff J_i. Specifically, when employing this algorithm, player i implements the following strategy:

$$u_i(t) = \hat{u}_i(t) + \mu_i(t), \tag{9.7}$$

$$\dot{\hat{u}}_i(t) = k_i \mu_i(t) J_i(t), \tag{9.8}$$

where $\mu_i(t) = a_i sin(\omega_i t + \varphi_i)$, J_i is given by (9.2), and $a_i, k_i, \omega_i > 0$. Figure 9.1 depicts a noncooperative game played by two players implementing the extremum seeking strategy (9.7)–(9.8) to attain a Nash equilibrium. Note how this strategy requires only the payoff value J_i to be known.

We select the parameters $k_i = \varepsilon \underline{\omega} K_i = O(\varepsilon \underline{\omega})$ where $\underline{\omega} = \min_i\{\omega_i\}$ and ε, $\underline{\omega}$ are small positive constants. As will be seen later, the perturbation signal gain a_i must also be small. Intuitively, $\underline{\omega}$ is small since the players' actions should evolve more slowly than the dynamic system, creating an overall system with two time scales. In the limiting case of an infinitely fast dynamic system, i.e., a static game where the players' actions are direct inputs to their payoff functions, $\underline{\omega}$ no longer needs to be small.

We denote the error relative to the Nash equilibrium as

$$\begin{aligned} \tilde{u}_i(t) &= u_i(t) - \mu_i(t) - u_i^*, \\ &= \hat{u}_i(t) - u_i^*. \end{aligned} \tag{9.9}$$

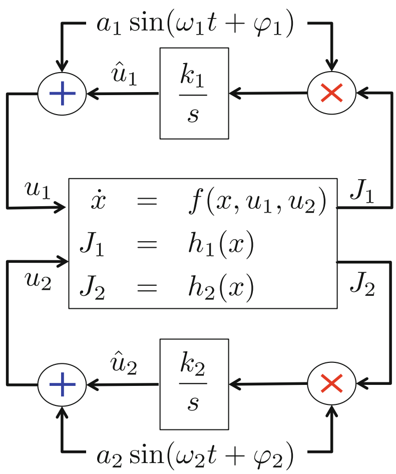


Fig. 9.1 Deterministic Nash seeking scheme employed by two players in a dynamic system with non-quadratic payoffs

and formulate the error system in the time scale $\tau = \underline{\omega} t$ as

$$\underline{\omega}\frac{\mathrm{d}x}{\mathrm{d}\tau} = f(x, u^* + \tilde{u} + \mu(\tau)), \tag{9.10}$$

$$\frac{\mathrm{d}\tilde{u}_i}{\mathrm{d}\tau} = \varepsilon K_i \mu_i(\tau) h_i(x), \tag{9.11}$$

where $\tilde{u} = [\tilde{u}_1, \ldots, \tilde{u}_N]$, $\mu(\tau) = [\mu_1(\tau), \ldots, \mu_N(\tau)]$ and $\mu_i(\tau) = a_i \sin(\omega_i \tau / \underline{\omega} + \varphi_i)$. The system (9.10)–(9.11) is in the standard singular perturbation form with $\underline{\omega}$ as a small parameter. Since ε is also small, we anlayze (9.10)–(9.11) using the general averaging theory for the quasi-steady state of (9.10), followed by the use of the singular perturbation theory for the full system.

9.4 General Averaging Analysis

For the averaging analysis, we first "freeze" x in (9.10) at its quasi-steady state

$$x = l(u^* + \tilde{u} + \mu(\tau)) \tag{9.12}$$

and substitute (9.12) into (9.11) to obtain the "reduced system,"

$$\frac{\mathrm{d}\tilde{u}_i}{\mathrm{d}\tau} = \varepsilon K_i \mu_i(\tau) p_i(u^* + \tilde{u} + \mu(\tau)), \tag{9.13}$$

where $p_i(u^* + \tilde{u} + \mu(\tau)) = (h_i \circ l)(u^* + \tilde{u} + \mu(\tau))$. This system's form allows for the use of general averaging theory [12, 21] and leads to the result:

Theorem 9.1. *Consider the system (9.13) for an N-player game under Assumptions 9.3 and 9.4 and where $\omega_i \neq \omega_j$, $\omega_i \neq \omega_j + \omega_k$, $2\omega_i \neq \omega_j + \omega_k$, and $\omega_i \neq 2\omega_j + \omega_k$ for all distinct $i,\ j,\ k \in \{1, \ldots, N\}$. There exist M, $m > 0$ and $\bar{\varepsilon}$, $\bar{a}$ such that, for all $\varepsilon \in (0, \bar{\varepsilon})$ and $a_i \in (0, \bar{a})$, if $|\Delta(0)|$ is sufficiently small, then for all $\tau \geq 0$,*

$$\left|\Delta(\tau)\right| \leq M \mathrm{e}^{-m\tau} \left|\Delta(0)\right| + O\left(\varepsilon + \max_i a_i^3\right), \tag{9.14}$$

where

$$\Delta(\tau) = \left[\tilde{u}_1(\tau) - \textstyle\sum_{j=1}^N c_{jj}^1 a_j^2, \ldots, \tilde{u}_N(\tau) - \sum_{j=1}^N c_{jj}^N a_j^2\right], \tag{9.15}$$

and

$$\begin{bmatrix} c_{jj}^1 \\ \vdots \\ c_{jj}^{j-1} \\ c_{jj}^j \\ c_{jj}^{j+1} \\ \vdots \\ c_{jj}^N \end{bmatrix} = -\frac{1}{4}\Lambda^{-1} \begin{bmatrix} \frac{\partial^3 p_1}{\partial u_1 \partial u_j^2}(u^*) \\ \vdots \\ \frac{\partial^3 p_{j-1}}{\partial u_{j-1} \partial u_j^2}(u^*) \\ \frac{1}{2}\frac{\partial^3 p_j}{\partial u_j^3}(u^*) \\ \frac{\partial^3 p_{j+1}}{\partial u_j^2 \partial u_{j+1}}(u^*) \\ \vdots \\ \frac{\partial^3 p_N}{\partial u_j^2 \partial u_N}(u^*) \end{bmatrix}. \tag{9.16}$$

Proof. As already noted, the form of (9.13) allows for the application of general averaging theory, which yields the average system,

$$\frac{\mathrm{d}\tilde{u}_i^{\mathrm{ave}}}{\mathrm{d}\tau} = \varepsilon K_i \lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i(\tau) p_i(u^* + \tilde{u}^{\mathrm{ave}} + \mu(\tau))\,\mathrm{d}\tau. \tag{9.17}$$

The equilibrium, $\tilde{u}^{\mathrm{e}} = [\tilde{u}_1^{\mathrm{e}}, \ldots, \tilde{u}_N^{\mathrm{e}}]$, of (9.17) satisifies

$$0 = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i(\tau) p_i(u^* + \tilde{u}^{\mathrm{e}} + \mu(\tau))\,\mathrm{d}\tau, \tag{9.18}$$

for all $i \in \{1, \ldots, N\}$, and we postulate that $\tilde{u}^{\mathrm{e}}$ has the form,

$$\tilde{u}_i^{\mathrm{e}} = \sum_{j=1}^N b_j^i a_j + \sum_{j=1}^N \sum_{k \ge j}^N c_{jk}^i a_j a_k + O\left(\max_i a_i^3\right). \tag{9.19}$$

By expanding p_i about u^* in (9.18) with a Taylor polynomial and substituting (9.19), the unknown coefficients b_j^i and c_{jk}^i can be determined. The Taylor polynomial approximation [2] requires p_i to be $k+1$ times differentiable, namely,

$$\begin{aligned} p_i(u^* + \tilde{u}^{\mathrm{e}} + \mu(\tau)) &= \sum_{|\alpha|=0}^{k} \frac{D^\alpha p_i(u^*)}{\alpha!}(\tilde{u}^{\mathrm{e}} + \mu(\tau))^\alpha \\ &\quad + \sum_{|\alpha|=k+1} \frac{D^\alpha p_i(\zeta)}{\alpha!}(\tilde{u}^{\mathrm{e}} + \mu(\tau))^\alpha, \\ &= \sum_{|\alpha|=0}^{k} \frac{D^\alpha p_i(u^*)}{\alpha!}(\tilde{u}^{\mathrm{e}} + \mu(\tau))^\alpha + O\left(\max_i a_i^{k+1}\right), \end{aligned} \tag{9.20}$$

where ζ is a point on the line segment that connects the points u^* and $u^*+\tilde{u}^{\mathrm{e}}+\mu(\tau)$. In (9.20), we have used multi-index notation, namely, $\alpha=(\alpha_1,\ldots,\alpha_N)$, $|\alpha|=\alpha_1+\cdots+\alpha_N$, $\alpha!=\alpha_1!\cdots\alpha_N!$, $u^\alpha=u_1^{\alpha_1}\cdots u_N^{\alpha_N}$, and $D^\alpha(h_i\circ l)=\partial^{|\alpha|}p_i/\partial u_1^{\alpha_1}\cdots\partial u_N^{\alpha_N}$. The second term on the last line of (9.20) follows by substituting the postulated form of $\tilde{u}^{\mathrm{e}}$ (9.19).

We select $k=3$ to capture the effect of the third order derivative on the system as a representative case. The effect of higher-order derivatives can be studied if the third order derivative is zero. Substituting (9.20) into (9.18) and computing the average of each term gives

$$\begin{aligned}0=\frac{a_i^2}{2}\Bigg[&\tilde{u}_i^{\mathrm{e}}\frac{\partial^2 p_i}{\partial u_i^2}(u^*)+\sum_{j\neq i}^{N}\tilde{u}_j^{\mathrm{e}}\frac{\partial^2 p_i}{\partial u_i\partial u_j}(u^*)\\&+\left(\frac{1}{2}(\tilde{u}_i^{\mathrm{e}})^2+\frac{a_i^2}{8}\right)\frac{\partial^3 p_i}{\partial u_i^3}(u^*)+\tilde{u}_i^{\mathrm{e}}\sum_{j\neq i}^{N}\tilde{u}_j^{\mathrm{e}}\frac{\partial^3 p_i}{\partial u_i^2\partial u_j}(u^*)\\&+\sum_{j\neq i}^{N}\left(\frac{1}{2}(\tilde{u}_j^{\mathrm{e}})^2+\frac{a_j^2}{4}\right)\frac{\partial^3 p_i}{\partial u_i\partial u_j^2}(u^*)\\&+\sum_{j\neq i}^{N}\sum_{\substack{k>j\\k\neq i}}^{N}\tilde{u}_j^{\mathrm{e}}\tilde{u}_k^{\mathrm{e}}\frac{\partial^3 p_i}{\partial u_i\partial u_j\partial u_k}(u^*)\Bigg]+O(\max_i a_i^5),\end{aligned}\tag{9.21}$$

where we have noted (9.4), utilized (9.19), and computed the integrals shown in the appendix.

Substituting (9.19) into (9.21) and matching first order powers of a_i gives

$$\begin{bmatrix}0\\\vdots\\0\end{bmatrix}=a_1\Lambda\begin{bmatrix}b_1^1\\\vdots\\b_1^N\end{bmatrix}+\cdots+a_N\Lambda\begin{bmatrix}b_1^1\\\vdots\\b_N^N\end{bmatrix},\tag{9.22}$$

which implies that $b_j^i=0$ for all i, j since Λ is nonsingular by Assumption 9.4. Similarly, matching second order terms of a_i, and substituting $b_j^i=0$ to simplify the resulting expressions, yields

$$\begin{bmatrix}0\\\vdots\\0\end{bmatrix}=\sum_{j=1}^{N}\sum_{k>j}^{N}a_ja_k\Lambda\begin{bmatrix}c_{jk}^1\\\vdots\\c_{jk}^N\end{bmatrix}+\sum_{j=1}^{N}a_j^2\left(\Lambda\begin{bmatrix}c_{jj}^1\\\vdots\\c_{jj}^N\end{bmatrix}+\frac{1}{4}\begin{bmatrix}\frac{\partial^3 p_1}{\partial u_1\partial u_j^2}(u^*)\\\vdots\\\frac{\partial^3 p_{j-1}}{\partial u_{j-1}\partial u_j^2}(u^*)\\\frac{1}{2}\frac{\partial^3 p_j}{\partial u_j^3}(u^*)\\\frac{\partial^3 p_{j+1}}{\partial u_j^2\partial u_{j+1}}(u^*)\\\vdots\\\frac{\partial^3 p_N}{\partial u_j^2\partial u_N}(u^*)\end{bmatrix}\right).\tag{9.23}$$

Thus, $c^i_{jk} = 0$ for all i, j, k when $j \neq k$, and c^i_{jj} is given by (9.16). Therefore, the equilibrium of the average system is

$$\tilde{u}_i^{\mathrm{e}} = \sum_{j=1}^{N} c^i_{jj} a_j^2 + O\left(\max_i a_i^3\right). \tag{9.24}$$

By again utilizing a Taylor polynomial approximation, one can show that the Jacobian $\Psi^{\mathrm{ave}} = [\psi_{i,j}]_{N\times N}$ of (9.17) at $\tilde{u}^e$ has elements given by

$$\begin{aligned}\psi_{i,j} &= \varepsilon K_i \lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i(\tau)\frac{\partial p_i}{\partial u_j}(u^* + \tilde{u}^{\mathrm{e}} + \mu(\tau))\,\mathrm{d}\tau, \\ &= \frac{1}{2}\varepsilon K_i a_i^2 \frac{\partial^2 p_i}{\partial u_i \partial u_j}(u^*) + O\left(\varepsilon \max_i a_i^3\right),\end{aligned} \tag{9.25}$$

and is Hurwitz by Assumptions 9.3 and 9.4 for sufficiently small a_i, which implies that the equilibrium (9.24) of the average system (9.17) is exponentially stable, i.e., there exist constants $M, m > 0$ such that

$$|\tilde{u}^{\mathrm{ave}}(\tau) - \tilde{u}^{\mathrm{e}}| \leq M\,\mathrm{e}^{-m\tau}\,|\tilde{u}^{\mathrm{ave}}(0) - \tilde{u}^{\mathrm{e}}|. \tag{9.26}$$

From the general averaging theory [12, 21], we have

$$|\tilde{u}(\tau) - \tilde{u}^{\mathrm{e}}| \leq M\,\mathrm{e}^{-m\tau}\,|\tilde{u}(0) - \tilde{u}^{\mathrm{e}}| + O(\varepsilon), \tag{9.27}$$

and defining $\Delta(\tau)$ as in Theorem 9.1 completes the proof. □

From Theorem 9.1, we see that u of reduced system (9.13) converges to a region that is biased away from the Nash equilibrium u^*. This bias is in proportion to the perturbation magnitudes a_i and the third derivatives of the payoff functions, which are captured by the coefficients c^i_{jj}. Specifically, $\hat{u}_i$ of the reduced system converges to $u_i^* + \sum_{j=1}^{N} c^i_{jj} a_j^2 + O(\varepsilon + \max_i a_i^3)$ as $t \to \infty$.

For a two-player game, Theorem 9.1 holds with the obvious change to omit any reference to ω_k and with the less obvious inclusion of the requirement $\omega_i \neq 3\omega_j$. The requirement $\omega_i \neq 3\omega_j$ is not explicitly stated in Theorem 9.1 since the combination of $\omega_i \neq \omega_j$ and $\omega_i \neq 2\omega_j + \omega_k$ for all distinct i, j, k implies that $\omega_i \neq 3\omega_j$. If the payoff functions were quadratic, rather than non-quadratic, the requirements for the perturbation frequencies would be simply, $\omega_i \neq \omega_j$, $\omega_i \neq \omega_j + \omega_k$ for the N-player game, and $\omega_i \neq \omega_j$, $\omega_i \neq 2\omega_j$ for the two-player game.

9.5 Singular Perturbation Analysis

We analyze the full system (9.10)–(9.11) in the time scale $\tau = \underline{\omega} t$ using singular perturbation theory. First, we note that by [12, Theorem 14.4] and Theorem 9.1 there exists an exponentially stable almost periodic solution $\tilde{u}^{\mathrm{a}} = [\tilde{u}_1^{\mathrm{a}}, \ldots, \tilde{u}_N^{\mathrm{a}}]$ such that

$$\frac{\mathrm{d}\tilde{u}_i^{\mathrm{a}}}{\mathrm{d}\tau} = \varepsilon K_i \mu_i(\tau) p_i(u^* + \tilde{u}^{\mathrm{a}} + \mu(\tau)). \tag{9.28}$$

Moreover, $\tilde{u}^{\mathrm{a}}$ is unique within a neighborhood of the average solution $\tilde{u}^{\mathrm{ave}}$ [16].

We define $z_i = \tilde{u}_i - \tilde{u}_i^{\mathrm{a}}$ and obtain

$$\frac{\mathrm{d}z_i}{\mathrm{d}\tau} = \varepsilon K_i \mu_i(\tau)\left[h_i(x) - p_i(u^* + \tilde{u}^{\mathrm{a}} + \mu(\tau))\right], \tag{9.29}$$

$$\underline{\omega}\frac{\mathrm{d}x}{\mathrm{d}\tau} = f(x, u^* + z + \tilde{u}^{\mathrm{a}} + \mu(\tau)), \tag{9.30}$$

which from Assumption 9.1, the quasi-steady state is

$$x = l(u^* + z + \tilde{u}^{\mathrm{a}} + \mu(\tau)). \tag{9.31}$$

Thus, the reduced model is given by

$$\frac{\mathrm{d}z_i}{\mathrm{d}\tau} = \varepsilon K_i \mu_i(\tau)\left[p_i(u^* + z + \tilde{u}^{\mathrm{a}} + \mu(\tau)) - p_i(u^* + \tilde{u}^{\mathrm{a}} + \mu(\tau))\right], \tag{9.32}$$

which has an equilibrium at $z = 0$ that is exponentially stable for sufficiently small a_i as shown in Sect. 9.4.

To formulate the boundary layer model, let $y = x - l(u^* + z + \tilde{u}^{\mathrm{a}} + \mu(\tau))$, and then in the time scale $t = \tau/\underline{\omega}$,

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}t} &= f(y + l(u^* + z + \tilde{u}^{\mathrm{a}} + \mu(\tau)), u^* + z + \tilde{u}^{\mathrm{a}} + \mu(\tau)),\\ &= f(y + l(u), u),\end{aligned} \tag{9.33}$$

where $u = u^* + \tilde{u} + \mu(\tau)$ should be viewed as a parameter independent of the time variable t. Since $f(l(u), u) = 0$, $y = 0$ is an equilibrium of (9.33) and is exponentially stable by Assumption 9.2.

With the exponential stability of the origin established for both the reduced model (9.32) and the boundary layer model (9.33), we apply Tikhonov's Theorem on the Infinite Interval [21, Theorem 11.2], where $\underline{\omega}$ is the singular perturbation parameter, to conclude that the solution $z(\tau)$ of (9.29) is $O(\underline{\omega})$-close to the solution of the reduced model (9.32). Consequently, $\tilde{u}(\tau)$ converges exponentially to the solution $\tilde{u}^{\mathrm{a}}(t)$, which is $O(\varepsilon)$-close to the equilibrium $\tilde{u}^{\mathrm{e}}$. Hence, as

$\tau \to \infty$, $\tilde{u}(\tau) = [\tilde{u}_1(\tau), \ldots, \tilde{u}_N(\tau)]$ converges to an $O(\underline{\omega} + \varepsilon)$-neighborhood of $\left[\sum_{j=1}^{N} c_{jj}^1 a_j^2, \ldots, \sum_{j=1}^{N} c_{jj}^N a_j^2\right] + O(\max_i a_i^3)$. Since $u(\tau) - u^* = \tilde{u}(\tau) + \mu(\tau) = \tilde{u}(\tau) + O(\max_i a_i)$, $u(\tau)$ converges to an $O(\underline{\omega} + \varepsilon + \max_i a_i)$-neighborhood of u^*.

Also from Tikhonov's Theorem on the Infinite Interval, the solution $x(\tau)$ of (9.10) satisfies

$$x(\tau) - l(u^* + \tilde{u}(\tau) + \mu(\tau)) - y(t) = O(\underline{\omega}), \tag{9.34}$$

where $y(t)$ is the solution to the boundary layer model (9.33). We can write

$$x(\tau) - l(u^*) = O(\underline{\omega}) + l(u^* + \tilde{u}(\tau) + \mu(\tau)) - l(u^*) + y(t). \tag{9.35}$$

From the convergence properties of $\tilde{u}(\tau)$ and because $y(t)$ is exponentially decaying, $x(\tau) - l(u^*)$ exponentially converges to an $O(\underline{\omega} + \varepsilon + \max_i a_i)$-neighborhood of the origin. Thus, $J_i = h_i(x)$ exponentially converges to an $O(\underline{\omega} + \varepsilon + \max_i a_i)$-neighborhood of the payoff value $(h_i \circ l)(u^*)$.

We summarize with the following theorem:

Theorem 9.2. *Consider the system (9.1)–(9.2), (9.7)–(9.8) for an N-player game under Assumptions 9.1–9.4 and where $\omega_i \neq \omega_j$, $\omega_i \neq \omega_j + \omega_k$, $2\omega_i \neq \omega_j + \omega_k$, and $\omega_i \neq 2\omega_j + \omega_k$ for all distinct i, j, $k \in \{1, \ldots, N\}$. There exists $\omega^* > 0$ and for any $\omega \in (0, \omega^*)$ there exist $\varepsilon^*, a^* > 0$ such that for the given ω and any $\varepsilon \in (0, \varepsilon^*)$, $\max_i a_i \in (0, a^*)$, the solution $(x(t), u_1(t), \ldots, u_N(t))$ converges exponentially to an $O(\omega + \varepsilon + \max_i a_i)$-neighborhood of the point $(l(u^*), u_1^*, \ldots, u_N^*)$, provided the initial conditions are sufficiently close to this point.*

Due to the Nash seeking strategy's continual perturbation of the players' actions, we achieve exponential convergence to a neighborhood of u^*, rather than u^* itself. The size of this neighborhood depends directly on the selected Nash seeking parameters, as seen by Theorem 9.2. Thus, smaller parameters lead to a smaller convergence neighborhood, but they also lead to slower convergence rates. (The reader is referred to [41] for detailed analysis of this design trade-off for extremum seeking controllers.) If another algorithm were used in parallel to detect convergence of a player's actions on the average, a player could either decrease the size of its perturbation, making the convergence neighborhood smaller, or choose a constant action based on its convergence detection. However, with a constant action, a player will not be able to adapt to any future changes in the game.

By achieving exponential convergence, the players are able to achieve convergence in the presence of a broader class of perturbations to the game than if convergence were merely asymptotic (see [21, Chap. 9]).

9.6 Numerical Example

For an example game with players that employ the extremum seeking strategy (9.7)–(9.8), we consider the system,

$$\dot{x}_1 = -4x_1 + x_1x_2 + u_1, \tag{9.36}$$

$$\dot{x}_2 = -4x_2 + u_2, \tag{9.37}$$

$$J_1 = -16x_1^2 + 8x_1^2x_2 - x_1^2x_2^2 - 4x_1x_2^2 + 15x_1x_2 + 4x_1, \tag{9.38}$$

$$J_2 = -64x_2^3 + 48x_1x_2 - 12x_1x_2^2, \tag{9.39}$$

whose equilibrium state is given by

$$\bar{x}_1 = \frac{4u_1}{16-u_2}, \quad \bar{x}_2 = \frac{1}{4}u_2. \tag{9.40}$$

The Jacobian at the equilibrium $(\bar{x}_1, \bar{x}_2)$ is

$$\begin{bmatrix} -\frac{1}{4}(16-u_2) & \frac{4u_1}{16-u_2} \\ 0 & -4 \end{bmatrix}, \tag{9.41}$$

which is Hurwitz for $u_2 < 16$. Thus, $(\bar{x}_1, \bar{x}_2)$ is locally exponentially stable, but not for all $(u_1, u_2) \in \mathbb{R}^2$, violating Assumption 9.2. However, as noted earlier, this restrictive requirement of local exponential stability for all $u \in \mathbb{R}^N$ was done merely for notational convenience; we actually only require this assumption to hold for the players' action sets. In this example, we restrict the players' actions to the set

$$U = \{(u_1, u_2) \in \mathbb{R}^2 \mid u_1, u_2 \geq 0, u_2 < 16\}. \tag{9.42}$$

At $x = \bar{x}$, the payoffs are

$$J_1 = -u_1^2 + u_1u_2 + u_1, \tag{9.43}$$

$$J_2 = -u_2^3 + 3u_1u_2, \tag{9.44}$$

which, on U, yield a Nash equilibrium: $(u_1^*, u_2^*) = (1,1)$. At u^*, this game satisfies Assumptions 9.3 and 9.4, implying the stability of the reduced model average system, which can be found explicitly according to (9.17) to be

$$\frac{d\tilde{u}_1^{\text{ave}}}{d\tau} = \varepsilon K_1 a_1^2 \left(-\tilde{u}_1^{\text{ave}} + \frac{1}{2}\tilde{u}_2^{\text{ave}} \right), \tag{9.45}$$

$$\frac{d\tilde{u}_2^{\text{ave}}}{d\tau} = \varepsilon K_2 a_2^2 \left(\frac{3}{2}\tilde{u}_1^{\text{ave}} - 3u_2^*\tilde{u}_2^{\text{ave}} - \frac{3}{2}(\tilde{u}_2^{\text{ave}})^2 - \frac{3}{8}a_2^2 \right), \tag{9.46}$$

with equilibria,

$$\tilde{u}_1^{\text{e}} = \frac{1}{8}(1-4u_2^*) \pm \frac{1}{8}\sqrt{(1-4u_2^*)^2 - 4a_2^2}, \tag{9.47}$$

$$\tilde{u}_2^{\text{e}} = 2\tilde{u}_1^{\text{e}}, \tag{9.48}$$

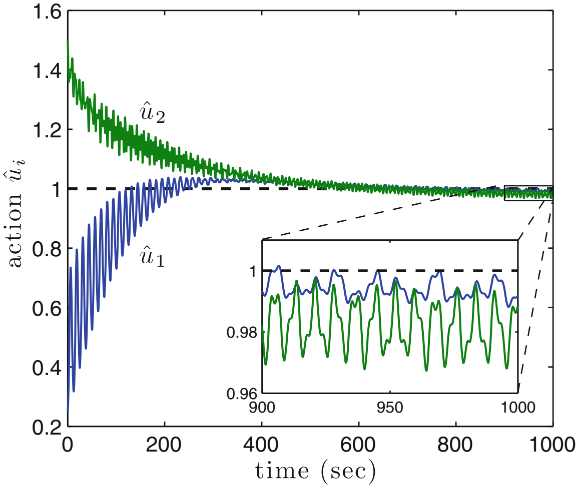


Fig. 9.2 Time history of the two-player game initialized at $(u_1,u_2) = (0.25,1.5)$

compared to the postulated form (9.19),

$$\tilde{u}_1^{\mathrm{e,p}} = \frac{2}{1-4u_2^*}a_2^2 + O(\max_i a_i^3), \tag{9.49}$$

$$\tilde{u}_2^{\mathrm{e,p}} = \frac{4}{1-4u_2^*}a_2^2 + O(\max_i a_i^3). \tag{9.50}$$

For sufficiently small a_2, $\tilde{u}^{\mathrm{e}} \approx (0,0)$, $(-3/4,-3/2)$, whereas the postulated form $\tilde{u}^{\mathrm{e,p}} \approx (0,0)$ only. The equilibrium at $(-3/4,-3/2)$ corresponds to the point $(1/4,-1/2)$, which lies outside of U and is an intersection of the extremals $\partial J_1/\partial u_1 = 0$, $\partial J_2/\partial u_2 = 0$.

The Jacobian Ψ^{ave} of the average system is

$$\Psi^{\mathrm{ave}} = \begin{bmatrix} -\kappa_1 & \frac{1}{2}\kappa_1 \\ \frac{3}{2}\kappa_2 & -3\kappa_2\left(\tilde{u}_2^{\mathrm{e}} + u_2^*\right) \end{bmatrix}, \tag{9.51}$$

where $\kappa_1 = \varepsilon K_1 a_1^2$ and $\kappa_2 = \varepsilon K_2 a_2^2$, and its characteristic equation is given by,

$$\lambda^2 + \underbrace{\left(\kappa_1 + 3\kappa_2\left(\tilde{u}_2^{\mathrm{e}} + u_2^*\right)\right)}_{\alpha_1}\lambda + \underbrace{3\kappa_1\kappa_2\left(\tilde{u}_2^{\mathrm{e}} + u_2^* - \frac{1}{4}\right)}_{\alpha_2} = 0.$$

Thus, Ψ^{ave} is Hurwitz if and only if α_1 and α_2 are positive. For sufficiently small a_2 (so that $\tilde{u}^{\mathrm{e}} \approx (0,0)$), α_1, $\alpha_2 > 0$, which implies that u^* is a stable Nash equilibrium.

For the simulations, we select $k_1 = 1.5$, $k_2 = 2$, $a_1 = 0.09$, $a_2 = 0.03$, $\omega_1 = 0.5$, and $\omega_2 = 1.3$, where the parameters are chosen to be small, in particular the perturbation frequencies ω_i, since the perturbation must occur at a time scale that is slower than fast time scale of the nonlinear system. Figures 9.2 and 9.3

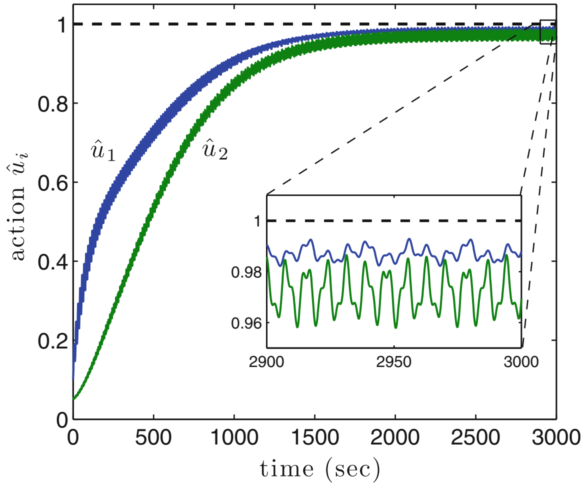


Fig. 9.3 Time history of the two-player game initialized at $(u_1, u_2) = (0.1, 0.05)$

depict the evolution of the players' actions $\hat{u}_1$ and $\hat{u}_2$ initialized at $(u_1(0), u_2(0)) = (\hat{u}_1(0), \hat{u}_2(0)) = (0.25, 1.5)$ and $(0.1, 0.05)$. The state (x_1, x_2) is initialized at the origin in both cases. We show $\hat{u}_i$ instead of u_i to better illustrate the convergence of the players' actions to a neighborhood about the Nash strategies since u_i contains the additive signal $\mu_i(t)$.

The slow initial convergence in Fig. 9.3 can best be explained by examining the phase portrait of the average of the reduced model $\hat{u}$-system, which can be shown to be

$$\dot{\hat{u}}_1^{\text{ave}} = \varepsilon K_1 a_1^2 \left(\frac{1}{2} - \hat{u}_1^{\text{ave}} + \frac{1}{2}\hat{u}_2^{\text{ave}} \right), \tag{9.52}$$

$$\dot{\hat{u}}_2^{\text{ave}} = \varepsilon K_2 a_2^2 \left(\frac{3}{2}\hat{u}_1^{\text{ave}} - \frac{3}{2}(\hat{u}_2^{\text{ave}})^2 - \frac{3}{8}a_2^2 \right), \tag{9.53}$$

and has equilibria given by

$$\begin{aligned} \hat{u}_1^{\text{e}} &= \frac{5}{8} \pm \frac{1}{8}\sqrt{9 - 4a_2^2}, \\ \hat{u}_2^{\text{e}} &= 2\hat{u}_1^{\text{e}} - 1. \end{aligned} \tag{9.54}$$

For $a_2 = 0.03$, $(\hat{u}_1^{\text{e}}, \hat{u}_2^{\text{e}}) = (0.9999, 0.9998)$ and $(0.2501, -0.4998)$. Figure 9.4 is the phase portrait of this system with the stable Nash equilibrium represented by a green circle, and the point $(1/4, -1/2)$ a red square, which is an unstable equilibrium in the phase portrait. The boundary of U is denoted by dashed red lines. The initial condition for Fig. 9.3 lies near the unstable point, so the trajectory travels almost

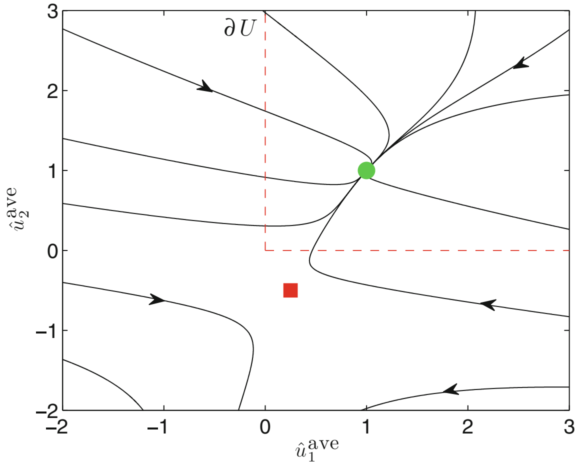


Fig. 9.4 Phase portrait for the $\hat{u}$-reduced model average system (9.52)–(9.53). The stable Nash equilibrium (*green circle*), the point $(1/4, -1/2)$ (*red square*), and the boundary of U (*red dashed lines*) are denoted

entirely along the eigenvector that points towards the stable equilibrium. We also see that the trajectories remain in U for initial conditions suitably close to the Nash equilibrium.

9.7 Conclusions

We have introduced a non-model-based approach for convergence to the Nash equilibria of noncooperative games with N players, where the game has a dynamic mapping from the players' actions to their payoff values. When employing this strategy each player measures only the value of its own payoff. The convergence is biased in proportion to both the payoff functions' higher derivatives and the perturbation signals' amplitudes for non-quadratic payoff functions, which give rise to the possibility of multiple, isolated Nash equilibria.

Even though we have considered only the case where the action variables of the players are scalars, the results equally apply to the vector case, namely $u_i \in \mathbb{R}^{n_i}$, by simply considering each different component of a player's action variable to be controlled by a different (virtual) player. In this case, the payoff functions of all virtual players corresponding to player i will be the same.

Acknowledgements This research was made with Government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a, and by grants from National Science Foundation, DOE, and AFOSR.

Appendix

The following integrals are computed to obtain (9.21), where we have assumed the frequencies satisfy $\omega_i \neq \omega_j$, $2\omega_i \neq \omega_j$, $3\omega_i \neq \omega_j$, $\omega_i \neq \omega_j + \omega_k$, $\omega_i \neq 2\omega_j + \omega_k$, $2\omega_i \neq \omega_j + \omega_k$, for distinct $i, j, k \in \{1, \ldots, N\}$ and defined $\gamma_i = \omega_i / \min_i\{\omega_i\}$:

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i(\tau)\,\mathrm{d}\tau &= \lim_{T\to\infty} \frac{a_i}{T}\int_0^T \sin(\gamma_i\tau + \varphi_i)\,\mathrm{d}\tau \\
&= 0,
\end{aligned} \tag{9.55}$$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i^2(\tau)\,\mathrm{d}\tau &= \lim_{T\to\infty} \frac{a_i^2}{2T}\int_0^T [1 - \cos(2\gamma_i\tau + 2\varphi_i)]\,\mathrm{d}\tau, \\
&= \frac{a_i^2}{2},
\end{aligned} \tag{9.56}$$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i^3(\tau)\,\mathrm{d}\tau &= \lim_{T\to\infty} \frac{a_i^3}{4T}\int_0^T [3\sin(\gamma_i\tau + \varphi_i) - \sin(3\gamma_i\tau + 3\varphi_i)]\,\mathrm{d}\tau, \\
&= 0,
\end{aligned} \tag{9.57}$$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i^4(\tau)\,\mathrm{d}\tau &= \lim_{T\to\infty} \frac{a_i^4}{8T}\int_0^T [3 - 4\cos(2\gamma_i\tau + 2\varphi_i) \\
&\quad + \cos(4\gamma_i\tau + 4\varphi_i)]\,\mathrm{d}\tau, \\
&= \frac{3a_i^4}{8},
\end{aligned} \tag{9.58}$$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i(\tau)\mu_j(\tau)\,\mathrm{d}\tau &= \frac{a_i a_j}{2T}\int_0^T [\cos((\gamma_i - \gamma_j)\tau + \varphi_i - \varphi_j) \\
&\quad - \cos((\gamma_i + \gamma_j)\tau + \varphi_i + \varphi_j)]\,\mathrm{d}\tau, \\
&= 0,
\end{aligned} \tag{9.59}$$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_i^2(\tau)\mu_j(\tau)\,\mathrm{d}\tau &= \frac{a_i^2 a_j}{2T}\int_0^T [\sin(\gamma_j\tau + \varphi_j) \\
&\quad - \cos(2\gamma_i\tau + 2\varphi_i)\sin(\gamma_j\tau + \varphi_j)]\,\mathrm{d}\tau, \\
&= \frac{a_i^2 a_j}{4T}\int_0^T [2\sin(\gamma_j\tau + \varphi_j) \\
&\quad - \sin((2\gamma_i + \gamma_j)\tau + 2\varphi_i + \varphi_j) \\
&\quad + \sin((2\gamma_i - \gamma_j)\tau + 2\varphi_i - \varphi_j)]\,\mathrm{d}\tau, \\
&= 0,
\end{aligned} \tag{9.60}$$

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\int_0^T \mu_i^3(\tau)\mu_j(\tau)\,\mathrm{d}\tau &= \frac{a_i^3 a_j}{4T}\int_0^T [3\sin(\gamma_i\tau+\varphi_j)\sin(\gamma_j\tau+\varphi_j) \\
&\quad -\sin(3\gamma_i\tau+3\varphi_i)\sin(\gamma_j\tau+\varphi_j)]\,\mathrm{d}\tau, \\
&= \frac{a_i^3 a_j}{8T}\int_0^T [3\cos((\gamma_i-\gamma_j)\tau+\varphi_i-\varphi_j) \\
&\quad -3\cos((\gamma_i+\gamma_j)\tau+\varphi_i+\varphi_j) \\
&\quad -\cos((3\gamma_i-\gamma_j)\tau+3\varphi_i-\varphi_j) \\
&\quad +\cos((3\gamma_i+\gamma_j)\tau+3\varphi_i+\varphi_j)]\,\mathrm{d}\tau, \\
&= 0, \qquad (9.61)
\end{aligned}$$

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\int_0^T \mu_i^2(\tau)\mu_j^2(\tau)\,\mathrm{d}\tau &= \lim_{T\to\infty}\frac{a_i^2 a_j^2}{8T}\int_0^T [2-2\cos(2\gamma_i\tau+2\varphi_i) \\
&\quad -2\cos(2\gamma_j\tau+2\varphi_j) \\
&\quad +\cos(2(\gamma_i-\gamma_j)\tau+2(\varphi_i-\varphi_j)) \\
&\quad +\cos(2(\gamma_i+\gamma_j)\tau+2(\varphi_i+\varphi_j))]\,\mathrm{d}\tau, \\
&= \frac{a_i^2 a_j^2}{4}, \qquad (9.62)
\end{aligned}$$

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\int_0^T \mu_i(\tau)\mu_j(\tau)\mu_k(\tau)\,\mathrm{d}\tau, &= \lim_{T\to\infty}\frac{a_i a_j a_k}{2T}\int_0^T [\cos((\gamma_i-\gamma_j)\tau+\varphi_i-\varphi_j) \\
&\quad -\cos((\gamma_i+\gamma_j)\tau+\varphi_i+\varphi_j)]\sin(\gamma_k\tau+\varphi_k)\,\mathrm{d}\tau, \\
&= \lim_{T\to\infty}\frac{a_i a_j a_k}{4T} \\
&\quad \times\int_0^T [\sin((\gamma_i-\gamma_j+\omega_k)\tau+\varphi_i-\varphi_j+\varphi_k) \\
&\quad -\sin((\gamma_i-\gamma_j-\gamma_k)\tau+\varphi_i-\varphi_j-\varphi_k) \\
&\quad -\sin((\gamma_i+\gamma_j+\gamma_k)\tau+\varphi_i+\varphi_j+\varphi_k) \\
&\quad +\sin((\gamma_i+\gamma_j-\gamma_k)\tau+\varphi_i+\varphi_j-\varphi_k)]\,\mathrm{d}\tau, \\
&= 0, \qquad (9.63)
\end{aligned}$$

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\int_0^T \mu_i(\tau)\mu_j^2(\tau)\mu_k(\tau)\,\mathrm{d}\tau, &= \lim_{T\to\infty}\frac{a_i a_j^2 a_k}{2T}\int_0^T \sin(\gamma_i\tau+\varphi_i) \\
&\quad \times(1-\cos(2\gamma_j\tau+2\varphi_j))\sin(\gamma_k\tau+\varphi_k)\,\mathrm{d}\tau, \\
&= \lim_{T\to\infty}\frac{a_i a_j^2 a_k}{4T}\int_0^T [\cos((\gamma_i-\gamma_k)\tau+\varphi_i-\varphi_k)
\end{aligned}$$

$$
\begin{aligned}
&-\cos((\gamma_i+\gamma_k)\tau+\varphi_i+\varphi_k)] \\
&\times(1-\cos(2\gamma_j\tau+2\varphi_j))\,\mathrm{d}\tau \\
=\lim_{T\to\infty}\frac{a_i a_j^2 a_k}{8T}\int_0^T &[2\cos((\gamma_i-\gamma_k)\tau+\varphi_i-\varphi_k) \\
&-\cos((\gamma_i-2\gamma_j-\gamma_k)\tau+\varphi_i-2\varphi_j-\varphi_k) \\
&-\cos((\gamma_i+2\gamma_j-\gamma_k)\tau+\varphi_i+2\varphi_j-\varphi_k) \\
&+\cos((\gamma_i-2\gamma_j+\gamma_k)\tau+\varphi_i-2\varphi_j+\varphi_k) \\
&+\cos((\gamma_i+2\gamma_j+\gamma_k)\tau+\varphi_i+2\varphi_j+\varphi_k) \\
&-2\cos((\gamma_i+\gamma_k)\tau+\varphi_i+\varphi_k)]\,\mathrm{d}\tau, \\
=0. & \qquad (9.64)
\end{aligned}
$$

The conditions $3\omega_i \neq \omega_j$, $\omega_i \neq 2\omega_j + \omega_k$, and $2\omega_i \neq \omega_j + \omega_k$, arise due to the payoff functions being non-quadratic and are not required for quadratic payoff functions.

References

1. Altman, E., Başar, T., Srikant, R.: Nash equilibria for combined flow control and routing in networks: asymptotic behavior for a large number of users. IEEE Trans. Autom. Control **47**, 917–930 (2002)
2. Apostol, T.M.: Mathematical Analysis, 2nd ed. Addison-Wesley, Reading (1974)
3. Ariyur, K.B., Kristic, M.: Real-Time Optimization by Extremum-Seeking Control. Wiley-Interscience, Hoboken (2003)
4. Başar, T.: Control and game-theoretic tools for communication networks (overview). Appl. Comput. Math. **6**, 104–125 (2007)
5. Başar, T., Olsder, G.J.: Dynamic Noncooperative Game Theory, 2nd ed. SIAM, Philadelphia (1999)
6. Bauso, D., Giarré, L., Pesenti, R.: Consensus in noncooperative dynamic games: a multiretailer inventory application. IEEE Trans. Autom. Control **53**, 998–1003 (2008)
7. Becker, R., King, R., Petz, R., Nitsche, W.: Adaptive closed-loop separation control on a high-lift configuration using extremum seeking, AIAA J. **45**, 1382–1392 (2007)
8. Carnevale, D., Astolfi, A., Centioli, C., Podda, S., Vitale, V., Zaccarian, L.: A new extremum seeking technique and its application to maximize RF heating on FTU. Fusing Eng. Design **84**, 554–558 (2009)
9. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press, New York (2006)
10. Cochran, J., Kanso, E., Kelly, S.D., Xiong, H., Krstic, M.: Source seeking for two nonholonomic models of fish locomotion. IEEE Trans. Robot. **25**, 1166–1176 (2009)
11. Cochran, J., Krstic, M.: Nonholonomic source seeking with tuning of angular velocity. IEEE Trans. Autom. Control **54**, 717–731 (2009)
12. Fink, A.M.: Almost Periodic Differential Equations, Lecture Notes in Mathematics, vol. 377. Springer, New York (1974)
13. Foster, D.P., Young, H.P.: Regret testing: learning to play Nash equilibrium without knowing you have an opponent. Theor. Econ. **1**, 341–367 (2006)

14. Fudenberg, D., Levine, D.K.: The Theory of Learning in Games. The MIT Press, Cambridge (1998)
15. Guay, M., Perrier, M., Dochain, D.: Adaptive extremum seeking control of nonisothermal continuous stirred reactors. Chem. Eng. Sci. **60**, 3671–3681 (2005)
16. Hale, J.K.: Ordinary Differential Equations. Wiley-Interscience, New York (1969)
17. Hart, S., Mansour, Y.: How long to equilibrium? The communication complexity of uncoupled equilibrium procedures. Games Econ. Behav. **69**, 107–126 (2010)
18. Hart, S., Mas-Colell, A.: Uncoupled dynamics do not lead to Nash equilibrium. Am. Econ. Rev. **95**, 1830–1836 (2003)
19. Hart, S., Mas-Colell, A.: Stochastic uncoupled dynamics and Nash equilibrium. Games Econ. Behav. **57**, 286–303 (2006)
20. Jafari, A., Greenwald, A., Gondek, D., Ercal, G.: On no-regret learning, fictitious play, and Nash equilibrium. In: Proceedings of the 18th International Conference on Machine Learning (2001)
21. Khalil, H.K.: Nonlinear Systems, 3rd ed. Prentice Hall, Upper Saddle River (2002)
22. Killingsworth, N.J., Aceves, S.M., Flowers, D.L., Espinosa-Loza, F., Krstic, M.: HCCI engine combustion-timing control: optimizing gains and fuel consumption via extremum seeking. IEEE Trans. Control Syst. Technol. **17**, 1350–1361 (2009)
23. Krstic, M., Frihauf, P., Krieger, J., Başar, T.: Nash equilibrium seeking with finitely- and infinitely-many players. In: Proceedings of the 8th IFAC Symposium on Nonlinear Control Systems, Bologna (2010)
24. Li, S., Başar, T.: Distributed algorithms for the computation of noncooperative equilibria. Automatica **23**, 523–533 (1987)
25. Luenberger, D.G.: Complete stability of noncooperative games. J. Optim. Theory Appl. **25**, 485–505 (1978)
26. Luo, L., Schuster, E.: Mixing enhancement in 2D magnetohydrodynamic channel flow by extremum seeking boundary control. In: Proceedings of the American Control Conference, St. Louis (2009)
27. MacKenzie, A.B., Wicker, S.B.: Game theory and the design of self-configuring, adaptive wireless networks. IEEE Commun. Mag. **39**, 126–131 (2001)
28. Marden, J.R., Arslan, G., Shamma, J.S.: Cooperative control and potential games. IEEE Trans. Syst. Man Cybern. B Cybern. **39**, 1393–1407 (2009)
29. Moase, W.H., Manzie, C., Brear, M.J.: Newton-like extremum-seeking part I: theory. In: Proceedings of the IEEE Conference on Decision and Control, Shanghai (2009)
30. Naimzada, A.K., Sbragia, L.: Oligopoly games with nonlinear demand and cost functions: two boundedly rational adjustment processes. Chaos Solitons Fract. **29**, 707–722 (2006)
31. Nešić, D., Tan, Y., Moase, W.H., Manzie, C.: A unifying approach to extremum seeking: adaptive schemes based on estimation of derivatives. In: Proceedings of the IEEE Conference on Decision and Control, Atlanta (2010)
32. Peterson, K., Stefanopoulou, A.: Extremum seeking control for soft landing of an electromechanical valve actuator. Automatica **29**, 1063–1069 (2004)
33. Rao, S.S., Venkayya, V.B., Khot, N.S.: Game theory approach for the integrated design of structures and controls. AIAA J. **26**, 463–469 (1988)
34. Rosen, J.B.: Existence and uniqueness of equilibrium points for concave N-person games. Econometrica **33**, 520–534 (1965)
35. Scutari, G., Palomar, D.P., Barbarossa, S.: The MIMO iterative waterfilling algorithm. IEEE Trans. Signal Process. **57**, 1917–1935 (2009)
36. Semsar-Kazerooni, E., Khorasani, K.: Multi-agent team cooperation: a game theory approach. Automatica **45**, 2205–2213 (2009)
37. Shamma, J.S., Arslan, G.: Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. IEEE Trans. Autom. Control **53**, 312–327 (2005)
38. Sharma, R., Gopal, M.: Synergizing reinforcement learning and game theory—a new direction for control. Appl. Soft Comput. **10**, 675–688 (2010)

39. Stanković, M.S., Johansson, K.H., Stipanović, D.M.: Distributed seeking of Nash equilibria in mobile sensor networks. In: Proceedings of the IEEE Conference on Decision and Control, Atlanta (2010)
40. Stanković, M.S., Stipanović, D.M.: Extremum seeking under stochastic noise and applications to mobile sensors. Automatica **46**, 1243–1251 (2010)
41. Tan, Y., Nešić, D., Mareels, I.: On non-local stability properties of extremum seeking control. Automatica **42**, 889–903 (2006)
42. Young, H.P.: Learning by trial and error. Games Econ. Behav. **65**, 626–643 (2009)
43. Zhang, C., Arnold, D., Ghods, N., Siranosian, A., Krstic, M.: Source seeking with nonholonomic unicycle without position measurement and with tuning of forward velocity. Syst. Control Lett. **56**, 245–252 (2007)
44. Zhu, M., Martínez, S.: Distributed coverage games for mobile visual sensors (I): Reaching the set of Nash equilibria. In: Proceedings of the IEEE Conference on Decision and Control, Shanghai, China (2009)

Chapter 10
A Uniform Tauberian Theorem in Optimal Control

Miquel Oliu-Barton and Guillaume Vigeral

Abstract In an optimal control framework, we consider the value $V_T(x)$ of the problem starting from state x with finite horizon T, as well as the value $W_\lambda(x)$ of the λ-discounted problem starting from x. We prove that uniform convergence (on the set of states) of the values $V_T(\cdot)$ as T tends to infinity is equivalent to uniform convergence of the values $W_\lambda(\cdot)$ as λ tends to 0, and that the limits are identical. An example is also provided to show that the result does not hold for pointwise convergence. This work is an extension, using similar techniques, of a related result by Lehrer and Sorin in a discrete-time framework.

Keywords Tauberian theorem • Optimal control • Asymptotic value • Game theory

10.1 Introduction

Finite horizon problems of optimal control have been studied intensively since the pioneer work of Stekhov, Pontryagin, Boltyanskii [27], Hestenes [18], Bellman [9] and Isaacs [19, 20] during the cold war—see for instance [7, 22, 23] for major references, or [14] for a short, clear introduction. A classical model considers the following controlled dynamic over $\mathbb{R}_+$

M. Oliu-Barton (✉)
Institut Mathématique de Jussieu, UFR 929, Université Paris 6, Paris, France
e-mail: miquel.oliu.barton@normalesup.org

G. Vigeral
CEREMADE, Université Paris-Dauphine, Paris, France
e-mail: vigeral@ceremade.dauphine.fr

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_10,

$$\begin{cases} y'(s) = f(y(s), u(s)) \\ y(0) = y_0 \end{cases} \tag{10.1}$$

where y is a function from $\mathbb{R}_+$ to $\mathbb{R}^n$, y_0 is a point in $\mathbb{R}^n$, u is the control function which belongs to $\mathcal{U}$, the set of Lebesgue-measurable functions from $\mathbb{R}_+$ to a metric space U and the function $f : \mathbb{R}^n \times U \to \mathbb{R}^n$ satisfies the usual conditions, that is: Lipschitz with respect to the state variable, continuous with respect to the control variable and bounded by a linear function of the state variable, for any control u.

Together with the dynamic, an objective function g is given, interpreted as the cost function which is to be minimized and assumed to be Borel-measurable from $\mathbb{R}^n \times U$ to $[0,1]$. For each finite horizon $t \in]0,+\infty[$, the average value of the optimal control problem with horizon t is defined as

$$V_t(y_0) = \inf_{u \in \mathcal{U}} \frac{1}{t} \int_0^t g(y(s,u,y_0), u(s))\, \mathrm{d}s. \tag{10.2}$$

It is quite natural to define, whenever the trajectories considered are infinite, for any discount factor $\lambda > 0$, the λ-discounted value of the optimal control problem, as

$$W_\lambda(y_0) = \inf_{u \in \mathcal{U}} \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda s} g(y(s,u,y_0), u(s))\, \mathrm{d}s. \tag{10.3}$$

In this framework the problem was initially to know whether, for a given finite horizon T and a given starting point y_0, a minimizing control u existed, solution of the optimal control problem (T, y_0). Systems with large, but fixed horizons were considered and, in particular, the class of "ergodic" systems (that is, those in which any starting point in the state space Ω is controllable to any point in Ω) has been thoroughly studied [2, 3, 5, 6, 8, 11, 25]. These systems are asymptotically independent of the starting point as the horizon goes to infinite. When the horizon is infinite, the literature on optimal control has mainly focussed on properties of given trajectories as the time tends to infinity. This approach corresponds to the uniform approach in a game theoretical framework and is often opposed to the asymptotic approach (described below), which we have considered in what follows, and which has received considerably less attention.

In a game-theoretical, discrete time framework, the same kind of problem was considered since [29], but with several differences in the approach: (1) the starting point may be chosen at random (a probability μ may be given on Ω, which randomly determines the point from which the controller will start the play); (2) the controllability-ergodicity condition is generally not assumed; (3) because of the inherent recursive structure of the process played in discrete time, the problem is generally considered for all initial states and time horizons.

For these reasons, what is called the "asymptotic approach"—the behavior of $V_t(\cdot)$ as the horizon t tends to infinity, or of $W_\lambda(\cdot)$ as the discount factor λ tends to zero—has been more studied in this discrete-time setup. Moreover, when it is considered in Optimal Control, in most cases [4, 10] an ergodic assumption is made which not only ensures the convergence of $V_t(y_0)$ to some V, but also forces the limit function V to be independent of the starting point y_0. The general asymptotic case, in which no ergodicity condition is assumed, has been to our knowledge studied for the first time recently. In [11, 28] the authors prove in different frameworks the convergence of $V_t(\cdot)$ and $W_\lambda(\cdot)$ to some non-constant function $V(y_0)$.

Some important, closely related questions are the following : does the convergence of $V_t(\cdot)$ imply the convergence of $W_\lambda(\cdot)$? Or vice versa? If they both converge, does the limit coincide? A partial answer to these questions goes back to the beginning of the twentieth century, when Hardy and Littlewood proved (see [17]) that for any sequence of bounded real numbers, the convergence of the Cesaro means is equivalent to the convergence of their Abel means, and that the limits are then the same :

Theorem 10.1 ([17]). *For any bounded sequence of reals $\{a_n\}_{n\geq 1}$, define $V_n = \frac{1}{n}\sum_{i=1}^{n} a_i$ and $W_\lambda = \lambda \sum_{i=1}^{+\infty}(1-\lambda)^{i-1}a_i$. Then,*

$$\liminf_{n\to+\infty} V_n \leq \liminf_{\lambda\to 0} W_\lambda \leq \limsup_{\lambda\to 0} W_\lambda \leq \limsup_{n\to+\infty} V_n.$$

Moreover, if the central inequality is an equality, then all inequalities are equalities.

Noticing that $\{a_n\}$ can be viewed as a sequence of costs for some deterministic (uncontrolled) dynamic in discrete-time, this results gives the equivalence between the convergence of V_t and the convergence of W_λ, to the same limit. In 1971, setting $V_t = \frac{1}{t}\int_0^t g(s)\,\mathrm{d}s$ and $W_\lambda = \lambda\int_0^{+\infty} \mathrm{e}^{-\lambda s} g(s)\,\mathrm{d}s$, for a given Lebesgue-measurable, bounded, real function g, Feller proved that the same result holds for continuous-time uncontrolled dynamics (particular case of Theorem 2, p. 445 in [15]).

Theorem 10.2 ([15]).

$$\liminf_{n\to+\infty} V_n \leq \liminf_{\lambda\to 0} W_\lambda \leq \limsup_{\lambda\to 0} W_\lambda \leq \limsup_{n\to+\infty} V_n.$$

Moreover, if the central inequality is an equality, then all inequalities are equalities.

In 1992, Lehrer and Sorin [24] considered a discrete-time controlled dynamic, defined by a correspondence $\Gamma : \Omega \rightrightarrows \Omega$, with nonempty values, and by g, a bounded real cost function defined on Ω. A feasible play at $z \in \Omega$ is an infinite sequence $\mathbf{y} = \{y_n\}_{n\geq 1}$ such that $y_1 = z$ and $y_{n+1} \in \Gamma(y_n)$. The average and discounted value functions are defined respectively by $V_n(z) = \inf \frac{1}{n}\sum_{i=1}^{n} g(y_i)$ and $W_\lambda(y_0) = \inf \lambda \sum_{i=1}^{+\infty}(1-\lambda)^{i-1} g(y_i)$, where the infima are taken over the feasible plays at z.

Theorem 10.3 ([24]).

$$\lim_{n\to+\infty} V_n(z) = V(z) \textit{ uniformly on } \Omega \iff \lim_{\lambda\to 0} W_\lambda(z) = V(z) \textit{ uniformly on } \Omega.$$

This result establishes the equivalence between uniform convergence of $W_\lambda(y_0)$ when λ tends to 0 and uniform convergence of $V_n(y_0)$ as n tends to infinity, in the general case where the limit may depend on the starting point y_0. The uniform condition is necessary: in the same article, the authors provide an example where only pointwise convergence holds and the limits differs.

In 1998, Arisawa (see [4]) considered a continuous-time controlled dynamic and proved the equivalence between the uniform convergence of W_λ and the uniform convergence of V_t in the specific case of limits independent of the starting point.

Theorem 10.4 ([4]). *Let $d \in \mathbb{R}$, then*

$$\lim_{t\to+\infty} V_t(z) = d, \textit{ uniformly on } \Omega \iff \lim_{\lambda\to 0+} W_\lambda(z) = d, \textit{ uniformly on } \Omega.$$

This does not settle the general case, in which the limit function may depend on the starting point.[1] For a continuous-time controlled dynamic in which $V_t(y_0)$ converges to some function $V(y_0)$, dependent on the state variable y_0, as t goes to infinity, we prove the following

Theorem 10.5. *$V_t(y_0)$ converges to $V(y_0)$ uniformly on Ω, if and only if $W_\lambda(y_0)$ converges to $V(y_0)$ uniformly on Ω.*

In fact, we will prove this result in a more general framework, as described in Sect. 10.2. Some basic lemmas which occur to be important tools will also be proven on that section. Section 10.3 will be devoted to the proof of our main result. Section 10.4 will conclude by pointing out, via an example, the fact that uniform convergence is a necessary requirement for the Theorem 10.5 to hold. A very simple dynamic is described, in which the pointwise limits of $V_t(\cdot)$ and $W_\lambda(\cdot)$ exist but differ. It should be noted that our proofs (as well as the counterexample in Sect. 10.4) are adaptations in this continuous-time framework of ideas employed in a discrete-time setting in [24]. In the appendix we also point out that an alternative proof of our theorem is obtained using the main theorem in [24] as well as a discrete/continuous equivalence argument.

For completeness, let us mention briefly this other approach, mentioned above as the uniform approach, and which has also been deeply studied, see for exemple [12, 13, 16]. In these models, the optimal average cost value is not taken over a finite period of time $[0,t]$, which is then studied for t growing to infinite, as in [4, 15, 17, 24, 28] or in our framework. On the contrary, only infinite trajectories are considered, among which the value $\overline{V_t}$ is defined as $\inf_{u\in\mathcal{U}} \sup_{\tau\geq t} \frac{1}{\tau}\int_0^\tau g(y(s,u,y_0),u(s))\,\mathrm{d}s$, or some other closely related variation. The asymptotic behavior, as t tends to infinity, of the function $\overline{V_t}$ has also been studied in that framework. In [16], both λ-discounted and average evaluations of an infinite trajectory are considered and their limits are compared. However, we stress

[1]Lemma 6 and Theorem 8 in [4] deal with this general setting, but we believe them to be incorrect since they are stated for pointwise convergence and, consequently, are contradicted by the example in Sect. 10.4.

out that the asymptotic behavior of those quantities is in general[2] not related to the asymptotic behavior of V_t and W_λ.

Finally, let us point out that in the framework of zero-sum differential games, that is when the dynamic is controlled by two players with opposite goals, a Tauberian theorem is given in the ergodic case by Theorem 2.1 in [1]. However, to our knowledge the general, non ergodic case is still an open problem.

10.2 Model

10.2.1 General Framework

We consider a deterministic dynamic programming problem in continuous time, defined by a measurable set of states Ω, a subset $\mathcal{T}$ of Borel-measurable functions from $\mathbb{R}_+$ to Ω, and a bounded Borel-measurable real-valued function g defined on Ω. Without loss of generality we assume $g : \Omega \to [0,1]$. For a given state x, define $\Gamma(x) := \{X \in \mathcal{T},\ X(0) = x\}$ the set of all feasible trajectories starting from x. We assume $\Gamma(x)$ to be non empty, for all $x \in \Omega$. Furthermore, the correspondence Γ is closed under concatenation: given a trajectory $X \in \Gamma(x)$ with $X(s) = y$, and a trajectory $Y \in \Gamma(y)$, the concatenation of X and Y at time s is

$$X \circ_s Y := \begin{cases} X(t) & \text{if } t \leq s \\ Y(t-s) & \text{if } t \geq s \end{cases} \tag{10.4}$$

and we assume that $X \circ_s Y \in \Gamma(x)$.

We are interested in the asymptotic behavior of the average and the discounted values. It is useful to denote the average payoff of a play (or trajectory) $X \in \Gamma(x)$ by:

$$\gamma_t(X) := \frac{1}{t}\int_0^t g(X(s))\,\mathrm{d}s \tag{10.5}$$

$$\nu_\lambda(X) := \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda s} g(X(s))\,\mathrm{d}s. \tag{10.6}$$

This is defined for $t, \lambda \in]0,+\infty[$. Naturally, we define the values as:

$$V_t(x) = \inf_{X \in \Gamma(x)} \gamma_t(X) \tag{10.7}$$

$$W_\lambda(x) = \inf_{X \in \Gamma(x)} \nu_\lambda(X). \tag{10.8}$$

---

[2] The reader may verify that this is indeed not the case in the example of Sect. 10.4.

Our main contribution is Theorem 10.5:

$$(A)\ W_\lambda \underset{\lambda\to 0}{\longrightarrow} V, \text{ uniformly on } \Omega \iff (B)\ V_t \underset{t\to\infty}{\longrightarrow} V, \text{ uniformly on } \Omega. \quad (10.9)$$

Notice that our model is a natural adaptation to the continuous-time framework of deterministic dynamic programming problems played in discrete time ; as it was pointed out during the introduction, this theorem is an extension to the continuous-time framework of the main result of [24], and our proof uses similar techniques.

This result can be applied to the model presented in Sect. 10.1: let $\widetilde{\Omega} = \mathbb{R}^d \times U$ and for any $(y_0, u_0) \in \widetilde{\Omega}$, define $\widetilde{\Gamma}(y_0, u_0) = \{(y(\cdot), u(\cdot))\ | u \in \mathcal{U}, u(0) = u_0$ and y is the solution of (10.1)$\}$. Then $\widetilde{\Omega}$, $\widetilde{\Gamma}$ and g satisfy the assumptions of this section. Defining $\widetilde{V}_t$ and $\widetilde{W}_\lambda$ as in (10.7) and (10.8) respectively, since the solution of (10.1) does not depend on $u(0)$ we get that

$$\widetilde{V}_t(y_0, u_0) = V_t(y_0)$$
$$\widetilde{W}_\lambda(y_0, u_0) = W_\lambda(y_0).$$

Theorem 10.5 applied to $\widetilde{V}$ and $\widetilde{W}$ thus implies that V_t converges uniformly to a function V in Ω if and only if W_λ converges uniformly to V in Ω.

10.2.2 Preliminary Results

We follow the ideas of [24], and start by proving two simple lemmas yet important tools, that will be used in the proof. The first establishes that the value increases along the trajectories. Then, we prove a convexity result linking the finite horizon average payoffs and the discounted evaluations on any given trajectory.

Lemma 10.1. *Monotonicity (compare with Proposition 1 in [24]). For all $X \in \mathcal{T}$, for all $s \geq 0$, we have*

$$\liminf_{t\to\infty} V_t(X(0)) \leq \liminf_{t\to\infty} V_t(X(s)) \quad (10.10)$$

$$\liminf_{\lambda\to 0} W_\lambda(X(0)) \leq \liminf_{\lambda\to 0} W_\lambda(X(s)). \quad (10.11)$$

Proof. Set $y := X(s)$ and $x := X(0)$. For $\varepsilon > 0$, take $T \in \mathbb{R}_+$ such that $\frac{s}{s+T} < \varepsilon$. Let $t > T$ and take an ε-optimal trajectory for V_t, i.e. $Y \in \Gamma(y)$ such that $\gamma_t(Y) \leq V_t(y) + \varepsilon$. Define the concatenation of X and Y at time s as in (10.4), where $X \circ_s Y$ is in $\in \Gamma(x)$ by assumption. Hence

$$\begin{aligned} V_{t+s}(x) \leq \gamma_{t+s}(X \circ_s Y) &= \frac{s}{t+s}\gamma_s(X) + \frac{t}{t+s}\gamma_t(Y) \\ &\leq \varepsilon + \gamma_t(Y) \\ &\leq 2\varepsilon + V_t(y). \end{aligned}$$

Since this is true for any $t \geq T$ the result follows.

Similarly, for the discounted case let $\lambda_0 > 0$ be such that

$$\lambda_0 \int_0^s \mathrm{e}^{-\lambda_0 r}\, \mathrm{d}r = 1 - \mathrm{e}^{\lambda_0 s} < \varepsilon.$$

Let $\lambda \in]0, \lambda_0]$ and take $Y \in \Gamma(y)$ an ε-optimal trajectory for $W_\lambda(y)$. Then:

$$\begin{aligned} W_\lambda(x) \leq \nu_\lambda(X \circ_s Y) &= \lambda \int_0^s \mathrm{e}^{-\lambda r} g(X(r))\, \mathrm{d}r + \lambda \int_s^{+\infty} \mathrm{e}^{-\lambda r} g(Y(r-s))\, \mathrm{d}r \\ &\leq \varepsilon + \mathrm{e}^{-\lambda s} \nu_\lambda(Y) \\ &\leq 2\varepsilon + W_\lambda(y). \end{aligned}$$

Again, this is true for any $\lambda \in]0, \lambda_0]$, and the result follows. □

Lemma 10.2. *Convexity (compare with Eq. (10.1) in [24]). For any play $X \in \mathcal{T}$, for any $\lambda > 0$:*

$$\nu_\lambda(X) = \int_0^{+\infty} \gamma_s(X) \mu_\lambda(s)\, \mathrm{d}s, \tag{10.12}$$

where $\mu_\lambda(s)\, \mathrm{d}s := \lambda^2 s \mathrm{e}^{-\lambda s} ds$ is a probability density on $[0, +\infty]$.

Proof. It is enough to notice that the following relation holds, by integration by parts:

$$\nu_\lambda(X) = \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda s} g(X(s))\, \mathrm{d}s = \lambda^2 \int_0^{+\infty} s \mathrm{e}^{-\lambda s} \left(\frac{1}{s} \int_0^s g(X(r))\, \mathrm{d}r \right) \mathrm{d}s,$$

and that $\int_0^{+\infty} \lambda^2 s \mathrm{e}^{-\lambda s} ds = 1$. □

The probability measure μ_λ plays an important role in the rest of the paper. Denoting

$$M(\alpha, \beta; \lambda) := \int_\alpha^\beta \mu_\lambda(s)\, \mathrm{d}s = \mathrm{e}^{-\lambda\alpha}(1 + \lambda\alpha) - \mathrm{e}^{-\lambda\beta}(1 + \lambda\beta),$$

we prove here two estimates that will be helpful in the next section.

Lemma 10.3. *The two following results hold (compare with Lemma 3 in [24]):*

(i) $\forall t > 0, \exists \varepsilon_0$ *such that* $\forall \varepsilon \leq \varepsilon_0, M\left((1-\varepsilon)t, t; \frac{1}{t}\right) \geq \frac{\varepsilon}{2\mathrm{e}}$.
(ii) $\forall \delta > 0, \exists \varepsilon_0$ *such that* $\forall \varepsilon \leq \varepsilon_0$, $\forall t > 0$, $M\left(\varepsilon t, (1-\varepsilon)t; \frac{1}{t\sqrt{\varepsilon}}\right) \geq 1 - \delta$.

Proof. Notice that in these particular cases, M does not depend on t:

(i) $M(t(1-\varepsilon), t; \frac{1}{t}) = (2-\varepsilon)\mathrm{e}^{-1+\varepsilon} - 2\mathrm{e}^{-1} = \frac{1}{\mathrm{e}}(\varepsilon + o(\varepsilon)) \geq \frac{\varepsilon}{2\mathrm{e}}$, for ε small enough.
(ii) $M(t\varepsilon, t(1-\varepsilon); \frac{1}{t\sqrt{\varepsilon}}) = (1+\sqrt{\varepsilon})\mathrm{e}^{-\sqrt{\varepsilon}} - (1 - 1/\sqrt{\varepsilon} + \sqrt{\varepsilon}) \exp\left(-1/\sqrt{\varepsilon} + \sqrt{\varepsilon}\right)$.
This expression tends to 1 as $\varepsilon \to 0$, hence the result. □

10.3 Proof of Theorem 10.5

10.3.1 From V_t to W_λ

Assume (B) : $V_t(\cdot)$ converges to some $V(\cdot)$ as t goes to infinity, uniformly on Ω. Our proof follows Proposition 4 and Lemmas 8 and 9 in [24].

Proposition 10.1. *For all $\varepsilon > 0$, there exists $\lambda_0 > 0$ such that $W_\lambda(x) \geq V(x) - \varepsilon$ for every $x \in \Omega$ and for all $\lambda \in]0, \lambda_0]$.*

Proof. Let T be such that $\|V_t - V\|_\infty \leq \varepsilon/2$ for every $t \geq T$. Choose $\lambda_0 > 0$ such that

$$\lambda^2 \int_T^{+\infty} s\mathrm{e}^{-\lambda s}\,\mathrm{d}s = 1 - (1+\lambda T)\mathrm{e}^{-\lambda T} \geq 1 - \frac{\varepsilon}{4}$$

for every $\lambda \in]0, \lambda_0]$. Fix $\lambda \in]0, \lambda_0]$ and take a play $Y \in \Gamma(x)$ which is $\varepsilon/4$-optimal for $W_\lambda(x)$. Since $\gamma_s(X) \geq 0$, the convexity formula (10.12) from Lemma 10.2 gives:

$$\begin{aligned}
W_\lambda(x) + \frac{\varepsilon}{4} \geq v_\lambda(Y) &\geq 0 + \lambda^2 \int_T^{+\infty} s\mathrm{e}^{-\lambda s}\gamma_s(Y)\,\mathrm{d}s \\
&\geq \lambda^2 \int_T^{+\infty} s\mathrm{e}^{-\lambda s} V_s(x)\,\mathrm{d}s \\
&\geq \left(1 - \frac{\varepsilon}{4}\right)\left(V(x) - \frac{\varepsilon}{2}\right) \\
&= V(x) - \frac{\varepsilon}{4}V(x) - \frac{\varepsilon}{2} + \frac{\varepsilon^2}{8} \\
&\geq V(x) - \frac{3\varepsilon}{4}.
\end{aligned}$$

□

Lemma 10.4. *$\forall \varepsilon > 0, \exists M$ such that for all $t \geq M, \forall x \in \Omega$, there is a play $X \in \Gamma(x)$ such that $\gamma_s(X) \leq V(x) + \varepsilon$ for all $s \in [\varepsilon t, (1-\varepsilon)t]$.*

Proof. By (B) there exists M such that $\|V_r - V\| \leq \varepsilon^2/3$ for all $r \geq \varepsilon M$. Given $t \geq M$ and $x \in \Omega$, let $X \in \Gamma(x)$ be a play (from x) such that $\gamma_t(X) \leq V_t(x) + \varepsilon^2/3$. For any $s \leq (1-\varepsilon)t$, we have that $t - s \geq \varepsilon t \geq \varepsilon M$ so Proposition 10.1 (Monotonicity) imply that

$$V_{t-s}(X(s)) \geq V(X(s)) - \frac{\varepsilon^2}{3} \geq V(x) - \frac{\varepsilon^2}{3}. \tag{10.13}$$

Since $V(x) + \varepsilon^2/3 \geq V_t(x)$, we also have:

$$\begin{aligned}
t\left(V(x) + \frac{2\varepsilon^2}{3}\right) &\geq t\left(V_t(x) + \frac{\varepsilon^2}{3}\right) \\
&\geq t\gamma_t(X) = \int_0^s g(X(r))\,\mathrm{d}r + \int_s^t g(X(r))\,\mathrm{d}r
\end{aligned}$$

$$\geq s\gamma_s(X) + (t-s)V_{t-s}(X(s))$$
$$\geq s\gamma_s(X) + (t-s)\left(V(x) - \frac{\varepsilon^2}{3}\right), \text{ by (10.13)}.$$

Isolating $\gamma_s(X)$ we get:

$$\gamma_s(X) \leq V(x) + \frac{t\varepsilon^2}{s}$$
$$\leq V(x) + \varepsilon, \quad \text{for } s/\varepsilon \geq t,$$

and we have proved the result for all $s \in [\varepsilon t, (1-\varepsilon)t]$. □

Proposition 10.2. *$\forall \delta > 0, \exists \lambda_0$ such that $\forall x \in \Omega$, for all $\lambda \in]0, \lambda_0]$, we have $W_\lambda(x) \leq V(x) + \delta$.*

Proof. By Lemma 10.3(ii), one can choose ε small enough such that

$$M\left(\varepsilon t, (1-\varepsilon)t; \frac{1}{t\sqrt{\varepsilon}}\right) \geq 1 - \frac{\delta}{2},$$

for any t. In particular, we can take $\varepsilon \leq \delta/2$. Using Lemma 10.4 with $\delta/2$, we get that for $t \geq t_0$ (and thus for $\lambda(t) := \frac{1}{t\sqrt{\varepsilon}} \leq \frac{1}{t_0\sqrt{\varepsilon}}$) and for any $x \in \Omega$, there exists a play $X \in \Gamma(x)$ such that

$$\nu_{\lambda(t)}(X) \leq \frac{\delta}{2} + \lambda(t)^2 \int_{\varepsilon t}^{(1-\varepsilon)t} s e^{s\lambda(t)} \gamma_s(X)\,\mathrm{d}s$$
$$\leq \frac{\delta}{2} + V(x) + \frac{\delta}{2}.$$
□

Propositions 10.1 and 10.2 establish the first part of Theorem 10.5: $(B) \Rightarrow (A)$.

10.3.2 From W_λ to V_t

Now assume (A) : $W_\lambda(\cdot)$ converges to some $W(\cdot)$ as λ goes to 0, uniformly on Ω. Our proof follows Proposition 2 and Lemmas 6 and 7 in [24]. Start by a technical Lemma:

Lemma 10.5. *Let $\varepsilon > 0$. For all $x \in \Omega$ and $t > 0$, and for any trajectory $Y \in \Gamma(x)$ which is $\varepsilon/2$-optimal for the problem with horizon t, there is a time $L \in [0, t(1 - \varepsilon/2)]$ such that, for all $T \in]0, t-L]$:*

$$\frac{1}{T}\int_L^{L+T} g(Y(s))\,\mathrm{d}s \leq V_t(x) + \varepsilon.$$

Proof. Fix $Y \in \Gamma(x)$ some $\varepsilon/2$-optimal play for $V_t(x)$. The function $s \mapsto \gamma_s(Y)$ is continuous on $]0,t]$ and satisfies $\gamma_t(Y) \leq V_t(x) + \varepsilon/2$. The bound on g implies that $\gamma_r(Y) \leq V_t(x) + \varepsilon$ for all $r \in [t(1-\varepsilon/2), t]$.

Consider now the set $\{s \in]0,t] \mid \gamma_s(Y) > V_t(x) + \varepsilon\}$. If this set is empty, then take $L = 0$ and observe that for any $r \in]0,t]$,

$$\frac{1}{r}\int_0^r g(Y(s))\,\mathrm{d}s \leq V_t(x) + \varepsilon.$$

Otherwise, let L be the superior bound of this set. Notice that $L < t(1-\varepsilon/2)$ and that by continuity $\gamma_L(Y) = V_t(x) + \varepsilon$. Now, for any $T \in [0, t-L]$,

$$\begin{aligned} V_t(x) + \varepsilon &\geq \gamma_{L+T}(Y) \\ &= \frac{L}{L+T}\gamma_L(Y) + \frac{T}{L+T}\left(\frac{1}{T}\int_L^{L+T} g(Y(s))\,\mathrm{d}s\right) \\ &= \frac{L}{L+T}(V_t(x)+\varepsilon) + \frac{T}{L+T}\left(\frac{1}{T}\int_L^{L+T} g(Y(s))\,\mathrm{d}s\right) \end{aligned}$$

and the result follows. □

Proposition 10.3. *$\forall \varepsilon > 0, \exists T$ such that for all $t \geq T$ we have $V_t(x) \geq W(x) - \varepsilon$, for all $x \in \Omega$.*

Proof. Let λ be such that $\|W_\lambda - W\| \leq \varepsilon/8$, and T such that

$$\lambda^2 \int_{T\varepsilon/4}^{+\infty} s\mathrm{e}^{-\lambda s}\,\mathrm{d}s < \frac{\varepsilon}{8}.$$

Proceed by contradiction and suppose that $\varepsilon > 0$ is such that for every T, there exists $t_0 \geq T$ and a state $x_0 \in \Omega$ such that $V_{t_0}(x_0) < W(x_0) - \varepsilon$.

Using Lemma 10.5 with $\varepsilon/2$, we get a play $Y \in \Gamma(x_0)$ and a time $L \in [0, t_0(1-\varepsilon/4)]$ such that, $\forall s \in [0, t_0 - L]$ (and, in particular, $\forall s \in [0, t_0\varepsilon/4]$),

$$\frac{1}{s}\int_L^{L+s} g(Y(r))\,\mathrm{d}r \leq V_{t_0}(x_0) + \frac{\varepsilon}{2} < W(x_0) - \frac{\varepsilon}{2}.$$

Thus,

$$\begin{aligned} W(Y(L)) - \frac{\varepsilon}{8} &\leq W_\lambda(Y(L)) \\ &\leq \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda s} g(Y(L+s))\,\mathrm{d}s \\ &\leq \lambda^2 \int_0^{t_0\varepsilon/4} s\mathrm{e}^{-\lambda s}\left(\frac{1}{s}\int_L^{L+s} g(Y(r))\,\mathrm{d}r\right)\mathrm{d}s + \frac{\varepsilon}{8} \end{aligned}$$

$$\leq W(x_0) - \frac{\varepsilon}{2} + \frac{\varepsilon}{8}$$
$$= W(x_0) - \frac{3\varepsilon}{8}.$$

This gives us $W(Y(L)) \leq W(x_0) - \varepsilon/4$, contradicting Proposition 10.1 (Monotonicity). □

Proposition 10.4. *$\forall \varepsilon > 0, \exists T$ such that for all $t \geq T$ we have $V_t(x) \leq W(x) + \varepsilon$, for all $x \in \Omega$.*

Proof. Otherwise, $\exists \varepsilon > 0$ such that $\forall T$, $\exists t \geq T$ and $x \in \Omega$ with $V_t(x) > W(x) + \varepsilon$. For any $X \in \Gamma(x)$ consider the (continuous in s) payoff function $\gamma_s(X) = \frac{1}{s}\int_0^s g(X(r))\,\mathrm{d}r$. Of course, $\gamma_t(X) \geq V_t(x) > W(x) + \varepsilon$. Furthermore, because of the bound on g,

$$\gamma_r(X) \geq W(x) + \varepsilon/2, \ \forall r \in [t(1-\varepsilon/2), t].$$

By Lemma 10.3, we can take ε small enough, so that for all t,

$$M\left(t(1-\varepsilon/2), t; \frac{1}{t} \geq \frac{\varepsilon}{4\mathrm{e}}\right)$$

holds. We set $\delta := \frac{\varepsilon}{4\mathrm{e}}$. By Proposition 10.3, there is a K such that $V_t \geq W(x) - \frac{\delta\varepsilon}{8}$, for all $t \geq K$. Fix K and consider

$$M(0, K; 1/t) = 1 - \mathrm{e}^{-K/t}(1 + K/t)$$

as a function of t. Clearly, it tends to 0 as t tends to infinity, so let t be such that this quantity is smaller than $\frac{\delta\varepsilon}{16}$. Also, let t be big enough so that $\|W_{1/t} - W\| < \frac{\delta\varepsilon}{5}$, which is a consequence of assumption (A).

We now set $\tilde{\lambda} := 1/t$ and consider the $\tilde{\lambda}$-payoff of some play $X \in \Gamma(x)$. We split $[0, +\infty]$ in three parts : $\mathcal{K} = [0, K], \mathcal{R} = [t(1-\varepsilon/2), t]$, and $(\mathcal{K} \cup \mathcal{R})^c$. The three parts are disjoint for t large enough, so by the Convexity formula (10.12), for any $\lambda > 0$,

$$v_{\tilde{\lambda}}(X) = \left(\int_{\mathcal{K}} \gamma_s(X)\mu_{\tilde{\lambda}}(\mathrm{d}s) + \int_{\mathcal{R}} \gamma_s(X)\mu_{\tilde{\lambda}}(\mathrm{d}s) + \int_{(\mathcal{K}\cup\mathcal{R})^c} \gamma_s(X)\mu_{\tilde{\lambda}}(\mathrm{d}s)\right)$$

where $\mu_\lambda(s)\,\mathrm{d}s = \lambda^2 s \mathrm{e}^{-\lambda s}\,\mathrm{d}s$. Recall that

$$\gamma_s(X)_{|\mathcal{K}} \geq 0$$
$$\gamma_s(X)_{|(\mathcal{K}\cup\mathcal{R})^c} \geq W(x) - \frac{\delta\varepsilon}{8}$$
$$\gamma_s(X)_{|\mathcal{R}} \geq W(x) + \frac{\varepsilon}{2}.$$

It is thus straightforward that

$$\begin{aligned} v_{\tilde{\lambda}}(X) &\geq 0+\delta\times\left(W(x)+\frac{\varepsilon}{2}\right)+\left(1-\delta-\frac{\delta\varepsilon}{16}\right)\times\left(W(x)-\frac{\delta\varepsilon}{8}\right) \\ &\geq W(x)+\delta\varepsilon\left(\frac{1}{2}-\frac{1}{16}-\frac{1}{8}-\frac{\delta}{8}+\frac{\delta\varepsilon}{64}\right) \\ &\geq W(x)+\frac{\delta\varepsilon}{8}. \end{aligned}$$

This is true for any play, so its infimum also satisfies $W_{\tilde{\lambda}}(x)\geq W(x)+\frac{\delta\varepsilon}{4}$, which is a contradiction, for we assumed that $W_{\tilde{\lambda}}(x)<W(x)+\frac{\delta\varepsilon}{5}$. □

Propositions 10.3 and 10.4 establish the second half of Theorem 10.5: $(A)\Rightarrow(B)$.

10.4 A Counter Example for Pointwise Convergence

In this section we give an example of an optimal control problem in which both $V_t(\cdot)$ and $W_\lambda(\cdot)$ converge pointwise on the state space, but to two different limits. As implied by Theorem 10.5, the convergence is not uniform on the state space.

Lehrer and Sorin were the first to construct such an example [24], in the discrete-time framework. We consider here one of its adaptations to continuous time, which was studied as Example 5 in [28],[3] where the notations are the same that in Sect. 10.1:

- The state space is $\Omega=\mathbb{R}_+^2$.
- The payoff function is given by $g(x,y)=0$ if $x\in[1,2]$, 1 otherwise.
- The set of control is $U=[0,1]$.
- The dynamic is given by $f(x,y,u)=(y,u)$ (thus Ω is forward invariant).

An interpretation is that the couple $(x(t),y(t))$ represents the position and the speed of some mobile moving along an axis, and whose acceleration $u(t)$ is controlled. Observe that since $U=[0,1]$, the speed $y(t)$ increases during any play. We claim that for any $(x_0,y_0)\in\mathbb{R}_+^2$, $V_t(x_0,y_0)$ (resp $W_\lambda(x_0,y_0)$) converges to $V(x_0,y_0)$ as t goes to infinity (respectively converges to $W(x_0,y_0)$ as λ tends to 0), where:

$$V(x_0,y_0)=\begin{cases} 1 & \text{if } y_0>0 \text{ or } x_0>2 \\ 0 & \text{if } y_0=0 \text{ and } 1\leq x_0\leq 2 \\ \frac{1-x_0}{2-x_0} & \text{if } y_0=0 \text{ and } x_0<1 \end{cases}$$

[3] We thank Marc Quincampoix for pointing out this example to us, which is simpler that our original one.

$$W(x_0,y_0) = \begin{cases} 1 & \text{if } y_0 > 0 \text{ or } x_0 > 2 \\ 0 & \text{if } y_0 = 0 \text{ and } 1 \leq x_0 \leq 2 \\ 1 - \frac{(1-x_0)^{1-x_0}}{(2-x_0)^{2-x_0}} & \text{if } y_0 = 0 \text{ and } x_0 < 1. \end{cases}$$

Here we only prove that $V(0,0) = \frac{1}{2}$ and $W(0,0) = \frac{3}{4}$; the proof for $y_0 = 0$ and $0 < x_0 < 1$ is similar and the other cases are easy.

First of all we prove that for any t or λ and any admissible trajectory (that is, any function $X(t) = (x(t), y(t))$ compatible with a control $u(t)$) starting from $(0,0)$, $\gamma_t(X) \geq \frac{1}{2}$ and $\nu_\lambda(X) \geq \frac{3}{4}$. This is clear if $x(t)$ is identically 0, so consider this is not the case. Since the speed $y(t)$ is increasing, we can define t_1 and t_2 as the times at which $x(t_1) = 1$ and $x(t_2) = 2$ respectively, and moreover we have $t_2 \leq 2t_1$. Then,

$$\begin{aligned} \gamma_t(X) &= \frac{1}{t}\left(\int_0^{\min(t,t_1)} \mathrm{d}s + \int_{\min(t,t_2)}^{t} ds\right) \\ &= 1 + \min\left(1, \frac{t_1}{t}\right) - \min\left(1, \frac{t_2}{t}\right) \\ &\geq 1 + \min\left(1, \frac{t_2}{2t}\right) - \min\left(1, \frac{t_2}{t}\right) \\ &\geq \frac{1}{2} \end{aligned}$$

and

$$\begin{aligned} \nu_\lambda(X) &= \int_0^{t_1} \lambda \mathrm{e}^{-\lambda s} \mathrm{d}s + \int_{t_2}^{+\infty} \lambda \mathrm{e}^{-\lambda s} \mathrm{d}s \\ &= 1 - \exp(-\lambda t_1) + \exp(-\lambda t_2) \\ &\geq 1 - \exp(-\lambda t_1) + \exp(-2\lambda t_1) \\ &\geq \min_{a>0}\{1 - a + a^2\} \\ &= \frac{3}{4}. \end{aligned}$$

On the other hand, one can prove [28] that $\limsup V_t(0,0) \leq 1/2$: in the problem with horizon t, consider the control "$u(s) = 1$ until $s = 2/t$ and then 0". Similarly one proves that $\limsup W_\lambda(0,0) \leq 3/4$: in the λ-discounted problem, consider the control "$u(s) = 1$ until $s = \lambda/\ln 2$ and then 0".

So the functions V_t and W_λ converge pointwise on Ω, but their limits V and W are different, since we have just shown $V(0,0) \neq W(0,0)$. One can verify that neither convergence is uniform on Ω by considering $V_t(1,\varepsilon)$ and $W_\lambda(1,\varepsilon)$ for small positive ε.

Remark 10.1. One may object that this example is not very regular since the payoff g is not continuous and the state space is not compact. However a related, smoother example can easily be constructed:

1. The set of controls is still $[0,1]$.
2. The continuous cost $g(x)$ is equal to 1 outside the segment [0.9,2.1], to 0 on [1,2], and linear on the two remainings intervals.
3. The compact state space is $\Omega = \{(x,y)|0 \leq y \leq \sqrt{2x} \leq 2\sqrt{2}\}$.
4. The dynamic is the same that in the original example for $x \in [0,3]$, and $f(x,y,u) = ((4-x)y,(4-x)u)$ for $3 \leq x \leq 4$. The inequality $y(t)y'(t) \leq x'(t)$ is thus satisfied on any trajectory, which implies that Ω is forward invariant under this dynamic.

With these changes the values $V_t(\cdot)$ and $W_\lambda(\cdot)$ still both converge pointwise on Ω to some $\widetilde{V}(\cdot)$ and $\widetilde{W}(\cdot)$ respectively, and $\widetilde{V}(0,0) \neq \widetilde{W}(0,0)$.

10.5 Possible Extensions

- We considered the finite horizon problem and the discounted one, but it should be possible to establish similar Tauberian theorems for other, more complex, evaluations of the payoff. This was settled in the discrete time case in [26].
- It would be very fruitful to establish necessary or sufficient conditions for uniform convergence to hold. In this direction we mention [28] in which sufficient conditions for the stronger notion of Uniform Value (meaning that there are controls that are nearly optimal no matter the horizon, provided it is large enough) are given in a general setting.
- In the discrete case an example is constructed in [26] in which there is no uniform value despite uniform convergence of the families V_t and W_λ. It would be of interest to construct such an example in continuous time, in particular in the framework of Sect. 10.1.
- It would be very interesting to study Tauberian theorems for dynamic systems that are controlled by two conflicting controllers. In the framework of differential games this has been done recently (Theorem 2.1 in [1]): an extension of Theorem 10.4 has been accomplished for two player games in which the limit of V_T or W_λ is assumed to be independent of the starting point. The similar result in the discrete time framework is a consequence of Theorems 1.1 and 3.5 in [21]. Existence of Tauberian theorems in the general setup of two-persons zero-sum games with no ergodicity condition remains open in both the discrete and the continuous settings.

Acknowledgements This article was done as part of the PhD of the first author. Both authors wish to express their many thanks to Sylvain Sorin for his numerous comments and his great help. We also thank Hélène Frankowska and Marc Quincampoix for helpful remarks on earlier drafts.

Appendix

We give here another proof[4] of Theorem 10.5 by using the analoguous result in discrete time [24] as well as an argument of equivalence between discrete and continuous dynamic.

Consider a deterministic dynamic programming problem in continuous time as defined in Sect. 10.2.1, with a state space Ω, a payoff g and a dynamic Γ. Recall that, for any $\omega \in \Omega$, $\Gamma(\omega)$ is the non empty set of feasible trajectories, starting from ω. We construct an associated deterministic dynamic programming problem in *discrete* time as follows.

Let $\widetilde{\Omega} = \Omega \times [0,1]$ be the new state space and let $\widetilde{g}$ be the new cost function, given by $\widetilde{g}(\omega,x) = x$. We define a multivalued-function with nonempty values $\widetilde{\Gamma} : \widetilde{\Omega} \rightrightarrows \widetilde{\Omega}$ by

$$(\omega,x) \in \widetilde{\Gamma}(\omega',x') \Longleftrightarrow \exists X \in \Gamma(\omega'), \quad \text{with } X(1) = \omega \quad \text{and} \quad \int_0^1 g(X(t))\,\mathrm{d}t = x.$$

Following [24], we define, for any initial state $\widetilde{\omega} = (\omega,x)$

$$v_n(\widetilde{\omega}) = \inf \frac{1}{n}\sum_{i=1}^{n} \widetilde{g}(\widetilde{\omega}_i)$$

$$w_\lambda(\widetilde{\omega}) = \inf \lambda \sum_{i=1}^{+\infty} (1-\lambda)^{i-1} \widetilde{g}(\widetilde{\omega}_i)$$

where the infima are taken over the set of sequences $\{\widetilde{\omega}_i\}_{i\in\mathbb{N}}$ such that $\widetilde{\omega}_0 = \widetilde{\omega}$ and $\widetilde{\omega}_{i+1} \in \widetilde{\Gamma}(\widetilde{\omega}_i)$ for every $i \geq 0$.

Theorem 10.5 is then the consequence of the following three facts.

Firstly, the main theorem of Lehrer and Sorin in [24], which states that uniform convergence (on $\widetilde{\Omega}$) of v_n to some v is equivalent to uniform convergence of w_λ to the same v.

Secondly, the concatenation hypothesis (10.4) on Γ implies that for any $(\omega,x) \in \widetilde{\Omega}$

$$v_n(\omega,x) = V_n(\omega)$$

where $V_t(\omega) = \inf_{X\in\Gamma(\omega)} \frac{1}{t}\int_0^n g(X(s))\,\mathrm{d}s$, as defined in equation (10.7). Consequently, because of the bound on g, for any $t \in \mathbb{R}_+$ we have

$$|V_t(\omega) - v_{\lfloor t \rfloor}(\omega,x)| \leq \frac{2}{\lfloor t \rfloor}$$

where $\lfloor t \rfloor$ stands for the integer part of t.

[4] We thank Frédéric Bonnans for the idea of this proof.

Finally, again because of hypothesis (10.4), for any $\lambda \in]0,1]$,

$$w_\lambda(\omega,x) = \inf_{X\in\Gamma(\omega)} \lambda \int_0^{+\infty} (1-\lambda)^{\lfloor t\rfloor} g(X(t))\,\mathrm{d}t.$$

Hence, by equation (10.8) and the bound on the cost function, for any $\lambda \in]0,1]$,

$$|W_\lambda(\omega) - w_\lambda(\omega,x)| \le \lambda \int_0^{+\infty} \left|(1-\lambda)^{\lfloor t\rfloor} - \mathrm{e}^{-\lambda t}\right| \mathrm{d}t$$

which tends uniformly (with respect to x and ω) to 0 as λ goes to 0 by virtue of the following lemma.

Lemma 10.6. *The function*

$$\lambda \mapsto \lambda \int_0^{+\infty} \left|(1-\lambda)^{\lfloor t\rfloor} - \mathrm{e}^{-\lambda t}\right| \mathrm{d}t$$

converges to 0 as λ tends to 0.

Proof. Since $\lambda \int_0^{+\infty}(1-\lambda)^{\lfloor t\rfloor} = \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda t}\mathrm{d}t = 1$, for any $\lambda > 0$, the lemma is equivalent to the convergence to 0 of

$$E(\lambda) := \lambda \int_0^{+\infty} \left[(1-\lambda)^{\lfloor t\rfloor} - \mathrm{e}^{-\lambda t}\right]_+ \mathrm{d}t$$

where $[x]_+$ denotes the positive part of x. Now, from the relation $1-\lambda \le \mathrm{e}^{-\lambda}$, true for any λ, one can easily deduce that, for any $\lambda > 0$, $t \ge 0$, the relation $(1-\lambda)^{\lfloor t\rfloor}\mathrm{e}^{\lambda t} \le \mathrm{e}^{\lambda}$ holds. Hence,

$$\begin{aligned} E(\lambda) &= \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda t}\left[(1-\lambda)^{\lfloor t\rfloor}\mathrm{e}^{\lambda t} - 1\right]_+ \mathrm{d}t \\ &\le \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda t}(\mathrm{e}^{\lambda} - 1)\,\mathrm{d}t \\ &= \mathrm{e}^{\lambda} - 1 \end{aligned}$$

which converges to 0 as λ tends to 0. □

References

1. Alvarez, O., Bardi, M.: Ergodic Problems in Differential Games. Advances in Dynamic Game Theory, pp. 131–152. Ann. Int'l. Soc. Dynam. Games, vol. 9, Birkhäuser Boston (2007)
2. Alvarez, O., Bardi, M.: Ergodicity, stabilization, and singular perturbations for Bellman-Isaacs equations. Mem. Am. Math. Soc. **960**(204), 1–90 (2010)
3. Arisawa, M.: Ergodic problem for the Hamilton-Jacobi-Bellman equation I. Ann. Inst. Henri Poincare **14**, 415–438 (1997)

4. Arisawa, M.: Ergodic problem for the Hamilton-Jacobi-Bellman equation II. Ann. Inst. Henri Poincare **15**, 1–24 (1998)
5. Arisawa, M., Lions, P.-L.: On ergodic stochastic control. Comm. Partial Diff. Eq. **23**(11–12), 2187–2217 (1998)
6. Artstein, Z., Gaitsgory, V.: The value function of singularly perturbed control systems. Appl. Math. Optim. **41**(3), 425–445 (2000)
7. Bardi, M., Capuzzo-Dolcetta, I.: Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA (1997)
8. Barles, G.: Some homogenization results for non-coercive Hamilton-Jacobi equations. Calculus Variat. Partial Diff. Eq. **30**(4), 449–466 (2007)
9. Bellman, R.: On the theory of dynamic programming. Proc. Natl. Acad. Sci. U.S.A, **38**, 716–719 (1952)
10. Bettiol, P.: On ergodic problem for Hamilton-Jacobi-Isaacs equations. ESAIM: COCV **11**, 522–541 (2005)
11. Cardaliaguet, P.: Ergodicity of Hamilton-Jacobi equations with a non coercive non convex Hamiltonian in $\mathbb{R}^2/\mathbb{Z}^2$. Ann. l'Inst. Henri Poincare (C) Non Linear Anal. **27**, 837–856 (2010)
12. Carlson, D.A., Haurie, A.B., Leizarowitz, A.: Optimal Control on Infinite Time Horizon. Springer, Berlin (1991)
13. Colonius, F., Kliemann, W.: Infinite time optimal control and periodicity. Appl. Math. Optim. **20**, 113–130 (1989)
14. Evans, L.C.: An Introduction to Mathematical Optimal Control Theory. Unpublished Lecture Notes, U.C. Berkeley (1983). Available at http:/math.berkeley.edu/~evans/control.course.pdf
15. Feller, W.: An Introduction to Probability Theory and its Applications, vol. II, 2nd ed. Wiley, New York (1971)
16. Grune, L.: On the Relation between Discounted and Average Optimal Value Functions. J. Diff. Eq. **148**, 65–99 (1998)
17. Hardy, G.H., Littlewood, J.E.: Tauberian theorems concerning power series and Dirichlet's series whose coefficients are positive. Proc. London Math. Soc. **13**, 174–191 (1914)
18. Hestenes, M.: A General Problem in the Calculus of Variations with Applications to the Paths of Least Time, vol. 100. RAND Corporation, Research Memorandum, Santa Monica, CA (1950)
19. Isaacs, R.: Games of Pursuit. Paper P-257. RAND Corporation, Santa Monica (1951)
20. Isaacs, R.: Differential Games. A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization. Wiley, New York (1965)
21. Kohlberg, E., Neyman, A.: Asymptotic behavior of nonexpansive mappings in normed linear spaces. Isr. J. Math. **38**, 269–275 (1981)
22. Kirk, D.E.: Optimal Control Theory: An Introduction. Englewood Cliffs, N.J. Prentice Hall (1970)
23. Lee, E.B., Markus, L.: Foundations of Optimal Control Theory. SIAM, Philadelphia (1967)
24. Lehrer, E., Sorin, S.: A uniform Tauberian theorem in dynamic programming. Math. Oper. Res. **17**, 303–307 (1992)
25. Lions, P.-L., Papanicolaou, G., Varadhan, S.R.S.: Homogenization of Hamilton-Jacobi Equations. Unpublished (1986)
26. Monderer, M., Sorin, S.: Asymptotic Properties in Dynamic Programming. Int. J. Game Theory **22**, 1–11 (1993)
27. Pontryiagin, L.S., Boltyanskii, V.G., Gamkrelidge: The Mathematical Theory of Optimal Processes. Nauka, Moskow (1962) (Engl. Trans. Wiley)
28. Quincampoix, M., Renault, J.: On the existence of a limit value in some non expansive optimal control problems. SIAM J. Control Optim. **49**, 2118–2132 (2011)
29. Shapley, L.S.: Stochastic games. Proc. Natl. Acad. Sci. **39**, 1095–1100 (1953)

Chapter 11
E-Equilibria for Multicriteria Games

Lucia Pusillo and Stef Tijs

Abstract In this paper we propose an equilibrium definition for non-cooperative multicriteria games based on improvement sets. Our new definition generalizes the idea of exact and approximate equilibria. We obtain existence theorems for some multicriteria games.

Keywords Multicriteria games • Pareto equilibria • Approximate solutions • Improvement sets

11.1 Introduction

We know that Game Theory studies conflicts, behaviour and decisions of more than one rational agent. Players choose their actions in order to achieve preferred outcomes. Often players have to "optimize" not one but more than one objective and these are often not comparable, so multicriteria games help us to make decisions in multi-objective problems. The first observation to make in studying these topics is that in general there is not an optimal solution from all points of view.

Let us consider, for example, an interactive decision between a seller and a buyer. The latter wishes to buy a car and he chooses among many models. He has to take into account the price, the power, the petrol consumption and the dimension of the car: it must be large enough to be comfortable for his family but not too large to

L. Pusillo (✉)
Dima – Department of Mathematics, University of Genoa, Via Dodecaneso 35, 16146 Genoa, Italy
e-mail: pusillo@dima.unige.it

S. Tijs
CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: s.h.tijs@uvt.nl

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_11,

have parking problems. The seller wishes to maximize his gain and to sell a good and useful car in order to satisfy the buyer so, in future, he will come back or he will recommend this car-dealer to his friends and thus the seller will have more buyers. We can consider this interaction as a game where the two players have to "optimize" many criteria. Starting from Vector Optimization (see [3–5]) the theory of multi-objective games helps us in these multi-objective situations.

Shapley, in 1959, gave a generalization of the classical definition of Nash equilibrium (the most widely accepted solution for non cooperative scalar games), to Pareto equilibrium (weak and strong), for non cooperative games with many criteria. Intuitively a feasible point in $\mathbb{R}^n$ is a *Pareto equilibrium* if there is no other feasible point which is larger in every coordinate (we will be more precise in the subsequent pages).

Let us consider the following game in matrix form, where player I has one criterion and player II has two criteria:

	C	D
A	$(2)\ (5,0)$	$(0)\ (0,1)$
B	$(0)\ (-1,0)$	$(0)\ (0,1)$

In pure strategies the weak Pareto equilibria of the game are:

$(A,C),(A,D),(B,D)$.

To know more about Pareto equilibria see also [1, 19].

In this paper we consider a new concept of equilibria for games: it is based on improvement sets and at the same time, presents the two aspects of approximate and exact equilibria. The idea of approximate equilibrium is very important, in fact some games have no equilibria but they have approximate ones. Let us consider the following example:

$$\begin{pmatrix} (1,2) & (1,2+\frac{1}{2}) & (1,2+\frac{3}{4}) & (1,2+\frac{n-1}{n}) & \cdots \\ (0,0) & (0,0) & (0,0) & (0,0) & \cdots \end{pmatrix}$$

In this game there are no Nash equilibria (player II does not reach the payoff 3) but there is an infinite number of ϵNE. Intuitively, approximate equilibria mean that deviations improve by at most ϵ.

There are also games without approximate equilibria, for example the game:

$$\begin{pmatrix} (1,2) & (1,3) & (1,4) & (1,5) & \cdots \\ (0,0) & (0,1) & (0,2) & (0,3) & \cdots \end{pmatrix}$$

This game has neither NE nor ϵNE.

Many auction situations lead to games without NE but with approximate NE (see [6]); in [18], the author proved some theorems about approximate solutions.

In a previous paper [13] the authors studied the extensions of approximate *NE* to approximate Pareto equilibria for a particular class of games: multicriteria potential games. Furthermore they extended the theory introduced in [10] for exact potential games with one objective, to include muti-objective games. For an approach to potential games see also [16]. In recent years much attention has been dedicated to multicriteria games to study applications for the real world. Neither exact nor approximate equilibrium has a natural generalization from the scalar to the vector case. However, some useful ideas can be found in [1, 15, 17] and references in it. The concept of approximate solution is also meaningful for the notion of Tikhonov well-posedness as made in [8, 9, 11].

The notion of equilibrium studied in this paper is inspired by a previous paper [2] dedicated to optimal points in vector optimization. This definition is based on improvement sets $E \subset \mathbb{R}^n$ with two properties:

(a) $0 \notin E$ (exclusion property).
(b) E is upper comprehensive i.e. $x \in E,\ y \geqq x$ implies $y \in E$ (comprehensive property).

We consider a multicriteria game G with n players and m objectives and we say that a strategy profile is an E-equilibrium for the game if it is an optimal point with respect to the improvement set E and the utility functions $u_i(.,\widehat{x_{-i}})$. We write $(\widehat{x_1},\dots,\widehat{x_n}) \in O^E(G)$ when it is clear from the context that G is a game.

The paper is structured as follows: in Sect. 11.2 we give some preliminary results; in Sect. 11.3 we give the notion of E-equilibrium for potential games and we prove some existence results; in Sect. 11.4 we generalize this notion to multicriteria games; in Sect. 11.5 we present the conclusions and some ideas for further researches.

11.2 Definitions and Preliminary Results

Given $a, b \in \mathbb{R}^n$ we consider the following inequalities on $\mathbb{R}^n$:

$$a \geqq b \Leftrightarrow a_i \geq b_i \forall i = 1,\dots,n;$$

$$a \geq b \Leftrightarrow a \geqq b \text{ and } a \neq b;$$

$$a > b \Leftrightarrow a_i > b_i \forall i = 1,\dots,n.$$

Obviously we mean:

$$a \leqq b \Longleftrightarrow -a \geqq -b;\ a \leq b \Longleftrightarrow -a \geq -b;\ a < b \Longleftrightarrow -a > -b.$$

We write $\langle a,b\rangle$ to mean the scalar product of two vectors $a, b \in \mathbb{R}^n$.

We say that $A \subset \mathbb{R}^n$ is upper bounded (u.b. for short) if there exists $b \in \mathbb{R}^n$ such that $a \leqq b\ \forall a \in A$.

By $\mathbb{R}^n_+$ we mean the points in $\mathbb{R}^n$ with all coordinates positive or null, by $\mathbb{R}^n_{++}$ we mean the points in $\mathbb{R}^n$ with all coordinates strictly positive.

Given a game $G = \langle (X_i)_{i\in N}, (u_i)_{i\in N}\rangle$, a strategy profile $x \in X = \prod_{k\in N} X_k$, let $(\widehat{x_i}, x_{-i})$ denote the profile $(x_1,\dots,x_{i-1},\widehat{x_i},x_{i+1},\dots,x_n)$.

We recall the definitions of weak and strong Pareto equilibrium:

Definition 11.1. Let $G = \langle (X_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a multicriteria game with n players, X_i is the strategy space for player $i \in N$, $X = \prod_{k\in N} X_k$, $u_i : X \to \mathbb{R}^{m_i}$ is the utility function for player i who have m_i criteria "to optimize". A strategy profile $x \in X$ is a

1. Weak Pareto equilibrium if

$$\nexists\, y \in X : \; u_i(y_i, x_{-i}) > u_i(x), \forall i \in N$$

2. Strong Pareto equilibrium if

$$\nexists\, y \in X : \; u_i(y_i, x_{-i}) \geq u_i(x), \forall i \in N$$

Intuitively a feasible point in $\mathbb{R}^n$ is a *weak Pareto equilibrium* if there is no other feasible point which is larger in each coordinate.

Similarly, a feasible point in $\mathbb{R}^n$ is *strong Pareto equilibrium* if there is no other feasible point which is larger in at least one coordinate and not smaller in all other coordinates. The set of weak and strong Pareto equilibria for the game G will be denoted by $\mathrm{wPE}(G)$ and $\mathrm{sPE}(G)$ respectively.

Definition 11.2. Let $E \subset \mathbb{R}^n$. We say that E is upper comprehensive if $x \in E$ and if $y \geqq x$ then also $y \in E$.

Let $E \subset \mathbb{R}^n \setminus \{0\}$, and E an upper comprehensive set. We shall call E an *improvement set*.

In Fig. 11.1 we see an improvement set, in Fig. 11.2 we see a non-improvement set.

We can define the optimal points of $A \subset \mathbb{R}^n$ w.r.t. the improvement set E (or the E-optimal points of A) and we shall write $O^E(A)$ in the following way:

$$a \in O^E(A) \text{ if and only if } a \in A \text{ and } (a+E)\bigcap A = \emptyset.$$

Obviously $O^{\emptyset}(A) = A \;\forall A \subset \mathbb{R}^n$, thus in the following we will consider $E \neq \emptyset$.

The definition is a generalization of the notions used in scalar optimization, and also of Pareto optimal points (for further details, see [2]).

Definition 11.3. Let $w \in \Delta = \left\{p \in \mathbb{R}^n_+ : \sum_{i=1}^n p_i = 1\right\}$ and $A \subset \mathbb{R}^n$. The set A is called *w-upper-bounded* if there is $k \in \mathbb{R}$ such that $\langle w, a\rangle \leq k$ for all $a \in A$.

Intuitively A is contained in a hyper-half space with w as normal of the "boundary" hyperplane.

It is obvious that the property of "upper boundedness" implies the "w-upper boundedness" property for all $w \in \Delta$

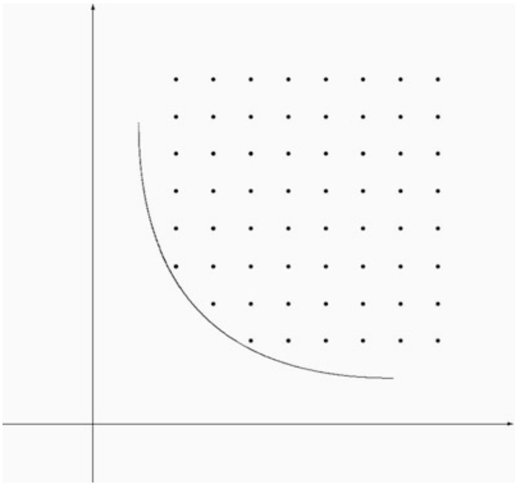


Fig. 11.1 Example of an improvement set E in $\mathbb{R}^2$. The set E is indicated by *dots*

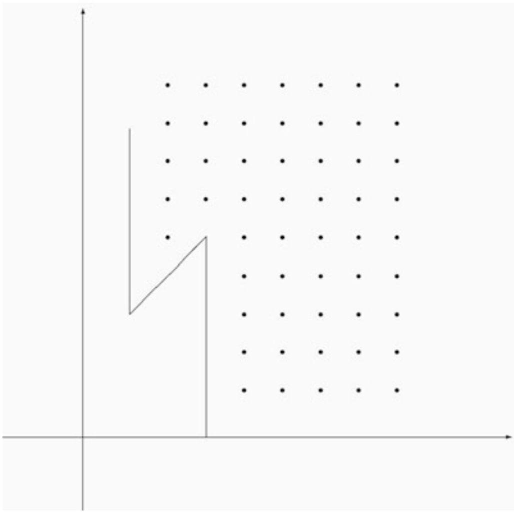


Fig. 11.2 Example of a non improvement set (indicated by *dots*)

Definition 11.4. Let $w \in \Delta$ and let E an improvement set. Then w is called a *separator* for E iff there exists a positive number t such that $\langle w, e \rangle > t \quad$ for each $e \in E$.

"A upper bounded subset of $\mathbb{R}^n$" is not a sufficient condition to have $O^E(A) \neq \emptyset$ as the following example proves:

Example 11.1. Let

$$A = \{(x,y) \in \mathbb{R}^2 : y < 1/x,\ x < 0\} \text{ and } E = E_+ \text{ then } O^E(A) = \emptyset.$$

This example illustrates also the following:

Proposition 11.1. *Let $A \subset \mathbb{R}^n$, A open set, E an improvement set such that $dist(E,0) = 0$, then $O^E(A) = \emptyset$.*

With $\text{dist}(E,0)$ we mean the usual distance in $\mathbb{R}^n$ between a set and a point.

Now we recall the definition of approximate NE for scalar games:

Definition 11.5. Given a scalar game $G = \langle (X_i)_{i\in N}, (u_i)_{i\in N} \rangle$, and $\epsilon \geq 0$, we remind that an approximate Nash equilibrium, for short ϵNE, is a strategy profile $\overline{x} \in X$ such that:

$$u_i(\overline{x}) \geq u_i(x_i, \overline{x}_{-i}) - \epsilon, \ \forall x_i \in X_i, \ \forall i \in N.$$

Intuitively for an ϵNE, deviations do not pay more than ϵ.

11.3 From Vector Optimization to Multicriteria Potential Games

Let us consider non-cooperative multicriteria potential games, that is games where the payoff functions have their values in $\mathbb{R}^m$ and there is a potential function P which relates them. Note that in a potential game the number of objectives is the same for all the players, so in this section $m_1 = m_2 = \ldots = m_n = m$ are assumed.

Mathematically a strategic multicriteria potential game is a tuple $G = \langle (X_i)_{i\in N}, \ (u_i)_{i\in N} \rangle$, where $N = 1,\ldots,n$ is the set of players, X_i is the strategy space for player $i \in N$, X is the cartesian product $\prod_{k\in N} X_k$ of the strategy spaces $(X_i)_{i\in N}$ and $u_i : X \to \mathbb{R}^m$ is the utility function for player i. We call G a *potential game* if there exists a map $P : X \to \mathbb{R}^m$ such that for all $i \in N, \ a_i, \ b_i \in X_i$, $a_{-i} \in X_{-i} := \prod_{j\in N\setminus\{i\}} X_j$, it turns out:

$$u_i(a_i, a_{-i}) - u_i(b_i, a_{-i}) = P(a_i, a_{-i}) - P(b_i, a_{-i}).$$

Essentially a multicriteria potential game can be seen as a multicriteria-optimization problem with P as objective function ($P : X \to \mathbb{R}^m$). To illustrate better the concept of multicriteria potential game, look at the following example:

Example 11.2. Consider the potential game with two players and three objectives as in the table below.

$(2,-2,0)$ $(1,0,0)$	$(0,0,0)$ $(0,2,1)$
$(1,3,1)$ $(0,3,1,)$	$(0,0,0)$ $(0,0,1)$

A potential function is

$P:$	1, 0, 0	0, 2, 1
	0, 5, 1	0, 2, 1

Given the game $G = \langle (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$, we mean $G^P = \langle (X_i)_{i \in N}, P) \rangle$, that is the game where the utility function is P for all players.

For a finite multicriteria potential game G the set of weak Pareto equilibria is not empty, $wPE(G) \neq \emptyset$ (see [13]).

Let us give the definition of optimal points of a function with respect to an improvement set.

Definition 11.6. Given a function $P : \prod X_i \to \mathbb{R}^m$, $X = X_i \times X_{-i}$, we say that $a \in O^E(P)$ that is a is an optimal point for the function P with respect to the improvement set E if $(a+E) \cap P(X) = \emptyset$.

Definition 11.7. A strategy profile $\widehat{x} \in X$ is an E-equilibrium for the multicriteria game G, where $E = (E_1, \ldots, E_i, \ldots, E_n)$, E_i is the improvement set for player i, if for each player i and for each $x_i \in X_i$ it turns out $u_i(x_i, \widehat{x_{-i}}) \notin u_i(\widehat{x}) + E_i$.

We write $\widehat{x} \in O^E(G)$.

Theorem 11.1. *Let $G = \langle (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$ be a multicriteria potential game and suppose that the potential functions are upper bounded. Let us suppose that there is a hyperplane which separates E from* 0.

Then $O^E(P(X)) \neq \emptyset$

Proof. The proof follows by considering the known results about separation theorems (see for example [3]) □

Remark that the condition "P upper bounded" is not equivalent to u_i upper bounded as the following example shows.

Example 11.3. Let us consider $G = (\mathbb{R}, \mathbb{R}, u_1, u_2)$ where $u_1(x,y) = \min\{x,y\} - y = u_2(y,x))$. So $u_1(x,y) \leq 0$ and $u_2(x,y) \leq 0$, but the potential function $P(x,y) = \min\{x,y\}$ is not upper bounded on $\mathbb{R}$.

For multicriteria potential games we have the following existence theorem:

Theorem 11.2. *Let $G = \langle (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$ be a multicriteria potential game and suppose that the potential function P is w-upper bounded. Furthermore w strongly separates E from* $\{0\}$. *Then $O^E(G) \neq \emptyset$.*

Proof. To make the notations easier, we write the proof only in the case of two players and two objectives, but the proof is analogous for n players with more then two objectives.

Let $\widehat{y} = P(\widehat{x})$ with $\langle w, \widehat{y} \rangle \geq \sup\{\langle w, a \rangle; a \in P(X)\} - t/2$.

I want to prove that $\widehat{x} = (\widehat{x_1}, \widehat{x_2}) \in O^E(G)$.

Let us suppose by contradiction that it is not true. Then

(i) $\exists x_1 \in X_1$ s.t. $u_1(x_1, \widehat{x_2}) \in u_1(\widehat{x_1}, \widehat{x_2}) + E_1$,
and/or
(ii) $\exists x_2 \in X_2$ s.t. $u_2(\widehat{x_1}, x_2) \in u_2(\widehat{x_1}, \widehat{x_2}) + E_2$.
Being G a game with exact potential P we have equivalently to (i) and (ii) respectively (iii) and (iv).
(iii) $\exists x_1 \in X_1$ s.t. $P(x_1, \widehat{x_2}) - P(\widehat{x_1}, \widehat{x_2})) \in E_1$
and/or
(iv) $\exists x_2 \in X_2$ s.t. $P(\widehat{x_1}, x_2) - P(\widehat{x_1}, \widehat{x_2}) \in E_2$.
Furthermore $\langle w, (P(x_1, \widehat{x_2}) - P(\widehat{x_1}, \widehat{x_2}))\rangle \geq t > 0$
that is:
$\sum_{k=1}^{2} w_k (P_k(x_1, \widehat{x_2}) - P_k(\widehat{x_1}, \widehat{x_2})) \geq t$ that is
$w_1 P_1(x_1, \widehat{x_2}) + w_2 P_2(x_1, \widehat{x_2}) \geq w_1 (P_1(\widehat{x_1}, \widehat{x_2}) + w_2 P_2(\widehat{x_1}, \widehat{x_2}) + t$
So
$\langle w, P(\widehat{x})\rangle + t \leq \sup\{\langle w, a\rangle . a \in P(X)\} \leq \langle w, P(\widehat{x})\rangle + t/2$ which is a contradiction; so $O^E(G) = \emptyset$. □

As it is already known in the scalar case, we have the following theorem:

Theorem 11.3. *Let $G = \langle (X_i)_{i \in N}, (u_i)_{i \in N}\rangle$ be a multicriteria potential game. Then $O^E(G) = O^E(G^P)$.*

Proof. Let $\widehat{x} \in O^E(G)$, that is:

For each $x_i \in X_i$ it turns out $u_i(x_i, \widehat{x_{-i}}) \notin u_i(\widehat{x}) + E_i$ for each $i \in N$
If and only if for each $x_i \in X_i$ it turns out $u_i(x_i, \widehat{x_{-i}}) - u_i(\widehat{x}) \notin E_i$
If and only if for each $x_i \in X_i$ it turns out $P(x_i, \widehat{x_{-i}}) - P(\widehat{x}) \notin E_i$
If and only if $\widehat{x} \in O^E(G^P)$. □

11.4 Multicriteria Games

In this section we study E-optimal points for multicriteria games. In general it is not easy to find the E-optimal points of a multicriteria game G, but in some important class of problems we can reduce the research of E-optimal points to search for the classical equilibria as shown in the following example.

Example 11.4. (a) If $E_1 = E_2 = (0, +\infty)$, $m_1 = m_2 = 1$ (that is the game is for one criterion), it turns out that the E-optimal points are the Nash equilibria for the game G, for short $O^E(G) = NE$.
(b) If $E_1 = (\epsilon, +\infty) = E_2$, $\epsilon > 0$, It turns out that the E-optimal points are the approximate Nash equilibria, for short $O^E(G) = \epsilon NE$.
(c) If $E_1 = \mathbb{R}^n_+$; $E_2 = \mathbb{R}^n_+$, then the E-optimal points are the strong Pareto equilibria of the game G, for short $O^E(G) = \text{sPE}(G)$.

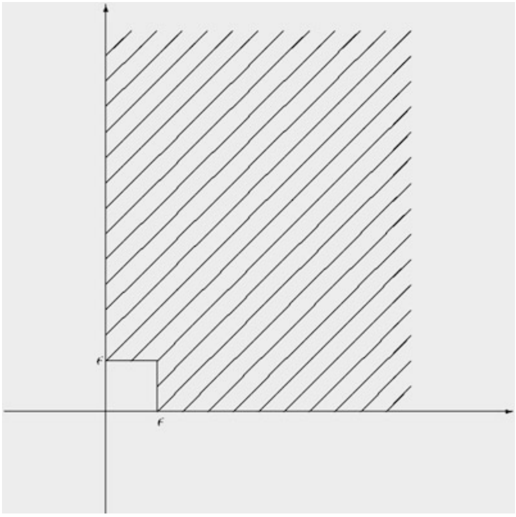


Fig. 11.3 The improvement set E in $\mathbb{R}^2$ (indicated by *shading*) is $\mathbb{R}^2 \backslash [0,\epsilon] \times [0,\epsilon]$

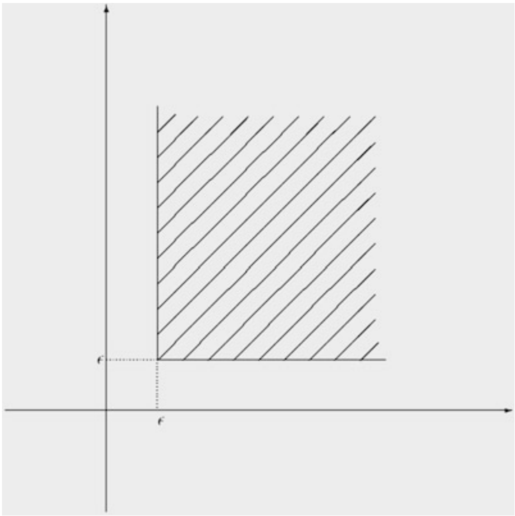


Fig. 11.4 The improvement set E in $\mathbb{R}^2$ (indicated by *shading*) is $\{x \in \mathbb{R}^2 : \ x_i \geq \epsilon_i > 0 \ (i = 1,2)\}$

(d) If $E_1 = \mathbb{R}^n_{++}$; $E_2 = \mathbb{R}^n_{++}$, then the E-optimal points are the weak Pareto equilibria of the game G, for short $O^E(G) = \text{wPE}(G)$.

(e) In the paper [13], the improvement set considered for multicriteria potential games is $E = \mathbb{R}^2_+ \setminus [0,\epsilon] \times [0,\epsilon]$ (see Fig. 11.3).

(f) In the paper [11] the improvement set considered for multicriteria games is

$$E = \{x \in \mathbb{R}^2 : x_i \geqq \epsilon_i, \ \epsilon_i \in \mathbb{R}_{++}, \ i = 1,2\} \text{ (see Fig. 11.4).}$$

Proposition 11.2. *A strategy profile $\widehat{x} \in X$ is an E-equilibrium for the multicriteria game G if, for each player i, $\widehat{x_i}$ is a E-optimal point for the vector function $u_i(.,\widehat{x_{-i}})$ with respect to the improvement set E that is*

$$\forall\, x_i \in X_i, C \cap (u_i(\widehat{x}) + E_i) = \emptyset \text{ where } C = \{z = u_i(x_i, \widehat{x_{-i}})\} \subset \mathbb{R}^n.$$

Proof. The proof follows from the definition of E-optimal point of a set. □

We remind that a relation $>_E$ defined on a set E is a *preorder* if the transitivity property is valid:

$$a >_E b, \;\; b >_E c \Rightarrow a >_E c \quad \forall a,b,c \in E.$$

A relation in E is called an *order* if the transitivity and asymmetric properties are valid:

$$a >_E b \Rightarrow \text{it is not valid } b >_E a, \quad \forall a,b \in E$$

We can define a relation on E in the following way:
$a >_E b$ if and only if $(a \in b + E)$ and we shall say "a is E-better then B". Intuitively the elements in $(b+E)$ are those *substantially better* than b. This relation is not an order in general because it is not *transitive*. If the improvement set has the property that $E + E \subset E$ then the transitive property is valid, for example this is true if $E = \mathbb{R}^m_+$, or if E is a convex set.

From what we have said we can ascertain the following:

Proposition 11.3. *Let $E = (E_1, E_2, \ldots, E_n)$ the improvement sets for the n players (E_i for the player i). The following conditions are equivalent:*

(i) *a strategy profile $\widehat{x} \in X$ is an E-equilibrium for the multicriteria game G if for each $i \in N$, $\nexists\, x_i \in X_i$ s.t. $u_i(x_i, \widehat{x_{-i}}) \in u_i(\widehat{x}) + E_i$.*
(ii) *if E_i $i = 1, \ldots n$ are convex sets, for all $x_i \in X_i$ $u_i(x_i, \widehat{x_{-i}})$ is not E_i-better then $u_i(\widehat{x})$ for each $i \in N$.*
(iii) *$\{u_i(x_i, \widehat{x_{-i}}) : x_i \in X_i\} \cap u_i(\widehat{x}) + E_i = \emptyset$ for each $i \in N$.*

Now let us define the *E-best reply* map.

Definition 11.8. E_i-$B_i(\widehat{x}_{-i}) = \{\widehat{x_i} \in X_i : \nexists x_i \in X_i : u_i(x_i, \widehat{x}_{-i}) \in u_i(\widehat{x}) + E_i\}$
This set is called the E-best reply of the player i with respect to the improvement set E_i, fixed the strategy of the other players. The map E-best reply is defined E-$BR =$ (E_1-B_1,...,E_n-B_n).

Proposition 11.4. *A strategy profile $\widehat{x} \in X$ is an E-equilibrium for the multicriteria game G if and only if $\widehat{x_i} \in E_i$-$B_i(\widehat{x_{-i}})$ for all $i \in N$*

Proof. The proof follows from the definitions and the properties given. □

11.5 Conclusions and Open Problem

In this paper we consider a new concept of equilibrium and approximate equilibrium for multicriteria games using improvement sets. We have also given, as a first step, an existence theorem for potential games. Some issues about open problems arise and would be interesting topics to develop further:

1. Existence theorems of E-equilibria for generic multicriteria games could be given.
2. In [7], one criterion Bayesian games were studied. Some results could be extended in the light of these new definitions to multicriteria bayesian ones.
3. A new concept of E- equilibria for multicriteria interval games could be interesting. Look at [14] for multicriteria interval games.
4. Perhaps we can study bargaining games through "improvement sets".
5. The approximate core for cooperative games (see [12]) could be defined through improvement sets.

Some of these issues are works in progress.

Acknowledgements This research was partially supported by MUR (Ministero Università Ricerca- Italia-) via a contract with S.Tijs. We thank Rodica Branzei for her helpful comments on a previous version of the paper. The valuable comments of the referees are gratefully acknowledged.

References

1. Borm, P.E.M., Tijs, S.H., van den Aarssen, J.C.M.: Pareto equilibria in multiobjective games. Methods Oper. Res. **60**, 303–312 (1989)
2. Chicco, M., Mignanego, F., Pusillo, L., Tijs, S.: Vector optimization problems via improvement sets. JOTA **150**(3), 516–529 (2011)
3. Ehrgott, M.: Multicriteria Optimization, 2nd ed. Springer, Berlin (2005)
4. Gutierrez, C., Jimenez, B., Novo, V.: A unified approach and optimatlity conditions for approximate solutions of vector optimization problems. Siam J. Optim. **17**(3), 688–710 (2006)
5. Loridan, P.: Well posedness in vector optimization. In: Lucchetti, R., Revalski, J. (eds.) Recent Developments in Well Posed-Variational Problems, pp. 171–192. Kluwer Academic, Dordrecht, The Netherland (1995)
6. Klemperer, P.: Auctions: Theory and Practice. Princeton University Press, Princeton, USA (2004)
7. Mallozzi, L., Pusillo, L., Tijs, S.: Approximate equilibria for Bayesian games. JMAA **342**(2), 1098–1102 (2007)
8. Margiocco, M., Pusillo, L.: Stackelberg well posedness and hierarchical potential games. In: Jorgensen, S., Quincampoix, M., Vincent, T. (eds.) Advances in Dynamic Games – Series: Annals of the International Society of Dynamic Games, pp. 111–128. Birkhauser, Boston (2007)
9. Margiocco, M., Pusillo, L.: Potential games and well-posedness. Optimization **57**, 571–579 (2008)
10. Monderer, D., Shapley, L.S.: Potential games. Games Econ. Behav. **14**, 124–143 (1996)

11. Morgan, J.: Approximations and well-posedness in multicriteria games. Ann. Oper. Res. **137**, 257–268 (2005)
12. Owen, G.: Game Theory, 3rd ed. Academic, New York (1995)
13. Patrone, F., Pusillo, L., Tijs, S.: Multicriteria games and potentials. TOP **15**, 138–145 (2007)
14. Pieri, G., Pusillo, L.: Interval values for multicriteria cooperative games. Auco Czech Econ. Rev. **4**, 144–155 (2010)
15. Puerto Albandoz, J., Fernandez Garcia F.: Teoria de Juegos Multiobjetivo, Imagraf Impresores S.A., Sevilla (2006)
16. Pusillo, L.: Interactive decisions and potential games. J. Global Optim. **40**, 339–352 (2008)
17. Shapley, L.S.: Equilibrium points in games with vector payoffs. Naval Res. Logistic Quart. **6**, 57–61 (1959)
18. Tijs S.H., ϵ equilibrium point theorems for two persons games. Methods Oper. Res. **26**, 755–766 (1977)
19. Voorneveld, M.: Potential games and interactive decisions with multiple criteria. Dissertation series n. 61. CentER of Economic Research- Tilburg University (1999)

Chapter 12
Mean Field Games with a Quadratic Hamiltonian: A Constructive Scheme

Olivier Guéant

Abstract Mean field games models describing the limit case of a large class of stochastic differential games, as the number of players goes to $+\infty$, were introduced by Lasry and Lions [C R Acad Sci Paris 343(9/10) (2006); Jpn. J. Math. 2(1) (2007)]. We use a change of variables to transform the mean field games equations into a system of simpler coupled partial differential equations in the case of a quadratic Hamiltonian. This system is then used to exhibit a monotonic scheme to build solutions of the mean field games equations.

Keywords Mean field games • Forward–backward equations • Monotonic schemes

12.1 Introduction

Mean field games (MFG) equations were introduced by Lasry and Lions [4–6] to describe the dynamic equilibrium of stochastic differential games involving a continuum of players.

Formally, we consider a continuum of agents, each agent being described by a position $X_t \in \overline{\Omega}$ [Ω being typically $(0,1)^d$] following a stochastic process $\mathrm{d}X_t = a_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$. In this stochastic process, a_t is controlled by the agent and W_t is a Brownian motion specific to the agent under investigation—this independence hypothesis being central in the sequel. These agents will interact in a mean field

O. Guéant (✉)
UFR de Mathématiques, Laboratoire Jacques-Louis Lions, Université Paris-Diderot, 175 rue du Chevaleret, 75013 Paris, France
e-mail: olivier.gueant@ann.jussieu.fr

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_12,

fashion, and we denote by $m(t,\cdot)$ the probability distribution function describing the distribution of the agents at time t.[1]

Each agent optimizes the same objective function, though possibly starting from a different position x_0:

$$\sup_{(a_t)_t} \mathbb{E}\left[\int_0^T (f(X_t, m(t,X_t)) - C(a_t))\,\mathrm{d}t + u_T(X_T)\middle| X_0 = x_0\right],$$

where f and u_T are functions whose regularity will be described subsequently and where C is a convex (cost) function.

To this problem we associate the so-called MFG equations. These equations consist in a backward Hamilton–Jacobi–Bellman (HJB) equation coupled with a forward Kolmogorov (K) transport equation:

$$\begin{aligned} &(\mathrm{HJB}) \qquad \partial_t u + \tfrac{\sigma^2}{2}\Delta u + H(\nabla u) = -f(x,m), \\ &(\mathrm{K}) \qquad \partial_t m + \nabla\cdot(mH'(\nabla u)) = \tfrac{\sigma^2}{2}\Delta m, \end{aligned}$$

with prescribed initial condition $m(0,\cdot) = m_0(\cdot) \geq 0$ and terminal condition $u(T,\cdot) = u_T(\cdot)$, where H is the Legendre transform of the cost function C.

In this paper, we focus on the particular case of the quadratic cost $C(a) = \frac{a^2}{2}$ and hence a quadratic Hamiltonian $H(p) = \frac{p^2}{2}$. In this special case, a change of variables was introduced by Guéant et al. in [3] to write the MFG equations as two coupled heat equations with similar source terms. If indeed we introduce $\phi = \exp\left(\frac{u}{\sigma^2}\right)$ and $\psi = m\exp\left(-\frac{u}{\sigma^2}\right)$, then the system reduces to

$$\begin{aligned} \partial_t\phi + \frac{\sigma^2}{2}\Delta\phi &= -\frac{1}{\sigma^2} f(x,\phi\psi)\phi, \\ \partial_t\psi - \frac{\sigma^2}{2}\Delta\psi &= \frac{1}{\sigma^2} f(x,\phi\psi)\psi, \end{aligned}$$

with $\phi(T,\cdot) = \exp\left(\frac{u_T(\cdot)}{\sigma^2}\right)$ and $\psi(0,\cdot) = \frac{m_0(\cdot)}{\phi(0,\cdot)}$.

We use this system to exhibit a constructive scheme for solutions to the MFG equations. This constructive scheme, proposed by Lasry and Lions in [7], starts with $\psi^0 = 0$ and builds recursively two sequences $(\phi^{n+\frac{1}{2}})_n$ and $(\psi^{n+1})_n$ using the following equations:

[1] In our case, this assumption consists only in assuming that the initial datum is a probability distribution function m_0.

$$\partial_t \phi^{n+\frac{1}{2}} + \frac{\sigma^2}{2}\Delta \phi^{n+\frac{1}{2}} = -\frac{1}{\sigma^2} f(x, \phi^{n+\frac{1}{2}} \psi^n)\phi^{n+\frac{1}{2}},$$

$$\partial_t \psi^{n+1} - \frac{\sigma^2}{2}\Delta \psi^{n+1} = \frac{1}{\sigma^2} f(x, \phi^{n+\frac{1}{2}} \psi^{n+1})\psi^{n+1},$$

with $\phi^{n+\frac{1}{2}}(T,\cdot) = \exp\left(\frac{u_T(\cdot)}{\sigma^2}\right)$ and $\psi^{n+1}(0,\cdot) = \frac{m_0(\cdot)}{\phi^{n+\frac{1}{2}}(0,\cdot)}$.

Then, ϕ and ψ are obtained as the monotonic limit of the two sequences $(\phi^{n+\frac{1}{2}})_n$ and $(\psi^n)_n$ under the usual assumptions on f.

In Sect. 12.2, we recall the change of variables and derive the associated system of coupled parabolic equations. Section 12.3 is devoted to the introduction of the functional framework, and we prove the main monotonicity properties of the system. Section 12.4 presents a constructive scheme and proves that we can have two monotonic sequences converging toward ϕ and ψ. Section 12.5 then gives additional properties on the constructive scheme regarding the absence of mass conservation.

12.2 From MFG Equations to a Forward–Backward System of Heat Equations with Source Terms

We consider the MFG equations introduced in [4–6] in the case of a quadratic Hamiltonian. These partial differential equations are considered on the domain $[0,T] \times \Omega$, where Ω stands for $(0,1)^d$, and consist in the following equations:

$$\text{(HJB)} \qquad \partial_t u + \frac{\sigma^2}{2}\Delta u + \frac{1}{2}|\nabla u|^2 = -f(x,m),$$

$$\text{(K)} \qquad \partial_t m + \nabla \cdot (m \nabla u) = \frac{\sigma^2}{2}\Delta m,$$

with:

- Boundary conditions: $\frac{\partial u}{\partial \mathbf{n}} = \frac{\partial m}{\partial \mathbf{n}} = 0$ on $(0,T) \times \partial\Omega$;
- Terminal condition: $u(T,\cdot) = u_T(\cdot)$, a given payoff whose regularity is to be specified;
- Initial condition: $m(0,\cdot) = m_0(\cdot) \geq 0$, a given positive function in $L^1(\Omega)$, typically a probability distribution function.

The change of variables introduced in [3] is recalled in the following proposition:

Proposition 12.1. *Let us consider a smooth solution (ϕ, ψ) of the following system $(\mathcal{S})$, with $\phi > 0$:*

$$\partial_t \phi + \frac{\sigma^2}{2}\Delta \phi = -\frac{1}{\sigma^2} f(x, \phi\psi)\phi \qquad (E_\phi),$$

$$\partial_t \psi - \frac{\sigma^2}{2}\Delta \psi = \frac{1}{\sigma^2} f(x, \phi\psi)\psi \qquad (E_\psi),$$

with:

- *Boundary conditions:* $\frac{\partial \phi}{\partial \mathbf{n}} = \frac{\partial \psi}{\partial \mathbf{n}} = 0$ *on* $(0,T) \times \partial \Omega$.
- *Terminal condition:* $\phi(T,\cdot) = \exp\left(\frac{u_T(\cdot)}{\sigma^2}\right)$.
- *Initial condition:* $\psi(0,\cdot) = \frac{m_0(\cdot)}{\phi(0,\cdot)}$.

Then $(u,m) = (\sigma^2 \ln(\phi), \phi\psi)$ *defines a solution of MFG.*

Proof. Let us start with (HJB):

$$\partial_t u = \sigma^2 \frac{\partial_t \phi}{\phi}, \qquad \nabla u = \sigma^2 \frac{\nabla \phi}{\phi} \qquad \Delta u = \sigma^2 \frac{\Delta \phi}{\phi} - \sigma^2 \frac{|\nabla \phi|^2}{\phi^2}.$$

Hence

$$\begin{aligned}
\partial_t u + \frac{\sigma^2}{2}\Delta u + \frac{1}{2}|\nabla u|^2 &= \sigma^2 \left[\frac{\partial_t \phi}{\phi} + \frac{\sigma^2}{2}\frac{\Delta \phi}{\phi}\right] \\
&= \frac{\sigma^2}{\phi}\left[-\frac{1}{\sigma^2} f(x,\phi\psi)\phi\right] \\
&= -f(x,m).
\end{aligned}$$

Now, for equation (K):

$$\partial_t m = \partial_t \phi \psi + \phi \partial_t \psi \qquad \nabla \cdot (\nabla u m) = \sigma^2 \nabla \cdot (\nabla \phi \psi) = \sigma^2 \left[\Delta \phi \psi + \nabla \phi \cdot \nabla \psi\right]$$
$$\Delta m = \Delta \phi \psi + 2\nabla \phi \cdot \nabla \psi + \phi \Delta \psi.$$

Hence

$$\begin{aligned}
\partial_t m + \nabla \cdot (\nabla u m) &= \partial_t \phi \psi + \phi \partial_t \psi + \sigma^2 \left[\Delta \phi \psi + \nabla \phi \cdot \nabla \psi\right] \\
&= \psi \left[\partial_t \phi + \sigma^2 \Delta \phi\right] + \phi \partial_t \psi + \sigma^2 \nabla \phi \cdot \nabla \psi \\
&= \psi \left[\frac{\sigma^2}{2}\Delta \phi - \frac{1}{\sigma^2} f(x,\phi\psi)\phi\right] \\
&\quad + \phi \left[\frac{\sigma^2}{2}\Delta \psi + \frac{1}{\sigma^2} f(x,\phi\psi)\psi\right] + \sigma^2 \nabla \phi \cdot \nabla \psi \\
&= \frac{\sigma^2}{2}\Delta \phi \psi + \sigma^2 \nabla \phi \cdot \nabla \psi + \frac{\sigma^2}{2}\phi \Delta \psi \\
&= \frac{\sigma^2}{2}\Delta m.
\end{aligned}$$

This proves the result since the boundary conditions and the initial and terminal conditions are coherent. □

Now we will focus our attention on the study of the preceding system of equations $(\mathcal{S})$ and use it to design a constructive scheme for the couple (ϕ,ψ) and thus for the couple (u,m) under regularity assumptions.

12.3 Properties of $(\mathcal{S})$

To study system $(\mathcal{S})$, we introduce several hypotheses on f: we suppose that it is a decreasing function of its second variable, continuous in that variable, and uniformly bounded. Moreover, to simplify the exposition,[2] we suppose that $f \leq 0$. The monotonicity hypothesis is to be linked to the usual proof of uniqueness for the MFG equations [6].

Now let us introduce the functional framework we are working in.

Let us denote by $\mathcal{P} \subset C([0,T],L^2(\Omega))$ the natural set of parabolic equations:

$$g \in \mathcal{P} \iff g \in L^2(0,T,H^1(\Omega)) \quad \text{and} \quad \partial_t g \in L^2(0,T,H^{-1}(\Omega)),$$

and let us introduce $\mathcal{P}_\epsilon = \{g \in \mathcal{P}, g \geq \epsilon\}$.

Proposition 12.2. *Suppose that $u_T \in L^\infty(\Omega)$. $\forall \psi \in \mathcal{P}_0$, there is a unique weak solution ϕ to the following equation (E_ϕ):*

$$\partial_t \phi + \frac{\sigma^2}{2}\Delta\phi = -\frac{1}{\sigma^2} f(x,\phi\psi)\phi \qquad (E_\phi),$$

with $\frac{\partial \phi}{\partial \mathbf{n}} = 0$ on $(0,T)\times\partial\Omega$ and $\phi(T,\cdot) = \exp\left(\frac{u_T(\cdot)}{\sigma^2}\right)$.
Hence $\Phi : \psi \in \mathcal{P}_0 \mapsto \phi \in \mathcal{P}$ is well defined.

Moreover, $\forall \psi \in \mathcal{P}_0, \phi = \Phi(\psi) \in \mathcal{P}_\epsilon$ for $\epsilon = \exp\left(-\frac{1}{\sigma^2}\left(\|u_T\|_\infty + \|f\|_\infty T\right)\right)$.

Proof. Let us consider $\psi \in \mathcal{P}_0$.

Existence of a weak solution ϕ:
Let us introduce $F_\psi : \varphi \in L^2(0,T,L^2(\Omega)) \mapsto \phi$ weak solution of the following linear parabolic equation:

$$\partial_t \phi + \frac{\sigma^2}{2}\Delta\phi = -\frac{1}{\sigma^2} f(x,\varphi\psi)\phi,$$

with $\frac{\partial \phi}{\partial \mathbf{n}} = 0$ on $(0,T)\times\partial\Omega$ and $\phi(T,\cdot) = \exp\left(\frac{u_T(\cdot)}{\sigma^2}\right)$.

By the classical theory of linear parabolic equations, ϕ is in $\mathcal{P} \subset L^2(0,T,L^2(\Omega))$.

Our goal is to use Schauder's fixed-point theorem on F_ψ.

[2] In terms of the initial MFG problem, the optimal control ∇u and the subsequent distribution m are not changed if we subtract $\|f\|_\infty$ to f.

12.3.1 Compactness

Common energy estimates [1] give that there exists a constant C that only depends on $\|u_T\|_\infty$, σ, and $\|f\|_\infty$ such that $\forall(\psi,\varphi)\in\mathcal{P}_0\times L^2(0,T,L^2(\Omega))$

$$\|F_\psi(\varphi)\|_{L^2(0,T,H^1(\Omega))}+\|\partial_t F_\psi(\varphi)\|_{L^2(0,T,H^{-1}(\Omega))}\le C.$$

Hence F_ψ maps the closed ball $B_{L^2(0,T,L^2(\Omega))}(0,C)$ to a compact subset of $B_{L^2(0,T,L^2(\Omega))}(0,C)$.

12.3.2 Continuity

Let us now prove that F_ψ is a continuous function.

Let us consider a sequence $(\varphi_n)_n$ of $L^2(0,T,L^2(\Omega))$ with $\varphi_n\to_{n\to\infty}\varphi$ in the $L^2(0,T,L^2(\Omega))$ sense. Let us write $\phi_n=F_\psi(\varphi_n)$. We know from the preceding compactness result that we can extract from $(\phi_n)_n$ a new sequence denoted $(\phi_{n'})_{n'}$ that converges in the $L^2(0,T,L^2(\Omega))$ sense toward a function ϕ. To prove that F_ψ is continuous, we then need to show that ϕ cannot be different from $F_\psi(\varphi)$.

Now, because of the energy estimates, we know that ϕ is in $\mathcal{P}$ and that we can extract another subsequence (still denoted $(\phi_{n'})_{n'}$) such that

- $\phi_{n'}\to\phi$ in the $L^2(0,T,L^2(\Omega))$ sense;
- $\nabla\phi_{n'}\rightharpoonup\nabla\phi$ weakly in $L^2(0,T,L^2(\Omega))$;
- $\partial_t\phi_{n'}\rightharpoonup\partial_t\phi$ weakly in $L^2(0,T,H^{-1}(\Omega))$;

and

- $\varphi_{n'}\to\varphi$ almost everywhere.

By definition, we have that $\forall w\in L^2(0,T,H^1(\Omega))$:

$$\int_0^T\langle\partial_t\phi_{n'}(t,\cdot),w(t,\cdot)\rangle_{H^{-1}(\Omega),H^1(\Omega)}\,\mathrm{d}t-\frac{\sigma^2}{2}\int_0^T\int_\Omega\nabla\phi_{n'}(t,x)\cdot\nabla w(t,x)\,\mathrm{d}x\,\mathrm{d}t$$
$$=-\frac{1}{\sigma^2}\int_0^T\int_\Omega f(x,\varphi_{n'}(t,x)\psi(t,x))\phi_{n'}(t,x)w(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

By weak convergence, the left-hand side of the preceding equality converges toward

$$\int_0^T\langle\partial_t\phi(t,\cdot),w(t,\cdot)\rangle_{H^{-1}(\Omega),H^1(\Omega)}\,\mathrm{d}t-\frac{\sigma^2}{2}\int_0^T\int_\Omega\nabla\phi(t,x)\cdot\nabla w(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

Since f is a bounded continuous function, and using the dominated convergence theorem, the right-hand side converges toward

$$-\frac{1}{\sigma^2}\int_0^T\int_\Omega f(x,\varphi(t,x)\psi(t,x))\phi(t,x)w(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

Hence $\phi = F_\psi(\varphi)$.

12.3.3 Schauder's Theorem

By Schauder's theorem, we then know that there exists a fixed-point ϕ to F_ψ and hence a weak solution to the nonlinear parabolic equation (E_ϕ).

Positiveness of ϕ:
Let us consider a solution ϕ as above. If $I(t) = \frac{1}{2}\int_\Omega(\phi(t,x)_-)^2\,\mathrm{d}x$, then:

$$\begin{aligned}
I'(t) &= -\int_\Omega \partial_t\phi(t,x)\phi(t,x)_-\,\mathrm{d}x\\
&= -\int_\Omega\left(\nabla\phi(t,x)\cdot\nabla(\phi(t,x)_-) - \frac{1}{\sigma^2}f(x,\phi(t,x)\psi(t,x))\phi(t,x)\phi(t,x)_-\right)\mathrm{d}x\\
&= -\int_\Omega\left(-|\nabla\phi(t,x)|^2 1_{\phi(t,x)\le 0} + \frac{1}{\sigma^2}f(x,\phi(t,x)\psi(t,x))(\phi(t,x)_-)^2\right)\mathrm{d}x\\
&= \int_\Omega |\nabla\phi(t,x)|^2 1_{\phi(t,x)\le 0} - \int_\Omega\frac{1}{\sigma^2}f(x,\phi(t,x)\psi(t,x))(\phi(t,x)_-)^2\,\mathrm{d}x\\
&\ge 0.
\end{aligned}$$

Since $I(T)=0$ and $I\ge 0$, we know that $I=0$. Hence, ϕ is positive.

Uniqueness:
Let us consider two weak solutions ϕ_1 and ϕ_2 to equation (E_ϕ).

Let us introduce $J(t) = \frac{1}{2}\int_\Omega(\phi_2(t,x)-\phi_1(t,x))^2\,\mathrm{d}x$. We have

$$\begin{aligned}
J'(t) &= \int_\Omega(\partial_t\phi_2(t,x)-\partial_t\phi_1(t,x))(\phi_2(t,x)-\phi_1(t,x))\,\mathrm{d}x\\
&= -\int_\Omega\frac{1}{\sigma^2}\left(f(x,\phi_2(t,x)\psi(t,x))\phi_2(t,x) - f(x,\phi_1(t,x)\psi(t,x))\phi_1(t,x)\right)\\
&\qquad\times(\phi_2(t,x)-\phi_1(t,x))\,\mathrm{d}x + \int_\Omega|\nabla\phi_2(t,x)-\nabla\phi_1(t,x)|^2\,\mathrm{d}x.
\end{aligned}$$

Because of our assumptions on f, the function $\xi\in\mathbb{R}_+\mapsto\frac{1}{\sigma^2}f(x,\psi\xi)\xi$ is a decreasing function.

Hence, since ϕ_1 and ϕ_2 are positive, $J'(t) \geq 0$. Since $J(T) = 0$ and $J \geq 0$, we know that $J = 0$. Hence, $\phi_1 = \phi_2$.

Lower bound to ϕ:

We can get a lower bound to ϕ through a subsolution taken as the solution of the following ordinary differential equation:

$$\underline{\phi}'(t) = \frac{1}{\sigma^2}\|f\|_\infty \underline{\phi}(t) \qquad \underline{\phi}(T) = \exp\left(-\frac{\|u_T\|_\infty}{\sigma^2}\right).$$

Let us indeed consider $K(t) = \frac{1}{2}\int_\Omega ((\underline{\phi}(t) - \phi(t,x))_+)^2 \, \mathrm{d}x$. We have

$$\begin{aligned}
K'(t) &= \int_\Omega (\underline{\phi}'(t) - \partial_t \phi(t,x))(\underline{\phi}(t) - \phi(t,x))_+ \mathrm{d}x \\
&= \int_\Omega \Bigg(\frac{1}{\sigma^2}\|f\|_\infty \underline{\phi}(t)(\underline{\phi}(t) - \phi(t,x))_+ + |\nabla\phi(t,x)|^2 1_{\underline{\phi}(t)-\phi(t,x)\geq 0} \\
&\quad + \frac{1}{\sigma^2}\left(f(x,\phi(t,x)\psi(t,x))\phi(t,x)(\underline{\phi}(t) - \phi(t,x))_+\right)\Bigg) \mathrm{d}x \\
&\geq \frac{1}{\sigma^2}\int_\Omega \left(\|f\|_\infty \underline{\phi}(t) + f(x,\phi(t,x)\psi(t,x))\phi(t,x)\right)(\underline{\phi}(t) - \phi(t,x))_+ \mathrm{d}x \\
&\geq \frac{1}{\sigma^2}\int_\Omega (\|f\|_\infty + f(x,\phi(t,x)\psi(t,x)))\, \phi(t,x)(\underline{\phi}(t) - \phi(t,x))_+ \mathrm{d}x \\
&\geq 0.
\end{aligned}$$

Since $K(T) = 0$ and $K \geq 0$, we know that $K = 0$. Hence, $\phi(t,x) \geq \underline{\phi}(t) = e^{-\frac{\|u_T\|_\infty}{\sigma^2}} \exp\left(-\frac{1}{\sigma^2}\|f\|_\infty (T-t)\right) \geq \epsilon$, and the result follows. □

Now we turn to a monotonicity result regarding Φ.

Proposition 12.3.

$$\forall \psi_1 \leq \psi_2 \in \mathcal{P}_0, \Phi(\psi_1) \geq \Phi(\psi_2).$$

Proof. Let us introduce $\phi_1 = \Phi(\psi_1)$ and $\phi_2 = \Phi(\psi_2)$.

Let us introduce $I(t) = \frac{1}{2}\int_\Omega ((\phi_2(t,x) - \phi_1(t,x))_+)^2 \mathrm{d}x$. We have

$$\begin{aligned}
I'(t) &= \int_\Omega (\partial_t \phi_2(t,x) - \partial_t \phi_1(t,x))(\phi_2(t,x) - \phi_1(t,x))_+ \mathrm{d}x \\
&= \int_\Omega \Bigg(|\nabla\phi_2(t,x) - \nabla\phi_1(t,x)|^2 1_{\phi_2(t,x)-\phi_1(t,x)\geq 0} - \frac{1}{\sigma^2}(\phi_2(t,x) - \phi_1(t,x))_+ \\
&\quad \times (f(x,\phi_2(t,x)\psi_2(t,x))\phi_2(t,x) - f(x,\phi_1(t,x)\psi_1(t,x))\phi_1(t,x)) \Bigg) \mathrm{d}x
\end{aligned}$$

$$\geq \frac{1}{\sigma^2}\int_\Omega (f(x,\phi_1(t,x)\psi_1(t,x))\phi_1(t,x) - f(x,\phi_2(t,x)\psi_2(t,x))\phi_2(t,x))$$
$$\times(\phi_2(t,x)-\phi_1(t,x))_+ \mathrm{d}x$$
$$\geq \frac{1}{\sigma^2}\int_\Omega (f(x,\phi_1(t,x)\psi_2(t,x))\phi_1(t,x) - f(x,\phi_2(t,x)\psi_2(t,x))\phi_2(t,x))$$
$$\times(\phi_2(t,x)-\phi_1(t,x))_+ \mathrm{d}x$$
$$\geq 0.$$

Hence, since $I(T)=0$ and $I \geq 0$, we know that $I=0$. Consequently, $\phi_1 \geq \phi_2$. □

We now turn to the second equation (E_ψ) of system $(\mathcal{S})$.

Proposition 12.4. *Let us fix $\epsilon > 0$ and suppose that $m_0 \in L^2(\Omega)$.*
$\forall \phi \in \mathcal{P}_\epsilon$, there is a unique weak solution ψ to the following equation (E_ψ):

$$\partial_t \psi - \frac{\sigma^2}{2}\Delta\psi = \frac{1}{\sigma^2} f(x,\phi\psi)\psi \qquad (E_\psi)$$

with $\frac{\partial\psi}{\partial\mathbf{n}} = 0$ on $(0,T)\times\partial\Omega$ and $\psi(0,\cdot) = \frac{m_0(\cdot)}{\phi(0,\cdot)}$.
Hence $\Psi : \phi \in \mathcal{P}_\epsilon \mapsto \psi \in \mathcal{P}$ is well defined.
Moreover, $\forall \phi \in \mathcal{P}_\epsilon, \psi = \Psi(\phi) \in \mathcal{P}_0$.

Proof. The proof of existence and uniqueness of a weak solution $\psi \in \mathcal{P}$ is the same as in Proposition 12.2. The only thing to notice is that the initial condition $\psi(0,\cdot)$ is in Ł$^2(\Omega)$ because $m_0 \in L^2(\Omega)$ and ϕ is bounded from below by $\epsilon > 0$.

Now, to prove that $\psi \geq 0$, let us introduce $I(t) = \frac{1}{2}\int_\Omega (\psi(t,x)_-)^2\,\mathrm{d}x$; then

$$I'(t) = -\int_\Omega \partial_t\psi(t,x)\psi(t,x)_-\mathrm{d}x$$
$$= -\int_\Omega \left(-\nabla\psi(t,x)\cdot\nabla(\psi(t,x)_-) + \frac{1}{\sigma^2}f(x,\phi(t,x)\psi(t,x))\psi(t,x)\psi(t,x)_-\right)\mathrm{d}x$$
$$= -\int_\Omega \left(|\nabla\psi(t,x)|^2 1_{\psi(t,x)\leq 0} - \frac{1}{\sigma^2}f(x,\phi(t,x)\psi(t,x))(\psi(t,x)_-)^2\right)\mathrm{d}x$$
$$\leq 0.$$

Since $I(0)=0$ and $I\geq 0$, we know that $I=0$. Hence, ψ is positive. □

Now we turn to a monotonicity result regarding Ψ.

Proposition 12.5.

$$\forall \phi_1 \leq \phi_2 \in \mathcal{P}_\epsilon, \Psi(\phi_1) \geq \Psi(\phi_2).$$

Proof. Let us introduce $\psi_1 = \Psi(\phi_1)$ and $\psi_2 = \Psi(\phi_2)$.

Let us introduce $I(t) = \frac{1}{2}\int_\Omega ((\psi_2(t,x) - \psi_1(t,x))_+)^2\,\mathrm{d}x$. We have

$$\begin{aligned}
I'(t) &= \int_\Omega (\partial_t \psi_2(t,x) - \partial_t \psi_1(t,x))(\psi_2(t,x) - \psi_1(t,x))_+ \mathrm{d}x \\
&= -\int_\Omega |\nabla\psi_2(t,x) - \nabla\psi_1(t,x)|^2 1_{\psi_2(t,x)-\psi_1(t,x)\geq 0}\,\mathrm{d}x + \frac{1}{\sigma^2}\int_\Omega (\psi_2(t,x) - \psi_1(t,x))_+ \\
&\quad \times (f(x,\phi_2(t,x)\psi_2(t,x))\psi_2(t,x) - f(x,\phi_1(t,x)\psi_1(t,x))\psi_1(t,x))\,\mathrm{d}x \\
&\leq \frac{1}{\sigma^2}\int_\Omega (f(x,\phi_2(t,x)\psi_2(t,x))\psi_2(t,x) - f(x,\phi_1(t,x)\psi_1(t,x))\psi_1(t,x)) \\
&\quad \times (\psi_2(t,x) - \psi_1(t,x))_+ \mathrm{d}x \\
&\leq \frac{1}{\sigma^2}\int_\Omega (f(x,\phi_1(t,x)\psi_2(t,x))\psi_2(t,x) - f(x,\phi_1(t,x)\psi_1(t,x))\psi_1(t,x)) \\
&\quad \times (\psi_2(t,x) - \psi_1(t,x))_+ \mathrm{d}x \\
&\leq 0.
\end{aligned}$$

Now, $I(0) = \frac{1}{2}\int_\Omega m_0(x)((\frac{1}{\phi_2(0,x)} - \frac{1}{\phi_1(0,x)})_+)^2\,\mathrm{d}x = 0$. Hence since $I \geq 0$, we know that $I = 0$. Consequently, $\psi_1 \geq \psi_2$. □

We will use these properties to design a constructive scheme for the couple (ϕ,ψ).

12.4 A Constructive Scheme to Solve System $(\mathcal{S})$

The scheme we consider involves two sequences $(\phi^{n+\frac{1}{2}})_n$ and $(\psi^n)_n$ that are built using the following recursive equations:

$$\psi^0 = 0,$$

$$\forall n \in \mathbb{N}, \phi^{n+\frac{1}{2}} = \boldsymbol{\Phi}(\psi^n),$$

$$\forall n \in \mathbb{N}, \psi^{n+1} = \boldsymbol{\Psi}(\phi^{n+\frac{1}{2}}).$$

Theorem 12.1. *Suppose that $u_T \in L^\infty(\Omega)$ and that $m_0 \in L^2(\Omega)$.*
Then, the preceding scheme has the following properties:

- $(\phi^{n+\frac{1}{2}})_n$ *is a decreasing sequence of $\mathcal{P}_\epsilon$, where ϵ is as in Proposition 12.2.*
- $(\psi^n)_n$ *is an increasing sequence of $\mathcal{P}_0$, bounded from above by $\Psi(\epsilon)$.*
- $(\phi^{n+\frac{1}{2}},\psi^n)_n$ *converges for almost every $(t,x) \in (0,T)\times\Omega$, and in $L^2(0,T,L^2(\Omega))$ toward a couple (ϕ,ψ).*
- $(\phi,\psi) \in \mathcal{P}_\epsilon \times \mathcal{P}_0$ *is a weak solution of $(\mathcal{S})$.*

Proof. By immediate induction, we obtain from Propositions 12.2 and 12.4 that the two sequences are well defined and in the appropriate spaces.

Now, as far as monotonicity is concerned, we have that $\psi^1 = \Psi(\phi^{\frac{1}{2}}) \geq 0 = \psi^0$. Hence, if for a given $n \in \mathbb{N}$ we have $\psi^{n+1} \geq \psi^n$, then Proposition 3 gives

$$\phi^{n+\frac{3}{2}} = \Phi(\psi^{n+1}) \leq \Phi(\psi^n) = \phi^{n+\frac{1}{2}}.$$

Applying now the function Ψ we obtain

$$\psi^{n+2} = \Psi(\phi^{n+\frac{3}{2}}) \geq \Psi(\phi^{n+\frac{1}{2}}) = \psi^{n+1}.$$

By induction, we then have that $(\phi^{n+\frac{1}{2}})_n$ is decreasing and $(\psi^n)_n$ is increasing.

Moreover, since $\phi^{n+\frac{1}{2}} \geq \epsilon$, we have that $\psi^{n+1} = \Psi(\phi^{n+\frac{1}{2}}) \leq \Psi(\epsilon)$.

Now this monotonic behavior allows us to define two limit functions ϕ and ψ in $L^2(0,T,L^2(\Omega))$, and the convergence is almost everywhere and in $L^2(0,T,L^2(\Omega))$.

Now we want to show that (ϕ,ψ) is a weak solution of $(\mathcal{S})$, and to this end we use the energy estimates of the parabolic equations.

We know that there exists a constant $C > 0$ that only depends on $\|u_T\|_\infty$, σ, and $\|f\|_\infty$ such that

$$\forall n \in \mathbb{N}, \qquad \|\phi^{n+\frac{1}{2}}\|_{L^2(0,T,H^1(\Omega))} + \|\partial_t \phi^{n+\frac{1}{2}}\|_{L^2(0,T,H^{-1}(\Omega))} \leq C.$$

Hence, $\phi \in \mathcal{P}$, and we can extract a subsequence $(\phi^{n'+\frac{1}{2}})_{n'}$ such that

- $\phi^{n'+\frac{1}{2}} \to \phi$ in the $L^2(0,T,L^2(\Omega))$ sense and almost everywhere.
- $\nabla \phi^{n'+\frac{1}{2}} \rightharpoonup \nabla \phi$ weakly in $L^2(0,T,L^2(\Omega))$.
- $\partial_t \phi^{n'+\frac{1}{2}} \rightharpoonup \partial_t \phi$ weakly in $L^2(0,T,H^{-1}(\Omega))$.

Now, for $w \in L^2(0,T,H^1(\Omega))$

$$\int_0^T \langle \partial_t \phi^{n'+\frac{1}{2}}(t,\cdot), w(t,\cdot)\rangle_{H^{-1}(\Omega),H^1(\Omega)} \,\mathrm{d}t - \frac{\sigma^2}{2}\int_0^T \int_\Omega \nabla \phi^{n'+\frac{1}{2}}(t,x)\cdot \nabla w(t,x)\,\mathrm{d}x\,\mathrm{d}t$$

$$= -\frac{1}{\sigma^2}\int_0^T \int_\Omega f(x,\phi^{n'+\frac{1}{2}}(t,x)\psi^{n'}(t,x))\phi^{n'+\frac{1}{2}}(t,x) w(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

Using the weak convergence stated previously, the continuity hypothesis on f, and the dominated convergence theorem for the last term, we get that $\forall w \in L^2(0,T,H^1(\Omega))$

$$\int_0^T \langle \partial_t \phi(t,\cdot), w(t,\cdot)\rangle_{H^{-1}(\Omega),H^1(\Omega)} \,\mathrm{d}t - \frac{\sigma^2}{2}\int_0^T \int_\Omega \nabla \phi(t,x)\cdot \nabla w(t,x)\,\mathrm{d}x\,\mathrm{d}t$$

$$= -\frac{1}{\sigma^2}\int_0^T \int_\Omega f(x,\phi(t,x)\psi(t,x))\phi(t,x) w(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

Hence, for almost every $t \in (0,T)$ and $\forall v \in H^1(\Omega)$,

$$\langle \partial_t \phi(t,\cdot), v\rangle_{H^{-1}(\Omega),H^1(\Omega)} - \frac{\sigma^2}{2}\int_\Omega \nabla\phi(t,x)\cdot\nabla v(x)\,\mathrm{d}x$$
$$= -\frac{1}{\sigma^2}\int_\Omega f(x,\phi(t,x)\psi(t,x))\phi(t,x)v(x)\,\mathrm{d}x,$$

and the terminal condition is obviously satisfied.

Obviously, we also have $\phi \geq \epsilon$.

Now, turning to the second equation, the proof works the same. The only additional thing to notice is that

$$\psi(0,\cdot) = \lim_{n\to\infty}\psi^{n+1}(0,\cdot) = \lim_{n\to\infty}\frac{m_0}{\phi^{n+\frac{1}{2}}(0,\cdot)} = \frac{m_0}{\phi(0,\cdot)},$$

where the limits are in the $L^2(\Omega)$ sense.

Hence (ϕ,ψ) is indeed a weak solution of $(\mathcal{S})$. □

12.5 Concluding Remarks

In this chapter, we exhibited a monotonic way to build a solution to the system $(\mathcal{S})$. To understand well the nature of the change of variables and of the constructive scheme we used, let us introduce the sequence $(m^{n+1})_n$, where $m^{n+1} = \phi^{n+\frac{1}{2}}\psi^{n+1}$. From Theorem 12.1, we know that $(m^{n+1})_n$ converges almost everywhere and in L^1 toward the function $m = \phi\psi$ for which we have the conservation of mass along the trajectory. However, this property is not true for m^{n+1} as it is stated in the following proposition.

Proposition 12.6. *Let us consider $n \in \mathbb{N}$ and let us denote by $M^{n+1}(t) = \int_\Omega m^{n+1}(t,x)\,\mathrm{d}x$ the total mass of m^{n+1} at date t.*

Then, there may be a loss of mass along the trajectory in the sense that:

$$\frac{\mathrm{d}}{\mathrm{d}t}M^{n+1}(t) = \int_\Omega \psi^{n+1}(t,x)\phi^{n+\frac{1}{2}}(t,x)\times\left(f\left(x,\psi^{n+1}(t,x)\phi^{n+\frac{1}{2}}(t,x)\right)\right.$$
$$\left. - f\left(x,\psi^{n}(t,x)\phi^{n+\frac{1}{2}}(t,x)\right)\right)\mathrm{d}x \leq 0.$$

Proof. From the regularity obtained previously we can write

$$\frac{\mathrm{d}}{\mathrm{d}t}M^{n+1}(t) = \langle\partial_t\phi^{n+\frac{1}{2}}(t,\cdot),\psi^{n+1}(t,\cdot)\rangle_{H^{-1}(\Omega),H^1(\Omega)}$$
$$+\langle\partial_t\psi^{n+1}(t,\cdot),\phi^{n+\frac{1}{2}}(t,\cdot)\rangle_{H^{-1}(\Omega),H^1(\Omega)}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{2}\int_\Omega \nabla\phi^{n+\frac{1}{2}}(t,x)\cdot\nabla\psi^{n+1}(t,x)\,\mathrm{d}x \\
&\quad -\frac{1}{\sigma^2}\int_\Omega \phi^{n+\frac{1}{2}}(t,x)\psi^{n+1}(t,x)f\left(x,\phi^{n+\frac{1}{2}}(t,x)\psi^{n}(t,x)\right)\mathrm{d}x \\
&\quad -\frac{\sigma^2}{2}\int_\Omega \nabla\phi^{n+\frac{1}{2}}(t,x)\cdot\nabla\psi^{n+1}(t,x)\,\mathrm{d}x \\
&\quad +\frac{1}{\sigma^2}\int_\Omega \phi^{n+\frac{1}{2}}(t,x)\psi^{n+1}(t,x)f\left(x,\phi^{n+\frac{1}{2}}(t,x)\psi^{n+1}(t,x)\right)\mathrm{d}x \\
&= \int_\Omega \psi^{n+1}(t,x)\phi^{n+\frac{1}{2}}(t,x) \\
&\quad \times\left(f\left(x,\psi^{n+1}(t,x)\phi^{n+\frac{1}{2}}(t,x)\right)-f\left(x,\psi^{n}(t,x)\phi^{n+\frac{1}{2}}(t,x)\right)\right) \\
&\le 0. \qquad \square
\end{aligned}
$$

This property shows that the constructive scheme is rather original since it basically consists in building probability distribution functions using sequences of functions in L^1 that only have the right total mass asymptotically. Despite this absence of mass conservation, a discrete counterpart of this constructive scheme is developed in a work in progress [2] to numerically compute approximations of the solutions.

References

1. Evans, L.C.: Partial Differential Equations (Graduate Studies in Mathematics, vol. 19). American Mathematical Society, Providence, RI (2009)
2. Guéant, O.: Mean field games equations with quadratic Hamiltonian: a specific approach. Math. Models Methods Appl. Sci., 22, (2012)
3. Guéant, O., Lasry, J.M., Lions, P.L.: Mean field games and applications. In: Paris Princeton Lectures on Mathematical Finance (2010)
4. Lasry, J.-M., Lions, P.-L.: Jeux champ moyen. I. Le cas stationnaire. C. R. Acad. Sci. Paris **343**(9), 619–625
5. Lasry, J.-M., Lions, P.-L.: Jeux champ moyen. II. Horizon fini et contrôle optimal. C. R. Acad. Sci. Paris **343**(10), 679–684 (2006)
6. Lasry, J.-M., Lions, P.-L.: Mean field games. Jpn. J. Math. 2(1), 229–260 (2007)
7. Lasry, J.-M., Lions, P.-L.: Cours au collège de france: théorie des jeux champs moyens. http://www.college-de-france.fr/default/EN/all/equ_der/audio_video.jsp. (2008)

Chapter 13
Differential Game-Theoretic Approach to a Spatial Jamming Problem

Sourabh Bhattacharya and Tamer Başar

Abstract In this work, we investigate the effect of an aerial jamming attack on the communication network of a team of UAVs flying in a formation. We propose a communication and motion model for the UAVs. The communication model provides a relation in the spatial domain for effective jamming by an aerial intruder. We formulate the problem as a zero-sum pursuit-evasion game. In our earlier work, we used Isaacs' approach to obtain motion strategies for a pair of UAVs to evade the jamming attack. We also provided motion strategies for an aerial intruder to jam the communication between a pair of UAVs. In this work, we extend the analysis to multiple jammers and multiple UAVs flying in a formation. We analyze the problem in the framework of differential game theory, and provide analytical and approximate techniques to compute non-singular motion strategies of the UAVs.

Keywords Pursuit-evasion games • Aerial vehicles • Jamming

13.1 Introduction

In the past few years, considerable research has been done to deploy multiple UAVs in a decentralized manner to carry out tasks in military as well as civilian scenarios. UAVs have shown promise in a wide range of applications. The recent availability of low-cost UAVs suggests the use of teams of vehicles to perform various tasks such as mapping, surveillance, search and tracking operations [10, 43]. For these applications, there has been much focus to deploy teams of multiple UAVs

***This work was supported in part by a grant from AFOSR.

S. Bhattacharya (✉) • T. Başar
Coordinated Science Laboratory, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA
e-mail: sbhattac@illinois.edu; basar1@illinois.edu

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_13,

in cooperative or competitive manners [31]. An extensive summary of important milestones and future challenges in networked control of multiple UAVs has been presented in [34].

In general, the mode of communication among UAVs deployed in a team mission is wireless. This renders the communication channel vulnerable to malicious attacks from aerial intruders flying in the vicinity. An example of such an intruder is an aerial jammer. Jamming is a malicious attack whose objective is to disrupt the communication of the victim network intentionally, causing interference or collision at the receiver side. Jamming attack is a well-studied and an active area of research in wireless networks. Many defense strategies have been proposed by researchers against jamming in wireless networks. In [47], Wu et al. have proposed two strategies to evade jamming. The first strategy, channel surfing, is a form of spectral evasion that involves legitimate wireless devices changing the channel that they are operating on. The second strategy, spatial retreats, is a form of special evasion whereby legitimate devices move away from the jammer. In [45], Wood et al. have presented a distributed protocol to map jammed region so that the network can avoid routing traffic through it. The solution proposed by Cagalj et al. [7] uses different worm holes (wired worm holes, frequency-hopping pairs, and uncoordinated channel hopping) that lead out of the jammed region to report the alarm to the network operator. In [46], Wood et al. have investigated how to deliberately avoid jamming in IEEE 802.15.4 based wireless networks. In [8], Chen has proposed a strategy to introduce a special node in the network called the anti-jammer to drain the jammer's energy. To achieve its goal, the anti-jammer configures the probability of transmitting bait packets to attract the jammer to transmit.

For a static jammer and mobile nodes, the optimal strategy for the nodes is to retreat away from the jammer after detecting jamming. In case of an aerial jamming attack, optimal strategies for retreat are harder to compute due to the mobility of the jammer and constraints in the kinematics of the UAVs. This attack can be modeled as a zero-sum game [1] between the jammer and the UAVs. Such dynamic games governed by differential equations can be analyzed using tools from differential game theory [19, 23]. In the past, differential game theory has been used as a framework to analyze problems in multi-player pursuit-evasion games. Solutions for particular multi-player games were presented by Pashkov and Terekhov [30], Levchenkov and Pashkov [22], Hagedorn and Breakwell [18], Breakwell and Hagedorn [6], and Shankaran et al. [35]. More general treatments of multi-player differential games were presented by Starr and Ho [37], Vaisbord and Zhukovskiy [44], Zhukovskiy and Salukvadze [48] and Stipanović et al. [40]. The inherent hardness in obtaining an analytical solution to Hamilton–Jacobi–Bellman–Isaacs equation has led to the development of numerical techniques for the computation of the value function. Recent efforts in this direction to compute approximate reachable sets have been provided by Mitchell and Tomlin [26] and Stipanović et al. [38, 39].

In contradistinction with the existing literature, our work analyzes the behavior of multiple UAVs in cooperative as well as non-cooperative scenarios in the presence of a malicious intruder in the communication network. In this paper, we envision a scenario in which aerial jammers intrude the communication network in a multiple

UAV formation. We model the intrusion as a continuous time pursuit-evasion game between the UAVs and the aerial jammers. In contradistinction to the previous work on pursuit-evasion games that formulate a payoff based on a geometric quantity in the configuration space of the system, we formulate a payoff based on the capability of the players in a team to communicate among themselves in the presence of a jammer in the vicinity. In particular, we are interested in computing strategies for spatial reconfiguration of a formation of UAVs in the presence of an aerial jammer to reduce the jamming on the communication channel.

In [2], we investigated the problem of finding motion strategies for two UAVs to evade jamming in the presence of an aerial intruder. We considered a differential game-theoretic approach to compute optimal strategies by a team of UAVs. We formulated the problem as a zero-sum pursuit-evasion game. The cost function was picked as the termination time of the game. We used *Isaacs'* approach to derive the necessary conditions to arrive at the equations governing the saddle-point strategies of the players. In [4], we extended the previous analysis to a team of heterogeneous vehicles, consisting of UAVs and autonomous ground vehicles (AGVs). In [3], we generalized the previous work to networks having an arbitrary number of agents possessing different dynamics. We modeled the problem as one of maintaining connectivity in a dynamic graph in which the existence of an edge between two nodes depends on the state of the nodes as well as the jammer. Due to the dependence of the combinatorial structure of the graph on the continuous-time dynamics of the nodes we used the notion of *state-dependent graphs*, introduced in [25], to model the problem. Applying tools from algebraic graph theory on the state-dependent graphs provided us with greedy strategies for connectivity maintenance for the agents as well as the jammer.

In this work, we extend our previous works to the case of multiple UAVs flying in a formation in the vicinity of multiple aerial jammers. We use Isaacs' conditions to obtain the saddle point strategies as well as the retrogressive path equations (RPE) for the UAVs as well as the jammers. Using tools from algebraic graph theory, we present an approximation of the terminal manifold of the game. The RPE can be integrated back in time from the terminal manifolds to provide an approximate value of the game.

The rest of the sections are organized as follows. In Sect. 13.2, we present the problem formulation, and provide a communication model and the kinematic model for the UAVs. In Sect. 13.3, the saddle-point strategy is computed for the UAVs and the jammers. In Sect. 13.4, we characterize the terminal manifolds appearing in the game. In Sect. 13.5, we present simulation results for scenarios involving different number of UAVs and jammers. Finally, Sect. 13.6 presents the conclusions.

13.2 Communication and Dynamic Model

In this section, we first introduce a communication model between two mobile nodes in the presence of a jammer. Then we present the mobility models for the nodes. We conclude the section by formally formulating the problems we study in the paper.

13.2.1 Jammer and Communication Model

Consider a mobile node (*receiver*) receiving messages from another mobile node (*transmitter*) at some frequency. Both communicating nodes are assumed to be lying on a plane. Consider a third node that is attempting to jam the communication channel in between the transmitter and the receiver by sending a high power noise at the same frequency. This kind of jamming is referred to as *trivial jamming*. Two other types of jamming are:

1. *Periodic jamming*: This refers to a periodic noise pulse being generated by the jammer irrespective of the packets that are put on the network.
2. *Intelligent jamming*: In this mode of jamming a jammer is put in a promiscuous mode to destroy primarily the control packets.

A variety of metrics can be used to compare the effectiveness of various jamming attacks. Some of these metrics are energy efficiency, low probability of detection, and strong *denial of service* [27, 29]. In this paper, we use the ratio of the jamming-power to the signal-power (JSR) as the metric. From [32], we have the following models for the JSR (ξ) at the receiver's antenna.

1. R^n model

$$\xi = \frac{P_{J_\mathrm{T}} G_\mathrm{JR} G_\mathrm{RJ}}{P_\mathrm{T} G_\mathrm{TR} G_\mathrm{RT}} 10^{n \log_{10}\left(\frac{D_\mathrm{TR}}{D_\mathrm{JR}}\right)}$$

2. Ground Reflection Propagation

$$\xi = \frac{P_{J_\mathrm{T}} G_\mathrm{JR} G_\mathrm{RJ}}{P_\mathrm{T} G_\mathrm{TR} G_\mathrm{RT}} \left(\frac{h_\mathrm{J}}{h_\mathrm{T}}\right)^2 \left(\frac{D_\mathrm{TR}}{D_\mathrm{JR}}\right)^4$$

3. Nicholson

$$\xi = \frac{P_{J_\mathrm{T}} G_\mathrm{JR} G_\mathrm{RJ}}{P_\mathrm{T} G_\mathrm{TR} G_\mathrm{RT}} 10^{4 \log_{10}\left(\frac{D_\mathrm{TR}}{D_\mathrm{JR}}\right)}$$

where P_{J_T} is the power of the jammer transmitting antenna, P_T is the power of the transmitter, G_TR is the antenna gain from transmitter to receiver, G_RT is the antenna gain from receiver to transmitter, G_JR is the antenna gain from jammer to receiver, G_RJ is the antenna gain from receiver to jammer, h_J is the height of the jammer antenna above the ground, h_T is the height of the transmitter antenna above the ground, D_TR is the Euclidean distance between the transmitter and the receiver, and D_JR is the Euclidean distance between the jammer and the transmitter. All the above models are based on the propagation loss depending on the distance of the jammer and the transmitter from the receiver. In all the above models JSR is dependent on the ratio $\frac{D_\mathrm{TR}}{D_\mathrm{JR}}$.

For digital signals, the jammer's goal is to raise the ratio to a level such that the *bit error rate* [33] is above a certain threshold. For analog voice communication,

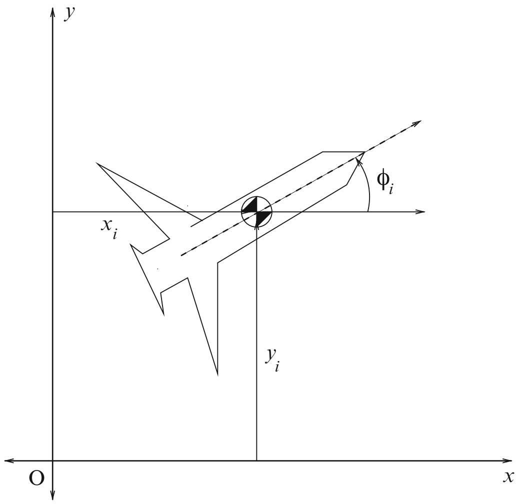


Fig. 13.1 Configuration of a UAV

the goal is to reduce the articulation performance so that the signals are difficult to understand. Hence we assume that the communication channel between a receiver and a transmitter is considered to be jammed in the presence of a jammer if $\xi \geq \xi_{tr}$ where ξ_{tr} is a threshold determined by many factors including application scenario and communication hardware. If all the parameters except the mutual distances between the jammer, transmitter and receiver are kept constant, we can conclude the following from all the above models: If the ratio $\frac{D_{\mathrm{TR}}}{D_{\mathrm{JR}}} \geq \eta$ then the communication channel between a transmitter and a receiver is considered to be jammed. Here η is a function of ξ, $P_{J_{\mathrm{T}}}$, P_{T}, G_{TR}, G_{RT}, G_{JR}, G_{RJ} and D_{TR}. Hence if the transmitter is not within a disc of radius ηD_{JR} centered around the receiver, then the communication channel is considered to be jammed. We call this disc as the *perception range*. The *perception range* for any node depends on the distance between the jammer and the node. For effective communication between two nodes, each node should be able to transmit as well as receive messages from the other node. Hence two nodes can communicate if they lie in each other's *perception range*.

We will adopt the above jamming and communication model, for the rest of the paper.

13.2.2 System Model

We now describe the kinematic model of the nodes. In our analysis, each node is a UAV. We assume that the UAVs are having a constant altitude flight. This assumption helps to simplify our analysis to a planar case. Referring to Fig. 13.1, the configuration of the ith UAV in the network can be expressed in terms of the variables (x_i, y_i, ϕ_i) in the global coordinate frame. The pair (x_i, y_i) represents the position of a reference point on UAV$_i$ with respect to the origin of the global reference frame and ϕ_i denotes the instantaneous heading of the UAV$_i$ in the global reference frame. Hence the state space for UAV$_i$ is $\mathbf{X}_i \simeq \mathbb{R}^2 \times \mathrm{S}^1$. In our analysis, we assume that the UAVs are a kinematic system and hence the dynamics of the UAVs are not taken into account in the differential equation governing the evolution of the system. The kinematics of the UAVs are assumed to be the following:

$$\dot{\mathbf{X}}_i := \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} \cos\phi_i \\ \sin\phi_i \\ \sigma_i^1 \end{bmatrix} =: f_i(\mathbf{x}_i, \sigma_i^1) \tag{13.1}$$

where, σ_i^1 is the angular speed of UAV$_i$. We assume that $\sigma_i^1 \in \mathcal{U}_i \simeq \{\phi : [0,t] \rightarrow [-1,+1] \mid \phi(\cdot) \quad \text{is measurable}\}$. Since the jammer is also an aerial vehicle we model its kinematics as a UAV. The motion of the jammer is governed by a set of equations similar to (13.1). The configuration of each jammer is expressed in terms of the variables (x_i', y_i', ϕ_i') and the kinematics is given by the following equation:

$$\dot{\mathbf{X}}_i := \begin{bmatrix} \dot{x}_i' \\ \dot{y}_i' \\ \dot{\phi}_i' \end{bmatrix} = \begin{bmatrix} \cos\phi_i' \\ \sin\phi_i' \\ \sigma_i^2 \end{bmatrix} =: f_i'(\mathbf{x}_i', \sigma_i^2)$$

where, $\sigma_i^2 \in \mathcal{U}_i$ as defined earlier. The state space for jammer i is $\mathbf{X}_i' \simeq \mathbb{R}^2 \times \mathrm{S}^1$. The state space of the entire system is $\mathbf{X} = \mathbf{X}_1 \times \cdots \times \mathbf{X}_n \times \mathbf{X}_1' \times \cdots \times \mathbf{X}_m' \simeq \mathbb{R}^{2(n+m)} \times (S^1)^{n+m}$. We use the following notation in order to represent the dynamics of the entire system:

$$\dot{\mathbf{x}} = f(\mathbf{x}, \{\sigma_i^1(\cdot)\}_{i=1}^n, \{\sigma_i^2(\cdot)\}_{i=1}^m) \tag{13.2}$$

where, $\mathbf{x} \in \mathbf{X}, f = [f_1^T, \cdots, f_n^T, f_1'^T, \cdots, f_m'^T]$.

13.2.3 Problem Formulation

From the communication model presented in the previous section, the connectivity of the network of UAVs depends on their position relative to the jammers. Given m UAVs in the network, we define a graph G on m vertices as follows. The vertex corresponding to UAV$_i$ is labeled as i. An edge exists between vertices i and j *iff* there is a communication link between UAV$_i$ and UAV$_j$. We define the communication network to be disconnected when G has more than one component. In this problem, the existence of a communication link between two nodes depends on the relative distance of the two nodes from the jammers. Using the above model for the communication network, we present the following problem statement.

Assume that G is initially connected. The jammers intend to move in order to disconnect the communication network in minimum amount of time possible. The UAVs must move in order to maintain the connectivity of the network for the maximum possible amount of time. We want to compute the motion strategies for the UAVs in the network. Our interest lies in understanding the spatial reconfiguration of the formation so that the jammers can be evaded.

13.3 Optimal Strategies

In this section, we introduce the concept of optimal strategies for the vehicles.

Given the control histories of the vehicles, $\{\sigma_i^1(\cdot)\}_{i=1}^n, \{\sigma_i^2(\cdot)\}_{i=1}^m$, the outcome of the game is denoted by $\pi : \mathbf{X} \times \mathcal{U}_i^{n+m} \to \mathbb{R}$ and is defined as the time of termination of the game:

$$\pi(\mathbf{x}_0, \{\sigma_i^1(\cdot)\}_{i=1}^n, \{\sigma_i^2(\cdot)\}_{i=1}^m) = t_f \tag{13.3}$$

where t_f denotes the time of termination of the game when the players play $(\{\sigma_i^1(\cdot)\}_{i=1}^n, \{\sigma_i^2(\cdot)\}_{i=1}^m)$ starting from an initial point $\mathbf{x}_0 \in \mathbf{X}$. The game terminates when the communication network gets disconnected. The objective of the jammer is to minimize the termination time and the objective of the UAVs is to maximize it.

Since the objective function of the team of UAVs is in conflict with that of the team of jammers, the problem can be formulated as a multi-player zero sum team game. A play $(\{\sigma_i^{1*}(\cdot)\}_{i=1}^n, \{\sigma_i^{2*}(\cdot)\}_{i=1}^m)$ is said to be the optimal for the players if it satisfies the following conditions:

$$\{\sigma_i^{1*}\}_{i=1}^n = \arg \max_{\{\sigma_i^1\}_{i=1}^n} \pi[\mathbf{x}_0, \{\sigma_i^1(\cdot)\}_{i=1}^n, \{\sigma_i^{2*}(\cdot)\}_{i=1}^m]$$

$$\{\sigma_i^{2*}\}_{i=1}^m = \arg \min_{\{\sigma_i^2\}_{i=1}^m} \pi[\mathbf{x}_0, \{\sigma_i^{1*}(\cdot)\}_{i=1}^n, \{\sigma_i^2(\cdot)\}_{i=1}^m]$$

The value of a game, denoted by the function $J : \mathbf{X} \to \mathbb{R}$, can then be defined as follows:

$$J(\mathbf{x}) = \pi[\mathbf{x}_0, \{\sigma_i^{1*}(\cdot)\}_{i=1}^n, \{\sigma_i^{2*}(\cdot)\}_{i=1}^m] \tag{13.4}$$

The value of the game is unique at a point $\mathbf{X}$ in the state-space. An important property satisfied by optimal strategies is the *Nash equilibrium* property. A set of strategies $(\{\widehat{\sigma}_i^1\}_{i=1}^n, \{\widehat{\sigma}_i^2\}_{i=1}^m)$ is said to be in Nash equilibrium if no unilateral deviation in strategy by a player can lead to a better outcome for that player. This can also be expressed mathematically by the following condition:

$$\pi[\mathbf{x}_0, \{\widehat{\sigma}_i^1\}_{i=1}^n, \{\widehat{\sigma}_i^2\}_{i=1}^m] \le \pi[\mathbf{x}_0, \{\widehat{\sigma}_i^1\}_{i=1}^n, \sigma_j^2, \widehat{\sigma}_{-j}^2], \quad \forall j \in \{1, \cdots, m\}$$

$$\pi[\mathbf{x}_0, \sigma_j^1, \widehat{\sigma}_{-j}^1, \{\widehat{\sigma}_i^2\}_{i=1}^n] \le \pi[\mathbf{x}_0, \{\widehat{\sigma}_i^1\}_{i=1}^n, \{\widehat{\sigma}_i^2\}_{i=1}^m], \quad \forall j \in \{1, \cdots, n\} \tag{13.5}$$

In the above expressions σ^1_{-j} is used to represent the controls of all the UAVs except UAV$_j$. Similarly, σ^2_{-j} is used to represent the controls of all the jammers except the jth jammer. From (13.5), we can conclude that there is no motivation for a player to deviate from its equilibrium strategy. In general, there may be multiple sets of strategies for the players that are in Nash equlibrium. Assuming the existence of a value, as defined in (13.4) and the existence of a unique Nash equilibrium, we can conclude that the Nash equilibrium concept of person-by-person optimality given in (13.5) is a necessary condition to be satisfied for the set of optimal strategies for the players and furthermore, computing the set of strategies that are in Nash equilibrium also gives us the set of optimal strategies. In the following analysis, we assume the aforementioned conditions in order to compute the optimal strategies.

The following theorem provides a relation between the optimal strategy of each player and the gradient of the value function, ∇J.

Theorem 13.1. *Assuming that $J(\mathbf{x})$ is a smooth function of $\mathbf{x}$, the optimal strategies $(\{\sigma_i^{1*}\}_{i=1}^n, \{\sigma_i^{2*}\}_{i=1}^m)$ satisfy the following condition:*

$$\begin{aligned} \sigma_i^{1*} &= \operatorname{sign} J_{\phi_i}, \quad i = 1, \cdots, n \\ \sigma_i^{2*} &= -\operatorname{sign} J_{\phi_i'}, \quad i = 1, \cdots, m \end{aligned} \tag{13.6}$$

Proof. The proof essentially follows the two-player version as provided in [19]. Let us consider a play after time t has elapsed from the beginning of the game at which point the players are at a state $\mathbf{x}$. The outcome functional is provided by the following expression:

$$\pi(\mathbf{x}(t), \{\sigma_i^1(\cdot)\}_{i=1}^n, \{\sigma_i^2(\cdot)\}_{i=1}^m) = \int_t^{t+h} \mathrm{d}t + J(\mathbf{x}(t+h))$$

Using Taylor series approximation of J we obtain the following relation:

$$J(\mathbf{x}(t+h)) - J(\mathbf{x}(t)) = J(\mathbf{x}(t) + f(\mathbf{x}(t), \{\sigma_i^1(t)\}_{i=1}^n, \{\sigma_i^2(t)\}_{i=1}^m)h + h\epsilon(h)) - J(\mathbf{x}(t)) \tag{13.7}$$

where $\epsilon(h)$ is a vector with each entry belonging to $o(h)$. Let δ be defined as follows:

$$\delta = f(\mathbf{x}(t), \{\sigma_i^1(t)\}_{i=1}^n, \{\sigma_i^2(t)\}_{i=1}^m)h + h\epsilon(h)$$

Using the Taylor series approximation for J around the point $\mathbf{x}(t)$, we get the following expression for the RHS of (13.7):

$$= \sum_{i=1}^{n+m} \nabla J \cdot \delta + |\delta| o(|\delta|)$$

$$= h\left[\sum_{i=1}^{n} (J_{x_i}\cos\phi_i + J_{y_i}\sin\phi_i + J_{\phi_i}\sigma_i^1) + \sum_{i=1}^{m} (J_{x_i'}\cos\phi_i' + J_{y_i'}\sin\phi_i' + J_{\phi_i'}\sigma_i^2) + \alpha(h)\right]$$

where $\lim_{h\to 0}\alpha(h)=0$.

$$\pi\left(\mathbf{x}(t),\{\sigma_i^1(\cdot)\}_{i=1}^n,\{\sigma_i^2(\cdot)\}_{i=1}^m\right)=J(\mathbf{x}(t))+h\left(1+\sum_{i=1}^n(J_{x_i}\cos\phi_i+J_{y_i}\sin\phi_i+J_{\phi_i}\sigma_i^1)\right.$$
$$\left.+\sum_{i=1}^m\left(J_{x_i'}\cos\phi_i'+J_{y_i'}\sin\phi_i'+J_{\phi_i'}\sigma_i^2\right)+\alpha(h)\right) \tag{13.8}$$

First, let us consider the controls of the jammer. From the Nash property, we can conclude that if $\sigma_j^1=\sigma_j^{1*},\quad \forall j\in[1,\cdots,n]$ and $\sigma_{-i}^2=\sigma_{-i}^{2*}$ then $\sigma_i^2=\sigma_i^{2*}$ minimizes the left hand side of the above equation. Therefore, we can conclude the following:

1. The optimal control satisfies the following condition:

$$\sigma_i^{2*}(t)=\arg\min_{\sigma_i^2} J_{\phi_i'}\sigma_i^2(t) \tag{13.9}$$

   In a similar manner the controls of the UAVs satisfy the following condition:

$$\sigma_i^{1*}(t)=\arg\max_{\sigma_i^1} J_{\phi_i}\sigma_i^1(t) \tag{13.10}$$

2. In the case, when $\sigma_i^j=\sigma_i^{j*}\quad\forall i,\quad j=1,2$ in (13.8), we obtain the following relation:

$$\underbrace{\pi\left(\mathbf{x}(t),\{\sigma_i^{1*}\}_{i=1}^n,\{\sigma_i^{2*}\}_{i=1}^m\right)}_{J(\mathbf{x}(t))} = J(\mathbf{x}(t))+h\left[1+\sum_{i=1}^n(J_{x_i}\cos\phi_i+J_{y_i}\sin\phi_i+J_{\phi_i}\sigma_i^1)\right.$$
$$\left.+\sum_{i=1}^m(J_{x_i'}\cos\phi_i'+J_{y_i'}\sin\phi_i'+J_{\phi_i'}\sigma_i^2)+\alpha(h)\right]$$
$$\Rightarrow h\left[1+\sum_{i=1}^n\left(J_{x_i}\cos\phi_i+J_{y_i}\sin\phi_i+J_{\phi_i}\sigma_i^1\right)\right.$$
$$\left.+\sum_{i=1}^m(J_{x_i'}\cos\phi_i'+J_{y_i'}\sin\phi_i'+J_{\phi_i'}\sigma_i^2)+\alpha(h)\right]=0$$
$$\Rightarrow 1+\sum_{i=1}^n\left(J_{x_i}\cos\phi_i+J_{y_i}\sin\phi_i+J_{\phi_i}\sigma_i^1\right)$$
$$+\sum_{i=1}^m\left(J_{x_i'}\cos\phi_i'+J_{y_i'}\sin\phi_i'+J_{\phi_i'}\sigma_i^2\right)+\alpha(h)=0$$

Taking $\lim_{h\to 0}$ on both sides leads to the following relation:

$$1+\sum_{i=1}^{n}(J_{x_i}\cos\phi_i+J_{y_i}\sin\phi_i+J_{\phi_i}\sigma_i^1)\sum_{i=1}^{m}(J_{x_i'}\cos\phi_i'+J_{y_i'}\sin\phi_i'+J_{\phi_i'}\sigma_i^2)=0 \quad (13.11)$$

(13.9), (13.10) and (13.11) extend the *Isaacs'* conditions that provide the optimal controls for two-player zero-sum differential games to the case of three-player zero-sum differential games. □

From [19], the Hamiltonian of the system is given by the following expression:

$$H\left(\mathrm{x},\nabla J,\{\sigma_i^{1*}\}_{i=1}^{n},\{\sigma_i^{2*}\}_{i=1}^{m}\right)=1+\sum_{i=1}^{n}\left(\mathrm{J}_{x_i}\cos\phi_i+\mathrm{J}_{y_i}\sin\phi_i+\mathrm{J}_{\phi_i}\sigma_i^1\right)$$
$$+\sum_{i=1}^{m}\left(\mathrm{J}_{x_i'}\cos\phi_i'+\mathrm{J}_{y_i'}\sin\phi_i'+\mathrm{J}_{\phi_i'}\sigma_i^2\right)$$

which is the left side of (13.11). Hence, (13.11) can equivalently be expressed as:

$$\mathcal{H}(\mathbf{x},\nabla J,\{\sigma_i^{1*}\}_{i=1}^{n},\{\sigma_i^{2*}\}_{i=1}^{m})=0 \quad (13.12)$$

In addition to the above conditions, the value function also satisfies the PDE given in the following theorem.

Theorem 13.2. *The value function follows the following partial differential equation (PDE) along the optimal trajectory, namely the retrogressive path equation (RPE)*

$$(\mathring{\nabla} J)=\frac{\partial\mathcal{H}}{\partial\mathbf{x}} \quad (13.13)$$

where $^\circ$ *denotes derivative with respect to retrograde time.*

Therefore, the above equation is called the *retrogressive path equations*. The proof of the above theorem follows the same lines as in the case of two-player zero-sum games [19].

Now for our specific problem, the RPE for the vehicles is given by the following set of equations:

- For UAV_i in the formation the following holds true:

$$\mathring{J}_{x_i}=0,\quad \mathring{J}_{y_i}=0$$
$$\mathring{J}_{\phi_i}=-J_{x_i}\sin\phi_i+J_{y_i}\cos\phi_i \quad (13.14)$$

- For the jammer i in the formation the following holds true:

$$\mathring{J}_{x_i'} = 0, \quad \mathring{J}_{y_i'} = 0$$
$$\mathring{J}_{\phi_i'} = -J_{x_i} \sin\phi_i' + J_{y_i} \cos\phi_i' \tag{13.15}$$

In many problems the value functions are not smooth enough to satisfy the Isaacs equations. Many papers have worked around this difficulty, especially Fleming [15, 16], Friedman [17], Elliott and Kalton [11, 12], Krassovski and Subbotin [21], and Subbotin [41]. In [9], the authors present a new notion of "viscosity" solution for Hamilton-Jacobi equations and prove the uniqueness of such solutions in a wide variety of situations. In [24], the author shows that the dynamic programming optimality condition for the value function in differential control theory problems implies that this value function is the viscocity solution of the associated HJB PDE. The foregoing conclusions turn out to extend to differential game theory. In [36], the authors show that in the context of differential games, the dynamic programming optimality conditions imply that the values are viscosity solutions of appropriate partial differential equations. In [13], the authors present a simplification of the previous work. This work is based on the smoothness assumption of the value function.

In the next section, we discuss the computation of the terminal manifold for the game.

13.4 Terminal Conditions

In this section, we present a computation of the termination conditions by modeling the communication network between the UAVs as a graph. In our problem, the connectivity of the network of UAVs depends on their position relative to the jammers. Given m UAVs in the network we define a graph on m vertices, G, as follows. The vertex corresponding to UAV$_i$ is labeled as i. An edge exists between vertices i and j *iff* there is a communication link between UAV$_i$ and UAV$_j$. In this problem, the existence of a communication link between two nodes depends on the relative distances of the two nodes from the jammers. Therefore, the adjacency relation between two nodes is dependent on the state of the system.

Now that we have a mapping from the state of the system to a graph G, we can investigate the connectivity of the communication network from the connectivity of G by defining the following quantity associated with graphs:

- Laplacian of a graph ($\mathcal{L}(G)$) : It is an $m \times m$ matrix with entries given as follows:

  1. $a_{ij} = \begin{cases} -1 & \text{if an edge exists between } i \text{ and } j \\ 0 & \text{if no edge exists between } i \text{ and } j \end{cases}$
  2. $a_{ii} = -\sum_{k=1,k\neq i}^{m} a_{ik}$

The second-smallest eigenvalue of $\mathcal{L}(G)$ is called the *Fiedler values*, denoted as $\lambda_2(\mathcal{L}(G))$. It is also called the algebraic connectivity of G. It has emerged as an important parameter in many systems problems defined over networks. In [14, 28, 42], it has also been shown to be a measure of the stability and robustness of the networked dynamic system. Since this work deals with connectivity maintenance in the presence of malicious intruders, $\lambda_2(\mathcal{L}(G))$ arises as a natural parameter of interest for both players.

Lemma 13.1 ([5]). *: A graph G is connected if and only if* $\lambda_2(\mathcal{L}(G)) > 0$.

Therefore, all disconnected graphs on m vertices belong to the following set:

$$\tilde{G} = \{G | \lambda_2(\mathcal{L}(G)) = 0\}$$

Let $F(\lambda, G) = \det(\mathcal{L}(G) - \lambda I_m)$ where, I_m denotes the identity matrix of dimension $m \times m$.

Theorem 13.3.

$$\tilde{G} = \{G \mid F_\lambda(\lambda, G)|_{\lambda=0} = 0\}$$

where F_λ denotes derivative with respect to λ.

Proof. $\Rightarrow$ The smallest eigenvalue of the Laplacian of any graph is zero, i.e., $\lambda_1(\mathcal{L}(G)) = 0$ [5]. Let G be a disconnected graph. From Lemma 13.1, we can conclude that $\lambda_2(\mathcal{L}(G)) = 0$. Therefore, $\lambda = 0$ is a repeated root of the equation $F(\lambda, G) = 0 \Rightarrow F_\lambda(\lambda, G) \mid_{\lambda=0} = 0$.

$\Leftarrow$ $F(0, G) = 0$ since $\lambda = 0$ is an eigenvalue of the Laplacian matrix. In addition, $F_\lambda(\lambda, G) \mid_{\lambda=0} = 0$. Therefore, $\lambda = 0$ is a multiple root of $F(\lambda, G)$ with algebraic multiplicity greater than 1. Hence $\lambda_2(\mathcal{L}(G)) = 0 \Rightarrow G$ is disconnected. ☐

From the expression for $F(\lambda, G)$, we can conclude the following:

$$\begin{aligned} F_\lambda(\lambda, G) \mid_{\lambda=0} &= \mathrm{tr}\left[\mathrm{adj}\left((\mathcal{L}(G) - \lambda I_m)\frac{\mathrm{d}((\mathcal{L}(G) - \lambda I_m))}{\mathrm{d}\lambda}\right)\right] \mid_{\lambda=0} \\ &= -\mathrm{tr}[\mathrm{adj}(\mathcal{L}(G) - \lambda I_m)] \mid_{\lambda=0} \\ &= -\mathrm{tr}[\mathrm{adj}(\mathcal{L}(G))] \\ &= -\sum_{i=1}^{m} M_{ii} \end{aligned}$$

where M_{ii} is the minor of $\mathcal{L}(G)$ corresponding to the diagonal element in the ith row. Substituting the above relation in Theorem 13.3 leads to the following equation for the variable $\{a_{ij}\}_{i,j=1}^{m}$ at the terminal manifold:

$$\sum_{i=1}^{m} M_{ii} = 0 \tag{13.16}$$

Since $a_{ij} \in [0,1]$ the decision problem of whether there exists a solution to the above equation is NP-complete [20]. Therefore, in the worst case, we have to enumerate all possible combinations of the edge variables and verify the above equation based on the assumption that $\mathcal{P} \neq \mathcal{NP}$. Let us assume that $\{a_{ij}\}_{i,j=1}^{m}$ is the set of edge variables that satisfies the above equation. The set of states of UAV$_i$, UAV$_j$ and jammers that satisfy the constraint $a_{ij} = 0$ or $a_{ij} = 1$ can be given by the following equation that represents a half-space:

$$g_{ij}(x_i, y_i, x_j, y_j, \{x'_k\}_{k=1}^{m}, \{y'_k\}_{k=1}^{m}) \geq 0$$

The set of states of the UAVs and the jammers that represent a disconnected communication network is given by the following expression:

$$\mathcal{R} = \bigcap_{i,j} (g_{ij}(x_i, y_i, x_j, y_j, \{x'_k\}_{k=1}^{m}, \{y'_k\}_{k=1}^{m}) \geq 0)$$

The terminal manifold of the game is given by the boundary of the region $\mathcal{R}$, $\partial\mathcal{R}$. The above expression characterizes the terminal manifold of the game. The value of the game at termination is identically zero. In this analysis, we compute the gradiant of the value at an interior point in a connected component of the terminal manifold. Since the terminal manifold is discontinuous, optimal trajectories emanating from them will give rise to *singular surfaces* [1, 23] which is a topic of ongoing research. Assuming that a single connected component of the terminal manifold, $\mathcal{M}$, is a hypersurface in $\mathbb{R}^{3(n+m)}$ we can parametrize it using $3(m+n)-1$ variables. Therefore, the tangent space at any point on $\mathcal{M}$ has a basis containing $3(m+n)-1$ elements, t_i. Since $J \equiv 0$ along $\mathcal{M}$ we obtain the following set of $3(m+n)-1$ equations:

$$\nabla J^0 \cdot t_i = 0 \quad \forall \quad i \tag{13.17}$$

From (13.12), ∇J^0 must satisfy the following equation:

$$\sum_{i=1}^{m} \left(\mathrm{J}^0_{x_i} \cos\phi_i + \mathrm{J}^0_{y_i} \sin\phi_i + \mathrm{J}^0_{\phi_i} \sigma_i^{1*} \right) + \sum_{i=1}^{n} \left(\mathrm{J}^0_{x'_i} \cos\phi'_i + \mathrm{J}^0_{y'_i} \sin\phi'_i + \mathrm{J}^0_{\phi'_i} \sigma_i^{2*} \right) = -1 \tag{13.18}$$

Given a termination condition, we can compute ∇J^0 from (13.17) and (13.18). This provides the boundary conditions for the RPE presented in the previous section.

In the next section, we present some examples.

13.5 Examples

In this section, we compute the optimal trajectories for the aerial vehicles in two scenarios involving different numbers of UAVs and jammers.

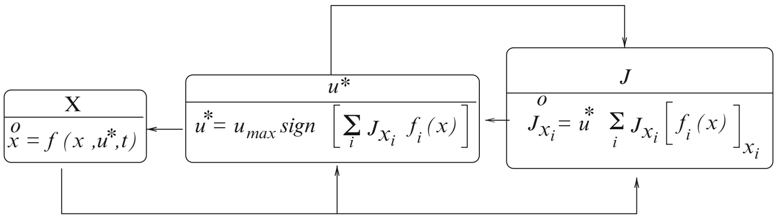


Fig. 13.2 The control loop for the system

13.5.1 $n = 3, m = 1$

First, we consider the case when a single jammer tries to disconnect the communication network in a team containing three UAVs. As stated in the previous section, the cost function of the game is the time of termination. The equations of motion of the vehicles are given as follows:

1. UAVs

$$\dot{x}_i = \cos\phi_i, \quad \dot{y}_i = \sin\phi_i, \quad \dot{\phi} = \sigma_i^1, \quad i = 1,2,3$$

2. Jammer

$$\dot{x}' = \cos\phi', \quad \dot{y}' = \sin\phi', \quad \dot{\phi}' = \sigma^2 \tag{13.19}$$

Under appropriate assumptions on the value function, let $J(\mathbf{x})$ represent the value at the point $\mathbf{x}$ in the state space.

From Theorem 13.1, the expression for the optimal controls can be given as follows:

$$\begin{aligned} \sigma_i^{1*} &= \text{sign}(J_{\phi_i}), \qquad i = 1,2,3 \\ \sigma^{2*} &= -\text{sign}(J_{\phi'}) \end{aligned} \tag{13.20}$$

The RPE for the system leads to the following equations.

$$\begin{aligned} \mathring{J}_{x_i} &= -\sigma_i^{1*} J_{y_i}, \quad \mathring{J}_{y_i} = \sigma_i^{1*} J_{x_i} \\ \mathring{J}_{\phi_i} &= -J_{x_i}\sin\phi_i + J_{y_i}\cos\phi_i \\ \mathring{J}_{x'} &= -\sigma^{2*} J_{y'}, \quad \mathring{J}_{y'} = \sigma^{2*} J_{x'} \\ \mathring{J}_{\phi'} &= -J_{x'}\sin\phi' + J_{y'}\cos\phi' \end{aligned} \tag{13.21}$$

where $^\circ$ denotes the derivative of ∇J with respect to retrograde time.

Figure 13.2 summarizes the control algorithm for each vehicle. The controller of each UAV takes as input the state variables and runs the RPE to compute the control.

Table 13.1 Table shows the value of $F_\lambda(0,G)$ for all possible combinations of (a_{12}, a_{13}, a_{23})

a_{12}	a_{13}	a_{23}	$F_\lambda(0,G)$
0	0	0	0
1	0	0	0
0	1	0	0
0	0	1	0
1	1	0	1
0	1	1	1
1	0	1	1
1	1	1	3

This control is then fed into the plant of the respective UAV. The plant updates the state variables based on the kinematic equations governing the UAV. Finally the sensors feed back the state variables into the controllers. In this case the sensors measure the position and the orientation of each UAV (Table 13.1).

The Laplacian of the graph representing the connectivity of the communication network is given by the following matrix:

$$\mathcal{L}(G) = \begin{bmatrix} -(a_{12}+a_{13}) & a_{12} & a_{13} \\ a_{12} & -(a_{12}+a_{23}) & a_{23} \\ a_{13} & a_{23} & -(a_{13}+a_{23}) \end{bmatrix}$$

From the above form of the Laplacian, we obtain the following expression for $F(\lambda, G)$.

$$F(\lambda,G) = -(\lambda^3 + 2\lambda^2(a_{12}+a_{13}+a_{23}) + 3\lambda(a_{12}a_{13}+a_{13}a_{23}+a_{23}a_{12}))$$

Differentiating the above expression with respect to λ and substituting $\lambda = 0$ provides the following equation:

$$F_\lambda(0,G) = 3(a_{12}a_{23}+a_{13}a_{12}+a_{13}a_{23}) = 0 \tag{13.22}$$

The set of triples (a_{12}, a_{23}, a_{13}) that satisfy Eq. (13.22) are $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ and $(0,0,0)$; see Table 13.1. The first three values represent the situation in which the communication exists only between a pair of UAVs at termination (Fig. 13.3). The last triple represents the scenario in which there is no communication between any pair of UAVs (Fig. 13.4).

Let us consider a termination situation corresponding to the triple $(1,0,0)$. Let D_r^i represent the closed disk of radius r centered at UAV$_i$. Let $\partial\mathcal{R}$ denote the boundary of a region $\mathcal{R}$. From the jamming model, we can infer that the jammer must lie in region $R_1 = (D^3_{\eta r_{31}} \cup D^1_{\eta r_{31}}) \cap (D^3_{\eta r_{32}} \cup D^2_{\eta r_{32}})/(D^1_{\eta r_{12}} \cup D^2_{\eta r_{12}})$. The termination manifold is represented by the hypersurfaces ∂R_1. An example of such a situation is when the jammer is on the boundary, $\partial D^3_{\eta r}$, where $r = \min\{r_{31}, r_{32}\}$. The terminal manifold is characterized by the following equation.

$$\sqrt{(x'-x_3)^2 + (y'-y_3)^2} = r \tag{13.23}$$

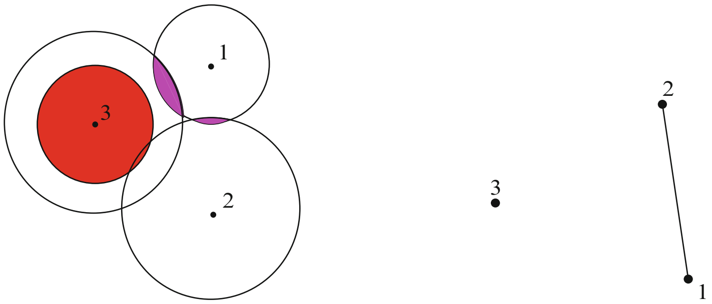


Fig. 13.3 The jammers can lie in the *shaded region* for a network graph of the form shown on the *right hand side*. (**a**) Single-pixel camera developed at Rice University. (**b**) Images taken by a single-pixel camera

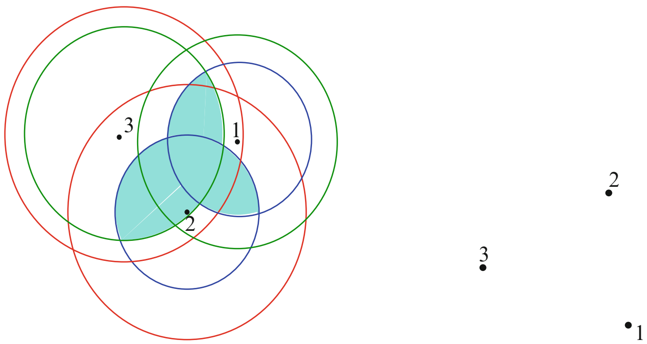


Fig. 13.4 The jammers can lie in the *shaded region* for a network graph of the form shown on the *right hand side*. (**a**) Single-pixel camera developed at Rice University. (**b**) Images taken by a single-pixel camera

This is an 11-dimensional manifold in a 12-dimensional state space. Therefore, we can characterize the manifold by using eleven independent variables. We let $x_1, y_1, \phi_1, x_2, y_2, \phi_2, x_3, y_3, \phi_3, x'$ and y' represent the independent variables. Let J^0 represent the value of the game on the terminal manifold. Since $J^0 \equiv 0$ on the terminal manifold, ∇J^0 satisfies the following equations at an interior point in the manifold:

$$J^0_{x_3} + J^0_{y'}\frac{\partial y'}{\partial x_3} = 0, \quad J^0_{y_3} + J^0_{y'}\frac{\partial y'}{\partial y_3} = 0, \quad J^0_{x'} + J^0_{y'}\frac{\partial y'}{\partial x'} = 0, \quad J^0_{\phi_3} = 0$$
$$\mathfrak{J}^0_{\phi'} = 0, \quad J^0_{x_1} = 0, \quad J^0_{y_1} = 0, \quad J^0_{\phi_1} = 0, \quad J^0_{x_2} = 0, \quad J^0_{y_2} = 0, \quad J^0_{\phi_2} = 0 \quad (13.24)$$

From Eq. (13.16), we can conclude the following:

$$\frac{\partial y'}{\partial x_3} = \frac{x' - x_3}{\sqrt{r^2 - (x' - x_3)^2}}$$

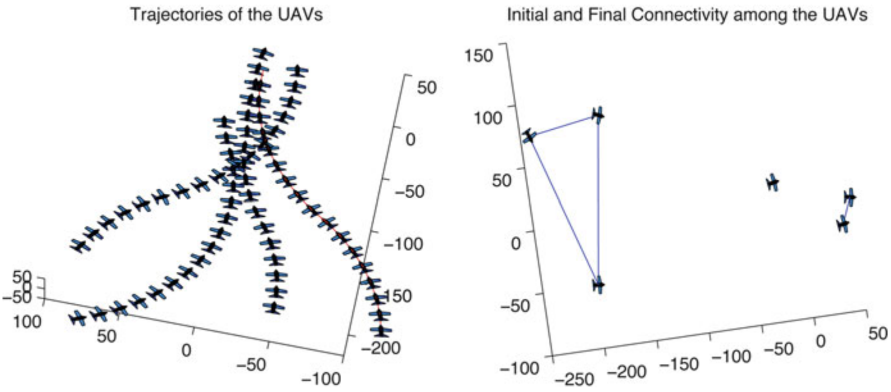


Fig. 13.5 Trajectories of the UAVs and Jammer

$$\frac{\partial y'}{\partial y_3} = 1$$
$$\frac{\partial y'}{\partial x'} = \frac{x' - x_3}{\sqrt{r^2 - (x' - x_3)^2}} \tag{13.25}$$

Substituting the above values of the gradients in Eq. (13.23), we get the following expression for $J^0_{y'}$:

$$J^0_{y'} = \frac{1}{\cos\phi_3 \frac{\partial y'}{\partial x_3} + \sin\phi_3 \frac{\partial y'}{\partial y_3} + \cos\phi' \frac{\partial y'}{\partial x'} - \sin\phi'} \tag{13.26}$$

Substituting (13.18) in (13.26), we obtain the following expression for $J^0_{y'}$:

$$J^0_{y'} = \frac{1}{\frac{x'-x_3}{\sqrt{r^2-(x'-x_3)^2}}(\cos\phi_3 + \cos\phi') + (\sin\phi_3 - \sin\phi')} \tag{13.27}$$

For the triple $(0,0,0)$, the jammer must lie in the region $R_2 = (D^3_{\eta r_{31}} \cup D^1_{\eta r_{31}}) \cap (D^3_{\eta r_{32}} \cup D^2_{\eta r_{32}}) \cap (D^1_{\eta r_{12}} \cup D^2_{\eta r_{12}})$. An analysis similar to the above can be carried out in order to compute the trajectories emanating back from the terminal conditions.

Figure 13.5 shows a simulation of the trajectories of the UAVs and the jammers from a terminal state. The final states (x, y, ϕ) of the three UAVs is given by $(20, -10, 0)$, $(40, 30, 0.14)$ and $(-20, 10, 0.15)$. The final state of the jammer is given by $(50, -10, -0.17)$. The figure on the left shows the trajectories of the UAVs. The jammer traces the path shown on the extreme right. The figure on the right shows the connectivity of the UAVs. The network of UAVs is initally connected. At termination, the jammers disconnect the network by isolating one of the UAVs.

Next, we consider the case when there is a couple of jammers trying to disconnect a team of four UAVs.

13.5.2 $n = 4, m = 2$

As in the earlier section, the equations of motion of the vehicles are given as follows:

1. UAVs

$$\dot{x}_i = \cos\phi_i, \quad \dot{y}_i = \sin\phi_i, \quad \dot{\phi} = \sigma_i^1, \quad i = 1,2,3,4$$

2. Jammer

$$\dot{x}_i' = \cos\phi_i', \quad \dot{y}_i' = \sin\phi_i', \quad \dot{\phi}_i' = \sigma_i^2, \quad i = 1,2$$

UAV_i is used to represent the ith UAV in the formation and UAV_i^J is used to represent the ith jammer in the formation. Under appropriate assumptions of the value function as discussed in Sect. 13.3, let $J(\mathbf{x})$ represent the value at the point $\mathbf{x}$ in the state space.

From Theorem 13.1, the expression for the optimal controls can be given as follows:

$$\begin{aligned} \sigma_i^{1*} &= \text{sign}(J_{\phi_i}), \qquad i = 1,2,3 \\ \sigma^{2*} &= -\text{sign}(J_{\phi'}) \end{aligned} \tag{13.28}$$

The RPE for the system leads to the following equations.

$$\begin{aligned} \mathring{J}_{x_i} &= -\sigma_i^{1*} J_{y_i}, \quad \mathring{J}_{y_i} = \sigma_i^{1*} J_{x_i} \\ \mathring{J}_{\phi_i} &= -J_{x_i}\sin\phi_i + J_{y_i}\cos\phi_i \\ \mathring{J}_{x'} &= -\sigma^{2*} J_{y'}, \quad \mathring{J}_{y'} = \sigma^{2*} J_{x'} \\ \mathring{J}_{\phi'} &= -J_{x'}\sin\phi' + J_{y'}\cos\phi' \end{aligned} \tag{13.29}$$

where "$\circ$" denotes derivative with respect to retrograde time. The Laplacian of the adjacency matrix is given by the following:

$$\mathcal{L}(G) = \begin{bmatrix} -(a_{12}+a_{13}+a_{14}) & a_{12} & a_{13} & a_{14} \\ a_{12} & -(a_{12}+a_{23}+a_{24}) & a_{23} & a_{24} \\ a_{13} & a_{23} & -(a_{13}+a_{23}+a_{34}) & a_{34} \\ a_{14} & a_{24} & a_{34} & -(a_{14}+a_{24}+a_{34}) \end{bmatrix}$$

From the above form of the Laplacian, we obtain the following expression for $F(\lambda, \mathbf{x})$.

$$\begin{aligned} F(\lambda, G) = \lambda^4 &+ 2\lambda^3(a_{12}+a_{13}+a_{14}+a_{23}+a_{24}+a_{34}) + \lambda^2(3a_{12}b + 3a_{12}a_{14} \\ &+ 3a_{12}a_{23} + 3a_{12}a_{24} + 4a_{12}a_{34} + 3a_{13}a_{14} + 3a_{13}a_{23} + 4a_{13}a_{24} + 3a_{13}a_{34} \\ &+ 4a_{14}a_{23} + 3a_{14}a_{24} + 3a_{14}a_{34} + 3a_{23}a_{24} + 3a_{23}a_{34} + 3a_{24}a_{34}) \end{aligned}$$

Table 13.2 Table shows the value of $F_\lambda(0.G)$ for all possible combinations of $(a_{12}, a_{13}, a_{23}, a_{24}, a_{34})$

a_{12}	a_{13}	a_{14}	a_{23}	a_{24}	a_{34}	$F_\lambda(0,\mathbf{x})$
0	0	0	0	0	0	0
1	0	0	0	0	0	0
0	1	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	1	0	0
0	0	0	0	0	1	0
1	1	0	0	0	0	0
1	0	1	0	0	0	0
1	0	0	1	0	0	0
1	0	0	0	1	0	0
1	0	0	0	0	1	0
0	1	1	0	0	0	0
0	1	0	1	0	0	0
0	1	0	0	1	0	0
0	1	0	0	0	1	0
0	0	1	1	0	0	0
0	0	1	0	1	0	0
0	0	1	0	0	1	0
0	0	0	1	1	0	0
0	0	0	1	0	1	0
0	0	0	0	1	1	0
0	0	0	1	1	1	0
1	1	1	0	0	1	0
1	0	1	0	1	0	0
1	1	0	1	0	0	0

$$
\begin{aligned}
&+4\lambda(a_{12}a_{13}a_{14}+a_{12}a_{13}a_{24}+a_{12}a_{13}a_{34}+a_{12}a_{14}a_{23}+a_{12}a_{14}a_{34}\\
&+a_{12}a_{23}a_{24}+a_{12}a_{23}a_{34}+a_{12}a_{24}a_{34}+a_{13}a_{14}a_{23}+a_{13}1_{14}a_{24}+a_{13}a_{23}a_{24}\\
&+a_{13}a_{24}a_{34}+a_{14}a_{23}a_{24}+a_{14}a_{23}a_{34}+a_{14}a_{24}a_{34})
\end{aligned}
$$

Evaluating the derivative of the above expression at $\lambda = 0$, we obtain the following equation:

$$
\begin{aligned}
F_\lambda(0,G) = 4(&a_{12}a_{13}a_{14}+a_{12}a_{13}a_{24}+a_{12}a_{13}a_{34}+a_{12}a_{14}a_{23}+a_{12}a_{14}a_{34}\\
&+a_{12}a_{23}a_{24}+a_{12}a_{23}a_{34}+a_{12}a_{24}a_{34}+a_{13}a_{14}a_{23}+a_{13}a_{14}a_{24}\\
&+a_{13}a_{23}a_{24}+a_{13}a_{24}a_{34}+a_{14}a_{23}a_{24}+a_{14}a_{23}a_{34}+a_{14}a_{24}a_{34}) = 0
\end{aligned}
$$

Table 13.2 enumerates all the combinations of $(a_{12}, a_{13}, a_{14}, a_{23}, a_{24}, a_{34})$ for which $F_\lambda(0,G) = 0$. From these values of the edge variables, we can construct all the graphs that are disconnected. Figure 13.6 shows all the equivalence classes of disconnected graphs under the equivalence relation of isomorphism.

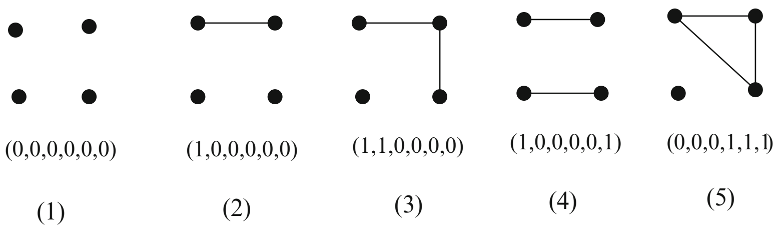


Fig. 13.6 All the equivalence classes of graphs that are disconnected under the equivalence relation of isomorphism

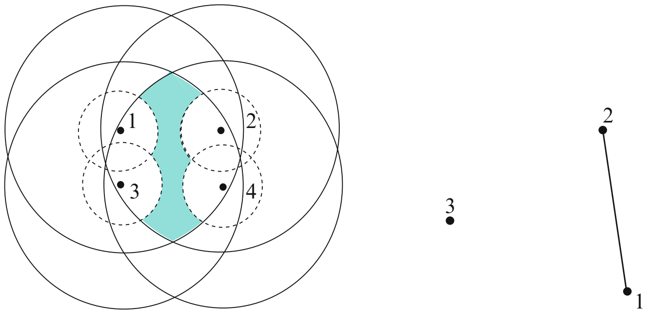


Fig. 13.7 The jammers can lie in the *shaded region* for a network graph of the form shown on the *right hand side*. (**a**) Single-pixel camera developed at Rice University. (**b**) Images taken by a single-pixel camera

Now we consider a situation in which the network graph G has the edge structure as shown in Fig. 13.6(4) at termination time. In order to attain the target network structure, at least one of the jammers has to lie in the shaded region shown in Fig. 13.7. Let $R = (D^1_{\eta r_{12}} \cap D^2_{\eta r_{12}} \cap D^3_{\eta r_{34}} \cap D^4_{\eta r_{34}})/(D^1_{\eta r_{13}} \cup D^2_{\eta r_{24}} \cup D^3_{\eta r_{13}} \cup D^4_{\eta r_{24}})$. Consider a termination situation in which $\text{UAV}^J_1 \in \partial D^2_{\eta r_{12}} \cap \partial R$ and $\text{UAV}^J_2 \notin (D^1_{\eta r_{12}} \cup D^2_{\eta r_{12}} \cup D^3_{\eta r_{34}} \cup D^4_{\eta r_{34}})$. In other words, only UAV^J_1 is responsible for disconnecting the communication network.

The terminal manifold is characterized by the following equation:

$$\sqrt{(x'_1 - x_2)^2 + (y'_1 - y_2)^2} = r \tag{13.30}$$

This is a 17-dimensional manifold in an 18-dimensional state space. Therefore, we can characterize the manifold by using seventeen independent variables. We let $x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x'_2$, y'_2, ϕ_1, ϕ_2, ϕ_3, ϕ_4, ϕ'_1, ϕ'_2 and x'_1 represent the independent variables. Let J^0 represent the value of the game at the termination manifold given by Eq. (13.30). Since $J^0 \equiv 0$ on the terminal manifold, ∇J satisfies the following equations at an interior point in the manifold:

$$J^0_{x_2} + J^0_{y'_1}\frac{\partial y'_1}{\partial x_2} = 0, \quad J^0_{y_2} + J^0_{y'_1}\frac{\partial y'_1}{\partial y_2} = 0, \quad J^0_{x'_1} + J^0_{y'_1}\frac{\partial y'_1}{\partial x'_1} = 0, \quad J^0_{\phi'_2} = 0$$
$$J^0_{\phi_3} = 0, \quad J^0_{\phi'_1} = 0, \quad J^0_{x_1} = 0, \quad J^0_{y_1} = 0, \quad J^0_{\phi_1} = 0, \quad J^0_{x_4} = 0, \quad J^0_{y_4} = 0, \quad J^0_{x'_2} = 0$$
$$J^0_{\phi_2} = 0, \quad J^0_{x_3} = 0, \quad J^0_{y_3} = 0, \quad J^0_{\phi'_2} = 0, \quad J^0_{x_3} = 0, \quad J^0_{y_3} = 0, \quad J^0_{\phi_4} = 0, \quad J^0_{y'_2} = 0 \tag{13.31}$$

From (13.30), we obtain the following expressions for the derivatives of the dependent variable y'_1:

$$\begin{aligned}\frac{\partial y'_1}{\partial x_2} &= \frac{x'_1 - x_2}{\sqrt{r^2 - (x'_1 - x_3)^2}}\\ \frac{\partial y'_1}{\partial y_2} &= 1\\ \frac{\partial y'_1}{\partial x'_1} &= \frac{x'_1 - x_2}{\sqrt{r^2 - (x'_1 - x_2)^2}}\end{aligned} \tag{13.32}$$

Substituting the above values of the gradients in Eq. (13.11), we get the following expression for $J^0_{y'_1}$:

$$J^0_{y'_1} = \frac{1}{\cos\phi_2 \frac{\partial y'_1}{\partial x_2} + \sin\phi_2 \frac{\partial y'_1}{\partial y_2} + \cos\phi'_1 \frac{\partial y'_1}{\partial x'_1} - \sin\phi'_1} \tag{13.33}$$

Substituting (13.32) in (13.33), we obtain the following expression for $J^0_{y'_1}$:

$$J^0_{y'_1} = \frac{1}{\frac{x'_1 - x_2}{\sqrt{r^2-(x'_1-x_2)^2}}(\cos\phi_2 + \cos\phi'_1) + (\sin\phi_2 - \sin\phi'_1)} \tag{13.34}$$

Figure 13.8 shows a simulation of the trajectories of the UAVs and the jammers from a terminal state. The final states (x, y, ϕ) of the four UAVs are given by $(40, 10, 0)$, $(20, 20, 0.14)$, $(20, -20, -0.25)$ and $(40, -10, -0.55)$. The final states of the four jammers are given by $(30, 0, 0.15)$ and $(10, 0, -0.17)$. The figure on the left shows the trajectories of the UAVs. The two UAVs on the extreme right represent the jammers. The figure on the right shows the connectivity of the UAVs. The network of UAVs is initally connected. At termination, the jammers disconnect the network into two disjoint components.

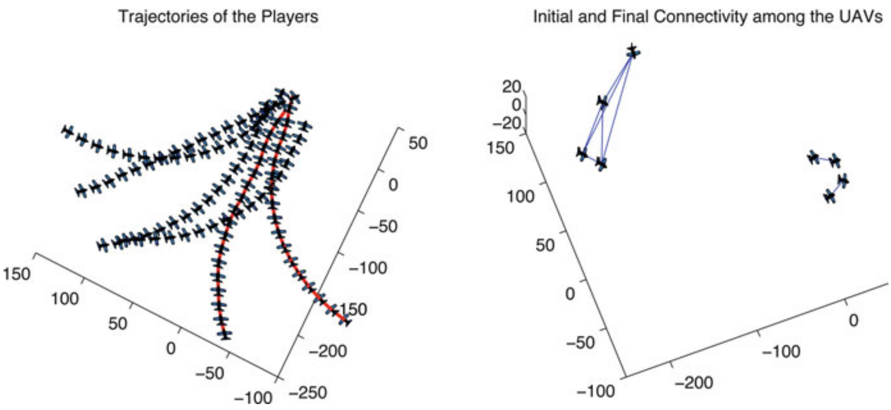


Fig. 13.8 Trajectories of the UAVs and Jammer

13.6 Conclusions

In this work, we have investigated the effect of an aerial jamming attack on the communication network of a team of UAVs flying in a formation. We have introduced a communication and motion model for the UAVs. The communication model provided a relation in the spatial domain for effective jamming by an aerial intruder. We have analysed a scenario in which multiple aerial intruders are trying to jam a communication channel among UAVs flying in a formation. We have formulated the problem of jamming as a zero-sum pursuit-evasion game, and analyzed it in the framework of differential game theory. We have used Isaacs' approach to compute the saddle-point strategies of the UAVs as well as the jammers and used tools from algebraic graph theory to characterize the termination manifold. Finally, we have provided simulation results for two scenarios involving different number of UAVs and jammers.

Future work will involve a simulation study on the construction of globally optimal trajectories. This would also require analysis and construction of singular surfaces.

References

1. Başar, T., Olsder, G.J.: Dynamic Noncooperative Game Theory, 2nd ed. SIAM Series in Classics in Applied Mathematics, Philadelphia (1999)
2. Bhattacharya, S., Başar, T.: Game-theoretic analysis of an aerial jamming attack on a UAV communication network. In: Proceedings of the American Control Conference, pp. 818–823. Baltimore, Maryland (2010)
3. Bhattacharya, S., Başar, T.: Graph-theoretic approach to connectivity maintenance in mobile networks in the presence of a jammer. In: Proceedings of the IEEE Conference on Decision and Control, pp. 3560–3565. Atlanta (2010)

4. Bhattacharya, S., Başar, T.: Optimal strategies to evade jamming in heterogeneous mobile networks. Autonomous Robots, **31**(4), 367–381 (2011)
5. Biggs, N.: Algebraic Graph Theory. Cambridge University Press, Cambridge, U.K., 1993.
6. Breakwell, J.V., Hagedorn, P.: Point capture of two evaders in succession. J. Optim. Theory Appl. **27**, 89–97 (1979)
7. Cagalj, M., Capcun, S., Hubaux, J.P.: Wormhole-based anti-jamming techniques in sensor networks. IEEE Trans. Mobile Comput. **6**, 100–114 (2007)
8. Chen, L.: On selfish and malicious behaviours in wireless networks – a non-cooperative game theoretic approach. PhD Thesis. Ecole Nationale Superieure des Telecommunications, Paris (2008)
9. Crandall, M.G., Lions, P.L.: Viscosity solutions of Hamilton–Jacobi equations. Trans. Am. Math. Soc. **277**(1), 1–42 (1976)
10. Ding, X.C., Rahmani, A., Egerstedt, M.: Optimal multi-UAV convoy protection. Conf. Robot Commun. Config. **9**, 1–6 (2009)
11. Elliott, R.J., Kalton, N.J.: Boundary value problems for nonlinear partial differential operators. J. Math. Anal. Appl. **46**, 228–241 (1974)
12. Elliott, R.J., Kalton, N.J.: Cauchy problems for certain Isaacs–Bellman equations and games of survival. Trans. Am. Math. Soc. **198**, 45–72 (1974)
13. Evans, L.C., Souganidis, P.E.: Differential games and representation formulas for solutions of Hamilton–Jacobi–Isaacs equations. Indiana Univ. Math. J. **33**(5), 773–797 (1984)
14. Fax, J.A., Murray, R.M.: Information flow and cooperative control of vehicle formations. IEEE Trans. Autom. Control **9**, 1465–1474 (2004)
15. Fleming, W.: The Cauchy problem for degenerate parabolic equations. J. Math. Mech. **13**, 987–1008 (1964)
16. Fleming, W.H.: The convergence problem for differential games. Ann. Math. **52**, 195–210 (1964) [Princeton University Press, Princeton]
17. Friedman, A.: Differential Games. Wiley, New York (1971)
18. Hagedorn, P., Breakwell, J.V.: A differential game of approach with two pursuers and one evader. J. Optim. Theory Appl. **18**, 15–29 (1976)
19. Isaacs, R.: Differential Games. Wiley, New York (1965)
20. Karp, R.: Reducibility among combinatorial problems. Complexity of Computer Computations, pp. 85–103. Plenum, New York (1972)
21. Krassovski, N., Subbottin, A.: Jeux Differentiels. Mir Press, Moscow (1977)
22. Levchenkov, A.Y., Pashkov, A.G.: Differential game of optimal approach of two inertial pursuers to a noninertial evader. J. Optim. Theory Appl. **65**, 501–518 (1990)
23. Lewin, J.: Differential Games: Theory and Methods for Solving Game Problems with Singular Surfaces. Springer, London (1994)
24. Lions, P.L.: Generalized Solutions of Hamilton–Jacobi Equations. Pitman, Boston (1982)
25. Mesbahi, M.: On state-dependent dynamic graphs and their controllability properties. IEEE Trans. Autom. Control **50**, 387–392 (2005)
26. Mitchell, I.M., Tomlin, C.J.: Overapproximating reachable sets by Hamilton–Jacobi projections. J. Sci. Comput. **19**, 323–346 (2003)
27. Noubir, G., Lin, G.: Low power denial of service attacks in data wireless lans and countermeasures, ACM SIGMOBILE Mobile Computing and Communications Review, **7**(3), 29–30 (2003)
28. Olfati-Saber, R., Murray, R.M.: Consensus problems in networks of agents with switching topology and time delay. IEEE Trans. Autom. Control, **49**(9), 1520–1533 (2004)
29. Papadimitratos, P., Haas, Z.J.: Secure routing for mobile ad hoc networks In: Proceedings of SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), San Antonio, Texas, **1**, 27–31 (2002)
30. Pashkov, A.G., Terekhov, S.D.: A differential game of approach with two pursuers and one evader. J. Optim. Theory Appl. **55**, 303–311 (1987)
31. Pavone, M., Savla, K., Frazzoli, E.: Sharing the load: mobile robotic networks in dynamic environments. IEEE Robot. Autom. Mag. **16**, 52–61 (2009)

32. Poisel, R.A.: Modern Communication Jamming Principles and Techniques. Artech, Norwood (2004)
33. Proakis, J.J., Salehi, M.: Digital Communications. McGraw-Hill, New York (2007)
34. Samad, T., Bay, J.S., Godbole, D.: Network-centric systems for military operations in urban terrian: the role of UAVs. Proc. IEEE, **95**(1), 92–107 (2007)
35. Shankaran, S., Stipanović, D., Tomlin, C.: Collision avoidance strategies for a three player game. In: Preprints of the 9th International Symposium on Dynamic Games and Applications. Wroclaw, Poland (2008)
36. Souganidis, P.E.: Approximation schemes for viscosity solutions of Hamilton–Jacobi equations. PhD Thesis. University of Wisconsin-Madison (1983)
37. Starr, A.W., Ho, Y.C.: Nonzero-sum differential games. J. Optim. Theory Appl. **3**, 184–206 (1969)
38. Stipanović, D.M., Hwang, I., Tomlin, C.J.: Computation of an over-approximation of the backward reachable set using subsystem level set functions. In: Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis, **11**, 399–411 (2004)
39. Stipanović, D.M., Shankaran, S., Tomlin, C.: Strategies for agents in multi-player pursuit-evasion games. In: Preprints of the Eleventh International Symposium on Dynamic Games and Applications. Tucson, Arizona (2004)
40. Stipanović, D.M., Melikyan, A.A., Hovakimyan, N.V.: Some sufficient conditions for multi-player pursuit evasion games with continuous and discrete observations. Ann. Int. Soc. Dyn. Games **10**, 133–145 (2009)
41. Subbottin, A.: A generalization of the basic equation of the theory of differential games. Soviet Math. Doklady **22**, 358–362 (1980)
42. Tanner, A.J.H., Jadbabaie, A., Pappas, G.: Flocking in fixed and switching networks. IEEE Trans. Autom. Control **5**, 863–868 (2007)
43. Tisdale, J., Kim, Z., Hedrick, J.: Autonomous UAV path planning and estimation. IEEE Robot. Autom. Mag. **16**, 35–42 (2009)
44. Vaisbord, E.M., Zhukovskiy, V.I.: Introduction to Multi-player Differential Games and Their Applications. Gordon and Breach, New York (1988)
45. Wood, A.D., Stankovic, J.A., Son, S.H.: Jam: a jammed-area mapping service for sensor networks. In: Proceedings of 24th IEEE Real-Time Systems Symposium (RTSS 03), pp. 286–297, December 2003
46. Wood, A.D., Stankovic, J.A., Zhou, G.: Deejam: defeating energy-efficient jamming in IEEE 802.15.4 basedwireless networks. In: Proceedings of 4th Annual IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 07), pp. 60–69 (2007)
47. Wu, X., Wood, T., Trappe, W., Zhang, Y.: Channel surfing and spatial retreats: defenses against wireless denial of service. In: Proceedings of 3rd ACM Workshop on Wireless Security (WiSe 04), pp. 80–89 (2004)
48. Zhukovskiy, V.I., Salukvadze, M.E.: The Vector Valued Maxmin, vol. 193. Academic, San Diego (1994)

Chapter 14
Study of Linear Game with Two Pursuers and One Evader: Different Strength of Pursuers

Sergey A. Ganebny, Sergey S. Kumkov, Stéphane Le Ménec, and Valerii S. Patsko

Abstract The paper deals with a problem of pursuit-evasion with two pursuers and one evader having linear dynamics. The pursuers try to minimize the final miss (an ideal situation is to get exact capture), the evader counteracts them. Results of numerical construction of level sets (Lebesgue sets) of the value function are given. A feedback method for producing optimal control is suggested. The paper includes also numerical simulations of optimal motions of the objects in various situations.

Keywords Game theory • Differential games • Group pursuit-evasion games • Maximal stable bridges • Numerical schemes for differential games

14.1 Introduction

Group pursuit-evasion games (several pursuers and/or several evaders) are studied intensively in the theory of differential games [2, 4, 6, 7, 11, 16, 17, 20].

From a general point of view, a group pursuit-evasion game (without any hierarchy among players) can be often treated as an antagonistic differential game where all pursuers are joined into one player, whose objective is to minimize some functional, and, similarly, all evaders are joined into another player, who is the opponent to the first one. The theory of differential games gives an existence theorem for the value function of such a game. But, usually, any more concrete results

S.A. Ganebny • S.S. Kumkov • V.S. Patsko (✉)
Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia
e-mail: ganebny@imm.uran.ru; sskumk@gmail.com; patsko@imm.uran.ru

S. Le Ménec
EADS/MBDA, Paris, France
e-mail: stephane.le-menec@mbda-systems.com

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_14,

(for example, concerning effective construction of the value function) cannot be obtained. This is due to high dimension of the state vector of the corresponding game and absence of convexity of time sections of level sets (Lebesgue sets) of the value function. Just these reasons can explain why group pursuit-evasion games are very difficult and are usually investigated by means of specific methods and under very strict assumptions.

In this paper, we consider a pursuit-evasion game with two pursuers and one evader. Such a model formulation arises from analysis of an applied problem where two aircrafts (or missiles) intercept another one in the horizontal plane. The peculiarity of the game explored in the paper is that solvability sets (the sets wherefrom the interception can be guaranteed with a miss which is not greater than some given value) and optimal feedback controls of the objects can be built numerically in the framework of a one-to-one antagonistic game. Such an investigation is the aim of this paper.

The paper is based on the problem formulation suggested in [12, 13]. In these works, a case is studied when each pursuer is "stronger" than the evader. In our paper, we research the game without this assumption.

14.2 Formulation of Problem

In Fig. 14.1, one can see a possible initial location of the pursuers and evader when they move towards each other. Also, the evader can move from both pursuers, or from one of them, but towards another pursuer.

Let us assume that the initial velocities are parallel and quite large, and control accelerations affect only lateral components of object velocities. Thus, one can suppose that instants of passages of the evader by each of the pursuers are fixed. Below, we call them termination instants and denote by T_{f1} and T_{f2}, respectively. We consider both cases of equal and different termination instants. The players' controls define the lateral deviations of the evader from the first and second pursuers at the termination instants. Minimum of absolute values of these deviations is called *the resulting miss*. The objective of the pursuers is minimization of the resulting miss, the evader maximizes it. The pursuers generate their controls by a coordinated effort (from one control center).

Fig. 14.1 Schematic initial positions of the pursuers and evader

In the relative linearized system, the dynamics is the following (see [12, 13]):

$$\begin{aligned} \ddot{y}_1 &= -a_{P1} + a_E, & \ddot{y}_2 &= -a_{P2} + a_E, \\ \dot{a}_{P1} &= (A_{P1}u_1 - a_{P1})/l_{P1}, & \dot{a}_{P2} &= (A_{P2}u_2 - a_{P2})/l_{P2}, \\ \dot{a}_E &= (A_E v - a_E)/l_E. \end{aligned} \tag{14.1}$$

Here, y_1 and y_2 are the current lateral deviations of the evader from the first and second pursuers; a_{P1}, a_{P2}, a_E are the lateral accelerations of the pursuers and evader; u_1, u_2, v are the players' command controls; A_{P1}, A_{P2}, A_E are the maximal values of the accelerations; l_{P1}, l_{P2}, l_E are the time constants describing the inertiality of servomechanisms.

The controls have bounded absolute values:

$$|u_1| \le 1, \quad |u_2| \le 1, \quad |v| \le 1.$$

The linearized dynamics of the objects in the problem under consideration is typical (see, for example, [19]).

Consider new coordinates x_1 and x_2 which are the values of y_1 and y_2 forecasted to the corresponding termination instants T_{f1} and T_{f2} under zero players' controls. One has

$$x_i = y_i + \dot{y}_i\tau_i - a_{Pi}l_{Pi}^2 h(\tau_i/l_{Pi}) + a_E l_E^2 h(\tau_i/l_E), \quad i = 1,2.$$

Here, x_i, y_i, a_{Pi}, and a_E depend on t, and

$$\tau_i = T_{fi} - t, \quad h(\alpha) = \mathrm{e}^{-\alpha} + \alpha - 1.$$

We have $x_i(T_{fi}) = y_i(T_{fi})$.

Passing to a new dynamics in "equivalent" coordinates x_1 and x_2 (see [12, 13]), we obtain

$$\begin{aligned} \dot{x}_1 &= -A_{P1}l_{P1}h(\tau_1/l_{P1})u_1 + A_E l_E h(\tau_1/l_E)v, \\ \dot{x}_2 &= -A_{P2}l_{P2}h(\tau_2/l_{P2})u_2 + A_E l_E h(\tau_2/l_E)v. \end{aligned} \tag{14.2}$$

Join both pursuers $P1$ and $P2$ into one player which will be called the *first player*. The evader E is the *second player*. The first player governs the controls u_1 and u_2; the second one governs the control v. We introduce the following payoff functional:

$$\varphi\big(x_1(T_{f1}), x_2(T_{f2})\big) = \min\big(|x_1(T_{f1})|, |x_2(T_{f2})|\big). \tag{14.3}$$

It is minimized by the first player and maximized by the second one. Thus, we get a standard antagonistic game with dynamics (14.2) and payoff functional (14.3). This game has [1, 8–10] the value function $V(t,x)$, where $x = (x_1, x_2)$. For each initial position (t_0, x_0), the value $V(t_0, x_0)$ of the value function V equals the payoff guaranteed for the first (second) player by its optimal feedback control. Each level set

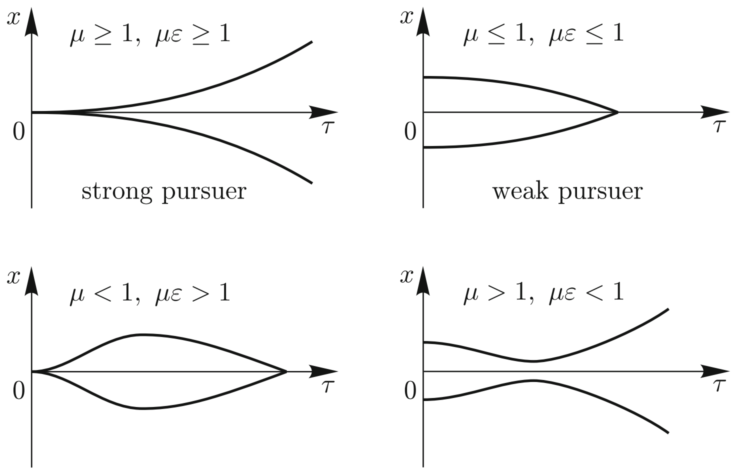


Fig. 14.2 Various variants of the stable bridge evolution in an individual game

$$W_c = \{(t,x) : V(t,x) \leq c\}$$

of the value function coincides with the maximal stable bridge (see [9, 10]) built from the target set

$$M_c = \{(t,x) : t = T_{f1},\ |x_1| \leq c\} \cup \{(t,x) :\ t = T_{f2},\ |x_2| \leq c\}.$$

The set W_c can be treated as the solvability set for the pursuit-evasion game with the result c.

When $c = 0$, we have the situation of the exact capture. The exact capture implies equality to zero of at least one of y_i at the instant T_{fi}, $i = 1, 2$.

The works [12, 13] consider only cases with the exact capture, and pursuers "stronger" than the evader. The latter means that the parameters A_{Pi}, A_E, and l_{Pi}, l_E ($i = 1, 2$) are such that the maximal stable bridges in the individual games ($P1$ vs. E and $P2$ vs. E) grow monotonically in the backward time.

Considering individual games of each pursuer vs. the evader, one can introduce parameters [18] $\mu_i = A_{Pi}/A_E$ and $\varepsilon_i = l_E/l_{Pi}$. They and only they define the structure of the maximal stable bridges in the individual games. Namely, depending on values of μ_i and $\mu_i\varepsilon_i$, there are four cases of the bridge evolution (see Fig. 14.2):

- Expansion in the backward time (a strong pursuer)
- Contraction in the backward time (a weak pursuer)
- Expansion of the bridge until some backward time instant and further contraction
- Contraction of the bridge until some backward time instant and further expansion (if the bridge still has not broken).

Respectively, given combinations of pursuers' capabilities and individual games durations (equal/different), there are significant number of variants for the problem with two pursuers and one evader. Some of them are considered below.

The main objective of this paper is to construct the sets W_c for typical cases of the game under consideration. The difficulty of the problem is that the *time sections* $W_c(t)$ of these sets are non-convex. Constructions are made by means of an algorithm for constructing maximal stable bridges worked out by the authors for problems with two-dimensional state variable. The algorithm is similar to one used in [15]. Another objective is to build optimal feedback controls of the first player (that is, of the pursuers $P1$ and $P2$) and the second one (the evader E).

14.3 Idea of Numerical Method

As it was mentioned above, a level set W_c of the value function is the maximal stable bridge for dynamics (14.2) built in the space t, x from the target set M_c. A time section $W_c(t)$ of the bridge W_c at the instant t is a set in the plane of two-dimensional variable x.

To be definite, let $T_{f1} \geq T_{f2}$. Then for any $t \in (T_{f2}, T_{f1}]$, the set $W_c(t)$ is a vertical strip around the axis x_2. Its width along the axis x_1 equals the width of the bridge in the individual game $P1$ vs. E at the instant $\tau = T_{f1} - t$ of the backward time. At the instant $t = T_{f1}$, the half-width of $W_c(T_{f1})$ is equal to c.

Denote by $W_c(T_{f2}+0)$ the right limit of the set $W_c(t)$ as $t \to T_{f2}+0$. Then the set $W_c(T_{f2})$ is cross-like obtained by union of the vertical strip $W_c(T_{f2}+0)$ and a horizontal one around the axis x_1 with the width equal $2c$ along the axis x_2.

When $t \leq T_{f2}$, the backward construction of the sets $W_c(t)$ is made starting from the set $W_c(T_{f2})$.

The algorithm which is suggested by the authors for constructing the approximating sets $\widetilde{W}_c(t)$, uses a time grid in the interval $[0, T_{f1}]$: $t_N = T_{f1}$, t_{N-1}, $\ldots$, $t_S = T_{f2}$, t_{S-1}, t_{S-2}, $\ldots$. For any instant t_k from the taken grid, the set $\widetilde{W}_c(t_k)$ is built on the basis of the previous set $\widetilde{W}_c(t_{k+1})$ and a dynamics obtained from (14.2) by fixing its value at the instant t_{k+1}. So, dynamics (14.2) which varies in the interval $(t_k, t_{k+1}]$ is changed by a dynamics with simple motions [8]. The set $\widetilde{W}_c(t_k)$ is regarded as a collection of all positions at the instant t_k, wherefrom the first player guarantees guiding the system to the set $\widetilde{W}_c(t_{k+1})$ under "frozen" dynamics (14.2) and discrimination of the second player, that is, when the second player announces its constant control v, $|v| \leq 1$, in the interval $[t_k, t_{k+1}]$.

Due to symmetry of dynamics (14.2) and the sets $W_c(T_{f1})$, $W_c(T_{f2})$ with respect to the origin, one gets that for any $t \leq T_{f1}$ the time section $W_c(t)$ is symmetric also.

Up to now, different workgroups suggested many algorithms for constructing the value function in differential games of quite generic type (see, for example, [3, 5, 14, 21]). The problem under consideration has linear dynamics and the second order on the phase variable. Due to this, we use a specific method. This allows us to make very fast computations of many variants of the game.

14.4 Strong Pursuers, Equal Termination Instants

Add dynamics (14.2) by a "cross-like" target set

$$M_c = \{|x_1| \le c\} \cup \{|x_2| \le c\}, \quad c \ge 0,$$

at the instant $T_f = T_{f1} = T_{f2}$. Then we get a standard linear differential game with fixed termination instant and non-convex target set. The collection $\{W_c\}$ of maximal stable bridges describes the value function of the game (14.2) with payoff functional (14.3).

For the considered case of two stronger pursuers, choose the following parameters:

$$\begin{array}{lll} A_{P1} = 2, & A_{P2} = 3, & A_E = 1, \\ l_{P1} = 1/2, & l_{P2} = 1/0.857, & l_E = 1, \\ T_{f1} = T_{f2} = 6. & & \end{array}$$

1. *Structure of maximal stable bridges.* Figure 14.3 shows results of constructing the set $W = W_0$ (that is, with $c = 0$). In the figure, one can see several time sections $W(t)$ of this set. The bridge has a quite simple structure. At the initial instant $\tau = 0$ of the backward time (when $t = 6$), its section coincides with the

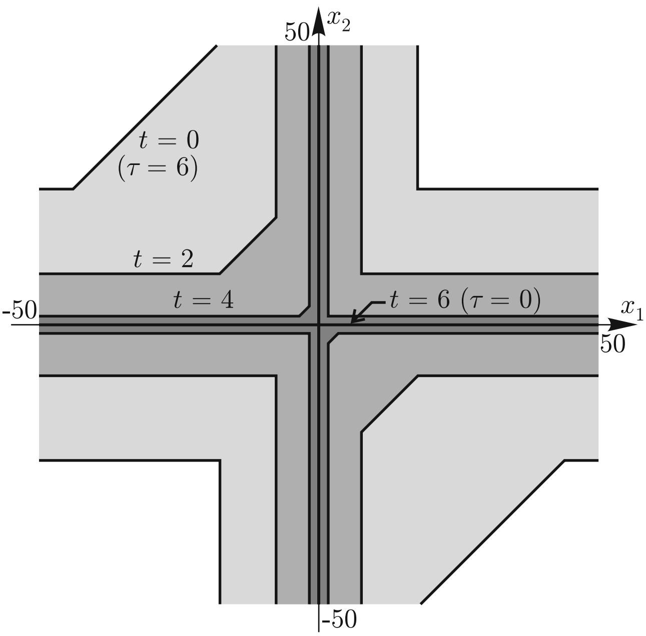


Fig. 14.3 Two strong pursuers, equal termination instants: time sections of the bridge W

target set M_0 which is the union of two coordinate axes. Further, at the instants $t = 4, 2, 0$, the cross thickens, and two triangles are added to it. The widths of the vertical and horizontal parts of the cross correspond to sizes of the maximal stable bridges in the individual games with the first and second pursuers. These triangles are located in the II and IV quadrants (where the signs of x_1 and x_2 are different, in other words, when the evader is between the pursuers) and give the zone where the capture is possible only under collective actions of both pursuers (trying to avoid one of the pursuer, the evader is captured by another one).

These additional triangles have a simple explanation from the point of view of problem (14.1). Their hypotenuses have slope equal to 45°, that is, are described by the equation $|x_1| + |x_2| = \text{const}$. Consider the instant τ when the hypotenuse reaches a point (x_1, x_2). It corresponds to the instant when the pursuers cover together the distance $|x_1(0)| + |x_2(0)|$ which is between them at the initial instant $t = 0$. Therefore, at this instant, both pursuers come to the same point. Since the evader was initially between the pursuers, it is captured at this instant.

The set W (maximal stable bridge) built in the coordinates of system (14.2) coincides with the description of the solvability set obtained analytically in [12, 13]. The solvability set for system (14.1) is defined as follows: if in the current position of system (14.1) at the instant t, the forecasted coordinates x_1, x_2 are inside the time section $W(t)$, then under the controls u_1, u_2 the motion is guided to the target set M_0; on the contrary, if the forecasted coordinates are outside the set $W(t)$, then there is an evader's control v which deviates system (14.2) from the target set. Therefore, there is no exact capture in the original system (14.1).

Time sections $W_c(t)$ of other bridges W_c, $c > 0$, have the shape similar to $W(t)$. In Fig. 14.4, one can see the sections $W_c(t)$ at $t = 2$ ($\tau = 4$) for a collection $\{W_c\}$ corresponding to some series of values of the parameter c. For other instants t, the structure of the sections $W_c(t)$ is similar. The sets $W_c(t)$ describe the value function $x \to V(t,x)$.

2. *Feedback control of the first player.* Rewrite system (14.2) as

$$\dot{x} = \mathcal{D}_1(t)u_1 + \mathcal{D}_2(t)u_2 + \mathcal{E}(t)v,$$
$$|u_1| \le 1,\ |u_2| \le 1,\ |v| \le 1.$$

Here, $x = (x_1, x_2)$; vectors $\mathcal{D}_1(t)$, $\mathcal{D}_2(t)$, and $\mathcal{E}(t)$ look like

$$\mathcal{D}_1(t) = \big(-A_{P1}l_{P1}h((T_{f1}-t)/l_{P1}), 0\big), \quad \mathcal{D}_2(t) = \big(0, -A_{P2}l_{P2}h((T_{f2}-t)/l_{P2})\big),$$
$$\mathcal{E}(t) = \big(A_E l_E h((T_{f1}-t)/l_E), A_E l_E h((T_{f2}-t)/l_E)\big).$$

We see that the vector $\mathcal{D}_1(t)$ ($\mathcal{D}_2(t)$) is directed along the horizontal (vertical) axis; when $T_{f1} = T_{f2}$, the angle between the axis x_1 and the vector $\mathcal{E}(t)$ equals 45°; when $T_{f1} \ne T_{f2}$, the angle changes in time.

Analyzing the change of the value function V along a horizontal line in the plane x_1, x_2 for a fixed instant t, one can conclude that the minimum of the function is reached in the segment of intersection of this line and the set $W(t)$.

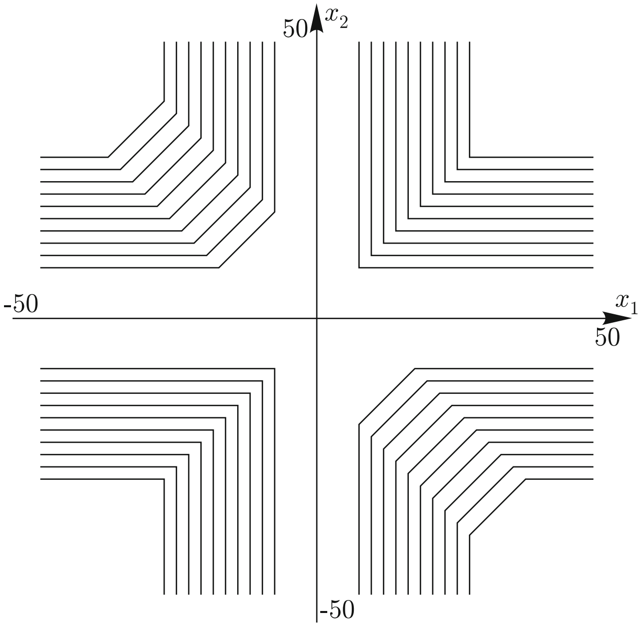


Fig. 14.4 Two strong pursuers, equal termination instants: level sets of the value function, $t = 2$

With that, the function is monotonic at both sides of the segment. For points at the right (at the left) from the segment, the control $u_1 = 1$ ($u_1 = -1$) directs the vector $\mathcal{D}_1(t)u_1$ to the minimum.

Splitting the plane into horizontal lines and extracting for each line the segment of minimum of the value function, one can gather these segments into a set in the plane and draw a *switching line* through this set which separates the plane into two parts at the instant t. At the right from this switching line, we choose the control $u_1 = 1$, and at the left the control is $u_1 = -1$. On the switching line, the control u_1 can be arbitrary obeying the constraint $|u_1| \leq 1$. The easiest way is to take the vertical axis x_2 as the switching line.

In the same way, using the vector $\mathcal{D}_2(t)$, we can conclude that the horizontal axis x_1 can be taken as the switching line for the control u_2.

Thus,

$$u_i^*(t,x) = \begin{cases} 1, & \text{if } x_i > 0, \\ -1, & \text{if } x_i < 0, \\ \text{any } u_i \in [-1,1], & \text{if } x_i = 0. \end{cases} \tag{14.4}$$

The switching lines (the coordinate axes) at any t divide the plane x_1, x_2 into 4 cells. In each of these cells, the optimal control (u_1^*, u_2^*) of the first player is constant.

The vector control $\left(u_1^*(t,x), u_2^*(t,x)\right)$ is applied in a discrete scheme (see [9, 10]) with some time step Δ: a chosen control is kept constant during a time step Δ. Then, on the basis of the new position, a new control is chosen, etc. When $\Delta \to 0$, this control guarantees to the first player a result not greater than $V(t_0, x_0)$ for any initial position (t_0, x_0).

3. *Feedback control of the second player.* Now let us describe the optimal control of the second player. When $T_{f1} = T_{f2}$, the vectogram $\left\{\mathcal{E}(t)v : v \in [-1,1]\right\}$ of the second player in system (14.2) is a segment parallel to the diagonal of I and III quadrants. Thus, the second player can shift the system along this line only.

   Using the sets $W_c(t)$ at some instant t, let us analyze the change of the function $x \to V(t,x)$ along the lines parallel to this diagonal. Consider an arbitrary line from this collection such that it passes through the II quadrant. One can see that local minima are attained at points where the line crosses the axes Ox_1 and Ox_2, and a local maximum is in the segment where the line passes through the rectilinear diagonal part of the boundary of some level set of the value function. The situation is similar for lines passing through the IV quadrant.

   Thus, the switching lines for the second player's control v can be constructed from three parts: the axes Ox_1 and Ox_2, and some slope line $\Pi(t)$. The latter has two half-lines passing through the middles of the diagonal parts on the level set boundaries in the II and IV quadrants. In our case, when the position of the system is on the switching line, the control v can take arbitrary values $|v| \leq 1$. Inside each of 6 cells, to which the plane is divided by the switching lines, the control is taken either $v = +1$, or $v = -1$. Such a control pulls the system towards the points of maximum. Applying this control in a discrete scheme with time step Δ, the second player guarantees that the result will be not less than $V(t_0, x_0)$ for any initial position (t_0, x_0) as $\Delta \to 0$.

   Note. Since $W(t) \neq \varnothing$, then the global minimum of the function $x \to V(t,x)$ is attained at any $x \in W(t)$ and equal 0. Thus, when the position (t,x) of the system is such that $x \in W(t)$, the players can choose, generally speaking, any controls under their constraints. If $x \notin W(t)$, the choices should be made according to the rules described above and based on the switching lines.
4. *Optimal motions.* In Fig. 14.5, one can see the results of optimal motion simulations. This figure contains time sections $W(t)$ (thin solid lines; the same sections as in Fig. 14.3), switching lines $\Pi(0)$ at the initial instant and $\Pi(6)$ at the termination instant of the direct time (dotted lines), and two trajectories for two different initial positions: $\xi_{\mathrm{I}}(t)$ (thick solid line) and $\xi_{\mathrm{II}}(t)$ (dashed line). The motion $\xi_{\mathrm{I}}(t)$ starts from the point $x_1^0 = 40$, $x_2^0 = -25$ (marked by a black circle) which is inside the initial section $W(0)$ of the set W. So, the evader is captured: the endpoint of the motion (also marked by a black circle) is at the origin. The initial point of the motion $\xi_{\mathrm{II}}(t)$ has coordinates $x_1^0 = 25$, $x_2^0 = -50$ (marked by a star). This position is outside the section $W(0)$, and the evader escapes from the exact capture: the endpoint of the motion (also marked by a star) has non-zero coordinates.

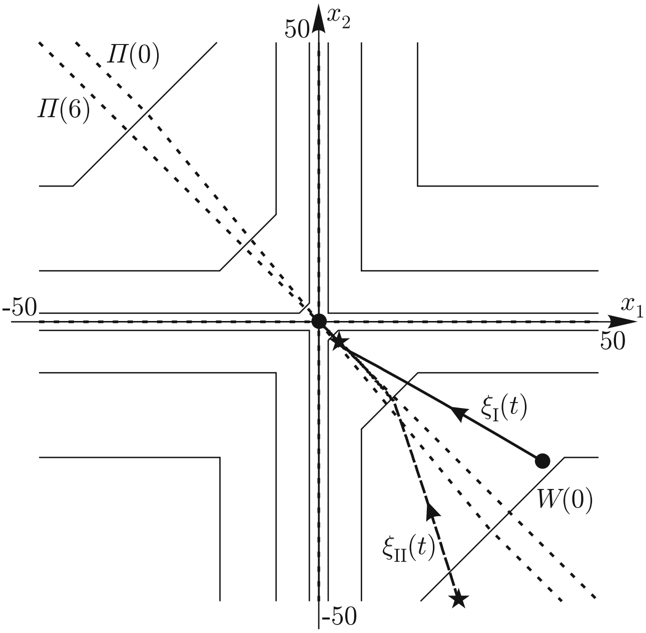


Fig. 14.5 Two strong pursuers, equal termination instants: result of optimal motion simulation

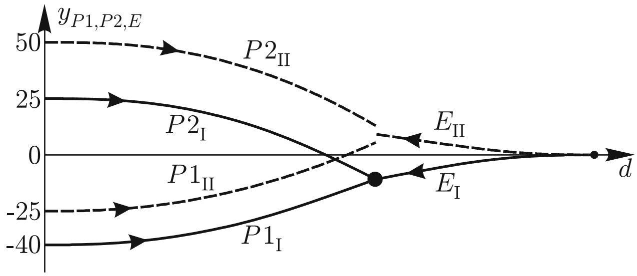


Fig. 14.6 Two strong pursuers, equal termination instants: trajectories in the original space

Figure 14.6 gives the trajectories of the objects in the original space. Values of longitudinal components of the velocities are taken such that the evader moves towards the pursuers. For all simulations here and below, we take

$$y_1^0 = -x_1^0, \quad y_2^0 = -x_2^0, \quad \dot{y}_1^0 = \dot{y}_2^0 = 0, \quad a_{P1}^0 = a_{P2}^0 = a_E^0 = 0.$$

Solid lines correspond to the first case when the evader is successfully captured (at the termination instant, the positions of both pursuers are the same as the position of the evader). Dashed lines show the case when the evader escapes: at the termination instant no one of the pursuers superposes with the evader. In this case, one can see that the evader aims itself to the middle between the terminal positions of the pursuers (this guarantees the maximum of the payoff functional φ).

14.5 Strong Pursuers, Different Termination Instants

Take the parameters as in the previous section, except the termination instants. Now they are $T_{f1} = 7$ and $T_{f2} = 5$. Investigation results are shown in Figs. 14.7–14.9.

The maximal stable bridge $W = W_0$ for system (14.2) with the taken target set

$$M_0 = \{t = T_{f1},\ x_1 = 0\} \cup \{t = T_{f2},\ x_2 = 0\}$$

is built in the following way. At the instant $\tau_1 = 0$ (that is, $t = T_{f1}$), the section of the bridge coincides with the vertical axis $x_1 = 0$. At the instant $\tau_1 = 2$ (that is, $t = T_{f2}$), we add the horizontal axis $x_2 = 0$ to the bridge expanded during passed time period. Further, the time sections of the bridge are constructed using standard procedure under relation $\tau_2 = \tau_1 - 2$.

In the same way, bridges W_c, $c > 0$, corresponding to the target sets

$$M_c = \{t = T_{f1},\ |x_1| \le c\} \cup \{t = T_{f2},\ |x_2| \le c\}$$

can be built.

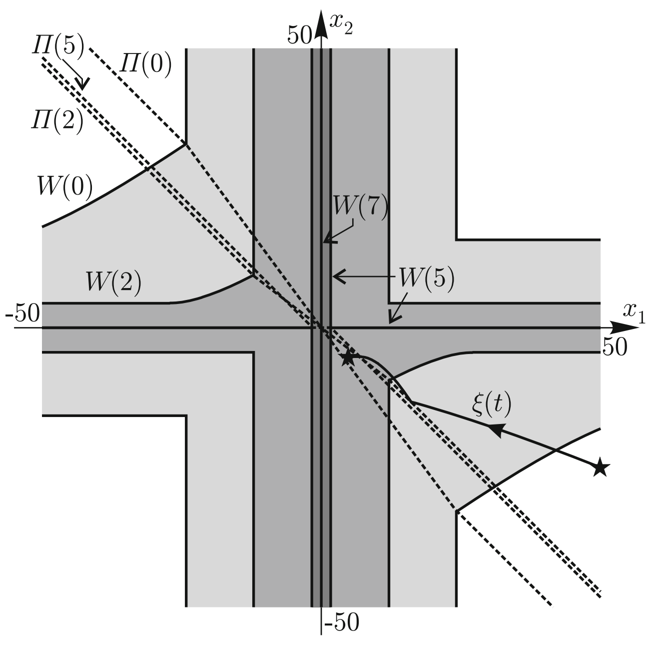


Fig. 14.7 Two strong pursuers, different termination instants: the bridge W and optimal motions

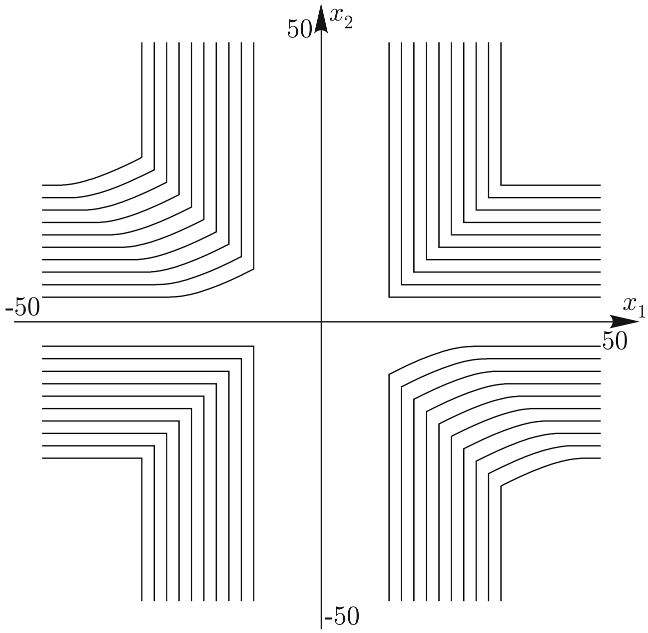


Fig. 14.8 Two strong pursuers, different termination instants: level sets of the value function, $t = 2$

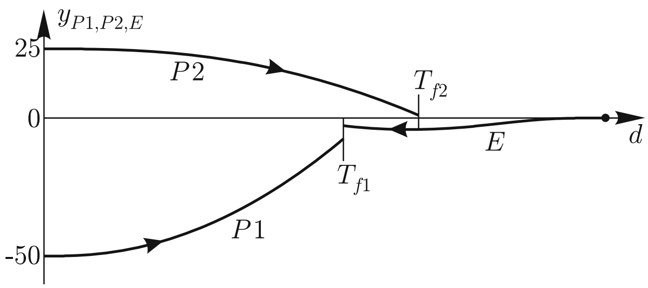


Fig. 14.9 Two strong pursuers, different termination instants: trajectories in the original space

Results of construction of the set W are given in Fig. 14.7. When $\tau_1 > 2$, time sections $W(t)$ grow both horizontally and vertically; two additional triangles appear, but now they are curvilinear. Analytical description of these curvilinear parts of the boundary is difficult. Due to this, in [12, 13], there is only an upper estimation for the solvability set for this variant of the game.

Total structure of the sections $W_c(t)$ at $t = 2$ ($\tau_1 = 5$, $\tau_2 = 3$) is shown in Fig. 14.8. Optimal feedback controls of the pursuers and evader are constructed in the same way as in the previous example, except that the switch line $\Pi(t)$ for the evader is formed by the corner points of the additional curvilinear triangles of the sets $W_c(t)$, $c \geq 0$.

In Fig. 14.7, the trajectory for the initial point $x_1^0 = 50$, $x_2^0 = -25$ is shown as a solid line between two points marked by starts. The trajectories in the original space are shown in Fig. 14.9. One can see that at the beginning the evader escapes from the second pursuer and goes down, after that the evader's control is changed to escape from the first pursuer and the evader goes up.

14.6 Two Weak Pursuers, Different Termination Instants

Now we consider a variant of the game when both pursuers are weaker than the evader. Let us take the parameters

$$A_{P1} = 0.9, \quad A_{P2} = 0.8, \quad A_E = 1, \quad l_{P1} = l_{P2} = 1/0.7, \quad l_E = 1,$$

and different termination instants $T_{f1} = 7$, $T_{f2} = 5$.

Since in this variant, the evader is more maneuverable than the pursuers, they cannot guarantee the exact capture.

Fix some level of the miss, namely, $\left|x_1(T_{f1})\right| \le 2.0$, $\left|x_2(T_{f2})\right| \le 2.0$. Time sections $W_{2.0}(t)$ of the corresponding maximal stable bridge are shown in Fig. 14.10. The upper-left subfigure corresponds to the instant $t = 7$ when the first pursuer stops to act. The upper-right subfigure shows the picture for the instant $t = 5$ when the second pursuer finishes its pursuit. At this instant, the horizontal strip is added which is a little wider than the vertical one contracted during the passed period of the backward time. Then, the bridges contracts both in horizontal and vertical directions, and two additional curvilinear triangles appear (see middle-left subfigure). The middle-right subfigure gives the view of the section when the vertical strip collapses, and the lower-left subfigure shows the configuration just after the collapse of the horizontal strip. At this instant, the section loses connectivity and disjoins into two parts symmetrical with respect to the origin. Further, these parts continue to contract (as it can be seen in the lower-right subfigure) and finally disappear.

Time sections $\{W_c(t)\}$ and corresponding switching lines of the first player are given in Fig. 14.11 at the instant $t = 0$ ($\tau_1 = 7$, $\tau_2 = 5$). The dashed line is the switching line for the component u_1; the dotted one is for the component u_2. The switching lines are obtained as a result of the analysis of the function $x \to V(t,x)$ in horizontal (for u_1) and vertical (for u_2) lines. In some region around the origin, the switching line for u_1 (respectively, for u_2) differs from the vertical (horizontal) axis. If in the considered horizontal (vertical) line the minimum of the value function is attained in a segment, then the middle of such a segment is taken as a point for the switching line. Arrows show directions of components of the control in four cells. Similarly, in Fig. 14.12, switching lines and optimal controls are displayed for the second player. Here, the switching lines are drawn with thick solid lines. We have four cells where the second player's control is constant.

For simulations, let us take the initial position $x_1^0 = 12, x_2^0 = -12$ for system (14.2). In Fig. 14.13, trajectories of the objects are shown in the original space. At the

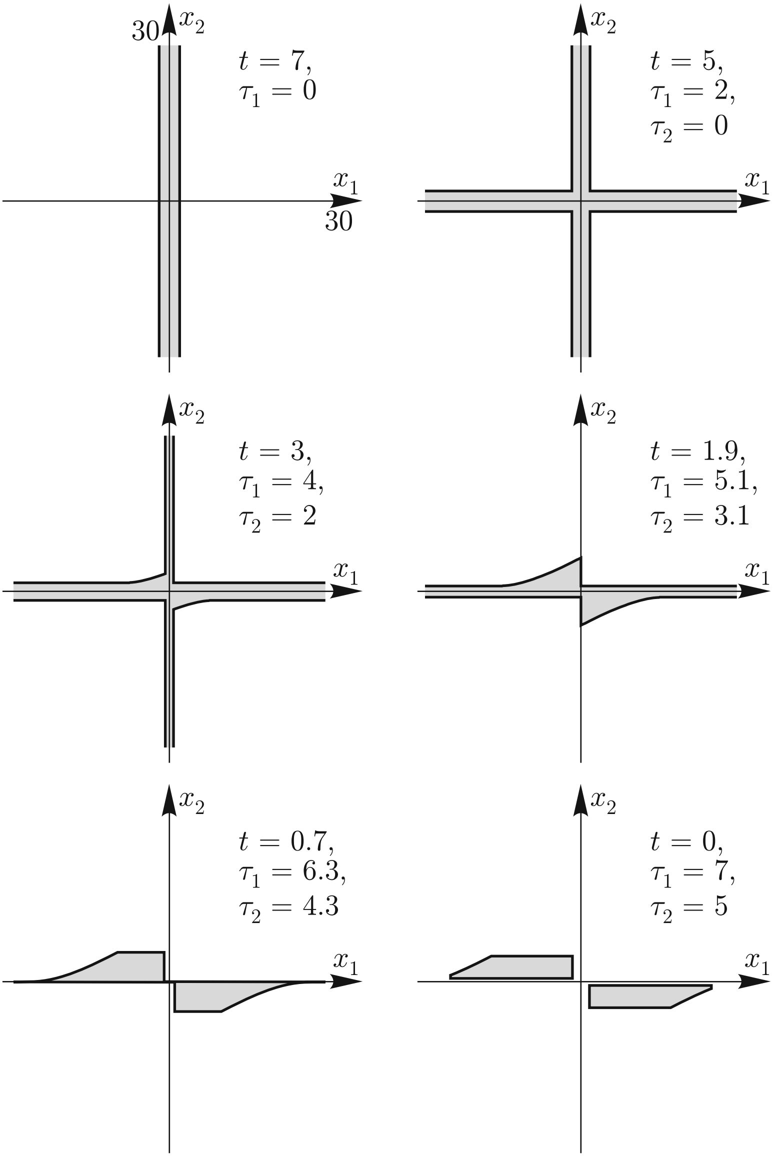


Fig. 14.10 Two weak pursuers, different termination instants: time sections of the maximal stable bridge $W_{2.0}$

beginning of the pursuit, the evader closes to the first (lower) pursuer. It is done to increase the miss from the second (upper) pursuer at the instant T_{f2}. Further closing is not reasonable, and the evader switches its control to increase the miss from the first pursuer at the instant T_{f1}.

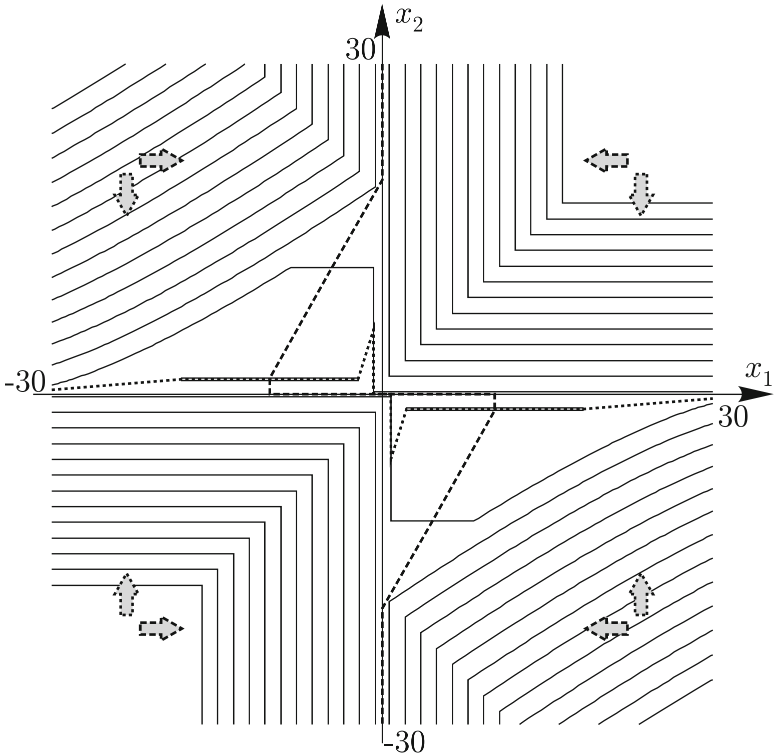


Fig. 14.11 Two weak pursuers, different termination instants: switching lines and optimal controls for the first player (the pursuers), $t = 0$

14.7 One Strong and One Weak Pursuers, Different Termination Instants

Let us change the parameters of the second pursuer in the previous example and take the following parameters of the game:

$$A_{P1} = 2, \quad A_{P2} = 1, \quad A_E = 1, \quad l_{P1} = 1/2, \quad l_{P2} = 1/0.3, \quad l_E = 1.$$

Now the evader is more maneuverable than the second pursuer, and an exact capture by this pursuer is unavailable. Assume $T_{f1} = 5$, $T_{f2} = 7$.

In Fig. 14.14, there are sections of the maximal stable bridge $W_{5.0}$ (that is, for $c = 5.0$) for six instants: $t = 7.0$, 5.0, 2.5, 1.4, 1.0, 0.0. The horizontal part of its time section $W_{5.0}(\tau)$ decreases with growth of τ, and breaks further. The vertical part grows. Even after breaking the individual stable bridge of the second pursuer (and respective collapse of the horizontal part of the cross), additional capture zones still exist and are kept in time.

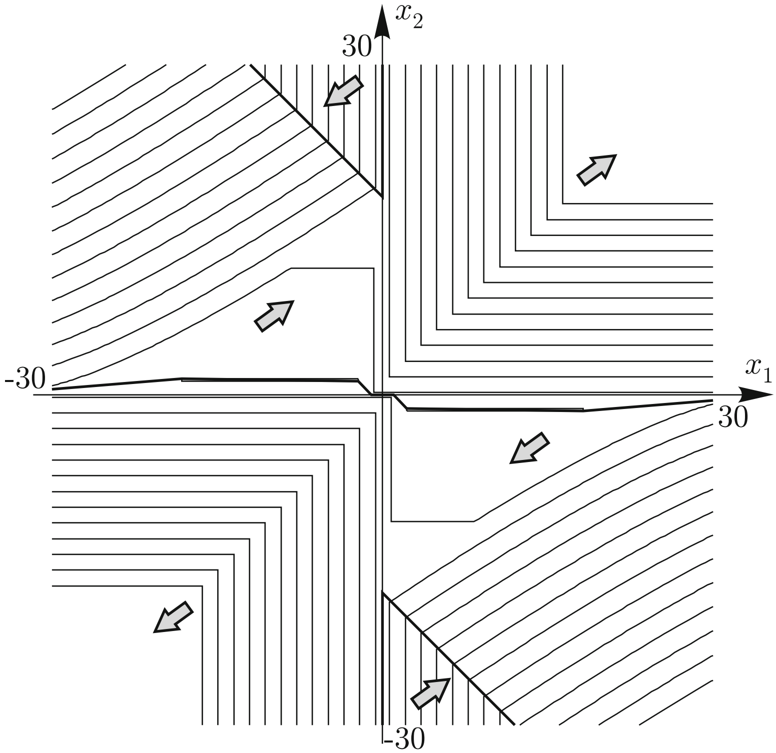


Fig. 14.12 Two weak pursuers, different termination instants: switching lines and optimal controls for the second player (the evader), $t = 0$

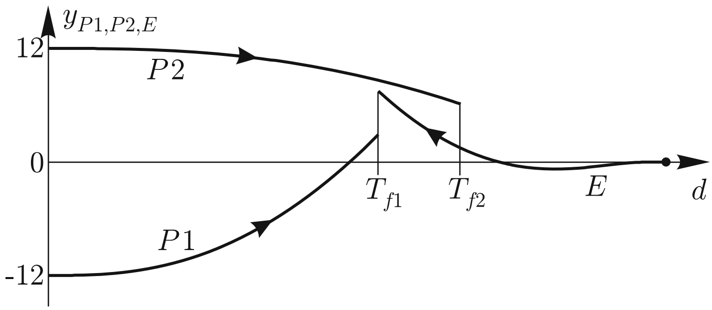


Fig. 14.13 Two weak pursuers, different termination instants: trajectories of the objects in the original space

Switching lines of the first and second players for the instant $t = 1$ are given in Figs. 14.15 and 14.16. These lines are obtained by processing collection $\{W_c(t = 1)\}$ computed for different values of c. In comparison with the previous case of two weak pursuers, the switching lines for the first player have simpler structure.

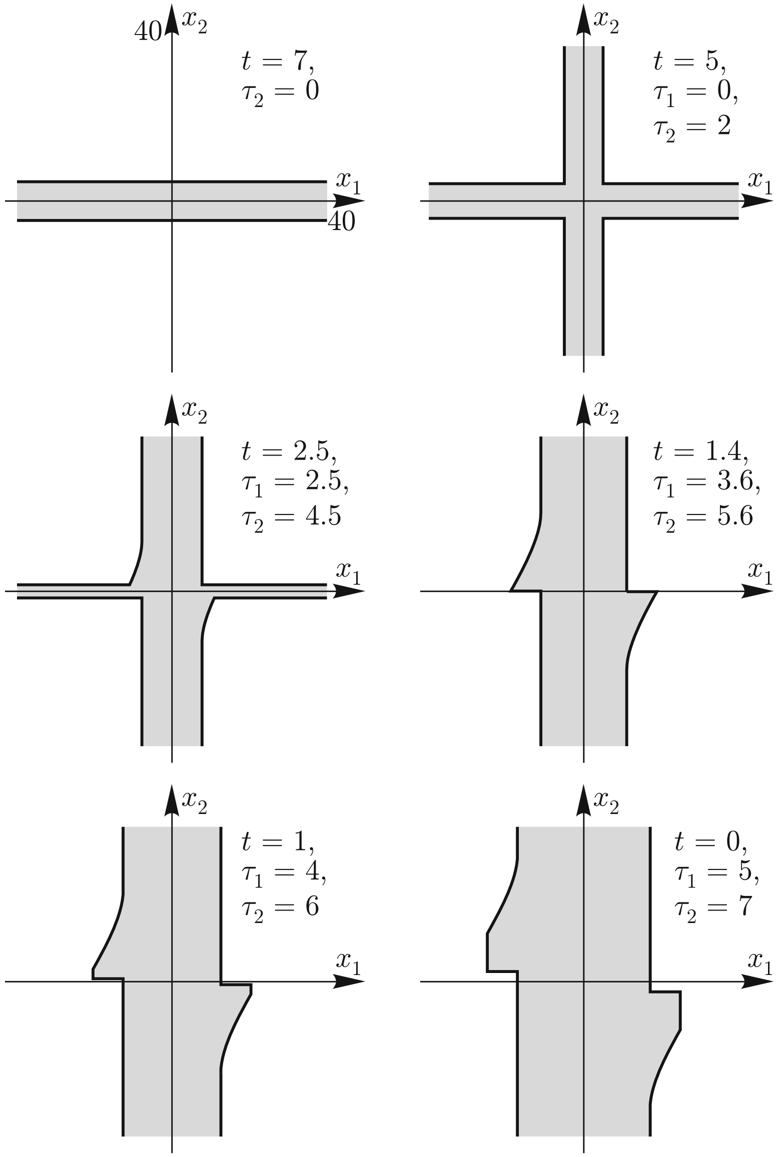


Fig. 14.14 One strong and one weak pursuers, different termination instants: time sections of the maximal stable bridge $W_{5.0}$

Here, as in the previous section, the trajectories of the objects are drawn in the original space only (see Fig. 14.17). For simulations, the initial lateral deviations are taken as $x_1^0 = 20$, $x_2^0 = -20$. Longitudinal components of the velocities are such that the evader moves towards one pursuer, but from another.

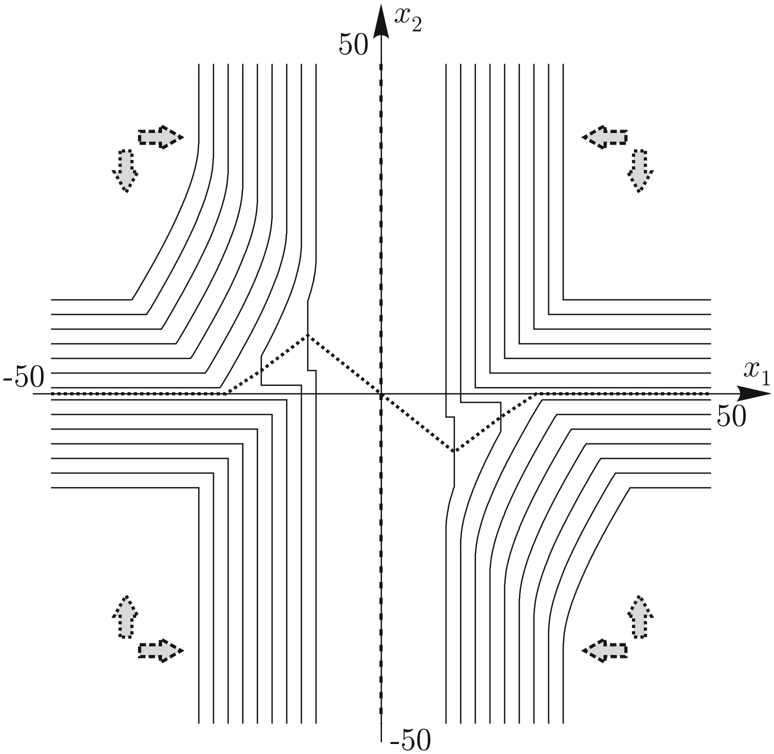


Fig. 14.15 One strong and one weak pursuers, different termination instants: switching lines and optimal controls for the first player (the pursuers), $t = 1$

14.8 Varying Advantage of Pursuers, Equal Termination Instants

Another interesting case is when the pursuers have equal capabilities such that, at the beginning of the backward time, the bridges in the individual games contract and further expand. That is, at the beginning of the direct time, the pursuers have advantage over the evader, but at the final stage the evader is stronger.

Parameters of the game are taken as follows:

$$A_{P1} = A_{P2} = 1.5, \quad A_E = 1, \quad l_{P1} = l_{P2} = 1/0.3, \quad l_E = 1.$$

Termination instants are equal: $T_{f1} = T_{f2} = 10$.

In Fig. 14.18, time sections of the maximal stable bridge $W_{1.5}$ built for $c = 1.5$ are shown for six instants: $t = 10.0, 7.0, 5.7, 4.5, 1.3, 0.0$. At the termination instant, the terminal set is taken as a cross (the upper-left subfigure).

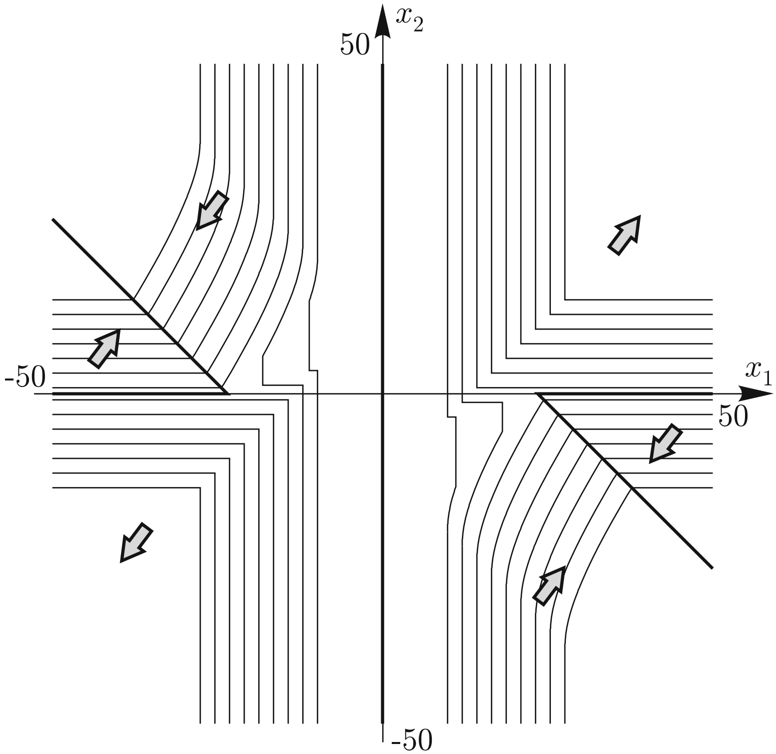


Fig. 14.16 One strong and one weak pursuers, different termination instants: switching lines and optimal controls for the second player (the evader), $t = 1$

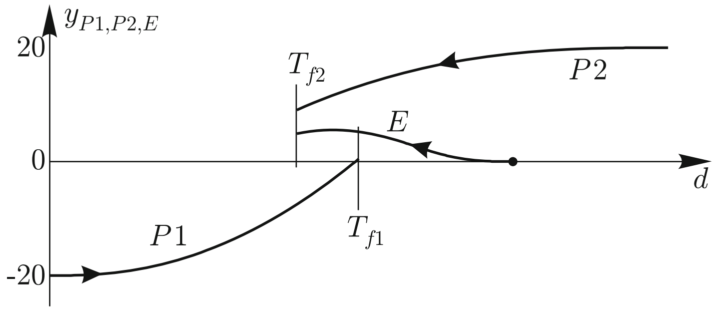


Fig. 14.17 One strong and one weak pursuers, different termination instants: trajectories of the objects in the original space

At the beginning of backward time, the structure of the bridges is similar to the case of two weak pursuers: widths of both vertical and horizontal strips of the "cross" decreases, and two straight-linear additional triangles of joint capture zone appear (the upper-right subfigure). Then at some instant, both strips collapse, and

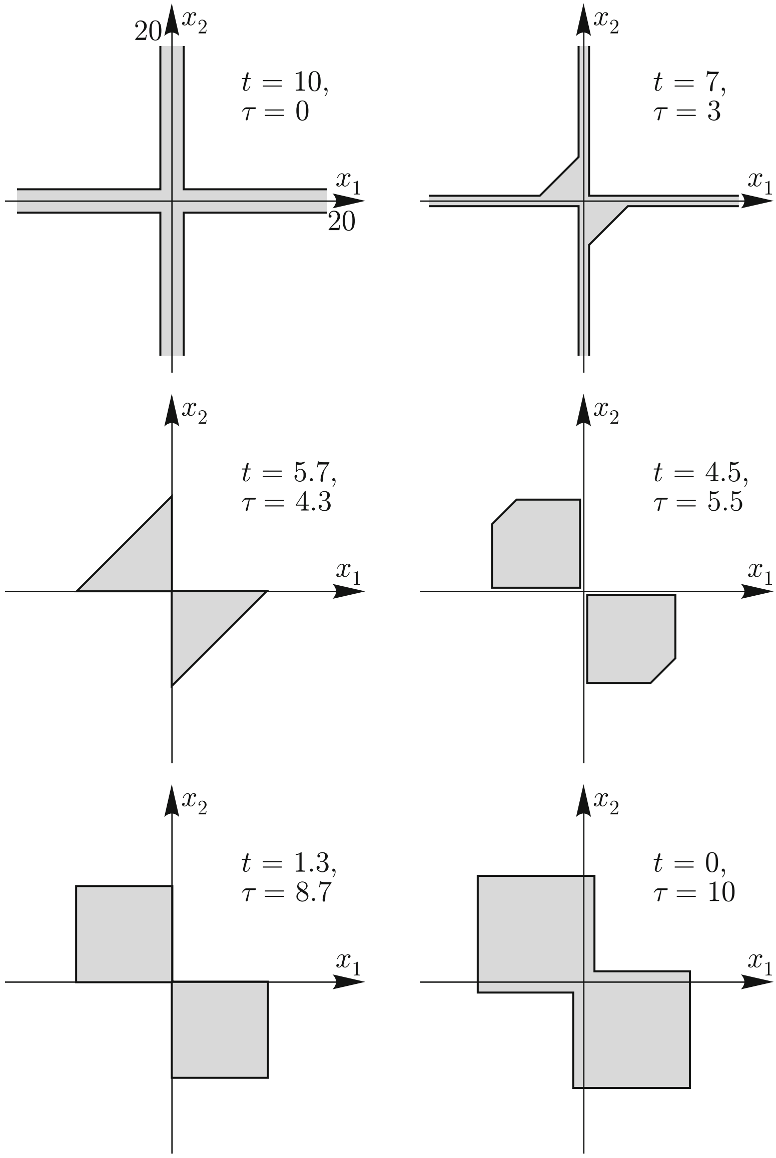


Fig. 14.18 Varying advantage of the pursuers, equal termination instants: time sections of the maximal stable bridge $W_{1.5}$

only the triangles constitute the time section of the bridge (the central left subfigure). Further, the triangles continue to contract, so they become two pentagons separated by an empty space near the origin (the central right subfigure in Fig. 14.18). Transformation to pentagons can be explained in the following way: the first player using

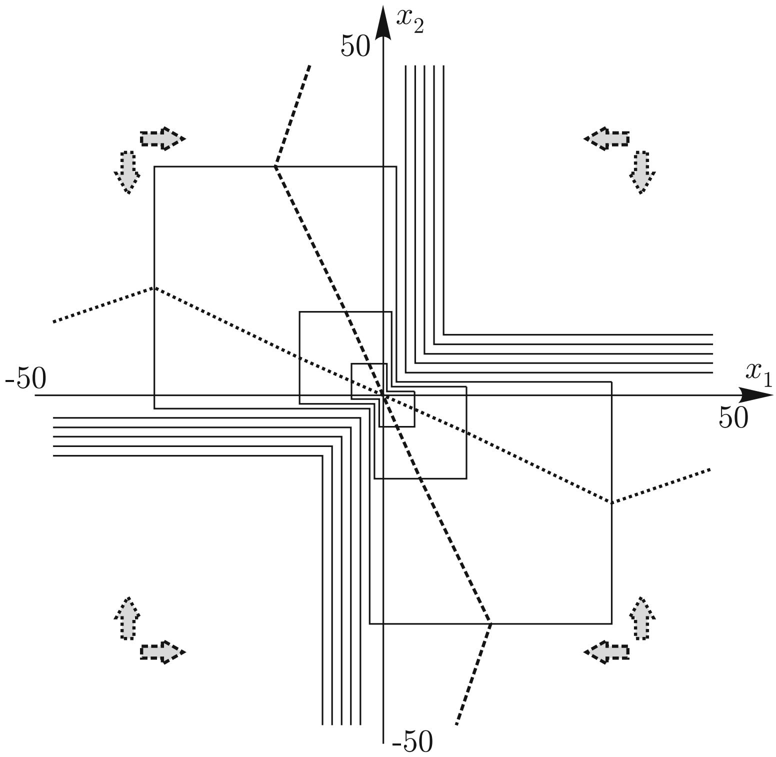


Fig. 14.19 Varying advantage of the pursuers, equal termination instants: switching lines and optimal controls for the first player (the pursuers), $t = 0$

its controls expands the triangles vertically and horizontally, and the second player contracts them in diagonal direction. So, vertical and horizontal edges appear, but the diagonal part becomes shorter. Also, in general, size of each figure decreases slowly.

Due to action of the second player, at some instant, the diagonal disappears, and the pentagons convert to squares (this is not shown in Fig. 14.18). After that, the pursuers take advantage, and total contraction is changed by growth: the squares start to enlarge. When some time passes, due to the growth, the squares touch each other at the origin (the lower-left subfigure in Fig. 14.18). Since the enlargement continues, their sizes grow, and the squares start to overlap forming one “eight-like” shape (the lower-right subfigure in Fig. 14.18).

Figures 14.19 and 14.20 show time sections of a collection of maximal stable bridges and switching lines for the first and second players, respectively, for the instant $t = 0$.

As above, the simulated trajectories are shown in the original space only. For simulation, the following initial conditions are taken: $x_1^0 = 5$, $x_2^0 = -20$. Longitudinal components of the velocities are such that the evader moves from both pursuers.

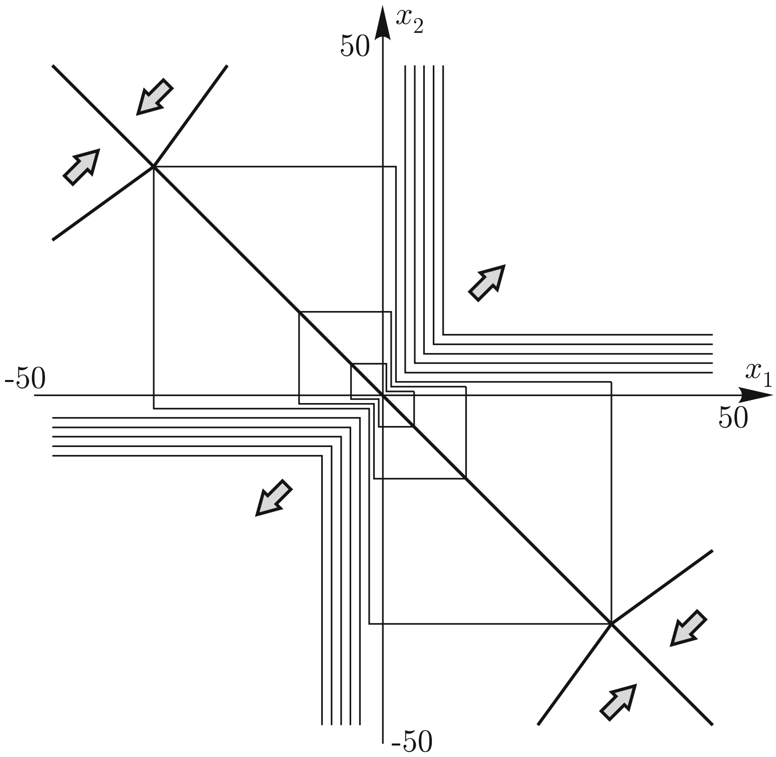


Fig. 14.20 Varying advantage of the pursuers, equal termination instants: switching lines and optimal controls for the second player (the evader), $t = 0$

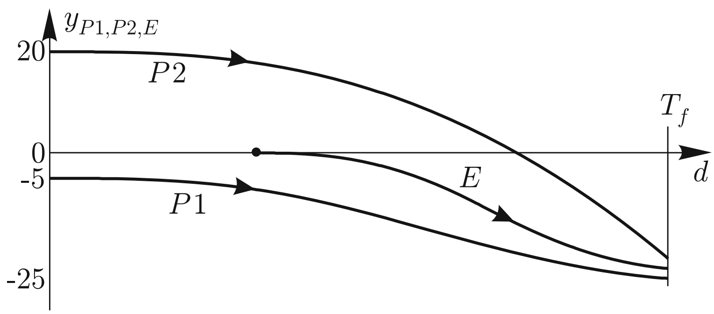


Fig. 14.21 Varying advantage of the pursuers, equal termination instants: trajectories of the objects in the original space

The computed trajectories are given in Fig. 14.21. As it was said earlier, since at the final stage of interception the pursuers are weaker than the evader, they cannot guarantee the exact capture but only some non-zero level of the miss.

14.9 Conclusion

Presence of two pursuers acting together and minimizing the miss from the evader leads to non-convexity of time sections of the value function when the situation is considered as a standard antagonistic differential game where both pursuers are joined into one player. In the paper, results of numerical study of this problem are given for several variants of the parameters. The structure of the solution depends on the presence or absence of dynamic advantage of one or both pursuers over the evader. Optimal feedback control methods of the pursuers and evader are built by preliminary construction and processing the level (Lebesgue) sets of the value function (maximal stable bridges) for some quite fine grid of values of the payoff. Switching lines obtained for each scalar component of controls depend on time, and only they, not the level sets, are used for generating controls. Optimal controls are produced at any current instant depending on the location of the state point respectively to the switching lines at this instant. Accurate proof of the suggested optimal control method needs for some additional study.

Acknowledgements This work was supported by Program of Presidium RAS "Dynamic Systems and Control Theory" under financial support of UrB RAS (project no. 12-*П*-1-1002) and also by the Russian Foundation for Basic Research under grants nos. 10-01-96006 and 11-01-12088.

References

1. Bardi, M., Capuzzo-Dolcetta, I.: Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations. Birkhauser, Boston (1997)
2. Blagodatskih, A.I., Petrov, N.N.: Conflict Interaction Controlled Objects Groups. Udmurt State University, Izhevsk, Russia (2009). (in Russian)
3. Cardaliaguet, P., Quincampoix, M., Saint-Pierre, P.: Set-valued numerical analysis for optimal control and differential games. In: Bardi, M., Raghavan, T.E., Parthasarathy, T. (eds.) Annals of the International Society of Dynamic Games, vol. 4, pp. 177–247. Birkhauser, Boston (1999)
4. Chikrii, A.A.: Conflict-Controlled Processes, Mathematics and its Applications, vol. 405. Kluwer Academic Publishers Group, Dordrecht (1997)
5. Cristiani, E., Falcone, M.: Fully-discrete schemes for the value function of pursuit-evasion games with state constraints. In: Annals of the International Society of Dynamic Games, vol. 10: Advances in Dynamic Games and Applications, pp. 177–206. Birkhauser, Boston (2009)
6. Grigorenko, N.L.: The problem of pursuit by several objects. In: Differential Games—Developments in Modelling and Computation (Espoo, 1990), Lecture Notes in Control and Inform. Sci., vol. 156, pp. 71–80. Springer, Berlin (1991)
7. Hagedorn, P., Breakwell, J.V.: A differential game with two pursuers and one evader. J. Optim. Theory Appl. **18**(2), 15–29 (1976)
8. Isaacs, R.: Differential Games. Wiley, New York (1965)
9. Krasovskii, N.N., Subbotin, A.I.: Positional Differential Games. Nauka, Moscow (1974). (in Russian)
10. Krasovskii, N.N., Subbotin, A.I.: Game-Theoretical Control Problems. Springer-Verlag, New York (1988)
11. Levchenkov, A.Y., Pashkov, A.G.: Differential game of optimal approach of two inertial pursuers to a noninertial evader. J. Optim. Theory Appl. 65, 501–518 (1990)

12. Le Ménec, S.: Linear differential game with two pursuers and one evader. In: Abstracts of 13th International Symposium on Dynamic Games and Applications, pp. 149–151. Wroclaw University of Technology, Wroclaw (2008)
13. Le Ménec, S.: Linear differential game with two pursuers and one evader. In: Annals of the International Society of Dynamic Games, vol. 11: Advances in Dynamic Games and Applications, pp. 209–226. Birkhauser, Boston (2011)
14. Mitchell, I.: Application of level set methods to control and reachability problems in continuous and hybrid systems. PhD Thesis. Stanford University (2002)
15. Patsko, V.S., Turova, V.L.: Level sets of the value function in differential games with the homicidal chauffeur dynamics. Int. Game Theory Rev. **3**(1), 67–112 (2001)
16. Petrosjan L.A.: Differential Games of Pursuit, Leningrad University, Leningrad (1977). (in Russian)
17. Pschenichnyi, B.N.: Simple pursuit by several objects. Kibernetika **3**, 145–146 (1976). (in Russian)
18. Shima, T., Shinar, J.: Time varying linear pursuit-evasion game models with bounded controls. J. Guidance Control Dyn. **25**(3), 425–432 (2002)
19. Shinar, J., Shima, T.: Non-orthodox guidance law development approach for the interception of maneuvering anti-surface missiles. J. Guidance Control Dyn. **25**(4), 658–666 (2002)
20. Stipanovic, D.M., Melikyan, A.A., Hovakimyan, N.: Some sufficient conditions for multiplayer pursuit-evasion games with continuous and discrete observations. In: Bernhard, P., Gaitsgory, V., Pourtallier, O. (eds.) Annals of the International Society of Dynamic Games, vol. 10: Advances in Dynamic Games and Applications, pp. 133–145. Springer, Berlin (2009)
21. Taras'ev, A.M., Tokmantsev, T.B., Uspenskii, A.A., Ushakov, V.N.: On procedures for constructing solutions in differential games on a finite interval of time. J. Math. Sci. **139**(5), 6954–6975 (2006)

Chapter 15
Salvo Enhanced No Escape Zone

Stéphane Le Ménec

Abstract This project proposes an innovative algorithm for designing target allocation strategies and missile guidance laws for air defense applications. This algorithm is localized in the control station of the area to protect; i.e. runs in a centralized manner. It has been optimized according to cooperative principles in a way to increase the defense team performances, which are interception of all threats and interception as soon as possible. Scenarios in naval and ground context have been defined for performance analysis by comparison to a benchmark target allocation policy. The cooperative target allocation algorithm is based on the following features: No Escape Zones (differential game NEZ) computation to characterize the defending missile capturability characteristics; In Flight (re) Allocation (IFA algorithm, late committal guidance) capability to deal with target priority management and pop up threats; capability to generate and counter alternative target assumptions based on concurrent beliefs of future target behaviors, i.e. Salvo Enhanced No Escape Zone (SENEZ) algorithm. The target trajectory generation has been performed using goal oriented trajectory extrapolation techniques. The target allocation procedure is based on minimax strategy computation in matrix games.

Keywords Game theory • Differential games • Minimax techniques • Guidance systems • Co-operative control • Prediction methods • Missiles

15.1 Introduction

This research program has focused on the problem of naval-based air defense systems which must defend against attacks from multiple targets. Modern anti-air warfare systems, capable of tackling the most sophisticated anti ship missiles are

S. Le Ménec (✉)
EADS/MBDA, 1 Avenue Réaumur, 92 358 Le Plessis-Robinson Cedex, France
e-mail: stephane.le-menec@mbda-systems.com

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_15,

based on homing missiles which employ inertial navigation with low frequency or no command update during their mid course phase before becoming autonomous, employing an active seeker for the terminal phase. Technology developments in the field of modular data links may allow the creation of a multi-link communication network to be established between anti-air missiles and the launch platform. The future prospect of such ad hoc networks makes it possible to consider cooperative strategies for missile guidance. Many existing guidance schemes are developed on the basis of one-on-one engagements which are then optimized for many-on-many scenarios [6, 8]. A priori allocation rules and natural missile dispersion can allow a salvo of missiles to engage a swarm of targets; however, this does not always avoid some targets leaking through the salvo, whilst other targets may experience overkill. Cooperative guidance combines a number of guidance technology strands and these have been studied as part of the research program underline:

- Prediction of the target behavior;
- Mid-course guidance to place the missile in position to acquire and engage the target;
- Allocation/re-allocation processes based on estimated target behavior and NEZ;
- Terminal homing guidance to achieve an intercept.

In the terminal phase, guidance has been achieved by handover to the DGL guidance law [16] based on the differential game theory [7]. Two approaches to missile allocation have been considered. The first one relates to Earliest Interception Geometry (EIG) concepts [15]. This article focus on the second one exploiting the NEZ defined by the linear differential game (DGL) guidance law which either acts to define an Allocation Before Launch (ABL) plan or refine an earlier plan to produce an In-Flight Allocation (IFA) plan.

A statement of the problem is given in Sect. 15.2 *SENEZ Concept*. In Sect. 15.6 *Matrix Game Target Allocation Algorithm* details of pre-flight and in-flight allocation planning are described. Missile guidance, both mid-course and terminal, is discussed in Sect. 15.7 *Guidance Logics*. The simulation results from a Simulink 6DoF (6 degree of freedom) model are reviewed in Sect. 15.9 *SENEZ Results*. Sections 15.10 *SENEZ Perspective* and 15.11 *Conclusion* are about the study conclusions and some remarks concerning the exploitation of these cooperative guidance technologies. Finally, Sect. 15.12 *Acronyms and notations* summarizes the meaning of the abbreviations we use and the variables we have in various mathematical formulas.

15.2 SENEZ Concept

There are occasions when the weapon system policy for defending against threats involves firing two or more missiles at the same target. Without any action taken, the missiles will naturally disperse en-route to the target, each arriving at the point of homing with a slightly different geometry. In such a case, there will be

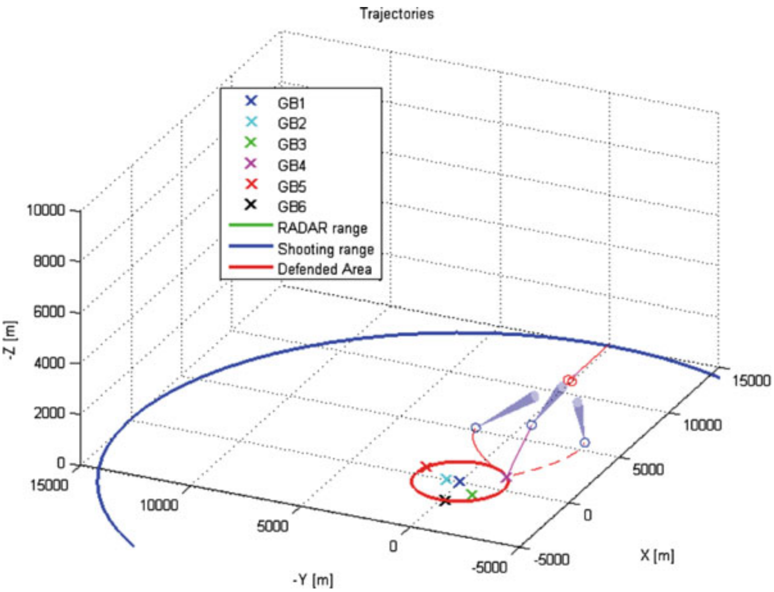


Fig. 15.1 Example of multi-shoot in the SENEZ firing policy; we optimize the management of missile salvos (target allocation process and the guidance laws) to cover uncertainties about the evolution of targets; indeed, at the beginning of the flight the target positions are updated at a low cadence; the missile control systems are provided with target measurements at high data rate only after target acquisition by their on-board sensor; the missile seekers with limited range are depicted by blue cones

a significant overlap of the NEZ. A SENEZ was introduced to optimized this type of engagement, with the cooperating missiles increasing their chances of at least one missile intercepting the target (Fig. 15.1).

In the naval or ground application, it is often the case that a number of assets may be situated in close vicinity to each other. In this situation, it may be difficult to predict which asset an inbound threat is targeting. In the case of air-to-air engagements, there are various break manoeuvres which a target aircraft could execute to avoid an interceptor. These paths can be partitioned into a small number of bundles determined by the number of missiles in the salvo.

By selecting well chosen geometric paths it should be possible to direct the defending missiles in such a way that each partition of the possible target trajectory bundles falls within the NEZ of at least one missile. Consider a naval case of a two missile salvo, and a threat that is initially heading straight towards the launch vessel; there is a possibility that the threat may break left or right at some point. One defending missile can be directed to cover the break right and straight-on possibilities; the second missile would defend against the break left and straight-on possibilities. By guiding to bundle partitions prior to the start of homing, the NEZ of the firing is enhanced. At least one of the missiles will be able to intercept the target. This SENEZ firing policy differs from the more standard shoot–look–shoot policy which considers the sequential firing of missiles where a kill assessment is performed before firing each new missile launch.

15.3 Goal Oriented Target Prediction

Different approaches have been studied to predict target positions [14]. Results detailed in the following are based on the version implementing the goal oriented approach; which is based on the hypothesis that the target will guide to a goal.

The target trajectories have been classified into three categories: threats coming from the left (with respect to the objective), from the front and from the right (Fig. 15.2). We generate these three assumption target trajectories defining one way-point per trajectory class. We compute the trajectories that lead to the threat object passing by the way-points using Trajectory Shaping Guidance (TSG) [17]. The basic TSG is similar to PN (Proportional Navigation) with a constraint on the final Line-Of-Sight (LOS) angle in addition. This means that near impact, the LOS angle λ equals a desired value λ_F. A 3D version of this law is applied from the threat's initial position to the way-point. When the way-point is reached, a switch is made from TSG to standard PN to guide on the objective. The LOS final angle of the TSG law is chosen to bring the threat aiming directly at the objective when it reaches the way-point. Figure 15.3 illustrates how assumption target trajectories have been generated.

A set of three way-points per target is defined using polar parameters (angle ψ_{wpt} and radius R_{wpt}). All way-points belong to a circle of radius R_{wpt} centered on the supposed objective. Way-points are then spread with ψ_{wpt} as an angular gap, using

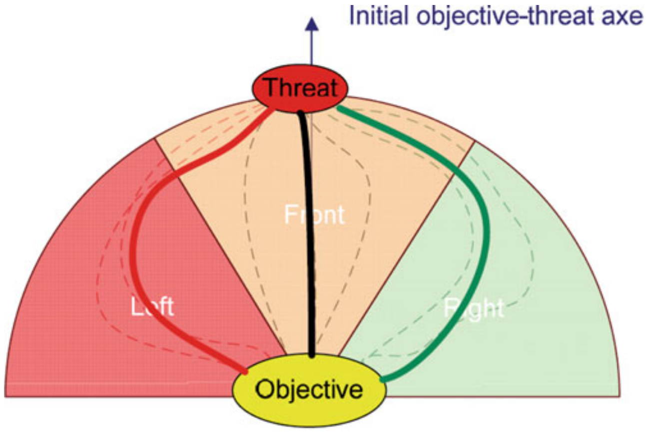


Fig. 15.2 Trajectory classification using three classes

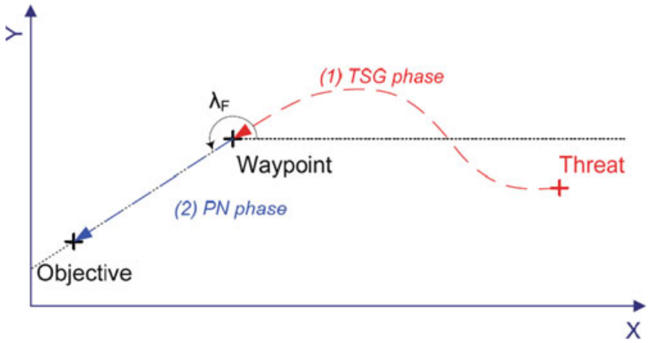


Fig. 15.3 2D target trajectory generation using way-points, TSG and PN as terminal homing guidance

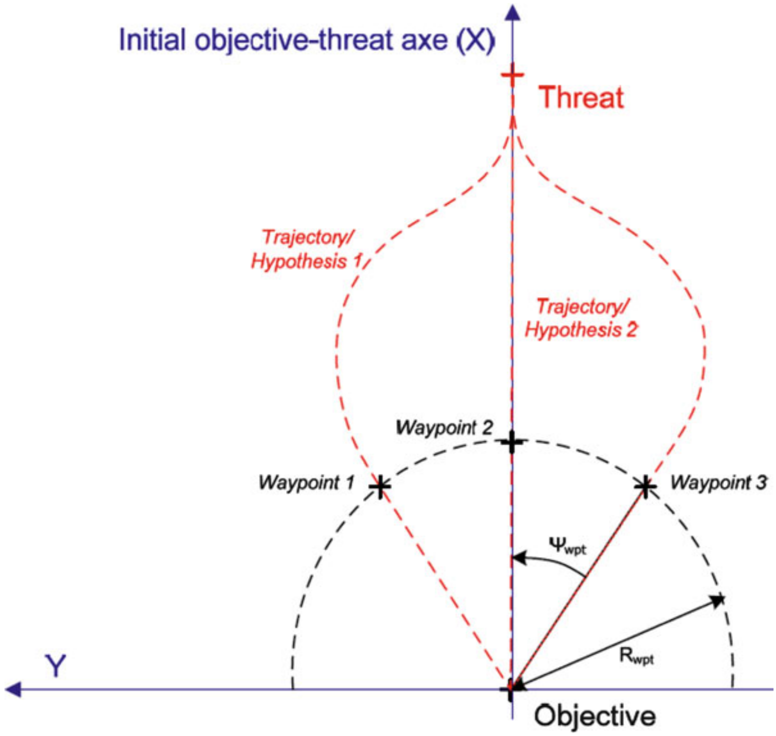


Fig. 15.4 Way-points geometry for target trajectory generation

the initial objective-threat line as a symmetry line defined at RADAR detection. In this way there is one trajectory per hypothesis, as seen in Fig. 15.4.

Way-points are defined for each target depending on its position at the time it is detected. To avoid a high disturbance of the defending missiles guidance, it is

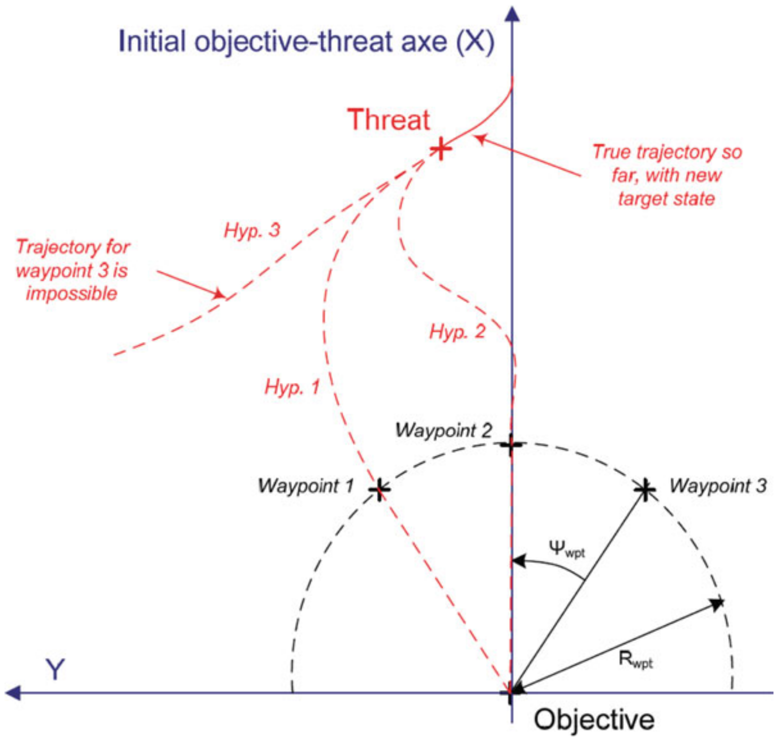


Fig. 15.5 Evolution of the engagement; way-points do not move; way-points trajectories become impossible

assumed that these way-points do not change as the engagement evolves. Some hypotheses will become progressively less likely to be true and others appear to be a good approximation of reality. In due course, some hypotheses will become unachievable and will be discarded during the cost computation process (Fig. 15.5). The SENEZ target allocation algorithm is in charge of evaluating all missile-target-hypothesis engagements [9, 11]. This means the algorithm must be able to tell for each case if successful interceptions are possible and to give a cost on a scale that enables comparisons.

Usage of the following letters is now reserved:

- W is the number of way-points considered;
- N is the number of defensive missile that can be allocated to a target (i.e. that are not already locked on a target, or destroyed);
- P is the number of active and detected threats.

We will now use the following notation to name engagements (i.e. guidance hypotheses):

$$M_i T_j H_k \tag{15.1}$$

Table 15.1 Costs over target trajectory alternatives and defending missile beliefs

		What the target does T_1H_1	T_1H_2	T_1H_3
Guidance hypothesis chosen	$M_1T_1H_1$	$Cost_1$	$Cost_2$	$Cost_3$
	$M_1T_1H_2$	$Cost_4$	$Cost_5$	$Cost_6$
	$M_1T_1H_3$	$Cost_7$	$Cost_8$	$Cost_9$

This means we are talking of the engagement of Missile i ($1 \leq i \leq N$) against Target j ($1 \leq j \leq P$), assuming it is behaving as described by hypothesis k ($1 \leq k \leq W$). Another useful notation is on the other hand:

$$T_j H_k \tag{15.2}$$

This is used to name what the target does (in this case target j is following hypothesis k). Based on the assumption that the target and missile may guide in three different ways H_1, H_2 and H_3, a three by three matrix leading to nine costs can be presented in the following manner (Table 15.1). As the number of missiles and targets increase, the size of the matrix will grow accordingly.

15.4 No Escape Zone

The target allocation algorithms developed during this study require the evaluation of many tentative engagements, considering both various target behavior assumptions and different defending missile assignments. The mid course trajectories are extrapolated using simulation models. However, after seeker acquisition, NEZ are used for the homing engagement kinematic evaluation. The well known classical DGL1 model [16] is used, except that time varying control bounds are considered to account for defending missile drag. Acceleration control bounds have been computed in accordance with 6DoF simulation runs. The time derivative of the standard DGL1 NEZ boundary, as a function of the normalized time to go θ is given by:

$$\frac{\mathrm{d}Z_{\mathrm{limit}}}{\mathrm{d}\theta}(\theta) = \tau_{\mathrm{P}}^2\, a_{\mathrm{E\,max}} \left[\varepsilon_0\, h(\theta) - \mu_0\, h\left(\frac{\theta}{\varepsilon_0}\right)\right] \tag{15.3}$$

$$\theta = \frac{t_{\mathrm{go}}}{\tau_{\mathrm{P}}} \tag{15.4}$$

$$t_{\mathrm{go}} = t_{\mathrm{F}} - t \tag{15.5}$$

$$h(\alpha) = \mathrm{e}^{-\alpha} + \alpha - 1 \tag{15.6}$$

$$\mu_0 = \frac{a_{\text{P max}}}{a_{\text{E max}}} \tag{15.7}$$

$$\varepsilon_0 = \frac{\tau_{\text{E}}}{\tau_{\text{P}}} \tag{15.8}$$

$$Z(\theta) = y + \dot{y}\,\tau - \ddot{y}_{\text{P}}\,\tau_{\text{P}}^2\,h(\theta) + \ddot{y}_{\text{E}}\,\tau_{\text{E}}^2\,h\left(\frac{\theta}{\varepsilon_0}\right) \tag{15.9}$$

where y is the perpendicular miss respect to the initial Line Of Sight (LOS) direction and $\dot{y}$ is the first order time derivative (perpendicular velocity). $\ddot{y}_{\text{P}}$ and $\ddot{y}_{\text{E}}$ are respectively the missile and target components of the acceleration perpendicular to the initial LOS. Z is the ZEM (Zero Effort Miss), which is a well known concept in missile guidance [17]. $a_{\text{P max}}$ is the maximum missile acceleration. $a_{\text{E max}}$ is the maximum target acceleration. τ_{P} and τ_{E} are respectively the pursuer and the evader time constants. μ_0 is the pursuer to evader maneuvering ratio and ε_0 the ratio of time constants (evader to pursuer). t is the regular forward time and t_{F} is the final time; fixed terminal time defined by longitudinal (along the initial LOS) missile target range equal to 0 (see [16] for more details).

By integration in backward time of Eq. (15.3) with initial condition $Z(\theta = 0) = 0$, the NEZ limits can be computed as described by the upper and lower symmetric boundaries of Fig. 15.6. Then, a simple model for the maneuverability has been introduced as a linear function of θ.

$$\mu(\theta) = \mu_0 + \nu\,\theta \qquad (\nu \leq 0) \tag{15.10}$$

The meaning of this equation is that μ which is the ratio of the maximum Pursuer acceleration $a_{\text{P max}}(\theta)$ over the maximum Evader acceleration $a_{\text{E max}}(\theta)$ is increasing as the time to go θ decreases (the missile gets nearer to the target; μ_0 value of μ at $t = t_{\text{F}}$). If thinking about vehicle's manoeuvring drag, then we assume that this phenomenon has more impact on the Evader than on the Pursuer. After integration of this equation, one obtains the new NEZ limits (upper positive limit, the negative one is symmetric):

$$Z_{\text{limit}}(\theta) = Z_{\text{limit}(\mu_0,\varepsilon_0)}(\theta) + \nu\,\tau_{\text{P}}^2\,a_{\text{E max}}\left[\frac{\theta^3}{3} - \frac{\theta^2}{2} - (\theta+1)\,\text{e}^{-\theta} + 1\right] \tag{15.11}$$

The term $Z_{\text{limit}(\mu_0,\varepsilon_0)}$ on the left side is the standard DGL1 bound. The term on the right side is the correcting term we obtain due to the linear variation of μ. When $\nu < 0$; this term actually closes the NEZ at a certain time as shown in Fig. 15.6.

There still exist other refinements for considering non constant velocity profiles using DGL1 kinematics [12]. When running 3D Simulink simulations, the attainability calculus is performed by considering two orthogonal NEZs, associated to the horizontal and to the vertical planes.

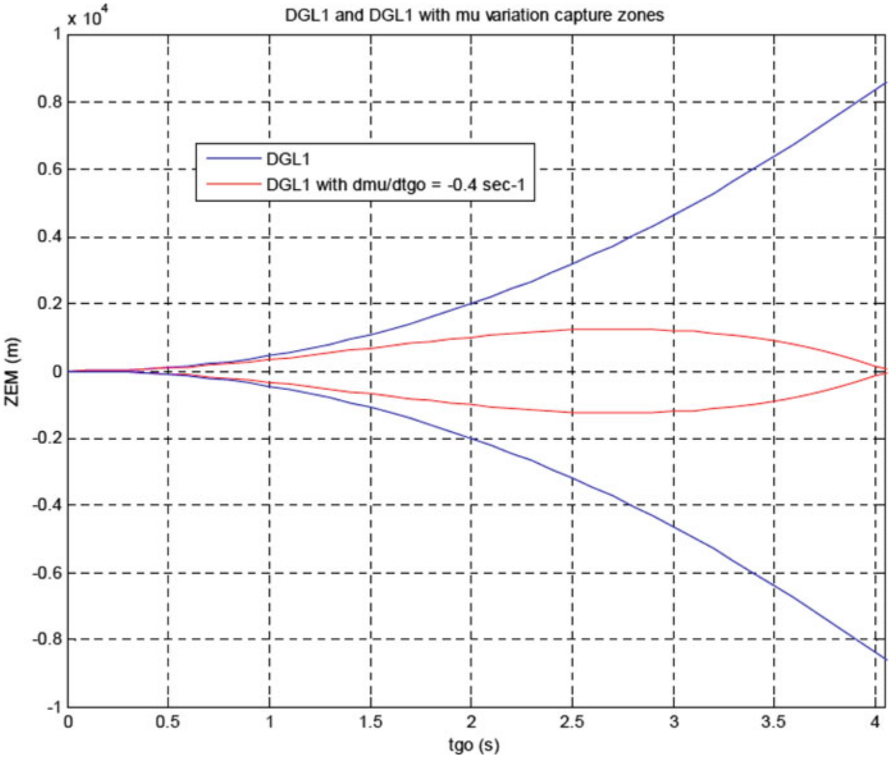


Fig. 15.6 Example of NEZ with variable μ

15.5 Cost Computations

Costs are computed through trajectory extrapolations and NEZ considerations. Target trajectories are extrapolated as explained previously using TSG and PN. Missile trajectories are extrapolated using PN guidance on a Predicted Interception Point (PIP); however, other mid-course guidance laws such as DGGL [13] can be considered. Coordinates of this point are computed using the time to go:

$$t_{\mathrm{go}} = \frac{R_{MT}}{V_c} \tag{15.12}$$

Where R_{MT} is the missile-target distance and V_c the closing velocity. Then, for any time t of the trajectory:

$$XYZ_{\mathrm{PIP}}(t) = XYZ_{\mathrm{T}}\left(t + t_{\mathrm{go}}\right) \tag{15.13}$$

Where XYZ_T are the target coordinates, in inertial frame. The PIP is assumed to have both its velocity and acceleration equal to 0. For every time sample of the target's trajectory, the PIP coordinates are calculated, then the PN command of the missile and finally integrating this command generates the missile states at next sample time. For initial extrapolations, i.e. when missiles are not already in flight, it is assumed that their velocity vector is aimed directly at the way-point of the hypothesis chosen. This is also used in the model when actually shooting missiles. PN on PIP objective makes use of the assumed knowledge of the target's behavior and allows the SENEZ target allocation algorithm to launch several defending missiles against the same real threat following different mid course paths. The SENEZ principle is indeed to shoot multiple missiles to anticipate target's behavior such as doglegs, and new target detections. Once missile trajectories have been computed, the costs are evaluated. The NEZ concept is applied as well as a modeling of the field of view of the missile's seeker. Two zones are defined; the first zone determines if a target can be locked by the seeker (information); the second zone determines if the target can be intercepted (attainability). The cost is simply the relative time when the target enters the intersection of both zones. If it never happens, the cost value is infinite. If the threat is already in both zones at the first sample time, the cost is zero.

When guiding on a hypothesis such as $M_1 T_1 H_1$, it is supposed that the seeker always looks at the predicted position of threat T_1, hypothesis H_1. This gives at every sample time the aiming direction of the seeker. This seeker direction is tested against all other hypotheses to check if a target is within the field of view at this sample time. If positive, an interception test using the NEZ evaluates whether interception is possible. As soon as a target enters the field of view and becomes reachable for a hypothesis, the cost is updated to the trajectory's current time. The cost computation concludes when all costs, i.e. of all hypotheses, have been computed, or when the last trajectory sample has been reached.

This cost logic has been chosen because of the following:

- It takes into account what the missile can or cannot lock on (seeker cone).
- It takes into account the missile's ability to reach the threats (NEZ).
- In most cases, it can be assumed that low costs imply short interception times.

15.6 Matrix Game Target Allocation Algorithm

After costs have been computed, the algorithm has to find the best possible allocation plan. This means we need to construct allocation plans and combine costs. The overall criterion for allocation plan discrimination is about minimizing the time to intercept all the threats which is likely equivalent to maximize the range between the area to protect and the closest location of threat interception. Consider the following illustrative example. One threat T_1 attacks one objective, with three possible hypotheses H_1, H_2 and H_3. Two missiles M_1 and M_2 are allocated to

Table 15.2 Allocation plan cost matrix

	$T_1 H_1$	$T_1 H_2$	$T_1 H_3$	$C_{i,j}$
$M_1 T_1 H_1 - M_2 T_1 H_2$	1.5	5.2	1.0	5.2
$\mathbf{M_1 T_1 H_1 - M_2 T_1 H_3}$	**1.5**	**1.8**	**1.0**	**1.8**
$M_1 T_1 H_2 - M_2 T_1 H_1$	2.1	5.5	1.2	5.5
$M_1 T_1 H_2 - M_2 T_1 H_3$	∞	1.8	1.0	∞
$M_1 T_1 H_3 - M_2 T_1 H_1$	2.1	1.8	1.5	2.1
$M_1 T_1 H_3 - M_2 T_1 H_2$	∞	1.8	1.0	∞

this target. First, it is necessary to determine the possible combinations, excluding options where the two missiles cover the same target hypothesis. We also compute the cost matrix of each missile as described in Sects. 15.3/15.5. Remember that low cost values imply hopefully early interceptions. Infinite values mean interception is not possible.

Using combinations of min max operators we construct the whole problem's cost matrix (Table 15.2) and advice the best one (mini max game equilibrium, [1]). The best allocation plan $(C_{i*,j*})$ is the plan that minimizes the cost value whatever is the target trajectory.

$$\min_{i,j}(C_{i,j}) = \min_{i,j}(\max_{k}(\min(C_{M_1 T_1 H_i | T_1 H_k}, C_{M_2 T_1 Hj | T_1 H_k}))) \tag{15.14}$$

Where i, j are target way-point beliefs defining the defending missile strategies (mid course trajectories) and k is the way-point number defining the threat strategies (trajectories).

The best allocation plan of this simple case is thus $M_1 T_1 H_1 - M_2 T_1 H_3$, $(i^* = 1; j^* = 3)$ which means guiding M_1 based on hypothesis H_1 of T_1 and M_2 on hypothesis H_3 of the same target. By playing this plan, the second hypothesis is covered with a satisfactory cost of 1.8, and no additional missile is needed.

This algorithm could also be used to optimize the number of missile to be involved. i.e. if no satisfactory solution exists, i.e. if the costs are higher than a threshold, the procedure can re-start with an additional missile, three missiles in this case.

The same principle applies when there are more than two missiles, and more than one target (the SENEZ algorithm has been written and evaluated in general scenarios). The mathematical formula for the construction and optimization of allocation plans cost matrix then becomes as follow:

$$\text{find}(A, B) \mid \min_{A,B} C_{A,B} = \min_{A,B}(\max_{i,j}(\min_{k}(C_{M_k T_{A(k)} H_{B(k)} | T_i H_j}))) \tag{15.15}$$

where

- k is the missile number (between 1 and N; maximum number of defending missile).
- $A(k)$ is the index of the allocated target (to missile k).

- $B(k)$ is the index of the hypothesis used for target $A(k)$.
- i $(1 \le i \le T)$ and j $(1 \le j \le W)$ so that T_i is an incoming target and H_j one of the possible hypotheses.

Obviously, when looking for the maximum (in the previous formulae), one scans all possible $T_i H_j$. An A, B vector pair represents one allocation plan. To be valid, one allocation plan must comply with the following constraints:

- All incoming targets should appear at least one time in A.
- A target-hypothesis (target number/way-point number) cannot appear more than one time per allocation plan.

The algorithm has then to find among all possible plans (A, B combinations), the plan that minimizes $C_{A,B}$. By defining heuristics, it is possible to prune potential allocation plans and to focus the algorithm on the most promising solutions (A^*, Dijkstra algorithms). This kind of algorithm has been tested in scenarios involving a large number of threats; i.e. saturating attacks [13].

15.7 Guidance Logics

The two diagrams Figs. 15.7 and 15.8 summarize the defending guidance phases (mid-course, homing phase) and explain how the 6DoF Simulink Common Model operates.

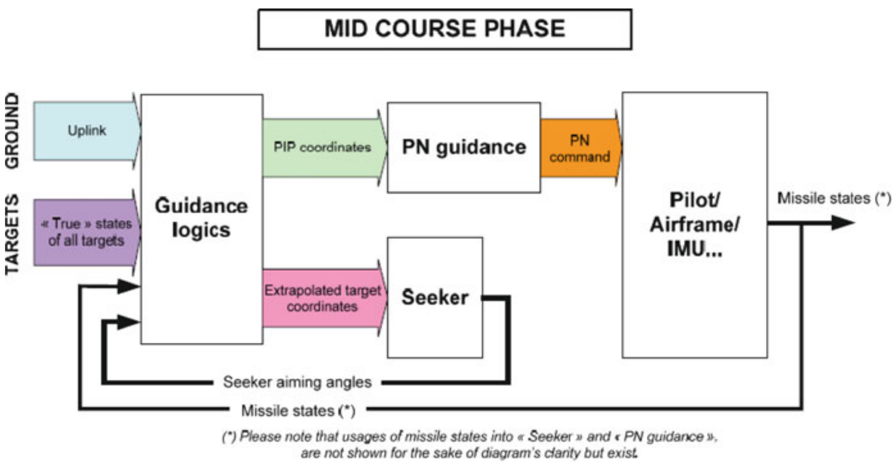


Fig. 15.7 During mid course, the guidance logics block extrapolates targets states and PIP coordinates. It also determines if the seeker locks on one of the targets

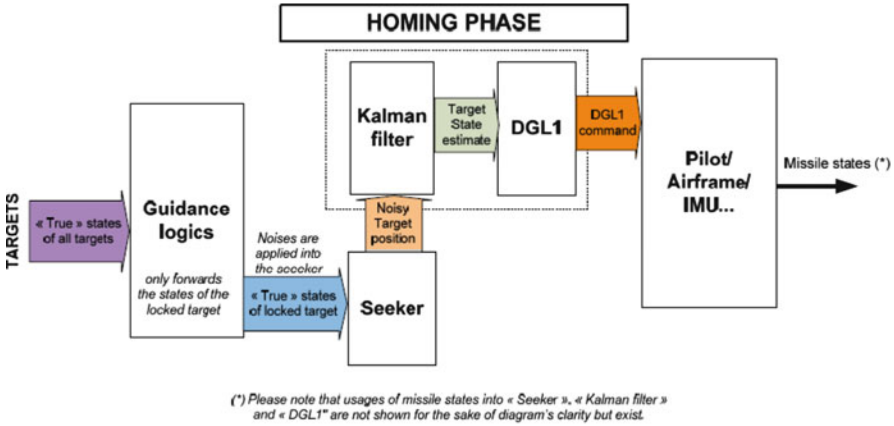


Fig. 15.8 In homing phase, guidance logics sends true states of locked target. The seeker block applies noises for measurement computation. Kalman filter estimates target's states. Finally a DGL1 command is applied

15.8 Scenario Description

Several scenarios for air defense in the ground and naval context have been defined. A target allocation benchmark policy, with neither re-allocation, nor SENEZ features, has been defined for comparison purpose. Scenario 3 (Fig. 15.9) deals with ground defense where Air Defense Units (ADUs) are located around (Defended Area, circle) the objective to be protected (RADAR, diamond mark in the center of the Defended Area). A threat aircraft launches a single missile and then escapes the radar zone. The aircraft and missile are supersonic.

The benchmark policy consists in launching a defending missile as soon as a threat appears in the radar detection range. The benchmark algorithm starts by launching one missile on the merged target. When both targets split, a second missile is shot. This second defending missile will intercept the attacking missile. Due to the sharp escape manoeuvres of the aircraft the first defending missile misses the aircraft. After missing the aircraft, the benchmark algorithm launches a third missile to chase the escaping aircraft. This last missile never reaches its target.

15.9 SENEZ Results

When the aircraft crosses the RADAR range the SENEZ algorithm launches two defending missiles (Fig. 15.10). In ground scenarios, several ADUs are considered, the algorithm automatically deciding by geometric considerations which ADU to use when launching defending missiles. For simplicity, in naval and ground scenarios only one location is considered as the final target goal (ground objective to protect,

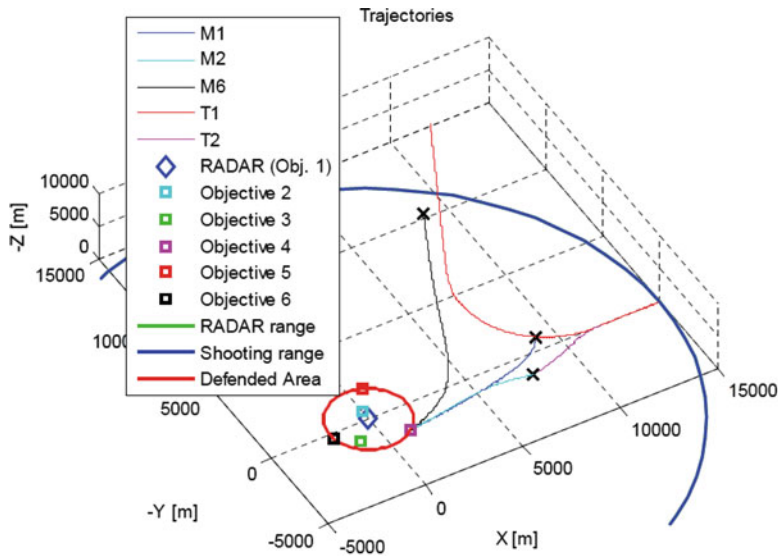


Fig. 15.9 Benchmark trajectories in scenario 3; the line coming from the right side which turns on its right side is the trajectory of an aircraft which launches a missile towards the area to protect (*circle*); the defense missiles are launched from the same launching base located on the border of the area to protect at *bottom* of the figure; the first missile misses the aircraft; the second one intercepts the threat/missile and the last one also misses reaching the aircraft

RADAR diamond mark). Simple waypoints are used to generate target trajectory assumptions, even if it is possible to extend the concept to more sophisticated target trajectory assumptions.

Figure 15.10 explains what happens when using the SENEZ algorithm and what the improvements with respect to the benchmark policy are. The defending missiles are $M2$ (on the left) and $M3$ (on the right). The aircraft trajectory is T1 turning on the right side. T2 is the missile launched by the aircraft. The defending missiles intercept when the threat trajectories switch from plain to dot lines. The dot lines describe what happens when using the benchmark policy in place of the SENEZ algorithm. The remaining dot lines are the target trajectory assumptions continuously refined during the engagement. A straight line assumption was considered by the algorithm, however defended missiles assigned to the right and to the left threats are enough to cover the three way-point assumptions elaborated when the initial threat appears. The SENEZ algorithm intercepts the attacking missile at longer distance than the benchmark algorithm, around a 1km improvement. Moreover, SENEZ only launches two defending missiles and also intercepts the launching aircraft which the benchmark algorithm fails to do. The fact that SENEZ directs missiles to the left and right sides, plus the fact that SENEZ launches earlier than the benchmark explains the SENEZ performance improvement. The Monte Carlo simulations confirm these explanations. The benchmark strategy as the

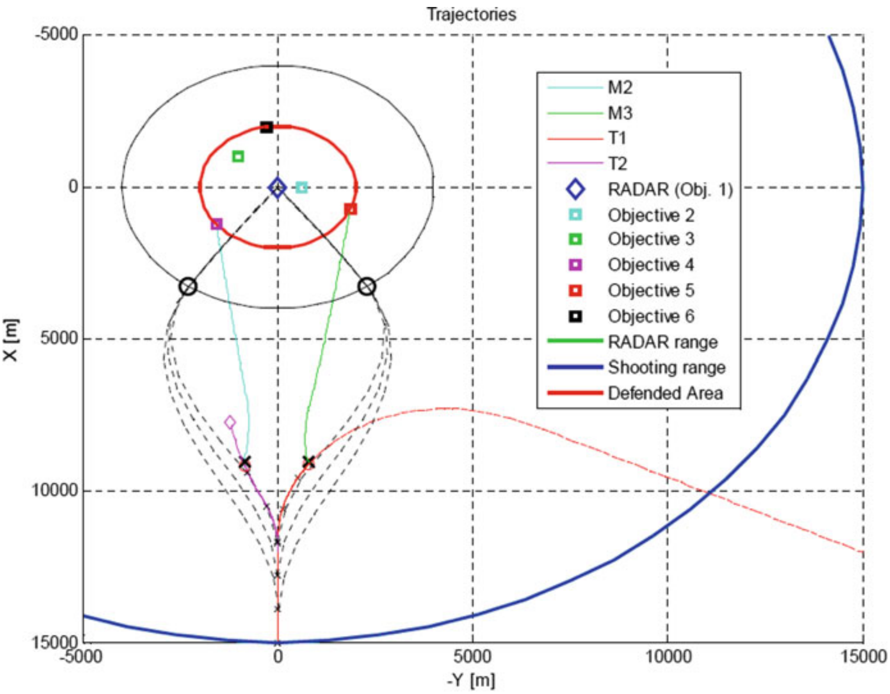


Fig. 15.10 SENEZ target allocation algorithm on scenario 3; the *dashed lines* are the target assumption trajectories that the algorithm considers for launching of defense missiles; a third hypothesis "flying straight" to counter the incoming aircraft is also taken into account, however two defense missiles are sufficient to cover the three hypotheses; the assumption "straight line" is then removed from the figure; the *black crosses* explain where the SENEZ interceptions occur; we superimpose the reference trajectories for purposes of comparison

SENEZ algorithm are able to intercept the missile/threat. The interception of the missile/threat occurs a little earlier in the SENEZ case. The main difference is that with the SENEZ algorithm the aircraft has been intercepted in most of the cases (see Fig. 15.11). The mean value for intercepting is around 20 s. with SENEZ and never happens before 20 s. with the benchmark strategy. Moreover, intercepting the aircraft often misses in the benchmark simulations (80 s is the maximum time of the simulation).

Monte-Carlo runs have been executed for all the scenarios, comparing interception times obtained with the benchmark model to those obtained with the SENEZ. Disturbances for these runs were as follow:

- Seeker noise;
- Initial position of the targets (disturbance with standard deviation equal to 50 m);
- Initial Euler angles of the target (disturbance with standard deviation equal to 2.5°).

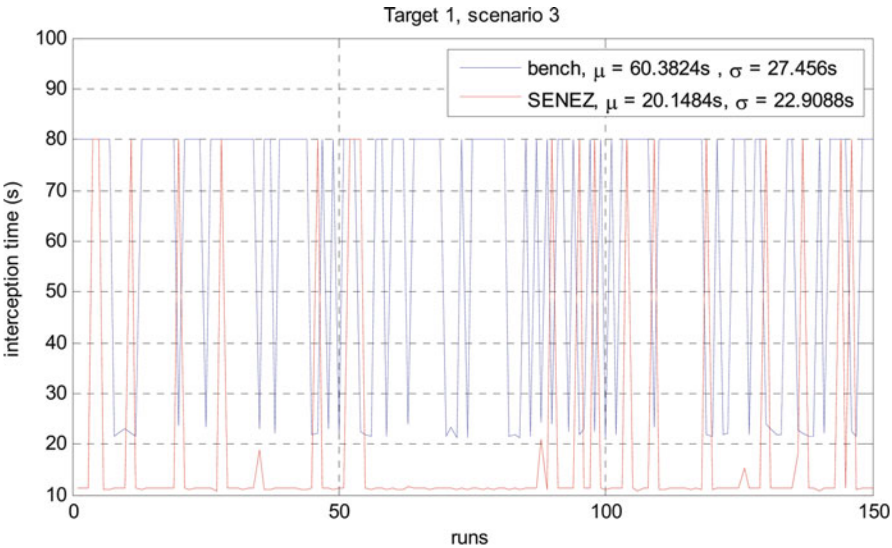


Fig. 15.11 Monte Carlo results on scenario 3; mean (μ) and standard deviation (σ) for the benchmark policy (*top*) and the SENEZ policy (*down*) when intercepting the target T1 (aircraft threat)

Performance analyses have also been executed on various other scenarios for ground and naval applications contexts. Moreover, parametric studies have been conducted on the following aspects:

- Waypoint placements;
- The drag coefficient of the defensive missiles;
- The radius and range of the seekers;
- Plus some variations on the scenario definition as time of appearance of the second target in scenario 3.

Attention is also paid to finding way-point placements that would be convenient for all ground to air scenarios, or all surface to air scenarios. The optimal placement of the way-points highly depends on the scenario. This tends to prove there would be an advantage in increasing the number of way-points/ missiles corresponding to an increased number of SENEZ hypotheses.

Potential benefits were first illustrated on all the scenarios considered against targets performing highly demanding evasive manoeuvres as well as apparent single targets that resolve into two splitting targets. The trajectories obtained gave a better idea of the SENEZ behavior. However, the way that target hypotheses are issued proved to be critical. This has been demonstrated by the parametric studies as placement of the way-points changed greatly the results from one scenario to another. The sensitivity to parameters such as drag and seeker features has also been investigated. Results obtained during these parametric studies seem to show the initial number of way-points/hypotheses per target chosen three might be too low.

Statistical studies have also been conducted. While providing improved performances in terms of time of last interception in most cases, the standard deviation greatly increased in some scenarios due to misses among the first salvo. These misses may be due to the simplified Kalman estimator used in our model, to the choice of the mid-course guidance made (classical PN on PIP for these tests), to the logics used for seeker pointing, or to an insufficient number of way-points.

15.10 SENEZ Perspectives

SENEZ guidance attempts to embed the future possible target behavior into the guidance strategy by using goal oriented predictions of partitioned threat trajectories to drive missile allocation and guidance commands. As such the SENEZ approach offers an alternative to mid-course guidance schemes which guide the intercepting missile or missiles towards a weighted track. The general application of SENEZ would lead to a major change in weapon C2 philosophy for naval applications which may not be justifiable.

The SENEZ engagement plan requires that a missile be fired at each partitioned set of trajectories. This is different from many existing naval firing policies which would fire a single missile to the target at long range and would delay firing another missile until later when, if there were sufficient time, a kill assessment would be undertaken before firing a second round. Depending on the evolution of target behaviour, current C2 algorithms may fire a second missile before the potential interception by the first missile. So existing systems tend to follow a more sequential approach, the naval platform needing to preserve missile stocks so that salvo firings are limited; unlike air platform the naval platform cannot withdraw rapidly from an engagement. The proposed engagement plan is purely geometric in formation as opposed to current schemes which use probabilities that the target is making for a particular goal [2]. This latter type of engagement plan will generally result in fewer missiles being launched. In the SENEZ scheme, a missile salvo will be fired more often because the potential target trajectories are all equally likely. For instance, when the target is at long range, it is likely that its choice of asset to attack is of same probability, whereas at the inner range boundary, it is most likely that the target is straight-flying towards its intended target.

Despite these potentially negative assessments of the SENEZ concept, there will be occasions when current C2 algorithms will determine that it is necessary to launch a salvo against a particular threat. For instance, a particularly high value asset such as an aircraft carrier may be targeted and a high probability of successful interception is required. In such circumstances there could be merit in the SENEZ approach. Essentially, in the naval setting SENEZ may be considered as a possible enhancement for the salvo firing determined by the engagement planning function in existing C2 systems.

For air-to-air systems the scope for considering a SENEZ form of guidance may be greater. It is often policy for aircraft to fire two missiles at an opposing aircraft engaged at medium range. With a two aircraft patrol, the leader and the wing aircraft will each fire a missile at the target, there is an opportunity to shape the guidance so that possible break manoeuvres are covered. With separate platforms firing the missiles it would be necessary for inter-platform communication so that each missile could be allocated to a unique trajectory partition.

By the way, many extensions could be addressed. First of all, new mid course guidance scheme, so called particle guidance or trade-off mid course guidance could be considered to guide on several tracks rather than assuming one unique target [3]. Moreover, by considering allocation plans with two defending missiles on one target NEZ as computed in [5, 10] it could be possible to involve defending missile with diminished performances (no up link during the mid-course, less kinematics performance, low cost seeker). Considering inter-missile communication capabilities the current centralized algorithm could be improved with decentralized features. Decentralization would allow to distribute the processing to individual missiles rather than concentrating the process computation in one unique location; i.e. the frigate to protect [4].

15.11 Conclusion

Cooperative guidance is a technique which is likely to emerge as a technology in future weapon systems. Future weapon system scenarios will include the need to engage multiple threats which places greater demands on the guidance chain compared with one-on-one. This project has developed various component technologies supporting the concept of cooperative guidance.

For the terminal phase, differential game guidance laws were applied where the NEZ was used to characterize the ability of the missile to capture the target. The focus of this article is concentrated on the way in which some of these technologies are combined to provide an enhanced capability when salvos are launched to deal with target threats, the SENEZ concept. Allocation algorithms have been extended to consider the future possible behavior of the target; the technique can determine how many missiles to fire and provide the initialization for the missiles in the salvo.

Initial results have demonstrated the potential of the SENEZ concept where in some cases this technique has produced results that were better than the baseline allocation algorithm. Although the potential has been demonstrated it remains to examine the full robustness of the approach in terms of range of scenarios and optimization of parameter setting.

15.12 Acronyms and Notations (Tables 15.3 and 15.4)

Table 15.3 Acronym table

NEZ	No Escape Zone
SENEZ	Salvo Enhanced No Escape Zone
EIG	Earliest Interception Geometry circle
ZEM	Zero Effort Miss
DGL	linear Differential game Guidance Law
DGL1	DGL guidance law version considering first order dynamics for both players
TSG	Trajectory Shaping Guidance law
PN	Proportional Navigation guidance law
DGGL	Differential Geometry Guidance Law
IFA	In Flight target Allocation algorithm
ABL	target Allocation Before Launch algorithm
LOS	Line Of Sight
PIP	missile target Predicted Interception Point
IMU	Inertial Measurement Unit
ADU	Air Defense Unit
C2	Command and Control systems
MCM ITP	Materials and Components for Missiles Innovation and Technology Partnership

Table 15.4 Notation table

R_{wpt}	Range from area to protect to waypoints
ψ_{wpt}	Heading angle of waypoints respect to the North direction three waypoints: $-\psi_{\text{wpt}}$, $+\psi_{\text{wpt}}$ and 0 (North)
$M_i T_j H_k$	Missile i mid course path considering Target j flying through waypoint Hypothesis k
$T_j H_k$	Target j flying through waypoint Hypothesis k
N	Maximum number of defending missiles
P	Maximum number of threats
W	Number of waypoints
Z	ZEM; Zero Effort Miss
Z_{limit}	Boundary of NEZ; maximum ZEM
t_{F}	Final time
t_{go}	Time to go $(t_{\text{F}} - t)$
θ	Normalized t_{go}
y	Evader pursuer perpendicular miss distance
$\dot{y}$	Evader pursuer perpendicular velocity
$\ddot{y}_{\text{P}}$	Perpendicular pursuer acceleration; close to pursuer Latax
$\ddot{y}_{\text{E}}$	Perpendicular evader acceleration
τ_{P}	Pursuer time lag
τ_{E}	Evader time lag
$a_{\text{P max}}$	Pursuer maximum acceleration
$a_{\text{E max}}$	Evader maximum acceleration
μ_0	Maximum pursuer acceleration Over maximum evader acceleration at $\theta = 0$
ϵ_0	Time lag of the evader over time lag of the pursuer at $\theta = 0$
ν	Changes in μ ratio according to θ
R_{MT}	Missile target range
$C_{A,B}$	Allocation plan cost

Acknowledgements This work was funded by the French—UK Materials and Components for Missiles—Innovation and Technology Partnership (MCM ITP) research programme.

References

1. Basar, T., Olsder, G.J.: Dynamic Non Cooperative Game Theory, 2nd ed. CLASSICS in Applied Mathematics, CL 23, Society for Industrial and Applied Mathematics, Philadelphia, ISBN 978-0-89871-429-6 (1999)
2. Bessière, P. and the BIBA – INRIA Research Group – Projet CYBERMOVE: Survey – Probabilistic Methodology and Techniques for Artefact Conception and Development, INRIA Research Report No. 4730, ISSN 0249-6399 ISRN INRIA/RR-4730-FR+ENG (February 2003)
3. Dionne, D., Michalska, H., Rabbath, C.-A.: Predictive guidance for pursuit-evasion engagements involving multiple decoys. J. Guidance Control Dyn. **30**(5), 1277–1286 (2007)
4. Farinelli, A., Rogers, A., Petcu A., Jennings, N.-R.: Decentralised coordination of low-power embedded devices using the max-sum algorithm. In: Padgham, Parkes, Müller, Parsons (eds.) Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems, AAMAS-08, May 12–16, 2008, Estoril, Portugal, pp. 639–646 (2008)
5. Ganebny, S.A., Kumkov, S.S., Le Ménec, S., Patsko, V.S.: Numerical study of two-on-one pursuit-evasion game. In: Preprints of the 18th IFAC World Congress, Milano, Italy, August 28–September 2, 2011, pp. 9326–9333 (2011)
6. Ge, J., Tang, L., Reimann, J., Vachtsevanos, G.: Suboptimal approaches to multiplayer pursuit-evasion differential games. In: AIAA 2006–676 Guidance, Navigation, and Control Conference, August 21–24, 2006, Keystone, Colorado (2006)
7. Isaacs, R.: Differential Games, a Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization. Wiley, New York (1965)
8. Jang, J.S., Tomlin C. .: Control strategies in multi-player pursuit and evasion game. In: AIAA 2005-6239 Guidance, Navigation, and Control Conference, August 15–18, 2005, San Francisco, California (2005)
9. Le Ménec, S.: Cooperative mid course guidance law based on attainability constrains: invited session: advanced methods for the guidance and control of autonomous vehicles. In: Proceedings of the European Control Conference, August 23–26, 2009, Hungary, MoA3.5, pp. 127–131, ISBN 978-963-311-369-1 (2009)
10. Le Ménec, S.: Linear differential game with two pursuers and one evader. In: Breton M., Szajowski, K. (eds.) Annals of the International Society of Dynamic Games, Advances in Dynamic Games, Theory, Applications, and Numerical Methods for Differential and Stochastic Games. Birkhäuser, Springer, New York, ISBN 978-0-8176-8088-6 (2011)
11. Le Ménec, S., Shin, H.-S., Tsourdos, A., White, B., Zbikowski, R., Markham K.: Cooperative missile guidance strategies for maritime area air defense. In: 1st IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys09), 24–26 September 2009, Venice, Italy (2009)
12. Shima, T., Shinar, J., Weiss, H.: New interceptor guidance law integrating time-varying and estimation-delay models. J. Guidance Control Dyn. **26**(2), 295–303 (2003)
13. Shin, H.-S., Le Ménec, S., Tsourdos, A., Markham, K., White, B., Zbikowski, R.: Cooperative guidance for naval area defence. In: 18th IFAC Symposium on Automatic Control in Aerospace, September 6–10, 2010, Nara, Japan (2010)
14. Shin, H.-S., Piet-Lahanier, H., Tsourdos, A., Le Ménec, S., Markham K., White B.-A.: Membership set-based mid course guidance: application to manoeuvring target interception. In: Preprints of the 18th IFAC World Congress, August 28–September 2, 2011 Milano, Italy, pp. 3903–3908 (2011)

15. Shin, H.-S., Tsourdos, A., Le Ménec, S., Markham, K., White, B.: Cooperative mid course guidance for area air defence. In: AIAA 2010–8056 Guidance, Navigation and Control, August 2–5, 2010, Toronto, Ontario, Canada (2010)
16. Shinar, J., Shima, T.: Non-orthodox guidance law development approach for the interception of maneuvering anti-surface missiles. J. Guidance Control Dyn. **25**(4), 658–666 (2002)
17. Zarchan, P.: Tactical and Strategic Missile Guidance, Progress in Astronautics and Aeronautics, vol. 219, 5th revised ed. (2007)

Chapter 16
A Method of Solving Differential Games Under Integrally Constrained Controls

Aleksandr A. Belousov, Aleksander G. Chentsov, and Arkadii A. Chikrii

Abstract This study deals with linear game problems under integral constraints on controls. The proposed scheme leans upon the ideas of the method of resolving functions [Chikrii, Conflict Controlled Processes. Kluwer, Boston (1997)]. The analog of the Pontryagin condition formulated in the paper, makes it feasible to derive sufficient conditions for the finite-time termination of differential game. Obtained results are illustrated with the typical game state of affairs “simple motion” and continue researches [Chikrii and Belousov, Mem. Inst. Math. Mech. Ural Div. Russ. Acad. Sci. 15(4):290–301 (2009) (in Russian); Nikol’sky, Diff. Eq. Minsk. 8(6):964–971 (1972) (in Russian); Subbotin and Chentsov, Optimization of Guarantee in Problems of Control. Nauka, Moscow (1981) (in Russian)].

Keywords Differential game • Pursuit game • Integral constraint • Set-valued mapping • Resolving function

A.A. Belousov (✉) • A.A. Chikrii
Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, Kiev, Ukraine
e-mail: dept165@insyg.kiev.ua; chik@insyg.kiev.ua

A.G. Chentsov
Institute of Mathematics and Mechanics, Ural Division of Russian Academy of Sciences, Ekaterinburg, Russia
e-mail: chentsov@imm.uran.ru

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_16,

16.1 Problem Statement

The dynamics of the game is given by the differential equation

$$\dot{z} = Az + Bu + Cv, \quad z(0) = z^0, \tag{16.1}$$

where $z \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $v \in \mathbb{R}^l$, A, B, C are constant matrices of dimensions $n \times n$, $n \times m$, $n \times l$, respectively.

Controls of the pursuing player $u(\cdot)$ and the evading one $v(\cdot)$ are Lebesgue measurable functions, satisfying integral constraints:

$$\int_0^\infty \|u(\tau)\|^2 \, \mathrm{d}\tau \leqslant 1, \quad \int_0^\infty \|v(\tau)\|^2 \, \mathrm{d}\tau \leqslant 1. \tag{16.2}$$

Such controls will be called admissible.

Terminal set M is a linear subspace from $\mathbb{R}^n$.

Definition 16.1. The game will be considered as terminated at the time $T = T(z^0)$ if for any admissible control of the evader $v(t)$ there exist admissible controls of the pursuer $u(t)$ that bring a solution to Eq. (16.1) $z(t)$, from the initial state z^0 to the terminal set at instant T exactly: $z(T) \in M$.

It is assumed that when constructing its control $u(t)$ at the instant t, the pursuer can use information on its adversary control that has been acquired up to this time $v(\tau)$, $\tau \in [0,t]$.

Denote by π the orthoprojector from $\mathbb{R}^n$ onto subspace L, which is a complement to M in $\mathbb{R}^n$. Let us introduce an assumption on the game parameters that may thought of as an analog of Pontryagin's condition [4] as applied to differential games with integrally constrained controls.

Assumption 16.1. *There exists a number* λ*,* $0 \leqslant \lambda < 1$*, such that for all positive* t *the following inclusion is fulfilled:*

$$\pi \mathrm{e}^{At} CV \subset \lambda \pi \mathrm{e}^{At} BU, \tag{16.3}$$

where $U = \{u \in \mathbb{R}^m : \|u\|^2 \leqslant 1\}$ *and* $V = \{v \in \mathbb{R}^l : \|v\|^2 \leqslant 1\}$ *are unit balls in the control domains.*

In the sequel, this assumption is thought to hold.

16.2 Auxiliary Statements

Let us fix an initial position z^0 and introduce a set-valued mapping

$$\Omega(t,\tau,v) = \left\{ \gamma \in \mathbb{R} : \gamma \pi \mathrm{e}^{At} z^0 + \pi \mathrm{e}^{A\tau} Cv \in \sqrt{(1-\lambda)\gamma + \lambda \|v\|^2}\, \pi \mathrm{e}^{A\tau} BU \right\}, \tag{16.4}$$

where $(t,\tau,v) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l$, $\quad \mathbb{R}_+ = [0,\infty)$. Consider an auxiliary function (a so-called resolving function [1]):

$$\gamma(t,\tau,v) = \sup \Omega(t,\tau,v). \tag{16.5}$$

In what follows, we investigate the properties of this set-valued mapping and function.

Lemma 16.1. *The following relationships hold:*

$\gamma(t,\tau,v) \geqslant 0$ *for all* $(t,\tau,v) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l$;
If $\pi e^{At} z^0 = 0$, *then* $\gamma(t,\tau,v) = +\infty$ *for all* $(\tau,v) \in \mathbb{R}_+ \times \mathbb{R}^l$;
If $\pi e^{At} z^0 \neq 0$, *then* $\gamma(t,\tau,v) < \infty$ *for all* $(\tau,v) \in \mathbb{R}_+ \times \mathbb{R}^l$.

Proof. To prove the first statement it is sufficient to show that $0 \in \Omega(t,\tau,v)$, that is,

$$\pi e^{A\tau} Cv \in \sqrt{\lambda \|v\|^2}\, \pi e^{A\tau} BU. \tag{16.6}$$

For $v = 0$ this inclusion is evidently satisfied. Taking into account (16.3) and the inclusion $0 \in U$ at $v \neq 0$ we have a chain of inclusions:

$$\pi e^{A\tau} C \frac{v}{\|v\|} \in \pi e^{A\tau} CV \subset \lambda \pi e^{A\tau} BU \subset \sqrt{\lambda}\, \pi e^{A\tau} BU,$$

which means that $\gamma(t,\tau,v) \geqslant 0$.

In view of (16.6) it is evident that when $\pi e^{At} z^0 = 0$, the inclusion $\gamma \in \Omega(t,\tau,v)$ holds for any positive number γ.

Let us fix the vector $(t,\tau,v) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l$ such that $\pi e^{At} z^0 \neq 0$. The norm $\|\gamma \pi e^{At} z^0 + \pi e^{A\tau} Cv\|$ grows in γ linearly (for sufficiently large γ) and the norm of the vector from the right-hand side of the inclusion in (16.4) is bounded by the function

$$\sqrt{(1-\lambda)\gamma + \lambda \|v\|^2} \max_{\|u\| \leqslant 1} \|\pi e^{A\tau} Bu\|,$$

which, as a function of γ, grows no faster than a square root. Therefore, for sufficiently large γ, the inclusion in (16.4) fails, that is, $\gamma(t,\tau,v) < \infty$. □

Lemma 16.2. *If for some positive number* γ *the inclusion* $\gamma \in \Omega(t,\tau,v)$ *is satisfied, then the interval* $[0,\gamma]$ *belongs to the set* $\Omega(t,\tau,v)$, $(t,\tau,v) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l$.

Proof. Suppose that for some positive number γ the inclusion $\gamma \in \Omega(t,\tau,v)$ is satisfied. Then for any number β, $\quad 0 < \beta < \gamma$, the following inclusion is fulfilled:

$$\beta \pi e^{At} z^0 + \frac{\beta}{\gamma} \pi e^{A\tau} Cv \in \frac{\beta}{\gamma} \sqrt{(1-\lambda)\gamma + \lambda \|v\|^2}\, \pi e^{A\tau} BU.$$

Furthermore, from inclusion (16.6) follows

$$\left(1-\frac{\beta}{\gamma}\right)\pi\,\mathrm{e}^{A\tau}Cv \in \left(1-\frac{\beta}{\gamma}\right)\sqrt{\lambda\,\|v\|^2}\,\pi\,\mathrm{e}^{A\tau}BU.$$

Whence, taking into account the convexity of the set $\pi\,\mathrm{e}^{A\tau}BU$, we obtain

$$\beta\pi\,\mathrm{e}^{At}z^0+\pi\,\mathrm{e}^{A\tau}Cv \in \frac{\beta}{\gamma}\sqrt{(1-\lambda)\gamma+\lambda\,\|v\|^2}\,\pi\,\mathrm{e}^{A\tau}BU+\left(1-\frac{\beta}{\gamma}\right)\sqrt{\lambda\,\|v\|^2}\,\pi\,\mathrm{e}^{A\tau}BU$$
$$\subset\left(\frac{\beta}{\gamma}\sqrt{(1-\lambda)\gamma+\lambda\,\|v\|^2}+\left(1-\frac{\beta}{\gamma}\right)\sqrt{\lambda\,\|v\|^2}\right)\pi\,\mathrm{e}^{A\tau}BU.$$

The function $f(\gamma)=\sqrt{(1-\lambda)\gamma+\lambda\,\|v\|^2}$ is concave since $f''(\gamma)<0$ for $\gamma>0$. That is why the following inequality in terms of function f is satisfied [5]:

$$\frac{f(\gamma)-f(0)}{\gamma}\leqslant\frac{f(\beta)-f(0)}{\beta}\quad\text{for}\quad\gamma>\beta>0,$$

whence we obtain

$$\frac{\beta}{\gamma}f(\gamma)+\left(1-\frac{\beta}{\gamma}\right)f(0)\leqslant f(\beta)$$

or

$$\frac{\beta}{\gamma}\sqrt{(1-\lambda)\gamma+\lambda\,\|v\|^2}+\left(1-\frac{\beta}{\gamma}\right)\sqrt{\lambda\,\|v\|^2}\leqslant\sqrt{(1-\lambda)\beta+\lambda\,\|v\|^2}.$$

Therefore, since $0\in\pi\,\mathrm{e}^{A\tau}BU$, we have

$$\beta\pi\,\mathrm{e}^{At}z^0+\pi\,\mathrm{e}^{A\tau}Cv\in\sqrt{(1-\lambda)\beta+\lambda\,\|v\|^2}\,\pi\,\mathrm{e}^{A\tau}BU,$$

and consequently $\beta\in\Omega(t,\tau,v)$ for any $\beta\in[0,\gamma]$. □

Lemma 16.3. *If $\pi\,\mathrm{e}^{At}z^0\neq 0$, then the upper bound in the definition of $\gamma(t,\tau,v)$ (16.5) is attained. Function $\gamma(t,\tau,v)$ is Borel measurable jointly in $(t,\tau,v)\in\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}^l$.*

Proof. Let us study the function of the distance between the vector and the compact, entering the definition of $\Omega(t,\tau,v)$ (16.4):

$$\delta(\gamma,t,\tau,v)=\min_{u\in U}\|\gamma\pi\,\mathrm{e}^{At}z^0+\pi\,\mathrm{e}^{A\tau}Cv-\sqrt{(1-\lambda)\gamma+\lambda\,\|v\|^2}\,\pi\,\mathrm{e}^{A\tau}Bu\|.$$

This function is continuous on the set $\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}^l$. The inclusion $\gamma\in\Omega(t,\tau,v)$ is equivalent to the equation $\delta(\gamma,t,\tau,v)=0$. Then, by the function defined in (16.5), it follows that there exists a sequence of numbers γ_i such that

$$\delta(\gamma_i,t,\tau,v)=0\quad\text{and}\quad\gamma_i\xrightarrow[i\to\infty]{}\gamma(t,\tau,v).$$

Whence, taking into account the finiteness of $\gamma(t,\tau,v)$ (if $\pi e^{At}z^0 \neq 0$), we conclude that $\delta(\gamma(t,\tau,v),t,\tau,v)=0$, and the upper bound in (16.5) is attained.

Let us consider the level set of function $\gamma(t,\tau,v)$:

$$\Lambda_a = \Big\{(t,\tau,v) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l : \quad \gamma(t,\tau,v) < a\Big\}.$$

We will show that this set is open and, therefore, Borel for any positive number a. This will be imply the Borel measureability of the function $\gamma(t,\tau,v)$.

Let us fix a positive number a and let $(\bar{t},\bar{\tau},\bar{v})$ be an arbitrary point of Λ_a. Consequently, $a \notin \Omega(\bar{t},\bar{\tau},\bar{v})$ and $\delta(a,\bar{t},\bar{\tau},\bar{v}) > 0$. The continuity of the function $\delta(\cdot)$ ensures the existence of a neighborhood Δ of the point $(\bar{t},\bar{\tau},\bar{v})$ such that the inequality $\delta(a,t,\tau,v) > 0$ holds for all $(t,\tau,v) \in \Delta$, that is, $\gamma(t,\tau,v) < a$ for all $(t,\tau,v) \in \Delta$. This implies that the set Λ_a is open, as required. □

16.3 Main Theorem

We now formulate the sufficient conditions to guarantee bringing a solution of (16.1), (16.2) to the terminal set M, beginning from the initial state z^0.

Theorem 16.1. *Let Assumption 16.3 on the parameters of the game (16.1), (16.2) hold.*

Suppose that there exists a moment $T = T(z^0)$ such that either $\pi e^{AT}z^0 = 0$ or $\pi e^{AT}z^0 \neq 0$ and for all admissible controls $v(\cdot)$ the following inequality is satisfied:

$$\int_0^T \gamma(T,T-\tau,v(\tau))\,d\tau \geqslant 1. \tag{16.7}$$

Then the differential game can be terminated at time T.

Proof. Let us fix the moment of time T that fulfills the assumption of the theorem. Let us first analyze the case $\pi e^{AT}z^0 \neq 0$.

In view of Lemma 16.3 the resolving function $\gamma(T,\tau,v)$ is Borel measurable, and for all $(\tau,v) \in \mathbb{R}_+ \times \mathbb{R}^l$ the following inclusion is satisfied:

$$\gamma(T,\tau,v)\pi e^{AT}z^0 + \pi e^{A\tau}Cv \in \bigcup_{u\in U} \sqrt{(1-\lambda)\gamma(T,\tau,v) + \lambda\|v\|^2}\,\pi e^{A\tau}Bu.$$

The mappings on the left- and right-hand sides of this inclusion are Borel measurable in (τ,v) and continuous in u, $u \in U$.

By the theorem of Kuratowski and Ryll-Nardzewski [2, 3] there exists a Borel-measurable selection, that is, a Borel-measurable mapping $w(\tau,v) \in U$, such that

$$\gamma(T,\tau,v)\pi e^{AT}z^0 + \pi e^{A\tau}Cv = \sqrt{(1-\lambda)\gamma(T,\tau,v) + \lambda\|v\|^2}\,\pi e^{A\tau}Bw(\tau,v)$$

for all $(\tau, v) \in \mathbb{R}_+ \times \mathbb{R}^l$. Also, from this theorem it may be concluded that there exists a Borel-measurable mapping $\tilde{w}(\tau, v) \in U$ such that

$$\pi \mathrm{e}^{A\tau} C v = \sqrt{\lambda \|v\|^2}\, \pi \mathrm{e}^{A\tau} B \tilde{w}(\tau, v)$$

for all $(\tau, v) \in \mathbb{R}_+ \times \mathbb{R}^l$.

Let us assume that on the interval $[0, T]$ the evader applies the control $v(\tau)$, which is a Lebesgue-measurable function satisfying the integral inequality

$$\int_0^T \|v(\tau)\|^2 \, \mathrm{d}\tau \leqslant 1.$$

By condition (16.7) of the theorem there exists an instant $T^* = T^*(z^0, v(\cdot))$ such that

$$\int_0^{T^*} \gamma(T, T - \tau, v(\tau)) \, \mathrm{d}\tau = 1.$$

Then the control of the pursuer on the interval $[0, T]$ prescribed by the formula

$$u(\tau) = \begin{cases} -\sqrt{(1-\lambda)\gamma(T, T-\tau, v(\tau)) + \lambda \|v(\tau)\|^2} \cdot w(T-\tau, v(\tau)), & \text{for } \tau \in [0, T^*], \\ -\sqrt{\lambda \|v(\tau)\|^2} \cdot \tilde{w}(T-\tau, v(\tau)), & \text{for } \tau \in (T^*, T]. \end{cases} \tag{16.8}$$

In essence, such a control law of the pursuer is a countercontrol with a single switching.

Note that superposition of Borel and Lebesgue functions is a Lebesgue-measurable function [2]. That is why control (16.8) is Lebesgue measurable for any measurable control $v(\tau)$.

Let us show that under such a control choice of the pursuer a solution of (16.1) hits the terminal set at instant T:

$$\begin{aligned} \pi z(T) &= \pi \mathrm{e}^{AT} z^0 + \int_0^T \pi \mathrm{e}^{A(T-\tau)} [B u(\tau) + C v(\tau)] \, \mathrm{d}\tau \\ &= \pi \mathrm{e}^{AT} z^0 + \int_0^{T^*} \pi \mathrm{e}^{A(T-\tau)} B u(\tau) \, \mathrm{d}\tau + \int_{T^*}^{\mathrm{T}} \pi \mathrm{e}^{A(T-\tau)} B u(\tau) \, \mathrm{d}\tau \\ &\quad + \int_0^T \pi \mathrm{e}^{A(T-\tau)} C v(\tau) \, \mathrm{d}\tau = \pi \mathrm{e}^{AT} z^0 \\ &\quad - \int_0^{T^*} \sqrt{(1-\lambda)\gamma(T, T-\tau, v(\tau)) + \lambda \|v(\tau)\|^2}\, \pi \mathrm{e}^{A(T-\tau)} B w(T-\tau, v(\tau)) \, \mathrm{d}\tau \end{aligned}$$

$$+\int_{T^*}^{\mathrm{T}} \sqrt{\lambda \|v(\tau)\|^2}\, \pi\, \mathrm{e}^{A(T-\tau)} B\tilde{w}(T-\tau, v(\tau))\,\mathrm{d}\tau + \int_0^T \pi\, \mathrm{e}^{A(T-\tau)} Cv(\tau)\,\mathrm{d}\tau$$

$$= \pi\, \mathrm{e}^{AT} z^0 - \int_0^{T^*} [\gamma(T, T-\tau, v(\tau))\pi\, \mathrm{e}^{AT} z^0 + \pi\, \mathrm{e}^{A(T-\tau)} Cv(\tau)]\,\mathrm{d}\tau$$

$$+\int_{T^*}^{\mathrm{T}} \pi\, \mathrm{e}^{A(T-\tau)} Cv(\tau)\,\mathrm{d}\tau + \int_0^T \pi\, \mathrm{e}^{A(T-\tau)} Cv(\tau)\,\mathrm{d}\tau$$

$$= \pi\, \mathrm{e}^{AT} z^0 - \int_0^{T^*} \gamma(T, T-\tau, v(\tau))\,\mathrm{d}\tau \cdot \pi\, \mathrm{e}^{AT} z^0 = 0.$$

This equation proves that a solution of (16.1) is brought to the terminal set $z(T) \in M$.

Let us verify that a control $u(\tau)$ (16.8) constructed in such a way meets the integral constraints (16.2):

$$\int_0^T \|u(\tau)\|^2\,\mathrm{d}\tau = \int_0^{T^*} [(1-\lambda)\gamma(T, T-\tau, v(\tau)) + \lambda\|v(\tau)\|^2]\|w(T-\tau, v(\tau))\|^2\,\mathrm{d}\tau$$

$$+\int_{T^*}^{\mathrm{T}} \lambda\|v(\tau)\|^2\|\tilde{w}(T-\tau, v(\tau))\|^2\,\mathrm{d}\tau$$

$$\leqslant (1-\lambda)\int_0^{T^*} \gamma(T, T-\tau, v(\tau))\,\mathrm{d}\tau + \lambda\int_0^T \|v(\tau)\|^2\,\mathrm{d}\tau \leqslant 1.$$

The case $\pi\, \mathrm{e}^{AT} z^0 = 0$ is analyzed in a similar way. In so doing, the pursuer control on the interval $[0, T]$ is as follows:

$$u(\tau) = -\sqrt{\lambda\|v(\tau)\|^2} \cdot \tilde{w}(T-\tau, v(\tau)). \tag{16.9}$$

Analogously, it may be shown that in this case, too, the control (16.9) ensures bringing a solution of (16.1) to the terminal set M at moment T (for any admissible control $v(\tau)$), and control $u(\tau)$ meets integral constraint (16.2). □

Remark 16.1. The theorem is easily transferred to the case of general constraints on controls:

$$\int_0^\infty u^{\mathrm{T}}(\tau)Gu(\tau)\,\mathrm{d}\tau \leqslant \mu^2, \quad \int_0^\infty v^{\mathrm{T}}(\tau)Hv(\tau)\,\mathrm{d}\tau \leqslant \rho^2, \tag{16.10}$$

where G and H are symmetric, positive-definite matrices of dimensions $m \times m$ and $l \times l$, respectively, $u(\tau)$ and $v(\tau)$ are measurable functions, and μ and ρ are positive numbers, where the symbol $^{\mathrm{T}}$ means transposition.

The substitution

$$\tilde{u} = \frac{1}{\mu} \cdot G^{\frac{1}{2}} u, \quad \tilde{v} = \frac{1}{\rho} \cdot H^{\frac{1}{2}} v, \quad \tilde{B} = \mu \cdot BG^{-\frac{1}{2}}, \quad \tilde{C} = \rho \cdot CH^{-\frac{1}{2}}$$

transforms the differential game (16.1), (16.10) into the original form

$$\dot{z} = Az + \tilde{B}\tilde{u} + \tilde{C}\tilde{v}, \quad z \in \mathbb{R}^n, \quad \tilde{u} \in \mathbb{R}^m, \quad \tilde{v} \in \mathbb{R}^l, \quad z(0) = z^0,$$

$$\int_0^\infty \|\tilde{u}(\tau)\|^2 \, d\tau \leqslant 1, \quad \int_0^\infty \|\tilde{v}(\tau)\|^2 \, d\tau \leqslant 1.$$

In such a form the theorem can be transformed into the general case.

16.4 Example: Simple Motions

Below, the developed technique is applied to solve a specific problem.

The motions of the pursuer and the evader are described by the following differential equations:

$$\begin{aligned} &\dot{x} = u, \quad x(0) = x^0, \quad x \in \mathbb{R}^n, \quad u \in \mathbb{R}^n, \\ &\dot{y} = v, \quad y(0) = y^0, \quad y \in \mathbb{R}^n, \quad v \in \mathbb{R}^n. \end{aligned} \tag{16.11}$$

Constraints on the control have the form

$$\int_0^\infty \|u(\tau)\|^2 \, d\tau \leqslant \mu^2, \quad \int_0^\infty \|v(\tau)\|^2 \, d\tau \leqslant \rho^2, \quad \mu > \rho > 0. \tag{16.12}$$

A terminal set is given by the equality $x = y$.

Upon substitution

$$z = x - y, \quad \tilde{u} = \frac{u}{\mu}, \quad \tilde{v} = \frac{v}{\rho},$$

the game takes the standard form

$$\dot{z} = \mu\tilde{u} - \rho\tilde{v}, \quad z^0 = x^0 - y^0,$$

$$\int_0^\infty \|\tilde{u}(\tau)\|^2 \, d\tau \leqslant 1, \quad \int_0^\infty \|\tilde{v}(\tau)\|^2 \, d\tau \leqslant 1.$$

The terminal set is $M = \{0\}$, the operator π represents itself as an identical transformation.

It is easy to see that Assumption 16.3 is satisfied for the parameter $\lambda = \frac{\rho}{\mu} < 1$:

$$-\rho D \subset \lambda \cdot \mu D, \quad D = \{z \in \mathbb{R}^n : \quad \|z\|^2 \leqslant 1\}.$$

The resolving function $\gamma(\cdot)$ can be found from the formula for the set-valued mapping

$$\Omega(t,\tau,\tilde{v}) = \left\{\gamma \in \mathbb{R} : \gamma z^0 - \rho\tilde{v} \in \sqrt{(1-\lambda)\gamma + \lambda\|\tilde{v}\|^2} \cdot \mu D\right\}$$
$$= \left\{\gamma \in \mathbb{R} : (\gamma z^0 - \rho\tilde{v})^2 \leqslant \left[(1-\lambda)\gamma + \lambda\|\tilde{v}\|^2\right] \cdot \mu^2\right\}$$
$$= \left\{\gamma : F(\gamma,\tilde{v}) = \|z^0\|^2\gamma^2 - 2\gamma\left[\rho\langle z^0,\tilde{v}\rangle + \frac{\mu(\mu-\rho)}{2}\right] - \rho(\mu-\rho)\|\tilde{v}\|^2 \leqslant 0\right\}.$$

The function $F(\gamma,\tilde{v})$ represents itself as a quadratic polynomial with respect to γ with a positive coefficient at a higher degree term; therefore, $\gamma(\tilde{v})$ (16.5) is the largest root of the quadratic equation $F(\gamma,\tilde{v}) = 0$.

Note that $F(0,v) \leqslant 0$ for all $v \in \mathbb{R}^n$; therefore, the function $\gamma(v)$ is defined for all v and $\gamma(v) \geqslant 0$.

Let us find v^* that yields a minimum to the function $\gamma(v)$. To this end we differentiate the following equation:

$$\frac{\partial\gamma}{\partial v} = -\frac{\partial F/\partial v}{\partial F/\partial\gamma} = \frac{2\gamma\rho z^0 + 2\rho(\mu-\rho)v^*}{\partial F/\partial\gamma} = 0,$$

whence we obtain a unique extremum of the function $\gamma(\cdot)$:

$$v^* = -\frac{\gamma}{\mu-\rho} \cdot z^0.$$

The corresponding value v^* of function $\gamma(\cdot)$ can be found from the quadratic equation $F(\gamma,v^*) = 0$:

$$\gamma(v^*) = \frac{(\mu-\rho)^2}{\|z^0\|^2},$$

whence

$$v^* = -\frac{\mu-\rho}{\|z^0\|^2} \cdot z^0.$$

It can be easily shown that, in view of the fact that the function $\gamma(\cdot)$ is the largest root of the quadratic equation, the following inequality is true:

$$\gamma(v) \geqslant \frac{\sqrt{\rho(\mu-\rho)}\|v\|}{\|z^0\|^2} \to \infty \qquad \text{when } \|v\| \to \infty.$$

From this it follows that the unique extremum of function $\gamma(v)$ appears as its minimum.

By Theorem 16.1 the time of the game termination is defined by the relationships

$$\int_0^T \gamma(v(\tau))\,\mathrm{d}\tau \geqslant \int_0^T \gamma(v^*)\,\mathrm{d}\tau = \frac{(\mu-\rho)^2}{\|z^0\|^2}\cdot T = 1,$$

whence

$$T = \frac{\|z^0\|^2}{(\mu-\rho)^2}. \tag{16.13}$$

It should be noted that instant T coincides with the time of first absorption [4] for the game (16.11), that is, it coincides with the first moment when the attainability set of the pursuer x absorbs the attainability set of the evader y. This instant T is the minimal guaranteed time of the game (16.11) termination.

Let us present an explicit form of a countercontrol $u(v)$ of the pursuer on the interval $[0,T]$ that solves the problem of approach. The strategy of the pursuer is defined by the relationships

$$\gamma z^0 - \rho\tilde{v} = \sqrt{(1-\lambda)\gamma + \lambda\tilde{v}^2}\,\mu w, \qquad \|w\| \leqslant 1,$$
$$\tilde{u}(\tilde{v}) = -\sqrt{(1-\lambda)\gamma + \lambda\tilde{v}^2}\,w,$$

whence

$$\tilde{u}(\tilde{v}) = \frac{-\gamma z^0 + \rho\tilde{v}}{\mu}.$$

Then, using the substitution $u = \mu\tilde{u}$, $v = \rho\tilde{v}$ and the quadratic equation $F(\gamma,\tilde{v}) = 0$, we deduce that

$$u(v) = v - \gamma(v)\cdot z^0,$$

where

$$\gamma(v) = \frac{1}{\|z^0\|^2}\times\Bigg\{\langle z^0, v\rangle + \frac{\mu(\mu-\rho)}{2} + \sqrt{\left(\langle z^0, v\rangle + \frac{\mu(\mu-\rho)}{2}\right)^2 + \|z^0\|^2\|v\|^2\left(\frac{\mu-\rho}{\rho}\right)}\Bigg\}.$$

Thus, this control assures solving the problem (16.11), (16.12) no later than at time T (16.13).

References

1. Chikrii, A.A.: Conflict Controlled Processes. Kluwer, Boston (1997)
2. Kisielewicz M.: Differential Inclusions and Optimal Control. Kluwer, Boston (1991)
3. Kuratowski, K., Ryll-Nardzewski, C.: A general theorem on selectors. Bull. Acad. Pol. Sci. Warsaw **13**, 397–403 (1965)
4. Pontryagin, L.S.: Selected Scientific Papers, vol. 2. Nauka, Moscow (1988) (in Russian)
5. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)

Chapter 17
Anglers' Fishing Problem

Anna Karpowicz and Krzysztof Szajowski

Abstract The model considered here will be formulated in relation to the "fishing problem," even if other applications of it are much more obvious. The angler goes fishing, using various techniques, and has at most two fishing rods. He buys a fishing pass for a fixed time. The fish are caught using different methods according to renewal processes. The fish's value and the interarrival times are given by the sequences of independent, identically distributed random variables with known distribution functions. This forms the marked renewal–reward process. The angler's measure of satisfaction is given by the difference between the utility function, depending on the value of the fish caught, and the cost function connected with the time of fishing. In this way, the angler's relative opinion about the methods of fishing is modeled. The angler's aim is to derive as much satisfaction as possible, and additionally he must leave the lake by a fixed time. Therefore, his goal is to find two optimal stopping times to maximize his satisfaction. At the first moment, he changes his technique, e.g., by discarding one rod and using the other one exclusively. Next, he decides when he should end his outing. These stopping times must be shorter than the fixed time of fishing. Dynamic programming methods are used to find these two optimal stopping times and to specify the expected satisfaction of the angler at these times.

Keywords Stopping time • Optimal stopping • Dynamic programming • Semi-Markov process • Marked renewal process • Renewal–reward process • Infinitesimal generator • Fishing problem • Bilateral approach • Stopping game

---

A. Karpowicz
Bank Zachodni WBK, Rynek 9/11, 50-950 Wrocław, Poland
e-mail: a.m.karpowicz@gmail.com

K. Szajowski (✉)
Institute of Mathematics and Computer Science, Wybrzeże
Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: Krzysztof.Szajowski@pwr.wroc.pl

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_17,

17.1 Introduction

Before we start our analysis of the double optimal stopping problem (cf. the idea of multiple stopping for stochastic sequences in Haggstrom [8] and Nikolaev [16]) for the marked renewal process related to the angler's behavior, let us present the so-called fishing problem. One of the first authors to consider the basic version of this problem was Starr [19], and further generalizations were done by Starr and Woodroofe [21], Starr et al. [20], and Kramer and Starr [14]. A detailed review of papers related to the fishing problem was presented by Ferguson [7]. A simple formulation of the fishing problem, where the angler changes his location or technique before leaving his fishing spot, was done by Karpowicz [12]. We extend the problem to a more advanced model by taking into account the various fishing techniques used at the same time (the parallel renewal–reward processes or the multivariate renewal–reward process). This is motivated by the natural, more precise models of the known, real applications of the fishing problem. The typical process of software testing consists in checking subroutines. Initially, many kinds of bugs are searched for. Consecutive stopping times are moments when the expert stops general testing of modules and starts checking the most important, dangerous types of errors. Similarly, in proofreading, it is natural to look for typographic and grammatical errors at the same time. Next, we look for errors in language use.

Since various works are done by different groups of experts, it is natural that they would compete against each other. If the first period work comprises one group and the second period requires other experts, then they can be players in a game among themselves. In this case, the proposed solution is to find the Nash equilibrium where player strategies are the stopping times.

The applied techniques of modeling and finding the optimal solution are similar to those used in the formulation and solution of the optimal stopping problem for the risk process. Both models are based on the methodology explicated by Boshuizen and Gouweleeuw [1]. The background mathematics for further reading are the monographs by Brémaud [3], Davis [4], and Shiryaev [18]. The optimal stopping problems for the risk process are considered in papers by Jensen [10], Ferenstein and Sierociński [6], and Muciek [15]. A similar problem for a risk process with a disruption (i.e., when the probability structure of the considered process is changed at one moment θ) was analyzed by Ferenstein and Pasternak-Winiarski [5]. The model of the last paper brings to mind the change in fishing methods considered here. However, such a change is considered as made by a decision maker and it is not uncontrolled, automatically simply consequence on the basis of environment type.

The following two sections present details of the model. A slight modification of the background assumption by the adoption of multivariate tools (two rods) and the possible control of their numbers in use yields a different structure of the base model (the underlying process, sets of strategies—admissible filtrations and stopping times). This modified structure allows the introduction of a new kind of knowledge selection, which consequently leads to a game model of the angler's expedition problem in Sects. 17.1.2 and 17.2.2. Following a general formulation of the problem, a version of the problem for a detailed solution will be chosen. However, the solution is presented as a scalable procedure dependent on parameters that depends on various circumstances. It is not difficult to adopt a solution to a wide range of natural cases.

17.1.1 Single Angler's Expedition

An angler goes fishing. He buys a fishing pass for a fixed time t_0, which gives him the right to use at most two rods. The total cost of fishing depends on the real time of each equipment usage and the number of rods used simultaneously. The angler starts fishing with two rods up to moment s. The effect on each rod can be modeled by the renewal processes $\{N_i(t), t \geq 0\}$, where $N_i(t)$ is the number of fish caught with rod i, $i \in \mathfrak{A} := \{1,2\}$, during time t. Let us combine them together into a marked renewal process. The usage of the ith rod by time t generates cost $c_i : [0,t_0] \to \Re$ (when a rod is used simultaneously with other rods, it will be denoted by an index depending on the set of rods, e.g., $\mathfrak{a}$, $c_i^{\mathfrak{a}}$), and the reward is represented by independent, identically distributed (i.i.d.) random variables $X_1^{\{i\}}, X_2^{\{i\}}, \ldots$ (*the value of the fish caught with the ith rod*) with cumulative distribution function H_i.[1] The streams of two kinds of fish are mutually independent and are independent of the sequence of random moments when the fish have been caught. The two-vector process $\overrightarrow{N}(t) = (N_1(t), N_2(t))$, $t \geq 0$, can be represented also by a sequence of random variables T_n taking values in $[0,\infty]$ such that

$$\begin{aligned} T_0 &= 0, \\ T_n < \infty &\Rightarrow T_n < T_{n+1}, \end{aligned} \tag{17.1}$$

for $n \in \mathbb{N}$, and a sequence of $\mathfrak{A}$-valued random variables $\mathfrak{z}_n$ for $n \in \mathbb{N} \cup \{0\}$ (Chap. 2 in Brémaud [3]; Jacobsen [9]). The random variable T_n denotes the moment of catching the nth fish ($T_0 = 0$) of any kind and the random variable $\mathfrak{z}_n$ indicates the class of fish to which the nth fish belongs. The processes $N_i(t)$ can be defined by the sequence $\{(T_n, \mathfrak{z}_n)\}_{n=0}^{\infty}$ as

$$N_i(t) = \sum_{n=1}^{\infty} \mathbb{I}_{\{T_n \leq t\}} \mathbb{I}_{\{\mathfrak{z}_n = i\}}. \tag{17.2}$$

Both the two-variate process $\overrightarrow{N}(t)$ and the double sequence $\{(T_n, \mathfrak{z}_n)\}_{n=0}^{\infty}$ are called a *two-variate renewal process*. Optimal stopping problems for the compound risk process based on the *two-variate renewal process* were considered by Szajowski [22].

[1] The following convention is used throughout the paper: $\overrightarrow{x} = (x_1, x_2, \ldots, x_s)$ for the ordered collection of the elements $\{x_i\}_{i=1}^{s}$.

Let us define, for $i \in \mathfrak{A}$ and $k \in \mathbb{N}$, the sequence

$$n_0^{\{i\}} = 0,$$
$$n_{k+1}^{\{i\}} = \inf\{n > n_k^{\{i\}} : \mathfrak{z}_n = i\} \tag{17.3}$$

and set $T_k^{\{i\}} = T_{n_k^{\{i\}}}$. Let us define random variables $S_n^{\{i\}} = T_n^{\{i\}} - T_{n-1}^{\{i\}}$ assuming that they are i.i.d. with the continuous, cumulative distribution function $F_i(t) = \mathbf{P}(S_n^{\{i\}} \leq t)$ and the conditional distribution function $F_i^s(t) = \mathbf{P}(S_n^{\{i\}} \leq t | S_n^{\{i\}} \geq s)$. In Sect. 17.2.1 the alternative representation of the two-variate renewal process will be proposed. There is also a mild extension of the model in which the stream of events after some moment changes to another stream of events.

Remark 17.1. In various procedures, it is necessary to localize the events in a group of renewal processes. Let $\mathfrak{C}$ be the set of indices related to such a group. The sequence $\{n_k^{\mathfrak{C}}\}_{k=0}^{\infty}$ such that $n_0^{\mathfrak{C}} = 0$, $n_{k+1}^{\mathfrak{C}} := \inf\{n > n_k^{\mathfrak{C}} : \mathfrak{z}_n \in \mathfrak{C}\}$ has an obvious meaning.

Analogously, $n^{\mathfrak{C}}(t) := \inf\{n : T_n > t, \mathfrak{z}_n \in \mathfrak{C}\}$.

Let $i, j \in \mathfrak{A}$. The angler's satisfaction measure (*the net reward*) at the period $\mathfrak{a}$ from rod i is the difference between the utility function $g_i^{\mathfrak{a}} : [0,\infty)^2 \times \mathfrak{A} \times \mathbb{R}^+ \to [0, G_i^{\mathfrak{a}}]$ which can be interpreted as the reward from the ith rod when the last success was on rod j and, additionally, is dependent on the value of the fish caught, the moment of the result evaluation, and the cost function $c_i^{\mathfrak{a}} : [0,t_0] \to [0, C_i^{\mathfrak{a}}]$ reflecting the cost of duration of the angler's expedition. We assume that $g_i^{\mathfrak{a}}$ and $c_i^{\mathfrak{a}}$ are continuous and bounded and, additionally, $c_i^{\mathfrak{a}}$ are differentiable. Each fishing method evaluation is based on different utility functions and cost functions. In this way, the angler's relative opinion about them is modeled.

The angler can change his method of fishing at moment s and decide to use only one rod. It could be one of the rods used up to moment s or the other one. Even though the rod used after s is the one chosen from those used before s, its effectiveness could be different before and after s. After moment s the modeling process is the renewal–reward one with the stream of i.i.d. random variables $X_n^{\{3\}}$ at moments $T_n^{\{3\}}$ [i.e., appearing according to the renewal process $N_3(t)$]. Following these arguments, the mathematical model of catching fish, and their value after s, could (and in practice should) be different from those for rods used before s. The reason for the reduction in the number of rods could be their better effectiveness. The value of the fish that have been caught up to time t, if the change in fishing technique took place at time s, is given by

$$M_t^s = \sum_{i \in \mathfrak{A}} \sum_{n=1}^{N_i(s \wedge t)} X_n^{\{i\}} + \sum_{n=1}^{N_3((t-s)^+)} X_n^{\{3\}} = M_{s \wedge t} + \sum_{n=1}^{N_3((t-s)^+)} X_n^{\{3\}},$$

where $M_t^{\{i\}} = \sum_{n=1}^{N_i(t)} X_n^{\{i\}}$ and $M_t = \sum_{i=1}^{2} M_t^{\{i\}}$. We denote $\overrightarrow{M}_t = (M_t^{\{1\}}, M_t^{\{2\}})$. Let $Z(s,t)$ denote the angler's payoff for stopping at time t (end of expedition) if the change in the fishing method took place at time s. The natural filtration related to the double-indexed process $Z(s,t)$ is $\mathcal{F}_t^s = \sigma\{0 \leq u \leq s \leq v \leq t : Z(u,v)\}$. If the effect of extending the expedition after s is described by $g_j^{\mathfrak{b}} : \mathbb{R}^{+2} \times \mathfrak{A} \times [0,t_0] \times \mathbb{R} \times [0,t_0] \to [0,G_j^{\mathfrak{b}}]$, $j \in \mathfrak{B}$, minus the additional cost of time $c_j^{\mathfrak{b}}(\cdot)$, where $c_j^{\mathfrak{b}} : [0,t_0] \to [0,C_j^{\mathfrak{b}}]$ [when card$(\mathfrak{B}) = 1$, then index j will be abandoned, and $c^{\mathfrak{b}} = \sum_{j\in\mathfrak{B}} c_j^{\mathfrak{b}}$ will be used, which will be adequate]. Then the payoff can be expressed as

$$Z(s,t) = \begin{cases} g^{\mathfrak{a}}(\overrightarrow{M}_t, \mathfrak{z}_{N(t)}, t) - c^{\mathfrak{a}}(t) & \text{if } t < s \leq t_0, \\ g^{\mathfrak{a}}(\overrightarrow{M}_s, \mathfrak{z}_{N(s)}, s) - c^{\mathfrak{a}}(s) & \\ \quad + g^{\mathfrak{b}}(\overrightarrow{M}_s, \mathfrak{z}_{N(s)}, s, M_t^s, t) - c^{\mathfrak{b}}(t-s) & \text{if } s \leq t \leq t_0, \\ -C & \text{if } t_0 < t, \end{cases} \tag{17.4}$$

where the function $c^{\mathfrak{a}}(t)$, $g^{\mathfrak{a}}(\overrightarrow{m}, i, t)$ and the constant C can be taken as follows: $c^{\mathfrak{a}}(t) = \sum_{i=1}^2 c_i^{\mathfrak{a}}(t)$, $g^{\mathfrak{a}}(\overrightarrow{M}_s, j, t) = \sum_{i=1}^2 g_i^{\mathfrak{a}}(\overrightarrow{M}_t, j, t)$, $C = C_1^{\mathfrak{a}} + C_2^{\mathfrak{a}} + C^{\mathfrak{b}}$. With the notation $w^{\mathfrak{b}}(\overrightarrow{m}, i, s, \widetilde{m}, t) = w^{\mathfrak{a}}(\overrightarrow{m}, i, s) + g^{\mathfrak{b}}(\overrightarrow{m}, i, s, \widetilde{m}, t) - c^{\mathfrak{b}}(t-s)$ and $w^{\mathfrak{a}}(\overrightarrow{m}, i, t) = g^{\mathfrak{a}}(\overrightarrow{m}, i, t) - c^{\mathfrak{a}}(t)$, formula (17.4) is reduced to

$$Z(s,t) = Z^{\{\mathfrak{z}_{N(t)}\}}(s,t)\mathrm{I}_{\{t<s\leq t_0\}} + Z^{\{\mathfrak{z}_{N(s)}\}}(s,t)\mathrm{I}_{\{s\leq t\}},$$

where

$$Z^{\{i\}}(s,t) = \mathrm{I}_{\{t<s\leq t_0\}} w^{\mathfrak{a}}(\overrightarrow{M}_t, i, t) + \mathrm{I}_{\{s\leq t\leq t_0\}} w^{\mathfrak{b}}(\overrightarrow{M}_s, i, s, M_t^s, t) - \mathrm{I}_{\{t_0<t\}} C.$$

17.1.2 The Competitive Fishing

When *the methods of fishing are operating by separate anglers*, a stopping random field can be built based on the structure of the marked renewal–reward process as a model of the competitive expedition results. One possible definition of payoff is based on the assumption that each player has his own account related to the exploration of the fishery. The states of the accounts depend on who forces the first stop for changing the technique, under what circumstances, and what techniques they choose. The first stopping moment, the minimum of stopping moments chosen by the players, is, after the moment of the event (catching fish), T_n with rod $\mathfrak{z}_n$, and the reward functions depend on the type of fishing that gives recent caught fish (i.e., j, where $j = \mathfrak{z}_n$). The player's payoff $w_i^{\mathfrak{a}}(\overrightarrow{m}, j, t) = g_i^{\mathfrak{a}}(\overrightarrow{m}, j, t) - c_i^{\mathfrak{a}}(t)$. The part of the payoff that depends on the second chosen moment, which stops the expedition, is different for the player who forces the change in fishing methods (the leader) by himself and the other for the opponent. The leader is the one responsible for determining the expedition deadline.

Let us assume for a while that the ith player, $i = 1, 2$, takes his opponent's rod and gives his own rod to his opponent. *It is not a crucial assumption at any rate, and the method of fishing after the change can be different from both available methods* *before the considered moment.* The method of treatment of the case without this assumption will be explained later (p. 335), when the behavior of the player in the second part of the expedition is formulated. Define the function

$$\tilde{w}_i^{\mathfrak{b}}(\overrightarrow{m},j,s,k,\widetilde{m},t) = \tilde{w}_i^{\mathfrak{a}}(\overrightarrow{m},j,s) + \tilde{g}_i^{\mathfrak{b}}(\overrightarrow{m},j,s,k,\widetilde{m},t) - c^{\mathfrak{b}}(t-s)$$

for $j \in \mathfrak{A}$, $k \in \mathfrak{B}$, where j is the rod with which the fish had been caught just before the moment of the first stop and k is the technique used by the ith player after the change (the denotation $-k$ is used for a complementary rod or player who has decided what is appropriate). It describes the case where the player deciding to change the method chooses the perspective technique of fishing as the first one. Presumably he will explore the best methods with improvements and the second angler will use the rod that is not used by the leader. The payoff of the players, when the ith player is the one who forces the first stop, has the following form:

$$Z_i(j,s,t) = \mathrm{I}_{\{t\le s\le t_0\}}\tilde{g}_i^{\mathfrak{a}}(\overrightarrow{M}_t,j,t) + \mathrm{I}_{\{s<t\le t_0\}}\tilde{w}_i^{\mathfrak{b}}(\overrightarrow{M}_s,i,s,-i,M_t^s,t) - \mathrm{I}_{\{t_0<t\}}C, \tag{17.5}$$

$$Z_{-i}(j,s,t) = \mathrm{I}_{\{t\le s\le t_0\}}\tilde{g}_{-i}^{\mathfrak{a}}(\overrightarrow{M}_t,j,t) + \mathrm{I}_{\{s<t\le t_0\}}\tilde{w}_{-i}^{\mathfrak{b}}(\overrightarrow{M}_s,i,s,i,M_t^s,t) - \mathrm{I}_{\{t_0<t\}}C. \tag{17.6}$$

In the preceding payoffs it is assumed that the final stop can be declared at any moment. Each players declares changes in techniques right after an event with his rod (catching fish with his rod) as long as on the opponent's rod there is no event. The details of the strategy sets and the solution concept are formulated in subsequent parts of this paper.

The extension considered here is motivated by the natural, more precise models of known real applications of the fishing problem. The typical process of software testing consists of checking subroutines. Various types of bugs can be discovered in this way. Each problem with subroutines generates the cost of bug removal and increases the value of the software. It depends on the types of bugs found. Preliminary testing requires various types of experts. The stable version of subroutines can be kept by less advanced computer scientists. The consecutive stopping times are moments when the expert of a certain class stops testing one module and another tester starts checking. The procedure for proofreading is similar.

17.2 The Optimization Problem and a Two-Person Game

17.2.1 Filtrations and Markov Moments

Let the sequences of pairs $\{(T_n,\mathfrak{z}_n)\}_{n=0}^{\infty}$ be a two-variate renewal process ($\mathfrak{A}$-marked renewal process) defined on $(\Omega,\mathcal{F},\mathbf{P})$. According to the denotation of the previous section, there are three renewal processes $\{T_n^{\{i\}}\}_{n=0}^{\infty}$, $i=1,2,3$, and they are denoted by $T_n = T^{\{\mathfrak{z}_n\}}_{N_{\mathfrak{z}_n}(T_n)}$. There are also three renewal–reward processes $\{(T^{\{i\}}_n, X^{\{i\}}_n)\}_{n=0}^{\infty}$, $i = 1,2,3$. By convention let us denote $X_n = X^{\{\mathfrak{z}_n\}}_{N_{\mathfrak{z}_n}(T_n)}$. The following σ-field generated by the history of the $\mathfrak{A}$-marked renewal processes are defined by

$$\mathcal{F}_t = \mathcal{F}^{\mathfrak{A}}_t = \sigma(X_0, T_0, \mathfrak{z}_0 \ldots, X_{N(t)}, T_{N(t)}, \mathfrak{z}_{N(t)}) \tag{17.7}$$

for $t \geq 0$. This σ-field can be defined as

$$\mathcal{F}^{\mathfrak{A}}_t = \sigma\{(\overrightarrow{N}(s), X_{N(s)}, \mathfrak{z}_{N(s)}), 0 \leq s \leq t,\ i \in \mathfrak{A}\}.$$

Definition 17.1. Let $\mathcal{T}$ be a set of stopping times with respect to σ-fields $\{\mathcal{F}_t\}$, $t \geq 0$, defined by (17.7). The restricted sets of stopping times are

$$\mathcal{T}_{n,K} = \{\tau \in \mathcal{T} :\ \tau \geq 0, T_n \leq \tau \leq T_K\} \tag{17.8}$$

for $n \in \mathbb{N}$ and $n < K$ are subsets of $\mathcal{T}$. The elements of $\mathcal{T}_{n,K}$ are denoted $\tau_{n,K}$.

The stopping times $\tau \in \mathcal{T}$ have a nice representation that will be helpful in the solution of the optimal stopping problems for the renewal processes [3]. A crucial role in our subsequent considerations will be played by such a representation. The following lemma is for unrestricted stopping times.

Lemma 17.1. *If $\tau \in \mathcal{T}$, then there exist $R_n \in Mes(\mathcal{F}_n)$ such that the condition $\tau \wedge T_{n+1} = (T_n + R_n) \wedge T_{n+1}$ on $\{\tau \geq T_n\}$ a.s. is fulfilled.*

Various restrictions in the class of admissible stopping times will change this representation. Some examples of subclasses of $\mathcal{T}$ are formulated here (Lemma 17.1). Only a few of them are used in the optimization problems investigated in this paper (Corollary 17.1).

Let $\mathcal{F}_{s,t} = \sigma(\mathcal{F}^{\mathfrak{A}}_s, X^{\{3\}}_0, T^{\{3\}}_0, \ldots, X^{\{3\}}_{N_3((t-s))^+}, T^{\{3\}}_{N_3((t-s)^+)})$ be the σ-field generated by all events up to time t if the switch at time s from a two-variate renewal process to another renewal process took place. For simplicity of notation we set $\mathcal{F}^{\{i\}}_n := \mathcal{F}_{T^{\{i\}}_n}$, $\mathcal{F}_n := \mathcal{F}_{T_n}$, $\mathcal{F}^s_n := \mathcal{F}_{s,T^{\{3\}}_n}$.[2] Let $Mes(\mathcal{F}_n)$ ($Mes(\mathcal{F}^{\{i\}}_n)$) denote the set of nonnegative and $\mathcal{F}_n$ ($\mathcal{F}^{\{i\}}_n$)-measurable random variables. Henceforth, $\mathcal{T}$ and $\mathcal{T}^s$ will stand for the sets of stopping times with respect to σ-fields $\mathcal{F}_s$ and $\{\mathcal{F}_{s,t}, 0 \leq s \leq t\}$, respectively. Furthermore, we can define for $n \in \mathbb{N}$ and $n \leq K$ the sets

1. $\mathcal{T}^{\{i\}}_{n,K} = \{\tau \in \mathcal{T} : \tau \geq 0,\ T^{\{i\}}_n \leq \tau \leq T_K\}$;
2. $\mathcal{T}^{\{i\}}_n = \{\tau \in \mathcal{T} : \tau \geq T^{\{i\}}_n\}$;

[2]For the optimization problem there are two epochs: before the first stop, where there are some payoffs, the model of stream of events, and after the first stop, when there are other payoffs and different streams of events. In Sect. 17.3, this will be emphasized by adopting adequate denotations.

3. $\bar{\mathcal{T}}_{n,K}^{\{i,\mathfrak{A}\{-i\}\}} = \left\{\tau \in \mathcal{T} : \tau \geq 0,\ T_n^{\{i\}} \leq \tau \leq T_K, \forall_k \tau \notin \left[T_k^{\mathfrak{A}^{-i}}, T_{k+1}^{\mathfrak{A}^{-i}} \vee T_{n^{\{i\}}(T_k^{\mathfrak{A}^{-i}})}^{\{i\}}\right]\right\}$,
   where $\mathfrak{A}^{\{-i\}} := \mathfrak{A} \setminus \{i\}$, $T_k^{\mathfrak{A}^{-i}} := \min_{\{j \in \mathfrak{A}\{-i\}\}} \left\{T_{n^{\{j\}}(T_k^{\{i\}})}^{\{j\}}\right\}$;
4. $\bar{\mathcal{T}}_n^{\{i\}} = \left\{\tau \in \mathcal{T} : \tau \geq T_n^{\{i\}}, \forall_k \tau \notin \left[T_k^{\mathfrak{A}^{-i}}, T_{k+1}^{\mathfrak{A}^{-i}} \vee T_{n^{\{i\}}(T_k^{\mathfrak{A}^{-i}})}^{\{i\}}\right]\right\}$;
5. $\mathcal{T}_{n,K}^s = \{\tau \in \mathcal{T}^s : 0 \leq s \leq \tau,\ T_n^{\{3\}} \leq \tau \leq T_K\}$.

The stopping times $\tau \in \mathcal{T}^{\{i\}}$ and $\tau \in \bar{\mathcal{T}}^{\{i\}}$ can also be represented as shown in Lemma 17.1.

Lemma 17.2. *Let the index $i \in \mathfrak{A}$ be chosen and fixed.*

1. *For every $\tau \in \mathcal{T}^{\{i\}}$ and $n \in \mathbb{N}$ there exist $R_n^{\{i\}} \in Mes(\mathcal{F}_n^{\{i\}})$ such that $\tau \wedge T_{n+1}^{\{i\}} = (T_n^{\{i\}} + R_n^{\{i\}}) \wedge T_{n+1}^{\{i\}}$ on $\{\tau^{\{i\}} \geq T_n^{\{i\}}\}$ a.s. is fulfilled.*
2. *If $\tau \in \bar{\mathcal{T}}^{\{i\}}$ and $n \in \mathbb{N}$, then there exist $R_n^{\{i\}} \in Mes(\mathcal{F}_n^{\{i\}})$ such that the condition $\tau \wedge T_{n+1}^{\{i\}} = (T_n^{\{i\}} + R_n^{\{i\}}) \wedge T_{n+1}^{\{i\}}$ on $\{\tau \geq T_n^{\{i\}}\}$ a.s. is fulfilled.*

Obviously the angler wants to have as much satisfaction as possible and must leave the lake before the fixed time. Therefore, his goal is to find two optimal stopping times $\tau^{\mathfrak{a}^*}$ and $\tau^{\mathfrak{b}^*}$ so that the expected gain is maximized:

$$\mathbf{E}Z(\tau^{\mathfrak{a}^*}, \tau^{\mathfrak{b}^*}) = \sup_{\tau^{\mathfrak{a}} \in \mathcal{T}} \sup_{\tau^{\mathfrak{b}} \in \mathcal{T}^{\tau^{\mathfrak{a}}}} \mathbf{E}Z(\tau^{\mathfrak{a}}, \tau^{\mathfrak{b}}), \tag{17.9}$$

where $\tau^{\mathfrak{a}^*}$ corresponds to the moment when he should change to the more effective rod and $\tau^{\mathfrak{b}^*}$ to the moment when he should stop fishing. These stopping moments should appear before the fixed time of fishing t_0. The process $Z(s,t)$ is piecewise-deterministic and belongs to the class of semi-Markov processes. The optimal stopping time of similar semi-Markov processes was studied by Boshuizen and Gouweleeuw [1] and the multivariate point process by Boshuizen [2]. Here the structure of multivariate processes is revealed and their importance for the model is shown. We use dynamic programming methods to find these two optimal stopping times and to specify the expected satisfaction of the angler. The method of the solution is similar to those used by Karpowicz and Szajowski [13], Karpowicz [12], and Szajowski [22]. Let us first observe that by the properties of conditional expectation we have

$$\mathbf{E}Z(\tau^{\mathfrak{a}^*}, \tau^{\mathfrak{b}^*}) = \sup_{\tau^{\mathfrak{a}} \in \mathcal{T}} \mathbf{E}\{\mathbf{E}\left[Z(\tau^{\mathfrak{a}}, \tau^{\mathfrak{b}^*})|\mathcal{F}_{\tau^{\mathfrak{a}}}\right]\} = \sup_{\tau^{\mathfrak{a}} \in \mathcal{T}} \mathbf{E}J(\tau^{\mathfrak{a}}),$$

where

$$J(s) = \mathbf{E}\left[Z(s, \tau^{\mathfrak{b}^*})|\mathcal{F}_s\right] = \operatorname*{ess\,sup}_{\tau^{\mathfrak{b}} \in \mathcal{T}^s} \mathbf{E}\left[Z(s, \tau^{\mathfrak{b}})|\mathcal{F}_s\right]. \tag{17.10}$$

Therefore, to find $\tau^{\mathfrak{a}^*}$ and $\tau^{\mathfrak{b}^*}$, we must calculate $J(s)$ first. The process $J(s)$ corresponds to the value of the revenue function in one stopping problem if the observation starts at moment s.

17.2.2 Angler Games

Based on the consideration of Sect. 17.1.2, a version of competitive fishing is formulated here. There are two anglers, each using one method of fishing at the beginning of an expedition and an additional fishing period after a certain moment by another method up to the moment chosen by a certain rule. The random field that is the model of payoffs in such a case is given by (17.5) and (17.6). The final segment starts at the moment when one of the anglers wants it. Let $\tau_i \in \bar{\mathcal{T}}^{\{i\}}$, $i = \mathfrak{A}$, be the strategies of the players for stopping their individual fishing period and switching to the time segment that is stopped at moment σ determined by one angler (let us call that angler the *leader*). The payoffs of the players are

$$\psi_i(\tau_1,\tau_2) = Z_i(\mathfrak{z}_{N(\tau_1\wedge\tau_2)},\tau_1\wedge\tau_2,\sigma^{\tau_1\wedge\tau_2})\mathrm{I}_{\{\tau_1\neq\tau_2\}} + Z_i(\mathfrak{z}_{N(\tau_1\wedge\tau_2)}\wedge\mathfrak{z}_{N(\tau_1\wedge\tau_2)},\tau_1\wedge\tau_2,\sigma^{\tau_1\wedge\tau_2})\mathrm{I}_{\{\tau_1=\tau_2\}}. \quad (17.11)$$

Assignment of the leader in the case $\tau_1 = \tau_2$ is arbitrary but defined. The aim is to find a pair $(\tau_1^\star,\tau_2^\star)$ of stopping times such that for $i \in \{1,2\}$ we have

$$\mathbf{E}\psi_i(\tau_i^\star,\tau_{-i}^\star) \geq \mathbf{E}\psi_i(\tau_i,\tau_{-i}^\star). \quad (17.12)$$

The optimization problem of the angler and the game between two anglers will involve the construction of the optimal second stopping moment.

17.3 Construction of the Optimal Second Stopping Time

In this section, we will find the solution to one stopping problem defined by (17.10). We will first solve the problem for a fixed number of fish caught and then consider the case with an infinite stream of fish caught. In this section we fix s, the moment when the change took place, and $m = M_s$, the mass of the fish at time s. Taking into account various models of fishing after the first stop, it is necessary to admit various models of event streams. Assume that the moments of successive fish caught after the first stop are $T_n^{\{3\}}$ and the times between the events are i.i.d. with continuous, cumulative distribution function $\mathrm{F}(t)$ with the density function $\mathrm{f}(t)$and the fish's value represented by i.i.d. random variables with distribution function $\mathrm{H}(t)$ [for convenience this part of the expedition is modeled by the renewal process denoted $(T_n^{\{3\}},X_n^{\{3\}})$].

17.3.1 Fixed Number of Fish Caught

In this subsection we are looking for the optimal stopping time $\tau_{0,K}^{\mathfrak{b}^*} := \tau_K^{\mathfrak{b}^*}$

$$\mathbf{E}\left[Z(s,\tau_K^{\mathfrak{b}^*})|\mathcal{F}_s\right] = \operatorname*{ess\,sup}_{\tau_K^{\mathfrak{b}}\in\mathcal{T}_{0,K}^s}\mathbf{E}\left[Z(s,\tau_K^{\mathfrak{b}})|\mathcal{F}_s\right], \tag{17.13}$$

where $s \geq 0$ is a fixed time when the position is changed and K is the maximum number of fish that can be caught. Let us define

$$\Gamma_{n,K}^s = \operatorname*{ess\,sup}_{\tau_{n,K}^{\mathfrak{b}}\in\mathcal{T}_{n,K}^s}\mathbf{E}\left[Z(s,\tau_{n,K}^{\mathfrak{b}})|\mathcal{F}_n^s\right] = \mathbf{E}\left[Z(s,\tau_{n,K}^{\mathfrak{b}^*})|\mathcal{F}_n^s\right], \quad n = K,\ldots,1,0 \tag{17.14}$$

and observe that $\Gamma_{K,K}^s = Z(s,T_K^{\{3\}})$. In subsequent considerations, we will use the representation of stopping time formulated in Lemmas 17.1 and 17.2. The exact form of the stopping strategies is given in the following corollary.

Corollary 17.1. *Let $i \in \mathfrak{A}$. If $\tau^{\mathfrak{a}} \in \mathcal{T}^{\{i\}}$, $\tau^{\mathfrak{b}} \in \mathcal{T}^s$, then there exist $R_n^{\mathfrak{a}} \in Mes(\mathcal{F}_n^{\{i\}})$ and $R_n^{\mathfrak{b}} \in Mes(\mathcal{F}_n^s)$, respectively, such that for conditions $\tau^{\mathfrak{a}} \wedge T_{n+1}^{\{i\}} = (T_n^{\{i\}} + R_n^{\mathfrak{a}}) \wedge T_{n+1}^{\{i\}}$ on $\{\tau^{\mathfrak{a}} \geq T_n^{\{i\}}\}$ a.s. and $\tau^{\mathfrak{b}} \wedge T_{n+1}^{\{3\}} = (T_n^{\{3\}} + R_n^{\mathfrak{a}}) \wedge T_{n+1}^{\{3\}}$ on $\{\tau^{\mathfrak{a}} \geq s \wedge T_n^{\{3\}}\}$ a.s. are valid.*

Now we can derive the dynamic programming equations satisfied by $\Gamma_{n,K}^s$. To simplify the notation, we can write $M_t = M_t^s$ for $t \leq s$, $\widehat{M}_n^{\{1\}} = M_{T_n^1}$, $M_n^s = M_{T_n^{\{3\}}}^s$, and $\bar{\mathrm{F}}_i = 1 - \mathrm{F}_i$. The payoff functions are simplified here to $\hat{g}^{\mathfrak{a}}(m) = g^{\mathfrak{a}}(m_1,m_2,i,t)\mathrm{I}_{\{m_1+m_2=m\}}(m_1,m_2)$, $\hat{g}^{\mathfrak{b}}(m) = g^{\mathfrak{b}}(m_1,m_2,i,s,\widetilde{m},t)\mathrm{I}_{\{\widetilde{m}-m_1-m_2=m\}}$.

Lemma 17.3. *Let $s \geq 0$ be the moment of changing the fishing place. For $n = K-1$, $K-2,\ldots,0$*

$$\Gamma_{K,K}^s = Z\left(s,T_K^{\{3\}}\right),$$

$$\Gamma_{n,K}^s = \operatorname*{ess\,sup}_{R_n^{\mathfrak{b}}\in Mes(\mathcal{F}_n^s)} \vartheta_{n,K}(M_s,s,M_n^s,T_n^{\{3\}},R_n^{\mathfrak{b}}) \text{ a.s.,} \tag{17.15}$$

where

$$\vartheta_{n,K}(m,s,\widetilde{m},t,r) = \mathrm{I}_{\{t\leq t_0\}}\left\{\bar{F}(r)\left[\mathrm{I}_{\{r\leq t_0-t\}}\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t+r) - C\mathrm{I}_{\{r>t_0-t\}}\right]\right.$$
$$\left. + E\left[\mathrm{I}_{\{S_{n+1}^{\{3\}}\leq r\}}\Gamma_{n+1,K}^s|\mathcal{F}_n^s\right]\right\} - C\mathrm{I}_{\{t>t_0\}}$$

and there exists $R_n^{\mathfrak{b}\star} \in Mes(\mathcal{F}_n^s)$ *such that*

$$\Gamma_{n,K}^s = \vartheta_{n,K}(M_s, s, M_n^s, T_n^{\{3\}}, R_n^{\mathfrak{b}\star}) \text{ a.s.}, \tag{17.16}$$

$$\tau_{n,K}^{\mathfrak{b}^*} = \begin{cases} \tau_{n+1,K}^{\mathfrak{b}^*} & \text{if } R_n^{\mathfrak{b}^*} \geq S_{n+1}^{\{3\}}, \\ T_n^{\{3\}} + R_n^{\mathfrak{b}^*} & \text{if } R_n^{\mathfrak{b}^*} < S_{n+1}^{\{3\}}, \end{cases} \tag{17.17}$$

$\tau_{K,K}^{\mathfrak{b}^*} = T_K^{\{3\}}$, *and* $\hat{w}^{\mathfrak{b}}(m,s,\tilde{m},t) = \hat{w}^{\mathfrak{a}}(m,s) + \hat{g}^{\mathfrak{b}}(\tilde{m}-m) - c^{\mathfrak{b}}(t-s)$ *where* $\hat{w}^{\mathfrak{a}}(m,t) = \hat{g}^{\mathfrak{a}}(m) - c^{\mathfrak{a}}(t)$.

Remark 17.2. Let $\{R_n^{\mathfrak{b}^*}\}_{n=1}^K$, $R_K^{\mathfrak{b}^*} = 0$, be a sequence of $\mathcal{F}_n^s$-measurable random variables, $n = 1, 2, \ldots, K$, and $\eta_{n,K}^{\star s} = K \wedge \inf\{i \geq n : R_i^{\mathfrak{b}\star} < S_{i+1}^{\{3\}}\}$. Then $\Gamma_{n,K}^s = \mathbf{E}\left[Z(s, \tau_{n,K}^{\mathfrak{b}^*})|\mathcal{F}_n^s\right]$ for $n \leq K-1$, where $\tau_{n,K}^{\mathfrak{b}^*} = T_{\eta_{n,K}^{\star s}} + R_{\eta_{n,K}^{\star s}}^{\mathfrak{b}\star}$.

Proof of Remark 17.2. It is a consequence of an optimal choice $R_n^{\mathfrak{b}\star}$ in (17.15). □

Proof of Lemma 17.3. The form of $\Gamma_{n,K}^s$ for the case $T_n^{\{3\}} > t_0$ is obvious from (17.4) and (17.14). Let us assume (17.15) and (17.16) for $n+1, n+2, \ldots, K$. For any $\tau \in \mathcal{T}_{n,K}^s$ (i.e., $\tau \geq T_n^{\{3\}}$) we have $\{\tau < T_{n+1}^{\{3\}}\} = \{\tau \wedge T_{n+1}^{\{3\}} < T_{n+1}^{\{3\}}\} = \{T_n^{\{3\}} + R_n^{\mathfrak{b}} < T_{n+1}^{\{3\}}\}$. This implies

$$\left\{\tau < T_{n+1}^{\{3\}}\right\} = \left\{S_{n+1}^{\{3\}} > R_n^{\mathfrak{b}}\right\}, \quad \left\{\tau \geq T_{n+1}^{\{3\}}\right\} = \left\{S_{n+1}^{\{3\}} \leq R_n^{\mathfrak{b}}\right\}. \tag{17.18}$$

Suppose that $T_{K-1}^{\{3\}} \leq t_0$ and take any $\tau_{K-1,K}^{\mathfrak{b}} \in \mathcal{T}_{K-1,K}^s$. According to (17.18) and the properties of conditional expectation,

$$\begin{aligned} \mathbf{E}[Z(s,\tau)|\mathcal{F}_n^s] &= \mathbf{E}\left[\mathrm{I}_{\left\{S_{n+1}^{\{3\}} \leq R_n^{\mathfrak{b}}\right\}} \mathbf{E}\left[Z\left(s, \tau \vee T_{n+1}^{\{3\}}\right) \middle| \mathcal{F}_{n+1}^s\right] \middle| \mathcal{F}_n\right] \\ &\quad + \mathbf{E}\left[\mathrm{I}_{\left\{S_{n+1}^{\{3\}} > R_n^{\mathfrak{b}}\right\}} Z\left(s, \tau \wedge T_{n+1}^{\{3\}}\right) \middle| \mathcal{F}_n^s\right] \\ &= \mathrm{I}_{\left\{R_n^{\mathfrak{b}} \leq t_0 - T_n\right\}} \bar{\mathrm{F}}(R_n) \hat{w}^{\mathfrak{b}}\left(M_s, s, M_n^s, T_n^{\{3\}} + R_n^{\mathfrak{b}}\right) \\ &\quad + \mathbf{E}\left[\mathrm{I}_{\left\{S_{n+1}^{\{3\}} \leq R_n^{\mathfrak{b}}\right\}} \mathbf{E}\left[Z\left(s, \tau \vee T_{n+1}^{\{3\}}\right) \middle| \mathcal{F}_{n+1}^s\right] \middle| \mathcal{F}_n^s\right]. \end{aligned}$$

Let $\sigma \in \mathcal{T}_{n+1}^{\mathfrak{b}}$. For every $\tau \in \mathcal{T}_n^s$ we have

$$\tau = \begin{cases} \sigma & \text{if } R_n^{\mathfrak{b}} \geq S_{n+1}^{\{3\}}, \\ T_n^{\{3\}} + R_n^{\mathfrak{b}} & \text{if } R_n^{\mathfrak{b}} < S_{n+1}^{\{3\}}. \end{cases}$$

We also have

$$\begin{aligned}\mathbf{E}[Z(s,\tau)|\mathcal{F}_n^s] &= \mathbf{E}\left[\mathrm{I}_{\{S_{n+1}^{\{3\}}\leq R_n^{\mathfrak{b}}\}}\mathbf{E}[Z(s,\sigma)|\mathcal{F}_{n+1}^s]|\mathcal{F}_n\right]\\ &\quad+\mathrm{I}_{\{R_n^{\mathfrak{b}}\leq t_0-T_n\}}\bar{\mathrm{F}}\left(R_n^{\mathfrak{b}}\right)\hat{w}^{\mathfrak{b}}\left(M_s,s,M_n^s,T_n^{\{3\}}+R_n^{\mathfrak{b}}\right)\\ &\leq \sup_{R\in Mes(\mathcal{F}_n^s)}\left\{\mathbf{E}\left[\mathrm{I}_{\left\{S_{n+1}^{\{3\}}\leq R\right\}}\Gamma_{n+1,K}^s\middle|\mathcal{F}_n\right]\right.\\ &\quad\left.+\mathrm{I}_{\{R\leq t_0-T_n\}}\bar{\mathrm{F}}(R)\hat{w}^{\mathfrak{b}}\left(M_s,s,M_n^s,T_n^{\{3\}}+R\right)\right\}=\mathbf{E}\left[Z\left(s,\tau_{n,K}^{\star}\right)\middle|\mathcal{F}_n^s\right]\end{aligned}$$

It follows that $\sup_{\tau\in\mathcal{T}_n^s}\mathbf{E}[Z(s,\tau)|\mathcal{F}_n^s]\leq\mathbf{E}[Z(s,\tau_{n,K}^{\star})|\mathcal{F}_n^s]\leq\sup_{\tau\in\mathcal{T}_n^{\mathfrak{b}}}\mathbf{E}[Z(s,\tau)|\mathcal{F}_n^s]$, where the last inequality is due to the fact that $\tau_{n,K}^{\star}\in\mathcal{T}_{n,K}^s$. We apply the induction hypothesis, which completes the proof. □

Lemma 17.4. *$\Gamma_{n,K}^s=\gamma_{K-n}^{s,M_s}(M_n^s,T_n^{\{3\}})$ for $n=K,\ldots,0$, where the sequence of functions $\gamma_j^{s,m}$ is given recursively as follows:*

$$\begin{aligned}\gamma_0^{s,m}(\widetilde{m},t)&=\mathrm{I}_{\{t\leq t_0\}}\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t)-C\mathrm{I}_{\{t>t_0\}},\\ \gamma_j^{s,m}(\widetilde{m},t)&=\mathrm{I}_{\{t\leq t_0\}}\sup_{r\geq 0}\kappa_{\gamma_{j-1}^{s,m}}^{\mathfrak{b}}(m,s,\widetilde{m},t,r)-C\mathrm{I}_{\{t>t_0\}},\end{aligned}\tag{17.19}$$

where

$$\begin{aligned}\kappa_\delta^{\mathfrak{b}}(m,s,\widetilde{m},t,r)&=\bar{F}(r)\left[\mathrm{I}_{\{r\leq t_0-t\}}\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t+r)-C\mathrm{I}_{\{r>t_0-t\}}\right]\\ &\quad+\int_0^r\mathrm{d}F(z)\int_0^\infty\delta\left(\widetilde{m}+x,t+z\right)\mathrm{d}H(x).\end{aligned}$$

Proof of Lemma 17.4. Since the case for $t>t_0$ is obvious, let us assume that $T_n^{\{3\}}\leq t_0$ for $n\in\{0,\ldots,K-1\}$. According to Lemma 17.3, we obtain $\Gamma_{K,K}^s=\gamma_0^{s,M_s}(M_K^s,T_K^{\{3\}})$, and thus the proposition is satisfied for $n=K$. Let $n=K-1$; then Lemma 17.3 and the induction hypothesis lead to

$$\begin{aligned}\Gamma_{K-1,K}^s=&\operatorname*{ess\,sup}_{R_{K-1}^{\mathfrak{b}}\in Mes(\mathcal{F}_{s,K-1})}\left\{\bar{\mathrm{F}}\left(R_{K-1}^{\mathfrak{b}}\right)\left[\mathrm{I}_{\left\{R_{K-1}^{\mathfrak{b}}\leq t_0-T_{K-1}^{\{3\}}\right\}}\hat{w}^{\mathfrak{b}}\left(M_s,s,M_{K-1}^s,T_{K-1}^{\{3\}}+R_{K-1}^{\mathfrak{b}}\right)\right.\right.\\ &\left.\left.-C\mathrm{I}_{\left\{R_{K-1}^{\mathfrak{b}}>t_0-T_{K-1}^{\{3\}}\right\}}\right]+\mathbf{E}\left[\mathrm{I}_{\left\{S_K^{\{3\}}\leq R_{K-1}^{\mathfrak{b}}\right\}}\gamma_0^{s,M_s}\left(M_K^s,T_K^{\{3\}}\right)\middle|\mathcal{F}_{s,K-1}\right]\right\}\ a.s.,\end{aligned}$$

where $M_K^s=M_{K-1}^s+X_K^{\{3\}}$, $T_K^{\{3\}}=T_{K-1}^{\{3\}}+S_K^{\{3\}}$, and the random variables $X_K^{\{3\}}$ and $S_K^{\{3\}}$ are independent of $\mathcal{F}_{s,K-1}$. Moreover, $R_{K-1}^{\mathfrak{b}}$, M_{K-1}^s, and $T_{K-1}^{\{3\}}$ are $\mathcal{F}_{s,K-1}$-measurable. It follows that

$$\Gamma^s_{K-1,K} = \operatorname*{ess\,sup}_{R^{\mathfrak{b}}_{K-1}\in Mes(\mathcal{F}_{s,K-1})} \left\{ \bar{\mathrm{F}}\left(R^{\mathfrak{b}}_{K-1}\right)\left[\mathrm{I}_{\{R^{\mathfrak{b}}_{K-1}\le t_0 - T^{\{3\}}_{K-1}\}} \hat{w}^{\mathfrak{b}}\left(M_s, s, M^s_{K-1}, T^{\{3\}}_{K-1} + R^{\mathfrak{b}}_{K-1}\right) \right.\right.$$
$$\left.\left. - C\mathrm{I}_{\left\{R^{\mathfrak{b}}_{K-1} > t_0 - T^{\{3\}}_{K-1}\right\}}\right] + \int_0^{R^{\mathfrak{b}}_{K-1}} \mathrm{dF}(z) \int_0^\infty \gamma_0^{s,M_s}\left(M^s_{K-1}+x, T^{\{3\}}_{K-1}+z\right)\mathrm{dH}(x)\right\}$$
$$= \gamma_1^{s,M_s}\left(M^s_{K-1}, T^{\{3\}}_{K-1}\right) \ a.s.$$

Let $n \in \{1,\ldots,K-1\}$ and suppose that $\Gamma^s_{n,K} = \gamma^{s,M_s}_{K-n}(M^s_n, T^{\{3\}}_n)$. As was done previously, we conclude by Lemma 17.3 and the induction hypothesis that

$$\Gamma^s_{n-1,K} = \operatorname*{ess\,sup}_{R^{\mathfrak{b}}_{n-1}\in Mes(\mathcal{F}^s_{n-1})} \left\{ \bar{\mathrm{F}}\left(R^{\mathfrak{b}}_{n-1}\right)\left[\mathrm{I}_{\left\{R^{\mathfrak{b}}_{n-1}\le t_0 - T^{\{3\}}_{n-1}\right\}} \hat{w}^{\mathfrak{b}}\left(M_s, s, M^s_{n-1}, T^{\{3\}}_{n-1} + R^{\mathfrak{b}}_{n-1}\right) \right.\right.$$
$$\left.\left. - C\mathrm{I}_{\left\{R^{\mathfrak{b}}_{n-1} > t_0 - T^{\{3\}}_{n-1}\right\}}\right] + \int_0^{R^{\mathfrak{b}}_{n-1}} \mathrm{dF}(s) \int_0^\infty \gamma_{K-n}^{s,M_s}\left(M^s_{n-1}+x, T^{\{3\}}_{n-1}+s\right)\mathrm{dH}(x)\right\} \ a.s.$$

Therefore, $\Gamma^s_{n-1,K} = \gamma^{s,M_s}_{K-(n-1)}(M^s_{n-1}, T^{\{3\}}_{n-1})$. □

Henceforth we will use α_i to denote the hazard rate of the distribution F_i (i.e., $\alpha_i = f_i/\bar{\mathrm{F}}_i$), and to shorten the notation, we set $\Delta^\bullet(a) = \mathbf{E}\left[\hat{g}^\bullet(a + X^{\{i\}}) - \hat{g}^\bullet(a)\right]$, where $\bullet$ can be $\mathfrak{a}$ or $\mathfrak{b}$.

Remark 17.3. The sequence of functions $\gamma_j^{s,m}$ can be expressed as

$$\gamma_0^{s,m}(\widetilde{m}, t) = \mathrm{I}_{\{t\le t_0\}}\hat{w}^{\mathfrak{b}}(m, s, \widetilde{m}, t) - C\mathrm{I}_{\{t>t_0\}},$$
$$\gamma_j^{s,m}(\widetilde{m}, t) = \mathrm{I}_{\{t\le t_0\}}\left\{\hat{w}^{\mathfrak{b}}(m, s, \widetilde{m}, t) + y_j^{\mathfrak{b}}(\widetilde{m} - m, t - s, t_0 - t)\right\}, -C\mathrm{I}_{\{t>t_0\}}$$

and $y_j^{\mathfrak{b}}(a,b,c)$ is given recursively as follows:

$$y_0^{\mathfrak{b}}(a,b,c) = 0$$
$$y_j^{\mathfrak{b}}(a,b,c) = \max_{0\le r\le c} \phi^{\mathfrak{b}}_{y^{\mathfrak{b}}_{j-1}}(a,b,c,r),$$

where $\phi^{\mathfrak{b}}_\delta(a,b,c,r) = \int_0^r \bar{\mathrm{F}}(z)\{\alpha(z)\left[\Delta^{\mathfrak{b}}(a) + \mathbf{E}\delta(a + X^{\{3\}}, b+z, c-z)\right] - c^{\mathfrak{b}\prime}(b+z)\}\mathrm{d}z$, and F is the c.d.f. of $S^{\{3\}}$ [$\alpha(t)$ is the hazard rate of the distribution of $S^{\{3\}}$].

Proof of Remark 17.3. Clearly

$$\int_0^r \mathrm{dF}(s)\int_0^\infty \gamma_{j-1}^{s,m}(\widetilde{m}+x, t+s)\,\mathrm{dH}(x) = \mathbf{E}\left[\mathrm{I}_{\{S^{\{3\}}\le r\}}\gamma_{j-1}^{s,m}\left(\widetilde{m} + X^{\{3\}}, t + S^{\{3\}}\right)\right],$$

where $X^{\{3\}}$ has the *c.d.f.* H. Since F is continuous and $\kappa^{\mathfrak{b}}_{\gamma^{s,m}_{j-1}}(m,s,\widetilde{m},t,r)$ is bounded and continuous for $t \in \mathbb{R}^+ \setminus \{t_0\}$, the supremum in (17.19) can be changed into a maximum. Let $r > t_0 - t$; then

$$\begin{aligned}
\kappa^{\mathfrak{b}}_{\gamma^{s,m}_{j-1}}(m,s,\widetilde{m},t,r) &= \mathbf{E}\left[\mathrm{I}_{\{S^{\{3\}} \le t_0 - t\}}\gamma^{s,m}_{j-1}\left(\widetilde{m}+X^{\{3\}},t+S^{\{3\}}\right)\right] - C\bar{\mathrm{F}}(t_0-t) \\
&\le \mathbf{E}\left[\mathrm{I}_{\{S^{\{3\}} \le t_0 - t\}}\gamma^{s,m}_{j-1}\left(\widetilde{m}+X^{\{3\}},t+S^{\{3\}}\right)\right] + \bar{\mathrm{F}}(t_0-t)\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t_0) \\
&= \kappa^{\mathfrak{b}}_{\gamma^{s,m}_{j-1}}(m,s,\widetilde{m},t,t_0-t).
\end{aligned}$$

The preceding calculations cause that $\gamma^{s,m}_j(\widetilde{m},t) = \mathrm{I}_{\{t \le t_0\}} \max_{0 \le r \le t_0 - t} \varphi_j(m,s,\widetilde{m},t,r)$ $-C\mathrm{I}_{\{t>t_0\}}$, where $\varphi_j(m,s,\widetilde{m},t,r) = \bar{\mathrm{F}}(r)\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t+r) + \mathbf{E}\Big[\mathrm{I}_{\{S^{\{3\}} \le r\}}\gamma^{s,m}_{j-1}$ $(\widetilde{m}+X^{\{3\}},t+S^{\{3\}})\Big]$. Obviously for $S^{\{3\}} \le r$ and $r \le t_0 - t$ we have $S^{\{3\}} \le t_0$; therefore, we can consider the cases $t \le t_0$ and $t > t_0$ separately. Let $t \le t_0$; then $\gamma^{s,m}_0(\widetilde{m},t) = \hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t)$, and the hypothesis is true for $j=0$. The task is now to calculate $\gamma^{s,m}_{j+1}(\widetilde{m},t)$ given $\gamma^{s,m}_j(\cdot,\cdot)$. The induction hypothesis implies that for $t \le t_0$

$$\begin{aligned}
\varphi_{j+1}(m,s,\widetilde{m},t,r) &= \bar{\mathrm{F}}(r)\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t+r) + \mathbf{E}\left[\mathrm{I}_{\{S^{\{3\}} \le r\}}\gamma^{s,m}_j(\widetilde{m}+X^{\{3\}},t+S^{\{3\}})\right] \\
&= \hat{g}^{\mathfrak{a}}(m) - c^{\mathfrak{a}}(s) + \bar{\mathrm{F}}(r)\left[\hat{g}^{\mathfrak{b}}(\widetilde{m}-m) - c^{\mathfrak{b}}(t-s+r)\right] \\
&\quad + \int_0^r \mathrm{f}(z)\{\mathbf{E}\hat{g}^{\mathfrak{b}}(\widetilde{m}-m+X^{\{3\}}) - c^{\mathfrak{b}}(t-s+z) \\
&\quad + \mathbf{E}y^{\mathfrak{b}}_j(\widetilde{m}-m+X^{\{3\}},t-s+z,t_0-t-z)\}\mathrm{d}z.
\end{aligned}$$

It is clear that for any a and b

$$\begin{aligned}
\bar{\mathrm{F}}(r)\left[\hat{g}^{\mathfrak{b}}(a) - c^{\mathfrak{b}}(b+r)\right] &= \hat{g}^{\mathfrak{b}}(a) - c^{\mathfrak{b}}(b) - \int_0^r \{\mathrm{f}(z)\left[\hat{g}^{\mathfrak{b}}(a) - c^{\mathfrak{b}}(b+z)\right] \\
&\quad + \bar{\mathrm{F}}(z)c^{\mathfrak{b}\prime}(b+z)\}\mathrm{d}z;
\end{aligned}$$

therefore,

$$\begin{aligned}
\varphi_{j+1}(m,s,\widetilde{m},t,r) &= \hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t) + \int_0^r \bar{\mathrm{F}}(z)\{\alpha(z)[\Delta^{\mathfrak{b}}(\widetilde{m}-m) \\
&\quad + \mathbf{E}y^{\mathfrak{b}}_j(\widetilde{m}-m+X^{\{3\}},t-s+z,t_0-t-z)] - c^{\mathfrak{b}\prime}(t-s+z)\}\mathrm{d}z,
\end{aligned}$$

which proves the theorem. The case for $t > t_0$ is trivial. □

Following the methods of Ferenstein and Sierociński [6], we find the second optimal stopping time. Let $\mathbb{B} = \mathfrak{B}([0,\infty) \times [0,t_0] \times [0,t_0])$ be the space of all bounded, continuous functions with the norm $\|\delta\| = \sup_{a,b,c} |\delta(a,b,c)|$. It is easy to check that $\mathbb{B}$ with the norm supremum is complete space. The operator $\boldsymbol{\Phi}^{\mathfrak{b}} : \mathbb{B} \to \mathbb{B}$ is defined by

$$(\boldsymbol{\Phi}^{\mathfrak{b}} \delta)(a,b,c) = \max_{0 \le r \le c} \phi_{\delta}^{\mathfrak{b}}(a,b,c,r). \tag{17.20}$$

Let us observe that $y_j^{\mathfrak{b}}(a,b,c) = (\boldsymbol{\Phi}^{\mathfrak{b}} y_{j-1}^{\mathfrak{b}})(a,b,c)$. Remark 17.3 now implies that there exists a function $r_j^{\mathfrak{b}*}(a,b,c)$ such that $y_j^{\mathfrak{b}}(a,b,c) = \phi_{y_{j-1}^{\mathfrak{b}}}^{\mathfrak{b}}(a,b,c,r^{\mathfrak{b}*}{}_j(a,b,c))$, and this gives

$$\begin{aligned} \gamma_j^{s,m}(\widetilde{m},t) \;=\; \mathrm{I}_{\{t \le t_0\}} \Big\{ & \hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t) + \phi_{y_{j-1}^{\mathfrak{b}}}^{\mathfrak{b}}(\widetilde{m}-m, t-s, t_0-t, r_j^{\mathfrak{b}*} \\ & \times (\widetilde{m}-m, t-s, t_0-t)) \Big\} - C\mathrm{I}_{\{t > t_0\}}. \end{aligned}$$

The consequence of the foregoing considerations is a theorem that determines the optimal stopping times $\tau_{n,K}^{\mathfrak{b}*}$ in the following manner.

Theorem 17.1. *Let $R_i^{\mathfrak{b}*} = r_{K-i}^{\mathfrak{b}*}(M_i^s - M_s, T_i^{\{3\}} - s, t_0 - T_i^{\{3\}})$ for $i = 0,1,\ldots,K$. Moreover, $\eta_{n,K}^s = K \wedge \inf\{i \ge n : R_i^{\mathfrak{b}*} < S_{i+1}^{\{3\}}\}$; then the stopping time $\tau_{n,K}^{\mathfrak{b}*} = T_{\eta_{n,K}^s}^{\{3\}} + R_{\eta_{n,K}^s}^{\mathfrak{b}*}$ is optimal in the class $\mathcal{T}_{n,K}^s$ and $\Gamma_{n,K}^s = \mathbf{E}\left[Z(s,\tau_{n,K}^{\mathfrak{b}*})|\mathcal{F}_n^s\right]$.*

17.3.2 Infinite Number of Fish Caught

The task is now to find the function $J(s)$ and the stopping time $\tau^{\mathfrak{b}*}$ that is optimal in the class $\mathcal{T}^s$. To obtain the solution of one stopping problem for an infinite number of fish caught, it is necessary to set the restriction $\mathrm{F}(t_0) < 1$.

Lemma 17.5. *If $F(t_0) < 1$, then the operator $\boldsymbol{\Phi}^{\mathfrak{b}} : \mathbb{B} \to \mathbb{B}$ defined by (17.20) is a contraction.*

Proof of Lemma 17.5. Let $\delta_i \in \mathbb{B}$ for $i \in \{1,2\}$. There exists ρ_i such that $(\boldsymbol{\Phi}^{\mathfrak{b}} \delta_i)(a,b,c) = \phi_{\delta_i}^{\mathfrak{b}}(a,b,c,\rho_i)$. We thus obtain

$$\begin{aligned} (\boldsymbol{\Phi}^{\mathfrak{b}} \delta_1)(a,b,c) - (\boldsymbol{\Phi}^{\mathfrak{b}} \delta_2)(a,b,c) &= \phi_{\delta_1}^{\mathfrak{b}}(a,b,c,\rho_1) - \phi_{\delta_2}^{\mathfrak{b}}(a,b,c,\rho_2) \\ &\le \phi_{\delta_1}^{\mathfrak{b}}(a,b,c,\rho_1) - \phi_{\delta_2}^{\mathfrak{b}}(a,b,c,\rho_1) \\ &= \int_0^{\rho_1} \mathrm{dF}(z) \int_0^{\infty} [\delta_1 - \delta_2](a+x, b+z, c-s)\, \mathrm{dH}(x) \\ &\le \int_0^{\rho_1} \mathrm{dF}(z) \int_0^{\infty} \sup_{a,b,c} |[\delta_1 - \delta_2](a,b,c)| \mathrm{dH}(x) \\ &\le \mathrm{F}(c)\, \|\delta_1 - \delta_2\| \le \mathrm{F}(t_0)\, \|\delta_1 - \delta_2\| \le \mathbf{C}\, \|\delta_1 - \delta_2\|, \end{aligned}$$

where $0 \leq \mathbf{C} < 1$. Similarly, as before, $(\Phi^{\mathfrak{b}}\delta_2)(a,b,c) - (\Phi^{\mathfrak{b}}\delta_1)(a,b,c) \leq \mathbf{C}\,\|\delta_2 - \delta_1\|$. Finally, we conclude that $\left\|\Phi^{\mathfrak{b}}\delta_1 - \Phi^{\mathfrak{b}}\delta_2\right\| \leq \mathbf{C}\,\|\delta_1 - \delta_2\|$, which completes the proof. □

Applying Remark 17.3, Lemma 17.5, and the fixed point theorem we conclude the following remark

Remark 17.4. There exists $y^{\mathfrak{b}} \in \mathbb{B}$ such that $y^{\mathfrak{b}} = \Phi^{\mathfrak{b}} y^{\mathfrak{b}}$ and $\lim_{K\to\infty} \|y^{\mathfrak{b}}_K - y^{\mathfrak{b}}\| = 0$.

According to the preceding remark, $y^{\mathfrak{b}}$ is the uniform limit of $y^{\mathfrak{b}}_K$ when K tends to infinity, which implies that $y^{\mathfrak{b}}$ is measurable and $\gamma^{s,m} = \lim_{K\to\infty} \gamma^{s,m}_K$ is given by

$$\gamma^{s,m}(\widetilde{m},t) = \mathrm{I}_{\{t\leq t_0\}}\left[\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t) + y^{\mathfrak{b}}(\widetilde{m}-m, t-s, t_0-t)\right] - C\mathrm{I}_{\{t>t_0\}}. \qquad (17.21)$$

We can now calculate the optimal strategy and the expected gain after changing locations.

Theorem 17.2. *If $F(t_0) < 1$ and has the density function f, then:*

(i) *For $n \in \mathbb{N}$ the limit $\tau^{\mathfrak{b}\star}_n = \lim_{K\to\infty} \tau^{\mathfrak{b}*}_{n,K}$ a.s. exists and $\tau^{\mathfrak{b}\star}_n \leq t_0$ is an optimal stopping rule in the set $\mathcal{T}^s \cap \{\tau \geq T^{\{3\}}_n\}$;*
(ii) $\mathbf{E}\left[Z(s,\tau^{\mathfrak{b}\star}_n)|\mathcal{F}^s_n\right] = \gamma^{s,m}(M^s_n, T^{\{3\}}_n)$ *a.s.*

Proof. (i) Let us first prove the existence of $\tau^{\mathfrak{b}\star}_n$. By the definition of $\Gamma^s_{n,K+1}$, we have

$$\begin{aligned}\Gamma^s_{n,K+1} &= \operatorname*{ess\,sup}_{\tau\in\mathcal{T}^s_{n,K+1}} \mathbf{E}\left[Z(s,\tau)|\mathcal{F}^s_n\right] = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}^s_{n,K}} \mathbf{E}\left[Z(s,\tau)|\mathcal{F}^s_n\right] \vee \operatorname*{ess\,sup}_{\tau\in\mathcal{T}^s_{K,K+1}} \mathbf{E}\left[Z(s,\tau)|\mathcal{F}^s_n\right] \\ &= \mathbf{E}\left[Z(s,\tau^{\mathfrak{b}*}_{n,K})|\mathcal{F}^s_n\right] \vee \mathbf{E}\left[Z(s,\sigma^*)|\mathcal{F}^s_n\right],\end{aligned}$$

and thus we observe that $\tau^{\mathfrak{b}*}_{n,K+1}$ is equal to $\tau^{\mathfrak{b}*}_{n,K}$ or σ^*, where $\tau^{\mathfrak{b}*}_{n,K} \in \mathcal{T}^s_{n,K}$ and $\sigma^* \in \mathcal{T}^s_{K,K+1}$, respectively. It follows that $\tau^{\mathfrak{b}*}_{n,K+1} \geq \tau^{\mathfrak{b}*}_{n,K}$, which implies that the sequence $\tau^{\mathfrak{b}*}_{n,K}$ is nondecreasing with respect to K. Moreover, $R^{\mathfrak{b}*}_i \leq t_0 - T^{\{3\}}_i$ for all $i \in \{0,\ldots,K\}$; thus $\tau^{\mathfrak{b}*}_{n,K} \leq t_0$, and therefore $\tau^{\mathfrak{b}\star}_n \leq t_0$ exists.

Let us now look at the process $\xi^s(t) = (t, M^s_t, V(t))$, where s is fixed and $V(t) = t - T^{\{3\}}_{N_3(t)}$. $\xi^s(t)$ is a Markov process with the state space $[s,t_0] \times [m,\infty) \times [0,\infty)$. In a general case, the infinitesimal operator for ξ^s is given by

$$\begin{aligned}Ap^{s,m}(t,\widetilde{m},v) = {} & \frac{\partial}{\partial t} p^{s,m}(t,\widetilde{m},v) + \frac{\partial}{\partial v} p^{s,m}(t,\widetilde{m},v) \\ & + \alpha(v)\left\{\int_{\mathbb{R}^+} p^{s,m}(t,x,0)\,\mathrm{dH}(x) - p^{s,m}(t,\widetilde{m},v)\right\},\end{aligned}$$

where $p^{s,m}(t,\widetilde{m},v) : [0,\infty)\times[0,\infty)\times[0,\infty)\to\mathbb{R}$ is continuous, bounded, and measurable with bounded left-hand derivatives with respect to t and v [1, 17]. For $t\geq s$ the process $Z(s,t)$ can be expressed as $Z(s,t)=p^{s,m}(\xi^s(t))$, where

$$p^{s,m}(\xi^s(t))=\begin{cases}\hat{g}^{\mathfrak{a}}(M_s)-c^{\mathfrak{a}}(s)+\hat{g}^{\mathfrak{b}}(M_t^s-M_s)-c^{\mathfrak{b}}(t-s) & \text{if } s\leq t\leq t_0,\\ -C & \text{if } t_0<t.\end{cases}$$

It follows easily that in our case $Ap^{s,m}(t,\widetilde{m},v)=0$ for $t_0<t$ and

$$Ap^{s,m}(t,\widetilde{m},v)=\alpha(v)[\mathbf{E}\hat{g}^{\mathfrak{b}}(\widetilde{m}+X^{\{3\}}-m)-\hat{g}^{\mathfrak{b}}(\widetilde{m}-m)]-c^{\mathfrak{b}\prime}(t-s) \quad (17.22)$$

for $s\leq t\leq t_0$. The process $p^{s,m}(\xi^s(t))-p^{s,m}(\xi^s(s))-\int_s^t(Ap^{s,m})(\xi^s(z))\,\mathrm{d}z$ is a martingale with respect to $\sigma\{\xi^s(z),z\leq t\}$, which is the same as $\mathcal{F}_t^s$. This can be found in [4]. Since $\tau_{n,K}^{\mathfrak{b}*}\leq t_0$, applying Dynkin's formula we obtain

$$\mathbf{E}\left[p^{s,m}(\xi^s(\tau_{n,K}^{\mathfrak{b}*}))|\mathcal{F}_n^s\right]-p^{s,m}(\xi^s(T_n^{\{3\}}))=\mathbf{E}\left[\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}(Ap^{s,m})(\xi^s(z))\,\mathrm{d}z|\mathcal{F}_n^s\right] \; a.s. \quad (17.23)$$

From (17.22) we conclude that

$$\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}(Ap^{s,m})(\xi^s(z))\,\mathrm{d}z=\left[\mathbf{E}\hat{g}^{\mathfrak{b}}(M_n^s+X^{\{3\}}-m)-\hat{g}^{\mathfrak{b}}(M_n^s-m)\right]\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}\alpha(z-T_n^{\{3\}})\,\mathrm{d}z$$
$$-\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}c^{\mathfrak{b}\prime}(z-s)\,\mathrm{d}z.$$

Moreover, let us check that

$$\left|\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}\alpha\left(z-T_n^{\{3\}}\right)\mathrm{d}z\right|\leq\frac{1}{\bar{\mathrm{F}}(t_0)}\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}\mathrm{f}\left(z-T_n^{\{3\}}\right)\mathrm{d}z\leq\frac{1}{\bar{\mathrm{F}}(t_0)}<\infty,$$
$$\left|\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}c^{\mathfrak{b}\prime}(z-s)\,\mathrm{d}z\right|=\left|c^{\mathfrak{b}}\left(\tau_{n,K}^{\mathfrak{b}*}-s\right)-c^{\mathfrak{b}}\left(T_n^{\{3\}}-s\right)\right|<\infty,$$
$$\left|\mathbf{E}\hat{g}^{\mathfrak{b}}\left(M_n^s+X^{\{3\}}-m\right)-\hat{g}^{\mathfrak{b}}\left(M_n^s-m\right)\right|<\infty,$$

where the last two inequalities result from the fact that the functions $\hat{g}^{\mathfrak{b}}$ and $c^{\mathfrak{b}}$ are bounded. On account of the preceding observation we can use the dominated convergence theorem and

$$\lim_{K\to\infty}\mathbf{E}\left[\int_{T_n^{\{3\}}}^{\tau_{n,K}^{\mathfrak{b}*}}(Ap^{s,m})(\xi^s(z))\,\mathrm{d}z|\mathcal{F}_n^s\right]=\mathbf{E}\left[\int_{T_n^{\{3\}}}^{\tau_n^{\mathfrak{b}\star}}(Ap^{s,m})(\xi^s(z))\,\mathrm{d}z|\mathcal{F}_n^s\right]. \quad (17.24)$$

Since $\tau_n^{\mathfrak{b}\star} \le t_0$, applying Dynkin's formula to the left-hand side of (17.24) we conclude that

$$\mathbf{E}\left[\int_{T_n^{\{3\}}}^{\tau_n^{\mathfrak{b}\star}} (Ap^{s,m})\,(\xi^s(z))\,\mathrm{d}z \Big| \mathcal{F}_n^s\right] = \mathbf{E}\left[p^{s,m}\left(\xi^s\left(\tau_n^{\mathfrak{b}\star}\right)\right)|\mathcal{F}_n^s\right] - p^{s,m}\left(\xi^s\left(T_n^{\{3\}}\right)\right) \quad a.s. \quad (17.25)$$

Combining (17.23)–(17.25) we can see that

$$\lim_{K\to\infty} \mathbf{E}\left[p^{s,m}\left(\xi^s\left(\tau_{n,K}^{\mathfrak{b}^*}\right)\right)|\mathcal{F}_n^s\right] = \mathbf{E}\left[p^{s,m}\left(\xi^s\left(\tau_n^{\mathfrak{b}^*}\right)\right)|\mathcal{F}_n^s\right], \quad (17.26)$$

hence $\lim_{K\to\infty}\mathbf{E}[Z(s,\tau_{n,K}^{\mathfrak{b}^*})|\mathcal{F}_n^s] = \mathbf{E}[Z(s,\tau_n^{\mathfrak{b}^*})|\mathcal{F}_n^s]$. We next prove the optimality of $\tau_n^{\mathfrak{b}^*}$ in the class $\mathcal{T}^s \cap \{\tau_n^{\mathfrak{b}} \ge T_n^{\{3\}}\}$. Let $\tau \in \mathcal{T}^s \cap \{\tau_n^{\mathfrak{b}} \ge T_n^{\{3\}}\}$, and it is clear that $\tau \wedge T_K^{\{3\}} \in \mathcal{T}_{n,K}^s$. As $\tau_{n,K}^{\mathfrak{b}^*}$ is optimal in the class $\mathcal{T}_{n,K}^s$, we have

$$\lim_{K\to\infty} \mathbf{E}\left[p^{s,m}(\xi^s(\tau_{n,K}^{\mathfrak{b}^*}))|\mathcal{F}_n^s\right] \ge \lim_{K\to\infty} \mathbf{E}\left[p^{s,m}(\xi^s(\tau \wedge T_K^{\{3\}}))|\mathcal{F}_n^s\right]. \quad (17.27)$$

From (17.26) and (17.27) we conclude that $\mathbf{E}\left[p^{s,m}(\xi^s(\tau_n^{\mathfrak{b}^*}))|\mathcal{F}_n^s\right] \ge \mathbf{E}[p^{s,m}(\xi^s(\tau))|\mathcal{F}_n^s]$ for any stopping time $\tau \in \mathcal{T}^s \cap \{\tau \ge T_n^{\{3\}}\}$, which implies that $\tau_n^{\mathfrak{b}^*}$ is optimal in this class.

(ii) Lemma 17.4 and (17.26) lead to $\mathbf{E}[Z(s,\tau_n^{\mathfrak{b}^*})|\mathcal{F}_n^s] = \gamma^{s,M_s}(M_n^s, T_n^{\{3\}})$. □

The remainder of this section will be devoted to the proof of the left-hand differentiability of the function $\gamma^{s,m}(m,s)$ with respect to s. This property is necessary to construct the first optimal stopping time. First, let us briefly denote $\delta(0,0,c) \in \mathbb{B}$ by $\bar{\delta}(c)$.

Lemma 17.6. *Let* $\bar{v}(c) = \Phi^{\mathfrak{b}}\bar{\delta}(c)$, $\bar{\delta}(c) \in \mathbb{B}$ *and* $\left|\bar{\delta}'_+(c)\right| \le A_1$ *for* $c \in [0,t_0)$; *then* $\left|\bar{v}'_+(c)\right| \le A_2$.

Proof of Lemma 17.6. The derivative $\bar{v}'_+(c)$ exists because $\bar{v}(c) = \max_{0\le r\le c}\bar{\phi}^{\mathfrak{b}}(c,r)$, where $\bar{\phi}^{\mathfrak{b}}(c,r)$ is differentiable with respect to c and r. Fix $h \in (0,t_0-c)$ and define $\bar{\delta}_1(c) = \bar{\delta}(c+h) \in \mathbb{B}$ and $\bar{\delta}_2(c) = \bar{\delta}(c) \in \mathbb{B}$. Obviously, $\|\Phi^{\mathfrak{b}}\bar{\delta}_1 - \Phi^{\mathfrak{b}}\bar{\delta}_2\| \ge |\Phi^{\mathfrak{b}}\bar{\delta}_1(c) - \Phi^{\mathfrak{b}}\bar{\delta}_2(c)| = |\Phi^{\mathfrak{b}}\bar{\delta}(c+h) - \Phi^{\mathfrak{b}}\bar{\delta}(c)|$, and on the other hand using Taylor's formula for the right-hand derivatives we obtain

$$\left\|\bar{\delta}_1 - \bar{\delta}_2\right\| = \sup_c\left|\bar{\delta}(c+h) - \bar{\delta}(c)\right| \le h\sup_c\left|\bar{\delta}'_+(c)\right| + |o(h)|.$$

From the foregoing and Remark 17.8 it follows that

$$-\mathbf{C}\left\{\sup_c\left|\bar{\delta}'_+(c)\right| + \frac{|o(h)|}{h}\right\} \le \frac{\bar{v}(c+h) - \bar{v}(c)}{h} \le \mathbf{C}\left\{\sup_c\left|\bar{\delta}'_+(c)\right| + \frac{|o(h)|}{h}\right\},$$

and letting $h \to 0^+$ gives $\left|\bar{v}'_+(c)\right| \le \mathbf{C}A_1 = A_2$. □

The significance of Lemma 17.6 is such that the function $\bar{y}(t_0 - s)$ has a bounded left-hand derivative with respect to s for $s \in (0, t_0]$. The important consequence of this fact is shown by the following remark.

Remark 17.5. The function $\gamma^{s,m}$ can be expressed as $\gamma^{s,m}(m,s) = \mathrm{I}_{\{s \le t_0\}} u(m,s) - C\mathrm{I}_{\{s>t_0\}}$, where $u(m,s) = \hat{g}^{\mathfrak{a}}(m) - c^{\mathfrak{a}}(s) + \hat{g}^{\mathfrak{b}}(0) - c^{\mathfrak{b}}(0) + \bar{y}^{\mathfrak{b}}(t_0 - s)$ is continuous, bounded, and measurable with the bounded left-hand derivatives with respect to s.

At the end of this section, we determine the conditional value function of the second optimal stopping problem. According to (17.10), Theorem 17.2, and Remark 17.5, we have

$$J(s) = \mathbf{E}\left[Z(s, \tau^{\mathfrak{b}^*})|\mathcal{F}_s\right] = \gamma^{s,M_s}(M_s, s) \text{ a.s.} \tag{17.28}$$

17.4 Construction of the Optimal First Stopping Time

In this section, we formulate the solution of the double stopping problem. In the first epoch of the expedition, the admissible strategies (stopping times) depend on the formulation of the problem. For the optimization problem the most natural strategies are the stopping times from $\mathcal{T}$ (see the relevant problem considered in Szajowski [22]). However, when the bilateral problem is considered, the natural class of admissible strategies depends on who uses the strategy. It should be $\mathcal{T}^{\{i\}}$ for the ith player. Here the optimization problem with restriction to the strategies from $\mathcal{T}^{\{1\}}$ in the first epoch is investigated.

The function $u(m,s)$ has similar properties to those of the function $\hat{w}^{\mathfrak{b}}(m,s,\widetilde{m},t)$ and the process $J(s)$ has a similar structure to that of the process $Z(s,t)$. By this observation one can follow the calculations of Sect. 17.3 to obtain $J(s)$. Let us define again $\Gamma_{n,K} = \operatorname{ess\,sup}_{\tau^{\mathfrak{a}} \in \mathcal{T}_{n,K}} \mathbf{E}\left[J(\tau^{\mathfrak{a}})|\mathcal{F}_n\right]$, $n = K, \ldots, 1, 0$, which fulfills the following representation:

Lemma 17.7. *$\Gamma_{n,K} = \gamma_{K-n}(\widehat{M}_n^{\{1\}}, T_n^{\{1\}})$ for $n = K, \ldots, 0$, where the sequence of functions γ_j can be expressed as*

$$\begin{aligned}
\gamma_0(m,s) &= \mathrm{I}_{\{s \le t_0\}} u(m,s) - C\mathrm{I}_{\{s>t_0\}}, \\
\gamma_j(m,s) &= \mathrm{I}_{\{s \le t_0\}} \left\{ u(m,s) + y_j^{\mathfrak{a}}(m,s,t_0 - s) \right\}, -C\mathrm{I}_{\{s>t_0\}}
\end{aligned}$$

and $y_j^{\mathfrak{a}}(a,b,c)$ is given recursively as follows:

$$\begin{aligned}
y_0^{\mathfrak{a}}(a,b,c) &= 0, \\
y_j^{\mathfrak{a}}(a,b,c) &= \max_{0 \le r \le c} \phi_{y_{j-1}^{\mathfrak{a}}}^{\mathfrak{a}}(a,b,c,r),
\end{aligned}$$

where

$$\phi_{\delta}^{\mathfrak{a}}(a,b,c,r) = \int_0^r \bar{F}_1(z)\left\{\alpha_1(z)\left[\Delta^{\mathfrak{a}}(a)+\mathbf{E}\delta(a+x^{\{1\}},b+z,c-z)\right]\right.$$
$$\left. - \left(\bar{y}^{\mathfrak{b}\prime}_{-}(c-z)+c^{\mathfrak{a}\prime}(b+z)\right)\right\}\mathrm{d}z.$$

Lemma 17.7 corresponds to the combination of Lemma 17.4 and Remark 17.3 from Sect. 17.3.1. Let the operator $\Phi^{\mathfrak{a}}:\mathbb{B}\to\mathbb{B}$ be defined by

$$(\Phi^{\mathfrak{a}}\delta)(a,b,c) = \max_{0\le r\le c}\phi_{\delta}^{\mathfrak{a}}(a,b,c,r). \tag{17.29}$$

Lemma 17.7 implies that there exists a function $r_{1,j}^{*}(a,b,c)$ such that

$$\gamma_j(m,s) = \mathrm{I}_{\{s\le t_0\}}\left\{u(m,s)+\phi^{\mathfrak{a}}_{y^{\mathfrak{a}}_{j-1}}(m,s,t_0-s,r^{*}_{1,j}(m,s,t_0-s))\right\} - C\mathrm{I}_{\{s>t_0\}}.$$

We can now state the analog of Theorem 17.1.

Theorem 17.3. *Let $R_i^{\mathfrak{a}^*} = r^{\mathfrak{a}^*}_{K-i}(M_i,T_i^{\{1\}},t_0-T_i^{\{1\}})$ and $\eta_{n,K} = K\wedge\inf\{i\ge n : R_i^{\mathfrak{a}*} < S_{i+1}^{\{1\}}\}$; then $\tau^{\mathfrak{a}^*}_{n,K} = T^{\{1\}}_{\eta_{n,K}} + R^{\mathfrak{a}^*}_{\eta_{n,K}}$ is optimal in the class $\mathcal{T}_{n,K}$ and $\Gamma_{n,K} = E\left[J(\tau^{\mathfrak{a}^*}_{n,K})|\mathcal{F}_n\right]$.*

The following results may be proved in much the same way as in Sect. 17.3.

Lemma 17.8. *If $F_1(t_0)<1$, then the operator $\Phi^{\mathfrak{a}}:\mathbb{B}\to\mathbb{B}$ defined by (17.29) is a contraction.*

Remark 17.6. There exists $y^{\mathfrak{a}}\in\mathbb{B}$ such that $y^{\mathfrak{a}} = \Phi^{\mathfrak{a}}y^{\mathfrak{a}}$ and $\lim_{K\to\infty}\|y_K^{\mathfrak{a}} - y^{\mathfrak{a}}\| = 0$.

The preceding remark implies that $\gamma = \lim_{K\to\infty}\gamma_K$ is given by

$$\gamma(m,s) = \mathrm{I}_{\{s\le t_0\}}\left[u(m,s)+y^{\mathfrak{a}}(m,s,t_0-s)\right] - C\mathrm{I}_{\{s>t_0\}}. \tag{17.30}$$

We can now formulate our main results.

Theorem 17.4. *If $F_1(t_0)<1$ and has the density function f_1, then*

(i) *For $n\in\mathbb{N}$ the limit $\tau_n^{\mathfrak{a}^*} = \lim_{K\to\infty}\tau^{\mathfrak{a}^*}_{n,K}$ a.s. exists and $\tau_n^{\mathfrak{a}^*}\le t_0$ is an optimal stopping rule in the set $\mathcal{T}\cap\{\tau\ge T_n^{\{1\}}\}$,*
(ii) $\mathbf{E}\left[J(\tau_n^{\mathfrak{a}^*})|\mathcal{F}_n\right] = \gamma(M_n,T_n^{\{1\}})$ *a.s.*

Proof. The proof follows the same method as in Theorem 17.2. The difference lies in the form of the infinitesimal operator. Define the processes $\xi(s) = (s,M_s,V(s))$, where $V(s) = s - T^{\{1\}}_{N_1(s)}$. As was the case previously, $\xi(s)$ is the Markov process with the state space $[0,\infty)\times[0,\infty)\times[0,\infty)$. Notice that $J(s) = p(\xi(s))$ and $p(s,m,v)$: $[0,t_0]\times[0,\infty)\times[0,\infty)\to\mathbb{R}$ are continuous, bounded, and measurable with the bounded left-hand derivatives with respect to s and v. It is easily seen that $Ap(s,m,v) = \alpha_1(v)[\mathbf{E}\hat{g}^{\mathfrak{a}}(m+x^{\{1\}}) - \hat{g}^{\mathfrak{a}}(m)] - \left[\bar{y}_{-}^{\mathfrak{b}\prime}(t_0 - s) + c^{\mathfrak{a}\prime}(s)\right]$ for $s \leq t_0$. The rest of the proof is the same as in the proof of Theorem 17.2. □

Summarizing, the solution of a double stopping problem is given by

$$\mathbf{E}Z(\tau^{\mathfrak{a}^*}, \tau^{\mathfrak{b}^*}) = \mathbf{E}J(\tau^{\mathfrak{a}^*}) = \gamma(M_0, T_0^{\{1\}}) = \gamma(0,0),$$

where $\tau^{\mathfrak{a}^*}$ and $\tau^{\mathfrak{b}^*}$ are defined according to Theorems 17.2 and 17.4, respectively.

17.5 Examples

The form of the solution results in the fact that it is difficult to calculate the solution in an analytic way. In this section, we will present examples of the conditions for which the solution can be calculated explicitly.

Remark 17.7. If the process $\zeta_2(t) = Ap^{s,m}(\xi^s(t))$ has decreasing paths, then the second optimal stopping time is given by $\tau_n^{\mathfrak{b}^*} = \inf\{t \in \left[T_n^{\{3\}}, t_0\right] : Ap^{s,m}(\xi^s(t)) \leq 0\}$; on the other hand, if $\zeta_2(t)$ has nondecreasing paths, then the second optimal stopping time is equal to t_0.

Similarly, if the process $\zeta_1(s) = Ap(\xi(s))$ has decreasing paths, then the first optimal stopping time is given by $\tau_n^{\mathfrak{a}^*} = \inf\{s \in \left[T_n^{\{1\}}, t_0\right] : Ap(\xi(s)) \leq 0\}$; on the other hand, if $\zeta_1(s)$ has nondecreasing paths, then the first optimal stopping time is equal to t_0.

Proof. From (17.25) we obtain $\mathbf{E}[Z(s,\tau_n^{\mathfrak{b}^*})|\mathcal{F}_n^s] = Z(s,T_n^{\{3\}}) + \mathbf{E}[\int_{T_n^{\{3\}}}^{\tau_n^{\mathfrak{b}^*}} (Ap^{s,m})(\xi^s(z))\,\mathrm{d}z]$ a.s., and the application results of Jensen and Hsu [11] complete the proof. □

Corollary 17.2. *If $S^{\{3\}}$ has an exponential distribution with constant hazard rate α, the function $\hat{g}^{\mathfrak{b}}$ is increasing and concave, the cost function $c^{\mathfrak{b}}$ is convex, and $t_{2,n} = T_n^{\{3\}}$, $m_n^s = M_n^s$, then*

$$\tau_n^{\mathfrak{b}^*} = \inf\{t \in [t_{2,n}, t_0] : \alpha[\mathbf{E}\hat{g}^{\mathfrak{b}}(m_n^s + x^{\{3\}} - m) - \hat{g}^{\mathfrak{b}}(m_n^s - m)] \leq c^{\mathfrak{b}\prime}(t-s)\}, \quad (17.31)$$

where s is the moment when the location is changed. Moreover, if $S^{\{1\}}$ has an exponential distribution with constant hazard rate α_1, $\hat{g}^{\mathfrak{a}}$ is increasing and concave, $c^{\mathfrak{a}}$ is convex, and $t_{1,n} = T_n^{\{1\}}$, $m_n = \widehat{M}_n^{\{1\}}$, then

$$\tau_n^{\mathfrak{a}^*} = \inf\{s \in [t_{1,n}, t_0] : \alpha_1 \left[\mathbf{E}\hat{g}^{\mathfrak{a}}(m_n + x^{\{1\}}) - \hat{g}^{\mathfrak{a}}(m_n)\right] \leq c^{\mathfrak{a}\prime}(s)\}.$$

Proof. The form of $\tau^{\mathfrak{a}*}_n$ and $\tau^{\mathfrak{b}*}_n$ is a consequence of Remark 17.7. Let us observe that by our assumptions $\zeta_2(t) = \alpha\Delta^{\mathfrak{b}}(M^s_t - m) - c^{\mathfrak{b}\prime}(t-s)$ has decreasing paths for $t \in [T^{\{3\}}_n, T^{\{3\}}_{n+1})$. It suffices to prove that $\zeta_2(T^{\{3\}}_n) - \zeta_2(T^3_{n-1}) = \alpha[\Delta^{\mathfrak{b}}(M^s_n - m) - \Delta^{\mathfrak{b}}(M^s_{n-1} - m)] < 0$ for all $n \in \mathbb{N}$.

It remains to check that $\bar{y}^{\mathfrak{b}\prime}_{-}(t_0 - s) = 0$. We can see that $\tau^{\mathfrak{b}*} = \tau^{\mathfrak{b}*}(s)$ is deterministic, which is clear from (17.31). If $s \le t_0$, then combining (17.25), (17.26), and (17.28) gives $\gamma^{s,m}(m,s) = \mathbf{E}\big[Z(s,\tau^{\mathfrak{b}^*})|\mathcal{F}_s\big] = Z(s,s) + \mathbf{E}\big[\int_s^{\tau^{\mathfrak{b}^*}}(Ap^{s,m})(\xi^s(z))\, \mathrm{d}z|\mathcal{F}_s\big]$. By Remark 17.5, it follows that

$$\bar{y}^{\mathfrak{b}}(t_0 - s) = \mathbf{E}\left[\int_s^{\tau^{\mathfrak{b}*}(s)} (Ap^{s,m})(\xi^s(z))\,\mathrm{d}z\right] = \int_s^{\tau^{\mathfrak{b}*}(s)} \left[\alpha\Delta^{\mathfrak{b}}(0) - c_2'(z-s)\right] dz,$$

and this yields

$$\begin{aligned}\bar{y}^{\mathfrak{b}\prime}_{-}(t_0 - s) &= \int_s^{\tau^{\mathfrak{b}*}(s)} c_2''(z-s)\,\mathrm{d}z + \tau^{\mathfrak{b}*\prime}(s)\left[\alpha\Delta^{\mathfrak{b}}(0) - c_2'(\tau^{\mathfrak{b}}{}^*_2(s) - s)\right] \\ &\quad - \left[\alpha\Delta^{\mathfrak{b}}(0) - c_2'(0)\right] \\ &= c_2'(\tau^{\mathfrak{b}*}(s) - s) - c_2'(0) - \left[\alpha\Delta^{\mathfrak{b}}(0) - c_2'(0)\right] = 0. \qquad (17.32)\end{aligned}$$

The rest of proof runs as previously. □

Corollary 17.3. *If for $i = 1$ and $i = 2$ the functions $g_i^{\bullet}$ are increasing and convex, c_i are concave, and $S^{\{i\}}$ have an exponential distribution with constant hazard rate α_i (i.e., $\alpha = \alpha_2$), then $\tau^{\mathfrak{a}^*}_n = \tau^{\mathfrak{b}^*}_n = t_0$ for $n \in \mathbb{N}$.*

Proof. This is also the straightforward consequence of Remark 17.7. It suffices to check that $\bar{y}^{\mathfrak{b}\prime}_{-}(t_0 - s)$ is nonincreasing with respect to s. First observe that $\tau^{\mathfrak{b}^*}(s) = t_0$. Considering (17.32), it is obvious that $\bar{y}^{\mathfrak{b}\prime}_{-}(t_0 - s) = \alpha_2\Delta^{\mathfrak{b}}(0) - c_2'(t_0 - s)$, and this completes the proof. □

17.6 Conclusions

This article presents a solution of the double stopping problem in the "fishing model" for a finite horizon. The analytical properties of the reward function in one stopping problem played a crucial rule in our considerations and allowed us to obtain a solution to the extended double stopping problem. Repeating considerations from Sect. 17.4, we can easily generalize our model and the solution to the multiple stopping problem, but the notation can be inconvenient. The construction of the equilibrium in the two-person non-zero-sum problem formulated in Sect. 17.2 can be reduced to the two double optimal stopping problems in the case where the payoff structure is given by (17.5), (17.6), and (17.11). The key assumptions related to the properties of the distribution functions. Assuming general distributions and an infinite horizon, one can obtain the extensions of the foregoing model.

References

1. Boshuizen, F., Gouweleeuw, J.: General optimal stopping theorems for semi-Markov processes. Adv. Appl. Probab. **4**, 825–846 (1993)
2. Boshuizen, F.A.: A general framework for optimal stopping problems associated with multivariate point processes, and applications. Sequential Anal. **13**(4), 351–365 (1994)
3. Brémaud, P.: Point Processes and Queues. Martingale Dynamics. Springer, Berlin (1981)
4. Davis, M.H.A.: Markov Models and Optimization. Chapman & Hall, New York (1993)
5. Ferenstein, E., Pasternak-Winiarski, A.: Optimal stopping of a risk process with disruption and interest rates. In: Brèton, M., Szajowski, K. (eds.) Advances in Dynamic Games: Differential and Stochastic Games: Theory, Application and Numerical Methods, Annals of the International Society of Dynamic Games, vol. 11, 18 pp. Birkhäuser, Boston (2010)
6. Ferenstein, E., Sierociński, A.: Optimal stopping of a risk process. Appl. Math. **24(3)**, 335–342 (1997)
7. Ferguson, T.: A Poisson fishing model. In: Pollard, D., Torgersen, E., Yang, G. (eds.) Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics. Springer, Berlin (1997)
8. Haggstrom, G.: Optimal sequential procedures when more then one stop is required. Ann. Math. Stat. **38**, 1618–1626 (1967)
9. Jacobsen, M.: Point process theory and applications. Marked point and piecewise deterministic processes. In: Prob. and its Applications, vol. 7. Birkhäuser, Boston (2006)
10. Jensen, U.: An optimal stopping problem in risk theory. Scand. Actuarial J. **2**, 149–159 (1997)
11. Jensen, U., Hsu, G.: Optimal stopping by means of point process observations with applications in reliability. Math. Oper. Res. **18(3)**, 645–657 (1993)
12. Karpowicz, A.: Double optimal stopping in the fishing problem. J. Appl. Prob. **46**(2), 415–428 (2009). DOI 10.1239/jap/1245676097
13. Karpowicz, A., Szajowski, K.: Double optimal stopping of a risk process. GSSR Stochast. Int. J. Prob. Stoch. Process. **79**, 155–167 (2007)
14. Kramer, M., Starr, N.: Optimal stopping in a size dependent search. Sequential Anal. **9**, 59–80 (1990)
15. Muciek, B.K., Szajowski, K.: Optimal stopping of a risk process when claims are covered immediately. In: Mathematical Economics, Toru Maruyama (ed.) vol. 1557, pp. 132–139. Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606-8502 Japan Kôkyûroku (2007)
16. Nikolaev, M.: Obobshchennye posledovatelnye procedury. Litovski**ui** Mat. Sb. **19**, 35–44 (1979)
17. Rolski, T., Schmidli, H., Schmidt, V., Teugels, J.: Stochastic Processes for Insurance and Finance. Wiley, Chichester (1998)
18. Shiryaev, A.: Optimal Stopping Rules. Springer, Berlin (1978)
19. Starr, N.: Optimal and adaptive stopping based on capture times. J. Appl. Prob. **11**, 294–301 (1974)
20. Starr, N., Wardrop, R., Woodroofe, M.: Estimating a mean from delayed observations. Z. f ür Wahr. **35**, 103–113 (1976)
21. Starr, N., Woodroofe, M.: Gone fishin': optimal stopping based on catch times. Report No. 33, Department of Statistics, University of Michigan, Ann Arbor, MI (1974)
22. Szajowski, K.: Optimal stopping of a 2-vector risk process. In: Stability in Probability, Jolanta K. Misiewicz (ed.), Banach Center Publ. **90**, 179–191. Institute of Mathematics, Polish Academy of Science, Warsaw (2010), doi:10.4064/bc90-0-12

Chapter 18
A Nonzero-Sum Search Game with Two Competitive Searchers and a Target

Ryusuke Hohzaki

Abstract In this paper, we deal with a nonzero-sum three-person noncooperative search game, where two searchers compete for detection of a target and the target tries to evade the searchers. We verify that there occurs cooperation between two searchers against the target in the game with a stationary target and for a special case of the game with a moving target. Using these characteristics, we can partially regard the three-person nonzero-sum game as an equivalent two-person zero-sum game with the detection probability of target as a payoff. For a general game with a moving target, however, there could be many Nash equilibria. We propose a numerical algorithm for a Nash equilibrium in the general case. The discussion on the nonzero-sum search game in this paper could help us to step forward to a cooperative search game, where a coalition of some searchers and the rest of searchers compete against each other for detection of target as the future work.

Keywords Search theory • Game theory • Nonzero-sum noncooperative game • Search allocation game

18.1 Introduction

Search theory starts from military affairs. As an application of game theory to search problem, Morse and Kimball [20] discuss a position planning of the patrol line in the straits by an anti-submarine warfare (ASW) airplane to block the passages of submarines. For several decades after the research, they focus on one-sided problems for an optimal search under the assumption that the stochastic rule on the behavior of the submarine is known [18].

R. Hohzaki (✉)
Department of Computer Science, National Defense Academy, 1-10-20 Hashirimizu, Yokosuka 239-8686, Japan
e-mail: hozaki@cc.nda.ac.jp

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_18,

Since then, many researchers take one of the game models, named "hide-and-search game", where a stationary target hides at a position. Norris [24] deals with a two-person zero-sum (TPZS) noncooperative game, where a target hides in a box at first and then a searcher sequentially looks into boxes with possible overlooking and a payoff is the expected number of looks until the detection of the target. Baston and Bostock [2] and Garnaev [7] study an ASW game, where an ASW airplane drops some depth charges to destroy a hidden submarine. They adopt the destruction probability of the target as a payoff. We can quote Nakai [23], Kikuta [17] and Iida et al. [16] as the other studies on the hide-and-search game. Their models are still the TPZS noncooperative game but they adopt a variety of payoff functions: detection probability, expected reward, expected time until detection and others.

The hide-and-search game is extended to the game with a moving target, named "evade-and-search game". Meinardi [19] analyzes the evade-and-search game, where a target moves on a line in a diffusive fashion and a searcher looks at a point on the line sequentially as time elapses. The target is informed of the history of searched points and then the game is modeled as a multi-stage TPZS game. Washburn [29] and Nakai [22] adopt the multi-stage model with the payoff of the moving distance of the searcher until detection of target. Their models are similar to Meinardi's one. Danskin [5] discusses a one-shot game played by a submarine and an ASW airplane. The submarine chooses a point to move to and the airplane selects a point to drop his dipping sonar buoys for detection of the submarine. Eagle and Washburn [6] also study the one-shot game, where a searcher moves in a search space as well as in Danskin.

Hohzaki [11, 12], Iida et al. [15] and Dambreville and Le Cadre [4] deal with an optimal distribution of searching resources for a searcher and an optimal moving strategy for a target by a one-shot game called "search allocation game (SAG)". For the one-shot SAG, Washburn [30] and Hohzaki [8, 13, 14] take account of practical constraints such as geographical restriction or energy limitation on moving. Hohzaki [9] proposes a method to derive an optimal solution for a multi-stage SAG.

As we reviewed the previous research on the search game above, almost all researchers handle the TPZS noncooperative game of a target versus a searcher although we can itemize small number of special game models such as Baston and Garnaev [3], who study a nonzero-sum game with a searcher and a protector of taking the distribution strategies of resources. However, we can think of cooperative search situation, in which several searchers cooperate for an effective search for the target or a drifting person in a shipwreck would take a cooperative action to a search and rescue (S&R) team, such as firing a distress signal or a flashlight to be easily detected. Hohzaki [10] is one of few researches on the search defined by the cooperative game (refer to Owen [26] or Myerson [21]). Using the framework of the SAG, Hohzaki models the search game with some cooperative searchers against a moving target under the criterion of detection probability of target. The discussion includes the imputation of the obtained reward among cooperative searchers of a team or a coalition, which is a common theme for an ordinary coalition game. There are other types of cooperative search problems. Alpern and Gal [1] have been studying the so-called rendezvous search problem, where players try to meet each other as soon as possible without knowing the exact position of another player. In the context of graph theoretic problem, we can enumerate further works. Parsons [27, 28] studies how many searchers are required to find a mobile hider in a graph. O' Rourke [25] theoretically or algorithmically derives the minimum number of guards to watch over the interior of a polygon-shaped art gallery by computational geometry. The problem on security by watchmen or guards are named "art gallery problem".

In the Hohzaki's model [10], he proves that searchers have the incentive to construct a grand coalition and develop his theory based on the assumption that only the members of the coalition join the search operation for the target. However there could be a competition between the coalition's members and nonmembers. In the treasure hunting from shipwreck, the nonmembers would be going to outwit the coalition for the preemptive detection of the treasure. In this paper, we discuss a three-person nonzero-sum noncooperative search game, where two teams or two coalitions of searchers compete for the detection of a target and the target tries to evade the teams, and derive a Nash-equilibrium (NE). The results of this paper would help us step forward to an other type of cooperative game or coalition game, where several groups of searchers compete each other, other than the Hohzaki's model and discuss the incentive of the groups to a larger group or a grand coalition beyond their competition.

As a preliminary, we consider a search game for a stationary target by a three-person nonzero-sum noncooperative game model in the next section. In Sect. 18.3, we discuss a game with a moving target. Because it would be difficult to derive a NE, we try to do it for a small size of problem at first and propose a computational algorithm for the NE of the general game with a moving target. In Sect. 18.4, we analyze the characteristics of the NE by some numerical examples.

18.2 A Nonzero-Sum Search Game with a Stationary Target

We consider the search game where two searchers compete against each other to get the value of a target while the target evades from them. The problem is formulated as a three-person nonzero-sum noncooperative game.

(A1) A search space is discrete and it consists of n cells denoted by $\mathbf{K} = \{1, 2, \cdots, n\}$. A target with value 1 chooses one cell to hide himself.

(A2) Searcher 1 has the amount Φ_1 of searching resources in hand and distributes them in the search space to detect the target while Searcher 2 has the amount Φ_2 of resources.

(A3) If the target is in cell i and the searcher scatters x resources there, the searcher can detect the target with probability $f_i(x) = 1 - \exp(-\alpha_i x)$, where parameter α_i indicates the effectiveness of unit resource for detection. The event of detection by one searcher is independent of the detection by the other.

(A4) If only one searcher detects the target, the detector monopolizes the value of the target 1. If both of them do, Searcher 1 gets a reward δ_1 and Searcher 2 δ_2, where $\delta_1 + \delta_2$ does not necessarily equal 1. The target is given 1 only if he is not detected.

We denote a mixed strategy of the target by $\mathbf{p} = \{p_1, p_2, \cdots, p_n\}$, where p_i is the probability of hiding in cell i. Let us denote a pure strategy of Searcher 1 or 2 by $\mathbf{x} = \{x_i,\ i \in \mathbf{K}\}$ or $\mathbf{y} = \{y_i,\ i \in \mathbf{K}\}$, respectively. x_i or y_i is the respective amount of resource distributed in cell i by Searcher 1 or 2. We denote feasible regions of players' strategies $\mathbf{p}$, $\mathbf{x}$ and $\mathbf{y}$ by Π, Ψ_1 and Ψ_2, which are given by

$$\Pi \equiv \left\{ \mathbf{p} \in \mathbf{R}^n | p_i \geq 0, i \in \mathbf{K},\ \sum_{i \in \mathbf{K}} p_i = 1 \right\}$$

$$\Psi_1 \equiv \left\{ \mathbf{x} \in \mathbf{R}^n | x_i \geq 0, i \in \mathbf{K},\ \sum_{i \in \mathbf{K}} x_i \leq \Phi_1 \right\}$$

$$\Psi_2 \equiv \left\{ \mathbf{y} \in \mathbf{R}^n | y_i \geq 0, i \in \mathbf{K},\ \sum_{i \in \mathbf{K}} y_i \leq \Phi_2 \right\}.$$

The searchers had obviously better use up all resources because the detection function $f_i(x)$ is monotone increasing for x, as stated in (A3). Therefore, we can replace inequality signs with equality signs in the definitions of Ψ_1 and Ψ_2.

From the independency of detection events in Assumption (A3), three players of the target, Searcher 1 and 2 would have the expected rewards or payoffs $Q(\mathbf{p}, \mathbf{x}, \mathbf{y})$, $R_1(\mathbf{p}, \mathbf{y}, \mathbf{x})$ and $R_2(\mathbf{p}, \mathbf{y}, \mathbf{x})$, expressed by

$$Q(\mathbf{p}, \mathbf{x}, \mathbf{y}) = \sum_{i \in \mathbf{K}} p_i \exp(-\alpha_i (x_i + y_i)) \tag{18.1}$$

$$R_1(\mathbf{p}, \mathbf{x}, \mathbf{y}) = \sum_{i \in \mathbf{K}} p_i (1 - \exp(-\alpha_i x_i)) \{\exp(-\alpha_i y_i) + \delta_1 (1 - \exp(-\alpha_i y_i))\} \tag{18.2}$$

$$R_2(\mathbf{p}, \mathbf{y}, \mathbf{x}) = \sum_{i \in \mathbf{K}} p_i (1 - \exp(-\alpha_i y_i)) \{\exp(-\alpha_i x_i) + \delta_2 (1 - \exp(-\alpha_i x_i))\}. \tag{18.3}$$

The payoff $R_1(\cdot)$ is strictly concave for variable $\mathbf{x}$ and $R_2(\cdot)$ is also the same for $\mathbf{y}$. The feasible regions Ψ_1 and Ψ_2 are closed convex sets. Therefore, if there is a Nash equilibrium (NE), we can find it among pure strategies of the searchers. From here, we consider maximization problems for these expected payoffs and derive an optimal response of a player to others.

1. Optimal response of the target
   We can transform a maximization problem of the target's payoff $Q(\mathbf{p}, \mathbf{x}, \mathbf{y})$, which is the non-detection probability of the target, as follows.

$$\max_{\mathbf{p}\in\Pi} Q(\mathbf{p},\mathbf{x},\mathbf{y}) = \max_{\mathbf{p}\in\Pi} \sum_{i\in\mathbf{K}} p_i \exp(-\alpha_i(x_i+y_i))$$
$$= \max_{i\in\mathbf{K}} \exp(-\alpha_i(x_i+y_i)) = \exp(-\min_{i\in\mathbf{K}} \alpha_i(x_i+y_i)).$$

As seen by the transformation from the second expression to the third, an optimal target's strategy $\mathbf{p}^* \in \Pi$ is given by $p_i^* = 0$ for $i \notin I^*$ and arbitrary p_i^* for $i \in I^*$, using a set of cells $I^* \equiv \{i \in \mathbf{K} | \alpha_i(x_i+y_i) = \nu \equiv \min_{s\in\mathbf{K}} \alpha_s(x_s+y_s)\}$ or $I^* \equiv Arg\min_{s\in\mathbf{K}} \alpha_s(x_s+y_s)$.

2. Optimal response of Searcher $j = 1,2$

If i does not belong to I^*, namely, $\alpha_i(x_i+y_i) > \nu$, the analysis of Item 1 above tells us that it must be $p_i = 0$ and then both searchers $j = 1,2$ should not distribute any resource in the cell i, i.e. $x_i^* = y_i^* = 0$, from Eqs. (18.2) and (18.3). This contradicts the first assumption of $i \notin I^*$. Therefore, it should be $\alpha_i(x_i+y_i) = \nu(const)$ for every $i \in \mathbf{K}$ and we have

$$\nu = \frac{1}{\sum_{s\in\mathbf{K}} 1/\alpha_s}(\Phi_1+\Phi_2) \tag{18.4}$$

$$x_i + y_i = \frac{1/\alpha_i}{\sum_{s\in\mathbf{K}} 1/\alpha_s}(\Phi_1+\Phi_2),\ i \in \mathbf{K}. \tag{18.5}$$

To determine the optimal response of a searcher, we consider a basic problem with one searcher to maximize the detection probability or minimize the non-detection probability equivalently by the limited amount of searching resource, Φ, given that the target's strategy p_i. The problem is formulated as follows.

$$\min_{\mathbf{x}} \sum_{i\in\mathbf{K}} p_i \exp(-\alpha_i x_i),\ \ s.t.\ \ x_i \geq 0, i \in \mathbf{K},\ \sum_{i\in\mathbf{K}} x_i \leq \Phi. \tag{18.6}$$

By the help of the definition of a Lagrange function $L(\mathbf{x};\lambda,\mu) \equiv \sum_{i\in\mathbf{K}} p_i \exp(-\alpha_i x_i) + \lambda(\sum_i x_i - \Phi) - \sum_i \mu_i x_i$, we can obtain necessary and sufficient conditions for optimality as the Karush–Kuhn–Tucker conditions. The conditions for optimal $\mathbf{x}$ are unified into

$$x_i = \frac{1}{\alpha_i}\left[\log \frac{p_i\alpha_i}{\lambda}\right]^+, \tag{18.7}$$

where $[x]^+$ means $[x]^+ \equiv \max\{0,x\}$. An optimal Lagrangian multiplier λ is uniquely determined from $\sum_i x_i = \Phi$ and Eq. (18.7).

Noting that we can generate Eq. (18.2) by replacing p_i in the objective function (18.6) with $p_i D_i^1(y)$, where

$$D_i^1(y) \equiv \exp(-\alpha_i y_i) + \delta_1(1-\exp(-\alpha_i y_i)) = (1-\delta_1)\exp(-\alpha_i y_i) + \delta_1,$$

we apply Eq. (18.7) to the original problem with two searchers to derive an optimal response **x** for Searcher 1 as

$$x_i = \frac{1}{\alpha_i}\left[\log \frac{p_i \alpha_i D_i^1(y)}{\lambda_1}\right]^+ = \frac{1}{\alpha_i}\left[\log \frac{p_i \alpha_i}{\lambda_1} + \log\{(1-\delta_1)\exp(-\alpha_i y_i) + \delta_1\}\right]^+, \tag{18.8}$$

given other strategies **p** and **y**. Similarly, we have an optimal response **y** for Searcher 2 given strategies **p** and **x**, as follows:

$$y_i = \frac{1}{\alpha_i}\left[\log \frac{p_i \alpha_i}{\lambda_2} + \log\{(1-\delta_2)\exp(-\alpha_i x_i) + \delta_2\}\right]^+. \tag{18.9}$$

Optimal Lagrangian multipliers λ_1 and λ_2 in Eqs. (18.8) and (18.9) are determined by conditions $\sum_i x_i = \Phi_1$ and $\sum_i y_i = \Phi_2$, respectively.

As a result, we organize the necessary and sufficient conditions for optimal **x**, **y** and **p** in the following system of equations.

$$x_i = \frac{1}{\alpha_i}\left[\log \frac{p_i \alpha_i}{\lambda_1} + \log\{(1-\delta_1)\exp(-\alpha_i y_i) + \delta_1\}\right]^+, \ i \in \mathbf{K} \tag{18.10}$$

$$y_i = \frac{1}{\alpha_i}\left[\log \frac{p_i \alpha_i}{\lambda_2} + \log\{(1-\delta_2)\exp(-\alpha_i x_i) + \delta_2\}\right]^+, \ i \in \mathbf{K} \tag{18.11}$$

$$x_i + y_i = \frac{1/\alpha_i}{\sum_{j \in \mathbf{K}} 1/\alpha_j}(\Phi_1 + \Phi_2), \ i \in \mathbf{K} \tag{18.12}$$

$$\sum_{i \in \mathbf{K}} x_i = \Phi_1 \ \ or \ \ \sum_{i \in \mathbf{K}} y_i = \Phi_2 \tag{18.13}$$

$$\sum_{i \in \mathbf{K}} p_i = 1. \tag{18.14}$$

We need only one of equations (18.13) for a full system because we can derive the other equation of (18.13) from Eq. (18.12). The total number of variables **x**, **y**, **p**, λ_1 and λ_2 is $3|\mathbf{K}| + 2$, which is the same as the number of equations contained in the system. If all equations of the system are independent, optimal variables are uniquely determined.

We can show that the following solution satisfies the conditions above.

$$x_i^* = \frac{1/\alpha_i}{\sum_{j \in \mathbf{K}} 1/\alpha_j}\Phi_1 \tag{18.15}$$

$$y_i^* = \frac{1/\alpha_i}{\sum_{j \in \mathbf{K}} 1/\alpha_j}\Phi_2 \tag{18.16}$$

$$p_i^* = \frac{1/\alpha_i}{\sum_{j \in \mathbf{K}} 1/\alpha_j}. \tag{18.17}$$

We can easily see that the above strategies satisfy the conditions (18.12)∼(18.14). Noting that $\alpha_i x_i^*$, $\alpha_i y_i^*$, and $\alpha_i p_i^*$ do not depend on the cell number i, we can derive Lagrangian multipliers λ_1 and λ_2 by substituting Eqs. (18.15)∼(18.17) in (18.10) and (18.11).

$$\lambda_1^* = \exp\left(-\frac{\Phi_1}{\sum_j 1/\alpha_j}\right)\left\{(1-\delta_1)\exp\left(-\frac{\Phi_2}{\sum_j 1/\alpha_j}\right)+\delta_1\right\} \Big/ \sum_j \frac{1}{\alpha_j} \tag{18.18}$$

$$\lambda_2^* = \exp\left(-\frac{\Phi_2}{\sum_j 1/\alpha_j}\right)\left\{(1-\delta_2)\exp\left(-\frac{\Phi_1}{\sum_j 1/\alpha_j}\right)+\delta_2\right\} \Big/ \sum_j \frac{1}{\alpha_j} \tag{18.19}$$

Let us note that the optimal strategies (18.15), (18.16) and (18.17) are also optimal for the TPZS game, where a searcher with $\Phi_1 + \Phi_2$ resources and a target fight against each other for the payoff of non-detection probability. We can easily verify the optimality of $x_i^* + y_i^*$ and p_i^* for the TPZS game by solving the following minimax or maximin optimization where a searcher's strategy is $z_i = x_i + y_i$ generated by the original strategies of two searchers, x_i and y_i.

$$\min_{\{z_i\}} \max_{\{p_i\}} \sum_{i\in\mathbf{K}} p_i \exp(-\alpha_i z_i) = \max_{\{p_i\}} \min_{\{z_i\}} \sum_{i\in\mathbf{K}} p_i \exp(-\alpha_i z_i).$$

If $\delta_1 + \delta_2 = 1$, there is the relation of $R_1(\mathbf{p},\mathbf{x},\mathbf{y}) + R_2(\mathbf{p},\mathbf{y},\mathbf{x}) = 1 - Q(\mathbf{p},\mathbf{x},\mathbf{y})$ between $R_1(\mathbf{p},\mathbf{x},\mathbf{y})$ and $R_2(\mathbf{p},\mathbf{y},\mathbf{x})$. Therefore, the minimization of the non-detection probability $Q(\mathbf{p},\mathbf{x},\mathbf{y})$ has Pareto optimality for two searchers in the case of $\delta_1 + \delta_2 = 1$ but does not necessarily have it in other cases. Nevertheless, the searchers' strategies of the NE have a direction to the minimization of the non-detection probability. The property is caused by the target strategy aiming the maximization of the non-detection probability, as explained just before deriving Eq. (18.5). The target strategy forces both searchers to have $x_i + y_i = (\Phi_1 + \Phi_2) / \alpha_i / \sum_s 1/\alpha_s$, which is an optimal representative response of both searchers against the target in the TPZS game.

18.3 A Search Game with a Moving Target

Here we consider the nonzero-sum game with two searchers and a moving target. Two searchers play in a noncooperative manner for the detection of target. The target moves in a search space to avoid the detection. The game with a moving target is modeled as follows:

(B1) A search space consists of a discrete cell space $\mathbf{K} = \{1, \cdots, K\}$ and a discrete time space $\mathbf{T} = \{1, \cdots, T\}$.

(B2) A target chooses one among a set of routes Ω to move along. His position on a route $\omega \in \Omega$ at time $t \in \mathbf{T}$ is represented by $\omega(t) \in \mathbf{K}$.

(B3) Two searchers distribute their searching resources to detect the target. $\Phi_k(t)$ resources are available at each time t for Searcher $k = 1,2$. Searchers can start distributing resource from time $\tau \in \mathbf{T}$.

(B4) The detection of the target by the distribution of x resources in cell i occurs with probability $1 - \exp(-\alpha_i x)$ only if the target is in the cell i. The parameter α_i indicates the effectiveness of unit resource in the cell i. The events of detection by two searchers are independent of each other.

If a searcher detects the target, the detector monopolizes the value of the target 1. If both searchers detect, Searcher 1 and 2 get reward δ_1 and δ_2 ($0 \leq \delta_1,\ \delta_2 \leq 1$), respectively. The game is terminated on detection of the target or at the last time T. The target is given 1 only if he is not detected.

(B5) Players do not know any information about the behavior of other players and the situation of the search in the process of the game. Therefore, all players make their plans or strategies in advance of the game.

Let $\widehat{\mathbf{T}} = \{\tau, \tau+1, \cdots, T\}$ be an available time period for searching. We denote a distribution plan of Searcher $k = 1,2$ by $\varphi_k = \{\varphi_k(i,t),\ i \in \mathbf{K},\ t \in \widehat{\mathbf{T}}\}$, where $\varphi_k(i,t) \in \mathbf{R}$ is the amount of searching resources distributed in cell i at time t, and a mixed strategy of the target by $\pi = \{\pi(\omega),\ \omega \in \Omega\}$, where $\pi(\omega)$ is the probability of taking path $\omega \in \Omega$.

When the target takes a path ω and the searchers adopt their strategies φ_1 and φ_2, non-detection probability $Q(\omega, \varphi_1, \varphi_2)$ is given by

$$Q(\omega, \varphi_1, \varphi_2) = \exp\left(-\sum_{t=\tau}^{T} \alpha_{\omega(t)}(\varphi_1(\omega(t),t) + \varphi_2(\omega(t),t))\right). \tag{18.20}$$

The detection at time t is conditioned that no detection occurs before t, the probability of which is

$$\exp\left(-\sum_{\zeta=\tau}^{t-1} \alpha_{\omega(\zeta)}(\varphi_1(\omega(\zeta),\zeta) + \varphi_2(\omega(\zeta),\zeta))\right).$$

Taking account of the detection probability by Searcher k at time t, $1 - \exp(-\alpha_{\omega(t)} \varphi_k(\omega(t),t))$, and the detection event by the other searcher, the reward that the searcher k expects at time t is

$$\exp\left(-\sum_{\zeta=\tau}^{t-1} \alpha_{\omega(\zeta)}(\varphi_1(\omega(\zeta),\zeta) + \varphi_2(\omega(\zeta),\zeta))\right)(1 - \exp(-\alpha_{\omega(t)}\varphi_k(\omega(t),t)))$$
$$\times \left\{\exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t)) + \delta_k(1 - \exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t)))\right\}.$$

Because the detection event is exclusive at each time, the total expected reward of Searcher k, $R_k(\omega, \varphi_k, \varphi_j)$, $(k,j) \in \{(1,2),(2,1)\}$, is given

$$R_k(\omega,\varphi_k,\varphi_j) = \sum_{t=\tau}^{T} \exp\left(-\sum_{\zeta=\tau}^{t-1} \alpha_{\omega(\zeta)}(\varphi_k(\omega(\zeta),\zeta)+\varphi_j(\omega(\zeta),\zeta))\right)$$
$$\times (1-\exp(-\alpha_{\omega(t)}\varphi_k(\omega(t),t))) \times \{\exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t))$$
$$+ \delta_k(1-\exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t)))\} \quad (18.21)$$

for the target path ω. As a result, we have the payoffs of the target and Searcher k, $Q(\pi,\varphi_1,\varphi_2)$ and $R_k(\pi,\varphi_k,\varphi_j)$, as follows:

$$Q(\pi,\varphi_1,\varphi_2) = \sum_{\omega\in\Omega} \pi(\omega)Q(\omega,\varphi_1,\varphi_2) \quad (18.22)$$

$$R_k(\pi,\varphi_k,\varphi_j) = \sum_{\omega\in\Omega} \pi(\omega)R_k(\omega,\varphi_k,\varphi_j),\ (k,j)=(1,2),(2,1). \quad (18.23)$$

Now that we have formulated the three-person nonzero-sum game with a target and two searchers, the next thing to do is to obtain a NE which maximizes $Q(\pi,\varphi_1,\varphi_2)$, $R_1(\pi,\varphi_1,\varphi_2)$ and $R_2(\pi,\varphi_2,\varphi_1)$ with respect to π, φ_1 and φ_2, respectively. The optimality conditions of the NE, $(\pi^*,\varphi_1^*,\varphi_2^*)$, are

$$Q(\pi^*,\varphi_1^*,\varphi_2^*) \geq Q(\pi,\varphi_1^*,\varphi_2^*),\ R_1(\pi^*,\varphi_1^*,\varphi_2^*) \geq R_1(\pi^*,\varphi_1,\varphi_2^*),$$
$$R_2(\pi^*,\varphi_2^*,\varphi_1^*) \geq R_2(\pi^*,\varphi_2^*,\varphi_1) \quad (18.24)$$

for arbitrary $\pi\in\Pi$, $\varphi_1\in\Psi_1$ and $\varphi_2\in\Psi_2$, where Π and Ψ_k are the feasible regions of players' strategies π and φ_k $(k=1,2)$, given by

$$\Pi \equiv \left\{\pi|\pi(\omega)\geq 0,\omega\in\Omega,\ \sum_{\omega\in\Omega}\pi(\omega)=1\right\} \quad (18.25)$$

$$\Psi_k \equiv \left\{\varphi_k|\varphi_k(i,t)\geq 0, i\in\mathbf{K}, t\in\widehat{\mathbf{T}},\ \sum_{i\in\mathbf{K}}\varphi_k(i,t)=\Phi_k(t), t\in\widehat{\mathbf{T}}\right\}. \quad (18.26)$$

All are closed convex sets. We can see that the payoff function, $Q(\pi,\varphi_1,\varphi_2)$ is linear for π and convex for φ_1 and φ_2. We are going to prove the strictly concavity of $R_k(\omega,\varphi_k,\varphi_j)$ for φ_k.

Using a notation

$$\beta_j(\omega,t) \equiv \exp\left(-\sum_{\zeta=\tau}^{t-1}\alpha_{\omega(\zeta)}\varphi_j(\omega(\zeta),\zeta)\right) \times \{\exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t))$$
$$+\delta_k\,(1-\exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t)))\},$$

we can transform $R_k(\omega,\varphi_k,\varphi_j)$ like

$$
\begin{aligned}
&R_k(\omega,\varphi_k,\varphi_j)\\
&=\sum_{t=\tau}^{T}\beta_j(\omega,t)\exp\left(-\sum_{\zeta=\tau}^{t-1}\alpha_{\omega(\zeta)}\varphi_k(\omega(\zeta),\zeta)\right)(1-\exp(-\alpha_{\omega(t)}\varphi_k(\omega(t),t)))\\
&=\sum_{t=\tau}^{T}\beta_j(\omega,t)\left\{\exp\left(-\sum_{\zeta=\tau}^{t-1}\alpha_{\omega(\zeta)}\varphi_k(\omega(\zeta),\zeta)\right)-\exp\left(-\sum_{\zeta=\tau}^{t}\alpha_{\omega(\zeta)}\varphi_k(\omega(\zeta),\zeta)\right)\right\}\\
&=\beta_j(\omega,\tau)-\sum_{t=\tau}^{T-1}(\beta_j(\omega,t)-\beta_j(\omega,t+1))\exp\left(-\sum_{\zeta=\tau}^{t}\alpha_{\omega(\zeta)}\varphi_k(\omega(\zeta),\zeta)\right)\\
&\quad-\beta_j(\omega,T)\exp\left(-\sum_{\zeta=\tau}^{T}\alpha_{\omega(\zeta)}\varphi_k(\omega(\zeta),\zeta)\right).
\end{aligned}
\tag{18.27}
$$

Noting

$$
\begin{aligned}
\beta_j(\omega,t+1)&=\exp\left(-\sum_{\zeta=\tau}^{t}\alpha_{\omega(\zeta)}\varphi_j(\omega(\zeta),\zeta))\right)\\
&\qquad\times\left\{(1-\delta_k)\exp(-\alpha_{\omega(t+1)}\varphi_j(\omega(t+1),t+1))+\delta_k\right\}\\
&\le\exp\left(-\sum_{\zeta=\tau}^{t-1}\alpha_{\omega(\zeta)}\varphi_j(\omega(\zeta),\zeta))\right)\\
&\qquad\times\left\{(1-\delta_k)\exp(-\alpha_{\omega(t)}\varphi_j(\omega(t),t))+\delta_k\right\}\\
&=\beta_j(\omega,t),
\end{aligned}
$$

the last expression of Eq. (18.27) is proved to be strictly concave.

From the strictly concavity of $R_k(\cdot)$ for φ_k and the closed convexity of the region Ψ_k, an optimal response of the searcher k is uniquely determined given other players' strategies π and φ_j $(j\neq k)$. As a lemma, we state the optimality conditions of $\pi\in\Pi$ of maximizing the non-detection probability, which is easily derived from Eqs. (18.20), (18.22) and (18.25).

Lemma 18.1. *An optimal selection probability of target path is* $\pi(\omega)\geq 0$ *for* $\omega\in\Omega_M(\varphi_1+\varphi_2)$ *and* $\pi(\omega)=0$ *for* $\omega\notin\Omega_M(\varphi_1+\varphi_2)$*, where* $\Omega_M(\varphi_1+\varphi_2)$ *is defined by*

$$\Omega_M(\varphi_1+\varphi_2)\equiv\{\omega_m\in\Omega\,|\,g(\omega_m,\varphi_1+\varphi_2)=\min_{\omega\in\Omega}\ g(\omega,\varphi_1+\varphi_2)\}$$

and $g(\omega,\varphi_1+\varphi_2)$ *is the weighted amount of resources cumulated on the path* $\omega\in\Omega$ *defined by*

$$g(\omega,\varphi_1+\varphi_2)\equiv\sum_{t\in\widehat{\mathbf{T}}}\alpha_{\omega(t)}(\varphi_1(\omega(t),t)+\varphi_2(\omega(t),t)).\tag{18.28}$$

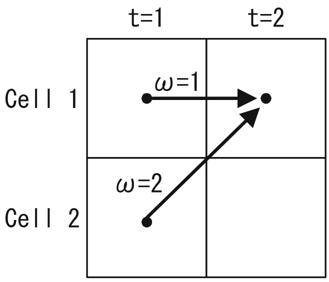


Fig. 18.1 A search space and target paths

18.3.1 A Special Case of $(K,T)=(2,2)$

We have so far discussed a general theory about the three-person nonzero-sum search game. Here we apply it to a small case of two cells $(K=2)$ and two time points $(T=2)$ to find a concrete form of NE. The case is illustrated in Fig. 18.1. The path route $\{\omega(t), t=1,2\}$ is set to be $\{1,1\}$ and $\{2,1\}$ for target path #1 and #2, respectively, and both paths rendezvous in cell 1 at time 2. That is why both searchers distribute all their resources $\Phi_1(2)$ and $\Phi_2(2)$ into cell 1 at time 2 and an optimal distribution at time 2 is determined to be $\varphi_1^*(1,2)=\Phi_1(2)$, $\varphi_1^*(2,2)=0$, $\varphi_2^*(1,2)=\Phi_2(2)$, $\varphi_2^*(2,2)=0$. Using two variables x and y, we denote the resource distribution at time 1 by $\varphi_1(1,1)=x$, $\varphi_1(2,1)=\Phi_1(1)-x$ for Searcher 1, and $\varphi_2(1,1)=y$, $\varphi_2(2,1)=\Phi_2(1)-y$ for Searcher 2. We are going to solve the maximization problem of the payoff function given by Eq. (18.23). The followings are the derivatives of $Q(\pi,x,y)$ and $R_k(\pi,x,y)$, which would be helpful for the coming analysis.

$$Q(\pi,x,y)=\pi(1)\exp\left(-\alpha_1(x+y+\Phi_1(2)+\Phi_2(2))\right)+\ \pi(2)\exp\left(-\alpha_2(\Phi_1(1)\right.$$
$$\left.+\Phi_2(1)-x-y)-\alpha_1(\Phi_1(2)+\Phi_2(2))\right) \tag{18.29}$$

$$\begin{aligned}\frac{\partial R_1}{\partial x}=&\ \{\pi(1)\alpha_1\exp(-\alpha_1(x+y))-\pi(2)\alpha_2\exp(-\alpha_2(\Phi_1(1)+\Phi_2(1)-x-y))\}\\&\times\{(1-\delta_1)-(1-\exp(-\alpha_1\Phi_1(2)))\\&\quad(\exp(-\alpha_1\Phi_2(2))+\delta_1(1-\exp(-\alpha_1\Phi_2(2))))\}\\&+\delta_1\{\pi(1)\alpha_1\exp(-\alpha_1x)-\pi(2)\alpha_2\exp(-\alpha_2(\Phi_1(1)-x))\}\end{aligned} \tag{18.30}$$

$$\begin{aligned}\frac{\partial R_2}{\partial y}=&\ \{\pi(1)\alpha_1\exp(-\alpha_1(x+y))-\pi(2)\alpha_2\exp(-\alpha_2(\Phi_1(1)+\Phi_2(1)-x-y))\}\\&\times\{(1-\delta_2)-(1-\exp(-\alpha_1\Phi_2(2)))\\&\quad(\exp(-\alpha_1\Phi_1(2))+\delta_2(1-\exp(-\alpha_1\Phi_1(2))))\}\\&+\delta_2\{\pi(1)\alpha_1\exp(-\alpha_1y)-\pi(2)\alpha_2\exp(-\alpha_2(\Phi_2(1)-y))\}.\end{aligned} \tag{18.31}$$

1. Optimal response of the target
   We can do the maximization of $\max_\pi Q(\pi,x,y)$ by comparing the values in the shoulder of the exponential function in Eq. (18.29) and obtain an optimal response of the target, using a notation

$$\Phi' \equiv \frac{\alpha_2}{\alpha_1+\alpha_2}(\Phi_1(1)+\Phi_2(1)),$$

   as follows:

$$\text{(i)}\quad \pi(1)=1 \text{ and } \pi(2)=0, \text{ if } x+y<\Phi'. \tag{18.32}$$

$$\text{(ii)}\quad \pi(1)=0 \text{ and } \pi(2)=1, \text{ if } x+y>\Phi'. \tag{18.33}$$

$$\text{(iii)}\quad \pi(1) \text{ and } \pi(2)\geq 0 \text{ satisfying } \pi(1)+\pi(2)=1,\ if\ x+y=\Phi'. \tag{18.34}$$

2. Optimal response of the searcher
   Please note that the value in parenthesis { } in the second line of Eq. (18.30) changes its sign of positiveness or negativeness depending on δ_1. It is positive if $\delta_1<\delta_1^*$ and negative if $\delta_1>\delta_1^*$, where

$$\delta_1^* = \frac{1-(1-\exp(-\alpha_1\Phi_1(2)))\exp(-\alpha_1\Phi_2(2))}{1+(1-\exp(-\alpha_1\Phi_1(2)))(1-\exp(-\alpha_1\Phi_2(2)))}.$$

   Similarly, the value in the parenthesis { } in the second line of Eq. (18.31) changes its sign with a threshold

$$\delta_2^* = \frac{1-(1-\exp(-\alpha_1\Phi_2(2)))\exp(-\alpha_1\Phi_1(2))}{1+(1-\exp(-\alpha_1\Phi_2(2)))(1-\exp(-\alpha_1\Phi_1(2)))}$$

   for δ_2.

We are going to prove that $x+y\neq\Phi'$ must not hold for any optimal x and y by classifying δ_1 and δ_2 into four cases. In the process of proof, we might refer to Eqs. (18.29)∼(18.34).

(i) Case of $\delta_1<\delta_1^*$ and $\delta_2<\delta_2^*$:
   If $x+y<\Phi'$, it must be $\pi(1)=1$ from Eq. (18.32) and then $R_1(\cdot)$ monotonically increases for x because $\partial R_1/\partial x>0$ from Eq. (18.30). $R_2(\cdot)$ is also monotone increasing for y. Therefore, x and y is never optimal within $x+y<\Phi'$. If $x+y>\Phi'$, it must be $\pi(1)=0$, from which $\partial R_1/\partial x<0$ and $\partial R_2/\partial y<0$ hold and then smaller x and y are much better for the searcher. The condition $x+y>\Phi'$ is never valid for any optimal x and y.

(ii) Case of $\delta_1<\delta_1^*$ and $\delta_2>\delta_2^*$:
   If $x+y<\Phi'$, we have $\pi(1)=1$ and $\partial R_1/\partial x>0$. This implies that larger x is desirable for Searcher 1. Concerning $\partial R_2/\partial y$ of Eq. (18.31), we have

$$\frac{\partial R_2(\pi,y,x)}{\partial y} \geq \left.\frac{\partial R_2(\pi,y,x)}{\partial y}\right|_{x=0} = \alpha_1 \exp(-\alpha_1 y)$$
$$\times\{1-(1-\exp(-\alpha_1\Phi_2(2)))(\exp(-\alpha_1\Phi_1(2))$$
$$+\delta_2(1-\exp(-\alpha_1\Phi_1(2))))\}$$
$$> 0 \tag{18.35}$$

and then larger y is desirable for Searcher 2. As a result, both searchers increase x and y until reaching $x+y=\Phi'$.

If $x+y>\Phi'$, we have $\pi(1)=0$, $\partial R_1/\partial x<0$ and

$$\frac{\partial R_2(\pi,y,x)}{\partial y} \leq \left.\frac{\partial R_2(\pi,y,x)}{\partial y}\right|_{x=\Phi_1(1)} = -\alpha_2 \exp(-\alpha_2(\Phi_2(1)-y))$$
$$\times\{1-(1-\exp(-\alpha_1\Phi_2(2)))(\exp(-\alpha_1\Phi_1(2))$$
$$+\delta_2(1-\exp(-\alpha_1\Phi_1(2))))\}$$
$$< 0. \tag{18.36}$$

Therefore, the searchers decrease x and y until $x+y=\Phi'$.

(iii) Case of $\delta_1>\delta_1^*$ and $\delta_2<\delta_2^*$:
By analogy to the case of (ii), we can say that $x+y\neq\Phi'$ must not hold for optimality.

(iv) Case of $\delta_1>\delta_1^*$ and $\delta_2>\delta_2^*$:
If $x+y<\Phi'$, it must be $\pi(1)=1$. In the similar manner to the transformation (18.35), we verify $\partial R_1/\partial x \geq \partial R_1/\partial x|_{y=0}>0$ and $\partial R_2/\partial y \geq \partial R_2/\partial y|_{x=0}>0$. If $x+y>\Phi'$ and then $\pi(1)=0$, we have $\partial R_1/\partial x<0$ and $\partial R_2/\partial y<0$ by analogy to the transformation (18.36). $x+y>\Phi'$ never holds for optimal x and y.

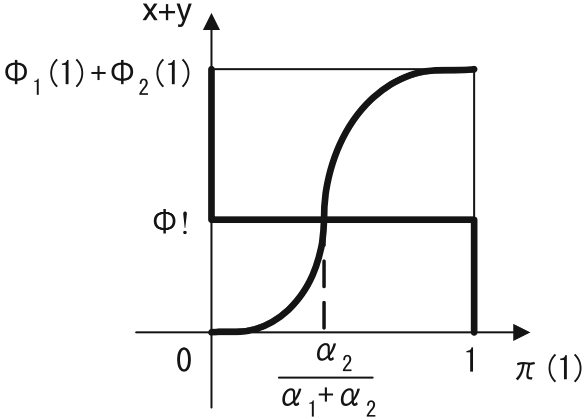
Considering Cases (i)$-$(iv), we must search for a NE under the condition $x+y=\Phi'$. By the substitution of the condition, we can transform Eqs. (18.30) and (18.31) to

$$\frac{\partial R_1}{\partial x} = (\pi(1)\alpha_1-\pi(2)\alpha_2)A_1$$
$$+\delta_1\{\pi(1)\alpha_1\exp(-\alpha_1 x)-\pi(2)\alpha_2\exp(-\alpha_2(\Phi_1(1)-x))\} \tag{18.37}$$
$$\frac{\partial R_2}{\partial y} = (\pi(1)\alpha_1-\pi(2)\alpha_2)A_2$$
$$+\delta_2\{\pi(1)\alpha_1\exp(-\alpha_1 y)-\pi(2)\alpha_2\exp(-\alpha_2(\Phi_2(1)-y))\}. \tag{18.38}$$

In the above expressions, we use notation

$$A_k \equiv \exp(-\alpha_1\Phi')\left[(1-\delta_k)\right.$$
$$\left.-(1-\exp(-\alpha_1\Phi_k(2)))\left\{\exp(-\alpha_1\Phi_j(2))+\delta_k(1-\exp(-\alpha_1\Phi_j(2)))\right\}\right],$$

Fig. 18.2 Optimal $x+y$ and optimal $\pi(1)$



where index (k, j) is one of $(1,2)$ or $(2,1)$. Zero point $(x^*,\ y^*)$ of equations $\partial R_1/\partial x = \partial R_2/\partial y = 0$ gives us the NE of maximizing both payoffs R_1 and R_2. To clarify the relation among optimal x^*, y^* and $\pi(1)$, we solve these equations with respect to $\pi(1)$ using $\pi(2) = 1 - \pi(1)$, and then we have

$$\pi(1) = \frac{\alpha_2(A_1 + \delta_1 \exp(-\alpha_1(\Phi_1(1) - x^*)))}{\alpha_2(A_1 + \delta_1 \exp(-\alpha_1(\Phi_1(1) - x^*))) + \alpha_1(A_1 + \delta_1 \exp(-\alpha_1 x^*))} < 1 \tag{18.39}$$

$$\pi(1) = \frac{\alpha_2(A_2 + \delta_2 \exp(-\alpha_1(\Phi_2(1) - y^*)))}{\alpha_2(A_2 + \delta_2 \exp(-\alpha_1(\Phi_2(1) - y^*))) + \alpha_1(A_2 + \delta_2 \exp(-\alpha_1 y^*))} < 1 \tag{18.40}$$

The functions in the right-hand sides of Eqs. (18.39) and (18.40) are monotone increasing for x^*, y^* and then $x^* + y^*$ is increasing for $\pi(1)$. When we draw the function $x^* + y^*$ and a horizontal line of Φ' on the axis of $\pi(1)$, a crossing point between these two curves gives us optimal $\pi^*(1)$. Figure 18.2 shows the function $x^* + y^*$ with respect to $\pi(1)$ and the optimal response of the target (18.32)–(18.34), in a general way. We might recall that the function $x^* + y^*$ is derived from Eqs. (18.37) and (18.38) under the condition $x + y = \Phi'$. Basically, we should have used the function $x + y$ of $\pi(1)$ directly derived from simultaneous equations $\partial R_1/\partial x = 0$ and $\partial R_2/\partial y = 0$ using Eqs. (18.30) and (18.31). But this derivation would be difficult. Anyway, we obtain the same NE both ways. From Eqs. (18.37) and (18.38), we can make sure that $\partial R_1/\partial x = 0$ and $\partial R_2/\partial y = 0$ hold for variables $\pi(1)$, $\pi(2)$, x and y satisfying $\pi(1)\alpha_1 = \pi(2)\alpha_2$, $\alpha_1 x = \alpha_2(\Phi_1(1) - x))$ and $\alpha_1 y = \alpha_2(\Phi_2(1) - y)$. An equation $x + y = \Phi'$ is also valid. Therefore, we have the following conclusion about the NE and the non-detection probability although we do not dare to present the searchers' payoffs $R_1(\pi^*,x^*,y^*)$ and $R_2(\pi^*,y^*,x^*)$ because of their long expressions.

$$\pi^*(1) = \frac{\alpha_2}{\alpha_1+\alpha_2}, \;\; \pi^*(2) = \frac{\alpha_1}{\alpha_1+\alpha_2} \tag{18.41}$$

$$x^* = \frac{\alpha_2}{\alpha_1+\alpha_2}\Phi_1(1) \tag{18.42}$$

$$y^* = \frac{\alpha_2}{\alpha_1+\alpha_2}\Phi_2(1) \tag{18.43}$$

$$Q(\pi^*,x^*,y^*) = \exp\left(-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}(\Phi_1(1)+\Phi_2(1)) - \alpha_1(\Phi_1(2)+\Phi_2(2))\right). \tag{18.44}$$

The optimal variables π^* and x^*+y^* also give us a NE for the TPZS game with the non-detection probability $Q(\pi,x,y)$ as a payoff, where two searchers are regarded as one player against the target. Thus we might pay attention to the equivalence between the nonzero-sum game with three persons and the zero-sum game with two persons. In the original nonzero-sum game, two searchers do not need to cooperate in searching for the target because parameters δ_1 and δ_2 are not necessarily set to be $\delta_1+\delta_2=1$. Nevertheless, they are possibly motivated to be cooperative by the selfish behavior of the target aiming to minimize the non-detection probability, as we see in this special case of the moving target problem. We can present another case in Sect. 18.4, where the target could exploit the noncooperative behavior between two searchers to direct the situation to the better with less detection probability.

18.3.2 A General Search Game with a Moving Target

We can formulate the problem of finding an optimal response of Searcher k by the following problem $P_k(\varphi_j;\pi)$, given the strategies of the target and the other searcher $j(\neq k)$, π and φ_j.

$$P_k(\varphi_j;\pi): \;\; \max_{\varphi_k,\lambda} \; R_k(\pi,\varphi_k,\varphi_j)$$

$$\text{s.t.} \;\; \sum_{i\in\mathbf{K}} \varphi_k(i,t) \le \Phi_k(t), \; t\in\widehat{\mathbf{T}} \tag{18.45}$$

$$\varphi_k(i,t) \ge 0, \; i\in\mathbf{K}, \; t\in\widehat{\mathbf{T}} \tag{18.46}$$

$$g(\omega,\varphi_k+\varphi_j) = \lambda, \;\; \omega\in\Omega^+(\pi) \tag{18.47}$$

$$g(\omega,\varphi_k+\varphi_j) \ge \lambda, \;\; \omega\in\Omega\backslash\Omega^+(\pi), \tag{18.48}$$

where $\Omega^+(\pi) \equiv \{\omega\in\Omega \,|\, \pi(\omega)>0\}$ and $\Omega\backslash\Omega^+(\pi) = \{\omega\in\Omega \,|\, \pi(\omega)=0\}$.

Conditions (18.47) and (18.48) are necessary to keep π be optimal for φ_1 and φ_2, as seen from Lemma 18.1. We have a theorem for the NE.

Theorem 18.1. *If a sequence of solutions converges to some* $(\varphi_1^*,\ \varphi_2^*)$ *by the repetition of solving Problem* $P_k(\varphi_j;\pi)$ *with fixed* π *for* $(k,j)=(1,2),(2,1)$*, the solution of* π*,* φ_1^**, and* φ_2^* *is a Nash-equilibrium. There exists a Nash-equilibrium for any target strategy* π *if Problem* $P_k(\varphi_j;\pi)$ *is well-defined for* $(k,j)=(1,2),(2,1)$.

Proof. The strategy φ_k^* is evidently an optimal response to other players' strategies π and φ_j^*. The rest we have to prove is to verify the optimality of π for φ_1^* and φ_2^*. Let λ^* be an optimal multiplier λ of Problem $P_k(\varphi_j;\pi)$. The non-detection probability becomes

$$
\begin{aligned}
Q(\pi,\varphi_1^*,\varphi_2^*) &= \sum_{\omega\in\Omega}\pi(\omega)\exp(-g(\omega,\varphi_1^*+\varphi_2^*)) \\
&= \sum_{\omega\in\Omega^+(\pi)}\pi(\omega)\exp(-g(\omega,\varphi_1^*+\varphi_2^*)) \\
&\quad + \sum_{\omega\in\Omega\setminus\Omega^+(\pi)}\pi(\omega)\exp(-g(\omega,\varphi_1^*+\varphi_2^*)) \\
&= \sum_{\omega\in\Omega^+(\pi)}\pi(\omega)\exp(-\lambda^*) = \exp(-\lambda^*).
\end{aligned}
$$

Noting that $Q(\pi',\varphi_1^*,\varphi_2^*)=\exp(-\lambda^*)$ holds for arbitrary $\pi'\in\Pi$ of $\Omega^+(\pi')=\Omega^+(\pi)$. For arbitrary $\pi'\in\Pi$ of $\Omega^+(\pi')\neq\Omega^+(\pi)$, we have the following inequality

$$
\begin{aligned}
&Q(\pi',\varphi_1^*,\varphi_2^*) \\
&\quad = \sum_{\omega\in\Omega^+(\pi)}\pi'(\omega)\exp(-\lambda^*) + \sum_{\omega\in\Omega\setminus\Omega^+(\pi)}\pi'(\omega)\exp(-g(\omega,\varphi_1^*+\varphi_2^*)) \\
&\quad \le \sum_{\omega\in\Omega^+(\pi)}\pi'(\omega)\exp(-\lambda^*) + \sum_{\omega\in\Omega\setminus\Omega^+(\pi)}\pi'(\omega)\exp(-\lambda^*) = \exp(-\lambda^*).
\end{aligned}
$$

This implies that the target does not have any incentive to change his current strategy π.

The problem $P_k(\varphi_j;\pi)$ has a unique solution from their strictly concavity if the problem is well-defined or its feasible region is not empty. A sequence of the solutions is a mapping of a new point (φ_1',φ_2') from an old one (φ_1,φ_2) by solving problems $\varphi_1'=\arg\max_{\varphi_1}R_1(\pi,\varphi_1,\varphi_2)$ and $\varphi_2'=arg\max_{\varphi_2}R_2(\pi,\varphi_2,\varphi_1)$. The mapping is closed from the continuity of functions $R_1(\cdot)$, $R_2(\cdot)$ and the closed convexity of the feasible region defined by conditions (18.45)–(18.48), and therefore it has a fixed point from the Kakutani's fixed-point theorem, that is, (φ_1',φ_2') coincides with (φ_1,φ_2). Therefore, there exists a Nash equilibrium for any target strategy π. □

The iterative algorithm of finding an optimal solution as a convergence point is often used in many problems. However, we also observe that such a direct methodology sometimes fails to find the convergence point by the swing of the temporary solutions in the process of calculation. To avoid the vibration of the solutions, the objective with penalty function could be effective. Let us substitute such function

$$\widetilde{R_k}(\pi, \varphi_k, \varphi_j) \equiv R_k(\pi, \varphi_k, \varphi_j) - \gamma ||\varphi_k - \widehat{\varphi_k}||^2$$

for the original objective function in problem $P_k(\varphi_j;\pi)$ $(k=1,2)$, and denote the renewed problem by $\widetilde{P_k}(\varphi_j;\pi)$. $\widehat{\varphi_k}$ is the current solution of Searcher k's strategy and *gamma* is a parameter for adjustment. If we find a convergence point mentioned in Theorem 18.1, the point is the NE, aside from the algorithmic idea for the practical computation of the NE. We can anticipate that there would be many NEs from Theorem 18.1. We are going to propose a reasonable target strategy π, based on which we can derive the convergence point $(\varphi_1^*,\ \varphi_2^*)$ for optimal searchers' strategies.

A thoughtful target would think of the worst case that searchers' strategies φ_1 and φ_2 are totally against his interest to make the non-detection probability $Q(\pi, \varphi_1, \varphi_2)$ as small as possible. The target has to respond optimally to the worst case that two searchers cooperate in minimizing $Q(\pi, \varphi_1, \varphi_2)$. We can regard the case as a TPZS game with non-detection probability as a payoff. In the game, the target chooses one path $\omega \in \Omega$ as a maximizer and a team of two searchers makes a plan of distribution $\varphi(i,t) = \varphi_1(i,t) + \varphi_2(i,t)$ as a minimizer. The non-detection probability or the payoff is given by

$$Q(\omega, \varphi) = \exp\left(-\sum_{t=\tau}^{T} \alpha_{\omega(t)} \varphi(\omega(t), t) \right),$$

which is modified from Eq. (18.20). Fortunately, we already have a research on this kind of TPZS search game, by Hohzaki [10]. It says that we obtain an optimal strategy of searchers φ^* from the following linear programming formulation:

$$\begin{aligned} P_{\mathrm{S}}:\ & w = \max_{\varphi, \eta}\ \eta \\ \text{s.t.}\ & \sum_{t \in \widehat{\mathbf{T}}} \alpha_{\omega(t)} \varphi(\omega(t), t) \geq \eta,\ \omega \in \Omega \\ & \sum_{i \in \mathbf{K}} \varphi(i,t) = \Phi_1(t) + \Phi_2(t),\ t \in \widehat{\mathbf{T}} \\ & \varphi(i,t) \geq 0,\ i \in \mathbf{K},\ t \in \widehat{\mathbf{T}}, \end{aligned}$$

and an optimal strategy of target π^* from the following problem, which is dual to Problem (P_{S}) above.

$$D_{\mathrm{T}}:\ w = \min_{\nu, \pi} \sum_{t \in \widehat{\mathbf{T}}} \nu(t)(\Phi_1(t) + \Phi_2(t))$$

Table 18.1 Cells in target paths

ω	t 1	2	3	4	5	6	7	8	9	10
ω_1	1	1	1	1	1	1	1	1	1	1
ω_2	2	2	2	2	2	2	2	2	2	2
ω_3	3	3	3	3	3	3	3	3	3	3
ω_4	4	4	4	4	3	3	3	3	3	3
ω_5	1	2	3	3	3	3	3	3	3	3

$$\text{s.t.} \quad \sum_{\omega\in\Omega} \pi(\omega) = 1$$

$$\pi(\omega) \geq 0,\ \omega \in \Omega$$

$$\alpha_i \sum_{\omega\in\Omega_{it}} \pi(\omega) \leq \nu(t),\ i \in \mathbf{K},\ t \in \widehat{\mathbf{T}},$$

where Ω_{it} is a set of paths passing through cell i at time t and is defined by $\Omega_{it} \equiv \{\omega \in \Omega | \omega(t) = i\}$. The resulting non-detection probability is calculated by $\exp(-w)$ using the optimal value w of the problem above.

At the end of this section, we incorporate the discussion so far into an algorithm to derive a NE for the original three-person nonzero-sum search game.

Algorithm AL_{2S}

(i) Solve Problem D_T to derive an optimal target strategy π^*. We also solve Problem P_S for an optimal strategy $\{\varphi^*(i,t)\}$ of the unified searcher. Generate initial temporary strategies for individual searcher k by

$$\varphi_k^0(i,t) = \varphi^*(i,t)\frac{\Phi_k(t)}{\Phi_k(t)+\Phi_j(t)},\ i \in \mathbf{K},\ t \in \widehat{\mathbf{T}}.$$

(ii) Using $\pi = \pi^*$, repeat solving convex problem $\widetilde{P}_k(\varphi_j;\pi)$ for $(k,j) = (1,2)$ and $(2,1)$ by turns. If their solutions φ_1^* and φ_2^* converge, the obtained π, φ_1^* and φ_2^* are a Nash equilibrium. The resulting payoff of each player are given by $Q(\pi,\varphi_1^*,\varphi_2^*)$ and $R_k(\pi,\varphi_k^*,\varphi_j^*)$, $(k,j) = (1,2),(2,1)$.

18.4 Numerical Examples

We took a small size of problem in Sect. 18.3.1 to derive an analytical form of NE. Here we take a larger size of problem to numerically analyze the property of the NE by applying the algorithm proposed in Sect. 18.3.2.

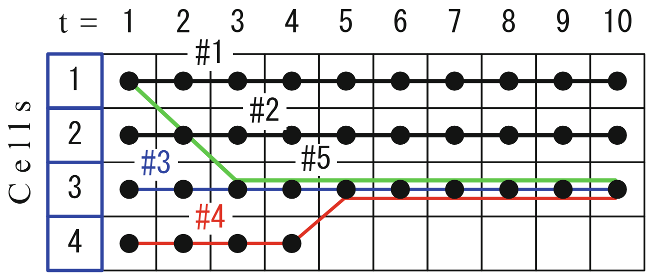


Fig. 18.3 Target paths

Table 18.2 Initial temporary strategy of a searcher

	t									
Cells	1	2	3	4	5	6	7	8	9	10
1	0.173	0.327	0.250	0.250	0.111	0.111	0.111	0.111	0.111	0.111
2	0.327	0.173	0.250	0.250	0.111	0.111	0.111	0.111	0.111	0.111
3	0	0	0	0	0.278	0.278	0.278	0.278	0.278	0.278
4	0	0	0	0	0	0	0	0	0	0
$\Phi_k(t)$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

We consider five target paths, $\Omega = \{1,\cdots,5\}$, running in a search space of four cells $\mathbf{K} = \{1,\cdots,4\}$ and 10 time points $\mathbf{T} = \widehat{\mathbf{T}} = \{1,\cdots,10\}$. Table 18.1 and Fig. 18.3 show the route of each path $\{\omega(t),\ t \in \mathbf{T}\}$ in the space of cells and time points. Path 1, 2 and 3 always stay in Cell 1, 2 and 3, respectively. Path 4 stays in Cell 4 in the early time but moves to Cell 3 at time 5. Path 5 moves accross some cells and stays in Cell 3 after time 3. The effectiveness of searching resource is the same in all cells, $\alpha_i = 0.2$ $(i = 1,\cdots,4)$, and two searchers have the same amount of available resources at each time, $\Phi_1(t) = \Phi_2(t) = 0.5$ $(t \in \widehat{\mathbf{T}})$.

Solving problem D_{T}, we have a target strategy $(\pi^*(1),\ \pi^*(2),\ \pi^*(3),\ \pi^*(4),\ \pi^*(5)) = (1/3,\ 1/3,\ 1/6,\ 1/6,\ 0)$. Path 5 has many crossing points between other paths. The crossing point is a good place for the searchers to efficiently focus their resource on because the resource distributed there can cover several paths simultaneously. That is why the target avoids taking the path and its selection probability becomes zero. Without Path 5, two paths 1 and 2 are running independently of any other paths. Path 3 and 4 are in the same situation that they runs independently during a time period from time 1 to 4 and meet at time $t = 5$ to stay in Cell 3 after then. Considering these situation, the selection probability of path π^* stated above is persuadable. From the solution of problem P_{S}, we generate the initial temporary strategy of Searcher $k = 1,2$, φ_k^0, shown in Table 18.2.

As the weighted amount of searching resource,

$$\{g(\omega_k,\varphi^0), k = 1,\ldots,5\} = \{0.667, 0.667, 0.667, 0.667, 0.805\}$$

are distributed on five target paths $\omega_1,\ldots,\omega_5$, respectively. The values correspond to the target strategy $\pi^*(\omega)$, as mentioned in Lemma 18.1. Using π^* and φ_k^0 $(k = 1,2)$,

Table 18.3 Optimal distribution of resource (Case 2: $\pi = \widetilde{\pi} = (0.4,\ 0.4, 0,\ 0,\ 0.2)$, order of $k = 1, 2$)

	t									
Cells	1	2	3	4	5	6	7	8	9	10
Searcher 1										
1	0	0.254	0.200	0.200	0.149	0.149	0.150	0.150	0.150	0.150
2	0.255	0	0.200	0.199	0.149	0.149	0.149	0.150	0.150	0.150
3	0.173	0.173	0	0	0.202	0.201	0.201	0.201	0.201	0.200
4	0.072	0.073	0.100	0.101	0	0	0	0	0	0
$\Phi_1(t)$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Searcher 2										
1	0.174	0.328	0.250	0.250	0.111	0.111	0.111	0.111	0.111	0.111
2	0.326	0.172	0.250	0.250	0.111	0.111	0.111	0.111	0.111	0.111
3	0	0	0	0	0.277	0.278	0.278	0.278	0.278	0.278
4	0	0	0	0	0	0	0	0	0	0
$\Phi_2(t)$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

we apply Algorithm AL_{2S} to three cases of $(\delta_1,\ \delta_2) = (1, 1)$ (Case 1), $(0.8, 0.2)$ (Case 2) and $(0, 0)$ (Case 3) to derive the NEs. Optimal distributions of searching resource are almost the same as Table 18.1 with some small differences in the three cases. They give the target almost the same non-detection probability 0.513 although the detailed probabilities are 0.51346, 0.51347 and 0.51348 for Case 1, 2 and 3, respectively, reasonably reflecting the favorableness of the simultaneous detection based on δ-value. By these NEs, Searcher 1 and 2 get the rewards $(0.249, 0.249)$ (Case 1), $(0.247, 0.240)$ (Case 2) and $(0.238, 0.238)$ (Case 3). The reward tends to decrease as the value δ_k gets smaller. However, we can say that the influence of the simultaneous detection by both searchers on the reward is not so large as the total detection probability by either searcher. That is why, for any case, the optimal distribution of resource is near to the initial distribution φ_k^0, which is derived from Problem P_S under the criterion of total detection probability.

We check another target strategy $\pi = \widetilde{\pi} = (0.4,\ 0.4,\ 0,\ 0,\ 0.2)$ different from π^*, in Case 2. After applying Algorithm AL_{2S} to this case, we have Table 18.3 as an optimal distribution φ_k^* for two searchers $(k = 1, 2)$.

In this case, the resulting non-detection probability is 0.525 and the rewards are 0.214 and 0.261 for Searcher 1 and 2. Searcher 2 can expect more reward than Searcher 1 in spite of $\delta_1 = 0.8$ and $\delta_2 = 0.2$. The results are advantageous to the target and Searcher 2 but disadvantageous to Searcher 1, comparing with the results by the original target strategy π^*. The advantage depends on the order of calculation in Step (ii) of Algorithm AL_{2S}. The results above are brought by the order $k = 1, 2$. If we change the order to $k = 2, 1$, we have the distribution obtained by exchanging two distribution plans for two searchers in Table 18.3. The results also have the same non-detection probability as the above but bring the expected rewards 0.266 and 0.208 to Searcher 1 and 2, respectively. The rewards become advantageous to Searcher 1 but disadvantageous to Searcher 2. These phenomena often appear in the repeated game or the game with a leader and a follower, where the leading player with the declaration of his intention or his strategy is usually in the favoring position. In Algorithm AL_{2S}, $\varphi_k^0(i,t)$ is declared first by Searcher 2 and used in the first calculation as fixed parameters in the case of order $k = 1,2$. In the case of order $k = 2,1$, the first declaration is done by Searcher 1. Anyway, these distributions are both the NEs or there are two NEs at least for the target strategy $\widetilde{\pi}$. Both these NEs are more favorite than the NE for π^* for the target. Now we may have the lesson that the target could lead the game to more advantageous Nash-equilibrium points if it lets the searchers carry such a belief on the target strategy as $\widetilde{\pi}$.

18.5 Conclusion

In this paper, we deal with a three-person nonzero-sum noncooperative search game, where a target and two searchers compete against one another. For the game with a stationary target, we show that there always exists the Nash-equilibrium (NE) point in which cooperation occurs between the searchers against the target under the common criterion of non-detection probability. We also demonstrate a special example of the moving-target game, which has a Nash-equilibrium in the same situation with the cooperation of two searchers as the stationary-target game. In the situation, the game can be regarded as a two-person zero-sum game between the target and a group of two searchers with the payoff of non-detection probability of the target. In a general case, we prove the existence of NE for any target strategy on path-selection if the problem is well-defined, in Theorem 18.1. We can anticipate that there would be many NEs for a general game with a moving target. We propose a computational algorithm to derive a NE. Applying the algorithm to some examples, we see that there could be the situation in which the target has to fight a coalition of two searchers. At the same time, there could be another situation that the target leads the game to his advantage by exploiting the noncooperative relation between two searchers.

We originally started this research on the game expecting the extension of our model to a cooperative game with two groups of searchers and a target, who compete against one another. Through the evaluation of reward given to searcher's groups, we could define the characteristic function for the coalition of searchers and discuss the complicated search operation with many searchers chasing a target as an ordinary coalition game. This is our work in the near future.

Acknowledgements We appreciate the Ministry of Education, Culture, Sports, Science and Technology of Japan, which partially supports the present research by Grants-in-Aid for Scientific Research (C), Grant No. 22510169 (2010–2012).

References

1. Alpern, S., Gal, S.: The Theory of Search Games and Rendezvous. Kluwer Academic, Boston (2003)
2. Baston, V.J., Bostock, F.A.: A one-dimensional helicopter-submarine game. Naval Res. Logistics **36**, 479–490 (1989)
3. Baston, V.J., Garnaev, A.Y.: A search game with a protector. Naval Res. Logistics. **47**, 85–96 (2000)
4. Dambreville, F., Le Cadre, J.P.: Search game for a moving target with dynamically generated informations. In: Proceedings of the 5th International Conference on Information Fusion (FUSION'2002), pp. 243–250 (2002)
5. Danskin, J.M.: A helicopter versus submarine search game. Oper. Res. **16**, 509–517 (1968)
6. Eagle, J.N., Washburn, A.R.: Cumulative search-evasion games. Naval Res. Logistics **38**, 495–510 (1991)
7. Garnaev, A.Y.: A remark on a helicopter-submarine game. Naval Res. Logistics **40**, 745–753 (1993)
8. Hohzaki, R.: Search allocation game. Eur. J. Oper. Res. **172**, 101–119 (2006)
9. Hohzaki, R.: A multi-stage search allocation game with the payoff of detection probability. J. Oper. Res. Soc. Jpn. **50**, 178–200 (2007)
10. Hohzaki, R.: A cooperative game in search theory. Naval Res. Logistics **56**, 264–278 (2009)
11. Hohzaki, R., Iida, K.: A search game with reward criterion. J. Oper. Res. Soc. Jpn. **41**, 629–642 (1998)
12. Hohzaki, R., Iida, K.: A solution for a two-person zero-sum game with a concave payoff function. In: Takahashi, W., Tanaka, T. (eds.) Nonlinear Analysis and Convex Analysis, pp. 157–166. World Science Publishing Co., London (1999)
13. Hohzaki, R., Iida, K., Komiya, T.: Discrete search allocation game with energy constraints. J. Oper. Res. Soc. Jpn. **45**, 93–108 (2002)
14. Hohzaki, R., Washburn, A.: An approximation for a continuous datum search game with energy constraint. J. Oper. Res. Soc. Jpn. **46**, 306–318 (2003)
15. Iida, K., Hohzaki, R., Furui, S.; A search game for a mobile target with the conditionally deterministic motion defined by paths. J. Oper. Res. Soc. Jpn. **39**, 501–511 (1996)
16. Iida, K., Hohzaki, R., Sato, K.: Hide-and-search game with the risk criterion. J. Oper. Res. Soc. Jpn. **37**, 287–296 (1994)
17. Kikuta, K.: A search game with traveling cost. J. Oper. Res. Soc. Jpn. **34**, 365–382 (1991)
18. Koopman, B.O.: The theory of search. III. The optimum distribution of searching effort. Oper. Res. **5**, 613–626 (1957)
19. Meinardi, J.J.: A sequentially compounded search game. In: Mensch A. (ed.) Theory of Games: Techniquea and Applications, pp. 285–299. The English Universities Press, London (1964)
20. Morse, P.M., Kimball, G.E.: Methods of Operations Research. MIT, Cambridge (1951)
21. Myerson, R.B.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge (1991)
22. Nakai, T.: A sequential evasion-search game with a goal. J. Oper. Res. Soc. Jpn. **29**, 113–122 (1986)
23. Nakai, T.: Search models with continuous effort under various criteria. J. Oper. Res. Soc. Jpn. **31**, 335–351 (1988)
24. Norris, R.C.: Studies in search for a conscious evader. MIT Technical Report, No. 279 (1962)
25. O'Rourke, J.: Art Gallery Theorems and Algorithms. Oxford University Press, New York (1987)
26. Owen, G.: Game Theory, pp. 212–224. Academic, New York (1995)
27. Parsons, T.D.: Pursuit-Evasion in a Graph. In: Alavi, Y., Lick, D.R. (eds.) Theory and Applications of Graphs. Springer, Berlin (1976)

28. Parsons, T.D.: The search number of a connected graph. In: Proceedings of the 10th Southeastern Conference on Combinatorics, Graph Theory and Computing, pp. 549–554 (1978)
29. Washburn, A.R.: Search-evasion game in a fixed region. Oper. Res. **28**, 1290–1298 (1980)
30. Washburn, A.R., Hohzaki, R.: The diesel submarine flaming datum problem. Milit. Oper. Res. **4**, 19–30 (2001)

Chapter 19
Advertising and Price to Sustain The Brand Value in a Licensing Contract

Alessandra Buratto

Abstract One of the reasons that induce a brand owner to issue a licensing contract is that of improving the value of his brand. In this paper, we look at a fashion licensing agreement where the licensee produces and sells a product in a complementary business. The value of a fashion brand is sustained by both the advertising efforts of the licensor and the licensee. We assume that demand is proportional to the brand value and decreases with the price. The licensor wants to maximize his revenue coming from the royalties and to minimize his advertising costs. Moreover, he does not want his brand to be devalued at the end of the selling season. On the other hand, the licensee plans her advertising campaign in order to invest in the brand value and maximize the sales revenue. The aim of this paper is to analyze the different strategies the licensor can adopt to sustain his brand. To this end, we determine the optimal advertising policies by solving a Stackelberg differential game, where the owner of the brand acts as the leader and the licensee as the follower. We determine the equilibrium policies of the two players assuming that advertising varies over time and price is constant. We also determine a minimum selling price which guarantees brand sustainability without advertising too much.

Keywords Game theory • Advertising • Licensing • Brand value

19.1 Introduction

Let's consider a licensing contract between the owner of a brand (licensor) and a manufacturer (licensee) who produces and sells a product with the licensor's brand. We focus on a particular type known as complementary business licensing which

A. Buratto (✉)
Department of Mathematics, Via Trieste 63, I-35131 Padova, Italy
e-mail: buratto@math.unipd.it

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_19,

is, for instance, the case with an owner of a brand of clothes that licenses his brand to a producer of accessories or perfumes [19]. This type of contract is very common nowadays. In fact, "Brands allow larger firms to diversify from clothing into other markets, outside their core business: perfumes, accessories ..." [17]. Licensing itself can be considered to be a brand extension strategy according to Keller's definition [12].

A licensing contract may last for several years, even decades. Nevertheless, at the beginning of every selling season for each new product, the two agents involved have to come to an agreement in setting the selling price and must plan the advertising campaigns they are respectively going to carry out. Advertising coordination is quite important in any vertical channel and becomes crucial in a licensing agreement [18].

The importance of the brand value is well known, especially in the fashion business, [1]. Several papers study brand dilution and imitation in the fashion industry, see [5, 6, 9]. Its links to the price have been studied since the early fifties, with the analysis of the different effects of price over demand, [13]. More recently in the game theory context such effects have been formalized. For example, in [2, 15] it is stressed that "In fashion goods price increases the brand value." In [5] the authors say that the price of prestige goods should not be too low and licensing is a mechanism for expanding sales; however, one of its risks is brand dilution. On the other hand, advertising may increase the brand value and a good advertising campaign can be useful in guaranteeing brand sustainability.

Here we tackle the issue in a game theoretical context, taking into account production costs too. In this paper we wish to determine the equilibrium policies of a licensor assuming that advertising varies over time and price is constant. We conduct a sensitivity analysis with respect to the price parameter, just in order to see if some particular prices can guarantee brand sustainability. Imposing to the licensee a given price, among these values, is one of the licensor's strategies to sustain the brand itself.

The aim of this paper is to provide a guideline for the owner of a brand who is concerned about the sustainability of his brand. We analyze different approaches he can follow in order to achieve this task, in view of the different dynamics between the two agents involved in the agreement. We consider the selling price as already fixed and we tackle the problem of determining the advertising campaigns by solving a Stackelberg game with the Licensor as the leader and the licensee as the follower. In fact, the licensor lays down the law in a licensing contract and can sever the agreement whenever the licensee does not meet his target [18]. Each player determines his optimal advertising strategy, maximizing the profits coming from the sale of the product and minimizing the advertising costs. Similar approaches, within differential game theory, are common in order to determine the advertising strategies in a vertical channel, see [8, 11], and they have been used in the franchising context too (see f.e. [7, 14]). An attempt has been made in [4] for a licensing contract without considering brand sustainability.

In the following we will analyze the brand sustainability problem from the licensor's point of view.The licensor can achieve the sustainability of his brand through advertising, either cooperating with the licensee in maximizing the sum of their profits and then both of them have to take into account the production costs and to take care of the brand value sustainability or in a competitive context, such as:

- By sharing with the licensee the necessary increase of the advertising effort.
- By increasing his usual advertising effort, without binding the licensee to do the same.
- By forcing the licensee to increase her advertising effort, considering the sustainability constraint in her advertising plan.

Alternatively

- He can wonder if there exists a minimum selling price to impose to the licensee, such that the brand value is sustained without spending too much for advertising.

We will analyze the licensor's brand sustainability problem in Sect. 19.2, in Sect. 19.3 we calculate the advertising strategies of the two players and we give an operational rule for the licensor to choose which is the optimal advertising effort with respect to the selling price value. Moreover we determine a minimum selling price which guarantees brand sustainability. In Sect. 19.4 we consider the fully cooperative solution and compare it to the Stackelberg solution, wondering if the former type of strategy is effectively the best option for the licensor.

19.2 Brand Value Evolution

Let's denote by $[0,T]$ the selling season, during which the two players make their advertising campaigns. In view of the short-term nature of the problem we do not discount future profits. We model the brand value using the goodwill state variable introduced by Nerlove and Arrow [16]. Let's denote by $G(t)$ the brand value, goodwill, at time $t \in [0,T]$ and by $G_0 > 0$ the initial brand value which we assume to be sufficiently high, just because only famous brands are licensed.

We assume that the brand value is increased by both the advertising efforts, $a_L(t) \geq 0$, $a_l(t) \geq 0$ and the price difference, as follows

$$\dot{G}(t) = \gamma_L a_L(t) + \gamma_l a_l(t) - \delta G(t) + \beta\,(p - p_R)\,, \quad t \in [0,T]\,, \tag{19.1}$$

where $\gamma_L > 0$ and $\gamma_l > 0$ are the advertising efficiencies of the two advertising messages and $\delta > 0$ is the decaying rate. Observe that the index $i = L$ stands for the licensor, while the index $i = l$ stands for the licensee.

The additive term $\beta\,(p - p_R)$ represents the *snob effect* by which the brand value increases with the selling price, p; while p_R is a reference price, known as *regular price* [15]; it is the price the licensor considers fair/proper for this type of branded product. If the selling price is greater than the reference price, then the brand value will be increased. On the other hand, if the selling price is too low, $p < p_R$, then the brand is under valued. Therefore $\beta \geq 0$ is called price sensitivity toward the brand value.

From now on, we will refer to *brand value sustainability* as the request of having a brand value level greater than or equal to the initial one at the end of the selling season. Being

$$G(0) = G_0, \tag{19.2}$$

brand value sustainability can be formalized by means of the constraint

$$G(T) \geq G_0. \tag{19.3}$$

19.3 Stackelberg Game

Let's consider a differential game played à la Stackelberg, with the licensor as the leader and the licensee as the follower and let's determine the equilibrium advertising strategies.

19.3.1 The Licensee's Point of View

The licensee's objective is to determine the intensity of the advertising campaign in order to maximize his profit, defined as the difference between sales revenue and production and advertising costs.

We assume that demand is increasing with the brand value and decreasing with the price,

$$D(G) = G(t) - \theta p, \tag{19.4}$$

where $\theta > 0$ is the price sensitivity of the demand. The revenue rate from the sales is then

$$p\left(G(t) - \theta p\right), \quad t \in [0, T]. \tag{19.5}$$

For granting the use of the brand, the licensee has to pay the royalties to the licensor, they generally consist of a percentage of the sales, that is

$$rp\left(G(t) - \theta p\right), \quad t \in [0, T], \tag{19.6}$$

where $r \in (0, 1)$ is the royalty percentage. Observe that the royalty percentage of the revenues from the sales that the licensee has to pay to the licensor is also exogenous and constant, and as a consequence, the unique strategic marketing instruments are the advertising efforts of the licensor and the licensee. The licensee's revenue after paying the royalties is therefore

$$(1 - r)p\left(G(t) - \theta p\right), \quad t \in [0, T]. \tag{19.7}$$

Production costs are linked to the demand according to the following linear form

$$C_{pr}(G(t)) = c\,(G(t) - \theta p)\,, \quad t \in [0,T]\,, \tag{19.8}$$

with $c > 0$, marginal production cost. We assume that

$$(1-r)p \geq c,$$

this means requiring that marginal revenues are greater than marginal costs and such an assumption is reasonable for any manufacturer.

We consider quadratic advertising costs

$$C^l_{pu}(a_l) = \frac{1}{2}c_l a_l^2 \quad t \in [0,T]\,, \tag{19.9}$$

with $c_l > 0$ as licensor's cost parameter.

The licensee has to solve the following problem

$$P^1_l\,:\; \max\,_{a_l \geq 0} J^1_l\,(a_L, a_l, p) = \int_0^T \left[((1-r)p - c)(G(t) - \theta p) - \frac{c_l}{2}a_l^2\,(t)\right]\,\mathrm{d}t, \tag{19.10}$$

under constraints (19.1)–(19.3).

Observe that we denote by P^1_i the optimization problem for agent $i \in \{L,l\}$ which maximizes his/her profit and considers brand sustainability, that is $P^1_i = \max J_i$ subject to (19.1)–(19.3). Similarly we denote by P^0_i the optimization problem for agent $i \in \{L,l\}$ which maximizes his/her profit without considering brand sustainability, that is $P^0_i = \max J_i$ subject to (19.1) and (19.2) only. Let $J^{0/1}_i, i \in \{L,l\}$ be the profits associated to problems $P^{0/1}_i$ respectively.

19.3.2 The Licensor's Point of View

The licensor's has his own advertising costs, supposed quadratic too

$$C^L_{pu}(a_L) = \frac{1}{2}c_L a_L^2, \quad t \in [0,T]\,, \tag{19.11}$$

and he obtains from the licensee the royalties, given in (19.6). The licensor's problem is

$$P^1_L\,:\; \max\,_{a_L \geq 0}\, J^1_L(a_L, a_l, p) = \int_0^T \left[rp\,(G(t) - \theta p) - \frac{c_L}{2}a_L^2(t)\right]\,\mathrm{d}t \tag{19.12}$$

under constraints (19.1)–(19.3).

In order to write the advertising efforts which constitute the Stackelberg Equilibrium for this game, it turns out convenient to introduce the following parameter

$$\eta_0 = \frac{1}{1+\mathrm{e}^{-\delta T}}\left\{2(\delta G_0 - \beta(p-p_R)) - (1-\mathrm{e}^{-\delta T})\left[\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)p-c}{\delta}\right]\right\}. \tag{19.13}$$

Let's observe that $\eta_0 \leq 0$ if and only if

$$p \geq p_\mathrm{s} = \frac{G_0 + \frac{\beta}{\delta}p_R + c\frac{\gamma_l^2}{c_l}\left(\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\right)}{\frac{\beta}{\delta} + \left[\frac{\gamma_L^2}{c_L}r + \frac{\gamma_l^2}{c_l}(1-r)\right]\left(\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\right)}. \tag{19.14}$$

Theorem 19.1 (Stackelberg Equilibria). *(I) If $p \geq p_\mathrm{s}$, then the advertising efforts which constitute the Stackelberg equilibrium are*

$$a_L^{0*}(t) = \frac{\gamma_L}{c_L}\left[\frac{rp}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right)\right] \tag{19.15}$$

and

$$a_l^{0*}(t) = \frac{\gamma_l}{c_l}\frac{(1-r)p-c}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right). \tag{19.16}$$

(II) If $p < p_\mathrm{s}$, then the advertising efforts $a_L^{0*}(t)$ and $a_l^{0*}(t)$ are not sufficient to sustain the brand value. In order to do so, the licensor can choose among the following three different options.

- *He can share with the licensee the additional advertising effort necessary to sustain the brand. The optimal advertising efforts are respectively*

$$a_L^{11*}(t) = a_L^{0*}(t) + \eta_L\frac{\gamma_L}{c_L}\mathrm{e}^{-\delta(T-t)}, \tag{19.17}$$

$$a_l^{11*}(t) = a_l^{0*}(t) + \eta_l\frac{\gamma_l}{c_l}\mathrm{e}^{-\delta(T-t)}, \tag{19.18}$$

where η_L and η_l satisfy the equation

$$\frac{\gamma_L^2}{c_L}\eta_L + \frac{\gamma_l^2}{c_l}\eta_l = \eta_0 \tag{19.19}$$

The two players join their efforts in advertising in order to sustain the brand and infinitely many solutions may exist.

- *He can increase his advertising adopting the optimal advertising effort*

$$a_L^{1*}(t) = a_L^{0*}(t) + \frac{\eta_0}{\gamma_L}\mathrm{e}^{-\delta(T-t)}, \tag{19.20}$$

while the licensee's advertising effort is the minimum one given in (19.16)

$$a_l^{0*}(t) = \frac{\gamma_l}{c_l}\frac{(1-r)p-c}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right).$$

Observe that in this scenario, the licensee does not take into account the brand sustainability.

- *He can bind the licensee to consider the sustainability constraint in her advertising plan, whereas he can neglect it. The licensor's optimal advertising effort is the one given in (19.15)*

$$a_L^{0*}(t) = \frac{\gamma_L}{c_L}\left[\frac{rp}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right)\right],$$

while the licensee's optimal advertising effort is

$$a_l^{1*}(t) = a_l^{0*}(t) + \frac{\eta_0}{\gamma_l}\mathrm{e}^{-\delta(T-t)}. \tag{19.21}$$

Proof. See Appendix.

It can be easily proved that the open loop Stackeberg Equilibrium in all cases coincides with the Markovian Nash Equilibrium, so that it is time consistent. Observe that if $p_s < c/(1-r)$, then only case I can occur and therefore brand sustainability is assured with the optimal advertising efforts as in (19.15) and (19.16). □

Theorem 19.1 states that if the selling price, which we assume to be constant, is greater than or equal to a minimum price p_s, then the players can limit their advertising effort, as if they were neglecting the constraint on the final brand value. In fact, in such a case, the optimal advertising strategies coincide with the ones they would have obtained without considering the brand sustainability. Note that if the licensor imposes the minimum selling price p_s to the licensee, then he can be sure that his brand value will be automatically increased, without any further advertising.

In such a situation, both the licensor's and the licensee's advertising efforts are positive, concave, decrease with time and vanish at the end of the selling season, in fact

$$\frac{\mathrm{d}a_L^{0*}(t)}{\mathrm{d}t} = -\frac{\gamma_L rp}{c_L}\mathrm{e}^{-\delta(T-t)} < 0, \qquad \frac{\mathrm{d}^2 a_L^{0*}(t)}{\mathrm{d}t^2} = \delta\frac{\mathrm{d}a_L^{0*}(t)}{\mathrm{d}t} < 0,$$
$$\frac{\mathrm{d}a_l^{0*}(t)}{\mathrm{d}t} = -\frac{((1-r)p-c)\gamma_l}{c_l}\mathrm{e}^{-\delta(T-t)} < 0, \quad \frac{\mathrm{d}^2 a_l^{0*}(t)}{\mathrm{d}t^2} = \delta\frac{\mathrm{d}a_L^{0*}(t)}{\mathrm{d}t} < 0.$$

On the other hand, if the selling price is lower than the threshold p_s then at least one of the two actors has to increase his/her advertising effort. In the case in which only one player increases his/her advertising, the derivatives of the advertising effort of the player who considers brand sustainability are respectively

$$\frac{\partial a_L^{1*}(t)}{\partial t} = \delta\,\mathrm{e}^{-\delta(T-t)}\left(\frac{\eta_0}{\gamma_L} - \frac{\gamma_L}{c_L}\frac{rp}{\delta}\right), \qquad \frac{\partial^2 a_L^{1*}(t)}{\partial t^2} = \delta\frac{\partial a_L^{1*}(t)}{\partial t},$$

$$\frac{\partial a_l^{1*}(t)}{\partial t} = \delta\,\mathrm{e}^{-\delta(T-t)}\left(\frac{\eta_0}{\gamma_l} - \frac{\gamma_l}{c_l}\frac{(1-r)p-c}{\delta}\right), \qquad \frac{\partial^2 a_l^{1*}(t)}{\partial t^2} = \delta\frac{\partial a_l^{1*}(t)}{\partial t}.$$

The advertising efforts turn out to be either increasing and convex, this for very low prices, that is for $p < \hat{p}_i,\ i \in \{L,l\}$, or decreasing and concave, for intermediate price values, that is for $p > \hat{p}_i$, or finally constant if $p = \hat{p}_i,\ i \in \{L,l\}$ where

$$\hat{p}_L = \frac{G_0 + \frac{c}{\delta^2}\frac{\gamma_l^2}{c_l}\left(\frac{1-\mathrm{e}^{-\delta T}}{2}\right) + \frac{\beta p_R}{\delta}}{\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\left[\frac{\gamma_L^2}{c_L}r + \frac{\gamma_l^2}{c_l}(1-r)\right] + \frac{\beta}{\delta} + \frac{\gamma_L^2 r}{c_L}\frac{1+\mathrm{e}^{-\delta T}}{2\delta^2}} < p_{\mathrm{s}}$$

and

$$\hat{p}_l = \frac{G_0 + \frac{c}{\delta^2}\frac{\gamma_l^2}{c_l} + \frac{\beta p_R}{\delta}}{\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\left[\frac{\gamma_L^2}{c_L}r + \frac{\gamma_l^2}{c_l}(1-r)\right] + \frac{\beta}{\delta} + \frac{\gamma_L^2(1-r)}{c_l}\frac{1+\mathrm{e}^{-\delta T}}{2\delta^2}}.$$

It is interesting to conduct a sensitivity analysis of the optimal minimum selling price, p_{s}, with respect to values of the problem's parameters. Speaking about prices, it seems reasonable to compare the minimum selling price with the regular price p_{R}. The minimum selling price is linear in p_{R} and monotonically increasing, that is, the higher the regular price, the higher the minimum selling price must be to sustain the brand. Moreover if the regular price is low, more precisely smaller that a given threshold $\underline{p_{\mathrm{R}}}$,

$$\underline{p_{\mathrm{R}}} = \frac{G_0 + c\frac{\gamma_l^2}{c_l}\left(\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\right)}{\left[\frac{\gamma_L^2}{c_L}r + \frac{\gamma_l^2}{c_l}(1-r)\right]\left(\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\right)},$$

then the optimal selling price must necessarily be greater than the regular price p_{R} itself, whereas with a high regular price, greater than the threshold $\underline{p_{\mathrm{R}}}$, then the selling price can be lower than p_{R}.

For what concerns the dependence on the royalty coefficient, r, if the licensor's advertising effectiveness is greater than the licensee's one, that is if $\frac{\gamma_L^2}{c_L} > \frac{\gamma_l^2}{c_l}$, then the minimum selling price decreases with the royalty coefficient. The opposite happens if the licensor's advertising is less effective than the licensee's one. Observe that the asymmetry of the game influences its solution; if we had assumed $\gamma_L = \gamma_l$ and $c_L = c_l$, then a substantial difference on the equilibria would hold, for example the minimum price p_{s} would not depend on the royalty's percentage r at all.

Other behaviors are summarized in the following table (sign "+" means increasing)

	G_0	c	γ_L	c_L
p_s		++	−	+

19.4 Fully Cooperative Solution

Here we consider the situation in which the licensor cooperates with the licensee in maximizing the sum of the profits and both of them take care of the brand value sustainability. They have to solve the following optimal control problem

$$P_C^1 = \max_{a_L,a_l \geq 0} J_C^1 = \max_{a_L,a_l \geq 0} J_L^1(a_L,a_l,p) + J_l^1(a_L,a_l,p)$$
$$= \max_{a_L,a_l \geq 0} \int_0^T \left[(p-c)(G(t)-\theta p) - \left(\frac{c_L}{2}a_L^2(t) + \frac{c_l}{2}a_l^2(t)\right)\right] \mathrm{d}t$$

subject to (19.1)–(19.3).

Theorem 19.2 (Cooperative Equilibrium). *The coordinated optimal advertising efforts are*

$$a_{LC}^*(t) = \frac{\gamma_L}{c_L}\left[\frac{p-c}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right) + \eta_C \mathrm{e}^{-\delta(T-t)}\right], \tag{19.22}$$

$$a_{lC}^*(t) = \frac{\gamma_l}{c_l}\left[\frac{p-c}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right) + \eta_C \mathrm{e}^{-\delta(T-t)}\right]. \tag{19.23}$$

where

$$\eta_C = \max\left\{\frac{2(\delta G_0 - \beta(p-p_{\mathrm{R}}))}{\left(\frac{\gamma_L^2}{c_L}+\frac{\gamma_l^2}{c_l}\right)(1+\mathrm{e}^{-\delta T})} - \frac{p-c}{\delta}\frac{1-\mathrm{e}^{-\delta T}}{1+\mathrm{e}^{-\delta T}}, 0\right\}.$$

Proof. See Appendix.

Observe that the equilibrium advertising efforts are strictly monotonically increasing in η_C. We can adopt the same argument as in the Stackelberg problem and obtain that $\eta_C \leq 0$ if and only if

$$p \geq p_C = \frac{G_0 + \frac{\beta}{\delta}p_R + \left(\frac{\gamma_L^2}{c_L}+\frac{\gamma_l^2}{c_l}\right)c\left(\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\right)}{\frac{\beta}{\delta} + \left(\frac{\gamma_L^2}{c_L}+\frac{\gamma_l^2}{c_l}\right)\left(\frac{1-\mathrm{e}^{-\delta T}}{2\delta^2}\right)}.$$

and therefore,

- If $p \geq p_C$, then $\eta_C = 0$ and therefore the cooperative advertising efforts $a^*_{LC}(t)$ and $a^*_{lC}(t)$ reduce to the minimum, just because of their monotonicity in η_C. Nevertheless they are greater than the minimum advertising efforts $a^{0*}_L(t)$ and $a^{0*}_l(t)$ obtained without considering brand sustainability.
- If $p < p_C$, then $\eta_C > 0$, and the players have to increase their advertising efforts. The cooperative advertising efforts in (19.22) and (19.23) are all the more reason greater than the minimum advertising efforts $a^{0*}_L(t)$ and $a^{0*}_l(t)$. Nonetheless, it's not possible to determine a priori neither if they are greater or lower than the increased advertising efforts in (19.17), (19.20), (19.18) and (19.21). In order to compare p_s and p_C, many parameters influence their values and generally not all the presented scenarios are practicable. It is easy to prove that cases $c/1-r < p_s < p_c$ and $p_s < c/1-r < p_c$ never occur, in fact from $p_s < p_c$, it follows that $p_c < c/1-r$. Furthermore, let be $T = 30$, $c_L = 0.15$, $c_l = 0.2$, $c = 8.1$, $p_R = 8$, $\beta = 0.5$, $\gamma_L = 0.75$, $\gamma_l = 0.7$, $r = 0.1$ and $\delta = 0.1$; then $c/1-r = 9$ and according to the value of G_0, we have the following results
  - If $G_0 = 50$, then $p_c = 8.265239040$, $p_s = 8.095855766$ and therefore
  $$p_s < p_c < c/1-r;$$
  - If $G_0 = 100$, then $p_c = 8.432147162$, $p_s = 8.487774502$ and therefore
  $$p_c < p_s < c/1-r;$$
  - If $G_0 = 200$, then $p_c = 8.765963407$, $p_s = 9.271611973$ and therefore
  $$p_c < c/1-r < p_s;$$
  - If $G_0 = 300$, then $p_c = 9.099779651$, $p_s = 10.05544945$ and therefore
  $$c/1-r < p_c < p_s. \qquad \square$$

Observe that situation $c/1-r < p_c < p_s$ occurs for high initial brand values: only the brand owner of a well known brand can take into account the free-riding situation of binding the licensee to sustain the brand.

Obviously the leader will adopt the strategy which leads him a greater profit. With this task, let's denote by J^{*kw}_L, with $k, w \in \{0,1\}$, the optimal profit of the licensor if he adopts the advertising strategy a^k_L and the follower adopts the advertising effort a^w_l. Analogously, let's denote by J^{*1}_{LC}, the optimal profit of the licensor while the players adopt the cooperation advertising strategies a^*_{LC} and a^*_{lC} in problem P^1_C.

A possible rational rule to obtain the players' profits in a cooperation context is the *Nash Bargaining solution,* see [3, 11], in any case it must be $J^{*1}_{LC} \leq J^{*1}_C$.

Turning back to the licensor's decision, it's a well known result that cooperation leads to greater profits, that is $J_{LC}^{*1} > J_L^{*10}$. It can be easily proved also that each player earns more when the other one takes care about the brand sustainability constraint, just because the goodwill is increased by the effect of the other player's additional advertising effort. This can be formalized as follows: $J_L^{*10} < J_L^{*01}$.

It remains to check if it turns out convenient to the leader to cooperate or to bind the licensee to care about brand sustainability. This is not possible to determine a priori, as this comparison depends on the values of the many problem's parameters. Nevertheless, once considered a particular instance of the problem, it can be easily found the licensor's best choice by comparing the optimal profits J_{LC}^{*1} and J_L^{*01} evaluated with the particular values of the parameters which characterize such an instance.

An interesting result is that the cooperative solution doesn't always lead to a greater profit for the licensor. In fact, let be $T = 30$, $c_L = 0.15$, $c_l = 0.2$, $c = 8.1$, $p_{\mathrm{R}} = 8$, $\beta = 0.5$, $\gamma_L = 0.75$, $\gamma_l = 0.7$, $r = 0.1$, $\delta = 0.1$, $\theta = 0.5$ and $p = 10$; either if $G_0 = 1{,}543$, or $G_0 = 1{,}545$, we have that $9 = c/1 - r < p < p_c < p_{\mathrm{s}}$ and therefore it makes sense to consider the free-riding situation. In the former case $J_C^{*1} - J_L^{*01} = 28.48809 > 0$, while in the latter $J_C^{*1} - J_L^{*01} = -28.50707 < 0$. Being $J_{LC}^{*1} < J_C^{*1}$, by definition, we have proved that there exists, at least, one situation in which the licensor's strategy of binding the licensee to take care about brand sustainability is, for him, more profitable than cooperating, that is $J_{LC}^{*1} < J_L^{*01}$.

19.5 Conclusions and Further Developments

We have tackled the problem of determining the optimal advertising strategies in a licensing contract, in order to maximize the profits of the two players involved in the contract and guaranteeing the brand sustainability. We considered the selling price as constant and we have analyzed different scenarios. We have found out that there exists a minimum selling price the brand owner can impose to the licensee to assure brand sustainability. In the case that the licensor either cannot bind the licensee to fix such a minimum price, or he doesn't want to do it, he can still guarantee the brand sustainability through an additive advertising. He can share with the licensee the additional advertising effort necessary the sustain the brand, he can increase his usual advertising effort, without binding the licensee to do the same, finally he can force the licensee to increase her advertising effort considering the sustainability constraint in her advertising plan, without doing himself such a strategy. We have considered the fully cooperative solution too and considered it as a benchmark to compare with the previous strategies. We have found out that in this particular problem the cooperative solution doesn't always lead to a greater profit for the licensor, as he can use the strategy of binding the licensee to do more advertising. That's why we do not consider an incentive strategy approach to obtain the cooperative solution as an equilibrium. This can be the idea for a further analysis, of course only for the cases in which the cooperative solution do effectively brings the licensor to a greater profit. Another interesting, and non trivial, approach consists in considering price as a constant decision variable to be determined using the theory of optimal processes with parameters. Such an approach has been used in [10] for an optimal control problem, but to the best of my knowledge nothing similar has ever been applied to a Stackelberg differential game. Considering the pricing problem requires to analyze the problem also from the follower's point of view; this can be done determinating the optimal selling price which takes into account production's costs and maximizes the licensee's profit.

Appendix. Proofs

Proof of Theorem 19.1

Let's determine the licensee's best response. The follower problem is

$$\begin{aligned}
\max_{a_l \geq 0} \ & J_l(a_L, a_l, p) = \int_0^T \left[((1-r)p - c)(G(t) - \theta p) - \frac{c_l}{2} a_l^2(t)\right] \mathrm{d}t, \\
\text{subject to} \ & \dot{G}(t) = \gamma_L a_L(t) + \gamma_l a_l(t) - \delta G(t) + \beta(p - p_R), \\
& G(0) = G_0, \\
& G(T) \geq G_0, \\
& a_l(t) \geq 0, \quad \forall t \in [0, T].
\end{aligned}$$

The Hamiltonian function is

$$\begin{aligned}
H_l(G, a_l, \lambda_l, t) = & \left[((1-r)p - c)(G - \theta p) - \tfrac{c_l}{2} a_l^2\right] \\
& + \lambda_l \left(\gamma_L a_L + \gamma_l a_l - \delta G + \beta\,(p - p_R)\right).
\end{aligned} \tag{19.24}$$

it's derivative w.r.t. a_l is

$$\frac{\partial H_l(G, a_l, \lambda_l, t)}{\partial a_l} = -c_l a_l + \gamma_l \lambda_l \tag{19.25}$$

and the stationary point is

$$a_l(t) = \frac{\lambda_l(t)\gamma_l}{c_l}, \quad t \in [0, T]. \tag{19.26}$$

The Hamiltonian function is concave in (G, a_l), therefore Mangasarian sufficiency theorem holds.

The co-state equation is

$$\dot{\lambda}_l(t) = -\frac{\partial H_l}{\partial G} = -((1-r)p - c) + \delta\lambda_l(t), \tag{19.27}$$

and solved it gives

$$\lambda_l(t) = \frac{(1-r)p - c}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_l(T)\,\mathrm{e}^{-\delta(T-t)}, \tag{19.28}$$

where $\lambda_l(T)$ satisfies the transversality conditions

$$\lambda_l(T) \geq 0 \quad \text{and} \quad \lambda_l(T)\,(G(T) - G_0) = 0. \tag{19.29}$$

Recalling that we assumed

$$(1-r)p \geq c, \tag{19.30}$$

it follows that the $\lambda_l(t) > 0$ for all $t \in [0,T]$, therefore the optimum advertising effort for the licensee is

$$\begin{aligned} a_l^*(t) &= \max\left\{\frac{\lambda_l(t)\gamma_l}{c_l}, 0\right\} \\ &= \frac{\gamma_l}{c_l}\left[\frac{(1-r)p - c}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_l(T)\,\mathrm{e}^{-\delta(T-t)}\right]. \end{aligned} \tag{19.31}$$

If condition (19.30) didn't hold, then it would not be convenient for the licensee to produce at all. Therefore, it would not even be convenient to advertise.

Let's substitute the follower optimal strategy into the state equation of the leader's problem

$$\dot{G}(t) = \gamma_L a_L(t) + \gamma_l a_l^*(t) - \delta G(t) + \beta\,(p - p_R)\,.$$

His Hamiltonian function is

$$\begin{aligned} H_L(G, a_L, \lambda_L, t) = &\left[rp\,(G - \theta p) - \frac{c_L}{2}a_L^2\right] \\ &+ \lambda_L\,(\gamma_L a_L + \gamma_l a_l^*(t) - \delta G + \beta\,(p - p_R))\,. \end{aligned} \tag{19.32}$$

Its derivative w.r.t. a_L is

$$\frac{\partial H_L(G^*(t), a_L, \lambda_L, t)}{\partial a_L} = -c_L a_L + \gamma_L \lambda_L \tag{19.33}$$

and the stationary point is

$$a_L(t) = \frac{\lambda_L(t)\gamma_L}{c_L}, \quad \forall t \in [0,T]. \tag{19.34}$$

Mangasarian's Theorem holds for the licensor's solution too, in fact his Hamiltonian function (19.32) is concave in (G, a_L).

The co-state equation is

$$\dot{\lambda}_L(t) = -\frac{\partial H_L}{\partial G} = -rp + \lambda_L(t)\delta; \tag{19.35}$$

solved it gives

$$\lambda_L(t) = \left(\lambda_L(T) - \frac{rp}{\delta}\right) \mathrm{e}^{-\delta(T-t)} + \frac{rp}{\delta} = \frac{rp}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_L(T)\,\mathrm{e}^{-\delta(T-t)} \geq 0, \tag{19.36}$$

where $\lambda_L(T)$ satisfies the transversality conditions

$$\lambda_L(T) \geq 0 \quad \text{and} \quad \lambda_L(T)\,(G(T) - G_0) = 0. \tag{19.37}$$

So that the optimum advertising effort for the licensor is

$$\begin{aligned} a_L^*(t) &= \max\left\{\frac{\gamma_L}{c_L}\left[\frac{rp}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_L(T)\,\mathrm{e}^{-\delta(T-t)}\right], 0\right\} \\ &= \frac{\gamma_L}{c_L}\left[\frac{rp}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_L(T)\,\mathrm{e}^{-\delta(T-t)}\right]. \end{aligned} \tag{19.38}$$

The advertising efforts $a_l^*(t)$ and $a_L^*(t)$ given in (19.31) and (19.38) with $\lambda_l(T) \geq 0$ and $\lambda_L(T) \geq 0$ such that $\lambda_l(T)(G(T) - G_0) = 0$ and $\lambda_L(T)(G(T) - G_0) = 0$ constitute a *Stackelberg equilibrium* and it can be proved that such equilibrium is time consistent as it coincides with the Markovian Nash Equilibrium.

In order to determine the values of parameters $\lambda_L(T)$ and $\lambda_l(T)$ from transversality condition, let's solve the motion equation with the initial condition

$$\begin{cases} \dot{G}(t) = \gamma_L a_L^*(t) + \gamma_l a_l^*(t) - \delta G(t) + \beta\,(p - p_R), & t \in [0,T], \\ G(0) = G_0. \end{cases} \tag{19.39}$$

The differential equation in (19.39), can be rewritten as

$$\begin{aligned} \dot{G}(t) &= \frac{\gamma_L^2}{c_L}\left[\frac{rp}{\delta}(1 - \mathrm{e}^{-\delta(T-t)}) + \lambda_L(T)\,\mathrm{e}^{-\delta(T-t)}\right] \\ &\quad + \frac{\gamma_l^2}{c_l}\left[\frac{(1-r)\,p - c}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_l(T)\,\mathrm{e}^{-\delta(T-t)}\right] - \delta G(t) + \beta\,(p - p_R) \\ &= -\delta G(t) + H\,\mathrm{e}^{-\delta(T-t)} + K, \end{aligned}$$

where

$$\begin{aligned}
\eta_L &= \lambda_L(T), \qquad \eta_l = \lambda_l(T),\\
H &= \left(\frac{\gamma_L^2}{c_L}\eta_L + \frac{\gamma_l^2}{c_l}\eta_l\right) - \left(\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)p-c}{\delta}\right),\\
K &= \frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)\,p-c}{\delta} + \beta\,(p-p_R).
\end{aligned}$$

The solution gives the brand value function

$$\begin{aligned}
G(t) &= G_0\,\mathrm{e}^{-\delta t} + \frac{\mathrm{e}^{-\delta(T-t)} - \mathrm{e}^{-\delta(T+t)}}{2\delta}H + \frac{1-\mathrm{e}^{-\delta t}}{\delta}K\\
&= G_0\,\mathrm{e}^{-\delta t} + \frac{\mathrm{e}^{-\delta(T-t)} - \mathrm{e}^{-\delta(T+t)}}{2\delta}\\
&\quad\times\left[\left(\frac{\gamma_L^2}{c_L}\eta_L + \frac{\gamma_l^2}{c_l}\eta_l\right) - s\left(\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)p-c}{\delta}\right)\right]\\
&\quad+\frac{1-\mathrm{e}^{-\delta t}}{\delta}\left[\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)\,p-c}{\delta} + \beta\,(p-p_R)\right], \qquad (19.40)
\end{aligned}$$

whose value at the final time is

$$\begin{aligned}
G(T) &= G_0\,\mathrm{e}^{-\delta T} + \frac{1-\mathrm{e}^{-2\delta T}}{2\delta}\left[\left(\frac{\gamma_L^2}{c_L}\eta_L + \frac{\gamma_l^2}{c_l}\eta_l\right) - \left(\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)p-c}{\delta}\right)\right]\\
&\quad+\frac{1-\mathrm{e}^{-\delta T}}{\delta}\left[\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)\,p-c}{\delta} + \beta\,(p-p_R)\right]. \qquad (19.41)
\end{aligned}$$

Let's observe that $G(T)$ is linear in η_L and η_l, furthermore $G(T) \geq G_0$ if and only if $(\frac{\gamma_L^2}{c_L}\eta_L + \frac{\gamma_l^2}{c_l}\eta_l) \geq \eta_0$, where

$$\eta_0 = \frac{1}{1+\mathrm{e}^{-\delta T}}\left\{2(\delta G_0 - \beta(p-p_R)) - (1-\mathrm{e}^{-\delta T})\left[\frac{\gamma_L^2}{c_L}\frac{rp}{\delta} + \frac{\gamma_l^2}{c_l}\frac{(1-r)p-c}{\delta}\right]\right\}.$$

From the transversality conditions the following situations may occur:

- $G(T) > G_0$, therefore $\eta_L = \eta_l = 0$. The advertising efforts are (19.15) and (19.16). This happens if $p > p_s$ where p_s is given in (19.14);
- $\eta_L > 0$ and $\eta_l > 0$, therefore $G(T) = G_0$. The advertising efforts are (19.17) and (19.18) with η_L and η_l such that $(\frac{\gamma_L^2}{c_L}\eta_L + \frac{\gamma_l^2}{c_l}\eta_l) = \eta_0$;

- $\eta_L > 0$, therefore $G(T) = G_0$ and $\eta_l = 0$ so that $\eta_L = \frac{c_L}{\gamma_L^2}\eta_0 \geq 0$, and the advertising efforts are (19.20) and (19.16);
- $\eta_l > 0$, therefore $G(T) = G_0$ and $\eta_L = 0$ so that $\eta_l = \frac{c_l}{\gamma_l^2}\eta_0 \geq 0$, and the advertising efforts are (19.15) and (19.21).

Proof of Theorem 19.2

The Hamiltonian function is

$$H_C(G,a_L,a_l,\lambda_C,t) = \left[(p-c)(G-\theta p) - \frac{c_L a_L^2 + c_l a_l^2}{2}\right] \\ +\lambda_C\left(\gamma_L a_L + \gamma_l a_l - \delta G + \beta\left(p - p_R\right)\right), \tag{19.42}$$

it's derivatives w.r.t. a_L and a_l are

$$\frac{\partial H_C(G,a_L,\lambda_C,t)}{\partial a_L} = -c_L a_L + \gamma_L \lambda_C, \tag{19.43}$$

$$\frac{\partial H_C(G,a_l,\lambda_C,t)}{\partial a_l} = -c_l a_l + \gamma_l \lambda_C \tag{19.44}$$

and the stationary point is

$$(a_L(t),a_l(t)) = \left(\frac{\lambda_C(t)\gamma_L}{c_L}, \frac{\lambda_C(t)\gamma_C}{c_l}\right), \quad t \in [0,T]. \tag{19.45}$$

Mangasarian's Theorem holds for the licensor's solution too, in fact his Hamiltonian function (19.42) is concave in (G,a_L,a_l).

The co-state equation is

$$\dot{\lambda}_C(t) = -\frac{\partial H_C}{\partial G} = -(p-c) + \lambda_C(t)\delta, \tag{19.46}$$

solved it gives

$$\lambda_C(t) = \frac{p-c}{\delta}\left(1 - \mathrm{e}^{-\delta(T-t)}\right) + \lambda_C(T)\,\mathrm{e}^{-\delta(T-t)}, \tag{19.47}$$

where $\lambda_C(T)$ satisfies the transversality conditions

$$\lambda_C(T) \geq 0 \quad \text{and} \quad \lambda_C(T)\left(G(T) - G_0\right) = 0. \tag{19.48}$$

Let's observe that $\lambda_C(t) \geq 0$ as $p \geq c/(1-r) \geq c$, therefore the optimal advertising efforts are

$$a^*_{LC}(t) = \frac{\gamma_L}{c_L}\left[\frac{p-c}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right)+\lambda_C(T)\,\mathrm{e}^{-\delta(T-t)}\right], \tag{19.49}$$

$$a^*_{lC}(t) = \frac{\gamma_l}{c_l}\left[\frac{(p-c}{\delta}\left(1-\mathrm{e}^{-\delta(T-t)}\right)+\lambda_C(T)\,\mathrm{e}^{-\delta(T-t)}\right]. \tag{19.50}$$

The motion equation can be rewritten as

$$\dot{G}(t) = -\delta G(t) + M\,\mathrm{e}^{-\delta(T-t)} + N,$$

where

$$\begin{aligned} \eta_C &= \lambda_C(T), \\ M &= \left(\frac{\gamma_L^2}{c_L}+\frac{\gamma_l^2}{c_l}\right)\left(\eta_C - \frac{p-c}{\delta}\right), \\ N &= \left(\frac{\gamma_L^2}{c_L}+\frac{\gamma_l^2}{c_l}\right)\frac{p-c}{\delta} + \beta\,(p-p_R)\,. \end{aligned}$$

It's solution, with the initial condition $G(0) = G_0$ is

$$G(t) = G_0\,\mathrm{e}^{-\delta t} + \frac{\mathrm{e}^{-\delta(T-t)} - \mathrm{e}^{-\delta(T+t)}}{2\delta}M + \frac{1-\mathrm{e}^{-\delta t}}{\delta}N, \tag{19.51}$$

whose value at the final time is

$$G(T) = G_0\,\mathrm{e}^{-\delta T} + \frac{1-\mathrm{e}^{-2\delta T}}{2\delta}M + \frac{1-\mathrm{e}^{-\delta T}}{\delta}N.$$

From the transversality condition we obtain

$$\eta_C = \max\left\{\frac{2(\delta G_0 - \beta(p-p_{\mathrm{R}}))}{\left(\frac{\gamma_L^2}{c_L}+\frac{\gamma_l^2}{c_l}\right)(1+\mathrm{e}^{-\delta T})} - \frac{p-c}{\delta}\frac{1-\mathrm{e}^{-\delta T}}{1+\mathrm{e}^{-\delta T}}, 0\right\}.$$

References

1. Aaker D.A.: Managing Brand Equity. The Free Press, New York (1991)
2. Amaldoss, W., Jain., S.: Pricing of conspicuous goods: a competitive analysis of social effects. J. Marketing Res. **42**(1), 30–42 (2005)

3. Binmore, K., Osborne, M.L., Rubinstein, A.: Noncooperative models of bargaining. In: Aumann, R.J., Hart, S. (eds.) Handbook of Game Theory with Economic Applications. North-Holland, Amsterdam (1992)
4. Buratto, A., Zaccour, G.: Coordination of advertising strategies in a fashion licensing contract. J. Opt. Theory Appl. **142**(1), 31–53 (2009)
5. Caulkins, J.P., Feichtinger, G., Kort P., Hartl, R.F.: Brand image and brand dilution in the fashion industry. Automatica 42, 1363–1370 (2006)
6. Caulkins, J.P., Hartl, R.F., Feichtinger, G.: Explaining fashion cycles: imitators chasing innovators in product space. J. Econ. Dyn. Control 31, 1535–1556 (2007)
7. Chintagunta, P., Sigué, J.P.: Advertising strategies in a franchise system. Eur. J. Oper. Res. 198, 655-665 (2009)
8. Dockner, E.J., Jørgensen, S., Van Long, N., Sorger, G.: Differential Games in Economics and Management Science. Cambridge University Press, Cambridge (2000)
9. Jørgensen, S., Di Liddo, A.: Design imitation in the fashion industry. In: Annals of the International Society of Dynamic Games, vol. 9, pp. 569–586. Birkhauser, Boston (2007)
10. Jørgensen, S., Kort P., Zaccour, G.: Optimal pricing and advertising policies for an entertainment event. J. Econ. Dyn. Control 33, 583–596 (2009)
11. Jørgensen, S., Zaccour, G.: Differential Games in Marketing. Kluwer Academic, Boston (2004)
12. Keller, K.L.: Strategic Brand Management: Building, Measuring, and Managing Brand Equity, 3rd ed. Prentice-Hall, New York (2007)
13. Leibenstein, H., Bandwagon, Snob, Veblen: Effects in the Theory of Consumers' Demand. Quart. J. Econ. 64(2), 183–207 (1950)
14. Martín-Herrán, G., Sigué, S.P., Zaccour, G.: Strategic interactions in traditional franchise systems: Are franchisors always better off? Eur. J. Oper. Res. **213**(3), 526–537 (2011)
15. Martín-Herrán, G., Taboubi, S., Zaccour, G.: On yopia in a dynamic marketing channel. G-2006-37, GERAD (2006)
16. Nerlove, M., Arrow, K.J.: Optimal advertising policy under dynamic conditions. Economica **39**(114), 129–142 (1962)
17. Power, D., Hauge, A.: No man's brand – brands, institutions, and fashion. Growth Change **39**(1), 123–143 (2008)
18. Raugust, K.: The Licensing Business Handbook, 8th Ed. EPM Communications, New York (2012)
19. White, E.P.: Licensing. A Strategy for Profits. KEW Licensing Press, Chapel Hill (1990)

Chapter 20
Cost–Revenue Sharing in a Closed-Loop Supply Chain

Pietro De Giovanni and Georges Zaccour

Abstract We consider a closed-loop supply chain (CLSC) with a single manufacturer and a single retailer. We characterize and compare the feedback equilibrium results in two scenarios. In the first scenario, the manufacturer invests in green activities to increase the product-return rate while the retailer controls the price. A Nash equilibrium is sought. In the second scenario, the players implement a cost–revenue sharing (CRS) contract in which the manufacturer transfers part of its sales revenues, and the retailer pays part of the cost of the manufacturer's green activities program that aims at increasing the return rate of used products. A feedback-Stackelberg equilibrium is adopted, with the retailer acting as the leader. Our results show that a CRS is successful only under particular conditions. While the retailer is always willing to implement such a contract, the manufacturer is better off only when the product return and the remanufacturing efficiency are sufficiently large, and the sharing parameter is not too high.

Keywords Closed-loop supply chain • Differential game • Product return • Green activities • Support program • Incentive • Contract

P. De Giovanni (✉)
Department of Information, Logistics and Innovation, VU University Amsterdam, de Boelelaan 1105, 3A-31, 1081 HV Amsterdam, The Netherlands
e-mail: pietro.degiovanni@vu.nl

G. Zaccour
GERAD, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada
e-mail: georges.zaccour@gerad.ca

P. Cardaliaguet and R. Cressman (eds.), *Advances in Dynamic Games*, Annals of the International Society of Dynamic Games 12, DOI 10.1007/978-0-8176-8355-9_20,

20.1 Introduction

A closed-loop supply chain (CLSC) combines forward and reverse supply-chain activities into a single system, to improve the environmental performance [38], create new economic opportunities, and provide competitive advantages to participants [15]. Forward activities include new product development, product design and engineering, procurement and production, marketing, sales, distribution, and after-sale service [45]. Reverse activities refer to all those needed to close the loop, such as product acquisition, reverse logistics, points of use and disposal, testing, sorting, refurbishing, recovery, recycling, re-marketing, and re-selling [16,25] . The integration of these activities makes it possible to recover of residual value from used products, thus reducing the amount of resources needed for production while also conserving landfill space and reducing air pollution [3].

Firms are interested in closing the loop when producing with used components is less costly than manufacturing with new materials [43]. Several empirical and case studies (see, e.g., [18, 45]) have already highlighted the relevance of CLSC for business and government, while the comprehensive reviews provided by Fleischmann et al. [17], Dekker et al. [11] and Atasu et al. [2] report on what has so far been achieved, and on the issues that still need to be addressed.

Because it has the most to gain from a CLSC, it is the manufacturer (or remanufacturer) that closes the loop, by managing, often exclusively, the product-return process. Guide [23] reports that 82 % of the sample firms collect directly from customers. However, this percentage does not always need to be that high, and other members of the supply chain can also play a significant role. Savaskan et al. [43] characterize four configurations of a channel in which a manufacturer, a retailer, or an OEM (original equipment manufacturer) does the collection, and they demonstrate that all supply chain members can be involved in the return process as long as the manufacturer provides the right incentives. Bhattacharya et al. [6] model a three-player game where a remanufacturer always does the collecting. Depending on the type of contract, several coordination mechanisms may align the players' goals.

In a general CLSC configuration, the manufacturer invests in green-activity programs for the main purpose of increasing product returns [23]. These activities may include marketing expenditures (e.g., advertising) to increase the customers' knowledge about the return policy [30] as well as operational activities, such as collection, inspection, reprocessing, disposal, and redistribution [38], that aim to increase remanufacturing efficiency and create a suitable logistics network. The integration of these activities leads to a new strategy for product returns, that emphasizes not only operational aspects but also social and environmental issues, through a suitable asset-recovery policy [27]. This is the strategy currently being used by some firms, such as Kodak, which advertises the return of single-use cameras to create customer knowledge about the resulting environmental and social benefits (Kodak.com). The manufacturer may decide either to manage all green activities exclusively and reap all of the cost saving, or to involve other players and share the economic benefits [43]. In the latter case, the manufacturer should design an adequate contract, provide attractive incentives for collaborating in closing the loop, and properly share the economic advantages of remanufacturing [8, 13].

This paper contributes to this research area by developing a dynamic CLSC game where a cost-sharing program for green activities is introduced along with a reverse-revenue-sharing contract (RRSC). As reported by Geng and Mallik [19], an RRSC is a good option when the upstream player wants to involve the downstream player in a specific activity. For instance, Savaskan et al. [43] show that, when a retailer is involved in the product-return process, the CLSC performs better. We confine our interest to a single-manufacturer-single-retailer case and characterize and contrast the equilibrium strategies and outcomes in two scenarios. In the first scenario, referred to as Benchmark scenario, the two firms choose non-cooperatively and simultaneously their strategies. In the second scenario, referred to as *CRS*, the players share the manufacturer's sales revenues and the cost of the green activities.[1] In both cases, the manufacturer controls the rate of green activities and the retailer controls the price. By contrasting the results of the two scenarios, we will be able to assess the impact of implementing an active approach to increasing consumers' environmental awareness, and, by the same token, the return rate of used products. When the retailer contributes to the manufacturer's activities, the game is played la Stackelberg. This game structure is common in the literature on marketing channels (see, e.g., the books [31, 35] operations (e.g., [29]), as well as environmental management (e.g., [43]).

There is a growing game-theoretic literature that deals with CLSCs, see, e.g., [1, 3, 10, 14, 21, 24, 39, 43]. While these contributions investigate CLSCs in static or two-period games, here we seek to evaluate the CLSC in a dynamic setting. Guide et al. [27] emphasize the importance of time in managing product returns, which are subject to time-value decay. Ray et al. [42] evaluate profits and pricing policy under time-dependent and time-independent scenarios—namely, age-dependent and age-independent differentiation—and show that the attractiveness of remanufacturing changes substantially. Finally, Savaskan et al. [43] advise researchers that the CLSC should be investigated as a dynamic phenomenon, as the influence of dynamic returns changes channel decisions. Our paper takes up this challenge and proposes a differential game to analyze equilibrium returns and pricing strategies in the two scenarios described above.

Our main results can be summarized as follows:

A1. A CRS alleviates the double-marginalization problem in the supply chain. The consumer pays a lower retail price and demands more product.

A2. The investment in green activities and the return rate of used products are higher in the CRS scenario than in the benchmark game. The environment also benefits from the implementation of a CRS contract.

---

[1] What we have in mind here is similar to cooperative advertising programs, where typically, a manufacturer pays part of the cost of promotion and advertising activities conducted locally by its retailers. Cooperative advertising programs have been studied in the marketing literature, in a static setting (e.g., [4, 5, 36]), as well as in a dynamic context (e.g., [32–34]).

A3. The retailer always prefers the CRS scenario to the benchmark scenario. The manufacturer does the same, under certain conditions involving the revenue-sharing parameter, the return rate, and the level of cost reduction attributable to remanufacturing. The conclusion is that a CRS is not always Pareto improving.

The paper is organized as follows. In Sect. 20.2 we state the differential game model and in Sect. 20.3 we characterize the equilibria in the two scenarios. Section 20.4 compares strategies and outcomes. Section 20.5 briefly concludes.

20.2 Model and Scenarios

Consider a supply chain made up of one manufacturer, player M, and one retailer, player R. Let time t be continuous and assume an infinite planning horizon.[2] The manufacturer can produce its single good using new materials or old materials extracted from returned past-sold products. This second option is common practice in many industries.[3] Managing returns effectively represents one key edge that is required for CLSC to succeed. Guide et al. [24] presents two streams of practices for managing returns: a passive and an active approach. The passive approach to returns (waste-stream approach) consists of waiting and hoping that the customers return their products. An active (market-driven) approach instead implies that CLSC participants are continuously involved in controlling and influencing the return rate by setting appropriate strategies. An active approach makes it possible to manage the forward activities as a further source of economic benefits [24].

The literature in CLSC can be divided into three streams in terms of modeling the return rate of used products. The first stream adopted a passive approach and assumed the return rate to be exogenous (see, e.g., [3, 14, 16, 21, 27]). The second stream also adopted a passive approach, but modelled the return rate as a random variable, e.g., an independent Poisson (see, e.g., [1]). The third group of studies considered an active approach, with the return rate being a function of a player's strategy (see, e.g., [43, 44]). We follow this last view. More precisely, we suppose that the manufacturer can increase the rate of return for previously sold products by investing in a "green" activity program (GAP). Examples of such activities include advertising and communications campaigns about the firm's recycling policies,

[2] The assumption of infinite-planning horizon is mainly for tractability.

[3] To illustrate, the remanufacturing sector in the US has reached over \$53 billion in sales, and includes over 70,000 firms and 480,000 employees [28]. Large retailers can have return rates in excess of 10 % of sales, and manufacturers, such as Hewlett-Packard, report product returns that exceed 2 % of total outbound sales [3]. Xerox practices asset recovery and remanufacturing for its photocopiers and toner cartridges in the US and abroad; it estimates its total cost-savings at over \$20 million per year [40]. The company saves 40–65 % in manufacturing costs through the re-use of parts and materials [22]. Kodak collects almost 60 % of the single-use cameras it sells worldwide and satisfies 58 % of its demand with cameras containing re-used components and almost 80 % of the materials may be re-used [21, 26].

logistics services, monetary and symbolic incentives, employees-training programs, etc. Denote by $A(t)$ the level of these activities at time t, $t \in [0,\infty)$, and assume that their cost can be well approximated by the increasing convex function

$$C(A(t)) = \frac{u_M (A(t))^2}{2}, \tag{20.1}$$

where $u_M > 0$ is a scaling parameter.[4] We suppose that the return rate $r(t)$, depends on the whole history, and not only on the current level of green activities. A common hypothesis in such a context is to assume that $r(t)$ corresponds to a continuous and weighted average of past green activities with an exponentially decaying weighting function. This assumption is intuitive because the return rate is related to environmental awareness, which is a "mental state" that consumers acquire over time, not overnight. This process is captured by defining $r(t)$ as a state variable whose evolution is governed by the linear-differential equation

$$\dot{r}(t) = A(t) - \delta r(t), \quad r(0) = r_0 \geq 0, \tag{20.2}$$

where $\delta > 0$ is the decay rate and r_0 is the initial rate of return.

The main economic benefit of the CLSC for the manufacturer is given by the saved cost (see, e.g., [39]). Following [43], we adopt the unit-production cost function:

$$C(r(t)) = c_n(1 - r(t)) + c_u r(t), \tag{20.3}$$

where $c_n > 0$ is the cost of producing one unit with new raw materials, and $c_u > 0$ is the cost to produce one unit with used material from returned products, with $c_u < c_n$. The above equation can be rewritten as

$$C(r(t)) = c_n - (c_n - c_u)\, r(t),$$

and, therefore, the difference $c_n - c_u$ is the marginal remanufacturing efficiency (cost saving) of returned products. The manufacturer incurs the highest unit cost c_n when $r(t) = 0$, and the lowest unit cost c_u is achieved when all previously purchased products are returned, i.e., for $r(t) = 1$. In (20.2)–(20.3), we implicitly assume that products may be returned independently of their condition, and that a good can be remanufactured an infinite number of times. In practice, this clearly does not hold true. For instance, Kodak's camera frame, metering system, and flash circuit are designed to be used up to six times [37] and any additional use compromises the product's reliability. Therefore, our functional forms in (20.2)–(20.3) are meant to be rough approximations of return dynamics and cost savings. In the conclusion we discuss some (necessarily much more complicated) avenues that are worth exploring in future investigations.

---

[4]Think of $A(t)$ as a composite index of the green activities.

Denote by $p(t)$ the retail price controlled by the retailer. We suppose that the demand for the manufacturer's product is given by

$$D(p(t)) = \alpha - \beta p(t), \tag{20.4}$$

where $\alpha > 0$ is the market potential and $\beta > 0$ represents the marginal effect of pricing on current sales. To have nonnegative demand, we assume that $p(t) \leq \alpha/\beta$. Two comments are in order regarding this demand function. First, following a long tradition in economics, we have chosen a linear form. In addition to being tractable, this choice is typically justified by the fact that such a demand form is derivable from the consumer-utility function. Second, we follow [43] and suppose that $D(\cdot)$ is independent of the return rate. Put differently, we are assuming here that the CLSC's main purpose is as a cost-saving rather than a demand-enhancing mechanism. Denote by ω the constant wholesale price charged by the manufacturer, with $c_n < \omega < p(t) \leq \alpha/\beta$. The lower bound ensures that the manufacturer's margin is positive even when there is no recycling. The second inequality ensures that the retailer's margin is nonnegative.

Up to now, our formulation states that the manufacturer is taking care of the CLSC's operational features, and that the marketing decisions (represented by pricing) are left to the retailer, who is not at all involved in recycling. Although the players follow an individual profit-maximization objective, they still may attempt to link their activities to achieve higher economic benefits for both of them. For instance, IBM and Xerox coordinate their recovery activities with their suppliers in order to increase their profitability [18, 24]. IBM gives the responsibility for managing all product returns worldwide to a dedicated business unit called Global Asset Recovery Services, that collects, inspects, and assigns a particular recovery option (resale or remanufacturing), and that maximizes the chain's efficiency by coordinating its activities with IBM refurbishment centres worldwide [18].

Here we explore a setting where: (a) the retailer financially supports the manufacturer's GAP; and, (b) the manufacturer designs an incentive mechanism to compensate the retailer for this participation, and to better coordinate the CLSC. Denote by $B(t), 0 \leq B(t) \leq 1$, the support rate, to be chosen by the retailer, in the total cost of the GAP. Consequently, the retailer pays $B(t)C(A(t))$ and the manufacturer contributes the remaining portion, i.e., $(1 - B(t))C(A(t))$. The rationale for the retailer to participate in the manufacturer's GAP is the premise that the combined efforts of the two players would lead to a higher return rate for used products, and consequently, to a lower production cost and wholesale price.

Denote by $I(r(t))$ the state-dependent incentive provided by the manufacturer to the retailer. The incentive assumes the traditional form as presented by [7, 9, 20], where the manufacturer transfers a share of his revenues to the retailer in order to modify her strategies. This way of modeling the incentive differs from the traditional scheme elaborated in the literature of CLSC. Typically, the incentive schemes assume the form of payment, where the manufacturer pays a certain per-unit amount when another player returns a product [43, 44]. Alternatively, rebates on new sales can also coordinate a CLSC [21]. Other valid alternative contract schemes link the incentive to some operational features. For instance, Guide et al. [24] characterized a quality-dependent price incentive for used products; Guide et al. [27] suggest the integration between return management with inventory management (VMI) and resource management (employees). Our way of modeling the incentive is analogous to the incentive modeled by ReCellular Inc. and presented by [24], where the manufacturer offers a two part incentive that is formed of a fix (direct) per unit component as well as of a variable (indirect) component that depends of the operational (collecting) costs.

Similarly, our incentive consists of a share of the manufacturer's revenues that is transferred to the retailer and that is formed of a fix state-independent part, as well as of a variable state-dependent component. In this sense, instead of focusing only on its main strengths—reduction of the double-marginalization effect, lower price and higher demand (see, e.g., [7, 9, 20]—a two-parameter contract implemented in a CLSC enhances collaboration in product return management. In [7, 9, 20] the incentive depends only on the sharing parameters, wholesale price, and production cost, while in our model it also is a function the remanufacturing cost and the return rate.

Assuming profit-maximization behavior, the players' objective functionals are then given by

$$J_M = \int_0^{\infty} \mathrm{e}^{-\rho t} \left\{ (\alpha - \beta p(t))(\omega - C(r(t)) - I(r(t))) - \frac{u_M}{2}(1 - B(t))A(t)^2 \right\} \mathrm{d}t, \tag{20.5}$$

$$J_R = \int_0^{\infty} \mathrm{e}^{-\rho t} \left\{ (\alpha - \beta p(t))(p(t) - \omega + I(r(t))) - \frac{u_M}{2} B(t)A(t)^2 \right\} \mathrm{d}t, \tag{20.6}$$

where ρ is the common discount rate.

In summary, by (20.2), (20.5), and (20.6) we have defined a two-player differential game with three control variables $A(t) \geq 0$, $p(t) \geq 0$, and $0 \leq B(t) \leq 1$, and one state variable $r(t)$, with $0 \leq r(t) \leq 1$.

20.2.1 The Scenarios

We shall characterize and compare equilibrium strategies and outcomes for two scenarios. In both of them, the assumption is that the players use Markovian strategies, i.e., strategies that are functions of the state variable. Further, we restrict ourselves to stationary strategies, that is, strategies that only depend on the current value of the state variable, and not explicitly on time.

Benchmark Scenario: The retailer does not participate in the green activities program of the manufacturer, and the latter does not offer any incentive to coordinate the CLSC, i.e., $B(t) \equiv 0$ and $I(r(t)) \equiv 0, \forall t \in [0, \infty)$. The game is played noncooperatively and a feedback-Nash equilibrium is sought. Equilibrium strategies and outcomes will be superscripted with N (for *N*ash).

Cost-Revenue Sharing Scenario: We assume that the retailer is the leader and announces its support rate for the green activities conducted by the manufacturer, who acts as the follower. The right (subgame-perfect) equilibrium concept in such a setting is the feedback-Stackelberg equilibrium. Equilibrium strategies and outcomes will be superscripted with S (for *S*haring or *S*tackelberg). Denote by ϕ the percentage of revenues transferred from the manufacturer to the retailer to stimulate green investments and coordinate the CLSC. Under CRS, we have

$$I^S(r) = \phi\left[\omega - C(r(t))\right] = \phi(\omega - c_n) + \phi(c_n - c_u)\, r(t). \tag{20.7}$$

As a consequence of this transfer, the manufacturer's margin ($m_M^S(r)$) and the retailer's margin ($m_R^S(r)$) become

$$\begin{aligned} m_M^S(r) &= (1-\phi)(\omega - C(r(t))), \\ m_R^S(r) &= p(t) - \omega(1-\phi) - \phi C(r(t)). \end{aligned}$$

The incentive scheme in (20.7) is made of two parts, with one being independent of the return rate ($\phi(\omega - c_n)$), and the other being a positive and increasing function in the return rate ($\phi(c_n - c_u)\, r(t)$). This shows that the retailer has a vested interest in contributing to a higher return rate.

From now on, we will omit the time argument when no ambiguity may arise.

20.3 Equilibria

In the following two subsections, we characterize the equilibria in the two scenarios described above.

20.3.1 Benchmark Equilibrium

Recall that in this scenario, the players choose, simultaneously, and independently their strategies to maximize their individual profits, with $B(t) \equiv 0$ and $I(r(t)) \equiv 0, \forall t \in [0, \infty)$.

Proposition 20.1. *The equilibrium GAP and price strategies are given by*

$$A^N = \frac{(c_n - c_u)(\alpha - \beta\omega)}{2u_M(\rho + \delta)} > 0, \tag{20.8}$$

$$p^N = \frac{\alpha + \beta\omega}{2\beta} > 0. \tag{20.9}$$

The value functions of the two players are given by

$$V_M^N(r) = \frac{(\alpha - \beta\omega)}{2}\left(\frac{(c_n - c_u)}{(\rho + \delta)} r + \frac{(\omega - c_n)}{\rho} + \frac{(c_n - c_u)^2 (\alpha - \beta\omega)}{4u_M (\rho + \delta)^2 \rho}\right), \tag{20.10}$$

$$V_R^N(r) = \frac{(\alpha - \beta\omega)^2}{4\beta\rho}. \tag{20.11}$$

Proof. See Appendix. □

The strategies in (20.8) and (20.9) are constant, i.e., independent of the state variable (return rate). This is a by-product of the linear-state structure of the differential game played in this scenario. Further, as the retailer's objective functional is independent of the state, its pricing strategy is obtained by optimizing the short-term (or current) profit. Consequently, its value function is state-independent and is given as a discounted stream of constant profits, i.e.,

$$V_R^N(r) \equiv V_R^N = \int_0^\infty \mathrm{e}^{-\rho t} \left(\frac{(\alpha - \beta\omega)^2}{4\beta}\right) \mathrm{d}t,$$

where $\frac{(\alpha-\beta\omega)^2}{4\beta}$ is the instantaneous payoff. The manufacturer's value function is increasing in the return rate. Indeed,

$$\frac{\partial V_M^N}{\partial r} = \frac{(\alpha - \beta\omega)(c_n - c_u)}{2(\rho + \delta)} > 0.$$

This result provides the rationale for the next scenario. Indeed, as it is in the best interest of the manufacturer to increase the level of used-product returns, it is tempting to provide an incentive to the retailer to induce a greater contribution to the green-activity program. It remains to be seen under which conditions this incentive is profitable for the manufacturer.

Substituting for green expenditures in the state dynamics (20.2) and solving gives the following trajectory for the return rate:

$$r^N(t) = \frac{\left(1 - \mathrm{e}^{-\delta t}\right)}{\delta} A^N + \mathrm{e}^{-\delta t} r_0 > 0.$$

The steady-state value is strictly positive and given by

$$r_\infty^N = \frac{A^N}{\delta} = \frac{(c_n - c_u)(\alpha - \beta\omega)}{2u_M (\rho + \delta)\delta} > 0. \tag{20.12}$$

From now on we assume (and check for in the numerical simulations) that the parameters are such that $r^N(t) \le 1, \forall t \in [0, \infty)$.

20.3.2 Cost-Revenue Sharing Equilibrium

The RRSC was recently introduced by [19] to coordinate a supply chain in which the upstream players have an economic incentive to coordinate. The traditional revenue sharing contract (RSC) fits with the implementation of a coordination strategy that is mainly driven by the retailer [7, 12]. The RSC, in fact, mitigates the double-marginalization effect and creates efficiency along the chain. While the retailer transfers a share of its revenues to the manufacturer, it also buys at a lower wholesale price. Consequently, price decreases and demand increases.

In the RRSC, however, the retailer receives a share of the manufacturer's net revenues. The manufacturer wishes to influence the retailer's strategies by offering an attractive economic incentive. This type of contract fits adequately with the CLSC's targets where the manufacturer has the highest incentive to close the loop. Further, in the marketing literature dealing with cooperative advertising programs, the context is one of a manufacturer helping his retailer by paying part of the cost of the local advertising or promotional efforts conducted by the retailer. Here, the situation is reversed and it is the retailer who is contributing to the manufacturer's GAP. Therefore, the retailer plays the role of leader and the manufacturer, the role of follower. The following proposition characterizes the equilibrium strategies.

Proposition 20.2. *Assuming an interior solution, the feedback-Stackelberg equilibrium price, the green activities, and the participation rate are given by*

$$p^S = \frac{\alpha + \beta\left(\omega\left(1-\phi\right) + c_n\left(1-r\right)\phi + c_u\phi r\right)}{2\beta}, \tag{20.13}$$

$$A^S = \frac{\left(2\mu_1 + \varphi_1\right)r + 2\mu_2 + \varphi_2}{2u_M}, \tag{20.14}$$

$$B^S = \frac{\left(2\mu_1 - \varphi_1\right)r + 2\mu_2 - \varphi_2}{\left(2\mu_1 + \varphi_1\right)r + 2\mu_2 + \varphi_2}, \tag{20.15}$$

where φ_1, φ_2, μ_1 and μ_2 are the coefficients of the quadratic value functions

$$V_M^S(r) = \frac{\varphi_1}{2}r^2 + \varphi_2 r + \varphi_3,$$

$$V_R^S(r) = \frac{\mu_1}{2}r^2 + \mu_2 r + \mu_3,$$

determined numerically later on.

Proof. See Appendix. □

The retail price, the investment in green activities, and the retailer's support rate are linear in the return rate. This result is expected in view of the linear-quadratic structure of the differential game. The results in the above proposition are obtained under the assumption of an interior solution. It is easy to see that for $0 \leq r \leq 1$, the price is strictly positive and decreasing in the return rate. As the coefficients of the value functions cannot be obtained analytically (the six Ricatti equations are highly coupled—see Appendix), we shall verify in the numerical simulations that the GAP strategy is nonnegative and that the support rate B^S and the return rate r are between 0 and 1. Unlike with the previous scenario, the strategies are now state-dependent, with the price being a decreasing function of the return rate. This is intuitive because a higher rate leads to a lower production cost. Further, the higher the percentage ϕ of revenues transferred from M to R, the lower the retail price. Indeed, we have $\frac{\partial p^S}{\partial \phi} = \frac{-\omega-(c_n-c_u)r}{2} < 0$. Therefore, as in the literature on revenue-sharing contracts (see, e.g., [7]), this parameter also lessens the double-marginalization problem in RRSCs.

20.4 Numerical Results

In the S scenario, the coefficients of the value functions cannot be obtained analytically; therefore, we proceed with numerical simulations to answer our research questions. This section is divided into two parts. In the first, we conduct a sensitivity analysis on the steady-state values of the control variables (price, green activities, and retailer's support) and on the return rate of used products. In the second subsection, we compare the strategies and the players' payoffs obtained in the two scenarios.

As a base case, we adopt the following values for the different parameters:

Demand parameters:	$\alpha = 1, \beta = 1,$
Cost parameters:	$c_n = 0.5, c_u = 0.1, u_M = 1,$
Contract parameters:	$\omega = 0.7, \phi = 0.4,$
Dynamic parameters:	$\rho = 0.2, \delta = 0.3.$

20.4.1 Sensitivity Analysis

Table 20.1 provides the results of a sensitivity analysis of the strategies and the state variable with respect to the main model's parameters. A positive (negative) sign indicates that the value taken by a variable increases (decreases) when we increase the value of the parameter. A "0" indicates that the variable is independent of that parameter, and n.a. means not applicable. The reported results for the benchmark game are analytical and hold true for all admissible parameter values, not only for those shown in the table. In the S scenario, when we vary the value of a parameter, the values of all other parameters remain at their base-case levels. Note that the selected parameters' values satisfy nonnegativity conditions for price, green

Table 20.1 Sensitivity analysis

	p_∞		B_∞	A_∞		r_∞	
	N	S	S	N	S	N	S
α	+	+	+	+	+	+	+
β	−	−	−	−	−	−	−
u_M	0	+	+	−	−	−	−
c_u	0	+	+	−	−	−	−
c_n	0	−	−	+	+	+	+
ω	+	+	+	−	+	−	+
ϕ	n.a.	−	+	n.a.	+	n.a.	+

activities, and demand in both scenarios. They also satisfy the requirement that the support rate and the return rate be bounded between zero and one. The results allow for the following intuitive comments:

A1. Varying α and β yields the same qualitative impact in both scenarios for all variables. Regarding the effect on the support rate provided by the retailer to the manufacturer's GAP, we obtain that a larger demand (through a higher market potential α or a lower consumer-price sensitivity β) induces the retailer to increase its support.

A2. A higher u_M means an upward shift in the cost of the green-activity program. Consequently, the manufacturer reduces its effort. Although the retailer increases its support rate to compensate for the manufacturer's lower effort, the final outcome is a lower return rate in the steady state. In short, the higher the cost of green activities, the lower the environmental and economic performance of the CLSC. The same qualitative results are obtained when the remanufacturing cost, c_u, is increased. Under such circumstances, the manufacturer is less interested in closing the loop since the savings from producing with used parts are lower.

A3. The higher the production cost, c_n, the higher the interest of the manufacturer in introducing used parts into production. Hence, the positive relationship between c_n and the investment in green activities. Consequently the return rate is increasing in c_n. The retailer benefits from the cost reduction and reduces the retail price, which in turn, feeds the demand and the returns of used products. A high c_n is therefore an incentive to implement an environmental policy. In the N scenario, the price is constant, as the production cost does not influence the retailer's strategy. In the S scenario, the support rate decreases in c_n. The economic incentive decreases with the production cost; and thus, the retailer's willingness to implement a coop program decreases accordingly.

A4. A higher wholesale price leads to a higher retail price and a lower demand. In turn, the pool of used products is smaller and green activities become less attractive. Consequently, the rate of return decreases.

A5. To interpret the results regarding ϕ, the revenue-sharing parameter in the S scenario, we recall the margins of the two players:

$$m_M^S(r) = (1-\phi)(\omega - C(r(t))),$$

$$m_R^S(r) = p(t) + \phi(\omega - C(r(t))).$$

Therefore, a higher ϕ means a higher margin for the retailer and a lower one for the manufacturer. This incentive is achieving its goal, that is, the retailer increases its support with ϕ, and consequently the manufacturer invests more in GAP, which leads to a higher a return rate.

A6. A higher decay rate leads to lower investments in GAP, and consequently, to a lower return rate. Also, increasing ρ, which amounts to giving more weight to short-term profits, leads to a lower investment in GAP.

20.4.2 Comparison

We turn to the analysis of the players' strategies and outcomes. As most of the comparisons need to be carried out numerically, we have to limit the number of parameters that we will let vary. It seems quite reasonable to focus on the most important parameters in our model, namely, the incentive parameter ϕ and the reduction in marginal cost due to manufacturing with used parts, i.e., $c_n - c_u$. All other parameter values are kept at their benchmark levels.[5]

Retail-Price Strategies: Recall that the Nash and Stackelberg equilibrium prices are given by

$$p^N = \frac{\alpha + \beta\omega}{2\beta}, \quad p^S = \frac{\alpha + \beta\left(\omega\left(1-\phi\right) + c_n\left(1-r\right)\phi + c_u r\phi\right)}{2\beta}.$$

Without resorting to numerical simulations, we can make two observations: First, the Stackelberg price is decreasing in the return rate ($\frac{\partial p^S}{\partial r} = -\frac{(c_n - c_u)\phi}{2}$); and, second, we have $p^S < p^N$, for all parameter values. To see this, we note that the two equilibrium prices are related as follows:

$$p^S\left(r\right) = p^N - \frac{I^S(r)}{2\beta}.$$

By the nonnegativity of the incentive $I^S(r)$, we then have $p^S < p^N$. Similarly to [24] and [10], the higher the remanufacturing efficiency, then the higher the incentive provided by the manufacturer, and the lower the retail price, and consequently, the higher the demand. Therefore, implementing a CRS contract alleviates the double-marginalization problem and is beneficial to the consumer.

GAP Strategies: The investments in green activities in the two scenarios are given by

$$A^N = \frac{\left(c_n - c_u\right)\left(\alpha - \beta\omega\right)}{2u_M\beta\left(\rho+\delta\right)}, \quad A^S = \frac{\left(2\mu_1 + \varphi_1\right)r + 2\mu_2 + \varphi_2}{2u_M}.$$

[5] We ran other simulations without noticing any significant qualitative changes in the results.

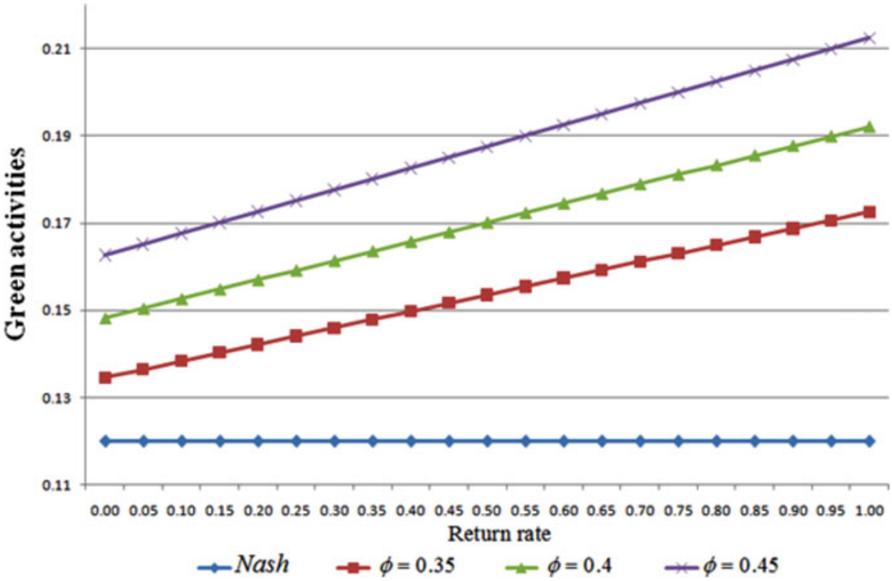


Fig. 20.1 Green-activity strategy for different values of cost savings

As for the retail price, the GAP strategy is constant in the Nash scenario, and it is linear in the return rate in the S scenario. Figures 20.1 and 20.2 display the green-activity strategy for different values of $c_n - c_u$ and ϕ. The following observations can be made: First, the higher the return rate, the higher the manufacturer's investment in GAP. This result is partly opposed to that of [3]; they advise OEMs not to invest into increasing the return rate if it is already high, but to focus on other activities, e.g., the collection system's efficiency. This would, in spirit, include the manufacturer's GAPs, which focus on the marketing and operational aspects of the return process. One interpretation of our result is that, when a CLSC achieves a high return rate, GAP investments are also required to keep up in terms of operations (e.g., logistics network, remanufacturing process, quality-control activities) and marketing (informing, promoting and advertising to a larger customer base). This is in line with [21] who suggest decreasing the investment in remanufacturing activities (e.g., product durability) when the return rate is low. Second, for any given return rate, the manufacturer invests more in GAP in the S scenario than in the Nash equilibrium. This result has also been reached in the cooperative-advertising literature cited previously. The fact that the manufacturer is sharing the cost of the green activities is in itself an incentive to do more. The implication is that the steady-state value of the return rate in the S scenario is higher than its Nash counterpart. Therefore, from an environmental point of view, as in [43], coordination in a CLSC is preferable to the benchmark game. Third, for any admissible value of r, shifting up ϕ leads to a higher investment in green activities.

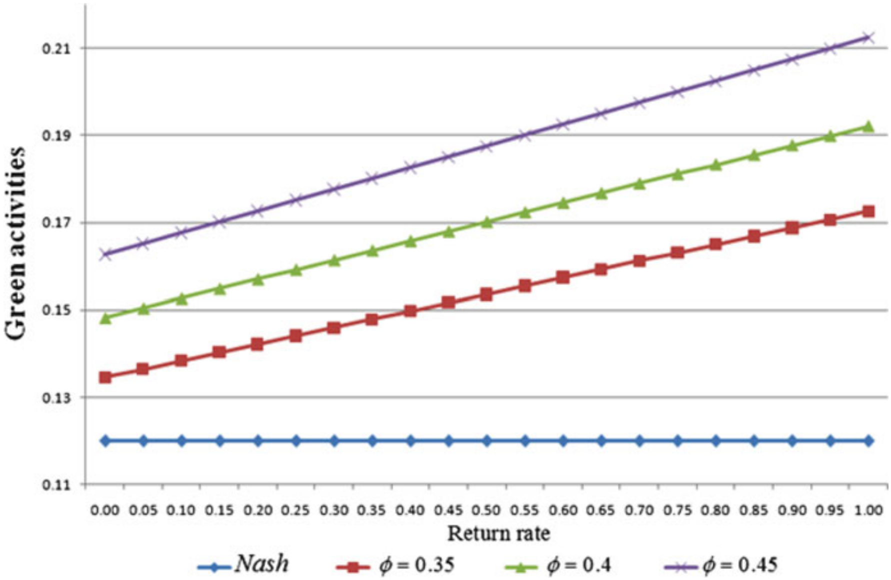


Fig. 20.2 Green-activity strategy for different values of the sharing parameter

Support-Rate Strategies: The support rate in the S game is given by

$$B^S = \frac{(2\mu_1 - \varphi_1)\,r + 2\mu_2 - \varphi_2}{(2\mu_1 + \varphi_1)\,r + 2\mu_2 + \varphi_2}.$$

Figure 20.3 shows that B^S is decreasing in the return rate. When we combine result with the previous one, namely, that GAP is increasing in r, then it is appealing to conjecture that the manufacturer and retailer control-variables are strategic substitutes. This can be seen by rewriting the support rate as

$$B^S = \frac{2u_M\left[(2\mu_1 - \varphi_1)\,r + 2\mu_2 - \varphi_2\right]}{A^S}.$$

Therefore, a higher A^S leads the retailer to lower its support rate (but not necessarily the total amount of the subsidy given to the manufacturer). Figure 20.4 reveals that the support rate increases with ϕ. This result is somehow expected, given that the incentive provided by the manufacturer to the retailer is precisely to (hopefully) drive up the retailer's participation in the green-activity program [19]. Still, it is interesting to mention the very significant effect that ϕ has on the support rate. Indeed, increasing ϕ, for instance by less than 15 % (i.e., from 0.35 to 0.40) more than triples the support provided by the retailer. Note that the positive impact of ϕ on the investment in GAP is much less limited (see Fig. 20.2). Further, the higher the production-cost saving resulting from recycling (higher $c_n - c_u$), the steeper the

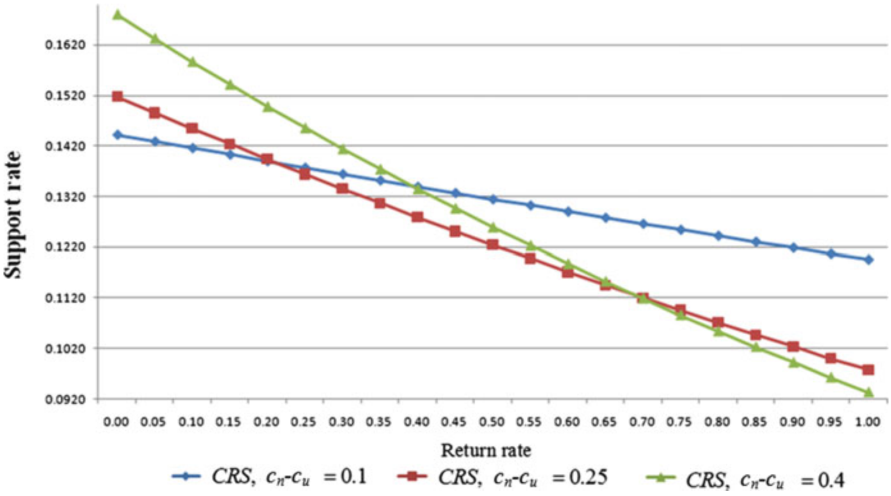


Fig. 20.3 Support-rate strategy for different values of cost savings

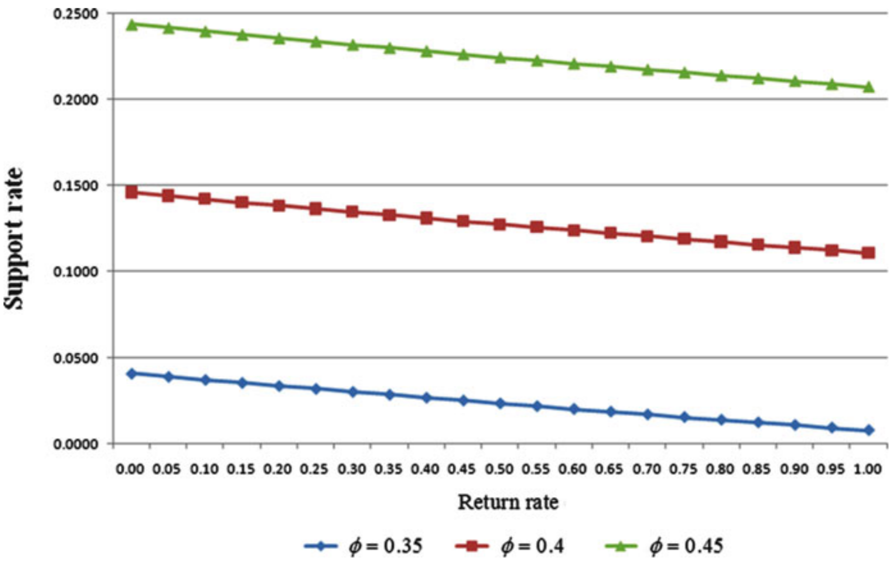


Fig. 20.4 Support-rate strategy for different values of sharing parameter

decline in the rate of support provided by the retailer to the manufacturer. Actually, when the manufacturer is highly efficient and the return rate of used products is also high, the retailer's support is simply less important. Note that the impact of varying $c_n - c_u$ on the support rate is proportionally much less visible than the impact of ϕ. However, the level of GAP is very sensitive to the value of $c_n - c_u$.

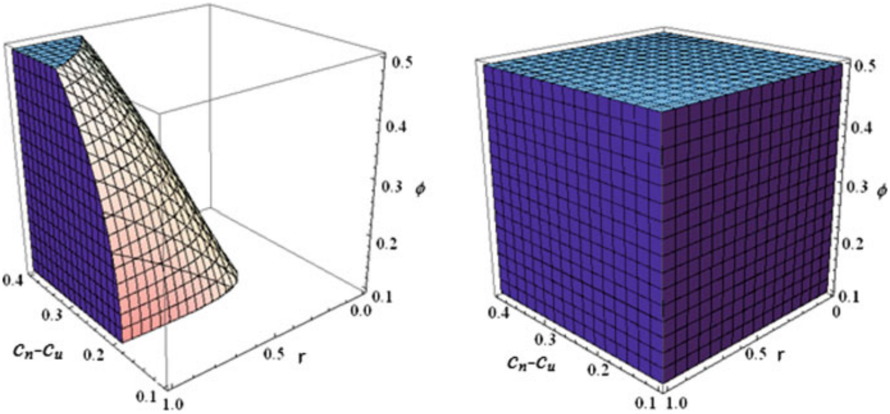


Fig. 20.5 Regions where manufacturer (*left*) and retailer (*right*) S profits are higher than Nash

Players' Payoffs: We now turn to the crucial issue of the profitability of a coop program along with a RRSC. Based on the previous results, the following proposition characterizes the form of the players' value functions in the S scenario.

Proposition 20.3. *If the form of the strategies are as in Figs. 20.1–20.4 and are invariant to changes in the parameter values, then the two players' value functions, i.e.,*

$$V_M^S(r) = \frac{\varphi_1}{2}r^2 + \varphi_2 r + \varphi_3,$$

$$V_R^S(r) = \frac{\mu_1}{2}r^2 + \mu_2 r + \mu_3,$$

are convex increasing, and assume their minimum levels at $r = 0$ *and maximal values at* $r = 1$.

Proof. See Appendix. □

This result is a clear invitation for members of a supply chain to implement policies to increase the return rate. That being said, it remains to be seen if the chain members benefit individually and collectively from the implementation of a CRS mechanism. Figure 20.5 shows, for each player, the region where the difference between the Stackelberg and Nash payoffs is positive. For the retailer, the result is clear cut. Indeed, for all feasible values of the return rate and all plausible values of cost reduction $(c_n - c_u)$ and ϕ, the retailer always prefers the S game to the N game. This shows that a two-parameter contract with costless monitoring always makes the player who receives a share better off (see [7]).

As for the manufacturer, the conclusion depends on the return rate and the parameter values. Roughly speaking, the manufacturer realizes a higher payoff in the S equilibrium than in the benchmark scenario when (a) the return rate is "sufficiently high;" (b) the cost reduction resulting from recycling is "sufficiently high;" and, (c) the incentive parameter ϕ is "not too high." If these conditions are met, then a CRS contract is Pareto payoff-improving. As the cost savings and the return rate are expected to vary significantly across firms and industries, it is difficult to make a general statement about the feasibility of Pareto-optimality in practice. Based on the following data, it seems reasonable to believe that this result is achievable. Indeed, regarding the return rate, Guide [23] reports that, if firms collect the products themselves, or provide incentives to other CLSC participants and adopt a market-driven approach, then the return rate can be as high as 82 %. If this example is representative of what can be realized, then we are in the zone of a "sufficiently high return rate." Concerning the cost savings, [18] report that remanufacturing costs at IBM are much lower than those for buying new parts, sometimes 80 % lower. Similarly, Xerox saves 40–65 % of its manufacturing costs by reusing parts and materials from returned products [43]. For these firms, the cost reduction due to remanufacturing is clearly "sufficiently high." However, these examples are (good) business exceptions, and no one would expect these levels of cost reduction to be very common. According to [13], most firms do not adopt closed-loop practices because of the small savings and inefficient remanufacturing. Nevertheless, other strategic motivations could still lead those firms to close the loop. For instance, to avoid and reduce remanufacturing competition, Lexmark has introduced the Prebate program, whereby customers who return an empty printer cartridge obtain a discount on a new cartridge; Lexmark does not remanufacture these used cartridges because the cost savings are low, but recycling them instead allows to reduce competition (www.atlex.com2003).

The last determinant of Pareto-optimality is the sharing-revenue parameter ϕ. The literature on contracting and coordination has already established the appropriateness of a revenue-sharing contract in a CLSC, and highlighted the critical role played by the sharing parameter [8]. Its actual value depends on the players' bargaining power, and therefore, no general statement can be made.

20.5 Conclusion

To the best of our knowledge, this study is the first attempt to assess, in a dynamic setting, the impact on a closed-loop supply chain of a CRS contract. Our starting point is that firms can influence the return rate of used products by carrying out green activities, and that this return rate is an inherently dynamic process. The optimality of a CRS can be assessed from the consumer's, the environmental, and the firms' points of view. We wrap up our main results in the following series of claims on these different points of view:

Claim 1 *Compared to the benchmark scenario, a CRS contract leads to a lower retail price and higher demand.*

Claim 2 *Compared to the benchmark scenario, a CRS contract leads to higher investments in green activities and a higher used-product-return rate.*

Claim 3 *A CRS contract is Pareto-improving with respect to the benchmark scenario only under some conditions. However, the retailer is always better off with such program.*

The conclusion is that the consumer and the retailer will vote in favor of such a CRS contract and that the environment is always in better off with one. For the manufacturer, the results are not clear cut.

As future research direction, the following extensions are worth considering:

A1. An analysis of the same game, but assuming a finite horizon. Indeed, our assumption of infinite horizon is a very strong one and were made mainly for tractability, i.e., to solve a system of algebraic Ricatti equations instead of having to deal with a highly coupled differential-equations system. A first step could be to analyze a two-stage game, where in the second period, the manufacturer produces with used parts recycled from first-period sales.

A2. An analysis of a multi-retailer situation, in which a manufacturer cooperates with different retailers while the retailers compete in the same market. This type of multi-agent configuration has been shown to be extremely important when evaluating a contract in supply-chain management. For instance, [7] evaluate a RRSC in a one-manufacturer–one-retailer chain configuration, and demonstrated its effectiveness for mitigating the double-marginalization effect and for making players better off. Later, they model a multi-retailer situation, and show that the positive effects of a two-parameter contract vanish whenever retailer competition occurs.

A3. A competitive setting where a manufacturer and an original equipment manufacturer (OEM) compete in the collection process. In this context, the manufacturer has more reasons to collect the end-of-use products, where the reverse flows need to be managed not only to appropriate some of the returns' residual value, but also to deter new entrants into the industry [13]. This context has also been described by [3] with real applications (e.g., Bosch). They report that remanufacturing can be really effective in a competitive context because remanufactured products may cannibalize competitors' products. However, the literature has overlooked competition in dynamic-CLSC settings, where players compete while adopting an active return policy.

A4. An evaluation of the impact of a green brand image on remanufacturing. In a CLSC, marketing and operations interface to ensure high remanufacturing efficiency while goodwill not only plays the traditional role of increasing sales (marketing role) but it also increases product returns (operational role). The main assumption here is that customer returns depend on the stock of (green) goodwill, which acts as a sustainable lever. Several companies, such as Coca-Cola, HP, and Panasonic, are modifying their brands, changing the colors and style to increase customers' green consciousness. Firms know that customers are concerned about the environment and are willing to buy and return green products, and that an appropriate brand strategy may provide superior performance. CLSCs seek to use goodwill not only to increase sales but also to induce customers to adopt sustainable behavior. By returning end-of-use products, customers contribute to conserving landfill space, reducing air pollution and preserving the environment. Therefore, green goodwill acts as a sustainable lever with the dual purpose of increasing both sales and returns.

A5. The integration of some quality features in our assumptions. The product remanufacturability, in fact, decreases over time, thereby also reducing the attractiveness. The quality of a return governs disassembly, recovery, and disposal operations to be carried out after closing the loop [38]. When a return is in good conditions, it possesses high residual value, and remanufacturing turns out to be an extremely appropriate operational strategy. Consequently, firms in the CLSC are committed to reducing the residence time (which is the time a product stays with customers) while increasing product remanufacturability (the high number of times a return may be used in a remanufacturing process). One our paper's main assumptions is that a return can be remanufactured an infinite number of times. Despite the obvious limitations, applications do exist in several industries (e.g., glass industry). Research in CLSCs has investigated remanufacturability in terms of product durability [3,21], highlight the trade-off this implies. High product durability maximizes cost savings but it considerably extends product life, thereby lowering demand [10]. Moreover, since durability is a quality feature, it impacts directly on production costs, reducing the players' unit-profits margin. Product durability has been also investigated as a dynamic phenomenon [41], where the stock of durability decreases over time, influencing both operational decisions and sales in future periods. While incorporating durability substantially increases the complexity of the model, addressing this trade-off determines the CLSC's success and improves decision-making process.

Acknowledgements We wish to thank the anonymous reviewer for his/her very helpful comments. Research supported by NSERC, Canada.

Appendix

Proof of Proposition 20.1

We need to establish the existence of a pair of value functions $V_M(r)$, $V_R(r)$ such that there exists a unique solution r to (20.2) and that the Hamilton–Jacobi–Bellman (HJB) equations

$$\rho V_M(r) = (\alpha - \beta p)\,(\omega - c_n - r\,(c_u - c_n)) - \frac{u_M}{2}A^2 + V_M'(A - \delta r), \tag{20.16}$$

$$\rho V_R = (\alpha - \beta p)\,(p - \omega), \tag{20.17}$$

are satisfied for all $r \geq 0$.

Since the retailer's objective does not depend on the state, only the manufacturer's value function is continuous in r. Maximizing the right-hand side (RHS) of the HJB for the GAP expenditures and price gives the following strategies:

$$A = \frac{V_M'}{u_M}, \tag{20.18}$$

$$p = \frac{\alpha + \beta\omega}{2\beta}. \tag{20.19}$$

Inserting (20.18) and (20.19) into (20.16) and (20.17) provides

$$\rho V_M = \left(\frac{\alpha - \beta\omega}{2\beta}\right)(\omega - c_n - r\,(c_u - c_n)) + V_M'\left(\frac{V_M'}{2u_M} - \delta r\right), \tag{20.20}$$

$$\rho V_R = \frac{(\alpha - \beta\omega)^2}{4\beta}. \tag{20.21}$$

We show that a linear value function satisfies (20.20) and (20.21). We define $V_M = \varsigma_1 r + \varsigma_2$, where ς_1 and ς_2 are the constants. Substituting V_M and its derivative into (20.20) we obtain:

$$\rho\,(\varsigma_1 r + \varsigma_2) = \left(\frac{\alpha - \beta\omega}{2}\right)(\omega - c_n - r\,(c_u - c_n)) + \varsigma_1\left(\frac{\varsigma_1}{2u_M} - \delta r\right). \tag{20.22}$$

By identification, we obtain

$$\varsigma_1 = \frac{(c_n - c_u)\,(\alpha - \beta\omega)}{2\,(\rho + \delta)},$$

$$\varsigma_2 = \frac{1}{2\rho}\left(\frac{u_M(\alpha - \beta\omega)(\omega - c_n) + \beta\varsigma_1^2}{u_M}\right),$$

$$V_M = \frac{(c_n - c_u)\,(\alpha - \beta\omega)}{2\,(\rho + \delta)}r + \frac{1}{2\rho}\left(\frac{u_M(\alpha - \beta\omega)(\omega - c_n)}{u_M} + \frac{(c_n - c_u)^2\,(\alpha - \beta\omega)^2}{4u_M\,(\rho + \delta)^2}\right).$$

By simple substitutions, we obtain the results in Proposition 20.1.

Proof of Proposition 20.2

We need to establish the existence of bounded and continuously differentiable value functions $V_M(r)$, $V_R(r)$ such that there exists a unique solution $r(t)$ to (20.2) and the HJB equations. To obtain a Stackelberg equilibrium, first we determine the manufacturer's green expenditures as a function of the retailer's controls p and B. The manufacturer's HJB is

$$\rho V_M(r) = \max_{A\geq 0}\{(1-\phi)(\alpha-\beta p)(\omega-c_n+r(c_n-c_u)) - \frac{u_M}{2}(1-B)A^2 + V_M'(r)(A-\delta r)\}. \tag{20.23}$$

Maximization of the RHS yields:

$$A = \frac{V_M'}{u_M(1-B)}. \tag{20.24}$$

Then, we substitute (20.24) into the retailer's HJB equation to obtain

$$\rho V_R(r) = \max_{p\geq 0, 0\leq B\leq 1}\left\{(\alpha-\beta p)[p-\omega+\phi(\omega-c_n+r(c_n-c_u))] - \frac{B^2(V_M')^2}{2u_M(1-B)^2} + V_R'\left(\frac{V_M'}{u_M(1-B)}-\delta r\right)\right\}. \tag{20.25}$$

Performing the maximization of the right-hand side we obtain

$$p = \frac{\alpha+\beta[\omega-\phi(\omega-c_n+r(c_n-c_u))]}{2\beta}, \tag{20.26}$$

$$B = \frac{2V_R'-V_M'}{2V_R'+V_M'}. \tag{20.27}$$

Inserting (20.26) and (20.27) inside the HJB we get

$$\rho V_M(r) = \left\{(1-\phi)\left(\frac{\alpha-\beta(\omega-\phi(\omega-c_n+r(c_n-c_u)))}{2}\right)(\omega-c_n+r(c_n-c_u)) + V_M'\left(\frac{2V_R'+V_M'}{4u_M}-\delta r\right)\right\}, \tag{20.28}$$

$$\rho V_R(r) = \left\{ \left(\frac{(\alpha - \beta\left[\omega - \phi\left(\omega - c_n + r\left(c_n - c_u\right)\right)\right])^2}{4\beta} \right) \right.$$
$$\left. + \frac{V_M'}{8u_M} + V_R'\left(\frac{V_R' + V_M'}{2u_M} - \delta r\right) \right\}. \tag{20.29}$$

Because the differential game is of the linear-quadratic variety, we make the informed guess that the value functions are quadratic, i.e.,

$$V_M(r) = \frac{\varphi_1}{2} r^2 + \varphi_2 r + \varphi_3,$$
$$V_R(r) = \frac{\mu_1}{2} r^2 + \mu_2 r + \mu_3,$$

where $\varphi_1, \varphi_2, \varphi_3, \mu_1, \mu_2$ and μ_3 are parameters to be determined. Let

$$a_1 = 2u_M(1-\phi)\beta\phi\left(c_u - c_n\right)^2,$$
$$a_2 = 2u_M\left(\rho + 2\delta\right),$$
$$a_3 = u_M(1-\phi)\left(c_n - c_u\right)\left(\alpha - \beta\left(\omega - 2\phi\left(\omega - c_n\right)\right)\right),$$
$$a_4 = 4u_M\left(\delta + \rho\right),$$
$$a_5 = 2u_M(1-\phi)\left(\omega - c_n\right)\left[\alpha - \beta\left(\omega - \phi\left(\omega - c_n\right)\right)\right],$$
$$a_6 = 4u_M\rho,$$
$$a_7 = 2u_M\phi^2\beta\left(c_u - c_n\right)^2,$$
$$a_8 = 2u_M\phi\left(c_n - c_u\right)\left(\alpha - \beta\left(\omega - \phi\left(\omega - c_n\right)\right)\right),$$
$$a_9 = 2u_M\left(\alpha - \beta\left[\omega - \phi\left(\omega - c_n\right)\right]\right)^2.$$

Inserting V_M and V_R and their derivatives in (20.28) and (20.29), we obtain the following six algebraic Ricatti equations:

$$a_1 + \varphi_1\left(2\mu_1 + \varphi_1\right) - a_2\varphi_1 = 0, \tag{20.30}$$
$$a_3 + \varphi_1\mu_2 + \varphi_2\left(\varphi_1 + \mu_1 - a_4\right) = 0, \tag{20.31}$$
$$a_5 + \varphi_2\left(2\mu_2 + \varphi_2\right) - a_6\varphi_3 = 0, \tag{20.32}$$
$$a_7 + 2\left(2\mu_1 - a_2\right)\mu_1 + \left(\varphi_1 + 4\mu_1\right)\varphi_1 = 0, \tag{20.33}$$
$$a_8 + \left(\varphi_1 + 2\mu_1\right)\varphi_2 + 2\left(2\mu_1 + \varphi_1 - a_4\right)\mu_2 = 0, \tag{20.34}$$
$$a_9 + \beta\left[\varphi_2^2 + 4\mu_2\left(\mu_2 + \varphi_2\right) - 2a_6\mu_3\right] = 0, \tag{20.35}$$

with the first three equations corresponding to the manufacturer and the next three to the retailer.

We briefly describe the procedure used to reduce the solution of that system into the solution of one non-linear equation to be solved numerically using Maple 10. From (20.30), we can obtain μ_1 as a function of $\varphi_1 : \mu_1 = f(\varphi_1)$, where

$$f_1(\varphi_1) = \Omega_1 = \frac{(a_2 - \varphi_1)\,\varphi_1 - a_1}{2\varphi_1}. \tag{20.36}$$

Replacing (20.36) for (20.31) and (20.34), we can obtain both φ_2 and μ_2 as function of φ_1

$$\varphi_2 = f_2(\varphi_1) = -\frac{a_3\Omega_3 + \varphi_1\Omega_2}{(\varphi_1 + \Omega_1 - a_4)\,\Omega_3} = \Omega_4 \tag{20.37}$$

$$\mu_2 = f_3(\varphi_1) = \frac{\Omega_2}{\Omega_3} = \Omega_5 \tag{20.38}$$

with

$$\Omega_2 = a_3\,(\varphi_1 + 2\Omega_1) - a_8\,(\varphi_1 + \Omega_1 - a_4)\,,$$
$$\Omega_3 = 2\,(2\Omega_1 + \varphi_1 - a_4)\,(\varphi_1 + \Omega_1 - a_4) - \varphi_1\,(\varphi_1 + 2\Omega_1)\,.$$

Similarly, we use (20.36)–(20.38) to derive φ_3 and μ_3 as a function of φ_1:

$$\varphi_3 = f_4(\varphi_1) = \frac{a_5 + (2\Omega_5 + \Omega_4)\,\Omega_4}{a_6}, \tag{20.39}$$

$$\mu_3 = f_5(\varphi_1) = \frac{a_9 + \beta\Omega_4^2 + 4\beta\,(\Omega_5 + \Omega_4)\,\Omega_5}{2a_6\beta}. \tag{20.40}$$

Finally, replacing (20.36) into (20.33) gives a non-linear equation that unfortunately cannot be solved analytically. We use the Maple "fsolve" function that uses numerical-approximation techniques to find a decimal approximation to a solution of an equation or a system of equations.

Proof of Proposition 20.3

Figure 20.1 shows that the advertising strategy

$$A^S = \frac{(2\mu_1 + \varphi_1)\,r + 2\mu_2 + \varphi_2}{2u_M},$$

is increasing in r, with $A^S(0) > 0$. Therefore, we conjecture the following relationships between the coefficients:

$$2\mu_1 + \varphi_1 > 0 \text{ and } 2\mu_2 + \varphi_2 > 0. \tag{20.41}$$

From the positivity of A^S, Fig. 20.2, and the expression of the support rate, we conclude that

$$2\mu_1 - \varphi_1 > 0 \text{ and } 2\mu_2 - \varphi_2 > 0 \ldots \tag{20.42}$$

Combining (20.41) and (20.42), we conclude that μ_1 and μ_2 are positive. The fact that the support rate is decreasing in the return rate leads to

$$\frac{\partial B^S}{\partial r} = \frac{4\left(\mu_1\varphi_2 - \mu_2\varphi_1\right)}{\left((2\mu_1+\varphi_1)\,r + 2\mu_2 + \varphi_2\right)^2} < 0 \Rightarrow \mu_1\varphi_2 - \mu_2\varphi_1 < 0. \tag{20.43}$$

Further, the positivity of A^S and $B^S(0) \leq 1$, imply $\varphi_2 > 0$. From the positivity of φ_2, μ_1 and μ_2, and the condition in (20.43), we conclude that $\varphi_1 > 0$.

Finally, it suffices to note that, since all other parameters involved in equations (20.32) and (20.35) are positive, a necessary condition for these equations to hold is to have φ_3 and μ_3 positive.

References

1. Aras, N., Boyaci, T., Verter, V.: The effect of categorizing returned products in remanufacturing. IIE Trans. **36**(4), 319–331 (2004)
2. Atasu, A., Guide, V.D.R., Van Wassenhove, L.N.: Product reuse economics in closed-loop supply chain research. Prod. Oper. Manage. **17**(5), 483–497 (2008)
3. Atasu, A., Sarvary, M., Van Wassenhove, L.N.: Remanufacturing as a marketing strategy. Manage. Sci. **54**(10), 1731–1746 (2008)
4. Berger, M.: Vertical cooperative advertising ventures. J. Marketing Res. **9**, 309–312 (1972)
5. Bergen, M., John, G.: Understanding cooperative advertising participation rates in conventional channels. J. Marketing Res. **46**, 357–369 (1997)
6. Bhattacharya, S., Guide, V.D.R., Van Wassenhove, L.N.: Optimal order quantities with remanufacturing across new product generations. Prod. Oper. Manage. J. **15**(3), 421–431 (2006)
7. Cachon, G.P., Lariviere, M.A.: 2005. Supply chain coordination with revenue sharing contracts: strength and limitations. Manage. Sci. **51**, 30–44 (2005)
8. Corbett, C.J., Savaskan, R.C.: Contracting and coordination in closed-loop supply chains. In: Daniel, V., Guide, R., Van Wassenhove, L.N. (eds.) Business Aspects of Closed-Loop Supply Chains: Exploring the Issues. Carnegie Mellon University Press, Pittsburgh (2003)
9. Dana, Jr., J.D., Spier, K.E.: Revenue sharing and vertical control in the video rental industry. J. Ind. Econ. **49**(3), 223–245 (2001)
10. Debo, L.G., Toktay, L.B., Van Wassenhove, L.N.: Market segmentation and product technology selection for remanufacturable products. Manage. Sci. **51**, 1193–1205 (2005)
11. Dekker, R., Fleischmann, M., Van Wassenhove, L.N. (eds.) Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains. Springer, Berlin (2004)
12. El Ouardighi, F., Jørgensen, S., Pasin, F.: A dynamic game of operations and marketing management in a supply chain. Int. J. Game Theory Rev. **34**, 59–77 (2008)
13. Ferguson, M.E., Toktay, L.B.: The effect of competition on recovery strategies. Prod. Oper. Manage. **15**(3), 351–368 (2006)
14. Ferrer, G., Swaminathan, J.M.: Managing new and remanufactured products. Manage. Sci. **52**(1), 15–26 (2006)

15. Ferrer, G., Whybark, C.: Material planning for a remanufacturing facility. Prod. Oper. Manage. **10**, 112–124 (2001)
16. Fleischmann, M., Beullens, P., Bloemhof-Ruwaard, J., Van Wassenhove, L.N.: The impact of product recovery on logistics network design. Prod. Oper. Manage. **10**(2), 156–173 (2001)
17. Fleischmann, M., Bloemhof-Ruwaard, J., Dekker, R., van der Laan, E., Van Wassenhove, L.N.: Quantitative models for reverse logistics: a review. Eur. J. Oper. Res. **103**, 1–17 (1997)
18. Fleischmann, M., van Nunen, J., Grave, B.: Integrating closed-loop supply chain and spare parts management at IBM. ERIM Report Series Research in Management, ERS-2002-107-LIS (2002)
19. Geng, Q., Mallik, S.: Inventory competition and allocation in a multi-channel distribution system. Eur. J. Oper. Res. **182**(2), 704–729 (2007)
20. Gerchak, Y., Wang, Y.: Revenue-sharing vs. wholesale-price contracts in assembly systems with random demand. Prod. Oper. Res. **13**(1), 23–33 (2004)
21. Geyer, R., Van Wassenhove, L.N., Atasu, A.: The economics of remanufacturing under limited component durability and finite product life cycles. Manage. Sci. **53**(1), 88–100 (2007)
22. Ginsburg, J.: Once is ot enough. Business Week, April 16 (2001)
23. Guide, Jr., V.D.R.: Production planning and control for remanufacturing: industry practice and research needs. J. Oper. Manage. **18**, 467–483 (2000)
24. Guide, Jr., V.D.R., Jayaraman, V., Linton, J.D.: Building contingency planning for closed-loop supply chains with product recovery. J. Oper. Manage. **21**, 259–279 (2003)
25. Guide, Jr., V.D.R., Van Wassenhove, L.N.: The evolution of closed-loop supply chain research. Oper. Res. **57**(1), 10–18 (2009)
26. Guide, Jr., V.D.R., Van Wassenhove, L.N.: Managing product return for remanufacturing. Prod. Oper. Manage. **10**(2), 142–155 (2001)
27. Guide, Jr., V.D.R., Souza, G.C., Van Wassenhove, L.N., Blackburn, J.D.: Time of value of commercial product returns. Manage. Sci. **52**(8), 1200–1214 (2006)
28. Hauser, W.M., Lund, R.T.: The Remanufacturing Industry: Anatomy of Giant. Boston University, Boston (2003)
29. He, X., Prasad, A., Sethi, S.: Cooperative advertising and pricing in a dynamic stochastic supply chain: feedback Stackelberg strategies. Prod. Oper. Manage. **18**, 78–94 (2009)
30. Hussain, S.S.: Green consumerism and ecolabelling: a strategic behavioural model. J. Agric. Econ. **51**(1), 77–89 (2000)
31. Ingene, C.A., Parry, M.E.: Mathematical Models of Distribution Channels. Kluwer Academic, Dordrecht (2004)
32. Jørgensen, S., Sigué, S.P., Zaccour, G.: Dynamic cooperative advertising in a channel. J. Retail. **76**(1), 71–92 (2000)
33. Jørgensen, S., Sigué, S.P., Zaccour, G.: Stackelberg leadership in a marketing channel. Int. Game Theory Rev. **3**(1), 13–26 (2001).
34. Jørgensen, S., Taboubi, S., Zaccour, G.: Retail promotions with negative brand image effects: is cooperation possible? Eur. J. Oper. Res. **150**, 395–405 (2003)
35. Jørgensen, S., Zaccour, G.: Differential Games in Marketing. International Series in Quantitative Marketing. Kluwer Academic, Boston, MA (2004)
36. Karray, S., Zaccour, G.: Could co-op advertising be a manufacturer's counterstrategy to store brands? J. Bus. Res. **59**, 1008–1015 (2006)
37. Kodak: Corporate Environmental Annual Report. The Kodak Corporation, Rochester (1999)
38. Krikke, H., Le Blanc, I., van de Velde, S.: Product modularity and the design of closed-loop supply chains. Calif. Manage. Rev. **46**(2), 23–39 (2004)
39. Majumder, P., Groenevelt, H.: Competition in remanufacturing. Prod. Oper. Manage. **10**, 125–141 (2001)
40. Mantovalli, J.: The producer pays. Environ. Mag. **8**(3), 36–42 (1997)
41. Muller, E., Peles, Y.C.: Optimal dynamic durability. J. Econ. Dyn. Control **14**(3–4), 709–719 (1990)
42. Ray, S., Boyaci, T., Aras, N.: Optimal prices and trade-in rebates for durable, remanufacturable products. Manuf. Serv. Oper. Manage. **7**(3), 208–228 (2005)

43. Savaskan, R.C., Bhattacharya, S., Van Wassenhove, L.N.: Closed loop supply chain models with product remanufacturing. Manage. Sci. **50**, 239–252 (2004)
44. Savaskan, R.C., Van Wassenhove, L.N.: Reverse channel design: the case of competing retailers. Manage. Sci. **52**(1), 1–14 (2006)
45. Talbolt, S., Lefebvre, E., Lefebvre, L.A.: Closed-loop supply chain activities and derived benefits in manufacturing SMEs. J. Manuf. Technol. Manage. **18**(6), 627–658 (2007)