

# **RANDOM FIELDS AND THEIR GEOMETRY**

**Robert J. Adler**

Faculty of Industrial Engineering and Management  
Technion – Israel Institute of Technology  
Haifa, Israel

e-mail: robert@ieadler.technion.ac.il

**Jonathan E. Taylor**

Department of Statistics  
Stanford University  
Stanford, U.S.A.

e-mail: jtaylor@stat.stanford.edu

December 24, 2003



# Contents

<b>1</b>	<b>Random fields</b>	<b>1</b>
1.1	Random fields and excursion sets . . . . .	1
1.2	Gaussian fields . . . . .	2
1.3	The Brownian family of processes . . . . .	6
1.4	Stationarity . . . . .	11
1.4.1	Stochastic integration . . . . .	13
1.4.2	Moving averages . . . . .	16
1.4.3	Spectral representations on $\mathbb{R}^N$ . . . . .	19
1.4.4	Spectral moments . . . . .	24
1.4.5	Constant variance . . . . .	27
1.4.6	Isotropy . . . . .	27
1.4.7	Stationarity over groups . . . . .	32
1.5	Non-Gaussian fields . . . . .	34
<b>2</b>	<b>Gaussian fields</b>	<b>39</b>
2.1	Boundedness and continuity . . . . .	40
2.2	Examples . . . . .	49
2.2.1	Fields on $\mathbb{R}^N$ . . . . .	49
2.2.2	Differentiability on $\mathbb{R}^N$ . . . . .	52
2.2.3	Generalised fields . . . . .	54
2.2.4	Set indexed processes . . . . .	61
2.2.5	Non-Gaussian processes . . . . .	66
2.3	Borell-TIS inequality . . . . .	67
2.4	Comparison inequalities . . . . .	74

2.5	Orthogonal expansions . . . . .	77
2.5.1	Karhunen-Loève expansion . . . . .	84
2.6	Majorising measures . . . . .	87
<b>3</b>	<b>Geometry</b>	<b>95</b>
3.1	Excursion sets . . . . .	95
3.2	Basic integral geometry . . . . .	97
3.3	Excursion sets again . . . . .	103
3.4	Intrinsic volumes . . . . .	112
3.5	Manifolds and tensors . . . . .	117
3.5.1	Manifolds . . . . .	118
3.5.2	Tensors and exterior algebras . . . . .	122
3.5.3	Tensor bundles and differential forms . . . . .	128
3.6	Riemannian manifolds . . . . .	129
3.6.1	Riemannian metrics . . . . .	129
3.6.2	Integration of differential forms . . . . .	133
3.6.3	Curvature tensors and second fundamental forms . . . . .	138
3.6.4	A Euclidean example . . . . .	142
3.7	Piecewise smooth manifolds . . . . .	146
3.7.1	Piecewise smooth spaces . . . . .	148
3.7.2	Piecewise smooth submanifolds . . . . .	154
3.8	Intrinsic volumes again . . . . .	157
3.9	Critical Point Theory . . . . .	160
3.9.1	Morse theory for piecewise smooth manifolds . . . . .	161
3.9.2	The Euclidean case . . . . .	166
<b>4</b>	<b>Gaussian random geometry</b>	<b>171</b>
4.1	An expectation meta-theorem . . . . .	172
4.2	Suitable regularity and Morse functions . . . . .	186
4.3	An alternate proof of the meta-theorem . . . . .	190
4.4	Higher moments . . . . .	191
4.5	Preliminary Gaussian computations . . . . .	193
4.6	Mean Euler characteristics: Euclidean case . . . . .	197
4.7	The meta-theorem on manifolds . . . . .	208
4.8	Riemannian structure induced by Gaussian fields . . . . .	213
4.8.1	Connections and curvatures . . . . .	214
4.8.2	Some covariances . . . . .	216
4.8.3	Gaussian fields on $\mathbb{R}^N$ . . . . .	218
4.9	Another Gaussian computation . . . . .	220
4.10	Mean Euler characteristics: Manifolds . . . . .	222
4.10.1	Manifolds without boundary . . . . .	223
4.10.2	Manifolds with boundary . . . . .	225
4.11	Examples . . . . .	230
4.12	Chern-Gauss-Bonnet Theorem . . . . .	235

0	Contents	
<b>5</b>	<b>Non-Gaussian geometry</b>	<b>237</b>
5.1	Basics	237
5.2	Conditional expectations of double forms	237
<b>6</b>	<b>Suprema distributions</b>	<b>243</b>
6.1	The basics	243
6.2	Some Easy Bounds	246
6.3	Processes with a Unique Point of Maximal Variance	248
6.4	General Bounds	253
6.5	Local maxima	257
6.6	Local maxima above a level	258
<b>7</b>	<b>Volume of tubes</b>	<b>263</b>
7.1	Preliminaries	264
7.2	Volume of tubes for finite Karhunen-Loève Gaussian processes	264
7.2.1	Local geometry of $\text{Tube}(M, \rho)$	266
7.3	Computing $F_{j,r}^*(\Omega_r)$	269
7.3.1	Case 1: $\widehat{M} = \mathbb{R}^l$	270
7.3.2	Case 2: $\widehat{M} = S_\lambda(\mathbb{R}^l)$	276
7.3.3	Volume of tubes for finite Karhunen-Loève Gaussian processes revisited	278
7.4	Generalized Lipschitz-Killing curvature measures	280
	<b>References</b>	<b>281</b>

# 1

## Random fields

### 1.1 Random fields and excursion sets

If you have not yet read the Preface, then please do so now.

Since you have read the Preface, you already know two important things about this book:

- The “random fields” of most interest to us will be random mappings from subsets of Euclidean spaces or, more generally, from Riemannian manifolds to the real line. However, since it is often no more difficult to treat far more general scenarios, they may also be real valued random mappings on any measurable space.
- Central to much of what we shall be looking is the geometry of *excursion sets*.

**Definition 1.1.1** *Let  $f$  be a measurable, real valued function on some measurable space and  $T$  be a measurable subset of that space. Then, for each  $u \in \mathbb{R}$ ,*

$$(1.1.1) \quad A_u \equiv A_u(f, T) \stackrel{\Delta}{=} \{t \in T : f(t) \geq u\} \\ \equiv f^{-1}([u, \infty)),$$

*is called the excursion set of  $f$  in  $T$  over the level  $u$ . We shall also occasionally write excursion sets as  $f|_T^{-1}[u, \infty)$ .*

Of primary interest to us will be the setting in which the function  $f$  is a random field.

**Definition 1.1.2** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space and  $T$  a topological space. Then a measurable mapping  $f : \Omega \rightarrow \mathbb{R}^T$  is called a real valued random field<sup>1</sup>. Measurable mappings from  $\Omega$  to  $(\mathbb{R}^T)^d$ ,  $d > 1$ , are called vector valued random fields. If  $T \subset \mathbb{R}^N$ , we call  $f$  an  $(N, d)$  random field, and if  $d = 1$  simply an  $N$ -dimensional random field.

We shall generally not distinguish between

$$f_t \equiv f(t) \equiv f(t, \omega) \equiv (f(\omega))(t),$$

etc., unless there is some special need to do so. Throughout, we shall demand that all random fields are separable, a property due originally to Doob [22], which implies conditions on both  $T$  and  $X$ .

**Definition 1.1.3** An  $\mathbb{R}^d$ -valued random field  $f$ , on a topological space  $T$ , is called separable if there exists a countable dense subset  $D \subset T$  and a fixed event  $N$  with  $\mathbb{P}\{N\} = 0$  such that, for any closed  $B \subset \mathbb{R}^d$  and open  $I \subset T$ ,

$$\{\omega : f(t, \omega) \in B, \forall t \in I\} \Delta \{\omega : f(t, \omega) \in B, \forall t \in I \cap D\} \subset N.$$

Here  $\Delta$  denotes the usual symmetric difference operator, so that

$$(1.1.2) \quad A \Delta B = (A \cap B^c) \cup (A^c \cap B),$$

where  $A^c$  is the complement of  $A$ .

Since you have read the Preface, you also know that most of this book centres on Gaussian random fields. The next section is devoted to defining these and giving some of their basic properties. Fortunately, most of these have little to do with the specific geometric structure of the parameter space  $T$ , and after decades of polishing even proofs gain little in the way of simplification by restricting to special cases such as  $T = \mathbb{R}^N$ . Thus, at least for a while, we can and shall work in as wide as possible generality. Only when we get to geometry, in Chapter 3, will we need to specialise, either to Euclidean  $T$  or to Riemannian manifolds.

## 1.2 Gaussian fields

The centrality of Gaussian fields to this book is due to two basic factors:

- Gaussian processes have a rich, detailed and very well understood general theory, which makes them beloved by theoreticians.

---

<sup>1</sup>On notation: While we shall follow the standard convention of denoting random variables by upper case Latin characters, we shall use lower case to denote random functions. The reason for this will become clear in Chapter 3, where we shall need the former for tangent vectors.

- In applications of random field theory, as in applications of almost any theory, it is important to have specific, explicit formulae that allow one to predict, to compare theory with experiment, etc. As we shall see, it will be only for Gaussian (and related, cf. Section 1.5) fields that it is possible to derive such formulae in the setting of excursion sets.

The main reason behind both these facts is the convenient analytic form of the multivariate Gaussian density, and the related definition of a Gaussian process.

A real-valued random variable  $X$  is said to be *Gaussian* (or *normally distributed*) if it has the density function

$$\varphi(x) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}, \quad x \in \mathbb{R},$$

for some  $m \in \mathbb{R}$  and  $\sigma > 0$ . It is elementary calculus that the mean of  $X$  is  $m$  and the variance  $\sigma^2$ , and that the characteristic function is given by

$$\phi(\theta) = \mathbb{E}\{e^{i\theta X}\} = e^{i\theta m - \theta^2/2\sigma^2}.$$

We abbreviate this by writing  $X \sim N(m, \sigma^2)$ . The case  $m = 0$ ,  $\sigma^2 = 1$  is rather special and in this situation we say that  $X$  has a *standard* normal distribution. In general, if a random variable has zero mean we call it *centered*.

Since the indefinite integral of  $\varphi$  is not a simple function, we also need notation ( $\Phi$ ) for the distribution and ( $\Psi$ ) tail probability functions of a standard normal variable:

$$(1.2.1) \quad \Phi(x) \triangleq 1 - \Psi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx.$$

While  $\Phi$  and  $\Psi$  may not be explicit, there are simple, and rather important, bounds which hold for every  $x > 0$  and become sharp very quickly as  $x$  grows. In particular, in terms of  $\Psi$  we have<sup>2</sup>

$$(1.2.2) \quad \left(\frac{1}{x} - \frac{1}{x^3}\right) \varphi(x) < \Psi(x) < \frac{1}{x} \varphi(x),$$

An  $\mathbb{R}^d$ -valued random variable  $X$  is said to be *multivariate Gaussian* if, for every  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$ , the real valued variable  $\langle \alpha, X' \rangle = \sum_{i=1}^d \alpha_i X_i$  is Gaussian<sup>3</sup>. In this case there exists a mean vector  $m \in \mathbb{R}^d$

<sup>2</sup>The inequality (1.2.2) follows from the observation that

$$\left(1 - \frac{3}{x^4}\right) \varphi(x) < \varphi(x) < \left(1 + \frac{1}{x^2}\right) \varphi(x),$$

followed by integration over  $x$ .

<sup>3</sup>Note: Throughout the book, vectors are taken to be row vectors and a prime indicates transposition. The inner product between  $x$  and  $y$  in  $\mathbb{R}^d$  is denoted by  $\langle x, y \rangle$  or, occasionally, by  $x \cdot y$ .

with  $m_j = \mathbb{E}\{X_j\}$  and a non-negative definite<sup>4</sup>  $d \times d$  covariance matrix  $C$ , with elements  $c_{ij} = \mathbb{E}\{(X_i - m_i)(X_j - m_j)\}$ , such that the probability density of  $X$  is given by

$$(1.2.3) \quad \varphi(x) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} e^{-\frac{1}{2}(x-m)C^{-1}(x-m)'},$$

where  $|C| = \det C$  is the determinant<sup>5</sup> of  $C$ . Consistently with the one-dimensional case, we write this as  $X \sim N(m, C)$ , or  $X \sim N_d(m, C)$  if we need to emphasise the dimension.

In view of (1.2.3) we have that Gaussian distributions are completely determined by their first and second order moments and that uncorrelated Gaussian variables are independent. Both of these facts will be of crucial importance later on.

While the definitions are fresh, note for later use that it is relatively straightforward to check from (1.2.3) that the characteristic function of a multivariate Gaussian  $X$  is given by

$$(1.2.4) \quad \phi(\theta) = \mathbb{E}\{e^{i\langle\theta, X'\rangle}\} = e^{i\langle\theta, m'\rangle - \frac{1}{2}\theta C \theta'}.$$

where  $\theta \in \mathbb{R}^d$ .

One consequence of the simple structure of  $\phi$  is the fact that if  $\{X_n\}_{n \geq 1}$  is an  $L^2$  convergent<sup>6</sup> sequence of Gaussian vectors, then the limit  $X$  must also be Gaussian. Furthermore, if  $X_n \sim N(m_n, C_n)$ , then

$$(1.2.5) \quad |m_n - m|^2 \rightarrow 0, \quad \text{and} \quad \|C_n - C\|^2 \rightarrow 0,$$

as  $n \rightarrow \infty$ , where  $m$  and  $C$  are the mean and covariance matrix of the limiting Gaussian. The norm on vectors is Euclidean and that on matrices any of the usual. The proofs involve only (1.2.4) and the continuity theorem for convergence of random variables.

One immediate consequence of either (1.2.3) or (1.2.4) is that if  $A$  is any  $d \times d$  matrix and  $X \sim N_d(m, C)$ , then

$$(1.2.6) \quad AX \sim N(mA, A'CA).$$

---

<sup>4</sup>A  $d \times d$  matrix  $C$  is called non-negative definite (or positive semi-definite) if  $xCx' \geq 0$  for all  $x \in \mathbb{R}^d$ . A function  $C : T \times T \rightarrow \mathbb{R}$  is called non-negative definite if the matrices  $(C(t_i, t_j))_{i,j=1}^n$  are non-negative definite for all  $1 \leq n < \infty$  and all  $(t_1, \dots, t_n) \in T^n$ .

<sup>5</sup>Just in case you have forgotten what was in the Preface, here is a one-time reminder: The notation  $||$  denotes any of 'absolute value', 'Euclidean norm', 'determinant' or 'Lebesgue measure', depending on the argument, in a natural fashion. The notation  $\| \|$  is used only for either the norm of complex numbers or for special norms, when it usually appears with a subscript.

<sup>6</sup>That is, there exists a random vector  $X$  such that  $\mathbb{E}\{|X_n - X|^2\} \rightarrow 0$  as  $n \rightarrow \infty$ .

A judicious choice of  $A$  then allows us to compute conditional distributions as well. If  $n < d$  make the partitions

$$\begin{aligned} X &= (X^1, X^2) = ((X_1, \dots, X_n), (X_{n+1}, \dots, X_d)), \\ m &= (m^1, m^2) = ((m_1, \dots, m_n), (m_{n+1}, \dots, m_d)), \\ C &= \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \end{aligned}$$

where  $C_{11}$  is an  $n \times n$  matrix. Then each  $X^i$  is  $N(m^i, C_{ii})$  and the conditional distribution<sup>7</sup> of  $X^i$  given  $X^j$  is also Gaussian, with mean vector

$$(1.2.7) \quad m_{i|j} = m^i + C_{ij}C_{jj}^{-1}(X^j - m^j)'$$

and covariance matrix

$$(1.2.8) \quad C_{i|j} = C_{ii} - C_{ij}C_{jj}^{-1}C_{ji}.$$

We can now define a real valued *Gaussian random field* to be a random field  $f$  on a parameter set  $T$  for which the (finite dimensional) distributions of  $(f_{t_1}, \dots, f_{t_n})$  are multivariate Gaussian for each  $1 \leq n < \infty$  and each  $(t_1, \dots, t_n) \in T^n$ . The functions  $m(t) = \mathbb{E}\{f(t)\}$  and

$$C(s, t) = \mathbb{E}\{(f_s - m_s)(f_t - m_t)\}$$

are called the *mean* and *covariance functions* of  $f$ . Multivariate<sup>8</sup> Gaussian fields taking values in  $\mathbb{R}^d$  are fields for which  $\langle \alpha, f_t' \rangle$  is a real valued Gaussian field for every  $\alpha \in \mathbb{R}^d$ .

In fact, one can also go in the other direction as well. Given *any* set  $T$ , a function  $m : T \rightarrow \mathbb{R}$ , and a non-negative definite function  $C : T \times T \rightarrow \mathbb{R}$  there exists<sup>9</sup> a Gaussian process on  $T$  with mean function  $m$  and covariance function  $C$ .

Putting all this together, we have the important principle that, *for a Gaussian process, everything about it is determined by the mean and covariance functions*. The fact that no real structure is required of the parameter

---

<sup>7</sup>To prove this, take

$$A = \begin{pmatrix} \mathbb{1}_n & -C_{12}C_{22} \\ 0 & \mathbb{1}_{d-n} \end{pmatrix}$$

and define  $Y = (Y^1, Y^2) = AX$ . Check using (1.2.6) that  $Y^1$  and  $Y^2 \equiv X^2$  are independent and use this to obtain (1.2.7) and (1.2.8) for  $i = 1, j = 2$ .

<sup>8</sup>Similarly, Gaussian fields taking values in a Banach space  $B$  are fields for which  $\alpha(f_t)$  is a real valued Gaussian field for every  $\alpha$  in the topological dual  $B^*$  of  $B$ . The covariance function is then replaced by a family of operators  $C_{st} : B^* \rightarrow B$ , for which  $Cov(\alpha(f_t), \beta(f_s)) = \beta(C_{st}\alpha)$ , for  $\alpha, \beta \in B^*$ .

<sup>9</sup>This is a consequence of the Kolmogorov existence theorem, which, at this level of generality, can be found in Dudley [27]. Such a process is a random variable in  $\mathbb{R}^T$  and may have terrible properties, including lack of measurability in  $t$ . However, it will *always* exist.

space  $T$  is what makes Gaussian fields such a useful model for random processes on general spaces. To build an appreciation for this, we need to look at some examples. The following Section looks at entire class of examples that are generated from something called ‘Gaussian noise’, that we shall also exploit in Section 1.4 to develop the notion of stationarity of random fields. Many more examples can be found in Chapter 2, which looks at Gaussian fields and their properties in far more depth.

### 1.3 The Brownian family of processes

Perhaps the most basic of all random fields is a collection of independent Gaussian random variables. While it is simple to construct such random fields for finite and even countable parameter sets, deep technical difficulties obstruct the construction for uncountable parameter sets. The path that we shall take around these difficulties involves the introduction of random measures which, at least in the Gaussian case, are straightforward to formulate.

Let  $(T, \mathcal{T}, \nu)$  be a  $\sigma$ -finite measure space and denote by  $\mathcal{T}_\nu$  the collection of sets of  $\mathcal{T}$  of finite  $\nu$  measure. A *Gaussian noise* based on  $\nu$ , or ‘Gaussian  $\nu$ -noise’ is a random field  $W : \mathcal{T}_\nu \rightarrow \mathbb{R}$  such that, for all  $A, B \in \mathcal{T}_\nu$ ,

$$(1.3.1) \quad W(A) \sim N(0, \nu(A)).$$

$$(1.3.2) \quad A \cap B = \emptyset \Rightarrow W(A \cup B) = W(A) + W(B) \text{ a.s.}$$

$$(1.3.3) \quad A \cap B = \emptyset \Rightarrow W(A) \text{ and } W(B) \text{ are independent.}$$

Property (1.3.2) encourages one to think of  $W$ <sup>10</sup> as a random (signed) measure, although it is not generally  $\sigma$ -finite. We describe (1.3.3) by saying that  $W$  has *independent increments*.

**Theorem 1.3.1** *If  $(T, \mathcal{T}, \nu)$  is a measure space, then there exists a real valued Gaussian noise, defined for all  $A \in \mathcal{T}_\nu$ , satisfying (1.3.1)–(1.3.3).*

**Proof.** In view of the closing remarks of the preceding Section all we need do is provide an appropriate covariance function on  $\mathcal{T}_\nu \times \mathcal{T}_\nu$ . Try

$$(1.3.4) \quad C_\nu(A, B) \stackrel{\Delta}{=} \nu(A \cap B).$$

---

<sup>10</sup>While the notation ‘ $W$ ’ is inconsistent our determination to use lower case Latin characters for random functions, we retain it as a tribute to Norbert Wiener, who is the mathematical father of these processes.

This is positive definite, since for any  $A_i \in \mathcal{T}_\nu$  and  $\alpha_i \in \mathbb{R}$

$$\begin{aligned} \sum_{i,j} \alpha_i C_\nu(A_i, A_j) \alpha_j &= \sum_{i,j} \alpha_i \alpha_j \int_T I_{A_i}(x) I_{A_j}(x) \nu(dx) \\ &= \int_T \left( \sum_i \alpha_i I_{A_i}(x) \right)^2 \nu(dx) \\ &\geq 0. \end{aligned}$$

Consequently there exists a centered Gaussian random field on  $\mathcal{T}_\nu$  with covariance  $C_\nu$ . It is immediate that this field satisfies (1.3.1)–(1.3.3) and so we are done.  $\square$

A particularly simple example of a Gaussian noise arises when  $T = \mathbb{Z}$ . Take  $\nu$  a discrete measure of the form  $\nu(A) = \sum_k a_k \delta_k(A)$ , where the  $a_k$  are non-negative constants and  $\delta_k$  is the Dirac measure on  $\{k\}$ . For  $\mathcal{T}$  take all subsets of  $\mathbb{Z}$ . In this case, the Gaussian noise can actually be defined on points  $t \in T$  and extended to a signed measure on sets in  $\mathcal{T}_\nu$  by additivity. What we get is a collection  $\{W_k\}_{k \in \mathbb{Z}}$  of independent, centered, Gaussian variables, with  $\mathbb{E}\{W_k^2\} = a_k$ . If the  $a_k$  are all equal, this is classical Gaussian ‘white’ noise on the integers.

A more interesting case is  $T = \mathbb{R}^N$ ,  $\mathcal{T} = \mathcal{B}^N$ , the Borel  $\sigma$ -algebra on  $\mathbb{R}^N$  and  $\nu(\cdot) = |\cdot|$ , Lebesgue measure. This gives us a Gaussian white noise defined on the Borel subsets of  $\mathbb{R}^N$  of finite Lebesgue measure, which is also a field with orthogonal increments, in the sense of (1.3.3). It is generally called the *set indexed Brownian sheet*. It is not possible, in this case, to assign non-trivial values to given points  $t \in \mathbb{R}^N$ , as was the case in the previous example.

It also turns out that working with the Brownian sheet on all of  $\mathbb{B}^N$  is not really the right thing to do, since, as will follow from Theorem 1.3.3, this process is rather badly behaved. Restricting the parameter space to various classes of subsets of  $\mathbb{B}^N$  is the right approach. Doing so gives us a number of interesting examples, with which the remainder of this Section is concerned.

As a first step, restrict  $W$  to rectangles of the form  $[0, t] \subset \mathbb{R}_+^N$ , where  $t \in \mathbb{R}_+^N = \{(t_1, \dots, t_N) : t_i \geq 0\}$ . It then makes sense to define a random field on  $\mathbb{R}_+^N$  itself via the equivalence

$$(1.3.5) \quad W(t) = W([0, t]).$$

$W_t$  is called the Brownian sheet on  $\mathbb{R}_+^N$ , or *multiparameter Brownian motion*. It is easy to check that this field is the centered Gaussian process with covariance

$$(1.3.6) \quad \mathbb{E}\{W_s W_t\} = (s_1 \wedge t_1) \times \cdots \times (s_N \wedge t_N),$$

where  $s \wedge t \triangleq \min(s, t)$ .

When  $N = 1$ ,  $W$  is the standard Brownian motion on  $[0, \infty)$ . When  $N > 1$ , if we fix  $N - k$  of the indices, it is a scaled  $k$ -parameter Brownian sheet in the remaining variables. (This is easily checked via the covariance function.) Also, when  $N > 1$ , it follows immediately from (1.3.6) that  $W_t = 0$  when  $\min_k t_k = 0$ ; i.e. when  $t$  is on one of the axes. It is this image, with  $N = 2$ , of a sheet tucked in at two sides and given a good shake, that led Ron Pyke [76] to introduce the name.

A simple simulation of a Brownian sheet, along with its contour lines, is in Figure 1.3.

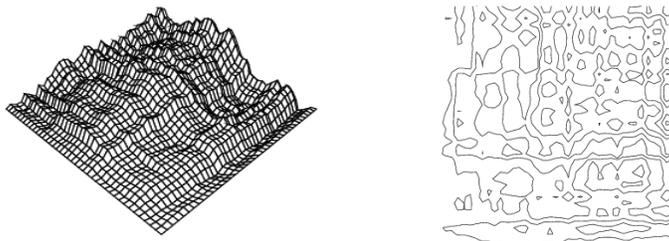


FIGURE 1.3.1. A simulated Brownian sheet on  $[0, 1]^2$ , along with its contour lines at the zero level.

One of the rather interesting aspects of the contour lines of Figure 1.3 is that they are predominantly parallel to the axes. There is a rather deep reason for this, and it has generated a rather massive literature. Many fascinating geometrical properties of the Brownian sheet have been discovered over the years (e.g. [18, 19, 19, 19, 20] and references therein) and a description of the potential theoretical aspects of the Brownian sheet is well covered in [51] where you will also find more references. Nevertheless, the geometrical properties of fields of this kind fall beyond the scope of our interests, since we shall be concerned with the geometrical properties of smooth (i.e. at least differentiable) processes only. Since the Brownian motion on  $\mathbb{R}_+^1$  is well known to be non-differentiable at all points, it follows from the above comments relating the sheet to the one dimensional case that Brownian sheets too are far from smooth.

Nevertheless, we shall still have need of these processes, primarily since they hold roughly the same place in the theory of multi-parameter stochastic processes that the standard Brownian motion does in one dimension. The Brownian sheet is a multi-parameter martingale (e.g. [18, 104, 105]) and forms the basis of the multiparameter stochastic calculus. There is a nice review of its basic properties in [100], which also develops its central rôle in the theory of stochastic partial differential equations, and describes in what sense it is valid to describe the derivative

$$\frac{\partial^N W(t_1, \dots, t_N)}{\partial t_1 \dots \partial t_N}$$

as Gaussian white noise.

The most basic of the sample path properties of the Brownian sheets are the continuity results of the following theorem, which we shall prove in Chapter 2, when we have all the tools. (cf. Corollary 2.2.2.) Introduce a partial order on  $\mathbb{R}^k$  by writing  $s \prec (\preceq) t$  if  $s_i < (\leq) t_i$  for all  $i = 1, \dots, k$ , and for  $s \preceq t$  let  $\Delta(s, t) = \prod_1^N [s_i, t_i]$ . Although  $W(\Delta(s, t))$  is already well defined via the original set indexed process, it also helpful to think of it as the ‘increment’ of the point indexed  $W_t$  over  $\Delta(s, t)$ ; viz.

$$(1.3.7) \quad W(\Delta(s, t)) = \sum_{\alpha \in \{0,1\}^N} (-1)^{N - \sum_{i=1}^N \alpha_i} W \left( s + \sum_{i=1}^N \alpha_i (t - s)_i \right).$$

We call  $W(\Delta(s, t))$  the *rectangle indexed* version of  $W$ .

**Theorem 1.3.2** *The point and rectangle indexed Brownian sheets are continuous over compact  $T \subset \mathbb{R}^N$ .*

In the framework of the general set indexed sheet, this result states that the Brownian sheet is continuous<sup>11</sup> over  $\mathcal{A} = \{\text{all rectangles in } T\}$  for compact  $T$ , and so bounded. This is far from a trivial result, for enlarging the parameter set, for the *same* process, can lead to unboundedness. The easiest way to see this is with an example.

An interesting, but quite simple example is given by the class of lower layers in  $[0, 1]^2$ . A set  $A$  in  $\mathbb{R}^N$  is called a *lower layer* if  $s \prec t$  and  $t \in A$  implies  $s \in A$ . In essence, restricted to  $[0, 1]^2$  these are sets bounded by the two axes and a non-increasing line. A specific example is given in Figure 1.3, which is part of the proof of the next theorem.

**Theorem 1.3.3** *The Brownian sheet on lower layers in  $[0, 1]^2$  is discontinuous and unbounded with probability one.*

**Proof.** We start by constructing some examples of lower layers. Write a generic point in  $[0, 1]^2$  as  $(s, t)$  and let  $T_{01}$  be the upper right triangle of  $[0, 1]^2$ ; i.e. those points for which  $s \leq 1$  and  $t \leq 1 \leq s + t$ . Let  $C_{01}$  be the largest square in  $T_{01}$ ; i.e. those points for which  $\frac{1}{2} < s \leq 1$  and  $\frac{1}{2} \leq t \leq 1$ .

Continuing this process, for  $n = 1, 2, \dots$ , and  $j = 1, \dots, 2^n$ , let  $T_{nj}$  be the right triangle defined by  $s + t \geq 1$ ,  $(j - 1)2^{-n} \leq s < j2^{-n}$ , and  $1 - j2^{-n} < t \leq 1 - (j - 1)2^{-n}$ . Let  $C_{nj}$  be the square filling the upper right corner of  $T_{nj}$ , in which  $(2j - 1)2^{-(n+1)} \leq s < j2^{-n}$  and  $1 - (2j - 1)2^{-(n+1)} \leq t < 1 - (j - 1)2^{-n}$ .

<sup>11</sup>For continuity, we obviously need a metric on the sets forming the parameter space. The symmetric difference metric of (1.1.2) is natural, and so we use it.

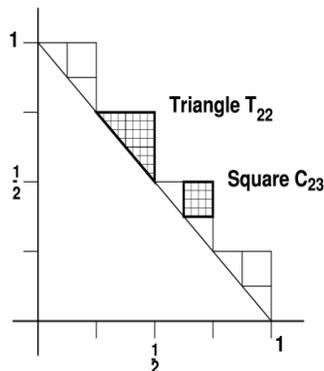


FIGURE 1.3.2. Construction of some lower layers.

The class of lower layers in  $[0, 1]^2$  certainly includes all sets made up by taking those points that lie between the axes and one of the step-like structures of Figure 1.3, where each step comes from the horizontal and vertical sides of some  $T_{nj}$  with, perhaps, different  $n$ .

Note that since the squares  $C_{nj}$  are disjoint for all  $n$  and  $j$ , the random variables  $W(C_{nj})$  are independent. Also  $|C_{nj}| = 4^{-(n+1)}$  for all  $n, j$ .

Let  $D$  be the negative diagonal  $\{(s, t) \in [0, 1]^2 : s + t = 1\}$  and  $L_{nj} = D \cap T_{nj}$ . For each  $n \geq 1$ , each point  $p = (s, t) \in D$  belongs to exactly one such interval  $L_{n, j(n, p)}$  for some unique  $j(n, p)$ .

For each  $p \in D$  and  $M < \infty$  the events

$$E_{np} \triangleq \{W(C_{n, j(n, p)}) > M2^{-(n+1)}\}$$

are independent for  $n = 0, 1, 2, \dots$ , and since  $W(C_{nj})/2^{-(n+1)}$  is standard normal for all  $n$  and  $j$  they also have the same positive probability. Thus, for each  $p$  we have that, for almost all  $\omega$ , the events  $E_{np}$  occur for all but finitely many  $n$ . Let  $n(p) = n(p, \omega)$  be the least such  $n$ .

Since the events  $E_{np}(\omega)$  are measurable jointly in  $p$  and  $\omega$ , Fubini's theorem implies that, with probability one, for almost all  $p \in D$  (with respect to Lebesgue measure on  $D$ ) some  $E_{np}$  occurs, and  $n(p) < \infty$ . Let

$$\begin{aligned} V_\omega &= \bigcup_{p \in D} T_{n(p), j(n(p), p)}, \\ A_\omega &\triangleq \{(s, t) : s + t \leq 1\} \cup V_\omega, \\ B_\omega &\triangleq A_\omega \setminus \bigcup_{p \in D} C_{n(p), j(n(p), p)}. \end{aligned}$$

Then  $A_\omega$  and  $B_\omega$  are lower layers. Furthermore, almost all  $p \in D$  belong to an interval of length  $2^{\frac{1}{2}-n(p)}$  which is the hypotenuse of a triangle with the square  $C_p = C_{n(p),j(n(p),p)}$  in its upper right corner, for which  $2W(C_p) > M2^{-n(p)}$ . Consequently,

$$W(A_\omega) - W(B_\omega) \geq \sum_p M2^{-n(p)}/2,$$

where the sum is over those  $p \in D$  corresponding to distinct intervals  $L_{n(p),j(n(p),p)}$ . Since the union of the countably many such intervals is almost all of the diagonal, the sum of  $\sum 2^{-n(p)}$  is precisely 1.

Hence  $W(A_\omega) - W(B_\omega) \geq M/2$ , implying that  $\max\{|W(A_\omega)|, |W(B_\omega)|\} \geq M/4$ . Sending  $M \rightarrow \infty$  we see that  $W$  is unbounded and so, *a fortiori*, discontinuous with probability one over lower layers in  $[0, 1]^2$ .  $\square$

The above argument is due to Dudley [25] and a similar argument shows that  $W$  is unbounded over the convex subsets of  $[0, 1]^3$ . Furthermore,  $W$  is also unbounded over the convex subsets of  $[0, 1]^k$  for all  $k \geq 4$  and (just to make sure that you don't confuse sample path properties with topological properties of the parameter space)  $W$  is continuous over convex subsets of the unit square. For details see [28].

These examples should be enough to convince you that the relationship between a Gaussian process and its parameter space is, as far as continuity and boundedness are concerned, an important and delicate subject. It therefore makes sense to look at this more carefully, before we look at further examples. Both of these tasks are therefore postponed to Chapter 2. In the meantime, we shall use Gaussian noise to look at some other less delicate, but nevertheless important, issues.

## 1.4 Stationarity

Stationarity has always been the backbone of almost all examples in the theory of Gaussian processes for which specific computations were possible. As described in the Preface, one of the main reasons we will be studying Gaussian processes on manifolds is to get around this assumption. Nevertheless, stationarity is an important concept and even if it were not true that it is a widely exhibited phenomenon, it would be worth studying to provide test cases for a more general theory. We anticipate that many of our readers will be familiar with most of the material of this Section and so will skip it and return only when specific details are required later<sup>12</sup>. For the newcomers, you should be warned that our treatment is only full enough

---

<sup>12</sup>The most important of these are concentrated in Section 1.4.4, and even the expert might want to look at these now.

to meet our specific needs and that both style and content are occasionally a little eclectic. In other words, you should go elsewhere for fuller, more standard treatments. References will be given along the way.

Although our primary interest lies in the study of real valued random fields it is mathematically more convenient to discuss stationarity in the framework of complex valued processes. Hence, unless otherwise stated, we shall assume throughout this section that  $f(t) = (f_R(t) + if_I(t))$  takes values in the complex plane  $\mathcal{C}$  and that  $\mathbb{E}\{\|f(t)\|^2\} = \mathbb{E}\{f_R^2(t) + f_I^2(t)\} < \infty$ . (Both  $f_R$  and  $f_I$  are, obviously, to be real valued.) As for a definition of normality in the complex scenario, we first define a complex random variable to be Gaussian if the vector of its two components is bivariate Gaussian<sup>13</sup>. A complex process  $f$  is Gaussian if  $\sum_i \alpha_{t_i} f'_{t_i}$  is a complex Gaussian variable for all sequences  $\{t_i\}$  and complex  $\{\alpha_{t_i}\}$ .

We also need some additional structure on the parameter space  $T$ . In particular, we need that it have a group structure<sup>14</sup> and an operation with respect to which the field is stationary. Consequently, we now assume that  $T$  has such a structure, '+' represents the binary operation on  $T$  and '-' represents inversion. As usual,  $t - s = t + (-s)$ . For the moment, we need no further assumptions on the group.

Since  $f_t \in \mathcal{C}$ , it follows that the mean function  $m(t) = \mathbb{E}\{f(t)\}$  is also complex valued, as is the covariance function, which we redefine for the complex case as

$$(1.4.1) \quad C(s, t) \triangleq \mathbb{E} \left\{ [f(s) - m(s)] [\overline{f(t) - m(t)}] \right\},$$

with the bar denoting complex conjugation.

Some basic properties of covariance functions follow immediately from (1.4.1):

- $C(s, t) = \overline{C(t, s)}$ , which becomes the simple symmetry  $C(s, t) = C(t, s)$  if  $f$  (and so  $C$ ) is real valued.
- For any  $k \geq 1$ ,  $t_1, \dots, t_k \in T$  and  $z_1, \dots, z_k \in \mathcal{C}$ , the Hermitian form  $\sum_{i=1}^k \sum_{j=1}^k C(t_i, t_j) z_i \bar{z}_j$  is always real and non-negative. We summarise this, as before, by saying that  $C$  is non-negative definite.

(The second property follows from the equivalence of the double sum to  $\mathbb{E}\{\|\sum_{i=1}^k [f(t_i) - m(t_i)] z_i\|^2\}$ .)

Suppose for the moment that  $T$  is Abelian. A random field  $f$  is called *strictly homogeneous* or *strictly stationary* over  $T$ , with respect to the group

<sup>13</sup>It therefore follows that a complex Gaussian variable  $X = X_R + iX_I$  is defined by five parameters:  $\mathbb{E}\{X_I\}$ ,  $\mathbb{E}\{X_R\}$ ,  $\mathbb{E}\{X_I^2\}$ ,  $\mathbb{E}\{X_R^2\}$  and  $\mathbb{E}\{X_I X_R\}$ .

<sup>14</sup>If stationarity is a new concept for you, you will do well by reading this Section the first time taking  $T = \mathbb{R}^N$  with the usual notions of addition and subtraction.

operation  $+$ , if its finite dimensional distributions are invariant under this operation. That is, for any  $k \geq 1$  and any set of points  $\tau, t_1, \dots, t_k \in T$

$$(1.4.2) \quad (f(t_1), \dots, f(t_k)) \stackrel{\mathcal{L}}{=} (f(t_1 + \tau), \dots, f(t_k + \tau)),$$

where  $\stackrel{\mathcal{L}}{=}$  indicates equivalence in distribution (law).

An immediate consequence of strict stationarity is that  $m(t)$  is constant and  $C(s, t)$  is a function of the difference  $s - t$  only.

Going the other way, if for a random field with  $\mathbb{E}\{\|f(t)\|^2\} < \infty$  for all  $t \in T$ , we have that  $m(t)$  is constant and  $C(s, t)$  is a function of the difference  $s - t$  only, then we call  $f$  simply *stationary* or *homogeneous*, occasionally adding either of the two adjectives ‘weakly’ or ‘second-order’.

Note that none of the above required the Gaussianity we have assumed up until now on  $f$ . If, however, we do add the assumption of Gaussianity, it immediately follows from the structure (1.2.3) of the multivariate Gaussian density<sup>15</sup> that a weakly stationary Gaussian field will also be strictly stationary if  $C'(s, t) = \mathbb{E}\{[f(s) - m(s)][f(t) - m(t)]\}$  is also a function only of  $s - t$ . If  $f$  is real valued, then since  $C \equiv C'$  it follows that all weakly stationary real valued Gaussian fields are also strictly stationary and the issue of qualifying adjectives is moot<sup>16</sup>.

If  $T$  is not Abelian we must distinguish between left and right stationary. We say that a random field  $f$  on  $T$  is *right-stationary* if (1.4.2) holds and that  $f$  is *left-stationary* if  $f'(t) \stackrel{\Delta}{=} f(-t)$  is right-stationary. The corresponding conditions on the covariance function change accordingly.

In order to build examples of stationary processes, we need to make a brief excursion into (Gaussian) stochastic integration.

### 1.4.1 Stochastic integration

We return to the setting of Section 1.3, so that we have a  $\sigma$ -finite<sup>17</sup> measure space  $(T, \mathcal{T}, \nu)$ , along with the Gaussian  $\nu$ -noise  $W$  defined over  $\mathcal{T}$ . Our aim

<sup>15</sup>Formally, (1.2.3) is not quite enough. Since we are currently treating complex valued processes, the  $k$ -dimensional marginal distributions of  $f$  now involve  $2k$ -dimensional Gaussian vectors. (The real and imaginary parts each require  $k$  dimensions.)

<sup>16</sup>The reason for needing the additional condition in the complex case should be intuitively clear:  $C$  alone will not even determine the variance of each of  $f_I$  and  $f_R$  but only their sum. However, together with  $C'$  both of these parameters, as well as the covariance between  $f_I$  and  $f_R$ , can be computed. See [69] for further details.

<sup>17</sup>Note that ‘ $\sigma$ -finite’ includes ‘countable’ and ‘finite’, in which case  $\nu$  will be a discrete measure and all of the theory we are about to develop is really much simpler. As we shall see later, these are more than just important special cases and so should not be forgotten as you read this Section. Sometimes it is all too easy to lose sight of the important simple cases in the generality of the treatment.

will be to establish the existence of integrals of the form

$$(1.4.3) \quad \int_T f(t) W(dt),$$

for *deterministic*  $f \in L^2(\nu)$  and, eventually, complex  $W$ . Appropriate choices of  $f$  will give us examples of stationary Gaussian fields, many of which we shall meet in the following Subsections.

Before starting, it is only fair to note that we shall be working with two and a bit assumptions. The ‘bit’ is that for the moment we shall only treat real  $f$  and  $W$ . We shall, painlessly, lift that assumption soon. Of the other two, one is rather restrictive and one not, but neither of importance to us. The non-restrictive assumption is the Gaussian nature of the process  $W$ . Indeed, since all of what follows is based only on  $L^2$  theory, we can, and so shall, temporarily drop the assumption that  $W$  is a Gaussian noise and replace conditions (1.3.1)–(1.3.3) with the following three requirements for all  $A, B \in \mathcal{T}$ .

$$(1.4.4) \quad \mathbb{E}\{W(A)\} = 0, \quad \mathbb{E}\{W(A)\}^2 = \nu(A).$$

$$(1.4.5) \quad A \cap B = \emptyset \Rightarrow W(A \cup B) = W(A) + W(B) \text{ a.s.}$$

$$(1.4.6) \quad A \cap B = \emptyset \Rightarrow \mathbb{E}\{W(A)W(B)\} = 0.$$

Note that in the Gaussian case (1.4.6) is really equivalent to the seemingly stronger (1.3.3), since zero covariance and independence are then equivalent.

The second restriction is that the integrand  $f$  in (1.4.3) is deterministic. Removing this assumption would lead us to having to define the Itô integral which is a construction for which we shall have no need.

Since, by (1.4.5),  $W$  is a finitely additive (signed) measure, (1.4.3) is evocative of Lebesgue integration. Consequently, we start by defining the stochastic version for simple functions

$$(1.4.7) \quad f(t) = \sum_1^n a_i \mathbb{1}_{A_i}(t),$$

where  $A_1, \dots, A_n$  are disjoint,  $\mathcal{T}$  measurable sets in  $T$ , by writing

$$(1.4.8) \quad W(f) \equiv \int_T f(t) W(dt) \triangleq \sum_1^n a_i W(A_i).$$

It follows immediately from (1.4.4) and (1.4.6) that in this case  $W(f)$  has zero mean and variance given by  $\sum a_i^2 \nu(A_i)$ . Think of  $W(f)$  as a mapping from simple functions in  $L^2(T, \mathcal{T}, \nu)$  to random variables<sup>18</sup> in

---

<sup>18</sup>Note that if  $W$  is Gaussian, then so is  $W(f)$ .

$L^2(\mathbb{P}) \equiv L^2(\Omega, \mathcal{F}, \mathbb{P})$ . The remainder of the construction involves extending this mapping to a full isomorphism from  $L^2(\nu) \equiv L^2(T, \mathcal{T}, \nu)$  onto a subspace of  $L^2(\mathbb{P})$ . We shall use this isomorphism to define the integral.

Let  $\mathcal{S} = \mathcal{S}(T)$  denote the class of simple functions of the form (1.4.7) for some finite  $n$ . Note first there is no problem with the consistency of the definition (1.4.8) over different representations of  $f$ . Furthermore,  $W$  clearly defines a linear mapping on  $\mathcal{S}$  which preserves inner products. To see this, write  $f, g \in \mathcal{S}$  as

$$f(t) = \sum_1^n a_i \mathbb{1}_{A_i}(t), \quad g(t) = \sum_1^n b_i \mathbb{1}_{A_i}(t);$$

in terms of the same partition, to see that the  $L^2(\mathbb{P})$  inner product between  $W(f)$  and  $W(g)$  is given by

$$\begin{aligned} (1.4.9) \quad \langle W(f), W(g) \rangle_{L^2(\mathbb{P})} &= \left\langle \sum_1^n a_i W(A_i), \sum_1^n b_i W(A_i) \right\rangle_{L^2(\mathbb{P})} \\ &= \sum_1^n a_i b_i \mathbb{E}\{[W(A_i)]^2\} \\ &= \int_T f(t)g(t) \nu(dt) \\ &= \langle f, g \rangle_{L^2(\nu)} \end{aligned}$$

the second line following from (1.4.6) and the second last from (1.4.4) and the definition of the Lebesgue integral.

Since  $\mathcal{S}$  is dense in  $L^2(\nu)$  (e.g. [81]) for each  $f \in L^2(\nu)$  there is a sequence  $\{f_n\}$  of simple functions such that  $\|f_n - f\|_{L^2(\nu)} \rightarrow 0$  as  $n \rightarrow \infty$ . Using this, for such  $f$  we define  $W(f)$  as the mean square limit

$$(1.4.10) \quad W(f) \triangleq \lim_{n \rightarrow \infty} W(f_n).$$

It follows from (1.4.9) that this limit exists and is independent of the approximating sequence  $\{f_n\}$ . Furthermore, the mapping  $W$  so defined is linear and preserves inner products; i.e. the mapping is an isomorphism. We take this mapping as our definition of the integral and so we now not only have existence, but from (1.4.9) we have that, for all  $f, g \in L^2(\nu)$ ,

$$(1.4.11) \quad \mathbb{E}\{W(f)W(g)\} = \int_T f(t)g(t) \nu(dt).$$

Note also that since  $L^2$  limits of Gaussian random variables remain Gaussian (cf. (1.2.5) and the discussion above it) under the additional assumption that  $W$  is a Gaussian noise it follows that  $W(f)$  is also Gaussian.

With our integral defined, we can now start looking at some examples of what can be done with it.

## 1.4.2 Moving averages

We now return to the main setting of this Section, in which  $T$  is an Abelian group under the binary operation  $+$  and  $-$  represents inversion. Let  $\nu$  be a Haar measure on  $(T, \mathcal{T})$  (assumed to be  $\sigma$ -finite) and take  $F : T \rightarrow \mathbb{R}$  in  $L^2(\nu)$ . If  $W$  is a  $\nu$ -noise on  $\mathcal{T}$ , then the random process

$$(1.4.12) \quad f(t) \triangleq \int_T F(t-s) W(ds),$$

is called a *moving average* of  $W$ , and we have the following simple result:

**Lemma 1.4.1** *Under the preceding conditions,  $f$  is a stationary random field on  $T$ . Furthermore, if  $W$  is a Gaussian noise, then  $f$  is also Gaussian.*

**Proof.** To establish stationarity we must prove that

$$\mathbb{E}\{f(t)f(s)\} = C(t-s)$$

for some  $C$ . However, from (1.4.11) and the invariance of  $\nu$  under the group operation,

$$\begin{aligned} \mathbb{E}\{f(t)f(s)\} &= \int_T F(t-u)F(s-u) \nu(du) \\ &= \int_T F(t-s+v)F(v) \nu(dv) \\ &\triangleq C(t-s), \end{aligned}$$

and we are done.

If  $W$  is Gaussian, then we have already noted when defining stochastic integrals that  $f(t) = W(F(t-\cdot))$  is a real-valued Gaussian random variable for each  $t$ . The same arguments also show that  $f$  is Gaussian as a *process*.  $\square$

A similar but slightly more sophisticated construction also yields a more general class of examples, in which we think of the elements  $g$  of a group  $G$  acting on the elements  $t$  of an underlying space  $T$ . This will force us to change notation a little and, for the argument to be appreciated in full, to assume that you also know a little about manifolds. If you do not, then you can return to this example later, after having read Chapter 3, or simply take the manifold to be  $\mathbb{R}^N$ . In either case, you may still want to read the very concrete and quite simple examples at the end of this subsection now.

Thus, taking the elements  $g$  of a group  $G$  acting on the elements  $t$  of an underlying space  $T$ , we denote the identity element of  $G$  by  $e$  and the left and right multiplication maps by  $L_g$  and  $R_g$ . We also write  $I_g = L_g \circ R_g^{-1}$  for the inner automorphism of  $G$  induced by  $g$ .

Since we are now working in more generality, we shall also drop the commutativity assumption that has been in force so far. This necessitates some

additional definitions, since we must distinguish between left and right stationarity. We say that a random field  $f$  on  $G$  is strictly *left-stationary* if for all  $n$ , all  $(g_1, \dots, g_n)$ , and any  $g_0$ ,

$$(f(g_1), \dots, f(g_n)) \stackrel{\mathcal{L}}{=} (f \circ L_{g_0}(g_1), \dots, f \circ L_{g_0}(g_n)).$$

It is called strictly *right-stationary* if  $f'(g) \triangleq f(g^{-1})$  is strictly left-stationary and strictly *bi-stationary*, or simply strictly *stationary*, if it is both left and right strictly stationary. As before, if  $f$  is Gaussian and has constant mean and covariance function  $C$  satisfying

$$(1.4.13) \quad C(g_1, g_2) = C'(g_1^{-1}g_2),$$

for some  $C' : G \rightarrow \mathbb{R}$ , then  $f$  is strictly left-stationary. Similarly, if  $C$  satisfies

$$(1.4.14) \quad C(g_1, g_2) = C''(g_1g_2^{-1})$$

for some  $C''$ , then  $f$  is right-stationary. If  $f$  is not Gaussian, but has constant mean and (1.4.13) holds, then  $f$  is *weakly* left-stationary. Weak right-stationarity and stationarity are defined analogously.

We can now start collecting the building blocks of the construction, which will be of a left-stationary Gaussian random field on a group  $G$ . An almost identical argument will construct a right-stationary field. It is then easy to see that this construction will give a bi-stationary field on  $G$  only if it is unimodular, i.e. if any left Haar measure on  $G$  is also right invariant.

We first add the condition that  $G$  be a *Lie group*; i.e. a group that is also a  $C^\infty$  manifold such that the maps taking  $g$  to  $g^{-1}$  and  $(g_1, g_2)$  to  $g_1g_2$  are both  $C^\infty$ . We say  $G$  has a smooth ( $C^\infty$ ) (left) action on a smooth ( $C^\infty$ ) manifold  $T$  if there exists a map  $\theta : G \times T \rightarrow T$  satisfying, for all  $t \in T$  and  $g_1, g_2 \in G$ ,

$$\begin{aligned} \theta(e, t) &= t \\ \theta(g_2, \theta(g_1, t)) &= \theta(g_2g_1, t). \end{aligned}$$

We write  $\theta_g : T \rightarrow T$  for the partial mapping  $\theta_g(t) = \theta(g, t)$ . Suppose  $\nu$  is a measure on  $T$ , let  $\theta_{g*}(\nu)$  be the *push-forward* of  $\nu$  under the map  $\theta_g$ ; i.e.  $\theta_{g*}(\nu)$  is given by

$$\int_A \theta_{g*}(\nu) = \int_{g^{-1}A} \nu.$$

Furthermore, we assume that  $\theta_{g*}(\nu)$  is absolutely continuous with respect to  $\nu$ , with Radon-Nikodym derivative

$$(1.4.15) \quad D(g) \triangleq \frac{d\theta_{g*}(\nu)}{d\nu}(t),$$

independent of  $t$ . We call such a measure  $\nu$  *left relatively invariant* under  $G$ . It is easy to see that  $D(g)$  is a  $C^\infty$  homomorphism from  $G$  into the multiplicative group of positive real numbers, i.e.  $D(g_1g_2) = D(g_1)D(g_2)$ . We say that  $\nu$  is *left invariant* with respect to  $G$  if, and only if, it is left relatively-invariant and  $D \equiv 1$ .

Here, finally, is the result.

**Lemma 1.4.2** *Suppose  $G$  acts smoothly on a smooth manifold  $T$  and  $\nu$  is left relatively invariant under  $G$ . Let  $D$  be as in (1.4.15) and let  $W$  be Gaussian  $\nu$ -noise on  $T$ . Then, for any  $F \in L^2(T, \nu)$ ,*

$$f(g) = \frac{1}{\sqrt{D(g)}} W(F \circ \theta_{g^{-1}}),$$

*is a left stationary Gaussian random field on  $G$ .*

**Proof.** We must prove that

$$\mathbb{E} \{f(g_1)f(g_2)\} = C(g_1^{-1}g_2)$$

for some  $C : G \rightarrow \mathbb{R}$ . From the definition of  $W$ , we have,

$$\begin{aligned} \mathbb{E} \{f(g_1)f(g_2)\} &= \frac{1}{\sqrt{D(g_1)D(g_2)}} \int_T F(\theta_{g_1^{-1}}(t)) F(\theta_{g_2^{-1}}(t)) \nu(dt) \\ &= \frac{1}{\sqrt{D(g_1)D(g_2)}} \int_T F(\theta_{g_1^{-1}}(\theta_{g_2}(t))) F(t) \theta_{g_2^*}(\nu)(dt) \\ &= \frac{D(g_2)}{\sqrt{D(g_1)D(g_2)}} \int_T F(\theta_{g_1^{-1}g_2}(t)) F(t) \nu(dt) \\ &= \sqrt{D(g_1^{-1}g_2)} \int_M F(\theta_{g_1^{-1}g_2}(t)) F(t) \nu(dt) \\ &\triangleq C(g_1^{-1}g_2) \end{aligned}$$

This completes the proof.  $\square$

It is easy to find simple examples to which Lemma 1.4.2 applies. The most natural generic example of a Lie group acting on a manifold is its action on itself. In particular, any right Haar measure is left relatively-invariant, and this is a way to generate stationary processes. To apply Lemma 1.4.2 in this setting one needs only to start with a Gaussian noise based on a Haar measure on  $G$ . In fact, this is the example (1.4.12) with which we started this Section.

A richer but still concrete example of a group  $G$  acting on a manifold  $T$  is given by  $G = GL(N, \mathbb{R}) \times \mathbb{R}^N$  acting<sup>19</sup> on  $T = \mathbb{R}^N$ . For  $g = (A, t)$  and

---

<sup>19</sup>Recall that  $GL(N, \mathbb{R})$  is the (general linear) group of transformations of  $\mathbb{R}^N$  by rotation.

$s \in \mathbb{R}^N$ , set

$$\theta(g, t)(s) = As + t.$$

In this example it is easy to see that Lebesgue measure  $\lambda_N(dt) = dt$  is relatively invariant with respect to  $G$  with  $D(g) = \det A$ . For an even more concrete example, take compact, Borel  $B \subset \mathbb{R}^N$  and  $F(s) = \mathbb{1}_B(s)$ . It then follows from Lemma 1.4.2 that

$$f(A, t) = \frac{W(A^{-1}(B - t))}{\sqrt{|\det(A)|}}$$

is a stationary process with variance  $\lambda_N(B)$  and covariance function

$$C((A_1, t_1), (A_2, t_2)) = \frac{|A_1 A_2^{-1}(B - (t_1 - t_2))|}{\sqrt{|\det(A_1 A_2^{-1})|}},$$

where we have adopted the usual notations that  $B + t = \{s + t : s \in B\}$  and  $AB = \{At : t \in B\}$ .

Examples of this kind have been used widely in practice. For examples involving the statistics of brain mapping see, for example, [86, 87].

### 1.4.3 Spectral representations on $\mathbb{R}^N$

The moving averages of the previous Subsection gave us examples of stationary fields that were rather easy to generate in quite general situations from Gaussian noise. Now, however, we want to look at a general way of generating *all* stationary fields, via the so-called *spectral representation*. This is quite a simple task when the parameter set is  $\mathbb{R}^N$ , but rather more involved when a general group is taken as the parameter space and issues of group representations arise. Thus we shall start with the Euclidean case which we treat in detail and then discuss some aspects of the general case in the following Subsection. In both cases, while an understanding of the spectral representation is a powerful tool for understanding stationarity and a variety of sample path properties of stationary fields, it is not necessary for what comes later in the book.

We return to the setting of complex valued fields, take  $T = \mathbb{R}^N$ , and assume, as usual, that  $\mathbb{E}\{f_i\} = 0$ . Furthermore, since we are now working only with stationary processes, it makes sense to misuse notation somewhat and write

$$C(t - s) = C(s, t) = \mathbb{E}\{f(s)\overline{f(t)}\}.$$

We call  $C$ , which is now a function of one parameter only, non-negative definite if  $\sum_{i,j=1}^n z_i C(t_i - t_j) \overline{z_j} \geq 0$  for all  $n \geq 1$ ,  $t_1, \dots, t_n \in \mathbb{R}^N$  and  $z_1, \dots, z_n \in \mathbb{C}$ . Then we have the following result, which dates back to Bochner [12], in the setting of (non-stochastic) Fourier analysis, a proof of which can be found in almost any text on Fourier analysis.

**Theorem 1.4.3 (Spectral distribution theorem)** *A continuous function  $C : \mathbb{R}^N \rightarrow \mathcal{C}$  is non-negative definite (i.e. a covariance function) if and only if there exists a finite measure  $\nu$  on  $\mathbb{B}^N$  such that*

$$(1.4.16) \quad C(t) = \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} \nu(d\lambda),$$

for all  $t \in \mathbb{R}^N$ .

With randomness in mind, we write  $\sigma^2 = C(0) = \nu(\mathbb{R}^N)$ . The measure  $\nu$  is called the *spectral measure* (for  $C$ ) and the function  $F : \mathbb{R}^N \rightarrow [0, \sigma^2]$  given by

$$F(\lambda) \triangleq \nu \left( \prod_{i=1}^N (-\infty, \lambda_i] \right), \quad \lambda = (\lambda_1, \dots, \lambda_N) \in \mathbb{R}^N,$$

is called the *spectral distribution function*<sup>20</sup>. When  $F$  is absolutely continuous the corresponding density is called the *spectral density*.

The spectral distribution theorem is a purely analytic result and would have nothing to do with random fields were it not for the fact that covariance functions are non-negative definite. Understanding of the result comes from the spectral representation theorem (Theorem 1.4.4) for which we need some preliminaries.

Let  $\nu$  be a measure on  $\mathbb{R}^N$  and  $W_R$  and  $W_I$  be two *independent* ( $\nu/\sqrt{2}$ )-noises, so that (1.4.4)–(1.4.6) hold for both  $W_R$  and  $W_I$  with  $\nu/\sqrt{2}$  rather than  $\nu$ . (The factor of  $1/\sqrt{2}$  is so that (1.4.18) below does not need a unaesthetic factor of 2 on the right hand side.). There is no need at the moment to assume that these noises are Gaussian. Define a new,  $\mathcal{C}$ -valued noise  $W$  by writing

$$W(A) \triangleq W_R(A) + iW_I(A),$$

for all  $A \in \mathbb{B}^N$ . Since  $E\{W(A)\overline{W(B)}\} = \nu(A \cap B)$ , we call  $W$  a *complex  $\nu$ -noise*. For  $f : \mathbb{R}^N \rightarrow \mathcal{C}$  with  $\|f\| \in L^2(\nu)$  we can now define the complex integral  $W(f)$  by writing

$$(1.4.17) \quad \begin{aligned} W(f) &\equiv \int_{\mathbb{R}^N} f(\lambda) W(d\lambda) \\ &\equiv \int_{\mathbb{R}^N} (f_R(\lambda) + if_I(\lambda)) (W_R(d\lambda) + iW_I(d\lambda)) \\ &\triangleq [W_R(f_R) - W_I(f_I)] + i[W_I(f_R) + W_R(f_I)] \end{aligned}$$

---

<sup>20</sup>Of course, unless  $\nu$  is a probability measure, so that  $\sigma^2 = 1$ ,  $F$  is not a distribution function in the usual usage of the term.

and the terms in the last line are all well defined as in Subsection 1.4.1. From the above and (1.4.11) it is trivial to check that

$$(1.4.18) \quad \mathbb{E}\{W(f)\overline{W(g)}\} = \int_{\mathbb{R}^N} f(\lambda)\overline{g(\lambda)}\nu(d\lambda)$$

for  $\mathcal{C}$ -valued  $f, g \in L^2(\nu)$ .

**Theorem 1.4.4 (Spectral representation theorem)** *Let  $\nu$  be a finite measure on  $\mathbb{R}^N$  and  $W$  a complex  $\nu$ -noise. Then the complex valued random field*

$$(1.4.19) \quad f(t) = \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} W(d\lambda)$$

has covariance

$$(1.4.20) \quad C(s, t) = \int_{\mathbb{R}^N} e^{i\langle (s-t), \lambda \rangle} \nu(d\lambda)$$

and so is (weakly) stationary. If  $W$  is Gaussian, then so is  $f$ .

Furthermore, to every mean square continuous, centered, (Gaussian) stationary random field  $f$  on  $\mathbb{R}^N$  with covariance function  $C$  and spectral measure  $\nu$  (cf. Theorem 1.4.3) there corresponds a complex (Gaussian)  $\nu$ -noise  $W$  on  $\mathbb{R}^N$  such that (1.4.19) holds in mean square for each  $t \in \mathbb{R}^N$ .

In both cases,  $W$  is called the spectral process corresponding to  $f$ .

**Proof.** The fact that (1.4.19) generates a stationary field with covariance (1.4.20) is an immediate consequence of the definition (1.4.17) and the relationship (1.4.11). What needs to be proven is the statement that *all* stationary fields can be represented as in (1.4.19). We shall only sketch the basic idea of the proof, leaving the details to the reader. (They can be found in almost any book on time series – our favourite is [16] – for processes on either  $\mathbb{Z}$  or  $\mathbb{R}$  and the extension to  $\mathbb{R}^N$  is trivial.)

For the first step, set up a mapping  $\Theta$  from<sup>21</sup>  $\mathcal{H} \triangleq \text{span}\{f_t, t \in \mathbb{R}^N\} \subset L^2(\mathbb{P})$  to  $\mathcal{K} \triangleq \text{span}\{e^{it\cdot}, t \in \mathbb{R}^N\} \subset L^2(\nu)$  via

$$(1.4.21) \quad \Theta \left( \sum_{j=1}^n a_j f(t_j) \right) = \sum_{j=1}^n a_j e^{it_j\cdot}$$

for all  $n \geq 1$ ,  $a_j \in \mathcal{C}$  and  $t_j \in \mathbb{R}^N$ . A simple computation shows that this gives an isomorphism between  $\mathcal{H}$  and  $\mathcal{K}$ , which can then be extended to an isomorphism between their closures  $\overline{\mathcal{H}}$  and  $\overline{\mathcal{K}}$ .

<sup>21</sup>  $\mathcal{H}$  is the closure in  $L^2(\mathbb{P})$  of all sums of the form  $\sum_{i=1}^n a_i f(t_i)$  for  $a_i \in \mathbb{C}$  and  $t_i \in T$ , thinned out by identifying all elements indistinguishable in  $L^2(\mathbb{P})$ : i.e. elements  $U, V$  for which  $\mathbb{E}\{\|(U - V)\|^2\} = 0$ .

Since indicator functions are in  $\overline{\mathcal{K}}$ , we can define a process  $W$  on  $\mathbb{R}^N$  by setting

$$(1.4.22) \quad W(\lambda) = \Theta^{-1}(\mathbb{1}_{(-\infty, \lambda]}),$$

where  $(-\infty, \lambda] = \prod_{j=1}^N (-\infty, \lambda_j]$ . Working through the isomorphisms shows that  $W$  is a complex  $\nu$ -noise and that  $\int_{\mathbb{R}^N} \exp(i\langle t, \lambda \rangle) W(d\lambda)$  is ( $L^2(\mathbb{P})$  indistinguishable from)  $f_t$ .  $\square$

There is also an inverse<sup>22</sup> to (1.4.19), expressing  $W$  as an integral involving  $f$ , but we shall have no need of it and so now turn to some consequences of Theorems 1.4.3 and 1.4.4.

When the basic field  $f$  is real, it is natural to expect a ‘real’ spectral representation, and this is in fact the case, although notationally it is still generally more convenient to use the complex formulation. Note firstly that if  $f$  is real, then the covariance function is a symmetric function ( $C(t) = C(-t)$ ) and so it follows from the spectral distribution theorem (cf. (1.4.16)) that the spectral measure  $\nu$  must also be symmetric. Introduce three<sup>23</sup> new measures, on  $\mathbb{R}_+ \times \mathbb{R}^{N-1}$ , by

$$\begin{aligned} \nu_1(A) &= \nu(A \cap \{\lambda \in \mathbb{R}^N : \lambda_1 > 0\}) \\ \nu_2(A) &= \nu(A \cap \{\lambda \in \mathbb{R}^N : \lambda_1 = 0\}) \\ \mu(A) &= 2\nu_1(A) + \nu_2(A) \end{aligned}$$

We can now rewrite<sup>24</sup> (1.4.16) in real form, as

$$(1.4.23) \quad C(t) = \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \cos(\langle \lambda, t \rangle) \mu(d\lambda).$$

There is also a corresponding real form of the spectral representation (1.4.19). The fact that the spectral representation yields a real valued processes also implies certain symmetries<sup>25</sup> on the spectral process  $W$ . In particular, it turns out that there are two independent *real* valued  $\mu$ -noises,

<sup>22</sup>Formally, the inverse relationship is based on (1.4.22), but it behaves like regular Fourier inversion. For example, if  $\Delta(\lambda, \eta)$  is a rectangle in  $\mathbb{R}^N$  which has a boundary of zero  $\nu$  measure, then

$$W(\Delta(\lambda, \eta)) = \lim_{K \rightarrow \infty} (2\pi)^{-N} \int_{-K}^K \dots \int_{-K}^K \prod_{k=1}^N \frac{e^{-i\lambda_k t_k} - e^{-i\eta_k t_k}}{-it_k} f(t) dt.$$

As is usual, if  $\nu(\Delta(\lambda, \eta)) \neq 0$ , then additional boundary terms need to be added.

<sup>23</sup>Note that if  $\nu$  is absolutely continuous with respect to Lebesgue measure (so that there is a spectral density) one of these,  $\nu_2$ , will be identically zero.

<sup>24</sup>There is nothing special about the half-space  $\lambda_1 \geq 0$  taken in this representation. Any half space will do.

<sup>25</sup>To rigorously establish this we really need the inverse to (1.4.19), expressing  $W$  as an integral involving  $f$ , which we do not have.

$W_1$  and  $W_2$ , such that<sup>26</sup>

$$(1.4.24) \quad f_t = \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \cos(\langle \lambda, t \rangle) W_1(d\lambda) \\ + \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \sin(\langle \lambda, t \rangle) W_2(d\lambda).$$

It is easy to check that  $f$  so defined has the right covariance function.

The real representation goes a long way to helping one develop a good understanding of what the spectral representation theorem says, and so we devote a few paragraphs on this. While it is not necessary for the rest of the book, it does help develop intuition.

One way to think of the integral in (1.4.24) is via the approximating sum

$$(1.4.25) \quad \sum_i \{ \cos(\langle \lambda_i, t \rangle) W_1(\Lambda_i) + \sin(\langle \lambda_i, t \rangle) W_2(\Lambda_i) \}$$

where the  $\{\Lambda_i\}$  give a partition of  $\mathbb{R}_+ \times \mathbb{R}^{N-1}$  and  $\lambda_i \in \Lambda_i$ . Indeed, this sum will be exact if the spectral measure is discrete with atoms  $\lambda_i$ . In either case, what (1.4.25) does is to express the random field as the sum of a large number of ‘sinusoidal’ components.

In the one-dimensional situation the basic components in (1.4.25) are simple sine and cosine waves of (random) amplitudes  $|W_2(\Lambda_i)|$  and  $|W_1(\Lambda_i)|$ , respectively, and wavelengths equal to  $2\pi/\lambda_i$ . In higher dimensions the elementary components are slightly harder to visualize. Consider the two-dimensional case. Dropping the subscript on  $\lambda_i$  for the moment, we have that an elementary cosine wave is of the form  $\cos(\lambda_1 t_1 + \lambda_2 t_2)$ . The  $\lambda_i$  are fixed and the point  $(t_1, t_2)$  ranges over  $\mathbb{R}^2$ . This gives a sequence of waves travelling in a direction which makes an angle

$$(1.4.26) \quad \theta = \arctan(\lambda_2/\lambda_1)$$

with the  $t_1$  axis and having the wavelength

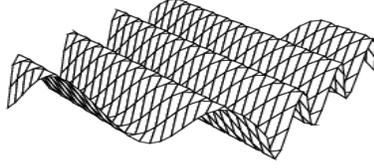
$$(1.4.27) \quad \lambda = \frac{2\pi}{\sqrt{\lambda_1^2 + \lambda_2^2}}$$

as the distance between troughs or crests, as measured along the line perpendicular to the crests. An example is given in Figure 1.4.3.

The corresponding sine function is exactly the same, except that its crests lie on the top of the troughs of the cosine function and vice versa. That

---

<sup>26</sup>In one dimension, it is customary to take  $W_1$  as a  $\mu$ -noise and  $W_2$  as a  $(2\nu_1)$ -noise, which at first glance is different to what we have. However, noting that, when  $N = 1$ ,  $\sin(\lambda t)W_2(d\lambda) = 0$  when  $\lambda = 0$ , it is clear that the two definitions in fact coincide in this case.

FIGURE 1.4.1. The elementary wave form  $\cos(\lambda_1 t_1 + \lambda_2 t_2)$  in  $\mathbb{R}^2$ .

is, the two sets of waves are out of phase by half a wavelength. As in the one-dimensional case, the amplitudes of the components  $\cos(\langle \lambda_i, t \rangle)$  and  $\sin(\langle \lambda_i, t \rangle)$  are given by the random variables  $|W_1(\Lambda_i)|$  and  $|W_2(\Lambda_i)|$ . Figure 1.4.3 shows what a sum of 10 such components looks like, when the  $\lambda_i$  are chosen randomly in  $(-\pi, \pi]^2$  and the  $W_j(\lambda_i)$  are independent  $N(0, 1)$ .

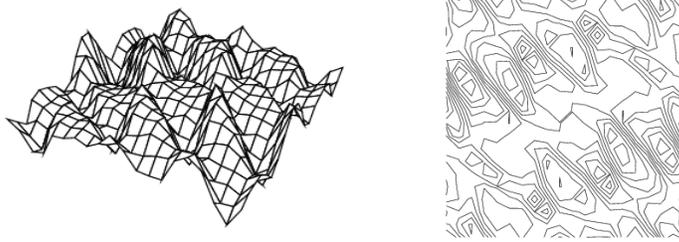


FIGURE 1.4.2. A more realistic surface based on (1.4.25), along with contour lines at the zero level.

#### 1.4.4 Spectral moments

Since they will be very important later on, we now take a closer look at spectral measures and, in particular, their moments. It turns out that these contain a lot of simple, but very useful, information. Given the spectral representation (1.4.20); viz.

$$(1.4.28) \quad C(t) = \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} \nu(d\lambda),$$

we define the *spectral moments*

$$(1.4.29) \quad \lambda_{i_1 \dots i_N} \triangleq \int_{\mathbb{R}^N} \lambda_1^{i_1} \cdots \lambda_N^{i_N} \nu(d\lambda),$$

for all  $(i_1, \dots, i_N)$  with  $i_j \geq 0$ . Recalling that stationarity implies that  $C(t) = C(-t)$  and  $\nu(A) = \nu(-A)$ , it follows that the odd ordered spectral

moments, when they exist, are zero; i.e.

$$(1.4.30) \quad \lambda_{i_1 \dots i_N} = 0, \quad \text{if } \sum_{j=1}^N i_j \text{ is odd.}$$

Furthermore, it is immediate from (1.4.28) that successive differentiation of both sides with respect to the  $t_i$  connects the various partial derivatives of  $C$  at zero with the spectral moments. To see why this is important, we need first to define the  $L^2$ , or *mean square (partial) derivatives* of a random field.

Choose a point  $t \in \mathbb{R}^N$  and a sequence of  $k$  'directions'  $t'_1, \dots, t'_k$  of  $\mathbb{R}^N$ , so that  $t' = (t'_1, \dots, t'_k) \in \otimes^k \mathbb{R}^N$ , the tensor product  $\mathbb{R}^N \times \dots \times \mathbb{R}^N$ . We say that  $f$  has a  $k$ -th order derivative at  $t$ , in the direction  $t'$ , which we denote by  $D_{L^2}^k f(t, t')$ , if the limit

$$(1.4.31) \quad D_{L^2}^k f(t, t') \triangleq \lim_{h \rightarrow 0} F(t, ht')$$

exists in  $L^2$ , where where  $F(t, t')$  is the symmetrized difference

$$F(t, t') = \frac{1}{\prod_{i=1}^k |t'_i|} \sum_{s \in \{0,1\}^k} (-1)^{k - \sum_{i=1}^k s_i} f\left(t + \sum_{i=1}^k s_i t'_i\right).$$

A simple sufficient condition for  $L^2$  differentiability of order  $k$  in all directions and throughout a region  $T \in \mathbb{R}^N$  is that

$$(1.4.32) \quad \lim_{|t'|, |s'| \rightarrow 0} \mathbb{E} \{F(t, t') F(s, s')\}$$

exists<sup>27</sup> for all  $s, t \in T$  and sequences  $s', t' \in \otimes^k \mathbb{R}^N$ . Note that if  $f$  is Gaussian then so are its  $L^2$  derivatives, when they exist.

By choosing  $t' = (e_{i_1}, \dots, e_{i_k})$ , where  $e_i$  is the vector with  $i$ -th element 1 and all others zero, we can talk of the mean square derivatives

$$\frac{\partial^k}{\partial t_{i_1} \dots \partial t_{i_k}} f(t) \triangleq F(t, (e_{i_1}, \dots, e_{i_k}))$$

of  $f$  of various orders.

It is then straightforward to see that the the covariance function of such partial derivatives must be given by

$$(1.4.33) \quad \mathbb{E} \left\{ \frac{\partial^k f(s)}{\partial t_{i_1} \partial t_{i_1} \dots \partial t_{i_k}} \frac{\partial^k f(t)}{\partial t_{i_1} \partial t_{i_1} \dots \partial t_{i_k}} \right\} = \frac{\partial^{2k} C(s, t)}{\partial s_{i_1} \partial t_{i_1} \dots \partial s_{i_k} \partial t_{i_k}}.$$

---

<sup>27</sup>This is an immediate consequence of the fact that a sequence  $X_n$  of random variables converges in  $L^2$  if, and only if,  $\mathbb{E}\{X_n X_m\}$  converges to a constant as  $n, m \rightarrow \infty$ .

The corresponding variances have a nice interpretation in terms of spectral moments when  $f$  is stationary. For example, if  $f$  has mean square partial derivatives of orders  $\alpha + \beta$  and  $\gamma + \delta$  for  $\alpha, \beta, \gamma, \delta \in \{0, 1, 2, \dots\}$ , then<sup>28</sup>

$$(1.4.34) \quad \mathbb{E} \left\{ \frac{\partial^{\alpha+\beta} f(t)}{\partial^{\alpha} t_i \partial^{\beta} t_j} \cdot \frac{\partial^{\gamma+\delta} f(t)}{\partial^{\gamma} t_k \partial^{\delta} t_l} \right\} = (-1)^{\alpha+\beta} \frac{\partial^{\alpha+\beta+\gamma+\delta}}{\partial^{\alpha} t_i \partial^{\beta} t_j \partial^{\gamma} t_k \partial^{\delta} t_l} C(t) \Big|_{t=0} \\ = (-1)^{\alpha+\beta} \int_{\mathbb{R}^N} \lambda_i^{\alpha} \lambda_j^{\beta} \lambda_k^{\gamma} \lambda_l^{\delta} \nu(d\lambda).$$

Here are some important special cases of the above, for which we adopt the shorthand  $f_j = \partial f / \partial t_j$  and  $f_{ij} = \partial^2 f / \partial t_i \partial t_j$  along with a corresponding shorthand for the partial derivatives of  $C$ .

- (i)  $f_j$  has covariance function  $-C_{jj}$  and thus variance  $\lambda_{2e_j} = -C_{jj}(0)$ , where  $e_j$ , as usual, is the vector with a 1 in the  $j$ -th position and zero elsewhere.
- (ii) In view of (1.4.30), and taking  $\beta = \gamma = \delta = 0$ ,  $\alpha = 1$  in (1.4.34)

$$(1.4.35) \quad f(t) \text{ and } f_j(t) \text{ are uncorrelated,}$$

for all  $j$  and all  $t$ . If  $f$  is Gaussian, this is equivalent to independence. Note that (1.4.35) does *not* imply that  $f$  and  $f_j$  are uncorrelated *as processes*. In general, for  $s \neq t$ , we will have that  $\mathbb{E}\{f(s)f_j(t)\} = -C_j(s-t) \neq 0$ .

- (iii) Taking  $\alpha = \gamma = \delta = 1$ ,  $\beta = 0$  in (1.4.34) gives that

$$(1.4.36) \quad f_i(t) \text{ and } f_{jk}(t) \text{ are uncorrelated}$$

for all  $i, j, k$  and all  $t$ .

This concludes our initial discussion of stationarity for random fields on  $\mathbb{R}^N$ . In the following Section we investigate what happens under slightly weaker conditions and after that what happens when the additional assumption of isotropy is added. In Section 1.4.7 we shall see how all of this is really part of a far more general theory for fields defined on groups. In Section 2.5 we shall see that, at least for Gaussian fields restricted to bounded regions, that spectral theory is closely related to a far wider theory of orthogonal expansions that works for general – i.e. not necessarily Euclidean – parameter spaces

---

<sup>28</sup>If you decide to check this for yourself using (1.4.19) and (1.4.20) – which is a worthwhile exercise – make certain that you recall the fact that the covariance function is defined as  $\mathbb{E}\{f(s)\overline{f(t)}\}$ , or you will make the same mistake RJA did in [1] and forget the factor of  $(-1)^{\alpha+\beta}$  in the first line. Also, note that although (1.4.34) seems to have some asymmetries in the powers, these disappear due to the fact that all odd ordered spectral moments, like all odd ordered derivatives of  $C$ , are identically zero.

### 1.4.5 Constant variance

It will be important for us in later that some of the relationships of the previous Section continue to hold under a weaker condition than stationarity. Of particular interest is knowing when (1.4.35) holds; i.e. when  $f(t)$  and  $f_j(t)$  are uncorrelated.

Suppose that  $f$  has constant variance,  $\sigma^2 = C(t, t)$ , throughout its domain of definition, and that its  $L^2$  first-order derivatives all exist. In this case, analogously to (1.4.34), we have that

$$(1.4.37) \quad \begin{aligned} \mathbb{E}\{f(t) f_j(t)\} &= \left. \frac{\partial}{\partial t_j} C(t, s) \right|_{s=t} \\ &= \left. \frac{\partial}{\partial s_j} C(t, s) \right|_{s=t}. \end{aligned}$$

Since constant variance implies that  $\partial/\partial t_j C(t, t) \equiv 0$ , the above two equalities imply that both partial derivatives there must be identically zero. That is,  $f$  and its first order derivatives are uncorrelated.

One can, of course, continue in this fashion. If first derivatives have constant variance, then they, in turn, will be uncorrelated with second derivatives, in the sense that  $f_i$  will be uncorrelated with  $f_{ij}$  for all  $i, j$ . It will not necessarily be true, however, that  $f_i$  and  $f_{jk}$  will be uncorrelated if  $i \neq j$  and  $i \neq k$ . This will, however, be true if the covariance matrix of all first order derivatives is constant.

### 1.4.6 Isotropy

An interesting special class of homogeneous random fields on  $\mathbb{R}^N$  that often arises in applications in which there is no special meaning attached to the coordinate system being used is the class of *isotropic* fields. These are characterized<sup>29</sup> by the property that the covariance function depends only on the Euclidean length  $|t|$  of the vector  $t$  so that

$$(1.4.38) \quad C(t) = C(|t|).$$

Isotropy has a number of surprisingly varied implications for both the covariance and spectral distribution functions, and is actually some more limiting than it might at first seem. For example, we have the following result, due to Matérn [67].

**Theorem 1.4.5** *If  $C(t)$  is the covariance function of a centered, isotropic random field  $f$  on  $\mathbb{R}^N$ , then  $C(t) \geq -C(0)/N$  for all  $t$ . Consequently,  $C$  is never negative.*

---

<sup>29</sup>Isotropy can also be defined in the non-stationary case, the defining property then being that the distribution of  $f$  is invariant under rotation. Under stationarity, this is equivalent to (1.4.38). Without stationarity, however, this definition does not imply (1.4.38). We shall treat isotropy *only* in the scenario of stationarity.

**Proof.** Isotropy implies that  $C$  can be written as a function on  $\mathbb{R}_+$  only. Let  $\tau$  be any positive real. We shall show that  $C(\tau) \geq -C(0)/N$ .

Choose any  $t_1, \dots, t_{N+1}$  in  $\mathbb{R}^N$  for which  $|t_i - t_j| = \tau$  for all  $i \neq j$ . Then, by (1.4.38),

$$\mathbb{E} \left\{ \left| \sum_{k=1}^{N+1} X(t_k) \right|^2 \right\} = (N+1)[C(0) + C(\tau)].$$

Since this must be positive, the result follows.  $\square$

The restriction of isotropy also has significant simplifying consequences for the spectral measure  $\nu$  of (1.4.16). Let  $\theta : \mathbb{R}^N \rightarrow \mathbb{R}^N$  be a rotation, so that  $|\theta(t)| = |t|$  for all  $t$ . Isotropy then implies  $C(t) = C(\theta(t))$  and so the Spectral Distribution Theorem implies

$$\begin{aligned} (1.4.39) \quad \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} \nu(d\lambda) &= \int_{\mathbb{R}^N} e^{i\langle \theta(t), \lambda \rangle} \nu(d\lambda) \\ &= \int_{\mathbb{R}^N} e^{i\langle t, \theta(\lambda) \rangle} \nu(d\lambda) \\ &= \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} \nu_\theta(d\lambda), \end{aligned}$$

where  $\nu_\theta$  is the push-forward of  $\nu$  by  $\theta$  defined by  $\nu_\theta(A) = \nu(\theta^{-1}A)$ . Since the above holds for all  $t$  it follows that  $\nu \equiv \nu_\theta$ ; i.e.  $\nu$ , like  $C$ , is invariant under rotation. Furthermore, if  $\nu$  is absolutely continuous, then its density is also dependent only on the modulus of its argument.

An interesting consequence of this symmetry is that an isotropic field cannot have all the probability of its spectral measure concentrated in one small region in  $\mathbb{R}^N$  away from the origin. In particular, it is not possible to have a spectral measure degenerate at one point, unless that point is the origin. The closest the spectral measure of an isotropic field can come to this sort of behaviour is to have all its probability concentrated in an annulus of the form

$$\{\lambda \in \mathbb{R}^N : a \leq |\lambda| \leq b\}.$$

In such a case it is clear from (1.4.26) and (1.4.27) that the field itself is then composed of a ‘sum’ of waves travelling in all directions but with wavelengths between  $2\pi/b$  and  $2\pi/a$  only.

Another consequence of isotropy is the spherical symmetry of the spectral measure significantly simplifies the structure of the spectral moments and so the correlations between various derivatives of  $f$ . In particular, it follows immediately from (1.4.34) that

$$(1.4.40) \quad \mathbb{E}\{f_i(t)f_j(t)\} = -\mathbb{E}\{f(t)f_{ij}(t)\} = \lambda_2\delta_{ij}$$

where  $\delta_{ij}$  is the Kronecker delta and  $\lambda_2 \triangleq \int_{\mathbb{R}^N} \lambda_i^2 \nu(d\lambda)$ , which is independent of the value of  $i$ . Consequently, if  $f$  is Gaussian, then the first order derivatives of  $f$  are independent of one another, as they are of  $f$  itself.

Since isotropy has such a limiting effect in the spectrum, it is natural to ask how the spectral distribution and representation theorems are affected under isotropy. The following result, due originally to Schoenberg [85] (in a somewhat different setting) and Yaglom [109], describes what happens.

**Theorem 1.4.6** *For  $C$  to be the covariance function of a mean square continuous, isotropic, random field on  $\mathbb{R}^N$  it is necessary and sufficient that*

$$(1.4.41) \quad C(t) = \int_0^\infty \frac{J_{(N-2)/2}(\lambda|t|)}{(\lambda|t|)^{(N-2)/2}} \mu(d\lambda),$$

where  $\mu$  is a finite measure on  $\mathbb{R}_+$  and  $J_m$  is the Bessel function of the first kind of order  $m$ ; viz.

$$J_m(x) = \sum_{k=0}^{\infty} (-1)^k \frac{(x/2)^{2k+m}}{k! \Gamma(k+m+1)}.$$

**Proof.** The proof consists in simplifying the basic spectral representation by using the symmetry properties of  $\nu$ .

We commence by converting to polar coordinates,  $(\lambda, \theta_1, \dots, \theta_{N-1})$ ,  $\lambda \geq 0$ ,  $(\theta_1, \dots, \theta_{N-1}) \in S^{N-1}$ , where  $S^{N-1}$  is the unit sphere in  $\mathbb{R}^N$ . Define a measure  $\mu$  on  $\mathbb{R}_+$  by setting  $\mu([0, \lambda]) = \nu(B^N(\lambda))$ , and extending as usual, where  $B^N(\lambda)$  is the  $N$ -ball of radius  $\lambda$  and  $\nu$  is the spectral measure of (1.4.16).

Then, on substituting into (1.4.16) with  $t = (|t|, 0, \dots, 0)$  and performing the coordinate transformation, we obtain

$$C(|t|) = \int_0^\infty \int_{S^{N-1}} \exp(i|t|\lambda \cos \theta_{N-1}) \sigma(d\theta) \mu(d\lambda)$$

where  $\sigma$  is surface area measure on  $S^{N-1}$ . Integrating out  $\theta_1, \dots, \theta_{N-2}$  it follows that

$$C(|t|) = s_{N-2} \int_0^\infty \int_0^\pi e^{i\lambda|t| \cos \theta_{N-1}} (\sin \theta_{N-1})^{N-2} d\theta_{N-1} \mu(d\lambda)$$

where

$$(1.4.42) \quad s_N \triangleq \frac{2\pi^{N/2}}{\Gamma(N/2)}, \quad N \geq 1,$$

is the surface area<sup>30</sup> of  $S^{N-1}$ .

<sup>30</sup>When  $N = 1$ , we are thinking of the ‘boundary’ of the ‘unit sphere’  $[-1, 1] \in \mathbb{R}$ . This is made up of the two points  $\pm 1$ , which, in counting measure, has measure 2. Hence  $s_1 = 2$  makes sense.

The inside integral can be evaluated in terms of Bessel functions to yield

$$\int_0^\pi e^{i\lambda|t|\cos\theta} \sin^{N-2}\theta \, d\theta = \frac{J_{(N-2)/2}(\lambda|t|)}{(\lambda|t|)^{(N-2)/2}}$$

which, on absorbing  $s_N$  into  $\mu$ , completes the proof.  $\square$

For small values of the dimension  $N$ , (1.4.41) can be simplified even further. For example, substituting  $N = 2$  into (1.4.41) yields that in this case

$$C(t) = \int_0^\infty J_0(\lambda|t|) \mu(d\lambda),$$

while substituting  $N = 3$  and evaluating the inner integral easily yields that in this case

$$C(t) = \int_0^\infty \frac{\sin(\lambda|t|)}{\lambda|t|} \mu(d\lambda).$$

Given the fact that the covariance function of an isotropic field takes such a special form, it is natural to seek a corresponding form for the spectral representation of the field itself. Such a representation does in fact exist and we shall now describe it, albeit without giving any proofs. These can be found, for example, in the book by Wong [103], or as special cases in the review by Yaglom [110], which is described in Section 1.4.7 below. Another way to verify it would be to check that the representation given in Theorem 1.4.7 below yields the covariance structure of (1.4.41). Since this is essentially an exercise in the manipulation of special functions and not of intrinsic probabilistic interest, we shall avoid the temptation to carry it out.

The spectral representation of isotropic fields on  $\mathbb{R}^N$  is based on the so-called *spherical harmonics*<sup>31</sup> on the  $(N-1)$ -sphere, which form an orthonormal basis for the space of square integrable functions on  $S^{N-1}$  equipped with the usual surface measure. We shall denote them by  $\{h_{ml}^{(N)}, l = 1, \dots, d_m, m = 0, 1, \dots\}$  where the  $d_m$  are known combinatorial coefficients<sup>32</sup>.

---

<sup>31</sup>We shall also avoid giving details about spherical harmonics. A brief treatment would add little to understanding them. The kind of treatment required to, for example, get the code correct in programming a simulation of an isotropic field using the representations that follow will, in any case, send you back to the basic reference of Erdélyi [33] followed by some patience in sorting out a software help reference. A quick web search will yield you many interactive, coloured examples of these functions within seconds.

<sup>32</sup>The spherical harmonics on  $S^{N-1}$  are often written as  $\{h_{m, l_1, \dots, l_{N-2}, \pm l_{N-1}}^{(N)}\}$ , where  $0 \leq l_{N-1} \leq \dots \leq l_1 \leq m$ . The constants  $d_m$  in our representation can be computed from this.

Now use the spectral decomposition

$$f(t) = \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} W(d\lambda)$$

to define a family of noises on  $\mathbb{R}_+$  by

$$W_{ml}(A) = \int_A \int_{S^{N-1}} h_{ml}^{(N-1)}(\theta) W(d\lambda, d\theta)$$

where, once again, we work in polar coordinates. Note that since  $W$  is a  $\nu$ -noise, where  $\nu$  is the spectral measure, information about the covariance of  $f$  has been coded into the  $W_{ml}$ . From this family, define a family of mutually uncorrelated, stationary, one-dimensional processes  $\{f_{ml}\}$  by

$$f_{ml}(r) = \int_0^\infty \frac{J_{m+(N-2)/2}(\lambda r)}{(\lambda r)^{(N-2)/2}} W_{ml}(d\lambda),$$

where, as in the spectral representation (1.4.19), one has to justify the existence of this  $L^2$  stochastic integral. These are all the components we need in order to state the following.

**Theorem 1.4.7** *A centered, mean square continuous, isotropic random field on  $\mathbb{R}^N$  can be represented by*

$$(1.4.43) \quad f(t) = f(r, \theta) = \sum_{m=0}^{\infty} \sum_{l=1}^{d_m} f_{ml}(r) h_{ml}^{(N-1)}(\theta).$$

In other words, isotropic random fields can be decomposed into a countable number of mutually uncorrelated stationary processes with a one-dimensional parameter, a result which one would not intuitively expect. As noted above, there is still a hidden spectral process in (1.4.43), entering via the  $W_{ml}$  and  $f_{ml}$ . This makes for an important difference between (1.4.43) and the similar looking Karhunen-Loève expansion which we shall meet in Section 2.5.1.

An interesting corollary of Theorem 1.4.7 is obtained by fixing  $r$  in (1.4.43). We then have a simple representation in terms of uncorrelated random coefficients  $f_{ml}(r)$  and spherical harmonics of an isotropic random field on the  $N$ -sphere of radius  $r$ . If the random field is Gaussian, then the coefficients are actually independent, and we will, essentially, have generated a Karhunen-Loève expansion.

One can keep playing these games for more and more special cases. For example, it is not uncommon in applications to find random fields that are functions of ‘space’  $x$  and ‘time’  $t$ , so that the parameter set is most conveniently written as  $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^N$ . Such processes are often stationary in  $t$  and isotropic in  $x$ , in the sense that

$$\mathbb{E}\{f(s, u) \overline{f(s+t, u+x)}\} = C(t, |x|),$$

where  $C$  is now a function from  $\mathbb{R}^2$  to  $\mathbb{C}$ . In such a situation the methods of the previous proof suffice to show that  $C$  can be written in the form

$$C(t, x) = \int_{-\infty}^{\infty} \int_0^{\infty} e^{it\nu} G_N(\lambda x) \mu(d\nu, d\lambda),$$

where

$$G_N(x) = \left(\frac{2}{x}\right)^{(N-2)/2} \Gamma\left(\frac{N}{2}\right) J_{(N-2)/2}(x)$$

and  $\mu$  is a measure on the half-plane  $\mathbb{R}_+ \times \mathbb{R}^N$ .

By now it should be starting to become evident that all of these representations must be special cases of some general theory, that might also be able to cover non-Euclidean parameter spaces. This is indeed the case, although for reasons that will soon be explained the general theory is such that, ultimately, each special case requires almost individual treatment.

#### 1.4.7 Stationarity over groups

We have already seen in Section 1.4.2 that the appropriate setting for stationarity is when the parameter set has a group structure. In this case it made sense, in general, to talk about left and right stationarity (cf. (1.4.13) and (1.4.14)). Simple ‘stationarity’ requires both of these and so makes most sense if the group is Abelian (commutative).

In essence, the spectral representation of a random field over a group is intimately related to the representation theory of the group. This, of course, is far from being a simple subject. Furthermore, its level of difficulty depends very much on the group in question and so it is correspondingly not easy to give a general spectral theory for random fields over groups. The most general results in this area are in the paper by Yaglom [110] already mentioned above and the remainder of this Subsection is taken from there<sup>33</sup>.

We shall make life simpler by assuming for the rest of this Subsection that  $T$  is a *locally compact, Abelian* (LCA) group. As before, we shall denote the binary group operation by  $+$  while  $-$  denotes inversion. The Fourier analysis of LCA groups is well developed (e.g. [80]) and based on *characters*. A homomorphism  $\gamma$  from  $T$  to the multiplicative group of complex numbers is called a character if  $\|\gamma(t)\| = 1$  for all  $t \in T$  and if

$$\gamma(t + s) = \gamma(t) \gamma(s), \quad s, t \in T.$$

---

<sup>33</sup>There is also a very readable, albeit less exhaustive, treatment in Hannan [43]. In addition, Letac [60] has an elegant exposition based on Gelfand pairs and Banach algebras for processes indexed by unimodular groups, which, in a certain sense, give a generalisation of isotropic fields over  $\mathbb{R}^N$ . Ylisen [111] has a theory for noncommutative locally compact groups that extends the results in [110].

If  $T = \mathbb{R}^N$  under the usual addition, then the characters are given by the family

$$\left\{ \gamma_\lambda(t) = e^{i\langle t, \lambda \rangle} \right\}_{\lambda \in \mathbb{R}^N}$$

of complex exponential functions, which were at the core of the spectral theory of fields over  $\mathbb{R}^N$ . If  $T = \mathbb{Z}^N$ , again under addition, the characters are as for  $T = \mathbb{R}^N$ , but  $\lambda$  is restricted to  $[-\pi, \pi]^N$ . If  $T = \mathbb{R}^N$  under rotation rather than addition, then the characters are the spherical harmonics on  $S^{N-1}$ .

The set of all continuous characters also forms a group,  $\Gamma$  say, called the *dual group* with composition defined by

$$(\gamma_1 + \gamma_2)(t) = \gamma_1(t) \gamma_2(t).$$

There is also a natural topology on  $\Gamma$  (cf. [80]) which gives  $\Gamma$  a LCA structure and under which the map  $(\gamma, t) \stackrel{\Delta}{=} \gamma(t) : \Gamma \times T \rightarrow \mathcal{C}$  is continuous. The spectral distribution theorem in this setting can now be written as

$$(1.4.44) \quad C(t) = \int_{\Gamma} (\gamma, t) \nu(d\gamma),$$

where the finite spectral measure  $\nu$  is on the  $\sigma$ -algebra generated by the topology on  $\Gamma$ . The spectral representation theorem can be correspondingly written as

$$(1.4.45) \quad f(t) = \int_{\Gamma} (\gamma, t) W(d\gamma),$$

where  $W$  is a  $\nu$ -noise on  $\Gamma$ .

Special cases now follow from basic group theoretic results. For example, if  $T$  is discrete, then  $\Gamma$  is compact, as we noted already for the special case of  $T = \mathbb{Z}^N$ . Consequently, the integral in the spectral representation (1.4.45) is actually a sum and as such will be familiar to every student of Time Series. Alternatively, if  $T$  is compact,  $\Gamma$  is discrete. This implies that  $\nu$  must be a sum of point masses and  $W$  is actually a discrete collection of independent random variables.

An interesting and important special case that we have already met is  $T = S^{N-1}$ . Treating  $T$  as both an  $N$ -sphere and the rotation group  $O(N)$  we write the sum  $\theta_1 + \theta_2$ , for  $\theta_1, \theta_2 \in S^{N-1}$ , for the rotation of  $\theta_1$  by  $\theta_2$ , the latter considered as an element of  $O(N)$ . In this case the characters are again the spherical harmonics and we have that

$$(1.4.46) \quad f(\theta) = \sum_{m=0}^{\infty} \sum_{l=1}^{d_m} W_{ml} h_{ml}^{(N-1)}(\theta)$$

where the  $W_{ml}$  are uncorrelated with variance depending only on  $m$ . This, of course, is simply (1.4.43) once again, derived from a more general setting. Similarly, the covariance function can be written as

$$(1.4.47) \quad C(\theta_1, \theta_2) = C(\theta_{12}) = \sum_{m=0}^{\infty} \sigma_m^2 C_m^{(N-1)/2}(\cos(\theta_{12})),$$

where  $\theta_{12}$  is the angular distance between  $\theta_1$  and  $\theta_2$ , and the  $C_m^N$  are the Gegenbauer polynomials.

Other examples for the LCA situation follow in a similar fashion from (1.4.44) and (1.4.45) by knowing the structure of the dual group  $\Gamma$ .

The general situation is much harder and, as has already been noted, relies heavily on knowing the representation of  $T$ . In essence, given a representation of  $G$  on a  $GL(H)$  for some Hilbert space  $H$ , one constructs a (left or right) stationary random field on  $G$  via the canonical white noise on  $H$ . The construction of Lemma 1.4.2 with  $T = G$  and  $H = L^2(G, \mu)$  can be thought of as a special example of this approach. For further details, you should go to the references we gave at the beginning of this Section.

## 1.5 Non-Gaussian fields

This section should be of interest to most readers and is crucial for those who care about applications,

Up until now, we have concentrated very heavily on Gaussian random fields. The one point where we departed somewhat from this theme was in the discussion on stationarity, where normality played a very limited rôle. In the following Chapter we shall concentrate exclusively on Gaussian fields.

Despite, and perhaps because of, this it is time to take a moment to explain both the centrality of Gaussian fields and how to best move away from them.

It will become clear as you progress through this book that while appeals to the Central Limit Theorem may be a nice way to justify concentrating on the Gaussian case, the real reason for this concentration is somewhat more mundane. The relatively uncomplicated form of the multivariate Gaussian density (and hence finite-dimensional distributions of Gaussian fields) makes it a reasonably straightforward task to carry out detailed computations and allows one to obtain explicit results and precise formulae for many facets of Gaussian fields. It is difficult to over-emphasise the importance of explicit results for applications. There is a widespread belief among modern pure mathematicians that the major contribution they have to make to ‘Science’ is the development of ‘Understanding’, generally at the expense of explicit results. Strangely enough, most subject matter scientists do not share the mathematicians’ enthusiasm for insight. They generally

know their subject well enough to develop their own insight. However, useful formulae are quite a different issue. Consequently, Gaussian fields are extremely important.

Nevertheless, there is clearly need for tools beyond the Gaussian in the modeller's box of tricks and there are two ways to go about this. The first is to somehow use the Gaussian theory as a building block for other theories. The second is to start from scratch. We describe the former first, since it will be more relevant for us.

Given a real-valued Gaussian field  $g$ , the easiest way to generate something non-Gaussian is to take a pointwise transformation; viz. to study processes of the form  $f(t) = F(g(t))$ , where  $F : \mathbb{R} \rightarrow \mathbb{R}$  has whatever smoothness properties required to carry the smoothness (e.g. continuity) of  $g$  to  $f$ . This is such a simple transformation that only very rarely does it require any serious additional analysis.

Of far more interest is a family of fields that have arisen in a wide variety of statistical applications, that involve the classical distributions of Statistics, in particular  $\chi^2$ ,  $F$  and  $T$ . To see how these work, let  $g : T \rightarrow \mathbb{R}^d$  be a centered vector valued Gaussian field, with *independent, identically distributed, constant variance* coordinate processes  $g^1, \dots, g^d$ . Take  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  and define

$$(1.5.1) \quad f(t) = F(g(t)) = F(g^1(t), \dots, g^d(t)).$$

When  $d = 1$  this is the transformation that we just treated with such disdain in the last paragraph. However, when  $d > 1$  this leads to some very interesting examples indeed. Here are three, for all of which  $\sigma^2 = \mathbb{E}\{(g^i(t))^2\}$ :

- (i)  $\chi^2$  field: Take  $F(x) = \sum_1^d x_i^2$ . Then the corresponding random field is always positive and has marginal distribution that of a  $\chi^2$  random variable, viz.

$$(1.5.2) \quad \varphi_{\chi^2}(x) \triangleq \frac{x^{(n-2)/2} e^{-x/2\sigma^2}}{\sigma^n 2^{n/2} \Gamma(n/2)}, \quad x \geq 0.$$

Consequently, it is called the ' $\chi^2$  random field' with  $d$  degrees of freedom.

- (ii)  $T$  field: Now take

$$F(x) = \frac{x_1 \sqrt{d-1}}{(\sum_2^d x_i^2)^{1/2}}.$$

The corresponding random field is known as the  $T$  field with  $d-1$  degrees of freedom and has marginal density, for  $x \in \mathbb{R}$ , given by

$$(1.5.3) \quad \varphi_T(x) \triangleq \frac{\Gamma(d/2)}{(2^{d-1} \pi (d-1))^{1/2} \Gamma((d-1)/2)} \left(1 + \frac{x^2}{d-1}\right)^{-d/2}$$

which is independent of  $\sigma^2$ .

(iii) *F field*: Take  $d = n + m$  and

$$F(x) = \frac{m \sum_1^n x_i^2}{n \sum_{n+1}^{n+m} x_i^2}.$$

The corresponding random field is known as the *F field* with  $n$  and  $m$  degrees of freedom and has marginal density, for  $x > 0$ , given by

$$(1.5.4) \quad \varphi_F(x) \triangleq \frac{n^{n/2} m^{m/2}}{B(n/2, m/2)} \frac{x^{n/2-1}}{(m + nx)^{(n+m)/2}},$$

where  $B$  is the usual Beta function.

Consider the issue of stationarity for each of these three examples. From their definitions, we could evaluate their covariance functions and check them for weak,  $L^2$  stationarity. However, this would be rather foolish, since the strong stationarity of their Gaussian components clearly implies strong stationarity for them as well. Sample path smoothness of various degrees will also follow immediately from any assumed smoothness on the paths of  $g$ , since the transformations  $F$  are  $C^\infty$  everywhere except at the origin<sup>34</sup>. Thus it should be reasonably clear that elementary properties of  $f$  pass over quite simply to  $f = F(g)$  as long as  $F$  is ‘nice’.

A more interesting question for us will be how to study the excursion sets (1.1.1) of  $f$ . There are two ways to develop this theory. In the past, the standard approach was to treat each particular  $F$  as a special case and to handle it accordingly. This invariably involved detailed computations, which were always related to the underlying Gaussian structure of  $g$  and to the specific transformation  $F$ . This approach led to a plethora of papers, which provided not only a nice theory but also a goodly number of PhD theses, promotions and tenure successes. In JT’s thesis [94] (see also [95]) another approach was taken, based, in essence, on the observation that

$$(1.5.5) \quad \begin{aligned} A_u(f, T) = A_u(F(g), T) &= \{t \in T : (F \circ g)(t) \geq u\} \\ &= \{t \in T : g(t) \in F^{-1}[u, \infty)\}. \end{aligned}$$

Thus, the excursion set of a *real valued* non-Gaussian  $f = F \circ g$  above a level  $u$  is equivalent to the excursion set for a *vector valued* Gaussian  $g$  in the manifold  $F^{-1}[u, \infty)$ . We shall see in Chapter 5 how this approach can be used to separate out probability computations from geometric computations based on the properties of  $F$  and so obtain an elegant theory for this non-Gaussian scenario.

---

<sup>34</sup>The pole at the origin will actually cause problems in some cases, such as  $T = \mathbb{R}^N$  and  $d = N$  in the  $\chi^2$  and  $T$  cases, as well as  $m = N$  in the  $F$  case, but we can worry about that later.

In other words, the reader interested in the non-Gaussian case can rest assured for the moment that the Gaussian emphasis of the next Chapter will not be too limiting, since it will ultimately be relevant in a far wider scenario.

An important question is what to do with excursion sets when dealing with a field that has no Gaussian basis. We have no satisfactory answer here, since, we know of no ‘purely’ non-Gaussian random field for which excursion sets have been successfully studied<sup>35</sup>. While this is undesirable, it is probably to be expected, since the sad fact is that even in the case of processes on  $\mathbb{R}$  very little is known in this generality. We therefore leave it as a (probably very hard) challenge for future generations.

---

<sup>35</sup>There is a partial exception to this statement, in that smoothed Poisson fields have been studied in [77]. The theory there, however, is nowhere near as rich as in the Gaussian case.



# 2

## Gaussian fields

The aim of this Chapter is to provide a basic coverage of the modern theory of Gaussian random fields on general parameter spaces. There will be no attempt to be exhaustive. There are now many books covering various aspects of this theory, including those by Bogachev [13], Dudley [28], Fernique [36], Hida and Hitsuda [44], Janson [48], Ledoux and Talagrand [57], Lifshits [61] and Piterbarg [75]. In terms of what will be important to us, [28] and [57] stand out from the pack, perhaps augmented with Talagrand’s review [92]. Finally, while not as exhaustive<sup>1</sup> as the others, you might find RA’s lecture notes [2], augmented with the corrections in Section 2.3 below, a user-friendly introduction to the subject.

There are four main theoretical results which will be of major importance to us. The first is encapsulated in various versions in Theorems 2.1.3 and 2.1.5 and their Corollaries. These give a sufficient condition, in terms of metric entropy, ensuring the sample path boundedness and continuity of a Gaussian process and provide information about moduli of continuity. While this entropy condition is also necessary for *stationary* fields, this is not the case in general, and so for completeness we look briefly at the “majorising measure” version of this theory in Section 2.6. However, it will be a rare reader of *this* book who will ever need the more general theory.

To put the seemingly abstract entropy conditions into focus, they will be immediately followed by a Section with a goodly number of extremely varied examples. Nevertheless, these cover only the tip of a very large iceberg.

---

<sup>1</sup>Nor, perhaps, as exhausting.

Their diversity shows the power of the abstract approach, in that all can be treated via the general theory without further probabilistic arguments. The reader who is not interested in the general Gaussian theory, and cares mainly about the geometry of fields on  $\mathbb{R}^N$ , need only read Sections 2.2.1 and 2.2.2 on continuity and differentiability in this scenario.

The next two important results are the Borell-TIS inequality and Slepian's inequality (and its newer relatives) in Sections 2.3 and 2.4 respectively. The Borell-TIS inequality gives a universal bound for the tail probability

$$\mathbb{P}\{\sup_{t \in T} f(t) \geq u\},$$

$u > 0$ , for *any* centered, continuous Gaussian field. As such, it is a truly basic tool of Gaussian processes, somewhat akin to Chebychev's inequality in Statistics or maximal inequalities in Martingale Theory. Slepian's inequality and its relatives are just as important and basic, and allow one to use relationships between covariance functions of Gaussian fields to compare the tail probabilities and expectations of their suprema.

The final major result of this Chapter is encapsulated in Theorem 2.5.1, which gives an expansion for a Gaussian field in terms of deterministic eigenfunctions with independent  $N(0, 1)$  coefficients. A special case of this expansion is the Karhunen-Loève expansion of Section 2.5.1, with which many readers will already be familiar. Together with the spectral representations of Section 1.4, they make up what are probably the most important tools in the Gaussian modeller's box of tricks. However, these expansions are also an extremely important theoretical tool, whose development has far reaching consequences.

## 2.1 Boundedness and continuity

The aim of this Section is to develop a useful *sufficient* condition for a centered Gaussian field on a parameter space  $T$  to be almost surely bounded and/or continuous; i.e. to determine conditions for which

$$\mathbb{P}\{\sup_{t \in T} |f(t)| < \infty\} = 1 \quad \text{or} \quad \mathbb{P}\{\lim_{s \rightarrow t} |f(t) - f(s)| = 0, \forall t \in T\} = 1.$$

Of course, in order to talk about continuity – i.e. for the notation  $s \rightarrow t$  above to have some meaning – it is necessary that  $T$  have some topology, so we assume that  $(T, \tau)$  is a metric space, and that continuity is in terms of the  $\tau$ -topology. Our first step is to show that  $\tau$  is irrelevant to the question of continuity<sup>2</sup>. This is rather useful, since we shall also soon show that

---

<sup>2</sup>However,  $\tau$  will come back into the picture when we talk about moduli of continuity later in this Section.

boundedness and continuity are essentially the same problem for Gaussian fields, and formulating the boundedness question requires no topological demands on  $T$ .

To start, define a new metric  $d$  on  $T$  by

$$(2.1.1) \quad d(s, t) \triangleq \left\{ \mathbb{E} \left[ (f(s) - f(t))^2 \right] \right\}^{\frac{1}{2}},$$

in a notation that will henceforth remain fixed. Actually,  $d$  is only a pseudo-metric, since although it satisfies all the other demands of a metric,  $d(s, t) = 0$  does not necessarily<sup>3</sup> imply that  $s = t$ . Nevertheless, we shall abuse terminology by calling  $d$  the *canonical metric* for  $T$  and/or  $f$ .

It will be convenient for us to always assume that  $d$  is continuous in the  $\tau$ -topology and we shall indeed do so, since in the environment of  $f$  continuity that will interest us,  $d$  continuity costs us nothing. To see this, suppose that  $\sup_T \mathbb{E}\{f_t^2\} < \infty$ , and that  $f$  is a.s. continuous. Then

$$\begin{aligned} \lim_{s \rightarrow t} d^2(s, t) &= \lim_{s \rightarrow t} \mathbb{E} \left\{ (f(s) - f(t))^2 \right\} \\ &= \mathbb{E} \left\{ \lim_{s \rightarrow t} (f(s) - f(t))^2 \right\} \\ &= 0, \end{aligned}$$

the exchange of expectation and integral coming from uniform integrability. In other words, a.s. continuity of  $f$  implies the continuity of  $d$ .

Here is the Lemma establishing the irrelevance of  $\tau$  to the continuity question.

**Lemma 2.1.1** *Let  $f$  be a centered Gaussian process on a compact metric space  $(T, \tau)$ . Then  $f$  is a.s. continuous with respect to the  $\tau$ -topology if, and only if, it is a.s. continuous with respect to the  $d$  (pseudo) topology. More precisely, with probability one, for all  $t \in T$ ,*

$$\lim_{s: \tau(s, t) \rightarrow 0} |f(s) - f(t)| = 0 \iff \lim_{s: d(s, t) \rightarrow 0} |f(s) - f(t)| = 0$$

**Proof.** Since  $d$  is (always) assumed continuous in the  $\tau$  topology, it is immediate that if  $f$  is  $d$ -continuous then it is  $\tau$ -continuous.

Suppose, therefore, that  $f$  is  $\tau$ -continuous. For  $\eta \geq 0$ , let

$$A_\eta = \{(s, t) \in T \times T : d(s, t) \leq \eta\}.$$

Since  $d$  is continuous, this is a  $\tau$ -closed subset of  $T \times T$ . Furthermore,  $\bigcap_{\eta > 0} A_\eta = A_0$ . Fix  $\varepsilon > 0$ . Then, by the  $\tau$ -compactness of  $T$ , there is a

---

<sup>3</sup>For a counter-example, think of a periodic process on  $\mathbb{R}$ , with period  $p$ . Then  $d(s, t) = 0$  implies no more than  $s - t = kp$  for some  $k \in \mathbb{Z}$ .

finite set  $B \subset A_0$  (the number of whose elements will in general depend on  $\varepsilon$ ) such that

$$\bigcup_{(s',t') \in B} \{(s,t) \in T \times T : \max(\tau(s,s'), \tau(t,t')) \leq \varepsilon\}$$

covers  $A_\eta$  for some  $\eta = \eta(\varepsilon) > 0$ , with  $\eta(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . That is, whenever  $(s,t) \in A_\eta$  there is a  $(s',t') \in B$  with  $\tau(s,s'), \tau(t,t') \leq \varepsilon$ . Note that

$$|f_t - f_s| \leq |f_s - f_{s'}| + |f_{s'} - f_{t'}| + |f_{t'} - f_t|.$$

Since  $(s',t') \in B \subset A_0$ , we have  $f(s') = f(t')$  a.s. Thus

$$\sup_{d(s,t) \leq \eta(\varepsilon)} |f_t - f_s| \leq 2 \sup_{\tau(s,t) \leq \varepsilon} |f_t - f_s|,$$

and the  $\tau$ -continuity of  $f$  implies its  $d$ -continuity.  $\square$

The astute reader will have noted that in the statement of Lemma 2.1.1 the parameter space  $T$  was quietly assumed to be compact, and that this additional assumption was needed in the proof. Indeed, from now on we shall assume that this is always the case, and shall rely on it heavily. Fortunately, however, it is not a serious problem. As far as continuity is concerned, if  $T$  is  $\sigma$ -compact<sup>4</sup> then a.s. continuity on its compact subsets immediately implies a.s. continuity over  $T$  itself. We shall not go beyond  $\sigma$ -compact spaces in this book. The same is not true for boundedness, nor should it be<sup>5</sup>. However, we shall see that, at least on compact  $T$ , boundedness and continuity are equivalent problems.

Now recall the Brownian noise processes from Section 1.3. There we saw that the *same* process could be continuous, or discontinuous, depending on how we specified its parameter set. We shall now see that the difference between the parameter sets had nothing to do with their geometrical properties, but rather their “size” as measured in terms of  $d$ , via the tool of metric entropy.

**Definition 2.1.2** *Let  $f$  be a centered Gaussian field on  $T$ , and  $d$  the canonical metric (2.1.1). Assume that  $T$  is  $d$ -compact, and write*

$$(2.1.2) \quad B_d(t, \varepsilon) \triangleq \{s \in T : d(s, t) \leq \varepsilon\}$$

for the  $d$  ball centered on  $t \in T$  and of radius  $\varepsilon$ . Let  $N(T, d : \varepsilon) \equiv N(\varepsilon)$  denote the smallest number of such balls that cover  $T$ , and set

$$(2.1.3) \quad H(T, d : \varepsilon) \equiv H(\varepsilon) = \ln(N(\varepsilon)).$$

<sup>4</sup> $T$  is  $\sigma$ -compact if it can be represented as the countable union of compact sets.

<sup>5</sup>Think of simple Brownian motion on  $\mathbb{R}_+$ . While bounded on every finite interval, it is a consequence of the law of the iterated logarithm that it is unbounded on  $\mathbb{R}_+$ .

Then  $H$  is called the metric entropy function for  $T$  (or  $f$ ). We shall refer to any condition or result based on  $N$  or  $H$  as an entropy condition/result.

Note that since we are assuming that  $T$  is  $d$ -compact, it follows that  $H(\varepsilon) < \infty$  for all  $\varepsilon > 0$ . The same need not be (nor generally is) true for  $\lim_{\varepsilon \rightarrow 0} H(\varepsilon)$ . Furthermore, note for later use that if we define

$$(2.1.4) \quad \text{diam}(T) = \sup_{s,t \in T} d(s,t),$$

then  $N(\varepsilon) = 1$  and so  $H(\varepsilon) = 0$  for  $\varepsilon \geq \text{diam}(T)$ .

Here then are the main results of this Section, all of which will be proven soon.

**Theorem 2.1.3** *Let  $f$  be a centered Gaussian field on a  $d$ -compact  $T$ ,  $d$  the canonical metric, and  $H$  the corresponding entropy. Then there exists a universal constant  $K$  such that*

$$(2.1.5) \quad \mathbb{E} \left\{ \sup_{t \in T} f_t \right\} \leq K \int_0^{\text{diam}(T)/2} H^{1/2}(\varepsilon) d\varepsilon.$$

This result has immediate consequences for continuity. Define the *modulus of continuity*  $\omega_F$  of a real-valued function  $F$  on a metric space  $(T, \tau)$  as

$$(2.1.6) \quad \omega_F(\delta) \equiv \omega_{F,\tau}(\delta) \triangleq \sup_{\tau(s,t) \leq \delta} |F(t) - F(s)|, \quad \delta > 0.$$

The modulus of continuity of  $f$  can be thought of as the supremum of the random field  $f_{s,t} = f_t - f_s$  over a certain neighbourhood of  $T \times T$ , in that

$$(2.1.7) \quad \omega_{f,\tau}(\delta) = \sup_{\substack{(s,t) \in T \times T \\ \tau(s,t) \leq \delta}} f(s,t).$$

(Note we can drop the absolute value sign since the supremum here is always non-negative.)

More precisely, write  $d^{(2)}$  for the canonical metric of  $f_{s,t}$  on  $T \times T$ . Then

$$\begin{aligned} d^{(2)}((s,t), (s',t')) &= \left[ \mathbb{E} \{ (f_t - f_s) - (f_{t'} - f_{s'})^2 \} \right]^{1/2} \\ &\leq 2 \max(d(s,t), d(s',t')), \end{aligned}$$

and so

$$N(T \times T, d^{(2)} : \delta) \leq N(T, d : \delta/2).$$

From these observations, it is immediate that Theorem 2.1.3 implies

**Corollary 2.1.4** *Under the conditions of Theorem 2.1.3 there exists a universal constant  $K$  such that*

$$(2.1.8) \quad \mathbb{E} \{ \omega_{f,d}(\delta) \} \leq K \int_0^\delta H^{1/2}(\varepsilon) d\varepsilon.$$

Note that this is not quite enough to establish the a.s. continuity of  $f$ . Continuity is, however, not far away, since the same construction used to prove Theorem 2.1.3 will also give us the following, which, with the elementary tools we have at hand at the moment<sup>6</sup>, neither follows from, nor directly implies, (2.1.8).

**Theorem 2.1.5** *Under the conditions of Theorem 2.1.3 there exists a random  $\eta \in (0, \infty)$  and a universal constant  $K$  such that*

$$(2.1.9) \quad \omega_{f,d}(\delta) \leq K \int_0^\delta H^{1/2}(\varepsilon) d\varepsilon,$$

for all  $\delta < \eta$ .

Note that (2.1.9) is expressed in terms of the  $d$  modulus of continuity. Translating this to a result for the  $\tau$  modulus is trivial.

We shall see later that if  $f$  is stationary then the convergence of the entropy integral is also necessary for continuity and that continuity and boundedness always occur together (Theorem 2.6.4). Now, however, we shall prove Theorems 2.1.3 and 2.1.5 following the approach of Talagrand [92]. The original proof of Theorem 2.1.5 is due to Dudley [26], and, in fact, things have not really changed very much since then. Immediately following the proofs, in Section 2.2, we shall look at examples, to see how entropy arguments work in practice. You may want to skip to the examples before going through the proofs first time around.

We start with the following almost trivial, but important, observations.

**Observation 2.1.6** *If  $f$  is a separable process on  $T$  then  $\sup_{t \in T} f_t$  is a well defined (i.e. measurable) random variable.*

Measurability follows directly from Definition 1.1.3 of separability which gave us a countable dense subset  $D \subset T$  for which

$$\sup_{t \in T} f_t = \sup_{t \in D} f_t.$$

The supremum of a countable set of measurable random variables is always measurable.

One can actually manage without separability for the rest of this Section, in which case

$$\sup \left\{ \mathbb{E} \left[ \sup_{t \in F} f_t \right] : F \subset T, F \text{ finite} \right\}$$

---

<sup>6</sup>See, however Theorem ?? below, to see what one can do with better tools.

should be taken as the definition of  $\mathbb{E}\{\sup_T f_t\}$ .

**Observation 2.1.7** *If  $f$  is a separable process on  $T$  and  $X$  a centered random variable (not necessarily independent of  $f$ ), then*

$$\mathbb{E}\left\{\sup_{t \in T} (f_t + X)\right\} = \mathbb{E}\left\{\sup_{t \in T} f_t\right\}.$$

As trite as this observation is, it ceases to be valid if, for example, we investigate  $\sup_t |f_t + X|$  rather than  $\sup_t (f_t + X)$ .

**Proof of Theorem 2.1.3** Fix a point  $t_o \in T$  and consider  $f_t - f_{t_o}$ . In view of Observation 2.1.7, we can work with  $\mathbb{E}\{\sup_T (f_t - f_{t_o})\}$  rather than  $\mathbb{E}\{\sup_T f_t\}$ . Furthermore, in view of separability, it will suffice to take these suprema over the countable separating set  $D \subset T$ . To save on notation, we might therefore just as well assume that  $T$  is countable, which we now do.

We shall represent the difference  $f_t - f_{t_o}$  via a telescoping sum, in what is called a *chaining* argument, and is in essence an approximation technique. We shall keep track of the accuracy of the approximations via entropy and simple union bounds.

To build the approximations, first fix some  $r \geq 2$  and choose the largest  $i \in \mathbb{Z}$  such that  $\text{diam}(T) \leq r^{-i}$ , where the  $\text{diam}(T)$  is measured in terms of the canonical metric of  $f$ . For  $j > i$ , take a finite subset  $\Pi_j$  of  $T$  such that

$$\sup_{t \in T} \inf_{s \in \Pi_j} d(s, t) \leq r^{-j},$$

(possible by  $d$ -compactness) and define a mapping  $\pi_j : T \rightarrow \Pi_j$  satisfying

$$(2.1.10) \quad \sup_{t \in T} d(t, \pi_j(t)) \leq r^{-j}.$$

For consistency, set  $\Pi_i = \{t_o\}$  and  $\pi_i(t) = t_o$  for all  $t$ . Consistent with the notations of Definition 2.1.2, we can choose  $\Pi_j$  to have no more than  $N_j \triangleq N(r^{-j})$  points, and so entropy has now entered into the argument.

The idea of this construction is that the points  $\pi_j(t)$  are successive approximations to  $t$ , and that as we move along the ‘chain’  $\pi_j(t)$  we have the decomposition

$$(2.1.11) \quad f_t - f_{t_o} = \sum_{j>i} f_{\pi_j(t)} - f_{\pi_{j-1}(t)}.$$

We need to check that this potentially infinite sum is well defined, with probability one. For this, recall (1.2.2), which implies

$$(2.1.12) \quad \mathbb{P}\{X \geq u\} \leq e^{-u^2/2\sigma^2},$$

for  $X \sim N(0, \sigma^2)$  and  $u > 0$ .

By this and (2.1.10), it follows that

(2.1.13)

$$\begin{aligned} \mathbb{P} \left\{ |f_{\pi_j(t)} - f_{\pi_{j-1}(t)}| \geq (r/\sqrt{2})^{-j} \right\} &\leq \exp \left( \frac{-(r/\sqrt{2})^{-2j}}{2(2r^{-j+1})^2} \right) \\ &= \exp(-2^j/8r^2), \end{aligned}$$

which is eminently summable. By Borel-Cantelli, and recalling that  $r \geq 2$ , we have that the sum in (2.1.11) converges absolutely, with probability one.

We now start the main part of the proof. Define

$$M_j = N_j N_{j-1}, \quad a_j = 2^{3/2} r^{-j+1} \sqrt{\ln(2^{j-i} M_j)}, \quad S = \sum_{j>i} a_j.$$

Then  $M_j$  is the maximum number of possible pairs  $(\pi_j(t), \pi_{j-1}(t))$  as  $t$  varies through  $T$  and  $a_j$  was chosen so as to make later formulae simplify. Applying (2.1.12) once again, we have, for all  $u > 0$ ,

(2.1.14)

$$\mathbb{P} \left\{ \exists t \in T : f_{\pi_j(t)} - f_{\pi_{j-1}(t)} > u a_j \right\} \leq M_j \exp \left( \frac{-u^2 a_j^2}{2(2r^{-j+1})^2} \right)$$

and so

$$\begin{aligned} \mathbb{P} \left\{ \sup_{t \in T} f_t - f_{t_o} \geq u S \right\} &\leq \sum_{j>i} M_j \exp \left( \frac{-u^2 a_j^2}{2(2r^{-j+1})^2} \right) \\ &= \sum_{j>i} M_j (2^{j-i} M_j)^{-u^2}. \end{aligned}$$

For  $u > 1$  this is at most

$$\begin{aligned} \sum_{j>i} (2^{j-i})^{-u^2} &\leq 2^{-u^2} \sum_{j>i} 2^{j-i+1} \\ &= 2 \cdot 2^{-u^2}. \end{aligned}$$

The basic relationship that, for non-negative random variables  $X$ ,

$$\mathbb{E}\{X\} = \int_0^\infty \mathbb{P}\{X \geq u\} du,$$

together with the observation that  $\sup_{t \in T} (f_t - f_{t_o}) \geq 0$  since  $t_o \in T$ , immediately yields

$$(2.1.15) \quad \mathbb{E} \left\{ \sup_{t \in T} f_t \right\} \leq K S,$$

with  $K = 2 \int_0^\infty 2^{-u^2} du$ . Thus, all that remains to complete the proof is to compute  $S$ .

Using the definition of  $S$ , along with the elementary inequality that  $\sqrt{ab} \leq \sqrt{a} + \sqrt{b}$ , gives

$$\begin{aligned} S &\leq 2^{3/2} \sum_{j>i} r^{-j+1} \left( \sqrt{j-i} \sqrt{\ln 2} + \sqrt{\ln N_j} + \sqrt{\ln N_{j-1}} \right) \\ &\leq K \left( r^{-i} + \sum_{j \geq i} r^{-j} \sqrt{\ln N_j} \right) \\ &\leq K \sum_{j \geq i} r^{-j} \sqrt{\ln N_j}, \end{aligned}$$

where  $K$  is a constant that may change from line to line, but depends only on  $r$ . The last inequality follows from absorbing the lone term of  $r^{-i}$  into the second term of the sum, possible since the very definition of  $i$  implies that  $N_{i+1} \geq 2$ , and changing the multiplicative constant  $K$  accordingly.

Recalling now the definition of  $N_j$  as  $N(r^{-j})$ , we have that

$$\varepsilon \leq r^{-j} \Rightarrow N(\varepsilon) \geq N_j.$$

Thus

$$\begin{aligned} \int_0^{r^{-i}} \sqrt{\ln N(\varepsilon)} d\varepsilon &\geq \sum_{j \geq i} (r^{-j} - r^{-j-1}) \sqrt{\ln N_j} \\ &= K \sum_{j \geq i} r^{-j} \sqrt{\ln N_j}. \end{aligned}$$

Putting this together with the bound on  $S$  and substituting into (2.1.15) gives

$$\mathbb{E} \left\{ \sup_{t \in T} f_t \right\} \leq K \int_0^{r^{-i}} H^{1/2}(\varepsilon) d\varepsilon.$$

Finally, note that  $N(\varepsilon) = 1$  (and so  $H(\varepsilon) = 0$ ) for  $\varepsilon \geq 2r^{-i} \geq \text{diam}(T)$ , to establish (2.1.5) and so complete the proof.  $\square$

**Proof of Theorem 2.1.5** The proof starts with the same construction as in the proof of Theorem 2.1.3. Note that from the same principles behind the telescoping sum (2.1.11) defining  $f_t - f_{t_0}$  we have that for all  $s, t \in T$  and  $J > i$ ,

(2.1.16)

$$\begin{aligned} f_t - f_s &= f_{\pi_J(t)} - f_{\pi_J(s)} \\ &\quad + \sum_{j>J} [f_{\pi_j(t)} - f_{\pi_{j-1}(t)}] - \sum_{j>J} [f_{\pi_j(s)} - f_{\pi_{j-1}(s)}]. \end{aligned}$$

From (2.1.12) we have that for all  $u > 0$ ,

$$\mathbb{P}\{|f_{\pi_j(t)} - f_{\pi_j(s)}| \geq u d(\pi_j(t), \pi_j(s))\} \leq e^{-u^2/2}.$$

Arguing as we did to obtain (2.1.14), and the line or two following, we now see that

$$\begin{aligned} \mathbb{P}\left\{\exists s, t \in T : |f_{\pi_j(t)} - f_{\pi_j(s)}| \geq \sqrt{2} d(\pi_j(t), \pi_j(s)) \sqrt{\ln(2^{j-i} N_j^2)}\right\} \\ \leq 2^{i-j}. \end{aligned}$$

Since this is a summable series, Borel-Cantelli gives the existence of a random  $j_o > i$  for which, with probability one,

$$j > j_o \Rightarrow |f_{\pi_j(t)} - f_{\pi_j(s)}| \leq \sqrt{2} d(\pi_j(t), \pi_j(s)) \sqrt{\ln(2^{j-i} N_j^2)}$$

for all  $s, t \in T$ .

Essentially the same argument also gives that

$$j > j_o \Rightarrow |f_{\pi_j(t)} - f_{\pi_{j-1}(t)}| \leq \sqrt{2} d(\pi_j(t), \pi_{j-1}(t)) \sqrt{\ln(2^{j-i} M_j)}$$

for all  $t \in T$ .

Putting these into (2.1.16) gives that

$$\begin{aligned} |f_t - f_s| \leq K d(\pi_{j_o}(t), \pi_{j_o}(s)) \sqrt{\ln(2^{j_o-i} N_{j_o}^2)} \\ + K \sum_{j>j_o} d(\pi_j(t), \pi_{j-1}(t)) \sqrt{\ln(2^{j-i} M_j)}. \end{aligned}$$

Note that  $d(\pi_j(t), \pi_{j-1}(t)) \leq 2r^{-j+1}$  and

$$\begin{aligned} d(\pi_{j_o}(t), \pi_{j_o}(s)) &\leq d(s, t) + 2r^{-j_o} \\ &\leq 3r^{-j_o} \end{aligned}$$

if we take  $d(s, t) \leq \eta \triangleq r^{-j_o}$ . The above sums can be turned into integrals just as we did at the end of the previous proof, which leads to (2.1.9) and so completes the argument.  $\square$

Before leaving to look at some examples, you should note one rather crucial fact: The *only* Gaussian ingredient in the preceding two proofs was the basic inequality (2.1.12) giving  $\exp(-u^2/2)$  as a tail bound for a single  $N(0, 1)$  random variable. The remainder of the proof used little more than the union bound on probabilities and some clever juggling. Furthermore, it does not take a lot of effort to see that the square root in the entropy integrals such as (2.1.5) is related to ‘inverting’ the square in  $\exp(-u^2/2)$  while the logarithm comes from ‘inverting’ the exponential. If this makes you feel that there is a far more general, non-Gaussian theory behind all this, and that it is not going to be very different to the Gaussian one, then you are right. A brief explanation of how it works is in Section 2.2.5.

## 2.2 Examples

### 2.2.1 Fields on $\mathbb{R}^N$

Returning to Euclidean space after the abstraction of entropy on general metric spaces, it is natural to expect that conditions for continuity and boundedness will become so simple to both state and prove that there was really no need to introduce such abstruse general concepts.

This expectation is both true and false. It turns out that avoiding the notion of entropy does not make it any easier to establish continuity theorems, and, indeed, reliance on the specific geometry of the parameter space often confounds the basic issues. On the other hand, the following important result is easy to state without specifically referring to any abstract notions. To state it, let  $f_t$  be a centered Gaussian process on a compact  $T \subset \mathbb{R}^N$  and define

$$(2.2.1) \quad p^2(u) = \sup_{|s-t| \leq u} \mathbb{E} \{|f_s - f_t|^2\},$$

where  $|\cdot|$  is the usual Euclidean metric. If  $f$  is stationary, then

$$(2.2.2) \quad p^2(u) = 2 \sup_{|t| \leq u} [C(0) - C(t)].$$

**Theorem 2.2.1** *If, for some  $\delta > 0$ , either*

$$(2.2.3) \quad \int_0^\delta (-\ln u)^{\frac{1}{2}} dp(u) < \infty \quad \text{or} \quad \int_\delta^\infty p(e^{-u^2}) du < \infty,$$

*then  $f$  is continuous and bounded on  $T$  with probability one. A sufficient condition for either integral in (2.2.3) to be finite is that, for some  $0 < C < \infty$  and  $\alpha, \eta > 0$ ,*

$$(2.2.4) \quad \mathbb{E} \{|f_s - f_t|^2\} \leq \frac{C}{|\log |s - t||^{1+\alpha}},$$

*for all  $s, t$  with  $|s - t| < \eta$ . Furthermore, there exists a constant  $K$ , dependent only on the dimension  $N$ , and a random  $\delta_o > 0$  such that, for all  $\delta < \delta_o$ ,*

$$(2.2.5) \quad \omega_f(\delta) \leq K \int_0^{p(\delta)} (-\ln u)^{\frac{1}{2}} dp(u),$$

*where the modulus of continuity  $\omega_f$  is taken with respect to the Euclidean metric. A similar bound, in the spirit of (2.1.8), holds for  $\mathbb{E}\{\omega_f(\delta)\}$ .*

**Proof.** Note first that since  $p(u)$  is obviously non-decreasing in  $u$ , the Riemann-Stieljes integral (2.2.3) is well defined. The proof that both integrals in (2.2.3) converge and diverge together and that the convergence of

both is assured by (2.2.4) is simple calculus and left to the reader. Of more significance is relating these integrals to the entropy integrals of Theorems 2.1.3 and 2.1.5 and Corollary 2.1.4. Indeed, all the claims of the Theorem will follow from these results if we show that

$$\int_0^\delta H^{1/2}(\varepsilon) d\varepsilon \leq K \int_0^{p(\delta)} (-\ln u)^{\frac{1}{2}} dp(u)$$

for small enough  $\delta$ .

Since  $T$  is compact, we can enclose it in a  $N$ -cube  $C_L$  of side length  $L = \max_{i=1, \dots, N} \sup_{s, t \in T} |t_i - s_i|$ . Since  $p$  is non-decreasing, there is no problem in defining

$$p^{-1}(u) \triangleq \sup\{t : p(t) \leq u\}.$$

Now note that, for each  $\varepsilon > 0$ , the cube  $C_L$ , and so  $T$ , can be covered by  $[1 + L\sqrt{N}/(2p^{-1}(\varepsilon))]^N$  (Euclidean)  $N$ -balls, each of which has radius no more than  $\varepsilon$  in the canonical metric  $d$ . Thus,

$$\begin{aligned} \int_0^\delta H^{1/2}(\varepsilon) d\varepsilon &\leq \sqrt{N} \int_0^\delta \left( \ln(1 + L\sqrt{N}/(2p^{-1}(\varepsilon))) \right)^{\frac{1}{2}} d\varepsilon \\ &= \sqrt{N} \int_0^{p(\delta)} \left( \ln(1 + L\sqrt{N}/2u) \right)^{\frac{1}{2}} dp(u) \\ &\leq 2\sqrt{N} \int_0^{p(\delta)} (-\ln u)^{\frac{1}{2}} dp(u) \end{aligned}$$

for small enough  $\delta$ . This completes the proof.  $\square$

The various sufficient conditions for continuity of Theorem 2.2.1 are quite sharp, but not necessary. There are two stages at which necessity is lost. One is simply that entropy conditions, in general, need not be necessary in the non-stationary case. The second is that something is lost in the passage from entropy to the conditions on  $p$ .

As an example of the latter, take the continuous, centered Gaussian process  $f$  on  $\mathbb{R}$  with covariance function  $C(t) = \exp(-t^2/2)$  and, for  $t > 0$  define  $g(t) = f(\ln(1/t))$ . It is easy to check that  $f$  is a.s. continuous, as is  $g$ , since it is obtained from  $f$  via a continuous transformation. It is also easy to check that  $f$  and  $g$  have identical entropy functions. However, while the  $p$  function for  $f$  satisfies (2.2.3), this is not true for that of  $g$ .

Despite these drawbacks, the results of Theorem 2.2.1 are, from a practical point of view, reasonably definitive. For example, we shall see below (Corollary 2.6.5) that if  $f$  is stationary and

$$(2.2.6) \quad \frac{K_1}{(-\ln |t|)^{1+\alpha_1}} \leq C(0) - C(t) \leq \frac{K_2}{(-\ln |t|)^{1+\alpha_2}},$$

for  $|t|$  small enough, then  $f$  will be sample path continuous if  $\alpha_2 > 0$  and discontinuous if  $\alpha_1 < 0$ .

Before leaving the Euclidean case, it is also instructive to see, at least for the stationary case, how the above conditions on covariance functions translate to conditions on the spectral measure  $\nu$  of (1.4.16). The translation is via standard Tauberian theory, which translates the behaviour of  $C$  at the origin to that of  $\nu$  at infinity. (cf. for example, [11]). A typical result is the following, again in the centered, Gaussian case: If the integral

$$\int_0^\infty (\log(1 + \lambda))^{1+\alpha} \nu(d\lambda)$$

converges for some  $\alpha > 0$  then  $f$  is a.s. continuous, while if it diverges for some  $\alpha < 0$  then  $f$  is a.s. discontinuous.

In other words, it is the “high frequency oscillations” in the spectral representation that are controlling the continuity/discontinuity dichotomy. This is hardly surprising. What is perhaps somewhat more surprising, since we have seen that for Gaussian processes continuity and boundedness come together, is that it is these same oscillations that are controlling boundedness as well.

Before leaving  $\mathbb{R}^N$  we have one debt left to fill: the proof of Theorem 1.3.2. We shall make it a Corollary to Theorem 2.2.1.

**Corollary 2.2.2** *The point and rectangle indexed Brownian sheets are continuous over compact  $T \subset \mathbb{R}^N$ .*

**Proof.** We need only consider the point indexed sheet, since by (1.3.7) its continuity immediately implies that of the rectangle indexed version. Furthermore, we lose nothing by taking  $T = [0, 1]^N$ . Thus, consider

$$d(s, t) = \mathbb{E} \{ |W(s) - W(t)|^2 \}$$

for  $s, t \in [0, 1]^N$ . We shall show that

$$(2.2.7) \quad d(s, t) \leq 2N |t - s|$$

from which it follows, in the notation of (2.2.1), that  $p^2(u) \leq 2Nu$ . Since  $\int_1^\infty e^{-u^2/2} du < \infty$ , Theorem 2.2.1 (cf. (2.2.3)) immediately yields the continuity and boundedness of  $W$ .

To establish (2.2.7) write  $u \vee v$  for  $\max(u, v)$  and note that

$$d(s, t) \leq 2 \prod_{i=1}^N (s_i \vee t_i) - 2 \prod_{i=1}^N (s_i \wedge t_i).$$

Set  $a = 2 \prod_{i=1}^{N-1} (s_i \vee t_i)$  and  $b = 2 \prod_{i=1}^{N-1} (s_i \wedge t_i)$ . Then  $2 \geq a > b$  and

$$d(s, t) \leq a(s_N \vee t_N) - b(s_N \wedge t_N).$$

If  $s_N > t_N$  the right-hand side is equal to

$$\begin{aligned} as_N - bt_N &= a(s_N - t_N) + t_N(a - b) \\ &\leq 2|s_N - t_N| + |a - b|. \end{aligned}$$

Similarly, if  $s_N < t_N$  the right-hand side equals

$$\begin{aligned} at_N - bs_N &= a(t_N - s_N) + s_N(a - b) \\ &\leq 2|t_N - s_N| + |a - b|, \end{aligned}$$

so that

$$d(s, t) \leq 2|t_N - s_N| + 2 \sum_{i=1}^{N-1} (s_i \vee t_i) - 2 \sum_{i=1}^{N-1} (s_i \wedge t_i).$$

Continuing this process yields

$$d(s, t) \leq 2 \sum_{i=1}^N |t_i - s_i| \leq 2N|t - s|,$$

which establishes (2.2.7) and so the Corollary.  $\square$

### 2.2.2 Differentiability on $\mathbb{R}^N$

We shall stay with Euclidean  $T \subset \mathbb{R}^N$  for the moment, and look at the question of a.s. differentiability of centered, Gaussian  $f$ . We have already considered the issue of  $L^2$  differentiability in Section 1.4.3. There, for  $(t, t') \in T \times \oplus^k \mathbb{R}^N$ , we defined the  $k$ -th order derivative in the direction  $t'$  as the  $L^2$  limit

$$D_{L^2}^k f(t, t') = \lim_{h \rightarrow 0} F(t, ht'),$$

where  $F(t, t')$  is the symmetrized difference

$$F(t, t') = \frac{1}{\prod_{i=1}^k |t'_i|} \sum_{s \in \{0,1\}^k} (-1)^{k - \sum_{i=1}^k s_i} f\left(t + \sum_{i=1}^k s_i t'_i\right).$$

We also noted there that when  $D_{L^2}^k f$  exists, it is a Gaussian field on  $T \times \oplus^k \mathbb{R}^n$ , since  $L^2$  limits of Gaussian variables are always Gaussian.

While  $L^2$  existence was fine for what we needed in Chapter 1, later on we shall need to know when a.s. derivatives exist, and whether or not they are a.s. continuous. The general structure that we have so far built actually makes this a very simple question to answer.

To see why, endow the space  $\mathbb{R}^N \times \oplus^k \mathbb{R}^N$  with the norm

$$\|(s, s')\|_{n,k} \triangleq |s| + \|s'\|_{\oplus^k \mathbb{R}^N} = |s| + \left( \sum_{i=1}^k |s'_i|^2 \right)^{1/2},$$

and write  $B_{n,k}(y, h)$  for the  $h$ -ball centered at  $y = (t, t')$  in the metric induced by  $\|\cdot\|_{n,k}$ . Furthermore, write

$$T_{k,\rho} \triangleq T \times \{t' : \|t'\|_{\oplus^k \mathbb{R}^N} \in (1 - \rho, 1 + \rho)\}$$

for the product of  $T$  with the  $\rho$ -tube around the unit sphere in  $\oplus^k \mathbb{R}^N$ . This is enough to allow us to formulate

**Theorem 2.2.3** *Suppose  $f$  is a centered Gaussian random field on an open  $T \in \mathbb{R}^N$ , possessing  $k$ -th order derivatives in the  $L^2$  sense. Suppose, furthermore, that there exists  $0 < K < \infty$ , and  $\rho, \delta, h_0 > 0$  such that for  $0 < \eta_1, \eta_2, h < h_0$ ,*

$$(2.2.8) \quad \mathbb{E} \left\{ [F(t, \eta_1 t') - F(s, \eta_2 s')]^2 \right\} \\ < K (-\ln(\|(t, t') - (s, s')\|_{n,j} + |\eta_1 - \eta_2|))^{-(1+\delta)},$$

for all

$$((t, t'), (s, s')) \in T_{j,\rho} \times T_{j,\rho} : (s, s') \in B_{n,j}((t, t'), h).$$

Then, with probability one,  $f$  is  $k$  times continuously differentiable; viz.  $f \in C^k(T)$ .

**Proof.** Recalling that we have assumed the existence of  $L^2$  derivatives, we can define the Gaussian field

$$\widehat{F}(t, t', \eta) = \begin{cases} F(t, \eta t') & \eta \neq 0, \\ D_{L^2}^k f(t, t') & \eta = 0, \end{cases}$$

on  $\widehat{T} \triangleq T_{k,\rho} \times (-h, h)$ , an open subset of the finite dimensional vector space  $\mathbb{R}^N \times \oplus^k \mathbb{R}^N \times \mathbb{R}$  with norm

$$\|(t, t', \eta)\|_{n,j,1} = \|(t, t')\|_{n,j} + |\eta|.$$

Whether or not  $f \in C^k(T)$  is clearly the same issue as whether or not  $\widehat{F} \in C(\widehat{T})$ , with the issue of the continuity of  $f$  really only being on the hyperplane where  $\eta = 0$ . But this puts us back into the setting of Theorem 2.2.1, and it is easy to check that condition (2.2.4) there translates to (2.2.8) in the current scenario.  $\square$

### 2.2.3 Generalised fields

We start with an example: Take a centered, Gaussian random field  $f$  on  $\mathbb{R}^N$ , with covariance function  $C$ . Let  $\mathcal{F}$  be a family of functions on  $\mathbb{R}^N$ , and for  $\varphi \in \mathcal{F}$  define

$$(2.2.9) \quad f(\varphi) = \int_{\mathbb{R}^N} \varphi(t) f(t) dt.$$

We thus obtain a centered Gaussian process indexed by functions in  $\mathcal{F}$ , whose covariance functional is given by

$$(2.2.10) \quad C(\varphi, \psi) = \mathbb{E}\{f(\varphi)f(\psi)\} = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \varphi(s)C(s, t)\psi(t) dsdt.$$

The above construction only makes sense when  $C(t, t) < \infty$ , for otherwise  $f$  has infinite variance and the integrand in (2.2.9) is not defined. Nevertheless, there are occasions when (2.2.10) makes sense, even though  $C(t, t) = \infty$ . In this case we shall refer to  $C$  as a *covariance kernel*, rather than covariance function.

Indeed, we have already been in this scenario twice. In Section 1.4.2 we looked at moving averages of Gaussian  $\nu$ -noise, in which case  $\mathcal{F}$  was made up of translations of the form  $\varphi(s) = F(t - s)$ , for  $F \in L^2(\nu)$ . When treating the spectral representation of stationary processes, we took  $\mathcal{F}$  as the complex exponentials  $\exp(it \cdot \lambda)$  for fields on  $\mathbb{R}^N$  and as the family of characters in Section 1.4.3 for fields on a general group. Consider the basic spectral distribution theorem, Theorem 1.4.3, which was written in the setting of  $\mathcal{C}$ -valued fields. For simplicity, assume that  $\nu$  has a spectral density  $g$ . Then (1.4.16) gave us that stationary covariances over  $\mathbb{R}^N$  can be formally written as

$$\begin{aligned} C(s, t) &= \int_{\mathbb{R}^N} e^{i(t-s)\cdot\lambda} g(\lambda) d\lambda \\ &= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} e^{it\cdot\lambda_1} \left[ g^{1/2}(\lambda_1) \delta(\lambda_1, \lambda_2) g^{1/2}(\lambda_2) \right] e^{-is\cdot\lambda_2} d\lambda_1 d\lambda_2, \end{aligned}$$

which is in the form of (2.2.10), with  $\varphi = e^{it\cdot}$ ,  $\psi = e^{-is\cdot}$ . The covariance kernel in the integrand now involves the Dirac delta function, which is certainly not finite on the diagonal. Although it was never formally acknowledged as such, it was this issue that led to our having to be careful in defining the stochastic integral in the spectral representation of Theorem 1.4.4.

While moving averages and stationary processes afford two classes of examples, there are many more, some of which we shall describe at the end of this Subsection. In particular, given any positive definite function  $C$  on  $\mathbb{R}^N \times \mathbb{R}^N$ , not necessarily finite, one can define a *function indexed* process on

$$\mathcal{F}_C \triangleq \left\{ \varphi : \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \varphi(s)C(s, t)\varphi(t) dsdt < \infty \right\}.$$

The proof requires no more than checking that given such a  $C$  on  $\mathbb{R}^N \times \mathbb{R}^N$ , the corresponding  $C$  defined by (2.2.10) determines a *finite* positive definite, and so covariance, function on  $\mathcal{F}_C \times \mathcal{F}_C$ .

In general, function indexed processes of this kind, for which the covariance kernel of (2.2.10) is infinite on the diagonal, are known as *generalised random fields*<sup>7</sup>.

The question that we shall now look at is when such processes are continuous and bounded. The answer involves a considerable amount of work, but it is worthwhile at some stage to go through the argument carefully. It is really the only non-Euclidean case for which we shall give an involved entropy calculation in reasonably complete detail.

To make our life a little simpler, while still covering most of the important examples, we shall assume that the divergence of the covariance kernel near the diagonal is bounded as follows:

$$(2.2.11) \quad C(s, t) \leq \frac{C}{|s - t|^\alpha},$$

for all  $|s - t| \leq \delta$ , for some  $\delta > 0$ ,  $C < \infty$  and  $\alpha > 0$ .

We now start describing potential function spaces to serve as parameter spaces for continuous generalised Gaussian fields with such covariance kernels.

Take  $T \subset \mathbb{R}^N$  compact,  $q > 0$ , and  $p = \lfloor q \rfloor$ . Let  $C_0, \dots, C_p$  and  $C_q$  be finite, positive constants, and let  $\mathcal{F}^{(q)} = \mathcal{F}^{(q)}(T, C_0, \dots, C_p, C_q)$  be the class of functions on  $T$  whose partial derivatives of orders  $1, \dots, p$  are bounded by  $C_0, \dots, C_p$ , and for which the partial derivatives of order  $p$  satisfy Hölder conditions of order  $q - p$  with constant  $C_q$ . Thus for each  $\varphi \in \mathcal{F}^{(q)}$  and  $t, t + \tau \in T$ ,

$$(2.2.12) \quad \varphi(t + \tau) = \sum_{n=0}^p \frac{\Phi_n(t, \tau)}{n!} + \Delta(t, \tau),$$

where each  $\Phi_n(t, \tau)$  is a homogeneous polynomial of degree  $n$  in  $\tau = (\tau_1, \dots, \tau_N)$  of the form

$$(2.2.13) \quad \Phi_n(t, \tau) = \sum_{j_1=1}^N \cdots \sum_{j_n=1}^N \frac{\partial^n \varphi(t)}{\partial t_{j_1} \cdots \partial t_{j_n}} \tau_{j_1} \cdots \tau_{j_n},$$

and where

$$(2.2.14) \quad \sup_{t \in T} \left| \frac{\partial^n \varphi(t)}{\partial t_{j_1} \cdots \partial t_{j_n}} \right| \leq C_n \quad \text{and} \quad |\Delta(t, \tau)| \leq C_q |\tau|^q.$$

---

<sup>7</sup>The terminology comes from deterministic analogues. There are many partial differential equations for which no pointwise solution exists, but “smoothed” or “weak” versions of the equations do have solutions. This is analogous to the non-existence of a pointwise-defined  $f(t)$  in (2.2.9), while a function indexed version of  $f$  does make sense. The function  $\varphi$  plays the rôle of a local “mollifier”, or smoothing function

Two things are obvious in the above setup, in which you should think of the dimension  $N$  as fixed. Firstly, the larger the  $\alpha$  in (2.2.11) the rougher the process with this covariance will be. Secondly, the larger  $q$  is the smaller the family  $\mathcal{F}^{(q)}$  will be, and thus the more likely that a Gaussian process defined on  $\mathcal{F}^{(q)}$  will be continuous. Thus it seems reasonable to expect that a result of the following kind should be true.

**Theorem 2.2.4** *A centered Gaussian process with covariance function satisfying (2.2.10) and covariance kernel satisfying (2.2.11) will be continuous on  $\mathcal{F}^{(q)}(T, C_0, \dots, C_p, C_q)$  if  $\alpha < N$  and*

$$q > \frac{1 + \alpha - N}{2}.$$

A few comments are in order before we start the proof. Firstly, note that since we have not specified any other metric on  $\mathcal{F}^{(q)}$ , the continuity claim of the Theorem is in relation to the topology induced by the canonical metric  $d$ . There are, of course, more natural metrics on  $\mathcal{F}^{(q)}$ , but recall from Lemma 2.1.1 that mere continuity is independent of the metric, as long as  $C(\varphi, \psi)$  is continuous in  $\varphi$  and  $\psi$ . More detailed information on moduli of continuity will follow immediately from Theorem 2.1.5, the relationship of the chosen metric to  $d$ , and the entropy bound (2.2.3) below.

Although we suggested above thinking of the dimension  $N$  as fixed, if you prefer to let it vary while keeping  $\alpha$  and  $q$  fixed, you will find that the larger  $N$  is, the less derivatives we require of our test functions to ensure continuity on  $\mathcal{F}^{(q)}$ . While at first this seems counter-intuitive, you should remember that as  $N$  increases the degree of the singularity in (2.2.11) decreases (for fixed  $\alpha$ ) and so the result is, in fact, reasonable.

Finally, while the reason for the assumption  $N > \alpha$  should be obvious from the proof, it is worthwhile noting already that it is precisely this condition that gives us a process with finite variance, since, for  $\varphi \in \mathcal{F}^{(q)}$  with  $\|\varphi\|_\infty < M$

$$\begin{aligned} \mathbb{E}\{f^2(\varphi)\} &= \int_T \int_T \varphi(s)C(s, t)\varphi(t) dsdt \\ &\leq M^2 \int_T \int_T C(s, t) dsdt \\ &\leq CM^2 \int_T \int_T |t - s|^{-\alpha} dsdt. \end{aligned}$$

Since  $T \subset \mathbb{R}^N$  is compact, a transformation to polar coordinates easily shows that the last integral is finite only if  $N > \alpha$ .

**Proof of Theorem 2.2.4** The proof relies on showing that the usual entropy integral converges, where the entropy of  $\mathcal{F}^{(q)}$  is measured in the

canonical metric  $d$ , where

(2.2.15)

$$d^2(\varphi, \psi) = \int_T \int_T (\varphi(s) - \psi(s)) C(s, t) (\varphi(t) - \psi(t)) dsdt.$$

We shall obtain a bound on the entropy by explicitly constructing, for each  $\varepsilon > 0$ , a finite family  $\mathcal{F}_\varepsilon^{(q)}$  of functions that serve as an  $\varepsilon$ -net for  $\mathcal{F}^{(q)}$  in the  $d$  metric. To make life notationally easier, we shall assume throughout that  $T = [0, 1]^N$ .

Fix  $\varepsilon > 0$  and set

$$(2.2.16) \quad \delta = \delta(\varepsilon) = \varepsilon^{1/(q+(N-\alpha)/2)}.$$

Let  $\mathbb{Z}_\delta$  denote the grid of the  $(1 + \lfloor \delta^{-1} \rfloor)^N$  points in  $[0, 1]^N$  of the form

$$(2.2.17) \quad t_\eta = (\eta_1 \delta, \dots, \eta_N \delta), \quad \eta_i = 0, 1, \dots, \lfloor \delta^{-1} \rfloor, \quad i = 1, \dots, N.$$

Set

$$\delta_n = \delta_n(\varepsilon) = \delta^{q-n}, \quad n = 0, \dots, p = \lfloor q \rfloor,$$

and for each  $\varphi \in \mathcal{F}^{(q)}$ ,  $n = 0, \dots, p$ , and  $t_\eta$  of the form (2.2.17) let  $A_\eta^{(n)}(\varphi)$  denote the vector formed by taking the integer part of  $\delta_n^{-1}$  times the partial derivatives of  $\varphi$  of order  $n$  evaluated at the point  $t_\eta$ . (The index  $\eta$  here is, of course,  $N$ -dimensional.) Thus, a typical element of  $A_\eta^{(n)}$  is of the form

$$(2.2.18) \quad \left\{ A_\eta^{(n)}(\varphi) \right\}_i = \left\lfloor \frac{\varphi^{(n_1, \dots, n_N)}(t_\eta)}{\delta_n} \right\rfloor, \quad n_1 + \dots + n_N = n,$$

where we have written  $\varphi^{(n_1, \dots, n_N)}$  for the derivative  $\partial^{n_1} \varphi / \partial^{n_1} x_1 \dots \partial^{n_N} x_N$ , and the index  $i$  runs from 1 to  $\binom{n+N-1}{N-1}$ , the number of partitions of  $n$  into  $N$  parts.

Finally, for  $\varphi \in \mathcal{F}^{(q)}$ , let  $A^{(n)} = A^{(n)}(\varphi)$  denote the vector valued function on  $\mathbb{Z}_\delta$  defined by  $A^{(n)}(t_\eta) = A_\eta^{(n)}(\varphi)$ . For each  $\varphi \in \mathcal{F}^{(q)}$ , let  $F_{A^{(n)}(\varphi)}$  denote the set of  $\psi \in \mathcal{F}^{(q)}$  with fixed matrix  $A^{(n)}(\varphi)$ . Our first task will be to show that the  $d$ -radius of  $F_{A^{(n)}(\varphi)}$  is not greater than  $C\varepsilon$ , where  $C$  is a constant dependent only on  $q$  and  $N$ . All that will then remain will be to calculate how many different collections  $F_{A^{(n)}(\varphi)}$  are required to cover  $\mathcal{F}^{(q)}$ . In other words, we need to find how many  $\varphi$ 's are needed to approximate, in terms of the metric (2.2.15), all functions in  $\mathcal{F}^{(q)}$ .

Thus, take  $\varphi_1, \varphi_2 \in F_{A^{(n)}(\varphi)}$ , and set

$$(2.2.19) \quad \varphi = \varphi_1 - \varphi_2.$$

Let  $\| \cdot \|_d$  be the norm induced on  $\mathcal{F}^{(q)}$  by the metric  $d$ , and  $\| \cdot \|_\infty$  the usual sup norm. Then

$$\|\varphi\|_d^2 = \int_{[0,1]^N} \int_{[0,1]^N} \varphi(s) C(s, t) \varphi(t) dsdt, \quad \|\varphi\|_\infty = \sup_{[0,1]^k} |\varphi(t)|.$$

We have to show that the  $\varphi$  of (2.2.19) has  $d$ -norm less than  $C\varepsilon$ .

Note first, however, that in view of the definition (2.2.18) of the matrix  $A_\delta$  we have that for each  $t_\eta \in \mathbb{Z}_\delta$  and each partial derivative  $\varphi^{(n_1, \dots, n_N)}$  of such a  $\varphi$  of order  $n_1 + \dots + n_N = n \leq p$  that

$$\varphi^{(n_1, \dots, n_N)}(t_\eta) \leq \delta_n.$$

Putting this inequality together with the Taylor expansion (2.2.12)–(2.2.13) we find that for all  $t \in [0, 1]^N$

$$\begin{aligned} |\varphi(t)| &\leq \sum_{n=0}^p \frac{N^n}{n!} \delta_n \delta^n + C\delta^q \\ &= C(N, p) \delta^q, \end{aligned}$$

the last line following from the definition of the  $\delta_n$  and the fact that each polynomial  $\Phi_n$  of (2.2.13) has less than  $N^n$  distinct terms.

Thus, for  $\varphi$  of the form (2.2.19),

$$(2.2.20) \quad \|\varphi\|_\infty \leq C\delta^q.$$

We now turn to  $\|\varphi\|_d$ . With  $\delta$  as above, set

$$D_\delta = \left\{ (s, t) \subset [0, 1]^N \times [0, 1]^N : \max_{i=1, \dots, N} |s_i - t_i| \leq \delta \right\}.$$

Then

$$\begin{aligned} (2.2.21) \quad \|\varphi\|_d^2 &= \int_{[0, 1]^N \times [0, 1]^N} \varphi(s) C(s, t) \varphi(t) ds dt \\ &= \int_{D_\delta} \varphi(s) C(s, t) \varphi(t) ds dt \\ &\quad + \int_{([0, 1]^N \times [0, 1]^N) \setminus D_\delta} \varphi(s) C(s, t) \varphi(t) ds dt \\ &= I_1(\delta) + I_2(\delta). \end{aligned}$$

Consider the first integral. Letting  $C$  change from line to line where necessary, we have from (2.2.11) and (2.2.20) that

$$\begin{aligned} (2.2.22) \quad I_1(\delta) &\leq C \delta^{2q} \int_{[0, 1]^N} ds \int \dots \int_{s_i - \delta \leq t_i \leq s_i + \delta} dt |s - t|^{-\alpha} \\ &\leq C \delta^{2q} \int_{[-\delta, \delta]^N} |t|^{-\alpha} dt \\ &\leq C \delta^{(2q + N - \alpha)}, \end{aligned}$$

the last inequality coming from an evaluation via polar coordinates, and requiring the condition  $\alpha < N$ .

Similarly, again relying on the fact that  $\alpha < N$ , it is easy to check that  $I_2(\delta)$  is also bounded above by  $C\delta^{(2q+N-\alpha)}$ . Substituting this fact and (2.2.22) into (2.2.21), and applying (2.2.16), we finally obtain that for  $\varphi$  satisfying (2.2.19)

$$(2.2.23) \quad \|\varphi\|_d < C\delta^{\frac{1}{2}(2q+N-\alpha)} = C\varepsilon.$$

That is, the  $d$ -radius of each set  $F_{A^{(n)}(\varphi)}$  is no greater than a uniform constant times  $\varepsilon$ .

It remains to determine how many collections  $F_{A^{(n)}(\varphi)}$  are required to cover  $\mathcal{F}^{(q)}$ . Since this is a calculation that is now independent of both Gaussian processes in general, and the above covariance function in particular, we shall only outline how it is done. The details, which require somewhat cumbersome notation, can be found in Kolmogorov and Tihomirov [53], which is a basic reference for general entropy computations.

Consider, for fixed  $\delta$ , the matrix  $A_\delta$ , parameterised, as in (2.2.17), by  $\eta_i = 0, 1, \dots, [\delta^{-1}]$ ,  $i = 1, \dots, N$ , and  $n = 0, 1, \dots, p$ . Fix, for the moment,  $\eta_2 = \dots = \eta_N = 0$ . It is clear from the restrictions (2.2.14), (2.2.18), the definition of  $\delta_n$ , and the fact that each vector  $A_\eta^{(n)}$  has no more than  $\binom{n+N-1}{N-1}$  distinct elements, that there are no more than

$$O\left(\frac{1}{\delta_0} \left(\frac{1}{\delta_1}\right)^{\binom{N}{N-1}} \dots \left(\frac{1}{\delta_p}\right)^{\binom{p+N}{N-1}}\right) = O(\delta^{-\xi})$$

(for an appropriate and eventually unimportant  $\xi$ ) ways to fill in the row of  $A_\delta$  corresponding to  $(n_1, \dots, n_N) = (0, \dots, 0)$ .

What remains to show is that because of the rigid continuity conditions on the functions in  $\mathcal{F}^{(q)}$ , there exists an absolute constant  $M = M(q, C_0, \dots, C_p, C_q)$ , such that once this first row is determined, there are no more than  $M$  ways to complete the row corresponding to  $(n_1, \dots, n_N) = (1, \dots, 0)$ , and similarly no more than  $M^2$  ways to complete the row corresponding to  $(n_1, \dots, n_N) = (2, \dots, 0)$ , etc. Thus, all told, there are no more than

$$(2.2.24) \quad O\left(\delta^{-\xi} M^{N(1+\delta^{-1})}\right)$$

ways to fill the matrix  $A_\delta$ , and thus we have a bound for the number of different collections  $F_{A^{(n)}(\varphi)}$ .

Modulo a constant, it now follows from (2.2.16), (2.2.23) and (2.2.24) that the log entropy function for our process is bounded above by

$$C_1 + \frac{C_2 \xi}{(q + (N - \alpha)/2)} \ln\left(\frac{1}{\varepsilon}\right) + C_3 \left(\frac{1}{\varepsilon}\right)^{1/(q+(N-\alpha)/2)}.$$

Since this is integrable if  $q > (1 + \alpha - N)/2$ , we are done.  $\square$

Before leaving function indexed processes, there are a number of comments that are worth making, that relate them to other problems both within and outside of the theory of Gaussian processes.

Firstly, in most of the literature pertaining to generalised Gaussian fields the parameter space used is the Schwartz space  $\mathcal{S}$  of infinitely differentiable functions decaying faster than any polynomial at infinity. Since this is a very small class of functions (at least in comparison to the classes  $\mathcal{F}^{(q)}$  that Theorem 2.2.4 deals with) continuity over  $\mathcal{S}$  is automatically assured and therefore not often explicitly treated. However, considerations of continuity and smaller parameter spaces are of relevance in the treatment of infinite dimensional diffusions arising as the solutions of stochastic partial differential equations, in which solutions over very specific parameter spaces are often sought. For more on this see, for example, [46, 101, 100].

Secondly, some words on our choice of (2.2.11) as a condition on the covariance kernel  $C(s, t)$ . When  $\alpha = N - 2$ ,  $N > 2$ , then the class of generalised fields that we are considering here includes the so called “free field” of Euclidean quantum field theory. (When  $k = 2$  the free field has a covariance kernel with a logarithmic singularity at 0, and when  $k = 1$  the free field is no longer generalised, but is the real valued, stationary Markov Gaussian process with covariance function  $C(t) = e^{-\beta|t|}$ , for some  $\beta > 0$ .) This process, along with a large number of related generalised fields whose covariance kernels satisfy similar conditions, possess a type of multi-dimensional Markov property. For details on this see, for example, Dynkin [31, 32, 29, 30] and Adler and Epstein [5] and references therein. For structural and renormalisation properties of generalised fields of this kind, presented among a much wider class of examples, see Dobrushin [21], who also treats a large variety of non-Gaussian fields.

Our divergence assumption (2.2.11) leaves out a class of examples important to the theory of empirical processes, in which the covariance kernel is the product of a Dirac delta  $\delta$  and a bounded “density”,  $g$ , in the sense that

$$\begin{aligned} \mathbb{E}\{f(\varphi)f(\psi)\} &= \int \varphi(t)\psi(t)g(t)dt \\ \text{“=”} &= \iint \varphi(s)\left[g^{1/2}(s)\delta(s,t)g^{1/2}(t)\right]\psi(t)dt. \end{aligned}$$

As we have already noted, such processes arose as the stochastic integrals  $W(\varphi)$  of Section 1.4.1, for which  $W$  was a Gaussian  $\nu$ -noise where  $\nu$  is (now) a probability measure with density  $g$ . For more on this setting, in which the computations are similar in spirit to those made above, see Dudley [28].

Finally, it is worth noting that much of what has been said above regarding generalised fields – i.e. function indexed processes – can be easily extended to Gaussian processes indexed by families of measures. For example, if we consider the function  $\varphi$  in (2.2.9) to be the (positive) density

of a measure  $\mu$  on  $\mathbb{R}^N$ , then by analogy it makes sense to write

$$f(\mu) = \int_{\mathbb{R}^N} f(t) \mu(dt),$$

with the corresponding covariance functional

$$C(\mu, \nu) = \mathbb{E}\{f(\mu)f(\nu)\} = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \mu(ds) C(s, t) \nu(dt).$$

Again, as was the case for generalised Gaussian fields, the process  $X(\mu)$  may be well defined even if the covariance kernel  $C$  diverges on the diagonal. In fact,  $f(\mu)$  will be well defined for all  $\mu \in M_C$ , where

$$M_C \triangleq \left\{ \mu : \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \mu(ds) C(s, t) \mu(dt) < \infty \right\}.$$

Similar arguments to those used above to characterise the continuity of a family of Gaussian fields on  $\mathcal{F}^{(q)}$  can be used to ascertain continuity of measure indexed processes on suitably smooth classes of measures. We leave both the details and an attempt to formulate the appropriate results to the interested reader<sup>8</sup>.

#### 2.2.4 Set indexed processes

We have already met some set indexed processes in dealing with the Brownian family of processes in Section 1.3, in which we concentrated on Gaussian  $\nu$ -noise (cf. (1.3.1)–(1.3.3)) indexed by sets. We saw, for example, that the while the Brownian sheet indexed by rectangles was continuous (Theorem 1.3.2) when indexed by lower layers in  $[0, 1]^2$  it was discontinuous and unbounded (Theorem 1.3.3).

In this Subsection we shall remain in the setting of  $\nu$ -noise, and look at two classes of set indexed processes. The first will be Euclidean, and the parameter spaces will be classes of sets in compact  $T \subset \mathbb{R}^N$  with smooth boundaries. For this example we also add the assumption that  $\nu$  has a density bounded away from zero and infinity on  $T$ . For the second example, we look at Vapnik-Červonenkis classes of sets, of singular importance in statistical learning theory and image processing, and characterised by certain combinatorial properties. Here the ambient space (in which the sets lie) can be any measure space. We shall skimp on proofs when they have nothing qualitatively new to offer. In any case, all that we have to say is

---

<sup>8</sup>Actually, to the best of our knowledge, this specific example has never been treated in the literature, and so it would be a rather interesting problem to work out how to optimally, and naturally, formulate the requisite smoothness conditions. The general idea of how to proceed can be gleaned from the treatment of set-indexed processes in the following Subsection.

done in full detail in Dudley [24, 28], where you can also find a far more comprehensive treatment and many more examples.

Our first family is actually closely related to the family  $\mathcal{F}^{(q)}$  of functions we have just studied in detail. While we shall need some of the language of manifolds and homotopy to describe this example, which will only be developed later in Chapter 3, it will be basic enough that the average reader should have no trouble following the argument.

With  $S^{N-1}$ , as usual, denoting the unit sphere in  $\mathbb{R}^N$ , recall the basic fact that we can cover it by two patches  $V_1$  and  $V_2$ , each of which maps via a  $C^\infty$  diffeomorphism  $F_j : V_j \rightarrow B^{N-1}$  to the open ball  $B^{N-1} = \{t \in \mathbb{R}^{N-1} : |t|^2 < 1\}$ .

Adapting slightly the notation of the previous example, let  $\mathcal{F}^{(q)}(V_j, M)$  be the set of all real valued functions  $\varphi$  on  $V_j$  such that

$$\varphi \circ F_j^{-1} \in \mathcal{F}^{(q)}(B^{N-1}, M, \dots, M)$$

(cf. (2.2.12)–(2.2.14)). Furthermore, let  $\mathcal{F}^{(q)}(S^{N-1}, M)$  denote the set of all real valued functions  $\varphi$  on  $S^{N-1}$  such that the restriction of  $\varphi$  to  $V_j$  is in  $\mathcal{F}^{(q)}(V_j, M)$ . Now taking the  $N$ -fold Cartesian product of copies of  $\mathcal{F}^{(q)}(S^{N-1}, M)$ , we obtain a family of functions from  $S^{N-1}$  to  $\mathbb{R}^N$ , which we denote by  $D(N, q, M)$ , where the “ $D$ ” stands for Dudley, who introduced this family in [23].

Each  $\varphi \in D(N, q, M)$  defines an  $(N - 1)$ -dimensional surface in  $\mathbb{R}^N$ , and a simple algebraic geometric construction<sup>9</sup> enables one to “fill in” the interior of this surface to obtain a set  $I_\varphi$ . We shall denote the family of sets obtained in this fashion by  $I(N, q, M)$ , and call them the “Dudley sets with  $q$ -times differentiable boundaries”.

**Theorem 2.2.5** *The Brownian sheet is continuous on a bounded collection of Dudley sets in  $\mathbb{R}^N$  with  $q$  times differentiable boundaries if  $q > N - 1 \geq 1$ . If  $N - 1 > 1 \geq q > 0$  or if  $N - 1 > q \geq 1$  then the Brownian sheet is unbounded with probability one.*

**Outline of Proof.** The proof of the unboundedness part of the result is beyond us, and so you are referred to [28]. As far as the proof of continuity is concerned, what we need are the following inequalities for the log-entropy, on the basis of which continuity follows from Theorem 2.1.5.

$$H(N(k, q, M), \varepsilon) \leq \begin{cases} C\varepsilon^{-2(N-1)/(Nq-N+1)} & (N-1)/N < q \leq 1, \\ C\varepsilon^{-2(N-1)/q} & 1 \leq q. \end{cases}$$

---

<sup>9</sup>The construction works as follows: For given  $\varphi : S^{N-1} \rightarrow \mathbb{R}^N$  in  $D(N, q, M)$ , let  $R_\varphi$  be its range and  $A_\varphi$  be the set of all  $t \in \mathbb{R}^N$ ,  $t \notin R_\varphi$ , such that among mappings of  $S^{N-1}$  into  $\mathbb{R}^N \setminus \{t\}$ ,  $\varphi$  is not homotopic (cf. Definition 3.2.3) to any constant map  $\psi(s) \equiv r \neq t$ . Then define  $I_\varphi = R_\varphi \cup A_\varphi$ . For an example, try untangling this description for  $\varphi$  the identity map from  $S^{N-1}$  to itself to see that what results is  $I_\varphi = B^N$ .

These inequalities rely on the “simple algebraic geometric construction” noted above, and so we shall not bring them in detail. The details are in [28]. The basic idea, however, requires little more than noting that there are basically as many sets in  $I(N, q, M)$  as there are functions in  $D(N, q, M)$ , and we have already seen, in the previous example, how to count the number of functions in  $D(N, q, M)$ .

There are also equivalent lower bounds for the log-entropy for certain values of  $N$  and  $q$ , but these are not important to us.  $\square$

We now turn to the so called *Vapnick-Červonenkis*, or VC, sets, due, not surprisingly, to Vapnick and Červonenkis [99, 98]. These sets arise in a very natural way in many areas including statistical learning theory and image analysis. The recent book [97] by Vapnick is a good place to see why.

The arguments involved in entropy calculations for VC classes of sets are of an essentially combinatoric nature, and so somewhat different to those we have met so far. We shall therefore look at them somewhat more closely than we did for Dudley sets. For more details, however, including a discussion of the importance of VC classes to the problem of finding “universal Donsker classes” in the theory of empirical processes, see [28].

Let  $(E, \mathcal{E})$  be a measure space. Given a class  $\mathcal{C}$  of sets in  $\mathcal{E}$  and a finite set  $F \subset E$ , let  $\Delta^{\mathcal{C}}(F)$  be the number of different sets  $A \cap F$  for  $A \in \mathcal{C}$ . If the number of such sets is  $2^{|F|}$ , then  $\mathcal{C}$  is said to *shatter*  $F$ . For  $n = 1, 2, \dots$ , set

$$m^{\mathcal{C}}(n) \triangleq \max\{\Delta^{\mathcal{C}}(F) : F \text{ has } n \text{ elements}\}.$$

Clearly,  $m^{\mathcal{C}}(n) \leq 2^n$  for all  $n$ . Also, set

(2.2.25)

$$V(\mathcal{C}) \triangleq \begin{cases} \inf\{n : m^{\mathcal{C}}(n) < 2^n\} & \text{if } m^{\mathcal{C}}(n) < 2^n \text{ for some } n, \\ \infty & \text{if } m^{\mathcal{C}}(n) = 2^n \text{ for all } n. \end{cases}$$

The class  $\mathcal{C}$  is called a Vapnik-Červonenkis class if  $m^{\mathcal{C}}(n) < 2^n$  for *some*  $n$ ; i.e. if  $V(\mathcal{C}) < \infty$ . The number  $V(\mathcal{C})$  is called the VC index of  $\mathcal{C}$ , and  $V(\mathcal{C}) - 1$  is the largest cardinality of a set shattered by  $\mathcal{C}$ .

Two extreme but easy examples which you can check for yourself are  $E = \mathbb{R}$  and  $\mathcal{C}$  all half lines of the form  $(-\infty, t]$ , for which  $m^{\mathcal{C}}(n) = n + 1$  and  $V(\mathcal{C}) = 2$ , and  $E = [0, 1]$  with  $\mathcal{C}$  all the open sets in  $[0, 1]$ . Here  $m^{\mathcal{C}}(n) = 2^n$  for all  $n$  and so  $V(\mathcal{C}) = \infty$  and  $\mathcal{C}$  is not a VC class.

A more instructive example, that also leads into the general theory we are after, is  $E = \mathbb{R}^N$  and  $\mathcal{C}$  is the collection of half-spaces of  $\mathbb{R}^N$ . Let  $\Phi(N, n)$  be the maximal number of components into which it is possible to partition  $\mathbb{R}^N$  via  $n$  hyperplanes. Then, by definition,  $m^{\mathcal{C}}(n) = \Phi(N, n)$ . It is not hard to see that  $\Phi$  must satisfy the recurrence relation

$$(2.2.26) \quad \Phi(N, n) = \Phi(N, n - 1) + \Phi(N - 1, n - 1),$$

with the boundary conditions  $\Phi(0, n) = \Phi(N, 0) = 1$ .

To see this, note that if  $\mathbb{R}^N$  has already been partitioned into  $\Phi(N, n-1)$  subsets via  $n-1$  ( $(N-1)$ -dimensional) hyperplanes,  $H_1, \dots, H_{n-1}$ , then adding one more hyperplane  $H_n$  will cut in half as many of these subsets as intersect  $H_n$ . There can be no more such subsets, however, than the maximal number of subsets formed on  $H_n$  by partitioning with the  $n-1$  ( $(N-2)$ -dimensional) hyperplanes  $H_1 \cap H_n, \dots, H_{n-1} \cap H_n$ ; i.e.  $\Phi(N-1, n-1)$ . Hence (2.2.26).

Induction then shows that

$$\Phi(N, n) = \begin{cases} \sum_{j=0}^N \binom{n}{j} & \text{if } n > N, \\ 2^n & \text{if } n \leq N, \end{cases}$$

where we adopt the usual convention that  $\binom{n}{j} = 0$  if  $n < j$ .

From either the above or (2.2.26) you can now check that

$$(2.2.27) \quad \Phi(N, n) \leq n^N + 1, \quad \text{for all } N, n > 0.$$

It thus follows, from (2.2.25), that the half-spaces of  $\mathbb{R}^N$  form a VC class for all  $N$ .

What is somewhat more surprising, however, is that an inequality akin to (2.2.27), which we developed only for this special example, holds in wide (even non-Euclidean) generality.

**Lemma 2.2.6** *Let  $\mathcal{C}$  be a collection of sets in  $\mathcal{E}$  such that  $V(\mathcal{C}) \leq v$ . Then*

$$(2.2.28) \quad m^{\mathcal{C}}(n) < \Phi(v, n) \leq n^v + 1, \quad \text{for all } n \geq v.$$

Since the proof of this result is combinatoric rather than probabilistic, and will be of no further interest to us, you are referred to either of [99, 28] for a proof.

The importance of Lemma 2.2.6 is that it enables us to obtain bounds on the entropy function for Gaussian  $\nu$ -noise over VC classes that are independent of  $\nu$ .

**Theorem 2.2.7** *Let  $W$  be the Gaussian  $\nu$ -noise on a probability space  $(E, \mathcal{E}, \nu)$ . Let  $\mathcal{C}$  be a Vapnik-Červonenkis class of sets in  $\mathcal{E}$  with  $V(\mathcal{C}) = v$ . Then there exists a constant  $K = K(v)$  (not depending on  $\nu$ ) such that for  $0 < \varepsilon \leq \frac{1}{2}$ , the entropy function for  $W$  satisfies*

$$N(\mathcal{C}, \varepsilon) \leq K\varepsilon^{-2v} |\ln \varepsilon|^v.$$

**Proof.** We start with a little counting, and then turn to the entropy calculation proper. The counting argument is designed to tell us something about the maximum number of  $\mathcal{C}$  sets that are a certain minimum distance from one another and can be packed into  $E$ .

Fix  $\varepsilon > 0$  and suppose  $A_1, \dots, A_m \in \mathcal{C}$ ,  $m \geq 2$ , and  $\nu(A_i \Delta A_j) \geq \varepsilon$  for  $i \neq j$ . We need an upper bound on  $m$ . Sampling with replacement, select  $n$  points at random from  $E$ . The  $\nu$ -probability that at least one of the sets  $A_i \Delta A_j$  contains none of these  $n$  points is at most

$$(2.2.29) \quad \binom{m}{2} (1 - \varepsilon)^n.$$

Choose  $n = n(m, \varepsilon)$  large enough so that this bound is less than 1. Then

$$P\{\text{all symmetric differences } A_i \Delta A_j \text{ are non-empty}\} > 0,$$

and so for at least one configuration of the  $n$  sample points the class  $\mathcal{C}$  picks out at least  $m$  distinct subsets. (Since, with positive probability, given any two of the  $A_i$  there is at least one point not in both of them.) Thus, by (2.2.28),

$$(2.2.30) \quad m \leq m^{\mathcal{C}}(n) \leq n^v = (n(m, \varepsilon))^v.$$

Take now the smallest  $n$  for which (2.2.29) is less than 1. For this  $n$  we have  $m^2(1 - \varepsilon)^{n-1} \geq 2$ , so that

$$n - 1 \leq \frac{2 \ln m - \ln 2}{|\ln(1 - \varepsilon)|},$$

and  $n \leq (2 \ln m)/\varepsilon$ . Furthermore, by (2.2.30),  $m \leq (2 \ln m)^v \varepsilon^{-v}$ .

For some  $m_0 = m_0(v) < \infty$ ,  $(2 \ln m)^v \leq m^{1/(v+1)}$  for  $m \geq m_0$ , and then  $m \leq \varepsilon^{-v-1}$ , so  $\ln m \leq (v+1)|\ln \varepsilon|$ . Hence

$$m \leq K(v) \varepsilon^{-v} |\ln \varepsilon|^v, \quad \text{for } 0 < \varepsilon \leq \frac{1}{2},$$

if  $K(v) = \max(m_0, 2^{v+1}(v+1)^v)$ .

This concludes the counting part of the proof. We can now do the entropy calculation. Recall that the canonical distance between on sets of  $\mathcal{E}$  is given by  $d_\nu(A, B) = [\nu(A \Delta B)]^{\frac{1}{2}}$ .

Fix  $\varepsilon > 0$ . In view of the above, there can be no more than  $m = K(v) \varepsilon^{2v} |\ln \varepsilon|^v$  sets  $A_1, \dots, A_m$  in  $\mathcal{C}$  for which  $d_\nu(A_i, A_j) \geq \varepsilon$  for all  $i, j$ . Take an  $\varepsilon$ -neighbourhood of each of the  $A_i$  in the  $d_\nu$  metric. (Each such neighbourhood is a collection of sets in  $\mathcal{E}$ .) The union of these neighbourhoods covers  $\mathcal{C}$ , and so we have constructed an  $\varepsilon$ -net of the required size, and are done.  $\square$

An immediate consequence of the entropy bound of Theorem 2.2.7 is

**Corollary 2.2.8** *Let  $W$  be Gaussian  $\nu$ -noise based over a probability space  $(E, \mathcal{E}, \nu)$ . Then  $W$  is continuous over any Vapnik-Červonenkis class of sets in  $\mathcal{E}$ .*

### 2.2.5 Non-Gaussian processes

A natural question to ask is whether or not the results and methods that we have seen in this Section extend naturally to non-Gaussian fields. We already noted, immediately after the proof of the central Theorem 2.1.5, that the proof there only used normality once, and so the general techniques of entropy should be extendable to a far wider setting.

For most of the processes that will concern us, this will not be terribly relevant. Back in Section 1.5 we already decided that these can be written in the form

$$f(t) = F(g^1(t), \dots, g^d(t)).$$

where the  $g^i$  are i.i.d. Gaussian and  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth. In this setting, continuity and boundedness of the non-Gaussian  $f$  follow deterministically from similar properties on  $F$  and the  $g^i$ , and so no additional theory is needed.

Nevertheless, there are many processes that are not attainable in this way, for which one might *a priori* expect that the random geometry of Chapter 4 might apply. In particular, we are thinking of the smooth “stable” fields on Samorodnitsky and Taqqu [82]. With this in mind, and for completeness, we state Theorem 2.2.9 below. However, other than the “function of Gaussian” non-Gaussian scenario, we know of no cases for which this random geometry has even the beginnings of a parallel theory.

To set up the basic result, let  $f_t$  be a random process defined on a metric space  $(T, \tau)$  and taking values in a Banach space  $(B, \|\cdot\|_B)$ . Since we are no longer in the Gaussian case, there is no reason to assume that there is any more a “canonical metric” of  $T$  to replace  $\tau$ . Recall that a function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is called a *Young function* if it is even, continuous, convex, and satisfies

$$\lim_{x \rightarrow 0} \frac{\varphi(x)}{x} = 0, \quad \lim_{x \rightarrow \infty} \frac{\varphi(x)}{x} = \infty.$$

**Theorem 2.2.9** *Take  $f$  as above, and assume that the real valued process  $\|f_t - f_s\|_B$  is separable. Let  $N_\tau$  be the metric entropy function for  $T$  with respect to the metric  $\tau$ . If there exists an  $\alpha \in (0, 1]$  and a Young function  $\varphi$  such that the following two conditions are satisfied, then  $f$  is continuous with probability one.*

$$\mathbb{E} \left\{ \phi \left( \frac{\|f(t) - f(s)\|_B^\alpha}{\tau(s, t)} \right) \right\} \leq 1,$$

$$\int_{N_\tau(u) > 1} \phi^{-1}(N_\tau(u)) \, du < \infty.$$

The best place to read about this is Ledoux and Talagrand [57].

## 2.3 Borell-TIS inequality

One of the first facts we learnt about Gaussian processes was that if  $X \sim N(0, \sigma^2)$  then, for all  $u > 0$ ,

$$(2.3.1) \quad \mathbb{P}\{X > u\} \leq \frac{\sigma}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2/\sigma^2}.$$

(cf. (1.2.2).) One immediate consequence of this is that

$$(2.3.2) \quad \lim_{u \rightarrow \infty} u^{-2} \ln \mathbb{P}\{X > u\} = -(2\sigma^2)^{-1}.$$

There is a classical result of Landau and Shepp [54] and Marcus and Shepp [66] that gives a result closely related to (2.3.2), but for the supremum of a general centered Gaussian process. If we assume that  $f_t$  is a.s. bounded, then they showed that

$$(2.3.3) \quad \lim_{u \rightarrow \infty} u^{-2} \ln \mathbb{P}\left\{\sup_{t \in T} f_t > u\right\} = -(2\sigma_T^2)^{-1},$$

where

$$\sigma_T^2 \triangleq \sup_{t \in T} \mathbb{E}\{f_t^2\}$$

is a notation that will remain with us throughout this Section. An immediate consequence of (2.3.3) is that, for all  $\varepsilon > 0$  and large enough  $u$ ,

$$(2.3.4) \quad \mathbb{P}\left\{\sup_{t \in T} f_t > u\right\} \leq e^{\varepsilon u^2 - u^2/2\sigma_T^2}.$$

Since  $\varepsilon > 0$  is arbitrary, comparing (2.3.4) and (2.3.1) we reach the rather surprising conclusion that the supremum of a centered, bounded Gaussian process behaves much like a single Gaussian variable with a suitably chosen variance.

In Chapter 7 we shall work very hard to close the gap between (2.3.1) and (2.3.4) (i.e. between  $u^{-1}$  and  $e^{\varepsilon u^2}$ ), however now we want to see from where (2.3.4) comes.

In fact, all of the above inequalities are special cases of a *non-asymptotic* result<sup>10</sup> due, independently, and with very different proofs, to Borell [15] and Tsirelson, Ibragimov and Sudakov (TIS) [96].

---

<sup>10</sup>Actually, Theorem 2.3.1 is not in the same form as Borell's original inequality, in which  $\mathbb{E}\{\|f\|\}$  was replaced by the median of  $\|f\|$ . However, the two forms are equivalent. For this and other variations of (2.3.5), including extensions to Banach space valued processes for which  $\|\cdot\|$  is the norm, see the more detailed treatments of [13, 28, 36, 57, 61]. To see how the Borell-TIS inequality fits into the wider theory of concentration inequalities, see the recent book [56] by Ledoux.

**Theorem 2.3.1 (Borell-TIS inequality)** *Let  $f_t$  be a centered Gaussian process, a.s. bounded on  $T$ . Write  $\|f\| = \|f\|_T = \sup_{t \in T} f_t$ . Then  $\mathbb{E}\{\|f\|\} < \infty$ , and, for all  $u > 0$ ,*

$$(2.3.5) \quad \mathbb{P}\{|\|f\| - \mathbb{E}\{\|f\|\}| > u\} \leq 2e^{-u^2/2\sigma_T^2}.$$

Before we look at the proof of (2.3.5), which we refer to as the Borell-TIS inequality, we take a moment to look at some of its consequences, which are many and major. It is no exaggeration to say that this inequality is today the single most important tool in the general theory of Gaussian processes.

An immediate and trivial consequence of (2.3.5) is that, for all  $u > \mathbb{E}\{\|f\|\}$ ,

$$\mathbb{P}\{\|f\| > u\} \leq 2e^{-(u - \mathbb{E}\{\|f\|\})^2/2\sigma_T^2},$$

so that both (2.3.3) and (2.3.4) follow from the Borell-TIS inequality.

Indeed, a far stronger result is true, for (2.3.4) can now be replaced by

$$\mathbb{P}\{\|f\| > u\} \leq e^{Cu - u^2/2\sigma_T^2},$$

where  $C$  is a constant depending only on  $\mathbb{E}\{\|X\|\}$ , and we know how to at least bound this quantity from Theorem 2.1.5.

Note that, despite the misleading notation,  $\|\cdot\| \equiv \sup$  is not a norm, and that very often one needs bounds on the tail of  $\sup_t |f_t|$ , which does give a norm. However, symmetry immediately gives

$$(2.3.6) \quad \mathbb{P}\left\{\sup_t |f_t| > u\right\} \leq 2\mathbb{P}\left\{\sup_t f_t > u\right\},$$

so that the Borell-TIS inequality helps out here as well.

Here is a somewhat more significant consequence of the Borell-TIS inequality.

**Theorem 2.3.2** *For  $f$  centered, Gaussian,*

$$(2.3.7) \quad \begin{aligned} \mathbb{P}\{\|f\| < \infty\} = 1 & \iff \mathbb{E}\{\|f\|\} < \infty \\ & \iff \mathbb{E}\left\{e^{\alpha\|f\|^2}\right\} < \infty \end{aligned}$$

*for sufficiently small  $\alpha$ .*

**Proof.** The existence of the exponential moments of  $\|f\|$  implies the existence of  $\mathbb{E}\{\|f\|\}$ , and this in turn implies the a.s. finiteness of  $\|f\|$ . Furthermore, since by Theorem 2.3.1 we already know that the a.s. finiteness of  $\|f\|$  entails that of  $\mathbb{E}\{\|f\|\}$ , all that remains is to prove is that the a.s. finiteness of  $\|f\|$  also implies the existence of exponential moments.

But this is an easy consequence of the Borell-TIS inequality, since, with both  $\|f\|$  and  $\mathbb{E}\{\|f\|\}$  now finite,

$$\begin{aligned}
\mathbb{E}\left\{e^{\alpha\|f\|^2}\right\} &= \int_0^\infty \mathbb{P}\left\{e^{\alpha\|f\|^2} > u\right\} du \\
&\leq \mathbb{E}\{\|f\|\} + \int_{\mathbb{E}\{\|f\|\}}^\infty \mathbb{P}\left\{\|f\| > \sqrt{\ln u^{1/\alpha}}\right\} du \\
&\leq \mathbb{E}\{\|f\|\} \\
&\quad + 2 \int_{\mathbb{E}\{\|f\|\}}^\infty \exp\left(\frac{-\left(\sqrt{\ln u^{1/\alpha}} - \mathbb{E}\{\|f\|\}\right)^2}{2\sigma_T^2}\right) du \\
&\leq \mathbb{E}\{\|f\|\} \\
&\quad + 4\alpha \int_0^\infty u \exp\left(\frac{-(u - \mathbb{E}\{\|f\|\})^2}{2\sigma_T^2}\right) \exp\{\alpha u^2\} du,
\end{aligned}$$

which is clearly finite for small enough  $\alpha$ .  $\square$

Recall Theorems 2.1.3 and 2.1.5 which established, respectively, the a.s. boundedness of  $\|f\|$  and a bound on the modulus of continuity  $\omega_{f,d}(\delta)$  under essentially identical entropy conditions. It was rather irritating back there that we had to establish each result independently, since it is “obvious” that one should imply the other. A simple application of the Borell-TIS inequality almost does this.

**Theorem 2.3.3** *Suppose that  $f$  is a.s. bounded on  $T$ . Then  $f$  is also a.s. uniformly continuous (with respect to the canonical metric  $d$ ) if, and only if,*

$$(2.3.8) \quad \lim_{\eta \rightarrow 0} \phi(\eta) = 0,$$

where

$$(2.3.9) \quad \phi(\eta) \triangleq \mathbb{E}\left\{\sup_{d(s,t) < \eta} (f_s - f_t)\right\}.$$

Furthermore, under (2.3.8), for all  $\varepsilon > 0$  there exists an a.s. finite random variable  $\eta > 0$  such that

$$(2.3.10) \quad \omega_{f,d}(\delta) \leq \phi(\delta) |\ln \phi(\delta)|^\varepsilon,$$

for all  $\delta \leq \eta$ .

**Proof.** We start with necessity. For almost every  $\omega$  we have

$$\lim_{\eta \rightarrow 0} \sup_{d(s,t) < \eta} |f_s(\omega) - f_t(\omega)| = 0.$$

But (2.3.8) now follows from dominated convergence and Theorem 2.3.2.

For sufficiency, note that from (2.3.8) we can find a sequence  $\{\delta_n\}$  with  $\delta_n \rightarrow 0$  such that  $\phi(\delta_n) \leq 2^{-n}$ . Set  $\delta'_n = \min(\delta_n, 2^{-n})$ , and consider the event

$$A_n = \left\{ \sup_{d(s,t) < \delta'_n} |f_s - f_t| > 2^{-n/2} \right\}.$$

The Borell-TIS inequality (cf. (2.3.6)) gives that, for  $n \geq 3$ ,

$$\begin{aligned} \mathbb{P}\{A_n\} &\leq 2 \exp\left(-\frac{1}{2}(2^{-n/2} - 2^{-n})^2/2^{-2n}\right) \\ &\leq K \exp(-2^{n-1}). \end{aligned}$$

Since  $\mathbb{P}\{A_n\}$  is an admirably summable series, Borel-Cantelli gives us that  $f$  is a.s. uniformly  $d$ -continuous, as required.

To complete the proof we need to establish the bound (2.3.10) on  $\omega_{f,d}$ . Note first that

$$\begin{aligned} \text{diam}(S) &= \sup_{s,t \in S} (\mathbb{E}\{|f_t - f_s|^2\})^{1/2} \\ &= \sqrt{2\pi} \sup_{s,t \in S} \mathbb{E}\{|f_t - f_s|\} \\ &\leq \sqrt{2\pi} \mathbb{E}\left\{ \sup_{s,t \in S} f_t - f_s \right\} \\ &= \sqrt{2\pi} \phi(\text{diam}(S)), \end{aligned}$$

where the second line is an elementary Gaussian computation (cf. Lemma 4.5.4 if it bothers you) and the third uses the fact that  $\sup_{s,t \in S} (f_t - f_s)$  is non-negative. Consequently, we have that

$$(2.3.11) \quad \delta \leq \sqrt{2\pi} \phi(\delta),$$

for all  $\delta > 0$ .

Now define the numbers

$$\delta_n = \inf\{\delta : \phi_\tau(\delta) \leq e^{-n}\}$$

and, for  $\varepsilon > 0$ , the events

$$B_n = \left\{ \sup_{d(s,t) < \delta_n} |f_s - f_t| > \phi(\delta_n) |\ln \phi(\delta_n)|^{\varepsilon/2} \right\}.$$

Then the Borell-TIS inequality gives

$$\begin{aligned} \mathbb{P}\{B_n\} &\leq 4 \exp\left\{-\frac{1}{2} \left(|\ln \phi(\delta_n)|^{\varepsilon/2} - 1\right)^2 \frac{\phi^2(\delta_n)}{\delta_n^2}\right\} \\ &\leq K_1 \exp\{-K_2 n^\varepsilon\}, \end{aligned}$$

the second inequality following from (2.3.11) and the definition of  $\delta_n$ .

Since  $\sum_n \mathbb{P}\{B_n\} < \infty$ , we have that for  $n \geq N(\omega)$

$$\omega_{f,d}(\delta_n) \leq \phi(\delta_n) |\ln \phi(\delta_n)|^{\varepsilon/2}.$$

Monotonicity of  $\omega_{f,d}$ , along with separability, complete the proof.  $\square$

We now turn to the proof of the Borell-TIS inequality. There are essentially three quite different ways to tackle this proof. Borell's original proof relied on isoperimetric inequalities. While isoperimetric inequalities may be natural for a book with the word "geometry" in its title, we shall avoid them, since they involve setting up a number of concepts for which we shall have no other need. The proof of Tsirelson, Ibragimov and Sudakov used Itô's formula from stochastic calculus. This is one of our<sup>11</sup> favourite proofs, since as one of the too few links between the Markovian and Gaussian worlds of stochastic processes, it is to be prized.

We shall, however, take a more direct route, which we learnt about from Marc Yor's excellent collection of exercises [113], although its roots are much older. The first step in this route involves the two following Lemmas, which are of independent interest.

**Lemma 2.3.4** *Let  $X$  and  $Y$  be independent  $k$ -dimensional vectors of centered, unit variance, independent, Gaussian variables. If  $f, g : \mathbb{R}^k \rightarrow \mathbb{R}$  are bounded  $C^2$  functions then*

(2.3.12)

$$\text{Cov}(f(X), g(X)) = \int_0^1 \mathbb{E} \left\{ \langle \nabla f(X), \nabla g(\alpha X + \sqrt{1-\alpha^2}Y) \rangle \right\} d\alpha,$$

where  $\nabla f(x) = \left( \frac{\partial f}{\partial x_i} f(x) \right)_{i=1, \dots, k}$ .

**Proof.** It suffices to prove the Lemma with  $f(x) = e^{i\langle t, x \rangle}$  and  $g(x) = e^{i\langle s, x \rangle}$ , with  $s, t, x \in \mathbb{R}^k$ . Standard approximation arguments (which is where the requirement that  $f$  is  $C^2$  appears) will do the rest. Write

$$\varphi(t) \triangleq \mathbb{E} \left\{ e^{i\langle t, X \rangle} \right\} = e^{t^2/2}.$$

(cf. (1.2.4).) It is then trivial that

$$\text{Cov}(f(X), g(X)) = \varphi(s+t) - \varphi(s)\varphi(t).$$

---

<sup>11</sup>Or at least one of RA's favourite proofs. Indeed this approach was used in RA's lecture notes [2]. There, however, there is a problem, for in the third line from the bottom of page 46 appear the words "To complete the proof note simply that...". The word "simply" is simply out of place, and, in fact, Lemma 2.2 there is false as stated. A correction (with the help Amir Dembo) appears, *en passant*, in the 'Proof of Theorem 2.3.1' below.

On the other hand, computing the integral in (2.3.12) gives

$$\begin{aligned}
& \int_0^1 \mathbb{E} \left\{ \left\langle \nabla f(X), \nabla g(\alpha X + \sqrt{1-\alpha^2}Y) \right\rangle \right\} d\alpha \\
&= \int_0^1 d\alpha \mathbb{E} \left\{ \left\langle \left( it_j e^{i(t,X)} \right)_j, \left( is_j e^{i(s,\alpha X + \sqrt{1-\alpha^2}Y)} \right)_j \right\rangle \right\} \\
&= - \int_0^1 d\alpha \sum_j \mathbb{E} \left\{ e^{i(t+\alpha s, X)} \right\} \mathbb{E} \left\{ e^{i(s, \sqrt{1-\alpha^2}Y)} \right\} \\
&= - \int_0^1 d\alpha \langle s, t \rangle e^{(|t|^2 + 2\alpha\langle s, t \rangle + |s|^2)/2} \\
&= -\varphi(s)\varphi(t) \left( 1 - e^{\langle s, t \rangle} \right) \\
&= \varphi(s+t) - \varphi(s)\varphi(t),
\end{aligned}$$

which is all that we need.  $\square$

**Lemma 2.3.5** *Let  $X$  be as in Lemma 2.3.4. If  $h : \mathbb{R}^k \rightarrow \mathbb{R}$  is Lipschitz, with Lipschitz constant 1 - i.e.  $|h(x) - h(y)| \leq |x - y|$  for all  $x, y \in \mathbb{R}^k$  - and if  $\mathbb{E}\{h(X)\} = 0$  then, for all  $t > 0$ ,*

$$(2.3.13) \quad \mathbb{E} \left\{ e^{th(X)} \right\} \leq e^{t^2/2}.$$

**Proof.** Let  $Y$  be an independent copy of  $X$  and  $\alpha$  a uniform random variable on  $[0, 1]$ . Define the pair  $(X, Z)$  via

$$(X, Z) \triangleq \left( X, \alpha X + \sqrt{1-\alpha^2}Y \right).$$

Take  $h$  as in the statement of the Lemma,  $t \geq 0$  fixed, and define  $g = e^{th}$ . Applying (2.3.12) (with  $f = h$ ) gives

$$\begin{aligned}
\mathbb{E} \{ h(X)g(X) \} &= \mathbb{E} \{ \langle \nabla h(X), \nabla g(Z) \rangle \} \\
&= t E \left\{ \langle \nabla h(X), \nabla h(Z) \rangle e^{th(Z)} \right\} \\
&\leq t E \left\{ e^{th(Z)} \right\},
\end{aligned}$$

using the Lipschitz property of  $h$ . Let  $u$  be the function defined by

$$e^{u(t)} = E \left\{ e^{th(X)} \right\}.$$

Then

$$\mathbb{E} \left\{ h(X)e^{th(X)} \right\} = u'(t) e^{u(t)},$$

so that from the preceding inequality  $u'(t) \leq t$ . Since  $u(0) = 0$  it follows that  $u(t) \leq t^2/2$  and we are done.  $\square$

The following Lemma gives the crucial step towards proving the Borell-TIS inequality.

**Lemma 2.3.6** *Let  $X$  be a  $k$ -dimensional vector of centered, unit variance, independent, Gaussian variables. If  $h : \mathbb{R}^k \rightarrow \mathbb{R}$  is Lipschitz, with Lipschitz constant  $\sigma$  then, for all  $u > 0$ ,*

$$(2.3.14) \quad \mathbb{P}\{|h(X) - \mathbb{E}\{h(X)\}| > u\} \leq 2e^{-\frac{1}{2}u^2/\sigma^2}.$$

**Proof.** By scaling it suffices to prove the result for  $\sigma = 1$ . By symmetry we can apply (2.3.13) also to  $-h$ . A Chebyshev argument and (2.3.13) therefore immediately yield that for every  $t, u > 0$

$$\mathbb{P}\{|h(X) - \mathbb{E}\{h(X)\}| > u\} \leq 2e^{\frac{1}{2}t^2 - tu}.$$

Taking the optimal choice of  $t = u$  gives (2.3.14) and we are done.  $\square$

We now have all we need for the

**Proof of Theorem 2.3.1** We have two things to prove. Firstly, Theorem 2.3.1 will follow immediately from Lemma 2.3.6 in the case of finite  $T$  and  $f$  having i.i.d. components once we show that  $\sup(\cdot)$ , or  $\max(\cdot)$  in this case, is Lipschitz. We shall show this, and lift the i.i.d. restriction, in one step. The second part of the proof involves lifting the result from finite to general  $T$ .

Thus, suppose  $T$  is finite, in which case we can write it as  $\{1, \dots, k\}$ . Let  $C$  be the  $k \times k$  covariance matrix of  $f$  on  $T$ , with components  $c_{ij} = \mathbb{E}\{f_i f_j\}$ , so that

$$\sigma_T^2 = \sup_{1 \leq i \leq k} c_{ii} = \sup_{1 \leq i \leq k} \mathbb{E}\{f_i^2\}.$$

Let  $A$  be such that  $A' \cdot A = C$ , so that  $f \stackrel{\mathcal{L}}{=} AW_1$ , and  $\max_i f_i \stackrel{\mathcal{L}}{=} \max_i (AW_1)_i$ .

Consider the function  $h(x) = \max_i (Ax)_i$ . Then

$$\begin{aligned} \left| \max_i (Ax)_i - \max_i (Ay)_i \right| &= \left| \max_i (e_i Ax) - \max_i (e_i Ay) \right| \\ &\leq \max_i |e_i A(x - y)| \\ &\leq \max_i |e_i A| \cdot \|x - y\|, \end{aligned}$$

where, as usual,  $e_i$  is the vector with 1 in position  $i$  and zeroes elsewhere. The first inequality above is elementary and the second is Cauchy-Schwartz. But

$$|e_i A|^2 = e_i' A' A e_i = e_i' C e_i = c_{ii},$$

so that

$$\left| \max_i (Ax)_i - \max_i (Ay)_i \right| \leq \sigma_T |x - y|.$$

In view of the equivalence in law of  $\max_i f_i$  and  $\max_i (AW_1)_i$  and Lemma 2.3.6, this establishes the Theorem for finite  $T$ .

We now turn to lifting the result from finite to general  $T$ . This is, almost, an easy exercise in approximation. For each  $n > 0$  let  $T_n$  be a finite subset of  $T$  such that  $T_n \subset T_{n+1}$  and  $T_n$  increases to a dense subset of  $T$ . By separability,

$$\sup_{t \in T_n} f_t \xrightarrow{a.s.} \sup_{t \in T} f_t,$$

and, since the convergence is monotone, we also have that

$$\mathbb{E} \left\{ \sup_{t \in T_n} f_t \right\} \rightarrow \mathbb{E} \left\{ \sup_{t \in T} f_t \right\}.$$

Since  $\sigma_{T_n}^2 \rightarrow \sigma_T^2 < \infty$ , (again monotonely) this would be enough to prove the general version of the Borell-TIS inequality from the finite  $T$  version if only we knew that the one worrisome term,  $\mathbb{E}\{\sup_T f_t\}$ , were definitely finite, as claimed in the statement of the Theorem. Thus if we show that the assumed a.s. finiteness of  $\|f\|$  implies also the finiteness of its mean, we shall have a complete proof to both parts of the Theorem.

We proceed by contradiction. Thus, assume  $E\{\|f\|\} = \infty$ , and choose  $u_o > 0$  such that

$$e^{-u_o^2/\sigma_T^2} \leq \frac{1}{4} \quad \text{and} \quad \mathbb{P} \left\{ \sup_{t \in T} f_t < u_o \right\} \geq \frac{3}{4}.$$

Now choose  $n \geq 1$  such that  $\mathbb{E}\{\|f\|_{T_n}\} > 2u_o$ , possible since  $\mathbb{E}\{\|f\|_{T_n}\} \rightarrow \mathbb{E}\{\|f\|_T\} = \infty$ . The Borell-TIS inequality on the finite space  $T_n$  then gives

$$\begin{aligned} \frac{1}{2} &\geq 2e^{-u_o^2/\sigma_T^2} &&\geq 2e^{-u_o^2/\sigma_{T_n}^2} \\ &&&\geq \mathbb{P}\{|\|f\|_{T_n} - \mathbb{E}\{\|f\|_{T_n}\}| > u_o\} \\ &&&\geq \mathbb{P}\{\mathbb{E}\{\|f\|_{T_n}\} - \|f\|_T > u_o\} \\ &&&\geq \mathbb{P}\{\|f\|_T < u_o\} \\ &&&\geq \frac{3}{4}. \end{aligned}$$

This provides the required contradiction, and so we are done.  $\square$

## 2.4 Comparison inequalities

The theory of Gaussian processes is rich in comparison inequalities, where by this term we mean results of the form “if  $f$  is a ‘rougher’ process than

$g$ , and both are defined over the same parameter space, then  $\|f\|$  will be ‘larger’ than  $\|g\|$ ”. The most basic of these is Slepian’s inequality.

**Theorem 2.4.1 (Slepian’s inequality)** *If  $f$  and  $g$  are a.s. bounded, centered Gaussian processes on  $T$  such that  $\mathbb{E}\{f_t^2\} = \mathbb{E}\{g_t^2\}$  for all  $t \in T$  and*

$$(2.4.1) \quad \mathbb{E}\{(f_t - f_s)^2\} \leq \mathbb{E}\{(g_t - g_s)^2\}$$

for all  $s, t \in T$ , then for all real  $u$

$$(2.4.2) \quad \mathbb{P}\{\|f\| > u\} \leq \mathbb{P}\{\|g\| > u\}.$$

Furthermore,

$$(2.4.3) \quad \mathbb{E}\{\|f\|\} \leq \mathbb{E}\{\|g\|\}.$$

Slepian’s inequality is so natural, that it hardly seems to require a proof, and hardly the rather analytic, non-probabilistic one that will follow. To see that there is more to the story than meets the eye, one need only note that (2.4.2) does *not* follow from (2.4.1) if we replace<sup>12</sup>  $\sup_T f_t$  by  $\sup_T |f_t|$  and  $\sup_T g_t$  by  $\sup_T |g_t|$ .

Slepian’s inequality is based on the following technical Lemma. the proof of which, in all its important details, goes back to Slepian’s original paper [88].

**Lemma 2.4.2** *Let  $f_1, \dots, f_k$  be centered Gaussian variables with covariance matrix  $C = (c_{ij})_{i,j=1}^k$ ,  $c_{ij} = \mathbb{E}\{f_i f_j\}$ . Let  $h : \mathbb{R}^k \rightarrow \mathbb{R}$  be  $C^2$ , and assume that, together with its derivatives, it satisfies a  $O(|x|^d)$  growth condition at infinity for some finite  $d$ <sup>13</sup>. Let*

$$(2.4.4) \quad \mathcal{H}(C) = \mathbb{E}\{h(f_1, \dots, f_k)\},$$

and assume that for a pair  $(i, j)$ ,  $1 \leq i < j \leq k$

$$(2.4.5) \quad \frac{\partial^2 h(x)}{\partial x_i \partial x_j} \geq 0$$

for all  $x \in \mathbb{R}^k$ . Then  $\mathcal{H}(C)$  is an increasing function of  $c_{ij}$ .

<sup>12</sup>For a counterexample to “Slepian’s inequality for absolute values” take  $T = \{1, 2\}$ , with  $f_1$  and  $f_2$  standard normal with correlation  $\rho$ . Writing  $P_\rho(u)$  for the probability under correlation  $\rho$  that  $\max(|f_1|, |f_2|) > u$ , it is easy to check that, for all  $u > 0$ ,  $P_{-1}(u) < P_0(u)$ , while  $P_0(u) > P_1(u)$ , which negates the monotonicity required by Slepian’s inequality.

<sup>13</sup>We could actually manage with  $h$  twice differentiable only in the sense of distributions. This would save the approximation argument following (2.4.7) below, and would give a neater, albeit slightly more demanding, proof of Slepian’s inequality, as in [57].

**Proof.** We have to show that

$$\frac{\partial \mathcal{H}(C)}{\partial c_{ij}} \geq 0$$

whenever  $\partial^2 h / \partial x_i \partial x_j \geq 0$ .

To make our lives a little easier we assume that  $C$  is non-singular, so that it makes sense to write  $\varphi(x) = \varphi_C(x)$  for the centered Gaussian density on  $\mathbb{R}^k$  with covariance matrix  $C$ . Straightforward algebra<sup>14</sup> shows that

$$(2.4.6) \quad \frac{\partial \varphi}{\partial c_{ii}} = \frac{1}{2} \frac{\partial^2 \varphi}{\partial x_i^2}, \quad \frac{\partial \varphi}{\partial c_{ij}} = \frac{\partial^2 \varphi}{\partial x_i \partial x_j}, \quad i \neq j.$$

Applying this and our assumptions on  $h$  to justify two integrations by parts, we obtain

$$\frac{\partial \mathcal{H}(C)}{\partial c_{ij}} = \int_{\mathbb{R}^k} h(x) \frac{\partial \varphi(x)}{\partial c_{ij}} dx = \int_{\mathbb{R}^k} \frac{\partial^2 h(x)}{\partial x_i \partial x_j} \varphi(x) dx \geq 0.$$

This completes the proof for the case of non-singular  $C$ . The general case can be handled by approximating a singular  $C$  via a sequence of non-singular covariance matrices.  $\square$

**Proof of Theorem 2.4.1** By separability, and the final argument in the proof of the Borell-TIS inequality, it suffices to prove (2.4.2) for  $T$  finite. Note that since  $\mathbb{E}\{f_t^2\} = \mathbb{E}\{g_t^2\}$  for all  $t \in T$ , (2.4.1) implies that  $\mathbb{E}\{f_s f_t\} \geq \mathbb{E}\{g_s g_t\}$  for all  $s, t \in T$ . Let  $h(x) = \prod_{i=1}^k h_i(x_i)$ , where each  $h_i$  is a positive non-increasing,  $C^2$  function satisfying the growth conditions placed on  $h$  in the statement of the Theorem, and  $k$  is the number of points in  $T$ . Note that, for  $i \neq j$

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} = h'_i(x_i) h'_j(x_j) \prod_{\substack{n \neq i \\ n \neq j}} h_n(x_n) \geq 0,$$

since both  $h'_i$  and  $h'_j$  are non-positive. It therefore follows from Lemma 2.4.2 that

$$(2.4.7) \quad \mathbb{E} \left\{ \prod_{i=1}^k h_i(f_i) \right\} \geq \mathbb{E} \left\{ \prod_{i=1}^k h_i(g_i) \right\}.$$

Now take  $\{h_i^{(n)}\}_{n=1}^\infty$  to be a sequence of positive, non-increasing,  $C^2$  approximations to the indicator function of the interval  $(-\infty, \lambda]$ , to derive that

$$\mathbb{P}\{\|f\| < u\} \geq \mathbb{P}\{\|g\| < u\},$$

---

<sup>14</sup>This is really the same algebra that needed to justify (??) in the proof of the Borell-TIS inequality. That is, it is the heat equation again.

which implies (2.4.2).

To complete the proof, all that remains is to show that (2.4.2) implies (2.4.3). But this is a simple consequence of integration by parts, since

$$\begin{aligned}\mathbb{E}\{\|f\|\} &= \int_0^\infty \mathbb{P}\{\|f\| > u\} du - \int_{-\infty}^0 \mathbb{P}\{\|f\| < u\} du \\ &\leq \int_0^\infty \mathbb{P}\{\|g\| > u\} du - \int_{-\infty}^0 \mathbb{P}\{\|g\| < u\} du \\ &= \mathbb{E}\|g\|.\end{aligned}$$

This completes the proof.  $\square$

As mentioned above, there are many extensions of Slepian's inequality, the most important of which is probably the following.

**Theorem 2.4.3 (Sudakov-Fernique inequality)** *Let  $f$  and  $g$  be a.s. bounded, centered Gaussian processes on  $T$ . Then (2.4.1) implies (2.4.3).*

In other words, a Slepian-like inequality holds without a need to assume identical variance for the compared processes. However, in this case we have only the weaker ordering of expectations of (2.4.3) and not the stochastic domination of (2.4.2).

Since we shall not need the Sudakov-Fernique inequality we shall not bother proving it. Proofs can be found in all of the references at the head of this Chapter. There you can also find out how to extend the above arguments to find conditions on covariance functions that allow statements of the form

$$\mathbb{P}\left\{\min_{1 \leq i \leq n} \max_{1 \leq j \leq m} X_{ij} \geq u\right\} \geq \mathbb{P}\left\{\min_{1 \leq i \leq n} \max_{1 \leq j \leq m} Y_{ij} > u\right\},$$

along with even more extensive variations due, originally, to Gordon [38]. Gordon [39] also shows how to extend the essentially Gaussian computations above to elliptically contoured distributions.

## 2.5 Orthogonal expansions

While most of what we shall have to say in this Section is rather theoretical, it actually covers one of the most important practical aspects of Gaussian modelling. The basic result of the Section is Theorem 2.5.1, which states that *every* centered Gaussian process with a continuous covariance function has an expansion of the form

$$(2.5.1) \quad f(t) = \sum_{n=1}^{\infty} \xi_n \varphi_n(t),$$

where the  $\xi_n$  are i.i.d.  $N(0, 1)$ , and the  $\varphi_n$  are certain functions on  $T$  determined by the covariance function  $C$  of  $f$ . In general, the convergence in (2.5.1) is in  $L^2(\mathbb{P})$  for each  $t \in T$ , but (Theorem 2.5.2) if  $f$  is a.s. continuous then the convergence is uniform over  $T$ , with probability one.

There are many theoretical conclusions that follow from this representation. For one example, note that since continuity of  $C$  will imply that of the  $\varphi_n$  (cf. Lemma (2.5.4)) it follows from (2.5.1) that sample path continuity of  $f$  is a “tail event” on the  $\sigma$ -algebra determined by the  $\xi_n$ , from which one can show that centered Gaussian processes are either continuous with probability one, or discontinuous with probability one. There is no middle ground.

The practical implications of (2.5.1) are mainly in the area of simulation. If one needs to simulate a stationary process on a Euclidean space, then the standard technique is to take the Spectral Representation of Theorem 1.4.4, approximate the stochastic integral there with a sum of sinusoids with random coefficients as in (1.4.25), taking as many terms in the sum as is appropriate<sup>15</sup>. However, not all parameter spaces are Euclidean, and perhaps more importantly, not all fields are stationary. In the latter case, in particular, another approach is needed.

Such an approach is furnished by (2.5.1). Again truncating the sum at a point appropriate to the problem at hand, one needs “only” to determine the  $\varphi_n$ . As we shall soon see, these arise as the orthonormal basis of a particular Hilbert space, and can generally be found by solving an eigenfunction problem involving  $C$ . In the latter case, and when  $T$  is a nice subset of  $\mathbb{R}^N$ , this leads to the Karhunen-Loève expansion of (2.5.12). While even in the Euclidean situation there are only a handful of situations for which this eigenfunction problem can be solved analytically, from the point of view of computing it is a standard problem, and the approach is practical. If one does not want to solve an eigenvalue problem, then it suffices to find an appropriate orthonormal basis, an easier task than it might seem at first. Indeed, the main practical problem here is the plethora of possible choices. In the stationary case, as we shall see, the complex exponentials are a natural choice, and then we return, albeit in a rather indirect fashion, to the results of spectral theory.

The first step towards establishing the expansion (2.5.1) lies in setting up the so-called *reproducing kernel Hilbert space* (RKHS) of a centered Gaussian process with covariance function  $C$ .

In essence, the RKHS is made up of functions that have about the same smoothness properties that  $C(s, t)$  has, as a function in  $t$  for fixed  $s$ , or vice

---

<sup>15</sup>What is “appropriate” depends, of course, on the problem one is working on. For example, for visualising a two-parameter field on a computer screen, there is no point taking more terms in the (two-dimensional) Fourier sum than there are pixels on the screen.

versa. Start with

$$S = \left\{ u : T \rightarrow \mathbb{R} : u(\cdot) = \sum_{i=1}^n a_i C(s_i, \cdot), \ a_i \text{ real}, \ s_i \in T, \ n \geq 1 \right\}.$$

Define an inner product on  $S$  by

$$\begin{aligned} (u, v)_H &= \left( \sum_{i=1}^n a_i C(s_i, \cdot), \sum_{j=1}^m b_j C(s_j, \cdot) \right)_H \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j C(s_i, s_j). \end{aligned}$$

The fact that  $C$  is non-negative definite implies  $(u, u)_H \geq 0$  for all  $u \in S$ . Furthermore, note that the inner product (2.5.2) has the following unusual property:

$$\begin{aligned} (2.5.2) \quad (u, C(t, \cdot))_H &= \left( \sum_{i=1}^n a_i C(s_i, \cdot), C(t, \cdot) \right)_H \\ &= \sum_{i=1}^n a_i C(s_i, t) \\ &= u(t). \end{aligned}$$

This is the *reproducing kernel* property.

For the sake of exposition, assume that the covariance function,  $C$ , is positive definite (rather than merely non-negative definite) so that  $(u, u)_H = 0$  if, and only if,  $u(t) \equiv 0$ . In this case (2.5.2) defines a norm  $\|u\|_H = (u, u)_H^{1/2}$ .

For  $\{u_n\}_{n \geq 1}$  a sequence in  $S$  we have

$$\begin{aligned} |u_n(t) - u_m(t)|^2 &= |(u_n - u_m, C(t, \cdot))_H|^2 \\ &\leq \|u_n - u_m\|_H^2 \|C(t, \cdot)\|_H^2 \\ &\leq \|u_n - u_m\|_H^2 C(t, t), \end{aligned}$$

the last line following directly from (2.5.2). Thus it follows that if  $\{u_n\}$  is Cauchy in  $\|\cdot\|_H$  then it is pointwise Cauchy. The closure of  $S$  under this norm is a space of real-valued functions, denoted by  $H(C)$ , and called the RKHS of  $f$  or of  $C$ , since every  $u \in H(C)$  satisfies (2.5.2) by the separability of  $H(C)$ . (The separability of  $H(C)$  follows from the separability of  $T$  and the assumption that  $C$  is continuous.)

Since all this seems at first rather abstract, consider two concrete examples. Take  $T = \{1, \dots, n\}$ , finite, and  $f$  centered Gaussian with covariance matrix  $C = (c_{ij})$ ,  $c_{ij} = \mathbb{E}\{f_i f_j\}$ . Let  $C^{-1} = (c^{ij})$  denote the inverse of  $C$ ,

which exists by positive definiteness. Then the RKHS of  $f$  is made up of all  $n$ -dimensional vectors  $u = (u_1, \dots, u_n)$  with inner product

$$(u, v)_H = \sum_{i=1}^n \sum_{j=1}^n u_i c^{ij} v_j.$$

To prove this, we need only check that the reproducing kernel property (2.5.2) holds<sup>16</sup>. But, with  $\delta(i, j)$  the Kronecker delta function, and  $C_k$  denoting the  $k$ -th row of  $C$ ,

$$\begin{aligned} (u, C_k)_H &= \sum_{i=1}^n \sum_{j=1}^n u_i c^{ij} c_{kj} \\ &= \sum_{i=1}^n u_i \delta(i, k) \\ &= u_k, \end{aligned}$$

as required.

For a slightly more interesting example, take  $f = W$  to be standard Brownian motion on  $T = [0, 1]$ , so that  $C(s, t) = \min(s, t)$ . Note that  $C(s, \cdot)$  is differentiable everywhere except at  $s$ , so that following the heuristics developed above we expect that  $H(C)$  should be made up of a subset of functions that are differentiable almost everywhere.

To both make this statement more precise, and prove it, we start by looking at the space  $S$ . Thus, let

$$u(t) = \sum_{i=1}^n a_i C(s_i, t), \quad v(t) = \sum_{i=1}^n b_i C(t_i, t),$$

be two elements of  $S$ , with inner product

$$(u, v)_H = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \min(s_i, t_j).$$

Since the derivative of  $C(s, t)$  with respect to  $t$  is  $\mathbb{1}_{[0, s]}(t)$ , the derivative of  $u$  is  $\sum_{i=1}^n a_i \mathbb{1}_{[0, s_i]}(t)$ . Therefore,

$$\begin{aligned} (u, v)_H &= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \int_0^1 \mathbb{1}_{[0, s_i]}(t) \mathbb{1}_{[0, t_j]}(t) dt \\ &= \int_0^1 \left( \sum_{i=1}^n a_i \mathbb{1}_{[0, s_i]}(t) \sum_{j=1}^n b_j \mathbb{1}_{[0, t_j]}(t) \right) dt \\ &= \int_0^1 \dot{u}(t) \dot{v}(t) dt. \end{aligned}$$

---

<sup>16</sup>A simple proof by contradiction shows that there can never be two different inner products on  $S$  with the reproducing kernel property.

With  $S$  under control, we can now look for an appropriate candidate for the RKHS. Define

$$(2.5.3) \quad H = \left\{ u : u(t) = \int_0^t \dot{u}(s) ds, \int_0^1 (\dot{u}(s))^2 ds < \infty \right\},$$

equipped with the inner product

$$(2.5.4) \quad (u, v)_H = \int_0^1 \dot{u}(s) \dot{v}(s) ds.$$

Since it is immediate that  $C(s, \cdot) \in H$  for  $t \in [0, 1]$ , and

$$(u, C(t, \cdot))_H = \int_0^1 \dot{u}(s) \mathbb{1}_{[0,t]}(s) ds,$$

it follows that the  $H$  defined by (2.5.3) is indeed our RKHS. This  $H$  is also known, in the setting of diffusion processes, as a *Cameron-Martin* space.

With a couple of examples under our belt, we can now return to our main task: setting up the expansion (2.5.1). The first step is finding a countable orthonormal basis for the separable Hilbert space  $H(C)$ . Recall, from the the proof of the Spectral Representation Theorem, the space  $\mathcal{H} = \text{span}\{f_t, t \in \mathbb{R}^N\}$ . Analogously to (1.4.21) define a linear mapping  $\Theta : S \rightarrow \mathcal{H}$  by

$$\Theta(u) = \Theta \left( \sum_{i=1}^n a_i C(t_i, \cdot) \right) = \sum_{i=1}^n a_i f(t_i).$$

Clearly  $\Theta(u)$  is Gaussian for each  $u \in S$ .

Since  $\Theta$  is norm preserving, it extends to all of  $H(C)$  with range equal to all of  $\mathcal{H}$ , with all limits remaining Gaussian. This extension is called the *canonical isomorphism* between these spaces.

Since  $H(C)$  is separable, we now also know that  $\mathcal{H}$  is, and proceed to to build an orthonormal basis for it. If  $\{\varphi_n\}_{n \geq 1}$  is an orthonormal basis for  $H(C)$ , then setting  $\xi_n = \Theta(\varphi_n)$  gives  $\{\xi_n\}_{n \geq 1}$  as an orthonormal basis for  $\mathcal{H}$ . In particular, we must have that the  $\xi_n$  are  $N(0, 1)$  and

$$(2.5.5) \quad f_t = \sum_{n=1}^{\infty} \xi_n \mathbb{E}\{f_t \xi_n\},$$

where the series converges in  $L^2(\mathbb{P})$ . Since  $\Theta$  was an isometry, it follows from (2.5.5) that

$$(2.5.6) \quad \begin{aligned} \mathbb{E}\{f_t \xi_n\} &= (C(t, \cdot), \varphi_n)_H \\ &= \varphi_n(t), \end{aligned}$$

the last equality coming from the reproducing kernel property of  $H(C)$ . Putting (2.5.6) together with (2.5.5) now establishes the following central result.

**Theorem 2.5.1** *If  $\{\varphi_n\}_{n \geq 1}$  is an orthonormal basis for  $H(C)$ , then  $f$  has the  $L^2$ -representation*

$$(2.5.7) \quad f_t = \sum_{n=1}^{\infty} \xi_n \varphi_n(t),$$

where  $\{\xi_n\}_{n \geq 1}$  is the orthonormal sequence of centered Gaussian variables given by  $\xi_n = \Theta(\varphi_n)$ .

The equivalence in (2.5.7) is only in  $L^2$ ; i.e. the sum is, in general, convergent, for each  $t$ , only in mean square. The following result shows that much more is true if we know that  $f$  is a.s. continuous.

**Theorem 2.5.2** *If  $f$  is a.s. continuous, then the sum in (2.5.7) converges uniformly on  $T$  with probability one<sup>17</sup>.*

We need two preliminary results before we can prove Theorem 2.5.2. The first is a convergence result due to Itô and Nisio [47] which, since it is not really part of a basic probability course we state in full, and the second an easy Lemma.

**Lemma 2.5.3** *Let  $\{Z_n\}_{n \geq 1}$  be a sequence of symmetric independent random variables, taking values in a separable, real Banach space  $B$ , equipped with the norm topology. Let  $X_n = \sum_{i=1}^n Z_i$ . Then  $X_n$  converges with probability one if, and only if, there exists a  $B$ -valued random variable  $X$  such that  $\langle X_n, x^* \rangle \rightarrow \langle X, x^* \rangle$  in probability for every  $x^* \in B^*$ , the topological dual of  $B$ .*

**Lemma 2.5.4** *Let  $\{\varphi_n\}_{n \geq 1}$  be an orthonormal basis for  $H(C)$ . Then each  $\varphi_n$  is continuous and*

$$(2.5.8) \quad \sum_{n=1}^{\infty} \varphi_n^2(t)$$

converges uniformly in  $t \in T$  to  $C(t, t)$ .

**Proof.** Note that

$$\begin{aligned} |\varphi_n(s) - \varphi_n(t)|^2 &= |(C(s, \cdot), \varphi_n(\cdot))_H - (C(t, \cdot), \varphi_n(\cdot))_H|^2 \\ &= |([C(s, \cdot) - C(t, \cdot)], \varphi_n(\cdot))_H|^2 \\ &\leq \|\varphi_n\|_H \|C(s, \cdot) - C(t, \cdot)\|_H \\ &= \|C(s, \cdot) - C(t, \cdot)\|_H \\ &= C(s, s) - 2C(s, t) + C(t, t), \end{aligned}$$

---

<sup>17</sup>There is also a converse to Theorem 2.5.2, that the a.s. uniform convergence of a sum like (2.5.7) implies the continuity of  $f$ , and some of the earliest derivations (e.g. [37]) of sufficient conditions for continuity actually used this approach. Entropy based arguments, however, turn out to give much easier proofs.

where the first and last equalities use the reproducing kernel property and the one inequality is Cauchy-Schwartz. Since  $C$  is pointwise continuous it now follows from the definition of  $\|\cdot\|_H$  on  $S$  that each  $\varphi_n$  is continuous over  $T$ .

To establish the uniform convergence of (2.5.8), note that the orthonormal expansion and the reproducing kernel property imply

$$(2.5.9) \quad \begin{aligned} C(t, \cdot) &= \sum_{n=1}^{\infty} \varphi_n(\cdot) (C(t, \cdot), \varphi_n)_H \\ &= \sum_{n=1}^{\infty} \varphi_n(\cdot) \varphi_n(t), \end{aligned}$$

convergence of the sum being in the  $\|\cdot\|_H$  norm. Hence,  $\sum_{n=1}^{\infty} \varphi_n^2(t)$  converges to  $C(t, t)$  for every  $t \in T$ . Furthermore, the convergence is monotone, and so it follows that it is also uniform ( $\equiv$  Dini's theorem).  $\square$

**Proof of Theorem 2.5.2** We know that, for each  $t \in T$ ,  $\sum_{n=1}^{\infty} \xi_n \varphi_n(t)$  is a sum of independent variables converging in  $L^2(\mathbb{P})$ . Thus, by Lemma 2.5.3, applied to real-valued random variables, it converges with probability one to a limit, which we denote by  $f_t$ . The limit process is, by assumption, almost surely continuous.

Now, consider both  $f$  and each function  $\xi_n \varphi_n$  as random variables in the Banach space  $C(T)$ , with sup-norm topology. Elements of the dual  $C^*(T)$  are therefore finite, signed, Borel measures  $\mu$  on  $T$ , and  $\langle f, \mu \rangle = \int f d\mu$ . Define

$$f_n(\cdot) = \sum_{i=1}^n \xi_i \varphi_i(\cdot) = \sum_{i=1}^n \Theta(\varphi_i) \varphi_i(\cdot).$$

By Lemma 2.5.3, it suffices to show that for every  $\mu \in C^*(T)$  the random variables  $\langle f_n, \mu \rangle$  converge in probability to  $\langle f, \mu \rangle$ . However,

$$\begin{aligned} \mathbb{E} \{ |\langle f_n, \mu \rangle - \langle f, \mu \rangle| \} &= \mathbb{E} \left\{ \left| \int_T (f_n(t) - f(t)) \mu(dt) \right| \right\} \\ &\leq \int_T \mathbb{E} \{ |f_n(t) - f(t)| \} |\mu|(dt) \\ &\leq \int_T \left[ \mathbb{E} (f_n(t) - f(t))^2 \right]^{\frac{1}{2}} |\mu|(dt) \\ &= \int_T \left( \sum_{j=n+1}^{\infty} \varphi_j^2(t) \right)^{\frac{1}{2}} |\mu|(dt), \end{aligned}$$

where  $|\mu|(A)$  is the total variation measure for  $\mu$ .

Since  $\sum_{j=n+1}^{\infty} \varphi_j^2(t) \rightarrow 0$  uniformly in  $t \in T$  by Lemma 2.5.4, the last expression above tends to zero as  $n \rightarrow \infty$ . Since this implies the convergence in probability of  $\langle f_n, \mu \rangle$  to  $\langle f, \mu \rangle$ , we are done.  $\square$

### 2.5.1 Karhunen-Loève expansion

As we have already noted, applying the orthogonal expansion (2.5.7) in practice relies on being able to find the orthonormal functions  $\varphi_n$ . When  $T$  is a compact subset of  $\mathbb{R}^N$  there is a special way in which to do this, leading to what is known as the Karhunen-Loève expansion.

For simplicity, take  $T = [0, 1]^N$ . Let  $\lambda_1 \geq \lambda_2 \geq \dots$ , and  $\psi_1, \psi_2, \dots$ , be, respectively, the eigenvalues and normalised eigenfunctions of the operator  $\mathcal{C} : L^2(T) \rightarrow L^2(T)$  defined by  $(\mathcal{C}\psi)(t) = \int_T C(s, t)\psi(s) ds$ . That is, the  $\lambda_n$  and  $\psi_n$  solve the integral equation

$$(2.5.10) \quad \int_T C(s, t)\psi(s) ds = \lambda\psi(t),$$

with the normalisation

$$\int_T \psi_n(t)\psi_m(t) dt = \begin{cases} 1 & n = m, \\ 0 & n \neq m. \end{cases}$$

These eigenfunctions lead to a natural expansion of  $C$ , known as Mercer's Theorem. (cf. [79, 114] for a proof.)

**Theorem 2.5.5 (Mercer)** *Let  $C$ ,  $\{\lambda_n\}_{n \geq 1}$  and  $\{\psi_n\}_{n \geq 1}$  be as above. Then*

$$(2.5.11) \quad C(s, t) = \sum_{n=1}^{\infty} \lambda_n \psi_n(s)\psi_n(t),$$

where the series converges absolutely and uniformly on  $[0, 1]^k \times [0, 1]^k$ .

The key to the Karhunen-Loève expansion is the following result.

**Lemma 2.5.6** *For  $f$  on  $[0, 1]^N$  as above,  $\{\sqrt{\lambda_n}\psi_n\}$  is a complete orthonormal system in  $H(C)$ .*

**Proof.** Set  $\varphi_n = \sqrt{\lambda_n}\psi_n$  and define

$$H = \left\{ h : h(t) = \sum_{n=1}^{\infty} a_n \varphi_n(t), t \in [0, 1]^N, \sum_{n=1}^{\infty} a_n^2 < \infty \right\}.$$

Give  $H$  the inner product

$$(h, g)_H = \sum_{n=1}^{\infty} a_n b_n,$$

where  $h = \sum a_n \varphi_n$  and  $g = \sum b_n \varphi_n$ .

To check that  $H$  has the reproducing kernel property, note that

$$\begin{aligned} (h(\cdot), C(t, \cdot))_H &= \left( \sum_{n=1}^{\infty} a_n \varphi_n(\cdot), \sum_{n=1}^{\infty} \sqrt{\lambda_n} \psi_n(t) \varphi_n(\cdot) \right) \\ &= \sum_{n=1}^{\infty} \sqrt{\lambda_n} a_n \psi_n(t) \\ &= h(t). \end{aligned}$$

It remains to be checked that  $H$  is in fact a Hilbert space, and that  $\{\sqrt{\lambda_n} \psi_n\}$  is both complete and orthonormal. But all this is standard, given Mercer's Theorem.  $\square$

We can now start rewriting things to get the expansion we want. Remaining with the basic notation of Mercer's theorem, we have that the RKHS,  $H(C)$ , consists of all square integrable functions  $h$  on  $[0, 1]^N$  for which

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} \left| \int_T h(t) \psi_n(t) dt \right|^2 < \infty,$$

with inner product

$$(h, g)_H = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \int_T h(t) \psi_n(t) dt \int_T g(t) \psi_n(t) dt.$$

The Karhunen-Loève expansion of  $f$  is obtained by setting  $\varphi_n = \lambda_n^{\frac{1}{2}} \psi_n$  in the orthonormal expansion (2.5.7), so that

$$(2.5.12) \quad f_t = \sum_{n=1}^{\infty} \lambda_n^{\frac{1}{2}} \xi_n \psi_n(t),$$

where the  $\xi_n$  are i.i.d.  $N(0, 1)$ .

As simple as this approach might seem, it is generally limited by the fact that it is usually not easy to analytically solve the integral equation (2.5.10). Here is one, classic, example – that of standard Brownian motion on  $[0, 1]$ , for which we have already characterised the corresponding RKHS as the Cameron-Martin space.

For Brownian motion (2.5.10) becomes

$$\begin{aligned} \lambda \psi(t) &= \int_0^1 \min(s, t) \psi(s) ds \\ &= \int_0^t s \psi(s) ds + t \int_t^1 \psi(s) ds. \end{aligned}$$

Differentiating both sides with respect to  $t$  gives

$$\begin{aligned}\lambda\psi'(t) &= \int_t^1 \psi(s) ds, \\ \lambda\psi''(t) &= -\psi(t),\end{aligned}$$

together with the boundary conditions  $\psi(0) = 0$  and  $\psi'(1) = 0$

The solutions of this pair of differential equations are given by

$$\psi_n(t) = \sqrt{2} \sin\left(\frac{1}{2}(2n+1)\pi s\right), \quad \lambda_n = \left(\frac{2}{(2n+1)\pi}\right)^2,$$

as is easily verified by substitution. Thus, the Karhunen-Loève expansion of Brownian motion on  $[0, 1]$  is given by

$$W_t = \frac{\sqrt{2}}{\pi} \sum_{n=0}^{\infty} \xi_n \left(\frac{2}{2n+1}\right) \sin\left(\frac{1}{2}(2n+1)\pi t\right).$$

For a class of important pseudo examples, pretend for the moment that the Karhunen-Loève expansion holds for a stationary process  $f$  defined on all of (non-compact)  $\mathbb{R}^N$ . As usual, in dealing with stationary processes, we also pretend that  $f$  is complex valued, so that the covariance function is  $C(s, t) = \mathbb{E}\{f_s \overline{f_t}\}$  which is a function of  $t - s$  only. It is then easy to find eigenfunctions for (2.5.10) via complex exponentials. Note that, for any  $\lambda \in \mathbb{R}^N$ , the function  $e^{i\lambda t}$  (a function of  $t \in \mathbb{R}^N$ ) satisfies

$$\begin{aligned}\int_{\mathbb{R}^N} C(s, t) e^{i\lambda s} ds &= \int_{\mathbb{R}^N} C(t - s) e^{i\lambda s} ds \\ &= e^{i\lambda t} \int_{\mathbb{R}^N} C(u) e^{-i\lambda u} du \\ &= K_\lambda e^{i\lambda t},\end{aligned}$$

for some, possibly zero,  $K_\lambda$ .

Suppose that  $K_\lambda \neq 0$  for only a countable number of  $\lambda$ . Then, comparing the Mercer expansion (2.5.11) and the Spectral Distribution Theorem (1.4.16) it is clear that we have recovered (1.4.16), for the case of a discrete spectral measure, from the Karhunen-Loève approach. The same is true of the spectral representation (1.4.19).

If  $K_\lambda \neq 0$  on an uncountable set then the situation becomes more delicate, but nevertheless actually yields the formal intuition behind setting up the stochastic integral that gave us the spectral representation of Theorem 1.4.4. In this sense, the Spectral Representation Theorem can be thought of as a special case of the Karhunen-Loève expansion.

As an aside, recall that we noted earlier that one of the uses of orthogonal expansions is as a simulation technique, in which case the expansion

is always taken to be finite. From the above comments, it follows that this provides a technique for simulating stationary fields which is exact if the spectral measure is concentrated on a finite number of points. It is interesting to note, however, that Karhunen-Loève expansions can never be finite if the field is also assumed to be isotropic. For a heuristic argument as to why this is the case, recall from Section 1.4.6 that under isotropy the spectral measure must be invariant under rotations, and so cannot be supported on a finite, or even countable, number of points. Consequently, one also needs a uncountable number of independent variables in the spectral noise process to generate the process via (1.4.19). However a process with a finite Karhunen-Loève expansion provides only a finite number of such variables, which can never be enough.

## 2.6 Majorising measures

Back in Section 2.1 we used the notion of entropy integrals to determine sufficient conditions for the boundedness and continuity of Gaussian fields. We claimed there that these arguments were sharp for stationary processes, but need not be sharp in general. That is, there are processes which *are* a.s. continuous, but whose entropy integrals diverge. We also noted that the path to solving these issues lay via *majorising measures*, and so we shall now explain what these are, and how they work.

Our plan here is to give only the briefest of introductions to majorising measures. You can find the full theory in the book by Ledoux and Talagrand [57] and the more recent papers [91, 92, 93] by Talagrand. In particular, [92] gives a very user friendly introduction to the subject, and the proof of Theorem 2.6.1 below is taken from there.

We include this Section for mathematical completeness<sup>18</sup>, and will not use its results anywhere else in the book. It is here mainly to give you an idea of how to improve on the weaknesses of entropy arguments, and to whet your appetite to turn to the sources for more. Here is the main result.

**Theorem 2.6.1** *If  $f$  is a centered Gaussian process on ( $d$ -compact)  $T$ , then there exists a universal constant  $K$  such that*

$$(2.6.1) \quad \mathbb{E} \left\{ \sup_{t \in T} f_t \right\} \leq K \inf_{\mu} \sup_{t \in T} \int_0^{\infty} \sqrt{\ln \frac{1}{\mu(B_d(t, \varepsilon))}} d\varepsilon,$$

where  $B_d$  is the  $d$ -ball of (2.1.2) and the infimum on  $\mu$  is taken over all probability measures  $\mu$  on  $T$ .

---

<sup>18</sup>“Mathematical completeness” should be understood in a relative sense, since our proofs here will most definitely be incomplete!

Furthermore, if  $f$  is a.s. bounded, then there exists a probability measure  $\mu$  and a universal constant  $K'$  such that

$$(2.6.2) \quad K' \sup_{t \in T} \int_0^\infty \sqrt{\ln \frac{1}{\mu(B_d(t, \varepsilon))}} d\varepsilon \leq \mathbb{E} \left\{ \sup_{t \in T} f_t \right\}.$$

A measure  $\mu$  for which the integrals above are finite for all  $t$  is called a “majorising measure”.

Note that the upper limit to the integrals in the Theorem is really  $\text{diam}(T)$ , since

$$\varepsilon > \text{diam}(T) \Rightarrow T \subset B_d(t, \varepsilon) \Rightarrow \mu(B_d(t, \varepsilon)) = 1,$$

and so the integrand is zero beyond this limit.

Theorem 2.6.1 is the majorising measure version of Theorem 2.1.3<sup>19</sup>, which gave an upper bound for  $\mathbb{E}\{\sup_t f_t\}$  based on an entropy integral, but which did not have a corresponding lower bound. Theorem 2.6.1, however, takes much more work to prove than its entropic relative. Nevertheless, by building on what we have already done, it is not all that hard to see where the upper bound (2.6.1) comes from.

**Outline of proof.** Start by re-reading the proof of Theorem 2.1.3 as far as (2.1.14). The argument leading up to (2.1.14) was that, eventually, increments  $f_{\pi_j(t)} - f_{\pi_{j-1}(t)}$  would be smaller than  $ua_j$ . However, on the way to this we could have been less wasteful in a number of our arguments.

For example, we could have easily arranged things so that

$$(2.6.3) \quad \forall s, t \in T, \quad \pi_j(t) = \pi_j(s) \Rightarrow \pi_{j-1}(t) = \pi_{j-1}(s),$$

which would have given us fewer increments to control. We could have also assumed that

$$(2.6.4) \quad \forall t \in \Pi_j, \quad \pi_j(t) = t,$$

so that, by (2.6.3),

$$(2.6.5) \quad \pi_{j-1}(t) = \pi_{j-1}(\pi_j(t)).$$

So let's assume both (2.6.3) and (2.6.4). Then controlling the increments  $f_{\pi_j(t)} - f_{\pi_{j-1}(t)}$  actually means controlling the the increments  $f_t - f_{\pi_{j-1}(t)}$  for  $t \in \Pi_j$ . There are only  $N_j$  such increments, which improves on our previous estimate of  $N_j N_{j-1}$ . This does not make much of a difference, but what does, and this is the core of the majorising measure argument, is

---

<sup>19</sup>There is a similiar extension of Theorem 2.1.5, but we shall not bother with it.

replacing the original  $a_j$  by families of non-negative numbers  $\{a_j(t)\}_{t \in \Pi_j}$ , and then to then ask when

$$(2.6.6) \quad \forall t \in \Pi_j, \quad f_t - f_{\pi_{j-1}(t)} \leq ua_j(t)$$

for large enough  $j$ . Note that under (2.6.6)

$$\forall t \in T, \quad f_t - f_{t_o} \leq uS,$$

where

$$S \triangleq \sup_{t \in T} \sum_{j>i} a_j(\pi_j(t)).$$

Thus

$$(2.6.7) \quad \mathbb{P} \left\{ \sup_{t \in T} (f_t - f_{t_o}) \geq uS \right\} \leq \sum_{j>i} \sum_{t \in \Pi'_j} \mathbb{P} \{ f_t - f_{\pi_{j-1}(t)} \geq ua_j(t) \} \\ \leq \sum_{j>i} \sum_{t \in \Pi'_j} 2 \exp \left( \frac{-u^2 a_j^2(t)}{(2r^{-j+1})^2} \right),$$

where  $\Pi'_j \triangleq \Pi_j \setminus \bigcup_{i < k \leq j-1} \Pi_k$ . (The move from the  $\Pi_j$  to the disjoint  $\Pi'_j$  is crucial, and made possible by (2.6.5).) This bound is informative only if the right hand side is less than or equal to one. Let's see how to ensure this when  $u = 1$ . Setting

$$(2.6.8) \quad w_j(t) = 2 \exp \left( \frac{-a_j^2(t)}{(2r^{-j+1})^2} \right)$$

we want to have  $\sum_j \sum_{t \in \Pi'_j} w_j(t) \leq 1$ . We are now getting close to our "majorising measure".

Recall that  $T$  has long ago been assumed countable. Suppose we have a probability measure  $\mu$  supported on  $T$  and for all  $j > i$  and all  $t \in \Pi_j$  set  $w_j(t) = \mu(\{t\})$ . Undo (2.6.8) to see that this means that we need to take

$$a_j(t) = 2r^{-j+1} \sqrt{\ln \frac{2}{\mu(\{t\})}}.$$

With this choice, the last sum in (2.6.7) is no more than  $2^{1-u^2}$ , and  $S$  is given by

$$S = 2 \sup_{t \in T} \sum_{j>i} r^{-j+1} \sqrt{\ln \frac{2}{\mu(\{\pi_j(t)\})}},$$

all of which ensures that for the *arbitrary* measure  $\mu$  we now have

$$(2.6.9) \quad \mathbb{E} \left\{ \sup_{t \in T} f_t \right\} \leq K \sup_{t \in T} \sum_{j>i} r^{-j+1} \sqrt{\ln \frac{2}{\mu(\{\pi_j(t)\})}}.$$

This is, in essence, the majorising measure upper bound. To make it look more like (2.6.1), note that each map  $\pi_j$  defines a partition  $\mathcal{A}_j$  of  $T$  comprising of the sets

$$A_t \triangleq \{s \in T : \pi_j(s) = t\}, \quad t \in \Pi_j.$$

With  $A_j(t)$  denoting the unique element of  $\mathcal{A}_j$  that contains  $t \in T$ , it is not too hard (but also not trivial) to reformulate (2.6.9) to obtain

$$(2.6.10) \quad \mathbb{E} \left\{ \sup_{t \in T} f_t \right\} \leq K \sup_{t \in T} \sum_{j>i} r^{-j+1} \sqrt{\ln \frac{2}{\mu(\{A_j(t)\})}}.$$

which is now starting to look a lot more like (2.6.1). To see why this reformulation works, you should go [92], which you should now be able to read without even bothering about notational changes. You will also find there how to turn (2.6.10) into (2.6.1), and also how to get the lower bound (2.6.2). All of this takes quite a bit of work, but at least now you should have some idea of how majorising measures arose.  $\square$

Despite the elegance of Theorem 2.6.1, it is not always easy, given a specific Gaussian process, to find the “right” majorising measure for it. To circumvent this problem, Talagrand recently [93] gave a recipe on how to wield the technique without the need to explicitly compute a majorising measure. However, we are already familiar with one situation in which there is a simple recipe for building majorising measures. This is when entropy integrals are finite.

Thus, let be  $H(\varepsilon)$  be our old friend (2.1.3), and set

$$(2.6.11) \quad g(t) \triangleq \sqrt{\ln \frac{1}{t}}, \quad 0 < t \leq 1.$$

Then here is a useful result linking entropy and majorising measures.

**Lemma 2.6.2** *If  $\int_0^\infty H^{1/2}(\varepsilon) d\varepsilon < \infty$ , then there exists a majorising measure  $\mu$  and a universal constant  $K$  such that*

$$(2.6.12) \quad \sup_{t \in T} \int_0^\eta g(\mu(B(t, \varepsilon))) d\varepsilon < K \left( \eta |\log \eta| + \int_0^\eta H^{1/2}(\varepsilon) d\varepsilon \right),$$

for all  $\eta > 0$ .

This Lemma, together with Theorem 2.6.1, provides an alternative proof of Theorems 2.1.3 and 2.1.5, since the additional term of  $\eta|\log \eta|$  in (2.6.12) can be absorbed into the integral (for small enough  $\eta$ ) by changing  $K$ . In other words, the Lemma shows how entropy results follow from those for majorising measures. What is particularly interesting, however, is the proof of the Lemma, which actually shows how to construct a majorising measure when the entropy integral is *a priori* known to be finite.

**Proof.** For convenience we assume that  $\text{diam}(T) = 1$ . For  $n \geq 0$ , let  $\{A_{n,1}, \dots, A_{n,N(2^{-n})}\}$  be a minimal family of  $d$ -balls of radius  $2^{-n}$  which cover  $T$ . Set

$$(2.6.13) \quad B_{n,k} = A_{n,k} \setminus \bigcup_{j < k} A_{n,j},$$

so that  $\mathcal{B}_n = \{B_{n,1}, \dots, B_{n,N(2^{-n})}\}$  is a partition of  $T$  and each  $B_i$  is contained in a  $d$ -ball of radius  $2^{-n}$ . For each pair  $n, k$  choose a point  $t_{n,k} \in B_{n,k}$  and then define a probability measure  $\mu$  on  $T$  by

$$\mu(A) = \sum_{n=0}^{\infty} 2^{-(n+1)} (N(2^{-n}))^{-1} \sum_{k=1}^{N(2^{-n})} \delta_{t_{n,k}}(A),$$

where  $\delta_{t_k}$  is the measure giving unit mass to  $t_k$ . This will be our majorising measure. To check that it satisfies (2.6.12), note first that if  $\varepsilon \in (2^{-(n+1)}, 2^{-n}]$  then

$$\mu(B(t, \varepsilon)) \geq \left(2^{n+1} N(2^{-(n+1)})\right)^{-1},$$

for all  $t \in T$ . Consequently,

$$\begin{aligned} \int_0^{2^{-n}} \sqrt{\ln \frac{1}{\mu(B(t, \varepsilon))}} d\varepsilon &\leq \sum_{k=n+1}^{\infty} 2^{-k} (\ln(2^k N(2^{-k})))^{\frac{1}{2}} \\ &\leq \sum_{k=n+1}^{\infty} 2^{-k} (k \ln 2)^{\frac{1}{2}} + 2 \int_0^{2^{-n}} (\ln(N(\varepsilon)))^{\frac{1}{2}} d\varepsilon \\ &\leq (n+2)2^{-n} \sqrt{\ln 2} + 2 \int_0^{2^{-n}} H^{1/2}(\varepsilon) d\varepsilon, \end{aligned}$$

the last line following from a little elementary algebra.

This establishes (2.6.12) for dyadic  $\eta$ . The passage to general  $\eta$  follows via a monotonicity argument.  $\square$

Another class of examples for which it is easy to find a majorising measure is given by that of stationary fields over compact Abelian groups. Here, not suprisingly, Haar measure does the job.

**Theorem 2.6.3** *If  $f$  is stationary over a compact Abelian group  $T$ , then (2.6.1) and (2.6.2) hold with  $\mu$  taken to be normalised Haar measure on  $T$ .*

A very similar result holds if  $T$  is only locally compact. You can find the details in [57].

**Proof.** Since (2.6.1) is true for any probability measure on  $T$ , it also holds for Haar measure. Thus we need only prove the lower bound (2.6.2).

Thus, assume that  $f$  is bounded, so that by Theorem 2.6.1 there exists a majorising measure  $\mu$  satisfying (2.6.2). We need to show that  $\mu$  can be replaced by Haar measure on  $T$ , which we denote by  $\nu$ .

Set

$$D_\mu \triangleq \sup \{ \eta : \mu(B(t, \eta)) < 1/2, \text{ for all } t \in T \},$$

with  $D_\nu$  defined analogously. With  $g$  as at (2.6.11), (2.6.2) can be rewritten as

$$\int_0^{D_\mu} g(\mu(B(t, \varepsilon))) d\varepsilon \leq K \mathbb{E} \left\{ \sup_{t \in T} f_t \right\},$$

for all  $t \in T$ . Let  $\tau$  be a random variable with distribution  $\nu$ ; i.e.  $\tau$  is uniform on  $T$ . For each  $\varepsilon > 0$ , set  $Z(\varepsilon) = \mu(B(\tau, \varepsilon))$ . Then, for any  $t_o \in T$ ,

$$\begin{aligned} \mathbb{E}\{Z(\varepsilon)\} &= \int_T \mu(B(t, \varepsilon)) \nu(dt) \\ &= \int_T \mu(t + B(t_o, \varepsilon)) \nu(dt) \\ &= \int_T \nu(t + B(t_o, \varepsilon)) \mu(dt) \\ &= \nu(B(t_o, \varepsilon)), \end{aligned}$$

where the second equality comes from the stationarity of  $f$  and the third and fourth from the properties of Haar measures.

Now note that  $g(x)$  is convex over  $x \in (0, \frac{1}{2})$ , so that it is possible to define a function  $\hat{g}$  that agrees with it on  $(0, \frac{1}{2})$ , is bounded on  $(\frac{1}{2}, \infty)$ , and convex on all of  $\mathbb{R}_+$ . By Jensen's inequality,

$$\hat{g}(\mathbb{E}\{Z(\varepsilon)\}) \leq \mathbb{E}\{\hat{g}(Z(\varepsilon))\}.$$

That is,

$$\hat{g}(\nu(B(t_o, \varepsilon))) \leq \int_T \hat{g}(\mu(B(t, \varepsilon))) \nu(dt).$$

Finally, set  $\Lambda = \min(D_\mu, D_\nu)$ . Then

$$\begin{aligned} \int_0^\Lambda \widehat{g}(\nu(B(t_o, \varepsilon))) d\varepsilon &\leq \int_0^\Lambda d\varepsilon \int_T \widehat{g}(\mu(B(t, \varepsilon))) \nu(dt) \\ &= \int_T \int_0^\Lambda \nu(dt) \widehat{g}(\mu(B(t, \varepsilon))) d\varepsilon \\ &\leq K \mathbb{E} \left\{ \sup_{t \in T} f_t \right\}. \end{aligned}$$

This is the crux of (2.6.2). The rest is left to you.  $\square$

This more or completes our discussion of majorising measures. However, we still have two promises to fulfill before completing this Chapter. One is to show why entropy conditions are necessary, as well as sufficient, for boundedness and continuity if  $f$  is stationary. This is Theorem 2.6.4 below. Its proof is really a little dishonest, since it relies on the main majorising measure result Theorem 2.6.1, which we have not proven. It could in fact be proven without resort to majorising measures, and was done so originally. However, for a Chapter that was supposed to be only an ‘introduction’ to the general theory of Gaussian processes, we have already used up more than enough space, and so shall make do with what we have at hand already.

Our last remaining responsibility will be to establish Corollary 2.6.5, which establishes the claims we made back at (2.2.6) and which is a simple consequence of the following Theorem.

**Theorem 2.6.4** *Let  $f$  be a centered stationary Gaussian process on a compact group  $T$ . Then the following three conditions are equivalent:*

$$(2.6.14) \quad f \text{ is a.s. continuous on } T,$$

$$(2.6.15) \quad f \text{ is a.s. bounded on } T,$$

$$(2.6.16) \quad \int_0^\infty H^{1/2}(\varepsilon) d\varepsilon < \infty.$$

**Proof.** That (2.6.14) implies (2.6.15) is obvious. That (2.6.16) implies (2.6.14) is Lemma 2.6.2 together with Theorem 2.6.1. Thus it suffices to show that (2.6.15) implies (2.6.16), which we shall now do.

Note firstly that by Theorem 2.6.3 we know that

$$(2.6.17) \quad \sup_{t \in T} \int_0^\infty g(\mu(B(t, \varepsilon))) d\varepsilon < \infty$$

for  $\mu$  normalised Haar measure on  $T$ . Furthermore, by stationarity, the value of the integral must be independent of  $t$ .

For  $\varepsilon \in (0, 1)$  let  $M(\varepsilon)$  be the maximal number of points  $\{t_k\}_{k=1}^{M(\varepsilon)}$  in  $T$  for which

$$\min_{1 \leq j, k \leq M(\varepsilon)} d(t_j, t_k) > \varepsilon.$$

It is easy to check that

$$N(\varepsilon) \leq M(\varepsilon) \leq N(\varepsilon/2).$$

Thus, since  $\mu$  is a probability measure, we must have

$$\mu(B(t, \varepsilon)) \leq (N(2\varepsilon))^{-1}.$$

Consequently, by (2.6.17) and, in particular, its independence on  $t$

$$\begin{aligned} \infty > \int_0^\infty g(\mu(B(t, \varepsilon))) d\varepsilon &\geq \int_0^\infty (\ln N(2\varepsilon))^{\frac{1}{2}} d\varepsilon \\ &= 2 \int_0^\infty H^{1/2}(\varepsilon) d\varepsilon, \end{aligned}$$

which proves the theorem.  $\square$

**Corollary 2.6.5** *If  $f$  is centered and stationary on open  $T \subset \mathbb{R}^N$  with covariance function  $C$ , and*

$$(2.6.18) \quad \frac{K_1}{(-\ln |t|)^{1+\alpha_1}} \leq C(0) - C(t) \leq \frac{K_2}{(-\ln |t|)^{1+\alpha_2}},$$

*for  $|t|$  small enough, then  $f$  will be sample path continuous if  $\alpha_2 > 0$  and discontinuous if  $\alpha_1 < 0$ .*

**Proof.** Recall from Theorem 2.2.1 that the basic energy integral in (2.6.16) converges or diverges together with

$$(2.6.19) \quad \int_\delta^\infty p(e^{-u^2}) du,$$

where

$$p^2(u) = 2 \sup_{|t| \leq u} [C(0) - C(t)].$$

(cf. (2.2.2).) Applying the bounds in (2.6.18) to evaluate (2.6.19) and applying the extension of Theorem 2.6.3 to non-compact groups easily proves the Corollary.  $\square$

# 3

## Geometry

For this Chapter we are going to leave the setting of probability and stochastic processes and deal only with classical, deterministic mathematics. In fact, we are going back to what is probably the oldest branch of mathematics, Geometry.

Our aim is to develop a framework for handling *excursion sets*, which we now redefine in a non-stochastic framework.

### 3.1 Excursion sets

**Definition 3.1.1** *Let  $f$  be a measurable, real valued, function on some measurable space, and  $T$  be a measurable subset of that space. Then, for each  $u \in \mathbb{R}$ ,*

$$(3.1.1) \quad A_u \equiv A_u(f, T) \triangleq \{t \in T : f(t) \geq u\},$$

*is called the excursion set of  $f$  in  $T$  over the level  $u$ .*

We need to study the geometric properties of such sets and somehow relate them to the properties of the function  $f$ . In particular, we shall need to adopt an approach that will, eventually, carry over to the case of  $f$  random. In order to be able to give geometric structure to excursion sets, it will be necessary for the ambient space to have some structure of its own. Consequently, throughout this Chapter we shall take  $T$  to either be a compact subset of  $\mathbb{R}^N$ , or an  $N$ -dimensional manifold. (Manifolds will be defined below in Section 3.5.)

We shall not make many demands on  $f$  beyond some simple smoothness conditions, which will translate to  $A_u(f, T)$  being almost any set with smooth boundaries. Thus, in preparation for this level of generality, we drop the explicit dependence on  $u$ ,  $T$  and  $f$  and for the moment simply look at ‘nice’ sets  $A$ .

This being the case, there are a number of ways to approach a description of the geometric properties of  $A$ , most centering around what is known as its *Euler*, or *Euler-Poincaré*, *characteristic* and its associated *intrinsic volumes* and *Minkowski functionals*, all of which will be carefully defined soon. In essence, the Euler-Poincaré characteristic is determined by the fact that it is the unique integer valued functional  $\varphi$ , defined on collections of nice  $A$ , satisfying

$$(3.1.2) \quad \varphi(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{if } A \text{ is ball-like,} \end{cases}$$

where by “ball-like” we mean homotopically equivalent<sup>1</sup> to the unit  $N$ -ball,  $B^N = \{t \in \mathbb{R}^N : |t| \leq 1\}$ , and with the additivity property

$$(3.1.3) \quad \varphi(A \cup B) = \varphi(A) + \varphi(B) - \varphi(A \cap B).$$

More global descriptions follow from this. For example, if  $A \subset \mathbb{R}^2$ , then  $\varphi(A)$  is simply the number of connected components of  $A$  minus the number of holes in it. In  $\mathbb{R}^3$ ,  $\varphi(A)$  is given by the number of connected components, minus the number of “handles”, plus the number of holes. Thus, for example, the Euler characteristics of a baseball, a tennis ball, and a coffee cup are, respectively, 1, 2, and 0.

One of the basic properties of the Euler characteristic is that it is determined by the homology class of a set. That is, smooth transformations that do not change the basic ‘shape’ of a set do not change its Euler characteristic. Finer geometric information, which will change under such transformations, lies in the Minkowski functionals, which we shall meet soon.

There are basically two different approaches to developing the geometry that we shall require. The first works well for sets in  $\mathbb{R}^N$  that are made up of the finite union of simple building blocks, such as convex sets. For many of our readers, we imagine that this will suffice, and so we treat it first, in a reasonably full fashion. The basic framework here is that of Integral Geometry.

The second, more fundamental approach is via the Differential Geometry of abstract manifolds. As described in the Preface, this more general setting has very concrete applications and, moreover, often provides more powerful and elegant proofs for a number of problems related to random

---

<sup>1</sup>The notion of homotopic equivalence is formalised below by Definition 3.2.4. However, for the moment, “ball-like” will be just as useful a concept.

fields even on the “flat” manifold  $\mathbb{R}^N$ . This approach is crucial if you want to understand the full theory. Furthermore, since some of the proofs of later results, even in the Integral Geometric setting, are more natural in the manifold scenario, it is essential if you need to see full proofs. However, the jump in level of mathematical sophistication between the two approaches is significant, so that unless you feel very much at home in the world of manifolds you are best advised to read the Integral Geometric story first.

## 3.2 Basic integral geometry

Quite comprehensive studies of integral geometry are available in the monographs of Hadwiger [41] and Santaló [83], although virtually all the results we shall derive can also be found in the paper by Hadwiger [42]. Other very useful and somewhat more modern and readable references are Schneider [84] and Klain and Rota [52].

Essentially, we shall be interested in the study of a class of geometric objects known as *basic complexes*. Later on, we shall show that, with probability one, the excursion sets of a wide variety of random fields belong to this class, and the concepts that we are about to discuss are relevant to random fields. We commence with some definitions and simple results, all of which are due to Hadwiger [42].

Assume that we have equipped  $\mathbb{R}^N$  with a Cartesian coordinate system, so that the  $N$  vectors  $e_j$  (with 1 in the  $j$ th position and zeros elsewhere) serve as an orthonormal basis. Throughout this and the following two Sections, everything that we shall have to say will be dependent on this choice of basis. The restriction to a particular coordinate system will disappear in the coordinate free approach based on manifolds.

We call a  $k$ -dimensional affine subspace of the form

$$E = \{t \in \mathbb{R}^N : t_j = a_j, j \in J, -\infty < t_j < \infty, j \notin J\}$$

a (*coordinate*)  $k$ -plane of  $\mathbb{R}^N$  if  $J$  is a subset of size  $N - k$  of  $\{1, \dots, N\}$  and  $a_1, \dots, a_{N-k}$  are fixed.

We shall call a compact set  $B$  in  $\mathbb{R}^N$  *basic* if the intersections  $E \cap B$  are simply connected for every  $k$ -plane  $E$  of  $\mathbb{R}^N$ ,  $k = 1, \dots, N$ . Note that this includes the case  $E = \mathbb{R}^N$ . These sets, as their name implies, will form the basic building blocks from which we shall construct more complicated and interesting structures. It is obvious that the empty set  $\emptyset$  is basic, as is any convex set. Indeed, convex sets remain basic under rotation, a property which characterizes them. Note that it follows from the definition that, if  $B$  is basic, then so is  $E \cap B$  for any  $k$ -plane  $E$ .

A set  $A \subset \mathbb{R}^N$  is called a *basic complex* if it can be represented as the union of a finite number of basic sets,  $B_1, \dots, B_m$ , for which the intersections  $B_{\nu_1} \cap \dots \cap B_{\nu_k}$  are basic for any combination of indices  $\nu_1, \dots, \nu_k$ ,  $k =$

$1, \dots, m$ . The collection of basic sets

$$p = p(A) = \{B_1, \dots, B_m\}$$

is called a *partition* of  $A$ , and their number,  $m$ , is called the *order* of the partition. Clearly, partitions are in no way unique.

The class of all basic complexes, which we shall denote by  $\mathcal{C}_B$ , or  $\mathcal{C}_B^N$  if we need to emphasize its dimension, possesses a variety of useful properties. For example, if  $A \in \mathcal{C}_B$  then  $E \cap A \in \mathcal{C}_B$  for every  $k$ -plane  $E$ . In fact, if  $E$  is a  $k$ -plane with  $k \leq N$  and  $A \in \mathcal{C}_B^N$ , we have  $E \cap A \in \mathcal{C}_B^k$ . To prove this it suffices to note that if

$$p = \{B_1, \dots, B_m\}$$

is a partition of  $A$  then

$$p' = \{E \cap B_1, \dots, E \cap B_m\}$$

is a partition of  $E \cap A$ , and, since  $E \cap B_k$  is a  $k$ -dimensional basic,  $E \cap A \in \mathcal{C}_B^k$ .

Another useful property of  $\mathcal{C}_B$  is that it is invariant under those rotations of  $\mathbb{R}^N$  that map the vectors  $e_1, \dots, e_N$  onto one another. It is not, however, invariant under all rotations, a point to which we shall return later<sup>2</sup>. Furthermore,  $\mathcal{C}_B$  is additive, in the sense that  $A, B \in \mathcal{C}_B$  and  $A \cap B = \emptyset$  imply<sup>3</sup>  $A \cup B \in \mathcal{C}_B$ .

These and similar properties make the class of basic complexes quite large, and, in view of the fact that convex sets are basic, implies that they include the convex ring. (i.e. the collection of all sets formed via finite union and intersection of convex sets.)

---

<sup>2</sup>A note for the purist: As noted earlier, our definition of  $\mathcal{C}_B$  is dependent on the choice of basis, which is what loses us rotation invariance. An easy counterexample in  $\mathcal{C}_B^2$  is the descending staircase  $\bigcup_{j=1}^{\infty} B_j$ , where  $B_j = \{(x, y) : 0 \leq x \leq 2^{1-j}, 1 - 2^{1-j} \leq y \leq 1 - 2^{-j}\}$ . This is actually a basic with relation to the natural axes, but not even a basic complex if the axes are rotated 45 degrees, since then it cannot be represented as the union of a *finite* number of basics. Hadwiger's [42] original definition of basic complexes was basis independent but covered a smaller class of sets. In essence, basic sets were defined as above (but relative to a coordinate system) and basic complexes were required to have a representation as a finite union of basics for *every* choice of coordinate system. Thus our descending staircase is not a basic complex in his scenario. While more restrictive, this gives a rotation invariant theory. The reasons for our choice will become clearer later on when we return to a stochastic setting. See, in particular, Theorem 4.2.3 and the comments following it.

<sup>3</sup>Note that if  $A \cap B \neq \emptyset$  then  $A \cup B$  is not necessarily a basic complex. For a counterexample, take  $A$  to be the descending staircase of the previous footnote, and let  $B$  be the line segment  $\{(x, y) : y = 1 - x, x \in [0, 1]\}$ . There is no way to represent  $A \cup B$  as the union of a *finite* number of basics, essentially because of the infinite number of single point intersections between  $A$  and  $B$ , or, equivalently, the infinite number of holes in  $A \cup B$ .

Now look back at the two requirements (3.1.2) and (3.1.3), and slightly rewrite them as

$$(3.2.1) \quad \varphi(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ 1, & \text{if } A \text{ is basic,} \end{cases}$$

and

$$(3.2.2) \quad \varphi(A \cup B) = \varphi(A) + \varphi(B) - \varphi(A \cap B),$$

whenever  $A, B, A \cup B, A \cap B \in \mathcal{C}_B$ .

An important result of Integral Geometry states that not only does a functional possessing these two properties exist, but it is *uniquely* determined by them. We shall prove this by obtaining an explicit formulation for  $\varphi$ , which will also be useful in its own right.

Let  $p = p(A)$  be a partition of order  $m$  of some  $A \in \mathcal{C}_B$  into basics. Define the *characteristic of the partition* to be

$$(3.2.3) \quad \begin{aligned} \kappa(A, p) = & \sum^{(1)} \epsilon(B_j) - \sum^{(2)} \epsilon(B_{j_1} \cap B_{j_2}) + \dots \\ & + (-1)^{n+1} \sum^{(n)} \epsilon(B_{j_1} \cap \dots \cap B_{j_n}) + \dots \\ & + (-1)^{m+1} \epsilon(B_1 \cap \dots \cap B_m) \end{aligned}$$

where  $\sum^{(n)}$  denotes summation over all subsets  $\{j_1, \dots, j_n\}$  of  $\{1, \dots, m\}$ ,  $1 \leq n \leq m$ , and  $\epsilon$  is an indicator function for basic sets, defined by

$$\epsilon(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ 1, & \text{if } A \text{ is basic.} \end{cases}$$

Then, if a functional  $\varphi$  satisfying (3.2.1) and (3.2.2) does in fact exist, it follows iteratively from these conditions and the definition of basic complexes that, for any  $A \in \mathcal{C}_B$  and any partition  $p$ ,

$$(3.2.4) \quad \varphi(A) = \kappa(A, p).$$

Thus, given existence, uniqueness of  $\varphi$  will follow if we can show that  $\kappa(A, p)$  is independent of the partition  $p$ .

**Theorem 3.2.1** *Let  $A \subset \mathbb{R}^N$  be a basic complex and  $p = p(A)$  a partition of  $A$ . Then the quantity  $\kappa(A, p)$  is independent of  $p$ . If we denote this by  $\varphi(A)$ , then  $\varphi$  satisfies (3.2.1) and (3.2.2), and  $\varphi(A)$  is called the Euler characteristic of  $A$ .*

**Proof.** The main issue is that of existence, which we establish by induction. When  $N = 1$ , basics are simply closed intervals or points, or the empty set.

Then if we write  $N(A)$  to denote the number of disjoint intervals and isolated points in a set  $A \subset \mathcal{C}_B^1$ , setting

$$\varphi(A) = N(A)$$

yields a function satisfying  $\varphi(A) = \kappa(A, p)$  for every  $p$  and for which (3.2.1) and (3.2.2) are clearly satisfied.

Now let  $N > 1$  and assume that for all spaces of dimension  $k$  less than  $N$  we have a functional  $\varphi^k$  on  $\mathcal{C}_B^k$  for which  $\varphi^k(A) = \kappa(A, p)$  for all  $A \in \mathcal{C}_B^k$  and every partition  $p$  of  $A$ . Choose one of the vectors  $e_j$ , and for  $x \in (-\infty, \infty)$  let  $E_x$  (which depends on  $j$ ) denote the  $(N - 1)$ -plane

$$(3.2.5) \quad E_x \triangleq \{t \in \mathbb{R}^N : t_j = x\}.$$

Let  $A \in \mathcal{C}_B^N$  and let  $p = p(A) = \{B_1, \dots, B_m\}$  be one of its partitions. Then clearly the projections onto  $E_0$  of the cross-sections  $A \cap E_x$  are all in  $\mathcal{C}_B^{N-1}$ , so that there exists a partition-independent functional  $\varphi_x$  defined on  $\{A \cap E_x, A \in \mathcal{C}_B^N\}$  determined by

$$\varphi_x(A \cap E_x) = \varphi^{N-1}(\text{projection of } A \cap E_x \text{ onto } E_0).$$

Via  $\varphi_x$  we can define a new partition-independent function  $f$  by

$$(3.2.6) \quad f(A, x) = \varphi_x(A \cap E_x).$$

However, by the induction hypothesis and (3.2.6), we have from (3.2.3) that

$$f(A, x) = \sum^{(1)} \epsilon(B_{j_1} \cap E_x) - \sum^{(2)} \epsilon(B_{j_1} \cap B_{j_2} \cap E_x) + \dots$$

Consider just one of the right-hand terms, writing

$$\epsilon(x) = \epsilon(B_{j_1} \cap \dots \cap B_{j_k} \cap E_x).$$

Since  $\epsilon(x)$  is zero when the intersection is empty and one otherwise, we have for some finite  $a$  and  $b$  that  $\epsilon(x) = 1$  if  $a \leq x \leq b$  and  $\epsilon(x) = 0$  otherwise. Thus  $\epsilon(x)$  is a step function, taking at most two values. Hence  $f(A, x)$ , being the sum of a finite number of such functions, is again a step function, with a finite number of discontinuities. Thus the left-hand limits

$$(3.2.7) \quad f(A, x^-) = \lim_{y \downarrow 0} f(A, x - y)$$

always exist. Now define a function  $h$ , which is non-zero at only a finite number of points  $x$ , by

$$(3.2.8) \quad h(A, x) = f(A, x) - f(A, x^-)$$

and define

$$(3.2.9) \quad \varphi(A) = \sum_x h(A, x),$$

where the summation is over the finite number of  $x$  for which the summand is non-zero. Note that since  $f$  is independent of  $p$  so are  $h$  and  $\varphi$ .

Thus we have defined a functional on  $\mathcal{C}_B$ , and we need only check that (3.2.1) and (3.2.2) are satisfied to complete this section of the proof. Firstly, note that if  $B$  is a basic and  $B \neq \emptyset$ , and if  $a$  and  $b$  are the extreme points of the linear set  $e_j \cap B$ , then  $f(B, x) = 1$  if  $a \leq x \leq b$  and equals zero otherwise. Thus  $h(B, a) = 1$ , while  $h(B, x) = 0$ ,  $x \neq a$ , so that  $\varphi(B) = 1$ . This is (3.2.1) since  $\varphi(\emptyset) = 0$  is obvious. Now let  $A, B, A \cup B, A \cap B$  all belong to  $\mathcal{C}_B$ . Then the projections onto  $E_0$  of the intersections

$$A \cap E_x, \quad B \cap E_x, \quad A \cup B \cap E_x, \quad A \cap B \cap E_x$$

all belong to  $\mathcal{C}_B^{N-1}$  and so by (3.2.6) and the induction hypothesis

$$f(A \cup B, x) = f(A, x) + f(B, x) - f(A \cap B, x).$$

Replacing  $x$  by  $x^-$  and performing a subtraction akin to that in (3.2.8) we obtain

$$h(A \cup B, x) = h(A, x) + h(B, x) - h(A \cap B, x).$$

Summing over  $x$  gives

$$\varphi(A \cup B) = \varphi(A) + \varphi(B) - \varphi(A \cap B)$$

so that (3.2.2) is established and we have the existence of our functional  $\varphi$ . It may, however, depend on the partition  $p$  used in its definition.

For uniqueness, note that since, by (3.2.4),  $\kappa(A, p') = \varphi(A)$  for any partition  $p'$ , we have that  $\kappa(A, p')$  is independent of  $p'$  and so that  $\varphi(A)$  is independent of  $p$ . That is, we have the claimed uniqueness of  $\varphi$ .

Finally, since  $\kappa(A, p)$  is independent of the particular choice of the vector  $e_j$  appearing in the proof, so is  $\varphi$ .  $\square$

The proof of Theorem 3.2.1 actually contains more than we have claimed in the statement of the result, since, in developing the proof, we actually obtained an alternative way of computing  $\varphi(A)$  for any  $A \in \mathcal{C}_B$ . This is given explicitly in the following theorem, for which  $E_x$  is as defined at (3.2.5).

**Theorem 3.2.2** *For basic complexes  $A \in \mathcal{C}_B^N$ , the Euler characteristic  $\varphi$ , as defined by (3.2.4), has the following equivalent iterative definition:*

$$(3.2.10) \quad \varphi(A) = \begin{cases} \text{number of disjoint intervals in } A, & \text{if } N = 1, \\ \sum_x \{\varphi(A \cap E_x) - \varphi(A \cap E_{x^-})\}, & \text{if } N > 1, \end{cases}$$

where

$$\varphi(A \cap E_{x-}) = \lim_{y \downarrow 0} \varphi(A \cap E_{x-y})$$

and the summation is over all real  $x$  for which the summand is non-zero.

This theorem is a simple consequence of (3.2.9) and requires no further proof. Note that it also follows from the proof of Theorem 3.2.1 (cf. the final sentence there) that the choice of vector  $e_j$  used to define  $E_x$  is irrelevant<sup>4</sup>.

The importance of Theorem 3.2.2 lies in the iterative formulation it gives for  $\varphi$ , for, using this, we shall show in a moment how to obtain yet another formulation that makes the Euler characteristic of a random excursion set amenable to probabilistic investigation.

Figure 3.2.1 shows an example of this iterative procedure in  $\mathbb{R}^2$ . Here the vertical axis is taken to define the horizontal 1-planes  $E_x$ . The values of  $\varphi(A \cap E_x)$  appear closest to the vertical axis, with the values of  $h$  to their left. Note in particular the set with the hole ‘in the middle’. It is on sets like this, and their counterparts in higher dimensions, that the characteristic  $\varphi$  and the number of connected components of the set differ. In this example they are, respectively, zero and one. For the moment, ignore the arrows and what they are pointing at.

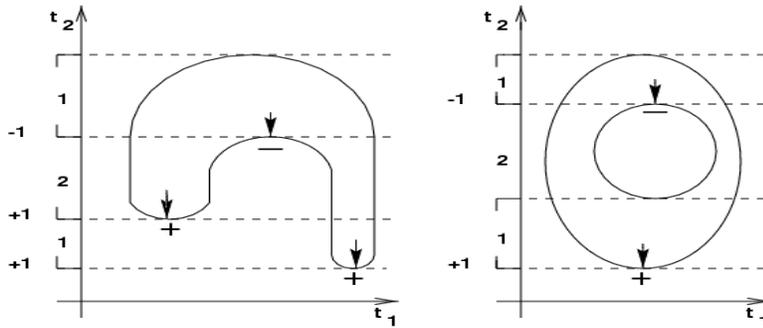


FIGURE 3.2.1. Computing the Euler characteristic.

To understand how this works in higher dimensions, you should try to visualise some  $N$ -dimensional examples to convince yourself that for the  $N$ -dimensional unit ball,  $B^N$ , and the  $N$ -dimensional sphere,  $S^{N-1}$ :

$$\varphi(B^N) = 1, \quad \varphi(S^{N-1}) = 1 + (-1)^{N-1}.$$

It is somewhat less easy (and, indeed, quite deep in higher dimensions) to see that, if  $K_{N,k}$  denotes  $B^N$  with  $k$  non-intersecting cylindrical holes

<sup>4</sup>It is also not hard to show that if  $A$  is a basic complex with respect to each of two coordinate systems (which are not simple relabellings of one another) then  $\varphi(A)$  will be the same for both. However, this is taking us back to original formulism of [42], which we have already decided is beyond what we need here.

drilled through it, then, since both  $K_{N,k}$  and its boundary belong to  $\mathcal{C}_B^N$ ,

$$\varphi(K_{N,k}) = 1 + (-1)^N k,$$

while

$$\varphi(\partial K_{N,k}) = [1 + (-1)^{N-1}](1 - k).$$

Finally, if we write  $\bar{K}_{N,k}$  to denote  $B^N$  with  $k$  ‘handles’ attached, then,

$$\varphi(\bar{K}_{N,k}) = 1 - k.$$

In view of (3.1.2), knowing the Euler characteristic of balls means that we know it for all basics. However, this invariance goes well beyond balls, since it can also be shown that the Euler characteristic is the same for all homotopically equivalent sets; i.e. sets which are geometry “alike” in that they can be deformed into one another in a continuous fashion. Here is a precise definition of this equivalence.

**Definition 3.2.3** *Let  $f$  and  $g$  be two mappings from  $A$  to  $B$ , both subsets of a Hausdorff space<sup>5</sup>  $X$ . If there exists a continuous mapping,  $F : A \times [0, 1] \rightarrow X$  with the following three properties, then we say  $f$  is homotopic or deformable to  $g$ :*

$$\begin{aligned} F(t, 0) &= f(t) & \forall t \in A, \\ F(t, \tau) &\in X & \forall t \in A, \tau \in [0, 1], \\ F(t, 1) &= g(t) & \forall t \in A. \end{aligned}$$

**Definition 3.2.4** *Let  $A$  and  $B$  be subsets of (possibly different) Hausdorff spaces. If there exist continuous  $f : A \rightarrow B$  and  $g : B \rightarrow A$  for which the composite maps  $g \circ f : A \rightarrow A$  and  $f \circ g : B \rightarrow B$  are deformable, respectively, to the identity maps on  $A$  and  $B$ , then  $A$  and  $B$  are called homotopically equivalent.*

### 3.3 Excursion sets again

We now return briefly to the setting of excursion sets. Our aim in this Section will be to find a way to compute their Euler characteristics directly from the function  $f$ , without having to look at the sets themselves. In particular, we will need to find a method that depends only on *local* properties of  $f$ , since later, in the random setting, this (via finite dimensional distributions) will be all that will be available to us for probabilistic computation.

---

<sup>5</sup>A Hausdorff space is a topological space  $T$  for which, given any two distinct points  $s, t \in T$  there are open sets  $U, V \subset T$  with  $s \in U$ ,  $t \in V$  and  $U \cap V = \emptyset$ .

Since the main purpose of this Section is to develop understanding and since we shall ultimately use the techniques of the critical point theory developed in Section 3.9 to redo everything that will be done here in far greater generality, we shall not give all the details of all the proofs. The interested reader can find most of them either in GRF [1] or the review [4]. Furthermore, we shall restrict the parameter set  $T$  to being a bounded rectangle of the form

$$(3.3.1) \quad T = [s, t] \triangleq \prod_{i=1}^N [s_i, t_i], \quad -\infty < s_i < t_i < \infty.$$

For our first definition, we need to decompose  $T$  into a disjoint union of open sets, starting with its interior, its faces, edges etc. More precisely, a *face*  $J$ , of dimension  $k$ , is defined by fixing a subset  $\sigma(J)$  of  $\{1, \dots, N\}$  of size  $k$  and a sequence of  $N - k$  zeroes and ones, which we write as  $\varepsilon(J) = \{\varepsilon_1, \dots, \varepsilon_{N-k}\}$ , so that

$$(3.3.2) \quad J = \{v \in T : v_j = (1 - \varepsilon_j)s_j + \varepsilon_j t_j, \text{ if } j \in \sigma(J), \\ s_j < v_j < t_j, \text{ if } j \in \sigma^c(J)\},$$

where  $\sigma^c(J)$  is the complement of  $\sigma(J)$  in  $\{1, \dots, N\}$ .

In anticipation of later notation, we write  $\partial_k T$  for the collection of faces of dimension  $k$  in  $T$ . This is known as the *k-skeleton* of  $T$ . Then  $\partial_2 T$  contains only  $T^\circ$ , while  $\partial_N T$  contains the  $2^N$  vertices of the rectangle. In general,  $\partial_k T$  has

$$(3.3.3) \quad J_k \triangleq 2^{N-k} \binom{N}{k}$$

elements. Note also that the usual boundary of  $T$ ,  $\partial T$ , is given by the disjoint union

$$J = \bigcup_{k=0}^{N-1} \bigcup_{J \in \partial_k T} J.$$

We start with some simple assumptions on  $f$  and, as usual, write the first and second order partial derivatives of  $f$  as  $f_j = \partial f / \partial t_j$ ,  $f_{ij} = \partial^2 f / \partial t_i \partial t_j$ .

**Definition 3.3.1** *Let  $T$  be a bounded rectangle in  $\mathbb{R}^N$  and let  $f$  be a real valued function defined on an open neighbourhood of  $T$ .*

*Then, if for a fixed  $u \in \mathbb{R}$  the following three conditions are satisfied for each face  $J$  of  $T$  for which  $N \in \sigma(J)$ , we say that  $f$  is suitably regular with*

respect to  $T$  at the level  $u$ .

(3.3.4)  $f$  has continuous partial derivatives of up to second order in an open neighborhood of  $T$ .

(3.3.5)  $f|_J$  has no critical points at the level  $u$ .

(3.3.6) If  $J \in \partial_k T$  and  $D_J(t)$  denotes the symmetric  $(k-1) \times (k-1)$  matrix with elements  $f_{ij}$ ,  $i, j \in \sigma^c(J) \setminus \{N\}$ , then there are no  $t \in J$  for which  $0 = f(t) - u = \det D_J(t) = f_j(t)$  for all  $j \in \sigma^c(J) \setminus \{N\}$ .

The first two conditions of suitable regularity are meant to ensure that the boundary  $\partial A_u = \{t \in T : f(t) = u\}$  of the excursion set be smooth in the interior  $T^\circ$  of  $T$  and that its intersections with  $\partial T$  also be smooth. The last condition is a little more subtle, since it relates to the curvature of  $\partial A_u$  both in the interior of  $T$  and on its boundary.

The main importance of suitable regularity is that it gives us the following:

**Theorem 3.3.2** *Let  $f : \mathbb{R}^N \rightarrow \mathbb{R}^1$  be suitably regular with respect to a bounded rectangle  $T$  at level  $u$ . Then the excursion set  $A_u(f, T)$  is a basic complex.*

The proof of Theorem 3.3.2 is not terribly long, but since it is not crucial to what will follow can be skipped at first reading. The reasoning behind it is all in Figure 3.3.1, and after understanding this you can skip to the examples immediately following the proof without losing much. In any case, this Lemma will reappear in a more general form as Lemma 4.1.12.

For those of you remaining with us, we start with a lemma:

**Lemma 3.3.3** *Let  $f : \mathbb{R}^N \rightarrow \mathbb{R}^1$  be suitably regular with respect to a bounded rectangle  $T$  at the level  $u$ . Then there are only finitely many  $t \in T$  for which*

$$(3.3.7) \quad f(t) - u = f_1(t) = \dots = f_{N-1}(t) = 0.$$

**Proof.** To start, let  $g = (g^1, \dots, g^N) : T \rightarrow \mathbb{R}^N$  be the function defined by

$$(3.3.8) \quad g^1(t) = f(t) - u, \quad g^j(t) = f_{j-1}(t), \quad j = 2, \dots, N.$$

Let  $t \in T$  satisfy (3.3.7). Then by (3.3.6)  $t \notin \partial T$ ; i.e.  $t$  is in the interior of  $T$ . We shall show that there is an open neighborhood of  $t$  throughout which no other point satisfies (3.3.7), which we rewrite as  $g(t) = 0$ . This would imply that the points in  $T$  satisfying (3.3.7) are isolated and thus, from the compactness of  $T$ , we would have that there are only a finite number of them. This, of course, would prove the lemma.

The inverse mapping theorem<sup>6</sup> implies that such a neighborhood will exist if the  $N \times N$  matrix  $(\partial g^i / \partial t_j)$  has a non-zero determinant at  $t$ . However, this matrix has the following elements:

$$\begin{aligned} \frac{\partial g^1}{\partial t_j} &= f_j(t), & \text{for } j = 1, \dots, N, \\ \frac{\partial g^i}{\partial t_j} &= f_{i-1}(t), & \text{for } i = 2, \dots, N, \quad j = 1, \dots, N. \end{aligned}$$

Since  $t$  satisfies (3.3.7) all elements in the first row of this matrix, other than the  $N$ -th, are zero. Expanding the determinant along this row gives us that it is equal to

$$(3.3.9) \quad (-1)^{N-1} f_N(t) \det D(t),$$

where  $D(t)$  is as defined at (3.3.6). Since (3.3.7) is satisfied, (3.3.5) and (3.3.6) imply, respectively, that neither  $f_N(t)$  nor the determinant of  $D(t)$  is zero, which, in view of (3.3.9), is all that is required.  $\square$

**Proof of Theorem 3.3.2** When  $N = 1$  we are dealing throughout with finite collections of intervals, and so the result is trivial.

Now take the case  $N = 2$ . We need to show how to write  $A_u$  as a finite union of basic sets.

Consider the set of points  $t \in T$  satisfying either

$$(3.3.10) \quad f(t) - u = f_1(t) = 0$$

or

$$(3.3.11) \quad f(t) - u = f_2(t) = 0.$$

For each such point draw a line containing the point and parallel to either the horizontal or vertical axis, depending, respectively, on whether (3.3.10) or (3.3.11) holds. These lines form a grid over  $T$ , and it is easy to check that the connected regions of  $A_u$  within each cell of this grid are basic. Furthermore, these sets have intersections which are either straight lines, points, or empty, and Lemma 3.3.3 guarantees that there are only a finite number of them, so that they form a partition for  $A_u$ . (An example of

---

<sup>6</sup> The inverse mapping theorem goes as follows: Let  $U \subset \mathbb{R}^N$  be open and  $g = (g^1, \dots, g^N) : U \rightarrow \mathbb{R}^N$  be a function possessing continuous first-order partial derivatives  $\partial g^i / \partial t_j$ ,  $i, j = 1, \dots, N$ . Then if the matrix  $(\partial g^i / \partial t_j)$  has a non-zero determinant at some point  $t \in U$ , there exist open neighborhoods  $U_1$  and  $V_1$  of  $t$  and  $g(t)$ , respectively, and a function  $g^{-1} : V_1 \rightarrow U_1$ , for which

$$g^{-1}(g(t)) = t \quad \text{and} \quad g(g^{-1}(s)) = s,$$

for all  $t \in U_1$  and  $s \in V_1$ .

this partitioning procedure is shown in Figure 3.3.1. The dots mark the points where either (3.3.10) or (3.3.11) holds.) This provides the required partition, and we are done.

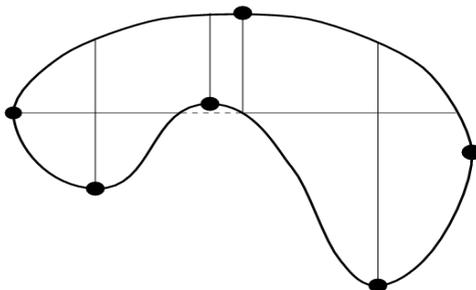


FIGURE 3.3.1. Partitioning excursion sets into basic components

For  $N > 2$  essentially the same argument works, using partitioning  $(N - 1)$  planes passing through points at which

$$f(t) - u = f_1(t) = \cdots = f_{N-1}(t).$$

Lemma 3.3.3 again guarantees the finiteness of the partition. The details are left to you<sup>7</sup>.  $\square$

We now attack the problem of obtaining a simple way of computing the Euler characteristic of  $A_u$ . As you are about to see, this involves looking at each  $A_u \cap J$ ,  $J \in \partial_k T$ ,  $0 \leq k \leq N$ , separately. We start with the simple example  $T = I^2$ , in which  $\partial_2 T = T^\circ$ ,  $\partial_1 T$  is composed of four open intervals parallel to the axes, and  $\partial_0 T$  contains the four vertices of the square. Since this is a particularly simple case, we shall pool  $\partial_1 T$  and  $\partial_0 T$ , and handle them together as  $\partial T$ .

Thus, let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^1$  be suitably regular with respect to  $I^2$  at the level  $u$ . Consider the summation (3.2.10) defining  $\varphi(A_u(f, I^2))$ ; viz.

$$(3.3.12) \quad \varphi(A_u) = \sum_{x \in (0,1]} \{\varphi(A_u \cap E_x) - \varphi(A_u \cap E_{x-})\},$$

where now  $E_x$  is simply the straight line  $t_2 = x$  and so  $n_x \triangleq \varphi(A_u \cap E_x)$  counts the number of distinct intervals in the cross-section  $A_u \cap E_x$ . The values of  $x$  contributing to the sum correspond to values of  $x$  where  $n_x$  changes.

It is immediate from the continuity of  $f$  that contributions to  $\varphi(A_u)$  can only occur when  $E_x$  is tangential to  $\partial A_u$  (Type I contributions) or when

<sup>7</sup>You should at least try the three dimensional case, to get a feel for the source of the conditions on the various faces of  $T$  in the definition of suitable regularity.

$f(0, x) = u$  or  $f(1, x) = u$  (Type II contributions). Consider the former case first.

If  $E_x$  is tangential to  $\partial A$  at a point  $t$ , then  $f_1(t) = 0$ . Furthermore, since  $f(t) = u$  on  $\partial A_u$ , we must have that  $f_2(t) \neq 0$ , as a consequence of suitable regularity. Thus, in the neighborhood of such a point and on the curve  $\partial A_u$ ,  $t_2$  can be expressed as an implicit function of  $t_1$  by

$$f(t_1, g(t_1)) = u.$$

The implicit function theorem<sup>8</sup> gives us

$$\frac{dt_2}{dt_1} = -\frac{f_1(t)}{f_2(t)},$$

so that applying what we have just noted about the tangency of  $E_x$  to  $\partial A_u$  we have for each contribution of Type I to (3.3.12) that there must be a point  $t \in I^2$  satisfying

$$(3.3.13) \quad f(t) = u,$$

and

$$(3.3.14) \quad f_1(t) = 0.$$

Furthermore, since the limit in (3.3.12) is one sided, continuity considerations imply that contributing points must also satisfy

$$(3.3.15) \quad f_2(t) > 0.$$

Conversely, for each point satisfying (3.3.13) to (3.3.15) there is a unit contribution of Type I to  $\varphi(A_u)$ . Note that there is no contribution of Type I to  $\varphi(A_u)$  from points on the boundary of  $I^2$  because of the regularity condition (3.3.5). Thus we have a one-one correspondence between unit contributions of Type I to  $\varphi(A_u)$  and points in the interior of  $I^2$  satisfying (3.3.13) to (3.3.15). It is also easy to see that contributions of +1

---

<sup>8</sup> The implicit function theorem goes as follows: Let  $U \subset \mathbb{R}^N$  be open and  $F : U \rightarrow \mathbb{R}$  possess continuous first-order partial derivatives. Suppose at  $t^* \in U$ ,  $F(t^*) = u$  and  $F_N(t^*) \neq 0$ . Then the equation

$$F(t_1, \dots, t_{N-1}, g(t_1, \dots, t_{N-1})) = u$$

defines an implicit function  $g$  which possesses continuous, first-order partial derivatives in some interval containing  $(t_1^*, \dots, t_{N-1}^*)$ , and such that  $g(t_1^*, \dots, t_{N-1}^*) = t_N^*$ . The partial derivatives of  $g$  are given by

$$\frac{\partial g}{\partial t_j} = -\frac{F_j}{F_N}, \quad \text{for } j = 1, \dots, N-1.$$

Furthermore, if  $F$  is  $k$  times differentiable, so is  $g$ .

will correspond to points for which  $f_{11}(t) < 0$  and contributions of  $-1$  to points for which  $f_{11}(t) > 0$ . Furthermore, because of (3.3.6) there are no contributing points for which  $f_{11}(t) = 0$ .

Consider now Type II contributions to  $\varphi(A)$ , which is best done by looking first at Figure 3.3.2.

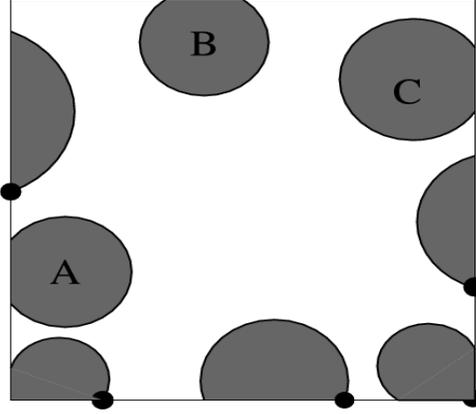


FIGURE 3.3.2. Contributing points to the Euler characteristic

The eight partial and complete disks there lead to a total Euler characteristic of 8. The three sets  $A$ ,  $B$  and  $C$  are accounted for by Type I contributions, since in each case the above analysis will count  $+1$  at their lowest points. We need to account for the remaining five sets, which means running along  $\partial I^2$  and counting points there. In fact, what we need to do is count  $+1$  at the points marked with a  $\bullet$ . There is never a need to count  $-1$  on the boundary. Note that on the two vertical sections of  $\partial I^2$  we count  $+1$  whenever we enter the set (at its intersection with  $\partial I^2$ ) from below. There is never a contribution from the top side of  $I^2$ . For the bottom, we need to count the number of connected components of its intersection with the excursion set, which can be done either on “entering” or “leaving” the set in the positive  $t_1$  direction. We choose the latter, and so must also count a  $+1$  if the point  $(1, 0)$  is covered.

Putting all this together, we have a Type II contribution to  $\varphi(A)$  whenever one of the following four sets conditions is satisfied:

$$(3.3.16) \quad \begin{cases} t = (t_1, 0), & f(t) = u, & f_1(t) < 0, \\ t = (0, t_2), & f(t) = u, & f_2(t) > 0, \\ t = (1, t_2), & f(t) = u, & f_2(t) > 0, \\ t = (1, 0), & f(t) > u. \end{cases}$$

The above argument has established the following, which in Section 3.9, with completely different techniques and a much more sophisticated lan-

guage, will be extended to parameter sets in  $\mathbb{R}^N$  and on  $C^2$  manifolds with piecewise smooth boundaries:

**Theorem 3.3.4** *Let  $f$  be suitably regular with respect to  $I^2$  at the level  $u$ . Then the Euler characteristic of its excursion set  $A_u(f, I^2)$  is given by the number of points  $t$  in the interior of  $I^2$  satisfying*

$$(3.3.17) \quad f(t) - u = f_1(t) = 0, \quad f_2(t) > 0, \quad \text{and} \quad f_{11}(t) < 0,$$

*minus the number of points satisfying*

$$(3.3.18) \quad f(t) - u = f_1(t) = 0, \quad f_2(t) > 0, \quad \text{and} \quad f_{11}(t) > 0,$$

*plus the number of points on the boundary of  $I^2$  satisfying one of the four sets of conditions in (3.3.16).*

This is what we have been looking for in a doubly simple case: The ambient dimension was only 2, and the set  $T$  a simple square. There is another proof of Theorem 3.3.4 in Section 3.9.2, built on Morse Theory. There you will also find a generalisation of this result to  $I^N$  for all finite  $N$ , although the final point set representation is a little different to that of (3.3.16). Morse Theory is also the key to treating far more general parameter spaces. Nevertheless, what we have developed so far, along with some ingenuity, does let us treat some more general cases, for which we give an example. You should be able to fill in the details of the computation by yourself, although some hand waving may be necessary. Here is the example:

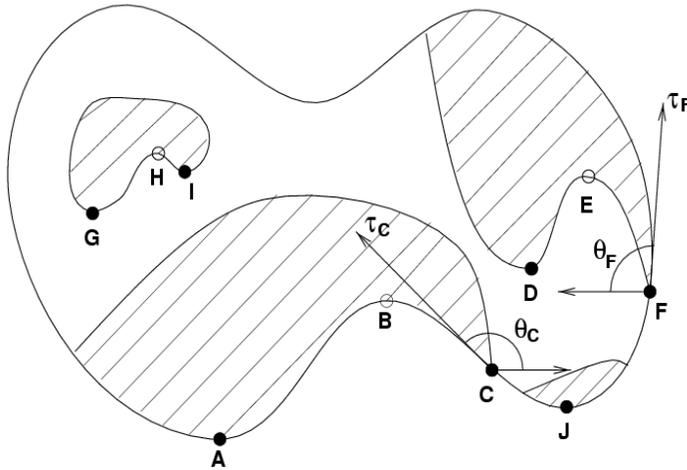


FIGURE 3.3.3. Computing the Euler characteristic for a general shape.

Consider the parameter space  $T$  to be the surrounding dumbbell shape of Figure 3.3.3. Of the four components of the excursion set (the hatched

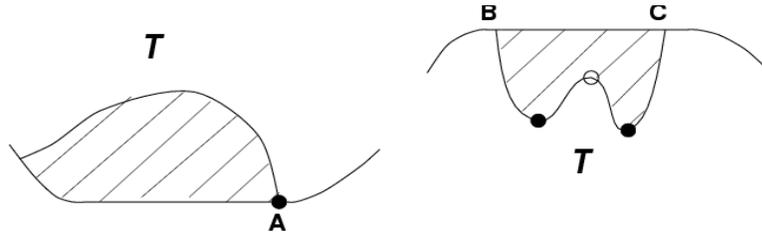


FIGURE 3.3.4. Points on a horizontal part of the boundary.

objects) three intersect  $\partial T$ . We already know how to characterise the small component in the top left: We count a +1 for each  $\bullet$ , -1 for for each  $\circ$ , and sum. The problem is what to do with the remaining two components. Worsley [107] showed that what needs to be done is to again subtract from the number of points in the interior of  $T$  that satisfy (3.3.17) the number satisfying (3.3.18) and then to go along  $\partial T$  and a +1 for each point marked with a  $\bullet$ .

More precisely, the rules for marking are as follows,

- (a) If  $t$  is in the interior of  $T$ , then apply the criteria (3.3.17) and (3.3.18) exactly as before, marking  $\bullet$  (+1) and  $\circ$  (-1) respectively.
- (b) If  $t \in \partial T \cap \partial A_u$ , and the tangent to  $\partial T$  is *not* parallel to the horizontal axis, then let  $f_{\text{up}}(t)$  be the derivative of  $f$  in the direction of the tangent to  $\partial T$  pointing in the positive  $t_2$  direction. (Two such tangent vectors appear as  $\tau_C$  and  $\tau_F$  in Figure 3.3.3.) Furthermore, take the derivative of  $f$  with respect to  $t_1$  in the direction pointing *into*  $T$ . Call this  $f_{\perp}$ . (It will equal either  $f_1$  or  $-f_1$ , depending on whether the angles  $\theta$  in Figure 3.3.3 from the horizontal to the  $\tau$  vector develop in a counter-clockwise or clockwise direction, respectively.) Now mark  $t$  as a  $\bullet$  (and so count as +1) if  $f_{\perp}(t) < 0$  and  $f_{\text{up}}(t) > 0$ . There are no  $\circ$  points in this class.
- (c) If  $t \in \partial T \cap \partial A_u$ , and the tangent to  $\partial T$  is parallel to the horizontal axis, but  $t$  is not included in an open interval  $\partial T$  which is parallel to this axis, then proceed as in (b), simply defining  $f_{\perp}$  to be  $f_1$  if the tangent is above  $\partial T$  and  $-f_1$  otherwise.
- (d) If  $t \in \partial T \cap \partial A_u$  belongs to an open interval of  $\partial T$  which is parallel to the horizontal axis (as in Figure 3.3.4) then mark it as a  $\bullet$  if  $T$  is above  $\partial T$  and  $f_1(t) < 0$ . (Thus, as in Figure 3.3.4, points such as  $B$  and  $C$  by which  $A$  “hangs” from  $\partial T$  will never be counted.)
- (e) Finally, if  $t \in \partial T \cap A_u$ , has not already been marked, and coincides with one of the points that contribute to the Euler characteristic of  $T$  itself (e.g.  $A, B$  and  $J$  in Figure 3.3.3) then mark it exactly as it was marked in computing  $\varphi(T)$ .

All told, this can be summarised as

**Theorem 3.3.5 (Worsley [107])** *Let  $T \subset \mathbb{R}^2$  be compact with boundary  $\partial T$  that is twice differentiable everywhere except, at most, at a finite number of points. Let  $f$  be suitably regular for  $T$  at the level  $u$ . Let  $\chi(A_u(f, T))$  be the number of points in the interior of  $T$  satisfying (3.3.17) minus the number satisfying (3.3.18).*

*Denote the number of points satisfying (b)–(d) above as  $\chi_{\partial T}$ , and denote the sum of the contributions to  $\varphi(T)$  of those points described in (e) by  $\varphi_{\partial T}$ . Then*

$$(3.3.19) \quad \varphi(A) = \chi(A) + \chi_{\partial T} + \varphi_{\partial T}.$$

Theorem 3.3.5 can be extended to three dimensions (also in [107]). In principle, it is not too hard to guess what has to be done in higher dimensions as well. Determining whether a point in the interior of  $T$  and on  $\partial A_u$  contributes a +1 or -1 will depend on the curvature of  $\partial A_u$ , while if  $t \in \partial T$  both this curvature and that of  $\partial T$  will have rôles to play.

It is clear that these kinds of arguments are going to get rather messy very quickly, and a different approach is advisable. This is provided via the critical point theory of Differential Topology, which we shall develop in Section 3.9. However, before doing this we want to develop some more geometry in the still relatively simple scenario of Euclidean space and describe how all of this relates to random fields.

### 3.4 Intrinsic volumes

The Euler characteristic of Section 3.2 arose as the unique additive functional (cf. (3.2.2)) that assigned the value 1 to sets homotopic to a ball, and 0 to the empty set. It turned out to be integer valued, although we did not demand this in the beginning and has an interpretation in terms of ‘counting’ the various topological components of a set. But there is more to life than mere counting and one would also like to be able to say things about the volume of sets, the surface area of their boundaries, their curvatures, etc. In this view, it is natural to look for a class of  $N$  additional, position and rotation invariant, non-negative, functionals  $\{\mathcal{L}_j\}_{j=1}^N$ , which are also additive in the sense of (3.2.2) and scale with dimensionality in the sense that

$$(3.4.1) \quad \mathcal{L}_j(\lambda A) = \lambda^j \mathcal{L}_j(A), \quad \lambda > 0,$$

where  $\lambda A \triangleq \{t : t = \lambda s, s \in A\}$ .

Such functionals exist, and together with  $\mathcal{L}_0 \triangleq \varphi$ , the Euler characteristic itself, make up what are known as the *intrinsic volumes* defined on a

suitable class of sets  $A$ . The reason why the intrinsic volumes are of crucial importance to us will become clear when we get to the discussion following Theorem 3.4.1, which shows that many probabilistic computations for random fields are intimately related to them. They can be defined in a number of ways, one of which is a consequence of *Steiner's formula* [52, 83], which, for *convex* subsets of  $\mathbb{R}^N$ , goes as follows:

For  $A \subset \mathbb{R}^N$  and  $\rho > 0$ , let

$$(3.4.2) \quad \text{Tube}(A, \rho) = \{x \in \mathbb{R}^N : d(x, A) \leq \rho\}$$

be the ‘tube’ of radius  $\rho$ , or ‘ $\rho$ -tube’, around  $A$ , where

$$d(x, A) \triangleq \inf_{y \in A} |x - y|$$

is the usual Euclidean distance from the point  $x$  to the set  $A$ . An example is given in Figure 3.4.1, in which  $A$  is the inner triangle and  $\text{Tube}(A, \rho)$  the larger triangular object with rounded-off corners.

With  $\lambda_N$  denoting, as usual, Lebesgue measure in  $\mathbb{R}^N$ , Steiner's formula states<sup>9</sup> that  $\lambda_N(\text{Tube}(A, \rho))$  has a finite expansion in powers of  $\rho$ . In particular,

$$(3.4.3) \quad \lambda_N(\text{Tube}(A, \rho)) = \sum_{j=0}^N \omega_{N-j} \rho^{N-j} \mathcal{L}_j(A),$$

where

$$(3.4.4) \quad \omega_j = \lambda_j(B(0, 1)) = \frac{\pi^{j/2}}{\Gamma(\frac{j}{2} + 1)} = \frac{s_j}{j}$$

is the volume of the unit ball in  $\mathbb{R}^j$ . (Recall that  $s_j$  was the corresponding surface area. cf. (1.4.42).)

We shall see a proof of (3.4.3) later on in Chapter 7 in a far more general scenario, but it is easy to see from Figure 3.4.1 from where the result comes.

To find the area (i.e. 2-dimensional volume) of the enlarged triangle, one needs only to sum three terms:

- The area of the original, inner triangle.
- The area of the three rectangles. Note that this is the perimeter (i.e. ‘surface area’) of the triangle multiplied by  $\rho$ .
- The area of the three corner sectors. Note that the union of these sectors will always give a disk of Euler characteristic 1 and radius  $\rho$ .

---

<sup>9</sup>There is a more general version of Steiner's formula for the case in which  $T$  and its tube are embedded in a higher dimensional space. See Theorem 7.3.6 for a version of this in the manifold setting.

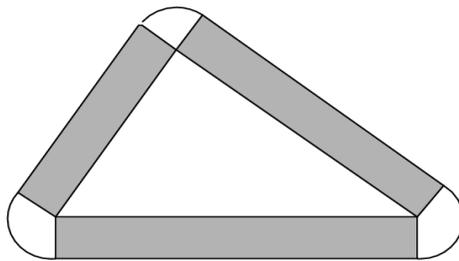


FIGURE 3.4.1. The tube around a triangle.

In other words,

$$\text{area}(\text{Tube}(A, \rho)) = \pi\rho^2\varphi(A) + \rho \text{perimeter}(A) + \text{area}(A).$$

Comparing this to (3.4.3) it now takes only a little thought to guess what the intrinsic volumes must measure. If the ambient space is  $\mathbb{R}^2$ , then  $\mathcal{L}_2$  measures area,  $\mathcal{L}_1$  measures boundary length, while  $\mathcal{L}_0$  gives the Euler characteristic. In  $\mathbb{R}^3$ ,  $\mathcal{L}_3$  that measures volume,  $\mathcal{L}_2$  measures surface area,  $\mathcal{L}_1$  is a measure of cross-sectional diameter, and  $\mathcal{L}_0$  is again the Euler characteristic. In higher dimensions, it takes some imagination, but  $\mathcal{L}_N$  and  $\mathcal{L}_{N-1}$  are readily seen to measure volume and surface area, while  $\mathcal{L}_0$  is always the Euler characteristic. Precisely why this happens, how it involves the curvature of the set and its boundary, and what happens in less familiar spaces forms much of the content of Section 3.8 and is treated again, in fuller detail, in Chapter 7.

In the meantime, you can try checking for yourself, directly from (3.4.3) and a little first-principles geometry that for a  $N$ -dimensional cube of side length  $T$  the intrinsic volumes are given by

$$(3.4.5) \quad \mathcal{L}_j([0, T]^N) = \binom{N}{j} T^j.$$

It is also not hard to see that for  $N$ -dimensional rectangles

$$(3.4.6) \quad \mathcal{L}_j\left(\prod_{i=1}^N [0, T_i]\right) = \sum T_{i_1} \dots T_{i_j},$$

where the sum is taken over the  $\binom{N}{j}$  different choices of subscripts  $i_1, \dots, i_j$ .

For handling  $B^N(T)$ , the ball of radius  $T$ , it useful to go beyond first principles. Noting that  $\text{Tube}(B^N(T), \rho) = B^N(T + \rho)$ , we have

$$\begin{aligned} \lambda_N(\text{Tube}(B^N(T), \rho)) &= (T + \rho)^N \omega_N \\ &= \sum_{j=0}^N \binom{N}{j} T^j \rho^{N-j} \omega_N \\ &= \sum_{j=0}^N \omega_{N-j} \rho^{N-j} \binom{N}{j} T^j \frac{\omega_N}{\omega_{N-j}}. \end{aligned}$$

Comparing this to Steiner's formula (3.4.3) it is immediate that

$$(3.4.7) \quad \mathcal{L}_j(B^N(T)) = \binom{N}{j} T^j \frac{\omega_N}{\omega_{N-j}}.$$

For  $S^{N-1}(T)$  the sphere of radius  $T$  in  $\mathbb{R}^N$ , a similar argument, using the fact that

$$\text{Tube}(S^{N-1}(T), \rho) = B^N(T + \rho) - B^N(T - \rho)$$

yields

$$(3.4.8) \quad \mathcal{L}_j(S^{N-1}(T)) = 2 \binom{N}{j} \frac{\omega_N}{\omega_{N-j}} T^j = 2 \binom{N-1}{j} \frac{s_N}{s_{N-j}} T^j$$

if  $N - j$  is even, and 0 otherwise.

Further examples can be found in [83].

A useful normalization of the intrinsic volumes are the so-called *Minkowski functionals*, defined as

$$(3.4.9) \quad \mathcal{M}_j(A) = (j! \omega_j) \mathcal{L}_{N-j}(A)$$

so that, when expressed in terms of the  $\mathcal{M}_j$ , Steiner's formula now reads like a Taylor series expansion

$$(3.4.10) \quad \lambda_N(\text{Tube}(A, \rho)) = \sum_{j=0}^N \frac{\rho^j}{j!} \mathcal{M}_j(A).$$

It is an important and rather deep fact, due to Weyl [102] (see also [40, 52]) that the  $\mathcal{L}_j$  are independent of the ambient space in which sets sit. Because of the reversed numbering system and the choice of constants, this is not true of the Minkowski functionals<sup>10</sup>.

In the current scenario of basic complexes there is alternate way to define intrinsic volumes based on the idea of *kinematic density*. Let  $G_N = \mathbb{R}^N \times O(N)$  be the group of rigid motions on  $\mathbb{R}^N$ , and equip it with Haar measure, normalized to be Lebesgue measure on  $\mathbb{R}^N$  and the invariant probability measure on  $O(N)$ . A formula of Hadwiger states that

$$(3.4.11) \quad \mathcal{L}_j(A) = \binom{N}{j} \int_{G_N} \varphi(A \cap gE_j) \mu_N(dg),$$

---

<sup>10</sup> To see why the  $\mathcal{M}_j$  are dependent on the ambient space, let  $i_{NM}$  be the standard inclusion of  $\mathbb{R}^N$  into  $\mathbb{R}^M$ ,  $M \geq N$ , defined by  $i_{NM}(x) = (x_1, \dots, x_N, 0, \dots, 0) \in \mathbb{R}^M$  for  $x \in \mathbb{R}^N$ . Consider  $M > N$ , and note that the polynomials  $\lambda_M(\text{Tube}(A, \rho))$  and  $\lambda_M(\text{Tube}(i_{NM}(A), \rho))$ , lead to different geometric interpretations. For example, for a curve  $C$  in  $\mathbb{R}^2$ ,  $\mathcal{M}_1(C)$  will be proportional to the arc length of  $C$  and  $\mathcal{M}_2(C) = 0$ , while  $\mathcal{M}_2(i_{2,3}(C))$ , rather than  $\mathcal{M}_1(i_{2,3}(C))$ , is proportional to arc length.

where  $E_j$  is any  $j$ -dimensional affine subspace of  $\mathbb{R}^N$  and

$$(3.4.12) \quad \binom{[N]}{[j]} = \frac{[N]!}{[N-j]![j]!} = \binom{N}{j} \frac{\omega_N}{\omega_{N-j} \omega_j},$$

where  $[N]! = N! \omega_N$ , and  $\varphi$  is our old friend, the Euler characteristic. This representation is important, and we shall return to it later.

The time has now come to explain what all of the above has to do with random fields. For this we need one more result from Integral Geometry, due to Hadwiger [41], which also seems to be the only place to find a proof. However, unless you have a weakness for classical German, you should turn to Schneider [84] to read more about it. In Klain and Rota [52], a proof is given for continuous, invariant functionals on the convex ring, which, as noted previously, is a subset of the basic complexes. Zahle [115] has the most general result in terms of the classes of sets covered, although with continuity replacing monotonicity among the conditions of the Theorem.

**Theorem 3.4.1** *Let  $\psi$  be a real valued function on basic complexes in  $\mathbb{R}^N$ , invariant under rigid motions, additive (in the sense of (3.2.2)) and monotone, in the sense that  $A \subseteq B \Rightarrow \psi(A) \leq \psi(B)$ . Then*

$$(3.4.13) \quad \psi(A) = \sum_{j=0}^N c_j \mathcal{L}_j(A),$$

where  $c_0, \dots, c_N$  are non-negative ( $\psi$ -dependent) constants.

Now take an isotropic field  $f$  on  $\mathbb{R}^N$  and consider the set-indexed functional

$$(3.4.14) \quad \psi(A) \triangleq \mathbb{P} \left\{ \sup_{t \in A} f(t) > u \right\}.$$

Then  $\psi$  is clearly monotone and rigid-motion invariant. Unfortunately, it is not quite additive, since even if  $A$  and  $B$  are disjoint,

$$(3.4.15) \quad \begin{aligned} \psi(A \cup B) &= \psi(A) + \psi(B) \\ &\quad - \mathbb{P} \left\{ \left( \sup_{t \in A} f(t) > u \right) \cup \left( \sup_{t \in B} f(t) > u \right) \right\}. \end{aligned}$$

However, if  $u$  is large then we expect each of the terms in (3.4.15) to be small, with the last term on the right hand side to be of smaller order than each of the others<sup>11</sup>. Arguing heuristically, it is therefore not unreasonable to expect that there might be an invariant, additive *and* monotone

---

<sup>11</sup>This would happen, for example, if  $A$  and  $B$  were sufficiently distant for the values of  $f$  in  $A$  and  $B$  to be close to independent. It will turn out that, at least for Gaussian  $f$ , these heuristics will be true even if  $A$  and  $B$  are close, as long as  $u$  is large enough. [THIS WILL BE A CONSEQUENCE OF THE SO-CALLED SLEPIAN MODELS OF CHAPTER 6.](#)

functional  $\tilde{\psi}_u$  for which

$$(3.4.16) \quad \mathbb{P} \left\{ \sup_{t \in A} f(t) > u \right\} = \tilde{\psi}_u(A) + o\left(\tilde{\psi}_u(A)\right),$$

in which case, by Theorem 3.4.1, we would be able to conclude that

$$(3.4.17) \quad \mathbb{P} \left\{ \sup_{t \in A} f(t) > u \right\} = \sum_{j=0}^N c_j(u) \mathcal{L}_j(A) + \text{lower order terms},$$

for some functions  $c_j(u)$ .

The fact that such a function does indeed exist, and is given by

$$\tilde{\psi}_u = \mathbb{E} \{ \varphi(A_u(f, A)) \},$$

where  $\varphi$  is the Euler characteristic and  $A_u$  an excursion set, is one of the main punch lines of this book and of the last few years of research in Gaussian random fields. Proving this, in wide generality, and computing the coefficients  $c_j$  in (3.4.17) as functions of  $u$  is what much of the next few chapters is about.

If you are interested mainly in nice Euclidean parameter spaces, and do not need to see the details of all the proofs, you can now comfortably skip the rest of this Chapter, with the exception of Section 3.9.2 which lifts Theorem 3.3.4 from two to general dimensions. The same is true if you have a solid background in differential geometry, even if you plan to follow all the proofs later on. You can return later when the need to confirm notation arises.

## 3.5 Manifolds and tensors

From now until the end of this Chapter, which is quite a while away, we move from the realm of Integral Geometry to that of Differential Geometry. Our basic parameter spaces will (eventually) become piecewise smooth Riemannian manifolds rather than simple subsets of  $\mathbb{R}^N$ . Our final results on the Morse theory of Section 3.9 will do in this setting what the results of Section 3.3 did up until now: provide point set representations for the Euler characteristic of excursion sets.

There will be much to do on the way, but the investment will pay dividends in later chapters where it will appear in solving problems that have no intrinsic connection to manifolds. We start, in this Section, with the basic definitions and structures of manifolds, and then turn to tensors and exterior algebras. Then, in Section 3.6, we add a Riemannian structure, which allows us to talk about curvature and integration over manifolds. Section 3.7 treats piecewise smooth manifolds (such as the  $N$  dimensional

cube) a level of generality necessary both for applications and to allow us to recoup the Integral Geometric results from the more general setting. After a short diversion to discuss intrinsic volumes in this setting in Section 3.8 we turn to Morse Theory in Section 3.9.

This is long and not particularly simple path. In particular, as opposed to what has gone before in this Chapter, it is going to be impossible to give full proofs of all that we claim and need. The Morse Theory alone would need its own volume to develop. On the other hand, essentially all that we have to say is ‘well known’, in the sense that it appears in textbooks of Differential Geometry. Thus, the reader familiar with these books will be able to skim the remainder of this Chapter, needing only to pick up our notation and emphases. For the first timer, the situation will be quite different. For him, we have added numerous footnotes and simple examples along the way that are meant to help him through the general theory. Hopefully, even the first timer will then be able to follow the computations of later chapters<sup>12</sup>. However, to go beyond merely following, and develop further results of his own, it will be necessary to learn the material from its classical sources<sup>13</sup>.

### 3.5.1 Manifolds

A differentiable manifold is mathematical generalisation, or abstraction, of objects such as curves and surfaces in  $\mathbb{R}^N$ . Intuitively, it is a smooth set with a locally defined, Euclidean, coordinate system. Thus, the place to start is with the construction of such a coordinate system.

We call  $M$  a *topological  $N$ -manifold* if  $M$  is a locally compact Hausdorff space and, for every  $t \in M$ , there exists an open  $U \subset M$  containing  $t$ , an open  $\tilde{U} \in \mathbb{R}^N$ , and a homeomorphism  $\varphi : U \rightarrow \tilde{U}$ .

To add a differential structure to such a manifold, and to talk about smoothness, we need the concepts of charts and atlases. A (*coordinate*) *chart* on  $M$  is simply a pair  $(\varphi, U)$ , where, as above,  $U \subset M$  is open and  $\varphi : U \rightarrow \varphi(U) \subset \mathbb{R}^N$  is a homeomorphism. A collection  $\{\varphi_i : U_i \rightarrow \mathbb{R}^N\}_{i \in I}$  of charts is called  *$C^k$  compatible* if

$$(3.5.1) \quad \varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \rightarrow \varphi_i(U_i \cap U_j)$$

is a  $C^k$  diffeomorphism<sup>14</sup> for every  $i, j \in I$  for which  $U_i \cap U_j \neq \emptyset$ .

---

<sup>12</sup>Actually, we will be satisfied with the exposition as long as we have not achieved the double failure of both boring the expert and bamboozling the novice.

<sup>13</sup>For the record, the books that we found most useful were Boothby, [14], Jost [49], Millman and Parker [70], Morita [71] and O’Neill [74]. The two recent books by Lee [58, 59] stand out from the pack as being particularly easy to read, and we highly recommend them as the right place to start.

<sup>14</sup>i.e. Both  $\varphi_i \circ \varphi_j^{-1}$  and its inverse  $\varphi_j \circ \varphi_i^{-1}$  are  $k$  times differentiable as functions from subsets of  $\mathbb{R}^N$  to subsets of  $\mathbb{R}^N$ .

If a collection of charts gives a covering of  $M$  – i.e.  $\bigcup_{i \in I} U_i = M$  – then it is called an ( $C^k$ ) *atlas* for  $M$ . An atlas is called *maximal* if it is not contained in any strictly larger atlas with homeomorphisms satisfying (3.5.1). Finally, we call a topological  $N$  manifold, together with a  $C^k$  maximal atlas, a  $C^k$  *differential manifold*. The maximal atlas is often referred to as the *differential structure* of  $M$ .

If some of the  $U_i$  are subsets of the ‘half-space’  $\mathbb{H}^N \triangleq \mathbb{R}^{N-1} \times \mathbb{R}_+$  in  $\mathbb{R}^N$ , then we talk about a *manifold with boundary*, rather than a simple manifold. A manifold with boundary can be thought of as a disjoint union of two manifolds:  $\partial M$ , its boundary, an  $N - 1$  dimensional manifold, and  $M^\circ$ , its interior, an  $N$ -dimensional manifold. For the moment, we shall concentrate on manifolds without boundary. Later on, however, boundaries will be of crucial importance for us.

The next step is to give a formalism for discussing the continuity and differentiability of functions on a  $C^k$  manifold. In essence this is straightforward, and a function  $f : M \rightarrow \mathbb{R}$  is said to be of class  $C^k$  if  $f \circ \varphi^{-1}$  is of class  $C^k$ , in the usual Euclidean sense, for every chart in the atlas.

What is somewhat more complicated, however, is the notion of tangent spaces and the notion of a derivative along vectors in these spaces.

For a manifold  $M$  embedded in  $\mathbb{R}^N$ , such as a curve or a surface, it is straightforward to envision what is meant by a *tangent vector* at a point  $t \in M$ . It is no more than a vector with origin at  $t$ , sitting in the tangent plane to  $M$  at  $t$ . Given such a vector,  $v$ , one can differentiate functions  $f : M \rightarrow \mathbb{R}$  along the direction  $v$ . Thus, to each  $v$  there corresponds a local derivative. For abstract manifolds, the basic notion is not that of these vectors, but rather that of differential operators.

To see how this works, we start with the simplest case, in which  $M = \mathbb{R}^N$  and everything reduces to little more than renaming familiar objects. Here we can manage with the atlas containing the single chart  $(M, i_{NN})$ , where  $i_{NN}$  is the inclusion map. We change notation after the inclusion, writing  $x = i_{NN}(t)$  and  $M' = i_{NN}(M)$  ( $= \mathbb{R}^N$ ). To every vector  $X_x$  with origin  $x \in M'$ , we can assign a linear map from  $C^1(M')$  to  $\mathbb{R}$  as follows:

If  $f \in C^1(M)$  and  $X_x$  is a vector of the form  $X_x = x + \sum_{i=1}^N a_i e_i$ , where  $\{e_i\}_{1 \leq i \leq N}$  is the standard basis for  $\mathbb{R}^N$ , we define the differential operator<sup>15</sup>  $X_x$  by its action on  $f$ :

$$(3.5.2) \quad X_x f = \sum_{i=1}^n a_i \left. \frac{\partial f}{\partial x_i} \right|_x.$$

It is elementary calculus that  $X_x$  satisfies the product rule  $X_x(fg) = fX_x g + gX_x f$ , and that for any two functions  $f, g$  that agree on a neigh-

<sup>15</sup>Hopefully, the standard usage of  $X_x$  to denote both a vector and a differential operator will not cause too much confusion. In this simple case, they clearly amount to essentially the same thing. In general, they are the same by definition.

borhood of  $x$  we have  $X_x f = X_x g$ . For each  $x \in U$ , identifying  $\frac{\partial}{\partial x_i} \Big|_x$  with  $e_i$ , we see that

$$\left\{ \frac{\partial}{\partial x_i} \Big|_x \right\}_{1 \leq i \leq N}$$

forms a basis for a  $N$ -dimensional space of first-order differential operators, which we call the *tangent space* at  $x$  and denote by  $T_x \mathbb{R}^N$ . Returning to  $M$ , and identifying  $T_t \mathbb{R}^N$  with  $T_x \mathbb{R}^N$ , we have now defined the tangent space here as well.

We now turn to the case of an abstract manifold  $M$ . The idea will be to take the vectors in the tangent spaces of  $\mathbb{R}^N$  and somehow push them ‘up’ to get the tangent spaces for  $M$ . More formally, if  $x \in \varphi(U) \subset \mathbb{R}^N$  for some chart  $(U, \varphi)$ , then we can lift the basis of  $T_x \mathbb{R}^N$  (built as in the Euclidean case) to the point  $t = \varphi^{-1}(x)$  via  $\varphi_*^{-1}$ , the so-called differential or *push-forward* of  $\varphi^{-1}$ . We define  $\varphi_*^{-1}$  by

$$(3.5.3) \quad (\varphi_*^{-1}(X_{\varphi(t)})) f = X_{\varphi(t)} (f \circ \varphi^{-1})$$

for any  $f \in C^1(M)$ . It is not hard to show that the set

$$(3.5.4) \quad \left\{ \varphi_*^{-1} \left( \frac{\partial}{\partial x_i} \Big|_{\varphi(t)} \right) \right\}_{1 \leq i \leq N}$$

is linearly independent and we define the space it spans to be  $T_t M$ , the *tangent space of  $M$  at  $t$* . Its elements  $X_t$ , while being differential operators, are called the *tangent vectors* at  $t$ .

With some abuse of notation, in a chart  $(U, \varphi)$  with a coordinate system  $(x_1, \dots, x_N)$  for  $\varphi(U)$  the basis (3.5.4) is usually written as

$$(3.5.5) \quad \left\{ \frac{\partial}{\partial x_i} \Big|_t \right\}_{1 \leq i \leq N},$$

and is referred to as the *natural basis* for  $T_t M$  in the chart  $(U, \varphi)$ .

Now that we know what vector fields are, we need to know how they transform under smooth transformations of manifolds. Specifically, given two manifolds  $M$  and  $N$  and any  $g \in C^1(M; N)$ <sup>16</sup> we can define its *push-forward*  $g_*$  by defining it in any charts  $(U, \varphi)$  and  $(V, \psi)$  of  $M$  and  $N$ , such that  $g(U) \cap V \neq \emptyset$ . We define  $g_*$  by

$$(g_* X_t) h = X_t (h \circ g)$$

---

<sup>16</sup> $C^k(M; N)$ , the space of ‘ $k$  times differentiable functions from  $M$  to  $N$ ’, is defined analogously to  $C^1(M)$ . Thus  $f \in C^k(M; N)$  if, for all  $t \in M$ , there is a chart  $(U_M, \varphi_M)$  for  $M$ , a neighbourhood  $V$  of  $t$  with  $V \subset U_M$ , such that  $f(V) \subset U_N$  for some chart  $(U_N, \varphi_N)$  for  $N$ , for which the composite map  $\varphi_N \circ f \circ (\varphi_M)^{-1} : \varphi_M(V) \rightarrow \varphi_N(f(V))$  is  $C^k$  in the usual, Euclidean, sense.

for any  $h \in C^1(N)$ . We can thus think of  $g_*$  as representing the usual chain rule in vector calculus.

So far, we have worked with single points on the manifold. However, since each  $X_t \in T_tM$  is a linear map on  $C^1(M)$  satisfying the product rule, we can construct a first-order differential operator  $X$ , called a *vector field*, that takes  $C^k$  functions ( $k \geq 1$ ) to real-valued functions, as follows:

$$(Xf)_t = X_t f \quad \text{for some } X_t \in T_tM.$$

In other words, a vector field is a map that assigns, to each  $t \in M$ , a tangent vector  $X_t \in T_tM$ . Thus, in a chart  $(U, \varphi)$  with coordinates  $(x_1, \dots, x_n)$ , a vector field can be represented (cf. (3.5.5)) as

$$(3.5.6) \quad X_t = \sum_{i=1}^N a_i(t) \left. \frac{\partial}{\partial x_i} \right|_{\varphi(t)}.$$

If the  $a_i$  are  $C^k$  functions, we can talk about  $C^k$  vector fields. Note that, for  $j \geq 1$ , a  $C^k$  vector field maps  $C^j(M)$  to  $C^{\min(j-1, k)}(M)$ .

The next step is to note that the collection of all tangent spaces  $T_tM$ ,  $t \in M$ , can be parametrized in a natural way into a manifold,  $T(M)$ , called the *tangent bundle*. In order to describe this, we need to first define the general notion of vector bundles.

A *vector bundle* is a triple  $(E, M, F)$ , along with a map  $\pi : E \rightarrow M$ , where  $E$  and  $M$  are  $N + q$  and  $N$  dimensional manifolds, respectively, and  $F$  is a  $q$  dimensional vector space. We require that  $E$  is locally a product. That is, every  $t \in M$  has a neighborhood  $U$  such that there is a homeomorphism  $\varphi_U : U \times F \rightarrow \pi^{-1}(U)$  satisfying  $(\pi \circ \varphi_U)(t, f_U) = t$ , for all  $f_U \in F$ . Furthermore, for any two such overlapping neighborhoods  $U, V$  with  $t \in U \cap V$ , we require that

$$\varphi_U(t, f_U) = \varphi_V(t, f_V) \iff f_U = g_{UV}(t)f_V,$$

where  $g_{UV}(t) : F \rightarrow F$  is a non-degenerate linear transformation. The functions  $g_{UV} : U \cap V \rightarrow GL(F)$  are called the *transition functions* of the bundle  $(E, M, F)$ . The manifold  $M$  is called the *base manifold* and the vector space  $\pi^{-1}(t)$ , which is isomorphic to  $F$  (as a vector space), is called the *fiber* over  $t$ . Usually, it is clear from the context what  $M$  and  $F$  are, so we refer to the bundle as  $E$ . A  $C^k$  *section* of a vector bundle is a  $C^k$  map  $s : M \rightarrow E$  such that  $\pi \circ s = \text{id}_M$ , where  $\text{id}_M$  is the identity map on  $M$ . In other words, if one thinks of a vector bundle as assigning, to each point of  $t \in M$ , the entire vector space  $\pi^{-1}(t)$ , then a  $C^k$  section is rule for choosing from this space in a smooth ( $C^k$ ) fashion.

In a similar fashion, one can define *fiber bundles* when the fibers  $F$  are not vector spaces, although we shall not go through such a construction yet. Two such fiber bundles which will be of interest to us are the *sphere bundle*  $S(M)$  and the *orthonormal frame bundle*  $\mathcal{O}(M)$  of a Riemannian

manifold  $(M, g)$ . These will be described below once we have the definition of a Riemannian manifold.

We now have the necessary vocabulary to define tangent bundles, which are, in fact, just a special case of the general construction of *tensor bundles* on  $M$ . We give an outline here to which we shall refer later when we construct tensor bundles in Section 3.5.3. In a chart  $(U, \varphi)$ , any tangent vector  $X_t$ ,  $t \in U$ , can be represented as in (3.5.6), so the set of all tangent vectors at points  $t \in U$  is a  $2N$ -dimensional space, locally parametrized by  $\tilde{\varphi}(X_t) = (x_1(t), \dots, x_N(t); a_1(t), \dots, a_N(t))$ . Call this  $E_t$ , and call the projection of  $E_t$  onto its last  $N$  coordinates  $F_t$ . Denote the union (over  $t \in M$ ) of the  $E_t$  by  $E$ , and of the  $F_t$  by  $F$ , and define the natural projection  $\pi : E \rightarrow M$  given by  $\pi(X_t) = t$ , for  $X_t \in E_t$ . The triple  $(E, M, F)$ , along with  $\pi$ , defines a vector bundle, which we call the *tangent bundle* of  $M$  and denote by  $T(M)$ .

The tangent bundle of a manifold is itself a manifold, and as such can be given a differential structure in the same way that we did for  $M$ . To see this, suppose  $M$  is a  $C^k$  manifold and note that an atlas  $\{U_i, \varphi_i\}_{i \in I}$  on  $M$  determines a covering on  $T(M)$ , the charts  $\{\pi^{-1}(U_i), \tilde{\varphi}_i\}_{i \in I}$  of which determine a topology on  $T(M)$ , the smallest topology such that  $\tilde{\varphi}_{|\pi^{-1}(U)}$  are homeomorphisms. In any two charts  $(U, \varphi)$  with coordinates  $(x_1, \dots, x_n)$  and  $(V, \psi)$  with coordinates  $(y_1, \dots, y_n)$  with  $U \cap V \neq \emptyset$ , a point  $X_t \in T(M)$  is represented by

$$\begin{aligned} X_t &= (x_1(t), \dots, x_N(t); a_1(t), \dots, a_N(t)) \\ &= (y_1(t), \dots, y_N(t); b_1(t), \dots, b_N(t)), \end{aligned}$$

where we have the relation

$$(3.5.7) \quad b_i(t) = \sum_{j=1}^N a_j(t) \frac{\partial y^i}{\partial x^j},$$

and  $\partial y^i / \partial x^j$  is the first order partial derivative of the real valued function  $y^i$  with respect to  $x^j$ .

Since  $\phi_U \circ \phi_V^{-1}$  is a  $C^k$  diffeomorphism, we have that  $\tilde{\phi}_U \circ \tilde{\phi}_V^{-1}$  is a  $C^{k-1}$  diffeomorphism, having lost one order of differentiation because of the partial derivatives in (3.5.7). In summary, the atlas  $\{U_i, \phi_i\}_{i \in I}$  determines a  $C^{k-1}$  differential structure on  $T(M)$  as claimed.

Now that we understand the basic structure of the tangent bundle  $T(M)$ , the next step will be to investigate how to work with it. For this, however, we require some preparatory material.

### 3.5.2 Tensors and exterior algebras

Tensors are basically linear operators on vector spaces. They have a multitude of uses, but there are essentially only two main approaches to setting

up the (somewhat heavy) definitions and notations that go along with them. When tensors appear in Applied Mathematics or Physics, the definitions involve high dimensional arrays that transform (as the ambient space is transformed) according to certain definite rules. The approach we shall adopt, however, will be the more modern Differential Geometric one, in which the transformation rules result from an underlying algebraic structure via which tensors are defined. This approach is neater and serves two of our purposes: We can fit everything we need into three pages, and the approach is essentially coordinate free<sup>17</sup>. The latter is of obvious importance for the manifold setting. The downside of this approach is that if this is the first time you see this material, it is very easy to get lost among the trees without seeing the forest. Thus, since one of our main uses of tensors will be for volume (and, later, curvature) computations on manifolds, we shall accompany the definitions with a series of footnotes showing how all of this relates to simple volume computations in Euclidean space. Since manifolds are locally Euclidean, this might help a first-timer get some feeling for what is going on.

To handle tensors, we shall need exterior algebras, which contain sets of rules for combining tensors of various kinds. We shall have need of these to operate on tangent bundles both now, when we are discussing the basic geometry of manifolds, and perhaps more surprisingly, later, when we finally get around to computing explicit expressions for the expectations of quite simple functionals of Gaussian random fields. The most useful parts of this Subsection are probably formulae (3.5.12) and (3.5.17) for the trace of certain tensors.

Now for the formalism. The examples, which provide the motivation, will come later.

Given an  $N$ -dimensional vector space  $V$  and its dual  $V^*$ , a map  $\omega$  is called a tensor of order  $(n, m)$  if

$$\omega \in L(\underbrace{V \oplus \cdots \oplus V}_n \oplus \underbrace{V^* \oplus \cdots \oplus V^*}_m; \mathbb{R})$$

where  $L(E; F)$  denotes the set of (multi) linear<sup>18</sup> mappings between two vector spaces  $E$  and  $F$ . We denote the space of tensors of order  $(n, m)$  by  $\mathcal{T}_m^n$ , where  $n$  is said to be the *covariant* order and  $m$  the *contravariant* order<sup>19</sup>.

<sup>17</sup>Ultimately, however, we shall need to use the array notation when we come to handling specific examples. It appears, for example, in the connection forms of (3.6.7) and the specific computations of Section 3.6.4.

<sup>18</sup>A multi-linear linear mapping is linear, separately, in each variable. In general, we shall not bother with the prefix.

<sup>19</sup>For a basic example, let  $V = \mathbb{R}^3$ . There are three, very elementary, covariant tensors of order 1 operating on vectors  $v = (v_1, v_2, v_3)$ . These are  $\theta_1(v) = v_1$ ,  $\theta_2(v) = v_2$  and  $\theta_3(v) = v_3$ . Thus  $\theta_j$  measures the 'length' of  $v$  in direction  $j$ , or, equivalently, the 'length'

Let  $\mathcal{T}(V) = \bigoplus_{i,j=1}^{\infty} \mathcal{T}_i^j(V)$  be the direct sum of all the tensor spaces  $\mathcal{T}_j^i(V)$ . Then we can define a bilinear associative product, the *tensor product*  $\otimes : \mathcal{T}(V) \times \mathcal{T}(V) \rightarrow \mathcal{T}(V)$ , by defining it on  $\mathcal{T}_j^i \times \mathcal{T}_l^k$  as

$$\begin{aligned} (\alpha \otimes \beta)(v_1, \dots, v_{i+k}, v_1^*, \dots, v_{j+l}^*) \\ \triangleq \alpha(v_1, \dots, v_i, v_1^*, \dots, v_j^*) \times \beta(v_{i+1}, \dots, v_{i+k}, v_{j+1}^*, \dots, v_{j+l}^*). \end{aligned}$$

We can split the algebra  $(\mathcal{T}(V), \otimes)$  into two subalgebras, the covariant and contravariant tensors  $\mathcal{T}^*(V) = \bigoplus_{i=1}^{\infty} \mathcal{T}_0^i(V)$  and  $\mathcal{T}_*(V) = \bigoplus_{i=1}^{\infty} \mathcal{T}_i^0(V)$ . Of special interest in differential geometry are the covariant tensors  $\mathcal{T}^*(V)$ , and so, for the rest of this section, we shall restrict our discussions to  $\mathcal{T}^*(V)$ .

A covariant tensor  $\gamma$  of order  $k$  (a  $(k, 0)$ -tensor) is said to be *alternating* if

$$\gamma(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \varepsilon_{\sigma} \gamma(v_1, \dots, v_k) \quad \forall \sigma \in S(k),$$

where  $S(k)$  is the symmetric group of permutations of  $k$  letters and  $\varepsilon_{\sigma}$  is the sign of the permutation  $\sigma$ . It is called *symmetric* if

$$\gamma(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \gamma(v_1, \dots, v_k) \quad \forall \sigma \in S(k).$$

Thus, for example, computing the determinant of the matrix formed from  $N$ -vectors gives an alternating covariant tensor of order  $N$  on  $\mathbb{R}^N$ .

For  $k \geq 0$ , we denote by  $\Lambda^k(V)$  (respectively,  $\text{Sym}(\mathcal{T}_0^k(V))$ ) the space of alternating (symmetric) covariant tensors on  $V$ , and by  $\Lambda^*(V) = \bigoplus_{k=0}^{\infty} \Lambda^k(V)$  ( $\text{Sym}^*(V)$ ) their direct sums. If  $k = 0$ , then both  $\Lambda^0(V)$  and  $\text{Sym}^0(V)$  are isomorphic to  $\mathbb{R}$ . Note that  $\Lambda^k(V) \equiv \{0\}$  if  $k > N$ , so that  $\Lambda^*(V)$  is actually a finite dimensional vector space. Furthermore, there are natural projections  $\mathcal{A} : \mathcal{T}^*(V) \rightarrow \Lambda^*(V)$  and  $\mathcal{S} : \mathcal{T}^*(V) \rightarrow \text{Sym}^*(V)$  defined on each  $\Lambda^k(V)$  by

$$\begin{aligned} \mathcal{A}\gamma(v_1, \dots, v_k) &= \frac{1}{k!} \sum_{\sigma \in S(k)} \varepsilon_{\sigma} \gamma(v_{\sigma(1)}, \dots, v_{\sigma(k)}), \\ \mathcal{S}\gamma(v_1, \dots, v_k) &= \frac{1}{k!} \sum_{\sigma \in S(k)} \gamma(v_{\sigma(1)}, \dots, v_{\sigma(k)}). \end{aligned}$$

Using  $\mathcal{A}$  we can define a bilinear, associative product called the *wedge product*<sup>20</sup>,  $\wedge : \Lambda^r(V) \times \Lambda^s(V) \rightarrow \Lambda^{r+s}(V)$  by defining it on  $\Lambda^r(V) \times \Lambda^s(V)$

---

of the projection of  $v$  onto the  $j$ -th axis, where ‘length’ is signed. The fact that the length may be signed is crucial to all that follows. Not also that, since  $\mathbb{R}^3$  is its own dual, we could also treat the same  $\theta_i$  as contravariant tensors of order 1.

<sup>20</sup>Continuing the example of Footnote 19, take  $u, v \in \mathbb{R}^3$  and check that this definition gives  $(\theta_1 \wedge \theta_2)(u, v) = 2(u_1 v_2 - v_1 u_2)$ , which is, up to a sign and a factor of 2, the area of the parallelogram with corners  $0, \pi(u), \pi(v)$  and  $\pi(u) + \pi(v)$ , where  $\pi(u)$  is the projection of  $u$  onto the plane spanned by the first two coordinate axes. That is, this particular

by

$$(3.5.8) \quad \alpha \wedge \beta = \frac{(r+s)!}{r!s!} \mathcal{A}(\alpha \otimes \beta).$$

The algebra  $(\wedge^*(V), \wedge)$  is called the *Grassman* or *exterior algebra* of  $V$ .

The next step is to note that there are some relatively simple relationships between the structures of  $\wedge^*(V)$  and those of  $V$  and its dual  $V^*$ . For example, if  $B_{V^*} = \{\theta_1, \dots, \theta_n\}$  is a basis for  $V^*$ , then

$$(3.5.9) \quad B_{\wedge^*(V)} = \bigcup_{j=1}^N \{\theta_{i_1} \wedge \dots \wedge \theta_{i_j} : i_1 < i_2 < \dots < i_j\}$$

is a basis<sup>21</sup> for  $\wedge^*(V)$ , a vector space of dimension  $\sum_{k=0}^N \binom{N}{k} = 2^N$ . Furthermore, given an inner product on  $V$ , there is a natural, corresponding inner product on  $\wedge^*(V)$ . To define it, fix an orthonormal basis  $B_V = \{v_1, \dots, v_N\}$  for  $V$ , which in turn uniquely determines a dual basis  $B_{V^*} = \{\theta_1, \dots, \theta_N\}$  for  $V^*$ . Carry out the construction in (3.5.9) to get a basis for  $\wedge^*(V)$ . Now take the unique inner product on  $\wedge^*(V)$  which makes this basis orthogonal. This is the ‘corresponding’ inner product to which we are referring.

In a similar fashion, inner products on  $V$  can be used to define corresponding inner products on any  $\mathcal{T}_m^n(V)$ , and thus on any direct sum of any finite collection of tensor spaces.

We are now converging to what will be, for us, one of the most important objects of this Subsection – the definition and basic properties of the trace of a tensor.

alternating covariant tensor of order 2 performs an area computation, where ‘areas’ may be negative. Now take  $u, v, w \in \mathbb{R}^3$ , and take the wedge product of  $\theta_1 \wedge \theta_2$  with  $\theta_3$ . A little algebra gives that that  $(\theta_1 \wedge \theta_2 \wedge \theta_3)(u, v, w) = 6 \times \det \langle u, v, w \rangle$  where  $\langle u, v, w \rangle$  is the matrix with  $u$  in the first column, etc. In other words, it is 6 times the (signed) volume of the parallelepiped three of whose edges are  $u, v$  and  $w$ . Extending this example to  $N$  dimensions, you should already be able to guess a number of important facts, including:

- (i) Alternating covariant tensors of order  $n$  have to a lot to do with computing  $n$ -dimensional (signed) volumes.
- (ii) The wedge product is a way of combining lower dimensional tensors to generate higher dimensional ones.
- (iii) Our earlier observation that  $\wedge^k(V) \equiv \{0\}$  if  $k > N$  translates, in terms of volumes, to the trivial observation that the  $k$ -volume of an  $N$ -dimensional object is zero if  $k > N$ .
- (iv) Since area computations and determinants are intimately related, so will be tensors and determinants.

<sup>21</sup>The notation of the example of Footnote 19 should now be clear. The 1-tensor  $\theta_j$  defined by  $\theta_j(v) = v_j$  is in fact the  $j$ -th basis element of  $V^*$  if  $V$  is given the usual basis  $\{e_j\}_{j=1}^N$ . Thus, in view of the fact that  $B_{\wedge^*(V)}$  is a basis for  $\wedge^*(V)$ , the examples of tensors of order 2 and 3 that we built are actually archetypal.

We start with  $\Lambda^{n,m}(V) \triangleq \Lambda^n(V) \otimes \Lambda^m(V)$  (i.e. the linear span of the image of  $\Lambda^n(V) \times \Lambda^m(V)$  under the map  $\otimes$ ) and let  $\Lambda^*(V) \otimes \Lambda^*(V) = \bigoplus_{n,m=0}^{\infty} \Lambda^{n,m}(V)$ . We call an element of  $\Lambda^{n,m}(V)$  a *double form of type*  $(n, m)$ . Note that a  $(n, m)$  double form is alternating in its first  $n$  and last  $m$  variables.

We can define a product  $\cdot$  on  $\Lambda^*(V) \otimes \Lambda^*(V)$  by first defining it on tensors of the form  $\gamma = \alpha \otimes \beta$  by

$$(\alpha \otimes \beta) \cdot (\alpha' \otimes \beta') = (\alpha \wedge \alpha') \otimes (\beta \wedge \beta'),$$

and then extending by linearity. For  $\gamma \in \Lambda^{n,m}(V)$  and  $\theta \in \Lambda^{p,q}(V)$  this gives  $\gamma \cdot \theta \in \Lambda^{n+p,m+q}(V)$  for which

(3.5.10)

$$\begin{aligned} & (\gamma \cdot \theta) ((u_1, \dots, u_{n+p}), (v_1, \dots, v_{m+q})) \\ &= \frac{1}{n!m!p!q!} \sum_{\substack{\sigma \in S(n+p) \\ \rho \in S(m+q)}} \varepsilon_{\sigma} \varepsilon_{\rho} \left[ \gamma((u_{\sigma_1}, \dots, u_{\sigma_n}), (v_{\rho_1}, \dots, v_{\rho_m})) \right. \\ & \quad \left. \times \theta((u_{\sigma_{n+1}}, \dots, u_{\sigma_{n+p}}), (v_{\rho_{m+1}}, \dots, v_{\rho_{m+q}})) \right]. \end{aligned}$$

Note that a double form of type  $(n, 0)$  is simply an alternating covariant tensor, so that, comparing (3.5.10) with (3.5.8), it is clear that in that case the dot product that we have just defined reduces to the usual wedge product.

We shall be most interested in the restriction of this product to

$$\Lambda^{*,*}(V) = \bigoplus_{j=0}^{\infty} \Lambda^j(V) \otimes \Lambda^j(V).$$

The pair  $(\Lambda^{*,*}(V), \cdot)$  is then a commutative (sub) algebra<sup>22</sup>.

For a double form  $\gamma \in \Lambda^{*,*}(V)$ , we define the polynomial  $\gamma^j$  as the product of  $\gamma$  with itself  $j$  times for  $j \geq 1$ , and set  $\gamma^0 = 1$ . If  $\gamma$  is of type  $(k, k)$  then  $\gamma^j$  is of type  $(jk, jk)$ . Furthermore, for powers (3.5.10) simplifies somewhat. In particular, if  $\gamma$  is a  $(1, 1)$  double form, then it is easy to check from (3.5.10) that

$$(3.5.11) \quad \gamma^j((u_1, \dots, u_j), (v_1, \dots, v_j)) = j! \det \left( \gamma(u_k, v_{\ell})_{k,\ell=1,\dots,j} \right).$$

Since, as described above, any inner product  $\langle \cdot, \cdot \rangle$  on  $V$  induces an inner product on  $\Lambda^*(V)$ ,  $\langle \cdot, \cdot \rangle$  also induces a real-valued map on  $\Lambda^{*,*}(V)$ , the *trace*, denoted by  $\text{Tr}$ . The trace is defined on tensors of the form  $\gamma = \alpha \otimes \beta$  by

$$\text{Tr}(\gamma) = \langle \alpha, \beta \rangle,$$

<sup>22</sup>The product is not commutative on all of  $\Lambda^*(V) \otimes \Lambda^*(V)$  since, in general, for  $\alpha \in \Lambda^{n,m}(V)$  and  $\beta \in \Lambda^{p,q}(V)$  we have  $\alpha \cdot \beta = (-1)^{np+mq} \beta \cdot \alpha$ .

and then extended by linearity to the remainder of  $\bigwedge^*(V)$ . Given an orthonormal basis  $(v_1, \dots, v_n)$  of  $V$ , a simple calculation shows that, for  $\gamma \in \Lambda^{k,k}(V)$ ,  $k \geq 1$ , we have the following rather useful expression for  $\text{Tr}$ :

$$(3.5.12) \quad \text{Tr}(\gamma) = \frac{1}{k!} \sum_{a_1, \dots, a_k=1}^N \gamma((v_{a_1}, \dots, v_{a_k}), (v_{a_1}, \dots, v_{a_k})).$$

If  $\gamma \in \Lambda^{0,0}(V)$ , then  $\gamma \in \mathbb{R}$  and we define  $\text{Tr}(\gamma) = \gamma$ . One can also check that while the above seemingly depends on the choice of basis, the trace operator is, in fact, basis independent, a property generally referred to as *invariance*.

There is also a useful extension to (3.5.12) for powers of symmetric forms in  $\gamma \in \Lambda^{1,1}$ . Using (3.5.11) to compute  $\gamma^j$ , and (3.5.12) to compute its trace, it is easy to check that

$$(3.5.13) \quad \text{Tr}(\gamma^j) = j! \text{detr}_j(\gamma(u_i, u_j))_{i,j=1, \dots, n},$$

where, for a matrix  $A$ ,

$$(3.5.14) \quad \text{detr}_j(A) \triangleq \text{Sum over all } j \times j \text{ principle minors of } A.$$

One last observation that we need later is that to each  $\gamma \in \Lambda^{k,k}(V)$  there corresponds a linear map  $T_\gamma : \Lambda^k(V) \rightarrow \Lambda^k(V)$ . The correspondence is unique, and so one can think of  $\gamma$  and  $T_\gamma$  as equivalent objects. To define  $T_\gamma$ , take a basis element of  $\Lambda^k(V)$  of the form  $\theta_{i_1} \wedge \dots \wedge \theta_{i_k}$  (cf. (3.5.9)) and define its image under  $T_\gamma$  by setting

$$(3.5.15) \quad T_\gamma(\theta_{i_1} \wedge \dots \wedge \theta_{i_k})(w_1, \dots, w_k) = \gamma(v_{i_1}, \dots, v_{i_k}, w_1, \dots, w_k)$$

and extending linearly to all of  $\Lambda^k(V)$ .

It is a simple computation to check that if  $k = 1$  and we write

$$(3.5.16) \quad I = \sum_{i=1}^N \theta_i \otimes \theta_i$$

then  $T_I$  is the identity from  $\Lambda^1(V)$  to  $\Lambda^1(V)$ . In general,  $T_{I^k}$  is the identity from  $\Lambda^k(V)$  to  $\Lambda^k(V)$ .

We shall need one useful formula (cf. [34], p. 425) later when we calculate the expected Euler characteristic of a Gaussian random field on  $M$ . Choose  $\gamma \in \Lambda^{k,k}(V)$ , and  $0 \leq j \leq N - k$ . Then

$$(3.5.17) \quad \text{Tr}(\gamma I^j) = \frac{(N - k)!}{(N - k - j)!} \text{Tr}(\gamma).$$

Finally, when there is more than one inner product space in consideration, say  $V_1$  and  $V_2$ , we shall denote their traces by  $\text{Tr}^{V_1}$  and  $\text{Tr}^{V_2}$ .

For a more complete description of the properties of traces, none of which we shall need, you could try Section 2 of [34]. This is not easy reading, but is worth the effort.

### 3.5.3 Tensor bundles and differential forms

In Section 3.5.1 we set up the tangent vector spaces  $T_t(M)$  of a differentiable manifold  $M$ , along with the tangent bundle  $T(M)$ . In Section 3.5.2 we saw how to define spaces of tensors and exterior algebras over vector spaces. The purpose of this brief Section is to put these two ideas together, thus giving exterior algebras over the tangent spaces of a manifold a differential structure, and to use this to define the notion of differential forms.

The basic observation is that, given a  $C^k$  manifold  $M$ , its tangent space  $T_tM$  with dual  $T_t^*M$ , we can carry out the constructions of tensors on  $V = T_tM$ , at every  $t \in M$ , exactly as we did in the previous Section. Then, just as we defined vector fields as maps  $t \mapsto X_t \in T_tM$ ,  $t \in M$ , we can define tensor fields, covariant tensor fields, alternating covariant tensor fields, etc. This level of generality also gives new ways of looking at things. For example, since  $T_tM$  is finite dimensional, we can identify  $T_tM$  with  $(T_tM)^{**}$  and thus treat vector fields as tensor fields of order  $(0, 1)$ .

It is also quite simple to associate a differential structure with these objects, following the argument at the end of Section 3.5.1. In particular, since tensor fields of order  $(n, m)$  are determined, locally, in the same way that vector fields are defined, any atlas  $\{U_i, \phi_i\}_{i \in I}$  for  $M$  determines<sup>23</sup> a  $C^{k-1}$  differential structure<sup>24</sup> on the collection

$$\mathcal{T}_m^n(M) \triangleq \bigcup_{t \in M} \mathcal{T}_m^n(T_tM),$$

which itself is called the *tensor bundle of order  $(n, m)$*  over  $M$ .

Constructing the Grassman algebra on each  $T_tM$  and taking the union

$$\Lambda^*(M) \triangleq \bigcup_{t \in M} \Lambda^*(T_tM)$$

gives what is known as the *Grassman bundle* of  $M$ . As before, one can take  $C^k$  sections of  $\Lambda^*(M)$ , and these are called the  $C^k$  *differential forms*<sup>25</sup> of mixed degree.

---

<sup>23</sup>The finite dimensionality of spaces  $\mathcal{T}_m^n(T_tM)$ , noted earlier, is crucial to make this construction work.

<sup>24</sup>Included in this ‘differential structure’ is a set of rules describing how tensor fields transform under transformations of the local coordinate systems, akin to we had for simple vector fields in (3.5.7).

In fact, there is a lot hidden in this seemingly simple sequence of constructions. In particular, recall that at the head of Section 3.5.2 we mentioned that the tensors of Applied Mathematics and Physics are defined via arrays which transform according to very specific rules. However, nowhere in the path we have chosen have these demands explicitly appeared. They are, however, implicit in the constructions that we have just carried out.

<sup>25</sup>The reason for the additional of the adjective ‘differential’ will be explained later. cf. (3.6.10) and the discussion following it.

Similarly, performing the same construction over the  $\Lambda^k(T_tM)$  gives the *bundle of (differential)  $k$ -forms* on  $M$ , while doing it for  $\Lambda^{*,*}(T_tM)$  gives the *bundle of double (differential) forms* on  $M$ .

## 3.6 Riemannian manifolds

### 3.6.1 Riemannian metrics

If you followed the footnotes while we were developing the notion of tensors, you will have noted that we related them to the elementary notions of area and volume in the simple Euclidean setting of  $M = \mathbb{R}^N$ . In general, of course, they are somewhat more complicated. In fact, if you think back, we do not as yet even have a notion of simple distance on  $M$ , let alone notions of volume. Developing this is our next task and, once it is in place, we can also talk about differentiating tensor fields and integrating differential forms.

The first step lies in understanding the notion of a  $C^k$  *Riemannian metric*  $g$  on  $M$ . Formally, this is a  $C^k$  section of  $\text{Sym}(\mathcal{T}_0^2(M))$ , such that for each  $t$  in  $M$ ,  $g_t$  is positive definite; viz.  $g_t(X_t, X_t) \geq 0$  for all  $t \in M$  and  $X_t \in T_tM$ , with equality if, and only if,  $X_t = 0$ . Thus a  $C^k$  Riemannian metric determines a family of smoothly ( $C^k$ ) varying inner products on the tangent spaces  $T_tM$ .

A  $C^{k+1}$  manifold  $M$  together with a  $C^k$  Riemannian metric  $g$  is called a  $C^{k+1}$  *Riemannian manifold*  $(M, g)$ .

The first thing that a Riemannian metric gives us is the size  $g_t(X_t, X_t)$  of tangent vectors  $X_t \in T_tM$ . This immediately allows us to define two rather important tangent bundles.

The *sphere bundle*  $S(M)$  of  $(M, g)$  is the set of all unit tangent vectors of  $M$ ; i.e. elements  $X_t \in T(M)$ , for some  $t \in M$ , with  $g_t(X_t, X_t) = 1$ . This is an example of a bundle with fibres that are not vector spaces, since the fibres are the spheres  $S(T_tM)$ . Another example is the *orthonormal frame bundle*,  $\mathcal{O}(M)$ , the set of all sets of  $N$  unit tangent vectors  $(X_t^1, \dots, X_t^N)$  of  $M$  such that  $(X_t^1, \dots, X_t^N)$  form an orthonormal basis for  $T_tM$ .

Despite its name, a Riemannian metric is not a true metric on  $M$ . However it does induce a metric  $\tau$  on  $M$ . Since  $g$  determines the length of a tangent vector, we can define the length of a  $C^1$  curve  $c: [0, 1] \rightarrow M$  by

$$L(c) = \int_{[0,1]} \sqrt{g_t(c', c')(t)} dt$$

and define the metric  $\tau$  by

$$(3.6.1) \quad \tau(s, t) = \inf_{c \in D^1([0,1]; M)_{(s,t)}} L(c)$$

where  $D^1([0, 1]; M)_{(s,t)}$  is the set of all piecewise  $C^1$  maps  $c : [0, 1] \rightarrow M$  with  $c(0) = s$ ,  $c(1) = t$ .

A curve in  $M$  connecting two points  $s, t \in M$ , along which the infimum in (3.6.1) is achieved, is called a *geodesic* connecting them. Geodesics need not be unique.

Now that we have a notion of distance on  $M$ , we can turn to the problem of differentiating vector fields (and, although we shall not need them, tensor fields) with respect to another vector field. This is considerably more subtle on manifolds than in simple Euclidean space. In essence, the problem is as follows, which is already interesting if  $M$  is a simple surface in  $\mathbb{R}^3$ :

Suppose  $X$  and  $Y$  are two vector fields, and we want ‘differentiate’  $Y$ , at the point  $t \in M$ , in the direction  $X_t$ . Following the usual procedure, we need to ‘move’ a distance  $\delta t$  in the direction  $X_t$  and compute the limit of the ratios  $(Y_t - Y_{t+\delta t})/\delta t$ . There are three problems here. The first is that there is no guarantee (and it will not generally be the case) that  $t + \delta t$  lies on  $M$ , so that  $Y_{t+\delta t}$  is not even defined. So we have to find a way of ‘moving in the direction  $X_t$ ’ while staying on  $M$ . This can, in essence, be done by moving along the unique geodesic with initial point  $t$  and tangent<sup>26</sup> proportional to  $X_t$

Even when this is done,  $Y_t$  and  $Y_{t+\delta t}$  may lie in different spaces, so that it is not clear what is meant by their difference. This problem can be solved by ‘moving’  $Y_{t+\delta t}$  ‘backwards’ along the geodesic joining  $t$  to  $t + \delta t$  so that it can be compared to  $Y_t$ . The last problem is that very often we shall want to work only with vectors in  $T(M)$ , and there is no guarantee (and, again, it will not generally be the case) that  $\lim_{\delta t \rightarrow 0} (Y_t - Y_{t+\delta t})/\delta t \in T_t M$ . This problem can be solved by taking the projection of the limit onto  $T_t M$  as our definition of derivative. All of this can be formalised and the resulting procedure is known as a *covariant differentiation*. It takes a few pages to set up and you can find it in any standard text on differential geometry. Thus we shall skip the details and go directly to what we need from this theory. Hopefully, however, this paragraph has helped at least set up some intuition for the remainder of the Section.

We have already seen that vector fields are actually differential operators taking smooth functions to continuous functions. Since the same is true of smooth sections of a vector bundle  $E$ , and since it will be useful for what we need later, we shall start by setting up a notion of differentiation for general vector bundles  $E$ , rather than just for vector fields.

---

<sup>26</sup>This description may seem somewhat circular, since it uses differentiation along geodesics while we are still attempting to define what we mean by differentiation. However, if you think for a moment about the scenario of  $\mathbb{R}^3$ , you will realise that defining derivatives of simple curves and derivatives of vector fields are quite different issues.

The basic concept is that of a *connection*. A connection on a vector bundle  $E$  with base manifold  $M$ , is a bilinear map

$$D : C^k(E) \times C^q(T(M)) \rightarrow C^{\min(k-1,q)}(E)$$

satisfying

$$(3.6.2) \quad D(f\omega, X) = (Xf)\omega + fD(\omega, X), \quad \forall f \in C^1(M),$$

$$(3.6.3) \quad D(\omega, fX) = fD(\omega, X), \quad \forall f : M \rightarrow \mathbb{R}.$$

If we specialise to the simple Euclidean case of  $M = \mathbb{R}^N$  with the bundle of  $C^1$  zero forms<sup>27</sup> then it should be obvious that we have done no more than formalise differentiation. In this case, if  $|X| = 1$ , then  $D(f, X)$  is precisely the derivative of  $f$  in the direction  $X$ . Thus, try to avoid being confused by the terminology. A connection is no more than a type of derivative<sup>28</sup>.

Every  $C^1$  Riemannian metric determines a unique connection  $\nabla$  on the tangent bundle  $T(M)$ , called the *Riemannian connection* or *Levi-Civita connection* of  $(M, g)$ . The Levi-Civita connection uses slightly different notation, and we write  $\nabla_X Y$  instead of  $D(Y, X)$ , where both  $X, Y \in T(M)$ . It satisfies the following two additional properties, which actually uniquely determine it.

$$(3.6.4) \quad \text{It is } \textit{torsion free}, \text{ i.e.}$$

$$\nabla_X Y - \nabla_Y X - [X, Y] = 0$$

$$(3.6.5) \quad \text{It is } \textit{compatible} \text{ with the metric } g, \text{ i.e.}$$

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z),$$

where  $[X, Y]f = XYf - YXf$  is the so called *Lie bracket*<sup>29</sup> of the vector fields  $X$  and  $Y$ . The Levi-Civita connection also has the rather useful property that  $X, Y \in T(M)$  implies that  $\nabla_X Y \in T(M)$ , so that this notion of differentiation of tangent vectors fields in the direction of tangent vectors never takes us out of the tangent bundle<sup>30</sup>. If  $M = \mathbb{R}^N$  and  $g$  is the usual

<sup>27</sup>The bundle of  $C^k$  zero forms over a manifold is simply the collection of  $C^k$  real valued functions on  $M$ . i.e.  $\omega$  in (3.6.2) and (3.6.2) is no more than a function  $h : M \rightarrow \mathbb{R}$ .

<sup>28</sup> Nevertheless, the name is not accidental. Connections provide a tool for moving vectors from one fibre to another so as to make them comparable. It is therefore intimately related to covariant differentiation, which, as described above, requires a method of moving vectors from one tangent space to another so that differences of the form  $Y_t - Y_{t+\delta t}$  are well defined. Although we shall not need to know how to do this, the 'parallel transport' of a vector  $X_0 \in T_s M$  to  $T_t M$  along a curve  $c(u)$ ,  $0 \leq u \leq 1$  from  $s$  to  $t$  is done by solving the so-called *Ricci differential equation*  $\nabla_{c'(u)} X_u = 0$ , the notation for which we are about to develop.

<sup>29</sup>The Lie bracket measures the failure of partial derivatives to commute, and is always zero in the familiar Euclidean case.

<sup>30</sup>In fact, if this were not true, (3.6.5) would make little sense, since  $g$  is defined only on  $T(M)$ . Note also that, since  $g(Y, Z) \in C^1(M)$ , the expression  $Xg(Y, Z)$  in (3.6.5) is

Euclidean inner product, then it is easy to check that  $\nabla_X Y$  is no more than the directional derivative of the vector field  $Y$  in the directions given by  $X$ . In this case  $\nabla_X Y$  is known as the *flat connection*.

It is a matter of simple algebra to derive from (3.6.4) and (3.6.5) that, for  $C^1$  vector fields  $X, Y, Z$

$$(3.6.6) \quad 2g(\nabla_X Y, Z) = Xg(Y, Z) + Yg(X, Z) - Zg(X, Y) \\ + g(Z, [X, Y]) + g(Y, [Z, X]) + g(X, [Z, Y]).$$

This equation is known as *Koszul's formula*. Its importance lies in the fact that the right hand side depends only on the metric  $g$  and differential-topological notions. Consequently, it gives a coordinate free formula that actually determines  $\nabla$ .

Since  $\nabla_X Y \in T(M)$  it must have a representation in terms of local coordinates. To write this out, we need the notion of a  $C^k$  *orthonormal frame field*  $\{E_i\}_{1 \leq i \leq N}$ , which is a  $C^k$  section of the orthonormal (with respect to  $g$ ) frame bundle  $\mathcal{O}(M)$ . Then there is a collection of  $N^2$  1-forms  $\{\theta_i^j\}_{i,j=1}^N$ , known as *connection forms*, which determine the Riemannian connection via the following two relationships and linearity:

$$(3.6.7) \quad \nabla_X E_i = \sum_{j=1}^N \theta_i^j(X) E_j,$$

$$(3.6.8) \quad \nabla_X(fY) = (Xf)Y + f\nabla_X Y, \quad f \in C^1(M).$$

We shall see in detail how to compute the  $\theta_i^j$  for some examples in Section 3.6.4. In general, they are determined by (3.6.13) below, in which  $\{\theta_i\}_{1 \leq i \leq N}$  denotes the orthonormal *dual frame* corresponding to  $\{E_i\}_{1 \leq i \leq N}$ . To understand (3.6.13) we need one more concept, that of the (*exterior*) *differential* of a  $k$ -form.

If  $f : M \rightarrow \mathbb{R}$  is  $C^1$ , then its *exterior derivative* or *differential*,  $df$ , is the 1-form defined by

$$(3.6.9) \quad df(E_{it}) = f_i(t) \triangleq E_{it}f.$$

Consequently, we can write  $df$ , using the dual frame, as

$$df = \sum_{i=1}^N f_i \theta_i.$$

---

also well defined. Actually, (3.6.5) is a little misleading, and would be more consistent with what has gone before if it were written as

$$X_t g(Y_t, Z_t) = g_t((\nabla_X Y)_t, Z_t) + g_t(Y_t, (\nabla_X Z)_t).$$

The preponderance of  $t$ 's here is what leads to the shorthand of (3.6.5).

If  $\theta = \sum_{i=1}^N h_i \theta_i$  is a 1-form, then its exterior derivative is the 2-form

$$(3.6.10) \quad d\theta \triangleq \sum_{i=1}^N dh_i \wedge \theta_i.$$

Note that the exterior derivative of a 0-form (i.e. a function) gave a 1-form, and that of a 1-form gave a 2-form. There is a general notion of exterior differentiation, which in general takes  $k$ -forms to  $(k+1)$ -forms<sup>31</sup>, but which we shall not need.

We do now, however, have enough to define the 1-forms  $\theta_i^j$  of (3.6.7)–(3.6.8). They are the unique set of  $N^2$  1-forms satisfying the following two requirements:

$$(3.6.11) \quad d\theta^i - \sum_{j=1}^N \theta^j \wedge \theta_i^j = 0,$$

$$(3.6.12) \quad \theta_i^j + \theta_j^i = 0.$$

The Riemannian metric  $g$  is implicit in these equations in that, as usual, it determines the notion of orthonormality and so the choice of the  $\theta_i$ . In fact, (3.6.11) is a consequence of the compatibility requirement (3.6.5), while (3.6.12) is a consequence of the requirement (3.6.4) that the connection be torsion free.

Since it will occasionally be useful, we note that (3.6.12) takes the following form if the tangent frame and its dual are not taken to be orthonormal:

$$(3.6.13) \quad dg_{ij} = \sum_{k=1}^N \left( \theta_k^i g_{kj} + \theta_k^j g_{ki} \right),$$

where  $g_{ij} = (E_i, E_j) = g(E_i, E_j)$ .

### 3.6.2 Integration of differential forms

We now have the tools and vocabulary to start making mathematical sense out of our earlier comments linking tensors and differential forms to volume computations. However, rather than computing only volumes, we shall need general measures over manifolds. Since a manifold  $M$  is a locally compact topological space, there is no problem defining measures over it and, by Riesz's theorem, these are positive, bounded, linear functionals on  $C_c^0(M)$ , the  $c$  denoting compact support. This description, however, does not have much of a geometric flavour to it and so we shall take a different approach. As usual, we need to start with additional terminology and notation.

---

<sup>31</sup>This is why we used the terminology of 'differential forms' when discussing Grassman bundles at the end of Section 3.5.3.

Recall that, for any two manifolds  $M$  and  $N$ , a  $C^1$  map  $g : M \rightarrow N$  induces a mapping  $g_* : T(M) \rightarrow T(N)$ , referred to as the push-forward, (cf. (3.5.3)). Just as  $g_*$  replaces the chain rule of vector calculus, another related mapping  $g^* : \bigwedge^*(T(N)) \rightarrow \bigwedge^*(T(M))$ , called the *pull-back*, replaces the change of variables formula. Actually, the pull-back is more than just a generalization of the change of variables formula, although for our purposes it will suffice to think of it that way. We define  $g^*$  on  $(k, 0)$ -tensors by

$$(3.6.14) \quad (g^* \alpha)(X_1, \dots, X_k) \triangleq \alpha(g_* X_1, \dots, g_* X_k).$$

The pull-back has many desirable properties, the main ones being that it commutes with both the wedge product of forms and the exterior derivative  $d$ . (cf. (3.6.10).)

With the notion of a pull-back defined, we can add one more small piece of notation. Take a chart  $(U, \varphi)$  of  $M$ , and recall the notation of (3.5.5) in which we used  $\{\partial/\partial x_i\}_{1 \leq i \leq N}$  to denote both the natural, Euclidean basis of  $\varphi(U)$  and its push-forward to  $T(U)$ . We now do the same with the notation

$$(3.6.15) \quad \{dx_1, \dots, dx_N\},$$

which we use to denote both the natural, dual coordinate basis in  $\mathbb{R}^N$  (so that  $dx_i(v) = v_i$ ) and its pull-back under  $\varphi$ .

Now we can start defining integration, all of which hinges on a single definition: If  $A \subset \mathbb{R}^N$ , and  $f : A \rightarrow \mathbb{R}$  is Lebesgue integrable over  $A$ , then we define

$$(3.6.16) \quad \int_A f dx_1 \wedge \dots \wedge dx_N \triangleq \int_A f(x) dx,$$

where  $\{dx_i\}_{1 \leq i \leq N}$ , as above, is the natural, dual coordinate basis in  $\mathbb{R}^N$ , and the right hand side is simply Lebesgue integration.

Since the wedge products in the left hand side of (3.6.16) generate all  $N$ -forms on  $\mathbb{R}^N$  (cf. (3.5.9)) this and additivity solves the problem of integrating  $N$ -forms on  $\mathbb{R}^N$  in full generality.

Now we turn to manifolds. For a given chart,  $(U, \varphi)$ , and an  $N$ -form  $\alpha$  on  $U$ , we define the integral of  $\alpha$  over  $V \subset U$  as

$$(3.6.17) \quad \int_V \alpha \triangleq \int_{\varphi(V)} (\varphi^{-1})^* \alpha$$

where the right hand side is defined by virtue of (3.6.16).

To extend the integral beyond single charts, we require a new condition, that of *orientability*. A  $C^k$  manifold  $M$  is said to be *orientable* if there is an atlas  $\{U_i, \varphi_i\}_{i \in I}$  for  $M$  such that, for any pair of charts  $(U_i, \varphi_i), (U_j, \varphi_j)$  with  $U_i \cap U_j \neq \emptyset$ , the Jacobian of the map  $\varphi_i \circ \varphi_j^{-1}$  is positive. For orientable manifolds it is straightforward to extend the integral (3.6.17) to general domains by additivity.

Given an oriented manifold  $M$ , with atlas as above, one can also define the notion of an *oriented (orthonormal) frame field* over  $M$ . This is a (orthonormal) frame field  $\{E_{1t}, \dots, E_{Nt}\}$  over  $M$  for which, for each chart  $(U, \varphi)$  in the atlas, and at each  $t \in U$ , the push-forward  $\{\varphi_* E_{1t}, \dots, \varphi_* E_{Nt}\}$  can be transformed to the standard basis of  $\mathbb{R}^N$  via a transformation with positive determinant.

Given an oriented orthonormal frame field, there is a unique volume form, which we denote by  $\Omega$ , or by  $\text{Vol}_g$  if we want to emphasise the dependence on the metric  $g$ , which plays the rôle of Lebesgue measure on  $M$ , and which is defined by the requirement that

$$(3.6.18) \quad (\text{Vol}_g)_t(E_{1t}, \dots, E_{Nt}) \equiv \Omega_t(E_{1t}, \dots, E_{Nt}) = +1.$$

The integral of  $\Omega$  is comparatively easy to compute. For a fixed (oriented) chart  $(U, \varphi)$ , with natural basis  $\{\partial/\partial x_i\}_{1 \leq i \leq N}$ , write

$$(3.6.19) \quad g_{ij}(t) = g_t\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right),$$

where  $g$  is the Riemannian metric. This determines, for a each  $t$ , a positive definite matrix  $(g_{ij}(t))_{1 \leq i, j \leq N}$ . Then, for  $A \subset U$ ,

$$(3.6.20) \quad \begin{aligned} \int_A \Omega &\equiv \int_{\varphi(A)} (\varphi^{-1})^* \Omega \\ &= \int_{\varphi(A)} \sqrt{\det(g_{ij} \circ \varphi^{-1})} dx_1 \wedge \dots \wedge dx_N \\ &= \int_{\varphi(A)} \sqrt{\det(g_{ij} \circ \varphi^{-1})} dx, \end{aligned}$$

where the crucial intermediate term comes from (3.6.16)–(3.6.19) and some algebra.

Retaining the orientation of  $M$  determined by  $\Omega$ , both this integral and that in (3.6.17) can be extended to larger subsets of  $M$  by additivity<sup>32</sup>.

For obvious reasons, we shall often write the volume form  $\Omega$  as  $dx_1 \wedge \dots \wedge dx_N$ , where the 1-forms  $dx_i$  are the (dual) basis of  $T^*(M)$ .

An important point to note is that  $\text{Vol}_g$  is also the  $N$ -dimensional Hausdorff measure<sup>33</sup> associated with the metric  $\tau$  induced by  $g$ , and so we shall also often write it as  $\mathcal{H}_N$ . In this case we shall usually write integrals as  $\int_M h(t) d\mathcal{H}_N(t)$  rather than  $\int_M h \mathcal{H}_N$ , which would be more consistent with our notation so far.

<sup>32</sup> The last line of (3.6.20), when written out in longhand, should be familiar to most readers as an extension of the formula giving the ‘surface area’ of a regular  $N$ -dimensional surface. The extension lies in the fact that an arbitrary Riemannian metric now appears, whereas in the familiar case there is only one natural candidate for  $g$ .

<sup>33</sup> If  $M$  is a  $N$ -manifold, treat  $(M, \tau)$  as a metric space, where  $\tau$  is the geodesic metric given by (3.6.1). The diameter of a set  $S \subset M$  is then  $\text{diam}(S) = \sup\{\tau(s, t) : s, t \in S\}$

We now return to the issue of orientability. In setting up the volume form  $\Omega$ , we first fixed an orthonormal frame field and then demanded that  $\Omega_t(E_{1t}, \dots, E_{Nt}) = +1$ , for all  $t \in M$ . (cf. (3.6.18).) We shall denote  $M$ , along with this orientation, by  $M^+$ . However, there is another orientation of  $M$  for which  $\Omega = -1$  when evaluated on the orthonormal basis. We write this manifold as  $M^-$ . On an orientable manifold, there are only two such possibilities.

In fact, it is not just  $\Omega$  which can be used to determine an orientation. Any  $N$ -form  $\alpha$  on an orientable manifold can be used to determine one of two orientations, with the orientation being determined by the sign of  $\alpha$  on the orthonormal basis at any point on  $M$ . We can thus talk about the ‘orientation induced by  $\alpha$ ’.

With an analogue for Lebesgue measure in hand, we can set up the analogues of Borel measurability and (Lebesgue) integrability in the usual ways. Furthermore, it follows from (3.6.16) that there is an inherent smoothness<sup>34</sup> in the above construction. In particular, for any continuous, non-vanishing  $N$ -form  $\alpha$  which induces the same orientation on  $M$  as does  $\Omega$ , there is an  $\Omega$ -integrable function  $d\alpha/d\Omega$  for which

$$\int_M \alpha = \int_M \frac{d\alpha}{d\Omega} \Omega.$$

For obvious reasons,  $d\alpha/d\Omega$  is called the *Radon-Nikodym derivative* of  $\alpha$ . For equally obvious reasons, if  $\alpha$  is a form for which  $d\alpha/d\Omega \geq 0$  and

$$\int_{M^+} \alpha = 1,$$

then we say it induces a probability on  $M$ .

The next step is to set up an important result, due to Federer [34] and known as his *coarea formula*, that allows us to break up integrals over

and, for integral  $n$ , the Hausdorff  $n$  measure of  $A \subset M$  is defined by

$$\mathcal{H}_n(A) \triangleq \omega_n 2^{-n} \liminf_{\varepsilon \downarrow 0} \sum_i (\text{diam } B_i)^n,$$

where, for each  $\varepsilon > 0$ , the infimum is taken over all collections  $\{B_i\}$  of open  $\tau$ -balls in  $M$  whose union covers  $A$  and for which  $\text{diam}(B_i) \leq \varepsilon$ . As usual,  $\omega_n = \pi^{n/2}/\Gamma(n/2 + 1)$  is the volume of the unit ball in  $\mathbb{R}^n$ . For the moment, we need only the case  $n = N$ . When both are defined, and the underlying metric is Euclidean, Hausdorff and Lebesgue measures agree.

Later we shall need a related concept, that of *Hausdorff dimension*. If  $A \subset M$ , the Hausdorff dimension of  $A$  is defined as

$$\dim(A) \triangleq \inf \left\{ \alpha : \liminf_{\varepsilon \downarrow 0} \sum_i (\text{diam } B_i)^\alpha = 0 \right\},$$

where the conditions on the infimum and the  $B_i$  are as above.

<sup>34</sup>This smoothness implies, for example, that  $N$ -forms cannot be supported, as measures, on lower dimensional sub-manifolds of  $M$ .

manifolds into iterated integrals over submanifolds of lower dimension. To see how this works, consider a differentiable map  $f : M \rightarrow N$  between two Riemannian manifolds, with  $m = \dim(M) \geq n = \dim(N)$ . Its push-forward,  $f_*$ , defined almost everywhere, is a linear map between two Hilbert spaces and can be written as

$$(3.6.21) \quad \sum_{i=1}^m \sum_{j=1}^n A_{ij}^f \theta_i(t) \otimes \tilde{\theta}_j(f(t))$$

for orthonormal bases  $\{\theta_1(t), \dots, \theta_m(t)\}$  and  $\{\tilde{\theta}_1(f(t)), \dots, \tilde{\theta}_n(f(t))\}$  of  $T_t M$  and  $T_{f(t)} N$ .

Similarly, the pull-back  $f^*$ , restricted to  $t \in M$ , is a linear map from  $\bigwedge^*(T_{f(t)} N)$  to  $\bigwedge^*(T_t M)$ . If we restrict  $f^*(t)$  to  $\Lambda^n(T_{f(t)} N)$ , it can be identified with an element of  $\Lambda^n(T_{f(t)} N) \otimes \Lambda^n(T_t M)$  and its norm  $Jf(t)$  can be calculated by taking the square root of the sum of squares of the determinants of the  $k \times k$  minors of the matrix  $A^f$  of (3.6.21). Note that if  $f$  is Lipschitz, then  $Jf$  is a bounded function.

Federer's coarea formula [34] states that for Lipschitz maps  $f$  and  $g \in L^1(M, \mathcal{B}(M), \mathcal{H}_m)$ ,

$$(3.6.22) \quad \int_M g(t) Jf(t) d\mathcal{H}_m(t) = \int_N d\mathcal{H}_n(u) \int_{f^{-1}(u)} g(s) d\mathcal{H}_{m-n}(s).$$

Simplifying matters a little bit, consider two special cases. If  $M = \mathbb{R}^N$  and  $N = \mathbb{R}$ , it is easy to see that  $Jf = |\nabla f|$ , so that

$$(3.6.23) \quad \int_{\mathbb{R}^N} g(t) |\nabla f(t)| dt = \int_{\mathbb{R}} du \int_{f^{-1}\{u\}} g(s) d\mathcal{H}_{N-1}(s).$$

There is not a great deal of simplification in this case beyond the fact that it is easy to see what the functional  $J$  is. On the other hand, if  $M = N = \mathbb{R}^N$ , then  $Jf = |\det \nabla f|$ , and

$$(3.6.24) \quad \begin{aligned} \int_{\mathbb{R}^N} g(t) |\det \nabla f(t)| dt &= \int_{\mathbb{R}^N} du \int_{f^{-1}\{u\}} g(s) d\mathcal{H}_0(s) \\ &= \int_{\mathbb{R}^N} \left( \sum_{t: f(t)=u} g(t) \right) du. \end{aligned}$$

We shall return to (3.6.24) in Section 4.3.

Before leaving integration for curvature, we note the following, which will be important for us later on.

If  $f : M \rightarrow N$  is a diffeomorphism between two oriented manifolds, we have, for any integrable form  $\alpha$  on  $N$ , that

$$\int_M f^* \alpha = \text{sign}(f^*) \int_N \alpha,$$

where  $\text{sign}(f^*)$  is determined as follows: Given any two volume forms  $\Omega_M, \Omega_N$  on  $M, N$  that determine the orientations  $M^+$  and  $N^+$ , we have

$$(3.6.25) \quad f^*\Omega_N = h\Omega_M,$$

where  $h$  is some non-vanishing function (because  $f$  is a diffeomorphism). We set

$$(3.6.26) \quad \text{sign}(f^*) = \begin{cases} +1, & \text{if } h > 0, \\ -1, & \text{if } h < 0. \end{cases}$$

If  $\text{sign}(f^*) = +1$  we say  $f$  *preserves orientation*, otherwise  $f$  *reverses orientation*. Given an oriented manifold  $M$ , an integrable  $n$ -form  $\alpha$  has a unique decomposition  $\alpha = \alpha^+ - \alpha^-$ , where  $\alpha^+$  and  $\alpha^-$  can be thought of as positive forms with respect to the orientation  $M^+$  determined by a volume form  $\Omega$ , and we write  $|\alpha| = \alpha^+ + \alpha^-$ . The decomposition is given by

$$\alpha^\pm = 1_{\pm d\alpha/d\Omega \geq 0} \alpha.$$

### 3.6.3 Curvature tensors and second fundamental forms

We now come to what is probably the most central of all concepts in Differential Topology, that of curvature. In essence, much of what we have done so far in developing the calculus of manifolds can be seen as no more than setting up the basic tools for handling the ideas to follow.

Curvature is the essence that makes manifolds inherently different from simple Euclidean space, where curvature is always zero. Since there are many very different manifolds, and many different Riemannian metrics, there are a number of ways to measure curvature. In particular, curvature can be measured in a somewhat richer fashion for manifolds embedded in ambient spaces of higher dimension than it can for manifolds for which no such embedding is given. A simple example of this is given in Figure 3.6.1, where you should think of the left hand circle  $S^1$  as being embedded in the plane while the right hand circle exists without any embedding. In the embedded case there are notions of ‘up’ and ‘down’, with the two arrows at the top and bottom of the circle pointing ‘up’. In one case the circle curves ‘away’ from the arrow, in the other, ‘towards’ it, so that any reasonable definition of curvature has to be different at the two points. However, for the non-embedded case, in which there is nothing external to the circle, the curvature must be the same everywhere. In what follows, we shall capture the first, richer, notion of curvature via the *second fundamental form* of the manifold, and the second via its *curvature tensor*.

We start with the (Riemannian) curvature tensor. While less informative than the second fundamental form, the fact that it is *intrinsic* (i.e. does not depend on an embedding) actually makes it a more central concept.

Much in the same way that the Lie bracket  $[X, Y] = XY - YX$  was a measure of the failure of partial derivatives to commute, the Riemannian

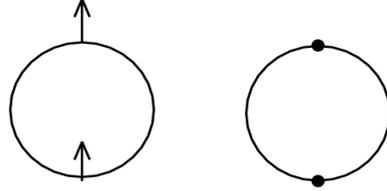


FIGURE 3.6.1. Embedded and intrinsic circles

curvature tensor,  $R$ , measures the failure of covariant derivatives to commute. A relatively simple computation shows that for vector fields  $X, Y$  it is not generally true that  $\nabla_X \nabla_Y - \nabla_Y \nabla_X = 0$ . However, rather than taking this difference as a measure of curvature, it is more convenient to define the (*Riemannian*) *curvature operator*<sup>35</sup>

$$(3.6.27) \quad R(X, Y) \triangleq \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X, Y]}.$$

The curvature operator is a multi-linear mapping from  $T(M)^{\otimes 2}$  to  $T(M)$ . Note that if  $[X, Y] = 0$ , as is the case when  $X_t$  and  $Y_t$  coordinate vectors in the natural basis of some chart, then  $R(X, Y) = \nabla_X \nabla_Y - \nabla_Y \nabla_X$ , and so is the first measure of lack of commutativity of  $\nabla_X$  and  $\nabla_Y$  mentioned above.

The (*Riemannian*) *curvature tensor*, also denoted by  $R$ , is defined by

$$(3.6.28) \quad \begin{aligned} R(X, Y, Z, W) &\triangleq g(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z, W) \\ &= g(R(X, Y)Z, W), \end{aligned}$$

where the  $R$  in the last line is, obviously, the curvature operator.

The definition (3.6.28) of  $R$  is not terribly illuminating, although one can read it as ‘the amount, in terms of  $g$  and in the direction  $W$ , by which  $\nabla_X$  and  $\nabla_Y$  fail to commute when applied to  $Z$ ’. To get a better idea of what is going on, we need the notion of planar sections.

For any  $t \in M$ , we call the span of two linearly independent vectors  $X_t, Y_t \in T_t M$  the *planar section* spanned by  $X_t$  and  $Y_t$ , and denote it by  $\pi(X_t, Y_t)$ . Such a planar section is determined by any pair of orthonormal vectors  $E_{1t}, E_{2t}$  in  $\pi(X_t, Y_t)$ , and we call

$$(3.6.29) \quad \kappa(\pi) \triangleq -R(E_{1t}, E_{2t}, E_{1t}, E_{2t})$$

the *sectional curvature* of the planar section. It is independent of the choice of basis. Sectional curvatures are somewhat easier to understand<sup>36</sup> than

<sup>35</sup>Note that the curvature operator depends on the underlying Riemannian metric  $g$  via the dependence of the connection on  $g$ .

<sup>36</sup>If  $\pi$  is a planar section at  $t$ , let  $M_t$  be an open, two-dimensional submanifold of  $M$  consisting of geodesic arcs through  $t$  and tangent at  $t$  to the section  $\pi$ . Then the sectional curvature of  $\pi$  is the Gaussian curvature of  $M_t$  at  $t$ . For a definition of Gaussian curvature, see page 141.

the curvature tensor, but essentially equivalent, since it is easy to check from the symmetry properties of the curvature tensor that it is uniquely determined by the sectional curvatures.

We shall later need a further representation of  $R$ , somewhat reminiscent of the representation (3.6.7) for the Riemannian connection. The way that  $R$  was defined in (3.6.28) it is clearly a covariant tensor of order 4. However, it is not a difficult computation, based on (3.6.7), to see that it can also be expressed as mixed tensor of order (2,2). In particular, if  $g$  is  $C^2$  and  $\{\theta_i\}_{1 \leq i \leq N}$  is the dual of a  $C^2$  orthonormal frame field, then  $R \in C^0(\Lambda^{2,2}(M))$  and can be written as

$$(3.6.30) \quad R = \frac{1}{2} \sum_{i,j=1}^N \Omega_{ij} \otimes (\theta_i \wedge \theta_j),$$

where the  $\Omega_{ij}$  are skew symmetric  $C^0$  differential 2-forms ( $\Omega_{ij} = -\Omega_{ji}$ ) known as the *curvature forms* for the section  $\{E_i\}_{1 \leq i \leq N}$  and are defined by

$$(3.6.31) \quad \Omega_{ij}(X, Y) = R(E_i, E_j, X, Y)$$

for vector fields  $X, Y$ .

This concludes the basics of what we shall need to know about the Riemannian curvature tensor. We shall get to the examples that interest us only in Sections 3.6.4 and 4.8 and for now turn to the notions of second fundamental forms and the shape operator.

For this, we take a Riemannian manifold  $(M, g)$  embedded<sup>37</sup> in an ambient Riemannian manifold  $(N, \hat{g})$  of co-dimension<sup>38</sup> at least one. We write  $\nabla$  for the connection on  $M$  and  $\hat{\nabla}$  for the connection on  $N$ . If  $t \in M$ , then the *normal space* to  $M$  in  $N$  at  $t$  is

$$(3.6.32) \quad T_t^\perp M \triangleq \{X_t \in T_t N : \hat{g}_t(X_t, Y_t) = 0 \text{ for all } Y_t \in T_t M\},$$

---

<sup>37</sup> We have used this term often already, albeit in a descriptive sense. The time has come to define it properly: Suppose  $f : M \rightarrow N$  is  $C^1$ . Take  $t \in M$  and charts  $(U, \phi)$  and  $(V, \psi)$  containing  $t$  and  $f(t)$ , respectively. The *rank* of  $f$  at  $t$  is defined to be the rank of the mapping  $\psi \circ f \circ \phi^{-1} : \phi(U) \rightarrow \psi(V)$  between Euclidean spaces. If  $f$  is everywhere of rank  $\dim M$  then it is called an *immersion*. If  $\dim M = \dim N$  then it is called a *submersion*. Note that this is a purely local property of  $M$ .

If, furthermore,  $f$  is a one-one homeomorphism of  $M$  onto its image  $f(M)$  (with its topology as a subset of  $N$ ) then we call  $f$  an *embedding* of  $M$  in  $N$  and refer to  $M$  as an *embedded (sub-)manifold* and to  $N$  as the *ambient manifold*. This is a global property, and amounts to the fact that  $M$  cannot ‘intersect’ itself on  $N$ .

Finally, let  $M$  and  $N$  be Riemannian manifolds with metrics  $g$  and  $\hat{g}$ , respectively. Then we say that  $(M, g)$  is a *embedded Riemannian manifold* of  $(N, \hat{g})$  if, in addition to the above,  $g = f^* \hat{g}$ , where  $f^* \hat{g}$  is the pull-back of  $\hat{g}$ . (cf. (3.6.14))

<sup>38</sup>The codimension of  $M$  and  $N$  is  $\dim(N) - \dim(M)$ .

and we also write  $T^\perp(M) = \bigcup_t T_t^\perp M$ . Note that since  $T_t N = T_t M \oplus T_t^\perp M$  for each  $t \in M$ , it makes sense to talk about tangential and normal components of an element of  $T_t N$  and so of the orthogonal projections  $P_{TM}$  of and  $P_{TM}^\perp$  of  $T(N)$  to  $T(M)$  and  $T^\perp(M)$ .

The *second fundamental form* of  $M$  in  $N$  can now be defined to be the operator  $S$  from  $T(M) \times T(M)$  to  $T^\perp(M)$  satisfying

$$(3.6.33) \quad S(X, Y) \triangleq P_{TM}^\perp \left( \widehat{\nabla}_Y \nu \right) = \widehat{\nabla}_X Y - \nabla_X Y.$$

Let  $\nu$  denote a unit normal vector field on  $M$ , so that  $\nu_t \in T_t^\perp M$  for all  $t \in M$ . Then the *scalar second fundamental form* of  $M$  in  $N$  for  $\nu$  is defined, for  $X, Y \in T(M)$ , by

$$(3.6.34) \quad S_\nu(X, Y) \triangleq \widehat{g}(S(X, Y), \nu),$$

where the internal  $S$  on the right hand side refers to the second fundamental form (3.6.33). Note that, despite its name, the scalar fundamental form is not a differential form, since it is symmetric (rather than alternating) in its arguments. When there is no possibility of confusion, we shall drop the adjective scalar, and refer also to  $S_\nu$  as the second fundamental form.

In view of the fact that  $\widehat{\nabla}$  is torsion free (cf. (3.6.4)) we also have that, for  $X, Y \in T(M)$ ,

$$(3.6.35) \quad S_\nu(X, Y) = \widehat{g}(\widehat{\nabla}_X Y, \nu) = -\widehat{g}(\widehat{\nabla}_Y X, \nu).$$

As we already noted,  $S_\nu$  is a symmetric 2 tensor, so that, as for the curvature tensor, we can view it as a symmetric section of  $\bigwedge^{**}(M)$ . As such, it contains a lot of information about the embedding of  $M$  in  $N$  and thus curvature information about  $M$  itself. For example, fix  $\nu$  and use the second fundamental form to define an operator  $\mathcal{S}_\nu : T(M) \rightarrow T(M)$  by  $\widehat{g}(\mathcal{S}_\nu(X), Y) = S_\nu(X, Y)$  for all  $Y \in T(M)$ . Then  $\mathcal{S}_\nu$  is known as the *shape operator*. It has  $N$  real eigenvalues, known as the *principal curvatures* of  $M$  in the direction  $\nu$ , and the corresponding eigenvalues are known as the *principal curvature vectors*.

All of the above becomes quite familiar and particularly useful if  $M$  is a simple surface determined by a mapping  $f$  from  $\mathbb{R}^N$  to the ambient manifold  $\mathbb{R}^{N+1}$  with the usual Euclidean metric. In this case, the principle curvatures are simply the eigenvalues of the Hessian  $(\partial^2 f / \partial x_i \partial x_j)_{i,j=1}^N$ . (cf. (3.9.4) below.) In particular, if  $M$  is a surface in  $\mathbb{R}^3$ , then the product of these eigenvalues is known as *Gaussian curvature* and is an intrinsic quantity of  $M$ : i.e. it is independent of the embedding in  $\mathbb{R}^3$ . In general, integrals<sup>39</sup> of such quantities over  $M$  are what yield intrinsic characteristics of the manifold<sup>40</sup>.

<sup>39</sup>See, for example Section 3.8 dealing with Lipschitz-Killing curvatures.

<sup>40</sup>In higher dimensions, one way to think of fundamental forms is as follows: If  $X_t$  is a unit tangent vector at  $t$ , then  $S(X_t, X_t)$  is the acceleration vector (in the ambient

Finally, we note one more important formula, which relates to the comments we made above about Riemannian curvatures being linked to second fundamental forms. With  $R^M$  and  $R^{\partial M}$  denoting, respectively, the curvature tensors on  $M$  and  $\partial M$ , and with the the second fundamental form defined by (3.6.35), we have the following simplified version of the Gauss formula<sup>41</sup>

$$(3.6.36) \quad S^2((X, Y), (Z, W)) = -2(R^{\partial M}(X, Y, Z, W) - R^M(X, Y, Z, W)).$$

### 3.6.4 A Euclidean example

The time has probably come to give a ‘concrete’ example, for which we take a compact  $C^2$  domain  $T$  in  $\mathbb{R}^N$  with a Riemannian metric  $g$ . We shall show how to explicitly compute both the curvature tensor  $R$  and the second fundamental form  $S$ , as well as traces of their powers, something that we shall need later.

We start with  $\{E_i\}_{1 \leq i \leq N}$ , the standard<sup>42</sup> coordinate vector fields on  $\mathbb{R}^N$ . This also gives the natural basis in the global chart  $(\mathbb{R}^N, i)$ , where  $i$  is the inclusion map. We now define<sup>43</sup> the so-called *Christoffel symbols of the first kind*,

$$(3.6.37) \quad \Gamma_{ijk} \triangleq g(\nabla_{E_i} E_j, E_k), \quad 1 \leq i, j, k \leq N.$$

We also need the functions

$$(3.6.38) \quad g_{ij} = g(E_i, E_j).$$

Despite the possibility of some confusion, we also denote the corresponding matrix function by  $g$ , doubling up on the notation for the metric.

space) for the geodesic starting at  $t$  with direction  $X_t$ . If, in addition,  $\nu_t$  is a unit normal vector at  $t$ , then  $S_{\nu_t}(X_t, X_t)$  has an interpretation closely related to usual curvature in the planar section spanned by  $X_t$  and  $\nu_t$ .

<sup>41</sup>The Gauss formula is a general result linking curvature tensors with second fundamental forms, given an ambient space. It can be expressed in a number of ways. Equation (3.6.36) is one of these and will be useful for our purposes.

<sup>42</sup>Note that while the  $E_i$  might be ‘standard’ there is no reason why they should be the ‘right’ coordinate system to use for a given  $g$ . In particular, they are not orthonormal any more, since  $g_{ij}$  of (3.6.38) need not be a Kronecker delta. Thus, although we start here, we shall soon leave this choice of basis for an orthonormal one.

<sup>43</sup>An alternative, and somewhat more motivated, definition of the  $\Gamma_{ijk}$  comes by taking the vector fields  $\{E_i\}$  to be orthonormal with respect to the metric  $g$ . In that case, they can be defined via their rôle in determining the Riemannian connection through the set of  $N^2$  equations  $\nabla_{E_i} E_j = \sum_{k=1}^N \Gamma_{ijk} E_k$ . Taking this as a definition, it is easy to see that (3.6.37) must also hold. In general, the Christoffel symbols are dependent on the choice of basis.

With this notation it now follows via a number of successive applications of (3.6.4) and (3.6.5) that

$$(3.6.39) \quad \Gamma_{ijk} = (E_j g_{ik} - E_k g_{ij} + E_i g_{jk}) / 2.$$

We need two more pieces of notation, the elements  $g^{ij}$  of the inverse matrix  $g^{-1}$  and the *Christoffel symbols of the second kind*, defined by

$$\Gamma_{ij}^k = \sum_{s=1}^n g^{ks} \Gamma_{ijs}.$$

It is an easy and standard exercise to show that, for any  $C^2$  Riemannian metric  $g$  on  $\mathbb{R}^N$  with curvature tensor  $R$ ,

$$(3.6.40) \quad \begin{aligned} R_{ijkl}^E &\triangleq R((E_i, E_j), (E_k, E_l)) \\ &= \sum_{s=1}^N [g_{sl} (E_i(\Gamma_{jk}^s) - E_j(\Gamma_{ik}^s)) + \Gamma_{isl} \Gamma_{jk}^s - \Gamma_{jst} \Gamma_{ik}^s] \\ &= E_i \Gamma_{jkl} - E_j \Gamma_{ikl} + \sum_{s,t=1}^N (\Gamma_{iks} g^{st} \Gamma_{jlt} - \Gamma_{jks} g^{st} \Gamma_{ilt}). \end{aligned}$$

Returning to the definition of the curvature tensor, and writing  $\{de_i\}_{1 \leq i \leq N}$  for the dual basis of  $\{E_i\}_{1 \leq i \leq N}$ , it now follows (after some algebra) that

$$(3.6.41) \quad R = \frac{1}{4} \sum_{i,j,k,l=1}^n R_{ijkl}^E (de_i \wedge de_j) \otimes (de_k \wedge de_l).$$

The next step is to develop a formula for the curvature tensor based on an arbitrary vector field and not just the  $E_i$ . To this end, let  $X = \{X_i\}_{1 \leq i \leq N}$  be a measurable section of  $\mathcal{O}(T)$ , having dual frames  $\{\theta_i\}_{1 \leq i \leq N}$ , so that

$$(3.6.42) \quad \theta_i = \sum_{i'=1}^n g_{ii'}^{\frac{1}{2}} de_{i'},$$

where  $g^{\frac{1}{2}}$  is given by

$$(g^{\frac{1}{2}})_{ij} = g(E_i, X_j)$$

and the notation comes from the easily verified fact that  $g^{\frac{1}{2}}(g^{\frac{1}{2}})' = g$ , so that  $g^{\frac{1}{2}}$  is a (measurable) square root of  $g$ .

It follows that

$$(3.6.43) \quad R = \frac{1}{4} \sum_{i,j,k,l=1}^n R_{ijkl}^X (\theta_i \wedge \theta_j) \otimes (\theta_k \wedge \theta_l),$$

where

$$\begin{aligned} R_{ijkl}^X &= \sum_{i',j',k',l'=1}^n R_{i'j'k'l'}^E g_{ii'}^{-\frac{1}{2}} g_{jj'}^{-\frac{1}{2}} g_{kk'}^{-\frac{1}{2}} g_{ll'}^{-\frac{1}{2}} \\ &= R((X_i, X_j), (X_k, X_l)) \end{aligned}$$

and you are free to interpret the  $g_{ij}^{-\frac{1}{2}}$  as either the elements of  $(g^{\frac{1}{2}})^{-1}$  or of a square root of  $g^{-1}$ .

In (3.6.43) we now have a quite computable representation of the curvature tensor for any orthonormal basis. Given this, we also have the curvature forms  $\Omega_{ij}(X, Y) = R(E_i, E_j, X, Y)$  of (3.6.30) and so, via (3.6.31), we can rewrite the curvature tensor as

$$R = \frac{1}{2} \Omega_{ij} \otimes (\theta_i \wedge \theta_j).$$

With the product and general notation of (3.5.10) we can thus write  $R^k$  as

$$R^k = \frac{1}{2^k} \sum_{i_1, \dots, i_{2k}=1}^N \left( \bigwedge_{l=1}^k \Omega_{i_{2l-1}i_{2l}} \right) \otimes \left( \bigwedge_{l=1}^k (\theta_{i_{2l-1}} \wedge \theta_{i_{2l}}) \right),$$

where

$$\begin{aligned} \bigwedge_{l=1}^k \Omega_{i_{2l-1}i_{2l}}(X_{a_1}, \dots, X_{a_{2k}}) &= \frac{1}{2^k} \sum_{\sigma \in S(2k)} \varepsilon_\sigma \prod_{l=1}^k \Omega_{i_{2l-1}i_{2l}}(X_{a_{\sigma(2l-1)}}, X_{a_{\sigma(2l)}}) \\ &= \frac{1}{2^k} \sum_{\sigma \in S(2k)} \varepsilon_\sigma \prod_{l=1}^k R_{i_{2l-1}i_{2l}a_{\sigma(2l-1)}a_{\sigma(2l)}}^X. \end{aligned}$$

It follows that

$$\begin{aligned} R^k((X_{a_1}, \dots, X_{a_{2k}}), (X_{a_1}, \dots, X_{a_{2k}})) &= \frac{1}{2^{2k}} \sum_{i_1, \dots, i_{2k}=1}^n \delta_{(i_1, \dots, i_{2k})}^{(a_1, \dots, a_{2k})} \left( \sum_{\sigma \in S(2k)} \varepsilon_\sigma \prod_{l=1}^k R_{i_{2l-1}i_{2l}a_{\sigma(2l-1)}a_{\sigma(2l)}}^X \right), \end{aligned}$$

where, for all  $m$ ,

$$\delta_{(b_1, \dots, b_m)}^{(c_1, \dots, c_m)} = \begin{cases} \varepsilon_\sigma & \text{if } c = \sigma(b), \text{ for some } \sigma \in S(m) \\ 0 & \text{otherwise.} \end{cases}$$

We are finally in a position to write down an expression for the trace of  $R^k$ , as defined by (3.5.12):

$$(3.6.44) \quad \begin{aligned} \text{Tr}(R^k) &= \frac{1}{(2k)!} \sum_{a_1, \dots, a_{2k}=1}^n R^k((X_{a_1}, \dots, X_{a_{2k}}), (X_{a_1}, \dots, X_{a_{2k}})) \\ &= \frac{1}{2^{2k}} \sum_{a_1, \dots, a_{2k}=1}^n \left( \sum_{\sigma \in S(2k)} \varepsilon_\sigma \prod_{l=1}^k R_{a_{2l-1} a_{2l} a_{\sigma(2l-1)} a_{\sigma(2l)}}^X \right). \end{aligned}$$

This is the equation we have been searching for to give a ‘concrete’ example of the general theory.

A little thought will show that while the above was presented as an example of a computation on  $\mathbb{R}^N$  it is, in fact, far more general. Indeed, you can reread the above, replacing  $\mathbb{R}^N$  by a general manifold and the  $E_i$  by a family of local coordinate systems that are in some sense ‘natural’ for computations. Then (3.6.44) still holds, as do all the equations leading up to it. Thus the title of this Section is somewhat of a misnomer, since the computations actually have nothing to do with Euclidean spaces!

Unfortunately, however, we are not yet quite done. While (3.6.44) handles the trace of powers of the curvature tensor, it would also be nice – and very important for what is to follow in Chapter 4 – to also know something about the second fundamental form  $S$  and, ultimately, about the trace of mixed powers of the form  $R^k S^j$ , ( $j \leq N - 2k$ ) on  $\partial T$ . Therefore, we now turn to this problem.

We start much as we did for the curvature tensor, by choosing a convenient set of bases. However, this time the ‘natural’ Euclidean basis is no longer natural, since our primary task is to parameterise the surface  $\partial T$  in a convenient fashion.

Thus, this time we start with  $\{E_i^*\}_{1 \leq i \leq N-1}$ , the natural basis determined by some atlas on  $\partial T$ . This generates  $T(\partial T)$ . It is then straightforward to enlarge this to a section  $E^* = \{E_i^*\}_{1 \leq i \leq N}$  of  $S(T)$ , the sphere bundle of  $T$ , in such a way that, on  $\partial T$ ,  $E_N^* = \nu$ , the inward pointing unit normal vector field on  $\partial T$ .

It then follows from the definition of the second fundamental form (cf. (3.6.34)) that

$$(3.6.45) \quad S = - \sum_{i=1}^{N-1} \alpha_{iN} \otimes de_i^*$$

where, by (3.6.35), the connection forms  $\alpha_{ij}$  (on  $\mathbb{R}^N$ ) satisfy

$$(3.6.46) \quad \alpha_{ij}(Y) = g(\nabla_Y E_i^*, E_j^*).$$

If we now define  $\Gamma_{ijk}^* = g(\nabla_{E_i^*} E_j^*, E_k^*)$ , then the connection forms  $\alpha_{iN}$  can be expressed<sup>44</sup> as

$$(3.6.47) \quad \alpha_{iN} = \sum_{j=1}^{N-1} \Gamma_{jiN}^* de_j^*.$$

If, as for the curvature tensor, we now choose a smooth section  $X$  of  $\mathcal{O}(T)$  with dual frames  $\theta$ , such that, on  $\partial T$ ,  $X_N = \nu$ , similar calculations yield that

$$S = \sum_{i,j=1}^{N-1} S_{ij}^X \theta_i \otimes \theta_j = \sum_{i=1}^{N-1} \theta_{iN} \otimes \theta_i,$$

where

$$S_{ij}^X \triangleq \sum_{i',j'=1}^{N-1} g_{ii'}^{-\frac{1}{2}} g_{jj'}^{-\frac{1}{2}} \Gamma_{j'i'N}, \quad \theta_{iN} \triangleq \sum_{i=1}^{N-1} S_{ij}^X \theta_j.$$

Finally, on setting  $p = 2k + j$ , it follows that

$$(3.6.48) \quad R^k S^j = \frac{1}{2^k} \sum_{a_1, \dots, a_p=1}^{N-1} \left( \bigwedge_{l=1}^k \Omega_{a_{2l-1}, a_{2l}} \right) \wedge \left( \bigwedge_{m=1}^j \theta_{a_{2k+m}N} \right) \otimes \left( \bigwedge_{l=1}^p \theta_{a_l} \right),$$

a formula we shall need in Chapter 4.

### 3.7 Piecewise smooth manifolds

So far, all that we have had to say about manifolds and calculus on manifolds has been of a local nature; i.e. it depended only on what was happening

---

<sup>44</sup> It is often possible to write things in a format that is computationally more convenient. In particular, if the metric is Euclidean and if it is possible to explicitly determine functions  $a_{ij}$  so that

$$E_{it}^* = \sum_{k=1}^N a_{ik}(t) E_{kt},$$

then it follows trivially from the definition of the  $\Gamma_{jiN}^*$  that

$$\Gamma_{jiN}^*(t) = \sum_{k,l,m=1}^N a_{jk}(t) \frac{\partial}{\partial t_k} (a_{Nl}(t) a_{im}(t) g_{ml}(t)).$$

in individual charts. However, looking back at what we did in Sections 3.2–3.4 in the setting of Integral Geometry, this is not going to solve our main problem, which is understanding the global structure of excursion sets of random fields now defined over manifolds.

One of issues that will cause us a heavy investment in notation will be the need to study what we shall call *piecewise smooth* manifolds, which primarily arise due to the fact that we want to develop a set of results that is not only elegant, but also useful. To understand this point, two simple examples will suffice, the sphere  $S^2$ , which is a  $C^\infty$  manifold without boundary, and the unit cube  $I^3$ , a flat manifold with a boundary that comprises of six faces which intersect at twelve edges, themselves intersecting at eight vertices. The cube, faces, edges and vertices are themselves flat  $C^\infty$  manifolds, of dimensions 3,2,1 and 0, respectively.

In the first case, if  $f \in C^k(S^2)$ , the excursion set  $A_u(S^2, f)$  is made of smooth subsets of  $S^2$ , each one bounded by a  $C^k$  curve. In the second case, for  $f \in C^k(I^3)$ , while the individual components of  $A_u(I^3, f)$  will have a  $C^k$  boundary away from  $\partial I^3$ , their boundaries will also have faces, edges and vertices where they intersect with  $\partial I^3$ . We already know from Section 3.3 that when we attempt to find point set representations for the Euler characteristics of excursion sets these boundary intersections are important. (e.g. (3.3.16) for the case of  $I^2$ .) This is even the case if the boundary of the parameter set is itself smooth. (e.g. Theorem 3.3.5) Consequently, as soon as we permit as parameter spaces manifolds with boundaries, we are going to require techniques to understand these boundaries and their intersections with excursion sets. The current Section will develop what we need in the setting of piecewise smooth manifolds<sup>45</sup>. If you are interested only in parameter spaces which are manifolds without a boundary then you can go directly to the following Section. However, you will then have to forgo fully understanding how to handle excursion sets over parameter spaces as simple as cubes, which, from the point of view of applications, is a rather significant loss.

To construct the objects of interest to us, we shall proceed in three stages. In Section 3.7.1 we start by constructing triangles of general dimension ('simplices') and show how to glue them together to make more general objects. We then smoothly perturb them to get 'piecewise smooth spaces'. In Section 3.7.2 we show how to give these spaces a differential structure, complete with with all the calculus that, to this point, we have developed only for smooth manifolds.

---

<sup>45</sup>An alternative approach is via 'sets of finite reach' as developed in [34, 35] and its generalisations as in [115]. While this gives an approach both powerful and mathematically elegant, it involves a heavier background investment than is justified for our purposes.

### 3.7.1 Piecewise smooth spaces

Soon we shall get to triangles but, as usual, we start with something more abstract: An *(abstract) simplicial complex*  $\mathcal{K}$  on  $\mathbb{N}$  is a collection of subsets of  $\mathbb{N}$  such that for  $A, B \subseteq \mathbb{N}$ ,

$$(3.7.1) \quad A \subseteq B \text{ and } B \in \mathcal{K} \Rightarrow A \in \mathcal{K}.$$

The *j-skeleton*  $\Sigma^j$  is the subcomplex of  $\mathcal{K}$  defined by

$$(3.7.2) \quad \Sigma^j = \{A \in \mathcal{K} : \#A \leq j + 1\},$$

where, for any set  $A$ ,  $\#A$  denotes the number of its elements. The one point subsets, which make up the 0-skeleton, are called the *vertices* of  $\mathcal{K}$ . The *link* of  $A$  in  $\mathcal{K}$  is the subcomplex defined by

$$L(A, \mathcal{K}) = \{B \setminus A : B \in \mathcal{K}, B \supseteq A\}.$$

The complex  $\mathcal{K}$  is called *locally finite* if  $\#L(A, \mathcal{K}) < \infty$  for all  $A \in \mathcal{K}$ .

An abstract simplicial complex  $\mathcal{K}$  is uniquely determined by the set of its *maximal elements*, where ‘maximal’ refers to usual set inclusion; i.e.  $A' \in \mathcal{K}$  is maximal if and only if  $A'$  is not a proper subset of any other  $A \in \mathcal{K}$ . If  $\mathcal{K}$  has a unique maximal element then it is called an (abstract) simplex and, if, furthermore,  $\#\mathcal{K} < \infty$ , then  $\mathcal{K}$  is isomorphic to the power set  $2^{\{1, 2, \dots, \#\mathcal{K}\}}$ , the set of all subsets of  $\{1, \dots, \#\mathcal{K}\}$ <sup>46</sup>.

Given an abstract simplicial complex, we can realise it as a concrete physical object. To see how to do this, first assume that  $\mathcal{K}$  is finite with  $k$  vertices and that we are given a linearly independent set  $B = \{v_1, \dots, v_k\} \subset \mathbb{R}^k$ . Then the so-called *geometric simplicial complex*  $\mathcal{K}_B$  associated to  $\mathcal{K}$  is the set

$$(3.7.3) \quad \mathcal{K}_B = \left\{ v \in \mathbb{R}^k : v = \sum_{j \in A} a_j v_j, a_j > 0, \right. \\ \left. \sum_{j \in A} a_j = 1, \text{ for some } A \in \mathcal{K} \right\}.$$

This construction is attractive, since realisations of 2-point sets are straight lines, while 3-point sets give equilateral triangles, 4-point sets give tetrahedra, etc. If  $\mathcal{K} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$  then its realisation is as in Figure 3.7.1 in which we take  $v_j = e_j$  in  $\mathbb{R}^3$ .

However, the set  $B$  need not be linearly independent<sup>47</sup>, which is rather convenient when we want to draw pictures on a two-dimensional page. It

<sup>46</sup>As an aside, we note that while the following discussion will assume  $\#\mathcal{K} < \infty$ , everything is also well defined under the assumption of only local finiteness.

<sup>47</sup>Nor, indeed, need it be a subset of  $\mathbb{R}^k$ . A  $k$ -dimensional subspace of a Hilbert space will do, in which case the inner products and norms appearing below should be interpreted in the Hilbert sense and references to Lebesgue measure should be replaced by Hausdorff measure. In fact, we shall soon move to this setting.

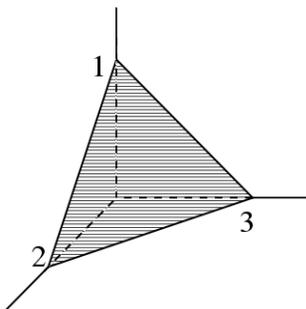


FIGURE 3.7.1. A geometric simplicial complex of 3 points, over the standard orthonormal basis of  $\mathbb{R}^3$ .

is enough to require for every  $A \in \mathcal{K}$ , the set

$$\{A\} \triangleq \left\{ v \in \mathbb{R}^k : \sum_{j \in A} a_j v_j, a_j > 0, \sum_{j \in A} a_j = 1 \right\}$$

has non-zero  $(\#A - 1)$ -dimensional Lebesgue measure; i.e. it is not a proper subset of any  $(\#A - 2)$ -plane in  $\mathbb{R}^k$ . We call sets  $\{A\} \subset \mathbb{R}^k$  of this kind the *face* of  $A$  in  $\mathcal{K}_B$ . Note that a face  $\{A\}$ , when restricted to the  $(\#A)$ -dimensional hyperplane containing it, is an open set. This is a trivial consequence of the strict positivity of the  $a_j$ . Figure 3.7.2 gives an example of a geometric realisation of the complex

$$\begin{aligned} \mathcal{K} = \{ & \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \\ & \{4, 6\}, \{5, 6\}, \{1, 2, 3\}, \{4, 5, 6\} \}. \end{aligned}$$

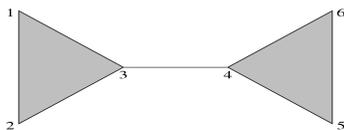


FIGURE 3.7.2. A geometric simplicial complex of 6 points.

An alternative expression for a geometric simplicial complex to that given by (3.7.3) and which we shall use later is as a union of faces, *viz.*

$$\mathcal{K}_B = \bigcup_{A \in \mathcal{K}} \left\{ \sum_{j \in A} a_j v_j : a_j > 0, \sum_{j \in A} a_j = 1 \right\} = \bigcup_{A \in \mathcal{K}} \{A\}.$$

Note that the union is disjoint since the individual faces are all open as subsets of the space they span.

Revert now to the usual orthonormal Euclidean basis. Then, looking at the above representation of  $\mathcal{K}_B$ , it is easy to see that it is a basic complex, in the Integral Geometric sense of Section 3.2, since each simplex is basic with respect to the usual Euclidean basis. As such,  $\mathcal{K}_B$  also has a well defined Euler characteristic and, in view of (3.2.3), it is not hard to see that it is given by

$$(3.7.4) \quad \varphi(\mathcal{K}_B) = \sum_{k=0}^{N-1} (-1)^k (\text{Number simplices in } \mathcal{K}_B \text{ with } (N-k) \text{ vertices}).$$

Our next step is to set up a metric on simplicial complexes. As subsets of  $\mathbb{R}^k$  they inherit the usual Euclidean metric. Under this metric, the distance between the vertices 1 and 6 in Figure 3.7.2 is less than that between 1 and 5. A more natural metric, which relies only on paths within the simplicial complex, would have these two distances equal. Such a metric is given by the ‘geodesic’ metric,

$$d_{\mathcal{K}_B}(x, y) \triangleq \inf \{L(c) : c \in C([0, 1], \mathcal{K}_B), c(0) = x, c(1) = y\},$$

where  $L(c)$  is the usual arc length of a curve in  $\mathbb{R}^k$ . Since every minimizing path will locally be a straight line, (i.e. when restricted to any of the faces  $\{A\}$ ) a little thought shows that this metric depends only on the Euclidean edge lengths of  $\mathcal{K}_B$ , i.e. on the distances

$$\{|v_i - v_j| : \{i, j\} \in \mathcal{K}\}.$$

This observation leads to an equivalence relation on the collection of all geometric simplicial complexes by setting  $\mathcal{K}'_{B'} \equiv \mathcal{K}_B$  if and only if there exists a bijection  $f : \mathbb{N} \rightarrow \mathbb{N}$  such that  $f(\mathcal{K}') = \mathcal{K}$  and

$$(3.7.5) \quad |v_i - v_j| = |v'_{f^{-1}(i)} - v'_{f^{-1}(j)}|$$

for all  $\{i, j\} \in \mathcal{K}$ , where  $B' = \{v'_1, \dots, v'_k\}$  is set of vectors corresponding to  $\mathcal{K}'_{B'}$ . The set of equivalence classes under this relation are referred to as *piecewise flat spaces*. Note that if we now take (3.7.4) as the *definition* of an Euler characteristic, for any choice of  $B$ , then  $\varphi$  will be the same for all members of a given equivalence class.

As we noted earlier, the above construction does not rely on the fact the vectors of  $B$  lie in  $\mathbb{R}^k$ , as any Hilbert space would do, in which case we simply interpret the norms in (3.7.5) as Hilbert space, rather than Euclidean, norms. Thus we now work in this generality, denoting the Hilbert space by  $H$ <sup>48</sup>.

---

<sup>48</sup> This will be important for us, since piecewise flat spaces will arise naturally in the applications in later chapters where they appear as discretizations of the parameter

Note that since distinct faces are disjoint, for any given  $t \in \mathcal{K}_B$ , there is a unique  $A \in \mathcal{K}$  for which  $t \in \{A\}$ . Also, if  $A \subseteq A'$ , then  $t \in \overline{\{A'\}}$ . Thus, to each  $A' \supseteq A$  we can associate a cone in the tangent space<sup>49</sup>  $T_t H$  given by

$$\mathcal{S}_t(A, A') \triangleq \left\{ v \in T_t H : \text{for } \varepsilon > 0 \text{ sufficiently small, } t + \varepsilon v \in \overline{\{A'\}} \right\},$$

called the *support cone* of  $A$  in  $A'$  at  $t \in \{A\}$ . We can enlarge the support cone to its supporting subspace in  $T_t H$ , denoted by

$$[A']_t = \text{span}\{\mathcal{S}_t(A, A')\}.$$

Denoting the unit sphere in  $[A']_t$  by  $S([A']_t)$ , we call the intersection  $\mathcal{S}_t(A, A') \cap S([A']_t)$  the *solid angle* of  $\mathcal{S}_t(A, A')$  in  $S([A']_t)$ . Furthermore, writing  $\mathcal{H}_k$  for  $k$ -dimensional Hausdorff measure on  $H$ , we compute the normalized  $(\#A' - 1)$ -dimensional Hausdorff volume of this angle as

$$(3.7.6) \quad \sigma(A, A') \triangleq \frac{\mathcal{H}_{\#A'-1}(\mathcal{S}_t(A, A') \cap S([A']_t))}{\mathcal{H}_{\#A'-1}(S([A']_t))}.$$

Note that for given  $A, A'$  and  $s, t \in A$ , the cones  $\mathcal{S}_s(A, A') \in T_s H$  and  $\mathcal{S}_t(A, A') \in T_t H$  are affine translations of one another. Thus, the right hand side of (3.7.6) is actually independent of  $t$ , and a function only of  $A, A'$  and the geometric realisation  $\mathcal{K}_B$ . Consequently, it was not an oversight that  $t$  does not appear on the left hand side. The ratio  $\sigma(A, A')$  is called the *internal angle* of  $A$  in  $A'$ .

Similarly, by taking the *dual* (or *normal*) *cone* of  $\mathcal{S}_t(A, A')$  defined by

$$(3.7.7) \quad N(\mathcal{S}_t(A, A')) \triangleq \{v \in [A']_t : \langle v, y \rangle \leq 0 \text{ for all } y \in \mathcal{S}_t(A, A')\}$$

and computing its normalized volume, we obtain  $\sigma^*(A, A')$  the *external angle* of  $A$  in  $A'$ . Both the support cone  $\mathcal{S}_t(A, A')$  and its dual are *simplicial* or *finitely generated* cones; *viz.* they are cones of the form

$$K = \left\{ v \in \tilde{H} : v = \sum_{i=1}^l c_i \tilde{v}_i, c_i \geq 0 \right\}$$

for some Hilbert space  $\tilde{H}$  and some linearly independent set  $\tilde{B} = \{v_1, \dots, v_l\}$ . Note that simplicial cones are necessarily convex.

spaces of random fields. In that scenario, natural Hilbert spaces will be the  $L^2$  and reproducing kernel Hilbert spaces of the fields. (cf. Section 2.5)

<sup>49</sup>Note that a Hilbert space  $H$  has a natural manifold structure given by mapping its basis elements to the natural Euclidean basis and pushing forward the Euclidean tangent spaces to give tangent spaces for  $H$ . This is obvious if  $H$  is finite dimensional, but also true if  $H$  is infinite dimensional. In any case, since both  $A$  and  $A'$  are finite dimensional, we can always restrict the  $v$  in the definition of  $\mathcal{S}_t(A, A')$  to a finite dimensional subset of  $H$  and so remain in the domain of objects that we have fully defined.

The time has come for some examples. In Figure 3.7.3 we take  $A' = \{1, 2, 3\}$  and, moving from (a) to (c), take  $t \equiv A$  to be the vertex  $\{1\}$ ,  $t$  a point on the edge  $A = \{1, 3\}$ , or  $t$  a point in the open triangle  $A = \{1, 2, 3\}$ . In each of these cases the support cone (truncated for the sake of the diagram) is the shaded area. In (a), the (truncated) normal cone is the bricked-in region, the sides of which are perpendicular to the sides of the triangle. In (b), the normal cone is the line emanating from  $t$  and perpendicular to the edge, while in (c) the normal cone is empty. Assuming an equilateral triangle for the geometric simplex, the corresponding internal angles are  $60^\circ$ ,  $180^\circ$  and  $360^\circ$ , while the external angles are  $120^\circ$ ,  $0^\circ$  for (a) and (b), and undefined for (c) since the normal cone is empty.

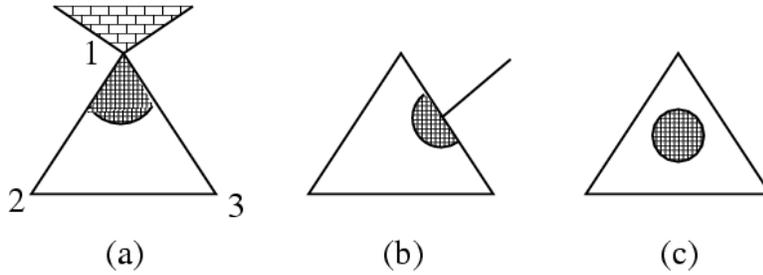


FIGURE 3.7.3. Support and normal cones for a triangle.

A more complicated example is given in Figure 3.7.4 for the tetrahedron based on four points. In this case we have shown the (truncated) normal cones for (a) the vertex  $\{2\}$ , (b) a point on the edge  $\{2, 4\}$ , and (c) a point on the face  $\{2, 3, 4\}$ . What remains is a point in the interior of the tetrahedron, for which the normal cone is empty.

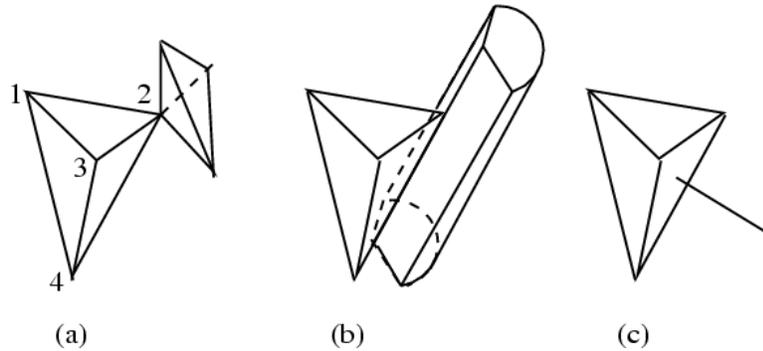


FIGURE 3.7.4. Normal cones for a tetrahedron.

We now return to  $L(A, \mathcal{K})$ , the link of  $A$  in  $\mathcal{K}$ . We modify this simplicial complex slightly, by first adjoining another singleton, representing  $A$ , to each maximal simplex in  $L(A, \mathcal{K})$ , and then adding any required additional

sets so that (3.7.1) is satisfied. We shall give an example in a moment. Referring to this new complex as  $\tilde{L}(A, \mathcal{K})$ , we can realise  $\tilde{L}(A, \mathcal{K})$  as a geometric subcomplex  $\tilde{L}(A, \mathcal{K})_B$  of the standard simplex in  $\mathbb{R}^{\#\Sigma^1(L(A, \mathcal{K}))}$  (cf. (3.7.2)) by taking  $B = \{0, e_1, \dots, e_{\#\Sigma^1(L(A, \mathcal{K}))}\}$  where 0 represents the singleton adjoined to each maximal simplex. Furthermore, each point  $t \in \{A\}$  has a neighbourhood  $N_t$ , homeomorphic to

$$\tilde{V} = \tilde{N}_1 \times \tilde{N}_2,$$

where  $\tilde{N}_1$  is a neighbourhood of the origin in  $\mathbb{R}^{\#A-1}$ ,  $\tilde{N}_2$  is a neighbourhood of the origin in  $\tilde{L}(A, \mathcal{K})_B$  and the topology is the one inherited as a subset of  $\mathbb{R}^{\#A-1} \times \mathbb{R}^{\#\Sigma^1(L(A, \mathcal{K}))}$ .

For a concrete example of the above procedure, consider the case when  $\mathcal{K}_B$  is the boundary of a tetrahedron, labelled as in Figure 3.7.4, and all geometric representations are in Euclidean space. Choosing one of the edges, say  $A = \{1, 2\}$  we have that

$$L(A, \mathcal{K}) = \{\{3\}, \{4\}, \{3, 4\}\}.$$

In this case, the maximal simplex is  $\{3, 4\}$ . If we adjoin a singleton, denoted by  $\{0\}$ , to the above we have the modified complex

$$\tilde{L}(A, \mathcal{K}) = \{\{0\}, \{3\}, \{4\}, \{0, 3\}, \{0, 4\}, \{0, 3, 4\}\}.$$

The geometric realisation of  $\tilde{L}(A, \mathcal{K})$  in  $\mathbb{R}^2$  is then simply a triangle, open on the side opposite 0 and closed on the other two sides. Consequently, every point in the edge  $\{A\}$  of the boundary of the tetrahedron is easily seen to have a neighbourhood homeomorphic to

$$(-\varepsilon, \varepsilon) \times \tilde{N}_2,$$

where  $\tilde{N}_2$  is some neighbourhood of the origin in the triangle.

Of course,  $\tilde{L}(A, \mathcal{K})$  need not be realised as a subset of the standard simplex, i.e. we can choose  $B$  to be any linearly independent set in any Hilbert space  $H$ , and the same statement concerning the neighbourhood of  $t \in \{A\}$  holds, i.e. if  $\tilde{v}_0$  is the singleton adjoined to each maximal simplex, then each  $t \in \{A\}$  has a neighbourhood  $N_t$  homeomorphic to

$$\tilde{N}_1 \times \tilde{N}_2$$

where  $\tilde{N}_1$  is, as before, a neighbourhood of the origin in  $\mathbb{R}^{\#A-1}$  and  $\tilde{N}_2$  is a neighbourhood of  $v_0$  in  $\tilde{L}(A, \mathcal{K})_B$ .

We now want to exploit the notion of piecewise flat spaces to develop the notion of piecewise smooth spaces, which are to be the precursors of piecewise smooth manifolds. Working in a fashion analagous to the definition of topological and differentiable manifolds, the first step is to establish

a notion of ‘local piecewise smoothness’. The second step, which we will undertake in the following Section, will be to add a differential structure.

Consider a fixed piecewise flat space  $\mathcal{X}$  as represented by  $\mathcal{K}_B$  (remember  $\mathcal{X}$  is an equivalence class) and an element  $A \in \mathcal{K}$ . As noted above, we can create two separate realisations of  $\tilde{L}(A, \mathcal{K})$ , say  $\tilde{L}(A, \mathcal{K})_{B_1}$  and  $\tilde{L}(A, \mathcal{K})_{B_2}$ . So that we can talk of differentiable mappings on these realisations, we take both to be in Euclidean spaces. For each  $t \in \{A\}$  we can find neighbourhoods  $N_{1,t}$ ,  $N_{2,t}$  and homeomorphisms

$$H_1 : N_{1,t} \rightarrow \tilde{N}_{1,1} \times \tilde{N}_{1,2}, \quad H_2 : N_{2,t} \rightarrow \tilde{N}_{2,1} \times \tilde{N}_{2,2},$$

where  $\tilde{N}_{i,1}$  and  $\tilde{N}_{i,2}$  correspond to  $\tilde{N}_1$  and  $\tilde{N}_2$  above. We can choose  $H_1$  as an affine map from  $\text{span}(\mathcal{K}_B)$  to  $\mathbb{R}^{\#A-1} \times \text{span}(B_1)$ , restricted to  $N_{1,t}$ , and similarly for  $H_2$ . With this choice of  $H_1$  and  $H_2$ , the composition  $H_1 \circ H_2^{-1}$  is an affine, invertible map. Therefore, the restriction of  $H_1 \circ H_2^{-1}$  to any neighbourhood of the form  $N_{2,1} \times$  ‘a face of  $\tilde{L}(A, \mathcal{K})_{B_2}$ ’ is a  $C^\infty$  diffeomorphism.

This leads us naturally to the notion of a *piecewise  $C^k$  space*, which we define to be a Hausdorff space  $M$  and a simplicial complex  $\mathcal{K}$ , such that each  $t \in M$  has a neighbourhood homeomorphic to  $\tilde{N}_1 \times \tilde{N}_2$  for some  $A \in \mathcal{K}$ . As above,  $\tilde{N}_1$  is a neighbourhood of the origin in  $\mathbb{R}^{\#A-1}$  and  $\tilde{N}_2$  is a neighbourhood of the origin in  $\tilde{L}(A, \mathcal{K})_B$ , for some realisation of  $\tilde{L}(A, \mathcal{K})$ . We further require that for any two such neighbourhoods the corresponding homeomorphisms  $H_1 \circ H_2^{-1}$  are  $C^k$  diffeomorphisms when restricted to the appropriate faces of the realisations of  $\tilde{L}(A, \mathcal{K})$ . The space is called *piecewise smooth* if  $k = \infty$ .

One closing, but important additional definition. Given a piecewise  $C^k$  space, we define its Euler characteristic to be the Euler characteristic of one of its representations  $\mathcal{K}_B$ , as given by (3.7.4).

### 3.7.2 Piecewise smooth submanifolds

As advertised, our next step is to lift the notion of piecewise  $C^k$  spaces to that of piecewise  $C^k$  smooth manifolds by adding a differential structure. In order to do this, we shall require that the manifold of interest sit inside something a little larger, so that there will be somewhere for the normal cones to live.

We therefore start with an ambient  $N$ -manifold  $\widetilde{M}$  and, as usual, write the chart at a point  $t \in \widetilde{M}$  as  $(U_t, \varphi_t)$ . For  $M$  to be a submanifold of  $\widetilde{M}$ , it will first need to be a manifold itself. We write its chart at a point  $t \in M$  as  $(V_t, \psi_t)$ . We now define a subset  $M$  of an  $N$ -manifold  $\widetilde{M}$  to be a *piecewise  $C^k$   $q$ -dimensional submanifold* of  $\widetilde{M}$  if  $M$  is a piecewise  $C^k$  space such that we can choose the neighbourhoods  $V_t$  to be of the form

$$V_t = U_t \cap M,$$

and the homeomorphisms  $\psi_t$  are such that

$$\psi_t = \pi_{1,\dots,q} \circ \varphi_{t|M},$$

where  $\pi_{1,\dots,q}$  represents projection onto the first  $q$  coordinates and  $\varphi_t$  is such that

$$\varphi_{t|M}(V_t) \subset \mathbb{R}^q \times \underbrace{(0, \dots, 0)}_{N-q \text{ times}}.$$

There are a number of consequences that follow from our definition of  $C^k$   $q$ -submanifolds and the results of the previous Subsection.

Firstly, note that, by refining the atlas if necessary, we can assume that for each  $t \in M$  there exists  $A \in \mathcal{K}$ , where  $\mathcal{K}$  is the simplicial complex corresponding to  $M$ , such that  $\psi_t(V_t) = \tilde{N}_1 \times \tilde{N}_2$ . Here  $\tilde{N}_1$  is a neighbourhood of the origin in  $\mathbb{R}^{\#A-1}$  and  $\tilde{N}_2$  is a neighbourhood of the origin in some geometric realisation of  $\tilde{L}(A, \mathcal{K})$ . We denote the space spanned by  $\tilde{N}_2$  as  $K_t$ .

Next, we define  $d_t = \#A - 1$ , the *dimension of  $M$  at  $t$* . If  $d_t = j$ , we say that  $t \in \partial_j M$ , the  *$j$ -dimensional boundary of  $M$* . Alternatively,

$$\partial_j M = \{t \in M : d_t = j\}.$$

It can then be shown that  $M$  has a unique decomposition

$$(3.7.8) \quad M = \bigcup_{j=0}^N \partial_j M,$$

where:

- (i) The union is disjoint.
- (ii)  $\partial_q M$  is the relative interior of  $M$  in the topology inherited from  $\tilde{M}$ .
- (iii) For each  $j$ ,  $\partial_j M$  is a  $j$ -dimensional embedded submanifold of  $\tilde{M}$ .
- (iv) For  $j > q$ ,  $\partial_j M = \emptyset$  and, for  $j \leq q$

$$\overline{\partial_j M} = \bigcup_{i=0}^j \partial_i M$$

is a  $C^k$  piecewise smooth space.

For a concrete, and indeed typical, example, take the cube  $[0, 1]^N$  and the point  $t = (1/2, 0, \dots, 0)$ . Then we can write  $A = \{1, 2\}$ , so that  $d_t = 1$ , and we can take  $K_t$  to be the positive orthant in  $\mathbb{R}^{N-1}$ , which is the simplicial cone generated by the standard basis in  $\mathbb{R}^{N-1}$ . Finally, we can take

$$\varphi_t(x_1, \dots, x_N) = (x_1 - 1/2, x_2, \dots, x_N).$$

In this decomposition of  $[0, 1]^N$ ,  $\partial_j M$  would obviously be the  $j$ -dimensional faces of  $[0, 1]^N$ , i.e.  $\partial_0 M$  is the set of vertices,  $\partial_1 M$  is the set (of the interiors) of all the edges, etc.

We now need to set up a differential structure for piecewise smooth submanifolds, despite the fact that they are not true manifolds. The main problem lies in defining a ‘tangent space’  $T_t M$  at each point.

By assumption, in a chart  $(U_t, \varphi_t)$  of the form described above,

$$U_t \cap M \subset \varphi_{t|N-q}^{-1}\{0\},$$

where  $\varphi_{t|N-q}$  denotes the last  $N - q$  coordinate functions of  $\varphi_t$ , so that  $t \in \varphi_{t|N-q}^{-1}\{0\}$ . By the implicit function theorem on manifolds  $\varphi_{t|N-q}^{-1}\{0\}$  is an embedded submanifold of  $U_t \subset \widetilde{M}$  and, naturally, we set  $T_t M = T_t \varphi_{t|N-q}^{-1}\{0\}$ .

Considering the cone  $\widetilde{K}_t = \mathbb{R}^{d_t} \times K_t \times (0, \dots, 0)$  as a subset of  $T_0 \mathbb{R}^q$ , a little thought shows that the push forward

$$(3.7.9) \quad \mathcal{S}_t(M) \triangleq \varphi_{t*}^{-1}|_{(0, \dots, 0)}(\widetilde{K}_t) \subset T_t M$$

is again a cone, which we call the *support cone* of  $M$  at  $t$ . Support cones  $\mathcal{S}_t(M)$  provide a natural replacement for the tangent spaces  $T_t(M)$ .

If  $\widetilde{M} = (\widetilde{M}, g)$  is a Riemannian manifold, in which case we call  $M$  a *piecewise smooth Riemannian submanifold*, we can also define the *normal cone*  $N_t(M)$  of  $M$  at  $t$  by

$$(3.7.10) N_t(M) \triangleq \left\{ X_t \in T_t \widetilde{M} : g_t(X_t, Y_t) \leq 0, \text{ for all } Y_t \in \mathcal{S}_t(M) \right\}.$$

While not bundles in the strict sense, since the fibers are not in general isomorphic (nor even necessarily of the same dimensions) the unions

$$\begin{aligned} T(M) &= \bigcup_{t \in M} \mathcal{S}_t(M), \\ T^{\text{red}}(M) &= \bigcup_{t \in M} T_t \partial_{d_t} M, \\ N(M) &= \bigcup_{t \in M} N_t(M), \end{aligned}$$

are referred to as the *tangent bundle*, *reduced tangent bundle* and *normal bundle* of  $M$  in  $\widetilde{M}$ . All three, along with the projection  $\pi$  defined by  $\pi(X_t) = t$  for  $X_t \in \mathcal{S}_t(M)$  are well-defined. Of the three, the normal bundle is a conceptually new object for us, arising from the embedding of the  $j$ -dimensional boundaries in  $M$  and of  $M$  in  $\widetilde{M}$ . We had no corresponding objects when treating smooth manifolds. It will turn out that  $N(M)$  is crucial for the Morse Theory of Section 3.9 as well as for the derivation of Weyl’s Tube Formula in Chapter 7.

We can also construct corresponding extensions to all of the various tensor bundles we built over manifolds by simply replacing  $T_t M$  by  $\mathcal{S}_t(M)$  through the constructions. In particular, all covariant tensor bundles are bundles in the strict sense, since they are no more than collections of maps from  $\mathcal{S}_t(M)^k$  to  $\mathbb{R}$ . What has changed is merely their domain of definition; i.e. they have been restricted to a subset of  $T_t M$ .

Finally, note that if  $\partial_j M = \emptyset$  for  $j = 0, \dots, q - 1$ , then these definitions agree with our earlier ones for smooth manifolds since, in this case,  $\mathcal{S}_t(M)$  is a  $q$ -dimensional vector space for every  $t \in M$ .

The above structure allows us to make three more definitions. Using the projection  $\pi$ , we can define vector fields as we did in the smooth case, as maps  $X : M \rightarrow T(M)$  such that  $\pi \circ X = \text{id}_M$ . Furthermore, we write  $f \in C^k(M)$  if  $f : M \rightarrow \mathbb{R}$  is such that  $f|_{\partial_j M} \in C^k(\partial_j M)$  for  $j = 1, \dots, q$ .  $C^k$  sections of bundles can be defined analogously.

We close this Section with something completely new in the manifold setting, but somewhat reminiscent of the structure of basic complexes back in Section 3.2. In particular, it should remind you of the way that intersections of complexes gave objects of a similar structure.

We say that two piecewise smooth submanifolds,  $M_1$  and  $M_2$ , subsets of the same ambient  $N$ -manifold  $\widetilde{M}$ , *intersect transversally* if, for each pair  $(j, k)$  with  $0 \leq j \leq \dim(M_1)$  and  $0 \leq k \leq \dim(M_2)$ , and each  $t \in \partial_j M_1 \cap \partial_k M_2$ , the dimension of

$$(3.7.11) \quad \text{span} \{X_t + Y_t : X_t \in T_t^\perp \partial_j M_1, Y_t \in T_t^\perp \partial_k M_2\}$$

is equal to  $2N - j - k \geq 0$ .

For transversally intersecting manifolds, we always have that  $\partial_j M_1 \cap \partial_k M_2$  is a  $(j + k - N)$ -dimensional submanifold of  $\widetilde{M}$ . Furthermore,

$$(3.7.12) \quad M_1 \cap M_2 = \bigcup_{j=0}^N \bigcup_{k=0}^N \partial_{j+k} M_1 \cap \partial_{N-k} M_2.$$

That is,

$$(3.7.13) \quad \partial_j(M_1 \cap M_2) = \bigcup_{k=0}^N \partial_{j+k} M_1 \cap \partial_{N-k} M_2$$

where, as usual, for  $l > N$

$$\partial_l M_1 = \partial_l M_2 \triangleq \emptyset.$$

### 3.8 Intrinsic volumes again

Back in Section 3.4, in the context of Integral Geometry, we described the notions of *intrinsic volumes* and *Minkowski functionals* and how they

could be used, via Steiner's formula (3.4.3), to find an expression for the tube around a convex Euclidean set.

We now describe corresponding functionals for Riemannian manifolds, which will turn out to play a crucial rôle in the expressions developed in Chapter 4 for the mean Euler characteristic of Gaussian excursion sets and later in Chapter 6 when we look at extrema probabilities. At this point we give no more than the definitions. A more general version of Steiner's formula, known in the manifold setting as *Weyl's tube formula* will be described in detail in Chapter 7.

In the setting of Riemannian manifolds, the intrinsic volumes are known as the *Lipschitz-Killing curvature measures* or simply *curvature measures*. To define these, let  $(M, g)$  be a  $C^2$  Riemannian manifold, for the moment without boundary.

For each  $t \in M$ , the Riemannian curvature tensor  $R_t$  given by (3.6.28) is in  $\Lambda^{2,2}(T_t M)$ , so that for  $j \leq \dim(M)/2$  it makes sense to talk about the  $j$ -th power  $R_t^j$  of  $R_t$ . We can also take the trace

$$\mathrm{Tr}^M(R^j)(t) \triangleq \mathrm{Tr}^{T_t M}(R_t^j).$$

Integrating these powers over  $M$  gives the Lipschitz-Killing curvature measures of  $(M, g)$  defined as follows, for measurable subsets  $U$  of  $M$ :

$$(3.8.1) \quad \mathcal{L}_j(M, U) = \begin{cases} \frac{(-2\pi)^{-(N-j)/2}}{(N-j)!} \int_U \mathrm{Tr}^M(R^{(N-j)/2}) \mathrm{Vol}_g & \text{if } N-j \text{ is even,} \\ 0 & \text{if } N-j \text{ is odd.} \end{cases}$$

If  $(M, g)$  is a piecewise  $C^2$  Riemannian manifold, then the Lipschitz-Killing curvature measures require a little notation to define. For a start, we shall now require that  $M$  be embedded in an ambient Riemannian manifold  $\widetilde{M}$  of codimension at least one. Recall that  $M$  can be written as the disjoint union  $\bigcup \partial_k M$ .

Writing, as usual,  $N_t \partial_k M$  for the normal cone to  $\partial_k M$  in  $\widetilde{M}$  at  $t \in \partial_k M$ , we define  $S(N_t \partial_k M) = \{\nu \in N_t \partial_k M : |\nu| = 1\}$  to be the intersection of the sphere bundle of  $M$  with  $N_t \partial_k M$ . Note that  $g$  determines a volume form on both  $\partial_k M$  and  $N_t \partial_k M$ . We shall use  $\mathcal{H}_k$  or  $\mathrm{Vol}_{\partial_k M, g}$  to denote the first and  $\mathcal{H}_{N-k}$  for the second. (Of course  $\mathcal{H}_{N-k}$  really depends on  $t$ , but we have more than enough subscripts already.)

Finally, note that as we did above for the curvature tensor, we can also take powers of the second fundamental form  $S$ . We now have all we need

to define the Lipschitz-Killing curvature measures of  $M$  as

$$(3.8.2) \quad \begin{aligned} \mathcal{L}_j(M) &= \pi^{-(N-j)/2} \sum_{k=j}^N \sum_{l=0}^{\lfloor \frac{k-j}{2} \rfloor} \frac{(-1)^l \Gamma\left(\frac{N-j-2l}{2}\right) 2^{-(l+2)}}{l!(k-j-2l)!} \\ &\quad \times \int_{U \cap \partial_k M} \mathcal{H}_k(dt) \int_{S(N_t \partial_k M)} \text{Tr}^{\partial_k M}(S_\nu^{k-j-2l} R^l) \mathcal{H}_{N-k-1}(d\nu). \end{aligned}$$

In all cases, the  $j$ -th *Lipschitz-Killing curvature*, or *intrinsic volume*, of  $M$  is defined as

$$(3.8.3) \quad \mathcal{L}_j(M) \stackrel{\Delta}{=} \mathcal{L}_j(M, M).$$

Although we shall meet Lipschitz-Killing curvatures again in Chapter 7 and in much more detail, you might want to look at an example<sup>50</sup> now to see how the above formula works on familiar sets.

These formulae simplify considerably if  $M$  is a  $C^2$  domain of  $\mathbb{R}^N$  with the induced Riemannian structure, since then the curvature tensor is identically zero. In that case  $\mathcal{L}_N(M, U)$  is the Lebesgue measure of  $U$  and

$$(3.8.5) \quad \mathcal{L}_j(M, U) = \int_{\partial M \cap U} \frac{1}{s_{N-j}(N-1-j)!} \text{Tr}(S^{N-1-j}) \text{Vol}_{\partial M, g},$$

---

<sup>50</sup> Consider, for our first example, the unit cube in  $\mathbb{R}^N$  equipped with the standard Euclidean metric. Thus, we want to recover (3.4.5), viz.

$$(3.8.4) \quad \mathcal{L}_j([0, T]^N) = \binom{N}{j} T^j.$$

Since we are in the Euclidean scenario, both the Riemannian curvature and second fundamental form are identically zero, and so the only terms for which the final integral in (3.8.2) is non-zero are those for which  $l = k - j = 0$ . Thus all of the sums collapse and (3.8.2) simplifies to

$$\pi^{-(N-j)/2} \Gamma\left(\frac{N-j}{2}\right) 2^{-1} \int_{\partial_j M} \mathcal{H}_j(dt) \int_{S(N_t \partial_j M)} \mathcal{H}_{N-j-1}(d\nu).$$

Note the following:

- (i)  $S(N_t \partial_j M)$  is a  $(1/2^{N-j})$ -th part of the sphere  $S^{N-j-1}$  and the measure  $\mathcal{H}_{N-j-1}$  is surface measure on  $S^{N-j-1}$ . Consequently

$$\mathcal{H}_{N-j-1}(S(N_t \partial_j M)) = \frac{2\pi^{(N-j)/2} 2^{-(N-j)}}{\Gamma\left(\frac{N-j}{2}\right)}.$$

- (ii) There are  $2^{N-j} \binom{N}{j}$  disjoint components to  $\partial_j M$ , each one a  $j$  dimensional cube of edge length  $T$ .
- (iii) The volume form on  $\partial_j M$  is Lebesgue measure, so that each of the cubes in (ii) has volume  $T^j$ .

Now combine all of the above, and (3.8.4) follows.

for  $0 \leq j \leq N-1$ , where  $s_j$  is given by (1.4.42). To be sure you understand how things are working, you might want also to read the footnote<sup>51</sup> describing how to use this result to compute the Lipschitz-Killing curvatures of the  $N$ -dimensional ball and sphere.

In this setting (3.8.5) can also be written in another form that is often more conducive to computation. If we choose an orthonormal frame field  $(E_1, \dots, E_{N-1})$  on  $\partial M$ , and then extend this to one on  $M$  in such a way that  $E_N = \nu$ , then it follows from (3.5.13) that

$$(3.8.6) \quad \mathcal{L}_j(M, U) = \frac{1}{s_{N-j}} \int_{\partial M \cap U} \text{detr}_{N-1-j}(\text{Curv}) \text{Vol}_{\partial M, g},$$

where  $\text{detr}_j$  is given by (3.5.14) and the curvature matrix  $\text{Curv}$  is given by

$$(3.8.7) \quad \text{Curv}(i, j) \triangleq S_{E_N}(E_i, E_j).$$

It is important to note that while the elements of the curvature matrix may depend on the choice of basis,  $\text{detr}_{N-1-j}(\text{Curv})$  is independent of the choice, as will be  $\mathcal{L}_j(M, U)$ .

The above is, in essence, all we need to know about intrinsic volumes for the material of Chapter 4. If you want to see more now, you can skip directly to Chapter 7 which requires none of the intervening detail.

### 3.9 Critical Point Theory

Critical point theory, also known as Morse theory, is a technique for describing various global topological characteristics of manifolds via the local behaviour, at critical points, of functions defined over the sets. We have already seen a version of this back in Section 3.3, where we obtained point set representations for the Euler characteristic of excursion sets. (cf. Theorems 3.3.4 and 3.3.5, which gave point set representations for excursion sets in  $\mathbb{R}^2$ , over squares and over bounded sets with  $C^2$  boundaries.) Our aim now is to set up an analogous set of results for the excursion sets of

---

<sup>51</sup>The first thing to note is that the Lipschitz-Killing curvatures of the  $N$ -dimensional sphere follow from (3.8.1), while those of the ball follow from (3.8.5). We shall look only at the ball of radius  $T$  and attempt to recover (3.4.7). It is easy to check that the second fundamental form of  $S^{N-1}(T)$  (a sphere of radius  $T$ ) in  $B^N(T)$  is  $T^{-1}$  and constant over the sphere. Thus the term involving its trace can be taken out of the integral in (3.8.5) leaving only the volume of  $S^{N-1}(T)$ , given by  $s_N T^{N-1}$ . To compute the trace we use (3.5.17), to see that  $\text{Tr} S^{N-1-j} = T^{-(N-j-1)}(N-1)!/j!$ , so that

$$\mathcal{L}_j(B^N(T)) = \frac{T^{-j} s_N T^{N-1} T^{-(N-j-1)}(N-1)!}{s_{N-j}(N-1-j)!j!} = \binom{N-1}{j} \frac{s_N}{s_{N-j}} = \binom{N}{j} \frac{\omega_N}{\omega_{N-j}},$$

which is (3.4.7).

$C^2$  functions defined over  $C^3$  piecewise smooth manifolds. We shall show in Section 3.9.2 how to specialise these back down to the known, Eulidean case.

A full development of this theory, which goes well beyond what we shall need, is in the classic treatise of Morse and Cairns [72], but you can also find a very readable introduction to this theory in the recent monograph of Matsumoto [68]. The standard theory, however, concentrates on smooth, as opposed to piecewise smooth, manifolds. Nevertheless, it is the piecewise scenario which is crucial for us, and so we shall, from the very beginning, work there.

### 3.9.1 Morse theory for piecewise smooth manifolds

We begin with a general definition of critical points. Let  $M$  be a  $N$ -manifold, without boundary, embedded in an ambient Riemannian manifold  $(\widetilde{M}, g)$  with the metric induced by  $g$ . For  $f \in C^1(\widetilde{M})$ , a *critical point* of  $f$  in  $M$  is a point  $t \in M$  such that  $\nabla f_t = 0$ , where  $\nabla f$ , the *gradient* of  $f$ , is the unique continuous vector field on  $\widetilde{M}$  such that

$$(3.9.1) \quad g_t(\nabla f_t, X_t) = X_t f$$

for every vector field  $X$ <sup>52</sup>. Points which are not critical are called *regular*.

Now take  $M$  to be compact and  $C^2$  piecewise smooth, with a  $C^3$  ambient manifold  $(\widetilde{M}, g)$ . Extending the notion of critical points to  $M$  requires taking note of the fact that the various boundaries in  $M$  are of different dimension and so, in essence, involves repeating the above definition for each  $f|_{\partial_j M}$ . However, our heavy investment in notation now starts to pay dividends, since it is easy to see from the general definition that a point  $t \in \partial_j M$ , for some  $0 < j < N$ , is a critical point if, and only if,

$$(3.9.3) \quad \nabla f_t \in T_t^\perp \partial_j M.$$

(cf. (3.6.32).) Thus we need work only with the single function  $f$  and not, explicitly at least, with its various restrictions<sup>53</sup>.

---

<sup>52</sup>Hopefully, the double usage of  $\nabla$  for both gradient and Riemannian connection will not cause too many difficulties. Note that, like the connection, the gradient ‘knows’ about the Riemannian metric  $g$ . In fact, it is not hard to check that, in terms of the natural basis on  $\widetilde{M}$ , the gradient can be expressed as

$$(3.9.2) \quad \nabla f = \sum_{ij} g^{ij} \frac{\partial f}{\partial x_i} \frac{\partial}{\partial x_j},$$

where the  $g^{ij}$  are the elements of the inverse matrix to the matrix determined by the  $g_{ij}$  of (3.6.19).

<sup>53</sup> This assumes, however, that one remembers where all these spaces are sitting, or (3.9.3) makes little sense. Assuming that the ambient space  $\widetilde{M}$  is  $N$ -dimensional, we

Consistent with this, all points in  $\partial_0 M$  will be considered as critical points. Furthermore, critical points of  $f|_{\partial_N M} \equiv f|_{M^\circ}$  are just critical points of  $f$  in the sense of the initial definition. We call the set

$$\bigcup_{j=0}^N \{t \in \partial_j M : \nabla f_t \in T_t^\perp \partial_j M\}$$

the set of critical points of  $f|_M$ .

The next step is to define the Hessian operator  $\nabla^2$  and look at some of its properties, once again exploiting our notation to handle the boundaries of  $M$ .

The (*covariant*) Hessian  $\nabla^2 f$  of a function  $f \in C^2(\widetilde{M})$  on a Riemannian manifold  $(\widetilde{M}, g)$  is the bilinear symmetric map from  $C^1(T(\widetilde{M})) \times C^1(T(\widetilde{M}))$  to  $C^0(\widetilde{M})$  (i.e. double differential form) defined by<sup>54</sup>

$$(3.9.4) \quad \nabla^2 f(X, Y) \triangleq XYf - \nabla_X Yf = g(\nabla_X \nabla f, Y),$$

where  $\nabla_X$  is the usual Levi-Civita connection of  $(\widetilde{M}, g)$ . The equality here is a consequence of (3.6.5) and (3.9.1). Note that

$$\nabla^2 f \in \Gamma^0(\Lambda^{1,1}(\widetilde{M}))$$

and, at a critical point, it follows from (3.6.4) and (3.9.4) that the Hessian is independent of the metric  $g$ .

A critical point  $t \in \partial_j M$  of  $f|_M$  is called *non-degenerate* if the bilinear mapping  $\nabla^2 f|_{T_t \partial_j M}$  is non-degenerate. A function  $f \in C^2(\widetilde{M})$  is said to be non-degenerate on  $M$  if all the critical points of  $f|_M$  are non-degenerate. The *index* of a non-degenerate critical point  $t \in \partial_j M$  of  $f|_M$  is the dimension of the largest subspace  $L$  of  $T_t \partial_j M$ , such that  $\nabla^2 f|_L$  is negative definite. Thus, a point of index zero is a local minimum of  $f$  on  $\partial_j M$ , while a point of index  $j$  is a local maximum. Other indices correspond to saddle points of various kinds.

A function  $f \in C^2(\widetilde{M})$  is called a *Morse function* on  $M$  if, for each  $k$ ,  $f|_{\partial_k M}$  it is non-degenerate on  $\partial_k M$  and the restriction of  $f$  to  $\overline{\partial_k M} =$

---

have, on the one hand, that  $\nabla f$  is also  $N$ -dimensional, whereas  $T_t^\perp \partial_j M$  is  $(N - j)$ -dimensional. It is important, therefore, to think of  $T_t^\perp \partial_j M$  as a subspace of  $T_t^\perp M$  for the inclusion to make sense. Overall, all of these spaces are dependent on  $\widetilde{M}$  and its Riemannian metric.

<sup>54</sup> $\nabla^2 f$  could also have been defined to be  $\nabla(\nabla f)$ , in which case the first relationship in (3.9.4) becomes a consequence rather than a definition. Recall that in the simple Euclidean case the Hessian is defined to the  $N \times N$  matrix  $H_f = (\partial^2 f / \partial x_i \partial x_j)_{i,j=1}^N$ . Using  $H_f$  to define the two-form  $\nabla^2 f(X, Y) = XH_f Y'$ , (3.9.4) follows from simple calculus.

$\bigcup_{j=0}^k \partial_j M$  has no critical points on  $\bigcup_{j=0}^{k-1} \partial_j M$ . By the assumed non-degeneracy of  $f|_{\partial_k M}$ , this is equivalent to the same requirement in a neighbourhood of  $\bigcup_{j=0}^{k-1} \partial_j M$ .

Finally, and most importantly, we have

**Definition 3.9.1** *In the above setup, we call a point  $t \in M$  an extended inward critical point of  $f|_M$  if, and only if,*

$$-\nabla f_t \in N_t(M),$$

where  $N_t(M)$  is the normal cone given by (3.7.10).

Furthermore, with the notation

(3.9.5)

$$C_i(f, M) \triangleq \{\text{extended inward critical points of } f|_M \text{ of index } i\},$$

we define, for  $0 \leq i \leq N$ ,

$$\mu_i(f, M) \triangleq \# C_i(f, M),$$

the augmented type numbers of  $f$  on  $M$ .

Note that, if  $t \in M^\circ$ , then  $t$  is an extended inward critical point if, and only if, it is a critical point in the simple sense of (3.9.3). The main point of the definition is to handle critical points on the boundary of  $M$ . In relation to these it is important to realise that within the set  $C_i(f, M)$  may lie points from any of the  $\partial_j M$ , for which the geometric meaning of belonging to  $C_i$  may be different.

With all the definitions cleared up, we have the necessary ingredients to state the following version of Morse's Theorem, due to Takemura and Kuriki [90].

**Theorem 3.9.2 (Morse's Theorem)** *Let  $M$  be a  $C^2$  piecewise smooth,  $N$  dimensional submanifold of a  $C^3$  Riemannian manifold  $\widetilde{M}$ . Assume that all support cones of  $M$  are convex and that the normal cones at points in  $\partial M$  are non-empty. Let  $f \in C^2(\widetilde{M})$  be a Morse function on  $M$ . Then,*

$$(3.9.6) \quad \varphi(M) = \sum_{i=0}^N (-1)^i \mu_i(f, M)$$

where  $\varphi(M)$  is the Euler characteristic of  $M$ .

Recall that  $\varphi(M)$  has been defined via a simplicial representation for  $M$ , as in (3.7.4).

We have slipped in two new and rather strange looking additional assumptions in the statement of Morse's Theorem relating to support and

normal cones. Convexity is a property that appears, one way or another, throughout all of Integral Geometry and those parts of Differential Geometry which interest us. Back in Section 3.2 we relaxed convexity by looking at basic complexes, the main trick being that basics were a little more general than convex sets. Here too, we are not asking for full convexity. Consider, for example, the examples in Figure 3.9.1.

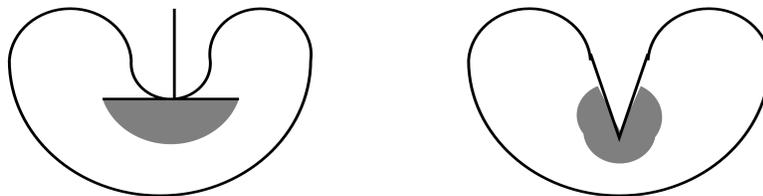


FIGURE 3.9.1. Convex and non-convex support cones

In the first example, in which the boundary is smooth, the support cone at the base of the concavity is a half plane, and so convex. The normal cone (assuming that the ambient manifold is  $\mathbb{R}^2$ ) is the outward normal to the surface. Thus this set fulfills the conditions of Morse's Theorem. On the other hand, in the second example the support cone at the base of the concavity is the union of two half planes and no longer convex. As well as this, the normal cone is empty. These two examples are generic, in that the import of our convexity assumption is not to force convexity itself, but rather to ensure that there are no 'concave cusps'.<sup>55</sup>

Morse's Theorem is a deep and important result in Differential Topology and is actually somewhat more general than as stated here, since in its full form it also gives a series of inequalities linking the Betti numbers<sup>56</sup> of  $M$  to augmented type numbers. We shall make no attempt to prove Morse's Theorem, which, given our definition of the Euler characteristic, relies on arguments of Homology Theory and Algebraic Geometry. As mentioned above, [72] and [68] have all the details. Theorem 3.9.2, as presented here, is essentially proven in [90], although the notation and terminology there is a little different to ours.

Soon, in Section 3.9.2, we shall see how all of this impacts on simple Euclidean examples, in which case we shall recover the Integral Geometric results of Section 3.3 for which we *do* have a proof.

<sup>55</sup>Thus, for example, while there is no problem with  $N$ -dimensional unit cubes, their  $(N - 1)$ -dimensional boundaries are not acceptable, since at the vertices the support cones are comprised of the non-convex union of  $N$  half-spaces of dimension  $N - 1$ .

<sup>56</sup>Betti numbers are additional geometrical invariants of  $M$  of somewhat less straightforward interpretation than the Euler characteristic. They are not unrelated to the Lipshitz-Killing curvatures of Section 3.8.

What we shall prove now is the following Corollary, which is actually what we shall be using in the future. The proof is included since, unlike that for Morse's Theorem itself, it does not seem to appear in the literature.

**Corollary 3.9.3** *Let  $M$  and  $\widetilde{M}$  be as in Morse's Theorem, with the added condition that  $\widetilde{M}$  not have a boundary. Let  $f \in C^2(\widetilde{M})$  be a Morse function on  $M$ , and let  $u \in \mathbb{R}$  be a regular value of  $f|_{\partial_j M}$  for all  $j = 0, \dots, N$ . Then,*

(3.9.7)

$$\varphi(M \cap f^{-1}[u, +\infty)) = \sum_{i=0}^N (-1)^i \#C_i(-f, M \cap f^{-1}(u, +\infty)).$$

**Proof.** As usual, write  $A_u = M \cap f^{-1}[u, +\infty)$ . Our first claim is that the conditions of the Corollary ensure that  $A_u$  has convex support cones. To see this, it suffices to note that both those of  $M$  and  $f^{-1}[u, +\infty)$  are convex (by assumption and the smoothness of  $f$ , respectively) and thus their intersections, which give the support cones of  $A_u$ , are also convex. It is also not hard to see that the normal cones to  $A_u$  are non-empty.

Consequently, if  $-f$  were a Morse function on  $A_u$ , and if we changed  $f^{-1}(u, +\infty)$  to  $f^{-1}[u, +\infty)$  on the right hand side of (3.9.7), then the Corollary would merely be a restatement of Morse's Theorem, and there would be nothing to prove. However, the change is not obvious, and  $-f$  is not a Morse function on  $A_u$ , since it is constant on  $f^{-1}\{u\}$ .

The bulk of the proof involves finding a Morse function  $\widehat{f}$  on  $A_u$  that agrees with  $-f$  on "most" of this set (thus solving the problem of the 'non-Morse-ness' of  $f$ ) and which, at the same time, ensures that there are no contributions to  $C_i(\widehat{f}, A_u)$  from  $M \cap f^{-1}\{u\}$ . (Thus allowing us to replace  $f^{-1}[u, +\infty)$  with  $f^{-1}(u, +\infty)$ .)

More formally, consider  $f$  as a function on  $\widetilde{M}$ , which, recall, is assumed to have no boundary. Then

$$\begin{aligned} \partial_N(f^{-1}[u, +\infty)) &= f^{-1}(u, +\infty), \\ \partial_{N-1}(f^{-1}[u, +\infty)) &= f^{-1}\{u\}, \\ \partial_j(f^{-1}[u, +\infty)) &= \emptyset, \quad \text{for } j = 0, \dots, N-2. \end{aligned}$$

Since  $f$  is a Morse function and  $u$  is a regular point for  $f|_{\partial_k M}$  for all  $k$ , it follows that  $M$  and  $f^{-1}[u, \infty)$  intersect transversally as subsets of  $\widetilde{M}$ . Therefore  $A_u$  is a piecewise smooth submanifold of  $\widetilde{M}$  and, as in (3.7.12),

it can be decomposed as

$$(3.9.8) \quad \begin{aligned} A_u &= \bigcup_{j=0}^N (\partial_j M \cap f^{-1}(u, +\infty)) \cup (\partial_{j+1} M \cap f^{-1}\{u\}) \\ &= \left( \bigcup_{j=0}^N \partial_j M \cap f^{-1}(u, +\infty) \right) \cup \left( \bigcup_{j=1}^N \partial_j M \cap f^{-1}\{u\} \right) \end{aligned}$$

We need to somehow get rid of the second term here.

Since  $f$  is a Morse function on  $M$ , it has only finitely many critical points inside a relatively compact neighborhood  $V$ , of  $M$ . Furthermore, again exploiting the fact that  $u$  is a regular value of  $f|_{\partial_k M}$  for every  $k$ , there exists an  $\varepsilon > 0$  such that

$$U_\varepsilon = f^{-1}(u - \varepsilon, u + \varepsilon) \cap M \cap V$$

contains no critical points of  $f|_M$ . It is standard fare that there exists a  $h \in C^2(\widetilde{M})$  which is a Morse function on  $f^{-1}\{u\}$  and which  $h$  is zero outside of  $U_\varepsilon$ . Furthermore, since  $\overline{V}$  is compact, there exist  $K_f$  and  $K_h$  such that  $|\nabla h| < K_h$  and

$$|\pi_{T_t \partial_j M} \nabla f| > K_f,$$

for all  $t \in \partial_j M \cap U_\varepsilon$ ,  $1 \leq j \leq N$ . (Here  $\pi$  is the usual projection operator.)

It then follows that the function

$$\widehat{f} \triangleq -f + \frac{K_f}{3K_h} h$$

is a Morse function on  $A_u$ . By our choice of  $h$ , the critical points of  $\widehat{f}|_{A_u}$  agree with those of  $f|_M$  on  $M \cap U_\varepsilon^c$ . Furthermore,  $\nabla f|_{\partial_j M} \equiv \pi_{T_t \partial_j M} \nabla f$  can never be zero on  $U_\varepsilon \cap \partial_k M$ , and so there are no critical points of  $\widehat{f}$  at all in this region. Consequently,

$$\begin{aligned} C_i(\widehat{f}, M \cap f^{-1}[u, +\infty)) &= C_i(-f, M \cap f^{-1}[u + \varepsilon, +\infty)) \\ &= C_i(-f, M \cap f^{-1}(u, +\infty)), \end{aligned}$$

from which the conclusion follows.  $\square$

### 3.9.2 The Euclidean case

With basic Morse Theory for piecewise smooth manifolds under our belt, it is now time to look at one rather important example for which everything becomes quite simple. The example is that of the  $N$ -dimensional cube  $I^N =$

$[0, 1]^N$ , and the ambient space is  $\mathbb{R}^N$  with the usual Euclidean metric. In particular, we want to recover Theorem 3.3.4, which gave a point set representation for the Euler characteristic of the excursion set of smooth function over the square.

To recover Theorem 3.3.4 for the unit square we use Morse's Theorem 3.9.2 in its undiluted version. Reserving the notation  $f$  for the function of interest which generates the excursion sets, write the  $f$  of Morse's Theorem as  $f_m$ . We are interested in computing the Euler characteristic of the set

$$A_u = \{t \in I^2 : f(t) \geq u\}$$

and will take as our Morse function the 'height function'

$$f_m(t) = f_m(t_1, t_2) = t_2.$$

Now assume that  $f$  is 'suitably regular' in the sense of Definition 3.3.1. This is almost enough to guarantee that  $f_m$  is a Morse function over  $I^2$  for the ambient manifold  $\mathbb{R}^2$ . Unfortunately, however, all the points along the top and bottom boundaries of  $I^2$  are degenerate critical points for  $f_m$ . We get around this by replacing  $I^2$  with a tilted version,  $I_\varepsilon^2$ , obtained by rotating the square through  $\varepsilon$  degrees, as in Figure 3.9.2.

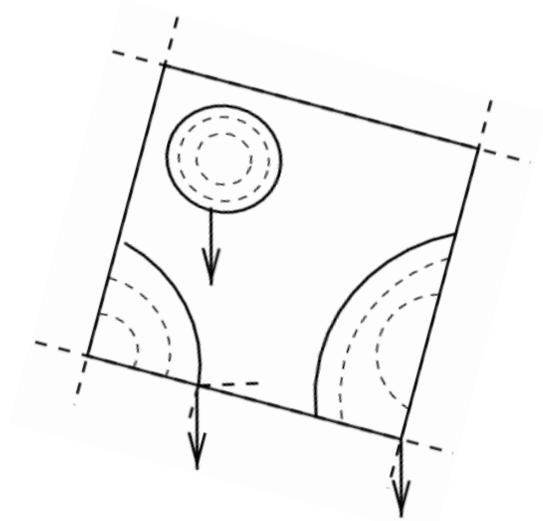


FIGURE 3.9.2. The Euler characteristic via Theorem 3.9.2

To compute  $\varphi(A_u)$  we now apply (3.9.6), and so need to characterise the various critical points and their indices. The first fact to note is that, using the usual coordinate system,  $\nabla f_m = (0, 1)$ , and so there are no critical points of  $f_m$  in  $A_u^\circ$ . Thus we can restrict interest to the boundary  $\partial A_u$  which we break into three parts:

- (i) Points  $t \in (I_\varepsilon^2)^\circ \cap \partial A_u$ .
- (ii) Points  $t \in \partial I_\varepsilon^2 \cap A_u$ , but not vertices of the square.
- (iii) The four vertices of the square.

An example of each of these three classes appears in Figure 3.9.2, where the excursion set of  $f$  appear along with contour lines in the interiors of the various components.

At points of type (i),  $f(t) = u$ . Furthermore, since the normal cone  $N_t(A_u)$  is then the one dimensional vector space normal to  $\partial A_u$ ,  $-\nabla f_m = (0, -1) \in N_t(A_u)$  at points for which  $\partial f/\partial t_1 = 0$  and  $\partial f/\partial t_2 > 0$ . Such a point is at the base of the arrow coming out of the disk in Figure 3.9.2. Differentiating between points which contribute +1 and -1 to the Euler characteristic involves looking at  $\partial^2 f/\partial t_1^2$ . Comparing with Theorem 3.3.4, we see we have characterised the contributions of (3.3.17) and (3.3.18) to the Euler characteristic.

We now turn to points of type (ii). Again due constancy of  $\nabla f_m$ , this time on  $\partial I_\varepsilon^2$ , there are no critical points to be counted on  $(\partial I_\varepsilon^2 \cap A_u)^\circ$ . We can therefore add the end points of the intervals making up  $\partial I_\varepsilon^2 \cap A_u$  to those of type (iii). One of these appears as the base of the left most arrow on the base of Figure 3.9.2. The rightmost arrow extends from such a vertex.

For points of these kinds, the normal cone is a closed wedge in  $\mathbb{R}^2$ , and it is left to you to check that that the contributions of these points correspond (on taking  $\varepsilon \rightarrow 0$ ) to those of those described by (3.3.16).

This gives us Theorem 3.3.4, which trivially extends to any rectangle in  $\mathbb{R}^2$ . You should now check that Theorem 3.3.5, which computed the Euler characteristic for a subset of  $\mathbb{R}^2$  with piecewise  $C^2$  boundary, also follows from Morse's Theorem, using the same  $f_m$ .

The above argument was really unnecessarily complicated, since it did not use Corollary 3.9.3 which we built specifically for the purpose of handling excursion sets. Nevertheless, it did have the value of connecting the Integral Geometric and Differential Geometric approaches.

Now we apply Corollary 3.9.3. We again assume that  $f$  is suitably regular at the level  $u$  in the sense of Definition 3.3.1, which suffices to guarantee that it is a Morse function over the  $C^2$  piecewise smooth  $I^N$  for the ambient manifold  $\mathbb{R}^N$  and that Conditions (i) and (ii) of the Corollary apply.

Write  $\mathcal{J}_k \equiv \partial_k I^N$  for be the collection of faces of dimension  $k$  in  $I^N$  (cf. (3.3.2)). With this notation, we can rewrite the sum (3.9.7) as

$$(3.9.9) \quad \varphi(A_u(f, I^N)) = \sum_{k=0}^N \sum_{J \in \mathcal{J}_k} \sum_{i=0}^k (-1)^i \mu_i(J),$$

where, for  $i \leq \dim(J)$ ,

$$\mu_i(J) \triangleq \#C_i(-f|_J, f|_J^{-1}(u, +\infty))$$

and the  $C_i$  are as in (3.9.5).

Recall that to each face  $J \in \mathcal{J}_k$  there corresponds a subset  $\sigma(J)$  of  $\{1, \dots, N\}$ , of size  $k$ , and a sequence of  $N - k$  zeroes and ones  $\varepsilon(J) = \{\varepsilon_1, \dots, \varepsilon_{N-k}\}$  so that

$$J = \{t \in I^N : t_j = \varepsilon_j, \text{ if } j \notin \sigma(J), \ 0 < t_j < 1, \text{ if } j \in \sigma(J)\}.$$

Set  $\varepsilon_j^* = 2\varepsilon_j - 1$ . Working with the definition of the  $C_i$ , it is then not hard to see that  $\mu_i(J)$  is given by the number of points  $t \in J$  satisfying the following four conditions:

$$(3.9.10) \quad f(t) \geq u,$$

$$(3.9.11) \quad f_j(t) = 0, \quad j \in \sigma(J)$$

$$(3.9.12) \quad \varepsilon_j^* f_j(t) > 0, \quad j \notin \sigma(J)$$

$$(3.9.13) \quad \text{Index}(f_{mn}(t))_{(m,n \in \sigma(J))} = k - i,$$

where, as usual, subscripts denote partial differentiation, and, consistent with the definition of the index of a critical point, we define the index of a matrix to be the number of its negative eigenvalues.

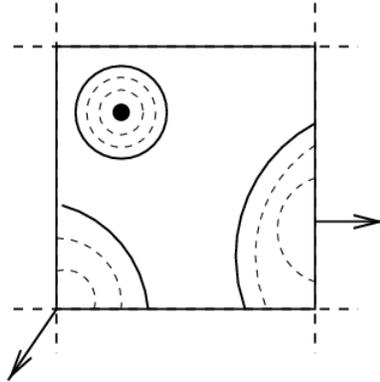


FIGURE 3.9.3. The Euler characteristic via Corollary 3.9.3

In Figure 3.9.3 there are three points which contribute to  $\varphi(A_u(f, I^2))$ . One, in the centre of the upper left disk, contributes via  $J = (I^2)^\circ \in \mathcal{J}_2$ . That on the right side contributes via  $J = \text{'right side'} \in \mathcal{J}_1$ , and that on the lower left corner via  $J = \{0\} \in \mathcal{J}_0$ .

The representation (3.9.9) of the Euler characteristic of an excursion set, along with the prescription in (3.9.10)–(3.9.13) as to how to count the contributions of various points to the sum, is clearly a tidier way of writing things than that we obtained via Integral Geometric methods. Nevertheless, it is now clear that the two are essentially different versions of the same basic result. However, it is the compactness of (3.9.9) that will be of importance to us in the upcoming computations for random excursion sets in Chapter 4.



# 4

## Gaussian random geometry

With the deterministic geometry of Chapter 3 behind us, we can now return to the stochastic setting. The main aim of this Chapter will be to obtain explicit formulae for the expected Euler characteristic of the excursion sets of Gaussian random fields.

In the same way that we divided the treatment of the geometry into two parts – for functions over Euclidean spaces and functions over general manifolds – we shall also have two treatments here. Unlike the case in Chapter 3, however, even if you are primarily interested in the manifold scenario you will need to read the Euclidean case first, since some of the manifold computations will be lifted from this case via atlas based arguments.

The Chapter is long, and develops in a number of distinct stages. Initially, we shall develop rather general results which give integral formulae for the expected number of points at which a vector valued random field takes specific values. These are Theorem 4.1.1 and its corollaries in the Euclidean setting and Theorem 4.7.1 in the manifold setting. In view of the results of Chapter 3, which relate the global topology of excursion sets of a function to its local behaviour, it should be clear what this has to do with Euler characteristic computations. However, the results are important beyond this setting and indeed beyond the setting of this book, so we shall develop them slowly and carefully.

Once these general results are established we return to the Euler characteristic scenario, where the main results are Theorems 4.6.2 (Euclidean) and 4.10.1 and 4.10.2 (manifolds).

In the final Section 4.12 of the Chapter we shall return to a purely deterministic setting and use our Gaussian field results to provide a probabilistic

proof of the classical Chern-Gauss-Bonnet Theorem of Differential Geometry using nothing<sup>1</sup> but Gaussian processes. This really has nothing to do with anything else in the book, but we like it too much not to include it.

## 4.1 An expectation meta-theorem

As promised, we start with a meta-theorem about the expected number of points at which a vector-valued random field takes values in some set. For the moment, we gain nothing by assuming that our fields are Gaussian and so do not do so. Here is the setting:

For some  $N, K \geq 1$  let  $f = (f^1, \dots, f^N)$  and  $g = (g^1, \dots, g^K)$ , respectively, be  $\mathbb{R}^N$  and  $\mathbb{R}^K$  valued  $N$ -parameter random fields. We need two sets,  $T \subset \mathbb{R}^N$  and  $B \subset \mathbb{R}^K$ .  $T$ , as usual, is a compact parameter set, but now we add the assumption that its boundary  $\partial T$  has finite  $\mathcal{H}_{N-1}$ -measure. (cf. Footnote 33 in Chapter 3.) While it is not essential, to simplify our proofs we shall assume that there exists a  $C^1$  function  $h_B : \mathbb{R}^K \rightarrow \mathbb{R}$  such that

$$(4.1.1) \quad B = \{x \in \mathbb{R}^K : h_B(x) \geq 0\}.$$

From this it follows that  $\partial B$  must be of (Hausdorff) dimension  $K-1$ , which is what we would assume if we were prepared to invest the time and space required to prove a fuller result.

As usual  $\nabla f$  denotes the gradient of  $f$ . Since  $f$  takes values in  $\mathbb{R}^N$  this is now a  $N \times N$  matrix of first-order partial derivatives of  $f$ ; i.e.

$$(\nabla f)(t) \equiv \nabla f(t) \equiv (f_j^i(t))_{i,j=1,\dots,N} \equiv \left( \frac{\partial f^i(t)}{\partial t_j} \right)_{i,j=1,\dots,N}.$$

All the derivatives here are assumed to exist in an almost sure sense<sup>2</sup>.

**Theorem 4.1.1** *Let  $f, g, T$  and  $B$  be as above. Assume that the following conditions are satisfied for some  $u \in \mathbb{R}$ :*

- (a) *All components of  $f, \nabla f$ , and  $g$  are a.s. continuous and have finite variances (over  $T$ ).*

---

<sup>1</sup>Of course, this cannot really be true. Establishing the Chern-Gauss-Bonnet Theorem without any recourse to algebraic geometry would have been a mathematical coup that might even have made Probability Theory a respectable topic within Pure Mathematics. What will be hidden in the small print is that everything relies on the Morse theory of Section 3.9 and this, in turn, uses algebraic geometry. However, our approach will save myopic probabilists from having to read the small print.

<sup>2</sup>This is probably too strong an assumption, since, at least in one dimension, Theorem 4.1.1 is known to hold under the assumption that  $f$  is absolutely continuous, and so has only a weak sense derivative. (cf. [65].) However, since we shall need a continuous sample path derivative later for other things, we assume it now.

- (b) For all  $t \in T$ , the marginal densities  $p_t(x)$  of  $f(t)$  (implicitly assumed to exist) are continuous at  $x = u$ .
- (c) The conditional densities<sup>3</sup>  $p_t(x|\nabla f(t), g(t))$  of  $f(t)$  given  $g(t)$  and  $\nabla f(t)$  (implicitly assumed to exist) are bounded above and continuous at  $x = u$ , uniformly in  $t \in T$ .
- (d) The conditional densities  $p_t(z|f(t) = x)$  of  $\det \nabla f(t)$  given  $f(t) = x$  are continuous for  $z$  and  $x$  in neighbourhoods of 0 and  $u$ , respectively, uniformly in  $t \in T$ .
- (e) The following moment condition holds:

$$(4.1.2) \quad \sup_{t \in T} \max_{1 \leq i, j \leq N} \mathbb{E} \left\{ |f_j^i(t)|^N \right\} < \infty.$$

- (f) The moduli of continuity with respect to the usual Euclidean norm (cf. (2.1.6)) of each of the components of  $f$ ,  $\nabla f$ , and  $g$  satisfy

$$(4.1.3) \quad \mathbb{P} \{ \omega(\eta) > \varepsilon \} = o(\eta^N), \quad \text{as } \eta \downarrow 0,$$

for any  $\varepsilon > 0$ .

Then, if

$$N_u \equiv N_u(T) \equiv N_u(f, g : T, B)$$

denotes the number of points in  $T$  for which

$$f(t) = u \in \mathbb{R}^N \quad \text{and} \quad g(t) \in B \subset \mathbb{R}^K,$$

and  $p_t(x, \nabla y, v)$  denotes the joint density of  $(f_t, \nabla f_t, g_t)$ , we have, with  $D = N(N+1)/2 + K$ ,

$$(4.1.4) \quad \mathbb{E}\{N_u\} = \int_T \int_{\mathbb{R}^D} |\det \nabla y| \mathbb{1}_B(v) p_t(u, \nabla y, v) d(\nabla y) dv dt.$$

It is sometimes more convenient to write this as

$$(4.1.5) \quad \mathbb{E}\{N_u\} = \int_T \mathbb{E} \left\{ |\det \nabla f(t)| \mathbb{1}_B(g(t)) \mid f(t) = u \right\} p_t(u) dt,$$

where the  $p_t$  here is the density of  $f(t)$ .

---

<sup>3</sup>Standard notation would demand that we replace the matrix  $\nabla y$  by the  $N(N+1)/2$  dimensional vector  $\text{vech}(\nabla y)$  both here and, even moreso, in (4.1.4) following, where the differential  $d(\nabla y)$ , which we shall generally write in even greater shorthand as  $d\nabla y$ , is somewhat incongruous. Nevertheless, on the basis that it is clear what we mean, and that it is useful to be able to easily distinguish between reals, vectors and matrices, we shall always work with this slightly unconventional notation.

While conditions (a)–(f) arise naturally in the proof of Theorem 4.1.1, they all but disappear in one of the cases of central interest to us, when the random fields  $f$  and  $g$  are Gaussian. In these cases all the marginal and conditional densities appearing in the conditions of Theorem 4.1.1 are also Gaussian and so their boundedness and continuity is immediate, as long as all the associated covariance matrices are non-degenerate, which is what we need to assume in this case. This will also imply that all variances are finite and that the moment condition (4.1.2) holds. Thus the only remaining conditions are the a.s. continuity of  $f$ ,  $\nabla f$  and  $g$ , and condition (4.1.3) on the moduli of continuity.

Note first, without reference to normality, that if  $\nabla f$  is continuous then so must<sup>4</sup> be  $f$ . Thus we have only the continuity of  $\nabla f$  and  $g$  to worry about. However, we spent a lot of time in Chapter 2 finding conditions which will guarantee this. For example, we can apply Theorem 2.2.1.

Write  $C_f^i = C_f^i(s, t)$  for the covariance function of  $f^i$ , so that  $C_{f_j}^i = \partial^2 C_f^i / \partial s_j \partial t_j$  is the covariance function of  $f_j^i = \partial f^i / \partial t_j$ . Similarly,  $C_g^i$  is the covariance function of  $g^i$ . Then, by (2.2.4),  $\nabla f$  and  $g$  will be a.s. continuous if

$$(4.1.6) \quad \max \left( \max_{i,j} \left| C_{f_j}^i(t) - C_{f_j}^i(s) \right|, \max_i \left| C_g^i(t) - C_g^i(s) \right| \right) \leq K |\ln |t - s||^{-(1+\alpha)},$$

for some finite  $K > 0$ , some  $\alpha > 0$  and all  $|t - s|$  small enough.

All that now remains to check is condition (4.1.3) on the moduli of continuity. Here the Borell-TIS Inequality – Theorem 2.3.1 – comes into play. Write  $h$  for any of the components of  $\nabla f$  or  $g$  and  $H$  for the random field on  $T \times T$  defined by  $H(s, t) = h(t) - h(s)$ . Then, writing

$$\omega(\eta) = \sup_{s,t: |t-s| \leq \eta} |H(s, t)|,$$

we need to show, under (4.1.6), that for all  $\varepsilon > 0$ ,

$$(4.1.7) \quad \mathbb{P} \{ \omega(\eta) > \varepsilon \} \leq o(\eta^N).$$

The Borell-TIS Inequality gives us that

$$(4.1.8) \quad \mathbb{P} \{ \omega(\eta) > \varepsilon \} \leq 2 \exp \left\{ -(\varepsilon - \mathbb{E} \{ \omega(\eta) \})^2 / 2\sigma_\eta^2 \right\},$$

where  $\sigma_\eta^2 = \sup_{s,t: |t-s| \leq \eta} \mathbb{E} \{ (H(s, t))^2 \}$ . But (4.1.6) immediately implies that  $\sigma_\eta^2 \leq K |\ln \eta|^{-(1+\alpha)}$ , while together with Theorem 2.2.1 it implies a

---

<sup>4</sup>The derivative can hardly exist, let alone be continuous, if  $f$  is not continuous! In fact, this condition was vacuous all along and was only included for ‘symmetry’ considerations.

bound of similar order for  $\mathbb{E}\{\omega(\eta)\}$ . Substituting this into (4.1.8) gives an upper bound of the form  $C_\varepsilon \eta^{|\ln \eta|^\alpha}$ , and so (4.1.7) holds with room to spare, in that holds for any  $N$  and not just  $N = \dim(T)$ .

Putting all of the above together, we have that Theorem 4.1.1 takes the following much more user friendly form in the Gaussian case.

**Corollary 4.1.2** *Let  $f$  and  $g$  be centered Gaussian fields over a  $T$  which satisfies the conditions of Theorem 4.1.1. If, for each  $t \in T$ , the joint distributions of  $(f(t), \nabla f(t), g(t))$  are non-degenerate, and if (4.1.6) holds, then so do (4.1.4) and (4.1.5).*

We now turn to the proof of Theorem 4.1.1. We shall prove it in a number of stages, firstly by setting up a result that rewrites the random variable  $N_u$  in an integral form, more conducive to computing expectations. Indeed, we could virtually end the proof here, if we were prepared to work at a level of rigour which would allow us to treat the Dirac delta function with gay abandon and exchange delicate orders of integration without justification. Since it is informative to see how such an argument works, we shall give it, before turning to a fully rigorous proof.

The rigorous proof comes in a number of steps. In the first, we use the integral representation to derive an upper bound to  $\mathbb{E}\{N_u\}$ , which actually gives the correct result. (The upper bound actually involves little more than replacing the 'gay abandon' mentioned above with Fatou's Lemma.) The second step involves showing that this upper bound is also a lower bound. The argument here is far more delicate, and involves locally linear approximations of the fields  $f$  and  $g$ . Both of these steps will involve adding conditions to the already long list of Theorem 4.1.1 and so the third and final step involves showing that these additional conditions can be lifted under the conditions of the Theorem.

To state the first result, let  $\delta_\varepsilon : \mathbb{R}^N \rightarrow \mathbb{R}$  be constant on the  $N$ -ball  $B(\varepsilon) = \{r \in \mathbb{R}^N : |t| < \varepsilon\}$ , zero elsewhere, and normalized so that

$$(4.1.9) \quad \int_{B(\varepsilon)} \delta_\varepsilon(t) dt = 1.$$

**Theorem 4.1.3** *Let  $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ ,  $g : \mathbb{R}^N \rightarrow \mathbb{R}^K$  be deterministic, and  $T$  and  $B^5$  as in Theorem 4.1.1. Suppose, furthermore, that the following conditions are all satisfied for  $u \in \mathbb{R}^N$ :*

- (a) *The components of  $f$ ,  $g$  and  $\nabla f$  are all continuous.*
- (b) *There are no points  $t \in T$  satisfying both  $f(t) = u$  and either  $g(t) \in \partial B$  or  $\det \nabla f(t) = 0$ .*

---

<sup>5</sup> Actually, we do not really need (4.1.1) at the moment, and it will not appear in the proof. However, it is needed for our main result, Theorem 4.1.1. cf. the comments preceding Lemma 4.1.10.

(c) There are no points  $t \in \partial T$  satisfying  $f(t) = u$ .

(d) There are only a finite number of points  $t \in T$  satisfying  $f(t) = u$ .

Then

$$(4.1.10) \quad N_u(f, g; T, B) = \lim_{\varepsilon \rightarrow 0} \int_T \delta_\varepsilon(f(t) - u) \mathbb{1}_B(g(t)) |\det \nabla f(t)| dt.$$

**Proof.** To ease on notation, and without any loss of generality, we take  $u = 0$ . Consider those  $t \in T$  for which  $f(t) = 0$ . Since there are only finitely many such points, and none lie in  $\partial T$ , each one can be surrounded by an open sphere, of radius  $\eta$ , say, in such a way that the spheres neither overlap nor intersect  $\partial S$ . Furthermore, because of (b), we can ensure  $\eta$  is small enough so that within each sphere  $g(t)$  always lies in either  $B$  or its complement, but never both.

Let  $\sigma(\varepsilon)$  be the sphere  $|f| < \varepsilon$  in the image space of  $f$ . From what we have just established we claim that we can now choose  $\varepsilon$  small enough for the inverse image of  $\sigma(\varepsilon)$  in  $T$  to be contained within the union of the  $\eta$  spheres. (In fact, if this were not so, we could choose a sequence of points  $t_n$  in  $T$  not belonging to any  $\eta$  sphere, and a sequence  $\varepsilon_n$  tending to zero such that  $f(t_n)$  would belong to  $\sigma(\varepsilon_n)$  for each  $n$ . Since  $T$  is compact the sequence  $t_n$  would have a limit point  $t^*$  in  $T$ , for which we would have  $f(t^*) = 0$ . Since  $t^* \notin \partial T$  by (c), we must have  $t^* \in T$ . Thus  $t^*$  is contained in the inverse image of  $\sigma(\varepsilon)$  for any  $\varepsilon$ , as must be infinitely many of the  $t_n$ . This contradiction establishes our claim.)

Furthermore, by (b) and the inverse mapping theorem (cf. Footnote 6 of Chapter 3) we can choose  $\varepsilon, \eta$  so small that, for each  $\eta$  sphere in  $T$ ,  $\sigma(\varepsilon)$  is contained in the  $f$  image of the  $\eta$  sphere, so that the restriction of  $f$  to such a sphere will be one-one. Since the Jacobian of the mapping of each  $\eta$  sphere by  $f$  is  $|\det \nabla f(t)|$  it follows that we can choose  $\varepsilon$  small enough so that

$$N_0 = \int_T \delta_\varepsilon(f(t)) \mathbb{1}_B(g(t)) |\det \nabla f(t)| dt.$$

This follows since each  $\eta$  sphere in  $T$  over which  $g(t) \in B$  will contribute exactly one unit to the integral, while all points outside the  $\eta$  spheres will not be mapped onto  $\sigma(\varepsilon)$ . Since the left-hand side of this expression is independent of  $\varepsilon$  we can take the limit as  $\varepsilon \rightarrow 0$  to obtain (4.1.10) and thus the theorem.  $\square$

Theorem 4.1.3 does not tell us anything about expectations. Ideally, it would be nice simply to take expectations on both sides of (4.1.10) and then, hopefully, find an easy way to evaluate the right-hand side of the resulting equation. While this requires justification and further assumptions, let us nevertheless proceed in this fashion, just to see what happens. We then

have

$$\begin{aligned}\mathbb{E}\{N_0\} &= \lim_{\varepsilon \rightarrow 0} \mathbb{E} \int_T \delta_\varepsilon(f(t)) \mathbb{1}_B(g(t)) |\det \nabla f(t)| dt \\ &= \int_T \int_{\mathbb{R}^{N(N+1)/2}} \int_{\mathbb{R}^K} \mathbb{1}_B(v) |\det \nabla y| \\ &\quad \times \left\{ \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^N} \delta_\varepsilon(x) p_t(x, \nabla y, v) dx \right\} d\nabla y dv dt,\end{aligned}$$

where the  $p_t$  are the obvious densities. Taking the limit in the innermost integral yields

$$\begin{aligned}(4.1.11) \quad \mathbb{E}\{N_0\} &= \int_T \int \int \mathbb{1}_B(v) |\det \nabla y| p_t(0, \nabla y, v) d\nabla y dv dt \\ &= \int_T \mathbb{E}\{|\det \nabla f(t)| \mathbb{1}_B(g(t)) \mid f(t) = 0\} p_t(0) dt.\end{aligned}$$

Of course, interchanging the order of integration and the limiting procedure requires justification. Nevertheless, at this point we can state the following tongue-in-cheek ‘corollary’ to Theorem 4.1.3.

**Corollary 4.1.4** *If the conditions of Theorem 4.1.3 hold, as well as ‘adequate’ regularity conditions, then*

$$\begin{aligned}(4.1.12) \quad E\{N_u\} &= \int_T \int_{\mathbb{R}^{N(N+1)/2}} \int_{\mathbb{R}^K} |\det \nabla x| \mathbb{1}_B(v) p_t(u, \nabla x, v) d\nabla x dv dt \\ &= \int_T \mathbb{E}\{|\det \nabla f(t)| \mathbb{1}_B(g(t)) \mid f(t) = u\} p_t(u) dt.\end{aligned}$$

As will be seen below, ‘adequate’ regularity conditions generally require no more than that the density  $p_t$  above be well behaved (i.e. continuous, bounded) and that enough continuous derivatives of  $f$  and  $g$  exist, with enough finite moments.

At this point we suggest that the reader who cares little about rigour move directly to Section 4.5, where we begin the preparations for using the above results. Indeed, there is probably much to be said in favour of even the mathematically inclined reader doing the same on his/her first reading.

We now turn to the rigorous upper bound, and start with a useful Lemma. It is easily proven via induction started from Hölder’s inequality.

**Lemma 4.1.5** *Let  $X_1, \dots, X_n$  be any real valued random variables. Then*

$$(4.1.13) \quad \mathbb{E}\{|X_1 \cdots X_n|\} \leq \prod_{i=1}^n [\mathbb{E}\{|X_i|^n\}]^{1/n}.$$

**Theorem 4.1.6** *Let  $f, g, B$  and  $T$  be as in Theorem 4.1.3, but with  $f$  and  $g$  random and conditions (a)–(d) there holding in an almost sure sense. Assume, furthermore, that Conditions (b)–(e) of Theorem 4.1.1 hold, with the notation adopted there. Then*

(4.1.14)

$$\mathbb{E}\{N_u(f, g; T, B)\} \leq \int_T dt \int_B dv \int_{\mathbb{R}^{N(N+1)/2}} |\det \nabla y| p_t(u, \nabla y, v) d\nabla y.$$

**Proof.** Again assume that  $u = 0$ . Start by setting, for  $\varepsilon > 0$ ,

$$N^\varepsilon = \int_T \delta_\varepsilon(f(t)) \mathbb{1}_B(g(t)) |\det \nabla f(t)| dt,$$

where  $\delta_\varepsilon$  is as in (4.1.9). By expanding the determinant, applying (4.1.13) and recalling the moment assumption (4.1.2), we have that everything is nicely finite, and so Fubini's Theorem gives us that

$$\begin{aligned} \mathbb{E}\{N^\varepsilon(T)\} &= \int_T dt \int_{\mathbb{R}^N \times \mathbb{R}^{N(N+1)/2} \times B} \delta_\varepsilon(x) |\det \nabla y| p_t(x, \nabla y, v) dx d\nabla y dv \\ &= \int_T dt \int_{\mathbb{R}^{N(N+1)/2} \times B} |\det \nabla y| p_t(\nabla y, v) d\nabla y dv \\ &\quad \times \int_{\mathbb{R}^N} \delta_\varepsilon(x) p_t(x | \nabla y, v) dx \end{aligned}$$

Since all densities are assumed continuous and bounded, the innermost integral clearly converges to

$$p_t(0 | \nabla y, v)$$

as  $\varepsilon \rightarrow 0$ . Furthermore,

$$\begin{aligned} \int_{\mathbb{R}^N} \delta_\varepsilon(x) p_t(x | \nabla y, v) dx &\leq \sup_x p_t(x | \nabla y, v) \int_{\mathbb{R}^N} \delta_\varepsilon(x) dx \\ &= \sup_x p_t(x | \nabla y, v), \end{aligned}$$

which, by assumption, is bounded.

Again noting the boundedness of  $\mathbb{E}\{|\det \nabla f_t|\}$ , it follows from (4.1.10), dominated convergence and Fatou's Lemma, that

$$\begin{aligned} E\{N_0\} &\leq \lim_{\varepsilon \rightarrow 0} E\{N^\varepsilon\} \\ &= \int_T dt \int_B dv \int_{\mathbb{R}^{N(N+1)/2}} |\det \nabla y| p_t(0, \nabla y, v) d\nabla y. \end{aligned}$$

This, of course, proves the lemma.  $\square$

We can now turn to the more difficult part of our problem: showing that the upper bound for  $\mathbb{E}\{N(T)\}$  obtained in the preceding Theorem also serves as a lower bound under reasonable conditions. We shall derive the following result.

**Theorem 4.1.7** *Assume now the setup and assumptions of Theorem 4.1.6, along with (4.1.3) of Theorem 4.1.1; viz. The moduli of continuity of each of the components of  $f$ ,  $\nabla f$ , and  $g$  satisfy*

$$(4.1.15) \quad \mathbb{P}\{\omega(\eta) > \varepsilon\} = o(\eta^N), \quad \text{as } h \downarrow 0,$$

for all  $\varepsilon > 0$ . Then (4.1.14) holds with the inequality sign reversed, and so is an equality.

Since the proof of this Theorem is rather involved, we shall start by first describing the principles underlying it. Essentially, the proof is based on constructing a pathwise approximation to the vector-valued process  $f$  and then studying the zeros of the approximating process. The approximation is based on partitioning  $T$  and replacing  $f$  within each cell of the partition by a hyperplane tangential to  $f$  at the cell's midpoint. We then argue that if the approximating process has a zero within a certain subset of a given cell then  $f$  has a zero somewhere in the full cell. Thus the number of zeros of the approximating process will give a lower bound to the number of zeros of  $f$ .

In one dimension, for example, we replace the real valued function  $f$  on  $T = [0, 1]$  by a series of approximations  $f^{(n)}$  given by

$$f^{(n)}(t) = f\left(\left(j + \frac{1}{2}\right)2^{-n}\right) + \left[t - \left(j + \frac{1}{2}\right)2^{-n}\right] f'\left(\left(j + \frac{1}{2}\right)2^{-n}\right),$$

for  $j2^{-n} \leq t < (j+1)2^{-n}$ , and study the zeros of  $f^{(n)}$  as  $n \rightarrow \infty$ . Although this is perhaps not the most natural approximation to use in one dimension, it generalizes easily to higher dimensions.

**Proof of Theorem 4.1.7** As usual, we take the level  $u = 0$ , and start with some notation. For each  $n \geq 1$  let  $\mathbb{Z}_n$  denote the lattice of points in  $\mathbb{R}^N$  whose components are integer multiples of  $2^{-n}$ ; i.e.

$$\mathbb{Z}_n = \{t \in \mathbb{R}^N : t_j = i2^{-n}, j = 1, \dots, N, t \in \mathbb{Z}\}.$$

Now fix  $\varepsilon > 0$ , and for each  $n \geq 1$  define two half-open hypercubes centred on an arbitrary point  $t \in \mathbb{R}^N$  by

$$\begin{aligned} \Delta_n(t) &= \{s \in \mathbb{R}^N : -2^{-(n+1)} \leq s_j - t_j < 2^{-(n+1)}\}, \\ \Delta_n^\varepsilon(t) &= \{s \in \mathbb{R}^N : -(1-\varepsilon)2^{-(n+1)} \leq s_j - t_j < (1-\varepsilon)2^{-(n+1)}\}. \end{aligned}$$

Set

$$I_{nt} = \begin{cases} 1, & \text{if } N_0(f, g; \Delta_n(t), B) \geq 1 \text{ and } \Delta_n(t) \subset T, \\ 0, & \text{otherwise,} \end{cases}$$

and define approximations  $N^n$  to  $N_0(f, g; T, B)$  by

$$N^n = \sum_{t \in \mathbb{Z}_n} I_{nt}.$$

Note (since it will be important later) that only those  $\Delta_n(t)$  which lie wholly within  $T$  contribute to the approximations. However, since the points being counted are, by assumption, almost surely isolated, and none lie on  $\partial T$ , it follows that

$$N^n \xrightarrow{a.s.} N_0(f, g; T, B) \quad \text{as } n \rightarrow \infty.$$

Since the sequence  $N^n$  is non-decreasing in  $n$ , monotone convergence yields

$$(4.1.16) \quad \begin{aligned} \mathbb{E}\{N_0\} &= \lim_{n \rightarrow \infty} \mathbb{E}\{N_n\} \\ &\geq \lim_{n \rightarrow \infty} \sum \mathbb{P}\{N(\Delta_n(t)) > 0\}, \end{aligned}$$

where the summation is over all  $t \in \mathbb{Z}_n$  for which  $\Delta_n(t) \subset T$ .

The remainder of the proof involves determining the limiting value of  $\mathbb{P}\{N(\Delta_n(t)) > 0\}$ . Fix  $\delta > 0$  small, and  $K > 0$  large. For given realizations of  $f$  and  $g$  define the function  $\omega_f^*(n)$  by

$$\omega_f^*(n) = \max \left[ \max_i \omega_{f^j}(2^{-n}), \max_{ij} \omega_{(\nabla f)_{ij}}(2^{-n}), \max_i \omega_{g^j}(2^{-n}) \right]$$

where the moduli of continuity are all taken over  $T$ . Furthermore, define  $M_f$  by

$$M_f = \max \left[ \max_{1 \leq j \leq N} \sup_{t \in T} |f^j(t)|, \max_{1 \leq i, j \leq N} \sup_{t \in T} |(\partial f^i / \partial t_j)(t)| \right].$$

Finally, set

$$\eta = \frac{\delta^2 \varepsilon}{2N!(K+1)^{N-1}}.$$

Then the conditions of the Theorem imply that, as  $n \rightarrow \infty$ ,

$$(4.1.17) \quad \mathbb{P}\{\omega^*(n) > \eta\} = o(2^{-Nn}).$$

Choose now a fixed  $n$  and  $t \in \mathbb{Z}_n$  for which  $\Delta_n(t) \subset T$ . Assume that

$$(4.1.18) \quad \omega_X^*(n) < \eta$$

and that the event  $G_{\delta K}$  occurs, where

$$G_{\delta K} = \left\{ |\det \nabla f(t)| > \delta, M_f < K/N, \right. \\ \left. g(t) \in \{x \in B : \inf_{y \in \partial B} \|x - y\|_1 > \delta\} \right\},$$

where  $\|x\|_1 = \sum_i |x^i|$  is the usual  $\ell^1$  norm. Finally, define  $t^*$  to be the solution of the following equation, when a unique solution in fact exists:

$$(4.1.19) \quad f(t) = (t - t^*) \cdot \nabla f(t).$$

We claim that if both  $\omega^*(n) < \eta$  and  $G_{\delta K}$  occurs, then, for  $n$  large enough,  $t^* \in \Delta_n^\varepsilon(t)$  implies

$$(4.1.20) \quad N_0(\Delta_n(t)) > 0$$

and

$$(4.1.21) \quad \|f(t)\|_1 \leq 2^{-n} K.$$

These two facts, which we shall establish in a moment, are enough to make the remainder of the proof quite simple. From (4.1.20), it follows that by choosing  $n$  large enough for (4.1.17) to be satisfied we have

$$\mathbb{P}\{N(\Delta_n(t)) > 0\} \geq \mathbb{P}\{G_{\delta K} \cap [t^* \in \Delta_n^\varepsilon(t)]\} + o(2^{-nN}).$$

Using this, making the transformation  $f(t) \rightarrow t^*$  given by (4.1.19), and using the notation of the Theorem, we obtain

$$\begin{aligned} & \mathbb{P}\{N_0(\Delta_n(t)) > 0\} \\ & \geq \int_{G_{\delta K} \cap \{t^* \in \Delta_n^\varepsilon(t)\}} |\det \nabla y| p_t((t - t^*) \nabla y, \nabla y, v) dt^* d\nabla y dv + o(2^{-nN}). \end{aligned}$$

Noting (4.1.21), the continuity and boundedness assumptions on  $p_t$ , and the boundedness assumptions on the moments of the  $f_j^i$ , it follows that, as  $n \rightarrow \infty$ , the last expression, summed as in (4.1.16), converges to

$$(1 - \varepsilon)^N \int_T dt \int_{G_{\delta K}} |\det \nabla y| p_t(0, \nabla y, v) d\nabla y dv.$$

Letting  $\varepsilon \rightarrow 0$ ,  $\delta \rightarrow 0$ ,  $K \rightarrow \infty$  and applying monotone convergence to the above expression we obtain, from (4.1.16), that

$$E\{N(T)\} \geq \int_T dt \int_B dv \int_{\mathbb{R}^{N(N+1)/2}} |\det \nabla y| p_t(0, \nabla y, v) d\nabla y.$$

This, of course, completes the proof, bar the issue of establishing that (4.1.20) and (4.1.21) hold under the conditions preceeding them.

Thus, assume that  $\omega^*(n) < \eta$ ,  $G_{\delta K}$  occurs, and the  $t^*$  defined by (4.1.19) satisfies  $t^* \in \Delta_n^\varepsilon(t)$ . Then (4.1.21) is immediate. The hard part is to establish (4.1.20). To this end, note that (4.1.19) can be rewritten as

$$(4.1.22) \quad t - f(t)[\nabla f(t)]^{-1} \in \Delta_n^\varepsilon(t).$$

Let  $\tau$  be any other point  $\Delta_n(t)$ . It is easy to check that

$$(4.1.23) \quad |\det \nabla f(\tau) - \det \nabla f(t)| < \frac{\varepsilon \delta^2}{2},$$

under the conditions we require. Thus, since  $\det \nabla f(t) > \delta$ , it follows that  $\det \nabla f(\tau) \neq 0$  for any  $\tau \in \Delta_n(t)$  and so the matrix  $\det \nabla f(\tau)$  is invertible throughout  $\Delta_n(t)$ . Similarly, one can check that  $g(\tau) \in B$  for all  $\tau \in \Delta_n(t)$ .

Consider first (4.1.20). We need to show that  $t^* \in \Delta_n^\varepsilon(t)$  implies the existence of at least one  $\tau \in \Delta_n(t)$  at which  $f(\tau) = 0$ .

The mean value theorem<sup>6</sup> allows us to write

$$(4.1.25) \quad f(\tau) - f(t) = (\tau - t) \cdot \nabla f(t^1, \dots, t^N)$$

for some points  $t^1, \dots, t^N$  lying on the line segment  $L(t, \tau)$ , for any  $t$  and  $\tau$ , where  $\nabla f(t^1, \dots, t^N)$  is the matrix function  $\nabla f$  with the elements in the  $k$ -th column evaluated at the point  $t^k$ . Using similar arguments to those used to establish (4.1.23) and the invertibility of  $\det \nabla f(\tau)$  throughout  $\Delta_n(t)$ , invertibility can be shown for  $\nabla f(t^1, \dots, t^N)$  as well. Hence we can rewrite (4.1.25) as

$$(4.1.26) \quad f(\tau)[\nabla f(t^1, \dots, t^N)]^{-1} = f(t)[\nabla f(t^1, \dots, t^N)]^{-1} + (\tau - t).$$

Suppose we could show  $f(t)[\nabla f(t^1, \dots, t^N)]^{-1} \in \Delta_n(0)$  if  $t^* \in \Delta_n^\varepsilon(t)$ . Then by the Brouwer fixed point theorem<sup>7</sup> it would follow that the continuous mapping of  $\Delta_n(t)$  into  $\Delta_n(t)$  given by

$$\tau \rightarrow t - f(t)[\nabla f(t^1, \dots, t^N)]^{-1}, \quad \tau \in \Delta_n(t),$$

---

<sup>6</sup>In the form we need it, here is the mean value theorem for Euclidean spaces. A proof can be found, for example, in [8].

**Lemma 4.1.8** (*Mean Value Theorem for  $\mathbb{R}^N$* .) *Let  $T$  be a bounded open set in  $\mathbb{R}^N$  and let  $f : T \rightarrow \mathbb{R}^N$  have first-order partial derivatives at each point in  $T$ . Let  $s$  and  $t$  be two points in  $T$  such that the line segment*

$$L(s, t) = \{u : u = \theta s + (1 - \theta)t, 0 < \theta < 1\}.$$

*is wholly contained in  $T$ . Then there exist points  $t^1, \dots, t^N$  on  $L(s, t)$  such that*

$$(4.1.24) \quad f(t) - f(s) = \nabla f(t^1, \dots, t^N) \cdot (t - s),$$

*where by  $\nabla f(t^1, \dots, t^N)$  we mean the matrix valued function  $\nabla f$  with the elements in the  $k$ -th column evaluated at the point  $t^k$ .*

<sup>7</sup>In the form we need it, the Brouwer fixed point theorem is as follows. Proofs of this result are easy to find (e.g. [89]).

**Lemma 4.1.9** *Let  $T$  be a compact, convex subset of  $\mathbb{R}^N$  and  $f : T \rightarrow T$  a continuous mapping of  $T$  into itself. Then  $f$  has at least one fixed point; i.e. there exists at least one point  $t \in T$  for which  $f(t) = t$ .*

has at least one fixed point. Thus, by (4.1.26), there would be at least one  $\tau \in \Delta_n(t)$  for which  $f(\tau) = 0$ . In other words,

$$\{\omega_X^*(n) < \eta, G_{\delta K}, t^* \in \Delta_n^\varepsilon(t), n \text{ large}\} \Rightarrow N(T \cap \Delta_n(t)) > 0.$$

But  $f(t)[\nabla f(t^1, \dots, t^N)]^{-1} \in \Delta_n(0)$  is easily seen to be a consequence of  $t^* \in \Delta_n^\varepsilon(t)$  and  $G_{\delta K}$  simply by writing

$$\begin{aligned} & f(t)[\nabla f(t^1, \dots, t^N)]^{-1} \\ &= f(t)[\nabla f(t)]^{-1} \nabla f(t) [\nabla f(t^1, \dots, t^N)]^{-1} \\ &= f(t)[\nabla f(t)]^{-1} (I + (\nabla f(t) - \nabla f(t^1, \dots, t^N)) [\nabla f(t^1, \dots, t^N)]^{-1}) \end{aligned}$$

noting (4.1.22) and bounding the rightmost expression using basically the same argument employed for (4.1.23). This completes the proof.  $\square$

We now complete the task of this Section – i.e. the proof of Theorem 4.1.1 – which gave the expression appearing in both Theorems 4.1.6 and 4.1.7 for  $\mathbb{E}\{N_u\}$ , but under seemingly weaker conditions than those we have assumed. What remains to show is that Conditions (b)–(d) of Theorem 4.1.3 are satisfied under the conditions of Theorem 4.1.1. Condition (c) follows immediately from the following rather intuitive result, taking  $h = f$ ,  $n = N - 1$ , and identifying the  $T$  of the Lemma with  $\partial T$  in the Theorems. The claim in (b) relating the points satisfying  $f(t) = 0$  and  $g(t) \in \partial B$  follows by taking  $h = (f, h_B(g))$ , where  $h_B$  is the function whose excursion set defines  $B$  (cf. (4.1.1)). Lemma 4.1.10 has roots going back to Bulinskaya [17].

**Lemma 4.1.10** *Let  $T$  be a compact set of Hausdorff dimension  $n$ . Let  $h : T \rightarrow \mathbb{R}^{n+1}$  be  $C^1$ , with a.s. bounded partial derivatives. Furthermore, assume that the univariate probability densities of  $h$  are bounded in a neighbourhood of  $u \in \mathbb{R}^{n+1}$ , uniformly in  $t$ . Then, for such  $u$ , there are a.s. no  $t \in T$  with  $h(t) = u$ . In the notation of inverses and excursion sets:*

$$(4.1.27) \quad \mathbb{P}\{h^{-1}(u) = \emptyset\} = \mathbb{P}\{\partial A_u(h, T) = \emptyset\} = 1.$$

**Proof.** We start with the observation that under the conditions of the Lemma, for any  $\varepsilon > 0$ , there exists a finite  $C_\varepsilon$  so that

$$\mathbb{P}\{\max_{ij} \sup_T \partial h^i(t) / \partial t_j > C_\varepsilon\} > 1 - \varepsilon.$$

Writing  $\omega_h$  for the modulus of continuity of  $h$ , it follows from the Mean Value Theorem that

$$(4.1.28) \quad \mathbb{P}\{E_\varepsilon\} > 1 - \varepsilon,$$

where

$$E_\varepsilon \triangleq \{\omega_h(\eta) \leq \sqrt{n}C_\varepsilon\eta, \text{ for small enough } \eta\}.$$

Take now a sequence  $\{\delta_m\}$  of positive numbers converging to 0 as  $m \rightarrow \infty$ . The fact that  $T$  has Hausdorff dimension  $n$  implies that for any any  $\delta > 0$ , and  $m$  large enough, there exists a collection of Euclidean balls  $\{B_{m_j}\}$  covering  $T$  such that

$$(4.1.29) \quad \sum_j (\text{diam}(B_{m_j}))^{n+\delta} < \delta_m.$$

Define the events

$$A = \{\exists t \in T : h(t) = u\}, \quad A_{m_j} = \{\exists t \in B_{m_j} : h(t) = u\}.$$

Fix  $\varepsilon > 0$  and note that

$$(4.1.30) \quad \mathbb{P}\{A\} \leq \sum_j \mathbb{P}\{A_{m_j} \cap E_\varepsilon\} + \mathbb{P}\{E_\varepsilon^c\}.$$

In view of (4.1.28) it suffices to show that the sum here can be made arbitrarily small.

To see this, let  $t_{m_j}$  be the centrepoint of  $B_{m_j}$ . Then, if both  $A_{m_j}$  and  $E_\varepsilon$  occur, it follows that, for large enough  $m$ ,

$$|h(t_{m_j}) - u| \leq \sqrt{n}C_\varepsilon \text{diam}(B_{m_j}).$$

Since  $h_t$  has a bounded density, it thus follows that

$$\mathbb{P}\{A_{m_j} \cap E_\varepsilon\} \leq M(\sqrt{n}C_\varepsilon \text{diam}(B_{m_j}))^{n+1},$$

where  $M$  is a bound on the densities. Substituting into (4.1.30) and noting (4.1.29) we are done.  $\square$

We now turn to the second part of Condition (b) of Theorem 4.1.3, relating to the points  $t \in T$  satisfying  $f(t) - u = \det \nabla f(t) = 0$ . Note firstly that this would follow easily from Lemma 4.1.10 if we were prepared to assume that  $f \in C^3(T)$ . This, however, is more than we are prepared to assume. That the conditions of Theorem 4.1.1 contain all we need is the content of the following Lemma.

**Lemma 4.1.11** *Let  $f$  and  $T$  be as in Theorem 4.1.1, with Conditions (a), (b), (d) and (f) of that Theorem in force. Then, with probability one, there are no points  $t \in T$  satisfying  $f(t) - u = \det \nabla f(t) = 0$ .*

**Proof.** As for the proof of the previous Lemma, we start with an observation. In particular, for any  $\varepsilon > 0$ , there exists a continuous function  $\omega_\varepsilon$  for which  $\omega_\varepsilon(\eta) \downarrow 0$  as  $\eta \downarrow 0$ , and a finite positive constant  $C_\varepsilon$ , such that  $\mathbb{P}\{E_\varepsilon\} > 1 - \varepsilon$ , where now

$$E_\varepsilon \triangleq \left\{ \max_{i,j} \sup_{t \in T} |f_j^i(t)| < C_\varepsilon, \max_i \omega_{f_j^i}(\eta) \leq \omega_\varepsilon(\eta), \text{ for } 0 < \eta \leq 1 \right\}.$$

To see this, the following simple argument<sup>8</sup> suffices: For ease of notation, set

$$\omega^*(\eta) = \max_{i,j} \omega_{f_j^i}(\eta).$$

It then follows from (4.1.3) that there are sequences  $\{c_n\}$  and  $\{\eta_n\}$ , both decreasing to zero, such that

$$\mathbb{P}\{\omega^*(\eta_n) < c_n\} > 1 - 2^{-n}\varepsilon.$$

Defining  $\omega_\varepsilon(\eta) = c_n$  for  $\eta_{n+1} \leq \eta < \eta_n$ , Borel-Cantelli then gives, for some  $\eta_1 > 0$ , that

$$\mathbb{P}\{\omega^*(\eta) < \omega_\varepsilon(\eta), 0 < \eta \leq \eta_1\} > 1 - \varepsilon/2.$$

If  $\eta_1 < 1$ , set  $\omega_\varepsilon = c_0$  for  $\eta_1 < \eta \leq 1$ , where  $c_0$  is large enough so that

$$\mathbb{P}\{\omega^*(\eta) < c_0, \eta_1 < \eta \leq 1\} > 1 - \varepsilon/2.$$

Now choose  $C_\varepsilon$  large enough so that

$$\mathbb{P}\left\{\sup_{t \in T} |f(t)| < C_\varepsilon\right\} > 1 - \varepsilon/2,$$

and we are done.

We now continue in a similar vein to that of the previous proof. Let the  $B_{m_j}$  and  $t_{m_j}$  be as defined there, although (4.1.29) is now replaced with

$$(4.1.31) \quad \sum_j (\text{diam} B_{m_j})^N \rightarrow c_N \lambda_N(T) < \infty$$

as  $m \rightarrow \infty$  for some dimension dependent constant  $c_N$ .

Define the events

$$A_{m_j} = \{\exists t \in B_{m_j} : f(t) - u = \det \nabla f(t) = 0\}.$$

Fixing  $\varepsilon > 0$ , as before, we now need only show that  $\sum_j \mathbb{P}\{A_{m_j} \cap E_\varepsilon\}$  can be made arbitrarily small.

Allowing  $C_N$  to be a dimension dependent constant that may change from line to line, if  $A_{m_j}$  and  $E_\varepsilon$  occur, then, by expanding the determinant, it is easy to see that

$$|\det \nabla f(t_{mk})| \leq C_N (\max(1, C_\varepsilon))^{N-1} \omega_\varepsilon(\sqrt{N} \text{diam}(B_{m_j})),$$

---

<sup>8</sup> Note that this event tells us less about the  $\omega_{f_j^i}$  than in the corresponding event in the previous Lemma, and so we shall have to work harder to show that  $\mathbb{P}\{E_\varepsilon\} > 1 - \varepsilon$ . The problem is that, this time, we are not assuming that the  $f_j^i$ , unlike the  $h^i$  of the previous Lemma, are  $C^1$ .

and that, as before,

$$|f(t_{m_j}) - u| \leq C_N C_\varepsilon \text{diam}(B_{m_j}).$$

It therefore follows that

$$\begin{aligned} & \mathbb{P}\{A_{m_j} \cap E_\varepsilon\} \\ & \leq \int \mathbb{P}\left\{|\det \nabla f(t_{m_j})| \leq (C_N (\max(1, C_\varepsilon))^{N-1} \omega_\varepsilon(\sqrt{N} \text{diam}(B_{m_j}))) \right. \\ & \quad \left. | f(t_{m_j}) = x \right\} p_{t_{m_j}}(x) dx, \end{aligned}$$

where  $p_t$  is the density of  $f(t)$  and the integral is over a region in  $\mathbb{R}^N$  of volume no greater than  $C_N \text{diam}(B_{m_j})^N$ .

By assumption (f) of Theorem 4.1.1 the integrand here tends to zero as  $m \rightarrow \infty$ , and so can be made arbitrarily small. Furthermore,  $p_t$  is uniformly bounded. Putting all of this together with (4.1.31) proves the result.  $\square$

We have one more result to establish before completing this Section and the proof of Theorem 4.1.1 – that Condition (d) of Theorem 4.1.3 holds. This is an immediate consequence of the following Lemma.

**Lemma 4.1.12** *Let the assumptions of Lemma 4.1.11 hold. Then, with probability one, there are only a finite number of points  $t \in T$  satisfying  $f(t) = u$ .*

**Proof.** Given the *conclusion* of Lemma 4.1.11, this is actually a reasonably standard deterministic result related to the Morse Theory of Chapter 3 (cf. the proof of Lemma 3.3.3). Thus we give only an outline of the proof.

Suppose that  $t$  is such that  $f(t) = u$ . (If there are no such  $t$ , there is nothing more to prove.) The fact that  $f \in C^1$  implies that there is a neighbourhood of  $t$  in which  $f$  is locally linear; *viz* it can be approximated by its tangent plane. Furthermore, since  $\det \nabla f(t) \neq 0$  (Lemma 4.1.11) not all the partial derivatives can be zero, and so the tangent plane cannot be at a constant ‘height’. Consequently, throughout this neighbourhood there is no other point at which  $f = u$ . Since  $T$  is compact, there can therefore be no more than a finite number of  $t$  satisfying  $f(t) = u$ , and we are done.  $\square$

## 4.2 Suitable regularity and Morse functions

Back in Chapter 3 we laid down a number of regularity conditions on deterministic functions  $f$  for the geometric analyses developed there to be valid. In the Integral Geometric setting they were summarised under the

‘suitable regularity’ of Definition 3.3.1. In the manifold setting, we moved to the ‘Morse functions’ of Section 3.9.

We shall soon (in Section 4.6) need to know when a random  $f$  is, with probability one, either suitably regular or a Morse function over a finite rectangle  $T \subset \mathbb{R}^N$  and later (in Section 4.10) when it is a Morse function over a piecewise smooth manifold.

Despite the fact that this is (logically) not the right place to be addressing these issues, we shall nevertheless do so, while the details of the previous Section are still fresh in your mind. In particular, without much further effort, Lemmas 4.1.10 and 4.1.11 will give us what we need.

We consider Morse functions over rectangles first, since here the conditions are tidiest. We start with a reminder (and slight extension) of notation: As usual,  $\partial_k T$  is the  $k$ -dimensional boundary, or skeleton, of  $T$ . Since  $T$  is a rectangle, each  $\partial_k T$  is made up of  $2^{N-k} \binom{N}{k}$  open rectangles, or ‘faces’, of dimension  $k$ .

Morse functions can then be defined via the following three characteristics:

- (i)  $f$  is  $C^2$  on an open neighbourhood of  $T$ .
- (ii) The critical points of  $f|_{\partial_k T}$  are non-degenerate for all  $k = 0, \dots, N$ .
- (iii)  $f|_{\partial_k T}$  has no critical points on  $\bigcup_{j=0}^{k-1} \partial_j T$  for all  $k = 1, \dots, N$ .

Recall that, in the current Euclidean setting, critical points are those  $t$  for which  $\nabla f(t) = 0$ , and ‘non-degeneracy’ means that the Hessian  $\nabla^2 f(t)$  has non-zero determinant. Note also that in both conditions (ii) and (iii) there is a strong dependence on dimension. In particular, in (ii) the requirement is that the  $\mathbb{R}^{k+1}$ -valued function  $(\nabla(f|_{\partial_k T}), \det \nabla^2(f|_{\partial_k T}))$  not have zeroes over a  $k$ -dimensional parameter set. Regarding (iii) the requirement is that the  $\mathbb{R}^k$ -valued function  $\nabla(f|_{\partial_k T})$  defined on a  $k$ -dimensional set not have zeroes on a subset of dimension  $k - 1$ .

In this light (ii) and (iii) are clearly related to Lemma 4.1.11 and we leave it to you to check the details that give us

**Theorem 4.2.1** *Let  $f$  be a real valued random field over a bounded rectangle  $T$  of  $\mathbb{R}^N$  with first and second order partial derivatives  $f_i$  and  $f_{ij}$ . Then  $f$  is, with probability one, a Morse function over  $T$  if the following conditions hold for each face  $J$  of  $T$ :*

- (a)  $f$  is, with probability one,  $C^2$  on an open neighbourhood of  $T$ , and all second derivatives have finite variance.
- (b) For all  $t \in J$ , the marginal densities  $p_t(x)$  of  $\nabla f|_J(t)$  are continuous at 0, uniformly in  $t$ .
- (c) The conditional densities  $p_t(z|x)$  of  $\det \nabla^2 f|_J(t)$  given  $\nabla f|_J(t) = x$  are continuous for  $(z, x)$  in a neighbourhood of 0, uniformly in  $t \in T$ .

(d) On  $J$ , the moduli of continuity of  $f$  and its first and second order partial derivatives all satisfy

$$\mathbb{P}\{\omega(\eta) > \varepsilon\} = o\left(\eta^{\dim(J)}\right), \quad \text{as } \eta \downarrow 0.$$

As usual, we also have a Gaussian Corollary (cf. Corollary 4.1.2). It is even simpler than usual, since Gaussianity allows us drop the references to the specific faces  $J$  that so cluttered the conditions of the Theorem.

**Corollary 4.2.2** *Let  $f$  be a centered Gaussian field over a finite rectangle  $T$ . If, for each  $t \in T$ , the joint distributions of  $(f_i(t), f_{ij}(t))_{i,j=1,\dots,N}$  are non-degenerate, and if, for some finite  $K$ ,*

$$(4.2.1) \quad \sup_{s,t \in T} \max_{i,j} |C_{f_{ij}}(t) - C_{f_{ij}}(s)| \leq K |\ln|t - s||^{-(1+\alpha)},$$

*then the sample functions of  $f$  are, with probability one, Morse functions over  $T$ .*

The next issue is to determine when sample functions are, with probability one, suitably regular in the sense of Definition 3.3.1. This is somewhat less elegant because of the asymmetry in the conditions of suitable regularity and their dependence on a particular coordinate system. Nevertheless, the same arguments as above work here as well and it is easy (albeit a little time consuming) to see that the following suffices to do the job.

**Theorem 4.2.3** *Under the conditions of Corollary 4.2.2 the sample functions of  $f$  are, with probability one, suitably regular over bounded rectangles.*

A little thought will show that the assumptions of this Theorem would seem to give more than is required. Consider the case  $N = 2$ . In that case, Condition (3.3.6) of suitable regularity requires that there are no  $t \in T$  for which  $f(t) - u = f_1(t) = f_{11}(t) = 0$ . This is clearly implied by Theorem 4.2.3. However, the Theorem also implies that  $f(t) - u = f_2(t) = f_{22}(t) = 0$  which is not something which we require. Rather, it is a consequence of a desire to write the conditions of the Theorem in a compact form.

In fact, Theorem 4.2.3 goes even further, in that it implies that for every *fixed* choice of coordinate system, the sample functions of  $f$  are, with probability one, suitably regular over bounded rectangles<sup>9</sup>.

We now turn to what is really the most important case, that of determining sufficient conditions for a random field to be, almost surely, a Morse function over a piecewise  $C^2$  Riemannian manifold  $(M, g)$ . Writing

---

<sup>9</sup>Recall that throughout our discussion of Integral Geometry there was a fixed coordinate system and that suitable regularity was defined relative to this system.

our manifold as usual as

$$(4.2.2) \quad M = \bigcup_{j=0}^N \partial_j M,$$

(cf. (3.7.8)) Conditions (i)–(iii) still characterise whether or not  $f$  will be a Morse function. The problem is how to generalise Theorem 4.2.1 to this scenario, since its proof was based on the results of the previous Section, all of which were established in a Euclidean setting. The trick, of course, is to recall that each of the three required properties is of a local nature. We can then argue as follows:

Choose a chart  $(U, \varphi)$  from a countable atlas covering  $M$ . Let  $t^* \in U$  be a critical point of  $f$ , and note that this property is independent of the choice of local coordinates. Working therefore with the natural basis for  $T_t M$  it is easy to see that  $\varphi(t^*)$  is also a critical point of  $\varphi(U)$ . Furthermore, the covariant Hessian  $\nabla^2 f(t^*)$  will be zero if and only if the same is true of the regular Hessian of  $f \circ \varphi^{-1}$ , and since  $\varphi$  is a diffeomorphism this implies that  $t^*$  will be a degenerate critical point of  $f$  if and only if  $t^*$  is for  $f \circ \varphi^{-1}$ . It therefore follows that, even in the manifold case, we can manage, with purely Euclidean proofs, to establish the following result. The straightforward but sometimes messy details are left to you.

**Theorem 4.2.4** *Let  $f$  be a real valued random field over a piecewise  $C^2$ , compact, Riemannian submanifold  $(M, g)$  of a  $C^3$  manifold  $\widetilde{M}$ . Assume that  $M$  has a countable atlas. Then  $f$  is, with probability one, a Morse function over  $M$  if the following conditions hold for each submanifold  $\partial_j M$  in (4.2.2). Throughout  $\nabla$  and  $\nabla^2$  are to be understood in their Riemannian formulations and implicitly assumed to have the dimension of the  $\partial_j M$  where they are being applied.*

- (a)  $f$  is, with probability one,  $C^2$  on an open neighbourhood of  $M$  in  $\widetilde{M}$ , and  $\mathbb{E}\{(XYf)^2\} < \infty$  for  $X, Y \in T_t M$ ,  $t \in M$ .
- (b) For each  $\partial_j M$ , the marginal densities  $p_t(x)$  of  $\nabla f|_{\partial_j M}(t)$  are continuous at 0, uniformly in  $t$ .
- (c) The densities  $p_t(z|x)$  of  $\text{Tr}^{\partial_j M} \nabla^2 f|_{\partial_j M}(t)$  given  $\nabla f|_{\partial_j M}(t) = x$  are continuous for  $(z, x)$  in a neighbourhood of 0, uniformly in  $t \in M$ ,
- (d) On  $\partial_j M$ , the modulus of continuity of  $\nabla^2 f|_{\partial_j M}(t)(X, Y)$  satisfies

$$(4.2.3) \quad \mathbb{P}\{\omega(\eta) > \varepsilon\} = o\left(\eta^{\dim(\partial_j M)}\right), \quad \text{as } \eta \downarrow 0,$$

for all  $X, Y \in S(M) \cap T_t \partial_j M$ , where  $S(M)$  is the sphere bundle of  $M$  and the modulus of continuity is taken with respect to the distance induced by the Riemannian metric  $g$ .

As usual, there is a Gaussian Corollary to Theorem 4.2.4 which requires only non-degeneracies and Condition (d). There are two ways that we could go about attacking Condition (d). The more geometric of the two would be to return to the entropy conditions of Section 2.1 and find a natural entropy condition for (4.2.3) to hold. This, however, involves adding a notion of ‘canonical distance’ (under the canonical metric  $d$  on  $M$ ) to the already existing distance corresponding to the Riemannian metric induced by  $f$ . The details of carrying this out, while not terribly difficult, would involve some work. Perhaps more importantly, the resulting conditions would not be in a form that would be easy to check in practice.

Thus, we take another route, given conditions that are less elegant but easier to establish and generally far easier to check in practice. As for Theorem 4.2.4 itself we leave it to you to check the details of the (straightforward) proof of the Corollary.

**Corollary 4.2.5** *Take the setup of Theorem 4.2.4 and let  $f$  be a centered Gaussian field over  $M$ . Let  $\mathcal{A} = (U_\alpha, \varphi_\alpha)_{\alpha \in I}$  be a countable atlas for  $M$  such that for every  $\alpha$  the Gaussian field  $f_\alpha = f \circ \varphi_\alpha^{-1}$  on  $\varphi_\alpha(U_\alpha) \subset \mathbb{R}^N$  satisfies the conditions of Corollary 4.2.2 with  $T = \varphi_\alpha(U_\alpha)$ ,  $f = f_\alpha$  and some  $K_\alpha > 0$ . Then the sample functions of  $f$  are, with probability one, Morse functions over  $M$ .*

### 4.3 An alternate proof of the meta-theorem

The proof of the ‘meta-theorem’ Theorem 4.1.1 given in the previous Section is but the latest tale in a long history.

The first proof of this kind is probably due to Rice [78] in 1945, who worked in the setting of real valued processes on the line. His proof was made rigorous by Itô [45] and Ylvisaker [112] in the mid 1960’s. Meanwhile, in 1957, Longuet-Higgins [62, 63] was the first to extend it to the setting of random fields and used it to compute the expectations of such variables as the mean number of local maxima of two- and three-dimensional Gaussian fields. Rigorous versions of Theorem 4.1.1, at various levels of generality and with various assumptions, appeared only in the early 1970’s as in [3, 6, 10, 64] with the proof of Section 4.1 being essentially that of [3].

Recently, however, Azais and Wschebor [9] have developed a new proof, based on Federer’s coarea formula, in the form (3.6.24). In the notation of the previous Section, this can be rewritten as

$$\int_{\mathbb{R}^N} \left( \sum_{t: f(t)=u} \alpha(t) \right) du = \int_{\mathbb{R}^N} \alpha(t) |\det \nabla f(t)| dt$$

assuming that  $f$  and  $\alpha : \mathbb{R}^N \rightarrow \mathbb{R}^N$  are sufficiently smooth.

Take  $\alpha(t) = \varphi(f(t))\mathbb{1}_T(t)$ , where  $\varphi$  is a smooth (test) function. (Of course,  $\alpha$  is now no longer smooth, but we shall ignore this for the moment.) The above then becomes

$$\int_{\mathbb{R}^N} \varphi(u) N_u(f : T) du = \int_T \varphi(f(t)) |\det \nabla f(t)| dt.$$

Now take expectations (assuming this is allowed) of both sides to obtain

$$\begin{aligned} \int_{\mathbb{R}^N} \varphi(u) \mathbb{E}\{N_u(f : T)\} du &= \int_T \mathbb{E}\{\varphi(f(t)) |\det \nabla f(t)|\} dt \\ &= \int_{\mathbb{R}^N} \varphi(u) \int_T \mathbb{E}\{|\det \nabla f(t)| | f(t) = u\} p_t(u) dt du. \end{aligned}$$

Since  $\varphi$  was arbitrary, this implies that for (Lebesgue) almost every  $u$ ,

$$(4.3.1) \quad \mathbb{E}\{N_u(f : T)\} = \int_T \mathbb{E}\{|\det \nabla f(t)| | f(t) = u\} p_t(u) dt,$$

which is precisely (4.1.5) of Theorem 4.1.1 with the  $g$  there identically equal to 1. Modulo this restriction on  $g$ , which is simple to remove, this is the result we have worked so hard to prove. The problem, however, is that since it is true only for almost every  $u$  one cannot be certain that it is true for a specific value of  $u$ .

To complete the proof, we need only show that both sides of (4.3.1) are continuous functions of  $u$  and that the assumptions of convenience made above are no more than that. This, of course, is not as trivial as it may sound. Going through the arguments actually leads to repeating many of the technical points we went through in the previous Section, and eventually Theorem 4.1.1 reappears with the same long list of conditions. However (and this is the big gain) the details have no need of the construction, in the proof of Theorem 4.1.7, of the linear approximation to  $f$ .

You can find all the details in [9] and decide for yourself which proof you prefer.

## 4.4 Higher moments

While not at all obvious at first sight, hidden away in Theorem 4.1.1 is another result, about higher moments of the random variable  $N_u(f, g : T, B)$ . To state it, we need, for integral  $k \geq 1$ , the  $k$ -th partial factorial of a real  $x$  defined by

$$(x)_k \triangleq x(x-1)\dots(x-k+1).$$

Then we have

**Theorem 4.4.1** *Let  $f, g, T$  and  $B$  be as in Theorem 4.1.1, and assume that Conditions (a), (e) and (f) there still hold. For  $k \geq 1$ , write*

$$\begin{aligned} T^k &= \{\tilde{t} = (t_1, \dots, t_k) : t_j \in T, \forall 1 \leq j \leq k\}, \\ \tilde{f}(\tilde{t}) &= (f(t_1), \dots, f(t_k)) : T^k \rightarrow \mathbb{R}^{Nk}, \\ \tilde{g}(\tilde{t}) &= (g(t_1), \dots, g(t_k)) : T^k \rightarrow \mathbb{R}^{Nk}. \end{aligned}$$

Replace the remaining assumptions of Theorem 4.1.1 by

- (b') For all  $\tilde{t} \in T^k$ , the marginal densities  $p_{\tilde{t}}$  of  $\tilde{f}(\tilde{t})$  are continuous at  $\tilde{u} \triangleq (u, \dots, u)$ .
- (c') The conditional densities of  $\tilde{f}(\tilde{t})$  given  $\tilde{g}(\tilde{t})$  and  $\nabla \tilde{f}(\tilde{t})$  are bounded above and continuous at  $\tilde{u}$ , uniformly in  $\tilde{t} \in T^k$ .
- (d') The conditional densities  $p_{\tilde{t}}(z | \tilde{f}(\tilde{t}) = x)$  of  $\det \nabla \tilde{f}(\tilde{t})$  given  $\tilde{f}(\tilde{t}) = x$  are continuous for  $z$  and  $x$  in neighbourhoods of 0 and  $\tilde{u}$ , respectively, uniformly in  $\tilde{t} \in T^k$ .

Then,

$$\begin{aligned} (4.4.1) \quad \mathbb{E}\{(N_u)_k\} &= \int_{T^k} \mathbb{E} \left\{ \prod_{j=1}^k |\det \nabla f(t_j)| \mathbf{1}_B(g(t_j)) \mid \tilde{f}(\tilde{t}) = \tilde{u} \right\} p_{\tilde{t}}(\tilde{u}) d\tilde{t} \\ &= \int_{T^k} \prod_{j=1}^k \int_{\mathbb{R}^{kD}} |\det D_j| \mathbf{1}_B(v_j) p_{\tilde{t}}(\tilde{u}, \tilde{D}, \tilde{v}) d\tilde{D} d\tilde{v} d\tilde{t} \end{aligned}$$

where

- (i)  $p_{\tilde{t}}(\tilde{x})$  is the density of  $\tilde{f}(\tilde{t})$ .
- (ii)  $p_{\tilde{t}}(\tilde{x}, \tilde{D}, \tilde{v})$  is the joint density of  $\tilde{f}(\tilde{t}), \tilde{D}(\tilde{t})$  and  $\tilde{g}(\tilde{t})$ .
- (iii)  $\tilde{D}(\tilde{t})$  represents the  $Nk \times Nk$  matrix  $\nabla \tilde{f}(\tilde{t})$ . Note that  $\tilde{D}(\tilde{t})$  is a diagonal block matrix, with the  $j$ -th block  $D_j(\tilde{t})$  containing the matrix  $\nabla f(t_j)$ , where  $t_j \in T$  is the  $j$ -th component of  $\tilde{t}$ .
- (iv)  $D = N(N+1)/2 + K$ .

**Proof.** For each  $\delta > 0$  define the domain

$$T_\delta^k \triangleq \{\tilde{t} \in T^k : |t_i - t_j| \geq \delta, \text{ for all } 1 \leq i < j \leq k\}.$$

Then the field  $\tilde{f}$  satisfies all the assumptions of Theorem 4.1.1 over the parameter set  $T_\delta^k$ .

It therefore follows from (4.1.4) and (4.1.5) that  $\mathbb{E}\{N_{\bar{u}}(\tilde{f}, \tilde{g} : T_{\delta}^k, B^k)\}$  is given by either of the integrals in (4.1.2), with the outer integrals taken over  $T_{\delta}^k$  rather than  $T^k$ .

Let  $\delta \downarrow 0$ . Then, using the fact that  $f = u$  only finitely often (cf. Lemma 4.1.12) it is easy to see that, with probability one,

$$N_{\bar{u}}(\tilde{f}, \tilde{g} : T_{\delta}^k, B^k) \uparrow (N_u(f, g : T, B))_k.$$

The monotonicity of the convergence then implies convergence of the expectations  $\mathbb{E}\{N_{\bar{u}}(\tilde{f}, \tilde{g} : T_{\delta}^k, B^k)\}$ .

On the other hand, the integrals in (4.1.2) are trivially the limit of the same expressions with the outer integrals taken over  $T_{\delta}^k$  rather than  $T^k$ , and so we are done.  $\square$

As for the basic expectation result, Theorem 4.4.1 takes a far simpler form if  $f$  is Gaussian, and we have

**Corollary 4.4.2** *Let  $f$  and  $g$  be centered Gaussian fields over a  $T$  which satisfies the conditions of Theorem 4.1.1. If, for each  $t \in T$ , the joint distributions of  $\{(f(t_j), \nabla f(t_j), g(t_j))\}_{1 \leq j \leq k}$  are non-degenerate for all choices of distinct  $t_j \in T$ , and if (4.1.6) holds, then so does (4.1.2).*

## 4.5 Preliminary Gaussian computations

In the following Section 4.6 we shall begin our computations of expectations for the Euler characteristics of excursion sets of Gaussian fields. In preparation, we need to collect a few facts about Gaussian integrals. Some are standard, others particularly tailored to our specific needs. However, since all are crucial components for the proofs of the main results of this Chapter, we shall give proofs for all of them.

The first, in particular, is standard fare:

**Lemma 4.5.1** *Let  $X_1, X_2, \dots, X_n$  be a set of real-valued random variables having a joint Gaussian distribution and zero means. Then for any integer  $m$*

$$(4.5.1) \quad \mathbb{E}\{X_1 X_2 \cdots X_{2m+1}\} = 0,$$

$$(4.5.2) \quad \mathbb{E}\{X_1 X_2 \cdots X_{2m}\} = \sum \mathbb{E}\{X_{i_1} X_{i_2}\} \cdots \mathbb{E}\{X_{i_{2m-1}} X_{i_{2m}}\},$$

where the sum is taken over the  $(2m)!/m!2^m$  different ways of grouping  $X_1, \dots, X_{2m}$  into  $m$  pairs.

Note that this result continues to hold even if some of the  $X_j$  are identical, so that the Lemma also applies to compute joint moments of the form  $\mathbb{E}\{X_1^{i_1} \cdots X_k^{i_k}\}$ .

**Proof.** Recall from (1.2.4) that the joint characteristic function of the  $X_i$  is

$$\phi(\theta) = \mathbb{E} \left\{ e^{i \sum_i \theta_i X_i} \right\} = e^Q,$$

where

$$Q = Q(\theta) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \theta_i \mathbb{E}\{X_i X_j\} \theta_j.$$

Following our usual convention of denoting partial derivatives by subscripting, we have, for all  $l, k, j$ ,

$$(4.5.3) \quad \begin{aligned} Q_j &= -\sum_{k=1}^n \mathbb{E}\{X_j X_k\} \theta_k, \\ Q_{kj} &= -\mathbb{E}\{X_j X_k\}, \\ Q_{lkj} &= 0. \end{aligned}$$

Successive differentiations of (4.5.3) yield

$$(4.5.4) \quad \begin{aligned} \phi_j &= \phi Q_j, \\ \phi_{kj} &= \phi_k Q_j + \phi Q_{kj}, \\ \phi_{lkj} &= \phi_{lk} Q_j + \phi_k Q_{lj} + \phi_l Q_{kj}, \\ &\vdots \\ \phi_{12\dots n} &= \phi_{12\dots(j-1)(j+1)\dots n} Q_j + \sum_{k \neq j} \phi_{r_1 \dots r_{n-2}} Q_{kj}, \end{aligned}$$

where, in the last equation, the sequence  $r_1, \dots, r_{n-2}$  does not include the two numbers  $k$  and  $j$ .

The moments of various orders can now be obtained by setting  $\theta = 0$  in the equations of (4.5.4). Since from (4.5.3) we have  $Q_j(0) = 0$  for all  $j$ , the last (and most general) equation in (4.5.4) thus leads to

$$\mathbb{E}\{X_1 \cdots X_n\} = \sum_{k \neq j} \mathbb{E}\{X_{r_1} \cdots X_{r_{n-2}}\} \mathbb{E}\{X_j X_k\}.$$

From this relationship and the fact that the  $X_j$  all have zero mean it is easy to deduce the validity of (4.5.1) and (4.5.2). It remains only to determine exactly the number,  $M$  say, of terms in the summation (4.5.2). We note first that while there are  $(2m)!$  permutations of  $X_1, \dots, X_{2m}$ , since the sum does not include identical terms,  $M < (2m)!$ . Secondly, for each term in the sum, permutations of the  $m$  factors result in identical ways of breaking up the  $2m$  elements. Thirdly, since  $\mathbb{E}\{X_j X_k\} = \mathbb{E}\{X_k X_j\}$ , an interchange of the order in such a pair does not yield a new pair. Thus

$$M(m!)(2^m) = (2m)!$$

implying

$$M = \frac{(2m)!}{m! 2^m}$$

as stated in the lemma.  $\square$

For the next Lemma we need some notation. Let  $\Delta_N$  be a symmetric  $N \times N$  matrix with elements  $\Delta_{ij}$ , such that each  $\Delta_{ij}$  is a zero mean normal variable with arbitrary variance but such that the following relationship holds:

$$(4.5.5) \quad \mathbb{E}\{\Delta_{ij}\Delta_{kl}\} = \mathcal{E}(i, j, k, l) - \delta_{ij}\delta_{kl},$$

where  $\mathcal{E}$  is a symmetric function of  $i, j, k, l$  and  $\delta_{ij}$  is the Kronecker delta. Write  $|\Delta_N|$  for the determinant of  $\Delta_N$ .

**Lemma 4.5.2** *Let  $m$  be a positive integer. Then, under (4.5.5),*

$$(4.5.6) \quad \mathbb{E}\{|\Delta_{2m+1}|\} = 0,$$

$$(4.5.7) \quad \mathbb{E}\{|\Delta_{2m}|\} = \frac{(-1)^m (2m)!}{m! 2^m}.$$

**Proof.** Relation (4.5.6) is immediate from (4.5.1). Now

$$|\Delta_{2m}| = \sum_P \eta(p) \Delta_{1i_1} \cdots \Delta_{2m, i_{2m}}$$

where  $p = (i_1, i_2, \dots, i_{2m})$  is a permutation of  $(1, 2, \dots, 2m)$ ,  $P$  is the set of the  $(2m)!$  such permutations, and  $\eta(p)$  equals  $+1$  or  $-1$  depending on the order of the permutation  $p$ . Thus by (4.5.2) we have

$$\mathbb{E}\{|\Delta_{2m}|\} = \sum_P \eta(p) \sum_Q \mathbb{E}\{\Delta_{1i_1}\Delta_{2i_2}\} \cdots \mathbb{E}\{\Delta_{2m-1, i_{2m-1}}\Delta_{2m, i_{2m}}\},$$

where  $Q$  is the set of the  $(2m)!/m! 2^m$  ways of grouping  $(i_1, i_2, \dots, i_{2m})$  into pairs without regard to order, keeping them paired with the first index. Thus, by (4.5.5),

$$\begin{aligned} \mathbb{E}\{|\Delta_{2m}|\} &= \sum_P \eta(p) \sum_Q \{\mathcal{E}(1, i_1, 2, i_2) - \delta_{1i_1}\delta_{2i_2}\} \times \cdots \\ &\quad \times \{\mathcal{E}(2m-1, i_{2m-1}, 2m, i_{2m}) - \delta_{2m-1, i_{2m-1}}\delta_{2m, i_{2m}}\}. \end{aligned}$$

It is easily seen that all products involving at least one  $\mathcal{E}$  term will cancel out because of their symmetry property. Hence

$$\begin{aligned} \mathbb{E}\{|\Delta_{2m}|\} &= \sum_P \eta(p) \sum_Q (-1)^m (\delta_{1i_1}\delta_{2i_2}) \cdots (\delta_{2m-1, i_{2m-1}}\delta_{2m, i_{2m}}) \\ &= \frac{(-1)^m (2m)!}{(m!) 2^m}, \end{aligned}$$

the last line coming from changing the order of summation and then noting that for only one permutation in  $P$  is the product of delta functions non-zero. This completes the proof.  $\square$

Note that (4.5.7) in no way depends on the specific (co)variances among the elements of  $\Delta_N$ . These all disappear in the final result due to the symmetry of  $\mathcal{E}$ .

Before stating the next result we need to introduce the family of *Hermite polynomials*. The  $n$ -th Hermite polynomial is the function

$$(4.5.8) \quad H_n(x) = n! \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{(-1)^j x^{n-2j}}{j! (n-2j)! 2^j}, \quad n \geq 1, \quad x \in \mathbb{R},$$

where  $\lfloor a \rfloor$  is the largest integer less than or equal to  $a$ . For convenience later, we also define

$$(4.5.9) \quad H_{-1}(x) = \sqrt{2\pi} \Psi(x) e^{x^2/2}, \quad x \in \mathbb{R},$$

where  $\Psi$  is the tail probability function for a standard Gaussian variable (cf. (1.2.1)). With the normalisation inherent in (4.5.8) the Hermite polynomials form an orthogonal (but not orthonormal) system with respect to standard Gaussian measure on  $\mathbb{R}$ , in that

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} H_n(x) H_m(x) e^{-x^2/2} dx = \begin{cases} n! & m = n, \\ 0 & m \neq n. \end{cases}$$

An alternative definition of the Hermite polynomials is via a generating function approach, which gives

$$(4.5.10) \quad H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} (e^{-x^2/2}), \quad n \geq 1.$$

From this it immediately follows that

$$(4.5.11) \quad \int_u^\infty H_n(x) e^{-x^2/2} dx = H_{n-1}(u) e^{-u^2/2}, \quad n \geq 1.$$

The centrality of Hermite polynomials for us lies in the following Corollary to Lemma 4.5.2.

**Corollary 4.5.3** *Let  $\Delta_N$  be as in Lemma 4.5.2, with the same assumptions in force. Let  $I$  be the  $N \times N$  unit matrix, and  $x \in \mathbb{R}$ . Then*

$$\mathbb{E}\{\det(\Delta_N - xI)\} = (-1)^N H_N(x).$$

**Proof.** It follows from the usual Laplace expansion of the determinant that

$$\begin{aligned} \det(\Delta_N - xI) &= (-1)^N [x^N - S_1(\Delta_N)x^{N-1} \\ &\quad + S_2(\Delta_N)x^{N-2} + \dots + (-1)^N S_N(\Delta_N)], \end{aligned}$$

where  $S_k(\Delta_N)$  is the sum of the  $\binom{N}{k}$  principal minors of order  $k$  in  $|\Delta_N|$ . The result now follows trivially from (4.5.6) and (4.5.7).  $\square$

The final lemma we require is elementary and is proven via integration by parts.

**Lemma 4.5.4** *Let  $X \sim N(\mu, \sigma^2)$ , and let  $\Phi = 1 - \Psi$ , as usual, denote the standard normal distribution function, and  $\varphi$  the corresponding density function. Then, with  $x^+ = \max(0, x)$ ,*

$$\begin{aligned}\mathbb{E}\{X^+\} &= \mu\Psi(-\mu/\sigma) + \sigma^2\varphi(\mu^2/2\sigma^2), \\ \mathbb{E}\{X^-\} &= \mu\Phi(-\mu/\sigma) + \sigma^2\varphi(\mu^2/2\sigma^2).\end{aligned}$$

## 4.6 Mean Euler characteristics: Euclidean case

We now assume  $f$  to be a centered, stationary, Gaussian process on a rectangle  $T \subset \mathbb{R}^N$  and satisfying the conditions of Corollary 4.2.2. As usual,  $C : \mathbb{R}^N \rightarrow \mathbb{R}$  denotes the covariance function, and  $\nu$  the spectral measure. Then  $f$  has variance  $\sigma^2 = C(0) = \nu(\mathbb{R}^N)$ . We change notation a little from Section 1.4.3, and introduce the *second-order spectral moments*

$$(4.6.1) \quad \lambda_{ij} = \int_{\mathbb{R}^N} \lambda_i \lambda_j \nu(d\lambda).$$

Since we shall use it often, denote the  $N \times N$  matrix of these moments by  $\Lambda$ .

Denoting also differentiation via subscripts, so that  $f_i = \partial f / \partial t_i$ ,  $f_{ij} = \partial^2 f / \partial t_i \partial t_j$ , etc. we have, from (1.4.34), that

$$(4.6.2) \quad \mathbb{E}\{f_i(t)f_j(t)\} = \lambda_{ij} = -C_{ij}(0).$$

Thus  $\Lambda$  is also the variance-covariance matrix of  $\nabla f$ . We could, of course, define the components of  $\Lambda$  using only derivatives of  $R$ , without ever referring to the spectrum.

The covariances among the second order derivatives can be similarly defined. However all we shall need is that

$$(4.6.3) \quad \mathcal{E}_{ijkl} \triangleq \mathbb{E}\{f_{ij}(t)f_{kl}(t)\} = \int_{\mathbb{R}^N} \lambda_i \lambda_j \lambda_k \lambda_\ell \nu(d\lambda)$$

is a symmetric function of  $i, j, k, \ell$ .

Finally, note, as shown in Section 1.4.3, that  $f$  and its first order derivatives are independent (at any fixed point  $t$ ) as are the first and second order derivatives (from one another). The field and its second order derivatives are, however, correlated, and

$$(4.6.4) \quad \mathbb{E}\{f(t)f_{ij}(t)\} = -\lambda_{ij}.$$

Finally, denote the  $N \times N$  Hessian matrix  $(f_{ij})$  by  $\nabla^2 f$ .

**Lemma 4.6.1** *Let  $f$  and  $T$  be as described above, and set*

$$(4.6.5) \quad \mu_k = \#\{t \in T : f(t) \geq u, \nabla f(t) = 0, \text{index}(\nabla^2 f) = k\}.$$

*Then, for all  $N \geq 1$ ,*

$$(4.6.6) \quad \mathbb{E} \left\{ \sum_{k=0}^N (-1)^k \mu_k \right\} = \frac{(-1)^N |T| |\Lambda|^{1/2}}{(2\pi)^{(N+1)/2} \sigma^N} H_{N-1} \left( \frac{u}{\sigma} \right) e^{-u^2/2\sigma^2}.$$

Before turning to the proof of the Lemma, there are some crucial points worth noting. The first is the rather surprising fact that the result depends on the covariance of  $f$  only through some of its derivatives at zero; viz. only through the variance and second order spectral moments. This is particularly surprising in view of the fact that the definition of the  $\mu_k$  depends quite strongly on the  $f_{ij}$ , whose distribution involves fourth order spectral moments.

As will become clear from the proof, the disappearance of the fourth order spectral moments has a lot to do with the fact that we compute the mean of the alternating sum in (4.6.6) and do not attempt to evaluate the expectations of the individual  $\mu_k$ . Doing so would indeed involve fourth order spectral moments. As we shall see in later Chapters, the fact that this is all we need is extremely fortunate, for it is actually *impossible* to obtain closed expressions for any of the  $\mathbb{E}\{\mu_k\}$ .

**Proof.** We start with the notationally simplifying assumption that  $\mathbb{E}\{f_t^2\} = \sigma^2 = 1$ . It is clear that if we succeed in establishing (4.6.6) for this case, then the general case follows by scaling. (Note that scaling  $f$  also implies scaling  $\nabla f$ , which, since  $\Lambda$  contains the variances of the elements of  $\nabla f$ , gives the factor of  $\sigma^{-N}$  in (4.6.6).)

Our second step is to simplify the covariance structure among the elements of  $\nabla f$ . Let  $Q$  be the orthogonal matrix which diagonalises  $\Lambda$  to the unit matrix  $I$ , so that

$$(4.6.7) \quad Q' \Lambda Q = \text{diag}(1, \dots, 1).$$

Note that  $\det Q = (\det \Lambda)^{-1/2}$ . Now take the transformation of  $\mathbb{R}^N$  given by  $t \rightarrow tQ$ , under which  $T \rightarrow T^Q = \{\tau : \tau = tQ^{-1} \text{ for some } t \in T\}$  and define  $f^Q : T^Q \rightarrow \mathbb{R}$  by

$$f^Q(t) \triangleq f(tQ).$$

The new process  $f^Q$  has covariance function  $C^Q(s, t) = C(sQ, tQ) = C((t-s)Q)$ , and so is still stationary, with constant, unit variance. Furthermore, simple differentiation shows that  $\nabla f^Q = (\nabla f)Q$ , from which it follows that

$$(4.6.8) \quad \begin{aligned} \Lambda^Q &\triangleq \mathbb{E} \{ ((\nabla f^Q)(t))' ((\nabla f^Q)(t)) \} \\ &= Q' \Lambda Q \\ &= I. \end{aligned}$$

That is, the first order derivatives of the transformed process are now uncorrelated and of unit variance. We now show that it is sufficient to work with this, much simpler, transformed process.

Firstly, it is crucial to note that the  $\mu_k$  of (4.6.5) for  $f$  over  $T$  are *identical* to those for  $f^Q$  over  $T^Q$ . Clearly, there is trivial a one-one correspondence between those points of  $T$  at which  $f(t) \geq u$  and those of  $T^Q$  at which  $f^Q(t) \geq u$ . We do, however, need to check more carefully what happens with the conditions on  $\nabla f$  and  $\nabla^2 f$ .

Since  $\nabla f^Q = (\nabla f)Q$ , we have that  $(\nabla f^Q)(t) = 0$  if, and only if,  $\nabla f(tQ) = 0$ . In other words, there is also a simple one-one correspondence between critical points. Furthermore, since  $\nabla^2 f^Q = Q' \nabla^2 Q$  and  $Q$  is a positive definite matrix  $\nabla^2 f^Q(t)$  and  $\nabla^2 f(tQ)$  have the same index.

Consequently, we can work now with  $f^Q$  rather than  $f$ , so that by (4.1.5) the expectation (4.6.6) is given by

$$(4.6.9) \quad \int_{T^Q} p_t(0) dt \sum_{k=0}^N (-1)^k \times \mathbb{E} \left\{ |\det \nabla^2 f^Q(t)| \mathbb{1}_{D_k}(\nabla^2 f^Q(t)) \mathbb{1}_{[u, \infty)}(f^Q(t)) \mid \nabla f^Q(t) = 0 \right\},$$

where  $p_t$  is the density of  $\nabla f^Q$  and  $D_k$  is set of square matrices of index  $k$ . Now note the following:

- (i) Since  $f$ , and so  $f^Q$ , are stationary, the integrand does not depend on  $t$ , and we can ignore the  $t$ 's throughout. The remaining integral then gives the Lebesgue volume of  $T^Q$ , which is simply  $|\det Q|^{-1}|T| = |\Lambda|^{1/2}|T|$ .
- (ii) The term  $p_t(0)$  is simply  $(2\pi)^{-N/2}$ , and so can be placed to the side for the moment.
- (iii) Most importantly, on the event  $D_k$ , the matrix  $\nabla^2 f^Q(t)$  has  $k$  negative eigenvalues, and so has sign  $(-1)^k$ . We can combine this with the factor  $(-1)^k$  coming immediately after the summation sign, and so remove both it and the absolute value sign on the determinant.
- (iv) Recall that from the discussion on stationarity in Section 1.4.3 (esp. (1.4.34)–(1.4.36)) we have the following relationships between the various derivatives of  $f^Q$ , for all  $i, j, k \in \{1, \dots, N\}$ .

$$\mathbb{E}\{f^Q(t)f_i^Q(t)\} = 0, \quad \mathbb{E}\{f_i^Q(t)f_{jk}^Q(t)\} = 0, \quad \mathbb{E}\{f^Q(t)f_{ij}^Q(t)\} = -\delta_{ij},$$

where  $\delta_{ij}$  is the Kronecker delta. The independence of the second derivatives from all the others means that the conditioning on  $\nabla f^Q$  in (4.6.9) can be ignored, and so all that remains is to evaluate

$$(4.6.10) \quad \int_u^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathbb{E}\{\det \Delta_x\} dx,$$

where, by (1.2.7) and (1.2.8)  $\Delta_x$  is a matrix of Gaussian random variables whose elements  $\Delta_{ij}$  have means  $-x\delta_{ij}$  and covariances

$$\begin{aligned} \mathbb{E}\{\Delta_{ij}\Delta_{k\ell}\} &= \mathbb{E}\{f_{ij}^Q(t)f_{k\ell}^Q(t)\} - \delta_{ij}\delta_{k\ell} \\ &= \mathcal{E}^Q(i, j, k, \ell) - \delta_{ij}\delta_{k\ell}. \end{aligned}$$

By (4.6.3),  $\mathcal{E}^Q$  is a symmetric function of its parameters.

This puts us directly into the setting of Corollary 4.5.3, from which it now follows that (4.6.10) is equivalent to

$$\frac{(-1)^N}{\sqrt{2\pi}} \int_u^\infty H_N(x) e^{-x^2/2} dx = \frac{(-1)^N}{\sqrt{2\pi}} H_{N-1}(u) e^{-u^2/2},$$

by (4.5.11).

Recalling (i) (ii) and (iv) above, substituting into (4.6.9) and lifting the restriction that  $\sigma^2 = 1$  gives us (4.6.6) and we are done.  $\square$

With the proof behind us, it should now be clear why it is essentially impossible to evaluate the individual  $\mathbb{E}\{\mu_k\}$ . In doing so, we would have had to integrate over the various subsets  $D_k \subset \mathbb{R}^{N(N+1)/2}$  of (4.6.9), and, with only the most rare exceptions<sup>10</sup>, such integrals do not have explicit forms.

A careful reading of the proof also shows that we never really used the full power of stationarity, but rather only the existence of the matrix  $Q$  of (4.6.7) and a number of relationships between  $f$  and its first and second order derivatives. Actually, given the simplicity of the final result, which depends only on  $\sigma^2$  and  $\Lambda$ , one is tempted to conjecture that in the non-stationary case one could replace  $Q$  by a family  $Q_t$  of such matrices. The final result might then be much the same, with the term  $|T||\Lambda|^{1/2}$  perhaps being replaced by an integral of the form  $\int_T |\Lambda_t|^{1/2} dt$  where  $\Lambda_t$  would be a local version of  $\Lambda$ , with elements  $(\Lambda_t)_{ij} = \mathbb{E}\{f_i(t)f_j(t)\}$ . That this is *not* the case will be shown later, when we do tackle the (far more complicated) non-stationary scenario as a corollary of results related to fields on Riemannian manifolds. (cf. the discussion of non-stationary fields in Section 4.11.)

We can now turn to our first mean Euler characteristic computation, for which we need to set up a little notation, much of it close to that of Section 3.9.2. Nevertheless, since there are some slight changes, we shall write it

---

<sup>10</sup>See Section 6.5 for one such exception, when  $N = 2$ . This is the only case we know of where the individual  $\mathbb{E}\{\mu_k\}$  can be explicitly computed. However, even here, the adjective ‘explicitly’ assumes that Legendre elliptic integrals can be considered ‘simple’ functions.

all out again. We start with

$$T = \prod_{i=1}^N [0, T_i],$$

a rectangle in  $\mathbb{R}^N$ . As in Section 3.9.2, write  $\mathcal{J}_k = \mathcal{J}_k(T)$  for the collection of the  $2^{N-k} \binom{N}{k}$  faces of dimension  $k$  in  $T$ . As opposed to our previous conventions, we take these faces as closed. Thus, all faces in  $\mathcal{J}_k$  are subsets of some face in  $\mathcal{J}_{k'}$  for all  $k' > k$ . (For example,  $\mathcal{J}_N$  contains only  $T$  while  $\mathcal{J}_0$  contains the  $2^N$  vertices of  $T$ .) Let  $\mathcal{O}_k$  denote the  $\binom{N}{k}$  elements of  $\mathcal{J}_k$  which include the origin.

We need one more piece of notation. Take  $J \in \mathcal{J}_k$ . With the  $\lambda_{ij}$  being the spectral moments of (4.6.1), write  $\Lambda_J$  for the  $k \times k$  matrix with elements  $\lambda_{ij}$ ,  $i, j \in \sigma(J)$ .

This is enough to state the following result.

**Theorem 4.6.2** *Let  $f$  be as described at the beginning of this Section and  $T$  as above. For real  $u$ , let  $A_u = A_u(f, T) = \{t \in T : f(t) \geq u\}$ , be an excursion set, and let  $\varphi$  be the Euler characteristic. Then*

(4.6.11)

$$\mathbb{E} \{ \varphi(A_u) \} = e^{-u^2/2\sigma^2} \sum_{k=1}^N \sum_{J \in \mathcal{O}_k} \frac{|J| |\Lambda_J|^{1/2}}{(2\pi)^{(k+1)/2} \sigma^k} H_{k-1} \left( \frac{u}{\sigma} \right) + \Psi \left( \frac{u}{\sigma} \right).$$

Note that the  $k$ -dimensional volume  $|J|$  of any  $J \in \mathcal{J}_k$  is given by  $|J| = \prod_{i \in \sigma(J)} T_i$ .

Since (4.6.11) and its extension to manifolds in Section 4.10 is, for our purposes, probably the single most important equation in this book, we shall take a little time to investigate some of its consequences, before turning to a proof. To do so, we first note that it simplifies somewhat if  $f$  is isotropic. In that case we have

**Corollary 4.6.3** *In addition to the conditions of Theorem 4.6.2, let  $f$  be isotropic and  $T$  the cube  $[0, T]^N$ . If  $\lambda_2$  denotes the variance of  $f_i$  (independent of  $i$  by isotropy) then*

(4.6.12)

$$\mathbb{E} \{ \varphi(A_u) \} = e^{-u^2/2\sigma^2} \sum_{k=1}^N \frac{\binom{N}{k} T^k \lambda_2^{k/2}}{(2\pi)^{(k+1)/2} \sigma^k} H_{k-1} \left( \frac{u}{\sigma} \right) + \Psi \left( \frac{u}{\sigma} \right).$$

The simplification follows immediately from the spherical symmetry of the spectral measure in this case, which (cf. (1.4.40)) implies that each matrix  $\Lambda_J = \lambda_2 I$ . In fact, looking back into the proof of Lemma 4.6.1, which is where most of the calculation occurs, you can see that the transformation

to the process  $f^Q$  is now rather trivial, since  $Q = \lambda^{-1/2}I$  (cf. (4.6.7)). Looked at in this light, it is clear that one of the key points of the proof was a transformation that made the first derivatives of  $f$  behave as if they were those of an isotropic process. We shall see this again, but at a far more sophisticated level, when we turn to the manifold setting.

Now consider the case  $N = 1$ , so that  $T$  is simply the interval  $[0, T]$ . Then, using the definition of the Hermite polynomials given by (4.5.8), it is trivial to check that

$$(4.6.13) \quad \mathbb{E} \{ \varphi(A_u(f, [0, T])) \} = \Psi(u/\sigma) + \frac{T\lambda_2^{1/2}}{2\pi\sigma} e^{-u^2/2\sigma^2}.$$

Figure 4.6.1 gives two examples, with  $\sigma^2 = 1$ ,  $\lambda_2 = 200$  and  $\lambda_2 = 1,000$ .

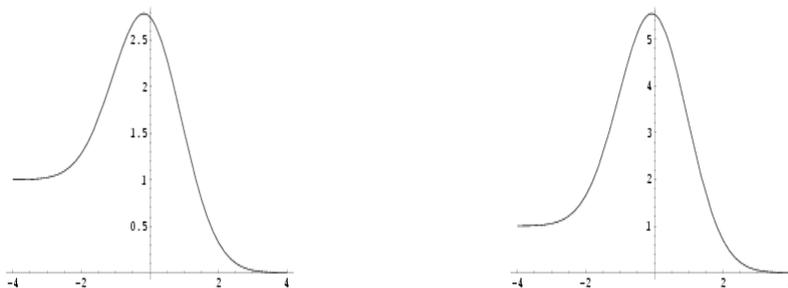


FIGURE 4.6.1.  $\mathbb{E} \{ \varphi(A_u) \} : N = 1$

There are at least two interesting facts that you should note from (4.6.13):

Firstly, as  $u \rightarrow -\infty$ ,  $\mathbb{E} \{ \varphi(A_u) \} \rightarrow 1$ . The excursion set geometry behind this is simple. Once  $u < \inf_T f(t)$  we have  $A_u \equiv T$ , and so  $\varphi(A_u) = \varphi(T)$  which, in the current case, is 1. This is, of course, a general phenomenon, independent of dimension or the topology of  $T$ .

To see this analytically, simply look at the expression (4.6.13), or even (4.6.11) for general rectangles. In both cases it is trivial that as  $u \rightarrow -\infty$  all terms other than  $\Psi(u/\sigma)$  disappear, while  $\Psi(u/\sigma) \rightarrow 1$ .

It thus seems not unreasonable to expect that when we turn to a more general theory (i.e. for  $T$  a piecewise smooth manifold) the term corresponding to the last term in (4.6.11) might be  $\varphi(T)\Psi(u/\sigma)$ . That this is in fact the case can be seen from the far more general results of Section 4.10 below.

Secondly, it is trivial to see that, still in dimension one,

$$\varphi(A_u) \equiv \mathbb{1}_{f_0 \geq u} + N_u^+(0, T)$$

where

$$\begin{aligned} N_u^+(0, T) &\triangleq \#\{t \in [0, t] : f(t) = u, f'(t) > 0\} \\ &= \text{the number of upcrossings of the level } u \text{ in } [0, T]. \end{aligned}$$

Putting this together with (4.6.13), we have recovered the famous *Rice formula*<sup>11</sup> for the expected number of upcrossings of a zero mean stationary process, viz;

$$(4.6.14) \quad \mathbb{E} \{N_u^+(0, T)\} = \frac{T\lambda_2^{1/2}}{2\pi\sigma} e^{-u^2/2\sigma^2}.$$

Written in terms of level crossings, the rôle of the second spectral moment  $\lambda_2$  becomes rather clear. For fixed variance  $\sigma^2$ , it follows from the spectral theory of Section 1.4.3 that increasing  $\lambda_2$  increases the ‘high frequency components’ of  $f$ . In other words, the sample paths become locally much rougher. For example, Figure 4.6.2 shows realisations of the very simple Gaussian process with discrete spectrum

$$(4.6.15) \quad \frac{a}{\sqrt{2}} \delta_{\pm\sqrt{200}}(\lambda) + \frac{\sqrt{1-a^2}}{\sqrt{2}} \delta_{\pm\sqrt{1,000}}(\lambda),$$

for  $a = 0$  and  $a = 5/13$ , which gives second spectral moments of  $\lambda_2 = 1,000$  and  $\lambda_2 = 200$ , respectively.

From the figure it should be reasonably clear what this means in sample path terms: higher second spectral moments correspond to an increased number of level crossings generated by local fluctuations. Similar phenomena occur also in higher dimensions, as we shall see soon.

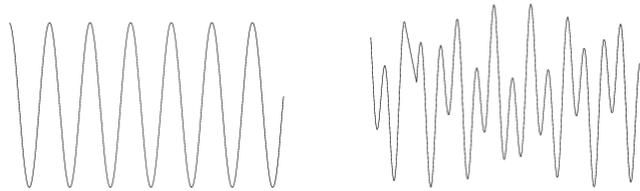


FIGURE 4.6.2. Realisations of Gaussian sample paths with the spectrum (4.6.15) and second spectral moments  $\lambda_2 = 200$  (left) and  $\lambda_2 = 1,000$  (right).

<sup>11</sup>Rice’s formula is really due to both Rice [78] and Kac [50] in the early 1940’s, albeit in slightly different settings. Over the years it underwent significant improvement, being proven under weaker and weaker conditions. In its final form, the only requirement on  $f$  (in the current stationary, zero mean Gaussian scenario) is that its sample paths be, almost surely, absolutely continuous with respect to Lebesgue measure. This is far less than we demand, since it does even require a continuous sample path derivative, let alone a continuous second derivative. Furthermore (4.6.14) turns out to hold whether or not  $\lambda_2$  is finite, whereas we have assumed  $\lambda_2 < \infty$ . For details see, for example, [55].

It is, by the way, hard to believe that the general ‘random field version of Rice’s formula’ given by (4.6.11) could ever be derived without the additional levels of smoothness that we have found it necessary to assume.

We now turn to two dimensions, in which case the right hand side of (4.6.12) becomes, for  $\sigma^2 = 1$ ,

$$(4.6.16) \quad \left[ \frac{T^2 \lambda_2}{(2\pi)^{3/2}} u + \frac{2T \lambda_2^{1/2}}{2\pi} \right] e^{-u^2/2} + \Psi(u).$$

Figure 4.6.3 gives two examples, again with  $\lambda_2 = 200$  and  $\lambda_2 = 1,000$ .

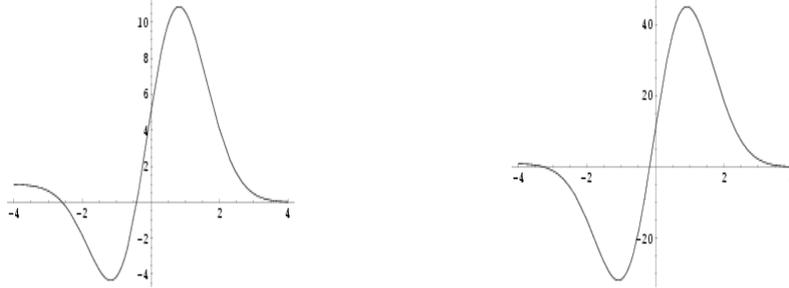


FIGURE 4.6.3.  $\mathbb{E}\{\varphi(A_u)\} : N = 2$

Many of the comments that we made for the one dimensional case have similar analogues here and we leave them to you. Nevertheless, we emphasise three points:

- (i) You should note, for later reference, how the expression before the exponential term can be thought of as one of a number of different power series; one in  $T$ , one in  $u$ , and one in  $\sqrt{\lambda_2}$ .
- (ii) The geometric meaning of the negative values of (4.6.16) are worth understanding. They are due to the excursion sets having, in the mean, more holes than connected components for (most) negative values of  $u$ .
- (iii) The impact of the spectral moments is not quite as clear in higher dimensions as it is in one. Nevertheless, to get a feel for what is happening, look back at the simulation of a Brownian sheet in Figure 1.3. The Brownian sheet is, of course, both non-stationary and non-differentiable, and so hardly belongs in our current setting. Nevertheless, in a finite simulation, it is impossible to ‘see’ the difference between non-differentiability and large second spectral moments<sup>12</sup>, so consider the simulation in the latter light. You can then see what is happening. Large spectral moments again lead to local fluctuations generating large numbers of small islands (or lakes, depending on the

<sup>12</sup>Ignore the non-stationarity of the Brownian sheet, since this has no qualitative impact on the discussion.

level at which the excursion set is taken) and this leads to larger variation in the values of  $\mathbb{E}\{\varphi(A_u)\}$ .

In three dimensions, the last case that we write out, (4.6.12) becomes, for  $\sigma^2 = 1$ ,

$$\left[ \frac{T^3 \lambda_2^{3/2}}{(2\pi)^2} u^2 + \frac{3T^2 \lambda_2}{(2\pi)^{3/2}} u + \frac{3T \lambda_2^{1/2}}{2\pi} - \frac{T^3 \lambda_2^{3/2}}{(2\pi)^2} \right] e^{-u^2/2} + \Psi(u).$$

Figure 4.6.4 gives the examples with  $\lambda_2 = 200$  and  $\lambda_2 = 1,000$ .

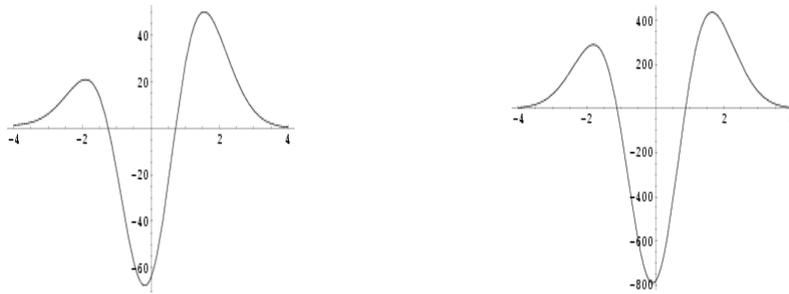


FIGURE 4.6.4.  $\mathbb{E}\{\varphi(A_u)\} : N = 3$

Note that once again there are a number of different power series appearing here, although now, as opposed to the two dimensional case, there is no longer a simple correspondence between the powers of  $T$ ,  $\sqrt{\lambda_2}$  and  $u$ .

The two positive peaks of the curve are due to  $A_u$  being primarily composed of a number of simply connected components for large  $u$  and primarily of simple holes for negative  $u$ . (Recall that in the three dimensional case the Euler characteristic of a set is given by the number of components minus the number of handles plus the number of holes.) The negative values of  $\mathbb{E}\{\varphi(A_u)\}$  for  $u$  near zero are due to the fact that  $A_u$ , at those levels, is composed mainly of a number of interconnected, tubular like regions; i.e. of handles.

An example<sup>13</sup> is given in Figure 4.6.5 which shows an impression of typical excursion sets of a function on  $I^3$  above high, medium and low levels.

We have one more – extremely important – observation to make before we turn to the proof of Theorem 4.6.2. Recall (3.4.5), which gave that the intrinsic volumes, or Lipschitz-Killing curvatures, of  $[0, T]^N$  were given by  $\binom{N}{j} T^j$ . With this in mind, simplifying to the case  $\sigma^2 = \lambda_2 = 1$ , (4.6.12)

<sup>13</sup>Taken, with permission, from Keith Worsley’s entertaining and illuminating *Chance* article [108].

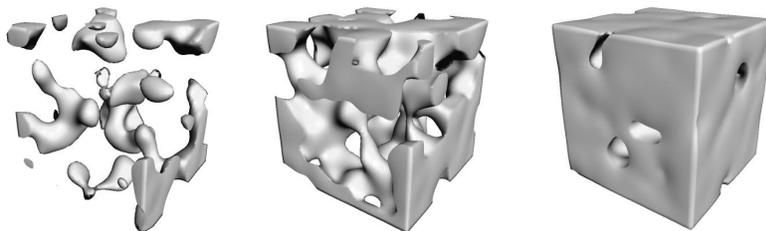


FIGURE 4.6.5. Three dimensional excursion sets above high, medium and low levels. For obvious reasons, astrophysicists refer to these three cases (from left to right) as ‘meatball’, ‘sponge’ and ‘bubble’ topologies.

can be written far more tidily as

$$(4.6.17) \quad \mathbb{E} \{ \varphi(A_u(f, T)) \} = \sum_{k=0}^N \mathcal{L}_k(T) \rho_k(u),$$

where

$$\rho_k(u) = (2\pi)^{-(k+1)/2} H_{k-1}(u) e^{-\frac{u^2}{2}} \quad k \geq 0,$$

since  $H_{-1}(u) = \Psi(u)$ , (cf.(4.5.9))

In fact, the above holds also when  $T$  is a  $N$ -dimensional, piecewise smooth manifold, and  $f$  has constant variance. (i.e. There are no assumptions of isotropy, or even stationarity.) The Lipschitz-Killing curvatures, however, will be somewhat more complex, depending on a Riemannian metric related to  $f$ . This will be the content of Sections 4.8 and 4.10 below. Now, however, we (finally) turn to the

**Proof of Theorem 4.6.2.** We start by recalling that each  $k$ -dimensional face  $J \in \mathcal{J}_k$  is determined by a subset  $\sigma(J)$  of  $\{1, \dots, N\}$ , of size  $k$ , and a sequence of  $N-k$  zeroes and ones, which we write as  $\varepsilon(J) = \{\varepsilon_1, \dots, \varepsilon_{N-k}\}$ , so that

$$J = \{t \in \mathbb{R}^N : t_j = \varepsilon_j T_j, \text{ if } j \notin \sigma(J), \quad 0 \leq t_j \leq T_j, \text{ if } j \in \sigma(J)\}.$$

Corresponding to each set  $\varepsilon(J)$  we define a set  $\varepsilon^*(J)$  of  $\pm 1$ 's, according to the rule  $\varepsilon_j^* = 2\varepsilon_j - 1$ .

Now recall<sup>14</sup> from Section 3.9.2 (cf. (3.9.9)–(3.9.13)) that the Euler characteristic characteristic of  $A_u$  is given by

$$(4.6.18) \quad \varphi(A_u(f, I^N)) = \sum_{k=0}^N \sum_{i=0}^k (-1)^i \sum_{J \in \mathcal{J}_k} \mu_i(J),$$

<sup>14</sup>‘Recall’ includes extending the results there from cubes to rectangles and changing the order of summation, but both these steps are trivial.

where, for  $i \leq \dim(J)$ ,  $\mu_i(J)$  is the number of  $t \in J$  for which

$$(4.6.19) \quad f(t) \geq u,$$

$$(4.6.20) \quad f_j(t) = 0, \quad j \in \sigma(J)$$

$$(4.6.21) \quad \varepsilon_j^* f_j(t) \geq 0, \quad j \notin \sigma(J)$$

$$(4.6.22) \quad I(t) \triangleq \text{Index}(f_{mn}(t))_{(m,n \in \sigma(J))} = k - i,$$

where, as usual, the index of a matrix to be the number of its negative eigenvalues.

Our first and main step will hinge on stationarity, which we exploit to replace the expectation of the sum over  $J \in \mathcal{J}_k$  in (4.6.18) by something simpler. Fix a particular  $J \in \mathcal{O}_k$  – i.e. a face containing the origin – and let  $\mathcal{P}(J)$  denote all faces in  $\mathcal{J}_k$  (including  $J$  itself) which are affine translates of (i.e. parallel to)  $J$ . There are  $2^{N-k}$  faces in each such  $\mathcal{P}(J)$ . We can then rewrite the right hand side of (4.6.18) as

$$(4.6.23) \quad \sum_{k=0}^N \sum_{i=0}^k (-1)^i \sum_{J \in \mathcal{O}_k} \sum_{J' \in \mathcal{P}(J)} \mu_i(J').$$

Consider the expectation of the innermost sum. By Theorem 4.1.1 (cf. (4.1.5)) we can rewrite this as

$$\sum_{J' \in \mathcal{P}(J)} \int_{J'} \mathbb{E} \left\{ \left| \det \nabla^2 f_{|J'}(t) \right| \mathbb{1}_{I(t)=k-i} \mathbb{1}_{\mathcal{E}_{J'}(t)} \mid \nabla f_{|J'}(t) = 0 \right\} p_{\nabla f_{|J'}}(0) dt,$$

where  $\mathcal{E}_J(t)$  denotes the event that (4.6.19) and (4.6.21) hold.

Further simplification requires one more item of notation. For  $J \in \mathcal{O}_k$ , let  $\mathcal{P}^\varepsilon(J)$  denote the collections of all sequences of  $\pm 1$ 's of length  $N - k$ . Note that  $\mathcal{P}^\varepsilon(J)$  is made up of the  $2^{N-k}$  sequences  $\varepsilon^*(J')$  with  $J' \in \mathcal{P}(J)$ .

With this notation, and calling on stationarity, we can replace the last expression by

$$\sum_{\varepsilon^* \in \mathcal{P}^\varepsilon(J)} \int_J \mathbb{E} \left\{ \left| \det \nabla^2 f_{|J}(t) \right| \mathbb{1}_{I(t)=k-i} \mathbb{1}_{\mathcal{E}_{\varepsilon^*}(t)} \mid \nabla f_{|J}(t) = 0 \right\} p_{\nabla f_{|J}}(0) dt,$$

where  $\mathcal{E}_{\varepsilon^*}(t)$  denotes the event that (4.6.19) and (4.6.21) hold for  $\varepsilon^*$ .

Now note the trivial fact that

$$\bigcup_{\varepsilon^* \in \mathcal{P}^\varepsilon(J)} \{ \varepsilon_j^* f_j(t) \geq 0 \}$$

is the sure event. Applying this to the last sum, we see that it simplifies considerably to

$$\int_J \mathbb{E} \left\{ \left| \det \nabla^2 f_{|J}(t) \right| \mathbb{1}_{I(t)=k-i} \mathbb{1}_{f(t) \geq u} \mid \nabla f_{|J}(t) = 0 \right\} p_{\nabla f_{|J}}(0) dt.$$

Going back to Theorem 4.1.1, we have that this is no more than the expected number of points in  $J$  for which

$$f(t) \geq u, \quad \nabla f(t) = 0, \quad \text{Index}(\nabla^2 f) = k - i.$$

If we call the number of points satisfying these conditions  $\mu'_{k-i}(J)$ , then putting all of the above together and substituting into (4.6.18) we see that we need the expectation of

$$\sum_{k=0}^N \sum_{J \in \mathcal{O}_k} (-1)^k \sum_{i=0}^k (-1)^{k-i} \mu'_{k-i}(J).$$

Lemma 4.6.1 gives us a precise expression for the expectation of the innermost sum, at least for  $k \geq 1$ ; *viz.*

$$\frac{(-1)^k |J| |\Lambda_J|^{1/2}}{(2\pi)^{(k+1)/2} \sigma^k} H_{k-1} \left( \frac{u}{\sigma} \right) e^{-u^2/2\sigma^2}.$$

It is left to you to check that for the case  $k = 0$  (i.e. when  $\mathcal{J}_k$  contains the  $2^N$  vertices of  $T$ ) that the remaining term is given by  $\Psi(u/\sigma)$ . Putting all this together immediately gives (4.6.11) and so the proof is complete.  $\square$

Before leaving this setting of random fields defined over  $N$ -dimensional rectangles, it is worthwhile, once again, to recall how crucial stationarity was for our ability to carry out the precise computations that lead to Theorem 4.6.2.

In fact, stationarity appeared twice. We first used some consequences of stationarity (although not its full force) in making the crucial transformation to the process  $f^Q$  (cf. (4.6.7)) in the proof of Lemma 4.6.1, which is where the really detailed Gaussian calculations were made. The second time, we exploited the full power of stationarity in the proof of Theorem 4.6.2 for handling the expectation of the awkward summation of (4.6.18). At this stage, our argument also relied rather heavily on the assumption that our parameter space was a rectangle.

It should be reasonably clear that handling non-stationary fields and/or non-rectangular domains is going to require new tools. This forms the subject matter of the next few Sections.

## 4.7 The meta-theorem on manifolds

In essence, this and the following four Sections will repeat, for random fields on manifolds, what we have already achieved in the Euclidean setting.

As there, our first step will be to set up a “meta-theorem” for computing the mean number of points at which a random field takes a certain value

under specific side conditions. This turns out to be rather easy to do, involving little more than taking the Euclidean result and applying it, chart by chart, to the manifold. This is the content of the current Section.

Actually computing the resulting expression for special cases – such as finding the mean Euler characteristic of excursion sets over piecewise smooth manifolds – turns out to be somewhat more complicated and will cover the other four Sections.

To formulate the meta-theorem for manifolds we need one small piece of notation.

Let  $(M, g)$  be an  $N$ -dimensional Riemannian manifold, and  $f : M \rightarrow \mathbb{R}^N$  be  $C^1$ . Fix an orthonormal frame field  $E$ . Then  $\nabla f_E$  denotes the vector field whose coordinates are given by

$$(4.7.1) \quad (\nabla f_E)_i \equiv \nabla f_{E_i} \triangleq (\nabla f)(E_i) \equiv E_i f,$$

where  $\nabla$  is the gradient operator defined at (3.9.1). If  $f = (f^1, \dots, f^N)$  takes values in  $\mathbb{R}^N$ , then  $\nabla f_E$  denotes the  $N \times N$  matrix with elements  $\nabla f_{E_i}^j$ .

Here, then, is the result:

**Theorem 4.7.1** *Let  $M$  be a compact, oriented,  $N$ -dimensional (possibly piecewise)  $C^3$  manifold with a  $C^2$  Riemannian metric  $g$ . Let  $f = (f^1, \dots, f^N) : M \rightarrow \mathbb{R}^N$  and  $h = (h^1, \dots, h^K) : M \rightarrow \mathbb{R}^K$  be random fields on  $M$ . For an open set  $B \subset \mathbb{R}^K$  of the form (4.1.1) and a point  $u \in \mathbb{R}^N$ , let*

$$N_u \equiv N_u(M) \equiv N_u(f, h; M, B)$$

*denote the number of points  $t \in M$  for which*

$$f(t) = u \quad \text{and} \quad h(t) \in B.$$

*Assume that the following conditions are satisfied for some orthonormal frame field  $E$ :*

- (a) *All components of  $f$ ,  $\nabla f_E$ , and  $h$  are a.s. continuous and have finite variances (over  $M$ ).*
- (b) *For all  $t \in M$ , the marginal densities  $p_t(x)$  of  $f(t)$  (implicitly assumed to exist) are continuous at  $x = u$ .*
- (c) *The conditional densities  $p_t(x|\nabla f_E(t), h(t))$  of  $f(t)$  given  $h(t)$  and  $(\nabla f_E)(t)$  (implicitly assumed to exist) are bounded above and continuous at  $x = u$ , uniformly in  $t \in M$ .*
- (d) *The conditional densities  $p_t(z|f(t) = x)$  of  $\det(\nabla f_{E_j}^i(t))$  given  $f(t) = x$  are continuous for  $z$  and  $x$  in neighbourhoods of 0 and  $u$ , respectively, uniformly in  $t \in M$ .*

(e) The following moment condition holds:

$$(4.7.2) \quad \sup_{t \in M} \max_{1 \leq i, j \leq N} \mathbb{E} \left\{ \left| \nabla f_{Ei}^j(t) \right|^N \right\} < \infty.$$

(f) The moduli of continuity with respect to metric induced by  $g$  (cf. (3.6.1)) of each component of  $h$ , each component of  $f$  and each  $\nabla f_{Ei}^j$  all satisfy

$$(4.7.3) \quad \mathbb{P} \{ \omega(\eta) > \varepsilon \} = o(\eta^N), \quad \text{as } \eta \downarrow 0,$$

for any  $\varepsilon > 0$ .

Then

$$(4.7.4) \quad \mathbb{E}\{N_u\} = \int_M \mathbb{E} \left\{ |\det(\nabla f_E)| \mathbb{1}_B(h) \mid f = u \right\} p(u) \text{Vol}_g,$$

where  $p$  is the density<sup>15</sup> of  $f$  and  $\text{Vol}_g$  the volume element on  $M$  induced by the metric  $g$ .

Before turning to the proof of the Theorem, there are a few points worth noting. The first is that the conditions of the Theorem do not depend on the choice of orthonormal frame field. Indeed, as soon as they hold for one such choice, not only will they hold for all orthonormal frame fields but also for any bounded vector field  $X$ . In the latter case the notation will change slightly, and  $\nabla f_{Ei}^j$  needs to be replaced by  $(Xf^j)_i$ .

Once this is noted, you should note that the only place that the metric  $g$  appears in the conditions is in the definition of the neighbourhoods  $B_r(t, h)$  in the final condition. A quick check of the proof to come will show that any neighbourhood system will in fact suffice. Thus the metric does not really play a rôle in the conditions beyond convenience.

Furthermore, the definition of the random variable  $N_u$  is totally unrelated to the metric. From this it follows that the same must be true of its expectation. Consequently, although we require a metric to be able to define the integration in (4.7.4), the final expression must actually yield a result that is independent of  $g$  and so be a function *only* of the ‘physical’ manifold and the distribution of  $f$ .

---

<sup>15</sup>Of course, what is implicit here is that, for each  $t \in M$ , we should really write  $p$  as  $p_t$ , since it is the density of  $f_t$ . There are also a number of additional places in (4.7.4) where we could append a  $t$ , but since it has been our habit to drop the subscript when working in the setting of manifolds, we leave it out here as well.

Note that it is implicitly assumed that the integrand in (4.7.4) is a well defined  $N$ -form on  $M$ , or, equivalently, that the expectation term is a well defined Radon-Nikodym derivative. That this is the case will follow from the proof.

**Proof.** Since  $M$  is compact it has a finite atlas. Let  $(U, \varphi)$  be one of its charts and consider the random fields  $\bar{f} : \varphi(U) \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$  and  $\bar{h} : \varphi(U) \subset \mathbb{R}^N \rightarrow \mathbb{R}^K$  defined by

$$\bar{f} \triangleq f \circ \varphi^{-1}, \quad \bar{h} \triangleq h \circ \varphi^{-1}.$$

It is immediate from the definition of  $N_u$  that

$$N_u(f, h; U, B) \equiv N_u(\bar{f}, \bar{h}; \varphi(U), B),$$

and so the expectations of both of these random variables are also identical.

Recall the comments made prior to the proof: All conditions in the Theorem that involve the orthonormal frame field  $E$  also hold for any other bounded vector field on  $U \subset M$ . In particular, they hold for the natural coordinate vector field  $\{\partial/\partial x_i\}_{1 \leq i \leq N}$  determined by  $\varphi$ .

Comparing conditions (a)–(f) of the current Theorem with those in Theorem 4.1.1, it is clear that  $\bar{f}$  and  $\bar{h}$  satisfy<sup>16</sup> the conditions of Theorem 4.1.1. Consequently,

$$\mathbb{E}\{N_u(f, h; U, B)\} = \int_{\varphi(U)} \mathbb{E} \left\{ \left| \det \nabla \bar{f}(x) \right| \mathbb{1}_B(\bar{h}(x)) \Big| \bar{f}(x) = u \right\} \bar{p}_x(u) dx,$$

where  $\bar{p}_x$  is the density of  $\bar{f}(x)$ .

All that remains is to show that this is equivalent to (4.7.4) with the domain of integration restricted to  $U$ , and that we can replace  $U$  by  $M$  throughout. Consider the right hand side above, and rewrite it in terms of an integral over  $U$ . To this end, note that

$$(\nabla \bar{f}^j(x))_i = \left( \frac{\partial}{\partial x_i} f^j \right)_{t=\varphi^{-1}(x)}$$

where  $\partial/\partial x_i$  is the push-forward under  $\varphi^{-1}$  of the natural basis on  $\varphi(U)$ . Together with the definition of integration of differential forms in Section 3.6.2 this gives us that

$$\begin{aligned} & \mathbb{E}\{N_u(f, h; U, B)\} \\ &= \int_U \mathbb{E} \left\{ \left| \det (\partial/\partial x_i f^j)_t \right| \mathbb{1}_B(h(t)) \Big| \bar{f}(t) = u \right\} p_t(u) \partial x_1 \wedge \cdots \wedge \partial x_N. \end{aligned}$$

The next step involves moving from the natural basis on  $U$  to the basis given by the orthonormal frame field  $E$ . Doing so generates two multiplicative factors, which fortunately cancel. The first comes from the move from

---

<sup>16</sup>The only condition that needs any checking is (4.1.3) on the moduli of continuity. It is here that the requirement that  $g$  be  $C^2$  over  $M$  comes into play. The details are left to you.

the form  $\partial x_1 \wedge \cdots \wedge \partial x_N$  to the volume form  $\text{Vol}_g$ , and generates a factor of  $(\det(g_{ij}))^{-1/2}$ , where  $g_{ij}(t) = g_t(\partial/\partial x_i, \partial/\partial x_j)$ . (cf. (3.6.20).)

The second factor comes from noting that

$$\begin{aligned} \frac{\partial}{\partial x_i} f^j &= \sum_k g \left( \frac{\partial}{\partial x_i}, E_k \right) E_k f^j \\ &= \sum_k g_{ik}^{1/2} E_k f^j, \end{aligned}$$

where  $g^{1/2} = (g(E_i, \partial x_j))_{1 \leq i, j \leq N}$  is a square root of the matrix  $g = (g_{ij})_{1 \leq i, j \leq N}$ . Consequently,

$$\det(\partial/\partial x_i f^j)_t = \sqrt{\det(g_{ij})} \det(\nabla f_E).$$

Putting the pieces together gives us

$$(4.7.5) \quad \mathbb{E}\{N_u(U)\} = \int_U \mathbb{E} \left\{ |\det(\nabla f_E)| \mathbb{1}_B(h) \mid f = u \right\} p(u) \text{Vol}_g,$$

for each chart  $(U, \varphi)$ .

To finish the proof, note that for each chart  $(U, \varphi)$  the conditions of the Theorem imply that there are only a finite number of points in  $\varphi(U)$  at which  $\bar{f} = u$  (cf. Lemma 4.1.12) and that there are no points of this kind on  $\partial\varphi(U)$ . (cf. Lemma 4.1.10.)

Consequently, the same is true of  $f$  over  $U$ . In particular, this means that we can refine a given atlas so that each point for which  $f = u$  appears in only one chart and no chart contains more than one point of this kind. If this is the case, the integrals in (4.7.5) are either zero or one, and so it is trivial to combine them to obtain a single integral over  $M$  and so the Theorem.  $\square$

As usual, we have the following Corollary for the Gaussian case:

**Corollary 4.7.2** *Let  $(M, g)$  be a Riemannian manifold satisfying the conditions of Theorem 4.7.1. Let  $f$  and  $h$  be centered Gaussian fields over  $M$ . Then if  $f$ ,  $h$  and  $\nabla f_E$  are a.s. continuous over  $M$ , and if, for each  $t \in M$ , the joint distributions of  $(f(t), \nabla f_E(t), h(t))$  are non-degenerate, then (4.7.4) holds.*

Ultimately, we shall apply the above Corollary to obtain, among other things, an expression for the expected Euler characteristic of Gaussian excursion sets over manifolds. Firstly, however, we need to set up some machinery.

## 4.8 Riemannian structure induced by Gaussian fields

Up until now, all our work with Riemannian manifolds has involved a general Riemannian metric  $g$ . Using this, back in Section 3.6 we developed a number of concepts, starting with connections and leading up to curvature tensors and shape operators, in corresponding generality.

For our purposes, however, it will turn out that, for each random field  $f$  on a piecewise  $C^2$  manifold  $M$ , there is only one Riemannian metric that we shall need. It is induced by the random field  $f$ , which we shall assume has zero mean and, with probability one, is  $C^2$  over  $M$ . It is defined by

$$(4.8.1) \quad g_t(X_t, Y_t) \triangleq \mathbb{E}\{(X_t f) \cdot (Y_t f)\},$$

where  $X_t, Y_t \in T_t M$ , the tangent manifold to  $M$  at  $t$ .

Since the notation of (4.8.1) is rather heavy, we shall in what follows generally drop the dependence on  $t$ . Thus (4.8.1) becomes

$$(4.8.2) \quad g(X, Y) = \mathbb{E}\{X f Y f\}.$$

We shall call  $g$  the *metric induced by the random field*<sup>17</sup>  $f$ . The fact that this definition actually gives a Riemannian metric follows immediately from the positive semi-definiteness of covariance functions.

Note that, at this stage, there is nothing in the definition of the induced metric that relies on  $f$  being Gaussian<sup>18</sup>. The definition holds for any  $C^2$  random field. Furthermore, there are no demands related to stationarity, isotropy, etc.

One way to develop some intuition for this metric is via the geodesic metric  $\tau$  that it induces on  $M$ . Since  $\tau$  is given by

$$(4.8.3) \quad \tau(s, t) = \inf_{c \in D^1([0,1]; M)_{(s,t)}} \int_{[0,1]} \sqrt{g_t(c', c')(t)} dt$$

(cf. (3.6.1)) it follows that the geodesic between two points on  $M$  is the curve along which the expected variance of the derivative of  $f$  is minimised.

---

<sup>17</sup>A note for the theoretician: Recall that a Gaussian process has associated with it a natural  $L^2$  space which we denoted by  $\mathcal{H}$  in Section 2.5. The inner product between two random variables in  $\mathcal{H}$  is given by their covariance. There is also a natural geometric structure on  $\mathcal{H}$ , perhaps seen most clearly through orthogonal expansions of the form (2.5.7). In our current scenario, in which  $f$  is indexed by a manifold  $M$ , it is easy to see that the Riemannian structure induced on  $M$  by  $f$  (i.e. via the associated metric (4.8.2)) is no more than the pull-back of the natural structure on  $\mathcal{H}$ .

<sup>18</sup>Or even on  $f$  being  $C^2$ . The induced metric is well defined for  $f$  which is merely  $C^1$ . However, it is not possible to go much further – e.g. to a treatment of curvature – without more derivatives.

It is obvious that  $g$  is closely related to the covariance function  $C(s, t) = \mathbb{E}(f_s f_t)$  of  $f$ . In particular, it follows from (4.8.1) that

$$(4.8.4) \quad g_t(X_t, Y_t) = X_s X_t C(s, t) \Big|_{s=t}$$

Consequently, it is also obvious that the tools of Riemannian manifolds – connections, curvatures, etc. – can be expressed in terms of covariances. In particular, in the Gaussian case, to which we shall soon restrict ourselves, all of these tools also have interpretations in terms of conditional means and variances. Since these interpretations will play a crucial rôle in the extension of the results of Sections 4.5 and 4.6 to Gaussian fields over manifolds we shall now spend some time developing them.

#### 4.8.1 Connections and curvatures

Our first step is to describe the Levi-Civita connection  $\nabla$  determined by the induced metric  $g$ . Recall from Chapter 3 that the connection is uniquely determined by Koszul's formula,

$$(4.8.5) \quad 2g(\nabla_X Y, Z) = Xg(Y, Z) + Yg(X, Z) - Zg(X, Y) \\ + g(Z, [X, Y]) + g(Y, [Z, X]) + g(X, [Z, Y]).$$

where  $X, Y, Z$  are  $C^1$  vector fields. (cf. (3.6.6).)

Since  $g(X, Y) = \mathbb{E}\{Xf \cdot Yf\}$ , it follows that

$$\begin{aligned} Zg(X, Y) &= Z\mathbb{E}\{Xf \cdot Yf\} \\ &= \mathbb{E}\{ZXf \cdot Yf + Xf \cdot ZYf\} \\ &= g(ZX, Y) + g(X, ZY). \end{aligned}$$

Substituting this into (4.8.5) yields

$$(4.8.6) \quad g(\nabla_X Y, Z) = \mathbb{E}\{(\nabla_X Yf)(Zf)\} = \mathbb{E}\{(XYf)(Zf)\},$$

and so we have a characterisation of the connection in terms of covariances. We shall see how to exploit this important relationship to obtain more explicit representations of  $\nabla$  when we turn to specific examples in a moment.

We now turn to the curvature tensor  $R$  of (3.6.28), given by

$$R(X, Y, Z, W) = g(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z, W).$$

In order to also write  $R$  in terms of covariances, we recall (cf. (3.9.4) the covariant Hessian of a  $C^2$  function  $f$ , viz.

$$(4.8.7) \quad \nabla^2 f(X, Y) = XYf - \nabla_X Yf.$$

It follows from the fact that  $\nabla$  is torsion free (cf. (3.6.4)) that  $\nabla^2 f(X, Y) = \nabla^2 f(Y, X)$ , and so  $\nabla^2$  is a symmetric form.

With this definition, we now prove the following useful result, which relates the curvature tensor  $R$  to covariances<sup>19</sup> and is crucial for later computations.

**Lemma 4.8.1** *If  $f$  is a zero mean,  $C^2$  random field on a  $C^3$  Riemannian manifold equipped with the metric induced by  $f$  then the curvature tensor  $R$  on  $M$  is given by*

$$(4.8.8) \quad -2R = \mathbb{E} \left\{ (\nabla^2 f)^2 \right\},$$

where the square of the Hessian is to be understood in terms of the dot product of tensors developed at (3.5.10).

**Proof.** Note that for  $C^1$  vector fields it follows from the definition<sup>20</sup> (4.8.7) that

$$\begin{aligned} & (\nabla^2 f)^2 ((X, Y), (Z, W)) \\ &= 2 [\nabla^2 f(X, Z) \nabla^2 f(Y, W) - \nabla^2 f(X, W) \nabla^2 f(Y, Z)] \\ &= 2 [(XZf - \nabla_X Zf)(YWf - \nabla_Y Wf) \\ &\quad - (XWf - \nabla_X Wf)(YZf - \nabla_Y Zf)]. \end{aligned}$$

Take expectations of this expression and exploit (4.8.6) to check (after a little algebra) that

$$\begin{aligned} \mathbb{E} \left\{ (\nabla^2 f)^2 ((X, Y), (Z, W)) \right\} &= 2 (\mathbb{E}[XZf \cdot YWf] - g(\nabla_X Z, \nabla_Y W) \\ &\quad - \mathbb{E}[XWf \cdot YZf] - g(\nabla_X W, \nabla_Y Z)). \end{aligned}$$

Now apply (3.6.5) along with (4.8.6) to see that the last expression is equal to

$$\begin{aligned} & 2 (X\mathbb{E}[Zf \cdot YWf] - \mathbb{E}[Zf \cdot XYWf] - g(\nabla_X Z, \nabla_Y W) \\ &\quad - Y\mathbb{E}[XWf \cdot Zf] + \mathbb{E}[Zf \cdot YXWf] + g(\nabla_X W, \nabla_Y Z)) \\ &= 2 (Xg(Z, \nabla_Y W) - g(\nabla_X Z, \nabla_Y W) - g(Z, \nabla_{[X, Y]} W) \\ &\quad - Yg(\nabla_X W, Z) - g(\nabla_X W, \nabla_Y Z)) \\ &= 2 (g(Z, \nabla_X \nabla_Y W) - g(\nabla_Y \nabla_X W, Z) - g(Z, \nabla_{[X, Y]} W)) \\ &= 2R((X, Y), (W, Z)) \\ &= -2R((X, Y), (Z, W)), \end{aligned}$$

<sup>19</sup>Keep in mind, however, that while (4.8.8) looks like it has only geometry on the left hand side and covariances on the right, the truth is a little more complicated, since  $\nabla^2$  involves the connection which depends on the metric which depends on covariances.

<sup>20</sup>Alternatively, apply (3.5.11), treating  $\nabla^2$  as a (1, 1) rather than (2, 0) tensor.

the first equality following from the definition of the Lie bracket, the second from (3.6.5), the third from the definition of the curvature tensor  $R$  and the last is trivial.

This establishes<sup>21</sup> (4.8.8) which is what we were after.  $\square$

#### 4.8.2 Some covariances

Many of the Euclidean computations of Section 4.6 were made possible as a result of convenient independence relationships between  $f$  and its first and second order derivatives. The independence of  $f$  and  $\nabla f$  followed from the fact that  $f$  had constant variance, while that of  $\nabla f$  and the matrix  $\nabla^2 f$  followed from stationarity. Computations were further simplified by a global transformation (cf. (4.6.7)) that transformed  $f$  to isotropic.

While we shall continue to assume that  $f$  has constant variance, we no longer can assume stationarity nor find easy transformations to isotropy. However, we have invested considerable effort in setting up the geometry of our parameter space with the metric induced by  $f$ , and now we are about to start profiting from this. We start with some general computations, which require no specific assumptions.

We start with the variance function

$$\sigma_t^2 = \mathbb{E}\{f_t^2\}.$$

Assuming, as usual, that  $f \in \mathcal{C}^2(M)$ , we also have that  $\sigma^2 \in \mathcal{C}^2(M)$ , in which case there are no problems in changing the order of differentiation and expectation to see that, for  $C^1$  vector fields  $X$  and  $Y$ ,

$$(4.8.9) \quad X\sigma^2 = X\mathbb{E}\{f^2\} = 2\mathbb{E}\{f \cdot Xf\}.$$

Continuing in this vein, we have

$$XY\sigma^2 = 2(\mathbb{E}\{XYf \cdot f\} + \mathbb{E}\{Xf \cdot Yf\})$$

and

$$XY\sigma^2 - \nabla_X Y\sigma^2 = 2(\mathbb{E}\{XYf \cdot f\} + \mathbb{E}\{Xf \cdot Yf\} - \mathbb{E}\{\nabla_X Yf \cdot f\}).$$

Rearranging the last line yields

$$(4.8.10) \quad \begin{aligned} \mathbb{E}\{\nabla^2 f(X, Y) \cdot f\} &= -\mathbb{E}\{Xf \cdot Yf\} + \frac{1}{2}\nabla^2 \sigma^2(X, Y) \\ &= -g(X, Y) + \frac{1}{2}\nabla^2 \sigma^2(X, Y) \end{aligned}$$

---

<sup>21</sup>If you are a stickler for detail, you may have noticed that since our assumptions only require that  $f$  is  $C^2$ , it is not at all clear that the terms  $XYWf$  and  $YXWf$  appearing in the derivation make sense. However, their difference,  $[X, Y]Wf$ , is well defined, and that is all we have really used.

Now note that  $Xf$  and  $\nabla^2 f(Y, Z)$  are uncorrelated (and so independent in our Gaussian scenario) since

$$(4.8.11) \quad \begin{aligned} \mathbb{E}\{Xf \cdot \nabla^2 f(Y, Z)\} &= \mathbb{E}\{Xf \cdot (YZf - \nabla_Y Zf)\} \\ &= 0 \end{aligned}$$

by (4.8.6). You should note that this result requires no assumptions whatsoever. It is an immediate consequence of the geometry that  $f$  induces on  $M$  via the induced metric and the fact that the covariant Hessian  $\nabla^2$  incorporates this metric in its definition.

Putting all the above together gives that

$$\mathbb{E}\left\{\nabla^2 f_t \mid f_t = x, \nabla f_t = v\right\} = -\frac{x}{\sigma_t^2} I + \frac{x}{2\sigma_t^2} \nabla^2 \sigma_t^2,$$

where where  $I$  is the identity double form determined by  $g$ , defined at (3.5.16).

Assume now that  $f$  has constant variance, which we take to be 1. Then  $X\sigma^2 \equiv 0$  and the last equality simplifies to give

$$(4.8.12) \quad \mathbb{E}\left\{\nabla^2 f_t \mid \nabla f_t = v, f_t = x\right\} = -xI.$$

The conditional variance of  $\nabla^2 f$  is also easy to compute, since combining (4.8.8) and (4.8.10) gives what is perhaps the most crucial equality for the detailed computations that we shall carry out in Section 4.10, viz.

$$(4.8.13) \quad \begin{aligned} &\mathbb{E}\left\{\left(\nabla^2 f - \mathbb{E}\left\{\nabla^2 f \mid \nabla f = v, f = x\right\}\right)^2 \mid \nabla f = v, f = x\right\} \\ &= -(2R + I^2) \end{aligned}$$

The above correlations change somewhat if, instead of concentrating on the covariant Hessian  $\nabla^2 f$ , we look at simple second derivatives. For example, it follows from (4.8.11) that

$$(4.8.14) \quad \begin{aligned} \mathbb{E}\{XYf \cdot Zf\} &= \mathbb{E}\{\nabla_X Yf \cdot Zf\} \\ &= g(\nabla_X Y, Z). \end{aligned}$$

Continuing to assume that  $f$  has unit variance, let  $E$  be an orthonormal frame field for the induced metric  $g$ . It then immediately follows from the above and the fact that  $g(E_i, E_j) = \delta_{ij}$  that

$$(4.8.15) \quad \begin{aligned} \mathbb{E}\left\{XYf_t \mid E_k f_t = v_k, k = 1, \dots, N, f_t = x\right\} \\ &= -xI + \sum_{k=1}^N v_k g(\nabla_X Y, E_k) \\ &= -xI + g(\nabla_X Y, v) \end{aligned}$$

Now might be a good time to take some time off to look at a few examples.

### 4.8.3 Gaussian fields on $\mathbb{R}^N$

An extremely important example, which can be treated in detail without too much pain is given by the differential structure induced on a compact domain  $M$  in  $\mathbb{R}^N$  by a zero mean,  $C^2$ , Gaussian field. For the moment we shall assume that  $M$  has a  $C^2$  boundary, although at the end of the discussion we shall also treat the piecewise  $C^2$  case.

We shall show how to explicitly compute both the curvature tensor  $R$  and the shape operator  $S$  in terms of the covariance function  $C$ , as well as traces of their powers. We shall also discuss the volume form  $\text{Vol}_g$ .

We shall not, in general, assume that  $f$  is either stationary or isotropic. In fact, one of the strengths of the manifold approach is that it handles the non-stationary case almost as easily as the stationary one.

The basis for our computations is Section 3.6.4, where we saw how to compute what we need after starting with a convenient basis. Not surprisingly, we start with  $\{E_i\}_{1 \leq i \leq N}$ , the standard coordinate vector fields on  $\mathbb{R}^N$ . This also gives the natural basis in the global chart  $(\mathbb{R}^N, i)$ , where  $i$  is the inclusion map. We give  $\mathbb{R}^N$  the metric  $g$  induced by  $f$ .

From Section 3.6.4 we know that, as far as the curvature operator is concerned, everything depends on two sets of functions, the covariances

$$(4.8.16) \quad g_{ij}(t) = g(E_{ti}, E_{tj}) = \left. \frac{\partial^2 C(r, s)}{\partial r_i \partial r_j} \right|_{(t,t)}$$

and the Christoffel symbols of the first kind,

$$(4.8.17) \quad \Gamma_{ijk} \triangleq g(\nabla_{E_i} E_j, E_k) = \left. \frac{\partial^3 C(r, s)}{\partial r_i \partial r_j \partial s_k} \right|_{(t,t)},$$

where the last equality is a trivial consequence of (4.8.6) and (1.4.34).

All other terms appearing in Section 3.6.4 are derivable from these two via either simple algebraic operations or by taking inverses and square roots of matrices. In other words, there is nothing that cannot be computed directly from the first three derivatives of the covariance function. Of course, for each example, considerable perseverance or, even better, computer algebra might come in handy to actually carry out the computations.

Nevertheless, there is one rather important case in which the expressions simplify considerably. If  $f$  is stationary, then the  $g_{ij}(t)$  are actually independent of  $t$ . Consequently, it follows from (1.4.34) and the symmetry of the spectral measure that

$$(4.8.18) \quad \Gamma_{ijk} \equiv 0, \quad \text{for all } i, j, k$$

and so the curvature tensor, its powers and their traces are identically zero. As a consequence, most of the complicated formulae of Section 3.6.4 simply disappear. The isotropic situation is, of course, simpler still, since then

$$(4.8.19) \quad g_{ij}(t) \equiv \lambda_2 \delta_{ij},$$

where  $\lambda_2$  is the variance of any first order directional derivative of  $f$ . (cf. (1.4.40).)

The computation of the shape operator  $S$  also follows from the considerations of Section 3.6.4. For a specific example, assume that  $\partial M$  is a  $C^2$  manifold of co-dimension one in  $M$ . Thus the normal space to  $\partial M$  in  $M$  consists of a one dimensional vector field, which we take to be generated by an inward pointing unit normal vector field  $\nu$  on  $\partial M$ .

As we saw in Section 3.6.4, the natural choice of basis in this setting is an orthonormal frame field  $E^* = \{E_i^*\}_{1 \leq i \leq N}$  chosen so that, on  $\partial M$ ,  $E_N^* = \nu$ . In this case, we only need to know how to compute the  $\Gamma_{jiN}^*$  of (3.6.47). While this is not always easy, if  $f$  is stationary and isotropic then as for the curvature tensor things do simplify somewhat. In particular, if it is possible to explicitly determine functions  $a_{ij}$  so that

$$E_{it}^* = \sum_{k=1}^N a_{ik}(t) E_{kt},$$

then, as in Footnote 44 of Chapter 3, we have that

$$(4.8.20) \quad \Gamma_{jiN}^*(t) = \lambda_2 \sum_{k,l=1}^N a_{jk}(t) \frac{\partial}{\partial t_k} (a_{Nl}(t) a_{il}(t)),$$

so that the summation has dropped one dimension. Far more significant, however, are the facts that the information about the Riemannian structure of  $(M, g)$  is now summarized in the single parameter  $\lambda_2$  and that this information has been isolated from the ‘physical’ structure of  $\partial M$  inherent in the functions  $a_{ik}$ .

In fact, this can also be seen directly from the definition of the shape operator. From (4.8.19) it is also easy to check that, for any vectors  $X, Y$ ,

$$g(X, Y) = \lambda_2 \langle X, Y \rangle,$$

where the right hand side denotes the Euclidean inner product of  $X$  and  $Y$ . Consequently, writing  $S^g$  for the shape operator under the induced Gaussian metric and  $S^E$  for the standard Euclidean one, we have

$$(4.8.21) \quad S^g(X, Y) = \lambda_2 S^E(X, Y),$$

a result that will be useful for us later.

There is another scenario in which significant simplification occurs. Suppose that  $A \subset \partial M$  is locally flat, in the sense that  $A$  is a subset of a  $N - 1$  dimensional hyperplane. In that case the  $a_{jk}(t)$  are constant over  $A$  and so it follows from (4.8.20)  $\Gamma_{jiN}^*(t) = 0$  for all  $t \in A$ .

The last issue that we need to look at for this class of examples is that of the form of the volume element  $\text{Vol}_g$  corresponding to the metric induced

by the Gaussian field. Since our parameter space is a compact domain  $M \subset \mathbb{R}^N$  we can take an atlas consisting of the single chart  $(M, i)$ , where  $i$  is the identity mapping on  $\mathbb{R}^N$ . This being the case, it is immediate from (3.6.20) that, for any  $A \subset \mathbb{R}^N$

$$(4.8.22) \quad \int_A \text{Vol}_g = \int_A |\det \Lambda_t|^{1/2} dt,$$

where  $\Lambda_t$  is the matrix with entries

$$\lambda_t(i, j) = \mathbb{E} \left\{ \frac{\partial f}{\partial t_i} \frac{\partial f}{\partial t_j} \right\} = \left. \frac{\partial^2 C(r, s)}{\partial r_i \partial s_j} \right|_{(r, s) = (t, t)}.$$

If  $f$  is stationary, then  $\Lambda_t$  is independent of  $t$  and is simply the matrix of second order spectral moments that we met at (4.6.1) and (4.6.2).

If  $f$  is also isotropic, then  $\Lambda_t = \lambda_2 I$ , where  $I$  is the identity matrix.

## 4.9 Another Gaussian computation

At the core of the calculation of the expected Euler characteristic in the Euclidean case were the results of Lemma 4.5.2 and Corollary 4.5.3 about mean values of determinants of Gaussian matrices. In the manifold case we shall need a somewhat more general result.

To start, recall the discussion following (3.5.12). If we view a  $N \times N$  matrix  $\Delta$  as representing a linear mapping  $T_\Delta$  from  $\mathbb{R}^N$  to  $\mathbb{R}^N$ , with  $\Delta_{ij} = \langle e_i, T_\Delta e_j \rangle$ , then  $\Delta$  can also be represented as a double form  $\gamma_\Delta \in \Lambda^{1,1}(\mathbb{R}^N)$ , and from the discussion in Section 3.5.2,

$$(4.9.1) \quad \det \Delta = \frac{1}{N!} \text{Tr}((\gamma_\Delta)^N).$$

Thus it should not be surprising that we now turn our attention the expectations of traces of random double forms, for which we need a little notation.

Let  $V$  be a vector space and  $\mu \in \Lambda^{1,1}(V)$  a double form on  $V$ . Furthermore, let  $\text{Cov} : (V \otimes V) \times (V \otimes V) \rightarrow \mathbb{R}$  be bilinear, symmetric and non-negative definite. If we think of  $\mu$  as a mean function and  $\text{Cov}$  as a covariance function, then we can define a random, Gaussian, 2-form  $W$  on  $V \otimes V$  with mean function  $\mu$  and covariance function  $\text{Cov}$ , so that

$$(4.9.2) \quad W(v_1, v_2) \sim N(\mu(v_1, v_2), \text{Cov}((v_1, v_2), (v_1, v_2)))$$

for all  $v_1, v_2 \in V$ .

**Lemma 4.9.1** *With the notation and conditions described above, and understanding all powers and products of double forms as being with respect to the dot product of (3.5.10),*

$$(4.9.3) \quad \mathbb{E} \{W^k\} = \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \frac{k!}{(k-2j)!j!2^j} \mu^{k-2j} C^j$$

in the sense that, for all  $v_1, \dots, v_k, v'_1, \dots, v'_k \in V$ ,

$$\begin{aligned} & \mathbb{E} \{W^k((v_1, \dots, v_k), (v'_1, \dots, v'_k))\} \\ &= \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \frac{k!}{(k-2j)!j!2^j} (\mu^{k-2j} C^j) ((v_1, \dots, v_k), (v'_1, \dots, v'_k)), \end{aligned}$$

where  $C \in \Lambda^{2,2}(V)$  is defined by

$$\begin{aligned} & C((v_1, v_2), (v'_1, v'_2)) \\ &= \mathbb{E} \left\{ (W - \mathbb{E}\{W\})^2((v_1, v_2), (v'_1, v'_2)) \right\} \\ &= 2 (\text{Cov}((v_1, v'_1), (v_2, v'_2)) - \text{Cov}((v_1, v'_2), (v_2, v'_1))). \end{aligned}$$

**Proof.** Since it is easy to check that the standard binomial expansion works also for dot products, the general form of (4.9.3) will follow from the special case  $\mu = 0$ , once we show that for this case

$$(4.9.4) \quad \mathbb{E} \{W^k\} = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \frac{(2j)!}{j!2^j} C^j & \text{if } k = 2j. \end{cases}$$

Thus, assume now that  $\mu = 0$ . The case of odd  $k$  in (4.9.4) follows immediately from (4.5.1) and so we have only the even case to consider.

Recalling the definition (3.5.10) of the dot product of double forms we have that

$$W^{2j}((v_1, \dots, v_{2j}), (v'_1, \dots, v'_{2j})) = \sum_{\pi, \sigma \in S(2j)} \varepsilon_\pi \varepsilon_\sigma \prod_{k=1}^{2j} W(v_{\pi(k)}, v'_{\sigma(k)}),$$

where, as usual,  $S(2j)$  is the symmetric group of permutations of  $2j$  letters and  $\varepsilon_\sigma$  is the sign of the permutation  $\sigma$ . It then follows immediately from (4.5.2) that the expectation on the right hand side is given by

$$\frac{(2j)!}{j!} \sum_{\pi, \sigma \in S(2j)} \varepsilon_\pi \varepsilon_\sigma \prod_{k=1}^j \mathbb{E} \left\{ W(v_{\pi(2k-1)}, v'_{\sigma(2k-1)}) W(v_{\pi(2k)}, v'_{\sigma(2k)}) \right\},$$

where the combinatorial factor comes from the different ways of grouping the vectors  $(v_{\pi(k)}, v'_{\sigma(k)})$ ,  $1 \leq k \leq 2j$ , into ordered pairs<sup>22</sup>.

The last expression can be rewritten as

$$\begin{aligned} & \frac{(2j)!}{j!2^j} \sum_{\pi, \sigma \in S(2j)} \varepsilon_{\pi} \varepsilon_{\sigma} \prod_{k=1}^j \mathbb{E} \left\{ W(v_{\pi(2k-1)}, v'_{\sigma(2k-1)}) W(v_{\pi(2k)}, v'_{\sigma(2k)}) \right. \\ & \qquad \qquad \qquad \left. - W(v_{\pi(2k)}, v'_{\sigma(2k-1)}) W(v_{\pi(2k-1)}, v'_{\sigma(2k)}) \right\} \\ & = \frac{(2j)!}{j!2^j} \sum_{\pi, \sigma \in S(2j)} \varepsilon_{\pi} \varepsilon_{\sigma} \prod_{k=1}^j C((v_{\pi(2k-1)}, v_{\pi(2k)}), (v'_{\sigma(2k-1)}, v'_{\sigma(2k)})) \\ & = \frac{(2j)!}{j!2^j} C^j((v_1, \dots, v_{2j}), (v'_1, \dots, v'_{2j})), \end{aligned}$$

which completes the proof. □

The following corollary is immediate from Lemma 4.9.1 and the definition (3.5.12) of the trace operator.

**Corollary 4.9.2** *With the notation and conditions of Lemma 4.9.1,*

$$(4.9.5) \quad \mathbb{E} \{ \text{Tr}(W)^k \} = \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \frac{k!}{(k-2j)!j!2^j} \text{Tr}(\mu^{k-2j} C^j).$$

## 4.10 Mean Euler characteristics: Manifolds

We now have everything we need to undertake the task of developing an explicit formula for  $\mathbb{E}\{\varphi(A_u(f, M))\}$  for a smooth, Gaussian random field  $f$  over a manifold  $M$ . We shall treat the cases in which  $M$  does and does not have a boundary separately, even though the first scenario is a special case of the second. Nevertheless, it is best to see the calculations, for the first time, in the simpler scenario. When we get around to adding in the all the parts of piecewise smooth manifolds the notation will become very heavy, although the main result, Theorem 4.10.2, will not look very different to the non-boundary case of Theorem 4.10.1.

---

<sup>22</sup>In comparing this with (4.5.2), note that there we had an extra summation over the groupings into *unordered* pairs rather than a simple multiplicative factor. We already have each possible grouping due to the summation over  $\pi$  and  $\sigma$  in  $S(2j)$ , and since we are keeping pairs ordered we also lose the factor of  $2^{-j}$  there.

### 4.10.1 Manifolds without boundary

Throughout this Section, we shall assume that  $M$  is a  $C^2$  submanifold of a  $C^3$  manifold. Here is the main result:

**Theorem 4.10.1** *Let  $f$  be a centered, unit variance Gaussian field on  $M$  and satisfying the conditions of Corollary 4.2.5. Then*

$$(4.10.1) \quad \mathbb{E} \{ \varphi(A_u) \} = \sum_{j=0}^N \mathcal{L}_j(M) \rho_j(u),$$

where the  $\rho_j$  are given by

$$(4.10.2) \quad \rho_j(u) = (2\pi)^{-(j+1)/2} H_{j-1}(u) e^{-\frac{u^2}{2}}, \quad j \geq 0,$$

$H_j$  is the  $j$ -th Hermite polynomial, given by (4.5.8) and (4.5.9), and the  $\mathcal{L}_j(M)$  are the Lipschitz-Killing curvatures (3.8.1) of  $M$ , calculated with respect to the metric (4.8.2) induced by  $f$ ; viz.

$$(4.10.3) \quad \begin{aligned} & \mathcal{L}_j(M, U) \\ &= \begin{cases} \frac{(-2\pi)^{-(N-j)/2}}{(N-j)!} \int_U \text{Tr}^M (R^{(N-j)/2}) \text{Vol}_g & \text{if } N-j \text{ is even,} \\ 0 & \text{if } N-j \text{ is odd.} \end{cases} \end{aligned}$$

**Proof.** The first consequence of the assumptions on  $f$  is that the sample functions of  $f$  are almost surely Morse functions over  $M$ . Thus Corollary 3.9.3 gives us that

$$\varphi(A_u) = \sum_{k=0}^N (-1)^k \# \{ t \in M : f_t \geq u, \nabla f_t = 0, \text{index}(-\nabla^2 f_t) = k \}.$$

To compute the expectation here first choose an orthonormal frame field  $E = (E_1, \dots, E_N)$  for  $M$  and apply Theorem 4.7.1  $N+1$  times. If we write the  $f$  there as  $f^*$ , then we set  $f^* = \nabla f_E$  and  $h$  to be  $(f, -(E_i E_j) f)$ , where, to save on subscripts, we shall write  $(E_i E_j f)$  to denote the matrix  $(E_i E_j f)_{i,j=1,\dots,N}$ .

We read off the components of  $(E_i E_j f)$  (and later of  $\nabla^2 f_E$ ) in lexicographic order to get a vector. To avoid confusion, we write everything out once more in full:

$$\nabla f_{E,k} = E_k f, \quad (E_i E_j f)_{kl} = E_k E_l f, \quad \nabla^2 f_{E,kl} = \nabla^2 f(E_k, E_l),$$

Finally, for the  $k$ -th application of Theorem 4.7.1 we set

$$B = (u, \infty) \times A_k \stackrel{\Delta}{=} (u, \infty) \times \{ H : \text{index}(-H) = k \} \subset \mathbb{R} \times \text{Sym}_{N \times N},$$

where  $\text{Sym}_{N \times N}$  is the space of symmetric  $N \times N$  matrices. Then the second consequence of our assumptions on  $f$  is that Theorem 4.7.1 is applicable in this setting. Applying it gives

$$\begin{aligned} \mathbb{E} \{ \varphi(A_u) \} &= \sum_{k=0}^N \int_M (-1)^k \mathbb{E} \{ |\det(E_i E_j f)| \\ &\quad \times \mathbb{1}_{A_k}(\nabla^2 f_E) \mathbb{1}_{(u, \infty)}(f) \mid \nabla f = 0 \} p_{\nabla f(0)} \text{Vol}_g, \end{aligned}$$

Recall now that  $\nabla^2 f(E_i, E_j) = E_i E_j f - \nabla_{E_i} E_j f$ . However, conditioning on  $\nabla f_E = 0$  gives  $\nabla^2 f(E_i, E_j) \equiv E_i E_j f$ , so that we can replace  $(E_i E_j f)$  by  $\nabla^2 f_E$  in the last equation above to obtain

$$\sum_{k=0}^N \int_M (-1)^k \mathbb{E} \left\{ |\det \nabla^2 f_E| \mathbb{1}_{A_k}(\nabla^2 f_E) \mathbb{1}_{(u, \infty)}(f) \mid \nabla f = 0 \right\} p_{\nabla f(0)} \text{Vol}_g.$$

If we now interchange summation and integration and bracket the factor of  $(-1)^k$  together with  $|\det \nabla^2 f_E|$  then we can drop the absolute value sign on the latter, although there is a remaining factor of  $-1$ . This allows us to exchange expectation with summation, and the factor of  $\mathbb{1}_{A_k}(\nabla^2 f_E)$  and the sum over  $k$  disappear completely.

Now recall that  $f$  has constant variance and note that since  $E$  is an orthonormal frame field (with respect to the induced metric  $g$ ) the components of  $\nabla f_E$  at any  $t \in M$  are all independent standard Gaussians. Furthermore, as we saw in Section 4.8.2, the constant variance of  $f$  implies that they are also independent of  $f(t)$ . Consequently, the joint probability density of  $(f, \nabla f_E)$  at the point  $(x, 0)$  is simply

$$\frac{e^{-x^2/2}}{(2\pi)^{(N+1)/2}}.$$

Thus, not only is it known, but it is constant over  $M$ .

Noting all this, conditioning on  $f$  and integrating out the conditioning allows us to rewrite the above in the much simpler format

$$\begin{aligned} (4.10.4) \quad \mathbb{E} \{ \varphi(A_u) \} &= (2\pi)^{-(N+1)/2} \int_u^\infty e^{-x^2/2} dx \\ &\quad \times \int_M \mathbb{E} \{ \det(-\nabla^2 f_E) \mid \nabla f_E = 0, f = x \} \text{Vol}_g. \end{aligned}$$

Recalling the definition of the trace (cf.(4.9.1)), the innermost integrand can be written as

$$\frac{1}{N!} \mathbb{E} \{ \text{Tr}((-\nabla^2 f)^N) \mid \nabla f_E = 0, f = x \}.$$

Since  $\nabla^2 f$  is a Gaussian double form, we can use Corollary 4.9.2 to compute the above expectation, once we recall (4.8.12) and (4.8.13) to give us the conditional mean and covariance of  $\nabla^2 f$ . These give

$$\begin{aligned}
& \frac{(-1)^N}{N!} \mathbb{E} \left\{ \text{Tr}^M (\nabla^2 f)^N \mid \nabla f_E = 0, f = x \right\} \\
&= \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \frac{(-1)^j}{(N-2j)!j!2^j} \text{Tr}^M ((xI)^{N-2j} (I^2 + 2R)^j) \\
&= \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \sum_{l=0}^j \frac{(-1)^j}{j!2^j(N-2j)!} x^{N-2j} \text{Tr}^M \left( I^{N-2l} \binom{j}{l} (2R)^l \right) \\
&= \sum_{l=0}^{\lfloor \frac{N}{2} \rfloor} \frac{(-1)^l}{l!} \text{Tr}^M \left( R^l \sum_{k=0}^{\lfloor \frac{N-2l}{2} \rfloor} \frac{(-1)^k}{2^k(N-2k-2l)!k!} x^{N-2k-2l} I^{N-2l} \right) \\
&= \sum_{l=0}^{\lfloor \frac{N}{2} \rfloor} \frac{(-1)^l}{l!} \text{Tr}^M (R^l) H_{N-2l}(x)
\end{aligned}$$

where, in the last line, we have used (3.5.17) and the definition (4.5.8) of the Hermite polynomials.

Substituting back into (4.10.4) we conclude that  $\mathbb{E} \{ \varphi(A_u) \}$  is given by

$$\begin{aligned}
& \sum_{l=0}^{\lfloor \frac{N}{2} \rfloor} \left[ \int_u^\infty (2\pi)^{-(N+1)/2} H_{N-2l}(x) e^{-x^2/2} dx \right] \frac{(-1)^l}{l!} \int_M \text{Tr}^M (R^l) \text{Vol}_g \\
&= \sum_{l=0}^{\lfloor \frac{N}{2} \rfloor} (2\pi)^{-(N+1)/2} H_{N-2l-1}(x) e^{-u^2/2} \frac{(-1)^l}{l!} \int_M \text{Tr}^M (R^l) \text{Vol}_g \\
&= \sum_{j=0}^N \mathcal{L}_j(M) \rho_j(u),
\end{aligned}$$

where the first equality follows from (4.5.11) and the second from the definitions (4.10.2) of the  $\rho_j$  and (4.10.3) for the  $\mathcal{L}_j$ , along with a little algebra.

That is, we have (4.10.1) and so the Theorem.  $\square$

#### 4.10.2 Manifolds with boundary

We now turn to the case of manifolds with boundaries, which will incorporate, in one main result, both the results of the previous Section where the manifold had no boundary, and the results of Section 4.6. There, as you will recall, the parameter space was a  $N$ -dimensional rectangle, and the Gaussian process was required to be stationary.

Thus, we return to the setting of Sections 3.7 and 3.9, and take our manifold  $M$  be a  $C^2$  piecewise smooth,  $N$ -dimensional submanifold of a  $C^3$  Riemannian manifold  $\widetilde{M}$ . We also require that all support cones of  $M$  are convex and that the normal cones at points in  $\partial M$  are non-empty.

**Theorem 4.10.2** *Let  $M$  be as above, and  $f$ , as in Theorem 4.10.1, a centered, unit variance Gaussian field on  $M$  satisfying the conditions of Corollary 4.2.5. Then  $\mathbb{E}\{\varphi(A_u)\}$  is again given by (4.10.1), so that*

$$(4.10.5) \quad \mathbb{E}\{\varphi(A_u)\} = \sum_{j=0}^N \mathcal{L}_j(M) \rho_j(u),$$

with the single change that the Lipschitz-Killing curvatures  $\mathcal{L}_j$  are now defined by (3.8.2), viz.

$$(4.10.6) \quad \begin{aligned} \mathcal{L}_j(M) &= \pi^{-(N-j)/2} \sum_{k=j}^N \sum_{l=0}^{\lfloor \frac{k-j}{2} \rfloor} \frac{(-1)^l \Gamma\left(\frac{N-j-2l}{2}\right) 2^{-(l+2)}}{l!(k-j-2l)!} \\ &\quad \times \int_{U \cap \partial_k M} \mathcal{H}_k(dt) \int_{S(N_t \partial_k M)} \text{Tr}^{\partial_k M} (S_\nu^{k-j-2l} R^l) \mathcal{H}_{N-k-1}(d\nu). \end{aligned}$$

**Proof.** In essence, although this result incorporates many others that have gone before, it will not be hard to prove. We have already done most of the hard work in proving some of the earlier cases, and so we now face a proof that is more concerned with keeping track of notation than needing any serious new computations. Nevertheless, there is something new here, and that is the way integration of normal cones is handled. This actually makes up most of the proof.

We start by recalling the setup, which implies that  $M$  has the unique decomposition

$$M = \bigcup_{j=0}^N \partial_j M,$$

as in (3.7.8). Morse's Theorem, as given in Corollary 3.9.3 and slightly rewritten, states that

$$(4.10.7) \quad \varphi(A_u) = \sum_{k=0}^N \sum_{j=0}^k (-1)^j \mu_{kj},$$

where

$$\mu_{kj} \triangleq \#\{t \in \partial_k M : f(t) \geq u, -\nabla f(t) \in N_t M, \text{index}(-\nabla^2 f|_{\partial_k M}(t)) = j\}$$

and  $N_t M$  is the normal cone given by (3.7.10). Since for  $t \in M^\circ$  the normal cone is  $\{0\}$ , the expectation of the  $k = N$  term in (4.10.7) has already been computed, since this is the computation of the previous proof. Nevertheless, we shall rederive it, *en passant*, in what follows. For this case, however, keep in mind that statements like “ $j \geq k + 1$ ” should be interpreted as “for no  $j$ ”. A similar interpretation should be given to statements like “ $1 \leq j \leq k$ ” for the other extreme, when  $k = 0$ .

Fix a  $k$ ,  $0 \leq k \leq N$  and choose an orthonormal frame field  $E$  so that  $E_{k+1}, \dots, E_N$  all point into  $M$ . With this choice, the condition  $-\nabla f_t \in N_t M$  is equivalent to  $E_i f_t = 0$  for  $1 \leq i \leq k$  and, for  $k < i \leq N$ ,  $E_i f_t \in N_t \partial_k M$ , the normal cone in  $T_t M$  to  $\partial_k M$ .

We shall use  $\mathcal{H}_k$  to denote the volume (Hausdorff) measures that  $g$  induces on  $\partial_k M$  and  $\mathcal{H}_{N-k}$  for the corresponding measure on  $N_t \partial_k M$ . (Of course  $\mathcal{H}_{N-k}$  really depends on  $t$ , but we have more than enough subscripts already.)

Since  $E$  is orthonormal with respect to the induced metric, the  $E_i f_t$  are all independent standard Gaussians for each  $t$ , so that arguing as in the proof of Theorem 4.10.1 (cf. the argument leading up to (4.10.4)) we have the following, in which  $\nu$  is a  $(N - k)$ -dimensional vector which, for convenience, we write as  $(\nu_{k+1}, \dots, \nu_N)$ .

$$\begin{aligned} \mathbb{E} \left\{ \sum_{j=0}^k (-1)^j \mu_{kj} \right\} &= (2\pi)^{-(N+1)/2} \int_u^\infty e^{-x^2/2} dx \\ &\quad \times \int_{\partial_k M} \mathcal{H}_k(dt) \int_{N_t \partial_k M} e^{-|\nu|^2/2} \theta_t^k(\nu, x) \mathcal{H}_{N-k}(d\nu), \end{aligned}$$

where  $\theta_t^k(\nu, x)$  is given by

$$\begin{aligned} \mathbb{E} \left\{ \det \left( -(\nabla^2 f(E_m, E_n))_{1 \leq n, m \leq k} \right) \right. \\ \left. \middle| f = x, E_m f = 0, 1 \leq m \leq k, E_m f = \nu_m, k + 1 \leq m \leq N \right\}. \end{aligned}$$

We need to compute a few basic expectations before we can continue. The conditional covariances of the Hessian remain the same as in the previous Section, as we are still conditioning on the vector  $(f, \nabla f_E)$ . Specifically,

$$\text{Var} \left( \nabla^2 f|_{\partial_k M} \middle| f = x, \nabla f_E = (0, \nu) \right) = -(2R + I^2),$$

where the zero vector here is of length  $k$ .

Conditional means, however, do change, and from (4.8.15) we have, for  $X, Y \in C^2(T(\partial_k M))$ ,

$$\begin{aligned} \mathbb{E} \left\{ XYf \mid f = x, \nabla f_E = (0, \nu) \right\} \\ &= \mathbb{E} \left\{ \nabla^2 f|_{\partial_k M}(X, Y) \mid f = x, \nabla f_E = (0, \nu) \right\} \\ &= -xg(X, Y) + g(\nabla_X Y, (0, \nu)) \\ &= -xg(X, Y) - S_\nu(X, Y), \end{aligned}$$

where  $S_\nu$  is the usual (scalar) second fundamental form, given by (3.6.33) and (3.6.35). Equivalently,

$$\mathbb{E} \left\{ \nabla^2 f|_{\partial_k M} \mid f = x, \nabla f_E = (0, \nu) \right\} = -xI - S_\nu.$$

We now have all we need to evaluate  $\theta_t^k(\nu, x)$ , which, following the argument of the preceding Section and using the above conditional expectations and variance is equal to

$$\begin{aligned} \frac{(-1)^k}{k!} \mathbb{E} \left\{ \text{Tr}^{\partial_k M} (\nabla^2 f|_{\partial_k M})^k \mid f = x, \nabla f_E = (0, \nu) \right\} \\ = \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^j}{(k-2j)!j!2^j} \text{Tr}^{\partial_k M} \left( (xI + S_\nu)^{k-2j} (I^2 + 2R)^j \right) \end{aligned}$$

by Lemma 4.9.1. Rearranging somewhat, this is equal to

$$\begin{aligned} \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \sum_{l=0}^j \frac{(-1)^j}{(k-2j)!} \frac{1}{2^{j-l}l!(j-l)!} \text{Tr}^{\partial_k M} \left( (xI + S_\nu)^{k-2j} I^{2j-2l} R^l \right) \\ = \sum_{j=0}^{\lfloor \frac{k}{2} \rfloor} \sum_{m=0}^{k-2j} \sum_{l=0}^j \frac{(-1)^j (k-2l-m)!}{(k-2j-m)!m!} \frac{1}{2^{j-l}l!(j-l)!} \\ \quad \times x^{k-2j-m} \text{Tr}^{\partial_k M} (S_\nu^m R^l) \end{aligned}$$

by (3.5.17). Further rearrangement gives that this is the same as

$$\begin{aligned}
 & \sum_{m=0}^k \sum_{j=0}^{\lfloor \frac{k-m}{2} \rfloor} \sum_{l=0}^j \frac{(-1)^j (k-2l-m)!}{(k-2j-m)!m!} \frac{1}{2^{j-l}l!(j-l)!} \\
 & \quad \times x^{k-2j-m} \mathrm{Tr}^{\partial_k M} (S_\nu^m R^l) \\
 &= \sum_{m=0}^k \sum_{l=0}^{\lfloor \frac{k-m}{2} \rfloor} \sum_{j=l}^{\lfloor \frac{k-m}{2} \rfloor} \frac{(-1)^j (k-2l-m)!}{(k-2j-m)!m!} \frac{1}{2^{j-l}l!(j-l)!} \\
 & \quad \times x^{k-2j-m} \mathrm{Tr}^{\partial_k M} (S_\nu^m R^l) \\
 &= \sum_{m=0}^k \sum_{l=0}^{\lfloor \frac{k-m}{2} \rfloor} \frac{(-1)^l \mathrm{Tr}^{\partial_k M} (S_\nu^m R^l)}{m!l!} \\
 & \quad \times \sum_{i=0}^{\lfloor \frac{k-2l-m}{2} \rfloor} \frac{(-1)^i (k-2l-m)!}{(k-2l-m-2i)!} \frac{1}{2^i i!} x^{k-2l-m-2i} \\
 &= \sum_{m=0}^k \sum_{l=0}^{\lfloor \frac{k-m}{2} \rfloor} \frac{(-1)^l}{l!m!} \mathrm{Tr}^{\partial_k M} (S_\nu^m R^l) H_{k-2l-m}(x) \\
 &= \sum_{m=0}^k \sum_{l=0}^{\lfloor \frac{m}{2} \rfloor} \frac{(-1)^l}{l!(m-2l)!} \mathrm{Tr}^{\partial_k M} (S_\nu^{m-2l} R^l) H_{k-m}(x) \\
 &= \sum_{j=0}^k H_j(x) \sum_{l=0}^{\lfloor \frac{k-j}{2} \rfloor} \frac{(-1)^l}{l!(k-j-2l)!} \mathrm{Tr}^{\partial_k M} (S_\nu^{k-j-2l} R^l).
 \end{aligned}$$

We now fix a  $t \in \partial_k M$  and concentrate on computing

$$(4.10.8) \quad \int_{N_t \partial_k M} e^{-|\nu|^2/2} \mathrm{Tr}^{\partial_k M} (S_\nu^{k-j-2l} R^l) \mathcal{H}_{N-k}(d\nu).$$

Firstly, write  $S(N_t \partial_k M) \triangleq \{\nu \in N_t \partial_k M : |\nu| = 1\}$  for the intersection of the sphere bundle of  $M$  with  $N_t \partial_k M$ . Make the usual identification between  $N_t \partial_k M$  and  $\mathbb{R}_+ \times S(N_t \partial_k M)$  by  $\nu \leftrightarrow (|\nu|, \nu/|\nu|)$ . Consequently, we can rewrite (4.10.8) as

$$\begin{aligned}
 & \int_{S(N_t \partial_k M)} \int_0^\infty e^{-r^2/2} r^{N-j-2l-1} \mathrm{Tr}^{\partial_k M} (S_\nu^{k-j-2l} R^l) dr \mathcal{H}_{N-k-1}(d\nu) \\
 &= \Gamma\left(\frac{N-j-2l}{2}\right) 2^{(N-j-2l-2)/2} \\
 & \quad \times \int_{S(N_t \partial_k M)} \mathrm{Tr}^{\partial_k M} (S_\nu^{k-j-2l} R^l) \mathcal{H}_{N-k-1}(d\nu),
 \end{aligned}$$

where, in the power of  $r$  in the first line, we get  $\max(0, N-k-1)$  from the change of variables and  $k-j-2l$  from the fact that  $S_\nu = |\nu|S_{\nu/|\nu|}$ . We are

also now using  $\mathcal{H}_{N-k-1}$  to denote the volume form induced on  $S(N_t\partial_k M)$  by  $g$ .

Collecting all the above, we finally have that

$$\begin{aligned} \mathbb{E}\{\varphi(A_u)\} &= \sum_{k=0}^N (2\pi)^{-(N+1)/2} \int_0^\infty e^{-x^2/2} \sum_{j=0}^k H_j(x) \\ &\quad \times \sum_{l=0}^{\lfloor \frac{k-j}{2} \rfloor} \frac{(-1)^l \Gamma\left(\frac{N-j-2l}{2}\right) 2^{(N-j-2l-2)/2}}{l!(k-j-2l)!} \int_{\partial_k M} \mathcal{H}_k(dt) \\ &\quad \times \int_{S(N_t\partial_k M)} \text{Tr}^{\partial_k M}(S_\nu^{k-j-2l} R^l) dx \mathcal{H}_{N-k-1}(d\nu) \\ &= \sum_{j=0}^N (2\pi)^{-(j+1)/2} e^{-u^2/2} H_{j-1}(u) \sum_{k=j}^N p_i^{-(N-j)/2} \\ &\quad \times \sum_{l=0}^{\lfloor \frac{k-j}{2} \rfloor} \frac{(-1)^l \Gamma\left(\frac{N-j-2l}{2}\right) 2^{-(l+1)}}{l!(k-j-2l)!} \int_{\partial_k M} \mathcal{H}_k(dt) \\ &\quad \times \int_{S(N_t\partial_k M)} \text{Tr}^{\partial_k M}(S_\nu^{k-j-2l} R^l) \mathcal{H}_{N-k-1}(d\nu), \end{aligned}$$

after integrating out  $x$  (via (4.5.11)) and changing the order of summation.

Comparing the last expression with the definitions (4.10.2) of the  $\rho_j$  and (4.10.6) of the  $\mathcal{L}_j$ , the proof is complete.  $\square$

### 4.11 Examples

With the hard work behind us, we can now look at some applications of Theorems 4.10.1, and 4.10.2. One of the most powerful implications of the formula

$$(4.11.1) \quad \mathbb{E}\{\varphi(A_u)\} = \sum_{j=0}^N \mathcal{L}_j(M) \rho_j(u),$$

is that, for any example, all that needs to be computed are the Lipschitz-Killing curvatures  $\mathcal{L}_j(M)$ , since the  $\rho_j$  are well defined by (4.10.2) and dependent neither on the geometry of  $M$  nor the covariance structure of  $f$ .

Nevertheless, this is not always easy and there is no guarantee that explicit forms for the  $\mathcal{L}_j(M)$  exist. In fact, more often than not, this will unfortunately be the case, and one needs to turn to a computer for assistance, performing either (or often both) symbolic or numeric evaluations.

However, there are some cases that are not too hard and so we shall look at these.

**Stationary fields over rectangles.** This is the example that we treated in Theorem 4.6.2 via the techniques of Integral rather than Differential Geometry.

Nevertheless, since  $N$ -dimensional rectangles are definitely piecewise  $C^2$  manifolds, we should be able to recover Theorem 4.6.2 from Theorem 4.10.2. Doing so is in fact quite easy, so we set  $M = \prod_{i=1}^N [0, T_i]$  and, to make life even easier, assume that  $f$  has unit variance and is isotropic with the variance of its derivatives given by  $\lambda_2$ . Thus, what we are trying to recover is (4.6.12) with  $\Lambda_J = \lambda_2 I$ , where  $I$  is the identity matrix.

The first point to note is that induced Riemannian  $g$  is given by

$$g(X, Y) = \mathbb{E}\{XfYf\} = \lambda_2 \langle X, Y \rangle,$$

where the last term is the simple Euclidean inner product. Thus  $g$  changes the usual flat Euclidean geometry of  $\mathbb{R}^N$  only by scaling, and so the geometry remains flat.

This being the case (3.4.6) gives us the necessary Lipschitz-Killing curvatures, although each  $\mathcal{L}_j$  of (3.4.6) needs to be multiplied by a factor of  $\lambda_2^{j/2}$ . Substituting this into (4.11.1) gives the required result; viz. (4.6.12).

The few lines of algebra needed along the way are left to you.

Also left to you is the non-isotropic case, which is not much harder, although you will need a slightly adapted version of (3.4.6) which allows for a metric that is a constant times the Euclidean metric on hyperplanes in  $\mathbb{R}^N$ , but for which the constant depends on the hyperplane. This will give Theorem 4.6.2 in its full generality.

**Isotropic fields over smooth domains.** In the previous example, we assumed isotropy only for convenience, and leaving the argument for the general stationary case to you was simply to save us having to do more complicated algebra.

However, once one leaves the setting of rectangles it is almost impossible to obtain simple closed form expressions for the  $\mathcal{L}_j$  if  $f$  is not isotropic. To see how isotropy helps, we now take  $f$  isotropic over a compact, piecewise  $C^2$  domain in  $\mathbb{R}^N$ , and also assume that  $\mathbb{E}\{f^2\} = \mathbb{E}\{(\partial f / \partial t_k)^2\} = 1$ , so as to save carrying through an awkward constant.

What makes the Lipschitz-Killing curvatures simpler in this case is not that their defining formula (4.10.6) changes in any appreciable way, but that the symbols appearing in it have much simpler meanings. In particular,  $\mathcal{H}_k$  is no more than standard Hausdorff measure over  $\partial_k M$ , while  $\mathcal{H}_{N-k-1}$  becomes surface measure on the unit sphere  $S^{N-k-1}$ . In other words, the  $\mathcal{L}_k$  no longer carry any information related to  $f$ .

The Riemannian curvature  $R$  is now zero, and the second fundamental form  $S$  simplifies considerably. To see how this works in an example, assume that  $M$  is a  $C^2$  domain, so that there is only one  $C^2$  component to its boundary, of dimension  $N - 1$ . Then  $\mathcal{L}_N(M) = \mathcal{H}_N(M)$  and from (3.8.6)

we have that, for  $0 \leq j \leq N - 1$ ,

$$(4.11.2) \quad \mathcal{L}_j(M) = \frac{1}{s_{N-j}} \int_{\partial M} \text{detr}_{N-1-j}(\text{Curv}) \mathcal{H}_{N-1}.$$

where  $\text{detr}_j$  is given by (3.5.14) and the curvature matrix  $\text{Curv}$  is given by (3.8.7)

A simple example was given in Figure 3.3.3, for which we discussed finding, via Integral Geometric techniques, the Euler characteristic of  $A_u(f, M)$ , where  $M$  was a  $C^2$  domain in  $\mathbb{R}^2$ . Although we found a point process representation for the  $\varphi(A_u)$  we never actually managed to use Integral Geometric techniques to find its expectation. The reason is that Differential Geometric techniques work much better, since applying (4.11.2) with  $N = 2$  immediately gives us the very simple result that

$$\mathcal{L}_2(M) = \text{Area}(M), \quad \mathcal{L}_1(M) = \frac{\text{length}(\partial M)}{2}, \quad \mathcal{L}_0(M) = 1,$$

which, when substituted into (4.11.1), gives the required expectation.

An historical note is appropriate here: It was the example of isotropic fields on  $C^2$  domains which, in Keith Worsley's paper [106], was really the genesis of the manifold approach to Gaussian geometry that has been the central point of this Chapter and the reason for writing this book.

Under isotropy, you should now be able to handle other examples yourself. All reduce to calculations of Lipschitz-Killing curvatures under a constant multiple of the standard Euclidean metric, and for many simple cases, such as balls and spheres, these have already been computed in Chapter 3. What is somewhat harder is the non-isotropic case.

**Stationary and non-stationary fields on  $\mathbb{R}^N$ .** It would be nice to be able to find nice formulae for non-stationary Gaussian fields over smooth domains, and even for stationary but non-isotropic fields, as we have just done for isotropic fields over smooth domains and stationary processes over rectangles. Unfortunately, although Theorems 4.10.1 and 4.10.2 allow us to do this in principle, it is not so simple to do in practice.

The basic reason for this can already be seen in the stationary scenario of Theorem 4.6.2, from which one can see that the Lipschitz-Killing curvatures for a stationary process over a rectangle  $T = \left(\prod_{i=1}^N [0, T_i]\right)$  are given by

$$(4.11.3) \quad \mathcal{L}_j(T) = \sum_{J \in \mathcal{O}_j} |J| |\Lambda_J|^{1/2}$$

the sum being over faces of dimension  $j$  in  $T$  containing the origin and the rest of the notation as in Theorem 4.6.2. What (4.11.3) shows is that there is no simple averaging over the boundary of  $T$  as there is in the isotropic case. Each piece of the boundary has its own contribution to make, with

its own curvature and second fundamental form. In the case of a rectangle this is not too difficult to work with. In the case of a general domain it is not so simple.

In Section 3.6.4 we saw how to compute curvatures and second fundamental forms on Euclidean surfaces in terms of Christoffel symbols. In Section 4.8.3 we saw how to compute Christoffel symbols for the metric induced by  $f$  in terms of its covariance function. For any given example, these two computations need to be coordinated and then fed into definitions such as (4.10.3) and (4.10.6) of Lipschitz-Killing curvatures. From here to a final answer is a long path, often leading through computer algebra packages.

There is, however, one negative result that is worth mentioning here, since it is sufficiently anti-intuitive that it has led many to an incorrect conjecture. As we mentioned following the proof of Lemma 4.6.1, a first guess at extending (4.11.3) to the non-stationary case would be to replace the terms  $|J| |\Lambda_J|^{1/2}$  there by integrals of the form  $\int_J |\Lambda_t|^{1/2} dt$  where the elements of  $\Lambda_t$  are the covariances  $\mathbb{E}\{f_i(t)f_j(t)\}$ . Without quoting references, it is a fact that this has been done more than once in the past. However, it is clear from the computations of the Christoffel symbols in Section 4.8.3 that this does not work, and additional terms involving third order derivatives of the covariance function enter into the computation. The fact that these terms are all identically zero in the stationary case is probably what led to the errors<sup>23</sup>.

**Stationary fields over Lie groups.** Lie groups provide an example which, while perhaps a little abstract, still yields a simple form for the expected Euler characteristic of excursion sets. We first met Lie groups back in Section 1.4.2, where we discussed them in the framework of stationarity.

Recall that a Lie group  $G$  is a group that is also a  $C^\infty$  manifold, such that the map taking  $g$  to  $g^{-1}$  is  $C^\infty$  and the map taking  $(g_1, g_2)$  to  $g_1 g_2$  is also  $C^\infty$ . We need a little more notation than we did in Section 1.4.2. We denote the identity element of  $G$  by  $e$ , the left and right multiplication maps by  $L_g$  and  $R_g$ , and the inner automorphism of  $G$  induced by  $g$  by  $I_g = L_g \circ R_g^{-1}$ .

Recall that a vector field  $X$  on  $G$  is said to be *left invariant* if for all  $g, g' \in G$ ,  $(L_g)_* X_{g'} = X_{gg'}$ . Similarly, a covariant tensor field  $\omega$  is said to be *left invariant* (*right invariant*) if, for every  $g_0, g$  in  $G$ ,  $L_{g_0}^* \omega_{g_0 g} = \omega_g$  ( $R_{g_0}^* \omega_{g g_0} = \omega_g$ ). As usual,  $\omega$  is said to be *bi-invariant* if it is both left and right invariant. If  $h$  is a (left, right, bi-)invariant Riemannian metric on

---

<sup>23</sup>On the other hand, if one thinks of the expression for  $\mathbb{E}\{\varphi(A_u)\}$  as  $e^{-u^2/2\sigma^2}$  multiplied by a power series in the level  $u$ , then it is correct to say that this conjecture gives the correct coefficient for the leading term of the power series.

$G$ , then it is clear that, for every  $g$ , the map  $(L_g, R_g, I_g)$  is an isometry<sup>24</sup> of  $(G, h)$ . In particular, the curvature tensor  $R$  of  $h$ , is (left, right, bi)-invariant. This means that for Gaussian random fields that induce such Riemannian metrics, the integrals needed to evaluate  $\mathbb{E}\{\varphi(A_u(f, M))\}$  are significantly easier to calculate.

**Theorem 4.11.1** *Let  $G$  be a compact  $N$ -dimensional Lie group and  $f$  a centered, unit variance Gaussian field on  $M$  satisfying the conditions of Corollary 4.2.5 for  $G$ . Suppose that the Riemannian metric  $g$  induced by  $f$  is (left, right, bi)-invariant. Then*

$$(4.11.4) \quad \mathbb{E}\{\varphi(A_u(f, M))\} = \text{Vol}_g(G) \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \frac{(-1)^k \rho_{N-2k}(u)}{(2\pi)^k k!} \text{Tr}^{T_e G}(R_e^k),$$

where  $T_e G$  is the tangent space to  $G$  at  $e$ .

**Proof.** This is really a corollary of Theorem 4.10.1. Applying that result, and comparing (4.11.4) with (4.10.1), it is clear that all we need to show is that

$$\int_G \text{Tr}^G(R^l)_{g'} \text{Vol}_g(dg') = \text{Tr}^{T_e G}(R_e^l) \text{Vol}_g(G).$$

Suppose  $X, Y, Z, W$  are left-invariant vector fields. Since  $g' \mapsto L_g g'$  is an isometry for every  $g$ , we have,

$$\begin{aligned} R_g((X_g, Y_g), (Z_g, W_g)) &= R_e((L_{g^{-1}*}X_g, L_{g^{-1}*}Y_g), (L_{g^{-1}*}Z_g, L_{g^{-1}*}W_g)) \\ &= R_e((L_{g*})^{-1}X_g, (L_{g*})^{-1}Y_g, (L_{g*})^{-1}Z_g, (L_{g*})^{-1}W_g) \\ &= R_e(X_e, Y_e, Z_e, W_e). \end{aligned}$$

Therefore, if  $(X_i)_{1 \leq i \leq n}$  is an orthonormal set of left-invariant vector fields,

$$\begin{aligned} (R_g)^l((X_{i_1 g}, \dots, X_{i_l g}), (X_{j_1 g}, \dots, X_{j_l g})) \\ = (R_e)^l((X_{i_1 e}, \dots, X_{i_l e}), (X_{j_1 e}, \dots, X_{j_l e})), \end{aligned}$$

from which it follows that,

$$\text{Tr}^{T_g G}((R_g)^l) = \text{Tr}^{T_e G}((R_e)^l),$$

which completes the proof.  $\square$

<sup>24</sup>An *isometry* between two  $C^k$  Riemannian manifolds  $(M, g)$  and  $(\hat{M}, \hat{g})$  is a  $C^{k+1}$  diffeomorphism  $F: M \rightarrow \hat{M}$  for which  $F^* \hat{g} = g$ .

### 4.12 Chern-Gauss-Bonnet Theorem

As promised at the beginning of the Chapter, we now give a purely probabilistic proof of the classical Chern-Gauss-Bonnet Theorem. Of course, ‘purely’ is somewhat of an overstatement, since the results on which our proof is based were themselves based on Morse’s critical point theory.

The Chern-Gauss-Bonnet Theorem is one of the most fundamental and important results of Differential Geometry, and gives a representation of the Euler characteristic of a deterministic manifold in terms of curvature integrals. While it has nothing to do with probability, it is, nevertheless, it is a simply corollary to Theorems 4.10.1 and 4.10.2.

**Theorem 4.12.1 (Chern-Gauss-Bonnet Theorem)** <sup>25</sup> *Let  $(M, g)$  be a  $C^3$  compact, orientable,  $N$  dimensional, Riemannian manifold, either without boundary or piecewise  $C^2$  and isometrically embedded in some  $C^3$  Riemannian manifold  $(\tilde{M}, \hat{g})$ . Then, if  $M$  has no boundary,*

$$(4.12.1) \quad \varphi(M) \equiv \mathcal{L}_0(M) = \frac{(-1)^{N/2}}{(2\pi)^{N/2} N!} \int_M \text{Tr}^M(R^{N/2}) \text{Vol}_g,$$

if  $N$  is even, and 0 if  $N$  is odd. In the piecewise smooth case,

$$\begin{aligned} \varphi(M) &= \pi^{-(N)/2} \sum_{k=0}^N \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^l \Gamma\left(\frac{N-2l}{2}\right) 2^{-(l+2)}}{l!(k-2l)!} \\ &\quad \times \int_{\partial_k M} \mathcal{H}_k(dt) \int_{S(N_t, \partial_k M)} \text{Tr}^{\partial_k M}(S_\nu^{k-2l} R^l) \mathcal{H}_{N-k-1}(d\nu), \end{aligned}$$

where we adopt the notation of (3.8.2).

**Proof.** Suppose  $f$  is a Gaussian random field on  $M$ , such that  $f$  induces<sup>26</sup> the metric  $g$ . Suppose, furthermore, that  $f$  satisfies the side all the conditions of either Theorem 4.10.1 or Theorem 4.10.2, depending in whether  $M$  does, or does not have, a boundary.

To save on notation, assume now that  $M$  does not have a boundary. Recall that in computing  $\mathbb{E}\{\varphi(A_u(f, M))\}$  we first wrote  $\varphi(A_u(f, M))$  as an alternating sum of  $N$  different terms, each one of the form

$$\mu_k(u) \triangleq \# \{t \in M : f_t > u, \nabla f_t = 0, \text{index}(-\nabla^2 f_t) = k\}.$$

<sup>25</sup>This result has a long and impressive history, starting in the early nineteenth century with simple Euclidean domains. Names were added to the result as the setting became more and more general. The form given here is very close to that proven in 1943 by Allendoerfer and Weil [7].

<sup>26</sup>Note, this is the opposite situation to that which we have faced until now. We have always started with the field  $f$  and used it to define the metric  $g$ . Now, however,  $g$  is given and we are assuming that we can find an appropriate  $f$ .

Clearly, as  $u \rightarrow -\infty$ , these numbers increase to

$$\mu_k \stackrel{\Delta}{=} \# \{t \in M : \nabla f_t = 0, \text{index}(-\nabla^2 f_t) = k\},$$

and  $\varphi(M)$  is given by an alternating sum of the  $\mu_k$ . Since  $\mu_k(u)$  is bounded by the total number of critical points of  $f$ , which is an integrable random variable, dominated convergence gives us that

$$\varphi(M) = \lim_{u \rightarrow -\infty} \mathbb{E} \{\varphi(A_u(f, M))\},$$

and the statement of the Theorem then follows by first using Theorem 4.10.1 to evaluate  $\mathbb{E} \{\varphi(A_u(f, M))\}$  and then checking that the right hand side of (4.12.1) is, in fact, the above limit.

If we do not know that  $g$  is induced by an appropriate Gaussian field, then we need to adopt a non-intrinsic approach via Nash's embedding theorem in order to construct one.

Nash's embedding theorem [73] states that for any  $C^3$ ,  $N$ -dimensional Riemannian manifold  $(M, g)$ , there is an isometry  $i_g : M \rightarrow i_g(M) \subset \mathbb{R}^{N'}$  for some finite  $N'$  depending only on  $N$ . More importantly,  $\mathbb{R}^{N'}$  is to be taken with the usual Euclidean metric.

The importance of this embedding is that it is trivial to find an appropriate  $f$  when the space is Euclidean with the standard metric. Any unit zero-mean unit variance isotropic Gaussian random field,  $f'$  say, on  $\mathbb{R}^{N'}$ , whose first partial derivatives have unit variance, and which satisfies the non-degeneracy conditions of Theorem 4.10.1, will do. If we now define

$$f = f'_{|i_g(M)} \circ i_g^{-1}$$

on  $M$  then it is easy to see that  $f$  induces the metric  $g$  on  $M$ , and so our construction is complete

Finally, we note that the case of piecewise smooth  $M$  follows exactly the same argument, simply appealing to Theorem 4.10.2 rather than Theorem 4.10.1.  $\square$

## References

- [1] R. J. Adler. *The Geometry of Random Fields*. John Wiley & Sons Ltd., Chichester, 1981. Wiley Series in Probability and Mathematical Statistics.
- [2] R. J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 12. Institute of Mathematical Statistics, Hayward, CA, 1990.
- [3] R.J. Adler. Excursions above a fixed level by  $n$ -dimensional random fields. *J. Applied Probability*, 13:276–289, 1976.
- [4] R.J. Adler. On excursion sets, tube formulae, and maxima of random fields. *Annals of Applied Prob.*, 10:1–74, 2000.
- [5] R.J. Adler and R. Epstein. Some central limit theorems for Markov paths and some properties of Gaussian random fields. *Stoch. Proc. Appls.*, 24:157–202, 1987.
- [6] R.J. Adler and A.M. Hasofer. Level crossings for random fields. *Annals of Probability*, 4:1–12, 1976.
- [7] Carl B. Allendoerfer and André Weil. The Gauss-Bonnet theorem for Riemannian polyhedra. *Trans. Amer. Math. Soc.*, 53:101–129, 1943.
- [8] T. M. Apostol. *Mathematical Analysis: A Modern Approach to Advanced Calculus*. Addison-Wesley Publishing Company, Inc., Reading, Mass., 1957.

- [9] J-M. Azaïs and M. Wschebor. On the distribution of the maximum of Gaussian field with  $d$  parameters. *xxxx*, *xxxx:xxxx*, *xxxx*. Preprint.
- [10] Yu.K. Belyaev. Point processes and first passage problems, Univ. of California Press, Berkeley. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, 2:1–17, 1972.
- [11] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1987.
- [12] S. Bochner. Monotone funktionen stieltjessche integrale and harmonische analyse. *Math. Ann.*, 108:378, 1933.
- [13] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.
- [14] W.M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press,, San Diego, 1986.
- [15] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30:205–216, 1975.
- [16] P.J Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 1991.
- [17] E. V. Bulinskaya. On the mean number of crossings of a level by a stationary Gaussian process. *Theor. Probability Appl*, 6:435–438, 1961.
- [18] R. Cairoli and Walsh J. B. Stochastic integrals in the plane. *Acta Math.*, 134:111–183, 1975.
- [19] R.C. Dalang and T. Mountford. Jordan curves in the level sets of additive Brownian motion. *Trans. Amer. Math. Soc*, 353:3531–3545, 2001.
- [20] R.C. Dalang and J.B. Walsh. The structure of a Brownian bubble. *Probab. Theory Related Fields*, 96:475–501, 1993.
- [21] R.L. Dobrushin. Gaussian and their subordinated self-similar random generalized fields. *Ann. Probability*, 7:1–28, 1979.
- [22] J.L. Doob. *Stochastic Processes*. Wiley, New York, 1953.
- [23] R. M. Dudley. Metric entropy of some classes of sets with differentiable boundaries. *J. Approximation Theory*, 10:227–236, 1974.
- [24] R. M. Dudley. Central limit theorems for empirical measures. *Ann. Probability*, 6:899–929, 1978.

- [25] R. M. Dudley. Lower layers in  $\mathbb{R}^2$  and convex sets in  $\mathbb{R}^3$  are not gb classes. *Lecture Notes in Math*, 709:97–102, 1978.
- [26] R.M. Dudley. Sample functions of the Gaussian process. *Ann. Probability*, 1:66–103, 1973.
- [27] R.M. Dudley. *Real Analysis and Probability*. Wadsworth, Belmont, 1989.
- [28] R.M. Dudley. *Uniform Central Limit Theorems*. Cambridge University Press, Cambridge, 1999.
- [29] E. B. Dynkin. Gaussian and non-Gaussian random fields associated with Markov processes. *J. Funct. Anal.*, 55(3):344–376, 1984.
- [30] E. B. Dynkin. Polynomials of the occupation field and related random fields. *J. Funct. Anal.*, 58(1):20–52, 1984.
- [31] E.B. Dynkin. Markov processes and random fields. *Bull. Amer. Math. Soc.*, 3:957–1000, 1980.
- [32] E.B. Dynkin. Markov processes as a tool in field theory. *J. Functional Anal.*, 50:167–187, 1983.
- [33] A. Erdélyi. *Higher Transcendental Functions*. McGraw-Hill, New York, 1953. Vol. II. Bateman Manuscript Proj.
- [34] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.
- [35] H. Federer. *Geometric Measure Theory*. Springer-Verlag, New York, 1969.
- [36] X. Fernique. *Fonctions aléatoires gaussiennes vecteurs aléatoires gaussiens*. CRM, Montréal, 1997.
- [37] A.M. Garsia, E. Rodemich, and H. Rumsey. A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana Univ. Math. J.*, 20:565–578, 1970.
- [38] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50:265–289, 1985.
- [39] Y. Gordon. Elliptically contoured distributions. *Prob. Theory Rel. Fields*, 76:429–438, 1987.
- [40] A. Gray. *Tubes*. Addison-Wesley, Redwood City, 1990.
- [41] H. Hadwiger. *Vorlesungen Über Inhalt, Oberfläche und Isoperimetrie*. Springer-Verlag, Berlin, 1957.

- [42] H. Hadwiger. Normale körper im euklidischen raum und ihre topologischen und metrischen eigenschaften. *Math. Zeitschrift*, 71:124–140, 1959.
- [43] E.J. Hannan. *Multiple Time Series*. Wiley, New York, 1970.
- [44] T. Hida and M. Hitsuda. *Gaussian Processes*. AMS, Providence, 1993.
- [45] K. Itô. The expected number of zeros of continuous stationay Gaussian processes. *J. Math. Kyoto Univ.*, 3:206–216, 1964.
- [46] K. Itô. Distribution valued processes arising from independent Brownian motions. *Math. Z.*, 182:17–33, 1983.
- [47] K. Itô and M. Nisio. On the convergence of sums of independent banach space valued random variables. *Osaka J. Math.*, 5:35–48, 1968.
- [48] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, Cambridge, 1997.
- [49] J. Jost. *Riemannian Geometry and Geometric Analysis, 2nd ed.* Springer-Verlag, Berlin, 1998.
- [50] M. Kac. On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.*, 43:314–320, 1943.
- [51] D. Khoshnevisan. *Multi-Parameter Process: An Introduction to Random Fields*. Springer-Verlag, New York, 2002.
- [52] D.A. Klain and G-C. Rota. *Introduction to geometric probability*. Cambridge University Press, Cambridge, 1997.
- [53] A. N. Kolmogorov and V. M. Tihomirov.  $\varepsilon$ -entropy and  $\varepsilon$ -capacity of sets in function spaces. *Uspehi Mat. Nauk*, 14(2 (86)):3–86, 1959. [English translation, (1961) *Am. Math. Soc. Transl.* 17:277-364.].
- [54] H. Landau and L.A. Shepp. On the supremum of a Gaussian process. *Sankya*, 32:369–378, 1970.
- [55] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York, 1983.
- [56] M. Ledoux. *The Concentration of Measure Phenomenon*. AMS, Providence, 2001.
- [57] M. Ledoux and M. Talagrand. *Probability in Banach spaces. Isoperimetry and Processes*. Springer-Verlag, Berlin, 1991.

- [58] John M. Lee. *Introduction to Topological Manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.
- [59] John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.
- [60] G. Letac. Problèmes classiques de probabilité sur un couple de gelfand. *Lecture Notes in Math.*, 861:93–120, 1981.
- [61] M.A. Lifshits. *Gaussian Random Functions*. Kluwer, Dordrecht, 1995.
- [62] M.S. Longuet-Higgins. On the statistical distribution of the heights of sea waves. *J. Marine Res.*, 11:245–266, 1952.
- [63] M.S. Longuet-Higgins. The statistical analysis of a random moving surface. *Phil. Trans. Roy. Soc.*, A249:321–387, 1957.
- [64] T.L. Malevich. A formula for the mean number of intersections of a surface by a random field (in russian). *Izv. Akad. Nauk. UzSSR*, 6:15–17, 1973.
- [65] M. B. Marcus. Level crossings of a stochastic process with absolutely continuous sample paths. *Ann. Probability*, 5:52–71, 1977.
- [66] M.B. Marcus and L.A. Shepp. Sample behaviour of Gaussian processes. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, 2:423–442, 1971.
- [67] B. Matérn. *Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations*. Meddelanden Fran Statens Skogsforskningsinstitut, Band 49, Nr. 5, Stockholm, 1960.
- [68] Y. Matsumoto. *An Introduction to Morse Theory*. AMS, 2001.
- [69] K. S. Miller. Complex random fields. *Info. Sciences*, 9:185–225, 1975.
- [70] Millman, R.S and G.D. Parker. *Elements of Differential Geometry*. Prentice-Hall, New Jersey, 1977.
- [71] S. Morita. *Geometry of Differential Forms*. Number 201 in Translations of Mathematical Monographs. AMS, Providence, 2001.
- [72] M. Morse and S. Cairns. *Critical Point Theory in Global Analysis and Differential Topology. An Introduction*. Academic Press, New York, 1969.
- [73] John Nash. The imbedding problem for Riemannian manifolds. *Ann. of Math. (2)*, 63:20–63, 1956.

- [74] B. O'Neill. *Elementary Differential Geometry, 2nd ed.* Academic, San Diego, 1997.
- [75] V.I. Piterbarg. *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* Number 148 in Translations of Mathematical Monographs. AMS, Providence, 1996.
- [76] R. Pyke. Partial sums of matrix arrays, and Brownian sheets. In *Stochastic analysis (a tribute to the memory of Rollo Davidson)*, pages 331–348. Wiley, London, 1973.
- [77] D. Rabinowitz and D. Siegmund. The approximate distribution of the maximum of a smoothed poisson random field. *Statist. Sinica*, 7:167–180, 1997.
- [78] S. O. Rice. Mathematical analysis of random noise. *Bell System Tech. J.*, 24:46–156, 1945. Also in Wax, N. (Ed.) (1954), *Selected Papers on Noise and Stochastic Processes*, Dover, New York.
- [79] R. Riesz and B. Sz-Nagy. *Functional Analysis.* Ungar, New York, 1955.
- [80] W. Rudin. *Fourier Analysis on Groups, Republished (1990).* Wiley, New York, 1962.
- [81] W. Rudin. *Real and Complex Analysis, Second Ed.* McGraw-Hill, New York, 1974.
- [82] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York, 1994.
- [83] L. A. Santalo. *Integral Geometry and Geometric Probability.* Encyclopedia of Mathematics and its Applications. Addison-Wesley, Reading, 1976.
- [84] R. Schneider. *Convex Bodies, The Brunn-Minkowski Theory.* Cambridge University Press, Cambridge, 1993.
- [85] I. J. Schoenberg. Metric spaces and completely monotone functions. *Ann. of Math.*, 39:811–841, 1938.
- [86] K. Shafie, B. Sigal, D. Siegmund, and K.J. Worsley. Rotation space random fields with an application to fMRI data. *Ann. Statist.*, xxx:xxx, xxx. To appear.
- [87] D.O. Siegmund and K.J. Worsley. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Statist.*, 23:608–639, 1995.

- [88] D. Slepian. On the zeroes of Gaussian noise. In M. Rosenblatt, editor, *Time Series Analysis*, pages 104–115. Wiley, New York, 1963.
- [89] D.R. Smart. *Fixed Point Theorems*. Cambridge University Press, Cambridge, 1974.
- [90] A. Takemura and S. Kuriki. On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. of Appl. Prob.*, 12(2):768–796, 2002.
- [91] M. Talagrand. A simple proof of the majorizing measure theorem. *Geom. Funct. Anal.*, 2:118–125, 1992.
- [92] M. Talagrand. Majorizing measures: The generic chaining. *Ann. Probab.*, 24:1049–1103, 1996.
- [93] M. Talagrand. Majorizing measures without measures. *Ann. Probab.*, 29:411–417, 2001.
- [94] J. E. Taylor. *Euler Characteristics for Gaussian Fields on Manifolds*. PhD thesis, McGill University, 2001.
- [95] J. E. Taylor. A Gaussian kinematic formula. *xxx*, xxxx, 2002. Submitted.
- [96] B. S. Tsirelson, I. A. Ibragimov, and V. N. Sudakov. Norms of Gaussian sample functions. *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975)*, 550:20–41, 1976.
- [97] V.N. Vapnik. *The Nature of Statistical Learning Theory, Second edition*,. Springer-Verlag, New York, 2000.
- [98] V.N. Vapnik and A.Ya. Cervonenkis. *Theory of Pattern Recognition: Statistical Problems in Learning*. Nauka, Moskow, 1974. [In Russian].
- [99] V.N. Vapnik and A.Ya. ~ Cervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probability. Appl.*, 16:264–280, 1971.
- [100] N. Walsh. An introduction to stochastic partial differential equations. *Lecture Notes in Math.*, 1180:265–437, 1986.
- [101] S. Watanabe. *Stochastic Differential Equations and Malliavan Calculus*. Springer-Verlag, Berlin, 1984. Tata Inst. of Fundamental Research.
- [102] H. Weyl. On the volume of tubes. *Amer. J. Math.*, 61:461–472, 1939.

- [103] E. Wong. *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York, 1971.
- [104] E. Wong and M. Zakai. Martingales and stochastic integrals for processes with a multi-dimensional parameter. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 29:109–122, 1974.
- [105] E. Wong and M. Zakai. Weak martingales and stochastic integrals in the plane. *Ann. Probability*, 4:570–587, 1976.
- [106] K.J. Worsley. Boundary corrections for the expected euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. Appl. Probab.*, 27:943–959, 1995.
- [107] K.J. Worsley. Estimating the number of peaks in a random field using the hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.*, 23:640–669, 1995.
- [108] K.J. Worsley. The geometry of random images. *Chance*, 9:27–40, 1997.
- [109] A.M. Yaglom. Some classes of random fields in  $n$ -dimensional space, related to stationary random processes. *Theor. Probability Appl.*, 2:273–320, 1957.
- [110] A.M. Yaglom. Second-order homogeneous random fields, Univ. of California Press, Berkeley. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 2:593–622, 1961.
- [111] K. Ylisen. Random fields on noncommutative locally compact groups. *Lecture Notes in Math.*, 1210:365–386, 1986.
- [112] N. D. Ylvisaker. The expected number of zeros of a stationary Gaussian process. *Ann. Math. Statist.*, 36:1043–1046, 1965.
- [113] M. Yor. *Exercises in Probability from Measure Theory to Random Processes, via Conditioning*. Springer-Verlag, New York, 2002. In print.
- [114] A.C. Zaanen. *Linear Analysis*. North Holland, Amsterdam, 1956.
- [115] M. Zahle. Approximation and characterization of generalized Lipshitz-Killing curvatures. *Ann. Global Anal. Geom.*, 8:249–260, 1990.