Note: Even if Google sent you here, what follows is **NOT** the book

**RANDOM FIELDS AND GEOMETRY**

published with Springer in 2007, but rather a companion volume, still under production, that gives a simpler version of the theory of the first book as well as lots of applications.

You can find the original *Random Fields and Geometry* on the Springer site.

Meanwhile, enjoy what is available of the second volume, and keep in mind that we are happy to receive all comments, suggestions and, most of all, corrections.

**Applications of**

# RANDOM FIELDS AND GEOMETRY

**Foundations and Case Studies**

**Robert J. Adler**

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

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

**Jonathan E. Taylor**

Department of Statistics
Stanford University
Stanford, California.

e-mail: jonathan.taylor@stanford.edu

**Keith J. Worsley**

Department of Mathematics and Statistics
McGill University
Montréal, Canada
and
Department of Statistics
University of Chicago
Chicago, Illinois.

e-mail: keith.worsley@mcgill.ca

October 22, 2015

# 0

# Preface


| Robert J. Adler | Jonathan E. Taylor[1] | Keith Worsley |
|---|---|---|
| Haifa, Israel | Stanford, California | Montreal, Canada |

ie.technion.ac.il/Adler.phtml
www-stat.stanford.edu/∼jtaylo
www.math.mcgill.ca/keith



[1]Research supported in part by NSF grant DMS-0405970

# Contents

# 1

# Introduction

## 1.1 An Initial Example

Suppose that you are given a conceptual task which, among other things, involves reading. To understand how your brain performs this task, it is scanned. Figure 1.1.1 shows the results of one of the first such brain mapping experiments REFERENCE NEEDED using PET Positron Emisson Tomography).

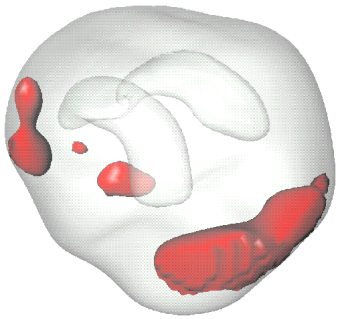

**Fig. 1.1.1.** A stimulated brain.

What this scan shows is your brain, its two connected lateral ventricles in the center, filled with cerebrospinal fluid, and four regions of high brain activity. These regions are defined by the fact that within them the blood flow, or, equivalently, the brain activity, is particularly high. The large region on the bottom right is the visual cortex, related to the fact that part of the task involved sight. The meaning of the three smaller regions is less clear, but since they are possibly related to language processing it is these which we would like to investigate.

Basic statistical theory, not to mention basic common sense, tells us that before we begin to offer learned explanations as to what is happening within

the three smaller regions, we should check as to whether or not they are consistent with random error, where the error may come from sampling mechanisms, instrument inaccuracies, or any other source.

The way to approach this is to assume a model. Indeed, denoting the brain by $T$, we could define some random function, $f$ say, over $T$, and ask if these three regions could also be generated, with a reasonable probability, by asking about the subsets of $T$ over which $f$ takes high values. In other words, we would ask if we could replace the physically meaningful brain activity level measurements by something completely random. If the answer to this question turned out to be positive, then clearly there is no information in the measurements and the three smaller regions may be considered as random sets of some kind. Thinking of them this way implies that in order to conduct the hypothesis test implicit in the above description, we shall need to know something about the behavior of random sets.

Actually, we now have enough motivation to introduce some notation, which will remain with us throughout this book. Parameter spaces, such as the brain, will generally be denoted by $T$, and a typical point in $T$ by $t$. The use of $T$ comes from the simplest of all cases, when the parameter space is an interval on the real line $\mathbb{R}$, and so $t$ stands for 'time'[1]. Almost always, $T$ will be a subset of some $N$-dimensional Euclidean space $\mathbb{R}^N$, $N \geq 1$.

We shall denote functions on $T$ by $f$, and when these are random, we shall call them *random fields*. This terminology comes from the pre-history of the subject, which dealt with yields of agricultural produce, so that the measurement of the density of yield gave a random (agricultural) field. Sometimes, usually when we are treating theory rather than practice, we shall talk about *stochastic processes* rather than random fields, but the two terms are completely synonymous. We shall also not distinguish between $f(t)$ and $f_t$ as the value of $f$ at $t \in T$.

The last component that we need is a formulation of the high activity regions of the brain, and for these we define the *excursion sets*[2] of $f$ over $T$ and above the level $u$ by

$$A_u \equiv A_u(f,T) \stackrel{\Delta}{=} \{t \in T : \, f(t) \geq u\}. \tag{1.1.1}$$

Excursion sets will play a major role in this book, whose central theme will be their study and use. In particular, we shall want to quantify various qualitative properties of these sets and understand these quantifiers, as well as possible, as random variables. We have already done this in quite some detail and at a rather abstract level in *RFG*[3]. Now, however, we want do it again,

[1] Later on, when we move to manifolds as parameter spaces, we shall emphasise this by replacing $T$ by $M$. Nevertheless, points in $M$ will still be denoted by $t$.

[2] In much of the mathematical literature, and in most of physics, excursion sets are known as *nodal domains*.

[3] One last reminder, in case you have not read the Preface, that *RFG* refers to the book *Random Fields and Geometry* [5], recently published by two of us.

with the accent on application and readability, rather than theory and rigor. However, this book is anything but a poor man's version of the earlier one. It will contain many results and topics that are not there at all, along with a wealth of additional material that will help you apply random field theory in a wide range of settings.

Perhaps the best way to convince you of this is via examples, and we shall choose one from astrophysics and one from brain mapping, the second enlarging on our opening example above. However, before we turn to these, we have to say just a little more about random fields and about geometry. The former is important so that we can agree on what is meant by the term 'random' and the latter is important since we shall be using it to study the behavior of excursion sets.

## 1.2 Gaussian and Related Random Fields

We shall define random fields in general and Gaussian ones in particular more carefully and in considerably more detail in Chapter 2, but since we need them to discuss the two examples below something needs to be said about them now.

In fact, Gaussian random fields are very easy to define. Given a parameter space $T$, a function $f : T \to \mathbb{R}$ which has the property that the collection

$$f(t_1), \ldots, f(t_k), \tag{1.2.1}$$

of random variables has a multivariate Gaussian distribution for all finite $k \geq 1$ and all collections $t_1, \ldots, t_k \in T$ is a *Gaussian random field*.

There is no reason not to extend this construction to random fields taking values in $\mathbb{R}^d$ for any $d \geq 1$, and in this case, and if $T$ has dimension $N$, we call $f$ an $(N, d)$ random field. Doing so allows us to define many different kinds of random fields which significantly enlarge the class of models which can be covered by a common theory. For example, if $f = (f_1, \ldots, f_d)$ is collection of $d$, independent, mean zero, unit variance Gaussian fields, then defining a new random field $\chi^2$ by setting

$$\chi^2(t) = \sum_{j=1}^{d} f_j^2(t)$$

gives a random field whose distribution is chi-squared at each point of the parameter space $T$. Random fields which we obtain this way – i.e. as functions of Gaussian fields – are what we call *Gaussian related*.

For the remainder of this chapter we shall work only with real valued Gaussian fields, and shall make a few simplifying assumptions. The first will be that $f$ has sample paths that are smooth, where we shall have more to say about what 'smooth' means in Section 2.5.

The next assumption, which is relatively harmless, is that $f$ have zero expectation at all points in $T$. A little more severe is the assumption that there be a common variance $\sigma^2$. Finally, we shall assume that $f$ is both *stationary* and *isotropic*, terms that will be defined carefully in Section 2.4. What these last two assumptions require is that $f$ behave the same, stochastically, in all regions and in all directions.

As a consequence, there is one more parameter that we can define, and this is the so-called *second spectral moment*, $\lambda$, which is simply the variance of any directional derivative of $f$.

Thus, in total, we really have only two free parameters defined by our Gaussian random field. The first, $\sigma$, determines the order of magnitude of the values that $f$ takes. The second, $\lambda$, describes how rapidly $f$ changes as $t$ changes, with the order of magnitude of these changes being given by $\lambda^{1/2}$. For the moment, these two parameters, among the many that could be used to describe the dependence structure of $f$ at different points in $T$, will suffice.

## 1.3 Shape and the Euler Characteristic

Returning to our example of brain activity in Section 1.1, we see that we are half way to setting up a statistical test, since we now know how to model randomness. Indeed, we can now consider a null hypothesis that the three small regions of high brain activity in Figure 1.1.1 are no more than the excursion sets of a Gaussian random field defined on the brain.

The next step is to determine which geometric aspects of these excursion sets can be measured, at least in a way that will be amenable to statistical analysis. Somewhat surprisingly, the solution comes from eighteenth century mathematics.

Among many other contributions to science, Leonhard Euler (1707–1783), arguably the most prolific mathematician of his century, discovered a very interesting and basic fact about convex polyhedra. Recall that a polyhedron is a solid object bounded by plane faces, such as a cube, pyramid, etc. Euler realized that if you count the numbers of faces ($F$), edges ($E$), and vertices (points, $P$) of a polyhedron, then

$$P - E + F = 2. \tag{1.3.1}$$

A cube, for example, has $F = 6$ faces, $E = 12$ edges and $P = 8$ vertices (see Figure 1.3.1(a)) so that $8-12+6 = 2$. For a solid that consists of $Q$ polyhedra, glued together on at least one common face, thereby creating a structure that is no longer necessarily convex, a more general version of (1.3.1) holds, and is given by

$$P - E + F - Q = 1. \tag{1.3.2}$$

See, for example, the L-shaped object in Figure 1.3.1(b).

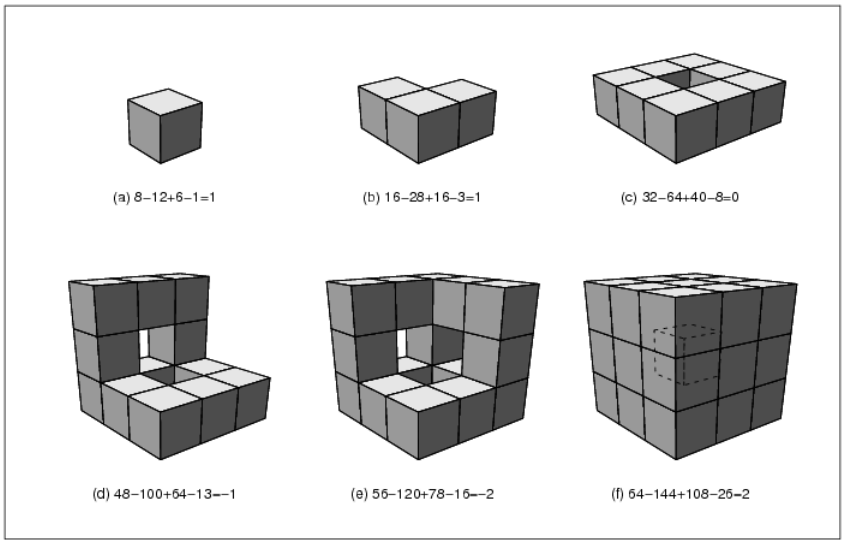


**Fig. 1.3.1.** The Euler characteristic of a solid is the number of vertices (points) − edges + faces − polyhedra. For a single polyhedron (a) or several joined together (b), the Euler characteristic is 1. If there is a hole in the solid (c) then the Euler characteristic is 0, and it decreases by 1 for each extra hole (d, e). In (f) the central cube is missing so the solid is hollow, which leads to an Euler characteristic of 2.

A little experimentation should convince you that this new formula works for all solids built from polyhedra, although it requires adjustment if the solid has a hole going through it, as in Figures 1.3.1(c)–(e). The adjustment is that we now have

$$P - E + F - Q = 1 - H,$$

where $H$ is the number of holes.

One more possibility that needs to be covered is that the set may have hollows, as in a tennis ball (which has only one). If the number of hollows is $H^*$, then, in fact

$$P - E + F - Q = 1 - H + H^*, \tag{1.3.3}$$

where $H^*$ is the number of hollows.

We have not quite covered all possibilities, for the solid may have a number of disjoint components. For example, consider all six structures in Figure 1.3.1 as if they were the six disjoint parts of one disconnected set. Let all the variables we have looked at so far now relate to the numbers in the entire structure, and let $C$ denote the number of disjoint connected components of the structure. (In Figure 1.3.1 $C = 6$.) Then (1.3.3) immediately gives us that

$$P - E + F - Q = C - H + H^*. \tag{1.3.4}$$

It turns out that the sum $P - E + F - Q$ is a topological invariant of the solid, and does not depend on the way it is subdivided into polyhedra. In fact, it has a name, and is known as the *Euler characteristic*. If $A$ is the solid, then we denote its Euler characteristic by $\varphi(A)$, so that

$$\varphi(A) \stackrel{\Delta}{=} P - E + F - Q, \tag{1.3.5}$$

for any subdivision of the solid into polyhedra. Avoiding the specific component polyhedra, by (1.3.4) we could also have adopted the definition

$$\varphi(A) \stackrel{\Delta}{=} C - H + H^*, \tag{1.3.6}$$

a more global, and usually preferable, definition.

A little thought shows that the the equivalence of the two definitions indicates an amazing balance between local and global phenomena. Each of the numbers $P$, $E$, $F$ and $Q$ are local, in the sense that if we change the way that $A$ is built up from polyhedra, then each of these numbers may change. However, since $C$, $H$ and $H^*$ (and $\varphi(A)$) are global, the four local parameters can only change in ways that keep their alternating sum constant.

Actually, (1.3.6) is but the tip of an iceberg, for the definition of the Euler characteristic works in all dimensions and not just in dimension three, as we have described it. Furthermore, it is well defined for sets that are far more general than those constructed from polyhedra. For example, suppose $A$ is a set in three dimensions with a smooth boundary. In this case we can cover $\mathbb{R}^3$ with many small cubes sitting on a cubical lattice. (Any collection of small, convex sets on a corresponding lattice would work, but cubes are simple and, at least for the moment, will suffice.) If the lattice is fine enough and the boundary of $A$ is smooth enough, then it can be shown that the Euler characteristic of the largest cube-based structure contained inside $A$ equals $\varphi(A)$ itself and the formula

$$\varphi(A) = C - H + H^*$$

still holds.

Digging deeper into the iceberg, we shall see, in Chapter 3, that the Euler characteristic is but one of a number of topological characteristics of sets, and that, in general, $N + 1$ natural quantifiers of geometry for sets of dimension

$N$. Some of these will tell us about volume, others about surface areas, and yet others about properties of cross-sections. However, it turns out that the Euler characteristic is the most basic, from which all others can be computed in ways which we shall describe in Chapter 3.

For now, however, it will suffice to work with the Euler characteristic. Some informative examples appear in Figure 1.4.2 which, while it deals with astrophysics, you can look at now. If our set is composed of many disconnected components, each containing relatively few holes or hollows (called a *meatball* topology in astrophysics) then the Euler characteristic is large and positive (Figure 1.4.2(a)). If the components are connected together by a network of 'bridges' or 'handles' to form many holes, then the Euler characteristic is negative. This is called a *sponge* topology (Figure 1.4.2(b)). If the network dominates to such an extent that the set consists of many surfaces surrounding isolated voids (rather like Swiss cheese), called a *bubble* topology (Figure 1.4.2(c)), then the hollows dominate and the Euler characteristic is once again positive.

With the tip of an iceberg firmly in view, we now actually have enough to look at the two promised applications.

## 1.4 The Large-Scale Structure of the Universe

One of the central problems of modern astrophysics is understanding the large-scale structure of the universe. Ever since the discovery of galaxies, it has been known that they are grouped together to form clusters, and these group together to form superclusters. With more accurate observations and larger galaxy surveys, it became evident that these structures themselves were not scattered throughout the universe in a uniform fashion, but rather in a fashion that varied considerably from region to region.

In fact, there are large voids throughout the universe, completely free of visible matter, surrounded by strings and even sheets of galaxies, one of which, relatively close to us, has been named the *Great Wall* by astrophysicists. It is one of the central challenges of modern cosmology to build models of the formation of the universe that explain this non-homogeneous structure, and to compare the results of these models to the structure we observe today.

There are many ways to make such comparisons, but most involve geometry and many involve the Euler characteristic and its relatives to which we just referred. In fact, in a series of articles in the Astrophysical Journal, starting in the mid-1980s, Richard Gott and his colleagues at Princeton used the Euler characteristic as a tool for describing the topology of the large scale structure of the universe.

### 1.4.1 Survey Data

The galaxy data studied by astronomers for understanding large scale structure are actually sets $A_u(f, T)$, as in (1.1.1). The parameter space $T$ is the

observable universe, and the function $f$ measures the 'density' of galaxies at a given point. To generate 'density' data, the atrophysicists first determine points $t_1, ...., t_n$ in $\mathbb{R}^3$ of galaxy centres, and place a 'mass' $w_i$ at each $t_i$. The masses are taken to be inversely proportional to the expected number of galaxies at distance $|t_i|$ from the Earth, which, at least on galactic scales, is the point from which measurements are taken. (This compensates for the fact that more distant galaxies appear fainter and are harder to detect, and so the survey data thins out at large distances.) These point masses are then smoothed with a Gaussian kernel, to give (up to a compensation for edge effects) the density field

$$f(t) = \sum_{i=1}^{n} w_i \frac{e^{-\alpha|t-t_i|^2}}{\sigma},$$

where $\alpha$ and $\sigma$ are scaling constants.

Varying the level $u$ for the excursion sets of $f$ shows regions of the universe with different galaxy densities, and so $u$ is actually no more than a *thresholding* parameter for determining the excursion sets.

Figure 1.4.2 shows high density regions from the Center for Astrophysics (CfA) survey containing $n = 10,506$ galaxies[4]. The density field $f$ has been 'Gaussianised', in the sense that the (univariate) marginal distribution of $f$ has been fixed to be standard normal, a procedure which we shall explain soon. For each of these regions the Euler characteristic was computed as the number of connected components minus the number of handles or holes plus the number of hollows, as in (1.3.6). The results give, as expected, high positive Euler characteristics for values of $u$ leading to regions where the topology of the excursion set is of 'meatball' or 'bubble' type, and low negative values when the topology is 'sponge'-like. In Figure 1.4.3 these observed Euler characteristics are shown (dashed line) for a full range of levels $u$.

### 1.4.2 Designing and Testing a Model

So far, all we have done with the galaxy data is to collect it, 'Gaussianise' it, and measure the Euler characteristics of excursion sets. All of these steps are simple, modulo, perhaps, the so-called Gaussianisation, which we now describe.

The first step is to compute the empirical distribution function, $\widehat{F}_n$, of the density $f$ at the basic grid of galaxy centers, viz. $\{f(t_1), \ldots, f(t_n)\}$. Letting $\Phi$ denote the distribution function of a standard normal random variable, define the 'Gaussianised', or sometimes just 'standardised', data to be

$$\widetilde{f}(t_j) \stackrel{\Delta}{=} \Phi^{-1}\left(\widehat{F}_n(f(t_j))\right), \qquad j = 1, \ldots, n. \tag{1.4.7}$$

[4] Completed in the mid-1990's, the CfA study was the first large-scale survey of the universe. Later on, in Chapter 10, we shall look at the most recent Sloan Digital Sky Survey (SDSS) which contains 141,300 galaxies.

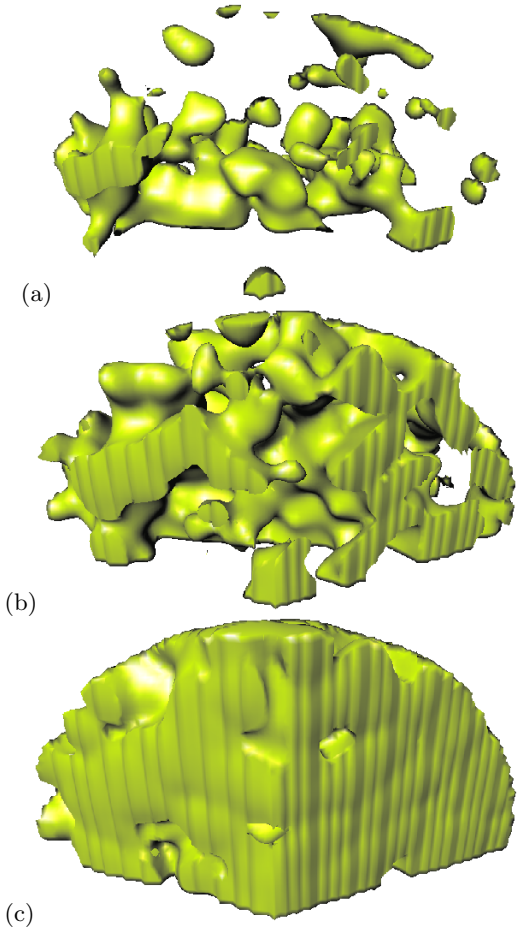


**Fig. 1.4.2.** Different topologies and examples of their Euler characteristics for the regions of high galaxy density from the CfA survey thresholded at the level $u$ on a Gaussian scale. There are 10,506 galaxies in the cone-shaped survey region, which extends out to 135 megaparsecs in the northern hemisphere, with the earth at the apex of the cone.
(a) At $u = 1$ there is a *meatball* topology of many disconnected components, each containing relatively few holes or hollows, giving an Euler characteristic of 23. The Great Wall is visible across the center of the cone.
(b) At $u = 0$ a *sponge* topology arises, in which the components are connected together by a network of 'bridges' or 'handles', to yield an Euler characteristic of $-4$.
(c) At $u = -1.3$ we find a *bubble* topology of surfaces surrounding isolated voids, giving an Euler characteristic of 8.

This standardised data is what appears in Figures 1.4.2 and 1.4.3.

We now turn to designing a stochastic model that might be able to explain the non-homogeneous nature of the CfA data, a model that we can then test. The basic model replaces the galaxy centers by the points of a uniform Poisson process in $\mathbb{R}^3$, following which the same construction of weighting and smoothing is adopted, the final stage being the Gaussianisation. The ensuing density field is now random, and it is easy to check that it is also stationary and isotropic.

We now have an *extremely important observation* to make, albeit a trivial one. This is that non-Gaussian random fields cannot generally be transformed to Gaussian by pointwise transformations. More specifically, if $f$ is a random field with distribution function $F_t$ at the point $t$, and $\Phi$ is the standard Gaussian distribution function as above, then the transformation

$$f(t) \to \widetilde{f}(t) \stackrel{\Delta}{=} \Phi^{-1}\left(F(f(t))\right) \tag{1.4.8}$$

does not generally imply that $\widetilde{f}$ is a Gaussian random *field*.

It does, of course, imply that the one-dimensional *marginal* distributions of $\widetilde{f}$ are standard Gaussian, but, as we saw at (1.2.1), for a random field to be Gaussian *all* of its finite dimensional distributions must be multi-variate normal, something which (1.4.8) cannot guarantee. This fact is so often overlooked in the applied literature that we shall devote the one and only displayed box of this book to emphasize that, with the exception of some rather trivial cases,

**Non-Gaussian fields cannot generally be transformed to Gaussian by localised transformations.**

A fortiori, a data driven transformation such as (1.4.7), while it does have the advantage of standardising the values of the the galaxy density data, does not make the data that of a Gaussian random field. Consequently, even after the transformation, the assumption of Gaussianity of the field for the standardised data is a very real assumption that needs to be tested statistically.

It turns out that a very effective way to test Gaussianity is via the Euler characteristic of the excursion sets of the galaxy density function. Since the density is random, so are the excursion sets, and therefore so are the Euler characteristics. However, it is possible to compute the expectations of these random variables, and a special case of far more general formulas that we shall meet time and again in this book gives that under stationarity and isotropy,

$$\mathbb{E}\left\{\varphi\left(A_u\right)\right\} = \left[\frac{\mathcal{L}_3\lambda^{3/2}(u^2-1)}{(2\pi)^2} + \frac{\mathcal{L}_2\lambda u}{(2\pi)^{3/2}} + \frac{\mathcal{L}_1\lambda^{1/2}}{2\pi}\right] e^{-u^2/2} + \mathcal{L}_0(1-\Phi(u)). \tag{1.4.9}$$

The constants

$$\mathcal{L}_j = \mathcal{L}_j(A)$$

should be thought of as measures of the '$j$-dimensional size' of the parameter space (that part of the universe scanned in the CfA survey) so that

$\mathcal{L}_3$ is the three dimensional volume of the surveyed space.
$\mathcal{L}_2$ is half the surface area of the surveyed space.
$\mathcal{L}_1$ is twice the *caliper diameter* of the surveyed space, where the caliper diameter of a convex solid is defined by placing the solid between two parallel planes (or calipers), measuring the distance between the planes, and averaging over all rotations of the solid i
$\mathcal{L}_0$ is the Euler characteristic of the parameter space, so that $\mathcal{L}_0(A) \equiv \varphi(A)$.

Thus, for example, for a rectangular box of size $a \times b \times c$, we have $\mathcal{L}_3 = abc$, $\mathcal{L}_2 = ab+bc+ac$, $\mathcal{L}_1 = a+b+c$ (half the 'volume' used by airlines to measure the size of luggage) and $\mathcal{L}_0 = 1$.

The remaining parameter $\lambda$ is the second spectral moment, or measure of rate of change, that we met back in Section 1.2, and this needs to be estimated from that data before we can use (1.4.9). We shall discuss a number of ways to do this in later chapters, and in the present case a good estimate turns out to be $\lambda = 0.00661$.

Computing the expected Euler characteristic from (1.4.9) and including the empirically observed Euler characteristics gives us Figure 1.4.3. The results are in rough agreement in that both yield meatball, sponge, and bubble topologies for, respectively, high, medium and low values of the thresholding variable $u$. However, we shall see later that this behavior is common to virtually all situations. More critical inspection shows that the empirical and theoretical curves are actually not very close, and that overall the empirical curve is smaller than expected, indicating (on the basis of (1.3.6)) fewer individual components in the excursion sets and more clumping of galaxies into clusters, strings, and walls. In short, the data are inconsistent with the assumption of Gaussianity. What are the implications of this fact, and what to do about it, will be discussed in some detail in Chapter 10.

Before leaving this example, it is important to note that while (1.4.9) arose in an astronomical setting, it is, in fact, a very general result. Indeed, it holds for all zero mean, stationary, isotropic, Gaussian random fields over three dimensional parameter spaces with unit variance. The extension to a general variance of $\sigma^2$ is done in the obvious way, by replacing $u$ by $u/\sigma$ throughout the formula. There are also versions of this result for all dimensions, for non-isotropic and (constant variance) non-stationary random fields, and also for a wide variety of non-Gaussian fields. We shall study these extensions in considerable detail in Chapter 4.

For the moment, however, we can remain in the simpler setting of (1.4.9) also while treating our next example.

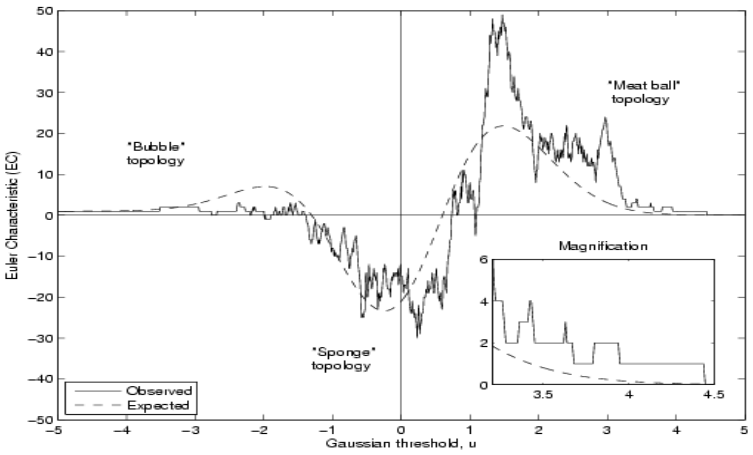


**Fig. 1.4.3.** The observed Euler characteristics of the set of high density regions of the CfA Galaxy survey plotted against the density threshold. High thresholds produce a *meatball* topology (magnification, and Figure 1.4.2(a)). Medium thresholds produce a *sponge* topology (Figure 1.4.2(b)), and low thresholds produce a *bubble* topology (Figure 1.4.2(c)). Also shown are the expected values of these Euler characteristics for galaxies generated as the excursion sets of Gaussian random fields.

## 1.5 Brain Imaging

One of the earliest experiments in brain imaging was conducted in 1990 at Montreal Neurological Institute (cf. [52].) In this experiment, subjects were injected with a radio isotope emitting positrons, which annihilate with nearby electrons to release gamma rays that are detected by Positron Emission Tomography (PET). By careful reconstruction, it was possible to build up an image of blood flow in the brain, a measure of brain activity. This opened up the possibility of actually seeing which regions of the brain were activated by different stimuli, and so to actually see the brain 'thinking'.

### 1.5.1 The Data

In 1992, in one of the first experiments of its kind, (cf. [32]) subjects were told to perform a linguistic task, involving the silent reading of words on a screen, during the imaging process. By subtracting an image in which each subject was 'at rest' looking at a blank screen, the experimenters were able to see evidence of increased blood flow in certain brain regions corresponding to the effort required for the task.

The images were, however, very blurred, and the signal (if any) was very weak compared to the background noise, so to increase the signal-to-noise ratio the experiment was repeated on 10 subjects. The brain images were

aligned in three dimensions, and the blood flow was averaged, as in Figure 1.5.4(a)–(c). These images seem to show increased activation, but this needs to be tested formally, which takes us, as in the previous example, into the realm of hypothesis testing.

Although Figure 1.5.4 seems to show a continuous image, the raw data was actually stored as values at $128 \times 128 \times 80$ voxels. At each such voxel there were 10 pairs of blood-flow values, one pair for each subject, one taken while the subject was performing the task, the other at rest. The aim is to test the null hypothesis of there being no difference between the rest and activation stages, and so it is elementary statistic methodology to perform a paired-difference $T$ test. This is calculated by taking the mean difference (Figure 1.5.4(d)), dividing by the standard deviation (Figure 1.5.4(e)), and multiplying by the square root of the number of subjects ($\sqrt{10}$). To increase the accuracy of the standard deviation, it was replaced by the average over all voxels to obtain a $Z$ statistic with an approximate Gaussian distribution. Proceeding in this fashion, we can make an image of $Z$ statistics, one for each voxel (Figure 1.5.4(f)).

### 1.5.2 Statistical Testing

The above construction of $Z$ statistics yields a simple statistical test for each voxel. However, all told there are about 300,000 voxels in the brain, and it makes no sense at all to carry out this number of tests.

For example, if we were to use the common 0.05 critical value of 1.64 for the standard normal variable, and carry out 300,000 tests, we would expect to see 15,000 voxels above this level, even if the null hypothesis of no difference between rest and activation were true. Thus a different threshold needs to be found[5].

Consequently, the procedure adopted is to scan the entire brain looking for high values of $Z$, and choose a threshold value $u$ which will be crossed only with some small, controllable probability. The null hypothesis of no difference between rest and activation stages is then rejected if this threshold is crossed. Under this global null hypothesis, the overall probability of error is the probability that any value of $Z$ will be higher than $u$ and is called the *family wise error rate*.

Let us rephrase this in terms of things we already know. The null hypothesis is now that as $Z$ varies over the brain it determines a Gaussian random

[5] The classical way of correcting for this would be to divide the 5% level by the number of tests, the so-called *Bonferroni correction*, which gives a critical value of $u = 5.10$. However, this usually over-corrects. To see why, note that the voxels are actually rather arbitrary, and if we subdivided each into eight voxels of half the size in each direction, we would divide now by 2,400,000 rather than 300,000 and obtain an even larger threshold ($u = 5.48$). Thus the Bonferroni threshold is not the best path to take.

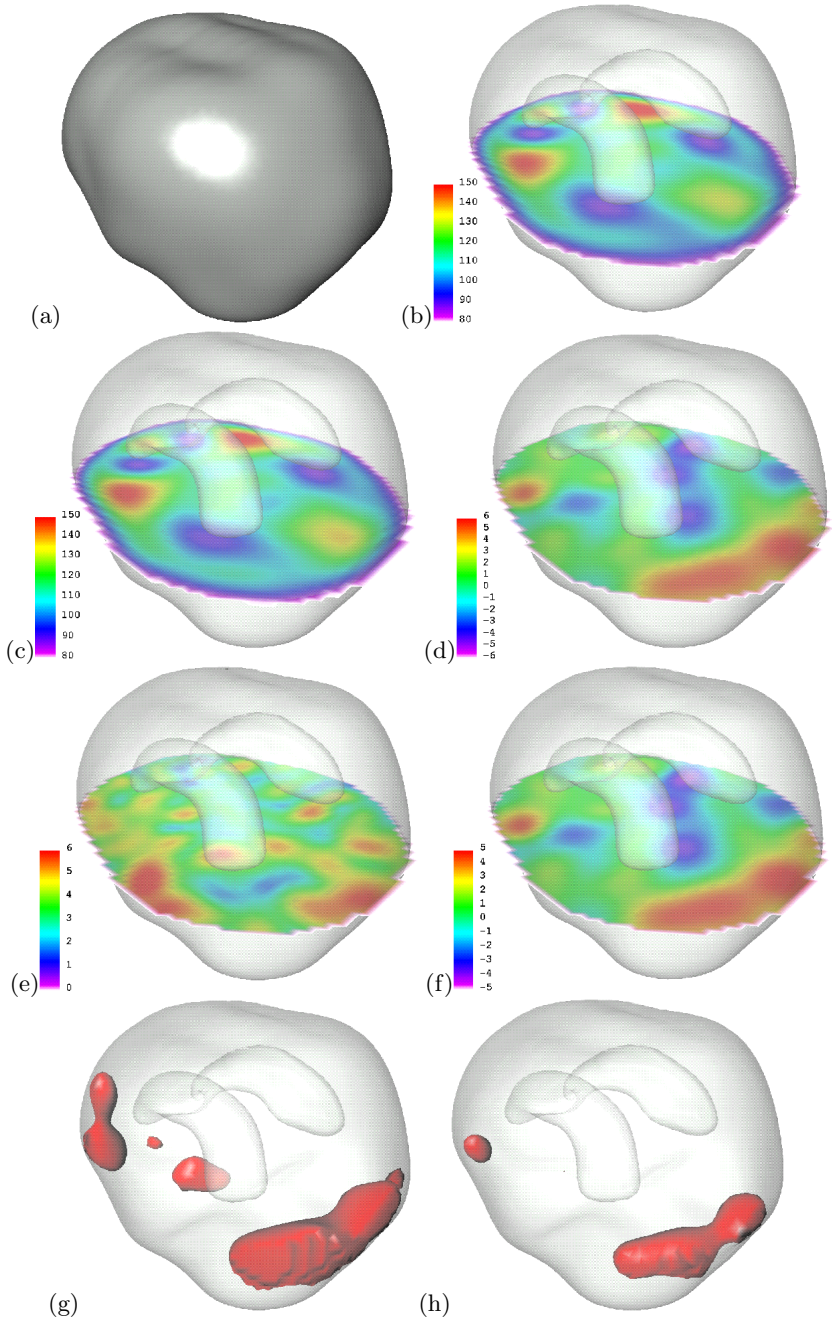


**Fig. 1.5.4.** A PET study showing regions of the brain activated by a reading task.

FIGURE 1.5.4, continued: The brain (a) has been rendered as a transparent solid with the rear left side facing the viewer. The ventricles in the center form a single connected hollow that gives the brain an Euler characteristic of 2, as in Figure 1.3.1(f). One of 80 slices through the brain is color-coded (red = high, purple = low) to show (b) average blood flow of 10 subjects under the rest condition and (c) under the task condition. The difference of the averaged blood flows, task – rest, is shown in (d) and the standard deviation of the 10 differences (9 degrees of freedom) in (e). The $Z$ statistic for testing for a significant increase in blood flow due to the task is (f) while (g) gives the excursion set of all $Z$ values greater than a threshold of $u = 3.3$, where we expect the Euler characteristic to be 1 if the image is pure noise (no activation). In fact the Euler characteristic is 4, due to the four components clearly visible, suggesting some evidence of activation. Changing the threshold to $u = 4.22$ (so that the expected Euler characteristic is 0.05 when no activation is present and all noise is excluded with a probability of 0.95) gives (h). Two regions remain (the Euler characteristic of the excursion set is 2), one in the left frontal lobe, near the language area of the brain and a broad region in the left visual cortex and extrastriate. These are the regions significantly activated by the task.

field[6] of zero mean, where we shall now go back to treating the brain as a continuous object rather than a collection of voxels of essentially arbitrary size. Thus we are seeking a threshold $u_{0.05}$ such that

$$\mathbb{P}\Big\{ \sup_{t\in T} Z(t) \geq u_{0.05} \Big\} = 0.05, \tag{1.5.10}$$

where the parameter set $T$ is the brain. For obvious reasons, we shall call a probability of this type an *exceedence probability* . Outside of the setting of Markov processes, which is definitely not our setting, exceedence probabilities are notoriously difficult to compute precisely. However, note that, for any $u$,

$$\mathbb{P}\Big\{ \sup_{t\in T} Z(t) \geq u \Big\} \equiv \mathbb{P}\Big\{ A_u(Z,T) \neq \emptyset \Big\}, \tag{1.5.11}$$

where $A_u$ is the excursion set (1.1.1).

Now for some handwaving: Suppose that $u$ is large, so that the probability in (1.5.11) is small. Then the excursion set $A_u$ is most likely to be of the meatball type; i.e. made up of a few isolated small regions, with neither holes, handles nor hollows. In fact, if $u$ is large, it is unlikely that there would be more than one component to this set. Now recall (1.3.6), which described the Euler characteristic of a set, to see that, for large $u$, with high probability,

$$\varphi(A_u(Z,T)) = \begin{cases} 1 & \text{if } A_u(Z,T) \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

[6] Of course, forming the $Z$ field in the way we have described again only ensures that the univariate marginals of the field are Gaussian, and does not say anything about general finite-dimensional marginals. Thus we are committing the unforgivable crime of ignoring the warning that we gave in the box of page 10. However, we shall test this part of the assumption en passant, in much the same way as we did for the CfA galaxy data.

Since the expectation of a 0-1 random variable is the probability that it takes the value 1, it follows immediately from this and (1.5.11) that

$$\mathbb{P}\Big\{ \sup_{t\in T} Z(t) \geq u \Big\} \simeq \mathbb{E}\left\{\varphi(A_u(Z,T))\right\}, \tag{1.5.12}$$

where by "$a(u) \simeq b(u)$" we mean that the $a$ and $b$ are, asymptotically, of the same order of magnitude as $u \to \infty$ in that their ratio converges to 1. Our hope is that these asymptotics are good enough to allow for approximations for 'reasonable' values of $u$ as well; i.e. long before we approach infinity. That this is in fact the case is something that we shall see later in Chapter 4, and so for the moment it is reasonable to take it as given.

With this approximation in hand, we can easily find $u_{0.05}$ if we are also prepared to assume that the $Z$ field is also stationary and isotropic. With these assumptions we can replace the probability on the left hand side of (1.5.10) by $\mathbb{E}\{\varphi(A_{u_{0.05}})\}$ and exploit the expression we have for this expectation in (1.4.9). That is, we need to solve for $u$ in the equation

$$\left[\frac{\mathcal{L}_3\lambda^{3/2}(u^2-1)}{(2\pi)^2} + \frac{\mathcal{L}_2\lambda u}{(2\pi)^{3/2}} + \frac{\mathcal{L}_1\lambda^{1/2}}{2\pi}\right] e^{-u^2/2} \;+\; \mathcal{L}_0(1-\Phi(u)) = 0.05. \tag{1.5.13}$$

But this is easy to do. Measuring the brain gives its volume, $\mathcal{L}_3$, half its surface area, $\mathcal{L}_2$, twice the caliper diameter, $\mathcal{L}_1$ and Euler characteristic, $\mathcal{L}_0$ as 1,064 cc, 1,077 cm$^2$, 0.1 cm, and 2, respectively[7]. All that remains is to estimate the second spectral moment $\lambda$, the details of which we again postpone to later. The estimate is given by $\lambda = 0.0693$. Substituting into (1.5.13) and solving for $u$ gives $u_{0.05} = 4.22$, much higher than the threshold of 1.64 that came from the most simple minded of tests, but not as high as the Bonferroni value of 5.10. (See Footnote 5.)

The excursion set above this level is shown in Figure 1.5.4(h), and these are the regions which show significant activation. One is in the visual cortex and adjacent left extrastriate, and one is in the left frontal cortex, near the language area of the brain.

However, we still have some work left to do. Along the way, we assumed that the $Z$ field was Gaussian (as a field), stationary, and isotropic. To both test these hypotheses and to obtain more information on the geometry of activation regions in the brain under the given task, we can use (1.4.9) and the parameter values given above to look at the entire curve of empirical and expected Euler characteristics of excursion sets. This is given in Figure 1.5.5, which plots the observed and expected Euler characteristics against the threshold, as for the galaxy data (Figure 1.4.3). The main feature of interest is the larger than expected Euler characteristic for the high thresholds ($u > 3$), attributable to the two regions of activation we have already discovered.

[7] The ventricles form a single large hollow in the center of the brain, visible in Figure 1.5.4, that increases the surface area, reduces the caliper diameter, and gives the brain an Euler characteristic of 2, as in Figure 1.3.1(f).

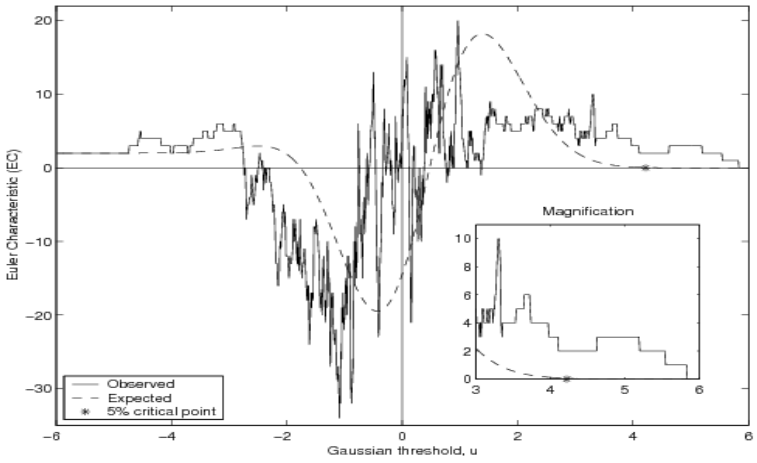


**Fig. 1.5.5.** Plot of the observed Euler characteristic of the set of high $Z$ regions for the PET data and the expected Euler characteristic from (1.4.9) if there is no activation due to the linguistic task, plotted against the threshold $u$. The most interesting part is when $u > 3$, showing a higher Euler characteristic than expected, confirming that some components of the excursion set are due to the linguistic task and not the random noise. In particular, at $u = 3.3$ we expect an Euler characteristic of 1, but we observe 4 (visible in Figure 1.5.4(g)). At the 5% critical value of $u = 4.22$, we expect 0.05 but we observe 2 components (visible in Figure 1.5.4(h)).

A remaining question is whether or not the empirical and expected curves in Figure 1.5.5 are sufficiently close to allow us to feel comfortable about the assumptions of Gaussianity of the field, and its stationarity and isotropy. The curves are certainly closer[8] in this case than they were for the CfA galaxy data in Figure 1.4.3 but we do not yet have sufficient experience to know if, in this case, "close enough is good enough". In fact, there are good physiological reasons for believing that neither stationarity nor isotropy are valid, and it will turn out that the graphs in Figure 1.5.5 actually provide evidence that this is the case. We shall see further details, and much more, in Chapter 9.

## 1.6 Beyond the Euler Characteristic

Throughout this chapter we have concentrated on two things: excursion sets and Euler characteristics.

It is clear that excursion sets are natural objects to study and that they are going to have many applications in a wide variety of scenarios. However, at least for a non-geometer, it is not as clear why we have chosen the Euler characteristic as the main quantifier of excursion set geometry.

[8] This is particularly true if one smoothes the empirical curve somewhat.

Two justifications have appeared along the way. The first is equation (1.4.9), which gave, at least in the stationary, isotropic, three-dimensional Gaussian scenario a very explicit formula for the expected Euler characteristic of excursion sets. We have already promised you that this was just a foretaste of things to come, and that in Chapter 4 we shall extend this formula considerably. The parameter spaces will be very general, the random fields will no longer necessarily be Gaussian, and neither stationarity nor isotropy will be assumed. Nevertheless, we shall still be able to give explicit (although somewhat more complicated) expressions for the expected Euler characteristics of excursion sets.

This is no mean feat in the theory and application of smooth random fields, where explicit formulae are few and far between. Thus this fact in itself justifies studying Euler characteristics .

The second reason lies in the approximation (1.5.12) between exceedence probabilities and expected Euler characteristics. It turns out that this also is true in wide generality, although this is much harder to prove out of the Gaussian scenario. Since exceedence probabilities are of central importance in a wide variety of applications, and essentially impossible to calculate directly, the ability to approximate them via expected Euler characteristics is another good reason to study them.

However, there are more quantifiers of geometry than just the Euler characteristic, and we have already met three of them in (1.4.9), which related to three-dimensional objects. We could rewrite (1.4.9) in the more generic form

$$\mathbb{E}\left\{\mathcal{L}_0\left(A_u\right)\right\} = e^{-u^2/2}\sum_{i=0}^{N} C_j(\lambda)\mathcal{L}_j(A)\rho_j(u), \tag{1.6.14}$$

with $N = 3$, noting that $\varphi(A) \equiv \mathcal{L}_0(A)$, and the rest of the notation being obvious.

Although we have only defined the $\mathcal{L}_j$ for $j \leq 3$, for a wide class of $N$-dimensional sets $A$ there are $N+1$ such quantifiers, $\mathcal{L}_N(A), \ldots, \mathcal{L}_0(A)$. It will always be true that $\mathcal{L}_N(A)$ is the $N$-dimensional Lebesgue measure (volume) of $A$, $\mathcal{L}_{N-1}(A)$ half the surface area of $A$, $\mathcal{L}_{N-2}(A), \ldots, \mathcal{L}_1(A)$ will measure the sizes of certain averaged measures of the sizes of cross-sections of various dimensions, and that $\mathcal{L}_0(A)$ will be the Euler characteristic. These quantifiers appear under a variety of other names, such as Quermassintegrales, Minkowski, Dehn and Steiner functionals, integral curvatures and intrinsic volumes, although in many of these cases the ordering and normalisations are different.

Given the existence of these additional quantifiers, it seems natural to also apply them to the excursion sets of random fields, and once again to ask for the expectations. That is, we would like to extend (1.6.14) to obtain $\mathbb{E}\{\mathcal{L}_j(A_u)\}$. Moreso, we would like a result without the restrictions of stationarity and isotropy, and to go beyond the Gaussian case. That this is in fact possible

was one of the main results of *RFG*, and we shall see from where they come in Chapter 4, and then later how to use them.

Firstly, however, we need to learn a little random field theory and a little geometry. This will make up the content of the next two chapters.

# Part I

## The Underlying Theory

# 2

# Random Fields

## 2.1 Stochastic Processes and Random Fields

As you read in the Preface, for us a *random field* is simply a stochastic process, taking values in a Euclidean space, and defined over a parameter space of dimensionality at least one. Actually, we shall be rather loose about exchanging the terms 'random field' and 'stochastic process'. In general, we shall use 'field' when the geometric structure of the parameter space is important to us, and shall use 'process' otherwise.

We shall usually denote parameter spaces by either $T$ or $M$, generally using $T$ when the parameter space is simple Euclidean domain, such as a cube, and $M$ when refering to manifolds, or surfaces. Elements of both $T$ and $M$ will be denoted by $t$, in a notation harking back to the early days of stochastic processes when the parameter was always one-dimensional 'time'.

Of course, we have yet to define what a stochastic process is. To do this properly, we should really talk about notions such as probability spaces, measurability, separability, and so forth, as we did in *RFG*. However, we shall send the purist to *RFG* to read about such things, and here take a simpler route.

**Definition 2.1.1.** *Given a parameter space $T$, a stochastic process $f$ over $T$ is a collection of random variables*

$$\{f(t) : t \in T\}.$$

*If $T$ is a set of dimension $N$, and the random variables $f(t)$ are all vector valued of dimension $d$, then we call the vector valued random field $f$ a $(N, d)$ random field.*

Note that while in what follows we shall adopt the standard convention of denoting random variables by upper case Latin characters, we shall use lower case to denote random processes. The reasons for this will become clear later, since we shall need objects such as $X(t)$ to denote geometric objects.

Examples of random fields abound, and we have already seen a few in Chapter 1, but perhaps the easiest to visualise is given by the ocean surface. In fact, this one example yields many, which together give a good introduction to many of the basic concepts of random field theory.

The height of an ocean surface above some nominal mean plane is, obviously, a function of both time and space, and so we acknowledge this by representing it as

$$f(t,x), \qquad t \in [0,\infty) \equiv \mathbb{R}_+, \quad x \in \mathbb{R}^2.$$

It is clear that the dependencies of $f$ on the two parameters $t$ and $x$ are quite different, so, for the moment, let us fix a time $t$ and consider only the ocean surface at this time, which we denote by $f(x)$.

The behavior of $f(x)$ as a function of $x$, however, depends very much on where the ocean is being observed. For example, at a sea shore we see mainly waves, and so $f$ has the structure of a cosine wave (cf. Figure 2.4.1) with local amplitudes and wavelengths that have been perturbed in a random fashion. However, despite additional random perturbations, these waves generally have a very well defined direction.

The same would not be true if we were to observe an ocean at its center, on a day with little wind. While we would still see waves, their directions would not be so well defined, and in fact would change randomly with time. This lack of preferred direction is called *isotropy*, a concept that we shall meet more formally in Section 2.4.7.

While waves at the sea shore do not exhibit this behavior, it is true that if they are neither very close nor very far from the shoreline, waves at any two positions behave much the same, in that the random perturbations to local amplitude and wavelength are stochastically similar, and the same dominant direction is preserved. This property is known as *stationarity*, and we shall look at it in more detail in Section 2.4.

Adding time back into the picture, you may now want to think about whether or not ocean surfaces are ever going to be stationary or isotropic in a joint space-time parameter space. (Answers will be given in Section ?????.) It is clear, however, that the different axes in this three-dimensional parameter space no longer play interchangeable roles, as, for example, the two components of space do for an ocean near its center, for a fixed time on a windless day.

The ocean surface also provides some very natural examples of vector valued random fields. Over half a century ago, Michael Longuet-Higgins, in a series of works that gave birth to the study of the stochastic properties of sea surfaces (e.g. [59, 60]) studied the $(2,2)$ gradient random field

$$\nabla f\,(x_1,x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right),$$

with the aim of understanding the 'specular points' of random waves, those points $x$ at which the two components of $\nabla f(x)$ take given values. For exam-

ple, depending on where one is standing on the sea shore, on where the sun is, and what the particular values of $\nabla f$ are, these might be the points at which the sea surface reflects the sun into one's eyes.

We shall meet many additional examples as we progress through this book, but the sea surface is probably the easiest and canonical example on which to try out most new ideas.

Before leaving this introductory section, there is one rather important, theoretical point that needs to be made. While Definition 2.1.1 may effectively define stochastic processes/random fields, it does not really tell us how to differentiate between them. A celebrated theorem of Kolmogorov, known as his *Consistency* or *Existence Theorem*, says that the distributional properties of a $(N, d)$ random field over $T$ are determined by its *finite-dimensional distributions*[1]. These are the distributions

$$\mathbb{P}\left\{f(t_1) \in B_1, \ldots, f(t_n) \in B_n\right\}, \tag{2.1.1}$$

for all $n \geq 1$ and all collections $\{t_j\}_{1 \leq j \leq n}$ and Borel $\{B_j\}_{1 \leq j \leq n}$ with $t_j \in T$ and $B_j \in \mathbb{R}^d$. If $f$ is a process which possesses joint probability densities, then the probabilities in (2.1.1) can be expressed, in a self-explanatory notation, as

$$\int_{B_1} \ldots \int_{B_n} p_{t_1,\ldots,t_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n, \tag{2.1.2}$$

and so Kolmogorov's theorem reduces to demanding that we know these densities.

## 2.2 Gaussian and Gaussian Related Random Fields

At the core of this book will be Gaussian and Gaussian-related random fields, and so it is appropriate that we define them before all others[2]. At the same time, we shall take the opportunity to collect a number of basic results about univariate and multivariate Gaussian random variables. We imagine that most readers will be familiar with these, and so can skip the next subsection on first reading, returning to it later for notational conventions.

### 2.2.1 Gaussian Variables

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

[1] Of course, Kolmogorov's theorem only holds under certain regularity conditions. However, since these can be found in any mathematically rigorous textbook on stochastic processes, we shall let you search for them there.

[2] In fact, we already did this informally in Section 1.2, but now the time has come to do it properly.

$$\varphi(x) \stackrel{\Delta}{=} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}, \qquad x \in \mathbb{R},$$

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

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

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 or process 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 function and ($\Psi$) for the tail probability function of a standard normal variable, so that

$$\Phi(x) \stackrel{\Delta}{=} 1 - \Psi(x) \stackrel{\Delta}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du. \tag{2.2.1}$$

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

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) < \Psi(x) < \frac{1}{x}\varphi(x). \tag{2.2.2}$$

(See Exercise 2.8.1.)

An $\mathbb{R}^d$-valued random variable $X = (X_1, \ldots, X_d)$ is said to be *multivariate Gaussian* if, for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, the real valued variable $\langle \alpha, X\rangle = \sum_{i=1}^d \alpha_i X_i$ is Gaussian[3]. In this case there exists a mean vector $m \in \mathbb{R}^d$ with $m_j = \mathbb{E}\{X_j\}$ and a non-negative definite[4] $d \times d$ 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

$$\varphi(x) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} e^{-\frac{1}{2}(x-m)C^{-1}(x-m)'}, \tag{2.2.3}$$

where $|C| = \det C$ is the determinant[5] 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.

[3] An important comment on notation: 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 usually denoted by $\langle x, y\rangle$, $(x, y)$ or, occasionally, by $x \cdot y$ and even $xy$ when there is no chance of confusion.

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

[5] Yet another important comment on (this time misleading and inconsistent) notation: At various places we shall use the notation $|\cdot|$ to denote any of 'absolute

In view of (2.2.3) 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 (2.2.3) that the characteristic function of a multivariate Gaussian $X$ is given by

$$\phi(\theta) = \mathbb{E}\{e^{i\langle\theta,X\rangle}\} = e^{i\langle\theta,m\rangle - \frac{1}{2}\theta C\theta'}, \tag{2.2.4}$$

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

An immediate consequence of either (2.2.3) or (2.2.4) is that if $A$ is any $d \times d$ matrix and $X \sim N_d(m, C)$, then

$$XA \ \sim \ N(mA,\, A'CA). \tag{2.2.5}$$

A judicious choice of $A$ (see Exercise 2.8.2) then allows us to compute conditional distributions as well. If $n < d$ make the partitions

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

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

$$m_{i|j} = m^i + (X^j - m^j)C_{jj}^{-1}C_{ji} \tag{2.2.6}$$

and covariance matrix

$$C_{i|j} = C_{ii} - C_{ij}C_{jj}^{-1}C_{ji}. \tag{2.2.7}$$

### 2.2.2 Gaussian Fields

We can now define a real valued *Gaussian (random) field* or *Gaussian (random) process* to be a random field $f$ on a parameter set $T$ for which the (finite dimensional) distributions (2.1.1) of $(f_{t_1}, \ldots, f_{t_n})$ are multivariate Gaussian for each $1 \le n < \infty$ and each $(t_1, \ldots, t_n) \in T^n$.

---

value', 'Euclidean norm', 'determinant' or 'Lebesgue measure', depending on the argument, in a natural fashion. The notation $\|\cdot\|$ is used only for either the norm of complex numbers or for special norms, when it usually appears with a subscript. This is unless it is used, as in Section 2.6 in particular, as $\|f\|$ to denote the supremum of a function $f$, which is not a norm at all. Despite this multitude of uses of a simple symbol, its meaning should always be clear from the context.

Since multivariate Gaussian distributions are determined by means and covariances, it is immediate that Gaussian random fields are determined[6] by their mean and covariance functions[7], defined by

$$m(t) = \mathbb{E}\{f(t)\} \tag{2.2.8}$$

and

$$C(s,t) = \mathbb{E}\left\{\left(f(s) - m(s)\right)\left(f(t) - m(t)\right)\right\}. \tag{2.2.9}$$

In fact, this is one of the main reasons, beyond ubiquitous but not always justified appeals to the central limit theorem, that Gaussian processes are such popular and useful choices for models for random processes on general spaces.

Multivariate 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 this case, $m(t)$ takes values in $\mathbb{R}^d$ and the covariance function of (2.2.9) is replaced by a function whose values are non-negative definite, $d \times d$, matrices. In particular,

$$C(s,t) = \mathbb{E}\left\{\left(f(s) - m(s)\right)'\left(f(t) - m(t)\right)\right\}, \tag{2.2.10}$$

so that the individual elements of $C$ are given by

$$C_{ij}(s,t) = \mathbb{E}\left\{\left(f_i(s) - m_i(s)\right)\left(f_j(t) - m_j(t)\right)\right\}.$$

As in the real valued case, for Gaussian $f$ the vector function $m$ and the matrix function $C$ determine all of its statistical properties.

### 2.2.3 Gaussian Related Fields

As convenient as Gaussian random fields may be for the mathematician, in that the form of the multivariate Gaussian distribution makes many things computable in the Gaussian case that are uncomputable otherwise, it would be a poor modeller or statistician who would work in a Gaussian scenario only.

---

[6] In fact, one can also go in the other direction as well. It is a consequence of this structure of Gaussian densities and the Kolmogorov consistency theorem that, given *any* set $T$, a function $m : T \to \mathbb{R}$, and a non-negative definite function $C : T \times T \to \mathbb{R}$ (see the following footnote) there exists a Gaussian process on $T$ with mean function $m$ and covariance function $C$.

[7] Covariance functions are always positive semi-definite; cf. Exercise 2.8.4. In the vector valued case there are two types of semi-definiteness to worry about. On the one hand, for each fixed pair $s, t$, the matrix $C(s,t)$ is positive semi-definite. On the other hand, for each pair $i, j$, the function $C_{ij}$ is positive semi-definite, as a function, on $T \times T$.

Leaving the Gaussian scenario is, however, not all that easy to do, and so we shall leave it in a fashion that, while somewhat limited, turns out to be broad enough to cover many, if not most, statistical applications. To be more precise, we shall call a random field $f : T \to \mathbb{R}^d$ a *Gaussian related* field if we can find a vector valued Gaussian random field,

$$g(t) = (g_1(t), \ldots, g_k(t)) : T \to \mathbb{R}^k,$$

and a function

$$F : \mathbb{R}^k \to \mathbb{R}^d,$$

such that $f$ has the same finite dimensional distributions as $F(g)$. That is,

$$f(t) \stackrel{\mathcal{L}}{=} F(g(t)) = F\left(g_1(t), \ldots, g_k(t)\right), \tag{2.2.11}$$

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

When $k = 1$, or, in general $k = d$ and $F$ is invertible, then the corresponding Gaussian related process is not much harder to study than the original Gaussian one. After all, what happens at the level $u$ for $f$ is precisely what happens at the uniquely defined level $F^{-1}(u)$ for $g$. However, in other cases, $f$ can provide a process that is qualitatively different to $g$.

For example, suppose that the $g_j$ are centred and of unit variance, and consider the following three choices for $F$, where in the third we set $k = n+m$.

$$\sum_1^k x_i^2, \qquad \frac{x_1\sqrt{k-1}}{(\sum_2^k x_i^2)^{1/2}}, \qquad \frac{m\sum_1^n x_i^2}{n\sum_{n+1}^{n+m} x_i^2}. \tag{2.2.12}$$

The corresponding random fields are known as $\chi^2$ fields with $k$ degrees of freedom, the $T$ field with $k-1$ degrees of freedom, and the $F$ field with $n$ and $m$ degrees of freedom. All of these will appear as important models in the applications we shall be looking at in Part IV of the book, and so we shall not attempt to motivate their usefulness now beyond stating that they are as fundamental to the statistical applications of random field theory as are their univariate distributions to standard statistical theory.

Before we leave Gaussian related processes, we want to make one comment about some interesting geometric problems associated with them.

Recall that one of the central themes of this book will be the study and application of the geometric properties of the excursion sets[8] of $f$ over $T$ and above the level $u$, viz.

$$A_u(f, T) = \{t \in T : f(t) \geq u\}.$$

[8] One final reminder, particularly for the physicists among our readers, that in much of the mathematical literature, and in most of physics, excursion sets are known as *nodal domains*.

For a Gaussian related field of the form (2.2.11), these can be rewritten as

$$\begin{aligned} A_u(f,T) \;=\; A_u(F(g),M) &= \{t \in T :\, (F \circ g)(t) \geq u\} \\ &= \{t \in T :\, g(t) \in F^{-1}[u,\infty)\} \\ &= T \cap g^{-1}(F^{-1}[u,+\infty)). \end{aligned} \tag{2.2.13}$$

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 $F^{-1}[u,\infty)$. This set will generally have an interesting (albeit deterministic) structure of its own. For example, in the case of a $\chi^2_k$ field it will be the complement, in $\mathbb{R}^k$, of a ball of radius $\sqrt{u}$ and centered at the origin. Thus, regardless of the simplicity of the underlying parameter space $T$, the moment we leave real valued Gaussian processes and turn to their Gaussian related relatives, geometry is going to be important to us.

We shall return to this point later in the book, once we know a little more about random fields and about geometry.

## 2.3 Examples of Random Fields

Before looking at specific examples, we need to make one small digression, to introduce the notions of stationarity and isotropy.

### 2.3.1 Stationary and Isotropic Fields

To formally define stationarity, which we actually already met in Section 1.2 and will learn a lot more about in Section 2.4 below, recall the definitions (2.2.8)–(2.2.10) of the mean and covariance functions of a stochastic process. These lead to two important definitions.

**Definition 2.3.1.** *Suppose that $f$ is an $(N,d)$ random field defined over all of $\mathbb{R}^N$. Suppose furthermore that the mean function $m(t)$ is constant, and that the covariance function $C(s,t)$ is a function of the difference $t-s$ only. Then we say that $f$ is homogeneous or stationary.*

*If it is also true that $C(s,t)$ is a function of the Euclidean distance $|t-s|$ only, then we say that $f$ is also isotropic*[9]*.*

Two comments are called for following on from this definition. The first is that, in fact, we have only really defined what is usually called *weak* or *second order* stationarity , which depends only on first and second moments.

[9] 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 the current definition. We shall treat isotropy *only* in the scenario of stationarity.

A stronger form of stationarity is that of the finite dimensional distributions, that requires that joint distributions of the form

$$\{f(t_1+\tau),\ldots,f(t_n+\tau)\} \tag{2.3.1}$$

be independent of $\tau$, for all $n \geq 1$ and all $t_j \in \mathbb{R}^N$. It is obvious that the two notions coincide when $f$ is Gaussian.

The second is more an issue of notation than a comment. When $f$ is stationary, we shall generally abuse notation and write

$$C(s,t) = C(s-t). \tag{2.3.2}$$

While the first $C$ here is defined on $T \times T$, the second is defined on $T$ only. In the isotropic case, we shall go even further, and write

$$C(s,t) = C(|s-t|),$$

where the second $C$ is defined on $\mathbb{R}_+$.

### 2.3.2 Random Polynomials

What appear at first sight to be the simplest of stochastic processes are the *random polynomials* on $\mathbb{R}$, functions of the form

$$f(t) = \sum_{k=0}^{n} \xi_k t^k,$$

for some finite $n$, where the random coefficients $\xi_k$ are generally taken to be either uncorrelated or independent. The random field version, on $\mathbb{R}^N$, is given by

$$f(t) = \sum_{k=0}^{n} \sum_{j_1+\cdots+j_N=k} \xi_{j_1\ldots j_N} t_1^{j_1} \cdots t_N^{j_N},$$

again with uncorrelated or independent coefficients.

Perhaps surprisingly (since polynomials really are the simplest of all non-random functions) random polynomials are actually very hard to study. For a start, they are clearly non-stationary, not even having constant variance. Nevertheless, despite this, or perhaps because of it, they are at the centre of considerable research activity at the moment, interest concentrating on the number of points in the zero set $\{t : f(t) = 0\}$ in the one-dimensional case, and the structure of this set in higher dimensions.

We shall take a look at this zero set using the tools we have developed, much later, in Section 12.1.

### 2.3.3 The Cosine Process

If random polynomials is where the theory of random processes might have begun, the true development of the theory began with what is perhaps the grandfather of all stationary processes, the so-called *cosine random process* on $\mathbb{R}$. It is defined as

$$f(t) \stackrel{\Delta}{=} \xi \cos \lambda t + \xi' \sin \lambda t, \tag{2.3.3}$$

where $\xi$ and $\xi'$ are uncorrelated, equidistributed, random variables and $\lambda$ is a positive constant.

The cosine process provides the simplest version of the the wave example that we met in Section 2.1, and we shall soon see that it also provides the elementary building block for general stationary processes.

It is elementary trigonometry to see that the cosine process can also be written as

$$f(t) = R \cos(\lambda t - \theta), \tag{2.3.4}$$

where $R^2 = \xi^2 + (\xi')^2 \geq 0$ and $\theta = \arctan(\xi'/\xi) \in (-\pi, \pi]$, from whence the name 'cosine process'. It is now obvious that the cosine process is no more than a cosine function with fixed wavelength $2\pi/\lambda$, but random amplitude and phase. Furthermore, supposing for convenience that $\mathbb{E}\{\xi\} = 0$, we have that its covariance function is given by

$$\begin{aligned} C(s,t) &= \mathbb{E}\{f(s)f(t)\} \\ &= \mathbb{E}\{(\xi \cos \lambda s + \xi' \sin \lambda s)(\xi \cos \lambda t + \xi' \sin \lambda t)\} \\ &= \mathbb{E}\{\xi^2\} \cos(\lambda(t-s)), \end{aligned}$$

on using the fact that $\xi$ and $\xi'$ are uncorrelated and equidistributed. Consequently, regardless of the distribution of $\xi$, the cosine process is stationary. In fact, as we shall soon see, it is the archetype of all stationary processes, all of which can be represented as a sum of cosine processes.

One of the nice aspects of the cosine process is that many things that are either difficult or impossible to compute for more general processes can be computed exactly, and from first principles, once some assumptions are made on the distribution of $\xi$. We shall therefore now assume that $\xi$ and $\xi'$ are independent, Gaussian variables, with zero mean and common variance $\sigma^2$.

Now consider, for $u > 0$, the exceedence probability

$$\mathbb{P}\left\{ \sup_{0 \leq t \leq T} f(t) \geq u \right\}, \tag{2.3.5}$$

which you know from reading the Preface is one of the important quantities of this book.

Under the Gaussian assumption, $R^2$ has an exponential distribution with mean $2\sigma^2$, $\theta$ has a uniform distribution on $(-\pi, \pi]$, and $R$ and $\theta$ are independent. We can use this information to compute some exceedence probabilities directly, and start by defining

$$N_u = N_u(f, T) = \#\{t \in [0, T] : f(t) = u \text{ and } df(t)/d(t) > 0\}.$$

Then $N_u$ is known as the *number of upcrossings* by $f$ of the level $u$ in time $[0, T]$, and the points being counted as *upcrossings*. Upcrossings and their generalizations will play a major rôle throughout this book.

It is trivial to see that the exceedence probability that we are after can now be written as

$$\begin{aligned} \mathbb{P}\left\{\sup_{0\le t\le T} f(t) \ge u\right\} &= \mathbb{P}\left\{f(0) \ge u\right\} + \mathbb{P}\left\{f(0) < u, N_u \ge 1\right\} \qquad (2.3.6) \\ &= \Psi\left(\frac{u}{\sigma}\right) + \mathbb{P}\left\{f(0) < u, N_u \ge 1\right\}. \end{aligned}$$

Obtaining an explicit expression for this process depends on the value of $T$. Consider first the case $T \le \pi/\lambda$, in which case the event $\{f(0) \ge u, N_u \ge 1\}$ is empty, implying that

$$\mathbb{P}\left\{f(0) < u, N_u \ge 1\right\} = \mathbb{P}\left\{N_u \ge 1\right\}.$$

Again using the fact that $T \le \pi/\lambda$, note that $N_u$ is either 0 or 1. In order that it be 1, two independent events must occur. Firstly, we must have $R > u$, with probability $e^{-u^2/2\sigma^2}$. Secondly (draw a picture) $\theta$ must fall in an interval of length $\lambda T$, so that the final result is

$$\mathbb{P}\left\{\sup_{0\le t\le T} f(t) \ge u\right\} = \Psi\left(\frac{u}{\sigma}\right) + \frac{\lambda T}{2\pi} e^{-u^2/2\sigma^2}, \qquad (2.3.7)$$

and the probability density of the supremum is given by

$$\frac{1}{\sigma}\varphi\left(\frac{u}{\sigma}\right) + \frac{\lambda T u}{2\pi\sigma^2} e^{-u^2/2\sigma^2}. \qquad (2.3.8)$$

For completeness, note that if $T \ge 2\pi/\lambda$, then the process will achieve the supremum $R$ at some point $0 \le t \le T$, and so the exceedence probability (2.3.5) is $P\{R > u\} = e^{-u^2/2\sigma^2}$. For $\pi/\lambda < T < 2\pi/\lambda$, the situation is more complicated.

The above computation was elementary, in the sense that it involved only two random variables and basic trigonometry. Thus one is tempted to believe that it must be easy to extend to many other processes. In fact, this is not the case, and the cosine process and field (see below) are the *only* differentiable, stationary, Gaussian processes for which the exceedence probabilities are explicitly known.

However, before we leave it, we can use it to nevertheless motivate a more general approach. Note first that since, as noted above, $N_u$ is either 0 or 1 when $T < \pi/\lambda$, we can rewrite (2.3.6) as

$$\mathbb{P}\left\{\sup_{0\le t\le T} f(t) \ge u\right\} = \Psi\left(\frac{u}{\sigma}\right) + \mathbb{E}\{N_u\}. \tag{2.3.9}$$

Thus, rather than arguing as above, we could concentrate on finding an expression for the mean number of upcrossings.

More importantly, note that for any $T$, and, indeed, for *any differentiable random process*, the argument leading to (2.3.6) and the fact that $P\{N_u \ge 1\} \le \mathbb{E}\{N_u\}$ give

$$\mathbb{P}\left\{\sup_{0\le t\le T} f(t) \ge u\right\} \le \mathbb{P}\left\{f(0) \ge u\right\} + \mathbb{E}\{N_u\}. \tag{2.3.10}$$

Thus there would seem to be a close relationship between exceedence probabilities and level crossing rates, that actually becomes exact for the cosine process over certain intervals. In fact, since, for a one dimensional set, its Euler characteristic is given by the number of its connected components, the right hand sides of both (2.3.9) and (2.3.10) could be written as $\mathbb{E}\{\varphi(A_u(f,T))\}$, where $\varphi$ is the Euler characteristic and $A_u(f,T)$ the excursion set $\{t \in [0,T] : f(t) \ge u\}$.

We shall investigate and exploit this, in detail, in Chapter 5.

### 2.3.4 The Cosine Field

The cosine field is a straightforward extension to $\mathbb{R}^N$ of the cosine process, and has the respresentation

$$f(t) = f(t_1, \ldots, t_N) \stackrel{\Delta}{=} \frac{1}{\sqrt{N}} \sum_{k=1}^{N} f_k(\lambda_k t_k), \tag{2.3.11}$$

where each $f_k$ is the process on $\mathbb{R}$ given by

$$f_k(t) = \xi_k \cos t + \xi_k' \sin t.$$

The $\lambda_k$ are fixed, and the $\xi_k$ and $\xi_k'$ are taken to be identically distributed and uncorrelated.

Again, it is a simple excercise to check that the cosine field is both stationary and isotropic (Exercise 2.8.5) but is a little harder to compute its exceedence probabilities. To see what can be done, we restrict attention to the cosine process on a rectangle of the form $T = \prod_{k=1}^{N}[0, T_k]$. Then, given the structure of the cosine field as a sum, it is immediate that

$$\sup_{t\in T} f(t) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \sup_{0\le t_k\le T_k} f_k(t).$$

If we assume that the $\xi_k$ and $\xi'_k$ are all independent $N(0,\sigma^2)$, then the suprema of the individual $f_k$ are also independent. Further assuming that each $T_k \in (0,\pi/\lambda_k]$, (2.3.7) and (2.3.8) give their individual distributions. The distribution of the supremum of the cosine field is then the convolution of these. In Exercise 2.8.6 you are asked to actually do the computation, to find that, if $p_N(u)$ is the density function of the supremum, $\varphi$ the usual standard Gaussian density and $\varphi^{(k)}$ its $k$-derivative, then there are simple constants, $C_{nk}$, depending only on $n$ and $k$, such that

$$p_N(u) = \varphi\left(\frac{u}{\sigma}\right) + \sum_{k=1}^{N}(-1)^k C_{nk}\varphi^{(k)}\left(\frac{u}{\sigma}\right)\sum_{j_1\dots j_k}\prod_{i=1}^{k}\frac{\lambda_{j_i}T_{j_i}}{\sigma}. \tag{2.3.12}$$

The inner sum here is over the $\binom{N}{k}$ subsets of size $k$ of $\{1,\dots,N\}$.

This is actually the archetype of a far more general result of a geometrical nature. To see it in its easiest formulation, take $\sigma^2 = 1$ and $\lambda_k = \lambda$ for all $k$, so that the product over the $\lambda_{j_i}$ becomes merely $\lambda^k$, and consider the sums

$$\sum_{j_1\dots j_k} T_{j_1}\cdots T_{j_k}.$$

When $k = N$, this gives the volume of the parameter space. When $k = N-1$, it gives its surface area. In fact, if you recall the $\mathcal{L}_k$ of Chapter 1, then this is precisely what these sums are, and we now define[10] the $\mathcal{L}_k$ of $N$-dimensional rectangles to be

$$\mathcal{L}_k\Big(\prod_1^N[0,T_j]\Big) = \sum_{j_1\dots j_k} T_{j_1}\cdots T_{j_k}, \tag{2.3.13}$$

where the sum is over the $\binom{N}{k}$ *distinct* choices of $k$ indices between 1 and $N$. Thus (2.3.12) becomes[11]

$$p_N(u) = \sum_{k=0}^{N}(-1)^k C'_{nk}\,\varphi^{(k)}(u)\,\lambda^k\mathcal{L}_j(T). \tag{2.3.15}$$

[10] Precisely why this is a useful definition will be made clear in Chapter 3. See also Exercise 3.6.4.

[11] Actually, (2.3.15) still makes sense if the $\lambda_j$ are not all identical and $\sigma^2 \neq 1$ but we define the $\mathcal{L}_k$ by

$$\mathcal{L}_k\Big(\prod_1^N[0,T_j]\Big) = \sum_{j_1\dots j_k}\prod_{i=1}^{k}\frac{\lambda_{j_i}T_{j_i}}{\sigma}, \tag{2.3.14}$$

which is based on a change of scale in (2.3.13). This is an example of what we shall call the 'inducing of a Riemannian metric on $\mathbb{R}^N$ via the random field $f$', something that we shall meet up with in some detail in Chapter 4 and that will be an important tool in later parts of the book.

Integrating over $u$, and applying some asymptotics[12], one finds that

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \geq u\right\} \tag{2.3.16}$$
$$= \Psi(u) + e^{-u^2/2}\sum_{k=1}^{N} C''_{nk} H_{k-1}(u)\lambda^k \mathcal{L}_j(T) + o\left(e^{-(1+\eta)u^2/2}\right),$$

for some $\eta > 0$, where the $H_k$ are Hermite polynomials, which we shall have cause to meet many times in the future. (cf. (3.3.20) for a definition.)

It turns out that the cosine process and field are the *only* smooth (non-degenerate) stationary Gaussian processes for which it has so far proven possible to actually compute the distribution of the supremum[13]. This means that in general we shall have to adopt other approaches, as we have mentioned before. However, the fact that there is at least one class of Gaussian fields for which the distribution of the supremum is explicitly known is extremely useful. We shall soon see that Slepian's Inequality (Theorem 2.6.3) will enable us to use the known exceedence probabilities of the cosine process and field to bound those of other Gaussian processes and fields as well.

### 2.3.5 Orthogonal Expansions

The idea of summing simple random fields to get more complicated ones, intrinsic to the definition of the cosine field, can be carried much further.

For example, suppose we have a collection $\{\varphi_n\}$ of real valued[14] functions on our parameter space $T$, and a collection of uncorrelated random variables $\{\xi_n\}$ with, for convenience, mean zero and variances $\sigma_n^2$. From these, we can form the sum

$$f(t) = \sum_{n=1}^{\infty} \xi_n \varphi_n(t), \tag{2.3.17}$$

which will have covariance function

[12] The asymptotics, while not being intrinsically difficult, are also not easy, and this is left to you as an exercise. You can find the details in Section 2.5 of Piterbarg's important monograph [75], which is where, to the best of our knowledge, (2.3.16) appeared for the first time.

[13] There are five non-smooth, essentially Markov, stationary Gaussian processes on $\mathbb{R}$ for which this is possible, but no other random fields. See Footnote 2 in Chapter 4 of *RFG* for a list.

[14] We are treating only the real valued case here for convenience. The extension to vector valued random fields in all that follows is straightforward. One needs only require that the $\varphi_n$ map $T$ to $\mathbb{R}^d$ and that the random variables $\xi_n$ are replaced by random matrices $\xi_n$ which are either independent or uncorrelated in the sense that $\mathbb{E}\{\xi_n'\xi_m\}$ is the zero matrix if $n \neq m$. The covariance function (2.3.18) is then replaced by the covariance matrix, $\sum_{n=1}^{\infty} \varphi_n'(s)\mathbb{E}\{\xi_n'\xi_n\}\varphi_n(t)$, etc.

$$\mathbb{E}\{f(s)f(t)\} = \sum_{n=1}^{\infty} \sigma_n^2 \, \varphi_n(s)\varphi_n(t), \tag{2.3.18}$$

and variance function

$$\mathbb{E}\{f^2(t)\} = \sum_{n=1}^{\infty} \sigma_n^2 \, \varphi_n^2(t), \tag{2.3.19}$$

as long as everything that needs to converge does in fact do so.

Random fields with the above structure are attractive to theoreticians and practitioners alike. For the theoretician, the mathematical structure of a summation of uncorrelated or independent variables is easy to handle. For the practitioner, this structure is even more useful. For a start, it generally matters little to a practitioner if the summations are finite or infinite. After all, even if they are infinite, the fact that they converge must mean that all but a finite number of terms are very small, or at least too small to see on a computer screen, so that they can be taken to be finite. Doing so means that one needs only a finite number of *deterministic* functions, and a finite number of random variables, to understand an entire random field. This, of course, has significant implications for simulation, which we shall look at in some detail in Chapter 7.

For the moment, however, we shall concentrate on the more theoretical aspects of fields defined by such sums. A particularly interesting question is how general they are.

Clearly, if the random coefficients in (2.3.17) are Gaussian, then so is the sum. However, there is also a converse, which states that *every* centered Gaussian process with a continuous covariance function has an expansion[15] of the form (2.3.17), in which the 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 $f$, and possess certain orthogonality properties[16].

[15] In general, the convergence in the expansion (2.3.17) is in $L^2(\mathbb{P})$ for each $t \in T$, but, if $f$ is Gaussian and continuous with probability one, then the convergence is uniform over $T$, and with probability one. For more details – and there are many – see Chapter 3 of *RFG* for an introduction to a rich and beautiful theory.

[16] Orthogonality is in the so-called *reproducing kernel Hilbert space* (RKHS) of $f$. 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 versa. To construct it, start with

$$S = \Big\{ u : T \to \mathbb{R} : \, u(\cdot) = \sum_{i=1}^{n} a_i C(s_i, \cdot), \ a_i \text{ real}, \ s_i \in T, \ n \geq 1 \Big\},$$

and define an inner product on $S$ by setting

$$(u,v)_H = \Big( \sum_{i=1}^{n} a_i C(s_i, \cdot), \sum_{j=1}^{m} b_j C(t_j, \cdot) \Big)_H = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j C(s_i, t_j).$$

To see how this works, we consider the case in which the parameter set $T$ is a compact domain in $\mathbb{R}^N$. The corresponding expansion of $f$, given by (2.3.22) below, is then known as the *Karhunen-Loève expansion*. To set up the expansion, suppose that $f$ has covariance function $C$ and define an operator $\mathcal{C}$, taking the space of square integrable functions on $T$ to itself, by

$$(\mathcal{C}\psi)(t) = \int_T C(s,t)\psi(s)\,ds.$$

Suppose that $\lambda_1 \geq \lambda_2 \geq \ldots$, and $\psi_1, \psi_2, \ldots$, are, respectively, the (ordered) eigenvalues and normalised eigenfunctions of the operator. That is, the $\lambda_n$ and $\psi_n$ solve the integral equation

$$\int_T C(s,t)\psi(s)\,ds \;=\; \lambda\psi(t), \tag{2.3.20}$$

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*, which states that

$$C(s,t) = \sum_{n=1}^{\infty} \lambda_n \psi_n(s)\psi_n(t)\,, \tag{2.3.21}$$

where the series converges absolutely and uniformly on $T \times T$.

The Karhunen-Loève expansion of $f$ is then obtained by setting $\varphi_n = \lambda_n^{\frac{1}{2}}\psi_n$ in the expansion (2.3.17), so that

$$f_t = \sum_{n=1}^{\infty} \lambda_n^{\frac{1}{2}}\,\xi_n\,\psi_n(t), \tag{2.3.22}$$

where the $\xi_n$ are i.i.d. $N(0,1)$. For an example of how this approach works for Brownian motion on the line, see Exercise 2.8.11.

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 (2.3.20), even for parameter sets as simple as the unit interval. Moreover, the handful

---

Note that this inner product has the unusual property that

$$(u, C(t,\cdot))_H = \Big(\sum_{i=1}^n a_i C(s_i,\cdot),\, C(t,\cdot)\Big)_H = \sum_{i=1}^n a_i C(s_i,t) = u(t).$$

This is called the *reproducing kernel* property, and the closure of $S$ under the corresponding norm is called the RKHS of $f$ or of $C$.

of examples for which it can be solved involve non-differentiable processes, well outside our range of interests. Nevertheless, it is nice to know that such expansions exist, and, as we shall soon see, they have a wealth of applications.

In particular, the same approach will work if $T$ is replaced by a finite set of points[17]. Then the integral in (2.3.20) becomes a finite sum, as does the expansion (2.3.22), and eigenfunctions become eigenvectors. Thus, even if no analytic solution is possible, we now have an eigenvector problem that will be easy to solve numerically, as long as the number of points is not too large.

We shall return to this in Chapter 7, when we discuss simulations. However, the important thing to note, already at this stage, is that, whether we are in the finite or infinite scenario, in order to simulate $f$ via the Karhunen-Loève expansion one needs *only once* to solve a deterministic eigenfunction or eigenvector problem, and then simulate nothing more difficult than a set of i.i.d. standard normals.

## 2.4 Stationarity and Isotropy

Although we have already met both stationarity and isotropy more than once, the time has now come to look at both of them in some detail. Understanding the structure of random fields with these properties is imperative for working in the subject at all, and understanding the analytic properties of stationary and isotropic covariance functions will be crucial for carrying out many of the calculations we shall need to make later.

The most important classic results of this chapter are the spectral distribution and spectral representation theorems for $\mathbb{R}^N$ which we shall meet soon in Section 2.4.2. However, the most important results for us will be some of the consequences of these theorems for relationships between spectral moments, and these are concentrated in Section 2.4.3. This is the one part of this section that you will almost definitely need to come back to, even if you have decided that you are familiar enough with stationarity to skip the remainder for the moment.

### 2.4.1 Basic Definitions

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[18]. Hence, unless otherwise stated, we shall assume

[17] Think of a continuous parameter space being replaced by a grid of points, and the smooth random field being replaced by the values of the original field on this grid. This kind of discretization is, of course, what one always has on a computer.

[18] The main reason for this is that it is so much more convenient to multiply complex exponentials than it is to multiply sines and cosines. No damage is done, since in dealing with real valued processes one just restricts to complex processes whose imaginary parts have mean and variance both zero.

throughout this section that $f(t) = (f_R(t) + if_I(t))$ takes values in the complex plane $\mathbb{C}$ and that

$$\mathbb{E}\left\{\|f(t)\|^2\right\} = \mathbb{E}\left\{f_R^2(t) + f_I^2(t)\right\} < \infty.$$

(Both $f_R$ and $f_I$ are, obviously, real valued.) The parameter space of $f$ will be $\mathbb{R}^N$ although it could be any space with a group structure without much change in notation. 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[19]. A complex process $f$ is Gaussian if $\sum_i \alpha_i f_{t_i}$ is a complex Gaussian variable for all sequences $\{t_i\}$ in $T$ and complex $\{\alpha_i\}$.

Since $f_t \in \mathbb{C}$, 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

$$C(s,t) \stackrel{\Delta}{=} \mathbb{E}\left\{[f(s) - m(s)]\,\overline{[f(t) - m(t)]}\right\}, \tag{2.4.1}$$

with the bar denoting complex conjugation.

Two basic properties of covariance functions follow immediately from (2.4.1).

(i) $C(s,t) = \overline{C(t,s)}$, which becomes the simple symmetry $C(s,t) = C(t,s)$ if $f$ (and so $C$) is real valued.

(ii) For any $k \geq 1$, $t_1, \ldots, t_k \in T$ and $z_1, \ldots, z_k \in \mathbb{C}$, the Hermitian form $\sum_{i=1}^k \sum_{j=1}^k C(t_i, t_j) z_i \overline{z}_j$ is always real and non-negative. We summarise this, as before, by saying that $C$ is non-negative definite.

Stationarity and homogeneity of $f$ are defined as in the real valued case, whether they be strict or second order, the latter requiring that the mean is constant and $C(s,t)$ is a function of the difference $s - t$ only.

The Gaussian situation is, however, a little different in the complex case. Whereas weakly stationary, real valued, Gaussian processes are also strongly stationary, in the complex case we also require that[20]

$$C'(s,t) = \mathbb{E}\{[f(s) - m(s)]\,[f(t) - m(t)]\}$$

is also a function only of $s - t$. If $f$ is real valued, $C \equiv C'$ and so we are back to the usual situation that strong and weak stationarity are equivalent for Gaussian processes.

Isotropy has the same definition in the complex case as in the real one.

[19] 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 $\mathbb{E}\{X_I X_R\}$.

[20] The reason for the additional condition in the complex case lies in the structure of the multivariate complex normal distribution. Following on from Footnote 19, in order to move from weak stationarity to strong stationarity we need that *all* covariances, including those between real and imaginary parts and among themselves. Knowing both $C$ and $C'$ is enough to do this, but knowing only one is not enough.

### 2.4.2 Spectral Distribution Theorem

We shall now investigate some of the basic properties of covariance functions, and what information can be easily gleaned from them, starting with an important characterization result.

The result, which dates back to Bochner [21], in the setting of (non-stochastic) Fourier analysis, is classical and a proof can be found in almost any text on Fourier analysis.

**Theorem 2.4.1 (spectral distribution theorem).** *A continuous function $C : \mathbb{R}^N \to \mathbb{C}$ is non-negative definite (i.e. a covariance function) if and only if there exists a finite measure $\nu$ on the Borel $\sigma$-field $\mathbb{B}^N$ of $\mathbb{R}^N$ such that*

$$C(t) = \int_{\mathbb{R}^N} e^{i\langle t,\lambda\rangle}\, \nu(d\lambda), \tag{2.4.2}$$

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

The "if" part of the theorem is easy to prove (see Excercise 2.8.7).

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 $F : \mathbb{R}^N \to [0, \sigma^2]$ given by

$$F(\lambda) \stackrel{\Delta}{=} \nu\left(\prod_{i=1}^{N}(-\infty, \lambda_i]\right), \qquad \lambda = (\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N, \tag{2.4.3}$$

is called the *spectral distribution function*[21]. When $F$ is absolutely continuous the corresponding density is called the *spectral density*. .

Note that if $C$ is the covariance of a real valued random field, then $C$ must also be real valued. In this case, it follows that $\nu$ must be a symmetric measure, in the sense that $\nu(A) = \nu(-A)$ for all $A \in \mathbb{B}^N$. In terms of spectral densities $p$, this implies that $p(x) = p(-x)$ for all $x \in \mathbb{R}^N$.

Similarly, if $C$ is isotropic then $\nu$ must be spherically symmetric, and the spectral density must satisfy $p(x) = p(xM)$ for all $x \in \mathbb{R}^N$ and any rotation matrix $M$.

It is important to note that 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 2.4.2) for which we offer two approaches, in Section 2.4.6.

However, before we turn to this more powerful result, we shall collect some easy but important facts that can be gleaned from covariance functions per se.

[21] 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.

### 2.4.3 Spectral Moments and Derivatives of Random Fields

We start by taking a closer look at spectral measures and, in particular, their moments. Given the the spectral representation (2.4.2) we define the *spectral moments*

$$\lambda_{i_1 \dots i_N} \stackrel{\Delta}{=} \int_{\mathbb{R}^N} \lambda_1^{i_1} \cdots \lambda_N^{i_N} \; \nu(d\lambda), \tag{2.4.4}$$

for all multi-indices $(i_1, \dots, i_N)$ with $i_j \geq 0$. Assuming that the underlying random field, and so the covariance function, are real valued, so that, as described above, 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.

$$\lambda_{i_1 \dots i_N} \; = \; 0 \qquad \text{if} \;\; \sum_{j=1}^{N} i_j \;\; \text{is odd.} \tag{2.4.5}$$

There are two ways to understand the meaning of the spectral moments of even order. One of these has to do with the 'high frequency components' of $f$, and relies on understanding the spectral representation theorem of Section 2.4.6. The other, which is already in reach, is related to the mean square, or $L^2$, derivatives of random fields.

To define $L^2$ derivatives, choose a point $t \in \mathbb{R}^N$ and a sequence of $k$ 'directions' $t'_1, \dots, t'_k$ in $\mathbb{R}^N$, and write these as $t' = (t'_1, \dots, t'_k)$. We say that $f$ has a *k-th order $L^2$ partial derivative* at $t$, in the direction $t'$, if the limit

$$D^k_{L^2} f(t, t') \stackrel{\Delta}{=} \lim_{h_1, \dots, h_k \to 0} \frac{1}{\prod_{i=1}^k h_i} \Delta^k f\,(t, t', h) \tag{2.4.6}$$

exists in mean square, where $h = (h_1, \dots, h_k)$. Here $\Delta^k f(t, t', h)$ is the symmetrized difference

$$\Delta^k f(t, t', h) = \sum_{s \in \{0,1\}^k} (-1)^{k - \sum_{i=1}^k s_i} \, f\left(t + \sum_{i=1}^k s_i h_i t'_i\right)$$

and the limit in (2.4.6) is interpreted sequentially, i.e. first send $h_1$ to 0, then $h_2$, etc. A simple sufficient condition for $L^2$ partial differentiability of order $k$ in all directions and throughout a region $T \subset \mathbb{R}^N$ is that

$$\lim_{h_1, \dots, h_k, \widehat{h}_1, \dots, \widehat{h}_k \to 0} \frac{1}{\prod_{i=1}^k h_i \widehat{h}_i} \mathbb{E}\left\{ \Delta^k f\,(t, t', h)\, \Delta^k f\left(s, s', \widehat{h}\right) \right\} \tag{2.4.7}$$

exists[22] for all $s, t \in T$, all directions $s', t'$, and all $h = (h_1, \ldots, h_k)$, $\widehat{h} = (\widehat{h}_1, \ldots, \widehat{h}_k)$, where the limits are again to be interpreted sequentially. Note that if $f$ is Gaussian then so are its $L^2$ derivatives, when they exist.

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

$$\frac{\partial^k}{\partial t_{i_1} \ldots \partial t_{i_k}} f(t) \stackrel{\Delta}{=} D^k_{L^2} f\,(t, (e_{i_1}, \ldots, e_{i_k}))$$

of $f$ of various orders.

It is then straightforward (Exercise 2.8.9) to see that the the covariance function of such partial derivatives of a (not necessarily stationary) random field must be given by

$$\mathbb{E}\left\{ \frac{\partial^k f(s)}{\partial s_{i_1} \partial s_{i_1} \ldots \partial s_{i_k}} \; \frac{\partial^k f(t)}{\partial t_{i_1} \partial t_{i_1} \ldots \partial t_{i_k}} \right\} = \frac{\partial^{2k} C(s,t)}{\partial s_{i_1} \partial t_{i_1} \ldots \partial s_{i_k} \partial t_{i_k}}. \quad (2.4.8)$$

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 $\alpha, \beta, \gamma, \delta \in \{0, 1, 2, \ldots\}$, then (still Exercise 2.8.9) (2.4.2) implies

$$\begin{aligned} \mathbb{E}\left\{ \frac{\partial^{\alpha+\beta} f(t)}{\partial^\alpha t_i \partial^\beta t_j} \; \frac{\partial^{\gamma+\delta} f(t)}{\partial^\gamma t_k \partial^\delta t_l} \right\} &= (-1)^{\alpha+\beta} \frac{\partial^{\alpha+\beta+\gamma+\delta}}{\partial^\alpha t_i \partial^\beta t_j \partial^\gamma t_k \partial^\delta t_l} C(t)\Big|_{t=0} \quad (2.4.9) \\ &= (-1)^{\alpha+\beta}\, i^{\alpha+\beta+\gamma+\delta} \int_{\mathbb{R}^N} \lambda_i^\alpha \lambda_j^\beta \lambda_k^\gamma \lambda_l^\delta \, \nu(d\lambda), \end{aligned}$$

where the $i$ in front of the integral is obviously $\sqrt{-1}$. Note that although this equation 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.

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 (2.4.5), and taking $\alpha = \gamma = \delta = 0,\ \beta = 1$ in (2.4.9)

$$f(t) \text{ and } f_j(t) \text{ are uncorrelated,} \quad (2.4.10)$$

for all $j$ and all $t$. If $f$ is Gaussian, this is equivalent to independence. Note that (2.4.10) does *not* imply that $f$ and $f_j$ are uncorrelated *as processes*. In general, for $s \neq t$, we will have that $\mathbb{E}\{f(s) f_j(t)\} = -C_j(s-t) \neq 0$.

[22] 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 $n, m \to \infty$.

(iii) Taking $\alpha = \gamma = \delta = 1,\ \beta = 0$ in (2.4.9) gives that

$$f_i(t) \text{ and } f_{kl}(t) \text{ are uncorrelated} \tag{2.4.11}$$

- for all $i, k, l$ and all $t$. Again, if $f$ is Gaussian, this is equivalent to independence.

Under the additional condition of isotropy, with its implication of spherical symmetry for the spectral measure, the structure of the spectral moments simplifies significantly, as do the correlations between various derivatives of $f$. In particular, it follows immediately from (2.4.9) that

$$\mathbb{E}\{f_i(t) f_j(t)\} \ = \ -\mathbb{E}\{f(t) f_{ij}(t)\} \ = \ \lambda_2 \delta_{ij} \tag{2.4.12}$$

where $\delta_{ij}$ is the Kronecker delta and $\lambda_2$ is the *second spectral moment*

$$\lambda_2 \stackrel{\Delta}{=} \int_{\mathbb{R}^N} \lambda_i^2 \, \nu(d\lambda),$$

which, because of isotropy, is independent of the value of $i$. Consequently, if $f$ is Gaussian, then the first order derivatives of $f$ are independent of one another, in addition to being independent of $f$ itself.

### 2.4.4 Constant Variance

It will be important for us in later chapters that some of the relationships of the previous section continue to hold under a condition much weaker than stationarity. Of particular interest is knowing when (2.4.10) 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, from (2.4.9), we have that

$$\mathbb{E}\{f(t)\, f_j(t)\} = \frac{\partial}{\partial t_j} C(t,s)\Big|_{s=t} = \frac{\partial}{\partial s_j} C(t,s)\Big|_{s=t}. \tag{2.4.13}$$

Since constant variance implies that $\partial/\partial t_j C(t,t) \equiv 0$, this and the equivalence of the above two partial derivatives implies that these must also be identically zero. Consequently, $f$ and its first order derivatives *at any given point* 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 $i \neq k$.

### 2.4.5 White Noise and Integration

With some easy, but extremely important, consequences of the spectral distribution theorem established, we now turn to understanding the structure of stationary random fields in a little more depth. This structure relies on some rather simple stochastic calculus. However, although stochastic calculus is a subject with an enormous and difficult literature, we shall need only a very small part of it, and even then we shall do it without detailed proofs.

The principal aim of the current section is to set up some of the basic machinery of stochastic integration. When this is done, we can turn to the spectral representation theorem, which will provide the deeper understanding of stationary that we are looking for. *En passant*, in this section we shall meet for the first time an extremely important class of random fields, known as moving averages.

We start with a collection of independent Gaussian random variables. While such a collections are easy to construct if they are finite or countable, serious technical difficulties obstruct the construction for uncountable collections. This is why white noise, which is essentially a collection of i.i.d. random variables indexed by the points $\mathbb{R}^N$, is such a delicate object. The way around this is to avoid giving the value of the noise at specific points, and to treat is as a sort of signed measure over subsets of $\mathbb{R}^N$.

More formally, suppose $\nu$ is a measure on $\mathbb{R}^N$. The classical example is Lebesgue measure, but any measure that can be written in the form

$$\nu(A) = \int_A p(x)\,dx, \tag{2.4.14}$$

will do. Here $A \subset \mathcal{B}^N$ and $p$ is a non-negative, but not necessarily integrable function on $\mathbb{R}^N$. For obvious reasons $p$ is called the (Radon-Nikodym) density of $\nu$, although, since it generally will not integrate to one (or, indeed, anything finite) it is not a probability density. If (2.4.14) holds we say that $\nu$ is continuous.

However, completely discrete $\nu$, that put all their mass on a finite or countable set of points, will also be important to us, and we shall even meet measures that distribute their mass smoothly over lower dimensional subsets of $\mathbb{R}^N$, such as spheres.

We can now define a *Gaussian noise*[23] $W$ based on $\nu$, or 'Gaussian $\nu$-noise' as a random process defined on the Borel subsets of $\mathbb{R}^N$ such that, for all $A, B \in \mathcal{B}^N$ with $\nu(A)$ and $\nu(B)$ finite,

$$W(A) \sim N(0, \nu(A)). \tag{2.4.15}$$

$$A \cap B = \emptyset \Rightarrow W(A \cup B) = W(A) + W(B) \text{ a.s.} \tag{2.4.16}$$

$$A \cap B = \emptyset \Rightarrow W(A) \text{ and } W(B) \text{ are independent.} \tag{2.4.17}$$

[23] While the notation '$W$' is inconsistent with our decision 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.

Property (2.4.17) is described by saying that $W$ has *independent increments*[24]. When $\nu$ is Lebesgue measure, $W$ is called *white noise* and it is closely connected to Brownian motion ($N = 1$) and the Brownian sheet. See Exercise 2.8.12.

Having defined Gaussian noises, our next step is to make sense out of the integral

$$\int_{\mathbb{R}^N} \varphi(t)\, W(dt), \tag{2.4.18}$$

for *deterministic* $\varphi$ with $\int \varphi^2(x)\, \nu(dx) < \infty$.

In principle, this is not hard to do. Roughly, the argument goes as follows: Start with simple functions

$$\varphi(t) = \sum_1^n a_i \mathbb{1}_{A_i}(t), \tag{2.4.19}$$

where $A_1, \ldots, A_n \subset \mathbb{R}^N$ are disjoint, and the $a_i$ are real, and define

$$W(\varphi) \equiv \int_T \varphi(t)\, W(dt) = \sum_1^n a_i\, W(A_i). \tag{2.4.20}$$

It follows immediately from (2.4.15) and (2.4.17) that in this case $W(\varphi)$ has zero mean and variance given by $\sum a_i^2\, \nu(A_i)$. Now think of $W$ as a mapping from simple functions to random variables, and extend it to all functions square integrable with respect to $\nu$. The extension is standard, and we send you to either *RFG* or any standard text for details.

An important point to note, however, is that if $\varphi$ and $\psi$ are the simple functions

$$\varphi(t) = \sum_1^n a_i \mathbb{1}_{A_i}(t), \quad \psi(t) = \sum_1^n b_i \mathbb{1}_{A_i}(t),$$

then, again by (2.4.15) and (2.4.17),

$$\begin{aligned} \mathbb{E}\{W(\varphi)W(\psi)\} &= \mathbb{E}\left\{ \sum_1^n a_i W(A_i) \cdot \sum_1^n b_i W(A_i) \right\} && (2.4.21) \\ &= \sum_1^n a_i b_i \mathbb{E}\left\{ [W(A_i)]^2 \right\} \\ &= \sum_1^n a_i b_i \nu(A_i) && (2.4.22) \\ &= \int_{\mathbb{R}^N} \varphi(t)\psi(t)\, \nu(dt), \end{aligned}$$

[24] Much of what follows can be done without the assumption of Gaussianity, in which case (2.4.15) is replaced by requiring that $\mathbb{E}\{W(A)\} = 0$ and $\mathbb{E}\{[W(A)]^2\} = \nu(A)$.

a result that extends, in general, to

$$\mathbb{E}\left\{\int_{\mathbb{R}^N}\varphi(t)\,W(dt)\int_{\mathbb{R}^N}\psi(t)\,W(dt)\right\}=\int_{\mathbb{R}^N}\varphi(t)\psi(t)\,\nu(dt). \quad (2.4.23)$$

Note also that since linear combinations and limits of Gaussian random variables remain Gaussian, it follows that $W(\varphi)$ is also Gaussian.

With our integral defined, we can now start looking at some examples of what can be done with it. One very important use is the definition of *moving average* random fields. These are now easily constructed by taking white noise with Lebesgue measure as control measure, choosing a square integrable function $\varphi:\mathbb{R}^N\to\mathbb{R}$, and defining

$$f(t)\stackrel{\Delta}{=}\int_{\mathbb{R}^N}\varphi(t-s)\,W(ds). \quad (2.4.24)$$

Since

$$\begin{aligned}\mathbb{E}\{f(t)f(s)\}&=\int_{\mathbb{R}^N}\varphi(t-u)\varphi(s-u)\,du\\&=\int_{\mathbb{R}^N}\varphi(t-s+v)\varphi(v)\,dv\\&\stackrel{\Delta}{=}C(t-s),\end{aligned}$$

moving averages are clearly stationary.

Furthermore, if $\varphi$ is spherically symmetric, so that it depends only on $|t|$, $f$ is isotropic. This is easiest to see as follows: let $M$ be a rotation matrix, then

$$\begin{aligned}C(sM,tM)&=\mathbb{E}\{f(tM)f(sM)\}\\&=\int_{\mathbb{R}^N}\varphi(tM-u)\varphi(sM-u)\,du\\&=\int_{\mathbb{R}^N}\varphi(tM-uM)\varphi(sM-uM)\,du\\&=\int_{\mathbb{R}^N}\varphi(t-u)\varphi(s-u)\,du\\&=C(s,t)\end{aligned}$$

### 2.4.6 Spectral Representation Theorem

Moving averages gave us examples of stationary fields that are rather easy to generate from white noise. Now, however, we want to use stochastic integrals to find a very general way of representing *all* stationary fields on $\mathbb{R}^N$, via the so-called *spectral representation*.

We require a minor extension of the integrals of the previous section, however, since we want to work in a complex valued scenario. Thus, given a

measure $\nu$ on $\mathbb{R}^N$, define a *complex* $\nu$-noise $W$ to be a $\mathbb{C}$-valued process satisfying

$$\mathbb{E}\{W(A)\} = 0, \quad \mathbb{E}\{W(A)\overline{W(A)}\} = \nu(A). \tag{2.4.25}$$

$$A \cap B = \emptyset \quad \Rightarrow \quad W(A \cup B) = W(A) + W(B) \text{ a.s.} \tag{2.4.26}$$

$$A \cap B = \emptyset \quad \Rightarrow \quad \mathbb{E}\{W(A)\overline{W(B)}\} = 0. \tag{2.4.27}$$

We can, in addition, assume that $W$ is Gaussian, in which case you should note that (2.4.25) does not specify all its parameters[25].

It is then a straightforward exercise to extend the construction of the previous section to an $L^2$ stochastic integral

$$W(\varphi) = \int_{\mathbb{R}^N} \varphi(\lambda)\, W(d\lambda)$$

for $\varphi : \mathbb{R}^N \to \mathbb{C}$ with $\int_{\mathbb{R}^N} \|\varphi\|^2\, d\nu < \infty$. It is immediate from the construction that

$$\mathbb{E}\{W(\varphi)\overline{W(\psi)}\} = \int_{\mathbb{R}^N} \varphi(\lambda)\overline{\psi(\lambda)}\, \nu(d\lambda). \tag{2.4.28}$$

This construction allows us to state the following important result.

**Theorem 2.4.2 (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*

$$f(t) = \int_{\mathbb{R}^N} e^{i\langle t,\lambda\rangle}\, W(d\lambda) \tag{2.4.29}$$

*has covariance*

$$C(s,t) = \int_{\mathbb{R}^N} e^{i\langle (s-t),\lambda\rangle}\, \nu(d\lambda), \tag{2.4.30}$$

*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 $\nu$ there corresponds a complex (Gaussian) $\nu$-noise $W$ on $\mathbb{R}^N$ such that (2.4.29) holds in mean square for each $t \in \mathbb{R}^N$.*

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

[25] See Footnote 19 above. In general, if we split a complex $\nu$-noise into its real and imaginary parts, $W_R$ and $W_I$ say, it does not follow from (2.4.27) that $A \cap B = \emptyset$ implies any of $\mathbb{E}\{W(A)W(B)\} = 0$, $\mathbb{E}\{W_R(A)W_R(B)\} = 0$, $\mathbb{E}\{W_I(A)W_I(B)\} = 0$, or $\mathbb{E}\{W_I(A)W_R(B)\} = 0$. Indeed, this is most definitely not the case for the complex $W$ of Theorem 2.4.2 when, for example, the stationary process $f$ there is real valued. See the discussion following (2.4.31) on 'real' spectral representations, in which the above problem is addressed by taking $W$ to be defined on a half space rather than all of $\mathbb{R}^N$.

In one direction, this theorem does need a proof. It is a consequence of the construction of the stochastic integral that the process $f$ defined by (2.4.29) has covariance function (2.4.30). The other direction is not so easy, although not hard. You can find the details in almost any book on time series – our favourite is [23] – for processes on either $\mathbb{Z}$ or $\mathbb{R}$, and the extension to $\mathbb{R}^N$ is trivial[26]. Note that the spectral representation theorem, Theorem 2.4.2, is more general than the spectral distribution theorem, Theorem 2.4.1. The spectral distribution theorem gives the Fourier expansion of the covariance function only, while the spectral representation theorem gives the Fourier expansion of the process itself.

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, *and computationally*, it is still generally more convenient to use the complex formulation. Nevertheless, the real representation is also useful. To describe it, note firstly that if $f$ is real, the symmetry of the spectral measure $\nu$ allows us to introduce three[27] new measures, on $\mathbb{R}_+ \times \mathbb{R}^{N-1}$, by setting[28]

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

We can now rewrite (2.4.2) in real form, as

$$C(t) = \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \cos(\langle \lambda, t \rangle)\, \mu(d\lambda). \tag{2.4.31}$$

There is also a corresponding real form of the spectral representation (2.4.29). The fact that the spectral representation yields a real valued process also implies certain symmetries on the spectral process $W$. In particular, it turns out that there are two independent *real* valued $\mu$-noises, $W_1$ and $W_2$, such that[29]

$$f_t = \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \cos(\langle \lambda, t \rangle)\, W_1(d\lambda) + \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \sin(\langle \lambda, t \rangle)\, W_2(d\lambda). \tag{2.4.32}$$

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

[26] There is also an inverse to (2.4.29), expressing $W$ as an integral involving $f$, but we shall have no need of it.

[27] Note that if $\nu$ is continuous (so that (2.4.14) holds) the second of these measures, $\nu_2$, will be identically zero.

[28] There is nothing special about the half-space $\lambda_1 \geq 0$ taken in this representation. Any half space will do.

[29] 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.

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 to this. While it is not necesary for the rest of the theory, it does help develop intuition and very much comes into its own when we turn to issues of simulation in Chapter 7.

One way to think of the integral in (2.4.32) is via the approximating sum

$$\sum_i \{\cos(\langle \lambda_i t \rangle) W_1(\Lambda_i) \ + \ \sin(\langle \lambda_i, t \rangle) W_2(\Lambda_i)\} \tag{2.4.33}$$

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 (2.4.33) 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 (2.4.33) 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 two dimensions let $\lambda$ be a generic $\lambda_i$ in (2.4.33) (i.e. drop the subscript on $\lambda_i$) with components $\lambda_1$ and $\lambda_2$, we have that an elementary cosine wave is of the form $\cos(\lambda_1 t_1 + \lambda_2 t_2)$. The points $\lambda_1$ and $\lambda_2$ 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 $\arctan(\lambda_2/\lambda_1)$ with the $t_1$ axis and having wavelength $2\pi/\sqrt{\lambda_1^2 + \lambda_2^2}$, the distance between troughs or crests, as measured along the line perpendicular to the crests. An example is given in Figure 2.4.1.

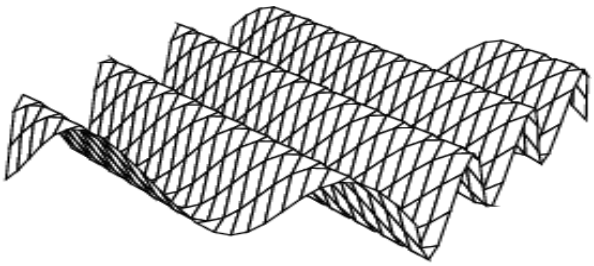

**Fig. 2.4.1.** The elementary wave form $\cos(\lambda_1 t_1 + \lambda_2 t_2)$ in $\mathbb{R}^2$.

The corresponding sine function is exactly the same, except for the obvious a phase shift of 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 2.4.2 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 $N(0,1)$. Higher dimensional situations behave analogously.

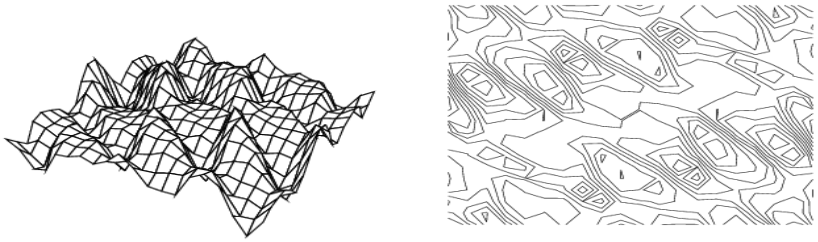

**Fig. 2.4.2.** A more realistic surface based on (2.4.33), along with contour lines at the zero level.

Another way to think of the spectral representation theorem is as a particular (rather non-rigorous) application of the Karhunen-Loève expansion of Section 2.3.5. Let's start by assuming that all that we wrote back there also works for complex valued processes and on the definitely unbounded region $\mathbb{R}^N$. Since we are in the stationary scenario, it is actually quite easy to find eigenfunctions for the integral equation (2.3.20), via complex exponentials. In fact, for any $\lambda \in \mathbb{R}^N$, the function $e^{i\langle t,\lambda\rangle}$ (as a function of $t \in \mathbb{R}^N$) satisfies

$$\begin{aligned}\int_{\mathbb{R}^N} C(s,t)\, e^{i\langle s,\lambda\rangle}\, ds &= \int_{\mathbb{R}^N} C(t-s)\, e^{i\langle s,\lambda\rangle}\, ds \\ &= e^{i\langle t,\lambda\rangle} \int_{\mathbb{R}^N} C(u)\, e^{-i\langle u,\lambda\rangle}\, du \\ &= K_\lambda\, e^{i\langle t,\lambda\rangle},\end{aligned}$$

for some, possibly zero, $K_\lambda$.

Suppose that $K_\lambda \neq 0$ for only a countable number of $\lambda \in \mathbb{R}^N$. Then the stationary, complex version of the Mercer expansion (2.3.21) can be written as

$$C(t) = \sum_\lambda K_\lambda\, e^{i\langle t,\lambda\rangle}, \tag{2.4.34}$$

while the Karhunen-Loève expansion becomes

$$f(t) = \sum_\lambda K_\lambda^{1/2} \xi_\lambda\, e^{i\langle t,\lambda\rangle}. \tag{2.4.35}$$

These are, respectively, special cases of the spectral distribution theorem (cf. (2.4.2)) and the spectral representation theorem (cf. (2.4.29)) when the spectral measure is discrete.

Despite the minor irregularity of assuming that $f$ is complex valued, the above argument is completely rigorous. On the other hand, what follows for the case in which $K_\lambda \neq 0$ on an uncountable set, is not. Nevertheless, it is still somewhat enlightening to look at. In this case, one could imagine replacing the summations in (2.4.34) and (2.4.35) by integrals, to obtain

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

and

$$f(t) = \int_{\mathbb{R}^N} K_\lambda^{1/2} \xi_\lambda \, e^{i\langle t,\lambda\rangle} \, d\lambda. \tag{2.4.36}$$

Everything is well defined in the first of these integrals, but in the second we have the problem that the $\xi_\lambda$ should be independent for each $\lambda$. The way to make mathematics of this is, of course, via stochastic integration, which is precisely what we did above. Thus, instead of (2.4.36), we could use (2.4.29) in the spectral representation theorem and write

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

where $W$ is Gaussian $\nu$-noise according to the measure defined by $\nu(d\lambda) = K_\lambda d\lambda$.

Despite the lack of rigor of the above ideas, they do give a useful way of thinking about the spectral representation theorem.

### 2.4.7 Isotropy

We now turn to the special case of stationary, isotropic random fields on $\mathbb{R}^N$, for which the covariance function $C(t)$ is a function only of $|t|$. Not surprisingly, isotropy implies significant simplifying consequences for the spectral distribution and representation theorems.

As we already noted in Section 2.4.2, spectral measures of isotropic random fields are spherically symmetric. Consequently, they cannot have all their mass 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 behavior is to have all its probability concentrated in an annulus of the form

$$\{\lambda \in \mathbb{R}^N : \, a \le |\lambda| \le b\}, \qquad a \le b.$$

In such a case it is not hard to see that 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.

Other, often unexpected, restrictions on covariance functions also arise from isotropy. For example, isotropic covariance functions cannot be very negative, a result due originally to the Matérn [62], who, in the 1950's, was one of the first researchers to actually employ random fields as a statistical modelling tool. In particular, he showed that the covariance functions of centered, isotropic random fields must satisfy (Exercise 2.8.14)

$$C(t) \geq -C(0)/N, \qquad \text{for all } t \in \mathbb{R}^N. \tag{2.4.37}$$

More important than the above, however, are the consequences of isotropy for the spectral distribution and representation theorems. We shall state these formally, and even show how they are derived, since the first response of a modern mathematician on seeing Bessel functions is to imagine that they arise from something quite mysterious. In fact, here quite the opposite is the case, and although the original statement of the spectral distribution theorem, with its complex exponentials, may look simpler than the version that follows, there are other consequences, at the level of the spectral representation theorem for isotropic processes (Theorem 2.4.4) that have quite significant practical importance. In particular, this will help set the scene for what happens when we look at random fields on spheres.

The following result, due originally to Schoenberg [83] (in a somewhat different setting) and Yaglom [102] describes what happens.

**Theorem 2.4.3.** *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*

$$C(t) \;=\; \int_0^\infty \frac{J_{(N-2)/2}(\lambda|t|)}{(\lambda|t|)^{(N-2)/2}}\,\mu(d\lambda), \tag{2.4.38}$$

*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 (2.4.29) by using the symmetry properties of $\nu$.

We commence by converting to polar coordinates, $(\lambda, \theta_1, \ldots, \theta_{N-1})$, $\lambda \geq 0$, $(\theta_1, \ldots, \theta_{N-1}) \in S^{N-1}$, where

$$S^{N-1} = \big\{t \in \mathbb{R}^N : |t| = 1\big\}$$

is the unit sphere in $\mathbb{R}^N$. Define a measure $\mu$ on $\mathbb{R}_+$ by setting $\mu([0,\lambda]) = \nu(B_\lambda^N)$, and extending as usual, where

$$B_\lambda^N = \big\{t \in \mathbb{R}^N : |t| \leq \lambda\big\}$$

is the $N$-ball of radius $\lambda$ and $\nu$ is the spectral measure of (2.4.2).

Then, on substituting into (2.4.2) with $t = (|t|, 0, \ldots, 0)$ 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 $\theta_1, \ldots, \theta_{N-2}$ 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

$$s_N \stackrel{\Delta}{=} \frac{2\pi^{N/2}}{\Gamma(N/2)}, \qquad N \geq 0, \tag{2.4.39}$$

is the surface area[30] of $S^{N-1}$.

The inside integral can be evaluated in terms of Bessel functions to yield, up to a multiplicative constant,

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

which, on absorbing all constants into $\mu$, completes the proof. □

For small values of the dimension $N$, (2.4.38) can be simplified even further. For example, substituting $N = 2$ into (2.4.38) 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 [98], or as special cases in the review by Yaglom [103]. Another way to verify it would be to check that the representation given in Theorem 2.4.4 below yields the covariance structure of (2.4.38). This is essentially an exercise in the manipulation of special functions.

The spectral representation of isotropic fields on $\mathbb{R}^N$ is based on the so-called *spherical harmonics* on the $(N-1)$-sphere, which form an orthonormal basis for the space of complex, square integrable functions on $S^{N-1}$ equipped with the usual surface measure. We shall denote them by $\{h_{ml}^{(N-1)}$, $l = 1, \ldots, d_m$, $m = 0, 1, \ldots\}$ where $d_m = \binom{N+m-2}{N-2}(N+2m-2)/(N+m-2)$.

[30] Including the case $N = 0$ here is not a mistake. We shall need it later. In any case, it makes perfect sense if we think of $\mathbb{R}^0$ as the integers with counting measure.

Figure 2.4.3 is a representation of the first few spherical harmonics, with $N = 3$ and $0 \leq m \leq 3$ in descending order in $m$. In each representation the distance from the underlying sphere indicates the size of $|h_{ml}|$, while the color/shading represents the ray in the complex plane on which $h_{ml}$ sits. The color/shading key is shown in the disc at top right. Thus $h_{01}$ takes a constant, positive, real value (1), while $h_{11}$ takes positive real values in the upper hemisphere and negative real values in the lower hemisphere, etc.

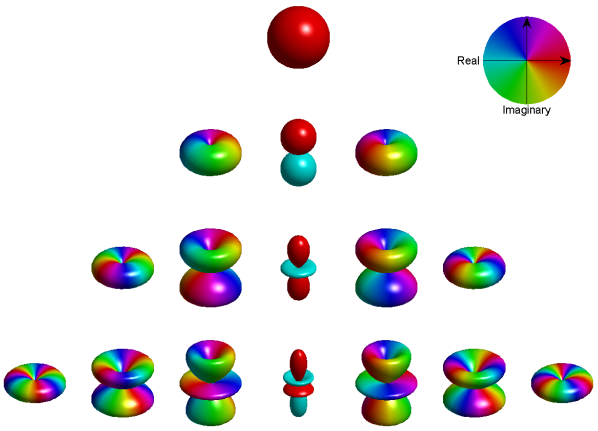


**Fig. 2.4.3.** Spherical harmonics, $N = 3$, $0 \leq m \leq 3$.

Given the spherical harmonics, we can 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 setting

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

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 (2.4.29), 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 2.4.4.** *A centered, mean square continuous, isotropic random field on $\mathbb{R}^N$ can be represented by*

$$f(t) \;=\; f(r,\theta) \;=\; \sum_{m=0}^{\infty}\sum_{l=1}^{d_m} f_{ml}(r)h_{ml}^{(N-1)}(\theta). \tag{2.4.40}$$

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 (2.4.40), entering via the $W_{ml}$ and $f_{ml}$. This makes for an important difference between (2.4.40) and the similar looking Karhunen-Loève expansion we met in Section 2.3.5. Another difference lies in the fact that while it is possible to truncate the expansion (2.4.40) to a finite number of terms and retain isotropy, this is not true of the standard Karhunen-Loève expansion. In particular, isotropic fields can never have finite Karhunen-Loève expansions. For a heuristic argument as to why this is the case, recall from 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 an uncountable number of independent variables in the spectral noise process to generate the process via (2.4.29). However a process with a finite Karhunen-Loève expansion provides only a finite number of such variables, which can never be enough.

We close this section on isotropy with a brief discussion of fields in 'space-time'. Taking the lead from the moving ocean waves with which we motivated this chapter, it is not at all 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 $\mathbb{R}\times\mathbb{R}_+$ to $\mathbb{C}$. In such a situation the methods that we used to study the purely isotropic case also 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 $\mathbb{R}_+\times\mathbb{R}^N$.

It is not hard to also develop a corresponding representation for the field itself.

### 2.4.8 Isotropic Fields on the Sphere and Other Homogenous Spaces

While the stationary and isotropic random fields that we have studied so far in this section have a very nice theory, they all had the common property that were defined over the Euclidean spaces $\mathbb{R}^N$. However, when we turn to applications we shall see that there are many random fields defined on more delicate parameter spaces, such as surfaces.

One classic example is electroencephalogram (EEG) data, neurophysiologic measurements of the electrical activity of the brain taking by recording from electrodes placed on the scalp. Unlike the fMRI data of Chapter 1, which is data taken over a solid domain in $\mathbb{R}^3$, EEG data is taken over a surface which, for the moment, we can think of as a sphere. Another example is provided by the Nobel Prize winning COBE (COsmic microwave Background Explorer) astrophysical data. This data is on the sphere of all directions away from the Earth, and measures an integral of microwave radiation throughout the universe, in each direction. We shall look at these two examples in some detail in Chapters 9 and 10.

However, since at the moment we are still doing theory rather than application, there is no reason to stay with the two dimensional sphere, and we introduce

$$S_\lambda^{N-1} = \{t \in \mathbb{R}^N : |t| = \lambda\}, \tag{2.4.41}$$

the sphere in $\mathbb{R}^N$ of radius $\lambda$, so that $S_1^{N-1}$ is the unit sphere $S^{N-1}$.

To define stationarity and isotropy for random fields defined on $S_\lambda^{N-1}$ we need a notion of translation, which is usually done by identifying each point $t \in S_\lambda^{N-1}$ with a rotation, and thinking of $t_1 + t_2$ as a composition of two rotations. Not that on a sphere, stationarity and isotropy are identical concepts, and amount to requiring that the covariance function is invariant under rotations.

Then an interesting corollary of Theorem 2.4.4 is obtained by fixing $r$ in (2.4.40). We then have, for a homogeneous and isotropic random field on $S_\lambda^{N-1}$, a simple representation in terms of uncorrelated random coefficients $f_{ml}(r)$ and spherical harmonics. If the random field is Gaussian, then the coefficients are actually independent, and we will, essentially, have generated a Karhunen-Loève expansion.

Furthermore, the covariance function can be written as

$$C(t_1, t_2) = C(\alpha_{12}) \;=\; \sum_{m=0}^{\infty} \sigma_m^2 C_m^{(N-1)/2}(\cos \alpha_{12}), \tag{2.4.42}$$

where $\alpha_{12}$ is the angular distance between $t_1$ and $t_2$, the $C_m^N$ are Gegenbauer polynomials, and $\sigma_m^2$ is the common variance of the $f_{ml}^2(r)$ in (2.4.40). An important special case of (2.4.42) occurs when $\sigma_m^2 = 0$ for all $m$ other than $m = 1$, and $\sigma_1^2 = 1/(N-1)$. In this case the covariance function is given by

$C(t_1, t_2) = \langle t_1, t_2 \rangle$ and the resulting random field is known as the canonical (isotropic) Gaussian field on the sphere. This field can be constructed directly as $f(t) = \langle t, \xi \rangle$, where $\xi \sim N(0, I_{N\times N})$, in which case the stationarity and isotropy are easy to show (Exercise 2.8.15). This canonical field plays a central rôle in the theory of smooth Gaussian fields, and is the point at which the Gaussian kinematic formulae of Chapter 4 meet the kinematic fundamental formula of Chapter 3. The details of the mathematics can be found in Chapter 15 of *RFG*.

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 quite non-Euclidean parameter spaces, as long as they have some group structure that mimics the ideas of translation and perhaps rotation. This is indeed the case, and what happens, in essence, is that the complex exponentials that have been at the core of all the representations of this section are replaced by the so-called 'characters' of the group.

You can find many more results of this kind, with much more detail, in the classic paper of Yaglom [103] or the book by Leonenko [54]. Other useful treatments, with various emphases, can be found in [14, 40, 55] and [104].

### 2.4.9 Transforming to Stationarity and Isotropy

The notions of stationarity and isotropy are clearly important, and particularly pleasant from a mathematical viewpoint, in that, as we have just seen, they allow for quite simple representations of random fields.

Nevertheless, Nature is not always as kind as to provide stationarity or isotropy, and it is natural to ask what can be done in that case. In particular, is would be natural to ask if there was a simple way to transform general random fields to stationarity and/or isotropy.

Despite the fact that the applied literature abounds with papers having titles which seem to indicate that this is indeed possible, the fact, sadly, is that in general it is not. We shall nevertheless take a moment to indicate what *is* known.

In a literature that seems to have begun with a statistical paper [80] by Sampson and Guttorp in 1992, there has been considerable interest in models of the form

$$f(t) = g(\varphi(t)),$$

where $g$ is a stationary, isotropic random field and $\varphi$ a bijective transformation of some kind. The now generally non-stationary and non-isotropic $f$ can therefore be transformed to stationarity and isotropy by studying $f(\varphi^{-1}(t))$.

In [80] a statistical procedure was developed, based on the so-called *variogram*[31]

[31] While the name is different, the variogram is no more than the square of the canonical metric that we shall meet and heavily use in the following section.

$$V(s,t) \stackrel{\Delta}{=} \mathbb{E}\left\{[f(t) - f(s)]^2\right\},$$

for estimating the transformation $\varphi$ in a non-parametric fashion. The procedure, while obviously useful, applies only for two-dimensional, real valued random fields, and does not seem to have an easy extension to higher dimensions.

There are some more recent papers [7, 8] which approach this problem a different way, by assuming the transformation $\varphi$ to be quasi-conformal. Once again, the approach is limited to two dimensions, and, in this case, the extension to higher dimensions seems completely blocked.

However, neither these procedures nor related ones (see [7, 8] for a recent bibliography) answer the general question we originally asked: Can general non-stationary or non-isotropic fields be easily treated within the stationary/isotropic framework; i.e. fields which we do not know a priori to be simple transforms?

Unfortunately, the answer to this question is negative. Nevertheless, there are some techniques that can help. For example, there is a notion of *local isotropy* that is often useful. To define this, recall from Section 2.4.3 that if a random field is stationary and isotropic, then, among many other properties, its first order partial derivatives are uncorrelated and have common variance. It turns out that many of the properties of Gaussian and related random fields rely only on these consequences, and so we shall define a *locally isotropic random field* to be one with constant variance and

$$\mathbb{E}\left\{f_i(t)f_j(t)\right\} = \lambda_2\delta_{ij}, \tag{2.4.43}$$

for some common second spectral moment $\lambda_2$, where $f_i(t) = \partial f(t)/\partial t_i$.

A natural question to now ask is how easy is it to transform general random fields to locally isotropic ones. For stationary random fields $f$ this is straightforward. If $\Lambda$ is the $N \times N$ matrix of second spectral moments $\lambda_{ij} = \mathbb{E}\left\{f_i(t)f_j(t)\right\}$, then it is trivial to check that the field defined by $f(\Lambda^{-1/2}t)$ is locally isotropic.

In general, however, there is no such simple transformation available. However, there is a trick, based on Riemannian geometry, that allows one to compute many things about constant variance Gaussian random fields as if they were, in fact, locally isotropic. It was this trick that, in many ways, was one of the most important themes of *RFG*. It will be explained in Section 4.3 and heavily used thereafter. However, we shall have to develop a few more concepts and notation before we can introduce it, so we leave it for the moment.

Finally, we note that there is a technique [34, 68, 69, 70] for making stationary fields from non-stationary ones, at the price of enlarging the parameter space. Suppose $f$ is a non-stationary random field $f$ on $T \subset \mathbb{R}^N$, and $\varphi : \mathbb{R}^N \to \mathbb{R}^{2N}$ an injective mapping. Consider the parameter space

---

While there, as the canonical metric, it will play an important rôle for developing theory, it plays no less an important rôle in the applied literature as the variogram.

$$\widehat{T} \stackrel{\Delta}{=} \{(t, \varphi(t)) : t \in T\} \subset \mathbb{R}^{2N},$$

and the random field $\widehat{f}$ on the manifold $\widehat{T}$ defined by $\widehat{f}(x, \varphi(x)) = f(\varphi^{-1}(x))$. Then for certain covariances (necessary and sufficient conditions can be found in the papers just referenced) one can find a $\varphi$ such that $\widehat{f}$ is 'stationary' on $\widehat{T}$ in the sense that

$$\mathbb{E}\left\{\widehat{f}(x, \varphi(x))\widehat{f}(y, \varphi(y))\right\}$$

is a function of $(x, \varphi(x)) - (y, \varphi(y))$ only. However, since it is an open (and hard) question as to whether $\widehat{f}$ can be extended to a stationary process on all of $\mathbb{R}^{2N}$, and the class of covariances for which the technique works is not large, in general it is hard to exploit this form of stationarity in a useful way.

## 2.5 Smoothness of Random Fields

Now that we have discussed the basic construction and structure of random fields, we turn to discussing basic sample path properties. For example, we would like to know when a random field $f$ is continuous, or continuously differentiable. These two issues are precisely what we shall look at in this section, mainly for Gaussian fields.

It is actually quite easy to describe what kinds of conditions are needed to ensure that these smoothness properties hold, at least in the Gaussian case. For simplicity, suppose that $f$ is Gaussian, centered, and stationary, with covariance function $C$. Since $C$ determines the finite dimensional distributions of $f$, and these determine its properties, ultimately we are looking for conditions on $C$.

Now, if $f$ is to be smooth, it must be true that for $s$ and $t$ close, the difference $f(t) - f(s)$ must be small. However, these differences have a known distribution, viz.

$$\begin{aligned} f(t) - f(s) &\sim N\left(0, \mathbb{E}\left\{\left|f(s) - f(t)\right|^2\right|\right\}) \\ &= N\left(0, 2[C(0) - C(t-s)]\right), \end{aligned}$$

the second line a consequence of stationarity. Thus, what we require is that $C$ itself is smooth in the neighborhood of the origin[32]. The only question is "exactly how smooth does it need to be?", and this is what we plan to answer.

Still in the stationary Gaussian scenario, one can also consider smoothness properties from the point of view of the spectral representation theorem. As we described in Section 2.4.6, one can think of stationary processes as a sum of infinitely many sinusoids. In particular, think of the (non-rigorous) Karhunen-Loève type representation (2.4.35), in terms of complex exponentials, in which we wrote

[32] As a consequence, it will be smooth everywhere. See Exercise 2.8.17.

$$f(t) = \sum_{\lambda} K_\lambda^{1/2} \xi_\lambda \, e^{i\langle t,\lambda\rangle}.$$

Obviously, since each function $e^{i\langle t,\lambda\rangle}$ is infinitely differentiable, the first few terms in the sum have no affect on the level of smoothness of $f$. In fact, no finite number of terms can affect it. Thus, the issue of smoothness is determined by the behavior of the tail of the sum, or, equivalently, by the behavior of $\mathbb{E}\{|K_\lambda|\}$ as $|\lambda| \to \infty$. Equivalently, it is determined by the variance of the high frequency components, or, in terms of the spectrum, by its decay rate at infinity.

These two considerations – the behavior of the covariance function at the origin and the behavior of the spectral measure at infinity – are actually the same mathematically, since they are linked by the Tauberian theorems of Fourier analysis. However, from a stochastic point of view, they give us two seemingly quite different ways to think about the smoothness of random fields.

To turn these ideas into mathematics is not trivial, and we shall decribe how to do it in the remainder of this section. The details (but not the results) can, however, be skipped by the reader who cares only about the 'what' rather than the 'why' of smoothness, or by the reader who feels that rigor leads only to *rigor mortis*. However, the reader who is prepared to accept that all his random fields satisfy regularity requirements, without checking, will do well to remember the principle of *caveat emptor*.

We start with the issue of continuity. Rather interestingly, and importantly, this turns out to be a problem which can be studied, at least for Gaussian processes, just as easily over very general parameter spaces as over $\mathbb{R}^N$, as long as they have metrics defined on them[33]. First of all, however, it is important to recall that there are several notions of continuity for random processes over a parameter space $T$, among them

*Continuity in probability:*

$$\lim_{s\to t} \mathbb{P}\left\{|f(t) - f(s)| \geq \varepsilon\right\} = 0, \quad \text{for each } t \in T \text{ and each } \varepsilon > 0.$$

*Continuity in mean square, or $L^2$ continuity:*

$$\lim_{s\to t} \mathbb{E}\left\{|f(t) - f(s)|^2\right\} = 0, \quad \text{for each } t \in T.$$

*Continuity with probability one, sample path, or almost sure (a.s.), continuity:*

$$\mathbb{P}\left\{\lim_{s\to t} |f(t) - f(s)| = 0, \text{ for all } t \in T\right\} = 1.$$

[33] In fact, not even a metric is required. It suffices that $T$ is metrizable, and one can then study everything in terms of the canonical metric (2.5.1). That the main questions of continuity are actually independent of the choice of metric is proven, for example, in Section 1.3 of *RFG*.

That these can be quite different is shown in Exercise 2.8.16. We are interested in the strongest of these, continuity with probability one.

### 2.5.1 The General Gaussian Theory

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

$$\mathbb{P}\left\{\sup_{t\in T}|f(t)|<\infty\right\}=1 \quad \text{or} \quad \mathbb{P}\left\{\lim_{s\to t}|f(t)-f(s)|=0, \text{ for all } t\in T\right\}=1.$$

To start, define a metric $d$ on $T$, known as the *canonical metric for $T$ induced by the field $f$*, by setting

$$d(s,t)\stackrel{\Delta}{=}\left\{\mathbb{E}\left[(f(s)-f(t))^2\right]\right\}^{\frac{1}{2}}, \qquad s,t\in T, \tag{2.5.1}$$

in a notation that will henceforth remain fixed[34].

The next thing we need is notation for the ball, of radius $\varepsilon$, in the canonical metric, centered at a point $t\in T$, which we denote by

$$B_d(t,\varepsilon)\stackrel{\Delta}{=}\{s\in T:\, d(s,t)\le\varepsilon\}. \tag{2.5.2}$$

It is important to realise that even if the underlying parameter space is as simple as $\mathbb{R}^2$ these 'balls' need not be 'round'. For example, consider the four balls in Figure 2.5.1, which are all assumed to have the same radius in the canonical metric. Balls $a$ and $b$ are physically (in a Euclidean sense) round, which indicates that in the regions in which they are placed the canonical distance between two points is proportional to the Euclidean distance. However, the constant of proportionality is different in the two regions. In both cases it makes sense to talk about some sort of local isotropy, at least as seen through the canonical metric. Of course, since this metric only takes into account second moments, such local isotropy is in the limited sense of Section 2.4.9.

To understand what this means about the random process, note from (2.5.1) that for, $s,t\in T$, the $d$-distance between any two points is the standard deviation of the difference $f(t)-f(s)$. Thus, in regions where this standard deviation is large, we expect the process to move around comparatively rapidly, and, as a consequence, the 'physical' balls of fixed $d$-radius will be smaller. Thus while both $a$ and $b$ are regions of local isotropy, we would expect $f$ to behave far more erratically in the former.

[34] 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 imply $s=t$. To see why, 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 integer $k$.

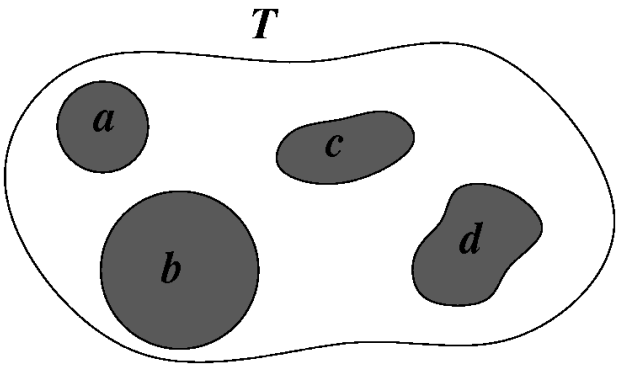


**Fig. 2.5.1.** Balls in the canonical metric

In regions $c$ and $d$ we have also lost local isotropy, and so the balls, albeit of the same $d$-radius, look quite different.

Once this is understood, it is not a big jump to realise that the number of balls needed to cover $T$ must, in some way, measure the behavior of $f$ over $T$, and this brings us to an important definition.

**Definition 2.5.1.** *Let $f$ be a centered Gaussian field on $T$, and $d$ the canonical metric (2.5.1). Assume that $T$ is $d$-compact, in that*

$$\mathrm{diam}(T) \stackrel{\Delta}{=} \sup_{s,t\in T} d(s,t) < \infty. \tag{2.5.3}$$

*Fix $\varepsilon > 0$ and let $N(T,d,\varepsilon) \equiv N(\varepsilon)$ denote the smallest number of $d$-balls of radius $\varepsilon$ whose union covers $T$. Set*

$$H(T,d,\varepsilon) \equiv H(\varepsilon) = \ln\left(N(\varepsilon)\right). \tag{2.5.4}$$

*Then $N$ and $H$ are called the (metric) entropy and log-entropy functions 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 $\mathrm{diam}(T)$ is finite, it follows that $H(\varepsilon) < \infty$ for all $\varepsilon > 0$. On the other hand, the same need not be (nor generally is) true for $\lim_{\varepsilon\to 0} H(\varepsilon)$, since as $\varepsilon \to 0$ the $d$-balls get smaller and smaller, and we need more and more of them to cover $T$. It is precisely the growth rate of $H$ at zero that captures the 'smoothness of the covariance at the origin' that we described in the setting of stationary fields at the beginning of this section.

Here then is the main result about Gaussian continuity and boundedness, due originally, more or less in the form given below, to Richard Dudley [30, 31]. However this result has a long and rich history, and, as we shall describe briefly below, is far from being the last word on the subject.

**Theorem 2.5.2.** *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*

$$\mathbb{E}\left\{\sup_{t\in T} f_t\right\} \leq K \int_0^{\operatorname{diam}(T)/2} H^{1/2}(\varepsilon)\, d\varepsilon, \tag{2.5.5}$$

*and*

$$\mathbb{E}\left\{\omega_{f,d}(\delta)\right\} \leq K \int_0^{\delta} H^{1/2}(\varepsilon)\, d\varepsilon, \tag{2.5.6}$$

*where*

$$\omega_f(\delta) \equiv \omega_{f,d}(\delta) \stackrel{\Delta}{=} \sup_{d(s,t)\leq\delta} |f(t) - f(s)|\,, \quad \delta > 0, \tag{2.5.7}$$

*is the modulus of continuity of $f$ on $T$ with respect to the canonical metric $d$. Furthermore, there exists a random $\eta \in (0,\infty)$ and a universal constant $K$ such that*

$$\omega_{f,d}(\delta) \ \leq K \int_0^{\delta} H^{1/2}(\varepsilon)\, d\varepsilon, \tag{2.5.8}$$

*for all $\delta < \eta$.*

We are not going to attempt to prove this general result here, since you can find a full proof in Section 1.3 of *RFG*. Furthermore, the proof, while not difficult, is a little too technical for our current purposes. Nevertheless, we encourage you to go to *RFG* for the details.

Reading Theorem 2.5.2 carefully, you will note that there is no claim that the sufficient conditions given there are also necessary and, indeed, they generally are not[35]. However, when $f$ is a stationary, centered, Gaussian process there is such a result, and then it can be shown that

$$\begin{aligned} f \text{ is a.s. continuous on } T &\iff f \text{ is a.s. bounded on } T \\ &\iff \int_0^{\infty} H^{1/2}(\varepsilon)\, d\varepsilon \ < \ \infty. \end{aligned} \tag{2.5.9}$$

We shall not go into further details for the general case, but now do want to see how to apply these results to Gaussian random fields on $\mathbb{R}^N$.

[35] There is a theory which gives necessary and sufficient conditions for boundedness and continuity of general Gaussian processes, based on the ideas of *majorising measures*. For details, see, for example, [53, 90].

### 2.5.2 Gaussian 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. Although we did not actually give detailed proofs here, 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 basic results for Gaussian processes on $\mathbb{R}^N$ are easy to state without specifically referring to any abstract notions. We shall do this first for continuity, and then for differentiability. In both, we assume that $f$ is a centered Gaussian process with continuous covariance function $C$ defined on a compact $T \subset \mathbb{R}^N$.

#### Continuity

We start by defining a function which captures the size of increments, by setting

$$\begin{aligned} p^2(u) &\stackrel{\Delta}{=} \sup_{|s-t|\le u} \mathbb{E}\left\{|f_s - f_t|^2\right\} \\ &= \sup_{|s-t|\le u} \big[C(t,t) + C(s,s) - 2C(s,t)\big]. \end{aligned} \tag{2.5.10}$$

If $f$ is stationary, then

$$p^2(u) \;=\; 2 \sup_{|t|\le u} [C(0) - C(t)], \tag{2.5.11}$$

and so one sees that the rate of convergence on $p$ to zero as $u \to 0$ is closely related to the smoothness of $C$ at the origin discussed above. Since $p$ is nondecreasing, there is no problem in defining

$$p^{-1}(\delta) \;\stackrel{\Delta}{=}\; \sup\{u : p(u) \le \delta\}, \qquad \delta > 0.$$

Here is the main result on continuity of Gaussian random fields on $\mathbb{R}^N$.

**Theorem 2.5.3.** *If, for some $\delta > 0$, either*

$$\int_0^\delta (-\ln u)^{\frac{1}{2}}\, dp(u) < \infty \quad \text{or} \quad \int_\delta^\infty p\left(e^{-u^2}\right)\, du < \infty, \tag{2.5.12}$$

*then $f$ is continuous and bounded on $T$ with probability one. A sufficient condition for either integral in* (2.5.12) *to be finite is that, for some $0 < K < \infty$ and $\alpha, \eta > 0$,*

$$\mathbb{E}\left\{|f_s - f_t|^2\right\} = C(t,t) + C(s,s) - 2C(s,t) \leq \frac{K}{|\ln|s-t|\,|^{1+\alpha}}, \tag{2.5.13}$$

*for all $s, t$ with $|s-t| < \eta$. Furthermore, there exists a constant $K'$, dependent only on the dimension $N$, and a random $\delta_0 > 0$, such that, for all $\delta < \delta_0$,*

$$\omega_f(\delta) \leq K' \int_0^{p^{-1}(\delta)} (-\ln u)^{\frac{1}{2}}\, dp(u), \tag{2.5.14}$$

*where the modulus of continuity $\omega_f$ is as in* (2.5.7)*, but taken with respect to the usual Euclidean metric rather than the canonical one. A similar bound, in the spirit of* (2.5.6)*, holds for $\mathbb{E}\{\omega_f(\delta)\}$.*

**Proof.** For once, we shall actually prove something in detail, the motivation being that it is worthwhile seeing how to convert the abstract entropy conditions to something more concrete.

Note first that since $p(u)$ is obviously non-decreasing in $u$, the Riemann-Stieljes integral (2.5.12) is well defined. The proof that both integrals in (2.5.12) converge and diverge together and that the convergence of both is assured by (2.5.13) is simple calculus and is Exercise 2.8.19. Of more significance is relating these integrals to the entropy integrals of Theorem 2.5.2. Indeed, all the claims of the theorem regarding the modulus of continuity $\omega_f$ in (2.5.14) will follow from these results if we show that

$$\int_0^{\delta} H^{1/2}(\varepsilon)\, d\varepsilon \;\leq\; K \int_0^{p^{-1}(\delta)} (-\ln u)^{\frac{1}{2}}\, dp(u) \tag{2.5.15}$$

for small enough $\delta$.

Since $T$ is compact, we can enclose it in a $N$-cube $C_L$ of side length $L = \max_{i=1,\ldots,N} \sup_{s,t\in T} |t_i - s_i|$, where $s = (s_1, \ldots, s_N)$ and $t = (t_1, \ldots, t_N)$. 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,

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

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

The various sufficient conditions for continuity of Theorem 2.5.3 are quite sharp, but not necessary. There are two stages at which necessity is lost. One is simply that, as we mentioned earlier, 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 $p$. For an example of the latter, see Exercise 2.8.18.

Despite these drawbacks, the results of Theorem 2.5.3 are, from a practical point of view, reasonably definitive. For example, if $f$ is stationary, then, following on from (2.5.9), it is possible to check that if

$$\frac{K_1}{(-\ln|t|)^{1+\alpha_1}} \;\le\; C(0)-C(t) \;\le\; \frac{K_2}{(-\ln|t|)^{1+\alpha_2}}, \tag{2.5.16}$$

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

In practical situations, it is rare indeed that one even gets close to the logarithmic behavior of (2.5.13) or (2.5.16). The more common situation in the applications considered in this book is that the covariance function has a power series representation of the form

$$C(s,t) = C(t,t) - (t-s)\Lambda_t(t-s)' + o\left(|t-s|^{2+\delta}\right), \tag{2.5.17}$$

for $|t-s|$ small and some $\delta > 0$, or, in the stationary case

$$C(t) = C(0) - t\Lambda t' + o\left(|t|^{2+\delta}\right), \tag{2.5.18}$$

for $t$ in the neighborhood of the origin. The matrices $\Lambda_t$ and $\Lambda$ are $N \times N$ and positive definite. In the stationary case, the elements of $\Lambda$ are the second order spectral moments (cf. (2.4.4)). In each of these cases the upper bound of (2.5.13) or (2.5.16) holds, with room to spare.

The condition (2.5.18) on the covariance function can easily be translated to spectral terms. It is standard Tauberian theory, which translates the behaviour of $C$ at the origin to that of the spectral measure $\nu$ at infinity, to see that if the integral

$$\int_{\mathbb{R}^N} \left(\log(1+|\lambda|)\right)^{1+\alpha} \nu(d\lambda) \tag{2.5.19}$$

converges for some $\alpha > 0$ then $f$ is continuous, while if it diverges for some $\alpha < 0$ then $f$ is discontinuous. (cf. Exercise 2.8.20.)

In other words, it is the behavior of the 'high frequency oscillations' in the spectral representation that are controlling the continuity/discontinuity dichotomy. This, of course, is what we suggested, on a purely heuristic basis, at the beginning of this section.

### Differentiability

The step beyond continuity for a function on $\mathbb{R}^N$ is differentiability, which we investigate now, again in the Gaussian setting.

As for continuity, there are various notions of differentiability. We have already seen one, in the mean square setting of (2.4.6). Now, however, we would like to know when the limits in (2.4.6) not only exist, for all points $t \in T$ and for all directions $t'$, with probability one, but when they are the same as what one gets by differentiating the sample paths of $f$ in the usual sense. Furthermore, we would like to know when these derivatives are continuous.

It turns out that, at least in the Gaussian scenario, this question can be handled within the framework of basic continuity since derivatives, if they exist, must still be Gaussian. To see how to do this, first endow the space $\mathbb{R}^N \times \otimes^k \mathbb{R}^N$ with the norm

$$\|(s,s')\|_{N,k} \stackrel{\Delta}{=} |s| + \|s'\|_{\otimes^k \mathbb{R}^N} = |s| + \left( \sum_{i=1}^{k} |s_i'|^2 \right)^{1/2},$$

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

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

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

**Theorem 2.5.4.** *Suppose $f$ is a centered Gaussian random field on an open $T \in \mathbb{R}^N$, possessing $k$-th order partial derivatives in the $L^2$ sense in all directions everywhere inside $T$. Suppose, furthermore, that there exists $0 < K < \infty$, and $\rho, \delta, h_0 > 0$ such that for $0 < \eta_1, \eta_2, h < h_0$,*

$$\mathbb{E}\left\{ \left[ \eta_1^{-k} \Delta^k f(t, \eta_1 t') - \eta_2^{-k} \Delta^k f(s, \eta_2 s') \right]^2 \right\} \tag{2.5.20}$$
$$< K \left| \ln \left( \|(t,t') - (s,s')\|_{N,k} + |\eta_1 - \eta_2| \right) \right|^{-(1+\delta)},$$

*for all*

$$((t,t'),(s,s')) \in T_{k,\rho} \times T_{k,\rho} : (s,s') \in B_{N,k}((t,t'),h),$$

*where $\Delta^k f(t,t')$ is the symmetrized difference*

$$\Delta^k f(t,t') = \sum_{s \in \{0,1\}^k} (-1)^{k - \sum_{i=1}^k s_i} f\left( t + \sum_{i=1}^{k} s_i t_i' \right)$$

*Then, with probability one, $f$ is $k$ times continuously differentiable.*

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

$$\widehat{f}(t,t',\eta) = \begin{cases} \eta^{-k}\Delta^k f(t,\eta t') & \eta \neq 0, \\ D^k_{L^2} f(t,t') & \eta = 0, \end{cases}$$

where $D^k_{L^2} f$ is the mean square derivative (2.4.6). This process is defined on the parameter space $\widehat{T} \stackrel{\Delta}{=} T_{k,\rho} \times (-h,h)$, an open subset of the finite dimensional vector space $\mathbb{R}^N \times \otimes^k \mathbb{R}^N \times \mathbb{R}$, with norm

$$\|(t,t',\eta)\|_{N,k,1} \;=\; \|(t,t')\|_{N,k} + |\eta|.$$

Whether or not $f$ is $k$ times differentiable on $T$ is clearly the same issue as whether or not $\widehat{f}$ is continuous in $\widehat{T}$, with the issue of the continuity of $\widehat{f}$ really being only on the hyperplane where $\eta = 0$. But this puts us back into the setting of Theorem 2.5.3, and it is easy to check that condition (2.5.13) there translates to (2.5.20) in the current scenario. □

The left hand side of (2.5.20) can, obviously, be written in terms of the covariance function $C$, although it becomes rather messy rather quickly. Consider what is perhaps the simplest of cases, in which $k = 1$, $t = s$, and $t' = s'$, so that we are looking at the simple, first order derivative of a random process $f$ on the real line. Then (2.5.20) becomes

$$C(t+h,t+h) - C(t+h,t) - C(t,t+h) + C(t,t) \leq \frac{K}{|\ln(|h|)|^{1+\delta}},$$

for all $h \in \mathbb{R}^1$ with $|h|$ sufficently small.

If we want to ensure that all first order partial derivatives of a random field on $\mathbb{R}^N$ exist and are continuous, then precisely the same condition suffices, with the understanding that now $t$ and $h$ are both in $\mathbb{R}^N$.

Second order partial derivatives of $f$ of general order $k$ require similar bounds, but the 'difference' of $C$ on the left hand side will always be a difference of order $2k$.

As for continuity, it is rare in practice to get close to the upper bound in (2.5.20), and this condition will easily be satisfied if, in analogy to (2.5.17) and (2.5.18), the covariance function has a Taylor series expansion of up to order $2k$ with a remainder of $o(|h|^{2k+\eta})$ for some $\eta > 0$. There is also a corresponding spectral result, which you are asked to prove in Exercise 2.8.20.

### 2.5.3 Non-Gaussian Processes

A natural question to ask is whether or not the results and methods relating to continuity and differentiabilty that we have seen so far only for Gaussian random fields extend naturally to the non-Gaussian scenario.

In fact, for most of the processes that will concern us, this will not be terribly relevant, since we plan to concentrate on Gaussian related fields which can be written in the form

$$f(t) \;=\; F\left(g^1(t), \ldots, g^d(t)\right),$$

where the $g^i$ are i.i.d. Gaussian and $F : \mathbb{R}^d \to \mathbb{R}$ 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, With these in mind, and for completeness, we state the following result, in which $f_t$ is a random field on a parameter space $T$ with a metric $\tau$. (This is usually something corresponding to the canonical metric $d$ of the Gaussian case.) Recall that a function $\varphi : \mathbb{R} \to \mathbb{R}$ is called a *Young function* if it is even, continuous, convex, and satisfies

$$\lim_{x\to 0} \frac{\varphi(x)}{x} = 0 \quad \text{and} \quad \lim_{x\to\infty} \frac{\varphi(x)}{x} = \infty.$$

A simple example of a Young function is given by $|x|^{1+\alpha}$, $\alpha > 0$.

**Theorem 2.5.5.** *Take $f$ as above and let $N_\tau$ be the metric entropy function for $T$ with respect to the metric $\tau$. If there exist an $\alpha \in (0, 1]$ and a Young function $\varphi$ such that*

$$\mathbb{E}\left\{\varphi\left(\frac{\|f(t) - f(s)\|_B^\alpha}{\tau(s,t)}\right)\right\} \le 1 \quad \textit{and} \quad \int_{N_\tau(u)>1} \varphi^{-1}\left(N_\tau(u)\right)\, du < \infty,$$

*then $f$ is continuous with probability one.*

The best place to read about this is in Ledoux and Talagrand [53].

Note that, in the Gaussian case, once we had solved the issue of continuity, continuous differentiability was in essence a corollary. This will not be true in general, since then derivatives need not belong to the same class of processes as the original one.

## 2.6 Gaussian Exceedence Probabilities

If you have read this far, you know that one of the things that we care about most are the exceedence probabilities

$$\mathbb{P}\left\{\sup_{t\in T} f_t > u\right\} \tag{2.6.1}$$

for Gaussian random fields.

There are at least four ways to approach these probabilities.

(i) Choose a particular random field, and use its special properties to calculate the exceedence probabilities from first principles. This is what we did for the cosine process and field in Section 2.3.4. However, we already stated there, that, as far as differentiable processes are concerned, these are the only cases for which this direct approach works.

(ii) Search for general inequalities, that will always work. After all, basic statistics has its Chebychev inequality, martingale theory has its maximal inequalities, Markov processes have large deviations, and so on, so surely there should be something for Gaussian processes. In fact, there is, and we shall meet it in Theorem 2.6.1 below. It was discovered independently, and established with very different proofs, by Borell [22] and Tsirelson, Ibragimov and Sudakov (TIS) [94]. For brevity, we shall call it the Borell-TIS inequality. However, like all blanket inequalities, while it provides an excellent tool for probabilists who wish to prove theorems, it is a poor tool for statisticians who need to see numbers.

(iii) Use comparison techniques. We do have exact results for the cosine processes. Perhaps these could be used as a basis for comparison. It seems reasonable, for example, that, other things (like means and variances) being equal, a random field with a weak correlation structure should have higher exceedence probabilities than a tightly correlated one. To a certain extent this is true, and the basic results quantifying this are Slepian's inequality, Theorem 2.6.3, and its extensions, below.

(iv) Finally, we could compromise. Rather than assuming everything about the process, as in (i), or nothing, as in (ii), we could assume perhaps a little more than we would like to, and obtain approximations to exceedence probabilities which, while not perfect, yield numbers that can be used by practicing statisticians.

In fact, (iv) is the path that we shall concentrate on later. However, since (ii) and (iii) are important, and often useful, we shall first invest a little time saying something about them, along with some of their extensions. We shall not give proofs. They can all be found in *RFG*.

### 2.6.1 Borell-TIS Inequality

In a notation that will remain fixed for the rest of the book, for any random field $f$ on any parameter space $T$, set

$$\sigma_T^2 \stackrel{\Delta}{=} \sup_{t\in T} \mathbb{E}\{f_t^2\}.$$

Then the Borell-TIS inequality[36] is

**Theorem 2.6.1 (Borell-TIS inequality).** *Let $f_t$ be a centered Gaussian process, a.s. bounded on $T$*[37]*. Write*

[36] Actually, Theorem 2.6.1 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.

[37] The parameter space $T$ in the Borell-TIS inequality is completely general, the only requirement being that it be compact with respect to the canonical metric $d$, in the sense of (2.5.3) .

$$\|f\| = \|f\|_T \stackrel{\Delta}{=} \sup_{t\in T} f_t.$$

*Then*

$$\mathbb{E}\{\|f\|\} \;<\; \infty,$$

*and, for all* $u > 0$,

$$\mathbb{P}\left\{\|f\| - \mathbb{E}\{\|f\|\} \;>\; u\right\} \leq e^{-u^2/2\sigma_T^2}. \tag{2.6.2}$$

Note that, despite the misleading notation, $\|\ \| \equiv \sup$ 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

$$\mathbb{P}\left\{\sup_t |f_t| \;>\; u\right\} \;\leq\; 2\,\mathbb{P}\left\{\sup_t f_t \;>\; u\right\}. \tag{2.6.3}$$

To obtain an appreciation of how tight this very general inequality is, note that it immediately implies that

$$\mathbb{P}\{\|f\|>u\} \;\leq\; e^{\mu_u - u^2/2\sigma_T^2}, \tag{2.6.4}$$

where $\mu_u = (2u\mathbb{E}\{\|f\|\} - [\mathbb{E}\{\|f\|\}]^2)/2\sigma_T^2$, which tells us that, for high levels $u$, the dominant behavior of *all* Gaussian exceedence probabilities is determined by $e^{-u^2/2\sigma_T^2}$.

This is somewhat surprising, for if $X$ is a single Gaussian variable with distribution $N(0, \sigma_T^2)$ then we already know (cf. (2.2.2)) that, for all $u > 0$,

$$\left(\frac{\sigma_T}{\sqrt{2\pi}u} - \frac{\sigma_T^3}{\sqrt{2\pi}u^3}\right) e^{-\frac{1}{2}u^2/\sigma_T^2} \;\leq\; \mathbb{P}\{X > u\} \;\leq\; \frac{\sigma_T}{\sqrt{2\pi}u}\, e^{-\frac{1}{2}u^2/\sigma_T^2}.$$

In other words, at high levels, the exceedence probability of a Gaussian random field is not that different from the exceedence probability of the field at the point of maximal variance.

For a stationary field, where the maximal variance is achieved at every point in $T$, this strengthens to the claim that, still at high levels, the exceedence probability of a Gaussian random field is not that different from the exceedence probability of the field *at any given point*.

Does this mean that we need go no further in studying Gaussian extrema? Not at all, for a number of reasons:

(i) First of all, while the exponent $\mu_u$ in (2.6.4) grows only linearly in $u$, $e^{\mu_u}$ grows much faster. It can therefore hardly be ignored, other than perhaps by a probabilist in the throes of a proof where "large $u$" is so large that linear terms are irrelevent when compared to quadratic, regardless of whatever constants may be around. Statisticians, however, will rarely feel this way.

(ii) If, in view of (i), we decide to keep track of the term $e^{\mu_u}$, we find it involves the expectation of the supremum, and this is hard to obtain. While it is true that under entropy conditions we did manage to find an upper bound for it (cf. Theorem 2.5.2) this bound involved an admittedly universal, but essentially unknown[38], constant.

(iii) As we shall see later, for most smooth random fields the factor of $e^{\mu_u}$, which is generally $O(e^{cu})$, can be replaced for a factor of the form $Cu^{\alpha}$, where both $C$ and $\alpha$ are explicitly computable. Furthermore, it can be shown that this is the correct order of growth. This is where expected Euler characteristics will come into their own, and justify the claims that we made back in Chapter 1 (cf. (1.5.12)).

As an example of how easy it is to improve on the Borell-TIS bound in the spirit of (ii), when a little more is assumed, consider the following, definitely sub-optimal, result.

**Theorem 2.6.2.** *Let be $f$ a centered, a.s. continuous Gaussian field over $T$ with entropy function $N$. If $N(\varepsilon) \le K\varepsilon^{-\alpha}$, then, for all sufficiently large $u$,*

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \ge u\right\} \le C_\alpha\, u^{\alpha+\eta}\, e^{-u^2/2\sigma_T^2}, \tag{2.6.5}$$

*for every $\eta > 0$, where $C_\alpha = C(K, \alpha, \sigma_T^2)$ is a finite constant.*

**Proof.** Take $\varepsilon > 0$ and define

$$\mu(t,\varepsilon) = \mathbb{E}\left\{\sup_{s\in B_d(t,\varepsilon)} f_s\right\},$$

and

$$\mu(\varepsilon) = \sup_{t\in T} \mu(t,\varepsilon),$$

where $B_d(t,\varepsilon)$ is a ball of radius $\varepsilon$ around $t$ in the canonical metric $d$ of (2.5.1). Since $N(\varepsilon)$ balls of radius $\varepsilon$ cover $T$, it is an immediate consequence of the Borell-TIS inequality that, for $u > \mu(\varepsilon)$,

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \ge u\right\} \le N(\varepsilon)\, e^{-\frac{1}{2}(u-\mu(\varepsilon))^2/\sigma_T^2}. \tag{2.6.6}$$

Allowing $C = C(\alpha)$ to denote a constant, dependent only on $\alpha$, that may change from line to line, we have from Theorem 2.5.2 that

[38] Actually, there are known values for this constant. However, they are so large as to yield bounds that are effectively valueless from the point of view of generating useful numerical bounds to exceedence probabilities.

$$\mu(t,\varepsilon) \le C \int_0^\varepsilon \left(\ln(N(\varepsilon))\right)^{\frac{1}{2}} d\varepsilon \le \; C \int_0^\varepsilon \left(\ln K + \alpha \ln(1/\varepsilon)\right)^{\frac{1}{2}} d\varepsilon,$$

so that, for small enough $\varepsilon$,

$$\mu(t,\varepsilon) \le C \int_0^\varepsilon \left(\ln(1/\varepsilon)\right)^{\frac{1}{2}} d\varepsilon \le C\,\varepsilon\sqrt{\ln(1/\varepsilon)}.$$

Set $\varepsilon = \varepsilon(u) = u^{-1}$, choose $u$ large enough so that $u > Cu^{-1}\sqrt{\ln u}$ and substitute into (2.6.6) to obtain

$$\begin{aligned}\mathbb{P}\left\{\sup_{t\in T} f(t) \ge u\right\} &\le C_1\, u^\alpha\, e^{-\frac{1}{2}\left(u - C_2 u^{-1}\sqrt{\ln u}\,\right)^2/\sigma_T^2} \\ &\le C_3\, u^\alpha\, e^{C_4\sqrt{\ln u}} e^{-u^2/2\sigma_T^2}.\end{aligned}$$

Since for $\eta > 0$ and $u$ large enough $e^{C\sqrt{\ln u}} < u^\eta$, this gives us (2.6.5) and so completes the proof. □

One can do much better than Theorem 2.6.2, by assuming a little more on the entropy function, or by working a little harder. However, the basic idea should be clear from this one simple result and its proof. In particular, in the situations in which we shall generally be working, of smooth, twice differentiable functions, power law behavior for the entropy function is always assured (cf. the proof of Theorem 2.5.3). Thus, in those cases, we can expect bounds on exceedence probabilities which are much better than those given by the Borell-TIS inequality. On the other hand, it is going to be difficult to get bounds with useful constants using only entropy arguments.

### 2.6.2 Comparison Inequalities

Having now seen that it is not going to be easy to get good, quantitative estimates for exceedence probabilities, we can now investigate the second path, that of using comparison with known cases.

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 generally be 'larger' than $\|g\|$", where, as in the previous section, we write $\|\cdot\|$ for supremum. The most basic of these results is Slepian's inequality, which, like the Borell-TIS inequality, holds for all parameter spaces $T$.

**Theorem 2.6.3 (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*

$$\mathbb{E}\left\{(f_t - f_s)^2\right\} \le \mathbb{E}\left\{(g_t - g_s)^2\right\}, \tag{2.6.7}$$

*for all $s, t \in T$, then for all real $u$*

$$\mathbb{P}\{\|f\| > u\} \leq \mathbb{P}\{\|g\| > u\}. \tag{2.6.8}$$

*Furthermore,*

$$\mathbb{E}\{\|f\|\} \leq \mathbb{E}\{\|g\|\}. \tag{2.6.9}$$

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 (2.6.8) does *not* follow from (2.6.7) if we replace $\sup_T f_t$ by $\sup_T |f_t|$ and $\sup_T g_t$ by $\sup_T |g_t|$ (cf. Exercise 2.8.21).

Furthermore, the proof of Slepian's inequality does not use the above 'obvious' heuristic argument at all, but uses little more than basic calculus and some approximation arguments. Since the proof is not long, and the result is important, we shall give a full proof. The proof is based on the following lemma, the proof of which, in all its important details, goes back to Slepian's original paper [85].

**Lemma 2.6.4.** *Let $f_1, ..., f_k$ be centered Gaussian variables with covariance matrix $C = (c_{ij})_{i,j=1}^k$, $c_{ij} = \mathbb{E}\{f_i f_j\}$. Let $h : \mathbb{R}^k \to \mathbb{R}$ be $C^2$, and assume that $h$ and its derivatives are all integrable with respect to the Gaussian density on $\mathbb{R}^k$. Let*

$$H(C) = \mathbb{E}\left\{h(f_1, \ldots, f_k)\right\}, \tag{2.6.10}$$

*and assume that for a pair $(i, j)$, $1 \leq i < j \leq k$*

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} \;\geq\; 0 \tag{2.6.11}$$

*for all $x \in \mathbb{R}^k$. Then $H(C)$ is an increasing function of $c_{ij}$.*

**Proof.** We have to show that

$$\frac{\partial H(C)}{\partial c_{ij}} \;\geq\; 0.$$

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 shows that[39]

$$\frac{\partial \varphi}{\partial c_{ii}} = \tfrac{1}{2}\frac{\partial^2 \varphi}{\partial x_i^2}, \qquad \frac{\partial \varphi}{\partial c_{ij}} = \frac{\partial^2 \varphi}{\partial x_i \partial x_j}, \qquad i \neq j. \tag{2.6.12}$$

Applying this and our assumptions on $h$ to justify two integrations by parts, we obtain, for $i \neq j$,

[39] This is, of course, little more that the heat equation of PDE theory.

$$\frac{\partial H(C)}{\partial c_{ij}} = \int_{\mathbb{R}^k} h(x)\,\frac{\partial \varphi(x)}{\partial c_{ij}}\,dx = \int_{\mathbb{R}^k} \frac{\partial^2 h(x)}{\partial x_i \partial x_j}\,\varphi(x)\,dx \geq 0.$$

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. □

**Proof of Theorem 2.6.3** Actually, we are only going to prove the main inequality, (2.6.8), for $T$ discrete and finite. The extension to general $T$ can be found in *RFG*, and while slightly technical, is not hard.

Note that since $\mathbb{E}\{f_t^2\} = \mathbb{E}\{g_t^2\}$ for all $t \in T$, (2.6.7) 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 Lemma 2.6.4, and $k$ is the number of points in (our finite, discrete) $T$. 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 2.6.4 that

$$\mathbb{E}\left\{\prod_{i=1}^k h_i(f_i)\right\} \geq \mathbb{E}\left\{\prod_{i=1}^k h_i(g_i)\right\}. \tag{2.6.13}$$

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 $(-\infty, u]$, to derive that

$$\mathbb{P}\{\|f\| < u\} \geq \mathbb{P}\{\|g\| < u\},$$

which implies (2.6.8).

To complete the proof, all that remains is to show that (2.6.8) implies (2.6.9). But this is a simple consequence of the tail formula for the mean, 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. □

There are many extensions of Slepian's inequality, the most important of which is probably the following, which we shall not attempt to prove.

**Theorem 2.6.5 (Sudakov-Fernique inequality).** *Let $f$ and $g$ be a.s. bounded Gaussian processes on $T$. If*

$$\mathbb{E}\{f_t\} = \mathbb{E}\{g_t\}$$

*and*

$$\mathbb{E}\left\{(f_t - f_s)^2\right\} \;\leq\; \mathbb{E}\left\{(g_t - g_s)^2\right\}$$

*for all $s, t \in T$, then*

$$\mathbb{E}\{\|f\|\} \leq \mathbb{E}\{\|g\|\}. \tag{2.6.14}$$

In other words, a Slepian-like inequality holds without a need to assume either zero mean or identical variance for the compared processes. However, in this case we have only the weaker ordering of expectations of (2.6.9) and not the stochastic domination of (2.6.8). Alternatively, while keeping the mean of a field constant, one can obtain a higher expected supremum by either increasing the variance or reducing the covariance.

### 2.6.3 Exceedence Probabilities for Smooth Processes

Now that you have had a little taste of the general theory, the time has come tp say something about the cases that will be of central interest to us, those in which the random field has smooth sample paths.

Perhaps the first result of this kind, in the setting of random fields, goes back to the Russian school of Belyaev [18, 19] and treats stationary, zero mean, Gaussian random fields on $\mathbb{R}^N$ with covariance functions which, near the origin, can be written as

$$C(t) = 1 - t\Lambda t' + o(|t|^2), \tag{2.6.15}$$

where $\Lambda$ is the matrix of second order spectral moments (cf. (2.4.4)) and, for convenience, we have also assumed unit variance. We already know that these random fields are both continuous with probability one, and differentiable in mean square[40].

In this case, one can show[41] that, for $N$-dimensional rectangles $T = \prod_{i=1}^{N}[0, T_i]$,

[40] If the $o(|t|^2)$ in (2.6.15) is small enough ($o(|t|^{2+\delta})$ will do) then they are also continuously differentiable, with probability one. See Theorem 2.5.4.

[41] You can learn a lot more about results of this kind, and how to prove them, in a variety of places. The book by Leadbetter, Lindgren and Rootzén [51] treats mainly processes on the real line, but is very readable. Piterbarg's monograph [75] is harder going, but treats random fields on $\mathbb{R}^N$ as well as processes on the line. There you will also find a detailed treatment of the so-called 'double-sum method' which can be used to compute results like (2.6.16) for a far wider collection of random fields.

$$\lim_{u\to\infty} \frac{\mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\}}{u^N \Psi(u)} = \frac{|T|\,|\Lambda|^{1/2}}{(2\pi)^{N/2}}, \tag{2.6.16}$$

where $|T| = \prod T_i$ is the volume of $T$, but $|\Lambda|$ is the determinant of $\Lambda$.

In view of what we know about the tail probabilities $\Psi(u)$ (cf. (2.2.2)) another way to write this result would be

$$\mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\} = u^{N-1} e^{-u^2/2} \left[\frac{|T||\Lambda|^{1/2}}{(2\pi)^{(N+1)/2}} + R(u)\right].$$

The remainder term $R(u)$ tends to 0 as $u \to \infty$, but otherwise (2.6.16) tells us nothing about it.

On the other hand, if one assumes just a little more, that $f$ is *twice* continuously differentiable, it is possible to show that there exist explicitly computable constants $\alpha$, $n$ and $C_j$ such that

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \;\geq\; u\right\} = \Psi(u) + u^\alpha\, e^{-u^2/2} \left[\sum_{j=0}^{n} C_j u^{-j} \;+\; \text{error}\right], \tag{2.6.17}$$

where the error is small for large $u$. Furthermore, one can get a good handle on the error term, which turns out to be much smaller than expected[42].

Indeed, while, for convenience, we have been assuming stationarity above, results like (2.6.17) hold in quite wide generality and also without the assumption that $T$ be a simple rectangle in $\mathbb{R}^N$.

To explain all of this properly, however, needs an excursion into geometry, which is the content of the next chapter, followed by some hard computation, which is the content of Chapter 4. We shall return to extremal problems, and results like (2.6.17), only in Chapter 5.

## 2.7 An Expectation Meta-Theorem: The Rice-Kac Formula

In this section, we quote a basic tool, the Rice-Kac formula, which we shall use later, in Chapter 4, to compute the expectations of certain functionals of the excursion sets of smooth random fields. Consider two vector-valued random fields $f = (f^1, \ldots, f^N)$ and $g = (g^1, \ldots, g^K)$ defined on some compact set $T \subset \mathbb{R}^N$ with non-empty interior. For $B \subset \mathbb{R}^K$, we need formulae for the expectations

$$\mathbb{E}\big\{\#\left\{t \in T :\, f(t) = u,\; g(t) \in B\right\}\big\}. \tag{2.7.1}$$

[42] Since (2.6.17) looks like the beginning of a power series expansion, one would 'expect' the error term to be $o(u^{-(n+1)})$. In fact, it turns out to be $o(e^{-\delta u^2})$, for an identifiable $\delta > 0$. This, of course, is much smaller, and is reminiscent of what we saw for the cosine field at (2.3.16).

Perhaps the most basic application of (2.7.1) is to prove the famous Rice, or Rice-Kac formula[43]. The Rice formula gives an expression for the number of upcrossings of the level $u$ of a real-valued process $h$ on the line where an upcrossing is defined as a point $t$ where $h(t) = u$ and $h(t)$ is increasing. In this example, we set $f = h$, $g = \dot{h}$ and $B = [0, +\infty)$. See Exercise 2.8.23.

However, we shall require these expectations for a number of reasons. For example, in Chapter 3 we shall see how, using Morse theory, critical points of random fields above a level $u$, which are local features, can be used to compute certain global properties of excursion sets. In turn, these can be related to the exceedence probabilities (2.6.1) of $f$ above the level $u$.

To see how (2.7.1) helps, note that that the critical points of a random field $Z$, say, are solutions to the equation $\nabla Z = 0$. If we are to count critical points of $Z$ above the level $u$, this suggests setting $f = \nabla Z$, $g = Z$ and $B = [u, +\infty)$ in (2.7.1). See Exercise 2.8.25 on applying this formula to maxima in one dimension.

We are now ready to quote the theorem. Below, $\nabla f$ denotes the derivative field 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.

$$(\nabla f)(t) \;\equiv\; \nabla f(t) \equiv \left(f_j^i(t)\right)_{i,j=1,\ldots,N} \;\equiv\; \left(\frac{\partial f^i(t)}{\partial t_j}\right)_{i,j=1,\ldots,N}.$$

**Theorem 2.7.1.** *Let $f$, $g$, $T$ and $B$ be as above, with the additional assumption that the boundaries of $T$ and $B$ have finite $N-1$ and $K-1$ dimensional measures, respectively. Furthermore, assume that the following conditions are satisfied for some $u \in \mathbb{R}^N$:*

*(a) All components of $f$, $\nabla f$, and $g$ are a.s. continuous and have finite variances (over $T$).*
*(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 $p_t(x|\nabla f(t), g(t))$ 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 conditional densities $p_t(z|f(t) = x)$ of $g(t)$ given $f(t) = x$, are continuous for all $z$ and for $x$ in a neighbourhood $u$, uniformly in $t \in T$.*
*(f) The following moment condition holds:*

$$\sup_{t\in T} \max_{1\le i,j\le N} \mathbb{E}\left\{\left|f_j^i(t)\right|^N\right\} < \infty. \tag{2.7.2}$$

[43] In fact, the name Rice-Kac formula comes from the fact that the first version of the above formula was used in exactly this way in Kac [45] and Rice [78], following on earlier work by Rice [77] in 1939.

*(g) The moduli of continuity with respect to the usual Euclidean norm (cf.* (2.5.7)*) of each of the components of $f$, $\nabla f$, and $g$ satisfy*

$$\mathbb{P}\{\,\omega(\eta) > \varepsilon\,\} = o\left(\eta^N\right), \qquad \text{as } \eta \downarrow 0, \tag{2.7.3}$$

*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*

$$f(t) = u \in \mathbb{R}^N \quad \text{and} \qquad g(t) \in B \subset \mathbb{R}^K,$$

*and $p_t(x, \nabla y, v)$ denotes the joint density of $(f_t, \nabla f_t, g_t)$, we have, with $D = N^2 + K$,*

$$\mathbb{E}\{N_u\} = \int_T \int_{\mathbb{R}^D} |\det \nabla y|\; \mathbb{1}_B(v)\, p_t(u, \nabla y, v)\, d(\nabla y)\, dv\, dt. \tag{2.7.4}$$

*It is sometimes more convenient to write this as*

$$\mathbb{E}\{N_u\} = \int_T \mathbb{E}\left\{\, |\det \nabla f(t)| \;\middle|\; \mathbb{1}_B(g(t)) \,\middle|\, f(t) = u \right\} p_t(u)\, dt, \tag{2.7.5}$$

*where $p_t$ here is the density of $f(t)$.*

In our applications of Theorem 2.7.1, the expression (2.7.5) will be the principal form used.

For a full proof, see *RFG*, where this result appears as Theorem 11.2.1. An outline of the beginning of this proof, that at least shows from where the result comes, will be given in a moment.

Conditions (a)–(g) are often tedious to check, but almost disappear when both $f$ and $g$ are either Gaussian, or simple functions of Gaussian (vector-valued) Gaussian random fields. In these situations, the primary consideration becomes one of sample path continuity and differentiability, which we already looked at in some detail in Section 2.5.1. Here is the Gaussian result.

**Corollary 2.7.2.** *Let $f$ and $g$ be centered Gaussian fields, and let $T$ and $B$ satisfy the conditions of Theorem 2.7.1. Assume that $f, \nabla f$ and $g$ are all a.s. continuous with finite variances over $T$ and that, for each $t \in T$, the joint distribution of $(f(t), \nabla f(t), g(t))$ is non-degenerate.*

*Write $C_f^i = C_f^i(s,t)$ for the covariance function of $f^i$, $C_{f_j}^i = \partial^2 C_f^i / \partial s_j \partial t_j$ for the covariance function of $f_j^i = \partial f^i / \partial t_j$, and $C_g^i$ for the covariance function of $g^i$. If*

$$\begin{aligned}\max_{i,j}\left|C^i_{f_j}(t,t)+C^i_{f_j}(s,s)-2C^i_{f_j}(s,t)\right| &\le K\,|\ln|t-s|\,|^{-(1+\alpha)}\,,\\ \max_{i}\left|C^i_{g}(t,t)+C^i_{g}(s,s)-2C^i_{g}(s,t)\right| &\le K\,|\ln|t-s|\,|^{-(1+\alpha)}\,,\end{aligned}\tag{2.7.6}$$

*for some finite $K > 0$, some $\alpha > 0$ and all $|t-s|$ small enough, then the conclusions of Theorem 2.7.1 hold.*

To close this Chapter, we look at a partial proof of Theorem 2.7.1. For a start, however, you should try Exercise 2.8.26, which shows that, if $T = [0,1]^N$, then, under conditions (a), (b) and (d) of Theorem 2.7.1, there are, with probability one, no points $t \in \partial T$ satisfying $f(t) = u$. We shall use this and other regularity properties of $f$ and $g$ in what follows. This is the easiest one to prove.

To start, let $\delta_\varepsilon : \mathbb{R}^N \to \mathbb{R}$ be an approximate delta function, constant on the $N$-ball $B(\varepsilon) = \{t \in \mathbb{R}^N : |t| < \varepsilon\}$, zero elsewhere, and normalized so that

$$\int_{B(\varepsilon)} \delta_\varepsilon(t)\,dt \;=\; 1. \tag{2.7.7}$$

We then claim that

$$N_u(f,g;T,B) \;=\; \lim_{\varepsilon\to 0}\int_T \delta_\varepsilon(f(t)-u)\,\mathbb{1}_B(g(t))\,|\det\nabla f(t)\,|\,dt. \tag{2.7.8}$$

If this is true, then, with no further pretense to rigor, take expectations on both sides and freely change the orders of limit and expectation to find that

$$\begin{aligned}\mathbb{E}\{N_u\} &= \lim_{\varepsilon\to 0}\mathbb{E}\int_T \delta_\varepsilon(f(t)-u)\,\mathbb{1}_B(g(t))\,|\det\nabla f(t)\,|\,dt\\ &= \int_T\int_{\mathbb{R}^{N(N+1)/2}}\int_{\mathbb{R}^K} \mathbb{1}_B(v)\,|\det\nabla y|\\ &\qquad\times\left\{\lim_{\varepsilon\to 0}\int_{\mathbb{R}^N}\delta_\varepsilon(x-u)\,p_t(x,\nabla y,v)\,dx\right\}d\nabla y\,dv\,dt,\end{aligned}$$

where the $p_t$ are the obvious densities. Taking the limit in the innermost integral yields

$$\begin{aligned}\mathbb{E}\{N_u\} &= \int_T\int\int \mathbb{1}_B(v)\,|\det\nabla y|\,p_t(u,\nabla y,v)\,d\nabla y\,dv\,dt\\ &= \int_T \mathbb{E}\{|\det\nabla f(t)|\,\mathbb{1}_B(g(t))\,\big|\,f(t)=u\}\,p_t(u)\,dt,\end{aligned}$$

which is what we wanted to show.

Of course, interchanging the order of integration and the limiting procedure requires justification, and not only is it far from trivial, it is in fact so hard to do that a fully rigorous proof requires a rather different approach. See Theorem 11.2.1 of *RFG* for details.

Now, however, we return to the proof of (2.7.8).

To save on notation, and without any loss of generality, we take $u = 0$. Consider those $t \in T$ for which $f(t) = 0$, of which we claim (without proof) that there is only a finite number. Furthermore, by Exercise 2.8.26, none of them lie in $\partial T$. Consequently, each one can be surrounded by an open ball, of radius $\eta$, say, in such a way that the balls neither overlap nor intersect $\partial T$. Furthermore (again as an unproven consequence of the assumptions) we can take $\eta$ small enough so that within each ball $g(t)$ always lies in either $B$ or the interior of its complement, but never both.

Let $\sigma(\varepsilon)$ be the ball $|f| < \varepsilon$ in the image space of $f$. From what we have just claimed follows the fact that we can also choose $\varepsilon$ small enough for the inverse image of $\sigma(\varepsilon)$ in $T$ to be contained within the union of the $\eta$ spheres.

Furthermore, by the inverse mapping theorem 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 $\varepsilon \to 0$ to obtain (2.7.8), as required.

## 2.8 Exercises

**Exercise 2.8.1.** Prove the basic, but very important, Gaussian inequality (2.2.2), that

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

for all $x > 0$.

**Exercise 2.8.2.** Show that the conditional distributions of multivariate Gaussian variables, are also Gaussian, with mean vectors given by (2.2.6) and covariance matrices given by (2.2.7).
Hint: Take

$$A = \begin{pmatrix} I_n & -C_{11}^{-1}C_{12} \\ 0 & I_{d-n} \end{pmatrix}$$

in (2.2.5) and define $Y = (Y^1, Y^2) = XA$, where $Y^1$ has length $n$. Check that $Y^1 \equiv X^1$ and $Y^2$ are independent and use this to obtain (2.2.6) and (2.2.7) for $i = 2,\ j = 1$.

**Exercise 2.8.3.** Prove the Wick product formula for Gaussian random variables, which says the following:

Let $X_1, X_2, \ldots, X_n$ be a set of real-valued random variables having a joint Gaussian distribution and zero means. Then, for any integer $m$,

$$\mathbb{E}\{X_1 X_2 \cdots X_{2m+1}\} = 0,$$
$$\mathbb{E}\{X_1 X_2 \cdots X_{2m}\} = \sum \mathbb{E}\{X_{i_1} X_{i_2}\} \cdots \mathbb{E}\{X_{i_{2m-1}} X_{i_{2m}}\},$$

where the sum is taken over the $(2m)!\,/m!\,2^m$ different ways of grouping $X_1, \ldots, X_{2m}$ into $m$ pairs.
Hint: Use characteristic functions.

**Exercise 2.8.4.** Using nothing but the definition (2.2.9) of the covariance function,

(i) Show that covariance functions of random processes must always be non-negative definite. (cf. Footnote 4 for a definition of non-negative definiteness.)
(ii) Find a simple non-degeneracy condition on a random process for its covariance function to be positive definite.

**Exercise 2.8.5.** Compute the covariance function of the cosine random field (2.3.11), and show that it is both stationary and isotropic.

**Exercise 2.8.6.** Let $f$ be the cosine field of (2.3.11) with the $\xi_k$ and $\xi'_k$ all independent $N(0, \sigma^2)$.

(i) Following the argument described in Section 2.3.4, derive the density (2.3.12) for the supremum, identifying the $C_{nk}$.
Hint: Use characteristic functions to handle the convolution.
(ii) Using (2.3.12) (and quite a lot of calculus) establish (2.3.16).

**Exercise 2.8.7.** Prove the "if" part of the spectral distribution theorem, Theorem 2.4.1. That is, prove that $C(t)$ in (2.4.2) is a non-negative definite function.

**Exercise 2.8.8.** In the statement of the spectral distribution theorem, Theorem 2.4.1, there is an implicit assumption that every non-negative definite function is also a covariance function. Show that this is true by showing that there exists a field with that covariance function.

**Exercise 2.8.9.** Prove that (2.4.8) and (2.4.9) hold, under condition (2.4.7).
Hint: It is easiest, but not necessary, to use both the spectral distribution theorem *and* the spectral representation theorem.
Warning: Don't forget that $i^2 = -1$ if you want to get the power of $-1$ correct!

**Exercise 2.8.10.** Let $f$ be a stationary, $C^2$, random field, $\nabla f$ the vector of its first order partial derivatives, and $\nabla^2 f(t)$ the matrix of its second order derivatives, written out as a vector for the following to make dimensional sense.

(i) Show that the covariance of $(f(t), \nabla f(t), \nabla^2 f(t))$ can be written as

$$\mathrm{Var}\begin{pmatrix} f(t) \\ \nabla f(t) \\ \nabla^2 f(t) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & -\Lambda \\ 0 & \Lambda & 0 \\ -\Lambda & 0 & \mathcal{E} \end{pmatrix} \tag{2.8.1}$$

where

$$\Lambda_{ij} = \mathbb{E}\{f_i(t) f_j(t)\}$$

and

$$\mathcal{E}_{ij,kl} \stackrel{\Delta}{=} \mathbb{E}\{f_{ij}(t) f_{kl}(t)\}.$$

(ii) Express $\Lambda$ and $\mathcal{E}$ in terms of the spectral measure (2.4.2) of $f$.
(iii) Show that $\mathcal{E}$ is symmetric in $i, j, k, l$.
(iv) How do these covariances simplify if $f$ is isotropic?

**Exercise 2.8.11.** Taking $W$ to be the Brownian motion on $[0,1]$ – i.e. the zero mean Gaussian process with covariance function $C(s,t) = s \wedge t \stackrel{\Delta}{=} \min(s,t)$ – show that integral equation defining the eigenvalues $\lambda_n$ and eigenfunctions $\psi_n$ needed for finding a Karhunen-Loève expansion for $W$ (i.e. (2.3.20)) is

$$\lambda\psi(t) = \int_0^1 \min(s,t)\,\psi(s)\,ds = \int_0^t s\,\psi(s)\,ds \;+\; t\int_t^1 \psi(s)\,ds.$$

Differentiate both sides twice with respect to $t$ to find a second order ordinary differential equation whose solution, together with the appropriate conditions, gives

$$\psi_n(t) \;=\; \sqrt{2}\sin\left(\tfrac{1}{2}(2n+1)\pi t\right), \quad \lambda_n \;=\; \left(\frac{2}{(2n+1)\pi}\right)^2.$$

**Exercise 2.8.12.** Let $W$ be a Gaussian white noise based on Lebesgue measure, and use it to define a random field on $\mathbb{R}^N_+ = \{(t_1, ..., t_N) : t_i \geq 0\}$ by setting

$$W(t) \;=\; W([0,t]), \tag{2.8.2}$$

where $[0,t]$ is the rectangle $\prod_{i=1}^N [0,t_i]$. $W_t$ is called the Brownian sheet on $\mathbb{R}^N_+$, or *multiparameter Brownian motion*. If $N = 1$ it is standard Brownian motion.

(i) Show that $W$ is a centered Gaussian field on $\mathbb{R}^N_+$ with covariance

$$\mathbb{E}\{W_s W_t\} \;=\; (s_1 \wedge t_1) \times \cdots \times (s_N \wedge t_N), \tag{2.8.3}$$

where $s \wedge t \stackrel{\Delta}{=} \min(s,t)$.

(ii) Suppose $N > 1$, and fix $N - k$ of the indices. Show that $W$ is a scaled $k$-parameter Brownian sheet in the remaining variables.
(iii) Using the result of Exercise 2.8.11, find a Karhunen-Loève expansion for $W$ on $[0,1]^N$.

**Exercise 2.8.13.** Show that for an isotropic random field on $\mathbb{R}^N$ the orthogonal expansion (2.3.17) can never be finite.

**Exercise 2.8.14.** By using the fact that covariance functions must be nonnegative definite, show that (2.4.37) is true.

**Exercise 2.8.15.**

(i) Let $f(t) = \langle t, \xi \rangle$, $t \in S^{N-1}$, be the canonical Gaussian field on the sphere in $\mathbb{R}^N$, where $\xi \sim N(0, I_{N\times N})$. Show that $f(t)$ is stationary and isotropic with covariance function $C(t_1, t_2) = \langle t_1, t_2 \rangle$.
(ii) Show that the restriction of a stationary isotropic random field in $\mathbb{R}^N$ to any fixed lower dimensional sphere is also stationary and isotropic on that sphere.

**Exercise 2.8.16.** Let $N(t)$, $t \geq 0$, be a standard, unit rate, Poisson process.

(i) Show that $N$ is continuous in probability and in mean square, but not with probability one.
(ii) Can you find an example of a stochastic process that is continuous in probability but not in mean square? What about the other way around?
(iii) Show that $M(t) \stackrel{\Delta}{=} \int_0^t N(s)\, ds$ is differentiable in mean square, but not with probability one.

**Exercise 2.8.17.** Let $f$ be any stochastic process on $\mathbb{R}^N$.

(i) Show that $f$ is mean square continuous if and only if its covariance function $C$ is continuous on $T \times T$.
(ii) Show that if $C$ is continuous at diagonal points $(t, t)$, then it is continuous everywhere on $T \times T$.

**Exercise 2.8.18.** Suppose $f$ is a continuous Gaussian process on $[0,1]$, $\varphi$ a homeomorphism of $[0,1]$, and $g$ a new process defined, also on $[0,1]$, by $g_t = f(\varphi(t))$.

(i) Show that $f$ and $g$ have identical entropy functions.
(ii) Show, by example, that it is possible for the covariance function of $f$ to satisfy (2.5.12) while that of $g$ does not.

Hint: Use Brownian motion as your process.

**Exercise 2.8.19.** Prove the equivalence of the finiteness of the two integrals in (2.5.12) and that the convergence of both is assured by (2.5.13).

**Exercise 2.8.20.** Given a stationary Gaussian field $f$ on $\mathbb{R}^N$ with spectral measure $\nu$,

(i) Show that the finiteness of the spectral integral (2.5.19) ensures that $f$ is sample path continuous.
(ii) Show that $f$ will be $k$ times continuously differentiable if

$$\int_{\mathbb{R}^N} |\lambda|^{2k} \big( \log(1+|\lambda|) \big)^{1+\alpha} \nu(d\lambda) < \infty,$$

for some $\alpha > 0$.

**Exercise 2.8.21.** Find a counterexample to "Slepian's inequality for absolute values" which, if it were correct, would claim that the inequality holds with absolute values.
Hint: It suffices to find a counterexample based on a parameter space with only two points.

**Exercise 2.8.22.** Suppose that $f$ is a stationary, zero mean, Gaussian random field on $\mathbb{R}^N$, with a covariance function $C(t)$ that can be expanded as

$$C(t) = C(0) + \tfrac{1}{2} t \Lambda t' + o(|t|^2),$$

for some matrix $\Lambda$ of second spectral moments, and for $|t|$ in some neighborhood of the origin.

(i) Using Slepian's inequality, show that there is (perhaps another) neighborhood of the origin throughout which the exceedence probabilities of $f$ can be bounded, above and below, by those of a cosine process.
(ii) Identify the parameters of the cosine process.

**Exercise 2.8.23 (Rice formula).** Let $f$ be a $C^1$ process defined on the real line. Let $N_u(T)$ denote the number of upcrossings by $f$ of the level $u$ in $[0,T]$, viz.

$$N_u(T) = \Big\{ \# \Big\{ t \in [0,T] : \, f(t) = u, \;\; \dot{f}(t) > 0 \Big\} \Big\}$$

(i) Use Theorem 2.7.1 to derive Rice's original formula:

$$\mathbb{E}\left\{ N_u(T) \right\} = \int_0^T \mathbb{E}\left\{ \dot{f}(t) \, \mathbb{1}_{\{\dot{f}(t)>0\}} \big| f(t) = u \right\} p_t(u) \, dt,$$

where $p_t(u)$ is the density of $f(t)$.
(ii) Assume that $f$ is stationary and Gaussian with zero mean and unit variance and show that

$$\mathbb{E}\left\{ N_u(T) \right\} = T \frac{\lambda_2^{1/2}}{2\pi} e^{-u^2/2}, \tag{2.8.4}$$

where $\lambda_2$ is the second spectral moment

$$\lambda_2 = \mathbb{E}\left\{\dot{f}(t)^2\right\} = \int_{\mathbb{R}} \lambda^2 \, \nu(d\lambda)$$

and $\nu$ is the spectral measure (2.4.2).

(iii) Assume that $f$ is Gaussian with zero mean and constant unit variance. Show that

$$\mathbb{E}\left\{N_u(T)\right\} = \frac{e^{-u^2/2}}{2\pi} \int_0^T \lambda_t^{1/2}(t)\, dt,$$

where $\lambda_t = \mathbb{E}\left\{\dot{f}(t)^2\right\}$.

**Exercise 2.8.24.** Let $f$ be the cosine process of (2.3.3). Show, from first principles, (i.e. without using the results of Section 2.7) that, for $T \leq \pi/\lambda$, the Rice formula (2.8.4) holds for this process.

**Exercise 2.8.25.** Let $f$ be a $C^2$ process defined on the real line, and let $M_u(T)$ be the number of local maxima of $f$ above the level $u$ in $[0, T]$, viz.

$$M_u(T) = \#\left\{t \in [0,T] : \ \dot{f}(t) = 0, \ \ddot{f}(t) < 0, \ f(t) \geq u\right\}.$$

Also, let $M(T) = M_{-\infty}(T)$ be the total number of local maxima in $[0, T]$.

(i) Apply Rice's formula to show that if $f$ is stationary and Gaussian with mean 0 and variance 1 then

$$\mathbb{E}\left\{M(T)\right\} = T\frac{\lambda_4^{1/2}}{2\pi\lambda_2^{1/2}},$$

where $\lambda_4$ is the fourth order spectral moment

$$\lambda_4 = \mathbb{E}\left\{\ddot{f}(t)^4\right\} = \int_{\mathbb{R}} \lambda^4 \, \nu(d\lambda),$$

and $\nu$ is the spectral measure (2.4.2).

(ii) For general $f$, use Theorem 2.7.1 to show that

$$\mathbb{E}\left\{M_u(T)\right\} = \int_0^T \mathbb{E}\left\{-\ddot{f}(t)\, \mathbb{1}_{\{\ddot{f}(t)<0\}}\, \mathbb{1}_{\{f(t)>u\}} \big| \dot{f}(t) = 0\right\} \dot{p}_t(0)\ dt,$$

where $\dot{p}_t$ is the density of $\dot{f}(t)$.

(iii) Again assuming that $f$ is stationary and Gaussian with mean 0 and variance 1, and using the above, show that

$$\mathbb{E}\left\{M_u(T)\right\} = T\frac{\lambda_4^{1/2}}{2\pi\lambda_2^{1/2}} \Psi\left(\frac{\lambda_4^{1/2} u}{\Delta^{1/2}}\right) - T\frac{\lambda_2^{1/2}}{\sqrt{2\pi}} \varphi(u) \Phi\left(\frac{\lambda_2 u}{\Delta^{1/2}}\right),$$

where $\Delta = \lambda_4 - \lambda_2^2$.

(iv) Using the above and Rice's formula show that, for $C^2$, stationary, zero mean and unit variance Gaussian processes on the real line,

$$\lim_{u\to\infty} \frac{\mathbb{E}\{M_u(T)\}}{\mathbb{E}\{N_u(T)\}} = 1.$$

**Exercise 2.8.26.** Suppose that $T = [0,1]^N$. Show that under conditions (a) and (b) of Theorem 2.7.1 there are no points $t \in \partial T$ satisfying $f(t) = u$.

# 3

# Geometry

In this chapter we shall introduce some of the basic ideas of integral geometry that will be needed later. Among these will be the Euler characteristic, intrinsic volumes and kinematic formulas of various kinds. Quite comprehensive studies of integral geometry are available in the monographs of Schneider [82] and Santaló [81]. A more complete treatment than the one following, with proofs and using the same notation, can be found in [5].

## 3.1 Basic Complexes and the Euler Characteristic

Before we can begin to talk about functionals such as the Euler characteristic, already discussed in Chapter 1, we have to make certain that they are well defined. Thus, we need first of all to find classes of sets which guarantee this, while at the same time being broad enough to include the excursion sets of random fields.

To build these sets, we begin with a class of geometric objects known as *basic complexes*. On these it will be rather straightforward to study the basic properties of the Euler characteristic. Also, while we shall not prove it, it is possible to show that, with probability one, excursion sets of a wide variety of random fields belong to this class.

We commence with some definitions and simple results, all of which are due to Hadwiger [38].

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 section, everything that we shall have to say will be dependent on this choice of basis. The restriction to a particular coordinate system disappears in the coordinate free approach based on Morse theory, which will be presented in Section 3.2.

We call a $k$-dimensional affine subspace of the form

$$E = \left\{ t \in \mathbb{R}^N : t_j = a_j,\ j \in J;\ -\infty < t_j < \infty,\ j \notin J \right\}$$

a *(coordinate) k-plane* of $\mathbb{R}^N$ if $J$ is a subset of size $N-k$ of $\{1,\ldots,N\}$ and $a_j,\ j \in J$, 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 $\mathbb{R}^N, k = 1,\ldots,N$. 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, and that in $\mathbb{R}^1$ basic sets are closed intervals or points. All closed convex sets in $\mathbb{R}^N$ are basic. 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, $B_1,\ldots B_m$, for which the intersections $B_{\nu_1} \cap \cdots \cap B_{\nu_k}$ are basic for any combination of indices $\nu_1,\ldots,\nu_k, k = 1,\ldots,m$. There is, obviously, no uniqueness in this representation, which is called a *partition* of $A$.

The class of $N$-dimensional basic complexes, which we denote by $\mathcal{C}_B^N$, is quite large, and, in view of the fact that convex sets are basic, includes the convex ring (i.e. the collection of all sets formed via finite union and intersection of convex sets.)

With a class of sets in mind, we can now begin our search for a way to describe the shape of sets by looking for an integer valued functional with the following two basic properties:

$$\varphi(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{if } A \neq \emptyset \text{ is basic,} \end{cases} \tag{3.1.1}$$

and

$$\varphi(A \cup B) = \varphi(A) + \varphi(B) - \varphi(A \cap B), \tag{3.1.2}$$

whenever $A,\ B,\ A \cup B,\ A \cap B \in \mathcal{C}_B^N$. That is, $\varphi$ is a finitely additive functional that assigns the value one to every basic set.

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. This functional is the Euler characteristic, and we shall continue to denote it by $\varphi$.

One way to compute the Euler characteristic that follows by iterating the additivity relationship (3.1.2) is to take a partition $B_1,\ldots,B_n$ of $A$ and note that

$$\varphi(A) = \sum\nolimits^{(1)} \varphi(B_j) - \sum\nolimits^{(2)} \varphi(B_{j_1} \cap B_{j_2}) \ + \ldots + (-1)^{n+1} \varphi(B_1 \cap \ldots \cap B_n), \tag{3.1.3}$$

where $\sum^{(k)}$ denotes summation over all subsets $\{j_1,\ldots,j_k\}$ of $\{1,\ldots,n\}$, $1 \leq k \leq n$. Since from the definition of a partition all the intersections in

(3.1.3) are either empty or basic sets, their Euler characteristics must zero or one, and so we can rewrite (3.1.3) as

$$\begin{aligned}\varphi(A) = \#\{&\text{Elements in the partition}\} \\ &-\#\{\text{Pairs in the partition with non-empty intersection}\} \\ &+\#\{\text{Triples in the partition with non-empty intersection}\} \\ &-\dots\end{aligned} \tag{3.1.4}$$

If you recall our first meeting with the Euler characteristic in the introductory Section 1.3, then this 'inclusion-exclusion' should remind you of the Euler characteristic we defined there for sets made up out of cubes. See Exercise 3.6.1.

An immediate consequence of (3.1.4) is that the Euler characteristic is an integer valued functional, something which was not explicitly required by our initial requirements (3.1.1) and (3.1.2).

The following theorem, which also follows from (3.1.3), albeit somewhat less directly, describes another algorithm for computing $\varphi(A)$ for $A \in \mathcal{C}_B^N$. It could actually be taken as a definition of the Euler characteristic for basic complexes

**Theorem 3.1.1.** *For basic complexes $A \in \mathcal{C}_B^N$, the Euler characteristic $\varphi$, as defined by* (3.1.1) *and* (3.1.2) *has the following equivalent iterative definition for any $1 \le j \le N$:*

$$\varphi(A) = \begin{cases} \text{number of disjoint intervals in } A & \text{if } N = 1, \\ \sum_x \{\varphi(A \cap E_x) - \varphi(A \cap E_{x^-})\} & \text{if } N > 1, \end{cases} \tag{3.1.5}$$

*where $E_x \stackrel{\Delta}{=} \{t \in \mathbb{R}^N : t_j = x\}$,*

$$\varphi(A \cap E_{x^-}) \stackrel{\Delta}{=} \lim_{y \downarrow 0} \varphi(A \cap E_{x-y}), \tag{3.1.6}$$

*and the summation in* (3.1.5) *is over the finite number of real $x$ for which the summand is non-zero.*

Figure 3.1.1 shows an example of this iterative procedure in $\mathbb{R}^2$. Here $j = 2$, so the lines $E_x$ are parallel to the horizontal $t_1$ direction. The values of $\varphi(A \cap E_{t_2})$ appear closest to the vertical axis, with those of $\varphi(A \cap E_{t_2}) - \varphi(A \cap E_{t_2^-})$ 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 Euler characteristic $\varphi$ and the number of connected components of the set differ. In this example they are, respectively, zero and one. The arrows indicate contributions to $\varphi$ as described in Theorem 3.1.1.

Although Theorem 3.1.1 gave a definition of the Euler characteristic based on the local behavior of sets at certain boundary points, it is not hard to see,

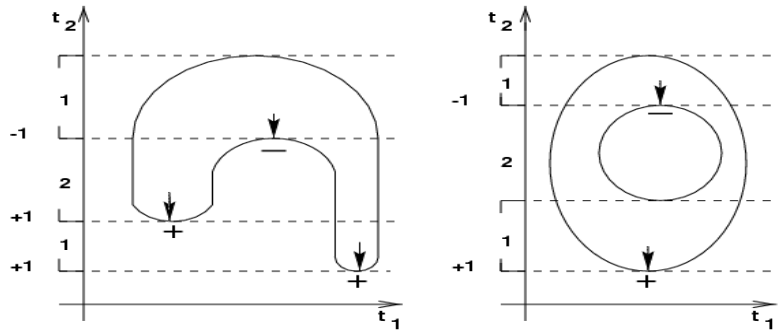


**Fig. 3.1.1.** Computing the Euler characteristic.

at least in two dimensions, that it also has a global, topological definition. In particular, if $A$ is a basic complex in $\mathbb{R}^2$, then

$\varphi(A) =$ Number of connected components in $A$ minus the number of holes.

The three dimensional case is a little harder, but you should still be able to convince yourself, via a few examples, that in this case,

$$\varphi(A) = \text{Number of connected components in } A \\ \text{minus the number of handles plus the number of holes.}$$

At this point, we can define 'holes' and 'handles' by example. A tennis ball has one hole in the middle. A teacup has one handle. One can replace 'handles' by 'tunnels' (think of the 'tunnel' going through the teacup's handle) or 'cylindrical holes drilled through $A$', as we shall in a moment.

To understand how this works in higher dimensions, you should try to visualize some $N$-dimensional examples to convince yourself that for the $N$-dimensional unit ball, $B^N$ and its boundary, the $(N-1)$-dimensional unit sphere, $S^{N-1}$:

$$\varphi(B^N) \;=\; 1, \qquad \varphi(S^{N-1}) \;=\; 1 + (-1)^{N-1}. \tag{3.1.7}$$

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 of dimension $N$ drilled completely 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,

$$\varphi(\bar{K}_{N,k}) \;=\; 1-k.$$

One final point to note about the Euler characteristic is that it possesses an important invariance property. Formally, homotopically equivalent sets have the same Euler characteristic. Less formally, sets which can be deformed into one another in a smooth fashion have the same Euler characteristic.

## 3.2 Excursion Sets and Morse Theory

You may have noticed that the points contributing to the Euler characteristic in Figure 3.1.1 are always maxima or minima of the height function $h(t_1, t_2) = t_2$, as one follows the values of $h$ along the boundary of the regions. Not all maxima and minima contribute, however. Neither of these facts is a coincidence, and, in fact, one can use the critical points of almost any twice differentiable function to compute the Euler characteristic. This is the basis of Morse theory, one of the topics of this section. The claim 'almost any' above is of course, rather imprecise, and a little searching for counter-examples will easily produce them. A class of functions for which this argument does work are the Morse functions which we shall define soon. With this definition, however, 'almost any' can be given a more precise meaning, since it turns out that, under mild assumptions, the sample paths of smooth random fields are Morse functions, with probability one.

Let us return briefly to the setting of excursion sets. To enable later probabilistic calculations, we would like to be able to compute their Euler characteristics directly from the function $f$, without having to look at the sets themselves. In other words, if we think of the sets in Figure 3.1.1 as the excursion sets of functions, then we need a way of describing which points contribute to the Euler characteristic of the excursion set, along with the contribution they make, in terms of local properties of $f$. To do this[1], we need some Morse theory.

We first state Morse's theorem for $C^2$ domains, i.e. closed compact sets in $\mathbb{R}^N$ bounded by $C^2$ hypersurfaces. The sets in Figure 3.1.1 satisfy this condition. However, before stating the theorem, we must first define the notion of non-degeneracy for both critical points of functions on domains and for their restrictions to hypersurfaces.

Suppose $M \subset \mathbb{R}^N$ is a $C^2$ domain and $f \in C^2(\mathbb{R}^N)$. We say a critical point of $f$, i.e. a point $t^*$ satisfying

$$\nabla f(t^*) = 0,$$

is non-degenerate if

$$\det\left(\nabla^2 f(t^*)\right) \neq 0,$$

[1] In fact, it turns out that Theorem 3.1.1 is no more that a special case of Morse's theorem (cf. Exercise 3.6.3).

where $\nabla^2 f$ is the standard $N \times N$ Hessian of second order partial derivatives of $f$.

A critical point of $f_{|\partial M}$, the restriction of $f$ to the boundary $\partial M$ of $M$, is a point $t^* \in \partial M$ at which $\nabla f$ is parallel to $\eta$, the unique outward pointing normal vector along $\partial M$. We write this as

$$\nabla f(t^*) \parallel \eta(t^*), \tag{3.2.1}$$

which is a statement about vectors in $\mathbb{R}^N$.

Another way to write this, which will be useful later, relies on the tangent spaces to $\partial M$. At a point $t \in \partial M$, the tangent space at $t$ is an $(N-1)$-dimensional linear approximation to $\partial M$, for which we can choose an orthonormal basis $\{e_1(t), \ldots, e_{N-1}(t)\}$. We can complete this to a local basis for $\mathbb{R}^N$ by adding $\eta$.

Using the basis of the tangent space, we can define the gradient $f_{|\partial M}(t)$ as the $(N-1)$-dimensional vector of derivatives of $f$ in the directions $e_1(t), \ldots, e_{N-1}(t)$, and rewrite (3.2.1) as

$$\nabla f_{|\partial M}(t^*) = 0. \tag{3.2.2}$$

The Hessian $\nabla^2 f_{|\partial M}(t)$ of $f$ (at a critical point) is similarly defined as the $(N-1)\times(N-1)$ matrix of second order derivatives with respect to this basis. Note that while, in general, both the gradient and Hessian are dependent on the choice of basis, neither the definition of a critical point nor the index of the Hessian depends on it.

The (tangential) index of the Hessian of a critical point is the index of it as a matrix, where we define the index of a symmetric matrix $A$ as

$$\mathrm{Ind}(A) = \text{Number of negative eigenvalues of } A. \tag{3.2.3}$$

Finally, a critical point of $f_{|\partial M}$ is the called non-degenerate if

$$\det\left(\nabla^2 f_{|\partial M}(t^*)\right) \neq 0.$$

Again, this does not depend on the choice of basis for the tangent space.

We now have all that we need to state Morse's theorem for $C^2$ domains.

**Theorem 3.2.1 (Morse's theorem for $C^2$ domains).** *Let $M \subset \mathbb{R}^N$ be a $C^2$ domain and $f \in C^2(\mathbb{R}^N)$ be such that*

*(i) $f$ has no critical points on $\partial M$.*
*(ii) $\nabla^2 f$, the Hessian of $f$, is non-degenerate at each critical point of $f$ in $M$.*
*(iii) $\nabla^2 f_{|\partial M}$, the Hessian of $f$ restricted to $\partial M$, is non-degenerate at each critical point of $f_{|\partial M}$.*

*Then,*

$$\varphi(M) = \sum_{\{t\in M:\nabla f(t)=0\}} (-1)^{\mathrm{Ind}(\nabla^2 f(t))} + \sum_{\{t\in\partial M:\nabla f_{|\partial M}(t)=0\}} (-1)^{\mathrm{Ind}(\nabla^2 f_{|\partial M}(t))} \mathbb{1}_{\{\langle \nabla f(t),\eta(t)\rangle<0\}}. \tag{3.2.4}$$

**Remark 3.2.2** *The condition that* $\langle \nabla f(t), \eta(t)\rangle < 0$ *in the indicator is related to the one-sided limit* (3.1.6) *in Theorem 3.1.1 and the counting of only the critical points on the lower boundary in Figure 3.1.1 when computing Euler characteristics, both of which result from a particular choice of* $f$*.*

*For details, see Exercise 3.6.3, which also confronts some of the limitations of Theorem 3.2.1, including the fact that it does not apply to domains as simple as rectangles.*

**Definition 3.2.3.** *We call* $C^2$ *functions satisfying conditions (i)–(iii) of Theorem 3.2.1 Morse functions.*

Theorem 3.2.1 gives a way of computing the Euler characteristic of a $C^2$ domain $M$ in terms of critical points of Morse functions restricted to $M$. Note that these functions have no particular relationship to the domain itself, and there is no uniqueness of the Morse function used in Theorem 3.2.1 to compute the Euler characteristic of $M$. In fact, this is one of the deepest aspects of the theorem.

Soon, however, we shall be interested in the Euler characteristic of excursion sets of the form

$$M \cap f^{-1}[u, +\infty),$$

so that now there *is* a connection between $f$ and the set whose Euler characteristic we want to calculate. Unfortunately, this set no longer fits into the framework of Theorem 3.2.1, since $M \cap f^{-1}[u, +\infty)$ is not generally a $C^2$ domain. The problem is that there are usually corners, edges, etc. where $\partial M$ meets $f^{-1}(u)$. Hence, the above version of Morse's theorem cannot be used for excursion sets. Furthermore, as we have already noted, Theorem 3.2.1, as it is currently formulated, cannot be applied to domains as simple as rectangles.

One way around these difficulties is to reinterpret the condition in the indicator in (3.2.4) as requiring that the vector $-\nabla f$ be directed to the exterior of the domain $M$ at boundary critical points. However, while it is clear what this means for $C^2$ domains, it is a somewhat more delicate in general, where, instead of there being a single outward normal direction, there may be many.

To describe these, we first need the *normal cone* and *support cone* at a point $t \in M$. To define these, let $V_t$ denote all those vectors whose base point is $t$, and define the support cone to be

$$\mathcal{S}_t \equiv \mathcal{S}_t M \stackrel{\Delta}{=} \{X \in V_t : \varepsilon X \in M \text{ for all } \varepsilon > 0 \text{ sufficiently small}\}. \tag{3.2.5}$$

In other words, these are the vectors pointing into the interior of $M$.

The normal cone at $t$ is defined as

$$N_t \equiv N_t M \stackrel{\Delta}{=} \{X \in V_t : \langle X, Y\rangle \leq 0 \text{ for all } Y \in \mathcal{S}_t(M)\}\,. \tag{3.2.6}$$

We shall often call vectors which lie in the normal cones *extended outward* vectors or directions, and boundary critical points for which $\nabla f(t) \in N_t M$ *extended outward critical points*. Note that with normal cones defined, the indicator in (3.2.4) can be rewritten as $\mathbb{1}_{\{-\nabla f(t) \in N_t M\}}$, and, as such, refers to extended *inward* critical points.

An example should make this definition quite clear. Consider a cube in $\mathbb{R}^3$, written as the disjoint union of the faces of all dimensions from three down to zero. Then,

(i) In the interior of the cube, the support cone is all of $\mathbb{R}^3$, the normal cone is empty, and so there are no extended outward directions.
(ii) Each point on a two dimensional face has a three dimensional half space as its support cone, and a unique outward pointing normal vector, as for the case of $C^2$ domains. This unique vector makes up the entire normal cone.
(iii) Along the one dimensional edges the support cones are three dimensional wedges. The normal cones are two dimensional, quarter-disks of directions within which the corresponding vectors are extended outward.
(iv) At the zero dimensional vertices, both the support cones and the normal cones are three-dimensional octants.

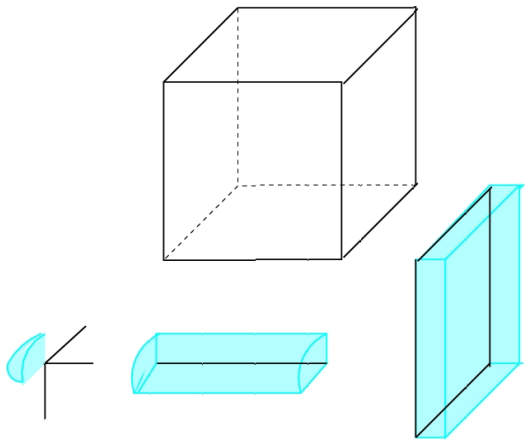

**Fig. 3.2.1.** The cube in $\mathbb{R}^3$ and some of its normal cones of extended outward directions.

Each point in a 3-dimensional cube $T$ can therefore be associated with a set of extended outward directions, depicted in Figure 3.2.1. The collection of

these directions at a point is the corresponding normal cone at $t$. For another example, a triangle in the plane, see Figure 3.3.1.

More formally, extending the example of the cube to a rectangle

$$T = \prod_{j=1}^{N} [0, T_j],$$

it is possible to find a simple, albeit notationally a little tedious, way of writing the normal cones. The notation provides a simple way to index the faces of $T$. In particular, a *face* $J$ of $T$, of dimension $k$, is defined by fixing a subset $\sigma(J)$ of $\{1, \dots, N\}$ of size $k$ and a sequence of $N - k$ zeros and ones, which we write as $\varepsilon(J) = \{\varepsilon_j(J), j \notin \sigma(J)\}$, so that

$$J = \{t \in T : t_j = \varepsilon_j(J) T_j, \text{ if } j \notin \sigma(J);\ 0 < t_j < T_j, \text{ if } j \in \sigma(J)\}. \quad (3.2.7)$$

In other words, each $k$-dimensional face $F$ is determined by fixing $N - k$ coordinates, and each of these coordinates can be fixed at either the top or bottom of the cube. Note that, with the above definition, faces are always open sets when considered as subsets of the affine subspaces in which they sit.

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 $\partial_N T$ contains only $T^\circ$, the interior of $T$, while $\partial_0 T$ contains the $2^N$ vertices of the rectangle. In general, $\partial_k T$ has $2^{N-k}\binom{N}{k}$ elements. Note also that the boundary of a $k$-dimensional face $J$, $\partial J = \overline{J} \setminus J$, is given by the disjoint union

$$\partial J = \bigcup_{k=0}^{k-1} \bigcup_{J \in \partial_k T} J. \quad (3.2.8)$$

With this notation, the normal cone at $t \in J$ is simply the set of directions $\nu$ that satisfy

$$\begin{aligned} \nu_i &= 0 && i \notin \sigma(J), \\ \operatorname{sgn}(\nu_i) &= \varepsilon_i^*(J) && i \in \sigma(J) \end{aligned}$$

where $\varepsilon_i^*(J) = 2\varepsilon_i(J) - 1$.

The following is a version of Morse's theorem for the excursion sets of a $C^2$ function $f$ defined on a rectangle.

**Theorem 3.2.4 (Morse's theorem for excursion sets over rectangles).** *Let $T = \prod_{j=1}^{N} [0, T_j]$ be a rectangle in $\mathbb{R}^N$ and $f \in C^2(\mathbb{R}^N)$ such that, for every face $J$ of $T$,*

*(i) $f_{|\overline{J}}$, the restriction of $f$ to the closure of $J$, has no critical points on $\partial J$.*
*(ii) $u$ is a regular value of $f_{|J}$, in the sense that $f_{|J}$ has no critical points at which $f_J(t) = u$.*
*(iii) $\nabla^2 f_{|J}$ is non-degenerate at each critical point of $f_{|J}$.*

*Then*

$$\varphi\left(T \cap f^{-1}[u,+\infty)\right) = \sum_{J \in T} \sum_{t \in J:\, \nabla f_{|J}(t)=0} (-1)^{\mathrm{Ind}(-\nabla^2 f_{|J}(t))} \mathbb{1}_{\{f(t)\geq u\}} \mathbb{1}_{\left\{\mathrm{sgn}\left(\frac{\partial f}{\partial t_i}\right)=\varepsilon_i^*(J),\forall i\in J\right\}}.$$

There is also a corresponding version of this result for $C^2$ domains.

**Theorem 3.2.5 (Morse's theorem for excursion sets over $C^2$ domains).** *Let $M \subset \mathbb{R}^N$ be a $C^2$ domain and $f \in C^2(\mathbb{R}^N)$ be such that*

*(i) $f$ has no critical points on $\partial M$.*
*(ii) $f$ has no critical points at which $f(t) = u$.*
*(iii) $\nabla^2 f$ is non-degenerate at each critical point of $f$ in $M$.*
*(iv) $\nabla^2 f_{|\partial M}$ is non-degenerate at each critical point of $f_{|\partial M}$.*

*Then,*

$$\begin{aligned}\varphi\left(M \cap f^{-1}[u,+\infty)\right) &= \sum_{t \in M:\nabla f(t)=0} (-1)^{\mathrm{Ind}(-\nabla^2 f(t))} \mathbb{1}_{\{f(t)\geq u\}} \\ &\quad + \sum_{t \in \partial M:\nabla f_{|\partial M}(t)=0} (-1)^{\mathrm{Ind}(-\nabla^2 f_{|\partial M}(t))} \mathbb{1}_{\{f(t)\geq u\}} \mathbb{1}_{\nabla f \in N_t M}.\end{aligned}$$

Since Theorems 3.2.4 and 3.2.5 cover parameter spaces that are, respectively, rectangles and $C^2$ domains, they treat completely disjoint classes of domains. Nevertheless, the final results are close enough that one expects that they should both be special cases of a more general theorem, where the domains are in part smooth, and yet are allowed to have edges and corners. This is in fact true, although stating the general result would involve us in more notation that we plan to develop in this book. Nevertheless, we shall discuss this briefly in Section 3.5 below, and you can find full details in Chapter 9 of *RFG*.

## 3.3 Volume of Tubes and Intrinsic Volumes

The Euler characteristic of Section 3.1 arose as the unique additive functional on basic complexes (cf. (3.1.1) and (3.1.2)) that assigned the value one to basics, and zero 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. However, the Euler characteristic by itself is of limited value. For example, the empty set, a disk with a hole, and 100 disks of which one contains 100 holes, all have the same Euler characteristic of zero, despite the fact that they are quite different. Also,

a disk of radius 1, and a disk of radius 100, have the same Euler characteristic of one.

It is therefore natural to want to go beyond the Euler characteristic and beyond mere counting, and begin to say things about the volume of sets, the surface area of their boundaries, their curvatures, etc. This is what we shall do now.

### 3.3.1 Euclidean Sets

In this section, we consider simple sets in Euclidean space, and describe $N$ new, position and rotation invariant functionals $\{\mathcal{L}_j\}_{j=1}^N$, which are also additive in the sense of (3.1.2) but scale with dimensionality in the sense that

$$\mathcal{L}_j(\lambda A) = \lambda^j \mathcal{L}_j(A), \qquad \lambda > 0, \tag{3.3.1}$$

where $\lambda A \stackrel{\Delta}{=} \{t : t = \lambda s,\ s \in A\}$.

Defining $\mathcal{L}_0$ to be the Euler characteristic, these functionals make up what are known as the *intrinsic volumes* and can be defined on quite large classes of sets. They can be defined in a number of ways, one of which is implicitly via *Steiner's formula* [49, 81] and its generalisations. To describe Steiner's formula, we first require the notion of a *tube* around a set.

For $A \subset \mathbb{R}^N$ and $\rho > 0$, let

$$\begin{aligned} \mathrm{Tube}(A,\rho) &= \{x \in \mathbb{R}^N : d(x,A) \le \rho\} \\ &= \left\{x \in \mathbb{R}^N : \exists\ y \in A \text{ such that } d(x,y) \le \rho\right\} \end{aligned} \tag{3.3.2}$$

be the *tube of radius* $\rho$, or $\rho$-*tube*, around $A$, where

$$d(x,A) \stackrel{\Delta}{=} \inf_{y \in A} |x - y|$$

is the usual Euclidean distance from the point $x$ to the set $A$. An example is given in Figure 3.3.1, in which $A$ is the inner triangle and $\mathrm{Tube}(A,\rho)$ the larger triangular object with rounded-off corners (which includes, of course, the original triangle as well).

However, unlike this particular example, it is not necessary that a set and its tube have the same dimension. For example, the tube around a simple one-dimensional curve in $\mathbb{R}^3$ will have dimension three.

With $\lambda_N$ denoting Lebesgue measure in $\mathbb{R}^N$, Steiner's formula states that, for convex $A$, the volume $\lambda_N(\mathrm{Tube}(A,\rho))$ is a polynomial in $\rho$ with terms of order $N,\ N-1,\ldots,N-\dim(A)$, where $\dim(A)$ is the dimension of $A$. In particular, if we write this polynomial in the form

$$\lambda_N\left(\mathrm{Tube}(A,\rho)\right) = \sum_{j=0}^{\dim A} \omega_{N-j}\rho^{N-j}\mathcal{L}_j(A), \tag{3.3.3}$$

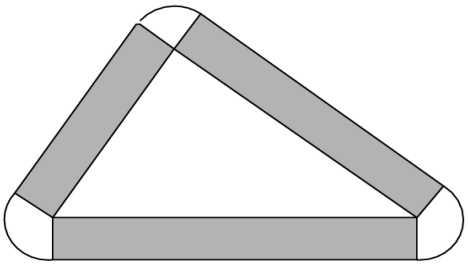

**Fig. 3.3.1.** The tube around a triangle.

where

$$\omega_j = \lambda_j(B(0,1)) = \frac{\pi^{j/2}}{\Gamma\left(\frac{j}{2}+1\right)} \tag{3.3.4}$$

is the volume of the unit ball in $\mathbb{R}^j$, then (3.3.3) defines the intrinsic volumes of $A$.

We shall not attempt to prove Steiner's formula, but rather motivate it by looking at some simple examples. For example, to find the area (i.e. 2-dimensional volume) of the tube about the triangle of Figure 3.3.1, one needs only 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 one and radius $\rho$.

Putting the above pieces[2] together we have

$$\text{area}\,(\text{Tube}(A,\rho)) \;=\; \pi\rho^2\varphi(A) \;+\; \rho\,\text{perimeter}(A) \;+\; \text{area}(A).$$

Comparing this to (3.3.3) it now takes only a little thought to guess what the intrinsic volumes must measure. If $A$ is two-dimensional, then

$$\begin{aligned} \mathcal{L}_2(A) &= \text{The area of } A, \\ \mathcal{L}_1(A) &= \tfrac{1}{2} \times (\text{The boundary length of } A) \\ \mathcal{L}_0(A) &= \text{The Euler characteristic of } A. \end{aligned} \tag{3.3.5}$$

If $A$ is three-dimensional, then

[2] Note that the above breakdown of the tube into its component parts is completely analagous to the breakdown of the different possibilities of the normal cone in the cube example in Figure 3.2.1.

$$\begin{aligned}
&\mathcal{L}_3(A) = \text{The volume of } A,\\
&\mathcal{L}_2(A) = \tfrac{1}{2} \times (\text{The surface area of } A),\\
&\mathcal{L}_1(A) = 2 \times (\text{The caliper diameter of } A) \quad (\text{cf. Section 1.4.2})\\
&\mathcal{L}_0(A) = \text{The Euler characteristic of } A.
\end{aligned} \tag{3.3.6}$$

Higher dimensions require a little more imagination, but it is reasonably obvious that $\mathcal{L}_N$ and $\mathcal{L}_{N-1}$ measure volume and surface area, while $\mathcal{L}_0$ is still the Euler characteristic. Note that $\mathcal{L}_j(A) = 0$ for all $j > \dim A$.

We now consider an $N$-dimensional example for which no hand-waving is needed. As usual, let $B_\lambda^N$ be the ball of radius $\lambda$ in $\mathbb{R}^N$. Noting that $\text{Tube}(B_\lambda^N, \rho) = B_{\lambda+\rho}^N$, we have

$$\begin{aligned}
\lambda_N\left(\text{Tube}\left(B_\lambda^N, \rho\right)\right) &= (\lambda+\rho)^N \omega_N \\
&= \sum_{j=0}^{N} \binom{N}{j} \lambda^j \rho^{N-j} \omega_N \\
&= \sum_{j=0}^{N} \omega_{N-j} \rho^{N-j} \binom{N}{j} \lambda^j \frac{\omega_N}{\omega_{N-j}}.
\end{aligned}$$

Comparing this to Steiner's formula (3.3.3) it is immediate that

$$\mathcal{L}_j\left(B_\lambda^N\right) = \binom{N}{j} \lambda^j \frac{\omega_N}{\omega_{N-j}}. \tag{3.3.7}$$

Three more important $N$-dimensional examples, of cubes, rectangles and spheres, can be found in Exercises 3.6.4 and 3.6.5.

It is sometimes useful consider a slightly different indexing and normalization of the intrinsic volumes, to obtain the so-called *Minkowski functionals*, a term which might be more familar to many readers. These are defined as

$$\mathcal{M}_j(A) = (j!\,\omega_j)\,\mathcal{L}_{N-j}(A), \qquad j = 0, \ldots, N, \tag{3.3.8}$$

so that $\mathcal{M}_j(A) = 0$ for $j < N - \dim A$.

When expressed in terms of the $\mathcal{M}_j$, Steiner's formula now reads like a Taylor series expansion, since then we have

$$\lambda_N\left(\text{Tube}(A, \rho)\right) = \sum_{j=0}^{N} \frac{\rho^j}{j!} \mathcal{M}_j(A). \tag{3.3.9}$$

Of course, this is a very special Taylor series, since it terminates after the $N$-th term. It is this fact that is actually one of the main points of Steiner's formula, along with the fact that the coefficients in the expansion have geometric meanings.

Given the intrinsic definition of the intrinsic volumes of $A$ as coefficients in a polynomial expansion of the volume of a tube around $A$, we have seen,

for some simple examples, how to compute them. Moving from these simple examples to more complicated ones is not always easy. If the more complicated sets can be expressed as unions and intersections of simple ones, then we can use the additivity properties of the $\mathcal{L}_j$.

However, this is not always possible, so we now give a result that describes how to go about computing intrinsic volumes for $C^2$ domains. For this, however, we need the notion of the (outward) curvature matrix $C(t)$ at $t \in \partial M$.

Roughly speaking, the curvature matrix is a $(N-1) \times (N-1)$ matrix of second order derivatives of the function that measures the perpendicular distance between $\partial M$ and the tangent space at $t$, where the perpendicularity is with respect to the tangent space. The formal definition requires the notions of covariant differentiation and second fundamental forms, , which are beyond the scope of this volume (but not, of course, of *RFG*, to where you can turn for details).

**Theorem 3.3.1 (Intrinsic volumes of $C^2$ domains).** *Let $M \subset \mathbb{R}^N$ be a compact $C^2$ domain with outward curvature matrix $C(t)$, $t \in \partial M$. Then, Steiner's formula, now called Weyl's tube formula, still holds, in that, for some $\rho_c > 0$ and for all $\rho < \rho_c$,*

$$\lambda_N\left(\mathrm{Tube}(M,\rho)\right) = \sum_{j=0}^{\dim A} \omega_{N-j}\rho^{N-j}\mathcal{L}_j(M), \tag{3.3.10}$$

*where the intrinsic volumes, for $0 \le j \le N-1$, can by computed as*

$$\mathcal{L}_j(M) = \frac{1}{s_{N-j}} \int_{\partial M} \mathrm{detr}_{N-1-j}(-C(s))\, ds. \tag{3.3.11}$$

*Here, for a square matrix $A$, $\mathrm{detr}_j(A)$ is the sum of the determinants of all $j \times j$ principal minors of $A$. The integration over $\partial M$ is with respect to the usual surface measure.*

The proof of Theorem 3.3.1 is of a debatable level of difficulty. In the paper [96] in which it was originally proven, the author, Hermann Weyl, claimed that it is hardly "more than what could have been accomplished by any student in a course of calculus."[3] On the other hand, most professors of mathematics have been searching for students of this caliber ever since. However, Weyl is undeniably correct in his implication that the proof requires little more than calculus. The basic approach involves parameterizing the tube of radius $\rho$ about $M$, but with the interior of $M$ removed, as the image of $\partial M \times [0,\rho]$ under the mapping

$$(t,r) \mapsto t + r \cdot \eta_t,$$

where $\eta_t$ is the unique outward pointing normal vector on $\partial M$. The rest of the work involves computing the Jacobian of this transformation. You can

[3] Actually, Weyl's statement was made in the more general setting of embedded submanifolds of $\mathbb{R}^N$, and the case of domains is considerably easier.

find a proof, in a more complicated scenario, but with the same notation, in Chapter 10 of *RFG*.

Before we move on to our next topic, there is one more fact that is worth noting, which testifies further to the centrality of intrinsic volumes in geometry.

Recall that the Euler characteristic was introduced at the very beginning of this chapter by demanding that it take the value one on basics – (3.1.1) – and satisfy the additivity condition (3.1.2). Suppose that we drop the first of these conditions, and look only for additive functionals $\psi$ on nice sets that are invariant under rigid motions and, either monotonic or continuous (in an appropriate[4] metric).

Then an extremely important result due to Hadwiger [37] for basic complexes, and others for more general classes of sets[5], is the following.

**Theorem 3.3.2 (Hadwiger's theorem).** *Let $\psi$ be a real valued function on nice classes of sets in $\mathbb{R}^N$, invariant under rigid motions, additive (in the sense of (3.1.2)) and monotone, in the sense that, for all pairs $A, B$, either $A \subseteq B \;\Rightarrow\; \psi(A) \leq \psi(B)$ or $A \subseteq B \;\Rightarrow\; \psi(A) \geq \psi(B)$. Then*

$$\psi(A) \;=\; \sum_{j=0}^{N} c_j \mathcal{L}_j(A), \tag{3.3.12}$$

*where $c_0, \ldots, c_N$ are ($\psi$-dependent) constants.*

Thus, studying intrinsic volumes is equivalent to studying a far wider class of functionals on sets. We shall have cause to apply this result on a number of occasions.

### 3.3.2 Intrinsic Volumes for Simplicial Complexes

Having described some geometry for basic complexes and $C^2$ domains, there are now essentially two directions in which we could continue. One involves moving to more general domains, such as stratified manifolds. This is important, but highly technical, and we shall discuss it briefly in Section 3.5. Now, however, we want to take an almost orthogonal direction, moving from Euclidean domains to something seemingly intrinsically simpler, simplicial complexes. The main reason for this is a practical one.

[4] The metric here is most definitely not one of the standard ones – such as the symmetric difference metric $d(A, B) = |A \cap B^c| + |A^c \cap B|$ – unless one makes additional assumptions, such as restricting the entire discussion to convex sets (but *not* the convex ring). See the original papers for details.

[5] The last word on this, at the time of writing, seems to be due to Zahle [105]. The proofs are not easy, although Klain and Rota [49] have a self-contained and readable proof for continuous invariant functionals on the convex ring.

In all of the applications in Part IV of this book, and, indeed, for all of the applications of which we aware, data never comes in the form of a continuous random field. Rather, by the very nature of computing, it is discrete in both the parameter space and the values it takes. The latter is of no real importance. Here the level of accuracy is so high that there is no practical difference between continuous and discrete. This is not always true of the discreteness in the parameter space, however, which is generally due more to limited sampling than to computing limitations, the basic data generally being made up of observations of a smooth random field on an underlying mesh.

These underlying meshes are usually examples of simplicial complexes.

**Definition 3.3.3.** *Given a set of points $V$, a simplicial complex $\mathcal{S}$ on $V$ is a collection of subsets of $V$ with the property that if $\mathcal{F} \in \mathcal{S}$, then so are all the subsets of $\mathcal{F}$.*

A simple example of a simplicial complex is given in Figure 3.3.2. An example in three dimensions is given by a tetrahedral mesh, by taking every tetrahedron in the mesh, along with its triangular faces, and edges and vertices. With each simplex $\mathcal{F} \in \mathcal{S}$ we associate the *physical simplex* $F$ consisting of the (closed) convex hull of the points in $\mathcal{F}$. With this convention we assume, in this section, that $S$ is the union of all the simplices in $\mathcal{S}$. For further details on general simplicial complexes see [49] and for statistical applications see [64, 65].

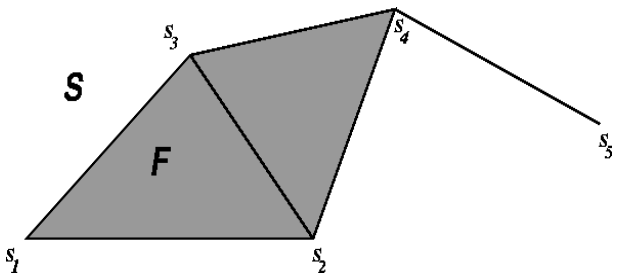


**Fig. 3.3.2.** Example of a simplicial complex. $\mathcal{S} = \{\{s_1\}, \{s_2\}, \{s_3\}, \{s_4\}, \{s_5\}, \{s_1, s_2\}, \{s_1, s_3\}, \{s_2, s_3\}, \{s_2, s_4\}, \{s_3, s_4\}, \{s_4, s_5\}, \{s_1, s_2, s_3\}, \{s_2, s_3, s_4\}\}$ with $V = \{s_1, \ldots, s_6\}$. $F$ is the solid triangle associated with the simplex $\mathcal{F} = \{s_1, s_2, s_3\}$. The union of all the points, edges and triangles in $\mathcal{S}$ is the set $S \subset \mathbb{R}^2$.

An immediate consequence of Definition 3.3.3 is that if we sample from a simplicial complex, then we once again have a simplicial complex. In particular, if $f$ is a function defined on a the vertices of a simplicial complex $\mathcal{S}$, then its excursion set is also a simplicial complex. That is, $\{\mathcal{F} \in \mathcal{S} : f(v) \geq u, \ \forall\ v \in \mathcal{F}\}$ is a simplicial complex.

It should now be obvious, in view of what has gone before, that intrinsic volumes of sets $S$ arising from simplicial complexes will be of interest to us.

Using the basic additivity properties of the intrinsic volumes, and the fact that all faces $F$ are convex, these are, in principle, easy to compute. Furthermore, the intrinsic volumes of any simplex $F$ depend only on the Lebesgue measure (volume, are, length, etc.) of the faces $F$, something we shall exploit later in Chapter ????? when we shall use distances between residual vectors as ohe methods for estimating curvatures.

However, these computations can be quite tedious. Fortunately, there are shortcuts.

For example, suppose that $F$ is a simplex in $\mathbb{R}^N$. It is not hard to see that, for any face $G \subset F$, the normal cone is identical (up to a translation parallel to $G$) for all $t \in G$. Define $\beta(G, F)$ to be the proportion of normal vectors to the span of $G$ that are in the normal cone, or, equivalently, the proportion of all normal vectors to a $t \in G$ that lie in the normal cone $N_t$. For the special case $G = F$, set $\beta(F, F) = 1$.

Then, writing $|G|$ for the $\dim(G)$-dimensional size of $G$, one can show (cf. Exercise 3.6.10) that

$$\mathcal{L}_j(F) = \sum_{G \subseteq F \,:\, \dim(G)=j} |G| \cdot \beta(G, F). \tag{3.3.13}$$

Note how much simpler this sum is than one would obtain via additivity, which would give an inclusion-exclusion type of sum with many more terms.

Another useful result (cf. Exercise 3.6.11) is that, for any simplicial complex $S$,

$$\mathcal{L}_d(S) = \sum_{F \in S:\ \dim(F) \geq d} (-1)^{\dim(F)-d} \mathcal{L}_d(F) \tag{3.3.14}$$

where the sum is over all simplices $F$ in $S$.

### 3.3.3 Volumes of Tubes on Spheres

So far, we have discussed only the usual Euclidean volumes of tubes of relatively simple sets in $\mathbb{R}^N$. However, the theory can be extended significantly, and there exist complete books devoted to the subject, the classic one being by Gray [35].

In this section we consider just one more case, in which the underlying sets are spherically convex subsets of $S_\lambda^{N-1}$, the sphere of radius $\lambda$ in $\mathbb{R}^N$. Spherically convex sets are those which are the intersection of a convex cone and $S_{N-1}^\lambda$.

In anticipation of later notation, let $\mathcal{H}_{N-1}$ be $(N-1)$-dimensional Hausdorff measure on $S_{N-1}^\lambda$, or, equivalently, surface measure. Thus

$$\mathcal{H}_{N-1}\left(S_{N-1}^\lambda\right) = s_N \lambda^{N-1} = \frac{2\pi^{N/2}}{\Gamma(N/2)} \lambda^{N-1}.$$

For a spherically convex subset $A$ of $S_\lambda^{N-1}$ the geodesic tube of radius $\rho$ around $A$ is defined as

$$\text{Tube}(A,\rho) = \left\{x \in S_\lambda^{N-1} : d(x,A) \le \rho\right\}, \tag{3.3.15}$$

where

$$d(x,y) = \lambda \cdot \theta(x,y) \stackrel{\Delta}{=} \lambda \cos^{-1}\left(\frac{\langle x,y\rangle}{\lambda^2}\right)$$

is geodesic distance on $S_{N-1}^\lambda$. Alternatively, we could write (3.3.15) as

$$\text{Tube}(A,\rho) = \left\{x \in S_\lambda^{N-1} : \exists\, y \in A \text{ such that } \frac{\langle x,y\rangle}{\lambda^2} \ge \cos(\rho/\lambda)\right\}. \tag{3.3.16}$$

The volume of tubes formula for $A \subset S_{N-1}^\lambda$ is

$$\mathcal{H}_{N-1}\left(\text{Tube}(A,\rho)\right) = \sum_{i=0}^{N-1} \lambda^{N-1-i} G_{i,N-1-i}(\rho/\lambda)\mathcal{L}_i^{\lambda^{-2}}(A), \tag{3.3.17}$$

where, for $\lambda \ge 0$, the

$$\mathcal{L}_i^\kappa(A) \stackrel{\Delta}{=} \sum_{n=0}^{\lfloor \frac{\dim A - i}{2} \rfloor} \frac{(-\kappa)^n}{(4\pi)^n} \frac{(i+2n)!}{n!i!} \mathcal{L}_{i+2n}(A) \tag{3.3.18}$$

are the intrinsic volumes for spaces of constant curvature $\kappa$, and

$$\begin{aligned} G_{a,b}(\rho) &\stackrel{\Delta}{=} \frac{\pi^{b/2}}{b\Gamma\left(\frac{b}{2}+1\right)} \int_0^\rho \cos^a(r)\sin^{b-1}(r)\, dr \\ &= \frac{\pi^{b/2}}{\Gamma\left(\frac{b}{2}\right)} \overline{IB}_{(a+1)/2,b/2}(\cos^2\rho), \end{aligned}$$

with

$$\overline{IB}_{(a+1)/2,b/2}(x) \stackrel{\Delta}{=} \int_x^1 x^{(a-1)/2}(1-x)^{(b-2)/2}\, dx,$$

the tail of the incomplete beta function.

Recall that Steiner's formula, which was a volume of tubes formula for convex sets in $\mathbb{R}^N$, held for all tube radii $\rho > 0$. On the other hand, Weyl's tube formula, as we presented it for $C^2$ domains in Theorem 3.3.1, only held for $\rho$ less than some critical $\rho_c$. On spheres the situation is a little more delicate, and even for spherically convex sets the tube formula only holds for small enough $\rho$. The problem, of course, is that as the radius of the tube grows, the tube can meet itself at the 'back' of the sphere.

We shall return to (3.3.17) in Chapter 4 when we discuss the suprema distributions of a special class of Gaussian processes known as finite Karhunen-Loève processes, processes whose Karhunen-Loève expanesion (2.3.22) terminates after a finite number of terms.

### 3.3.4 Probabilities of Tubes

Our entire discussion of tube sizes has, up until now, been based on their volumes, either as subsets of Euclidean space or of spheres. However, there is an interesting extension of these ideas, from volumes to probabilities. This extension will be of crucial importance for us in the following chapters.

Suppose that $\mathbb{P}$ is a probability measure on $\mathbb{R}^k$ with an analytic density with respect to Lebesgue measure, a canonical example being the Gaussian measure $\gamma_k$ corresponding to the distribution of a $N(0, I_{k\times k})$ random variable. As the derivation of the Steiner-Weyl formula (3.3.10) consists of integrating over hypersurfaces at distance $r$ from $A \subset \mathbb{R}^N$, it is reasonable to consider expanding the density of $\mathbb{P}$ in a power series in normal directions and integrating the density over these hypersurfaces. Doing so gives an expansion[6] of the form

$$\mathbb{P}\{\mathrm{Tube}(A,\rho)\} \stackrel{\Delta}{=} \sum_{j=0}^{\infty} \frac{\rho^j}{j!}\mathcal{M}_j^{\mathbb{P}}(A), \tag{3.3.19}$$

in which the $\mathcal{M}_j^{\mathbb{P}}(A)$ can be represented as curvature integrals. Finding explicit expressions for the $\mathcal{M}_j^{\mathbb{P}}$ via this construction is, however, generally not easy, and beyond the scope of this book. Details, for Gaussian $\mathbb{P}$, can be found in Chapter 10 of [5]. When $\mathbb{P} = \gamma_k$, the distribution of a $N(0, I_{k\times k})$ random vector, then the corresponding $\mathcal{M}_j^{\gamma_k}$ are known as the *Gaussian Minkowski functionals*.

Alternatively, given any other way for computing the probability content of a tube, so that the left hand side of (3.3.19) is known, the right hand side gives an implicit definition of the $\mathcal{M}_j^{\mathbb{P}}$, much as we originally defined the intrinsic volumes via Steiner's formula.

To see how this might work, consider the simple, one-dimensional example for which $\mathbb{P}$ is taken to be $\gamma = \gamma_1$, the distribution of a standard normal variable. For $A$ we take the semi-infinite interval $[u, \infty)$. The argument is then simple, although we need some notation first.

Let $H_n$ denote the $n$-th Hermite polynomial on $\mathbb{R}$, so that

$$H_n(x) = n! \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{(-1)^j x^{n-2j}}{j!\,(n-2j)!\,2^j}, \qquad n \geq 0,\ x \in \mathbb{R}, \tag{3.3.20}$$

[6] Note that there is a qualitative difference between (3.3.19) and the Steiner and Weyl formula (3.3.9). The latter are expansions with only a finite number of terms, whereas (3.3.19) is, in principle, and generally in practice, an infinite expansion.

Another difference lies in the fact that $\mathcal{M}_j^{\mathbb{P}}(A) \equiv 0$ for all $j < N - \dim(A)$, so that the sum in (3.3.19) actually starts at $j = N - \dim(A)$. You might want to try an example or two to see why this should be the case. For example, $A$ an interval in $\mathbb{R}^1$ is done below, but straight line in $\mathbb{R}^3$ is a little different and already instructive.

We also set

$$H_{-1}(x) \stackrel{\Delta}{=} \sqrt{2\pi} e^{x^2/2} \Psi(x),$$

where $\Psi$ is the tail probability (2.2.1) of a standard normal.

Then it is easy to check that

$$\frac{d^n}{dx^n} e^{-x^2/2} = (-1)^n H_n(x) e^{-x^2/2}, \qquad n \geq 0,\ x \in \mathbb{R}, \tag{3.3.21}$$

a fact that could also be taken as the actual definition of the Hermite polynomials.

We now turn to computing the $\mathcal{M}_j^\gamma([u,\infty))$, and proceed via a standard Taylor series expansion of the Gaussian distribution function $\Phi$:

$$\begin{aligned}
\gamma\big(\mathrm{Tube}([u,\infty),\rho)\big) &= 1 - \Phi(u-\rho) \\
&= 1 - \left(\Phi(u) + \sum_{j=1}^{\infty} \frac{(-\rho)^j}{j!} \frac{(-1)^{j-1}}{\sqrt{2\pi}} H_{j-1}(u) e^{-u^2/2}\right) \\
&= 1 - \Phi(u) + \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \frac{1}{\sqrt{2\pi}} H_{j-1}(u) e^{-u^2/2},
\end{aligned}$$

so that, on comparison with (3.3.19), we find that

$$\mathcal{M}_j^\gamma([u,\infty)) = \frac{1}{\sqrt{2\pi}} H_{j-1}(u) e^{-u^2/2}, \tag{3.3.22}$$

and we are done.

This was a particularly easy computation, but it turns out that in very many of the important cases to follow, in which $\mathbb{P} = \gamma_k$, similar arguments hold. For example, suppose that $F : \mathbb{R}^k \to \mathbb{R}$ is smooth enough for the sets $F^{-1}([u,\infty))$ to be smooth and locally convex. Suppose also that

$$\mathrm{Tube}\left(F^{-1}([u,\infty)),\, \rho\right) = F^{-1}([u-\rho,\infty)), \tag{3.3.23}$$

a relationship that we shall see holds surprisingly often in practice[7]. Then it follows immediately from the tube formula (3.3.19) that, for $A$ of this form,

$$\begin{aligned}
\mathcal{M}_j^{\mathbb{P}}(A) &= \frac{d^k}{d\rho^k} \mathbb{P}\left\{F^{-1}([u-\rho,\infty))\right\}\Big|_{\rho=0} \\
&= (-1)^k \frac{d^k}{dx^k} \mathbb{P}\left\{F(Z) \geq x\right\}\Big|_{x=u},
\end{aligned} \tag{3.3.24}$$

where $Z \sim \gamma_k$.

These kinds of sets appear often in applications, in which they take the form of rejection regions of a statistical test.

For an example, involving $\chi^2$ distributions, see Exercise 3.6.13.

[7] Alternatively, it might be the case that $\mathrm{Tube}\left(F^{-1}([u,\infty)),\, \rho\right) = F^{-1}([u+\rho,\infty))$, in which case (3.3.24) still holds, but without the alternating sign.

## 3.4 Some Integral Geometric Formulae

While the various volume of tubes formulas that we have seen so far are extremely important, they represent but a small sample of the formulas of integral geometry. Most of the others, however, require that intrinsic volumes are already defined, so we have not really wasted time or space by concentrating on tube formulas to introduce them.

The aim of this section is to describe some of these formulas. The most important of these is the *kinematic fundamental formula* (KFF). This is undoubtedly one of the most general and fundamental results in integral geometry, of which many other well known formulae are special cases or corollaries. For a full treatment of this result in a variety of scenarios you should turn to any of the classic references, including [20, 28, 33, 49, 81, 82].

The kinematic fundamental formula deals with the intersection of a fixed set with a random one. Suppose that $M_1$ and $M_2$ are two nice sets in $\mathbb{R}^n$ or $S^{n-1}_\lambda$. Then the kinematic fundamental formula is essentially a formula giving the 'average' intrinsic volumes of $M_1 \cap M_2$, as $M_2$ is moved about, and rotated, in a random fashion.

### 3.4.1 The Euclidean KFF

For the first kinematic fundamental formula we take two nice subsets $M_1$ and $M_2$ of $\mathbb{R}^n$ with finite intrinsic volumes, without being too specific about what 'nice' means. Convex sets, basic complexes and $C^2$ domains are all 'nice'.

In order to move $M_2$ 'randomly' in $\mathbb{R}^n$, we take $G_n$, the group of rigid motions, or isometry group of $\mathbb{R}^n$, and equip it with a Haar measure $\nu_n$. We normalize the Haar measure by demanding that, for any $x \in \mathbb{R}^n$ and any Borel set $A \subset \mathbb{R}^n$,

$$\nu_n\left(\{g_n \in G_n : \, g_n x \in A\}\right) = \lambda_n(A), \tag{3.4.1}$$

where $\lambda_n$, as usual, is Lebesgue measure. In other words, the set of all rigid motions that move $x$ into $A$ has the same measure as $A$ itself.

With this normalization, the kinematic fundamental formula states that

$$\begin{aligned}\int_{G_n} \mathcal{L}_i\left(M_1 \cap g_n M_2\right) \, d\nu_n(g_n) &= \sum_{j=0}^{n-i} \begin{bmatrix} i+j \\ i \end{bmatrix} \begin{bmatrix} n \\ j \end{bmatrix}^{-1} \mathcal{L}_{i+j}(M_1)\mathcal{L}_{n-j}(M_2) \\ &= \sum_{j=0}^{n-i} \frac{s_{i+1}s_{n+1}}{s_{i+j+1}s_{n-j+1}} \mathcal{L}_{i+j}(M_1)\mathcal{L}_{n-j}(M_2),\end{aligned} \tag{3.4.2}$$

where

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{[n]!}{[k]!\,[n-k]!}, \quad [n]! = n!\omega_n, \quad \omega_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}, \quad s_n = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} \tag{3.4.3}$$

The $\begin{bmatrix} n \\ k \end{bmatrix}$ are known as *flag coefficients*.

The kinematic fundamental formula is actually a generalization of the volume of tubes formulas that we have met so far. See Exercise 3.6.14.

### 3.4.2 The KFF on Spheres

What will actually be more important for us than the kinematic fundamental formula on Euclidean space, for reasons which will become clearer later, is a version of the KFF for subsets of spheres. Take $S_\lambda^{n-1}$, the sphere of radius $\lambda$ in $\mathbb{R}^n$, as our basic space, and let $M_1$ and $M_2$ be nice subsets of it. To move $M_2$ 'randomly', let $G_{n,\lambda}$ be the group of isometries (i.e. rotations) on $S_\lambda^{n-1}$.

We normalize Haar measure $\nu_{n,\lambda}$ on $G_{n,\lambda}$ so that, for any $x \in S_\lambda^{n-1}$ and every Borel $A \subset S_\lambda^{n-1}$,

$$\nu_{n,\lambda}\left(\{g_n \in G_{n,\lambda} : g_n x \in A\}\right) = \mathcal{H}_{n-1}(A), \tag{3.4.4}$$

where $\mathcal{H}_{n-1}$ is Hausdorff (surface) measure on $S_\lambda^{n-1}$, so that $\mathcal{H}_{n-1}(S_\lambda^{n-1}) = \lambda^{n-1} s_n$, with $s_n$ as in (3.4.3).

Then the kinematic fundamental formula on $S_\lambda^{n-1}$ states that

$$\begin{aligned} \int_{G_{n,\lambda}} & \mathcal{L}_i^{\lambda^{-2}}\left(M_1 \cap g_n M_2\right) d\nu_{n,\lambda}(g_n) \\ &= \sum_{j=0}^{n-1-i} \begin{bmatrix} i+j \\ i \end{bmatrix} \begin{bmatrix} n-1 \\ j \end{bmatrix}^{-1} \mathcal{L}_{i+j}^{\lambda^{-2}}(M_1) \mathcal{L}_{n-1-j}^{\lambda^{-2}}(M_2) \\ &= \sum_{j=0}^{n-1-i} \frac{s_{i+1} s_n}{s_{i+j+1} s_{n-j}} \mathcal{L}_{i+j}^{\lambda^{-2}}(M_1) \mathcal{L}_{n-1-j}^{\lambda^{-2}}(M_2), \end{aligned} \tag{3.4.5}$$

where the functionals $\mathcal{L}_i^\kappa$ are from the one parameter family defined in (3.3.18).

### 3.4.3 Crofton's Formula

In some sense, Crofton's formula is a version of the Euclidean kinematic fundamental formula when the set that is being moved around is a subspace of $\mathbb{R}^n$. It is not, however, a consequence of the kinematic fundamental formula and its consequences are somewhat different.

You may recall that, already in Chapter 1, we talked about the second intrinsic volume of a set in $\mathbb{R}^3$ as being related to its caliper diameter, which we defined by placing the set between two parallel planes (or calipers), measuring the distance between the planes, and averaging over all rotations of the set. Of course, we could have averaged over all rotations of the planes rather than the set itself, and nothing would have changed.

Crofton's formula does this in a more formal fashion, and relates all the intrinsic volumes of nice sets to properties of the cross-sections. In fact, it

goes further, since it shows that *all the intrinsic volumes can be computed by averaging the Euler characteristics of cross-sections*. Thus we were justified, at the beginning of the chapter, for singling out the Euler characteristic for special treatment.

In order to state Crofton's formula we need a little notation. In particular, we write $\mathrm{Graff}(n,k)$ for the *affine Grassmanian*, the set of $k$-dimensional affine subspaces of $\mathbb{R}^n$ (not necessarily passing through the origin).

Such subspaces can be parameterized by the angle and length of the perpendicular line between them and the origin, and this parameterization can be used to define an appropriate measure, $\lambda^n_{n-k}$, on $\mathrm{Graff}(n,k)$. See Section 12.4 of *RFG* for details.

Crofton's formula is now that, for nice $M$,

$$\int_{\mathrm{Graff}(n,n-k)} \mathcal{L}_j(M \cap V)\, d\lambda^n_{n-k}(V) = \begin{bmatrix} k+j \\ j \end{bmatrix} \mathcal{L}_{k+j}(M). \tag{3.4.6}$$

The special case $j = 0$ is generally known as *Hadwiger's formula*

$$\mathcal{L}_k(M) = \int_{\mathrm{Graff}(n,n-k)} \mathcal{L}_0(M \cap V)\, d\lambda^n_{n-k}(V). \tag{3.4.7}$$

This is the result that shows how to obtain all the intrinsic measures from the Euler characteristics of cross-sections of appropriate dimension.

## 3.5 Stratified Riemannian Manifolds

Before leaving our brief and rather selective excursion into geometry, we would be misleading you if we did emphasise how limited our treatment has been.

Throughout, we have concentrated on spaces and sets that were 'nice', which meant that they were convex, or basic complexes, or $C^2$ domains. From the point of view of geometry, these restrictions would not actually be considered severe. For example, there are entire books devoted to the geometry of the convex ring, and they are considered far from trivial. Indeed, many of results that we have discussed, but not proven, in this chapter are quite hard hard to prove even in the purely convex setting.

Nevertheless, we are missing something that is important from the points of view of both differential geometry and the study of the (random) excursion sets generated by random fields. What is missing is that all our sets, and all our measures of volume, are intrinsically Euclidean, and we have associated no truly Riemannian structure to them Although the reasons for this being important for the study of random fields will only become clearer in Chapter 4 we shall nevertheless describe at least the basic ideas now.

In usual Euclidean geometry, everything hinges on the usual Euclidean inner product $(u,v) = \sum_i u_i v_i$ between vectors $u$ and $v$. From the inner product we can compute the angle between two vectors with a common starting point,

or the length of a given vector. Although this is not the way we usually do it, the inner product is also a basic building block from which we can define the usual notions of area, volume, and so Lebesgue measure in any dimension.

In Riemannian geometry the Euclidean inner product is replaced by what is known as a *Riemannian metric*. If we are dealing with a manifold $M$, then a Riemannian metric is actually a collection of functions $\{g_t\}_{t\in M}$, which, for any $t \in M$ and any two vectors $X_t$ and $Y_t$ in the tangent space $\partial_t M$ gives a value $g_t(X_t, Y_t)$. Like the usual Euclidean metric, each $g_t$ must be non-negative definite and linear in each of its parameters.

Given a Riemannian metric on a manifold, one can use it to define notions of length, area, volume, etc. A *Riemannian manifold* is then a pair $(M, g)$; a 'physical' manifold $M$ endowed with a metric $g$. There is no uniqueness here. There are many metrics for any manifold, and some of them can be quite different. Examples abound, and, as we already intimated, we shall meet some important ones resulting from Gaussian random fields in later Chapters.

There was another problem with the geometry treated so far. This limitation is actually a very practical one, and we shall demonstrate it via *scale space problems*, some of which we shall meet in Chapters 4 and ???????. Rather than defining them twice, we give an example, based on the fMRI data of Chapter 1. There, we described data taken over a three dimensional brain, which could reasonably considered as being a $C^2$ domain in $\mathbb{R}^3$, which we denote by $B$.

However, we did not tell all of the story in Chapter 1. Due to the finite resolution of fMRI scanners, the data supposed collected at individual points are actually local averages of an underlying function $f$. Each such average is, effectively, taken over an ellipsoid. The ellipsoids are parameterized by two parameters, a point $x \in \mathbb{R}^3$ which determines the lengths of the three major axes, a direction $\theta$ on the sphere $S^2$ which gives the direction of the major axis, and another on $S^1$ which gives the direction of the minor axis. The axis lengths cannot be arbitrary, and so let us assume that they are chosen from within some three-dimensional rectangle $A$.

Thus, if what we thought we observed was the value of a function at some point $t \in B$, what we are really observing is a value of a function at a point

$$(t, x, \theta) \in B \times A \times S^2 \times S^1 \subset \mathbb{R}^9.$$

This product parameter space is certainly not a $C^2$ domain, and while it is a basic complex, it is a far more complicated one than any of the examples for which we managed to do computations earlier in the chapter.

On the other hand, it is a space that is made out of simple pieces, and this indicates to where we need to turn for a fuller theory.

Putting together the two issues described above, the approach taken in *RFG* was to base everything on parameter spaces that were stratified Riemannian manifolds, or, to be more specific, locally convex Whitney stratified Riemannian manifolds, satisfying some minor side conditions that we do not even want to hint at here.

To carefully define what these are is beyond the scope of the current book, but the idea is as follows:

- Define a manifold to be a smooth surface. Examples are given by a smooth curve (without its end points), an $l$-dimensional sphere, half of an $l$-dimensional sphere or part of a ball (but without their boundaries).
- Choose a collection of $l$-dimensional manifolds, and call their union $\partial_l M$.
- Join these unions together to form

$$M = \bigcup_{l=0}^{\dim M} \partial_l M. \tag{3.5.1}$$

Actually, we have already performed such constructions, having, for example, decomposed rectangles into their faces of various dimensions, as in (3.2.8). The only difference here is that instead of gluing together flat surfaces, edges and corners, we are now gluing together surfaces, or manifolds. The final object $M$ of (3.5.1) is a what we mean by the term *stratified manifold*.

The addition of the qualifier *Whitney* relates to rules as to what happens at the joins. Basically, the joins have to have some (very) minimal smoothness properties.

The term *locally convex* requires that the support cones (3.2.5) of $M$ all be convex. Figure 3.5.1 shows a part of the (shaded) support cones for two domains in $\mathbb{R}^2$, where the bases of the cones are the points at the base of the concavity in each domain. The smooth domain has convex support cones, while the domain with the concave cusp does not. In fact, $C^2$ domains are always locally convex. These examples are actually quite generic, since the main import of the convex support cone assumption is to exclude sharp, concave cusps. Similarly, while the $N$-cube $I^N$ is locally convex (and, indeed, convex) its boundary $\partial I^N$ is not.

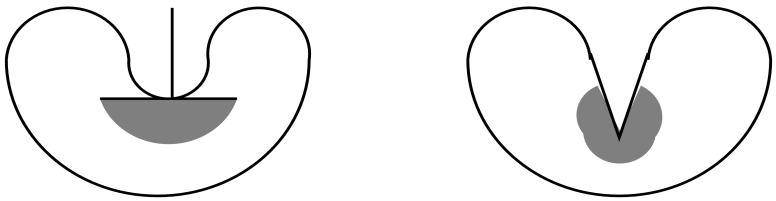

**Fig. 3.5.1.** Convex and non-convex support cones.

The last remaining qualifier was *Riemannian*, and this requires putting a Riemannian metric on each component of the stratified manifold, and then making sure that they can be joined together in a consistent fashion on the full set $M$.

How wide a class of sets is covered here? Among others, it includes

- Piecewise linear sets.
- Finite simplicial complexes.
- Riemannian polyhedra.
- Riemannian manifolds (with boundary).
- Closed semialgebraic (subanalytic) subsets of Euclidean spaces. (i.e. sets which are finite Boolean combinations of excursion sets of algebraic (analytic) functions.)
- Basic complexes with $C^2$ boundary.
- The fMRI example, $B \times A \times S^2 \times S^1 \subset \mathbb{R}^9$, above.

This then, is the framework, in which we could, and in *RFG* did, work.

Intrinsic volumes can also be defined in this framework, although they take on a slightly different meaning when the Riemannian structure is non-Euclidean. However, with or without the Riemannian structure they are generally referred to as *Lipschitz-Killing curvatures* after Rudolf Lipschitz and Wilhelm Killing, German mathematicians of the late 19-th and early 20-th centuries, respectively. The reason for the term 'curvature' can already be seen from (3.3.11), where the definition of the $\mathcal{L}_j$ for $C^2$ domains is based on integrals involving the curvature matrix.

To go into further detail would involve the heavy notation that too often makes *RFG* a challenge to read. However, we did want to give you some flavor of a more general theory, before we start looking at the remaining three parts of this book.

## 3.6 Exercises

**Exercise 3.6.1.** Suppose that $B_1, \ldots, B_n$ is a collection of unit squares, the vertices of which all sit on points of the integer lattice in $\mathbb{R}^2$, and let $A = \bigcup_{j=1}^n B_j$. Using only the 'inclusion-exclusion' description (3.1.4) of the Euler characteristic, show that

$$\varphi(A) = \#\{\text{Squares}\} - \#\{\text{Distinct sides }\} + \#\{\text{Distinct vertices}\}, \quad (3.6.1)$$

where by 'distinct' sides (vertices) we mean that a side (vertex) that appears in two different squares is counted only once.

State and prove a corresponding result for three dimensional sets made up of the union of unit cubes with vertices on the integer lattice in $\mathbb{R}^3$.

**Exercise 3.6.2.** Let $B_\lambda^N$ be the ball of radius $\lambda$ in $\mathbb{R}^N$, with boundary $S_\lambda^{N-1}$, the sphere of radius $\lambda$. Show that the outward curvature matrix (cf. Section 3.2) is given by

$$C(s) = -\frac{1}{\lambda} I_{(N-1)\times(N-1)}, \qquad \forall s \in S_\lambda^{N-1}.$$

**Exercise 3.6.3.** If $M$ is a $C^2$ domain, show that Theorem 3.1.1 is a special case of Morse's theorem, Theorem 3.2.1, by using the height function $f(t_1, \ldots, t_N) = t_N$. What additional conditions are required on the domain $M$?
Hint: Consider what happens if part of $\partial M$ lies in the plane $\{(t_1, \ldots, t_N) : t_N = 0\}$.

**Exercise 3.6.4.** Using Steiner's formula (3.3.3) show that

(i) For a $N$-dimensional cube of side length $T$ the intrinsic volumes are given by

$$\mathcal{L}_j\left([0,T]^N\right) = \binom{N}{j} T^j. \tag{3.6.2}$$

(ii) Generalize this to $N$-dimensional rectangles

$$\mathcal{L}_j\left(\prod_{i=1}^{N}[0,T_i]\right) = \sum T_{i_1} \ldots T_{i_j}, \tag{3.6.3}$$

where the sum is taken over the $\binom{N}{j}$ different choices of subscripts $i_1, \ldots, i_j$.
(iii) Let $T$ be a parallelogram in $\mathbb{R}^N$. Show that

$$\mathcal{L}_i(T) = \sum_{J:\,\dim(J)=i, 0\in J} |J|,$$

where, for an $i$-dimensional face $J$ of $T$, $|J|$ is its $i$-dimensional volume.

**Exercise 3.6.5.** Let $S_\lambda^{N-1}$ be the sphere of radius $\lambda$ in $\mathbb{R}^N$. Show, using Steiner's formula (3.3.3), that

$$\mathcal{L}_j\left(S_\lambda^{N-1}\right) \;=\; 2\binom{N}{j}\frac{\omega_N}{\omega_{N-j}}\lambda^j \;=\; 2\binom{N-1}{j}\frac{s_N}{s_{N-j}}\lambda^j \tag{3.6.4}$$

if $N-1-j$ is even, and 0 otherwise.

**Exercise 3.6.6.** Consider the $L$ and its tube in Figure 3.6.1.

(i) Calculate the intrinsic volumes of the $L$ by using the their definition for convex sets and additivity.
(ii) Compute the volume of the tube of radius $\rho$ about the $L$ for small $\rho$. For what values of $\rho$ does Steiner's formula hold for this set?
(iii) What does this say about the applicability of Steiner's formula to all non-convex sets?

**Exercise 3.6.7.** Apply Theorem 3.3.1 and Exercise 3.6.2 to find the intrinsic volumes of an $N$-dimensional ball of radius $\lambda$.

**Fig. 3.6.1.** A tube around the the letter L.

**Exercise 3.6.8.** Let $M \subset \mathbb{R}^2$ be a compact, simply connected (i.e. ball-like) $C^2$ domain in the plane. Show, using only Weyl's formula (3.3.10), that

$$\begin{aligned}\mathcal{L}_2(M) &= \lambda_2(M),\\ \mathcal{L}_1(M) &= \tfrac{1}{2} \times \text{the arclength of } \partial M.\end{aligned}$$

**Exercise 3.6.9.** Let $M \subset \mathbb{R}^3$ be a compact, simply connected (i.e. ball-like) $C^2$ domain in $\mathbb{R}^3$. Show, using only Weyl's formula (3.3.10), that

$$\begin{aligned}\mathcal{L}_3(M) &= \lambda_3(M),\\ \mathcal{L}_2(M) &= \tfrac{1}{2} \times \text{the surface area of } \partial M,\\ \mathcal{L}_1(M) &= \frac{1}{\pi}\int_{\partial M} \mathrm{Tr}(-C(s))\, ds,\end{aligned}$$

where $C$ is the outward curvature matrix (cf. Section 3.2).

**Exercise 3.6.10.** Let $F$ be a simplex in $\mathbb{R}^N$, and $G \subset F$ a simplex in $F$.

(i) Show that the normal cone is identical for all $t \in G$.
(ii) Prove (3.3.13).

**Exercise 3.6.11.** The following identity is due to Sommerville. For any simplex $F$ and any face $G \subseteq F$

$$\beta(G,F) = \sum_{H:\, G\subseteq H\subseteq F} (-1)^{\dim(H)-\dim(G)}\beta(G,H),$$

where $\beta(F,G)$ is the proportion of normal vectors to the span of $G$ that are in the normal cone.

Using this identity and (3.3.13) show that (3.3.14) holds.

**Exercise 3.6.12.** Using Hadwiger's theorem (Theorem 3.3.2) show that the following formula holds for nice $A \in \mathbb{R}^k$, $B \in \mathbb{R}^n$ and any $i \geq 0$:

$$\mathcal{L}_i(A \times B) = \sum_{j=0}^{i} \mathcal{L}_j(A)\mathcal{L}_{i-j}(B).$$

In fact, the same holds true for products of Riemannian manifolds when endowed with the product metric. We shall use this heavily in Chapter 4 when we treat specific examples of (non-Eculidean) Lipschitz-Killing curvatures.

**Exercise 3.6.13.**

(i) Find an expression for Gaussian Minkowski functionals $\mathcal{M}_j(A)$, akin to (3.3.24), for sets $A$ of the form $F^{-1}([u,\infty))$ when there is a smooth function $g : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$\text{Tube}\left(F^{-1}([u,\infty)),\, \rho\right) = F^{-1}([u - g(\rho), \infty)).$$

(Note that the special case $g(x) \equiv x$ establishes (3.3.24).)

(ii) Using the alternate version of (3.3.24) given in Footnote 7 and taking $\mathbb{P} = \gamma_k$, the law of $Z \sim N(0, I_{k\times k})$, show that, for $j \geq 1$,

$$\mathcal{M}_j^{\gamma_k}(B_\lambda^k) = \frac{d^k}{dx^k}\mathbb{P}\left\{\chi_k \leq x\right\}\Big|_{x=\lambda} \tag{3.6.5}$$

where $B_\lambda^k$ is the ball of radius $\lambda$ in $\mathbb{R}^k$ and $\chi_k$ is the square root of a $\chi_k^2$ random variable.

(iii) Using (3.3.24), compute an explicit expression for $\mathcal{M}_j^{\gamma_k}(B_\lambda^k)$ for $j \geq 1$. Hint: Look at the calculations for the $\chi^2$ example in Section 4.5.

**Exercise 3.6.14.** Under appropriate assumptions, the kinematic fundamental formula is actually a generalization of both Steiner's formula and the Weyl's volume of tubes formula. Show this in the following two cases, by taking the set $M_2$ in (3.4.2) to be a ball of radius $\rho$.

(i) Suppose that $A$ is a convex set in $\mathbb{R}^N$. Show that Steiner's formula holds for tubes of $A$ of all radii.

(ii) Suppose that $A$ is a $C^2$ domain. Show that Weyl's tube formula (3.3.10) holds for tubes of $A$ of small enough radius, and describe the local and global properties of $A$ that determine the critical radius.

# Part II

# Quantifiable Properties

# 4

# The Expected Euler Characteristic

Our aim in this chapter is to derive a formula for the mean value of a rather simple random variable, the Euler characteristic of the excursion set of a random field. That is, we are looking for

$$\mathbb{E}\left\{\varphi(A_u(f,T))\right\} = \mathbb{E}\left\{\varphi\{t \in T:\ f(t) \geq u\}\right\}.$$

We would like to do this for smooth Gaussian $f$, and also for the Gaussian related random fields that we met in Section 2.2.3. Furthermore, we should like to do this for $T$ as general as possible.

Although we shall end up doing more or less all of this, we shall only give a complete derivation for the case of isotropic, Gaussian $f$ over $N$-dimensional rectangles $T$. The stationary case will be left to you as an exercise, and then we shall call on a very general result from *RFG* to handle the remaining cases, which include non-stationary Gaussian fields and Gaussian-related fields. This general result is quite complex, but we shall show you how to apply it to a number of examples, so that by the end of this chapter you should be able to apply it to new random fields that have not yet been covered in the literature.

In this chapter we shall also see why we invested so much time discussing Lipschitz-Killing curvatures in Chapter 3. The Euler characteristic $\varphi$ itself is, of course, the same as the lowest order Lipschitz-Killing curvature $\mathcal{L}_0$ and, from many points of view, it is the most important of them. However, even if the Euler characteristic of excursion sets were all that we were interested in, it turns out that we would still need to study the remaining Lipschitz-Killing curvatures (of $T$), since they appear in the formula for $\mathbb{E}\left\{\varphi(A_u(f,T))\right\}$.

Beyond this, the other Lipschitz-Killing curvatures of excursion sets are themselves of interest. Thus, later in the chapter, in Sections 4.7 and 4.8, we shall also see how to compute their expectations as well.

Before we start in earnest, however, a brief reminder of one of the reasons we are interested in the mean Euler characteristic of excursion sets will help in motivating some of the results of this chapter. Recall that back in the Introduction, when, in Section 1.5.2, we discussed brain imaging and statistical testing, we mentioned that, in many cases, it was true that

$$\mathbb{P}\Big\{ \sup_{t\in T} f(t) \geq u \Big\} \simeq \mathbb{E}\left\{\varphi(A_u(f,T))\right\}, \tag{4.0.1}$$

an approximation that is often refered to as the *Euler characteristic heuristic.* We shall treat this in some detail in Chapter 5, but for now you should just keep it in the back of your mind as you read on.

## 4.1 Regularity Conditions

We are going to need two sets of regularity conditions in studying the Euler characteristics of excursion sets. The first will be on the random fields $f$, and these are neeeded to guarantee that the sample paths are smooth enough to be Morse functions in the sense of Definition 3.2.3, and that they are regular enough so that we can apply the Rice-Kac meta-theorem Theorem 2.7.1 when needed. The second will be on the subsets $T$ of $\mathbb{R}^N$ over which we compute Euler characteristics.

We shall introduce conditions on $T$ as needed, but as far as $f$ is concerned we already saw in Corollary 2.7.2 that the conditions required for applying the meta-theorem become quite simple in the Gaussian case, and it is not hard to see (cf. Section 11.3 of *RFG* for details) that the same holds for applying Morse Theory.

In fact, all that we require on $f$ for all the results of this chapter to hold is that, if $f$ is a real-valued, centered, Gaussian random field on $\mathbb{R}^N$ with mean square first and second order partial derivatives $f_i$ and $f_{ij}$, then

(i) With $\nabla f$ the vector of the $f_i$ and $\nabla^2 f$ the matrix of the $f_{ij}$, we require that the joint distributions of $(f(t), \nabla f(t), \nabla^2 f(t))$ are, for each $t \in T$, non-degenerate.

(ii) If $C_{ij}(s,t)$ is the covariance function of $f_{ij}$, then

$$\max_{i,j} |C_{ij}(t,t) + C_{ij}(s,s) - 2C_{ij}(s,t)| \leq K\,|\ln|t-s|\,|^{-(1+\alpha)},$$

for some finite $K > 0$, some $\alpha > 0$, and all $|t-s|$ small enough.

## 4.2 Stationary Gaussian Fields

To make the statement of our first result reasonably self-contained, we review some notation from Section 3.2. Starting with $T = \prod_{i=1}^N [0, T_i]$, a rectangle in $\mathbb{R}^N$, we write $\partial_k T$ for the collection of the $2^{N-k}\binom{N}{k}$ faces of dimension $k$ in $T$. As opposed to our previous conventions, in this chapter we take these faces as closed. Thus all faces in $\partial_k T$ are subsets of some face in $\partial_{k'} T$ for all $k' > k$.

Each $k$-dimensional face $J \in \partial_k T$ is determined by a subset $\sigma(J)$ of $\{1, \ldots, N\}$, of size $k$, corresponding to the coordinates spanning the face, and

a sequence of $N-k$ zeros and ones, which we write as $\varepsilon(J) = \{\varepsilon_j, j \notin \sigma(J)\}$, corresponding to the fixed coordinates that determine the face, so that

$$J = \left\{t \in \mathbb{R}^N : t_j = \varepsilon_j T_j, \text{ if } j \notin \sigma(J); \;\; 0 \le t_j \le T_j, \text{ if } j \in \sigma(J)\right\}. \quad (4.2.1)$$

As before, $\mathcal{O}_k$ denotes the $\binom{N}{k}$ elements of $\partial_k T$ which include the origin.

We need one more piece of notation. For a stationary random field $f$, and $1 \le i, j \le N$ let $\lambda_{ij}$ be the second order spectral moments of (2.4.4), so that

$$\lambda_{ij} = \mathbb{E}\left\{f_i(t) f_j(t)\right\}.$$

For a face $J \in \partial_k T$, write $\Lambda_J$ for the $k \times k$ matrix with elements $\lambda_{ij}$, $i, j \in \sigma(J)$.

We can now state the first expected Euler characteristic (EEC) result.

**Theorem 4.2.1 (EEC for stationary fields on rectangles).** *Let $f$ be a centered, stationary Gaussian field on $\mathbb{R}^N$ satisfying the conditions of Section 4.1 and with variance $\sigma^2$, and let $T \subset \mathbb{R}^N$ be a finite rectangle. For real $u$, let*

$$A_u = A_u(f, T) = \{t \in T : f(t) \ge u\} = T \cap f^{-1}[u, +\infty),$$

*be an excursion set, and let $\varphi$ be the Euler characteristic. Then*

$$\mathbb{E}\left\{\varphi\left(A_u\right)\right\} = e^{-u^2/2\sigma^2} \sum_{k=1}^{N} \sum_{J \in \mathcal{O}_k} \frac{|J|\,|\Lambda_J|^{1/2}}{(2\pi)^{(k+1)/2}\sigma^k} H_{k-1}\left(\frac{u}{\sigma}\right) + \Psi\left(\frac{u}{\sigma}\right), \quad (4.2.2)$$

*where $|J| = \prod_{i \in \sigma(J)} T_i$ is the usual $k$-dimensional measure of the face $J$, $H_k$ is the $k$-th Hermite polynomial* (3.3.20) *and $\Psi$ is the Gaussian tail probability* (2.2.1)*.*

We shall find other, more general, versions of (4.2.2) later on. In anticipation of these consider one special, but nevertheless illuminating case. Suppose that $f$ is also isotropic, and write $\lambda_2$ for the variance of $f_i$ (independent of $i$ by isotropy). Then each $\Lambda_J$ in (4.2.2) is a diagonal matrix with determinant $\lambda_2^{k/2}$, and so $|J|$ depends on $J$ only through its dimension. Furthermore, by Exercise 3.6.4 the sum $\sum_{J \in \mathcal{O}_k} |J|$ is precisely the Lipschitz-Killing curvature $\mathcal{L}_k(J)$ given by (3.6.3). Recalling the convention that $H_{-1}(x) \equiv \sqrt{2\pi} e^{x^2/2} \Psi(x)$, and substituting all of this into (4.2.2), allows us to rewrite it more compactly as

$$\mathbb{E}\left\{\varphi\left(A_u\right)\right\} = e^{-u^2/2\sigma^2} \sum_{k=0}^{N} \frac{\lambda_2^{k/2}}{(2\pi)^{(k+1)/2}\sigma^k} \mathcal{L}_k(T) H_{k-1}\left(\frac{u}{\sigma}\right). \quad (4.2.3)$$

As we shall see later, not only is this form[1] more compact, but it is the key to understanding far more general results, once the $\mathcal{L}_k$ are given an appropriate interpretation.

[1] If you have a good memory, this should remind you of the formula for the exceedence probability of a cosine random field, (2.3.16). As we shall see in Chapter 5, this is no coincidence.

(The graphics and discussion in pages 293-296 of RFG would be useful somewhere in this chapter.)

Now, however, let us look at the proof of Theorem 4.2.1, which begins by recalling from Theorem 3.2.4 that there is a way to write $\varphi(A_u(f,T)$ via the number of critical points of $f$ of various types on the various faces of $T$.

To write this out, we start by defining, corresponding to each sequence $\varepsilon(J)$ in (4.2.1), a set $\varepsilon^*(J)$ of $\pm 1$'s, according to the rule $\varepsilon_j^* = 2\varepsilon_j - 1$. Then, with a little rewriting, it follows from Theorem 3.2.4 that

$$\varphi\left(A_u(f,T)\right) = \sum_{k=0}^{N} \sum_{J \in \partial_k T} \sum_{i=0}^{k} (-1)^i \mu_i(J), \tag{4.2.4}$$

where, for $i \leq \dim(J)$, $\mu_i(J)$ is the number of $t \in J$ for which

$$f(t) \geq u, \tag{4.2.5}$$

$$f_j(t) = 0, \qquad j \in \sigma(J) \tag{4.2.6}$$

$$\mathrm{Ind}\left((-f_{mn}(t))_{(m,n\in\sigma(J))}\right) = i, \tag{4.2.7}$$

$$\varepsilon_j^* f_j(t) > 0, \qquad j \notin \sigma(J) \tag{4.2.8}$$

and, as usual, the index of a matrix is the number of its negative eigenvalues. (See Exercise 4.9.2 regarding the conditions needed for applying Theorem 3.2.4 here.)

Thus we can find the mean excursion set Euler characteristic by taking expectations of each term on the right of (4.2.4). We start with the term for $k = 0$, temporarily dropping the requirement (4.2.8).

**Lemma 4.2.2.** *Let $f$ and $T$ be as in Theorem 4.2.1. Let*

$$\mu_k = \#\{t \in T : f(t) \geq u,\ \nabla f(t) = 0,\ \mathrm{Ind}(-\nabla^2 f(t)) = k\}, \tag{4.2.9}$$

*be the number of critical points of index $k$ in $T$ at which $f$ exceeds the level $u$. Then*

$$\mathbb{E}\left\{\sum_{k=0}^{N} (-1)^k \mu_k\right\} = \frac{|T|\,|\Lambda|^{1/2}}{(2\pi)^{(N+1)/2}\sigma^N} H_{N-1}\left(\frac{u}{\sigma}\right) e^{-u^2/2\sigma^2}, \tag{4.2.10}$$

*where $\Lambda$ is the $N \times N$ matrix of all the second order spectral moments of $f$.*

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 via its value and second order derivatives at zero; *viz.* only through the variance and second order spectral moments. This is somewhat surprising for two reasons.

The first is that although the definition of the $\mu_k$ depends quite strongly on the $f_{ij}$, the distributions of which involve fourth order spectral moments,

these do not appear in the final expectation. 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 (4.2.10) and do not attempt to evaluate the expectations of the individual $\mu_k$. Doing so would indeed involve fourth order spectral moments. The fact that this is all we need is extremely fortunate, for it is actually *impossible* to obtain closed expressions for any of the $\mathbb{E}\{\mu_k\}$.

The second point of interest is that neither (4.2.10) nor any of the results that build on it depend in any way on the long term decay rate of the covariance function of $f$. It is not hard to see that this is an immediate consequence of the additive property (in the sense of (3.1.2)) of the random variables $\mu_k$.

**Proof.** We shall prove the lemma under the additional assumption of isotropy, so that $\Lambda$ is a diagonal matrix. Moving from isotropy to general stationarity is left to you in Exercise 4.9.4. To save on notation, we shall also assume that $\sigma = 1$, $\Lambda = I$. The extension to general $\sigma$ and $\Lambda$ is simple and is Exercise **??**.

Direct application of the Rice-Kac meta-theorem, Theorem 2.7.1, applied to each $\mu_k$ separately, yields that

$$\mathbb{E}\left\{\sum_{k=0}^{N}(-1)^k\mu_k\right\} = \sum_{k=0}^{N}\int_T (-1)^k\mathbb{E}\Big\{\left|\det(\nabla^2 f(t))\right| \mathbb{1}_{\{f(t)\geq u, \mathrm{Ind}(-\nabla^2 f(t))=k\}} \Big| \nabla f(t) = 0\Big\} p_{\nabla f(t)}(0)dt,$$

where

$$p_{\nabla f(t)}(0) = (2\pi)^{-N/2}$$

is the density of $\nabla f(t)$ at 0.

The fact that $f$ is isotropic implies that the pair $(f(t), \nabla^2 f(t))$ is independent of $\nabla f(t)$ (cf. Exercise 2.8.10) so we can remove the conditioning event. This greatly simplifies the calculation.

However, the small miracle that simplifies everything enormously (without which we would not get such a final simple answer) is the fact that

$$(-1)^k \left|\det(\nabla^2 f(t))\right| \mathbb{1}_{\{\mathrm{Ind}(-\nabla^2 f(t))=k\}} = \det(-\nabla^2 f(t))\, \mathbb{1}_{\{\mathrm{Ind}(-\nabla^2 f(t))=k\}},$$

which implies that

$$\sum_{k=0}^{N}(-1)^k|\det(\nabla^2 f(t))|\, \mathbb{1}_{\{\mathrm{Ind}(-\nabla^2 f(t))=k\}} = \det(-\nabla^2 f(t)).$$

Applying this, along with stationarity to integrate out $t$, then dropping the index $t$ on what remains, we see that

$$\begin{aligned}\mathbb{E}\left\{\sum_{k=0}^{N}(-1)^k\mu_k\right\} &= (2\pi)^{-N/2}|T|\,\mathbb{E}\left\{\det(-\nabla^2 f)\mathbb{1}_{\{f\geq u\}}\right\}\\ &= (2\pi)^{-N/2}|T|\,\mathbb{E}\left\{\det(-\nabla^2 f - fI + fI)\mathbb{1}_{\{f\geq u\}}\right\}\\ &= (2\pi)^{-N/2}|T|\,\mathbb{E}\left\{\sum_{j=0}^{N} f^{N-j}\mathrm{detr}_j(-\nabla^2 f - fI)\mathbb{1}_{\{f\geq u\}}\right\},\end{aligned}$$

where, as usual, $\mathrm{detr}_j(A)$ is the sum of determinants of all $j\times j$ principle minors of $A$. Here, we have used the standard expansion that, for an $N\times N$ matrix $A$,

$$\det(A+\lambda I) = \sum_{j=0}^{N}\lambda^j \mathrm{detr}_j(A).$$

Once again appealing to isotropy (and Exercise 2.8.10) it is simple to check that the matrix $\nabla^2 f + fI$ is independent of $f$. Therefore,

$$\mathbb{E}\left\{\sum_{k=0}^{N}(-1)^k\mu_k\right\} = (2\pi)^{-N/2}|T|\sum_{j=0}^{N}\mathbb{E}\left\{f^{N-j}\mathbb{1}_{\{f\geq u\}}\right\}\mathbb{E}\left\{\mathrm{detr}_j(-\nabla^2 f - fI)\right\}.$$

The computation of the expectation of the determinant here is left to you as Exercise 4.9.3, and applying it we find that

$$\begin{aligned}\mathbb{E}\left\{\sum_{k=0}^{N}(-1)^k\mu_k\right\} &= (2\pi)^{-N/2}|T|\,\mathbb{E}\left\{\sum_{j=0}^{\lfloor\frac{N}{2}\rfloor}\frac{(-1)^j N!}{(N-2j)!j!2^j}f^{N-2j}\mathbb{1}_{\{f\geq u\}}\right\}\\ &= (2\pi)^{-N/2}|T|\mathbb{E}\left\{H_N(f)\mathbb{1}_{\{f\geq u\}}\right\}\\ &= (2\pi)^{-(N+1)/2}|T|\int_u^\infty H_N(x)\,e^{-x^2/2}\,dx\\ &= (2\pi)^{-(N+1)/2}|T|H_{N-1}(u)\,e^{-u^2/2},\end{aligned} \tag{4.2.11}$$

where the last line follows directly from the basic properties of Hermite polynomials (cf. (3.3.21)). □

**Proof of Theorem 4.2.1.** As in the proof of Lemma 4.2.2, we shall only give a full proof under the additional assumptions of isotropy and $\sigma = 1$, $\Lambda = I$, leaving the rest to you as Exercise 4.9.4.

Consider conditions (4.2.5)–(4.2.7). If we restrict $f$ to the face $J$, then Lemma 4.2.2, modified only to allow for the dimension of $J$, actually gives the expected number of points satisfying these three conditions. However, we also have to allow for the additional constraint (4.2.8).

To do this, let $\tilde{\mu}_i(J)$ denote the number of points $t\in J$ satisfying (4.2.5)–(4.2.7) while $\mathcal{E}_J(t)$ denotes the event (4.2.8). We need to compute

$$\sum_{i=0}^{k}(-1)^i\mathbb{E}\left\{\tilde{\mu}_i(J)\mathbb{1}_{\mathcal{E}_J(t)}\right\},$$

and then sum over the faces $J$ of $T$. However, once again applying isotropy (or Exercise 2.8.10) gives us that the random variables in (4.2.8) are independent of those in (4.2.5)–(4.2.7). Therefore, by Lemma 4.2.2, when $k \geq 1$,

$$\begin{aligned}\sum_{i=0}^{k}(-1)^i\mathbb{E}\left\{\tilde{\mu}_i(J)\mathbb{1}_{\mathcal{E}_J(t)}\right\} &= \frac{|J|}{(2\pi)^{(k+1)/2}}H_{k-1}(u)e^{-u^2/2}\mathbb{P}\{\mathcal{E}_J(t)\}\\ &= \frac{|J|}{(2\pi)^{(k+1)/2}}H_{k-1}(u)e^{-u^2/2}\frac{1}{2^{N-k}},\end{aligned}$$

where the calculation of $\mathbb{P}\{\mathcal{E}_i(t)\}$ follows from symmetry considerations.

It is easy to check from first principles, using the connection between $H_{-1}$ and $\Psi$, that the above also holds when $k = 0$.

From this and (4.2.4) we finally have

$$\begin{aligned}\mathbb{E}\left\{\varphi\left(A_u(f,T)\right)\right\} &= \sum_{k=0}^{N}\sum_{J\in\partial_k T}\mathbb{E}\left\{\sum_{i=0}^{k}(-1)^i\,\tilde{\mu}_i(J)\mathbb{1}_{\mathcal{E}_J}(t)\right\}\\ &= \sum_{k=0}^{N}\sum_{J\in\partial_k T}\frac{|J|}{(2\pi)^{(k+1)/2}}H_{k-1}(u)e^{-u^2/2}\frac{1}{2^{N-k}}\\ &= \sum_{k=0}^{N}\sum_{J\in\mathcal{O}_k}\frac{|J|}{(2\pi)^{(k+1)/2}}H_{k-1}(u)e^{-u^2/2},\end{aligned}$$

where we have used the fact that, for each $J \in \mathcal{O}_k$, there are $2^{N-k}$ congruent faces of $T$ in $\partial_k T$.

This gives (4.2.2) (or, alternatively, (4.2.3)) and so the proof is complete (at least for the isotropic case with $\sigma^2 = 1$ and $\Lambda = I$.) □

## 4.3 Non-Stationary Gaussian Fields

While the conceptual move from the stationary random fields of the previous section to non-stationary fields is not a large one, the jump in the level of mathematics required is substantial.

Consider, for a moment, the crucial steps that made the proof of Theorem 4.2.1 work, as opposed to the detailed calculations that made up the bulk of the proof. For a start, we assumed that the random field was isotropic. This had the effect of making the random variables $f$ and $\nabla f$ independent, as was also true for $\nabla f$ and $\nabla^2 f$ and for $(f, \nabla^2 f)$ and $\nabla f$. Furthermore, the correlation structure between $f$ and $\nabla^2 f$ was quite simple.

This independence was crucial in terms of simplifying our computations. However, we did not need to assume isotropy for this. Stationarity would have sufficed[2], although it would have left us with a rather complicated correlation structure between $f$ and $\nabla^2 f$. If you completed the missing parts of the proof by doing Exercise 4.9.4, then you already know that the way to move from the isotropic case to the general stationary one is via a transformation, $t \to \Lambda^{1/2}t/\sigma$, of the parameter space, and that this made the transformed field locally isotropic in the sense of Section 2.4.9, in that the covariance matrix of the first order derivatives of the transformed random field was a unit matrix.

As we noted in Section 2.4.9, there is no simple transformation to local isotropy for general non-stationary fields, since one may need different transformations in different parts of the parameter space, and patching them together to make something which is globally well defined may not be easy. In fact, it is *not* easy, but under the assumption of constant variance there is a technique for doing it that works well for our needs, and this involves replacing the usual Euclidean inner product between vectors by an appropriate Riemannian metric. Thus, while we do not transform the parameter space at all, *we change the way we measure things on it.*

As we discussed briefly in Section 3.5, a Riemannian metric is defined on vectors in tangent planes of a manifold, but in our current situation, where everything is taking place in $\mathbb{R}^N$, it suffices to define it on unit vectors parallel to the axes. Thus, if $t \in T$ and we define the vector based at $t$ and parallel to the $i$-th axis as

$$\frac{\partial}{\partial t_i}(t) = t + e_i,$$

where, as usual, $e_j$ is the vector with 1 in the $i$-th position and 0 elsewhere, then a Riemannian metric which is suited to our purposes is given by

$$g_t\left(\frac{\partial}{\partial t_i}(t),\, \frac{\partial}{\partial t_j}(t)\right) = (\sigma^{-2}\Lambda(t))_{ij} \stackrel{\Delta}{=} \sigma^{-2}\,\mathrm{Cov}\left(\frac{\partial f(t)}{\partial t_i}, \frac{\partial f(t)}{\partial t_j}\right). \qquad (4.3.1)$$

We call $g_t$ the *metric induced by the random field* $f$. Note that to define this metric we have assumed that $f$ has constant variance $\sigma^2$. However, neither stationarity nor isotropy is needed for the definition to make sense.

In the isotropic and locally isotropic cases, the induced metric is the usual Euclidean metric, multiplied by a factor of $\lambda_2$, the second spectral moment. In the stationary case, for vectors $u, v \in \mathbb{R}^N$,

$$g_t(u, v) = (u\sigma^{-1}\Lambda^{1/2},\, v\sigma^{-1}\Lambda^{1/2}) = (u, v\sigma^{-2}\Lambda),$$

where the right hand inner products are the usual Euclidean one and $\Lambda$ is the usual matrix of second order spectral moments. In both the stationary and isotropic cases $g_t$ is independent of $t$.

[2] In fact, constant variance for the field and its first order derivatives would have been enough. See Section 2.4.4.

Note that the induced metric is scale independent, since if we multiply $f$ by $a$, say, then both the variance of $f$ and the covariances of its derivatves will change by a factor of $a^2$. Thus $g$ is measure of the spatial variation of $f$ only.

Recall now from Section 3.5 that Riemannian metrics determine notions of length, distance, area, volume, etc. In fact, given a Riemannian metric and a nice enough set $T$, it is also possible to define Lipschitz-Killing curvatures $\mathcal{L}_j(T)$ which correspond to the metric. Actually, we have already seen one such example when looking at the cosine random field in Section 2.3.4 (cf. (2.3.14)) where we defined generalized Lipschitz-Killing curvatures for rectangles as

$$\mathcal{L}_k\Big(\prod_1^N [0,T_j]\Big) = \sum_{j_1 \ldots j_k} \prod_{i=1}^k \frac{\lambda_{j_k} T_{j_k}}{\sigma}, \tag{4.3.2}$$

the sum being taken over the $\binom{N}{k}$ distinct choices of $j_1, \ldots, j_k$. (cf. (2.3.14).) In Section 2.3.4 the $\lambda_j$ were defined as parameters of the cosine field, but, in general, (4.3.2) is appropriate for a random field with a matrix of second order spectral moments of the form $\mathrm{diag}(\lambda_{jj}) = \mathrm{diag}(\lambda_j^2)$.

Note that, in the notation of Theorem 4.2.1, we can rewrite (4.3.2) as

$$\mathcal{L}_k\Big(\prod_1^N [0,T_j]\Big) = \sigma^{-k} \sum_{J \in \mathcal{O}_k} |J|\, |\Lambda_J|^{1/2}, \tag{4.3.3}$$

an expression that will help Theorem 4.3.1 below look more like the generalisation of Theorem 4.2.1 that it is.

Extending (4.3.2) or (4.3.3) to the case of non-isotropic and non-stationary random fields is not easy. In general, a concrete expression for the Lipschitz-Killing curvatures corresponding to a particular Riemannian metric involves the curvature tensor of the metric and so is beyond the scope of this book. You can find full details in *RFG*, with similar notation to that used here.

Ignoring for the moment the fact that, in general, these functionals are difficult to define, we can still state the following fundamental result which significantly generalises Theorem 4.2.1 and which we shall use heavily in the applications to follow.

**Theorem 4.3.1 (EEC for general Gaussian fields on general sets).** *Let $f$ be a centered Gaussian field on $\mathbb{R}^N$ satisfying the conditions of Section 4.1 and with constant variance $\sigma^2$. Then for nice*[3] *parameter spaces $T \subset \mathbb{R}^N$,*

[3] Technically, 'nice' here is defined to mean one of the locally convex Whitney stratified manifolds we discussed in Section 3.5, where we also gave a list of examples of nice sets. Recall that this list incorporated sets like $N$-dimensional rectangles and balls and anything that could be constructed from them either by smooth transformations or by glueing different sets of these kinds together.

$$\begin{aligned}\mathbb{E}\{\varphi(A_u(f,T)\} &= \sum_{k=0}^{\dim T} \mathcal{L}_k(T)\frac{H_{k-1}(u/\sigma)e^{-u^2/2\sigma^2}}{(2\pi)^{(k+1)/2}} \\ &\stackrel{\Delta}{=} \sum_{k=0}^{\dim T} \mathcal{L}_k(T)\rho_k^G(u),\end{aligned} \tag{4.3.4}$$

*where the Lipschitz-Killing curvatures $\mathcal{L}_k(T)$ are computed with respect to the Riemannian metric* (4.3.1)*.*

Our reason for introducing the functions $\rho_j^G$ in (4.3.4) ('$G$' is for 'Gaussian') will become clear later.

Now, however, a pertinent question might be how useful is the above result for someone who has neither read *RFG* nor wants to, and is interested in applications rather than the differential geometry of Riemannian manifolds. After all, we have just given a result in which the right hand side of the equation involves terms which we have not even fully defined. Beyond that, we have hinted that, as is indeed the case, that even with a proper definition the $\mathcal{L}_j(T)$ are far from easy to compute.

However, it turns out that, in practical situations, it is actually possible to estimate the $\mathcal{L}_j(T)$ of (4.3.4) from data, even when they are essentially impossible to calculate analytically. Some techniques for doing this will be described in Chapter ?????. For the moment, however, you will have to either wait patiently, or think in terms of the isotropic case, in which the $\mathcal{L}_j(T)$ of (4.3.4) are, up to a factor of $\lambda_2^{j/2}$, the simple, Euclidean Lipschitz-Killing curvatures which we saw how to compute in Section 3.3.

In the case of stationary processes, the Lipschitz-Killing curvatures $\mathcal{L}_j$ have a particularly simple form. In particular, if $\Lambda$ is the usual matrix of second order spectral moments (2.4.4), then

$$\mathcal{L}_j(T) = \mathcal{L}_j^E\big(\Lambda^{1/2}T\big), \tag{4.3.5}$$

where $\Lambda^{1/2}$ is a square root of $\Lambda$, $\Lambda^{1/2}T = \{\Lambda^{1/2}t,\ t \in T\}$, and the $\mathcal{L}_j^E$ are the standard Euclidean Lipschitz-Killing curvatures . (cf. 4.9.6.)

There is one case, however, for which (4.3.4) gives something quite simple and useful, even if $f$ is not stationary. Consider high levels $u$, a case of particular interest since then we know from the Euler characteristic heuristic (cf. (4.0.1)) that $\mathbb{E}\{\varphi(A_u)\}$ is a good approximation to an exceedence probability. Since the $H_k$ in (4.3.4) are polynomials, the leading term in $u$ there comes from the highest order polynomial. If $\dim(T) = N$, this gives

$$\mathbb{E}\{\varphi(A_u(f,T)\} \simeq \frac{\mathcal{L}_N(T)}{\sigma^{N-1}(2\pi)^{(N+1)/2}}u^{N-1}e^{-u^2/2\sigma^2}, \tag{4.3.6}$$

so that, at least as far as asymptotics are concerned, we need only know how to evaluate the highest order Lipschitz-Killing curvature corresponding to volume. It is not hard to see (cf. *RFG* for details) that

$$\mathcal{L}_N(T) = \sigma^{-N} \int_T |\det \Lambda(t)|^{1/2}\, dt, \tag{4.3.7}$$

where $\Lambda(t)$ is defined at (4.3.1).

It is also possible to derive a reasonable expression for $\mathcal{L}_{N-1}(T)$, in the case in which $T$ is a compact region in $\mathbb{R}^N$ with $C^2$ boundary $\partial T$. Then we have

$$\mathcal{L}_{N-1}(T) = \frac{1}{2\sigma^{(N-1)}} \int_{\partial T} |\det \Lambda_{\partial T}(t)|^{1/2}\, \mathcal{H}_{N-1}(dt), \tag{4.3.8}$$

where $\mathcal{H}_{N-1}$ is surface measure on $\partial T$. To define $\Lambda_{\partial T}(t)$, let $e_1(t), \ldots, e_{N-1}(t)$ be an orthonormal basis to the tangent space to $T$ at $t \in \partial T$. Then, in analogy to (4.3.1), we define

$$(\Lambda_{\partial T}(t))_{ij} = \mathrm{Cov}\left(\frac{\partial f(t)}{\partial e_i(t)}, \frac{\partial f(t)}{\partial e_j(t)}\right).$$

Note how well (4.3.7) and (4.3.8) relate to (4.3.3), even though the types of sets they describe are quite different.

## 4.4 The Gaussian Kinematic Formula, I

In the previous two sections we met two main results. Theorem 4.3.1 gave an expression for the expected Euler characteristic of the excursion sets $A_u(f, T)$ of stationary Gaussian fields on $N$-dimensional rectangles, and Theorem 4.3.1 extended this by replacing stationarity with the assumption of constant variance and by moving to fairly general parameter spaces.

In both cases we saw the dependence of the expected Euler characteristic on $T$ was via the Lipschitz-Killing curvatures of the parameter space, which, at least in the first case, we saw in Section 3.3 could be defined as coefficients in a volume of tubes expansion. In this section, we study the dependence of the expected Euler characteristic on the threshold $u$. To be more specific, if we write the excursion sets as

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

then our aim is to replace the real-valued $f$ by a random field taking values in $\mathbb{R}^k$ and $[u, \infty)$ by a general subset of $D$ of $\mathbb{R}^k$, so that the excursion sets become $f^{-1}(D)$.

We shall give a mathematical motivation for this problem later on in Section 4.8 (where we shall also explain from where the title of this section comes) but now we want to give a statistical motivation, more consistent with the applications of later chapters, and a harbinger of things to come.

Thus, consider the following rather classical statistical problem: We observe data from the linear regression model

$$Y(t) = X\beta(t) + \epsilon(t),$$

where $\beta(t) \in \mathbb{R}^p$, $X_{k\times p}$ is a design matrix that does not depend on $t$ and $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_k(t))_{t\in T}$ is an i.i.d. sample of size $k$ of some real-valued, centered, Gaussian field with constant variance $\sigma^2$. The only thing that is perhaps non-standard in this model is that we are going to let $t$ range over a general parameter space $T \subset \mathbb{R}^N$. However, for the moment fix a $t$, and suppose that we are interested in testing the hypothesis

$$H_{0,t} : \beta(t) = 0,$$

for this $t \in T$. Standard regression techniques indicate that we should reject $H_{0,t}$ if the $F$-statistic

$$\begin{aligned} f(t) &= \frac{1}{p\widehat{\sigma}^2(t)}\widehat{\beta}(t)'(X'X)\widehat{\beta}(t) \\ &= \frac{\epsilon(t)'X(X'X)^{-1}X'\epsilon(t)/p}{\epsilon(t)'(I - X(X'X)^{-1}X')\epsilon(t)/(k-p)} \\ &\stackrel{\Delta}{=} F(\epsilon(t)) \end{aligned} \tag{4.4.1}$$

is larger than some threshold $u$, where $\widehat{\sigma}^2$ and $\widehat{\beta}$ are the usual least squares estimates of $\sigma^2(t)$ and $\beta(t)$ for the chosen $t$.

If we now consider the compound hypothesis

$$H_0 = \bigcap_{t\in T} H_{0,t},$$

then, as $t$ varies over $T$ in (4.4.1), we obtain a random field of test statistics $\{f(t)\}_{t\in T}$, and a reasonable test statistic for the compound hypothesis would be $\sup_T f(t)$, as described in Chapter 1. To apply this we need to know, at least for large $u$, how to compute the exceedence probabilities $\mathbb{P}\{\sup_T f \geq u\}$.

However, the event $\{f(t) \geq u\} = \{F(\epsilon(t)) \geq u\}$ can be written as $\{\epsilon(t) \in F^{-1}[u, +\infty)\}$, and so, applying the Euler characteristic heuristic (4.0.1) we have

$$\begin{aligned} \mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\} &\simeq \mathbb{E}\left\{\varphi(A_u(f,T))\right\} \\ &= \mathbb{E}\left\{\varphi(T \cap f^{-1}[u,+\infty))\right\} \\ &= \mathbb{E}\left\{\varphi(T \cap \epsilon^{-1}(F^{-1}[u,+\infty)))\right\}. \\ &= \mathbb{E}\left\{\varphi\left(T \cap \epsilon^{-1}D\right)\right\} \end{aligned}$$

with $D \stackrel{\Delta}{=} F^{-1}[u, \infty)$.

Thus, for random fields of test statistics constructed out of an underlying i.i.d. sample of Gaussian random fields, which we met first as Gaussian related random fields in Section 2.2, the expected Euler characteristic of excursion sets has a lot to do with computing the probabilities of rejection regions.

Here then is a result that is general enough to cover the cases described above. It is the main tool that we shall use to examine a large number of important examples in the remainder of this chapter and, indeed, in the applications that follow later. We refer to it as the Gaussian kinematic formula [91, 92] (GKF), a name that we shall explain in Section 4.8.

**Theorem 4.4.1 (Gaussian kinematic formula: Basic form).** *Let $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_k(t))$ be a $\mathbb{R}^k$-valued Gaussian random field on $\mathbb{R}^N$, with independent components each of which satisfies the conditions of Section 4.1 and has constant unit variance. Suppose that both $T \subset \mathbb{R}^N$ and $D \subset \mathbb{R}^k$ are nice (cf. Footnote 3). Then*

$$\mathbb{E}\left\{\varphi\left(T \cap \epsilon^{-1}D\right)\right\} = \sum_{j=0}^{\dim T} (2\pi)^{-j/2} \mathcal{L}_j(T) \mathcal{M}_j^{\gamma_k}(D). \tag{4.4.2}$$

*where $\gamma_k$ is the law of a $N(0, I_{k\times k})$ variable, the functionals $\mathcal{M}_j^{\gamma_k}(D)$ are defined implicitly in* (3.3.19)*, and the Lipschitz-Killing curvatures $\mathcal{L}_j$ are defined with respect to the Riemannian metric induced by the $\epsilon_i$, as described in Section 4.3.*

While Theorem 4.4.1 is stated in somewhat wider generality than needed for the statistical motivation that preceeded it, in this book we shall almost always apply it when $D$ belongs to a family of rejection regions determined by some test statistic. This motivates us to adopt new notation and terminology, that was already anticipated in the simplest one-dimensional case in (4.3.4).

**Definition 4.4.2.** *Suppose $F : \mathbb{R}^k \to \mathbb{R}$ is such that, for each $u \in \mathbb{R}$, the set $F^{-1}[u, \infty)$ is nice*[4]*. Then the 'EC densities' of the random field $f(t) = F(\epsilon(t))$ are defined as the functions*

$$\rho_j^*(u) = (2\pi)^{-j/2} \mathcal{M}_j^{\gamma_k}(F^{-1}[u, \infty)). \tag{4.4.3}$$

*The asterisk in the $\rho_j^*$ is usually taken to be a descriptor of the random variable $F(Z)$ with $Z \sim N(0, I_{k\times k})$.*

With this definition, we can rewrite the general result (4.4.2) for this situation as

$$\mathbb{E}\left\{\varphi\left(A_u(f, T)\right)\right\} = \mathbb{E}\left\{\varphi\left(T \cap \epsilon^{-1}\left(F^{-1}[u, \infty)\right)\right)\right\} = \sum_{j=0}^{\dim T} \mathcal{L}_j(T) \rho_j^*(u). \tag{4.4.4}$$

[4] Although we have already defined 'nice' once before in Footnote 3, we are not going to be so rigid as to demand that 'nice' means the same thing every time it appears. For example, we shall soon apply this theorem, in Section 4.5.4, to the rejection region for an $F$ statistic. In that case the sets $F^{-1}[u, \infty)$ are never locally convex, having a singularity at the origin. However, the Gaussian kinematic formula still holds for $F$ random fields in the sense that the terms in the expansion (3.3.19) can still be used to compute the EC densities of an $F$ random field.

In the simplest of cases, in which $k = 1$ and $F(x) = x$, we are obviously back in the purely Gaussian scenario, and so with $* = G$ (for Gaussian) it is now clear from where we took the notation of (4.3.4). In a moment we shall see how to retrieve (4.3.4) from the Gaussian kinematic formula.

However, before we do this, there is something that is very important to note about the structure of (4.4.2) and (4.4.4), and this is the way that information on the various 'parameters' of the problem factors in the final answer. There are essentially three parameters: the parameter space $T$, the covariance structure of the random field $\epsilon$, and the transformation $F$. On the one hand, as we have already seen, the first two of these parameters appear together in the $\mathcal{L}_j(T)$, describing the geometry of $T$ as seen in a Riemannian sense through the metric induced by $\epsilon$. On the other hand, the transformation $F$ affects only the EC densities $\rho_j^*$, which are unaffected by either $T$ or the distribution of $\epsilon$.

The impact of this factorisation is very significant for computations, since it means we can separate the geometry of $T$ from that of $D$ or $F^{-1}[u, \infty)$. We have already spent some time on the geometry of $T$, discussing how to compute the $\mathcal{L}_j(T)$, at least in the simple Euclidean scenarios corresponding to locally isotropic random fields. We shall return to these calculations for some non-isotropic, and, indeed, non-stationary examples in Section 4.6. Now, however, we leave these considerations for a while and concentrate on the geometry of $D$, via the EC densities.

## 4.5 Euler Characteristic Densities

The aim of this section will be to show that although the EC densities described above actually represent quite abstract geometric quantities, they are generally not too hard to compute. We shall show this by considering a number of examples, via a number of techniques, occasionally applying more than one technique to a specific case. We shall not attempt to give much more than minimal motivation for our choice of examples at this stage. Later in the book, when we turn to applications, you will see many of them being applied in various scenarios.

Note that in working with EC densities the distributional properties of the random fields that are our central concern actually play a rather minimal rôle. We shall assume only that $\epsilon$ is a smooth Gaussian field from $T$ to $\mathbb{R}^k$, with independent components, all of which have mean zero and constant unit variance. The only parameter left is therefore the transformation $F : \mathbb{R}^k \to \mathbb{R}$, by which we define the random field $f = F \circ \epsilon$.

### 4.5.1 Gaussian Fields

In the simple Gaussian case $k = 1$ and $F$ is the identity function. Thus, up to a power of $2\pi$ the EC densities are the coefficients $\mathcal{M}_j^\gamma([u, \infty))$ in the tube

formula

$$\gamma\big(\mathrm{Tube}([u,\infty),\rho)\big) = \sum_{j=0}^{\infty} \frac{\rho^l}{j!} \mathcal{M}_j^{\gamma}([u,\infty)).$$

However, we already computed these coefficients as our first example of Gaussian tube formulas (cf. (3.3.22)), to find that

$$\mathcal{M}_j^{\gamma}([u,\infty)) = \frac{1}{\sqrt{2\pi}} H_{j-1}(u) e^{-u^2/2},$$

where the $H_j$ are Hermite polynomials. Thus the EC densities for Gaussian processes are given by

$$\rho_j^G(u) = \frac{1}{(2\pi)^{(j+1)/2}} H_{j-1}(u) e^{-u^2/2}.$$

Theorem 4.3.1 is now an immediate consequence of the Gaussian kinematic formula (4.4.2).

### 4.5.2 $\chi_k^2$ Fields

An example that we shall meet often in later applications is the $\chi_k^2$ random field defined as

$$f(t) = \chi_k^2(t) = \sum_{j=1}^{k} \epsilon_j^2(t).$$

Thus, the $F$ of our transformation is the squared norm $F(x) = |x|^2$.

To compute the EC densities for this example, we shall use the general result (3.3.24) which, in the current case, becomes, for $j \geq 1$,

$$\begin{aligned}
(2\pi)^{j/2} \rho_j^{\chi_k^2}(u) &= \mathcal{M}_j^{\gamma_k}\left(F^{-1}([u,\infty))\right) \qquad (4.5.1)\\
&= \frac{d^j}{d\rho^j} \mathbb{P}\left\{F^{-1}([u-\rho,\infty))\right\}\Big|_{\rho=0}\\
&= (-1)^j \frac{d^j}{dx^j} \mathbb{P}\left\{\chi_k^2 \geq x\right\}\Big|_{x=u}\\
&= (-1)^j \frac{d^j}{dx^j} \mathbb{P}\left\{\chi_k \geq x\right\}\Big|_{x=\sqrt{u}}\\
&= (-1)^{j-1} \frac{d^{j-1}}{dx^{j-1}} p_k(x)\Big|_{x=\sqrt{u}},
\end{aligned}$$

where $p_k$ is the density of a $\chi_k$ random variable[5] (i.e. the square root of a $\chi_k^2$ random variable).

[5] Of course, one has to check that condition (3.3.23) holds for this case, but you have effectively already shown that in Exercise 3.6.13.

Elementary computations give this probability density as

$$p_k(x) \;=\; \frac{1}{\Gamma(k/2)2^{(k-2)/2}} x^{k-1} e^{-x^2/2}.$$

Direct calculations, exploiting the basic property (3.3.21) of Hermite polynomials, show that

$$\begin{aligned}\frac{d^{j-1}p_k(x)}{dx^{j-1}} &= \frac{1}{\Gamma(k/2)2^{(k-2)/2}} \sum_{i=0}^{j-1} \binom{j-1}{i} (-1)^i \frac{d^{j-1-i}x^{k-1}}{dx^{j-1-i}} H_i(x) e^{-x^2/2} \\ &= \frac{e^{-x^2/2}}{\Gamma(k/2)2^{(k-2)/2}} \sum_{i=0}^{j-1} \binom{j-1}{i} (-1)^i \frac{d^{j-1-i}x^{k-1}}{dx^{j-1-i}} H_i(x).\end{aligned}$$

The summation can be rewritten as

$$\begin{aligned}&\sum_{i=0}^{j-1} \mathbb{1}_{\{k\ge j-i\}} \binom{j-1}{i} (-1)^i \frac{(k-1)!}{(k+i-j)!} x^{k+i-j} H_i(x) \\ &= x^{k-j} \sum_{i=0}^{j-1} \sum_{l=0}^{\lfloor i/2 \rfloor} \mathbb{1}_{\{k\ge j-i\}} \binom{j-1}{i} (-1)^{i+l} \frac{(k-1)!}{(k+i-j)!} \frac{i!}{(i-2l)!l!2^l} x^{2i-2l} \\ &= x^{k-j} \sum_{l=0}^{\lfloor \frac{j-1}{2} \rfloor} \sum_{i=2l}^{j-1} \mathbb{1}_{\{k\ge j-i\}} \binom{j-1}{i} (-1)^{i+l} \frac{(k-1)!}{(k+i-j)!} \frac{i!}{(i-2l)!l!2^l} x^{2i-2l} \\ &= x^{k-j} \sum_{l=0}^{\lfloor \frac{j-1}{2} \rfloor} \sum_{m=0}^{j-1-2l} \mathbb{1}_{\{k\ge j-m-2l\}} \binom{k-1}{j-1-m-2l} \frac{(-1)^{m+l}(j-1)!}{m!l!2^l} x^{2m+2l}.\end{aligned}$$

Note that as formidable as the above expression may look, it is no more than a polynomial in $x$.

Combining the above with the definition of the EC densities now easily leads to

**Theorem 4.5.1.** *The EC densities of the $\chi^2_k$ random field are given, for $j \ge 1$ and $u > 0$, by*

$$\begin{aligned}\rho_j^{\chi^2_k}(u) = \frac{u^{(k-j)/2} e^{-u/2}}{(2\pi)^{j/2}\Gamma(k/2)2^{(k-2)/2}} &\sum_{l=0}^{\lfloor \frac{j-1}{2} \rfloor} \sum_{m=0}^{j-1-2l} \\ &\times \mathbb{1}_{\{k\ge j-m-2l\}} \binom{k-1}{j-1-m-2l} \frac{(-1)^{j-1+m+l}(j-1)!}{m!l!2^l} u^{m+l}.\end{aligned} \tag{4.5.2}$$

Note that, when $j = 0$, $\rho_0^{\chi^2_k}(u)$ is simply $\mathbb{P}\left\{\chi^2_k \ge u\right\}$. Since this connection between $\rho_0$ and tail probabilities will always hold (cf. (4.5.1)) we shall not mention it explicitly in dealing with the remaining examples.

There are two important lessons to take home from the above example. The first is that although some of the formulas involved may have been long, there was nothing at all conceptually difficult in deriving them. In fact, despite the essentially geometric character of the $\rho^{\chi^2_k}$, all that was involved in their computation was simple calculus.

The second is that it is worthwhile noting that there is another approach to obtaining the $\chi^2_k$ EC densities, one that links both the $\chi^2_k$ random field and these densities to Gaussian ones. To see this, note first that we can write

$$f(t) = \sum_{j=1}^{k} \epsilon_j^2(t) = \sup_{v \in \mathbb{R}^k : |v|=1} \langle v, \epsilon(t) \rangle^2 \stackrel{\Delta}{=} \sup_{v \in \mathbb{R}^k : |v|=1} Z^2(t,v), \qquad (4.5.3)$$

where $Z$ is now a *Gaussian* random field on $T \times S^{k-1}$. It is easy to see (cf. Exercise 3.6.12) that

$$\varphi\left(\{t \in T : f(t) \geq u\}\right) = \tfrac{1}{2}\varphi\left(\left\{(t,v) \in T \times S^{k-1} : Z(t,v) \geq \sqrt{u}\right\}\right),$$

the factor of $1/2$ coming from the symmetry $Z^2(t,v) = Z^2(t,-v)$, and so taking expectations on both sides should yield $\chi^2_k$ EC densities in terms of Gaussian ones[6].

We shall not do this now, but the computation will be implicit in the calculations of the following example. Doing so[7] would yield a somewhat more compact form of (4.5.2) as

$$\rho_j^{\chi^2_k}(u) = \sum_{l=0}^{k-1} \mathcal{L}_l(S^{k-1}) \rho^G_{j+l}(\sqrt{u}), \qquad (4.5.4)$$

where we calculated the Lipschitz-Killing curvatures of the unit ball $S^{k-1}$ back in Section 3.3 (cf. (3.3.7)).

Another important consequence of (4.5.3) lies in the trivial observation that

$$\sup_{t \in T} f(t) = \sup_{(t,v) \in T \times S^{k-1}} Z^2(t,v) = \sup_{(t,v) \in T \times S^{k-1}} \left(Z_+(t,v)\right)^2, \qquad (4.5.5)$$

so that exceedence probabilities (at $u$) for the non-Gaussian $f$ are immediately computable from those for the Gaussian $Z$ (at $\sqrt{u}$), albeit on a slightly more

[6] This approach, of enlarging the parameter space and then taking a supremum over the added part, is a standard result in the study of Banach spaces, which states that the norm of an element $x$ is equivalent to the supremum of $\langle x, y \rangle$ over all $y$ of norm one in the dual space. Within the statistical literature it is often referred to as an application of 'Roy's intersection-union principle' having been applied by Roy [79] as a heuristic method of test construction for multiple hypotheses.

[7] To prove this, you can use (4.5.17) below and (4.5.1) above. However, it would be a good test of your understanding to try to do this using only what you know so far. This is Exercise 4.9.8.

complicated parameter space. We shall return to this point in Chapter 5 (cf. Section 5.3.2) when we look at exceedence probabilities in some detail.

In general, whereas for the $\chi^2_k$ process this approach seems, at the moment, to do little more than provide a novel diversion, in the following example we shall see that it can sometimes be the natural way to proceed.

### 4.5.3 $\bar{\chi}^2_K$ Fields

The $\bar{\chi}^2_K$ fields are random fields, with distributions depending on a parameter $K$, a cone in some Euclidean space. The simplest examples arise when $K$ is the positive orthant of $\mathbb{R}^k$, for some $k \geq 1$, in which case the field is of the form

$$f(t) \;=\; \sum_{j=1}^{k} \epsilon_j^2(t) \mathbb{1}_{\{\epsilon_j(t) \geq 0\}}, \tag{4.5.6}$$

These random fields arise in regression applications of random fields where some coefficients are assumed to be zero under the null hypothesis and positive under the alternative hypothesis, and (4.5.6) corresponds to an example in which a design matrix had orthogonal columns.

More generally, a $\bar{\chi}^2_k$ random variable is determined by a $k$-dimensional convex cone $K$ and can be expressed as

$$\bar{\chi}^2_K(Z) = \left( \sup_{v \in K : |v| = 1} \langle v, Z \rangle^+ \right)^2, \qquad Z \sim N(0, I_{k \times k})$$

where $x^+$ indicates the positive part of $x$. It is well known (e.g. [88]) that the distribution of a $\bar{\chi}^2_K$ random variable is that of a mixture of $\chi^2_j$ random variables for $0 \leq j \leq k$, the $j = 0$ case corresponding to the situation in which the supremum is negative, so the $\bar{\chi}^2_K$ random variable is zero.

The tail of the distribution function, therefore, can be written as

$$\mathbb{P}\left\{\chi^2_K(Z) \geq u\right\} = \sum_{j=1}^{k} p_j(K)\, \mathbb{P}\left\{\chi^2_j \geq u\right\} \tag{4.5.7}$$

for some weights $p_j(K)$ depending on the cone $K$. If $K$ is polyhedral, these weights have a simple interpretation: they are the probability that the projection of $Z \sim N(0, I_{k \times k})$ onto $K$ lies in a face of dimension $j$ of $K$.

A natural definition of a $\bar{\chi}^2_K$ random field is therefore

$$f(t) = \left( \sup_{v \in K : |v| = 1} \langle v, \epsilon(t) \rangle^+ \right)^2.$$

There are at least two approaches for computing the EC densities of this random field. We shall describe both of them, both to indicate the techniques

involved and then to obtain additional information that comes as a byproduct of the Gaussian kinematic formula.

The first approach follows the lines that we adopted for the Gaussian and $\chi^2_k$ examples. As usual, the first step lies in checking that (3.3.23) holds. This follows from Exercise 4.9.9, which shows that if we write the rejection region for a $\bar{\chi}^2_K$ random variable as

$$R_{K,u} = \Big\{ x \in \mathbb{R}^k : \sup_{v \in K : |v| = 1} \langle v, \epsilon(t) \rangle^+ \geq \sqrt{u} \Big\}$$

then

$$\mathrm{Tube}(R_{K,u}, \rho) = R_{K,(\sqrt{u}-\rho)^2}.$$

That is, a tube around the rejection region lies in the same family of rejection regions, and so we can use the arguments that we have used so far, based on a Taylor expansion of the distribution function. However, it then follows immediately from (4.5.7) that the EC densities of of $\bar{\chi}^2_K$ random fields have the following representation in terms of those of $\chi^2_k$ fields:

$$\rho_j^{\bar{\chi}^2_K}(u) = \sum_{l=0}^{k} p_l(K)\, \rho_j^{\chi^2_l}(u). \tag{4.5.8}$$

This completes the first approach. Note, however, that in order to use this result we need to know what the $p_l(K)$ are, and we have actually only described what they are for polyhedral $K$.

The second approach uses the representation of $\bar{\chi}^2_K$ as a supremum, a technique that will also appear in later examples, where it will be the most natural approach.

Note first that if $f$ is a $\bar{\chi}^2_K$ random field then

$$\begin{aligned} \sup_{t \in T} f(t) &= \sup_{t \in T} \Big( \sup_{v \in K : |v| = 1} \langle v, \epsilon(t) \rangle^+ \Big)^2 \\ &= \sup_{(x,v) \in T \times K : |v| = 1} \big( \langle v, \epsilon(t) \rangle^+ \big)^2 \\ &\stackrel{\Delta}{=} \sup_{(t,v) \in T \times K : |v| = 1} \big( (Z(t,v))^+ \big)^2 . \end{aligned}$$

Therefore, for $u \geq 0$,

$$\sup_{t \in T} f(t) \geq u \quad \Longleftrightarrow \quad \sup_{(t,v) \in T \times K : |v| = 1} Z(t,v) \geq \sqrt{u}, \tag{4.5.9}$$

linking suprema of the highly non-Gaussian $f$ on the simple parameter space $T$ to the supremum of the Gaussian $Z$ on the somewhat more complicated space $T \times \{v \in K : |v| = 1\}$.

This connection goes further than merely linking suprema. Applying Exercise 3.6.12 with $j = 0$ and noting that the Euler characteristic of the convex $K$ is one, we have that

$$\varphi\big(\{t \in T : f(t) \geq u\}\big) = \varphi\big(\big\{(t,v) \in T \times \{v \in K : |v| = 1\} : Z(t,v) \geq \sqrt{u}\big\}\big),$$

from which it follows that the corresponding expectations are identical as well.

From the definition of the EC densities $\rho_j^{\bar{\chi}_K^2}$ we know that

$$\mathbb{E}\left\{\varphi\big(\{t \in T : f(t) \geq u\}\big)\right\} = \sum_{j=0}^{\dim T} \mathcal{L}_j(T)\rho_j^{\bar{\chi}_K^2}(u), \tag{4.5.10}$$

where the $\mathcal{L}_j$ are now computed with respect to the metric induced by $\epsilon$.

On the other hand, applying Theorem 4.4.1 to the zero mean, constant variance Gaussian process $Z$ we have that

$$\begin{aligned}\mathbb{E}\big\{\varphi\left(\big\{(t,v) \in T \times \{v \in K : |v| = 1\} : Z(t,v) \geq \sqrt{u}\big\}\right)\big\} \qquad (4.5.11)\\ = \sum_{j=0}^{\dim(T\times\{v\in K:|v|=1\})} \mathcal{L}_j(T \times \{v \in K : |v| = 1\})\,\rho_j^G(\sqrt{u}).\end{aligned}$$

The $\rho_j^G$ are the EC densities (4.3.4) for the Gaussian case and the $\mathcal{L}_j$ are computed with respect to the Riemannian metric induced on $T \times \{v \in K : |v| = 1\}$ by the random field $Z$.

However, these Lipschitz-Killing curvatures can be simplified somewhat. An additional application[8] of Exercise 3.6.12 implies that

$$\mathcal{L}_j(T \times \{v \in K : |v| = 1\}) = \sum_{l=0}^{j} \mathcal{L}_l(T)\mathcal{L}_{j-l}\left(\{v \in K : |v| = 1\}\right). \tag{4.5.12}$$

Now, however, while the $\mathcal{L}_j(T)$ are still computed with respect to the Riemannian metric induced on $T$ by $\epsilon$, the remaining Lipschitz-Killing curvatures are computed with respect to the standard Euclidean metric on the unit sphere $S^{k-1}$.

Putting together (4.5.10)–(4.5.12) gives us the following result.

**Theorem 4.5.2.** *The EC densities of the $\bar{\chi}_K^2$ random field are given, for $j \geq 1$ and $u > 0$, by either*

$$\rho_j^{\bar{\chi}_K^2}(u) = \sum_{l=0}^{k-1} \mathcal{L}_l(K_1)\,\rho_{l+j}^G(\sqrt{u}), \tag{4.5.13}$$

*where $K_1 = \{v \in K : |v| = 1\} = K \cap S^{k-1}$ and the $\mathcal{L}_l(K_1)$ are the usual, Euclidean, Lipschitz-Killing curvatures, or by*

$$\rho_j^{\bar{\chi}_K^2}(u) = \sum_{l=0}^{k} p_l(K)\,\rho_j^{\chi_l^2}(u). \tag{4.5.14}$$

[8] Note that to apply Exercise 3.6.12 we need the version for Riemannian manifolds equipped with a product Riemannian metric. Our assumptions guarantee that the random field $Z(t,v)$ generates an appropriate structure.

The two expressions in the theorem for the EC densities of the $\bar{\chi}^2_K$ random field imply a relation between the $p_j(K)$ and the $\mathcal{L}_j(K_1)$. In particular, some algebra based on them and the explicit expressions for $\rho_j^G(\sqrt{u})$ and $\rho_j^{\chi^2_l}(u)$ shows that

$$p_j(K) = \frac{1}{2^j \pi^{\frac{j-1}{2}} \Gamma(\frac{j+1}{2})} \sum_{m=0}^{\lfloor (k-j)/2 \rfloor} \frac{(-1)^m (d+2m)!}{(4\pi)^m m!} \mathcal{L}_{j+2m-1}(K_1). \quad (4.5.15)$$

Thus the $p_j(K)$, which to this point were undefined, now have an explicit representation in terms of the Lipschitz-Killing curvatures of $K \cap S^{k-1}$.

Another interesting set of identities results from taking $K = \mathbb{R}^k$, in which case the $\bar{\chi}^2_K$ is actually $\chi^2_k$. Then (4.5.13) implies that the distribution of a $\chi_k$ random variable can be written as

$$\mathbb{P}\{\chi_k \geq u\} = \sum_{j=0}^{k-1} \mathcal{L}_j(S^{k-1}) \rho_j^G(u), \quad (4.5.16)$$

a result that we shall need later, and also that

$$\frac{\partial^l}{\partial u^l} \mathbb{P}\{\chi_k \geq u\} = (-1)^l \sum_{j=0}^{k-1} \mathcal{L}_j(S^{k-1}) (2\pi)^{l/2} \rho_{j+l}^G(u) \quad (4.5.17)$$

### 4.5.4 *F* Fields

We now consider perhaps the most important random field for statistical applications, the $F$ random field.

As in the previous sections, we observe $k$ i.i.d. samples, $\epsilon_i(t)$, $1 \leq i \leq k$, of some centered, unit variance random field. In the current setting, however, we take $n, m \geq 1$ with $n + m = k$, and define the $F$ random field with $n$ degrees of freedom in the numerator and $m$ in the denominator as

$$f(t) = H(\epsilon(t)) = \frac{\sum_{i=1}^n \epsilon_i^2(t)/n}{\sum_{i=n+1}^k \epsilon_i^2(t)/m} \stackrel{\Delta}{=} \frac{|U(\epsilon(t))|^2/n}{|L(\epsilon(t))|^2/m} \quad (4.5.18)$$

where

$$U(y) = (y_1, \ldots, y_n), \qquad L(y) = (y_{n+1}, \ldots, y_k),$$

are the coordinates corresponding to the numerator and denominator, respectively.

Some care must be taken to avoid points $x \in T$ for which the random field above is not defined due to a singularity of the type 0/0 [99]. However, for this to happen, the random field $\epsilon$ must hit the origin in $\mathbb{R}^k$. It is not hard to see[9] that if $k > \dim(T)$ then this will happen with probability zero, and so

[9] See Exercise 2.8.26 for a similar result, or Lemma 11.2.10 of *RFG* for a full argument.

can be ruled out. Thus for the remainder of this section we shall assume that $k > \dim(T)$.

Following the lead of the last three examples, we would like to use the Gaussian tube formula expansion in order to compute the EC densities. However, there is an intrinsic problem, for the rejection regions $H^{-1}[u, \infty)$ of the $F$ field, where $H$ is given by (4.5.18), are not 'nice'. While we have never been terribly precise in this book about what we mean by 'nice', we have emphasised that nice sets were locally convex, and this time this is not the case.

We need only consider the case $n = m = 1$, in which case $H^{-1}[u, \infty)$ is the double cone in $\mathbb{R}^2$ enclosed by the lines $L = \pm U/\sqrt{u}$. This is concave at $L = U = 0$. (The triangle marked $R_u$ in Figure 4.5.1 below is part of the intersection of this region with the positive quadrant $\mathbb{R}^2_+$.)

In order to get around this problem, and so be able to start the derivation of the EC densities, we need another way of describing tubes around $H^{-1}[u, \infty)$, which we develop in the following lemma.

**Lemma 4.5.3.** *In the above notation, let* $R_u \stackrel{\Delta}{=} H^{-1}[u, \infty)$ *denote the rejection region of the statistic*

$$\frac{|U(x)|^2}{|L(x)|^2}, \qquad x \in \mathbb{R}^k,\ k = n + m, \tag{4.5.19}$$

*at level* $u$. *Then,*

$$x \in \mathrm{Tube}(R_u, \rho) \iff |U(x)| \geq |L(x)|\sqrt{u} - \rho\sqrt{1 + u}, \tag{4.5.20}$$

*for all* $x$ *outside*

$$\Big\{x :\ |L(x)| \leq \rho\sqrt{\frac{u}{1+u}},\ |U(x)| \leq \sqrt{u}\,|L(x)|\Big\}. \tag{4.5.21}$$

*Furthermore, for fixed* $u$, *the exceptional set of* $x$*'s has Lebesgue measure, and so probability, of order* $O(\rho^k)$.

**Proof.** Note that, to ease the notation of the proof, the ratio in (4.5.19) is not $F$ distributed, since the normalisation by $n$ and $m$ are missing.

The main idea of the proof is captured in Figure 4.5.1, which tells the entire story for $n = m = 1$, and shows a two dimensional cross-section of the story in higher dimensions. However, we shall still give a full formal proof.

We consider two cases: $x \in R_u$ and $x \notin R_u$. For $x \in R_u$ it is trivial to check that (4.5.20) holds, so we concentrate on $x \notin R_u$. We begin with the direction ($\Leftarrow$). We need to check whether $x \in \mathrm{Tube}(R_u, \rho)$, that is, to determine whether there is some point $y$ within a distance $\rho$ from $x$ which satisfies $|U(y)|/|L(y)| \geq \sqrt{u}$. Let

$$\eta_U(x) = \frac{1}{\sqrt{U(x)}} \left(\epsilon_1(x), \ldots, \epsilon_n(x), 0, \ldots, 0\right),$$
$$\eta_L(x) = \frac{1}{\sqrt{L(x)}} \left(0, \ldots, 0, \epsilon_{n+1}(x), \ldots, \epsilon_k(x)\right),$$

be the unit direction vectors for the numerator and denominator, respectively, and define

$$\eta^u(x) = \frac{1}{\sqrt{1+u}} \left(-\eta_U(x) + \sqrt{u} \cdot \eta_L(x)\right). \tag{4.5.22}$$

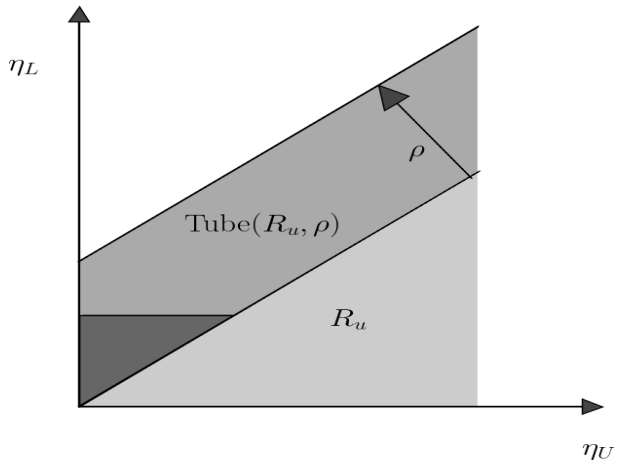


**Fig. 4.5.1.** A cross-section of $\mathrm{Tube}(R_u, \rho)$ related to the $F$-statistic in one quadrant. The lightly shaded region is part of the rejection region, while the other two are part of $\mathrm{Tube}(R_u, \rho) \setminus R_u$. The arrow indicates a displacement $\rho$ in the direction of the vector $\eta^u(x)$ of (4.5.22). The small dark triangle, whose rightmost point has coordinates $(\rho u/\sqrt{1+u},\ \rho\sqrt{u/(1+u)})$, represents the part of the tube excluding $R_u$ where the right hand side of (4.5.20) fails as a test of $x \in \mathrm{Tube}(R_u, \rho)$. The union of the dark triangles over all possible values of $U$ and $L$ is the set (4.5.21) and has Lebesgue measure, and hence probability, $O(\rho^k)$.

It is not hard to check that the normalized gradient of $|U(x)|^2/|L(x)|^2$ is just

$$\frac{1}{\sqrt{1+|U(x)|^2/|L(x)|^2}} \left(\eta_U(x) - \sqrt{\frac{|U(x)|^2}{|L(x)|^2}} \cdot \eta_L(x)\right).$$

Thus the outward pointing normal vector field along the boundary of $R_u$ is simply $\eta^u(x)$, and the tube can be written as

$$\mathrm{Tube}(R_u, \rho) = \{x + r \cdot \eta^u(x) : x \in R_u, 0 \le r \le \rho\}.$$

In order to verify whether $x \in \mathrm{Tube}(R_u, \rho)$ we must effectively invert the mapping $(x, r) \mapsto x + r \cdot \eta^u(x)$. The key to the inversion is the observation

that, if $x \in \text{Tube}(R_u, \rho) \setminus R_u$, then it will have left $R_u$ along a normal vector parallel to $\eta^u(x)$.

There is some subtlety at work here. For an arbitrary set $A$, to determine whether $x \in \text{Tube}(A, \rho) \setminus A$ we must find $x^* \in A$, the closest point to $x$ within $A$ (which is a function of $x$). The point $x^*$ is such that $x - x^*$ is parallel to $\eta(x^*)$, a unit outward normal of $\partial A$ at $x^*$. If $|x - x^*| < \rho$, then we can conclude that $x \in \text{Tube}(A, \rho)$. In the general case, therefore, the relevant normal vector in question is based at $x^*$, not $x$. Our observation above can be restated as $\eta^u(x^*) = \eta^u(x)$, meaning that we actually know which direction we must move $x$ to reach $R_u$ in a normal direction.

Finally, we must move $x$ a distance $\rho$ along $-\eta^u(x)$ and check whether $x - \rho\eta^u(x) \in R_u$. This amounts to verifying whether or not

$$\frac{|U(x - \rho \cdot \eta^u(x))|}{|V(x - \rho \cdot \eta^u(x))|} \geq \sqrt{u}.$$

A straightforward calculation shows that, as long as $x$ is outside the region described by (4.5.21),

$$|U(x - \rho \cdot \eta^u(x))| = |U(x)| + \frac{\rho}{\sqrt{1+u}},$$
$$|L(x - \rho \cdot \eta^u(x))| = |L(x)| - \rho\sqrt{\frac{u}{1+u}}.$$

The result now follows from some simple manipulations.

As for the converse ($\Rightarrow$), suppose $x \in \text{Tube}(R_u, \rho)$. Then, we can write

$$x = x^* + \delta \cdot \eta(x^*) = x^* + \delta \cdot \eta(x)$$

for some $x^* \in \partial R_u, 0 \leq \delta \leq \rho$ with

$$|U(x^*)| = \sqrt{u}\,|L(x^*)|.$$

Similar calculations to those above show that this is equivalent to

$$|U(x)| = |L(x)|\sqrt{u} - \delta\sqrt{1+u} \Rightarrow |U(x)| \geq |L(x)|\sqrt{u} - \rho\sqrt{1+u},$$

and so we are done. □

We now return to the computation of the EC densities for an $F$ random field. Note firstly that in the notation of Lemma 4.5.3, the rejection region for an $F$ statistic is $R_{un/m}$. Formally, then, the EC densities up to order $k-1$ of the $F$ random field are coefficients of the powers of $\rho$ in the expansion

$$\begin{aligned}
&\mathbb{P}\left\{|U(x)| \geq |L(x)|\sqrt{\frac{un}{m}} - \rho\sqrt{1+\frac{un}{m}}\right\} \\
&\quad = \mathbb{P}\left\{\chi_n \geq \chi_m\sqrt{\frac{un}{m}} - \rho\sqrt{1+\frac{un}{m}}\right\} \\
&\quad = \mathbb{E}\left\{\mathbb{P}\left\{\chi_n \geq \chi_m\sqrt{\frac{un}{m}} - \rho\sqrt{1+\frac{un}{m}} \;\middle|\; \chi_m\right\}\right\} \\
&\quad = \mathbb{E}\left\{\sum_{j=0}^{k-1}\mathcal{L}_j(S^{k-1})\,\rho_j^G\left(\chi_m\sqrt{\frac{un}{m}} - \rho\sqrt{1+\frac{un}{m}}\right)\right\},
\end{aligned}$$

the last equality being a consequence of (4.5.16). Now, from (4.5.1), check that

$$\begin{aligned}
\frac{\partial^l}{\partial x^l}\rho_j^G(x) &= \frac{\partial^l}{\partial x^l}(-1)^{j-1}\frac{\partial^{j-1}}{\partial x^{j-1}}((2\pi)^{-(j+1)/2})e^{-x^2/2} \\
&= (2\pi)^{-(j+1)/2}(-1)^{j-1}\frac{\partial^{l+j-1}}{\partial x^{l+j-1}}e^{-x^2/2} \\
&= (2\pi)^{l/2}(-1)^l\rho_{j+l}^G(x),
\end{aligned}$$

and use the power series expansion

$$\begin{aligned}
\rho_j^G(x-c\rho) &= \sum_{l=0}^{\infty}(-1)^l\frac{(c\rho)^l}{l!}\frac{\partial^l}{\partial x^l}\rho_j^G(x) \\
&= \sum_{l=0}^{\infty}\frac{(2\pi)^{l/2}(c\rho)^l}{l!}\rho_{j+l}^G(x)
\end{aligned}$$

to evaluate the final expectation above to find that it is equal to

$$\sum_{l=0}^{\infty}\frac{(2\pi)^{l/2}\rho^l}{l!}\left(1+\frac{un}{m}\right)^{l/2}\sum_{j=0}^{k-1}\mathcal{L}_j(S^{k-1})\,\mathbb{E}\left\{\rho_{j+l}^G\left(\chi_m\sqrt{\frac{un}{m}}\right)\right\}.$$

The EC densities are therefore given by

$$\left(1+\frac{un}{m}\right)^{j/2}\sum_{l=0}^{n-1}\mathcal{L}_l(S^{n-1})\,\mathbb{E}\left\{\rho_{l+j}^G\left(\chi_m\sqrt{\frac{un}{m}}\right)\right\}. \tag{4.5.23}$$

Evaluating the final expectation, we obtain

**Theorem 4.5.4.** *The EC densities of an $F$ random field on $T$ with $n$ and $m$ degrees of freedom and $n+m > \dim(T)$ are given, for $j \geq 1$, by*

$$\rho_j^F(u;n,m) = \left(1+\frac{un}{m}\right)^{j/2}\sum_{l=0}^{n-1}\mathcal{L}_l(S^{n-1})\,\mathbb{E}\left\{\rho_{l+j}^G\left(\chi_m\sqrt{\frac{un}{m}}\right)\right\}, \tag{4.5.24}$$

*where*

$$\mathbb{E}\left\{\rho_j^G\left(\chi_m\sqrt{\frac{un}{m}}\right)\right\} = \sum_{l=0}^{\lfloor\frac{j-1}{2}\rfloor} \frac{(-1)^l(j-1)!\Gamma\left(\frac{j-1-2l+m}{2}\right)}{\pi^{(j+1)/2}2^{2l+1}(j-1-2l)!l!\Gamma\left(\frac{m}{2}\right)} \times \left(\frac{un}{m}\right)^{(j-1-2l)/2}\left(1+\frac{un}{m}\right)^{-(m-1-2l+j)/2}. \tag{4.5.25}$$

Before leaving the $F$ case, there is an important aspect of the calculations we have made that deserves further explanation, or, to be more precise, justification. Recall that we already discussed the fact that the rejection rejection for the $F$ statistic is not locally convex, and so the general structure of tube probability expansions of Section 3.3.4 cannot be used. The way we solved this problem was to find an alternative way to evaluate tube probabilities, although only up to a certain order of accuracy, and then treating the coefficients in an expansion of these probabilities (up to the appropriate order) as if they were the $\mathcal{M}_j$ in a 'true' expansion.

There is no problem in doing this. The serious problem lies in whether or not the GKF at all continues to hold in the current scenario, with these coefficients replacing the first few coefficients from a 'true' expansion. This is certainly not clear from the way that Theorem 4.4.1 is stated. However, it does follow from the details of the proof of Theorem 4.4.1 (cf. Chapter 15 of *RFG*) that all is fine. The proof does not really use the formalism of the expansions of Section 3.3.4. Rather, it involves a direct evaluation of the $\mathcal{M}_j$ which is consistent with what we have done above. The tube probability expansions are then introduced to provide additional insight into the $\mathcal{M}_j$ and to provide a method for computing them in smooth cases.

If you are unhappy with the above arguments, then you can turn our version of the original proof of Theorem 4.5.4 in [99], which we give in Section 4.5.10. The proof there is what might be called a 'first principles' proof, in that it is a version of our proof of Theorem 4.2.1 for stationary Gaussian fields on rectangles extended to stationary $F$ fields. Whereas the proof there does not rely on the GKF, you will find the specific computations there even more involved than those above.

### 4.5.5 Student $T$ Fields

Retaining our previous notation, and letting $\epsilon$ takes values in $\mathbb{R}^{m+1}$, the Student $T$ field with $m$ degrees of freedom is defined as

$$T(t) = \frac{\epsilon_1(t)}{|L(\epsilon(t))|/\sqrt{m}},$$

where $L(\epsilon(t)) = (\epsilon_2(t), \ldots, \epsilon_{m+1}(t))$.

This, of course, is almost the square root of an $F$ field, with 1 and $m$ degrees of freedom, modulo a sign ambiguity. Handling the ambiguity by symmetry, it follows immediately that the corresponding EC densities are given by

$$\rho_j^{\mathrm{T}}(u;m) = \frac{1}{2}\rho_j^F(u^2;1,m) = \left(1+\frac{u^2}{m}\right)^{j/2} \mathbb{E}\Big\{\rho_j^G\Big(\chi_m \frac{u}{\sqrt{m}}\Big)\Big\}, \qquad (4.5.26)$$

where we have used (4.5.24) and the fact that $\mathcal{L}_0(S^0) = \mathcal{L}_0(\{-1,+1\}) = 2$. The expectation here has the expansion given by (4.5.25).

### 4.5.6 Hotelling $T^2$ Fields

In order to define the Hotelling $T^2$ field, we first need a matrix of i.i.d., zero mean, unit variance Gaussian fields,

$$\epsilon(t) = \epsilon_{ij}(t), \qquad 1 \le i \le n,\ 1 \le j \le p.$$

Let $\epsilon_{i\cdot}$ denote the $i$-th row of $\epsilon$, and let $\bar{\epsilon}(t)$ be the row vector of column means, defined by

$$\bar{\epsilon}(t) = \frac{1}{n}\sum_{i=1}^{n} \epsilon_{i\cdot}(t).$$

Also, let

$$W_{lm}(t) = \frac{1}{n-1}\sum_{i=1}^{n} \left(\epsilon_{il}(t) - \bar{\epsilon}_l(t)\right)\left(\epsilon_{im}(t) - \bar{\epsilon}_m(t)\right)$$

be the sample covariance estimator of one row of the matrix $\epsilon(t)$. The Hotelling $T^2$ random field with $n-1$ degrees of freedom in $p$ dimensions is then defined as

$$H(t) = n\,\bar{\epsilon}(t) W(t)^{-1} \bar{\epsilon}(t)'.$$

Finding the EC densities for $H$ directly from a tube formula expansion is not easy, since the rejection region is quite complicated. However, we can adopt the approach that we have already used once for $\bar{\chi}^2_K$ fields, of enlarging the parameter space somewhat to obtain a random field with a much simpler rejection region, and applying the GKF to this. To do this, note that

$$H(t) = \sup_{v:\,|v|=1} T_v^2(t)$$

where, for any unit vector $v \in \mathbb{R}^p$,

$$T_v^2(t) = \frac{(\bar{\epsilon}(t), v)^2}{\frac{1}{n-1}\sum_{i=1}^{n} \left((\epsilon_{i\cdot}(t) - \bar{\epsilon}(t)),\, v\right)^2}$$

is the $F$-random field formed from the vector-valued random field $(\epsilon(t), v)$.

Therefore, we can consider the maximum of the $T^2$ random field with $n-1$ degrees of freedom over $T$ as the maximum of an $F$ random field with $(1, n-1)$ degrees of freedom over $T \times S^{p-1}$.

However, as was the case for the $\bar{\chi}^2_K$ random field, we can go beyond this and compare actual excursion sets. In the present case the connection is

$$\varphi\left(\left\{(t,v):\, T_v^2(t) \geq u\right\}\right) = 2\varphi\left(\left\{t:\, H(t) \geq u\right\}\right),$$

the factor of 2 coming from the symmetry $T_v^2(t) = T_{-v}^2(t)$. (cf. Exercise 3.6.12.) We therefore have

$$\begin{aligned}
\mathbb{E}\left\{\varphi\left(\left\{t \in T:\, H(t) \geq u\right\}\right)\right\} &= \frac{1}{2}\mathbb{E}\left\{\varphi\left(\left\{(t,v) \in T \times S^{p-1}:\, T_v^2(t) \geq u\right\}\right)\right\} \\
&= \frac{1}{2}\sum_{j=0}^{\dim(T)+p-1} \mathcal{L}_j(T \times S^{p-1})\rho_j^F(u; 1, n-1) \\
&= \frac{1}{2}\sum_{j=0}^{\dim T}\mathcal{L}_j(T)\sum_{l=0}^{p-1}\mathcal{L}_l(S^{p-1})\rho_{j+l}^F(u; 1, n-1) \\
&= \sum_{j=0}^{\dim T}\mathcal{L}_j(T)\sum_{l=0}^{p-1}\mathcal{L}_l(S^{p-1})\rho_{j+l}^T(\sqrt{u}; n-1),
\end{aligned}$$

the second to last equality here following from Exercise 3.6.12[10] and the last from (4.5.26).

Since the GKF also tells us that the above can also be expressed as $\sum \mathcal{L}_j(T)\rho_j^H(t; p, n-1)$ where the $\rho_j^H$ are the EC densities of a Hotelling $T^2$ field, we have derived the following result.

**Theorem 4.5.5.** *The EC densities of the Hotelling $T^2$ random field with $n-1$ degrees of freedom in $p$ dimensions are given, for $j \geq 1$ and $u > 0$, by*

$$\rho_j^H(u; p, n-1) = \sum_{l=0}^{p-1}\mathcal{L}_l(S^{p-1})\rho_{j+l}^T(\sqrt{u}; n-1), \tag{4.5.27}$$

*where the $\rho^T$ are the EC densities of the Student $T$ random field, given by* (4.5.26).

(Should we be citing theses of students who worked these cases out?)

### 4.5.7 Roy's Maximum Root Field

As I understand it, these involve approximation arguments which would be nice to describe.

### 4.5.8 Conjunctions/Correlated

I think these are pretty much dead in the water in terms of applications – should we include them?

I think we should include a long list of everything that is known to date, whether there is a GKF proof or not. This will be most useful to applied readers.

[10] cf. Footnote 8 above, as applied to the $F$ random field $T_v^2(t)$.

### 4.5.9 Correlation Fields

QUOTE THESE FROM [27] IN CURRENT NOTATION.

### 4.5.10 Another Look at the *F* Field

I THOUGHT IN THIS SECTION WE MIGHT TRY TO WRITE SOME OF KEITH'S ORIGINAL $F$ COMPUTATION IN OUR NOTATION – MAYBE USE MY OTHER DERIVATION FROM FIRST PRINICPLES AND THE GKF, TOO. BUT, THIS DEVOTES AN AWFUL LOT OF TIME TO COMPUTING THE $F$ EC DENSITIES THAT WE HAVE ALREADY DONE – I'M NOT TOTALLY SURE THIS IS WORTH IT

## 4.6 Some Lipschitz-Killing Curvatures

The general Gaussian kinematic formula of Theorem 4.4.1 has, so far, been more general than we have been able to exploit. Although, in the previous section, we saw how to compute the Gaussian Minkowski functionals $\mathcal{M}_j^{\gamma_k}$ for many useful examples, we have not really seen how to compute the Lipschitz-Killing curvatures $\mathcal{L}_j$ in any examples beyond either locally isotropic fields or stationary fields over rectangles.

In this section we shall look at two such examples. The first, of time-space random fields, is quite easy to handle. The second, of scale space fields, is not.

However, both rely heavily on the fact that in the Gaussian kinematic formula the computation of the Lipschitz-Killing curvatures splits completely from the the calculation of the Gaussian Minkowski functionals $\mathcal{M}_j^{\gamma_k}$.

### 4.6.1 Time-Space Fields

As we noted already in Chapter 1, there are many interesting examples of random fields which develop in time, and so are best written as $f(t, x)$, with $t \in [t_l, t_u] \subset \mathbb{R}$ and $x \in T \subset \mathbb{R}^N$. Clearly, $t$ as the time parameter and $x$ is the spatial parameter.

Of particular interest, and easiest to work with analytically, are random fields for which the covariance function is of product form, in that (for zero mean fields)

$$\mathbb{E}\left\{f(t_1, x_1)f(t_2, x_2)\right\} = C_t(t_1, t_2)C_x(x_1, x_2), \tag{4.6.1}$$

where both $C_t$ and $C_x$ are constant variance covariance functions, on $\mathbb{R}$ and $\mathbb{R}^N$ respectively.

It is an immediate consequence of (4.6.1) that the first order temporal and spatial partial derivatives of $f$ are uncorrelated. Furthermore, the Riemannian metric (4.3.1) induced by such processes on the product space $M = [t_l, t_u] \times T$ is also of product form, so that, by Exercise 3.6.12

$$\mathcal{L}_i(A \times B) = \sum_{j=0}^{i} \mathcal{L}_j(A)\mathcal{L}_{i-j}(B) = \mathcal{L}_i(B) + |A|\mathcal{L}_{i-1}(B), \tag{4.6.2}$$

for an interval $A \subset \mathbb{R}$ and $B \subset T$.

Thus, for example, suppose that both $C_t$ and $C_x$ are stationary covariance functions, with variances $\sigma_t^2$ and $\sigma_x^2$ respectively, and with second spectral moments $\lambda_t$ in the first case and matrix of second spectral moments $\Lambda^x$ in the second. Assume furthermore that the set $T$ is a rectangle of the form $\prod_{i=j}^{N}[0, T_j]$. Then it follows simply from (4.6.2) and the arguments of Section 4.2 that the Lipschitz-Killing curvatures of $f$ on $M$ are given by

$$\mathcal{L}_i([t_l, t_u] \times T\Big) = \sigma_x^{-i} \sum_{J \in \mathcal{O}_i} |J|\,|\Lambda_J^x|^{1/2} + \frac{\lambda_t^{1/2}(t_u - t_l)}{\sigma_t \sigma_x^{i-1}} \sum_{J \in \mathcal{O}_{i-1}} |J|\,|\Lambda_J^x|^{1/2},$$

where, as usual, $\mathcal{O}_k$ denotes the $\binom{N}{k}$ elements of $\partial_k T$ which include the origin, with $\mathcal{O}_{-1} \triangleq \emptyset$, and $\Lambda_J^x$ is the usual minor of $\Lambda^x$. (cf. Theorem 4.2.1 and (4.3.3).)

The above is a nice, and not atypical, example of how it is often possible to compute the Lipschitz-Killing curvatures for non-stationary processes from stationary components.

### 4.6.2 Scale Space Fields

The Gaussian scale space random field is a conceptually simple random field obtained by smoothing white noise with an isotropic spatial filter over a range of filter widths or scales. The parameter space therefore includes both scale and location parameters [84, 100]. In essence the scale-space field is a continuous wavelet transform of white noise that is designed to be powerful at detecting a localised signal of unknown spatial scale as well as location. We shall see it in action later in Section ??????. For the moment, however, we shall treat it more abstractly.

We start with a definition. Let $W$ be Gaussian noise on $\mathbb{R}^N$ based on Lebesgue measure (cf. Section 2.4.5) and let $h(t)$ be a filter, Then the Gaussian scale space random field with filter $h$ is defined as

$$f(\sigma, t) = \sigma^{-N/2} \int_{\mathbb{R}^N} h\left(\frac{t-u}{\sigma}\right) W(du). \tag{4.6.3}$$

We shall let $t$ range over a nice subset $T$ of $\mathbb{R}^N$, and take $\sigma = [\sigma_l, \sigma_u]$, for some $0 < \sigma_l < \sigma_u < \infty$.

Assume that the filter is normalised so that $\int_{\mathbb{R}^N} h^2(t)\,dt = 1$, and that there exists a $\lambda > 0$ such that

$$\int_{\mathbb{R}^N} (\nabla h(t))' \nabla h(t)\,dt = \lambda I_{N \times N}. \tag{4.6.4}$$

Two common examples are given by the standard Gaussian kernel

$$h(t) = \frac{1}{\pi^{N/4}} e^{-|t|^2/2},$$

and the Marr or 'Mexican hat' wavelet

$$h(t) = \left[\frac{4N}{(N+2)\pi^{N/2}}\right]^{1/2} \left(1 - \frac{|t|^2}{N}\right) e^{-|t|^2/2}.$$

It is easy to check that in the Gaussian case (4.6.4) holds with $\lambda = \frac{1}{2}$ and in the wavelet case with $\lambda = (N+4)/(2N)$.

The covariance function is easily calculated to be (cf. Section 2.4.5)

$$\begin{aligned} C((\sigma_1, t_1), (\sigma_2, t_2)) &= \frac{1}{(\sigma_1\sigma_2)^{N/2}} \int_{\mathbb{R}^N} h\left(\frac{t_1 - u}{\sigma_1}\right) h\left(\frac{t_2 - u}{\sigma_2}\right) du \\ &= \frac{1}{(\sigma_1\sigma_2)^{N/2}} \int_{\mathbb{R}^N} h\left(\frac{t_1 - t_2 - v}{\sigma_1}\right) h\left(\frac{-v}{\sigma_2}\right) dv. \end{aligned}$$

Note that, for fixed $\sigma$, it follows from the last line above that $f$ is stationary in $t$. However, $f$ is most definitely not stationary as a process in the pair $(\sigma, t)$.

For notational convenience in what follows, we make the simple scale transformation $s = -\ln \sigma$, so that our random field is now denoted by $f(s,t)$ and has covariance function

$$C((s_1, t_1), (s_2, t_2)) = e^{N(s_1+s_2)/2} \int_{\mathbb{R}^N} h\left((t_1 - u)e^{s_1}\right) h\left((t_2 - u)e^{s_2}\right) du$$

Our choice of normalisation gives that the variance of $f$ is one. As for the the matrix of second order spectral moments, we can find these, as usual (cf. (2.4.9)) by differentiating the covariance function. Doing so, a little calculus gives that all first order partial derivatives of $f$ with respect to the space variables are uncorrelated with the first order derivative in the scale variable, and that

$$\kappa \stackrel{\Delta}{=} \mathrm{Var}\left(\frac{\partial f(s,t)}{\partial s}\right) = \int_{\mathbb{R}^N} \left[\langle u, \nabla h(u)\rangle + N h(u)/2\right]^2 du \qquad (4.6.5)$$

$$\Lambda_s \stackrel{\Delta}{=} \mathrm{Var}\left(\nabla f(s,t)\right) = \int_{\mathbb{R}^N} e^{2s} (\nabla h(u))' \nabla h(u)\, du, \qquad (4.6.6)$$

where $\nabla f$ denotes the spatial derivative of $f$ with respect to the elements of $t$ only. In view of (4.6.4), we have $\Lambda_s = \lambda e^{2s} I_{N\times N}$.

For the Gaussian and Marr wavelet kernels the values of $\kappa$ are $1/2$ and $N/2$, respectively.

### Calculating the $\mathcal{L}_j$

Since $f$ is not only non-stationary, but also not locally isotropic (since the matrix $\lambda_s$ in (4.6.6) is dependent on $s$) we once again cannot use the simple arguments of Euclidean geometry to calculate the $\mathcal{L}_j$ in the Gaussian kinematic formula.

There are two paths to overcoming this problem. The first was taken in [84, 100] which contained the first computation of the results below. The computation there was 'from first principles', in the sense that the expected Euler characteristic of the excursion sets of scale-space fields was computed much in the way we computed the original Gaussian result of Theorem 4.2.1, by using Morse theory and the expectation meta-theorem of Section 2.7.

The computations of this approach were quite long and rather intricate. On the other hand, they are quite within the mathematical level that we have been working at so far.

The second approach is the one that we shall take below, and which has not appeared elsewhere. It has the advantage of showing how the Gaussian kinematic formula can be applied in situations in which there is neither stationarity nor local isotropy. On the other hand, it has the disadvantage that it requires differential geometric arguments, which we have studiously avoided in this book, but did treat in detail in *RFG*.

We include it here to show how the general argument works and to attempt to motivate you to also master the material in *RFG*. The treatment is therefore unfortunately no longer self contained. However, the results can be applied without understanding the derivation, once you have read the following material on stratification and on what we need to compute. Ultimately, we shall write the Lipschitz-Killing curvatures for this problem, which are essentially Riemannian in nature, in terms of the Euclidean Lipschitz-Killing curvatures of $T$ and the parameters $\kappa$ and $\lambda$.

The derivation is purposely slow and methodical, with all the main steps carefully signposted. Our aim is not so much to show you how the following calculation works, but rather to set it up as a template for other examples that you might need to work out for yourself.

### Stratifying the parameter space

We start by adopting the approach of Section 3.5 by stratifying the parameter space into manifolds of common dimension. In the current case, this allows us to write the $(N+1)$-dimensional parameter space

$$M \stackrel{\Delta}{=} [s_l, s_u] \times T$$

as the disjoint union of of four types of pieces, arranged in three strata, according to their dimension:

$$\begin{aligned}\partial_{N+1}M = M^\circ &= (s_l, s_u) \times T^\circ,\\ \partial_N M &= (s_l, s_u) \times \partial T \ \cup \ \{s_u\} \times T^\circ \ \cup \ \{s_l\} \times T^\circ\\ &\stackrel{\Delta}{=} \text{side} \ \cup \ \text{top} \ \cup \text{bottom},\\ \partial_{N-1}M &= \{s_l\} \times \partial T \ \cup \ \{s_u\} \times \partial T.\end{aligned}$$

An example is given in Figure 4.6.1, in which $T$ is two-dimensional, and from which our terminology of 'top', 'bottom' and 'side' comes. Note that while this example is at the moment completely 'Euclidean', we shall soon give it a Riemannian structure with the metric induced by the random field $f$. In this case, as we shall see, one needs to think of the top as having positive (outward) curvature, while the bottom has negative (outward) curvature.

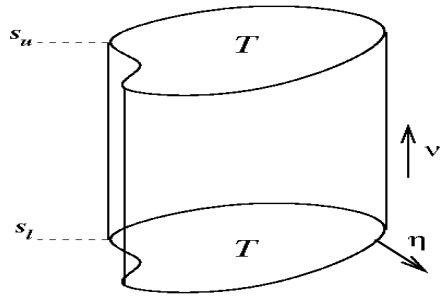


**Fig. 4.6.1.** The scale space parameter set

### What we need to compute

Ultimately, we need to compute the Lipschitz-Killing curvatures $\mathcal{L}_i(M)$ of $M$ under the Riemannian metric induced by the random field $f$. Rather than doing this directly, we shall compute certain other curvatures, denoted by $\mathcal{L}_i^\alpha(M)$ and defined via (3.3.18), specifically designed to be simpler to handle on sets of constant curvature. (We already met these in Section 3.3.3 in dealing with tube formulae on spheres, where they were convenient to use because of the constant curvature of spheres. We shall soon see that $M$ also has constant curvature, given by $-\kappa^{-1}$.)

There are simple relationships between the $\mathcal{L}_i(M)$ and $\mathcal{L}_i^\alpha(M)$, including the following, which is (10.5.12) of *RFG*.

$$\mathcal{L}_i(M) = \sum_{n=0}^{\lfloor \frac{N+1-i}{2} \rfloor} \frac{\alpha^n (i+2n)!}{(4\pi)^n n! i!} \mathcal{L}_{i+2n}^\alpha(M). \tag{4.6.7}$$

(The inverse relationship, which expresses the $\mathcal{L}_j^\alpha$ in terms of the $\mathcal{L}_j$, has already been given at (3.3.18).)

To compute the $\mathcal{L}_i^\alpha(M)$ themselves, we first write the decomposition

$$\mathcal{L}_i^\alpha(M) = \sum_{j=N-1}^{N+1} \mathcal{L}_i^\alpha(M; \partial_j M) \tag{4.6.8}$$

according to the three strata described above, and then restrict ourselves to the case $\alpha = -\kappa^{-1}$. Applying (10.7.10)[11] of *RFG* along with the (yet to be proven) fact that $M$ has constant curvature $-\kappa^{-1}$, gives that, for $i \leq j$,

$$\begin{aligned} &\mathcal{L}_i^{-\kappa^{-1}}(M; \partial_j M) \\ &\quad = \frac{1}{(2\pi)^{(j-i)/2}(j-i)!} \int_{\partial_j M} \mathbb{E}\left\{ \mathrm{Tr}^{T_t \partial_j M}(S_{Z_{j,t}}^{j-i}) \mathbb{1}_{N_t M}(Z_{j,t}) \right\} \mathcal{H}_j(dt). \end{aligned} \tag{4.6.9}$$

Here, for each $t \in \partial_j M$, $Z_{j,t}$ is a normally distributed random vector of dimension $N + 1 - j$ in the space $T_t \partial_j M^\perp$, the orthogonal complement of $T_t \partial_j M$ in $\mathbb{R}^{N+1}$, $S$ is the Riemannian shape operator of $M$, $N_t M$ is the normal cone to $M$ at $t$, $\mathcal{H}_j$ is the volume measure on $\partial_j M$ corresponding to the Riemannian metric induced by the random field $f$, and $\mathrm{Tr}^{T_t \partial_j M}$ is the trace restricted to the tangent space at $t$ to $\partial_j M$.

We shall now commence computing each term in (4.6.9), leaving you to compute the final Lipschitz-Killing curvatures by substituting into (4.6.8) and (4.6.7). Since all the formulas – as you shall soon see – are long and unintuitive, any explicit computation of them will involve a computer. This being the case, breaking down the problem into smaller pieces for later self-assembly generates no additional work at the application stage.

### The induced Riemannian metric

The first step in computing the Lipschitz-Killing curvatures is, obviously, determining the Riemannian metric $g$ that the random field induces on $M$. In order to describe this, however, we need first to choose a family of vector fields generating the tangent bundle of $M$.

We do this sequentially, starting with the set $T$, ignoring for the moment the scale component of the parameter space. For $T$, let $\eta$ be the (Euclidean) outward unit normal vector field on $\partial T$, and extend this to a full (Euclidean) orthonormal tangent vector field $X_1, \ldots, X_N$, on $T$, with $X_N = \eta$. Enlarge this to a vector field on all of $M$ by adding the vector field $\nu = \partial/\partial s$, the field of tangent vectors in the scale direction. See Figure 4.6.1.

The Riemannian metric $g$ at points $(s,t) \in M$ can now be calculated on pairs of vectors from the above vector field using the variances (4.6.5) and (4.6.6) and the independence discussed there. These yield

[11] There is a small, but, for us, rather significant typo in (10.7.10) of *RFG*, in that there is a minus sign missing before the $\kappa$ in $\kappa I^2/2$.

$$g_{(s,t)}(X_{i,(s,t)}, X_{j,(s,t)}) = \lambda e^{2s}\delta_{ij}, \tag{4.6.10}$$

$$g_{(s,t)}(X_{i,(s,t)}, \nu_{(s,t)}) = 0, \tag{4.6.11}$$

$$g_{(s,t)}(\nu_{(s,t)}, \nu_{(s,t)}) = \kappa, \tag{4.6.12}$$

where $\delta_{ij}$ is the Kronecker delta.

Note that the above three equations imply that the stucture of the normal cones to $M$ is well described by the initial, Euclidean, choice of vector fields. There is, of course, no normal cone at points $t \in \partial_{N+1}M$, the interior of $M$. For the sets in $\partial_N M$, one of the vector fields $\eta$ and $\nu$ describes the normal geometry. In particular, along the side the normal is a $s$-dependent multiple of $\eta$ and along the top and bottom the normals are constant multiples of $\nu$.

Along $\partial_{N-1}$ all normals are linear combinations of elements in the vector fields $\eta$ and $\nu$.

### The Levi-Civita connection

The next step towards computing the second fundamental forms in (4.6.9) requires identifying the Levi-Civita connection $\widetilde{\nabla}$ of $(M, g)$. For our purposes, we need only to know how it operates on normal vectors. Furthermore, since the Lie bracket

$$[X_i, \nu] = 0, \qquad \text{for all } 1 \le i \le N, \tag{4.6.13}$$

we need only compute $\widetilde{\nabla}_{X_i}X_j$, $\widetilde{\nabla}_{\nu}X_i$ and $\widetilde{\nabla}_{\nu}\nu$.

To start, note that a straightforward application of Koszul's formula[12] and (4.6.10) and (4.6.11) imply, for $i, j, k = 1, \ldots, N$, that

$$g(\widetilde{\nabla}_{X_i}X_j, X_k) = \lambda e^{2s}\langle \nabla_{X_i}X_j, X_k\rangle = g(\nabla_{X_i}X_j, X_k),$$

where $\nabla$ is the standard Euclidean connection, and

$$g(\widetilde{\nabla}_{X_i}X_j, \nu) = -\frac{\lambda}{2}\frac{\partial}{\partial s}e^{2s}\delta_{ij} = -\lambda e^{2s}\delta_{ij}. \tag{4.6.14}$$

Now note that for any tangent vectors $X, Y$ to $T$, $\widetilde{\nabla}_X Y$ is a vector in $\mathbb{R}^{N+1}$, and so can be written as $\sum_{j=1}^N a_j X_j + b\nu$ for appropriate coefficients. This fact, together with the last two equalities and (4.6.12) imply that for any two vector fields $X, Y$ on $T$,

$$\widetilde{\nabla}_X Y = \nabla_X Y - \kappa^{-1}\lambda e^{2s}\langle X, Y\rangle\nu, \tag{4.6.15}$$

giving us the first computation.

As far as $\widetilde{\nabla}_{\nu}\nu$ is concerned, note first that another easy consequence of Koszul's formula is that

[12] This is (7.3.12) of *RFG*, and in our case simplifies to $2g(\nabla_X Y, Z) = Xg(Y, Z) + Yg(X, Z) - Zg(X, Y)$.

$$g(\widetilde{\nabla}_\nu \nu, \nu) = \frac{1}{2}\nu(g(\nu,\nu)) = 0, \tag{4.6.16}$$

giving us $\widetilde{\nabla}_\nu \nu \equiv 0$.

All that remains is to compute $\widetilde{\nabla}_\nu X_i$. An application of the Weingarten equation[13] and (4.6.15) yield

$$g(\widetilde{\nabla}_\nu X_i, X_j) = \tfrac{\lambda}{2}\tfrac{\partial}{\partial s}e^{2s}\delta_{ij} = \lambda e^{2s}\delta_{ij} = g(X_i, X_j)$$

and

$$g(\widetilde{\nabla}_\nu X_i, \nu) = 0.$$

Applying now the torsion freeness of connections[14], along with (4.6.13) to the above, gives

$$\widetilde{\nabla}_\nu X_i = \widetilde{\nabla}_{X_i}\nu = X_i, \tag{4.6.17}$$

and we have the last of the three cases we were seeking.

### The second fundamental forms and the curvature matrix

With the relevant connections determined, we can now finally turn to computing second fundamental forms and curvature matrices along the three stratifications of $M$.

1. $\partial_{N+1}M$, the interior $(s_l, s_u) \times T^\circ$:

Since the normal space in the interior is empty, there are no normals, no second fundamental form and no curvature matrix here.

2. $\partial_N M$: The side $(s_l, s_u) \times \partial T$:

A convenient choice for an orthonormal (in the metric $g$) basis for the tangent space at any point on the side is given by

$$\left\{ \frac{X_1}{\lambda^{1/2}e^s}, \ldots, \frac{X_{N-1}}{\lambda^{1/2}e^s}, \frac{\nu}{\kappa^{1/2}} \right\},$$

with outward unit normal vector $\eta/\lambda^{1/2}e^s$.

The scalar second fundamental forms of interest are therefore

$$S_{\frac{\eta}{\lambda^{1/2}e^s}}\left(\frac{X_i}{\lambda^{1/2}e^s}, \frac{X_j}{\lambda^{1/2}e^s}\right), \quad S_{\frac{\eta}{\lambda^{1/2}e^s}}\left(\frac{X_i}{\lambda^{1/2}e^s}, \frac{\nu}{\kappa^{1/2}}\right), \quad S_{\frac{\eta}{\lambda^{1/2}e^s}}\left(\frac{\nu}{\kappa^{1/2}}, \frac{\nu}{\kappa^{1/2}}\right),$$

all of which can be computed directly using Weingarten's equation and (4.6.15)–(4.6.17). We summarise them in a curvature matrix, giving

[13] This is (7.5.12) of *RFG* and states that the scalar second fundamental form is given by $S_\nu(X,Y) = g(\widetilde{\nabla}_X Y, \nu) = -g(Y, \widetilde{\nabla}_X \nu)$.

[14] This is (7.3.10) of *RFG*, and states that $\nabla_X Y - \nabla_Y X - [X,Y] = 0$.

$$\begin{pmatrix} \lambda^{-1/2}e^{-s}C_t & 0 \\ 0 & 0 \end{pmatrix},$$

where $C_t$ is the $(N-1)\times(N-1)$ Euclidean curvature matrix of $\partial T$ in the basis $\{X_1,\dots,X_{d-1},\eta\}$. Recall for later use that, from Section 7.2 of *RFG*, in the current scenario, the trace in (4.6.9) can be replaced by

$$(j-i)!\mathrm{detr}_{j-i}C, \tag{4.6.18}$$

where $\mathrm{detr}_j$ is our usual sum of determinants of $j\times j$ principal minors.

3. $\partial_N M$, the bottom $\{s_l\}\times T^\circ$ and the top $\{s_u\}\times T^\circ$:

Starting with the bottom $\{s_l\}\times T^\circ$, the outward unit normal vector is $-\kappa^{-1/2}\nu$, and a convenient orthonormal basis is given by

$$\left\{\frac{X_1}{\lambda^{1/2}e^s},\dots,\frac{X_N}{\lambda^{1/2}e^s}\right\},$$

The curvature matrix is therefore $N\times N$ with entries

$$S_{-\kappa^{-1/2}\nu}\left(\frac{X_i}{\lambda^{1/2}e^s},\frac{X_j}{\lambda^{1/2}e^s}\right) = g\left(\widetilde{\nabla}_{\frac{X_i}{\lambda^{1/2}e^s}}(-\kappa^{-1/2}\nu),\frac{X_j}{\lambda^{1/2}e^s}\right) = \kappa^{-1/2}\delta_{ij},$$

by Weingarten's equation and (4.6.14).

Thus, for the bottom, the outward curvature matrix is $\kappa^{-1/2}I_{N\times N}$ while along the top the same arguments give it as $-\kappa^{-1/2}I_{N\times N}$.

4. $\partial_{N-1}M$: The edges $\{s_l\}\times\partial T$ and $\{s_u\}\times\partial T$:

For the edges, we need to consider the scalar second fundamental form itself, and not just the curvature matrix. As above, an orthonormal basis for the tangent space is

$$\left\{\frac{X_1}{\lambda^{1/2}e^s},\dots,\frac{X_{N-1}}{\lambda^{1/2}e^s},\frac{\nu}{\kappa^{1/2}}\right\},$$

but now an orthonormal basis for the normal space is $\{\lambda^{-1/2}e^{-s}\eta,\kappa^{-1/2}\nu\}$.

Applying the Weingarten equation, we need to compute

$$g\left(\widetilde{\nabla}_{\lambda^{-1/2}e^{-s}X_i}\left(\lambda^{-1/2}e^{-s}X_j\right),\ a\lambda^{-1/2}e^{-s}\eta + b\kappa^{-1/2}\nu\right),$$

for arbitrary $a,b$. Applying (4.6.15) gives that this is

$$a\,\lambda^{-1/2}e^{-s}C_{ij,t} + b\,\kappa^{-1/2}\delta_{ij},$$

where, with a minor abuse of notation, we now use $C_t$ to denote the $(N-1)\times(N-1)$ Euclidean curvature matrix of $\partial M$ at $t$.

**The curvature tensor**

We now have all the pieces we need to compute the Lipschitz-Killing curvatures $\mathcal{L}_j^{-\kappa^{-1}}$, and could actually proceed to the final computation. However, in justifying the formula (4.6.9) we used the fact that $M$ has constant negative curvature $-\kappa^{-1}$. Now we shall take a moment to prove this.

Let

$$R(X,Y) = \widetilde{\nabla}_X \widetilde{\nabla}_Y - \widetilde{\nabla}_Y \widetilde{\nabla}_X - \widetilde{\nabla}_{[X,Y]}$$

be the curvature operator. Using our previous calculations of the connection we have

$$\begin{aligned}
R(X_i,\nu)X_k &= \widetilde{\nabla}_{X_i}\widetilde{\nabla}_\nu X_k - \widetilde{\nabla}_\nu \widetilde{\nabla}_{X_i} X_k \\
&= \widetilde{\nabla}_{X_i} X_k - \widetilde{\nabla}_\nu(\nabla_{X_i} X_k - \delta_{ik}\kappa^{-1}\lambda e^{2s}\nu) \\
&= \nabla_{X_i} X_k - \delta_{ik}\kappa^{-1}\lambda e^{2s}\nu - \nabla_{X_i} X_k + \delta_{ik}\kappa^{-1}\frac{\partial}{\partial s}\lambda e^{2s}\nu \\
&= \delta_{ik}\kappa^{-1}\lambda e^{2s}\nu,
\end{aligned}$$

$$\begin{aligned}
R(X_i,\nu)\nu &= \widetilde{\nabla}_{X_i}\widetilde{\nabla}_\nu \nu - \widetilde{\nabla}_\nu \widetilde{\nabla}_{X_i}\nu \\
&= -\widetilde{\nabla}_\nu X_i \\
&= -X_i,
\end{aligned}$$

$$\begin{aligned}
R(X_i,X_j)\nu &= \widetilde{\nabla}_{X_i}\widetilde{\nabla}_{X_j}\nu - \widetilde{\nabla}_{X_j}\widetilde{\nabla}_{X_i}\nu - \widetilde{\nabla}_{[X_i,X_j]}\nu \\
&= \widetilde{\nabla}_{X_i} X_j - \widetilde{\nabla}_{X_j} X_i - [X_i,X_j] \\
&= \nabla_{X_i} X_j - \nabla_{X_j} X_i - [X_i,X_j] - \kappa^{-1}\lambda e^{2s}(\delta_{ij} - \delta_{ji})\nu \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
R(X_i,X_j)X_k &= \widetilde{\nabla}_{X_i}\widetilde{\nabla}_{X_j} X_k - \widetilde{\nabla}_{X_j}\widetilde{\nabla}_{X_i} X_k - \widetilde{\nabla}_{[X_i,X_j]} X_k \\
&= \widetilde{\nabla}_{X_i}(\nabla_{X_j} X_k - \delta_{jk}\kappa^{-1}\lambda e^{2s}\nu) - \widetilde{\nabla}_{X_j}(\nabla_{X_i} X_k - \delta_{ik}\kappa^{-1}\lambda e^{2s}\nu) \\
&\qquad - \nabla_{[X_i,X_j]} X_k + \kappa^{-1}\lambda e^{2s}\langle [X_i,X_j], X_k\rangle \nu \\
&= \nabla_{X_i}\nabla_{X_j} X_k - \delta_{jk}\kappa^{-1}\lambda e^{2s} X_i - \kappa^{-1}\lambda e^{2s}\langle \nabla_{X_j} X_k, X_i\rangle \nu \\
&\quad - \nabla_{X_j}\nabla_{X_i} X_k + \delta_{ik}\kappa^{-1}\lambda e^{2s} X_j - \kappa^{-1}\lambda e^{2s}\langle \nabla_{X_i} X_k, X_j\rangle \nu \\
&\qquad - \nabla_{[X_i,X_j]} X_k + \kappa^{-1}\lambda e^{2s}\langle [X_i,X_j], X_k\rangle \nu \\
&= \nabla_{X_i}\nabla_{X_j} X_k - \nabla_{X_j}\nabla_{X_i} X_k - \nabla_{[X_i,X_j]} X_k \\
&\quad + \kappa^{-1}\lambda e^{2s}\nu\left(\langle [X_i,X_j], X_k\rangle - \langle \nabla_{X_j} X_k, X_i\rangle + \langle \nabla_{X_i} X_k, X_j\rangle\right) \\
&\qquad - \delta_{jk}\kappa^{-1}\lambda e^{2s} X_i + \delta_{ik}\kappa^{-1}\lambda e^{2s} X_j \\
&= -\delta_{jk}\kappa^{-1}\lambda e^{2s} X_i + \delta_{ik}\kappa^{-1}\lambda e^{2s} X_j.
\end{aligned}$$

where the last equality follows from the flatness of Euclidean space and torsion freeness.

Since the curvature tensor is given by

$$R(X,Y,Z,W) = g(R(X,Y)Z,W),$$

it is now easy to use the above calculations to check cases and see that

$$R(X,Y,Z,W) = -\kappa^{-1} I^2(X,Y,Z,W)/2,$$

where $I$ is the identity form given by $I(X,Y) = g(X,Y)$, and, with the usual tensor product $I^2(X,Y,Z,W) \stackrel{\Delta}{=} I(X,Z)I(Y,W) - I(X,W)I(Y,Z)$.

From this, of course, follows our claim that $M$ is a space of constant curvature $-\kappa^{-1}$.

### The Lipschitz-Killing curvatures $\mathcal{L}^{-\kappa^{-1}}$

With all the preparation done, we can now begin the computation of the Lipschitz-Killing curvatures $\mathcal{L}_j^{-\kappa^{-1}}(M;\partial_k M)$, according to their definition in (4.6.9). As in the discussion of second fundamental forms, we divide the computation into separate sections, each corresponding to a different stratum in the stratification of $M$.

Throughout the following computation we take $Z_1$ and $Z_2$ to be two independent $N(0,1)$ random variables. We shall associate the $Z_j$ with the normal vector fields to obtain normal vectors of the form $Z_1\nu + Z_2\eta$.

We also adopt the notation

$$\mu_k \stackrel{\Delta}{=} \mathbb{E}\left\{Z_j^k \mathbb{1}_{Z_j \geq 0}\right\} = \begin{cases} \frac{2^{n-1}n!}{\sqrt{2\pi}} & \text{if } k = 2n+1 \text{ is odd} \\ \frac{(2n-1)(2n-3)\ldots}{2} & \text{if } k = 2n \text{ is even.} \end{cases}$$

1. $\partial_{N+1}M$, the interior $(s_l, s_u) \times T^\circ$:

Since the second fundamental form is zero in $\partial_{N+1}M$, the only non-zero Lipschitz-Killing curvature occurs when $N+1 = j = i$ in (4.6.9). In this case we have

$$\begin{aligned} \mathcal{L}_{N+1}^{-\kappa^{-1}}(M;\partial_{N+1}M) &= \mathcal{H}_{N+1}((s_l,s_u)\times T^\circ) \\ &= \kappa^{1/2}\lambda^{N/2}\int_{s_l}^{s_u}\int_{T^\circ} e^{Ns}dt\,ds \\ &= \kappa^{1/2}\lambda^{N/2}\frac{e^{Ns_u} - e^{Ns_l}}{N}\mathcal{L}_N(T;T^\circ), \end{aligned}$$

where $\mathcal{L}_N(T;T^\circ)$ is computed in the standard Euclidean sense.

In anticipation of a more general structure to follow, note that, for all $j \geq 1$, we could actually write

$$\mathcal{L}_j^{-\kappa^{-1}}(M;\partial_{N+1}M) = \kappa^{1/2}\lambda^{(j-1)/2}\frac{e^{(j-1)s_u} - e^{(j-1)s_l}}{(j-1)}\,\mathcal{L}_{j-1}^E(T;T^\circ), \quad (4.6.19)$$

allowing $\mathcal{L}_0(\cdot)/0 = 0$, and adding the superscript $E$ to emphasise the Euclidean nature of the Lipschitz-Killing curvatures on the right hand side. Of course, for $j \leq N$, both sides here are identically zero.

2. $\partial_N M$: The side $(s_l, s_u) \times \partial T$:

For this case, all the Lipschitz-Killing curvatures need to be computed, and we replace the trace in (4.6.9) by the determinants of the curvature matrix as in (4.6.18). Then

$$\begin{aligned}
&\mathcal{L}_j^{-\kappa^{-1}}(M; (s_l, s_u) \times \partial T) \qquad (4.6.20)\\
&\quad = (2\pi)^{-(N-j)/2}\mathbb{E}\big\{Z_2^{N-j}\mathbb{1}_{\{Z_2>0\}}\big\}\\
&\qquad\qquad \times \int_{(s_l,s_u)\times\partial T} \mathrm{detr}_{N-j}\begin{pmatrix} e^{-s}C_t & 0 \\ 0 & 0\end{pmatrix} d\mathcal{H}_N(s,t)\\
&\quad = (2\pi)^{-(N-j)/2}\kappa^{1/2}\mathbb{E}\big\{Z_2^{N-j}\mathbb{1}_{\{Z_2>0\}}\big\}\\
&\qquad\qquad \times \int_{s_l}^{s_u}\int_{\partial T} (\lambda^{-1/2}e^{-s})^{N-j}(\lambda^{1/2}e^{s})^{N-1}\mathrm{detr}_{N-j}(C_t)\, dtds\\
&\quad = \kappa^{1/2}\lambda^{(j-1)/2}\frac{e^{(j-1)s_u} - e^{(j-1)s_l}}{j-1}\,\mathcal{L}_{j-1}^E(T;\partial T),
\end{aligned}$$

in a parallel notation to (4.6.19).

3. $\partial_N M$, the bottom $\{s_l\} \times T^\circ$ and the top $\{s_u\} \times T^\circ$:

Beginning with the bottom, $\{s_l\} \times T^\circ$, for $0 \leq j \leq N$ we have

$$\begin{aligned}
&\mathcal{L}_j^{-\kappa^{-1}}(M; \{s_l\} \times T^\circ) \qquad (4.6.21)\\
&\quad = (2\pi)^{-(N-j)/2}\mathbb{E}\big\{Z_1^{N-j}\mathbb{1}_{\{Z_1>0\}}\big\}\\
&\qquad\qquad \times \int_{\{s_l\}\times T^\circ} \mathrm{detr}_{N-j}\big(\kappa^{1/2} I_{N\times N}\big) d\mathcal{H}_N(t)\\
&\quad = (2\pi\kappa^{-1})^{-(N-j)/2}\mu_{N-j}\binom{N}{j}\lambda^{N/2}e^{Ns_l}\mathcal{L}_N^E(T;T^\circ).
\end{aligned}$$

A similar result holds for the top, $\{s_u\} \times T^\circ$, viz.

$$\begin{aligned}
&\mathcal{L}_j^{-\kappa^{-1}}(M; \{s_u\} \times T^\circ) \qquad (4.6.22)\\
&\quad = (-1)^{N-j}(2\pi\kappa^{-1})^{-(N-j)/2}\mu_{N-j}\binom{N}{j}\lambda^{N/2}e^{Ns_u}\mathcal{L}_N^E(T;T^\circ).
\end{aligned}$$

4. $\partial_{N-1} M$: The edges $\{s_l\} \times \partial T$ and $\{s_u\} \times \partial T$:

We start with the top edge, $\{s_u\} \times \partial T$ .

$$\begin{aligned}
&\mathcal{L}_j^{-\kappa^{-1}}(M; \{s_u\} \times \partial T)\\
&\quad = (2\pi)^{-(N-1-j)/2}\int_{\{s_u\}\times\partial T} \mathbb{E}\big\{\mathrm{detr}_{N-1-j}(Z_1\kappa^{-1/2}I - Z_2\lambda^{-1/2}e^{-s}C_t)\\
&\qquad\qquad \times\mathbb{1}_{\{Z_1>0\}}\mathbb{1}_{\{Z_2>0\}}\big\} d\mathcal{H}_{N-1}(s,t).
\end{aligned}$$

Use now the easily checked expansion that, for $0 \le k \le n$,

$$\mathrm{detr}_k\left(\alpha I_{n\times n} + A_{n\times n}\right) = \sum_{m=0}^{k} \alpha^{k-m} \binom{n-m}{k-m} \mathrm{detr}_m(A),$$

to expand the detr term in the expectation above and see that

$$\begin{aligned}
&\mathcal{L}_j^{-\kappa^{-1}}(M; \{s_u\} \times \partial T) \qquad (4.6.23)\\
&\quad = (2\pi)^{-(N-1-j)/2} \sum_{m=0}^{N-1-j} \kappa^{-(N-1-j-m)/2} \binom{N-1-m}{j}\\
&\qquad \times \mathbb{E}\left\{ Z_1^{N-1-j-m} \mathbb{1}_{\{Z_1>0\}} \right\} (\lambda^{1/2} e^{s_u})^{N-1} \mathbb{E}\left\{ Z_2^m \mathbb{1}_{\{Z_2>0\}} \right\}\\
&\qquad\quad \times \int_{\partial T} \mathrm{detr}_m(\lambda^{-1/2} e^{-s_u} C_t)\, d\mathcal{H}_{N-1}(t)\\
&\quad = (2\pi\kappa^{-1})^{-(N-1-j)/2} \sum_{m=0}^{N-1-j} (\lambda^{1/2} e^{s_u})^{N-1-m} \binom{N-1-m}{j} \kappa^{m/2}\\
&\qquad \times \mu_{N-1-j-m} \mathbb{E}\{ Z_2^m \mathbb{1}_{\{Z_2>0\}} \} \int_{\partial T} \mathrm{detr}_m(C_t)\, d\mathcal{H}_{N-1}(t)\\
&\quad = \kappa^{(N-1-j)/2} (\lambda^{1/2} e^{s_u})^{N-1} \sum_{m=0}^{N-1-j} \binom{N-1-m}{j} (\lambda^{-1/2} e^{-s_u} \kappa^{1/2})^m\\
&\qquad \times \mu_{N-1-j-m} \mathcal{L}_{N-1-m}^E(T; \partial T).
\end{aligned}$$

A similar argument also works for the bottom edge $\{s_l\} \times \partial T$, the only change being that the condition $\mathbb{1}_{\{Z_1>0\}}$ becomes $\mathbb{1}_{\{Z_1<0\}}$, giving the final form

$$\begin{aligned}
&\mathcal{L}_j^{-\kappa^{-1}}(M; \{s_l\} \times \partial T) \qquad (4.6.24)\\
&\quad = \kappa^{(N-1-j)/2} (\lambda^{1/2} e^{s_l})^{N-1} \sum_{m=0}^{N-1-j} \binom{N-1-m}{j} (\lambda^{-1/2} e^{-s_l} \kappa^{1/2})^m\\
&\qquad \times (-1)^{N-1-j-m} \mu_{N-1-j-m} \mathcal{L}_{N-1-m}^E(T; \partial T).
\end{aligned}$$

Collecting now (4.6.19)–(4.6.24) gives us all the $\mathcal{L}_i^{-\kappa^{-1}}(M; \partial_j M)$, from which, via (4.6.7) and (4.6.8), we can compute the $\mathcal{L}_i(M)$, and so we are done.

### An alternative approach, built on existing results

At this point we could explain how to get the results from the Worlsey/Siegmund or Roy papers, once we are certain that the results are the same.

BELOW IS JONATHAN'S ORIGINAL CUT AND PASTE FROM THE ROY PAPER.

let $f(s)$ be a filter, normalised so that $\int f^2 = 1$, and scaled so that $\int \dot{f}\dot{f}' = I_{N\times N}$. The Gaussian scale space random field with filter $f$ is defined as

$$T(s,w) = w^{-N/2} \int_{\mathbb{R}^N} f((s-t)/w) dB(t). \tag{4.6.25}$$

Note that $T(s,w) \sim \mathrm{N}(0,1)$ and $\mathrm{Var}(\partial T/\partial s) = w^{-2N} I_{N\times N}$ at each point $s, w$. [84] and [100] show that for searching over a range of scales $w \in [w_1, w_2]$

$$\mathbb{E}\left\{\varphi\{s, w \in S \times [w_1, w_2] : T(s,w) \geq t\}\right\} = \sum_{i=0}^{N} \mu_i(S) \rho_i^{\mathrm{S}}(t) \tag{4.6.26}$$

where the Gaussian scale space EC density is

$$\rho_i^{\mathrm{S}}(t) = \frac{w_1^{-i} + w_2^{-i}}{2} \rho_i^{\mathrm{G}}(t) + \frac{w_1^{-i} - w_2^{-i}}{i} \sum_{j=0}^{\lfloor i/2 \rfloor} \frac{\kappa^{(1-2j)/2}\ (-1)^j\ i!}{(1-2j)(4\pi)^j j!(i-2j)!} \rho_{i+1-2j}^{\mathrm{G}}(t) \tag{4.6.27}$$

(we define $w^i/i$ as $\log(w)$ when $i = 0$).

the scale space result can be set in terms of the Lipschitz-Killing curvature of $S \times [w_1, w_2]$, as in (**??**):

$$\mathbb{E}(\varphi\{s, w \in S \times [w_1, w_2] : T(s,w) \geq t\}) = \sum_{i=0}^{N+1} \mathcal{L}_i(S \times [w_1, w_2]) \rho_i^{\mathrm{G}}(t). \tag{4.6.28}$$

Equating the two expressions (4.6.26) and (4.6.28) for the expected EC implies that $\mathcal{L}_0(S \times [w_1, w_2]) = \mu_0(S)$ and for $i \geq 1$

$$\begin{aligned} \mathcal{L}_i(S \times [w_1, w_2]) = & \frac{w_1^{-1} + w_2^{-1}}{2} \mu_i(S) + \sum_{j=0}^{\lfloor (N-i+1)/2 \rfloor} \frac{w_1^{-i-2j+1} - w_2^{-i-2j+1}}{i+2j-1} \\ & \times \frac{\kappa^{(1-2j)/2}\ (-1)^j\ (i+2j-1)!}{(1-2j)(4\pi)^j j!(i-1)!} \mu_{i+2j-1}(S). \end{aligned} \tag{4.6.29}$$

### 4.6.3 Rotation Space Fields

NOT QUITE SURE WHAT TO DO ABOUT THIS. THE 4-AUTHOR PAPER DOES THE 2-D CASE FOR A GAUSSIAN KERNEL, SO IT IS A VERY SPECIAL RESULT. ALSO, THE FINAL EEC IS COMPLICATED ENOUGH THAT I CANNOT SEE OFFHAND HOW TO 'UNWIND' THINGS TO PUT IT INTO THE FORM OF THE GKF. WHAT SHOULD WE DO?

## 4.7 Mean Lipschitz-Killing Curvatures under Isotropy

Throughout this chapter, we have concentrated on finding expressions for the mean value of the Euler characteristics of the random excursion sets $T \cap \epsilon^{-1} D$. Theorem 4.4.1, which we called the 'basic form' of the Gaussian kinematic formula gave a general expression for these expectation, and since then we have been busy treating special cases.

However, as interesting and important as the Euler characteristic of excursion sets may be, this is only one measure of their geometric structure. One could well ask about their size, the size of their boundaries, etc. The first of these is actually rather trivial, for writing $\lambda_N$ as usual for Lebesgue measure in $\mathbb{R}^N$, it is immediate that

$$\begin{aligned} \mathbb{E}\left\{\lambda_N\left(T \cap \epsilon^{-1} D\right)\right\} &= \mathbb{E}\left\{\int_T \mathbb{1}_D(\epsilon(t))\, dt\right\} \\ &= \int_T \mathbb{E}\left\{\mathbb{1}_D(\epsilon(t))\right\}\, dt \\ &= \int_T \mathbb{P}\left\{\epsilon(t) \in D\right\}\, dt. \end{aligned} \tag{4.7.1}$$

If $\epsilon$ is stationary, Gaussian, and has mean zero and unit variance, the last integral is trivial, and so we obtain

$$\mathbb{E}\left\{\lambda_N\left(T \cap \epsilon^{-1} D\right)\right\} = \lambda_N(T)\gamma_k(D).$$

Regardless of the distribution of $\epsilon$, stationary or not, Gaussian or not, even smooth or not, (4.7.1) is true for every random field for which the interchange of orders of integration and expectation is valid[15].

However, volume is just one measure of the size of sets, and we spent a good portion of our time in Chapter 3 setting up an entire family of such measures; viz. Lipschitz-Killing curvatures. An obvious question, therefore, is whether or not one can say as much about the mean Lipschitz-Killing curvatures of excursion sets as one can say about their mean Euler characeristics. The answer to this question is positive, and to see why we start with the simplest scenario, in which the underlying random field $\epsilon$ is isotropic with unit variance and unit second spectral moment. What is special about this case is that the Lipschitz-Killing curvatures appearing in the Gaussian kinematic formula are then the simple Euclidean Lipschitz-Killing curvatures of Section 3.3 measuring the usual volume, surface area, average cross-sectional diameter, and so forth.

Recall Crofton's formula, from Section 3.4.3, which said that, for nice $M$ of dimension $N$,

[15] In fact, since every term in (4.7.1) is non-negative and bounded, this interchange only breaks down in situations of non-measurability, these being situations of such irregularity that they are of absolutely no interest in this book.

$$\int_{\mathrm{Graff}(N,N-k)} \mathcal{L}_j(M \cap V)\, d\lambda^N_{N-k}(V) = \begin{bmatrix} k+j \\ j \end{bmatrix} \mathcal{L}_{k+j}(M), \qquad (4.7.2)$$

the average being taken over the affine Grassmanian $\mathrm{Graff}(N, N-k)$ of all hyperplanes $V$ of dimension $N-k$ in $\mathbb{R}^N$ and the combinatorial flag coefficients are defined by (3.4.3).

Now recall the Gaussian kinematic formula (4.4.2), that

$$\begin{aligned} \mathbb{E}\left\{\varphi\left(T \cap \epsilon^{-1}D\right)\right\} &= \mathbb{E}\left\{\mathcal{L}_0\left(T \cap \epsilon^{-1}D\right)\right\} \qquad (4.7.3) \\ &= \sum_{j=0}^{\dim T} \mathcal{L}_j(T)(2\pi)^{-j/2}\mathcal{M}_j^{\gamma_k}(D). \end{aligned}$$

Noting, as we just mentioned, that the $\mathcal{L}_j$ in (4.7.2) and (4.7.3) are identical, and putting these two results together, with the identification $M = T \cap \epsilon^{-1}D$, and with $j = 0$ in (4.7.2), argue as follows

$$\begin{aligned} \mathbb{E}\left\{\mathcal{L}_k\left(T \cap \epsilon^{-1}D\right)\right\} &= \int_{\mathrm{Graff}(N,N-k)} \mathbb{E}\left\{\mathcal{L}_0\left(T \cap \epsilon^{-1}D \cap V\right)\right\}\, d\lambda^N_{N-k}(V) \\ &= \sum_{l=0}^{N-k} (2\pi)^{-l/2}\mathcal{M}_l^{\gamma_k}(D) \int_{\mathrm{Graff}(N,N-k)} \mathcal{L}_l(T \cap V)\, d\lambda^N_{N-j}(V) \\ &= \sum_{l=0}^{N-k} \begin{bmatrix} k+l \\ l \end{bmatrix} \mathcal{L}_{k+l}(T)(2\pi)^{-l/2}\mathcal{M}_l^{\gamma_k}(D). \end{aligned}$$

Actually, there is no difficulty also carrying out the above argument for a general second spectral moment, and this leads to the following formal statement of the result.

**Theorem 4.7.1 (GKF: General form, under isotropy).** *Retain the setup of Theorem 4.4.1, with the added assumption that $\epsilon$ is isotropic with second spectral moment $\lambda_2$. Then, for every $0 \le j \le \dim(T)$,*

$$\mathbb{E}\left\{\mathcal{L}_j\left(T \cap \epsilon^{-1}D\right)\right\} = \sum_{l=0}^{\dim(T)-j} \begin{bmatrix} j+l \\ l \end{bmatrix} \lambda_2^{(j+l)/2}\mathcal{L}_{j+l}(T)(2\pi)^{-l/2}\mathcal{M}_l^{\gamma_k}(D), \qquad (4.7.4)$$

*where the $\mathcal{L}_j$, on both sides of the equation, are computed with respect to the standard Euclidean metric on $\mathbb{R}^N$.*

By now, you should recognise the terms $\lambda_2^{(j)/2}\mathcal{L}_j(T)$ in the right hand side of (4.7.4) as the Lipschitz-Killing curvatures of $T$, measured with respect to the Riemannian metric induced on $T$ by the components of $\epsilon$. This indicates that a more general version of Theorem 4.7.1 should also be true, and we shall meet it in a moment in the following section.

In the meantime, however, you should note that whereas Theorem 4.7.1 respresents a substantial improvement (in the isotropic case) over Theorem 4.4.1, which treated only mean Euler characteristics, there is nothing on the right hand side of (4.7.4) that we have not met before. *All* of these terms have already been studied in detail, when we looked at the mean Euler characteristic result.

## 4.8 The Gaussian Kinematic Formula, II

We have now finally arrived at a point where we can formulate the most general version of the Gaussian kinematic formula. The full proof is far from simple, and takes up a couple of chapters in *RFG*. Nevertheless, given all that has been done so far in this chapter, the result should by now at least be understandable and believable, even if you decide to bypass the natural temptation to turn to *RFG* to see how the proof goes.

The result is as follows:

**Theorem 4.8.1 (Gaussian kinematic formula: General form).** *Retain the setup of Theorem 4.4.1, Then, for every* $0 \leq j \leq \dim(T)$*,*

$$\mathbb{E}\left\{\mathcal{L}_j\left(T \cap \epsilon^{-1}D\right)\right\} = \sum_{l=0}^{\dim(T)-j} \begin{bmatrix} j+l \\ l \end{bmatrix} \mathcal{L}_{j+l}(T)(2\pi)^{-l/2}\mathcal{M}_l^{\gamma_k}(D), \quad (4.8.1)$$

*where the* $\mathcal{L}_j$*, on both sides of the equation, are computed with respect to the Riemannian metric on* $\mathbb{R}^N$ *induced by the components of* $\epsilon$*.*

The time has finally come to explain from where the term "Gaussian kinematic formula" comes.

Absorbing all the constants in (4.7.4) into constants $C_{jl}$, and writing the expectation as an integral, the equation can be rewritten as

$$\int_\Omega \mathcal{L}_j\left(T \cap (\epsilon(\omega))^{-1}D\right)\, d\mathbb{P}(\omega) = \sum_{l=0}^{\dim(T)-j} C_{jl}\mathcal{L}_{j+l}(T)\mathcal{M}_l^{\gamma_k}(D), \quad (4.8.2)$$

while the classical, Euclidean, kinematic formula (3.4.2) can be written as

$$\int_{G_N} \mathcal{L}_i\left(T \cap g_N D\right)\, d\nu_N(g_N) = \sum_{j=0}^{\dim(T)-i} C'_{jl}\mathcal{L}_{j+l}(T)\mathcal{L}_{N-j}(D), \quad (4.8.3)$$

for some (related) constants $C'_{lj}$.

Written this way, it is clear that the Gaussian kinematic formula and the kinematic fundamental formula must somehow be related, an observation that is very true. In fact, the full, and most elegant proof of the GKF uses the KFF. However, the GKF is not a simple consequence of the KFF. The proof is long

and involved. Furthermore, the GKF actually opens up a completely new class of results in Riemannian geometry. However, these are results in 'pure' mathematics, and this is a book about applications, so we shall say no more about these here.

Before we leave it, there is one more, rather important, aspect to the GKF that deserves noting, and is easily missed at first reading. Recall that $\epsilon$ is a (random) mapping of a $N$-dimensional parameter set into a $k$-dimensional parameter space. We have never said much about the dimension of the set $D$, and there is no need for it to be of full dimension in $\mathbb{R}^k$. If $\dim(D) = d$, then, since $\epsilon$ is smooth,

$$\dim\left(T \cap \epsilon^{-1}D\right) = N + k - d.$$

If $d \neq k$, then $T \cap \epsilon^{-1}D$ will be a submanifold of $\mathbb{R}^N$, and so the information given by its Lipschitz-Killing curvatures is of particular interest.

For example, consider the real valued case in which $k = 1$, and $D$ is a semi-infinite interval $[u, \infty)$. Then the GKF gives information about the mean geometry of the excursion set. However, the GKF can also be applied directly to the boundary $\partial D$ rather than $D$ itself, giving information about the boundary of the excursion set. This kind of problem has been studied in some detail by Wschebor and others (cf. [101]) with techniques somewhat different to those used to study the excursion sets themselves. The GKF, however, includes all these scenarios in a single, unified, result.

In this spirit, we therefore conclude this chapter by noting, once again, that although Theorem 4.8.1 goes far beyond the results on the mean Euler characteristic, there are no terms in the final formula that we have not met earlier in that simpler case. Consequently, no additional work needs to be done in order to apply it.

## 4.9 Exercises

**Exercise 4.9.1.** Derive Rice's formula (2.8.4) as an immediate corollary of Theorem 4.2.1.

**Exercise 4.9.2.** Show that a Gaussian random field $f$ satisfying the conditions of Section 4.1 also satisfies, with probability one, conditions (i) and (ii) of Theorem 3.2.4.
(In fact, it will also satisfy condition (iii) there, but this is somewhat harder to show, and so you should only try this if you are very keen.)

**Exercise 4.9.3.** Suppose $f$ is a centered, isotropic Gaussian random field on $\mathbb{R}^N$ with unit variance and $\mathrm{Var}(\nabla f(t)) = I$. Using Wick's formula (cf. Exercise 2.8.3) show that

$$\mathbb{E}\left\{\mathrm{detr}_j(-\nabla^2 f - fI)\right\} = \begin{cases} 0 & j \text{ is odd,} \\ \frac{(-1)^m N!}{(N-2m)!m!\,2^m} & j = 2m \text{ is even.} \end{cases}$$

Hint: Use the results of parts (iii) and (iv) of Exercise 2.8.10.

**Exercise 4.9.4.** Prove the general formula (4.2.2) without the assumptions $\sigma = 1$ and $\Lambda = I$, but using the fact that the result holds for this case.
Hint: Show first that

$$\mathbb{E}\left\{\varphi\left(A_u(f,T)\right)\right\} = \sum_{j=0}^{N} \mathcal{L}_j(\widetilde{T}) \frac{H_{j-1}(u/\sigma) e^{-u^2/2\sigma^2}}{(2\pi)^{(j+1)/2}},$$

where $\widetilde{T}$ is the paralellogram

$$\widetilde{T} = \left\{ \Lambda^{1/2} t/\sigma : t \in T \right\}$$

and $\Lambda^{1/2}$ is any square root of $\Lambda$. Then apply Exercise 3.6.4 (iii) to complete the proof.

**Exercise 4.9.5.** Use Hadwiger's theorem, Theorem 3.3.2, to show that the expected Lipschitz-Killing curvatures of the excursion set of an isotropic random field (not necessarily Gaussian) over a parameter set $T \in \mathbb{R}^N$ must be expressible as a linear combination of the Lipschitz-Killing curvatures of $T$.

**Exercise 4.9.6.** Prove the claim that, for stationary processes, the Lipschitz-Killing curvatures $\mathcal{L}_j$ appearing in the Gaussian kinematic formula have the form given by (4.3.5).

**Exercise 4.9.7.** Consider the non-central $\chi^2_k$ random field

$$f(t) = |\epsilon(t) + \alpha|^2, \tag{4.9.1}$$

with non-centrality parameter $\alpha \in \mathbb{R}^k$. Compute the EC densities of this random field.
Hint: Use the fact that the density $p_{\lambda,k}$ of the square root of a non-central $\chi^2_k$ random variable can be written as

$$p_{\lambda,k}(x) = \sum_{j=0}^{\infty} e^{-\alpha^2/2} \frac{\alpha^j}{2^j j!} p_k(x),$$

where $p_k$ is the density of a square root of a standard $\chi^2_k$ random variable.

**Exercise 4.9.8.** Carry out the suggestion of Footnote 7.

**Exercise 4.9.9.** For any subset $A \subset S^{k-1}$ of the unit sphere in $\mathbb{R}^k$ and $z > 0$, define

$$R_{A,z} = \left\{ y \in \mathbb{R}^k : \sup_{x \in A} \langle x, y \rangle^+ \geq z \right\}.$$

Show that, for $0 < \rho < u$

$$\mathrm{Tube}(R_{A,z}, \rho) = R_{A,z-\rho}.$$

**Exercise 4.9.10.** Show that the $F$ EC densities (4.5.23) can be written in terms of the Student $T$ EC densities as

$$\rho_d^{\mathrm{F}}(u;n,m) = \sum_{j=0}^{n-1} \mathcal{L}_j(S^{n-1})\, \rho_d^{\mathrm{T}}(\sqrt{un};m).$$

**Exercise 4.9.11.** There is an interesting result in Chapter 13 of *RFG* that we called a *Gaussian Crofton formula*. It is a version of the standard Crofton formula in which averages over intersecting hyperplanes are replaced by Gaussian averages over certain random submanifolds.

To state it, let $M$ be a nice set, $\epsilon : M \to \mathbb{R}^k$ as throughout the chapter, and $Z \sim N(0, I_{k\times k})$ independent of $\epsilon$. Furthermore, for $u \in \mathbb{R}^k$ define the (random) submanifold

$$D_u = \{t \in M : \epsilon(t) = u\}.$$

Then, for $0 \le j \le \dim(M) - k$,

$$\mathbb{E}\{\mathcal{L}_j(M \cap D_Z)\} = (2\pi)^{-k/2} \frac{[k+j]!}{[j]!} \mathcal{L}_{k+j}(M), \tag{4.9.2}$$

where the $\mathcal{L}_j$ on both sides of the equation are computed with respect to the Riemannian metric induced on $M$ by the components of $\epsilon$ and $[j]!$ is defined at (3.4.3). (Note that the expectation here is actually a double expectation, over both $\epsilon$ and $Z$.)

Use this result to prove the full form of the Gaussian kinematic formula, Theorem 4.8.1, from the basic version, Theorem 4.4.1.

Need more exercises here. Could include as exercises some of the EC densities that we quote but do not derive, or are these too hard? Also mean number of zeroes of $f : \mathbb{R}^N \to \mathbb{R}^N$.

# 5

# Exceedence Probabilities

In the previous chapter we spent a lot of time developing precise formulae for the mean values of various geometric characteristics of the excursion sets

$$A_u = A_u(f,T) = \{t \in T : \, f(t) \geq u\} = T \cap f^{-1}[u, +\infty),$$

of real valued Gaussian random fields. We also considered the extensions of these results to vector valued Gaussian fields, looking at excursion sets of the form

$$A_D = A_D(f,T) = \{t \in T : \, f(t) \in D\} = T \cap f^{-1}(D).$$

The corresponding result, the Gaussian kinematic formula of Sections 4.4 and 4.8, allowed us to also treat a wide variety of non-Gaussian fields.

These formulae are at the core of many applications of random field theory, primarily because, aside from a handful of special case results, *they are essentially the only closed form expressions holding in broad generality in the subject*, an important fact that it is hard to over-emphasise.

In this and the following chapter we plan to tackle a number of other objects of interest, among them the exceedence probabilities

$$\mathbb{P}\Big\{ \sup_{t \in T} f(t) \geq u \Big\} \tag{5.0.1}$$

and the mean number of local maxima $\mathbb{E}\{M_u\}$, where

$$M_u \;=\; M_u(f,T) \;=\; \#\{t \in T : \, t \text{ is a local maximum of } f\}. \tag{5.0.2}$$

What we shall see in both these cases is that, for large $u$, and smooth, real valued, centered Gaussian fields $f$, both of the above are very, very close to the mean Euler characteristic of $A_u$. To explain the extra "very' consider the exceedence probability (5.0.1). We shall see below, in Theorem 5.3.1, that

$$\left| \mathbb{P}\Big\{ \sup_{t \in T} f(t) \geq u \Big\} - \mathbb{E}\left\{ \varphi\left( A_u(f,T) \right) \right\} \right| \;<\; O\left( e^{-\alpha u^2 / 2\sigma^2} \right), \tag{5.0.3}$$

where $\sigma^2$ is the variance of $f$ (assumed constant) and $\alpha > 1$ is a constant that we shall often be able to identify. The explicit expressions for $\mathbb{E}\left\{\varphi\left(A_u(f,T)\right)\right\}$ developed in Chapter 4 show us that we can rewrite (5.0.3) as

$$\mathbb{P}\Big\{\sup_{t\in M} f(t) \geq u\Big\} = C_0\Psi\left(\frac{u}{\sigma}\right) + e^{-u^2/2\sigma^2}\sum_{j=1}^{N} C_j u^{N-j} + o\left(e^{-\alpha u^2/2\sigma^2}\right), \tag{5.0.4}$$

where the $C_j$ are constants depending on the parameters of $f$ and the geometry of $T$, and $N = \dim(T)$.

The final term in this expression is quite remarkable, for, if we think of the right hand side as an expansion of the form

$$C_0\Psi\left(\frac{u}{\sigma}\right) + e^{-u^2/2\sigma^2}\sum_{j=1}^{N} C_j u^{N-j} + \text{error}, \tag{5.0.5}$$

it would be natural to expect that the error term here would be the 'next' term of what seems like the beginning of an infinite expansion for the exceedence probability, and so of order $u^{-1}e^{-u^2/2\sigma^2}$. However (5.0.5) indicates that this is *not* the case. Since $\alpha > 1$, the error is actually *exponentially smaller* than this. Hence the "very, very close" above.

Remarkably, and far from coincidentally, an almost identical result to (5.0.3) holds when the exceedence probability is replaced by $\mathbb{E}\{M_u\}$.

There is a long history of formulae like (5.0.5), with an excellent treatment of the results and techniques in Piterbarg's monograph [75] and a more recent treatment in [13]. However, what is common to almost all of this history is the direction of proof, which aims to establish a formula like (5.0.5) directly, without linking the object of study to the mean Euler characteristic, the only expectation for which, we re-emphasise, a closed form, explicit expression is known.

Despite our intention to rely on excursion set techniques, we would be at the very least doing the reader a disservice, and at worst bordering on the misleading, if we did not emphasise now that excursion set techniques will only work when the underlying random fields are smooth enough for this theory to hold. In particular, as we demanded throughout Chapter 4, $f$ must be at least $C^2$. Since these are the random fields of most interest to us, we can replace all other techniques with those based on excursion sets. However, since not all random fields are $C^2$, and you may occassionally run across one or two, we shall briefly discuss some the other main techniques for computing exceedence probabilities at the end of the chapter.

Until then, however, we return to the main aim of this chapter, which is to present a number of approximations, all linked, in one way or another, to the mean Euler characteristic of excursion sets and to show, generally not in full detail, why the approximations work. It is important to note, however, that

while we shall not give all the details, there is no reason (other than tedium) why they cannot be given[1]. On the other hand, in the third part of this book we shall move from mathematically justifiable approximations to heuristic approximations, justified only on the basis that in real world applications one sometimes has to cut corners in order to develop procedures which seem to work most of the time.

We shall start by looking at the mean numbers of various types of critical points of Gaussian random fields.

## 5.1 The Mean Number of Maxima

Much of what we shall investigate in this chapter is related to the critical points of real valued random fields, viz. those points $t$ in the parameter space $T$ for which

$$\nabla f(t) = 0. \tag{5.1.1}$$

Among these, local maxima are characterised by the fact that the Hessian matrix $\nabla^2 f(t)$ is negative definite, while for local minima $\nabla^2 f(t)$ is positive definite[2].

It is not hard to write out an expression for the expected number of critical points $C = C(T)$ of $f$, assuming that appropriate regularity conditions apply. (cf. Exercise 5.6.1.) In fact, Theorem 2.7.1 easily gives us that

$$\mathbb{E}\{C(T)\} = \int_T \mathbb{E}\left\{ \left|\det \nabla^2 f(t)\right| \,\middle|\, \nabla f(t) = 0 \right\} p_{\nabla f(t)}(0)\, dt. \tag{5.1.2}$$

where $p_{\nabla f(t)}(0) = (2\pi)^{-N/2}$ is the density of $\nabla f(t)$ at 0.

If you read the details when we derived an expression for the mean value of the Euler characteristic of excursion sets back in Section 4.2, then the above expectation should look rather familiar. However, unlike that case, here it turns out to be effectively impossible to compute the expectation within the integral. The problem lies in the seemingly innocuous absolute value sign on the determinant of the Hessian. In the Euler characteristic case there was an alternating sum over different types of critical points (cf. (4.2.4)) that allowed us to drop the absolute value sign. Here we have no such luck.

Considering local maxima rather than critical points is no simpler, for writing $M = M(T)$ ($= M_{-\infty}(T)$ in the notation of (5.0.2)) (5.1.2) becomes

[1] For example, for a full proof of (5.0.3), at the level of generality at which $T$ is a stratified manifold, see Chapter 14 of *RFG*. However, be prepared for 30 pages of detailed technical arguments.

[2] Of course, $f$ will generally also have local maxima and minima on the boundary of $T$, which will not be critical points in the sense of (5.1.1). We shall treat these later, restricting ourselves for the moment to those local maxima and minima which are in the interior of $T$ and so critical points in the sense of (5.1.1).

$$\mathbb{E}\{M(T)\} = -\int_T \mathbb{E}\left\{\det\nabla^2 f(t)\mathbb{1}_{\mathcal{N}}(\nabla^2 f(t))\Big|\nabla f(t) = 0\right\} p_{\nabla f(t)}(0)\, dt \quad (5.1.3)$$

where $\mathcal{N}$ is the set of negative definite matrices. While we have now removed the absolute value sign, we have restricted the expectation in a way that makes it virtually impossible to compute.

Nevertheless, there is one case for which it is in fact possible to compute $\mathbb{E}\{M(T)\}$, which goes back to Longuet-Higgins [3] [60]. The case is that of centered, stationary, Gaussian $f$ over nice, two-dimensional domains $T$. We shall describe the result here, without attempting to prove it. A proof can be found in the original paper of Longuet-Higgins or, in a notation closer to ours, in [3]. In both cases the derivation, while not hard in principle, is a lot of work in practice, and amounts to a brute force computation of the right hand side of (5.1.3). The fact that this computation has, to the best of our knowledge, never been completed for any dimension greater than two should give an indication of its complexity, as should the actual form of the result.

To state the result, we need some notation. With $\Lambda$ be the usual matrix of of second order spectral moments, we let $V$ be the covariance matrix of the three distinct elements of $\nabla^2 f$, in the form

$$V = \begin{pmatrix} v_{40} & v_{31} & v_{22} \\ v_{31} & v_{22} & v_{13} \\ v_{22} & v_{13} & v_{04} \end{pmatrix}.$$

where the $v_{ij}$ are the fourth order spectral moments $\int_{\mathbb{R}^2} \lambda_1^i \lambda_2^j \, \nu(d\lambda)$. (cf. (2.4.9).) Next, let $A$ be the matrix given by

$$T = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & -1 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix}.$$

and let $d_1 > 0 > d_2 \geq d_3$ be the eigenvalues of $VT$, also given by the roots of (Why roots is you are giving them as explicit expressions?)

$$d_i = H^{1/2} \cos\theta_i,$$

where $\theta_1, \theta_2, \theta_3$ are the roots of

$$\cos 3\theta = \frac{\det V}{H^{3/2}}$$

and

$$H = \left(v_{40}v_{04} - 4v_{31}v_{13} + 3v_{22}^2\right)/3.$$

[3] Starting in the 1950's, Longuet-Higgins began a long and important series of papers on the modelling of sea waves as random fields which, still today, are a great place to search for specific formulae like (5.1.4). However, like (5.1.4), they are almost always limited to the stationary, two-dimensional case.

Now let $E$ and $F$ denote the Legendre elliptic integrals of the first and second kind, so that

$$E(k,\theta) = \int_0^\theta (1 - k^2 \sin^2 \alpha)^{1/2} d\alpha, \qquad F(k,\theta) = \int_0^\theta (1 - k^2 \sin^2 \alpha)^{-1/2} d\alpha.$$

Finally, define the function

$$G(\alpha) = [\alpha(1-\alpha)]^{1/2} \left[ \left( \frac{1+\alpha}{\alpha} \right)^{1/2} E(\alpha', \tfrac{1}{2}\pi) - \left( \frac{\alpha}{1+\alpha} \right)^{1/2} F(\alpha', \tfrac{1}{2}\pi) \right],$$

where

$$(\alpha')^2 = \frac{1-2\alpha}{1-\alpha^2}, \qquad 0 < \alpha \le \tfrac{1}{2}.$$

Then

$$\mathbb{E}\{M(T)\} = |T| \, \frac{d_1}{2\pi^2 |\Lambda|^{1/2}} G\left( -\frac{d_2}{d_1} \right). \tag{5.1.4}$$

If $f$ is also isotropic, then (5.1.4) simplifies considerably to

$$\mathbb{E}\{M(T)\} = |T| \, \frac{1}{6\pi\sqrt{3}} \frac{\nu_4}{\lambda_2}, \tag{5.1.5}$$

where $\lambda_2$ and $\nu_2 = v_{ii}$ are, respectively, the variances of any of the first and second order partial derivatives of $f$.

Note that we did not make any assumptions about the mean and variance of $f$ (other than being constant by stationarity) in formulating the above results, since they do not, in any way, impact on the total number of local maxima (they do not appear in (5.1.4)). They will, however, impact on $M_u(T)$, the number of local maxima above the level $u$.

It is easy to see, and it is Exercise 5.6.2 that, in both cases, (WHICH BOTH CASES?) the mean number of minima is equal to the mean number of maxima, while the mean numbers of saddle points and critical points are, respectively, two and four times this.

Carrying the arguments that give (5.1.4) and (5.1.5) over to higher dimensions has, to the best of our knowledge, never been fully successful, despite many valiant attempts. The three dimensional case is, however, amenable to some calculations, and some advance on computing $\mathbb{E}\{M(T)\}$ and $\mathbb{E}\{M_u(T)\}$ in this case been made in the astrophysics literature. The final answer is rather involved, and ultimately requires numerical integrations and approximations of integrands in order to get actual numbers. Details can be found, for example, in [15, 29], which also look at other expectations in the three dimensional, Gaussian setting.

Rather than describe further what the astrophysicists have done, we shall now begin to introduce our own approximations. These are somewhat simpler, describe better what is actually going on, yield approximations which are at least as good, and work equally well in all dimensions.

## 5.2 The Mean Number of Maxima and the Mean Euler Characteristic

A more interesting random variable than the total number of local maxima of a random field is the number of local maxima above a fixed level $u$. Regarding this, we have the following theorem.

**Theorem 5.2.1.** *Let $f$ be a centered Gaussian field on $\mathbb{R}^N$ satisfying the conditions of Section 4.1 and with constant variance $\sigma^2$, and let $T$ be a nice[4] subset of $\mathbb{R}^N$. Let $M_u(T)$ denote the number of local maxima, in $T$, of $f$ above the level $u$, be they critical points in the interior of $T$ or local maxima on the boundary. Then there are constants $\alpha > 1$, depending on the covariance function of $f$, and $C$, depending on the covariance and the geometry of $T$, such that, for large enough $u$,*

$$|\mathbb{E}\{M_u(T)\} - \mathbb{E}\{\varphi(A_u(T)\}| \leq Ce^{-\alpha u^2/2\sigma^2}. \tag{5.2.1}$$

Before proving this result, there are a number of remarks that are worth making. The first is that we shall not attempt to say anything about the magnitudes of the constants $C$ and $\alpha$. If you are interested, then you can find more information in Chapter 6 of Azaïs and Wschebor [13], where you can also find information ~~of~~ on the related distribution of $\sup_T f$, something about which we shall have more to say in Section 5.3. Not surprisingly, in order to be more explicit about the constants, one needs to make more assumptions on the random field $f$.

The second remark is merely to reiterate the opening comments of this chapter about how small is the difference in (5.2.1); viz. super-exponentially small.

The third and final remark is to give a heuristic argument as to why (5.2.1) might be true. Given a high enough level, $u$, it is, of course, unlikely that $f$ will exceed it anywhere within $T$. If it nevertheless does manage to do so, it is even more unlikely that it will do so in two or more separate regions. In fact, if we think of excursions as rare events, divide $T$ into two disjoint regions $T_1$ and $T_2$, then, to a first order approximation, we should have

$$\begin{aligned}\mathbb{P}\left\{\sup_{T_1} f \geq u,\ \sup_{T_2} f \geq u\right\} &\simeq \mathbb{P}\left\{\sup_{T_1} f \geq u\right\}\mathbb{P}\left\{\sup_{T_2} f \geq u\right\} \\ &= O\Big(\Big(\mathbb{P}\Big\{\sup_{T} f \geq u\Big\}\Big)^2\Big).\end{aligned} \tag{5.2.2}$$

In Section 5.4.2 we shall discuss making this approximation rigorous and indicate how it can be properly exploited, but, for the moment, we shall merely

[4] We use the same notion of 'nice' here that we did when computing the expected Euler characteristic of excursion sets in Theorem 4.3.1. See Footnote 3 of Chapter 4. In particular, we remind you that $T$ must be locally convex.

apply the basic argument to bolster our claim that if there is an exceedence of $f$ above $u$, it is likely to be localised in $T$.

Furthermore, such an exceedence will tend to have a rather simple shape, a point that we shall prove when discussing Slepian models in Section 6.2. This shape will be approximately parabolic, which will have two consequences:

(i) There will only be one local maximum over the excursion set, and there will be no other critical points.
(ii) The base of this parabolic excursion, which is the excursion set itself, will be a simple, convex, set, of Euler characteristic one.

Hence, for large $u$, with high probability, we have one of the two scenarios,

$$\sup_{t\in T} f(t) < u \Rightarrow M_u(T) = 0 = \varphi(A_u(T),$$
$$\sup_{t\in T} f(t) \geq u \Rightarrow M_u(T) = 1 = \varphi(A_u(T),$$

giving us, with much hand waving, that $\mathbb{E}\{M_u(T)\}$ and $\mathbb{E}\{\varphi(A_u(T)\}$ should be close.

To prove that this is indeed the case, and that $\mathbb{E}\{M_u(T)\}$ and $\mathbb{E}\{\varphi(A_u(T)\}$ are as close as (5.2.1) indicates, requires some work, which we now start with an important technical lemma.

**Lemma 5.2.2.** *Let $V$ be an $k \times k$ symmetric matrix and let $Y$ the vector of length $K = k(k+1)/2$ formed by placing the successive columns on and above the main diagonal of $V$ under one another. Let $\Lambda$ be a diagonal matrix with positive entries on the diagonal and let $D_x$ be the region in $\mathbb{R}^K$ defined by*

$$Y \in D_x \iff V - x\Lambda \text{ is negative definite.}$$

*Then there exists a finite $C$ dependent only on $\Lambda$, such that, for each $0 \leq r < Cx$, $D_x$ contains $B_K(r)$, the ball of radius $r$ in $\mathbb{R}^K$.*

**Proof.** Let $P$ be an orthogonal matrix such that $P'\Lambda P = I$, (No such $P$ exists in general, so the proof is wrong as it is. However, the lemma is only applied later with $\Lambda = \lambda_2 I$. If the statement of the lemma is changed to $\Lambda$ of the form $\lambda I$ then the proof works with $P = \lambda^{-1/2} I$, but then it can further simplified. Otherwise, I can offer an alternative proof for general diagonal $\Lambda$, but it seems unnecessary.) and $Q$ an orthonormal matrix reducing $P'VP$ to diagonal form, i.e.

$$Q'(P'VP)Q = \text{diag}(d_1, \ldots, d_k).$$

Then

$$(PQ)'(V - x\Lambda)(PQ) = \text{diag}(d_1 - x, \ldots, d_k - x).$$

Since the expression on the left-hand side will be a negative definite matrix if and only if $V - x\Lambda$ is , it follows that $V - x\Lambda$ will be negative definite provided $x > \max_j d_j$.

Since $Q$ is orthonormal, it is now easy to check that there exists a constant $c$, dependent only the elements of $P$, and so only on $\Lambda$, such that, for all $1 \le i \le N$,

$$|d_i| \le c \sup_{m,n} |v_{mn}|.$$

Thus, if $x > rc$ and $Y \in B(r)$, we have that $u > \max_j d_i$, implying that $V - x\Lambda$ is negative definite, which, with $C = c^{-1}$, proves the lemma. □

**Proof of Theorem 5.2.1.** In order to keep the proof simple, we shall make two sets of of additional assumptions. The first set require that $f$ is stationary and isotropic, with zero mean, constant unit variance, and second spectral moment $\lambda_2$. The second is that the parameter space $T$ is a rectangle in $\mathbb{R}^N$. If you read through the proofs of the previous chapter you will already know how to lift these assumptions to obtain a more general proof. If you have not (yet) read those proofs, then you should. In either case, the restrictions are of no significant consequence, since we are not going to try to identify the constants in the bound (5.2.1).

In this restricted setting, and recalling the results of Section 4.2, in particular (4.2.4), we can write $\varphi(A_u(f,T))$ as the alternating sum of numbers $\mu_i(J)$ of different types of critical points over the faces $J$ of $T$. In particular,

$$\varphi(A_u(f,T)) = \sum_{k=0}^{N} \sum_{J \in \partial_k T} \sum_{i=0}^{k} (-1)^i \mu_i(J), \tag{5.2.3}$$

where all terms are defined in (4.2.5)–(4.2.8).

An important fact, that we did not emphasise in Section 4.2, is that $\mu_{\dim J}(J)$ is always the number of local maxima of $f$ in $J$ above the level $u$. That is,

$$\{k = \dim(J)\} \quad \Rightarrow \quad \{\mu_k(J) = M_u(J)\},$$

for all faces $J$ of $T$. Consequently, if we separate the $k$-th term in the inner sum in (5.2.3), then the second term is bounded above by $M_u(T)$ and so

$$|\varphi(A_u(T)) - M_u(T)| \le \sum_{k=0}^{N} \sum_{J \in \partial_k T} \sum_{i=0}^{k-1} \mu_i(J).$$

The inner sum here is just the number of critical points of $f_{|J}$, above the level $u$, which are *not* local maxima, and so if we can prove that this has expectation of order $e^{-\alpha u^2/2}$ for some $\alpha > 1$ we will be done. But this is quite easy.

For the remainder of the argument fix a $J$, with $\dim(J) = k$. For notational convenience, renumber, if necessary, the free variables in $J$ as $t_1 \ldots, t_k$. It then follows immediately from Theorem 2.7.1 that, up to a multiplicative factor of $|J|$, the expectation we seek is equal to

$$\mathbb{E}\left\{ \left|\det \nabla^2 f_{|J}\right| \mathbb{1}_{\mathcal{N}_k^c}\left(\nabla^2 f_{|J}\right) \,\middle|\, \nabla f_{|J} = 0 \right\} \tag{5.2.4}$$
$$= \int_u^\infty dx \int_{\mathcal{N}_k^c} |\det \nabla^2 f_J| p(x, 0, y)\, dy,$$

where, in the first expression, $\mathcal{N}_k$ is the collection of $k \times k$ negative definite matrices, so that the indicator is over the complement of this set. Similarly, in the second expression, we let $Y$ be the random vector of length $K = k(k+1)/2$ made from the distinct elements of $\nabla^2 f_{|J}$, and then we interpret $\mathcal{N}_k$ to be the region of $y \in \mathbb{R}^K$ for which the corresponding matrix is negative definite. The density $p$ is then the joint density of $(f, \nabla f, Y)$.

To bound the integral, recall our simplifying assumption of isotropy, which makes the elements of $\nabla f$ uncorrelated and of common variance $\lambda_2$. Then Exercise 2.8.10 and, in particular, (2.8.1), give that $f$ and $Y$ are independent of $\nabla f$, and that the conditional mean of an $f_{ij}$ of $\nabla^2 f$ given $f = x$ is $-\lambda_2 x$. Consequently, making the change of variables

$$\widehat{f}_{ij} = f_{ij} + \delta_{ij}\lambda_2,$$

where $\delta_{ij}$ is the Kronecker delta, we can rewrite the integral in (5.2.4) as

$$\int_u^\infty p_1(x)dx \int_{D_x^c} \det\left(\nabla^2 \widehat{f}_{|J} - x\lambda_2 I\right) p_2(\widehat{y})\, d\widehat{y}, \tag{5.2.5}$$

where $\widehat{Y}$ bears the same relation to $\widehat{f}$ that $Y$ does to $f$, and $D_x$ is the region of $\widehat{y}$'s over which $\nabla^2 \widehat{f} - x\lambda_2 I$ is negative definite. Furthermore, $p_1$ is a standard Gaussian density and $p_2$ a multivariate Gaussian density of zero mean variables with variances that depend on fourth order spectral moments but which are unimportant for this proof. Note, from Lemma 5.2.2, that for some $C$ dependent only on $\lambda_2$, the region $D_x^c$ lies outside a ball of radius $Cx$.

Now expand the determinant in (5.2.5) in powers of $x$, so that the inner integral can be bounded by a sum of terms of the form

$$x^i \lambda_2^i \int_{B_K^c(Cx)} \widehat{y}_{j_1} \ldots \widehat{y}_{j_{k-i}} p_2(\widehat{y})\, d\widehat{y}.$$

It is now a simple exercise in Gaussian integrals to see that each such integral is bounded by an expression of the form $x^\beta e^{-\gamma x^2}$, where $\beta \le 2k$ and $\gamma > 0$ (and is dependent on the fourth order spectral moments).

Substituting this into (5.2.5) and carrying out the remaining $x$ integration completes the proof. □

We noted during the above proof that the bounds we obtained depended on fourth order spectral moments – the variances of the second order partial derivates of $f$ – despite the fact that we neglected to follow the details of this dependence. Doing so would have lead us closer to the results described above of [13], in which constants similar to those of Theorem 5.2.1 are identified.

In the next section we shall rely on some results from *RFG* to identify constants in a related, and actually more important, problem.

Before doing so, however, it is worth pointing out that a wide class of other results also follow from the above arguments and, in particular, from Lemma 5.2.2. The application of Lemma 5.2.2 in the above proof was, in essence, to establish the fact that if, at a point $t$, we had $f(t) > u$ for a large $u$, then, with high probability, the Hessian $\nabla^2 f(t)$ would be negative definite. In other words, at high levels the function $f$ will tend to be concave. Exercise 5.6.3 gives a simple example of this. A natural consequence of this concavity is that the individual connected components of the excursion sets $A_u$ will tend to be convex, for which Exercise 5.6.4 provides a simple example. All of this, however, will be made more precise in Section 6.2.

## 5.3 Expected Euler Characteristic Heuristic

The term "EEC (expected Euler characteristic) heuristic" is, nowadays, and at least as far as constant variance Gaussian fields are concerned, somewhat of a misnomer.

Originally, the EEC heuristic followed the lines of our heuristic explanation of Theorem 5.2.1 in the previous section, and was used to argue that, at least for Gaussian random fields, the EEC of excursion sets $A_u$ was close to the exceedence probability $\mathbb{P}\{\sup f \geq u\}$. Theorem 5.3.1, which is proven in Chapter 14 of *RFG*, shows that what was once the result of a heuristic argument is now a fully fledged theorem.

Nevertheless, we shall retain the terminology "EEC heuristic", since later in the book we shall claim that a similar result must hold also for a wide range of non-Gaussian, despite the lack of a detailed proof. Outside the purely Gaussian scenario, however, the bound of Theorem 5.3.1 will look somewhat difference. See, for example, Section 5.3.2, where we shall look at $\chi^2$ random fields, where we can, once again, say something precise and rigorous.

Here is the main Gaussian result, which, although it is stated for random fields on $\mathbb{R}^N$, actually holds in wider generality.

**Theorem 5.3.1.** *As in Theorem 5.2.1, let $f$ be a centered Gaussian field on $\mathbb{R}^N$ satisfying the conditions of Section 4.1 and with constant variance which, for convenience, we take to be one. Let $T$ again be a nice subset of $\mathbb{R}^N$. Then there exists a constant $\sigma_c^2 > 0$, dependent on the distribution of $f$ and the geometry of $T$, such that*

$$\liminf_{u\to\infty} -u^{-2}\log\left|\mathbb{E}\left\{\varphi\left(A_u(f,M)\right)\right\} - \mathbb{P}\left\{\sup_{t\in M} f(t)\geq u\right\}\right| \;\geq\; \frac{1}{2}\left(1+\frac{1}{\sigma_c^2}\right). \tag{5.3.1}$$

Equation (5.3.1) is a little more formal that its analogue (5.2.1) in Theorem 5.2.1, but it is trivial to unravel it to the same (weaker) form, which states that there exist constants $C$ and $\alpha > 1$ such that, for large enough $u$,

$$\left|\mathbb{E}\left\{\varphi\left(A_u(f,M)\right)\right\} - \mathbb{P}\left\{\sup_{t\in M} f(t)\geq u\right\}\right| \leq Ce^{-\alpha u^2/2}.$$

The parameter $\sigma_c^2$ of (5.3.1) actually has a generic characterisation, which we shall allude to below when we shall have something to say about the proof of Theorem 5.3.1. For details you will have to turn to *RFG*. Nevertheless, there are many examples given there for which it is easy to determine $\sigma_c^2$ in terms of the spectral moments of $f$.

For example, if $f$ is a centered, unit variance $C^2$ stationary Gaussian process on $[0,T]$ satisfying the conditions of the theorem and such that $\mathrm{Var}(\dot{f}(t)) = -\ddot{\rho}(0) = 1$. Then, for all $T > 0$,

$$\sigma_c^2 = \sup_{0\leq t\leq T} \frac{1 - C^2(t) - \dot{C}^2(t) + (\max(\dot{C}(t),0))^2}{(1-C(t))^2}, \tag{5.3.2}$$

where $C$ is the covariance (and correlation) function of $f$.

While this is a reasonably simple expression to compute, it simplifies considerably if we are prepared to assume that $C$ is monotone over $[0,T]$. In that case the max term disappears in (5.3.2) and the supremum is achieved at $t = 0$.

JONATHAN: CAN YOU SEE AN EASY WAY TO PROVE THIS CLAIM IN THE CURRENT RESTRICTED SETTING? THE PROOF IN THE BOOK IS HORRID AND TOO GENERAL.

Actually computing the ratio at $t = 0$ requires four applications of l'Hôpital's rule, which give

$$\sigma_c^2 = \mathrm{Var}\left(\frac{d^2 f(t)}{dt^2}\Big| f(t)\right) \;=\; \frac{d^4 C(t)}{dt^4}\Big|_{t=0} \;-\; 1. \tag{5.3.3}$$

In fact, a similar result holds if $f$ is an isotropic field on $\mathbb{R}^N$, with unit variance and second spectral moments and covariance decreasing monotonely ~~fcom~~ from the origin, and the derivatives on the right hand side of (5.3.3) are replaced with partial derivatives in any of the $N$ directions, viz.

$$\sigma_c^2 = \mathrm{Var}\left(\frac{\partial^2 f(t)}{\partial t_1^2}\Big| f(t)\right) \;=\; \frac{\partial^4 \rho(t)}{\partial t_1^4}\Big|_{t=0} \;-\; 1. \tag{5.3.4}$$

For other examples see *RFG* and Exercise 5.6.5.

### 5.3.1 About the proof

The full proof of Theorem 5.3.1 is long and technical, but it is worthwhile seeing how to at least start it, since the basic idea is a little surprising.

In fact, the basic idea is to turn the exceedence probability into an event involving a specific point process, and then use the Rice-Kac expectation meta-theorem of Section 2.7 to compute the mean number of points involved.

To start, with $C$ the covariance (and correlation) function of $f$ assume that $C(s,t)$ is *strictly* less than one for all $s \neq t$ and, for all $t \in T$ define the random field $f^t$ on $T$ by setting

$$f^t(s) \stackrel{\Delta}{=} \begin{cases} \frac{f(s)-C(s,t)f(t)}{1-C(s,t)} & s \neq t, \\ f(s) & s = t. \end{cases}$$

The motivation for defining this field is that, for each $t \in T$, $f(t)$ and $f^t(s)$ are independent for all $s \neq t$. Then note that points $t \in T$ will be local maximers of $f$ above the level $u$ if, and only if, they satisfy the following three conditions:

(i) $f(t) \geq u$.
(ii) $\nabla f(t) \in N_t$, where $N_t$ is the normal cone to $T$ at $t$. Thus $t$ is an extended outward critical point of $f$. (cf.(**??**))
(iii) $f(t) \geq \sup_{s \in T \setminus \{t\}} f^t(s)$.

If $t$ is the maximizer of $f$, then the necessity of conditions (i) and (ii) is obvious. As for the final condition, since $t$ is a maximizer we have $f(s) < f(t)$ for all $s \neq t$, and since it is also true that and $C(s,t) < 1$, it follows that

$$f^t(s) = \frac{f(s) - C(s,t)f(t)}{1 - C(s,t)} < f(t).$$

The reverse implication, that (i)–(iii) imply $t$ is a maximiser, is no harder and is left to you.

Assume now that $f$ has at most one[5] global maximum in $T$ above the level $u$. Then the expected number of points satisfying conditions (i)–(iii) is precisely the probability that $\sup f \geq u$, and so we have reduced the exceedence probability computation to one involving the mean number of points of a point process, viz.

$$\mathbb{P}\{\sup f(t) \geq u\} = \mathbb{E}\{\#\{t \in T : \text{ (i)–(iii) are satisfied for } t.\}\}.$$

In principle, this expectation can now be computed using the Rice-Kac meta-theorem. In practice, this is not so easy and approximations need to be made along the way.

[5] It is not hard to show that this must happen, with probability one, under the conditions that we are assuming. The argument is not dissimilar to that you would have used for Exercise 2.8.26.

The details of the argument rely on two main phenomena. The first is an application of arguments using facts like Lemma 5.2.2 (Do you mean Theorem 5.2.1?) which help link mean critical point counts to mean Euler characteristics.

The second uses the easily checked fact already mentioned that, for each $t \in T$, $f(t)$ and $f^t(s)$ are independent for all $s \neq t$. Hence the random variables $f(t)$ and $\sup_{s \in T \setminus \{t\}} f^t(s)$ in (iii) are also independent, and bounds on the tail of the second of these follow from the Borell-TIS inequality of Section 2.6.1.

The details (including why, despite the above promising start, $f^t$ is actually not a good process with which to work) which are unfortunately long and tedious, can be found in *RFG*.

### 5.3.2 $\chi^2$ and $\bar{\chi}^2_K$ Fields

As one might guess from even the sketchy proof given above, it is not easy to prove results like Theorem 5.3.1 for general, non-Gaussian random fields.

However, there are two classes of Gaussian related fields for which it is possible to derive a rigorous parallel, these being the $\chi^2$ fields of Section 4.5.2 and the $\bar{\chi}^2_K$ fields of Section 4.5.3. Recall that, in developing an explicit formula for the mean Euler characteristics of their excursion sets, we linked each of them to Gaussian processes. Rewriting (4.5.5) and (4.5.9) in an obvious notation, we showed the existence of Gaussian fields $Z_1$ and $Z_2$ for which

$$\sup_{t \in M} \chi^2(t) \geq u \quad \Longleftrightarrow \quad \sup_{(t,v) \in M \times S^{k-1}} Z_1(t,v) \geq \sqrt{u}, \tag{5.3.5}$$

and

$$\sup_{t \in M} \bar{\chi}^2_K(t) \geq u \quad \Longleftrightarrow \quad \sup_{(t,v) \in M \times K : |v|=1} Z_2(t,v) \geq \sqrt{u}. \tag{5.3.6}$$

These two equivalences and Theorem 5.3.1 immediately imply that for $\chi^2_k$ and $\bar{\chi}^2_K$ random fields there exists $\alpha > 1$ (dependent on the underlying Gaussian fields, $k$, $K$ and $M$) such that

$$\left| \mathbb{E}\left\{\varphi\left(A_u(f,M)\right)\right\} - \mathbb{P}\left\{\sup_{t \in M} f(t) \geq u\right\} \right| \;\leq\; e^{-\alpha u/2}, \tag{5.3.7}$$

where an analogous, more rigorous, formulation as in (5.3.1) is, obviously, also possible.

Thus, for these random fields, as for Gaussian ones, the Euler characteristic heuristic is expressible as a rigorous theorem. Unfortunately, however, this is not the case for other random fields, for which the heuristic, for the moment at least, remains precisely that.

### 5.3.3 Numerical Examples

Should we do some numerical examples here? Keith, I think you once said that you had done some simulations. I think it would be worthwhile, both for general knowledge and since there are some simulations in Azais-Wschebor [12, 13], based on work by Mercadier [63]. These do not show up the EC heuristic in a proper light, since they compare estimates based on my old DT characteristic which we now know is not as good.

It would seem odd to have simulations davka here when everything so far has been theoretical.

## 5.4 Alternative Approaches to Exceedence Probabilities

So far in this chapter we have concentrated on the close relationships between numbers of critical points, exceedence probabilities, and the Euler characteristic of excursion sets. However, the theory, and heuristics, that we have presented are far from the whole story, and there are a number of other techniques that you should know about, particularly if you are interested in exceedence probabilities.

The first of these is known as the *volume of tubes* approach, and, as we shall see, works only for constant variance Gaussian fields. Furthermore, it turns out to yield estimates of exceedence probabilties that are identical to those that come from the Euler characteristic results of the previous section. Nevertheless, it is a powerful technique and worth knowing about. It also helps explain why the Lipschitz-Killing curvatures , which are essentially geometric quantities, mysteriously appear in seemingly unrelated results about exceedence probabilities.

### 5.4.1 Volume of Tubes and Gaussian Exceedence Probabilities

The tube method, which dates back to papers by Hotelling [42] and Weyl [96] in 1939, is a geometric technique for computing Gaussian exceedence probabilities. In the setting in which we shall apply it, it was developed primarily in [44, 50, 87] and starts by replacing (almost) all Gaussian random fields, on (almost) all parameter spaces, by a very special process defined over a subset of a unit sphere.

To this end, as in Section 2.3.5, we express $f$ via an expansion of the form

$$f(t) = \sum_{j=1}^{l} \xi_j \varphi_j(t) = \langle \xi, \varphi(t), \rangle_{\mathbb{R}^l} , \tag{5.4.1}$$

where $\xi = (\xi_1, \ldots, \xi_l)$ is a vector of standard Gaussian variables and $\varphi = (\varphi_1, \ldots, \varphi_l)$ a vector of functions on the parameter space $M$.

Note that there is an important assumption inherent in (5.4.1), that $f$ has a *finite* expansion, something which is not in general the case[6].

Assuming that $f$ has constant unit variance, (5.4.1) immediately implies

$$|\varphi(t)|^2 = \sum_{j=1}^{l} \varphi_j^2(t) = 1, \tag{5.4.2}$$

for all $t$, so that $\varphi(t)$ is also in $S(\mathbb{R}^l)$. (Has this been defined? In the notation of Section 2.4.8 it would be called it $S^{l-1}$. Check notation hereafter.)

This being the case, we can define a map $\varphi : t \to (\varphi_1(t), \ldots, \varphi_l(t))$, an embedding of $M$ in $S(\mathbb{R}^l)$. More significantly, we can define a random field $\widetilde{f}$ on ~~$\varphi(M) \in S(\mathbb{R}^l)$~~ $\varphi(M) \subset S(\mathbb{R}^l)$ by setting

$$\widetilde{f}(x) \stackrel{\Delta}{=} f\left(\varphi^{-1}(x)\right), \tag{5.4.3}$$

for all ~~$x \in \psi(M)$~~ $x \in \varphi(M)$. Note that $\widetilde{f}$ has the simple covariance function

$$C(x, y) = \mathbb{E}\left\{\widetilde{f}(x)\widetilde{f}(y)\right\} = \langle x, y\rangle, \tag{5.4.4}$$

and thus there is no problem taking a version of it on all of $S(\mathbb{R}^l)$ with this covariance function. This process is known as the *canonical (isotropic) Gausssian field on* $S(\mathbb{R}^l)$ (cf. Exercise 2.8.15), and it has the simple expansion

$$\tilde{f}(x) = \langle \xi, x\rangle = \sum_{j=1}^{l} \xi_j x_j, \qquad x \in S(\mathbb{R}^l), \tag{5.4.5}$$

where $\xi \sim N(0, I_{l\times l})$, which corresponds to the covariance function (5.4.4).

Since it is trivial that

$$\sup_{t\in M} f(t) \equiv \sup_{x\in\varphi(M)} \widetilde{f}(x), \tag{5.4.6}$$

it is clear that in computing exceedence probabilities for unit variance, finite expansion fields, we lose no generality by treating only the canonical process over subsets of $S(\mathbb{R}^l)$. Thus we shall do so, from now on letting $f$ denote this process, (Is there a need to make this change of notation for only one page?) with the representation (5.4.5) taking ~~$M \in S(\mathbb{R}^l)$~~ $M \subset S(\mathbb{R}^l)$. Then we can argue as follows, writing $\mathbb{P}_{|\xi|}$ for the distribution of $|\xi|$:

[6] For example, isotropic random fields on $\mathbb{R}^N$ can never have a finite expansion. cf. Exercise 2.8.13.

$$\begin{aligned}
\mathbb{P}\left\{\sup_{t\in M} f_t \geq u\right\} &= \int_0^\infty \mathbb{P}\left\{\sup_{t\in M} f_t \geq u \middle| |\xi| = r\right\} \mathbb{P}_{|\xi|}(dr) \\
&= \int_0^\infty \mathbb{P}\left\{\sup_{t\in M} \langle \xi, t\rangle \geq u \middle| |\xi| = r\right\} \mathbb{P}_{|\xi|}(dr) \\
&= \int_u^\infty \mathbb{P}\left\{\sup_{t\in M} \langle \xi, t\rangle \geq u \middle| |\xi| = r\right\} \mathbb{P}_{|\xi|}(dr) \\
&= \int_u^\infty \mathbb{P}\left\{\sup_{t\in M} \langle \xi/r, t\rangle \geq u/r \middle| |\xi| = r\right\} \mathbb{P}_{|\xi|}(dr).
\end{aligned} \tag{5.4.7}$$

Consider the integrand here. Since $\xi$ is multivariate Gaussian, it is standard fare that the vector $\xi/|\xi|$ is uniformly distributed on $S(\mathbb{R}^l)$, independently of $|\xi|$, which is distributed as the square root of a $\chi^2_l$ random variable. If we now write $\eta_l$ to denote the uniform measure over $S(\mathbb{R}^l)$, we can rewrite the integrand as a simple volume computation, once we take a moment to consider tubes on spheres.

Extending our definition of tubes from the simple Euclidean setting of Section 3.3 to spheres, we take the standard geodesic metric on $S(\mathbb{R}^l)$ given by

$$\tau(x, y) = \cos^{-1}\left(\langle x, y\rangle\right),$$

and define a tube of radius $\rho$ around a closed set $M \in S(\mathbb{R}^l)$ as

$$\begin{aligned}
\mathrm{Tube}(M, \rho) &= \left\{t \in S(\mathbb{R}^l) : \tau(t, M) \leq \rho\right\} \\
&= \left\{t \in S(\mathbb{R}^l) : \exists\ s \in M \text{ such that } \langle s, t\rangle \geq \cos(\rho)\right\} \\
&= \left\{t \in S(\mathbb{R}^l) : \sup_{s\in M} \langle s, t\rangle \geq \cos(\rho)\right\}.
\end{aligned} \tag{5.4.8}$$

With this behind us, (Was it that awful?) we can now continue the development of ~~(5.4.9)~~ (5.4.8) to obtain

$$\mathbb{P}\left\{\sup_{t\in M} f_t \geq u\right\} = \int_u^\infty \eta_l\left(\mathrm{Tube}(M, \cos^{-1}(u/r))\right) \mathbb{P}_{|\xi|}(dr) \tag{5.4.9}$$

Thus, the exceedence probability that we seek is weighted average of the volume of tubes around $M$ of varying radii, and if we could ~~to~~ compute

$$\eta_l\left(\mathrm{Tube}(M, \rho)\right)$$

we would, basically, be done, since the averaging, over the square root of a $\chi^2_l$ random variable is, in principle, straightforward.

Not surprisingly, there are analogues of Steiner's formula (3.3.3) for subsets of spheres, with the Lipschitz-Killing curvatures appearing there replaced by their sperical counterparts. Thus, we have seen, *without ever mentioning the Euler characteristic and without any heuristic arguments* that exceedence

probabilities should be related to the Lipschitz-Killing curvatures of the parameter space, at least for the canonical isotropic process on the sphere.

However, this observation is, in fact, true in general, since if $M$ is not a subset of the sphere, and the process $f$ is quite general (but with a finite expansion) it follows from (5.4.6) that the exceedence probabilities for $f$ should be related to the Lipschitz-Killing curvatures of $\varphi(M)$ which *is* a subset of a sphere. Exactly how these Lipschitz-Killing curvatures relate to those of $M$ itself is a rather long story in Reimannian geometry, and you can find some of the details, en passant, in *RFG*.

Now, however, we need to explain why, if everything is so simple, is there a need for the Euler characteristic, or any, approximation to Gaussian exceedence probabilities.

There are two intrinsic problems with (5.4.9). The first is that it is not generally easy to evaluate tube volumes on spheres[7]. The second is that the tube formula for evaluating $\eta_l(\text{Tube}(M, \rho)$ breaks down for large $\rho$ (close to 1) or, equivalently, for large $r$ in (5.4.9) (cf. Exercise 5.6.7.).

Of course, large $r$ in (5.4.9) does not cause to much of a problem, since this is in the tail of a $\chi^2_l$ variable, and approximations can be made. The devil is in the details, but what is interesting is that when one does this properly *the approximation that results is identical to the Euler characteristic approximation*, a result proven in wide generality by Takemura and Kuriki in [89]. We shall not, however, go into the details.

Which approach, then, is to be preferred? As far as Gaussian processes are concerned, it may be a matter of taste. We believe that the Euler characteristic approach is easier to implement, mainly due to the existence of a closed form expression for the mean Euler characteristic of excursion sets[8]. The major advantage of the Euler characteristic approach, however, lies in that the volume of tube approach works only for Gaussian processes, and there is no obvious way to extend it, even heuristically, to non-Gaussian ones. This, as we have already seen, is not a limitation of the Euler characteristic approach, where the EC heuristic both makes sense and seems to work in practice.

### 5.4.2 The Double Sum Method

There is a technique that dates back at least to Pickands in the 1960's [71, 72, 73] for approximating Gaussian extremal probabilities which has come

[7] To be fair, we should point out there are people who feel that this is easier than playing with the computations surrounding the Euler characteristic approach. In some sense, both sets of computations are parallel, and, as we shall see below, generally lead to equivalent results, but *only* for Gaussian fields.

[8] Admittedly, one may have to do some work to compute the Lipschitz-Killing curvatures if the random field is non-stationary or the parameter space is complicated. Recall how hard we had to work for the scale space fields of Section 4.6.2.

to be known as the 'double sum method' and is extremely powerful[9]. In particular, whereas the Euler characteristic heuristic and rigorous results such as Theorem 5.3.1 can handle only smooth random fields, the double sum method also works for non-smooth fields[10].

The basic idea behind this technque is to break up the parameter space $T$ into a finite union of small sets $T_k$, where the size of the $T_k$ generally depends on the exceedence level $u$. The $T_k$ need not be disjoint, although any overlap should be small in relation to their sizes. While initially there is no need to assume any particular structure for $T$, this usually is needed for the most precise results.

It is then elementary that

$$\begin{aligned}\sum_k \mathbb{P}\left\{\sup_{t\in T_k} f(t)\ge u\right\} &\ge \mathbb{P}\left\{\sup_{t\in T} f(t)\ge u\right\}\\ &\ge \sum_k \mathbb{P}\left\{\sup_{t\in T_k} f(t)\ge u\right\} \qquad (5.4.10)\\ &\quad - \sum\sum\nolimits_{j\ne k} \mathbb{P}\left\{\sup_{t\in T_j} f(t)\ge u, \sup_{t\in T_k} f(t)\ge u\right\}\end{aligned}$$

The summands in the single summations are treated by choosing a point $t_k \in T_k$, and then writing, in an obvious notation,

$$\begin{aligned}\mathbb{P}\left\{\sup_{t\in T_k} f(t)\ge u\right\} &= \int_{-\infty}^{\infty} \mathbb{P}\left\{\sup_{t\in T_k} f(t)\ge u \big| f(t_k)=x\right\} p_{t_k}(x)\,dx\\ &= \int_{-\infty}^{0} + \int_{0}^{u-\varepsilon} + \int_{u-\varepsilon}^{\infty} \dots\end{aligned}$$

All of this, of course, holds regardless of the distribution of the random field $f$. If, however, $f$ is Gaussian, then the first term two integrals can be shown to be of smaller order than the third, via careful applications of the Borel-CIS inequality of Theorem 2.6.1. The third term is clearly bounded by $\Psi(u-\varepsilon)$, assuming constant unit variance for $f$. Optimizing $\varepsilon$ against the number of sets $T_k$, not unlike what we did in proving Theorem 2.6.2, one gets a good bound for the single sums in (5.4.10). The details, as one would expect, are considerably more complicated than this one-paragraph summary.

[9] See also [51] which has a detailed treatment of Gaussian and other extrema in the one dimensional case and [75], which has the most powerful applications of this technique and which also treats random fields in considerable detail.

[10] Of course, there is a price to pay for a more general technique. In this case the price is that for non-smooth fields only the first term of expansions such as (5.0.5) can usually be computed. There are expections to this, as, for example, in [41] and [76], where, for some very specific fields, one can also say something about the second term.

The crux of the double sum method, however, lies in showing that the remaining double sum in (5.4.10) is of lower order than the single sum and, where possible, estimating its size.

If one could write the joint probabilities in (5.4.10) as the product

$$\mathbb{P}\left\{\sup_{t\in T_j} f(t) \geq u, \sup_{t\in T_k} f(t) \geq u\right\} \;=\; \mathbb{P}\left\{\sup_{t\in T_j} f(t) \geq u\right\} \mathbb{P}\left\{\sup_{t\in T_k} f(t) \geq u\right\},$$

then we would be basically done, since then the double sum term would then easily be seen to be of lower order than the single sum. Such independence obviously does not hold (although we have already argued for it back at (5.2.2)) but if we could choose the sizes of the $T_k$ in such a fashion that a 'typical component' of an excursion set is considerably smaller than this size, and manage to show that high extrema are independent of one another, then we would be well on the way to a proof.

The details, which are heavy, are all in Piterbarg [75]. Piterbarg's monograph is also an excellent source of worked examples, and includes a number of rigorous computations of exceedence probabilities for many interesting examples of processes and fields. (How about Piterbarg's approximation to the distribution of the supremum? Does it come from here?)

### 5.4.3 Double sums of moments

Another technique, closely related to the double sum approach above, involves looking at moments of critical points. Again, for details we send you to Piterbarg's book [75] as well as the more recent monograph by Azaïs and Wschebor [13], but the basic inequality is simple enough to present here[11].

An example of this result is given in the following theorem, a more general version of which can be found in *RFG*.

**Theorem 5.4.1.** *Let $f$ be a Gaussian field on $\mathbb{R}^N$ satisfying the conditions of Section 4.1 and with constant variance which, for convenience, we take to be one. Let $T$ be a $N$-dimensional rectangle, and let $\mu_k$ be the number of extended outward critical points of $f$, at which $f \geq u$, on the interior of the $k$-dimensional face $\partial_k T$ of $T$. In terms of the $\mu_k(J)$ of* (5.2.3) *and* (4.2.5)–(4.2.8)*,*

$$\mu_k = \sum_{J\in\partial_k T} \mu_k(J).$$

[11] In [13] this approach is considerably developed in terms of what are known as *Rice series*. The two series providing bounds for the exceedence probability in (5.4.1) are examples of the first and second terms of such a series, the $n$-th term involving a factorial moment of $\mu_k$ of order $n$. Azaïs and Wschebor also develop this approach to obtain bounds on the *density* of the supremum, a topic that we have not touched on at all. See also [12] for further information.

*Furthermore, assume that the random variables $\mu_k$ have finite second moments for all $k$ and given $u$. Then*

$$\sum_{k=0}^{N} \mathbb{E}\{\mu_k\} \geq \mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\} \geq \sum_{k=0}^{N} \left[\mathbb{E}\{\mu_k\} - \tfrac{1}{2}\mathbb{E}\{\mu_k(\mu_k-1)\}\right]. \quad (5.4.1)$$

**Proof.** Since the notion of extended outward maxima includes the 0-dimensional sets $\partial_0 T$ (i.e. the 'corners' of $T$) it is not hard to see that that, if $\sup_{t\in T} f_t \geq u$, then $f$ must have at least one extended outward maximum on $T$ above the level $u$, and vice versa. Consequently, writing $\mu(T)$ for the total number of extended maxima above the level $u$, we have

$$\mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\} \;=\; \mathbb{P}\{\mu(T)\geq 1\} \;\leq\; \sum_{k=0}^{N}\mathbb{P}\{\mu_k \geq 1\} \;\leq\; \sum_{k=0}^{N}\mathbb{E}\{\mu_k\}\,.$$

This gives the upper bound of (5.4.1).

For the lower bound, note firstly that

$$\left\{\sup_{t\in T} f_t \geq u\right\} \iff \{\mu_N \geq 1,\ \widehat{\mu}_{N-1}=0\} \cup \{\widehat{\mu}_{N-1}\geq 1\}\,, \quad (5.4.2)$$

where $\widehat{\mu}_k \overset{\Delta}{=} \sum_{j=0}^{k}\mu_j$. Since $f$, as a function on an $N$-dimensional rectangle $T$, does not have any critical points on $\partial T$ (cf. Corollary 2.7.2) we can write

$$\{\mu_N \geq 1,\ \widehat{\mu}_{N-1}=0\} \;=\; \{\mu_N = 1,\ \widehat{\mu}_{N-1}=0\} \cup \{\mu_N \geq 2,\ \widehat{\mu}_{N-1}=0\}\,.$$

To compute the probabilities of the above two events, set

$$p_k = \mathbb{P}\{\mu_N = k\}\,,$$

and note that

$$\mathbb{E}\{\mu_N\} = \mathbb{P}\{\mu_N = 1\} + \sum_{k=2}^{\infty} k p_k,$$

so that

$$\begin{aligned}\mathbb{P}\{\mu_N \geq 1,\ \widehat{\mu}_{N-1}=0\} &= \mathbb{P}\{\mu_N = 1\} - \mathbb{P}\{\widehat{\mu}_{N-1}=1,\ \widehat{\mu}_{N-1}\geq 1\}\\ &= \mathbb{E}\{\mu_N\} - \sum_{k=2}^{\infty} k p_k - \mathbb{P}\{\widehat{\mu}_N = 1,\ \widehat{\mu}_{N-1}\geq 1\}\,.\end{aligned}$$

(Unclear, some subindeces may be wrong in the equation above and the equation below.) In a similar vein,

$$\mathbb{P}\{\widehat{\mu}_{N-1}\geq 2, \widehat{\mu}_{N-1}=0\} = \sum_{k=2}^{\infty} p_k - \mathbb{P}\{\mu_N \geq 2, \widehat{\mu}_{N-1}\geq 1\}\,.$$

Putting the last two equalities together and substituting into (5.4.2) gives us that

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\} \\
&\quad = \mathbb{E}\left\{\mu_N\right\} - \sum_{k=2}^{\infty}(k-1)p_k - \mathbb{P}\left\{\mu_N \geq 1, \widehat{\mu}_{N-1} \geq 1\right\} + \mathbb{P}\left\{\widehat{\mu}_{N-1} \geq 1\right\} \\
&\quad \geq \mathbb{E}\left\{M_u^E(M^\circ)\right\} - \tfrac{1}{2}\mathbb{E}\left\{M_u^E(M^\circ)[M_u^E(M^\circ)-1]\right\} + \mathbb{P}\left\{\widehat{\mu}_{N-1} \geq 1\right\},
\end{aligned}$$

on noting that $k-1 < k(k-1)/2$ for $k \geq 2$. Apply the same argument to the final term here, and iterate to complete the proof. □

If you read the proof, you will have noticed that we actually never used the fact that $f$ was Gaussian. In fact, it is not needed, and the only reason for adding the Gaussian assumption was so that we could avoid a long list of conditions under which the Theorem might hold in general. You can easily build these conditions yourself by applying the conditions of the (non-Gaussian) expectation meta-theorem 2.7.1 to the setting of Theorem 5.4.1. Consequently, this result offers a level of generality far beyond the Gaussian theory we have treated so far.

It also makes sense that it would be quite a general phenomenon that the negative terms in (5.4.1) will be of smaller order that the others. After all, for $M_u^E(\partial_k M)[M_u^E(\partial_k M)-1]$ to be non-zero, there must be at least two extended outward maxima of $f|_{\partial_k M}$ above the level $u$ on $\partial_k M$, and this is unlikely to occur if $u$ is large.

Thus Theorem 5.4.1 seems to hold a lot of promise for approximating extremal probabilities in general, assuming that we could actually compute explicit expressions for the expectations in (5.4.1), perhaps just bounding product term. Unfortunately, as we have already seen in the Gaussian case, in Section 5.2, the best we can usually do is to approximate these expectations, as a result of which we are back to essentially the same results given by the Euler characteristic approach.

## 5.5 Aproximately Gaussian Random Fields

Here I think something should be said about all the new results on approximately Gaussian fields. Jonathan? Keith?

## 5.6 Exercises

**Exercise 5.6.1.** Apply Theorem 2.7.1 to determine under which conditions the mean number of critical points is given by (5.1.2).

How do the conditions simplify in the Gaussian case?

**Exercise 5.6.2.** Let $f$ satisfy all the conditions of Theorem 4.2.1. Let $\varphi(A_u(T))$ be the usual Euler characteristic of an excursion set, $M(T)$ the number of local maxima of $f$ over $T$, $m(T)$ the number of local minima, $C(T)$ the number of critical points and $S(T)$ the number of saddle points. Show:

(i) $\mathbb{E}\{\varphi(A_0(T))\} = 0$ for all rectangles $T$.
(ii) $\mathbb{E}\{M(T)\} = \mathbb{E}\{m(T)\} = \frac{1}{2}\mathbb{E}\{S(T)\} = \frac{1}{4}\mathbb{E}\{C(T)\}$.

Hint: Use the Morse theoretic characterisation of Theorem 3.2.4 to see how $\varphi(A_u(T))$ relates to the other variables. Then apply (i) and symmetries to large (WHY?) rectangles to prove (ii).

**Exercise 5.6.3.** Suppose that $f$ satisfies the conditions of Theorem 5.2.1, and let $L_u(T)$ denote the Lebesgue of $A_u(T)$ and $M_u^-(T)$ the Lebesgue measure of the portion of $A_u$ over which $\nabla^2 f$ is negative definite. Using Lemma 5.2.2, or otherwise, show that

$$\frac{\mathbb{E}\{M_u^-\}}{E\{M_u\}} \to 1, \qquad \text{as } u \to \infty.$$

**Exercise 5.6.4.** Suppose that $f$ satisfies the conditions of Theorem 5.2.1 with the parameter space a subset of $\mathbb{R}^2$.

On the level curves of $f$ (i.e. the boundary $\partial A_u$ of the excursion set $A_u$) define the *curvature function*

$$\kappa = \frac{-(f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22})}{(f_1^2 + f_2^2)^{3/2}},$$

where subscripts refer to the usual derivatives. When $\kappa(t) > 0$ at a point $t \in \partial A_u$ then $A_u$ will be locally convex at that point.

Let $L_u$ denote the arc length of $\partial A_u$ and write $L_u^+$ for the arc length of these segments of $\partial A_u$ which have positive curvature.

Use Crofton's formula (ADD NUMBER REFERENCE) and Lemma 5.2.2, show that

$$\frac{\mathbb{E}\{L_u^+\}}{E\{L_u\}} \to 1 \qquad \text{as } u \to \infty.$$

**Exercise 5.6.5.** Consider the cosine random field $f$ of (2.3.11), with all $\lambda_k \equiv 1$ and with defined over the rectangle $[0, T]^N$. Show that

$$\text{Var}\left(\frac{\partial^2 f(t)}{\partial t_j^2}\Big| f(t)\right) = N - 1.$$

Although $f$ is not isotropic (why?) the result (5.3.4), which gave the parameter for the EEC heuristic of Theorem 5.3.1, can be shown to hold for the cosine field as well.

How does this relate to the results of Sections 2.3.3 and 2.3.4 relating to exceedence probabilities for cosine processes and fields?

**Exercise 5.6.6.** (Label was repeated so changed it (see Latex file)) Let $f$ be the cosine process $\xi \cos t + \xi' \sin t$ of (2.3.3) on the interval $M = [0, 2\pi]$.

Show that the mapping $\varphi$ of (5.4.3) maps $M$ to a great circle in $S(\mathbb{R}^2)$ (Check notation with text here and next exercise.).

**Exercise 5.6.7.** (i) Let $M \subset \mathbb{R}^N$ be a locally convex set, in the sense of Section 3.5. (The left hand example in Figure 3.5.1, or a simple pertubation of it, will suffice for this exercise.) Show that even if a tube volume formula holds for tubes $\text{Tube}(M, \rho)$ for small enough $\rho$, (which it does, see Chapter 10 of *RFG*), it cannot be expected to hold for all $\rho$ unless $M$ is convex.

(ii) Let $M \subset S(\mathbb{R}^l)$ be convex, with non-empty interior. Defining $\text{Tube}(M, \rho)$ for all $\rho$ as in (5.4.8), explain why a tube volume formula cannot hold for $\text{Tube}(M, \rho)$ for all $\rho$, and, in fact, can be expected to break down as $\rho$ grows towards 1. (Note that this is quite different to the Euclidean case, in which the tube formula holds for all $\rho$ if $M$ is convex.)

(iii) Argue that if $M$ is either a great circle in $S(\mathbb{R}^l)$, or a segment of a great circle which wraps at least half way around the sphere (Why?), then a tube formula cannot hold for all $\rho > 1$. (In view of the preceeding exercise, this implies that the tube formula approach to exceedence probabilities of Section 5.4.1 should give an exact result for the cosine process, as, in fact, both it and the Euler characteristic approach do, when the parameter space is restricted to $[0, T]$ for $T < \pi$. (cf. (2.3.9).)) (Unclear)

# 6

# The Structure of Excursion Sets

In the preceeding Chapter 5 we met some intruiging properties of the extrema and excursions of Gaussian random fields: The mean number of local maxima (or, indeed of all critical points) above a high level was very close to the mean Euler characteristic for the excursion set for that level. Furthermore, both of these were close to the corresponding exceedence probability. We gave a number of heuristic explanations as to why these results should be true, all boiling down to the idea that, at high levels, Gaussian random fields tended to take on very simple shapes. In the current chapter we shall make mathematics out of this idea, which will put the heuristic ideas of Chapter 5 on a firmer footing.

However, the results of this chapter have significant interest and application in their own right. After we have set a little preliminary background by describing Palm measures, we shall show how to describe the structure of a random field conditional on the occurrence of some special event, such as a local maximum, or level crossing of a particular kind, occurring at a chosen point. These conditional models are generally known as *Slepian models*, or *Slepian processes* after their discoverer, David Slepian. (cf. [46, 85]) for Gaussian processes on $\mathbb{R}$ and Lindgren [56] for Gaussian fields.)

We shall then apply the Slepian models, which are exact conditional models, to develop precise statements about the local structure of Gaussian fields in the neighborhood of high level maxima. For example, suppose that $f$ is a stationary, centered, Gaussian field on $\mathbb{R}^N$ with variance $\sigma^2$ and matrix $\Lambda$ of second spectral moments. Then, conditional on $f$ having a local maximum of height $u$ at the point $t = 0$, with probability approaching one as $u \to \infty$, it has the following representation in a neighborhood of the origin:

$$f(t) = u - \frac{u}{2\sigma} t\Lambda t' + O(1).$$

In other words, $f$ is approximately parabolic. Interestingly, in turns out that while, for Gaussian related fields it is hard to develop analogues of the exact

conditional Gaussian results, it is possible to find rigorous analogues of many of the high level results, and so we shall also describe these.

With this rigorous mathematics behind us, we shall turn to what David Aldous has called the *Poisson clumping heuristic* to begin making a large number of quite varied heuristically based claims about the structure of high level excursions, for both Gaussian and non-Gaussian fields, attempting not only to describe their mean shapes but also making claims about the full distributions of some of their geometric characteristics. In doing so, by the end of the chapter, you may find it necessary to somewhat sever yourself from the usual demands of mathematical rigor and submit rather to the demands of applications, where approximate answers are considered acceptable if they are all that it is available.

## 6.1 Palm Distributions and Conditioning

It is well known that one needs to be careful when talking about random events for stochastic processes, as unexpected biases can easily affect computations. Even the simplest of continuous time processes, the constant rate Poisson process on the line, has its biased sampling theorem[1]. Much in the same way that one needs to be careful in describing a 'typical' interval of constancy in a Poisson process, one needs to be careful about describing the behavior of a random field in the neighborhood of a 'typical' local maximum.

The first thing to note is that the notion of 'typical' only makes sense when talking about random fields that satisfy a condition of stationarity or, at least ergodicity. Otherwise, 'typical' requires a qualification, so that we would have to talk about 'typical for a specific part of the parameter space'. Since we do not want to do this, we shall, for the rest of this section and most of the chapter, assume that we are dealing only with stationary random fields on $\mathbb{R}^N$.

To begin to formalise things, define a *point process* $N$ on $\mathbb{R}^N$ to be a non-negative, integer valued, random measure on the bounded Borel sets of $\mathbb{R}^N$. The point process is called *simple* if

$$\mathbb{P}\big\{N(\{t\}) \leq 1 \text{ for all } t \in \mathbb{R}^N\big\} = 1,$$

and its *intensity measure* is given by $\mu(B) = \mathbb{E}\{N(B)\}$.

The point processes that we shall be most concerned with are those for which $N(B)$ is the number of local maxima, or other special points, of a

[1] To recall, it is well known that the times in between jumps of a unit rate Poisson process are mean one exponential variables. On the other hand, if $t > 0$ is some fixed time, then the size of the interval of constancy containing $t$ is stochastically larger than exponential, since it is made up of the time from $t$ to the next jump, which by Markovianess is mean one exponential, plus the time between the preceeding jump and $t$.

random field $f$ in the set $B$. Since we are assuming that $f$ is stationary, this will also be true of the pair $(f, N)$, where by stationarity for such a pair we mean that, for each $\tau \in \mathbb{R}^N$,

$$\theta_\tau(f, N) \stackrel{\Delta}{=} (\theta_\tau f, \theta_\tau N) \stackrel{\mathcal{L}}{=} (f, N), \tag{6.1.1}$$

where $\theta_\tau f(t) \stackrel{\Delta}{=} f(t+\tau)$ and $\theta_t N(B) \stackrel{\Delta}{=} N(B+\tau)$.

In general, if a random field $f$ and a point process $N$ are jointly stationary on $\mathbb{R}^N$, we can define the *Palm distribution*, or *Palm measure*, $\mathbb{P}_{f,N}$ of $(f, N)$ with respect to $N$ via the corresponding expectation

$$\mathbb{E}_{f,N}\{F(f, N)\} \stackrel{\Delta}{=} \frac{\mathbb{E}\left\{\int_B F(\theta_\tau(f, N))\, N(d\tau)\right\}}{\mathbb{E}\{N(B)\}}, \tag{6.1.2}$$

where $F \geq 0$ is any measurable function on the product of the spaces of sample paths of $f$ and of non-negative measures. The set $B$ can be any bounded Borel set in $\mathbb{R}^N$ with finite $\mu$ measure, and the definition can be shown to be independent of $B$[2].

Rewriting this in terms of finite dimensional distributions, what (6.1.2) says is that if $\bar{t} = (t_1, \ldots, t_n)$ is a collection of points in $\mathbb{R}^N$, then the Palm measure describes a new random field, $\widetilde{f}$, for which, for any $\bar{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\begin{aligned} F_{\bar{t}}(\bar{x}) &\stackrel{\Delta}{=} \mathbb{P}\left\{\widetilde{f}(t_1) \leq x_1, \ldots, \widetilde{f}(t_n) \leq x_n\right\} \\ &= \frac{\mathbb{E}\left\{\#\{s_j \in B :\, f(s_j + t_k) \leq x_k,\ k = 1, \ldots, n\}\right\}}{\mathbb{E}\{\#\{s_j \in B\}\}}, \end{aligned} \tag{6.1.3}$$

where the $s_j$ are the atoms, or points, of $N$.

To see why Palm measures are interesting, we can exploit the fact that, as already mentioned, their definition, in the stationary case of interest, is independent of the set $B$. Consider two special, and extreme, cases, both based on taking $B$ to be $B_\lambda^N$, the $N$-ball of radius $\lambda$.

Assume that $f$ is ergodic, which, for Gaussian fields, occurs if the spectral distribution function (2.4.3) is continuous (cf. [36, 61]). Then, taking $B = B_\lambda^N$ in (6.1.3) and sending $\lambda \to \infty$, both expectations can be replaced with averages, so that

$$F_{\bar{t}}(\bar{x}) = \lim_{\lambda \to \infty} \frac{\#\{s_j \in B_\lambda^N :\, f(s_j + t_k) \leq x_k,\ k = 1, \ldots, n\}}{\#\{s_j \in B_\lambda^N\}}. \tag{6.1.4}$$

The importance of this limit is the meaning it gives to the (Palm) distributions of the random field $\widetilde{f}$, which now has a frequency interpretation as the random field $f$, conditioned on $N$ having an atom at the origin.

[2] For this fact, and other details on point processes and Palm measures, see either of Kallenberg's books, [47, 48].

For the second case, we again take $B = B_\lambda^N$ in (6.1.3), but now consider the limit $\lambda \to 0$. In that case it can be shown (e.g. Theorem 11.5 of [48]) that

$$\lim_{\lambda\to 0} \frac{\mathbb{P}\{N(B_\lambda^N) = 1\}}{\mathbb{P}\{N(B_\lambda^N) > 0\}} \;=\; \lim_{\lambda\to 0} \frac{\mathbb{P}\{N(B_\lambda^N) = 1\}}{\mathbb{E}\{N(B_\lambda^N)\}} \;=\; 1.$$

Applying a similar result to the numerator in (6.1.3) and taking the limit we can rewrite[3] it as

$$F_{\bar{t}}(\bar{x}) = \mathbb{P}\left\{f(t_k) \le x_k,\ k = 1, \dots, n \,\middle|\, N \text{ has an atom at } 0\right\}. \quad (6.1.5)$$

This conditional probability coincides with the frequency interpretation of Palm distributions given above from the $\lambda \to \infty$ limit. It does not, however, require ergodicity, and so is more intrinsic than the frequency interpretation.

Having defined the conditional probability (6.1.5), there are two important things that you should always remember about it. The first is that there is a convenient, but potentially misleading notational transgression inherent in the equation, for although we have used the usual symbol $|$ for conditioning, this is not the usual conditioning event. For example, as we shall see soon, while, under the usual conditioning on $f(0) = u$, $\nabla f(0) = v$, the elements of $\nabla^2 f(0)$ have a multivariate Gaussian distribution (cf. Exercise 2.8.10) this is not true under the Palm distribution. The difference is due to a phenomenon not unrelated to the sample bias of the intervals of a Poisson process mentioned above.

The second thing to remember is something that (6.1.5) makes quite clear, and that is that we are conditioning on an event of probability zero. This, of course, is the event that a simple point process has an atom at particular point. From the technical point of view, it is this that forced us to be careful about how conditional probabilities were defined.

We can now turn to applying this tool to a point process of interest to us.

## 6.2 The Local Structure of Gaussian Maxima

In this section we shall study the structure of a Gaussian field $f$ conditional on there being a local maximum of height $u$ at some fixed point, which, without loss of generality, we take to be the origin of $\mathbb{R}^N$. Throughout, $f$ will be assumed to be centered and satisfying the regularity conditions of Section

[3] This actually requires some work, since substituting into (6.1.3) for fixed $\lambda > 0$ actually approximates probabilities of the form

$$\mathbb{P}\left\{f(s + t_k) \le x_k,\ k = 1, \dots, n \,\middle|\, N \text{ has an atom at } s \in B_\lambda^N\right\},$$

which does not immediately give (6.1.5) in the limit. For details which are relevant to the uses that we shall make of Palm distributions, see, for example, [56].

4.1. In addition, we assume that it is stationary, with covariance function $C$ and, for convenience, variance $C(0) = 1$.

We shall proceed in two stages. At first, we shall develop a model that gives precise information regardless of the level $u$ of the local maximum. In the second stage we shall see what this has to say about the behaviour of $f$ at high local maxima, where the sample functions of the random field take on a particularly simple structure, subject to small, stochastic perturbations.

### 6.2.1 The Exact Slepian Model

Rather than tackle the problem of maxima at a given level directly, we shall start by looking at $f$ conditioned on a local maximum at the origin, *above* the level $u$ (rather than *at* the level $u$). In terms of the Palm distributions of the previous section[4], the corresponding point process of interest, $N$, counts the number of local maxima of $f$ above the level $u$ and so the the conditional finite dimensional distributions of $f$ are given by the ratios

$$\frac{\mathbb{E}\big\{\#\{t \in B : \nabla f(t) = 0, \nabla^2 f(t) \in \mathcal{N}, f(t) \geq u, f(t + t_k) \leq x_k, k = 1, \ldots, n\}\big\}}{\mathbb{E}\big\{\#\{t \in B : \nabla f(t) = 0,\ \nabla^2 f(t) \in \mathcal{N},\ f(t) \geq u\}\big\}},$$

for all choices of $n \geq 1$, $t_k \in \mathbb{R}^N$, $x_k \in \mathbb{R}$ and where, as usual, $\mathcal{N}$ is the set of negative definite matrices. Since the ratio is independent of $B$, we assume that it is a set of $N$-dimensional volume one.

In then follows immediately from Theorem 2.7.1 (cf. also (5.1.2)) that the denominator above is given by

$$(-1)^N p_{\nabla f}(0)\, \mathbb{E}\Big\{\det \nabla^2 f\, \mathbb{1}_{\{\nabla^2 f \in \mathcal{N}\}}\, \mathbb{1}_{\{f \geq u\}} \,\Big|\, \nabla f = 0\Big\},$$

where $p_{\nabla f}$ is the density of $\nabla f$.

Similarly, the numerator in (6.1.3) is easily seen to be

[4] In much of the literature, the conditional distributions that we are about to study are described as arising from *horizontal window conditioning*. This conditioning arises from ~~we want to condition~~ conditioning on an event of probability zero, the event that a local maximim occurs at the origin. There are two natural ways to approximate this event. Ignoring the positive probability event that $\nabla^2 f(0) \in \mathcal{N}$, one way is to require that $\nabla f$ is in small ball around the origin and let the radius of this ball tend to zero. This is known as *vertical window conditioning* and in the limit, is equivalent to the usual conditioning. The other approach is to condition on the event that $\nabla f(t) = 0$ for $t$ in some small ball around the origin, and then let the radius of this ball tend to zero. This is known as horizontal window conditioning. (Think of the one dimensional case, when this ball is simply a horizontal interval, or 'window'.) The limits of these two types of conditioning are quite different, and it is only the second that gives the Palm distributions with which we are working. See Exercise 6.7.1 for an example of the differences.

$$(-1)^N p_{\nabla f}(0)\,\mathbb{E}\left\{\det\nabla^2 f\,\mathbb{1}_{\{\nabla^2 f\in\mathcal{N}\}}\,\mathbb{1}_{\{f\geq u\}}\prod_{k=1}^{n}\mathbb{1}_{\{f(t_k)\leq x_k\}}\,\Big|\,\nabla f=0\right\},$$

We need to write out these expectations in full, and so let $p(f,f',f'')$ denote the joint density of $f(0)$ and the elements of the vector $\nabla f(0)$ and the matrix $\nabla^2 f(0)$. As before, writing $\bar{t}=(t_1,\ldots,t_n)$, $\bar{x}=(x_1,\ldots,x_n)$,let $p_{\bar{t}}(f,f',f'',v)$ be the joint density of these variables along with $f(t_k)$, $k=1,\ldots,n$. Then it follows from the above that the Palm conditional distributions of $f$ given a local maximum at $t=0$ above the level $u$ have the form

$$F_{\bar{t}}(\bar{x})=\frac{\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}\int_{f=u}^{\infty}\int_{\mathcal{N}}\det(-f'')\,p_{\bar{t}}(f,0,f'',x)\,df\,df''\,dx}{\int_{f=u}^{\infty}\int_{\mathcal{N}}\det(-f'')\,p(f,0,f'')\,df''\,df}\ . \quad (6.2.1)$$

We now turn to conditioning on the maximum at the origin occurring at exactly the level $u$, a zero probability event.

Note that if in (6.3.22)(DOUBLE LABEL) we were to restrict the integration over $f$ to the range $(u,u+\delta)$ rather than $(u,\infty)$ the same argument that lead to (6.3.22) would show that the resulting ratio would give the conditional distributions for the random field $f$ conditioned on having a local maximum at the origin of height that lies in the interval $(u,u+\delta)$.

Sending $\delta\to 0$, therefore, turns the conditional distributions into those of $f$ conditioned to have a local maximum at the origin which is precisely at the level $u$. Taking this limit in (6.3.22), and moving to conditional densities rather than conditional distribution functions, establishes the following result.

**Lemma 6.2.1.** *With the conditions and notation given above the conditional (Palm) densities of $f(t_1),\ldots,f(t_n)$, given that $f$ has a local maximum at 0 with height $u$, are given by*

$$\varphi^u_{\bar{t}}(x)=\frac{\int_{\mathcal{N}}\det(-f'')\,p_{\bar{t}}(u,0,f'',x)\,df''}{\int_{\mathcal{N}}\det(-f'')\,p(u,0,\underline{x}f'')\,df''}\ . \quad (6.2.2)$$

Before continuing with investigating the implications of this lemma, we note for later use that nowhere in its derivation did we actually use the fact that $f$ was assumed Gaussian, other than to appeal to existing results which gave us simple conditions for all the arguments to work. In fact, (6.2.2) will hold for any random field for which all the integrals are well defined, and for which all the limits that we took along the way are justified. We shall return to this point later, when considering non-Gaussian fields in Section 6.4.

Now, however, we begin to investigate (6.2.2) a little more closely, looking for a structure that might be more informative. As a first step we rewrite it as

$$\varphi^u_{\bar{t}}(x)=\int_{\mathcal{P}}\psi_u(f'')\,p_{\bar{t}}(v|u,0,f'')\,df'', \quad (6.2.3)$$

(Where is $x$? What's $v$? Did the order of the arguments in $p$ change? Revise notation hereafter.) where $\mathcal{P}$ denotes the collection of positive definite matrices and

$$\psi_u(f'') \stackrel{\Delta}{=} \begin{cases} \frac{\det f'' \, p(f''|u,0)}{\int_{\mathcal{P}} \det f'' \, p(f''|u,0) df''} & \text{if } f'' \in \mathcal{P}, \\ 0 & \text{otherwise,} \end{cases} \tag{6.2.4}$$

In other words, we can rewrite the joint density $\varphi_{\bar{t}}^u(x)$ as a mixture of Gaussian densities (i.e. the $p_{\bar{t}}(v|u,0,f'')$) on the matrix-valued variable $f''$ which has the density $\psi_u$. From this it follows that the conditional random field is also a mixture of Gaussian fields in an appropriate sense. Note that the effect of the Palm measure has been to replace the 'natural' Gaussian distribution of $f''$, which started out as the Hessian $\nabla^2 f(0)$, with something quite different and quite far from Gaussian. (cf. Exercise 6.7.1.)

Continuing along these lines, we need to know more about the densities appearing in (6.2.3), and for this we require some new notation and a reminder about existing notation. As far as the latter is concerned, recall that we are are not only writing $f''$ to denote the matrix $\nabla^2 f$, but we also want to be able to consider it as vector of its elements, rearranged by first taking the $f_{ii}$ and then the $f_{ij}$, $i = 1, \ldots, N$, $j = i+1, \ldots, N$, subscripts representing partial derivatives. Then, slightly rewriting what you already derived in Exercise 2.8.10, and making the simplifying assumption that $f$ has unit variance, the covariance matrix of the $1 + N + N(N+1)/2$ vector $(f, f', f'')$ can be written in the form

$$\begin{pmatrix} 1 & 0 & \mu_{02} \\ 0 & \Lambda & 0 \\ \mu_{20} & 0 & \mu_{22} \end{pmatrix},$$

where $\Lambda$ is the usual matrix of second-order spectral moments, $\mu_{22}$ is made up of fourth-order spectral moments, and $\mu_{02} = \mu_{20}'$ contains second-order spectral moments. If we now introduce the two new vectors

$$\begin{aligned} \mu_1(t) &= (-C_1(t), \ldots, -C_N(t)) \\ \mu_2(t) &= (C_{ii}(t),\ i = 1, \ldots, N;\ C_{ij}(t),\ i = 1, \ldots, N,\ j = i+1, \ldots, N), \end{aligned}$$

where $C$, as usual, is the covariance function of $f$, then the $1 + N + N(N+1)/2 + n$ variates $f(0), \nabla f(0), \nabla^2 f(0), f(t_1), \ldots, f(t_n)$ have a joint normal distribution with zero mean and covariance matrix

$$\begin{pmatrix} 1 & 0 & \mu_{02} & C(t_1) & \ldots & C(t_n) \\ 0 & \Lambda & 0 & \mu_1'(t_1) & \ldots & \mu_1'(t_n) \\ \mu_{20} & 0 & \mu_{22} & \mu_2'(t_1) & & \mu_2'(t_n) \\ C(t_1) & \mu_1(t_1) & \mu_2(t_1) & 1 & & C(t_n - t_1) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ C(t_n) & \mu_1(t_n) & \mu_2(t_n) & C(t_1 - t_n) & & 1 \end{pmatrix}$$

Next, we need the real-valued functions $\alpha(t)$, $\gamma(s,t)$, and the vector-valued function $\beta(t) = (\beta_{ii}(t),\ i = 1,\dots,N; \beta_{ij}(t), i = 1,\dots,N,\ j = i+1,\dots,N$, defined by

$$(\alpha(t),\beta(t)) = \big(C(t),\mu_2(t)\big)\begin{pmatrix}1 & \mu_{02}\\ \mu_{20} & \mu_{22}\end{pmatrix}^{-1} \tag{6.2.5}$$

$$\gamma(s,t) = C(s-t) - \big(C(s),\mu_2(s)\big)\begin{pmatrix}1 & \mu_{02}\\ \mu_{20} & \mu_{22}\end{pmatrix}^{-1}\begin{pmatrix}C(t)\\ \mu_2'(t)\end{pmatrix} - \mu_1(s)\Lambda^{-1}\mu_1'(t). \tag{6.2.6}$$

It is now a simple matter to check that $p_{\bar{t}}(x|u,0,-f'')$ is a normal density with means $u(\alpha(t_i) + f''\cdot(\beta(t_i))'$ and covariances $\gamma(t_i,t_j)$. The density $\psi_u$ comes out similarly: $p(f''|u,0)$ is normal with mean

$$(u,0)\begin{pmatrix}1 & 0\\ 0 & \Lambda\end{pmatrix}^{-1}\begin{pmatrix}\mu_{02}\\ 0\end{pmatrix} \tag{6.2.7}$$

and covariance matrix

$$\mu_{2\cdot 0} \stackrel{\Delta}{=} \mu_{22} - (\mu_{20},0)\begin{pmatrix}1 & 0\\ 0 & \Lambda\end{pmatrix}^{-1}\begin{pmatrix}\mu_{02}\\ 0\end{pmatrix} = \mu_{22} - \mu_{20}\mu_{02}. \tag{6.2.8}$$

Thus, for $f'' \in \mathcal{N}$, we have

$$\psi_u(f'') = k_u^{-1}\det f''\exp[-\tfrac{1}{2}(f''+u\mu_{20})\mu_{2\cdot 0}^{-1}(f''+u\mu_{20})'\}], \tag{6.2.9}$$

where $k_u$ is the obvious normalizing constant.

Fortunately, this is all the notation we need, and with it we can compute the characteristic function of the density $\varphi_t^u$ of (6.2.2) as

$$\begin{aligned}\widehat{\varphi}_{\bar{t}}^{\,u}(\theta) &= \int_{\mathbb{R}}^{N} e^{i\langle\theta,x\rangle}\varphi_{\bar{t}}^{u}(x)\,dx\\ &= \int_{\mathcal{N}}\psi_u(f'')\left[\int_{\mathbb{R}^n} e^{i\langle\theta,x\rangle}p_{\bar{t}}(x|u,0,f'')\,dx\right]df''.\end{aligned}$$

Using the above information about the densities $\psi_u$ and $p_{\bar{t}}$ we obtain, on writing $A$ for the $n$ vector with the $j$-th element $\alpha(t_j)$, $B$ for the $N(N+1)/2\times n$ matrix whose $j$-th column contains the vector $\beta'(t_j)$, and $\Gamma$ for the $n\times n$ matrix with elements $\gamma(t_i,t_j)$, that

$$\begin{aligned}\widehat{\varphi}_{\bar{t}}^{\,u}(\theta) &= \int_{\mathcal{N}}\psi_u(f'')\,e^{u\langle\theta,A\rangle - \langle\theta,f''B\rangle - \frac{1}{2}\theta\Gamma\theta'}\,df''\\ &= e^{u\langle\theta,A\rangle - \frac{1}{2}\theta\Gamma\theta'}\int_{\mathcal{N}} e^{-\langle\theta,f''B\rangle}\,\psi_u\left(f''\right)\,df''.\end{aligned}$$

The first factor here is simply the characteristic function of an $n$ dimensional Gaussian variable with means $uA(t_i)$ and covariances $\Gamma(t_i,t_j)$. The

second factor is only marginally more complicated, being the characteristic function of the $n$ dimensional random variable $f''B$, where $f''$ has the density $\psi_u$. The matrix $B$, of course, is non-random. Since characteristic functions determine distributions and finite dimensional distributions determine a process, we have proven the following important result, which gives the promised Slepian model for $f$ in the neighborhood of local maxima.

**Theorem 6.2.2.** *Let $f$ be a mean zero, unit variance, stationary, ergodic, Gaussian field satisfying the regularity conditions of Section 4.1. Then, given that $f$ has a local maximum with height $u$ at 0, the conditional field has the same finite dimensional distributions as the field $\{f_u(t), t \in \mathbb{R}^N\}$ defined by*

$$f_u(t) = u\alpha(t) - Z_u\beta'(t) + g(t), \tag{6.2.10}$$

*where $g$ is a non-homogeneous, zero-mean, Gaussian field with the covariance function $\gamma$ given by* (6.2.6) *and $Z_u$ is an $N(N+1)/2$ dimensional random variable, independent of $g$ and with the density $\psi_u$ given by* (6.2.9).

In order to understand what the Slepian model has to say about the conditioned process, it is instructive to consider three regions, one for $t$ close to the origin, one for $t$ far from it, and the intermediate case. For the second of these, assume that the covariance function $C$, along with those of its partial derivatives of up to second order, tend to zero as $|t| \to \infty$. Then it immediately follows that both $\alpha(t) \to 0$ and $\beta(t) \to 0$ as $|t| \to \infty$ so that, well away from the conditioning event at the origin, $f_u$ behaves much the same as $g$ behaves. However, although $g$ is a non-stationary random field, it is immediate from (6.2.6) and the above assumptions on the decay of $C$ and its derivatives, that its covariance function converges to that of the unconditional field $f$ well away from the origin. In other words, well away from the origin the conditioned process has basically the same distributional properties as the unconditioned process, hardly a suprising result.

What happens when $t$ is small is far more informative, and actually quite simple, particularly when $u$ is large. We shall look at that case in the following subsection.

What happens for intermediate $t$ is somewhat harder than the other cases. In this situation the complicated forms of the functions $\alpha$ and $\beta$, along with the rather non-standard distribution that we have for $Z_u$, compounded by the non-stationarity of $g$ makes it hard to carry out computations. In general, therefore, numerical work is the order of the day, and there have been significant advances in this direction, leading to the development of computer packages for computing the distributions of variables of interest, such as the position of the nearest maximum to that at the origin.

We shall have more to say about this in Chapter 11, when we discuss appllications of these models to oceanography. Now, however, we turn the case of small $t$ and large $u$.

### 6.2.2 Excursions Above High Levels

We start by rewriting the Slepian model (6.2.10) a little, noting that by (6.2.5) it is equivalent to

$$f_u(t) = uC(t) - W_u\beta'(t) + g(t), \tag{6.2.11}$$

where $W_u \stackrel{\Delta}{=} Z_u + u\mu_{02}$.

Since the distribution of $g$ is independent of $u$, it is clear that, as $u \to \infty$, the random field $g$ exerts little influence on $f_u(t)$, whereas the first term of (6.2.11) is extremely important. The rôle played by the term $W_u\beta'$ is unclear. To clarify this, we prove the following lemma.

**Lemma 6.2.3.** *As $u \to \infty$ the vector $W_u$ of* (6.2.11) *converges in distribution to an $N(N+1)/2$ dimensional Gaussian random variable, independent of the field $g$, having zero mean and covariance matrix $\mu_{2\cdot 0}$ of* (6.2.8)*.*

**Proof.** The vector $W_u$ has the density $\psi_u^*(w) = \psi_u(w - u\mu_{20})$. Think of the vectors $w$ and $W_u$ also as symmetric matrices, formed from the vectors with the inverse of the operation we used to go from the matrix $f''$ to its vector version. Then the vector $w - u\mu_{02}$ is equivalent, in this sense, to the matrix $w - u\Lambda$. Thus it follows from (6.2.9) that,

$$u^{-N(N+1)/2}k_u\psi_u^*(w) = u^{-N(N+1)/2}\det(w - u\Lambda)\exp(-\tfrac{1}{2}w\mu_{2\cdot 0}^{-1}w')$$

for $w - u\Lambda$ negative definite, where on the right hand side we first interpret $w$ as a vector and then as a matrix. Clearly, as $u \to \infty$, the right-hand side tends pointwise and with dominated convergence to

$$\det(\Lambda)\exp(-\tfrac{1}{2}w\mu_{2\cdot 0}^{-1}w'),$$

while the region over which $w - u\Lambda$ is negative definite tends to the whole of $\mathbb{R}^{N(N+1)/2}$. Thus it follows that there exists a finite $k_\infty = \lim_{u\to\infty} u^{-N(N+1)/2}k_u$, from which it follows that

$$\psi_u^*(w) \to k_\infty^{-1}\exp(-\tfrac{1}{2}w\mu_{2\cdot 0}^{-1}w)$$

with dominated convergence, which is the content of the lemma. □

This lemma motivates one to expand the right hand side of (6.2.11) as a power series in $t$, and so, for small $|t|$ and large $u$ to claim that

$$f_u(t) = u - \frac{u}{2}t\Lambda t' + o(t^2) + O_P(1), \tag{6.2.12}$$

where, by $O_P(1)$, we mean terms which, with high probability, take bounded values. In other words, in the neighborhood of a high maximum, Gaussian

fields look much like a paraboloid the shape of which is a multiple of that of their covariance functions. More formally, in Exercise 6.7.5 you are asked to show that the following is true, where the convergence is for the finite dimensional distributions of both sides:

$$\lim_{u\to\infty} u\left(f_u\left(\frac{t}{u}\right)-u\right) \stackrel{\mathcal{L}}{=} -\ \tfrac{1}{2}t\Lambda t'. \tag{6.2.13}$$

One can go beyond the convergence of finite dimensional distributions, and extend this convergence to full weak convergence, and even almost sure, uniform convergence over compacts. For details see [56, 57].

With the stronger theorems, it becomes quite easy to obtain additional results. For example, it is not surprising that near a high maximum $f$ will not have any more critical points, of any kind. The formal statement, a proof of which is given in [56], goes as follows:

Let $M$ be a compact subset of $\mathbb{R}^N$, and set $M_\varepsilon = M \cap \{t \in \mathbb{R}^N : |t| \le \varepsilon\}$. Suppose that, for all $\varepsilon > 0$,

$$\inf_{t\in M_\varepsilon} \max_i |C_i(t)| > 0.$$

Then, as $u \to \infty$,

$$\mathbb{P}\{\nabla f_u = 0 \text{ for some } t \in M,\ t \neq 0\} \to 0.$$

That is, $f_u$ only has critical points when $C(t)$ does.

It should be clear by now that the existence of a Slepian process opens a Pandora's box of other results about the sample paths of Gaussian random fields, and that these seem to be of a particularly simple form when talking about high level excursions.

Soon we shall investigate this approach further, including the non-Gaussian scenario. While this is not quite as easy, a surprising amount of rigorous results exist, all of which can be put together with approximations to obtain results that can be very useful in practice. Firstly, however, we shall look at some more results for Gaussian fields.

## 6.3 More on the Local Structure of Gaussian Fields

The Palm distribution of the previous section can be used to obtain further interesting and useful information on the structure of Gaussian random fields, conditioned on various kinds of behavior. Here are some examples.

Once again, $f$ will be assumed to be centered and satisfying the regularity conditions of Section 4.1. In addition, we assume that it is stationary and ergodic, with covariance function $C$ and unit variance.

### 6.3.1 Low Maxima and High Minima

In the preceeding section, we developed a model for the behavior of a Gaussian field given that it had a local maximum, of height $u$, at the origin. This was Theorem 6.2.2. We then used this to show that if $u \to \infty$ the behavior of $f$ in a small neighborhood of the origin was essentially deterministic.

If, however, we condition on the occurence of a local minimum, rather than local maximum, at a high level then the Palm distributions are quite different. To see this, we start by looking at low maxima. The basic Slepian model still holds, and the density of the random vector $Z_u$ there is still given by (6.2.3). What does change, however, are limiting distributions, and the first one is that the random vector $Z = \lim_{u\to-\infty} |u| Z_u$, has the density

$$p(z) = \begin{cases} k \det(z)\, e^{-\mu_{02}\mu_{2\cdot0}^{-1} z'} & \text{if } z \in \mathcal{P}, \\ 0 & \text{otherwise,} \end{cases} \tag{6.3.14}$$

for an appropriate normalising constant $k$. (If $Z$ is a vector, how can it have a determinant and its density be defined over $\mathcal{P}$?)

Whereas (6.3.14) is quite easy to show (cf. Exercise 6.7.4) its consequence for the corresponding Slepian model is somewhat more complicated. Nevertheless, mimicking the arguments that led to (6.2.13), and requiring extra smoothness[5] on the covariance function one can show, in the notation of Theorem 6.2.2, that

$$\begin{aligned} \lim_{u\to-\infty} |u|^3 \left( f_u\left(\frac{t}{u}\right) - u \right) \stackrel{\mathcal{L}}{=} & -\frac{1}{4!} \sum_{ijkl} t_i t_j t_k t_l \frac{\partial^4 \alpha(t)}{\partial t_i \partial t_j \partial t_k \partial t_l}\Big|_{t=0} \\ & + \frac{1}{3!} \sum_{ijk} t_i t_j t_k \frac{\partial^3 g(t)}{\partial t_i \partial t_j \partial t_k}\Big|_{t=0} - \frac{1}{2} t Z t', \end{aligned} \tag{6.3.15}$$

(Again, here $Z$ must be a matrix, not a vector.) where, once again, the equivalence in law is that of all finite dimensional distributions and can be lifted to almost sure convergence on compacts.

This, of course, is quite different to the limiting model for high maxima, since it involves random terms, through $g$ and $Z$. Furthermore, the structure of the limit is that of a fourth order polynomial, rather than a quadratic one. For further details, see [56].

Turning from asymptotically low maxima back to asymptotically high minima is now an easy exercise due to the symmetry of Gaussian processes. In fact, if $f^u$ is the Slepian model process for a $f$ given a local minimum of height $u$ at the origin, then it is immediate from (6.3.16) that

[5] For the argument to work in the case, it is necessary to assume that $C$ has continuous partial derivatives of up to order six, or two more than we usually require. Given the quartic nature of the final limit in (6.3.16) the extra assumptions are not surprising.

$$\lim_{u\to\infty} u^3\left(f^u\left(\frac{t}{u}\right)-u\right) \tag{6.3.16}$$

has the same distribution as minus one times the right hand side of (6.3.16). Thus, high level maxima, and high level minima, behave quite differently.

### 6.3.2 High Excursions

While in Section 6.2 we concentrated on conditioning on the local maxima of random fields, there are many other conditioning events that we could have chosen. For example, in [97] a Slepian model was built for a $f : \mathbb{R}^N \to \mathbb{R}$ conditioned on the following events:

$$f(0) = u, \quad \frac{\partial f}{\partial t_i}(0) = 0,\ i = 1,\dots,N-1, \quad \frac{\partial f}{\partial t_N}(0) > 0. \tag{6.3.17}$$

These are the points we first met in Chapter 3 when we began computing Euler characteristics of excursion sets. They are the points marked with arrows in Figure 3.1.1, if one thinks of the sets there as being excursion sets of $f$.

When $N = 1$ the conditioning event reduces to the very simple event that $f$ have an upcrossing at the origin, the case with which the theory of Slepian processes first began in [46, 85]. Motivated by this case, and for lack of a better terminology, we shall say that $f$ has an upcrossing at the origin of the level $u$, in the $N$-th direction, if $f$ satisfies (6.3.17).

We shall let you turn to [97] for the details of the full model describing such upcrossings, recalling here just one of the high level results. This result describes the position of a the nearest local maximum to a level crossing, something which is of intrinsic applied interest when data comes in the form of level manifolds (contour lines when $N = 2$) and it is necessary to make inference about the unseen part of the field.

Thus, let $f_u^{\uparrow}$ be the process $f$ conditioned, in the Palm sense, to have an upcrossing of $u$ in the $N$-th direction at the origin. Then

$$\lim_{u\to\infty} u\left(f_u^{\uparrow}\left(\frac{t}{u}\right)-u\right) \stackrel{\mathcal{L}}{=} \eta t_N - \tfrac{1}{2} t\Lambda t', \tag{6.3.18}$$

where the convergence is, as usual, weak convergence that can be lifted to almost sure convergence over compacts. The positive random variable $\eta$ has Raleigh density

$$p(\eta) = \eta e^{-\eta^2/2}, \qquad \eta \ge 0.$$

An interesting and practically useful result that comes out of this model describes the distribution of the closest local maximum of $f$ to the origin, for which both its position and height can be deduced from this (although, for a fully rigorous proof, you will need the full Slepian model). Letting $\tau_u$ denote this maximum, we have

$$\lim_{u\to\infty}\left(u\tau_u,\, u\left(f_u^{\uparrow}(\tau_u)-u\right)\right) \stackrel{\mathcal{L}}{=} \left(\eta\left(-\bar{\Lambda}\bar{\lambda}',\, 1\right),\, \eta^2/2\right), \tag{6.3.19}$$

where $\bar{\Lambda}$ and $\bar{\lambda}$ come from partitioning the matrix $\Lambda$ as

$$\Lambda = \begin{pmatrix} \bar{\Lambda} & \bar{\lambda}' \\ \bar{\lambda} & \lambda_{NN} \end{pmatrix}.$$

The second of the random variables in (6.3.19) is called the *normalised overshoot* of $f$, being the additional height the random field goes beyond $u$ in the neighborhood of the conditioning point. Note that the normalised overshoot has an asymptotic exponential distribution. We shall return to this in a moment, in Section 6.3.3, and show how to derive overshoot distributions without recourse to a full Slepian model.

As you might guess, there are many other ways to condition a random field, and each will generate and appropriate Slepian model, from which, when appropriate, some simplified asymptotic model arises. Although the Slepian models are not always easily tractable, there are often shortcuts to the asymptotics. Nosko [66, 67], building on early and important work of Belyaev and his school (e.g. [16, 17, 18]) showed that if $f$ satisfies

$$f(0) = u, \quad \nabla f(0) = v, \tag{6.3.20}$$

then, in the usual sense,

$$\lim_{u\to\infty} u\left(f_u\left(\frac{t}{u}\right) - u - tv'\right) \stackrel{\mathcal{L}}{=} -\tfrac{1}{2}t\Lambda t'. \tag{6.3.21}$$

Comparing (6.3.20) to (6.2.13) for the case $v = 0$, we obtain precisely the same limiting model. In other words, at high levels, $f$ behaves the same conditioned on having a critical point of any kind as it does conditioned on having a local maximum. However, this should not be surprising, for we already saw in Section 5.2 that, with high probablility, the only critical points above high levels are local maxima.

### 6.3.3 Overshoot Distributions

We now want to return to the asymptotic overshoot distribution of (6.3.19), phrase the conditioning event slightly differently, and see how to obtain the distribution without formal recourse to a full Slepian model. The value of this approach will become clear in Section 6.4 when we look at non-Gaussian fields.

Suppose we observe that $f(0) \geq u$. Then we know that the origin belongs to a component of the excursion set of $f$ and that somewhere over this component $f$ must have a local maximum. What can be said about the (conditional) distribution of this maximum, if we require no information about its position? Clearly (WHY?) this distribution is equivalent to the conditional distribution

of the height of a local maximum at the origin, given that such a maximum occurred and exceeded $u$.

Write $O_u = f(0) - u$ for the overshoot of a local maximum of $f$ at the origin. Then the Palm distribution arguments of Section 6.1 imply that (the tail of) such a distribution must be

$$\begin{aligned}\mathbb{P}\left\{f(0) \geq u+v \,\middle|\, f \text{ has a local maximum at 0 of height at least } u\right\} \\ = \mathbb{P}\left\{O_u \geq v \,\middle|\, O_u \geq 0\right\} \\ = \frac{\mathbb{E}\{M_{u+v}\}}{\mathbb{E}\{M_u\}},\end{aligned}$$

(The event $\{O_u \geq v\}$ needs to be defined more carefully. As it is now, it reads as $\{f(0) - u \geq v\}$, but this is clearly wrong because the latter is Gaussian while the former is not.) where $M_u$ is the number of local maxima of $f$ above the level $u$ in some fixed, but unimportant, region, which for convenience we take to be $[0,1]^N$.

The problem now is that we have already seen, back in Chapter 5, that computing the above expectations is an essentially impossible task. On the other hand, in Theorem 5.2.1, we saw that, for large $u$, they are well approximated by the mean Euler characteristic of excursion sets, with an explicit bound on the error of the approximation (cf. (5.2.1)). Substituting this bound into the above, and preparing for a limit, we have that

$$\lim_{u\to\infty} \mathbb{P}\left\{uO_u \geq v \,\middle|\, O_u \geq 0\right\} = \lim_{u\to\infty} \frac{\mathbb{E}\{\varphi(A_{u+v/u}(f,[0,1]^N)\}}{\mathbb{E}\{\varphi(A_u(f,[0,1]^N))\}}. \tag{6.3.22}$$

However, we have an explicit form of the expected Euler characteristic of excursion sets in Theorem 4.2.1, and substituting this into the above gives

$$\begin{aligned}\lim_{u\to\infty} \mathbb{P}\left\{uO_u \geq v \,\middle|\, O_u \geq 0\right\} &= \lim_{u\to\infty} \frac{\exp(-(u+v/u)^2/2)}{\exp(-u^2/2)} \\ &= e^{-v},\end{aligned} \tag{6.3.23}$$

recovering the exponential tail of (6.3.19), as promised.

This time, however, we did not require the full Slepian model in order to obtain the normalised overshoot distribution. In fact, we hardly even needed the fact that $f$ was Gaussian. Equation (6.3.22) required nothing distributional, beyond regularity conditions that ensure that the terms are well defined and finite along with stationarity and ergodicity, the latter so that Palm arguments are appropriate. Only in the last stage of the argument, leading to ~~(6.3.23)~~(6.3.22) did we require Gaussianity. (Because Theorem 5.2.1 required Gaussianity.) This was to justify the claim that the mean number of maxima above a high level $u$ and the mean Euler characteristic of the corresponding excursion set were asymptotically the same, and then to compute the latter. In Section 6.4 we shall exploit these observations to treat non-Gaussian fields.

Similar arguments, using (6.3.23) above and (6.2.13) which describes the stucture of high maxima allow one to say even more about high level excursions, but we shall posptone this until Section 6.5.

### 6.3.4 Numerical Approaches

So far, our discussion of Palm conditioning and Slepian models has taken two related, but slightly different paths. In the first, we described specific, detailed and explicit Slepian models such as that of Theorem 6.2.2 for random fields given a local maximum at the origin of height $u$. Then either using the Slepian process or bypassing it, we obtained asymptotic results related to high levels.

Whereas it is true that what is of most interest to us is the high level behavior of random fields, there are many instances and areas of application in which moderate levels are also important. In those situations, as in high level problems, the explicit Slepian model is, in fact, just a first step towards a useful result.

The problem with a model such as that for local maxima is that its components are rather complicated. In particular, the non-homogeneous random field $g$ appearing in (6.2.10) has a complicated covariance function, and there is not much that one can say about its behavior. Consequently, there is not much that one can say about the behavior of the Slepian model for moderate levels $u$.

The only way to handle this difficulty is computationally, whether this be the numerical computation of distributions or simulation. We shall have more to say about this later in Chapters 10 and 11, but for the moment briefly describe a couple of approaches.

In the astrophysics literature, there is considerable interest in random variables such as the distribution of curvatures at the points of local maxima of two and three dimensional random fields. Since curvatures are functions of the eigenvalues of the Hessian $\nabla^2 f$, the starting point for these calculations is the (Palm) conditional density $\psi_u(f'')$ of (6.2.4)[6], that is,

$$\psi_u(f'') \stackrel{\Delta}{=} \begin{cases} \frac{\det f''\, p(f''|u,0)}{\int_{\mathcal{P}} \det f''\, p(f''|u,0) df''} & \text{if } f'' \in \mathcal{P}, \\ 0 & \text{otherwise.} \end{cases} \tag{6.3.24}$$

This is actually a rather complicated density. Its support $\mathcal{P}$ of positive definite matrices is not a simple subset of $\mathbb{R}^{N(N+1)/2}$, and consequently the normalising constant in the denominator cannot be explicitly evaluated. (Are you sure? Not even for isotropic fields?)

Nevertheless, in papers such as [15, 29], written for the astrophysics community, you will find a number of useful approximate formulas for distributions of curvatures and other variables, formulas which, not surprisingly, involve

[6] Actually, since we have not yet proven that this is the meaning of $\psi_u$, you should prove it as Exercise 6.7.6.

heavy numerical integrations. Also not surprisingly, in view of what we have been doing in this section, these formulas simplify considerably when the level $u$ becomes large. We shall have more to say about this in Section 10.1.

Another computational approach to Slepian and other Palm distributions has been undertaken by a large group of researchers based in Lund, motivated by applications in oceanography. Their interest lies in random variables such as wave period and crest-trough wave height, and they have developed a toolbox of Matlab routines for the statistical analysis and simulation of random waves and random loads. This software, known as *WAFO*, for *Wave Analysis for Fatigue and Oceanography* is freely distributed, and can handle the statistical modelling, calculation and analysis of random waves and wave characteristics and their statistical distributions. A description of this software can be found in [24], with the theoretical background, which has much in common with our current chapter, in [43, 58].

We shall return to WAFO, in more detail, in Section 11.4

## 6.4 The Local Structure of Some Non-Gaussian Fields

We now turn our attention to a number of non-Gaussian, but Gaussian related, random fields for which much of the analysis we have undertaken so far can, to a large extent, be replicated.

In principle, all that we have done for the Gaussian case is doable for non-Gaussian random fields as well. As we pointed out, there is nothing intrinsically Gaussian in the definition of Palm distributions themselves, and even basic objects such as the density $\psi_u$ of (6.2.4) and (6.3.24) are not Gaussian. Gaussianity only entered when we began the specifics and the asymptotics, and then exploited the simple form of the multivariate Gaussian distribution.

In principle, therefore, we should be able to do the same thing again, as long as we are prepared to work hard enough. In practice, however, doing this we quickly encounter intractable algebra and calculus, and so it is hard to develop, for example, useful Slepian models. Asymptotic results, however, are accessible.

In this section we shall show in some detail how to do this for $\chi^2$ random fields, and then give a listing of results for some other Gaussian related examples. The first results of these kinds for non-Gaussian random processes and two dimensional surfaces were in [9, 10, 11], but these were expanded in considerable detail in the McGill PhD thesis of Jin Cao [25] whose results form the basis of this section. (See also [26].)

### 6.4.1 $\chi^2_k$ fields

The $\chi^2_k$ random field, which we have met previously, is defined as

$$f(t) = \chi_k^2(t) = \sum_{j=1}^{k} \epsilon_j^2(t),$$

where the $\epsilon_j$ are all independent, identically distributed Gaussian random fields. We shall assume that the $\epsilon_j$ satisfy the assumptions we required above for the Gaussian case. Thus they are mean zero, unit variance, stationary and ergodic, and satisfy the regularity conditions of Section 4.1.

In order to replicate the Gaussian arguments, the first thing we need to understand is the joint distribution of $f$, $\nabla f$ and $\nabla^2 f$. This is not as simple as in the Gaussian case, when all such multivariate distributions were Gaussian, and all one had to do was to sort out the covariances~~, and t~~. The following lemma, due in its original form to Worsley [99], is the key tool in sorting this out.

**Lemma 6.4.1.** *If $f$ is a $\chi_k^2$ random field as above, then we can express the joint distribution of $\nabla f(t)$ and $\nabla^2 f(t)$ as follows:*

$$f \stackrel{\mathcal{L}}{=} U, \tag{6.4.1}$$

$$\nabla f \stackrel{\mathcal{L}}{=} 2U^{1/2} z, \tag{6.4.2}$$

$$\nabla^2 f \stackrel{\mathcal{L}}{=} 2\big(P + (zz')z'z - U\Lambda + U^{1/2}H\big), \tag{6.4.3}$$

*where $\Lambda$ is the usual $N \times N$ matrix of second order spectral moments of the $\epsilon_j$'s, $U \sim \chi_k^2$, $z \sim N(0, \Lambda)$, $P \sim \text{Wishart}_N(k-1, \Lambda)$[7]. With some abuse of notation, write $H \sim N(0, \mathcal{E}(\Lambda))$, indicating that the elements of the matrix $H$ are jointly normal with covariances*

$$\mathcal{E}_{ij,kl} = \mathbb{E}\{f_{ij}(t) f_{kl}(t)\} = \mathcal{E}(i,j,k,l) - \lambda_{ij}\lambda_{kl} \tag{6.4.4}$$

*where $\mathcal{E}$ is symmetric in its arguments.*

*Furthermore, $U$, $z$ $P$, and $H$ are all independent of one another.*

**Proof.** (The entire proof is inconsistent in the use of row vs. column vectors, leading to incorrect results. I can fix it if you tell me if $\nabla f$ should be a row or column vector. Fix a $t \in \mathbb{R}^N$.

[7] Recall that the $\text{Wishart}_N(n, \Sigma)$ distribution is defined as the distribution of a $N \times N$ matrix $W$ with elements of the form $W_{ij} = \sum_{m=1}^{n} X_{mi} X_{mj}$, where the $n$ vectors $X_m = (X_{m1}, \ldots, X_{mN})$ are independent, each distributed as $N(0, \Sigma)$. The density function is given by

$$\frac{|\det(W)|^{(N-n-1)/2}}{2^{nN/2} |\det \Sigma|^{n/2} \Gamma_N(n/2)} e^{-\text{Tr}(\Sigma^{-1} W)},$$

for $W \in \mathcal{P}$, and 0 otherwise, where $\Gamma_N(n/2) = \pi^{N(N-1)/4} \prod_{j=1}^{N} \Gamma[(n+1-j)/2]$.

Writing $\epsilon$ for the vector of the $\epsilon_j$ defining $f$, set $U = f(t) = \epsilon(t)\epsilon'(t)$. Note that $\partial f/\partial t_i = 2\sum_{l=1}^{k} \epsilon_l \partial\epsilon_l/\partial t_i$. Condition on $\epsilon$ so that we can write

$$\nabla f \;=\; 2(\nabla\epsilon)\epsilon' \;\sim\; N(0, 4U\Lambda).$$

Since this depends on $\epsilon$ only through $U$, it is also the conditional distribution when conditioning on $U$ alone. Setting $z' = U^{-1/2}(\nabla\epsilon)\epsilon'$ then gives (6.4.2).

As for (6.4.3), we start by noting that

$$\frac{\partial^2 f}{\partial t_i \partial t_j} = 2\sum_{l=1}^{k} \frac{\partial\epsilon_l}{\partial t_i}\frac{\partial\epsilon_l}{\partial t_j} + 2\sum_{l=1}^{k} \epsilon_l \frac{\partial^2\epsilon_l}{\partial t_i \partial t_j}$$

Condition now on both $\epsilon$ and $\nabla\epsilon$, and use the result of Exercise 2.8.10 (ii) and (iii), applied to each $\epsilon_j$, to see that, conditionally, the above equality implies

$$\nabla^2 f \;\sim\; N\big(2(\nabla\epsilon)'(\nabla\epsilon) - U\Lambda),\, 4U\mathcal{E}(\Lambda)\big),$$

where $\mathcal{E}(\Lambda)$ is given by (6.4.4). Consequently, still conditioning on both $\epsilon$ and $\nabla\epsilon$, we can write

$$\nabla^2 f = 2\big((\nabla\epsilon)'(\nabla\epsilon) - U\Lambda + U^{1/2}H\big), \tag{6.4.5}$$

with $H$ as in the statement of the lemma.

With the conditioning still in play, and writing $I$ for the $k \times k$ identity matrix, set

$$A = I - U^{-1}\epsilon'\epsilon, \qquad P = (\nabla\epsilon)A(\nabla\epsilon)'.$$

Since $z = U^{-1/2}\epsilon(\nabla\epsilon)'$, we can rewrite (6.4.5) as

$$\nabla^2 f = 2\big(P + z'z - U\Lambda + U^{1/2}H\big),$$

where $H$ is independent of everything else. This is the claim (6.4.3) of the lemma, although we still need to establish the distribution of $P$ and the independence of $P$, $z$ and $U$.

Fix now only $\epsilon$. Then $A$ is also fixed and of rank $k-1$ and so $P \sim$ Wishart$(k-1, \Lambda)$. NEED A REFERENCE FOR THIS. (I CAN CHECK ONCE I FIX THE PROOF AND UNDERSTAND WHAT $P$ IS.)

Regarding independence, recall first that if $M_1$ and $M_2$ are orthogonal matrices, in the sense that $M_1'M_2 = 0$, and $x$ and $y$ are any two Gaussian vectors, then the vectors $M_1x$ and $M_2y$ are independent. Take $M_1 = A = A'$ and let $M_2$ be the matrix with $U^{-1/2}\epsilon'$ in the first row and zeros elsewhere. Then, since $A\epsilon' = 0$, we also have $M_1'M_2 = 0$. Now take $x$ and $y$ to be any two rows of $\nabla\epsilon$. Then $M_1x$ is a column of $A(\nabla\epsilon)'$ and $M_2y$ is a column vector all of whose elements are zero bar one, which is one of the $z_i$. Since the rows we started with were arbitrary, this implies that, conditional on $\epsilon$, we have

independence between $z$ and $A(\nabla\epsilon)'$. Noting that $A^2 = A$ it follows that since $P$ is a quadratic form in $A(\nabla\epsilon)'$ it is also conditionally independent of $z$.

Finally, since the distributions of $P$ and $z$ depend neither on $\epsilon$ nor $U$, they are also independent unconditionally, and the proof is complete. □

There are some very useful results that follow straightforwardly from Lemma 6.4.1 which we shall now describe, along with a sketch of how to prove them. Details can be found in [25].

For a start, consider the Palm distribution of the Hessian matrix $f'' = \nabla^2 f$ given the occurrence of a local maximum at the origin. Going through the Palm arguments of Sections 6.1 and 6.2 it is straightforward to see that this is given by

$$\psi(f'') = \mu_u |\det f''| \, \psi\left(f'' \,\middle|\, \nabla f = 0, f = u\right) \tag{6.4.6}$$

for $f'' \in \mathcal{N}$ and zero otherwise. The factor $\mu_u$ includes information on the joint density of $\nabla f$ and $f$ as well as the mean number of local maxima of $f$ above the level $u$, but it is best now to think of it purely as a normalisation constant and concentrate on how the conditional density $\psi$ behaves as $u$ changes.

Firstly, consider $u^{-1}\nabla^2 f$. By Lemma 6.4.1 its distribution, given that $\nabla f = 0$ and $f = u$ is the same as that of the random matrix

$$2u^{-1}\left(P - u\Lambda + u^{1/2}H\right).$$

As $u \to \infty$ this random matrix converges in distribution to the constant matrix $-2\Lambda$, and applying this to (6.4.6) gives us that the Palm distribution of $u^{-1}\nabla^2 f$ at high local maxima converges (in law) to the matrix $-2\Lambda$.

Note that in order to carry out this limit we did not need to know very much about the explicit form of the density $\psi$. The structure provided by Lemma 6.4.1 sufficed.

On the other hand, now send $u \to 0$ (Did you mean $u \to -\infty$?) in (6.4.6), assuming a low minimum at the origin. (Of course, the domain of the density has to be changed to $f'' \in \mathcal{P}$ and the absolute value signs around the determinant can be dropped.) In this case, there is no need to normalise $\nabla^2 f$, and the same argument gives us that the limiting density of $\nabla^2 f/2$ is, in this case, proportional to

$$\det f'' \psi(f'').$$

However, given than $\psi$ here is a $\text{Wishart}_N(k-1, \Lambda)$ density, and given the form of this density (cf. Footnote 7) it follows that, as $u \to 0$, $\nabla^2 f \to 2 \times \text{Wishart}_N(k+1, \Lambda)$.

While the above arguments describe the structure of the matrix $\nabla^2 f$ at high maxima and low minima, it was the behavior of the entire process in a neighborhood of the origin that has been of more interest to us. Developing a

full Slepian model is not a useful way to proceed, but the same ideas, along with asymptotics, give useful results quite quickly.

The first step involves an approximation for the vector valued random field $\epsilon$, given the values of the vector $\epsilon$, the $k \times N$ matrix $\nabla\epsilon$ and the $k \times N \times N$ array $\nabla^2\epsilon$ at the origin. Note that at this point we are interested not in Palm conditioning, but in usual conditioning of some variables on others. The next step is to note the simplifying fact that since the $\epsilon_j$ were assumed independent, it suffices to work with one $\epsilon_j$ at a time. Since everything is jointly Gaussian, it is not too hard[8] to show that the conditional random fields $\epsilon_j$ have, up to $o(|t|^2)$, the same finite dimensional distributions as the fields

$$\epsilon_j(0) \,+\, t\nabla\epsilon_j(0) \,+\, \tfrac{1}{2}t(\nabla^2\epsilon_j(0))t'.$$

Consequently, under the same conditioning event, the $\chi^2_k$ field $f$, scaled in space, can be written as

$$\begin{aligned} f\left(\frac{t}{\sqrt{u}}\right) &\simeq \sum_{j=1}^{k}\left[\epsilon_j(0) \,+\, \frac{1}{\sqrt{u}}t\nabla\epsilon_j(0) \,+\, \frac{1}{2u}t(\nabla^2\epsilon_j(0))t'\right]^2 \\ &\simeq f(0) \,+\, \frac{1}{u}t\nabla\epsilon(0)(\nabla\epsilon(0))'t' \,+\, \frac{2}{\sqrt{u}}t\epsilon(0)(\nabla\epsilon(0))' \\ &\qquad\qquad + \frac{1}{u}\sum_{j=1}^{k}\epsilon_j(0)t(\nabla^2\epsilon_j(0))t' \\ &\stackrel{\mathcal{L}}{=} f(0) \,+\, \frac{1}{\sqrt{u}}t\,\nabla f(0) \,+\, \frac{1}{2u}t\,\nabla^2 f(0)t'. \end{aligned}$$

where by $\simeq$ we mean 'up to $o(|t|^2/u)$, in an appropriate sense, in law'.

Given this structure, now condition, in the Palm sense, on a local maximum of $f$ at height $u$ at the origin. Then $f(0) = u$, $\nabla f(0) = 0$ and, as $u \to \infty$, we established above that $u^{-1}\nabla^2 f(0) \to -2\Lambda$, giving us that

$$f\left(\frac{t}{\sqrt{u}}\right) - u \;\to\; -t\Lambda t'$$

as $u \to \infty$. A similar argument for low minima also works, and we summarise all the results of this section in the following theorem, which we state far more carefully than is justified given the rather detail challenged tone of the section so far. The theorem also includes a result about overshoots, which you should prove for yourself in Exercise 6.7.7.

**Theorem 6.4.2.** *Let $f = \sum_{j=1}^{k}\epsilon_j^2$ be a $\chi^2_k$ random field on $\mathbb{R}^N$, for which the Gaussian component processes $\epsilon_j$ are zero mean, unit variance, stationary*

[8] The argument involves writing out the full covariance matrix of all variables, and then expanding the various covariance terms as functions of $t$. In principle, this is not hard. In practice, it takes a while. The details are in [25] where you will also find a more serious and honest treatment of the error term.

*and ergodic, and satisfy the regularity conditions of Section 4.1. As usual, $\Lambda$ denotes the matrix of second order spectral moments of the $\epsilon_j$.*

*Then, conditioning, in the Palm sense, on a local maxima of $f$ of height $u$ at the origin, we have, as $u \to \infty$,*

$$\nabla^2 f(0) \to -2\Lambda \tag{6.4.7}$$

$$f\left(\frac{t}{\sqrt{u}}\right) - u \to -t\Lambda t', \tag{6.4.8}$$

*where, in both cases, the convergence is in distribution, and in* (6.4.8) *can also be taken to be almost sure convergence over compacts. Furthermore,*

$$\mathbb{P}\left\{f(0) - u \geq v \,\middle|\, \textit{maximum of height} \geq u \textit{ at } 0\right\} \to e^{-v/2}, \quad v \geq 0. \tag{6.4.9}$$

*Similarly, conditioning on a local minimum of height $u$ at the origin, and letting $u \to 0$, we have*

$$\nabla^2 f(0) \to 2 \times \textit{Wishart}_N(k+1, \Lambda) \tag{6.4.10}$$

$$f\left(\sqrt{u}t\right)/u \to 1 + tBt', \tag{6.4.11}$$

*where $B \sim \textit{Wishart}_N(k+1, \Lambda)$ and*

$$\mathbb{P}\left\{u - f(0) \geq uv \,\middle|\, \textit{minimum of height} \leq u \textit{ at } 0\right\} \to (1-v)^{(k-N)/2}, \tag{6.4.12}$$

*for $0 < v < 1$ and $k > N$.*

This is (almost) all we shall have to say about the extrema of $\chi^2_k$ random fields, and so now we turn to turn other examples of Gaussian related fields.

### 6.4.2 Student $T$ fields

In this and the following subsection we are going to state some results without any attempt at proofs. The proofs all follow the style of proof of the previous subsection, differing in detail and being somewhat more complicated in their preliminary stages.

Recall that the first stage of the previous derivation relied on finding a representation for the (non-Palm) distribution of $f$, $\nabla f$ and $\nabla^2 f$. We shall give these, which are due to Worsley [99], and then follow them with results analagous to those of Theorem 6.4.2, these being due to Cao [25, 26]. This should enable you to attempt proofs by yourself.

Throughout this and the following subsection we shall assume that the conditions of Theorem 6.4.2 hold for the random fields $\epsilon_j$ on $\mathbb{R}^N$.

The first result is that if

$$f(t) = \frac{\epsilon_1(t)}{m^{-1/2}\left(\sum_{j=2}^{m+1} \epsilon_j^2(t)\right)^{1/2}}$$

is a Student's $T$ field with $m$ degrees of freedom, then the joint distribution of $f(t)$, $\nabla f(t)$ and $\nabla^2 f(t)$ can be written as follows.

$$\begin{aligned} f &\stackrel{\mathcal{L}}{=} T, \\ \nabla f &\stackrel{\mathcal{L}}{=} m^{1/2}\left(1+T^2/m\right) S^{-1/2} z_1, \\ \nabla^2 f &\stackrel{\mathcal{L}}{=} m^{1/2}\left(1+T^2/m\right) S^{-1/2} \\ &\qquad \times \left[-m^{-1/2} T(Q - 2z_1' z_1) - z_1' z_2 - z_2' z_1 + S^{1/2} H\right] \end{aligned}$$

where $T$ has a $T$ distribution with $m$ degrees of freedom, $S \sim \chi^2_{m+1}$, $z_1, z_2 \sim N(0, \Lambda)$, $Q \sim \text{Wishart}_N(m-1, \Lambda)$ and $H \sim N(0, \mathcal{E}(\Lambda))$, as in (6.4.4). Furthermore, all the random variables are independent of one another.

Conditioning, in the Palm sense, on a local maxima of $f$ of height $u$ at the origin, the following holds as $u \to \infty$.

$$\begin{aligned} u^{-1}\left(1+u^2/m\right)\nabla^2 f(0) &\to -A^{-1}B \\ u^{-1} f\left(\frac{t}{u}\right) &\to \left(\frac{mA}{mA + tBt'}\right)^{1/2} \end{aligned}$$

where, $A \sim \chi^2_{m+1-N}$, $B \sim \text{Wishart}_N(m+1, \Lambda)$ are independent. Furthermore,

$$\mathbb{P}\left\{ f(0) \geq \frac{u}{\sqrt{1-v}} \,\middle|\, \text{maximum of height} \geq u \text{ at } 0 \right\} \to (1-v)^{(m-N)/2},$$

for $0 < v < 1$ and $m > N$.

### 6.4.3 *F* fields

Analgous results hold for $F$ fields, with $n$ and $m$ degrees of freedom, given by

$$f(t) = \frac{\sum_{i=1}^n \epsilon_i^2(t)/n}{\sum_{i=n+1}^k \epsilon_i^2(t)/m}.$$

The joint distribution of $f(t)$, $\nabla f(t)$ and $\nabla^2 f(t)$ can be written as follows.

$$\begin{aligned} f &\stackrel{\mathcal{L}}{=} (n/m)G, \\ \nabla f &\stackrel{\mathcal{L}}{=} 2G^{1/2}(1+G)W^{-1/2} z_1 \\ \nabla^2 f &\stackrel{\mathcal{L}}{=} 2(1+G)\Big[W^{-1/2}\left(P - GQ + (1+3G) z_1' z_1 - G^{1/2}\left(z_1' z_1 + z_2' z_2\right)\right) \\ &\qquad + G^{1/2} W^{-1/2} H\Big], \end{aligned}$$

where $G \sim (n/m) F_{n,m}$, $W \sim \chi^2_{m+n}$, $z_1, z_2 \sim N(0, \Lambda)$, $P \sim \text{Wishart}_N(n-1, \Lambda)$, $Q \sim \text{Wishart}_N(m-1, \Lambda)$, and $H \sim N(0, \mathcal{E}(\Lambda))$, as in (6.4.4). Furthermore, all the random variables are independent of one another.

Conditioning, in the Palm sense, on a local maxima of $f$ of height $u$ at the origin, the following holds as $u \to \infty$.

$$u^{-2}\nabla^2 f(0) \to \frac{-2n}{m} A^{-1} B$$
$$u^{-1} f\left(\frac{t}{\sqrt{u}}\right) \to \left(\frac{mA}{mA + ntBt'}\right)^{1/2},$$

where $A \sim \chi^2_{m+n-N}$, $B \sim \text{Wishart}_N(m+1, \Lambda)$ are independent. Furthermore,

$$\mathbb{P}\left\{ f(0) \geq \frac{u}{1-v} \,\middle|\, \text{maximum of height} \geq u \text{ at } 0 \right\} \to (1-v)^{(m-N)/2},$$

for $0 < v < 1$ and $m > N$.

The close connection between the results for $T$ and $F$ random fields comes, of course, from the fact that a $T$ field with $m$ degrees of freedom is equivalent to the square root of a $F$ field with 1 and $m$ degrees of freedom.

There are also results about low level minima, which, since they are of less importance for later chapters, we have not cited. They can be found in [25, 26].

### 6.4.4 On Computing the Mean Euler Characteristic

It is worthwhile breaking the flow for a moment to make a comment that is both of historical and practical interest. (This section is really out of place here.)

Recall the calculations of the formulas for the expected Euler characteristic of Chapter 4. We gave there a quite full derivation for the Gaussian case, and then resorted to the Gaussian kinematic formula for the case of Gaussian-related random fields, a result which we have not proven in this book.

Historically, however, this is not how things happenned. The Gaussian case was proven first, with the earliest results due to Adler in the 1970's (e.g. [1, 2, 3, 4]). Some non-Gaussian results then started appearing, and in 1994 Worsley [99] developed a new approach based on the representations appearing above for $f(0)$, $\nabla f(0)$ and $\nabla^2 f(0)$. The idea was simple: Looking back at the derivation of $\mathbb{E}\{\varphi(A_u)\}$ in Section 4.2 for the Gaussian case, all that one needs to compute are expectations involving these three sets of random variables. Hence, having useful representations for their joint distribution is a major help in carrying out the computations. In fact, you might like to try (in Exercise 6.7.9) doing things this way for the $\chi^2$ case, which is the easiest of the non-Gaussian ones.

As for the comment of practical interest, there is the question of how to derive an expected Euler characteristic formula for a random field which we have *not* treated in this book. Our best advice to you is to turn to learn how to use the Gaussian kinematic formula, turning to *RFG* if needed. As we have

seen for the example of the scale space field of Section 4.6.2 this is not always easy. However, once you have made the investment you will be well equipped to handle as many Gaussian-related fields as you are likely to encounter. On the other hand, if, against our best advice, you prefer to avoid the investment, then adopting Worsley's approach is generally the best second choice. It will, however, almost definitely involve you in heavier calculus.

## 6.5 The Size of an Excursion Set Component

We already know quite a lot about excursion sets of Gaussian and Gaussian related related random fields, the main result being the Gaussian kinematic formulas of Theorems 4.4.1 and 4.8.1, which gave us expressions for the mean values of all their Lipschitz-Killing curvatures. Among these was an expression for the mean value of their volume, a result which hardly needed such powerful techniques since, with $|\cdot|$ as usual indicating Lebesgue measure in the appropriate dimension.

$$\mathbb{E}\left\{\left|A_u(f,T)\right|\right\} = \mathbb{E}\Big\{\int_T \mathbb{1}_{[u,\infty)}(f(t))\,dt\Big\} = \int_T \mathbb{P}\{f(t)\geq u\}\,dt, \quad (6.5.1)$$

with the integral simplifying to $|T|\mathbb{P}\{f(0)\geq u\}$ in the stationary case.

However, none of these results gave us information about individual connected components of $A_u$, something which we wish to look at now.

This seems to be an almost impossible task, for individual components cannot be studied by any of the techniques that we have developed so far. Indeed, despite close to half a century of interest, there has been no real progress in computing either the mean number of components or the mean size of a specific component. (In view of the easy result (6.5.1) these are closely related problems.)

Nevertheless, if we are prepared to consider only high level excursion sets, then there are results that follow rather easily from the asymptotic Palm based models that have made up the bulk of this chapter so far. The basic idea is that since we know from these results that random fields take on specific shapes in the neighborhood of high maxima, and this implies something about the shape of that component of the excursion set under them. We start with the Gaussian case.

### 6.5.1 The Gaussian Case

We now take $f$ to be Gaussian, and, at least for a while, argue heuristically. Suppose that ~~at~~ $f(0) > u$, so that we know that the origin is part of the excursion set $A_u$. If $u$ is large, then somewhere, not too far from the origin, there is, with high probability, a local maximum of $f$, a claim justified by results such as (6.3.18). We know that this maximum has height greater than

$u$, and, by (6.3.23), the additional, overshoot, is distributed as $u^{-1}$ times an exponential variable, $X$ say. The picture is as follows:

PUT A FIGURE HERE

From (6.2.12) we know that, in the neighborhood of this local maximum, which for convenience we move to the origin, and which is now of height $u + u^{-1}X$, the random field behaves like

$$u + u^{-1}X \; - \; \frac{u + u^{-1}X}{2} t\Lambda t',$$

where $X \sim \exp(1)$. If we look only at this function above the level $u$, we have that it is given by

$$u^{-1}X \; - \; \frac{u + u^{-1}X}{2} t\Lambda t'.$$

To compute the size of the excursion set of this function above zero, which is what we are looking for, note firstly the easily checked fact that the excursion set above the zero level of the elliptic paraboloid

$$f(t) = u - t\Lambda t' \tag{6.5.2}$$

on $\mathbb{R}^N$ is an ellipsoid with $N$-dimensional Lebesgue measure

$$u^{N/2}\omega_N(\det\Lambda)^{-1/2}, \tag{6.5.3}$$

where, as usual, $\omega_N$ is the volume of the unit ball in $\mathbb{R}^N$.

If we now write

$$S_u \stackrel{\Delta}{=} \big|\, \{t : \; f(t) \geq u, \;\; t \text{ and } 0 \text{ in same connected component of } A_u\} \,\big|,$$

applying (6.5.3) to the random field suggests that

$$\begin{aligned} S_u &\simeq (u^{-1}X)^{N/2}\omega_N \left( \frac{u + u^{-1}X}{2} \right)^{-N/2} (\det\Lambda)^{-1/2} \\ &\simeq u^{-N}X^{N/2}2^{N/2}\omega_N(\det\Lambda)^{-1/2}, \end{aligned} \tag{6.5.4}$$

and we have an approximate distribution for the size of a single connected component of the excursion set $A_u$.

The above argument can actually be made completely rigorous, with the main technical tool being the fact, noted at the end of Section 6.2.2, that the high level convergence of the Slepian model to a quadratic function can be shown to be almost sure over compact sets. We do not want to get into these detailed arguments in this book[9] but shall suffice by stating a properly formulated theorem in a moment.

[9] We should point out, however, that these arguments are not conceptually difficult, and amount to little more than dotting the i's and crossing the t's (or introducing appropriate $\delta$'s and $\varepsilon$'s) in the heuristic arguments that we have relied on.

Before this, however, it is time for a reality check, to see if (6.5.4) makes sense in terms of other things we know, and to also get a feel for how good the approximation might be. To this end, note first that since $\mathbb{E}\{X^\alpha\} = \Gamma(1+\alpha)$,

$$\begin{aligned}\mathbb{E}\{S_u\} &\simeq u^{-N}\Gamma(1+N/2)2^{N/2}\omega_N(\det\Lambda)^{-1/2}\\ &= u^{-N}(2\pi)^{N/2}(\det\Lambda)^{-1/2}.\end{aligned}$$

Note also that the expected volume of the excursion set over some set $T$ is, by (6.5.1),

$$\mathbb{E}\{|A_u(T)|\} = |T|\Psi(u) \simeq \frac{|T|}{u\sqrt{2\pi}}e^{-u^2/2},$$

the approximation coming from (2.2.2), and being correct up to $o(u^{-1}e^{-u^2/2})$.

Finally, consider the mean Euler characteristic of $A_u(T)$ (e.g. Theorem 4.2.1), taking only the highest order term in $u$, to see that

$$\begin{aligned}\mathbb{E}\{\varphi(A_u(T)\} &\simeq e^{-u^2/2}|T|\,|\Lambda|^{1/2}(2\pi)^{-(N+1)/2}H_{N-1}(u)\\ &\simeq e^{-u^2/2}|T|\,|\Lambda|^{1/2}(2\pi)^{-(N+1)/2}u^{N-1},\end{aligned}$$

where, once again, the approximation is correct up to $o(u^{-1}e^{-u^2/2})$.

Combining the last three approximations immediately yields that

$$\mathbb{E}\{|A_u(T)|\} \simeq \mathbb{E}\{\varphi(A_u(T))\}\,\mathbb{E}\{\text{Size of one component of } A_u\}, \quad (6.5.5)$$

to a degree of accuracy of $o(u^{-1}e^{-u^2/2})$.

The approximation (6.5.5) would actually be an equality, and require no proof, if, for example, we knew *a priori*, that

- Each connected component of $A_u$ was also simply connected, with Euler characteristic of one. In this case $\varphi(A_u)$ would simply count the number of such components.
- The number of connected components and their sizes were independent random variables.

Neither of these conditions ~~is~~ actually holds, but we have often claimed that they should be satisfied in some approximate sense, at least for large $u$. The approximation (6.5.5) gives further justification to these claims, and will form the basis of the heuristic techniques of the following section.

Note, however, that the arguments we have given above yield much more that is available from (6.5.5), which only deals with expectations, since they also yield the actual (asymptotic) distribution of the size of a connected component of the excursion set.

Returning to more rigorous mathematics, here is the promised theorem[10], summarising the above discussion.

[10] The first version of this theorem, for two-dimensional random fields, is due to Nosko [66, 67].

**Theorem 6.5.1.** *Let $f$ be a mean zero, unit variance, stationary, ergodic, Gaussian field satisfying the regularity conditions of Section 4.1. Let $S_u$ denote the $N$-dimensional Lebesgue measure of the component of the excursion set $A_u(\mathbb{R}^N)$ containing the origin. Then, given that*[11] *$S_u > 0$ the conditional Palm distributions converge to an exponential limit, in that*

$$\lim_{u\to\infty} \mathbb{P}\left\{u^N 2^{-N/2}(\omega_N)^{-1}(\det\Lambda)^{1/2} S_u \geq v \,\middle|\, S_u > 0\right\} = e^{-v^{2/N}}, \quad v > 0 \tag{6.5.6}$$

### 6.5.2 The Non-Gaussian Case

Actually, if you look back over the Gaussian case, you will find that there was not very much in the main argument that had to do with the Gaussian assumption. Gaussianess only entered when we turned to the specific structure of the high level Slepian model. However, we already developed such models for a number of non-Gaussian random fields in Section 6.4, and all that one needs do now is substitute these models into the arguments of the previous subsection to obtain the following results (cf. Exercise 6.7.10). All hold only under stationarity and ergodicity.

If $f$ is a $\chi^2_k$ random field on $\mathbb{R}^N$, as in Section 6.4.1, and $S_u$ the size of the connected component of $A_u(f, \mathbb{R}^N)$ containing the origin, then, conditional on $S_u > 0$,

$$u^{N/2} S_u \Rightarrow \omega_N (\det\Lambda)^{-1/2} X^{N/2}, \tag{6.5.7}$$

where $X$ has an exponential distribution with mean 2 and the convergence is weak convergence of (conditional, Palm) distributions.

Similarly, if $f$ is a Student $T$ random field on $\mathbb{R}^N$ with $m > N$ degrees of freedom, as in Section 6.4.2, then, conditional on $S_u > 0$,

$$u^N S_u \Rightarrow \omega_N (\det\Lambda)^{-1/2} m^{N/2} X^{N/2} Y^{N/2} (\det A)^{-1/2}, \tag{6.5.8}$$

where $X \sim \text{Beta}(1, (m-N)/2))$, $Y \sim \chi^2_{m+1-N}$ and $A \sim \text{Wishart}_N(I, m+1)$, independently of one another.

If $f$ is a $F$ random field on $\mathbb{R}^N$, with $m > N$ and $n$ degrees of freedom as in Section 6.4.3, then, conditional on $S_u > 0$,

$$u^{N/2} S_u \Rightarrow \omega_N (\det\Lambda)^{-1/2} n^{-N/2} m^{N/2} X^{N/2} Z^{N/2} (\det A)^{-1/2}, \tag{6.5.9}$$

where $X$ and $A$ are as in (6.5.8) and $Z \sim \chi^2_{m+n-N}$, with all variables independent of one another.

With these results, and some computation, you can now verify that the relationship (6.5.5) holds for these non-Gaussian random fields as well. An explanation for this, as a general phenomenon, comes from the Poisson clumping heuristic.

[11] Note that $S_u > 0 \iff f(0) \geq u$.

## 6.6 The Poisson Clumping Heuristic and Excursions

Two decades ago, David Aldous wrote a superb monograph [6] which he called *Probability Approximations via the Poisson Clumping Heuristic*. The title was a terrible misnomer, since the word "heuristic" gives the misguided impression that "Poisson clumping" is a rough and ready way for illegitimately guessing answers that sometimes work~~s~~ and sometimes do~~es~~ not. Rather than "heuristic", the choice "principle" would have been more appropriate, since the idea of Poisson clumping manages to distill the essence of a large number of related problems to identify a common underlying principle, which almost always[12] works.

The philosophy behind the Poisson clumping heuristic (which we shall abbreviate to *PC heuristic*) is the approximate modelling of random sets (such as excursion sets) as 'mosaic processes'. Mosaic processes[13] have two components: a random subset, or 'clump' $B \subset \mathbb{R}^N$, and a Poisson point process on $\mathbb{R}^N$ with mean measure $\mu$. ~~The~~A mosaic process, $A$, is ~~then~~ determined by choosing a sequence $B_1, B_2 \ldots$, of i.i.d. copies of the random set, and a numbering $x_1, x_2, \ldots$, of the points of the Poisson process, and defining the random set

$$A = \bigcup_k (x_k \oplus B_k)\,, \tag{6.6.1}$$

where $x \oplus B = \{y \in \mathbb{R}^N : y = x + z, \text{ for some } z \in B\}$. As a consequence of the independence of the increments of Poisson processes, and the independence of the clumps $B_k$ many properties of mosaic processes are quite easy to study.

For example, suppose that the random sets $B_j$, are small enough, or the points of the Poisson process sparse enough, that the $B_j$ rarely, if ever, overlap. Writing, as usual, $|\cdot|$ for Lebesgue measure, we then have that for the random mosaic $A$ and any compact subset $T$ of $\mathbb{R}^N$

$$\mathbb{E}\{|A \cap T|\} \simeq \mu(T)\mathbb{E}\{|B|\}. \tag{6.6.2}$$

(Haven't defined $\mu(T)$. I suppose it is the number of expected points of the Poisson process, equal to the rate of the process times $|T|$.) In essence, this approximation, when treated as an equality, is the heuristic. It assumes sparce, Poisson points, small random clumps, and lots of independence.

However, we have seen results like (6.6.2) already, when the point process was the process of local maxima of a random field above a high level, the clumps were connected components of the excursion set, and the mosaic process was the excursion set, $A_u(T)$, considered as a function of $T$. The corresponding result was (6.5.5), viz.

[12] Of course, "almost always" here depends on the measure one places on spaces of problems.

[13] For further details on mosaic processes see Hall [39].

$$\mathbb{E}\{|A_u(T)|\} \simeq \mathbb{E}\{\varphi(A_u(T))\}\,\mathbb{E}\{\text{Size of one component of } A_u\}. \tag{6.6.3}$$

Arguing now backwards from this 'coincidence' leads to the conjecture that the random mosaic model does give an appropriate description of the structure of high level excursion sets, a fact which actually has a strong theoretical background.

In has been known for some time that the number of high level maxima of stationary Gaussian fields $M_u$ ha~~ves~~ an asymptotic Poisson limit, in the sense that, for all Borel $T \subset \mathbb{R}^N$,

$$M_u\left(f, \mu_u^{-1}T\right) \Rightarrow N(T), \tag{6.6.4}$$

(Define notation.)as $u \to \infty$, where $N$ is a Poisson process on $\mathbb{R}^N$ (On $T$ perhaps?) with constant, unit intensity measure,

$$\mu_u = \mathbb{E}\left\{\varphi\left(A_u(f, [0,1]^N\right)\right\} \simeq \mathbb{E}\left\{M_u\left(f, [0,1]^N\right)\right\},$$

and the convergence is in distribution. The formal theory is due to Piterbarg[14], and the only condition needed, beyond those which we have required throughout this chapter, is that

$$\lim_{|t|\to\infty} C(t)\log(t) = 0$$

It is to be presumed that similar Poisson limit theorems hold for the Gaussian related random fields that we studied in Sections 6.4 and 6.5, although, to the best of our knowledge, they have never been proven. However, we have seen that for these random fields (6.6.3) *does* hold, so it seems reasonable to proceed in general as if the excursion sets of these random fields are also described by a mosaic process.

In fact, for the remainder of this chapter we shall adopt the PC heuristic as always holding, and not just for the random fields for which we have already seen that (6.6.3). This will extend the range of application of the next two subsections far beyond the Gaussian, $\chi^2_k$, Student's $T$ and $F$ cases that we have treated so far.

Here are two examples of what can be done with such an assumption.

### 6.6.1 Exceedence Probabilities, Again

All me assume for this application is that we are dealing with a stationary random field $f$ on $\mathbb{R}^N$ for which

(i) We believe the excursion set $A_u$ is well modelled by a mosaic process of simply connected clumps positioned on the points of a Poisson process, of local maxima, with uniform intensity measure

[14] See [74, 75] for this result and [51] for the general theory of Poisson limits of this kind.

$$\mu_u(T) = \mathbb{E}\{M_u(f,T)\} = |T|\mathbb{E}\{M_u(f,[0,1]^N)\}.$$

This is generally a reasonable assumption if $u$ is large and the dependence structure of $f$ decays fast enough. (Recall that in the Gaussian case we require only $C(t)\log t \to 0$ as $|t| \to \infty$, which is a very mild condition.)

(ii) We have an easy way to compute, or at least approximate $\mu_u$ as a function of $u$. The approximation will generally be via the mean Euler characteristic of $A_u$, which is what we shall use in the following.

Then, for a fixed set $T$, and large $u$ (so that $\mu_u$ is small)

$$\begin{aligned}
\mathbb{P}\left\{\sup_{t\in T} f(t) \geq u\right\} &= 1 - \mathbb{P}\left\{A_u(T) = \emptyset\right\} \\
&= 1 - \mathbb{P}\{\text{no excursion clumps in } T\} \\
&\simeq 1 - e^{-\mu_u(T)} \\
&\simeq \mu_u(T) \\
&\simeq \mathbb{E}\{\varphi\left(A_u(f,T)\right)\},
\end{aligned}$$

and so we have re-established the EEC heuristic of Section 5.3, but *without any assumption of Gaussianity.*

Of course, there are prices to pay for not assuming Gaussianity. One is that the above argument gives little indication of the level of accuracy of the EEC approximation, which we did have from the purely Gaussian approach. The other is that the underlying assumption that the PC heuristic can be applied is essentially uncheckable.

On the other hand, it does indicate that the EEC heuristic should hold for a wide class of random fields.

### 6.6.2 The Largest Component of an Excursion Set

A random variable of considerable interest in recent applications of excursion theory to fMRI imaging is the size of the largest connected component of a medium to high level excursion set. We shall discuss the application itself in Chapter 9 and now just look at the theory.

Given an excursion set $A_u$, made up of a number $N = N_u$ of connected components, let $S_1^u, \ldots, S_N^u$ denote the sizes of the various components and $S_{\max}^u = \max_{1\leq i\leq N_u} S_i^u$ the largest of them. Assume that we know, or can at least approximate, the distributions of the $S_j^u$, typically via the high level Slepian models of Section 6.5. Furthermore, assuming that we know how to compute $\mathbb{E}\{\varphi(A_u)\}$, so that we can use the version (6.6.3) of the PC heuristic to compute $\mathbb{E}\{N_u\}$.

Then, using the assumed mosaic structure of the excursion set, we have

$$\begin{aligned}
&\mathbb{P}\Big\{S^u_{\max} \le v \,\Big|\, N_u \ge 1\Big\} \qquad (6.6.5)\\
&\quad = \sum_{n=1}^{\infty} \mathbb{P}\left\{S^u_i \le v,\ 1 \le i \le n,\ N_u = n \,\middle|\, N_u \ge 1\right\}\\
&\quad = \sum_{n=1}^{\infty} \mathbb{P}\left\{N_u = n \,\middle|\, N_u \ge 1\right\} \mathbb{P}\left\{S^u_i \le v,\ 1 \le i \le n,\ \middle|\, N_u = n\right\}\\
&\quad \simeq \frac{1}{\mathbb{P}\{N_u \ge 1\}} \sum_{n=1}^{\infty} \frac{(\mathbb{E}\{N_u\})^n}{n!} e^{-\mathbb{E}\{N_u\}} \left(\mathbb{P}\left\{S^u \le v\right\}\right)^n\\
&\quad = \frac{e^{-\mathbb{E}\{N_u\}\mathbb{P}\{S^u \le v\}} - e^{-\mathbb{E}\{N_u\}}}{1 - e^{-\mathbb{E}\{N_u\}}},
\end{aligned}$$

where $S^u$ is the size of a generic component of $A_u$ and the PC heuristic has been applied at the penultimate line.

This gives an approximation to the distribution of the size of the largest component, as required, in terms of quantities which are themselves estimable.

There has been a lot of approximation going on here, so a natural question to ask is whether or not what comes out in the end is reliable. We would argue that since these approximations have been heavily used in a number of areas over the last 30 years or so as if they were exact results, they have been proven, by democratic vote, to be as reliable as one can hope for in practice. After all, this is what is important. Nevertheless, it is always useful to see some numbers as well.

### 6.6.3 Numerical Experiments

(If there is going to be a chapter on simulations, this section should move there.)

As part of her thesis work, Jin Cao carried out a variety of simulations to check the accuracy of the approximations of the preceeding sections, for both Gaussian and non-Gaussian random fields on $\mathbb{R}^2$. These were reported on in [26] and we bring them here with minor only editorial changes[15].

The random variables studied were the number of connected components $N$ of an excursion set, the sizes $S$ of the components and $S_{\max}$, the size of the largest component, where we have dropped the "$u$" from these variables indicating the level involved. Simulations of $\chi^2$, $T$ and $F$ fields were obtained, in principle, by first generating the component Gaussian fields $\epsilon$ as moving averages of the form

$$\epsilon_j(t) = \int_{\mathbb{R}^2} e^{-|t-x|^2/2\sigma^2}\, W(dx), \qquad (6.6.6)$$

where $W$ is Gaussian white noise and $\sigma = 3.9$, and then combining these in the obvious way to obtain the non-Gaussian fields.

[15] That is, this subsection represents plagiarism in its purest form.

Of course, (6.6.6) is an idealization of what was actually done on $256 \times 256$ lattice, and repeated to give 5,000 realizations of each random field. For each realization an excursion set was obtained by thresholding the field at a prescribed level, and then recording the values of $N$, $S$, and $S_{\max}$. From these the empirical probabilities $\mathbb{P}\{N = n\}$, $\mathbb{P}\{S = s\}$ and $\mathbb{P}\{S_{\max} > s\}$ were computed, and then compared to the (asymptotic) theory of the previous sections. The results are summarised in the following four figures.

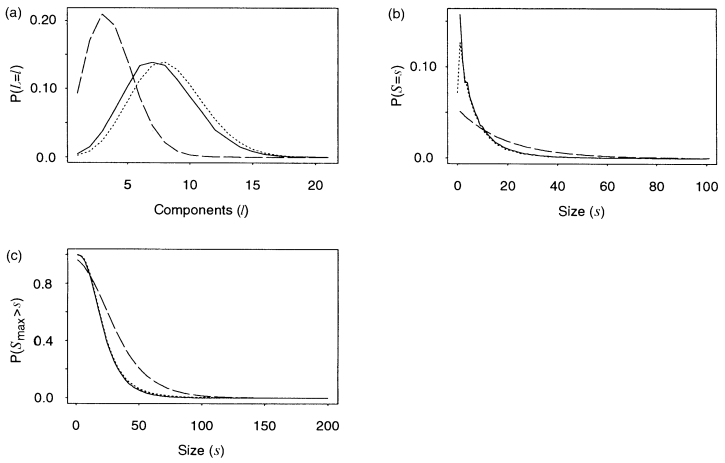


**Fig. 6.6.1.** Simulation of the size of connected components of the excursion set above threshold $u = 4.75$ for a Student's $T$ field with 7 degrees of freedom. (a) number of components $N$; (b) size of one component $S$, (c) size of the largest component $S_{\max}$. — empirical; $\cdots$ theoretical, - - -, Gaussianized.

Figure 6.6.1 shows the simulation result for a $T$ field with $m = 7$ and threshold $u = 4.75$. Figures 6.6.2 and 6.6.3 are for a $\chi^2_7$ field and thresholds $u = 22.04$ and $u = 0.7945$, respectively, and Figure 6.6.4 is for an $F$ field with $n = 4$ and $m = 5$ degrees of freedom, and threshold $u = 21.05$. The thresholds were chosen so that the tail probabilities of the corresponding distributions are 0.001 for the $T$ field and 0.025 for the rest. In all the figures, the solid lines are empirical estimates and the dotted lines are theoretical results.

There is an extra graph in Figure 6.6.1, respresenting what happens if one adopts the trivial Gaussianization of (1.4.8), viz. replacing the $T$ field $f$ by

$$\Phi^{-1}\left(F(f(t))\right),$$

where $F$ is the distribution function of a $T$ random variable with 7 degrees of freedom. As we warned you back in Chapter 1, while this transformation generates a random field with one-dimensional marginal distributions which are standard normal, the random field generated this way is generally quite far

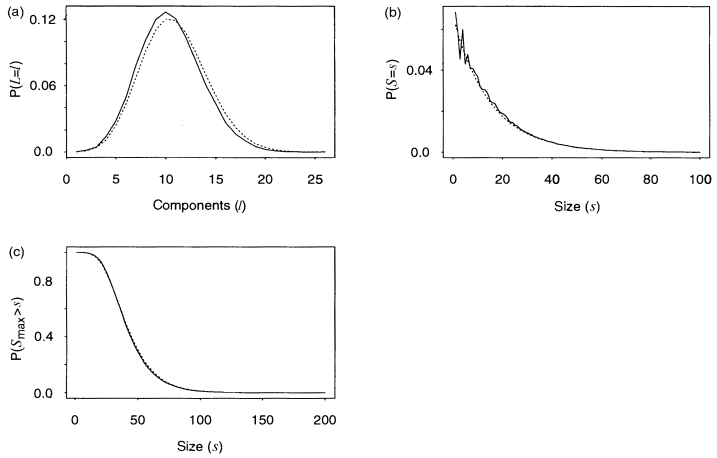


**Fig. 6.6.2.** Simulation of the size of connected components of the excursion set above the threshold $u = 22.04$ for a $\chi^2_7$ field. (a) number of components $N$; (b) size of one component $S$; (c) size of the largest component $S_{\max}$. — empirical; $\cdots$ theoretical.

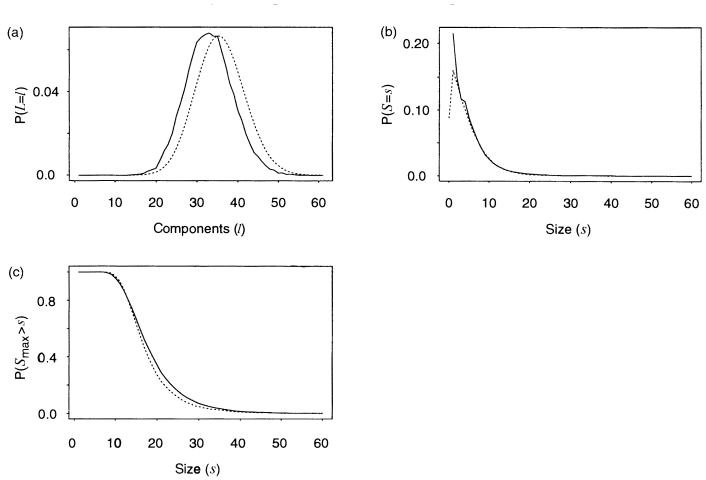


**Fig. 6.6.3.** Simulation of the size of connected components of the excursion set below the threshold $u = 0.7945$ for a $\chi^2_7$ field. (a) number of components $N$; (b) size of one component $S$; (c) size of the largest component $S_{\max}$. — empirical; $\cdots$ theoretical.

from Gaussian. The dashed line in Figure 6.6.1 comes from this procedure, and it is clear that it gives a result both qualitatively and quantitatively inconsistent with the data. However, this was to be expected given the warning you were given in Section 1.4.2 not to adopt such a procedure.

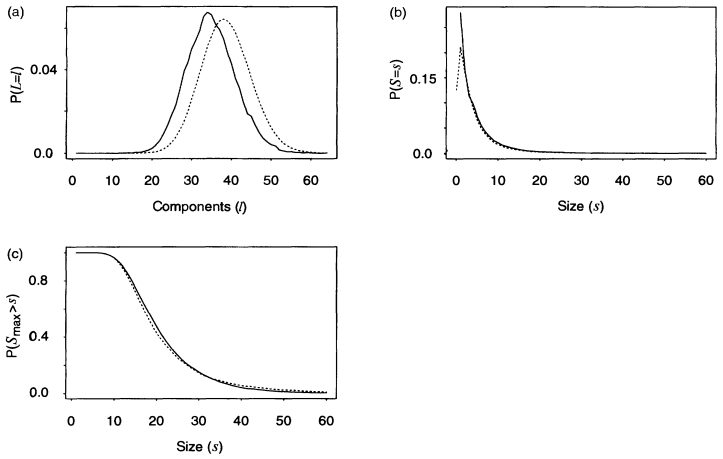


**Fig. 6.6.4.** Simulation of the size of connected components of the excursion set above threshold $u = 21.05$ for an $F$ field with $n = 4$ and $m = 5$ degrees of freedom. (a) number of components $N$; (b) size of one component $S$; (c) size of the largest component $S_{\max}$. — empirical; $\cdots$ theoretical.

As far as the other graphs are concerned, they give a level of accuracy far better than the rough and ready approximations of this chapter seem to deserve.

## 6.7 Exercises

(This section obviously needs work.)

**Exercise 6.7.1.** Something about how non-Palm conditioning would leave $f$ Gaussian. Perhaps in one dimension. Is this written up in LL&R? If not, should be easy.

See the next two exercises.

**Exercise 6.7.2.** Write out the Slepian model for maxima in one dimension and compare this to the vertical window conditioning.

**Exercise 6.7.3.** Write out the Slepian model for level crossings in one dimension and compare this to the vertical window conditioning.

**Exercise 6.7.4.** Prove (6.3.14) and also the asymptotic Slepian model (**??**) Refer [56].

**Exercise 6.7.5.** Prove (6.2.13).

**Exercise 6.7.6.** Show that the density $\psi_u(f'')$ of (6.2.4) gives the distribution of the elements of $\nabla^2 f$, conditioned, in the Palm sense, on $f$ having a local maximum of height $u$.

**Exercise 6.7.7.** Prove (6.4.9) and (6.4.9).
Hint. Note first that there are (at least) two ways to prove the results. Consider first (6.4.9), and start by showing that

$$\mathbb{P}\left\{f(0) - u \geq v \,\middle|\, \text{maximum of height} \geq u \text{ at } 0\right\} = \frac{\mathbb{E}\left\{M_{u+v}\right\}}{\mathbb{E}\left\{M_u\right\}}, \qquad (6.7.7)$$

where $M_u$ is the number of local maxima of $f$ above the level $u$ in some region.

The problem is that even in the Gaussian case we did not know how to explicitly compute $\mathbb{E}\{M_u\}$, and so it should be no surprise that the same is true in the $\chi^2_k$ case. Thus, two routes are open to you. One is to obtain an asymptotic formula for this expectation, which can then be substituted into (6.7.7) and the limit computed. The other is to behave like a true believer in the Euler characteristic heuristic, and simply assume (correctly) that one can replace $\mathbb{E}\{M_u\}$ by $\mathbb{E}\{\varphi(A_u)\}$ in (6.7.7), using the results of Chapter 4, and especially Theorem 4.5.1, to compute the Euler characteristic expectations. The advantage of this approach is that, once you join the community of believers, you can use the same argument for other random fields as well.

The proof of (6.4.9) follows similar lines, but you will have to think a little about how to use the Euler characteristic heuristic in this case.

**Exercise 6.7.8.** An example to show how to compute EEC from the representations of the distributions of $f$, $\nabla f$ and $\nabla^2 f$, as in Keith's original proofs. Also add a reference in EC chapter to this exercise, and thereby maybe kill the "additional $F$" section.
??????????????????????

**Exercise 6.7.9.** Compute the EEC for the $chi^2$ case using the representations (6.4.1)–(6.4.3) for $f(0)$, $\nabla f(0)$ and $\nabla^2 f(0)$ and the equations (which ones???) of Section 4.2

**Exercise 6.7.10.** Prove (6.5.7)–(6.5.9), results originally due to Jin Cao [25, 26].
Warning: The algebra gets harder as you go from example to example!

**Exercise 6.7.11.** Maybe another example? Prove the result, due originally to [66, 67], that (this is in 2-d only as stated)

$$\begin{aligned}\lim_{u\to\infty} \mathbb{P}\left\{\gamma u^2 S_u > v \,\middle|\, S_u > 0\right\} &= \lim_{u\to\infty} \mathbb{P}\left\{\left(2\gamma u^3 V_u\right)^{1/2} > v \,\middle|\, V_u > 0\right\} \\ &= \lim_{u\to\infty} \mathbb{P}\left\{uO_u > v \,\middle|\, O_u > 0\right\} \\ &= e^{-v}, \end{aligned} \qquad (6.7.8)$$

where $\gamma = (\det(\Lambda))^{1/2}/2\pi$.

# Part III

# Numerics

# 7

# Simulating Random Fields

# 8

# Discrete approximation

## 8.1 Estimating Intrinsic Volumes

Four methods at least

1. By estimating the covariance function
2. Worsley-Taylor JASA method
3. Sam-Kevin method
4. Methods from integral geometry, eg via Crofton and tube formulae. (Isotropic case only?)

## 8.2 Data on a Lattice/Triangulated Surface

## 8.3 Computing the Euler Characteristic

## 8.4 Continuity Correction for Supremum Distribution

## 8.5 Improved Bonferroni Inequality

# Part IV

# Applications

# 9

# Neuroimaging

## 9.1 Functional Data

### 9.1.1 Functional Magnetic Resonance Imaging (fMRI)

### 9.1.2 Positron Emission Tomography (PET)

## 9.2 Anatomical data

### 9.2.1 Gray Matter Density

### 9.2.2 Deformation Fields

### 9.2.3 Cortical Thickness

# 10

# Astrophysics

## 10.1 Approximations

## 10.2 COBE Data

For the moment, we note that typical papers in the area are [93, 95], and a popular account of the subject can be found in the book [86] by George Smoot, who, together with John Mather, received the 2006 Nobel Prize in physics for their discovery of the black body form and anisotropy of the cosmic microwave background radiation.

## 10.3 Galaxy Density

## 10.4 Discrete Models?

# 11

# Oceanography

## 11.1 The Problem

## 11.2 Speckles

Longuet-Higgins [59, 60] and others

## 11.3 Space/time Modelling

## 11.4 WAFO

In a random wave model, like that for Gaussian or transformed Gaussian waves, the distribution of wave characteristics such as wave period and crest-trough wave height can be calculated by high accuracy for almost any spectral type. WAFO is a third- generation package of Matlab routines for handling statistical modelling, calculation and analysis of random waves and wave characteristics and their statistical distributions. The package also contains routines for cycle counting and computation in ran- dom load models, in particular the rain ow counting often used in fatigue life prediction. Random wave distributions are notoriously di cult to ob- tain in explicit form from a random wave model, but numeri- cal algorithms, based on the so-called regression approximation, work well. This method to calculate wave distributions is the only known method that gives correct answers valid for general spectra. The theoretical background is reviewed in Lindgren and Rychlik (1991)LIND-RYCH and computational aspects and algorithms in Rychlik and Lindgren (1993). [43, 58] The algorithms are based on a speci cation of the random waves by means of their (uni-directional or directional) spec- trum, and on the underlying assumption of linear wave theory and Gaussian distribution. However, a transformation of sea el- evation data can be made to obtain a desired (horizontal) asym- metric marginal distribution.

(Wave Analysis for Fatigue and Oceanography [24]

WAFO is a toolbox of Matlab routines for statistical analysis and simulation of random waves and random loads. WAFO is freely redistributable software,

# 12

# Miscellaneous

## 12.1 Random Polynomials

## 12.2 Eigenvalues of random matrices

Start with Jonathan's and Kuriki/Takemura's papers.

## 12.3 Chemical Reactions

## 12.4 Knots in Random Fields

## 12.5 Quantum Chaos

## 12.6 Condition numbers of random matrices

# Notation Index

# Subject Index

# References


1. R.J. Adler. Excursions above a fixed level by $n$-dimensional random fields. J. Applied Probability, 13:276–289, 1976.
2. R.J. Adler. On generalising the notion of upcrossings to random fields. Adv. Applied Probability, 8:789–905, 1976.
3. R.J. Adler. The Geometry of Random Fields. John Wiley & Sons Ltd., Chichester, 1981. Reprinted in 2010 by SIAM, Classics of Applied Mathematics Series.
4. R.J. Adler and A.M. Hasofer. Level crossings for random fields. Ann. Probab., 4:1–12, 1976.
5. R.J. Adler and J.E. Taylor. Random Fields and Geometry. Springer, 2007.
6. D. Aldous. Probability Approximations via the Poisson Clumping Heuristic, volume 77 of Applied Mathematical Sciences. Springer-Verlag, New York, 1989.
7. E. Anderes and S. Sourav Chatterjee. Consistent estimates of deformed isotropic Gaussian random fields on the plane. 2007. arxiv:0710.0379.
8. E. Anderes and M.L. Stein. Estimating deformations of isotropic gaussian random fields on the plane. Ann. Statist., 2008. In press.
9. M. Aronowich and R.J. Adler. Behaviour of $\chi^2$ processes at extrema. Adv. in Appl. Probab., 17(2):280–297, 1985.
10. M. Aronowich and R.J. Adler. Extrema and level crossings of $\chi^2$ processes. Adv. in Appl. Probab., 18(4):901–920, 1986.
11. M. Aronowich and R.J. Adler. Sample path behaviour of $\chi^2$ surfaces at extrema. Adv. in Appl. Probab., 20(4):719–738, 1988.
12. J-M. Azaïs and M. Wschebor. A general expression for the distribution of the maximum of a Gaussian field and the approximation of the tail. Stoch. Proc. Appl, 118:1190–1218, 2008.
13. Jean-Marc Azaïs and Mario Wschebor. Level sets and extrema of random processes and fields. John Wiley & Sons Inc., Hoboken, NJ, 2009.
14. P. Baldi, D. Marinucci, and V.S. Varadarajan. On the characterization of isotropic Gaussian fields on homogeneous spaces of compact groups. Elec. J. Probab., 12:291–302, 2007.
15. J.M. Bardeen, J.R. Bond, N. Kaiser, and A.S. Szalay. The statistics of peaks of Gaussian random fields. The Astrophysical Journal, 304:15–61, 1986.
16. 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.

17. Yu.K. Belyaev, V.P. Nosko, and A.A. Sveˇ snikov. Probabilistic characteristics of excursions of a two-dimensional random field. Theor. Probability Math. Stat., 6:23–30, 1975.
18. Yu.K. Belyaev and V.I. Piterbarg. Asympotics of the average number of $a$-points of overshoot of a Gaussian field beyond a high level. *Dokl.* Akad. Nauk, SSSR, 203:309–313, 1972.
19. Yu.K. Belyaev and V.I. Piterbarg. The asymptotic formula for the mean number of $a$-points of excursions of Gaussian fields above high levels (in russian). In Bursts of Random Fields, pages 62–89. Moscow Univ. Press, Moscow, 1972.
20. A. Bernig and L. Bröcker. Lipschitz-Killing invariants. Math. Nachr., 245:5–25, 2002.
21. S. Bochner. Monotone Funktionen Stieltjessche Integrale and Harmonishe Analyse. Math. Ann., 108:378, 1933.
22. C. Borell. The Brunn-Minkowski inequality in Gauss space. Invent. Math., 30:205–216, 1975.
23. P.J Brockwell and R.A. Davis. Time Series: Theory and Methods. Springer-Verlag, New York, 1991.
24. P.A. Brodtkorb, P. Johannesson, G. Lindgren, I. Rychlik, J. Ryden, and E. Sjo. Wafo, a Matlab toolbox for analysis of random waves and loads. Proceedings of the 10th International Offshore and Polar Engineering Conference, III(4):343–350, 2000. Software available at http://www.maths.lth.se/matstat/wafo/index.html.
25. J. Cao. The excursion set of random fields with applications to human brain mapping. PhD thesis, McGill University, 1997.
26. J. Cao. The size of the connected components of excursion sets of $\chi^2$, $t$ and $F$ fields. Adv. in Appl. Probab., 31(3):579–595, 1999.
27. J. Cao and K. Worsley. The geometry of correlation fields with an application to functional connectivity of the brain. Ann. Appl. Probab., 9(4):1021–1057, 1999.
28. S.S. Chern. On the kinematic formula in integral geometry. J. Math. Mech., 16:101–118, 1966.
29. H.M.P. Couchman. The statistics of peaks in three-dimensional Gaussian noise: some simple results. Mon. Not. R. Ast. Soc., 225:777–794, 1987.
30. R.M. Dudley. The sizes of compact subsets of hilbert space and continuity of Gaussian processes. J. Funct. Anal., 1:290–330, 1967.
31. R.M. Dudley. Sample functions of the Gaussian process. Ann. Probab., 1:66–103, 1973.
32. A. C. Evans, S. Marret, P. Neelin, L. Collins, K.J. Worsley, W. Dai, S. Milot, E. Meyer, , and D. Bub. Anatomical mapping of functional activation in stereotactic coordinate space. NeuroImage, 1:43–53, 1992.
33. H. Federer. Geometric Measure Theory. Springer-Verlag, New York, 1969.
34. M.G. Genton and O. Perrin. On a time deformation reducing nonstationary stochastic processes to local stationarity. J. Appl. Probab., 41(1):236–249, 2004.
35. A. Gray. Tubes. Addison-Wesley, Redwood City, 1990.
36. U. Grenander. Stochastic processes and statistical inference. Ark. Mat., 1:195–277, 1950.
37. H. Hadwiger. Vorles¨ungen ¨Uber Inhalt, Oberfl¨ache und Isoperimetrie. Springer-Verlag, Berlin, 1957.

38. H. Hadwiger. Normale Körper im euklidischen Raum und ihre topologischen und metrischen Eigenschaften. Math. Zeitschrift, 71:124–140, 1959.
39. P. Hall. Introduction to the theory of coverage processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1988.
40. E.J. Hannan. Multiple Time Series. Wiley, New York, 1970.
41. M.L. Hogan and D.O. Siegmund. Large deviations for the maxima of some random fields. Adv. in Appl. Math., 7(1):2–22, 1986.
42. H. Hotelling. Tubes and spheres in $n$-spaces and a class of statistical problems. Amer. J. Math., 61:440–460, 1939.
43. Rychlik I. and Lindgren G. Crossreg, A technique for firrst passage and wave density analysis. Probability in the Engineering and Informational Sciences, 7:125–148, 1993.
44. Søren Johansen and Iain M. Johnstone. Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. Ann. Statist., 18(2):652–684, 1990.
45. M. Kac. On the average number of real roots of a random algebraic equation. Bull. Amer. Math. Soc., 43:314–320, 1943.
46. M. Kac and D. Slepian. Large excursions of Gaussian processes. Ann. Math. Statist., 30:1215–1228, 1959.
47. O. Kallenberg. Random measures. Akademie-Verlag, Berlin, fourth edition, 1986.
48. O. Kallenberg. Foundations of modern probability. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
49. D.A. Klain and G-C. Rota. Introduction to Geometric Probability. Cambridge University Press, Cambridge, 1997.
50. M. Knowles and D.O. Siegmund. On Hotelling's approach to testing for a nonlinear parameter in a regression. Int. Statist. Rev, 57:205–220, 1989.
51. M.R. Leadbetter, G. Lindgren, and H. Rootzén. Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York, 1983.
52. R. Leblanc, E. Meyer, D. Bub, R. J. Zatorre, and A. C. Evans. Language localization with activation positron emission tomography scanning. Neurosurgery, 31(2):369–372, 1992.
53. M. Ledoux and M. Talagrand. Probability in Banach Spaces. Isoperimetry and Processes. Springer-Verlag, Berlin, 1991.
54. N. Leonenko. Limit Theorems for Random Fields with Singular Spectrum, volume 465 of Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 1999.
55. G. Letac. Problèmes classiques de probabilité sur un couple de gelfand. Lecture Notes in Math., 861:93–120, 1981.
56. G. Lindgren. Local maxima of Gaussian fields. Ark. Mat., 10:195–218, 1972.
57. G. Lindgren. Extreme values and crossings for the $\chi^2$-process and other functions of multidimensional Gaussian processes, with reliability applications. Adv. in Appl. Probab., 12(3):746–774, 1980.
58. G. Lindgren and I. Rychlik. Slepian models and regression approximations in crossing and extreme value theory. International Statistical Review, 59:195–225, 1991.
59. M.S. Longuet-Higgins. On the statistical distribution of the heights of sea waves. J. Marine Res., 11:245–266, 1952.

60. M.S. Longuet-Higgins. The statistical analysis of a random moving surface. Phil. Trans. Roy. Soc., A249:321–387, 1957.
61. G. Maruyama. The harmonic analysis of stationary stochastic processes. Mem. Fac. Sci. Kyūsyū Univ. A., 4:45–106, 1949.
62. 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.
63. C. Mercadier. Numerical bounds for the distributions of the maxima of some one- and two-parameter Gaussian processes. Adv. in Appl. Probab., 38(1):149–170, 2006.
64. D.Q. Naiman and H.P. Wynn. Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. Ann. Statist., 20(1):43–76, 1992.
65. D.Q. Naiman and H.P. Wynn. Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. Ann. Statist., 25(5):1954–1983, 1997.
66. V.P. Nosko. Local structure of Gaussian random fields in the neighborhood of high-level shines. Dokl. Akad. Nauk SSSR, 189:714–717, 1969.
67. V.P. Nosko. Bursts of Gaussian stochastic fields. Vestnik Moskov. Univ. Ser. I Mat. Meh., 25(4):18–22, 1970.
68. O. Perrin and W. Meiring. Identifiability for non-stationary spatial structure. J. Appl. Probab., 36(4):1244–1250, 1999.
69. O. Perrin and R. Senoussi. Reducing non-stationary stochastic processes to stationarity by a time deformation. Statist. Probab. Lett., 43(4):393–397, 1999.
70. O. Perrin and R. Senoussi. Reducing non-stationary random fields to stationarity and isotropy using a space deformation. Statist. Probab. Lett., 48(1):23–32, 2000.
71. J. Pickands, III. Maxima of stationary Gaussian processes. Z. Wahrsch. Verw. Gebiete, 7:190–223, 1967.
72. J. Pickands, III. Asymptotic properties of the maximum in a stationary Gaussian process. Trans. Amer. Math. Soc., 145:75–86, 1969.
73. J. Pickands, III. Upcrossing probabilities for stationary Gaussian processes. Trans. Amer. Math. Soc., 145:51–73, 1969.
74. V.I. Piterbarg. The asymptotic Poisson character of the number of high excursions and the distribution of the maximum of homogeneous Gaussian fields. In Bursts of Random Fields, pages 90–118. Moscow Univ. Press, Moscow, 1972. In Russian.
75. V.I. Piterbarg. Asymptotic methods in the theory of Gaussian processes and fields. American Mathematical Society, Providence, RI, 1996. Translated from the Russian by V. V. Piterbarg, Revised by the author.
76. D. Rabinowitz and D.O. Siegmund. The approximate distribution of the maximum of a smoothed Poisson random field. Statist. Sinica, 7:167–180, 1997.
77. S. O. Rice. The Distribution of the Maxima of a Random Curve. Amer. J. Math., 61(2):409–416, 1939.
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. S.N. Roy. On a heuristic method of test construction and its use in multivariate analysis. Ann. Math. Statistics, 24:220–238, 1953.

80. P.D. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. J. Amer. Stat. Assoc., 87:108–119, 1992.
81. L.A. Santalo. Integral Geometry and Geometric Probability. Encyclopedia of Mathematics and its Applications. Addison-Wesley, Reading, 1976.
82. R. Schneider. Convex Bodies: the Brunn-Minkowski Theory, volume 44 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1993.
83. I.J. Schoenberg. Metric spaces and completely monotone functions. Ann. of Math., 39:811–841, 1938.
84. 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.
85. D. Slepian. On the zeros of Gaussian noise. In Proc. Sympos. Time Series Analysis (Brown Univ., 1962), pages 104–115. Wiley, New York, 1963.
86. G. Smoot and K. Davidson. Wrinkles in Time. William Morrow, New York, 1993.
87. Jiayang Sun. Tail probabilities of the maxima of Gaussian random fields. Ann. Probab., 21(1):34–71, 1993.
88. A. Takemura and S. Kuriki. Weights of $\overline{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives. Ann. Statist., 25(6):2368–2387, 1997.
89. 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.
90. M. Talagrand. The Generic Chaining: Upper and Lower Bounds of Stochastic Processes. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.
91. J.E. Taylor. A Gaussian kinematic formula. Ann. Probab., 34(1):122–158, 2006.
92. J.E. Taylor and R.J. Adler. Poincaré's limit, kinematic formulae and Gaussian processes. Ann. Probab., 37(3):1459–1482, 2009.
93. S. Torres. Topological analysis of COBE-DMR cosmic microwave background maps. Astrophysical Journal, 423:L9–L12, 1994.
94. 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.
95. M.S. Vogeley, C. Park, M.J. Geller, J.P. Huchin, and J.R. Gott. Topological analysis of the CfA redshift survey. Astrophysical Journal, 420:525–544, 1994.
96. H. Weyl. On the volume of tubes. Amer. J. Math., 61:461–472, 1939.
97. R. J. Wilson and R.J. Adler. The structure of Gaussian fields near a level crossing. Adv. in Appl. Probab., 14(3):543–565, 1982.
98. E. Wong. Stochastic Processes in Information and Dynamical Systems. McGraw-Hill, New York, 1971.
99. K.J. Worsley. Local maxima and the expected Euler characteristic of excursion sets of $\chi^2$, $f$ and $t$ fields. Adv. in Appl. Probab., 26:13–42, 1994.
100. K.J. Worsley. Testing for signals with unknown location and scale in a $\chi^2$ random field, with an application to fMRI. Adv. in Appl. Probab., 33(4):773–793, 2001.
101. M. Wschebor. Surfaces aléatoires, volume 1147 of Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1985. Mesure géométrique des ensembles de niveau. [The geometric measure of level sets].

102. A.M. Yaglom. Some classes of random fields in $n$-dimensional space, related to stationary random processes. Theor. Probability Appl., 2:273–320, 1957.
103. A.M. Yaglom. Second-order homogeneous random fields, Univ. of California Press, Berkey. Proc. Fourth Berkeley Symp. Math. Statist. Prob., 2:593–622, 1961.
104. K. Ylinen. Random fields on noncommutative locally compact groups. Lecture Notes in Math., 1210:365–386, 1986.
105. M. Zähle. Approximation and characterization of generalized Lipshitz-Killing curvatures. Ann. Global Anal. Geom., 8:249–260, 1990.