# Excursion sets of random fields with applications to human brain mapping

Jin Cao

Department of Mathematics and Statistics

McGill University, Montreal

1997

A Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

**Excursion sets of random fields with applications to human brain mapping**

by

Jin Cao

Submitted to the Department of Mathematics and Statistics
on Mar. 7, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

The statistical analysis of brain functional and structural change presents a formidable statistical problem. Computation of a test statistic indicating change at each location of the brain gives a 3-D statistical image. The global maximum of the statistical image has been proposed by Friston et al. (1991) and Worsley et al. (1992) to test for regions of high change based on the belief that the changes are localized, not distributed. To find the correct $p$ value, Worsley et al. (1992) modelled the statistical image as a stationary random field under the null hypothesis. The approximate $p$ value of the global maximum was then determined by the expected Euler characteristic (EC) of the excursion set, that is, the set of points with a value larger than a given threshold. The expected EC of excursion sets of stationary Gaussian random fields has been derived by Adler and Hasofer (1976) and Adler (1981) and extended to to $\chi^2$, $t$ and $F$ fields by Worsley (1994). The first part of the thesis aims to extend their results to Hotelling's $T^2$ fields and apply the theory to a real example from 3-D brain deformations.

Although the global maximum as a test statistic is more widely used, some workers have suggested an alternative: the size of the largest connected component of the excursion set, based on the belief that neural activity tends to stimulate changes over contiguous locations. Nosko (1969) derived the asymptotic distribution of the size of one connected component of the excursion set above high levels for a stationary Gaussian field. Applying this work, Friston et al. (1994) presented an approximate analysis for assessing the significance of focal changes by their spatial extent. Although the analysis was developed for Gaussian fields originally, it was suggested that it could also be applied to some non-Gaussian fields, like the $t$ field, by appropriate "Gaussianization", that is, pointwise transformation to a Gaussian distribution. However, the Gaussianzied field is not a Gaussian random field and hence the approach is not strictly valid. The second part of the thesis aims to derive the asymptotic distribution of the size of one connected component of the excursion set for $\chi^2$, $t$ and $F$ fields and apply this to an experiment on pain perception.

Thesis Supervisor: K.J. Worsley
Title: Professor

# Excursion sets of random fields with applications to human brain mapping

by

Jin Cao

Submitted to the Department of Mathematics and Statistics
on Mar. 7, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

L'étude statistique des changements fonctionnels et structuraux du cerveau humain est un bel example du problème de comparaisons multiples. Le calcul d'une statistique indiquant l'activité neuronale en chaque partie du cerveau donne une image statistique en 3-D. Le maximum global des images statistiques a été proposé, par Friston & al. (1991) et Worsley & al. (1992) pour tester les régions sujettes à des changements importants selon la croyance que ces changements sont locaux. Afin de calculer le seuil réel, Worsley & al. (1992) ont modélisé ces images statistiques comme la réalisation d'un champ aléatoire stationnaire sous l'hypothèse nulle. Le seuil approché du maximal global est alors déterminé par l'espérance de la caractéristique d'Euler (EC) de l'ensemble d'excursion, c'est-a-dire l'ensemble des points dont la mesure est plus grande qu'un niveau de tolérance préalablement fixé. Dans le cadre du champ gaussien, cette espérance EC a été fournie par Adler & Hasofer (1976) et Adler (1981) et étendue aux champs $\chi^2, t$ et $F$ par Worsley (1994). La première partie de cette thèse vise à étendre ces résultats aux champs $T^2$ de Hotelling, puis à les appliquer à un example concrèt de malformations cérébrales.

Bien que le maximum global comme test statistique soit d'un usage plus courant, certains auteurs ont proposé en échange la taille de la plus grande composante connexe de l'ensemble d'excursion. Cette proposition repose sur la croyance selon laquelle l'activité neuronale tend à favoriser des échanges entre parties contigues. Nosko (1969) a obtenu la distribution asymptotique de la taille d'une composante connexe de l'ensemble d'excursion. Appliquant ces travaux aux champs gaussiens, Friston & al. (1994) ont présenté une analyse approximative pour évaluer l'importance des activités par leur répartion spaciale. Malgré son cadre restrictif, les auteurs ont suggéré de l'étendre aux champs non gaussiens tel le champ t, par "Gaussianisation" appropriée c'est-à-dire en la convertissant point par point en une distribution gaussienne. Cependant le champ gaussianisé n'étant pas gaussien, cette approche n'est généralement pas bonne. Le but de la seconde partie de cette thèse est de déduire la distribution asymptotique de la taille d'une composante connexe de l'ensemble d'excursion pour les champs $\chi^2, t$

et $F$ et appliquer cela à l'étude de la perception de la douleur.

Thesis Supervisor: K.J. Worsley
Title: Professor

# Acknowledgements

I would like to thank my supervisor, Professor Keith J. Worsley, for his friendly advice and continuous encouragement since I joined the department, for providing me with first class resources and for reading the preliminary version of this thesis and giving me useful comments.

I would like to acknowledge the encouragement and assistance of L. Collins, A.C. Evans, J.D. MacDonald, T. Paus of the McConnell Brain Imaging Center, Montreal Neurological Institute and to thank them for permisson to use their data.

I would also like to thank Professor W.J. Anderson, Professor M. Gu and Professor D.B. Wolfson for their wonderful lectures and support through the years.

My thanks are also to the other professors, staff and fellow students who helped me in any way in my work. I am especially greatful to C.E. Mlan who helped me with the French translation of the abstract and to K. Shafie for discussions on the aspects of the work.

On a more practical note, I am indebted to the Department of Mathematics and Statistics at McGill that provided me with a financial support and a comfortable environment for my studies and thesis work.

Finally, for my family and friends, words cannot express what they mean to me. Their love, encouragement and their belief in me helped me keep going during the rough times. I would like to especially thank S. Li, for his understanding and constant support.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Functional and structural human brain mapping

One of the many exciting areas of current medical research is the study of the living human brain. Modern scientific medicine is armed with two powerful research tools for such studies, Positron Emission Tomography and Magnetic Resonance Imaging.

Positron Emission Tomography (PET) is an imaging technique that is used to produce a 3-D image of brain activity such as *regional cerebral blood flow* (rCBF). Magnetic Resonance Imaging (MRI) is used to produce an image of brain structure. An illustration of this is shown in Figure B-1, where the left panel shows a horizontal slice through the center of a color coded 3-D PET image of rCBF and the right panel shows the same slice through an MRI image.

PET has opened up the possibility of studying the function of the brain. In simple activation studies, a baseline stimulus is first applied to the subject so that the subject is in a *rest* condition. Then, a stimulus is applied so that the subject is in an *active* condition. Two scans are acquired under the rest and active conditions respectively, and we try to locate the brain region that responds to the stimulus, i.e., where there is an increase in brain activity. By subtracting the stimulus images from the rest image, an image is obtained which shows the difference in brain activity caused by the stimulus (see Figure B-2). If there is an evident increase of blood flow in some region, then that part of the brain is thought to be activated by the stimulus.

The differenced image usually has a low signal to noise ratio. To decrease the noise component, the experiment is always repeated on several subjects and kernel smoothing is applied to the differenced image (see Figure B-3).

Suppose there are $p$ subjects in the activation study. For each of the subjects, a 3-D image $Z_i(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$ ($N = 3$ in this case) representing the difference in brain activity caused by the stimulus, was obtained. Our primary interest is to identify brain regions which show a significant increase in brain activity. Usually some prior information of the possible location of the activated regions is available, and so the search for activation is confined to a *search region*, $C \subset \mathbb{R}^N$. To describe the effect of the stimulus on brain functional change at some unknown location in $C$, the following simple model can be built at each location $\mathbf{t}$:

$$Z_i(\mathbf{t}) = \mu(\mathbf{t}) + e_i(\mathbf{t}), \tag{1.1}$$

where $\mu(\mathbf{t})$ is the mean functional change at $\mathbf{t}$, and $e_i(\mathbf{t})$ is the error for subject $i$ at point $\mathbf{t}$, $\mathrm{E}(e_i(\mathbf{t})) = 0$. Hence, the following hypothesis test will be appropriate to serve the purpose:

$$\begin{aligned} &\mathrm{H}_0 : \mu(\mathbf{t}) = 0 \text{ for all } \mathbf{t} \in C \quad \text{(no functional change)} \\ &\mathrm{H}_1 : \mu(\mathbf{t}) > 0 \text{ (or } \mu(\mathbf{t}) \neq 0) \text{ for some } \mathbf{t} \in C \quad \text{(functional change).} \end{aligned} \tag{1.2}$$

The same kind of hypothesis testing problem arises in structural change. To align individual brain images, a local displacement is applied at each location of the individual brains so as to bring the local structure of the individual brain into alignment with a template brain (see Figure B-4). This produces a deformation field whose value at each location of the brain is a 3-D displacement vector (see Figure B-5 [1]). This opens up the possibility of studying brain structural differences, since

[1]The deformation field in the figure is not what is applied to Figure B-4

changes in brain structure will be reflected in changes in the deformations required to align that structure with the template.

Suppose the subjects are divided into two groups and we are interested in detecting any abnormal 3-D deformations between the groups. Then, a simple *multivariate* model can be built at each location of the brain, $\mathbf{t}$. Let $\boldsymbol{\Delta}_{ij}(\mathbf{t})$ be the 3-D displacement vector that was applied at location $\mathbf{t}$ for subject $j = 1, \ldots, p_i$ in group $i = 1, 2$ to match the template brain, then

$$\boldsymbol{\Delta}_{ij}(\mathbf{t}) = \mu_i(\mathbf{t}) + \mathbf{e}_{ij}(\mathbf{t}), \quad j = 1, \ldots, p_i; i = 1, 2, \tag{1.3}$$

where $\mu_i(\mathbf{t})$ is now the mean 3-D deformation applied at point $\mathbf{t}$ for the subjects in group $i$, and $\mathbf{e}_{ij}(\mathbf{t})$ is now the individual error vector. Again, to search for significant brain structural differences in region $C$, we are interested in testing the hypotheses:

$$\begin{aligned} &\mathrm{H}_0 : \mu_1(\mathbf{t}) - \mu_2(\mathbf{t}) = 0 \text{ for all } \mathbf{t} \in C \quad \text{(no structural change)} \\ &\mathrm{H}_1 : \mu_1(\mathbf{t}) - \mu_2(\mathbf{t}) \neq 0 \text{ for some } \mathbf{t} \in C \quad \text{(structural change).} \end{aligned} \tag{1.4}$$

## 1.2 Hypothesis testing

### 1.2.1 From hypothesis testing to statistical images

If our attention is focused at one location of the brain, $\mathbf{t}$, in the search region $C$, then a test statistic can easily be constructed to test for either brain functional change (1.2) or structural change between groups (1.4). For functional change, the sample mean $\bar{Z}(\mathbf{t})$ and sample variance $S^2(\mathbf{t})$ of the differenced images are:

$$\bar{Z}(\mathbf{t}) = \sum_{i=1}^{p} Z_i(\mathbf{t})/p, \qquad S^2(\mathbf{t}) = \sum_{i=1}^{p}(Z_i(\mathbf{t}) - \bar{Z}(\mathbf{t}))^2/(p-1), \tag{2.5}$$

and the statistic at point $\mathbf{t}$

$$T(\mathbf{t}) = p^{\frac{1}{2}} \bar{Z}(\mathbf{t})/S(\mathbf{t}) \tag{2.6}$$

can be used to test the null hypothesis of no change, $\mu(\mathbf{t}) = 0$. Let $\mathrm{N}(\mu, \sigma^2)$ represent the normal distribution with mean $\mu$ and variance $\sigma^2$ and let $t_\nu$ represent the Student-$t$ distribution with $\nu$ degrees of freedom. If $e_1(\mathbf{t}), \ldots, e_p(\mathbf{t}) \sim \mathrm{N}(0, \sigma^2(\mathbf{t}))$ independently, then $T(\mathbf{t}) \sim t_{p-1}$ under the null hypothesis. Furthermore, if the underlying noise is assumed to have stationary variance, that is $\sigma^2(\mathbf{t})$ is constant over all $\mathbf{t}$, then one may estimate the variance by pooling the sample variance $S^2(\mathbf{t})$ over all the locations in region $C$ (Worsley et al., 1992) to obtain

$$\bar{S}^2 = \int S^2(\mathbf{t}) d\mathbf{t} / \lambda(C),$$

where $\lambda(C)$ is the size of $C$. If $\lambda(C)$ is large enough so that $\bar{S}^2$ is approximately constant, then replacing $S(\mathbf{t})$ by $\bar{S}$ gives a Gaussian statistic:

$$X(\mathbf{t}) = p^{\frac{1}{2}} \bar{Z}(\mathbf{t}) / \bar{S}. \tag{2.7}$$

Under the null hypothesis, $X(\mathbf{t}) \sim \mathrm{N}(0, 1)$ approximately.

Structural change can be tested in a similar way. At each location $\mathbf{t}$, one can estimate the vector of mean deformations for the two groups by:

$$\bar{\mathbf{\Delta}}_1(\mathbf{t}) = \sum_{j=1}^{p_1} \mathbf{\Delta}_{1j}(\mathbf{t}) / p_1, \quad \bar{\mathbf{\Delta}}_2(\mathbf{t}) = \sum_{j=1}^{p_2} \mathbf{\Delta}_{2j}(\mathbf{t}) / p_2,$$

and the variance of the deformations by the $3 \times 3$ matrix:

$$\begin{aligned} \mathbf{V}(\mathbf{t}) = \frac{1}{p_1 + p_2 - 2} \Bigg\{ & \sum_{j=1}^{p_1} (\mathbf{\Delta}_{1j}(\mathbf{t}) - \bar{\mathbf{\Delta}}_1(\mathbf{t}))(\mathbf{\Delta}_{1j}(\mathbf{t}) - \bar{\mathbf{\Delta}}_1(\mathbf{t}))' \\ & + \sum_{j=1}^{p_2} (\mathbf{\Delta}_{2j}(\mathbf{t}) - \bar{\mathbf{\Delta}}_2(\mathbf{t}))(\mathbf{\Delta}_{2j}(\mathbf{t}) - \bar{\mathbf{\Delta}}_2(\mathbf{t}))' \Bigg\}. \end{aligned}$$

The Hotelling's $T^2$ statistic $(p_1 + p_2 - 2) Y(\mathbf{t})$, where

$$Y(\mathbf{t}) = \frac{p_1 p_2}{(p_1 + p_2)(p_1 + p_2 - 2)} [\bar{\mathbf{\Delta}}_1(\mathbf{t}) - \bar{\mathbf{\Delta}}_2(\mathbf{t})]' \mathbf{V}(\mathbf{t})^{-1} [\bar{\mathbf{\Delta}}_1(\mathbf{t}) - \bar{\mathbf{\Delta}}_2(\mathbf{t})], \tag{2.8}$$

can be used to test for a structural difference between the two groups at point $\mathbf{t}$ (Worsley et al., 1996b). Let $\text{Normal}_d(\mu, \mathbf{\Sigma})$ represent the multivariate normal distribution with mean $\mu$ and variance $\mathbf{\Sigma}$ and let $\chi^2_\nu$ represent the $\chi^2$ distribution with $\nu$ degrees of freedom. Then under the assumption of $\mathbf{e}_{ij}(\mathbf{t}) \sim \text{Normal}_3(\mathbf{0}, \mathbf{\Sigma}(\mathbf{t}))$, $(p_1 + p_2 - 4)Y(\mathbf{t})/3$ has an $F$ distribution with $3, p_1 + p_2 - 4$ numerator and denominator degrees of freedom respectively under the null hypothesis. If $\sigma(\mathbf{t})$ is constant, then the variance can be pooled as before, to give

$$\bar{\mathbf{V}} = \int \mathbf{V}(\mathbf{t}) d\mathbf{t} / \lambda(C).$$

Replacing $\mathbf{V}(\mathbf{t})$ by $\bar{\mathbf{V}}$ in (2.8) gives a $\chi^2$ statistic:

$$U(\mathbf{t}) = \frac{p_1 p_2}{(p_1 + p_2)} [\bar{\mathbf{\Delta}}_1(\mathbf{t}) - \bar{\mathbf{\Delta}}_2(\mathbf{t})]' \bar{\mathbf{V}}^{-1} [\bar{\mathbf{\Delta}}_1(\mathbf{t}) - \bar{\mathbf{\Delta}}_2(\mathbf{t})].$$

If $\lambda(C)$ is large enough so that $\bar{\mathbf{V}}$ is approximately constant, then $U(\mathbf{t}) \sim \chi^2_3$.

Computation of a statistic indicating evidence of change at each location gives a 3-D statistical image. It is straightforward to generalize models (1.1) and (1.3) to linear models of the form:

$$Z_i(\mathbf{t}) = \mathbf{x}_i' \beta(\mathbf{t}) + e_i(\mathbf{t}),$$

where $\mathbf{x}_i$ is a $k \times 1$ vector of external regression variables for subject $i$ (note that $\mathbf{x}_i$ does not depend on $\mathbf{t}$) and $\beta(\mathbf{t})$ is $k \times 1$ vector of regression parameters. The test of $\beta(\mathbf{t}) = \mathbf{0}$ or components of $\beta(\mathbf{t})$ equal to zero lead to $F$-statistic images in a straightforward way.

Based on the belief that changes are localised, not distributed, Friston et al. (1991) and Worsley et al. (1992) have used the global maximum of the statistical image to test for regions of changes at an unknown location. However, the test based on the global maximum takes no account of the spatial extent of an activation. An alternative, the size of the largest connected component of the excursion set, or the set of points with a value larger than a given threshold, has been proposed by Friston et al. (1994) based on the belief that areas of true neural activity will tend to stimulate changes over contiguous locations.

The two test statistics address different features of the activation foci. Friston et al. (1994) show that a small but intense signal is best detected by the global maximum, whereas a broad but diffuse signal is best detected by the size of the largest connected component in the excursion set.

To find the correct P-value of the two test statistics, Worsley et al. (1992) have modelled $\sigma(\mathbf{t})^{-1}e_i(\mathbf{t})$ in the model (1.1) as i.i.d. stationary Gaussian random fields with mean 0 and variance 1. Similarly, for model (1.3), it is assumed that $\mathbf{\Sigma}(\mathbf{t})^{-\frac{1}{2}}\mathbf{e}_{ij}(\mathbf{t})$ can be well approximated by a vector of three i.i.d. stationary Gaussian random fields with mean 0 and variance 1. Hence, under the null hypothesis, the statistic images mentioned before can be modelled as a Gaussian, $\chi^2, t, F$ or Hotelling's $T^2$ random field. A more precise definition of the random field will be given in the next section.

It is worth noting that the global maximum is also the likelihood ratio test statistic for testing brain functional change from model (1.1), under the assumption that $e_1(\mathbf{t}), \ldots, e_p(\mathbf{t})$ are i.i.d. stationary Gaussian random fields with mean 0 and variance $\sigma^2$, and for some $a, \tau$,

$$\mu(\mathbf{t}) = aR(\mathbf{t} - \tau),$$

(Siegmund and Worsley, 1995), where $R(\mathbf{t})$ is the covariance function (see (3.9) for the definition). For the multivariate model (1.1), it is straightforward to show that if $\mu(\mathbf{t}) = \mathbf{a}R(\mathbf{t} - \tau)$ and $\mathbf{e}_{ij}(\mathbf{t})$ is stationary, then the global maximum of $U(\mathbf{t})$ in (2.9) is the likelihood ratio. In other cases, the likelihood ratio test statistic cannot be found easily.

## 1.3 Random fields and excursion sets

### 1.3.1 Random fields

The determination of the correct P-value of the two test statistics relies on random field theory. A random field is a random function defined over some Euclidean space. The familiar time series is an example of a random field in 1-D. More precisely, a random field is a collection of random variables $Z(\mathbf{t}), \mathbf{t} \in \mathrm{I\!R}^N$, together with a collection of joint distribution functions of $Z(\mathbf{t}_1), \ldots, Z(\mathbf{t}_n)$ for any $\mathbf{t}_1, \ldots, \mathbf{t}_n \in \mathrm{I\!R}^N$. The collection of such joint distributions will be later referred as the finite-dimensional distributions, or the *fi-di* distributions. A celebrated result of Kolmogorov states that, under some general conditions, a random field $Z(\mathbf{t}), \mathbf{t} \in \mathrm{I\!R}^N$, is uniquely determined by its fi-di distributions.

The kind of random fields that are discussed in this thesis, are *stationary* random fields with *smooth sample paths*. *Stationary* means that the fi-di distributions are invariant under translations in the parameter $\mathbf{t}$. This ensures second-order stationarity, that is,

$$\mathrm{E}(Z(\mathbf{t})) = \mu, \qquad \mathrm{Cov}(Z(\mathbf{s}), Z(\mathbf{t})) = R(\mathbf{s} - \mathbf{t}), \tag{3.9}$$

where $R(\mathbf{t})$ is called the *covariance function*. *Smooth sample paths*, roughly speaking, is a property that the realizations of the random field are continuous and have sufficient derivatives. For a random field to have smooth sample paths, some regularity conditions, usually the conditions on the covariance function $R(\mathbf{t})$ in (3.9), have to be imposed (see Adler 1981, Chapter 3); in particular $R(\mathbf{t})$ must have a finite second derivative at $\mathbf{t} = \mathbf{0}$.

An important special class of random fields is the class of Gaussian fields. That is, the *fi-di* distribution of the field is *multivariate Gaussian*. Simulations of a stationary Gaussian random field with mean 0 and variance 1 in the case $N = 1$ and $N = 2$ are shown in Figure B-6. Gaussian random fields will be used as building blocks for defining other random fields, for example, $\chi^2, t, F$ and Hotelling's $T^2$ random fields.

These random fields are the main focus of this thesis.

### 1.3.2 Excursion sets of the random fields

To test for brain functional and structural change, the regions where the random field have high values are of main interest because those regions correspond to high change. This leads to the investigation of the *excursion set* of the random field.

The excursion set of the random field is the set of points where the random field exceeds a given threshold. Let $Z(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$, be a real-valued random field and let $C$ be a compact subset of $\mathbb{R}^N$. The excursion set of $Z(\mathbf{t})$ inside $C$ above the level $z$ is defined as $A_z(Z, C) = \{\mathbf{t} \in C : Z(\mathbf{t}) \geq z\}$. The excursion sets at level $1, 1.2, 1.6$ of the simulated Gaussian field on a unit square in 2-D are shown in Figure B-6, B-7 respectively.

The general theory about the geometry of the excursion set of random fields was pioneered by Adler (1981). In his book, the notion of *Euler characteristic* (EC) from differential topology was used as a descriptive tool for the excursion set. A brief description of the EC is necessary at this point. Obviously, the adjacent points in the excursion set form *clusters*, or *connected components*. In the case of $N = 1$, the geometry of these connected components is simple: they are just line segments. The EC in this case counts the number of connected components in the set. But in higher dimensions, the topology of these connected components becomes more complex. In 2D, the EC counts roughly speaking, the number of connected components minus the number of holes (see Figure B-8) and in 3D, the EC counts roughly speaking the number of connected components minus the number of holes plus the number of hollows in the set.

For high level excursion sets for smooth random fields, there is a close relation between the Euler characteristic and the number of local maxima. When the threshold gets higher, the excursion set of smooth random fields tends to have a simpler topology. The holes and hollows disappear and we are left with isolated regions each of which has one *local maximum* (see Figure B-8 for illustration). The EC in this case counts just the number of connected components, or counts the number of local maxima in the set. Define $\chi(A_z(Z,C))$ to be the EC of the excursion set and $M_z^+(Z,C)$ to be the number of local maxima in the excursion set $A_z(Z,C)$. Then when $z$ is high enough,

$$\chi(A_z(Z,C)) \approx \text{no. of connected components in } A_z(Z,C) \approx M_z^+(Z,C). \quad (3.10)$$

More than that, the EC can be used to approximate the P-value of the global maximum at high levels. The reason for this is as follows. When the threshold gets even higher, the chance of getting more than one connected component becomes a rare event. Hence, if the excursion set is not void, it would be composed of one connected component with one local maximum which is indeed the global maximum. Let $Z_{\max}$ be the global maximum of $Z$ inside $C$. Then Hasofer (1978) gave a heuristic argument that when $z \to \infty$,

$$\mathrm{P}\{Z_{\max} \geq z\} = \mathrm{P}(M_z^+(Z,C) \geq 1) \leq \mathrm{E}(M_z^+(Z,C)) \approx \mathrm{E}(\chi(A_z(Z,C))).$$

The importance of $\chi(A_z(Z,C))$, as opposed to $M_z^+(Z,C)$, is that despite its complex definition, it is more amenable to statistical analysis. Only asymptotic results can be obtained for $\mathrm{P}(Z_{\max} \geq z)$ and $\mathrm{E}(M_z^+(Z,C))$, but for $\mathrm{E}(\chi(A_z(Z,C)))$, Adler (1981) has derived an exact closed form for stationary Gaussian fields, and Worsley (1994) extended the result to $\chi^2, t$ and $F$ fields. Next, a review of this work will be given.

### 1.3.3 Review of the expected EC and local maxima

Let $Z(\mathbf{t})$ be a smooth stationary random field, and let $Z = Z(\mathbf{t})$. Let $\dot{\mathbf{Z}} = \partial Z/\partial \mathbf{t}$ be the gradient $N$-vector of $Z$ with $i$th element $\dot{Z}_i$, and let $\ddot{\mathbf{Z}} = \partial^2 Z/\partial \mathbf{t} \partial \mathbf{t}'$ be the $N \times N$ Hessian matrix of $Z$ with $(i,j)$th element $\ddot{Z}_{ij}$. For a matrix $D$, let $D^+$ be $D$ if $D$ is positive definite and zero otherwise and let $D^-$ be $D$ if $D$ is negative definite and zero otherwise. We shall require that the boundary $\partial C$ of $C$ must be twice differentiable, and let $\mathbf{c}$ be the $(N-1) \times (N-1)$ inside curvature matrix of $\partial C$. At a point $\mathbf{t} \in \partial C$, let $\dot{Z}_\perp$ be the gradient of $Z$ in the direction of the inside normal to $\partial C$, let $\dot{\mathbf{Z}}_{\mathrm{T}}$ be the gradient $(N-1)$-vector in the tangent plane to $\partial C$, let $\ddot{\mathbf{Z}}_{\mathrm{T}}$ be the $(N-1)\times(N-1)$ Hessian matrix in the tangent plane to $\partial C$. Let $\mathrm{sign}(z) = z/|z|$ if $z \neq 0$ and zero otherwise. We shall adopt the notation advocated by Knuth (1992) where a logical expression in parentheses takes the value one if it is true and zero if it is false. Finally, we shall define the moduli of continuity of $\dot{Z}_k$ and $\ddot{Z}_{kl}$ inside $C$ to be

$$\omega_k(h) = \sup_{||\mathbf{t}-\mathbf{s}||<h} |\dot{Z}_k(\mathbf{t}) - \dot{Z}_k(\mathbf{s})|, \quad \omega_{kl}(h) = \sup_{||\mathbf{t}-\mathbf{s}||<h} |\ddot{Z}_{kl}(\mathbf{t}) - \ddot{Z}_{kl}(\mathbf{s})|,$$

where the supremum is taken over all $\mathbf{t}, \mathbf{s} \in C, k, l = 1, \ldots, N$.

We shall now state Adler's results on the expectation of the local maxima in the excursion set for stationary random fields.

**Theorem 1.3.1** *Assume (i) that for any* $\epsilon > 0$

$$\mathrm{P}(\max_{k,l}\{\omega_k(h), \omega_{kl}(h)\} > \epsilon) = o(h^N) \quad \textit{as } h \downarrow 0.$$

*If (ii) all the second-order partial derivatives* $\ddot{Z}_{kl}$ *have finite variances conditional on* $Z, \dot{\mathbf{Z}}$*, and the density* $f_N(z, \dot{\mathbf{Z}})$ *of* $Z$ *and* $\dot{\mathbf{Z}}$ *is bounded above, then the expectation of* $M_z^+(Z, C)$ *is*

$$\mathrm{E}(M_z^+(Z,C)) = \lambda(C) \int_z^\infty \mathrm{E}\{-\det(\ddot{\mathbf{Z}}^-)|Z = y, \dot{\mathbf{Z}} = \mathbf{0}\} f_N(y, \mathbf{0}) dy.$$

Under the regularity conditions given in Theorem 1.3.1, Worsley (1995b) shows that, with probability one,

$$\begin{aligned} \chi(A_z(Z,C)) &= \sum_{\mathbf{t}\in C} (Z \geq z)\mathrm{sign}[\det(-\ddot{\mathbf{Z}})](\dot{\mathbf{Z}} = \mathbf{0}) \\ &+ \sum_{\mathbf{t}\in \partial C} (Z \geq z)(\dot{Z}_\perp < 0)\mathrm{sign}[\det(-\ddot{\mathbf{Z}}_{\mathrm{T}} - \dot{Z}_\perp \mathbf{c})](\dot{\mathbf{Z}}_{\mathrm{T}} = \mathbf{0}). \quad (3.11) \end{aligned}$$

The expectations of this point set representation ( 3.11) are given by Worsley

(1995b):

$$\begin{aligned} \mathrm{E}[\chi(A_z(Z,C))] &= \lambda(C)\mathrm{E}[(Z \geq z)\det(-\ddot{\mathbf{Z}}) \mid \dot{\mathbf{Z}} = \mathbf{0}]\theta(\mathbf{0}) \\ &+ \int_{\partial C} \mathrm{E}[(Z \geq z)(\dot{Z}_\perp < 0)\det(-\ddot{\mathbf{Z}}_\mathrm{T} - \dot{Z}_\perp \mathbf{c}) \mid \dot{\mathbf{Z}}_\mathrm{T} = \mathbf{0}]\theta_\mathrm{T}(\mathbf{0}), \end{aligned} \tag{3.12}$$

where $\theta(\cdot)$ and $\theta_\mathrm{T}(\cdot)$ are the densities of $\dot{\mathbf{Z}}$ and $\dot{\mathbf{Z}}_\mathrm{T}$, respectively.

It is possible to find a much simplified result for the expectations (3.12) when the field is isotropic in $\mathbf{t}$. First let $Z_{|j}(t_1, \ldots, t_j) = Z(t_1, \ldots, t_j, 0, \ldots, 0)$ be the restriction of $Z$ to a $j$-dimensional subspace of $\mathbf{t}$, $j = 0, \ldots, N$. Then $\dot{\mathbf{Z}}_{|j}$ is the vector of the first $j$ components of $\dot{\mathbf{Z}}$, and $\ddot{\mathbf{Z}}_{|j}$ is the $j \times j$ matrix of the first $j$ rows and columns of $\ddot{\mathbf{Z}}$, by an obvious extension of the notation. Then define the $j$-dimensional *EC intensity* as

$$\rho_j(z) = \mathrm{E}[(Z \geq z)\det(-\ddot{\mathbf{Z}}_{|j}) \mid \dot{\mathbf{Z}}_{|j} = \mathbf{0}]\theta_{|j}(\mathbf{0}), \tag{3.13}$$

where $\theta_{|j}(\cdot)$ is the density of $\dot{\mathbf{Z}}_{|j}$. Using this notation, the first term of the expectation (3.12) can be written as $\mathrm{E}[\chi(A_z(Z,C))] = \lambda(C)\rho_N(z)$. Thus $\rho_j(z)$ can be regarded as the intensity of the EC point-set process per unit volume of $\mathbb{R}^j$. Worsley (1995b) gives a simplification for $\rho_j(z)$ which involves less integration. Let $Z_j = \partial Z / \partial t_j$, and $x^+ = x$ if $x > 0$ and 0 otherwise. Then

$$\rho_j(z) = \mathrm{E}[\dot{Z}_j^+ \det(-\ddot{\mathbf{Z}}_{|j-1}) \mid \dot{\mathbf{Z}}_{|j-1} = \mathbf{0}, Z = z]\phi_{|j-1}(\mathbf{0}, z), \tag{3.14}$$

where $\phi_{|j-1}(\cdot,\cdot)$ is the joint density of $\dot{\mathbf{Z}}_{|j-1}$ and $Z$. This is the intensity of the DT characteristic, defined by Adler (1981), which equals the EC provided the excursion set does not touch the boundary of the search region $C$.

If $Z(\mathbf{t})$ is isotropic in $\mathbf{t}$ then the expected EC (3.12) depends on the random field $Z$ and the threshold $z$ only through the EC intensities $\rho_j(z)$. To do this, we need some more definitions. For a square $n \times n$ matrix $\mathbf{M}$ let $\mathrm{detr}_j(\mathbf{M})$ be the sum of all $j \times j$ principal minors of $\mathbf{M}$, so that $\mathrm{detr}_n(\mathbf{M}) = \det(\mathbf{M})$, $\mathrm{detr}_1(\mathbf{M}) = \mathrm{tr}(\mathbf{M})$ (trace of $\mathbf{M}$) and we define $\mathrm{detr}_0(\mathbf{M}) = 1$. Finally let $a_j = 2\pi^{j/2}/\Gamma(j/2)$ be the surface area of a unit $(j-1)$-sphere in $\mathbb{R}^j$. Then the expected EC (3.12) is given by:

$$\mathrm{E}[\chi(A_z(Z,C))] = \sum_{j=0}^{N-1} \left( \frac{1}{a_{N-j}} \int_{\partial C} \mathrm{detr}_{N-1-j}(\mathbf{c}) d\mathbf{t} \right) \rho_j(z) + \lambda(C)\rho_N(z). \qquad (3.15)$$

Note there is a simple interpretation of the term given in the bracket. When $j=0$, it is the Euler characteristic of $C$ and when $j=N-1$, it is half of $\lambda(\partial C)$ (see Worsley, 1995b for the case $N=3$). Our main task, therefore, is to evaluate $\rho_j(z)$ from (3.14).

A closed form of the EC density $\rho_N(z)$ for stationary Gaussian fields has been derived by Adler and Hasofer (1976) and Adler (1981) as follows. They showed that, under the same conditions as are required for Theorem 1.3.1 to be true, the mean value of the DT characteristic of the excursion set is given by:

$$\begin{aligned} \rho_N(z) &= (2\pi)^{-(N+1)/2} |\mathbf{\Lambda}|^{\frac{1}{2}} \exp(-z^2/2) \\ &\quad \times \sum_{j=0}^{\lfloor (N-1)/2 \rfloor} (-1)^j \frac{(2j)!2^j}{j!} \begin{pmatrix} N-1 \\ 2j \end{pmatrix} z^{N-1-2j}, \end{aligned}$$

where $\mathbf{\Lambda} = \mathrm{Var}(\dot{\mathbf{Z}})$.

We shall not state the corresponding results for $\chi^2, t$ and $F$ fields. Interested readers can find them in Worsley (1994).

### 1.3.4 The size of one connected component of the excursion set

The EC counts the number of connected components of the excursion sets at high levels, but there is another measure of these connected components, their sizes or spatial extent in $\mathbb{R}^N$ (area in 2D, volume in 3D). When the threshold gets higher, the connected components get more regular and become convex sets with one local maximum (see Figure B-7 for an illustration). Hence, a study of the shape of these convex sets will lead to an approximation for their sizes. Nosko (1969) investigated the shape of stationary Gaussian random fields above high levels. He found that Gaussian random fields have a deterministic quadratic form above high levels, which then implies that the connected components of the excursion set at high levels are ellipsoids. The following is a slightly different version of what he has shown:

**Theorem 1.3.2** *Suppose $Z(\mathbf{t})$ is a $N$-dimensional, real-valued, ergodic, stationary Gaussian field, with mean zero and satisfying the condition*

$$\max_{ij} \mathrm{E}\{|\ddot{Z}_{ij}(\mathbf{t}) - \ddot{Z}_{ij}(\mathbf{0})|^2\} \leq C\|\mathbf{t}\|^2,$$

*for some finite $C > 0$ and all $\mathbf{t}$ with $\mathbf{t}$ small enough. Given that the random field $Z(\mathbf{t})$ takes value $z$ at $\mathbf{t} = \mathbf{0}$, then with probability approaching one as $z \to \infty$, the field has the following representation over that component of the excursion set containing $\mathbf{0}$:*

$$Z(\mathbf{t}) = z - \frac{1}{2} z \mathbf{t}' \mathbf{\Lambda} \mathbf{t} + o(\frac{1}{z}).$$

By applying the above theorem, he further obtained an asymptotic distribution of the size $S$ of one connected component. His result states that $S^{\frac{2}{N}}$ follows an exponential distribution in the limit as $z \to 0$.

Some extensions of these results to $\chi^2$ fields in one and two dimensions have been done by Aronowich and Adler (1986, 1988). Their results state that $\chi^2$ random fields in one and two dimensions also have a deterministic quadratic form above high levels, while below low levels, $\chi^2$ fields have a random quadratic form.

## 1.4 P-value of the two test statistics

It is already mentioned that there are two test statistics proposed in the brain mapping literature, which are the *global maximum* and the *size of the largest connected component of the excursion set*.

The P-value for the global maximum can be approximated by the expected EC of the excursion set above the threshold. But determination of the P-value of the size of the largest connected component of the excursion set needs more explanation.

Suppose there are $L$ connected components in the excursion set with sizes $S_1, \ldots,$ $S_L$ and $S_{\max} = \max_{1 \leq i \leq N} S_i$. In Section 6.9, Adler (1981) pointed out that $L$ should have a limiting Poisson distribution. A Poisson distribution is intuitively sensible, in the sense that one can regard the centers of the connected components as multi-dimensional point processes with no 'memory' . Hence, $S_i, i = 1, \ldots, L$ are independent identically distributed as $S$ in the limiting sense. It would then be enough if one can find the distribution of the size of *one* connected component of the excursion set. In particular, we shall show that in Section 3.7, when $E(L)$ is large enough,

$$\mathrm{P}(S_{\max} \leq x | L \geq 1) \approx \exp\{-\mathrm{E}(L)\mathrm{P}(S \geq x)\}$$

and

$$\mathrm{E}(L) \approx \mathrm{E}(\chi(A_z(Z, C))). \tag{4.16}$$

Usually, only an asymptotic distribution for $S$ at high (or low levels) can be obtained. As we noted before in Section 1.3.2, for the derivation of this distribution, one needs to investigate the shape of the random field above high (or below low) levels.

Since only an asymptotic distribution for $S$ can be obtained, a correction to the mean size might achieve a better approximation. Denote $\lambda(A)$ to be the Lebesgue measure of the set $A$ as before, then Friston et al. (1994) corrected the mean size by ensuring that:

$$\mathrm{E}(S) = \mathrm{E}(\lambda(A_z(Z, C)))/\mathrm{E}(L).$$

Adler (1981), Chapter 1, shows that for any level $z$,

$$\mathrm{E}(\lambda(A_z(Z, C))) = \lambda(C)\Phi(-z) = \lambda(C)\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{4.17}$$

and $E(L)$ can be approximated by (4.16) (note that 4.17 is exact for all levels).

## 1.5 Objectives and Outline

The expectation of the EC of the excursion set for stationary Gaussian fields was derived by Adler and Hasofer (1976) and Adler (1981) and extended to $\chi^2, t$ and $F$ fields by Worsley (1994). Chapter 2 of the thesis aims to extend these results to Hotelling's $T^2$ fields, which are proposed by Worsley et al. (1996) to detect abnormal 3D brain deformations between groups.

Nosko (1969) has obtained an asymptotic distribution for the size of one connected component of the excursion set of stationary Gaussian fields. Applying Nosko's result, Friston et al. (1994) presented an approximate analysis for assessing the significance of the activation foci by their spatial extent. Although their approach originally aims to deal with images generated from Gaussian random fields, they also argued that it can be used on some non-Gaussian random fields, such as a $t$ field, through appropriate *Gaussianization*. However, their approach is only approximately valid in this case, since the Gaussianized field is not a Gaussian field. To find the distribution of $S_{\max}$ directly for non-Gaussian fields such as $\chi^2, t$ and $F$ fields is the focus of Chapter 3.

The thesis is organized as follows.

In Chapter 2, the Hotelling's $T^2$ field is rigorously defined, the expectation of the Euler characteristic of its excursion set is found, and the number of its extremal points as well as the conditional curvature at these points are discussed. The chapter concludes with an application of this work to a real example of 3-D brain deformations.

Chapter 3 focuses attention on the size of the connected components of the excursion set of $\chi^2, t$ and $F$ fields. In the chapter, results on $\chi^2, t$ and $F$ fields from Worsley (1994) are reviewed, the structure of these fields above high (or below low) levels are discussed, and the limiting distribution of the size of one connected component of the excursion set is obtained. The chapter closes with an approximate analysis of the distribution of the size of the largest connected component of the excursion set and an application of the result to PET images from a pain perception study.

The whole thesis concludes with a discussion of the directions for further work which is presented in Chapter 4.

# Chapter 2

# Local extrema and the expected Euler characteristic of excursion sets of a Hotelling's $T^2$ field

## 2.1 Introduction

The use of brain imaging techniques to study functional neuroanatomy via focal activation and brain mapping has made rapid progress in recent years. However, the ability to draw useful conclusions is limited by the anatomical variability among individual brains. To reduce the effect of the anatomic variability, spatial alignment of the images is performed. A 3D deformation field, whose value at each location is a 3-D displacement vector, is produced to bring the local structure of the individual brain into alignment with that of the template. The deformation can be achieved through a variety of ways, for example, landmark-driven local deformations, contour-driven deformations and volume-based deformations. Since the anatomic variability is especially pronounced for the cerebral cortex, surface-based deformations which aim to give a better alignment of cortical areas is also developed. This opens up the possibility of studying brain structural differences, since changes in brain structure will be reflected in changes in the deformations required to align that structure with the template. It can be applied to study the brain structural difference between groups or the effect of an explanatory variable on anatomical structure.

Worsley et al. (1996b) proposed the Hotelling's $T^2$ field to detect abnormal deformations between groups, or abnormal deformations associated with an external variable from a general linear model. They modelled the 3-D deformation field, at each location, by a 3-variate normal distribution with an arbitrary variance matrix that allows for correlation between the coordinates. A Hotelling's $T^2$ statistic which indicates evidence of significant difference in deformations between the groups, was calculated at each location to form a statistical image. Under the null hypothesis of no abnormal deformations, the statistical image can be well approximated by a Hotelling's $T^2$ field which we will define in Section 2.2. The global maximum of the random field can then be used to test the null hypothesis of no change.

Worsley et al. (1992) have used the Euler characteristic (EC) of the excursion sets of a random field as an estimator of the number of local non-zero signal components and its expected value as an approximate P-value for the global maximum. The expected EC of the excursion set of a stationary Gaussian field was derived by Adler and Hasofer (1976) and Adler (1981), and has been extended to $\chi^2, t$, and $F$ fields by Worsley (1994). The purpose of this chapter is to extend these results to the Hotelling's $T^2$ field.

The chapter is organized as follows. We start with the Wishart field in section 2.2 and then give the definition of the Hotelling's $T^2$ field and the representations of its first two derivatives in section 2.3. In section 2.4, we will focus on the EC of the Hotelling's $T^2$ field. In section 2.5 we shall consider the number of extremal points of the Hotelling's $T^2$ field and in section 2.6 we shall derive the limiting conditional distribution of the curvature of the field at these points. Finally in section 2.7 we shall apply this work to a real example from a 3-D brain deformation study.

At the end, we shall note that in this chapter, vectors and matrices are not going to appear in bold face unless we feel the need to do so.

## 2.2 Wishart field

### 2.2.1 Definition

Suppose $Z(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$ is a $\nu \times d$ matrix of i.i.d. Gaussian random fields on $\mathbb{R}^N$ with zero mean and unit variance. Define the Wishart random field with $\nu \geq d$ degrees of freedom as

$$W(\mathbf{t}) = Z(\mathbf{t})' Z(\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^N. \tag{2.1}$$

### 2.2.2 Representations of the derivatives

Before we proceed, we shall introduce some notations used throughout the thesis. Let $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, let $I_{d\times d}$ or simply $I$ be the $d \times d$ identity matrix, let the density function of a random variable $Y$ be $f_Y(y)$ or simply $f(y)$, let $\mathrm{Wishart}_d(\Sigma, \nu)$ represent the Wishart distribution of a $d \times d$ matrix with expectation $\nu\Sigma$ and degrees of freedom $\nu$, let $F_{n,m}$ represent the $F$ distribution with $n, m$ degrees of freedom for numerator and denominator respectively and let $\mathrm{Unif}(C)$ represent the uniform distribution over set $C$. Recall that $\mathrm{Normal}_d(\mu, \Sigma)$ represents the multivariate normal distribution on $\mathbb{R}^d$ with mean $\mu$ and variance $\Sigma$ and $\chi^2_\nu$ represents the $\chi^2$ distribution with $\nu$ degrees for freedom.

We shall assume, for the time being, the $N \times N$ identity matrix is the variance-covariance matrix of the first derivative of the component Gaussian field of $Z(\mathbf{t})$. The following lemma is about the first two derivatives of $Z(\mathbf{t})$.

**Lemma 2.2.1** **(a)** *Let $\dot{Z}_i$ be the first-order partial derivative of $Z$ with respect to $t_i$ and $\ddot{Z}_{ij}$ be the second-order partial derivative of $Z$ with respect to $t_i$ and $t_j$. Then $(\dot{Z}_1, \ldots, \dot{Z}_N)$ are joint normal with mean zero, covariance matrix*

$$\mathrm{Cov}(\dot{Z}_j, \dot{Z}_k) = \delta_{jk} I_{\nu d \times \nu d}, \quad j, k = 1, \ldots N$$

*independent of $Z$.*

**(b)** *For $j,k=1,\ldots,N$, $\ddot{Z}_{jk} \stackrel{D}{=} -\delta_{jk}Z + H_{jk}$, where $H_{jk}$ are independent of $Z$ and joint normal with mean zero and covariance matrix*

$$\mathrm{Cov}(H_{ij},H_{kl}) = (\epsilon(i,j,k,l) - \delta_{ij}\delta_{kl})I_{\nu d\times\nu d} \qquad (2.2)$$

*where $\epsilon(i,j,k,l)$ is symmetric in its arguments and $\stackrel{D}{=}$ is equality in law.*

The proof of this follows directly from the proof of Theorem 5.3.1 from Adler (1981). If $\nu = d = 1$ we shall refer the type of covariance matrix as in (2.2) by $M(I_{N\times N})$.

**Lemma 2.2.2** *We can write the first two derivatives of $W = W(\mathbf{t})$ in terms of independent random variables as follows, where $\stackrel{D}{=}$ is equality in law:*

**(a)** *for $j = 1,\ldots,N$, $\dot{W}_j \stackrel{D}{=} W^{\frac{1}{2}}A_j + (W^{\frac{1}{2}}A_j)'$;*

**(b)** *for $j,k=1,\ldots,N$,*

$$\ddot{W}_{jk} \stackrel{D}{=} -2\delta_{jk}W + (W^{\frac{1}{2}}H^*_{jk})' + W^{\frac{1}{2}}H^*_{jk} + V_{jk} + V_{kj} + A'_jA_k + A'_kA_j,$$

*where $W \sim \mathrm{Wishart}_d(0,I)$, $A = (A_1,\ldots,A_N) \sim \mathrm{Normal}_{d\times dN}(0,I)$, $V = (V_{jk}) \sim \mathrm{Wishart}_{dN}(I,\nu-d)$ and $H^*_{jk} \stackrel{D}{=} H_{jk}$ in Lemma 2.2.1 (b), all independently.*

*Proof.* For $j = 1,\ldots,N$, we have

$$\dot{W}_j = Z'\dot{Z}_j + \dot{Z}'_jZ.$$

Conditional on $Z$, $Z'\dot{Z}_j \stackrel{D}{=} (Z'Z)^{\frac{1}{2}}A_j$, where $A_j$ is multivariate normal with all of its elements i.i.d $N(0,1)$. Furthermore, $A_j$ and $A_k$ are independent if $j \neq k$. So $A = (A_1,\ldots,A_N) \sim \mathrm{Normal}_{d\times dN}(0,I)$ conditional on $Z$. Since the conditional distribution does not depend on $Z$, it is also the unconditional distribution. This proves (a).

For the second derivative, we have for $j = 1,\ldots,N$ and $k = 1,\ldots,N$,

$$\ddot{W}_{jk} = \ddot{Z}'_{jk}Z + \dot{Z}'_k\dot{Z}_j + \dot{Z}'_j\dot{Z}_k + Z'\ddot{Z}_{jk}.$$

Note by Lemma 2.2.1 (b), $\ddot{Z}_{jk} \stackrel{D}{=} -\delta_{jk}Z + H_{jk}$, where $H_{jk}$ is independent of $Z$. Replace $\ddot{Z}_{jk}$ by $-\delta_{jk}Z + H_{jk}$ and rewrite $I$ as $(I - Z(Z'Z)^{-1}Z') + Z(Z'Z)^{-1}Z'$ in the above expression,

$$\begin{aligned}\ddot{W}_{jk} &\stackrel{D}{=} (-\delta_{jk}Z + H_{jk})'Z + Z'(-\delta_{jk}Z + H_{jk}) + \dot{Z}_j'(I - Z(Z'Z)^{-1}Z')\dot{Z}_k \\ &\quad + \dot{Z}_k'(I - Z(Z'Z)^{-1}Z')\dot{Z}_j + \dot{Z}_j'Z(Z'Z)^{-1}Z'\dot{Z}_k + \dot{Z}_k'Z(Z'Z)^{-1}Z'\dot{Z}_j.\end{aligned}$$

Define $V_{jk} = \dot{Z}_j'(I - Z(Z'Z)^{-1}Z')\dot{Z}_k$ and $H_{jk}^* = W^{-\frac{1}{2}}Z'H_{jk}$, then

$$\begin{aligned}\ddot{W}_{jk} &\stackrel{D}{=} -2\delta_{jk}W + H_{jk}'Z + Z'H_{jk} + V_{jk} + V_{kj} + A_j'A_k + A_k'A_j \\ &= -2\delta_{jk}W + (W^{\frac{1}{2}}H_{jk}^*)' + W^{\frac{1}{2}}H_{jk}^* + V_{jk} + V_{kj} + A_j'A_k + A_k'A_j.\end{aligned}$$

It is easy to verify that $V = (V_{jk}) \sim \text{Wishart}_{\nu N\times d}(I, \nu - d)$ and $H_{jk}^* \stackrel{D}{=} H_{jk}$ independent of $A$ and $W$, hence (b). $\square$

## 2.3 Hotelling's $T^2$ field

### 2.3.1 Definition

Let $W$ be a Wishart random field defined in (2.1) and $z$ be a vector of $d$ i.i.d Gaussian random fields independent of $W$ and with the same distribution as the component Gaussian fields of $Z$ in (2.1). Hotelling's $T^2$ field is then defined by $\nu Y$, where

$$Y = z'W^{-1}z. \tag{3.3}$$

Although the Hotelling's $T^2$ statistic is well defined *at a point* provided $\nu \geq d$, it is not clear that this is so at all points in a compact subset $C$ of $\mathbb{R}^N$. Consider the case where $d = 1$, so that Hotelling's $T^2$ field is a scalar multiple of an $F$ field with 1 and $\nu$ degrees of freedom. Worsley (1994) shows that if $\nu < N$ then *at some point* inside a finite region both the numerator and the denominator of the $F$ field can be exactly zero, so that the $F$ field is not well defined.

We shall now show that, in general, we require $\nu \geq d + N - 1$. For the general case, note that we can write

$$Y = \det(W + zz')/\det(W) - 1.$$

$Y$ is well defined if and only if either $\det(W + zz') > 0$ or $\det(W) > 0$. Since the latter condition implies the former, then the latter is both necessary and sufficient. Now we can write

$$\det(W + zz') = \prod_{i=\nu+2-d}^{\nu+1} U_i, \quad U_i = \sum_{j=1}^{i} X_{ij}^2, \tag{3.4}$$

and $X_{ij}$ are independent Gaussian random variables. Hence $\det(W + zz') > 0$ if and only if $U_i > 0$ for each $i$. Now $U_i = 0$ if and only if $X_{ij} = 0$ for each $j = 1, \ldots, i$. Inspection of the derivation of (3.4) (e.g. Anderson, 1983) shows that if $z$ and $Z$ are spatially smooth, then so are $X_{ij}$. The set of points inside a compact subset $C$ of $\mathrm{I\!R}^N$ where $X_{ij} = 0$ for each $j = 1, \ldots, i$ is a manifold of dimension no greater than $N - i$, and empty if $i > N$, with probability one. Hence $i > N$ is a necessary and sufficient condition for $U_i > 0$ at all points inside $C$, with probability one. For all $U_i > 0$ in (3.4) we require $\nu + 2 - d > N$, or $\nu \geq N + d - 1$.

Note that if $\nu = N + d - 1$ then $\det(W) = 0$ is possible at isolated points, so that $Y$ may be infinite at isolated points; if $\nu > N + d - 1$ then $Y$ is almost surely finite everywhere inside $C$. On the other hand, if $d \leq N$ then all components of $z$ can be zero, and hence $Y$ can be zero, on a manifold of dimension $N - d$.

A look at this argument will show that the distributional properties of $z$ and $Z$ do not enter into consideration; the only requirement is independence of the components, smoothness of their realizations, and unbounded support.

### 2.3.2 Representation of derivatives

**Lemma 2.3.1** *We can write the first two derivatives of* $Y = Y(\mathbf{t})$ *in terms of independent variables as follows, where* $\stackrel{D}{=}$ *is equality in law:*

**(a)** $\dot{Y} \stackrel{D}{=} 2(1+Y)Y^{\frac{1}{2}}e'T^{-\frac{1}{2}}\xi$,

**(b)** $\ddot{Y} \stackrel{D}{=} 2(1+Y)\{e'T^{-1}e + \frac{3}{4(1+Y)^2}\dot{Y}\dot{Y}' - Y^{\frac{1}{2}}(\xi'T^{-1}\xi)^{\frac{1}{2}}[e'T^{-\frac{1}{2}}f + f'T^{-\frac{1}{2}}e - H]$
$- Y(\xi'T^{-1}\xi)U\}$,

*where* $(\nu-d+1)Y/d \sim F_{d,\nu-d+1}$, $e \sim \mathrm{Normal}_{d\times N}(0,I)$, $T \sim \mathrm{Wishart}_d(I,\nu+1)$, $\xi \sim \mathrm{Unif}(S(\mathbb{R}^N))$ *where* $S(\mathbb{R}^N))$ *is the unit sphere of* $\mathbb{R}^N$, $f \sim \mathrm{Normal}_{d\times N}(0,I)$, $U \sim \mathrm{Wishart}_N(I,\nu-d)$ *and* $H \sim \mathrm{Normal}_{N\times N}(0,M)$, *all independently.*

*Proof.* By Lemma 2.2.2, for $j=1,\ldots,N$, the first derivative of $Y$ with respect to $t_j$ is:

$$\begin{aligned}
\dot{Y}_j &= 2z'W^{-1}\dot{z}_j - z'W^{-1}\dot{W}_jW^{-1}z \stackrel{D}{=} 2z'W^{-1}\dot{z}_j - z'W^{-1}(W^{\frac{1}{2}}A_j + (W^{\frac{1}{2}}A_j)')W^{-1}z \\
&= 2z'W^{-1}(\dot{z}_j - A_j'W^{-\frac{1}{2}}z) = 2z'W^{-1}a_j, \qquad (3.5)
\end{aligned}$$

where $a_j = \dot{z}_j - A_j'W^{-\frac{1}{2}}z$. Conditional on $z, W$,

$$\mathrm{Cov}(a_j,a_k) = \mathrm{Cov}(\dot{z}_j,\dot{z}_k) + \mathrm{Cov}(A_j'W^{-\frac{1}{2}}z, A_k'W^{-\frac{1}{2}}z) = \delta_{jk}(1+Y)I_{d\times d}. \qquad (3.6)$$

Define $e_j = (1+Y)^{-\frac{1}{2}}a_j$, then $e_j$ is normally distributed with $\mathrm{Cov}(e_j,e_k) = \delta_{jk}I_{d\times d}$ and is independent of $z, W$. Let $T = W + zz'$, $g = T^{-\frac{1}{2}}z$, then

$$z = T^{\frac{1}{2}}g, \qquad W = T^{\frac{1}{2}}(I-gg')T^{\frac{1}{2}} \qquad (3.7)$$

and the joint distribution of $T$ and $g$ is

$$\begin{aligned}
f_{T,g}(T,g) &= f_{z,W}(z,W)\left|\frac{\partial(z,W)}{\partial(T,g)}\right| \\
&= \mathrm{const}\cdot|T|^{\frac{\nu-d-1}{2}}\exp(-\frac{1}{2}\mathrm{tr}T)(1-g'g)^{\frac{\nu-d-1}{2}}\left|\frac{\partial(z,W)}{\partial(T,g)}\right|.
\end{aligned}$$

Since

$$\left|\frac{\partial(z,W)}{\partial(T,g)}\right| = \left|\frac{\partial(z,W)}{\partial(z,T)}\right|\left|\frac{\partial(z,T)}{\partial(g,T)}\right| = |T^{\frac{1}{2}}|,$$

$T \sim \text{Wishart}_d(I, \nu+1)$ and $g$ is with density $f_g(g) \propto (1-g'g)^{\frac{\nu-d-1}{2}}$ independently. Substitute $z$ by $T^{\frac{1}{2}}g$, $W$ by $T^{\frac{1}{2}}(I-gg')T^{\frac{1}{2}}$ and $a_j$ by $(1+Y)^{\frac{1}{2}}e_j$ in (3.5), then

$$\dot{Y}_j \stackrel{D}{=} 2(1+Y)^{\frac{1}{2}}g'(I-gg')^{-1}e_j.$$

Using the fact that

$$(I-gg')^{-1} = I + \frac{gg'}{1-g'g}, \tag{3.8}$$

then

$$\dot{Y}_j \stackrel{D}{=} 2(1+Y)^{\frac{1}{2}}(1-g'g)^{-1}g'T^{-\frac{1}{2}}e_j.$$

Define $\xi = g/\|g\|$, since the distribution of $g$ only depends on its norm, $\xi$ is uniformly distributed on the unit sphere of $\mathbb{R}^N$ independent of $\|g\|$. Note that

$$1 + Y = (1-g'g)^{-1} \tag{3.9}$$

and define $e = (e_1, \ldots, e_N)$, then

$$\dot{Y} \stackrel{D}{=} 2(1+Y)Y^{\frac{1}{2}}e'T^{-\frac{1}{2}}\xi$$

which proves (a).

The second derivative of $Y$ with respect to $t_j$ and $t_k$, $j, k = 1, \ldots, N$ is

$$\begin{aligned}
\ddot{Y}_{jk} &= 2\ddot{z}_{jk}'W^{-1}z - 2\dot{z}_j'W^{-1}\dot{W}_kW^{-1}z + 2\dot{z}_j'W^{-1}\dot{z}_k - 2\dot{z}_k'W^{-1}\dot{W}_jW^{-1}z \\
&\quad +2z'W^{-1}\dot{W}_kW^{-1}\dot{W}_jW^{-1}z - z'W^{-1}\ddot{W}_{jk}W^{-1}z \\
&= 2\ddot{z}_{jk}'W^{-1}z + 2(\dot{z}_j - \dot{W}_jW^{-1}z)'W^{-1}(\dot{z}_k - \dot{W}_kW^{-1}z) - z'W^{-1}\ddot{W}_{jk}W^{-1}z.
\end{aligned}$$

Applying Lemma 2.2.1 to $z$ with $\nu$ replaced by 1, we can write $\ddot{z}_{jk}$ as $-\delta_{jk}z + h_{jk}$, where $h_{jk}$ is independent of $z$ with the distribution specified in Lemma 2.2.1. Substituting $\ddot{z}_{jk}$ by $-\delta_{jk}z + h_{jk}$, and $\dot{W}_j$, $\ddot{W}_{jk}$ by its representation in Lemma 2.2.2, we have

$$
\begin{aligned}
\ddot{Y}_{jk} \quad &\stackrel{D}{=} \quad 2h'_{jk}W^{-1}z + 2(\dot{z}_j - A'_jW^{-\frac{1}{2}}z - W^{\frac{1}{2}}A_jW^{-1}z)'W^{-1} \\
& \quad (\dot{z}_k - A'_kW^{-\frac{1}{2}}z - W^{\frac{1}{2}}A_kW^{-1}z) - 2z'W^{-\frac{1}{2}}H^*_{jk}W^{-1}z \\
& \quad -2z'W^{-1}V_{jk}W^{-1}z - 2z'W^{-1}A'_jA_kW^{-1}z. \qquad (3.10)
\end{aligned}
$$

Consider the combination of the first and third term in (3.10) which is

$$2(h'_{jk}W^{-1}z - z'W^{-\frac{1}{2}}H^*_{jk}W^{-1}z).$$

If we let

$$H_{jk} = (1+Y)^{-\frac{1}{2}}(z'W^{-2}z)^{-\frac{1}{2}}(h'_{jk}W^{-1}z - z'W^{-\frac{1}{2}}H^*_{jk}W^{-1}z),$$

then it can be easily verified that $H = (H_{jk}) \sim \mathrm{Normal}_{N\times N}(0, M)$ independent of $z, W$.

Consider the second to last term in (3.10) which is $-2z'W^{-1}V_{jk}W^{-1}z$ and define

$$U_{jk} = (z'W^{-2}z)^{-1}(z'W^{-1}V_{jk}W^{-1}z),$$

then it is easy to see $U = (U_{jk}) \sim \mathrm{Wishart}_N(I, \nu - d)$ independent of $z, W$.

Consider the combination of the second term and the last term in (3.10) which is

$$
\begin{gathered}
2(\dot{z}_j - A'_jW^{-\frac{1}{2}}z - W^{\frac{1}{2}}A_jW^{-1}z)'W^{-1}(\dot{z}_k - A'_kW^{-\frac{1}{2}}z - W^{\frac{1}{2}}A_kW^{-1}z) \\
-2z'W^{-1}A'_jA_kW^{-1}z.
\end{gathered}
$$

Define $b_j = W^{\frac{1}{2}}A_jW^{-1}z$, then conditional on $z, W$, $c_j = b_j - \mathrm{Cov}(b_j, a_j)\mathrm{Var}(a_j)^{-1}a_j$ is independent of $a_k$. Note that $\mathrm{E}(A_jMA_j) = M'$ for a matrix $M$, then

$$
\begin{aligned}
\mathrm{Cov}(b_j, a_k) &= -\delta_{jk}W^{\frac{1}{2}}\mathrm{E}(A_jW^{-1}zz'W^{-\frac{1}{2}}A_k) \\
&= -\delta_{jk}W^{\frac{1}{2}}[W^{-1}zz'W^{-\frac{1}{2}}]' \;=\; -\delta_{jk}zz'W^{-1}
\end{aligned}
$$

and from (3.6),

$$b_j = c_j - \frac{1}{1+Y} zz'W^{-1}a_j.$$

Therefore,

$$\begin{aligned}
& (\dot{z}_j - A_j'W^{-\frac{1}{2}}z - W^{\frac{1}{2}}A_jW^{-1}z)'W^{-1}(\dot{z}_k - A_k'W^{-\frac{1}{2}}z - W^{\frac{1}{2}}A_kW^{-1}z) \\
& \qquad -z'W^{-1}A_j'A_kW^{-1}z \\
= \; & (a_j - b_j)'W^{-1}(a_k - b_k) - b_j'W^{-1}b_k \; = \; a_j'W^{-1}a_k - a_j'W^{-1}b_k - b_j'W^{-1}a_k \\
= \; & a_j'W^{-1}a_k + \frac{2}{1+Y}a_j'W^{-1}zz'W^{-1}a_k - a_j'W^{-1}c_k - c_j'W^{-1}a_k \\
= \; & a_j'(W^{-1} - \frac{1}{1+Y}W^{-1}zz'W^{-1})a_k + \frac{3}{1+Y}a_j'W^{-1}zz'W^{-1}a_k - a_j'W^{-1}c_k \\
& -c_j'W^{-1}a_k \\
= \; & (1+Y)e_j'T^{-1}e_k + \frac{3}{4(1+Y)}\dot{Y}_j'\dot{Y}_k - a_j'W^{-1}c_k - c_j'W^{-1}a_k.
\end{aligned}$$

Since conditional on $z, W$,

$$\begin{aligned}
& \mathrm{Cov}(W^{-1}c_j, W^{-1}c_k) \\
= \; & \mathrm{Cov}(W^{-1}b_j, W^{-1}b_k) - \mathrm{Cov}(\frac{1}{1+Y}W^{-1}zz'W^{-1}a_j, \frac{1}{1+Y}W^{-1}zz'W^{-1}a_k) \\
= \; & \delta_{jk}\{(z'W^{-2}z)W^{-1} - \frac{1}{1+Y}(z'W^{-2}z)W^{-1}zz'W^{-1}\} \\
= \; & \delta_{jk}(z'W^{-2}z)T^{-1},
\end{aligned}$$

we can write $W^{-1}c_j \stackrel{D}{=} (z'W^{-2}z)^{\frac{1}{2}}T^{-\frac{1}{2}}f_j$ ,where $f = (f_1, \ldots, f_N)$ is normally distributed with $\mathrm{Cov}(f_j, f_k) = \delta_{jk}I_{d\times d}$ independent of $z, W, e_k$ and then

$$a_j'W^{-1}c_k \stackrel{D}{=} (z'W^{-2}z)^{\frac{1}{2}}(1+Y)^{\frac{1}{2}}e_j'T^{-\frac{1}{2}}f_k.$$

Now we can represent $\ddot{Y}_{jk}$ as:

$$\ddot{Y}_{jk} \quad \stackrel{D}{=} \quad 2\{(1+Y)e_j'T^{-1}e_k + \frac{3\dot{Y}_j'\dot{Y}_k}{4(1+Y)} - (z'W^{-2}z)^{\frac{1}{2}}(1+Y)^{\frac{1}{2}}$$

$$(e_j'T^{-\frac{1}{2}}f_k + e_k'T^{-\frac{1}{2}}f_j - H_{jk}) - (z'W^{-2}z)U_{jk}\}.$$

Note that by (3.7), (3.8) and (3.9),

$$z'W^{-2}z = g'(I-gg')^{-1}T^{-1}(I-gg')^{-1}g = Y(1+Y)\xi'T^{-1}\xi,$$

then we have

$$\begin{aligned}\ddot{Y} \quad &\stackrel{D}{=} \quad 2(1+Y)\{e'T^{-1}e + \frac{3}{4(1+Y)^2}\dot{Y}\dot{Y}' - Y^{\frac{1}{2}}(\xi'T^{-1}\xi)^{\frac{1}{2}}(e'T^{-\frac{1}{2}}f + f'T^{-\frac{1}{2}}e - H) \\ &\quad -Y(\xi'T^{-1}\xi)U\}. \end{aligned} \tag{3.11}$$

Hence (b). □

### 2.3.3 Comments

We know from standard multivariate statistics that at each point, $(\nu - d + 1)Y/d$, is a $F_{d,\nu-d+1}$ random variable. But we can see from Lemma 2.3.1 that if $d \geq 2$, it has a different representation for its derivatives from that of an $F$ field (Worsley, 1994), hence it is not an $F$ field.

### 2.3.4 Some properties of the independent variables $e, T, \xi$

If we look at the representations of the first- and second-derivatives of the Hotelling's $T^2$ field in Lemma 2.3.1, the independent variables $e, T, \xi$ are involved through terms $e'T^{-1}e, e'T^{-\frac{1}{2}}\xi$ and $\xi'T^{-1}\xi$. Therefore, it is worthwhile to explore the relationship between these three variables.

In what follows, we shall need the following lemma from Olkin and Rubin (1964), Theorem 4.1.

**Lemma 2.3.2** *Let $X : k \times p, k \leq p$, and $V : p \times p$ be independent random matrices, where the rows of $X$ are independently and identically distributed as* $\text{Normal}_p(0, \Sigma)$, *and* $V \sim \text{Wishart}_p(\Sigma, n)$*, then the density of* $G = XV^{-1}X'$ *is*

$$f(G) = c(p,k,n)\det(G)^{(p-k-1)/2}\det(I+G)^{-(n+k)/2}, \quad G > 0,$$

*where*

$$c(p,k,n) = \pi^{-k(k-1)/4}\prod_{i=1}^{k}\Gamma[\frac{1}{2}(n+k-i+1)]/\{\Gamma[\frac{1}{2}(p-i+1)]\Gamma[\frac{1}{2}(n+k-p-i+1)]\}. \tag{3.12}$$

We shall later refer to this type of distribution as Multi-$T^2(p,k,n)$.

The following lemma is about the distributional properties of $e'T^{-1}e, e'T^{-\frac{1}{2}}\xi$, and $\xi'T^{-1}\xi$. What we will achieve from this lemma is that we find a one-to-one transformation between random variables $e'T^{-1}e, e'T^{-\frac{1}{2}}\xi, \xi'T^{-1}\xi$ and three independent random variables shown in the lemma. Hence the joint distribution of $e'T^{-1}e, e'T^{-\frac{1}{2}}\xi, \xi'T^{-1}\xi$ can be easily obtained and furthermore, we can reduce the dimensionality a great deal from the original variables $e, T, \xi$.

**Lemma 2.3.3** *Let* $e : d \times N, d \geq N$*,* $T : d \times d$ *and* $\xi : d \times 1$ *be independent random matrices, where* $e \sim \text{Normal}_{d\times N}(0, I)$*,* $T \sim \text{Wishart}_d(I, \nu+1)$*,* $\xi \sim \text{Unif}(S(\mathbb{R}^N))$*. Then the three variables*

$$V_1 = e'T^{-1}e, \quad v_2 = (e'T^{-1}e)^{-\frac{1}{2}}e'T^{-\frac{1}{2}}\xi \quad \textit{and} \quad v_3 = \frac{\xi'T^{-1}\xi}{(\xi'T^{-\frac{1}{2}}e)(e'T^{-\frac{1}{2}}\xi)+1}$$

*are independently distributed as follows:*

**(a)** $V_1 \sim \text{Multi-}T^2(d, N, \nu+1)$.

**(b)** *If* $d > N$*, then* $f(v_2) = \dfrac{\Gamma(\frac{d}{2})}{\pi^{\frac{N}{2}}\Gamma(\frac{d-N}{2})}(1 - v_2'v_2)^{\frac{d-N}{2}-1}$*. If* $d = N$*, then* $v_2 \sim \text{Unif}(S(\mathbb{R}^N))$.

**(c)** $v_3^{-1} \sim \chi^2_{\nu+N-d+2}$.

*If* $d < N$*, then* $V_1$ *is singular, but (c) still holds.*

*Proof.* Since $\xi$ is distributed as $\mathrm{Unif}(S(\mathbb{R}^N))$, we can set $b = \xi \cdot L$, where $L$ is the square root of a $\chi_d^2$ random variable independent of $e, T, \xi$. Therefore $b \sim \mathrm{Normal}_d(0, I)$. If we define

$$a_1 = (T + ee')^{\frac{1}{2}} T^{-\frac{1}{2}} b, \quad a_2 = (T + ee')^{-\frac{1}{2}} e \quad \text{and} \quad A_3 = T + ee'$$

then

$$b = \{A_3^{\frac{1}{2}}(I - a_2 a_2') A_3^{\frac{1}{2}}\}^{\frac{1}{2}} A_3^{-\frac{1}{2}} a_1, \quad e = A_3^{\frac{1}{2}} a_2, \quad T = A_3^{\frac{1}{2}}(I - a_2 a_2') A_3^{\frac{1}{2}}$$

and the joint distribution of $a_1, a_2, A_3$ is

$$\begin{aligned} f_{a_1,a_2,A_3}(a_1, a_2, A_3) &= f_{b,e,T}(b, e, T) \left| \frac{\partial(b, e, T)}{\partial(a_1, a_2, A_3)} \right| \\ = \quad & \mathrm{const} \cdot \exp(-\frac{1}{2} a_1'(I - a_2 a_2') a_1) \cdot \exp(-\frac{1}{2} \mathrm{tr}(a_2' A_3 a_2)) |A_3|^{\frac{\nu-d}{2}} |I - a_2 a_2'|^{\frac{\nu-d}{2}} \\ & \cdot \exp(-\frac{1}{2} \mathrm{tr}(A_3 - a_2' A_3 a_2)) \left| \frac{\partial(b, e, T)}{\partial(a_1, a_2, A_3)} \right|. \end{aligned}$$

Since

$$\begin{aligned} \left| \frac{\partial(b, e, T)}{\partial(a_1, a_2, A_3)} \right| &= \left| \frac{\partial(b, e, T)}{\partial(a_1, e, T)} \right| \cdot \left| \frac{\partial(a_1, e, T)}{\partial(a_1, e, A_3)} \right| \cdot \left| \frac{\partial(a_1, e, A_3)}{\partial(a_1, a_2, A_3)} \right| \\ &= |T|^{\frac{1}{2}} \cdot |T + ee'|^{-\frac{1}{2}} \cdot |A_3|^{\frac{N}{2}} = |I - a_2 a_2'|^{\frac{1}{2}} \cdot |A_3|^{\frac{N}{2}}, \end{aligned}$$

hence,

$$\begin{aligned} & f_{a_1,a_2,A_3}(a_1, a_2, A_3) \\ = \quad & \mathrm{const} \cdot \exp(-\frac{1}{2} a_1'(I - a_2 a_2') a_1) \cdot \exp(-\frac{1}{2} \mathrm{tr} A_3) |A_3|^{\frac{\nu-d+N}{2}} |I - a_2 a_2'|^{\frac{\nu-d+1}{2}}. \end{aligned}$$

Now we can see that $A_3 \sim \mathrm{Wishart}_d(I, \nu + 1 + N)$ is independent of the joint distribution of $a_1, a_2$. Furthermore, we can observe that

$$V_1 = e'T^{-1}e = a_2' a_2 (I - a_2' a_2)^{-1}, \quad V_1^{\frac{1}{2}} v_2 = e' T^{-\frac{1}{2}} \xi = \frac{a_2' a_1}{(a_1' a_1 - a_1' a_2 a_2' a_1)^{\frac{1}{2}}}$$

and

$$v_3 = \frac{\xi' T^{-1} \xi}{(\xi' T^{-\frac{1}{2}} e)(e' T^{-\frac{1}{2}} \xi) + 1} = \frac{a_1' A_3^{-1} a_1}{a_1' a_1}.$$

For a given value of $a_1, a_2$, it follows from Anderson (1958) that $v_3^{-1} \sim \chi^2_{\nu+N-d+2}$ which does not depend on the actual value of $a_1, a_2$. Hence, $v_3$ is independent of the joint distribution of $V_1$ and $v_2$ since both of them are functions of $a_1$ and $a_2$.

Now we consider the joint distribution of $V_1$ and $v_2$. Let

$$x_1 = (e' T^{-1} e)^{-\frac{1}{2}} e' T^{-\frac{1}{2}} b, \quad x_2 = b'(I - T^{-\frac{1}{2}} e (e' T^{-1} e)^{-1} e' T^{-\frac{1}{2}}) b.$$

Then conditional on $e, T$,

$$x_1 \sim \mathrm{Normal}_N(0, I), \quad x_2 \sim \chi^2_{d-N} \quad \text{independently.}$$

Since both of the conditional distributions do not depend on the actual value of $e$ and $T$, then $x_1$, and $x_2$ are independent of $e, T$ and hence independent of $e' T^{-1} e$. So

$$v_2 = x_1 / (x_1' x_1 + x_2)^{\frac{1}{2}}$$

is independent of $V_1 = e' T^{-1} e$. Therefore, the joint distribution of $v_2$ and $x_2$ is:

$$\begin{aligned} f_{v_2, x_2}(v_2, x_2) &= f_{x_1, x_2}\left( \frac{v_2 x_2^{\frac{1}{2}}}{(1 - v_2' v_2)^{\frac{1}{2}}}, x_2 \right) \cdot \left| \frac{\partial(x_1, x_2)}{\partial(v_2, x_2)} \right| \\ &= \frac{1}{(2\pi)^{\frac{N}{2}} \Gamma(\frac{d-N}{2}) 2^{\frac{d-N}{2}}} \cdot \exp(-\frac{1}{2} \cdot \frac{x_2}{1 - v_2' v_2}) x_2^{\frac{d-N}{2} - 1} \left| \frac{\partial(x_1, x_2)}{\partial(v_2, x_2)} \right|. \end{aligned}$$

Since

$$\left| \frac{\partial(x_1, x_2)}{\partial(v_2, x_2)} \right| = x_2^{\frac{N}{2}} \left| \frac{I}{(1 - v_2' v_2)^{\frac{1}{2}}} + \frac{v_2 v_2'}{(1 - v_2' v_2)^{\frac{3}{2}}} \right|,$$

it turns out that

$$\left| \frac{\partial(x_1, x_2)}{\partial(v_2, x_2)} \right| = x_2^{\frac{N}{2}} (1 - v_2' v_2)^{-\frac{N}{2} - 1}.$$

Therefore,

$$f_{v_2,x_2}(v_2,x_2) = \frac{1}{(2\pi)^{\frac{N}{2}}\Gamma(\frac{d-N}{2})2^{\frac{d-N}{2}}} \exp(-\frac{1}{2}\cdot\frac{x_2}{1-v_2'v_2})x_2^{\frac{d}{2}-1}(1-v_2'v_2)^{-\frac{N}{2}-1}.$$

Integrating over $x_2$, we have (b). The distribution of $V_1 = e'T^{-1}e$ can be found to be (a) by using Lemma 2.3.2. □

In the following sections, we shall use the results in section 1.3.3 to find the expected EC and the expected number of local extrema of the Hotelling's $T^2$ field. We shall also discuss the limiting distribution of the curvature of a Hotelling's $T^2$ field at these extrema. In order to do so, we have to assume some regularity conditions on the random field. If we impose the appropriate regularity conditions given in Theorem 1.3.1 on each of the Gaussian component field in the definition of the Hotelling's $T^2$ field, then by similar arguments given by Worsley (1994, Lemma 4.2) for the $F$ field, we can prove that the Hotelling's $T^2$ field is suitably regular for (3.14) to hold. Similarly, for Theorem 2.5.2 and 2.6.1 to hold, we shall impose the regularity conditions in Theorem 1.3.2 on the component Gaussian fields in the definition of the Hotelling's $T^2$ field.

## 2.4 The EC of the Hotelling's $T^2$ field

### 2.4.1 The EC

The following theorem is about the EC density $\rho_N(y)$ of the Hotelling's $T^2$ field $\nu Y(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$. We shall use the expression (3.14) in Section 1.3.3 to evaluate this.

**Theorem 2.4.1** *For $\nu \geq d+N-1$, under the regularity conditions of Theorem 1.3.1 for each of the component Gaussian fields,*

$$\rho_N(y) = \pi^{-\frac{N+1}{2}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{d}{2})}\frac{2^{\nu-d}(N-1)!}{(\nu-d)!}(1+y)^{-\frac{\nu-1}{2}}y^{-\frac{N-d}{2}}Q_{\nu,d,N}(y), \tag{4.13}$$

where $Q_{\nu,d,N}(y)$ is a polynomial of degree $N-1$ in $y$ given by

$$Q_{\nu,d,N} = (-1)^{N-1} \sum_{j=0}^{\lfloor (N-1)/2 \rfloor} \sum_{k=0}^{N-1-2j} \sum_{m=0}^{\lfloor (N-1-2j-k)/2 \rfloor} \frac{(-1)^{j+k+m}\Gamma(\frac{\nu-d+N}{2}-j-k-m)}{j!m!(N-1-2j-k-2m)!2^{2j+k+m}}$$
$$y^{j+k+m} \begin{pmatrix} \nu-d \\ k \end{pmatrix} \begin{pmatrix} d-1 \\ N-2j-k-1-m \end{pmatrix} \begin{pmatrix} \nu-d+N-1-2j-k-m \\ N-1-2j-k-m \end{pmatrix}^{-1},$$

where division by the factorial of a negative integer is treated as multiplication by zero.

*Proof.* In this proof, we shall adopt all the notations established in the previous section unless otherwise stated.

(3.14) in Section 1.3.3 gives a general formula to calculate the EC density for random fields. We shall evaluate the expectation in (3.14) for the random field $Y(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$ by conditioning on $\xi' T^{-1}\xi$. Define $\dot{\mathbf{Y}}^* = (\dot{Y}_1, \ldots, \dot{Y}_{N-1})$, then we can write:

$$\begin{aligned} \rho_N(y) &= (-1)^{N-1}\mathrm{E}_{\xi' T^{-1}\xi}\{\mathrm{E}(\dot{Y}_N^+ \det(\ddot{\mathbf{Y}}^*)|Y=y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}^* = \mathbf{0}) \\ &\qquad f_{N-1}(\mathbf{0}|y, \xi T^{-1}\xi)\}f(y), \end{aligned}$$

where $\ddot{\mathbf{Y}}^*$ is the $(N-1)\times(N-1)$ matrix of second order partial derivatives of $Y$ with respect to $t_1, \ldots, t_{N-1}$, $f_{N-1}(\mathbf{0}|y, \xi T^{-1}\xi)$ is the density of $\dot{\mathbf{Y}}^*$ at $\mathbf{0}$ conditional on $Y = y$ and $\xi T^{-1}\xi$, $f(y)$ is the density of $Y$. Define

$$e^* = (e_1, \ldots, e_{N-1}), \quad f^* = (f_1, \ldots, f_{N-1}), \quad x^* = (e^{*'} T^{-1} e^*)^{-\frac{1}{2}} e^{*'} T^{-\frac{1}{2}} f^*,$$

and $U^*, H^*$ be the $(N-1)\times(N-1)$ matrix of the first $N-1$ rows and columns of $U, H$ respectively. Then $e^* \sim \mathrm{Normal}_{d\times(N-1)}(0, I), x^* \sim \mathrm{Normal}_{(N-1)\times(N-1)}(0, I), U^* \sim \mathrm{Wishart}_{N-1}(I, \nu-d), H^* \sim \mathrm{Normal}_{(N-1)\times(N-1)}(0, M^*(I))$ all independently. Furthermore, if we define

$$V_1^* = e^{*'} T^{-1} e^*, \quad a = y^{\frac{1}{2}}(\xi T^{-1}\xi)^{\frac{1}{2}}, \quad A = V_1^* - a[(V_1^*)^{\frac{1}{2}} x^* + x^{*'}(V_1^*)^{\frac{1}{2}}]$$

then by Lemma 2.3.1 conditional on $Y = y, \xi' T^{-1} \xi$ and $\dot{\mathbf{Y}}^* = \mathbf{0}$, we can write

$$\ddot{\mathbf{Y}}^* = 2(1+y)(A + aH^* - a^2 U^*).$$

Note by Lemma 2.3.1,

$$\begin{aligned}
\dot{Y}_N &= 2(1+Y)Y^{\frac{1}{2}} e'_N T^{-\frac{1}{2}} \xi \;=\; 2(1+Y)Y^{\frac{1}{2}} (\xi' T^{-1} \xi)^{\frac{1}{2}} \eta_N, \\
\dot{\mathbf{Y}}^* &= 2(1+Y)Y^{\frac{1}{2}} e^{*'} T^{-\frac{1}{2}} \xi, \qquad (4.14)
\end{aligned}$$

where $\eta_N$ is N$(0,1)$ independent of $Y, \xi, T, e^*, U$ and $H$, and hence $\dot{\mathbf{Y}}^*$. Therefore,

$$\begin{aligned}
&\mathrm{E}(\dot{Y}_N^+ \det(\ddot{\mathbf{Y}}^*) | Y = y, \xi' T^{-1} \xi, \dot{\mathbf{Y}}^* = \mathbf{0}) \\
= &\int \mathrm{E}\left\{ \dot{Y}_N^+ \det(\ddot{\mathbf{Y}}^*) | Y = y, \xi' T^{-1} \xi, \dot{\mathbf{Y}}^* = \mathbf{0}, V_1^* \right\} f(V_1^* | y, \xi' T^{-1} \xi, \mathbf{0}) dV_1^* \\
= &\; 2^N (1+y)^N a \int \mathrm{E}\left\{ \eta_N^+ \det(A + aH^* - a^2 U^*) | \dot{\mathbf{Y}}^* = \mathbf{0}, V_1^* \right\} f(V_1^* | y, \xi' T^{-1} \xi, \mathbf{0}) dV_1^*,
\end{aligned}$$

where $f(V_1^* | y, \xi' T^{-1} \xi, \mathbf{0})$ is the density of $V_1^*$ conditional on $Y = y, \xi' T^{-1} \xi$ and $\dot{\mathbf{Y}}^* = \mathbf{0}$ and in the last equality we are holding $a$ fixed.

It is easy to see that conditionally, $\eta_N^+$ and $A + aH^* - a^2 U^*$ are independent with $\mathrm{E}(\eta_N^+) = (2\pi)^{-\frac{1}{2}}$ and also note that $x^*, H^*, U^*$ are independent of $\dot{\mathbf{Y}}^*$, so that

$$\begin{aligned}
&\mathrm{E}(\dot{Y}_N^+ \det(\ddot{\mathbf{Y}}^*) | Y = y, \xi' T^{-1} \xi, \dot{\mathbf{Y}}^* = \mathbf{0}) \qquad (4.15) \\
= &\; (2\pi)^{-\frac{1}{2}} 2^N (1+y)^N a \int \mathrm{E}\left\{ \det(A + aH^* - a^2 U^*) | V_1^* \right\} f(V_1^* | y, \xi' T^{-1} \xi, \mathbf{0}) dV_1^*.
\end{aligned}$$

Now we shall find $\mathrm{E}\{\det(A + aH^* - a^2 U^*) | V_1^*\}$ first and then take expectations with respect to the conditional distribution $f(V_1^* | y, \xi' T^{-1} \xi, \mathbf{0})$. Holding $a$ fixed, we can apply Lemma A.0.5 to obtain

$$\begin{aligned}
&\mathrm{E}\left\{ \det(A + aH^* - a^2 U^*) | V_1^* \right\} \\
= &\sum_{j=0}^{\lfloor (N-1)/2 \rfloor} \sum_{k=0}^{N-1-2j} \sum_{m=0}^{\lfloor (N-1-2j-k)/2 \rfloor} \frac{(-1)^{j+k+m}}{2^j j!} a^{2j+2k+2m} \begin{pmatrix} \nu - d \\ k \end{pmatrix}
\end{aligned}$$

$$\frac{(2j+k+m)!(N-1-2j-k-m)!}{m!(N-1-2j-k-2m)!}\mathrm{detr}_{N-1-2j-k-m}(V_1^*), \tag{4.16}$$

Conditional on $Y=y$, recall

$$\dot{\mathbf{Y}}^* = 2(1+y)y^{\frac{1}{2}}e^{*\prime}T^{-\frac{1}{2}}\xi$$

by Lemma 2.3.1 and note that if we define

$$v_2^* = (e^{*\prime}T^{-1}e^*)^{-\frac{1}{2}}e^{*\prime}T^{-\frac{1}{2}}\xi, \qquad v_3^* = \frac{\xi' T^{-1}\xi}{(\xi' T^{-\frac{1}{2}}e^*)(e^{*\prime}T^{-\frac{1}{2}}\xi)+1},$$

then $V_1^*, v_2^*, v_3^*$ are independently distributed with the distribution shown in Lemma 2.3.3 by replacing $N$ by $N-1$. Hence,

$$f(V_1^*|y, \xi' T^{-1}\xi, \mathbf{0}) \propto f(V_1^*, \xi' T^{-1}\xi, e^{*\prime}T^{-\frac{1}{2}}\xi = \mathbf{0}) \propto f(V_1^*, v_2^* = \mathbf{0}, v_3^*) \cdot J,$$

where $J$ is the Jacobian and the constant of proportionality contains $y$. Since the Jacobian turn out to be $J = |V_1^*|^{-\frac{1}{2}}$, it can be easily seen that conditional on $Y=y$ and $\dot{\mathbf{Y}}^* = \mathbf{0}$, $V_1^*$ and $\xi' T^{-1}\xi$ are independent with

$$\begin{aligned} V_1^*|Y=y, \dot{\mathbf{Y}}^* = \mathbf{0} \quad &\sim \quad \text{Multi-}T^2(d-1, N-1, \nu+1), \quad \text{if} \quad d \geq N \\ (\xi' T^{-1}\xi)^{-1}|Y=y, \dot{\mathbf{Y}}^* = \mathbf{0} \quad &\sim \quad \chi^2_{\nu+N-d+1}. \end{aligned} \tag{4.17}$$

Hence for $k = 1, \ldots, N-1$,

$$(V_1^*)_k|Y=y, \dot{\mathbf{Y}}^* = \mathbf{0} \ \sim \ \text{Multi-}T^2(d-1, k, \nu+1) \ \text{ if } \ d-1 \geq k, \tag{4.18}$$

where $(V_1^*)_k$ is any $k \times k$ principle minor of $V_1^*$. So by Lemma A.0.1, we have

$$\mathrm{E}\left\{\det[(V_1^*)_k]|Y=y, \dot{\mathbf{Y}}^* = \mathbf{0}\right\} = \frac{(d-1)!(\nu-d)!}{(d-1-k)!(\nu-d+k)!} \ \text{ for any } \ k, \tag{4.19}$$

where division by the factorial of a negative integer is treated as multiplication by zero. Putting (4.15), (4.16), (4.19) together gives

$$
\begin{aligned}
& \mathrm{E}(\dot{Y}_N^+ \det(\ddot{\mathbf{Y}}^*)|Y=y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}^* = \mathbf{0}) \\
= \quad & (2\pi)^{-\frac{1}{2}} 2^N (1+y)^N \sum_{j=0}^{\lfloor (N-1)/2 \rfloor} \sum_{k=0}^{N-1-2j} \sum_{m=0}^{\lfloor (N-1-2j-k)/2 \rfloor} \begin{pmatrix} \nu - d \\ k \end{pmatrix} \\
& \begin{pmatrix} d-1 \\ N-1-2j-k-m \end{pmatrix} \begin{pmatrix} \nu - d + N - 1 - 2j - k - m \\ N-1-2j-k-m \end{pmatrix}^{-1} \\
& \frac{(-1)^{j+k+m}(N-1)!}{2^j j! m! (N-1-2j-k-2m)!} a^{2j+2k+2m+1}.
\end{aligned}
$$

Note that the density of $\dot{\mathbf{Y}}^*$ at $\mathbf{0}$ conditional on $Y=y$ and $\xi' T^{-1}\xi$ is

$$
f_{N-1}(\mathbf{0}|y, \xi' T^{-1}\xi) = \left\{ 8\pi(1+y)^2 y (\xi' T^{-1} \xi) \right\}^{-\frac{1}{2}(N-1)}
$$

as easily follows from (4.14), hence,

$$
\begin{aligned}
& \mathrm{E}(\dot{Y}_N^+ \det(\mathbf{D}_{N-1})|Y=y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}^* = \mathbf{0}) f_{N-1}(\mathbf{0}|y, \xi' T^{-1}\xi) \\
= \quad & (2\pi)^{-\frac{N}{2}} 2 (1+y) \sum_{j=0}^{\lfloor (N-1)/2 \rfloor} \sum_{k=0}^{N-1-2j} \sum_{m=0}^{\lfloor (N-1-2j-k)/2 \rfloor} \begin{pmatrix} \nu - d \\ k \end{pmatrix} \\
& \begin{pmatrix} d-1 \\ N-1-2j-k-m \end{pmatrix} \begin{pmatrix} \nu - d + N - 1 - 2j - k - m \\ N-1-2j-k-m \end{pmatrix}^{-1} \\
& \frac{(-1)^{j+k+m}(N-1)!}{2^j j! m! (N-1-2j-k-2m)!} a^{2j+2k+2m+2-N}.
\end{aligned}
$$

Since $(\xi' T^{-1}\xi)^{-1} \sim \chi^2_{\nu-d+2}$, using the fact that

$$
\mathrm{E}[(\xi' T^{-1}\xi)^l] = 2^{-l}\Gamma\{(\nu-d+2)/2 - l\}/\Gamma\{(\nu-d+2)/2\}
$$

and multiplying by the density of $Y$:

$$
f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu-d+1}{2})\Gamma(\frac{d}{2})} y^{\frac{1}{2}d-1}(1+y)^{-\frac{\nu+1}{2}}
$$

gives:

$$\rho_N(y) = \pi^{-\frac{N}{2}}(1+y)^{-\frac{\nu-1}{2}} y^{-\frac{N-d}{2}} \sum_{j=0}^{\lfloor (N-1)/2 \rfloor} \sum_{k=0}^{N-1-2j} \sum_{m=0}^{\lfloor (N-1-2j-k)/2 \rfloor}$$

$$\begin{pmatrix} \nu-d \\ k \end{pmatrix} \begin{pmatrix} d-1 \\ N-1-2j-k-m \end{pmatrix} \begin{pmatrix} \nu-d+N-1-2j-k-m \\ N-1-2j-k-m \end{pmatrix}^{-1}$$

$$\frac{(-1)^{N-1+j+k+m}(N-1)!}{2^{2j+k+m} j! m! (N-1-2j-k-2m)!} \frac{\Gamma(\frac{\nu-d+N}{2}-j-k-m)\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu-d}{2}+1)\Gamma(\frac{\nu-d+1}{2})\Gamma(\frac{d}{2})} y^{j+k}.$$

Then the theorem follows by using the factorization $x! = 2^x \Gamma\{(x+2)/2\}\Gamma\{(x+1)/2\}/\pi^{\frac{1}{2}}$. □

**Corollary 2.4.2** *If $N = 2$ and $\nu \geq d+1$ then*

$$\rho_N(y) = \pi^{-1} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{d}{2})\Gamma(\frac{\nu-d+1}{2})} (1+y)^{-\frac{\nu-1}{2}} y^{\frac{d}{2}-1} \left( -\frac{d-1}{\nu-d+1} + y \right),$$

*and if $N = 3$ and $\nu \geq d+2$ then*

$$\rho_N(y) = \pi^{-\frac{3}{2}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{d}{2})\Gamma(\frac{\nu-d}{2})} (1+y)^{-\frac{\nu-1}{2}} y^{\frac{d-3}{2}}$$
$$\left\{ \frac{(d-1)(d-2)}{(\nu-d+2)(\nu-d)} - \frac{2d-1}{\nu-d} y + y^2 \right\}.$$

### 2.4.2 Comments

Theorem 2.4.1 tells us a lot about the zeros (points $\mathbf{t}$ where the field is zero) and infinities (point $\mathbf{t}$ where the field is infinite) of a Hotelling's $T^2$ field. As $y \to \infty$, the EC counts the number of infinities whereas as $y \to 0$, the EC multiplied by $(-1)^{(N-1)}$, counts number of exact zeroes. Therefore, taking the limit when $y \to \infty$ in (4.13), we know that if $\nu \geq d+N$, then in the unit volume, $\rho_N(\infty) = 0$, with probability one, there are no points where the field is infinite. When $\nu = d+N-1$ there is a finite number of points where the field is infinite in a unit volume almost surely with the average

$$\rho_N(\infty) = \pi^{-\frac{N}{2}} \frac{\Gamma(\frac{d+N}{2})}{\Gamma(\frac{d}{2})}.$$

Taking the limit when $y \to 0$ in (4.13) shows that if $d > N$ the field has no zeros with probability one, if $d = N$ the field has almost surely finite zeros with average

$$\rho_N(0) = \pi^{-\frac{1}{2}(N+1)}\Gamma\{(N+1)/2\},$$

and if $d < N$ the field has almost surely infinitely many zeros in a unit volume with probability one. This is not surprising if we see that the behaviour of the Hotelling's $T^2$ field near infinity depends on the behaviour of the determinant of the Wishart field, $\det(W)$, near zero and the behaviour of the Hotelling's $T^2$ field near zero depends on the behaviour of the $\chi^2$ field, $z'z$, near zero. Similar discussions for $\chi^2, t$ and $F$ random fields can be found in Worsley (1994).

## 2.5 Expected number of the extremal points of the Hotelling's $T^2$ field

Theorem 1.3.1 has given a general formula to calculate the expectation of local extremal points. We shall apply the theorem to the Hotelling's $T^2$ field. Recall the following notations. For any symmetric matrix $D$, $D^- = D$ if $D$ is negative definite and zero otherwise, $D^+ = D$ if $D$ is positive definite and zero otherwise. For a random field $Y(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$, the number of local maxima of $Y(\mathbf{t})$ above $y$ inside a compact subset of $\mathbb{R}^N$ is denoted by $M_y^+(Y, C)$ and the number of local minima of $Y(\mathbf{t})$ below $y$ inside $C$ is denoted by $M_y^-(Y, C)$.

From now on, we shall assume that the component Gaussian fields of $Y$ satisfies the same regularity conditions as in 1.3.1.

**Theorem 2.5.1** *For a Hotelling's $T^2$ field $Y(\mathbf{t})$,*

**(a)** $\mathrm{E}(M_y^+(Y,C)) = \dfrac{\lambda(C)}{\pi^{\frac{N}{2}}} \dfrac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{d}{2})\Gamma(\frac{\nu-d-N+1}{2}+1)} y^{-\frac{\nu+1-d-N}{2}}\{1+O(y^{-\frac{1}{2}})\}$ *for* $\nu \geq d+N$,

**(b)** $\mathrm{E}(M_y^-(Y,C)) = \dfrac{\lambda(C)\Gamma(\frac{\nu+1}{2})(d-1)!}{\pi^{\frac{N}{2}}2^{N-1}\Gamma(\frac{d}{2})\Gamma(\frac{\nu+1+N-d}{2})(d-N)!} y^{\frac{d-N}{2}}\{1+O(y^{\frac{1}{2}})\}$ *for* $d > N$.

*Proof.* Rewriting the formula for the expectation of local extrema in Theorem 1.3.1 by conditioning on $\xi' T^{-1}\xi$ first, and then taking expectations gives:

$$\mathrm{E}(M_y^+(Y,C)) = \lambda(C)\int_y^\infty \mathrm{E}_{\xi' T^{-1}\xi}\{\mathrm{E}[|\det(\ddot{\mathbf{Y}}^-)|\,\big| Y=y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}=\mathbf{0}]$$
$$f_N(\mathbf{0}|y,\xi' T^{-1}\xi)\}f(y)dy,$$

where $f_N(\mathbf{0}|y,\xi' T^{-1}\xi)\}$ is the density of $\dot{\mathbf{Y}}$ at $\mathbf{0}$ conditional on $Y=y$ and $\xi' T^{-1}\xi$, and $f(y)$ is the density of $Y$. By Lemma 2.3.1, we can see that if $\dot{\mathbf{Y}}=\mathbf{0}$ then we can write

$$\ddot{\mathbf{Y}} = 2(1+Y)\left\{e' T^{-1}e - Y^{\frac{1}{2}}(\xi' T^{-1}\xi)^{\frac{1}{2}}[(e' T^{-1}e)^{\frac{1}{2}}x + x(e' T^{-1}e)^{\frac{1}{2}} - H]\right.$$
$$\left. -Y(\xi' T^{-1}\xi)U\right\},$$

where $U \sim \mathrm{Wishart}_N(I,\nu-d)$ is independent of $Y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}$. Thus $\det(\ddot{\mathbf{Y}}^-)$ is a polynomial in $Y^{\frac{1}{2}}$ of degree $2N$, multiplied by $(1+Y)^N$. Following the same arguments as in the proof of Theorem 6.3.1 of Adler (1981), page 134, we can see that as $Y\to\infty$, $\ddot{\mathbf{Y}}$ converges to the negative definite matrix $-2(1+Y)Y(\xi' T^{-1}\xi)U$ and so

$$\begin{aligned}&\mathrm{E}\left\{|\det(\ddot{\mathbf{Y}}^-)|\,\Big| Y=y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}=\mathbf{0}\right\}\\ =\;& 2^N(1+y)^N y^N(\xi' T^{-1}\xi)^N\mathrm{E}(\det(U))\{1+O(y^{-\frac{1}{2}})\}.\end{aligned}$$

From standard multivariate statistics we have $\mathrm{E}(\det(U)) = (\nu-d)!/(\nu-d-N)!$ and from Lemma 2.3.1(a) the density of $\dot{\mathbf{Y}}$ at $\mathbf{0}$ conditional on $Y, \xi' T^{-1}\xi$ is

$$f_N(\mathbf{0}|y,\xi' T^{-1}\xi) = (8\pi(1+y)^2 y)^{-\frac{N}{2}}(\xi' T^{-1}\xi)^{-\frac{N}{2}}. \tag{5.20}$$

Multiplying these together gives

$$\mathrm{E}\left\{|\det(\ddot{\mathbf{Y}}^-)|\,\Big| Y=y, \xi' T^{-1}\xi, \dot{\mathbf{Y}}=\mathbf{0}\right\} f_N(\mathbf{0}|Y,\xi' T^{-1}\xi)$$

$$= \quad (2\pi)^{-\frac{N}{2}} y^{\frac{N}{2}} \frac{(\nu-d)!}{(\nu-d-N)!} (\xi' T^{-1}\xi)^{\frac{N}{2}} \{1 + O(y^{-\frac{1}{2}})\}.$$

Using the fact that

$$\mathrm{E}[(\xi' T^{-1}\xi)^l] = 2^{-l}\Gamma\{(\nu-d+2)/2 - l\}/\Gamma\{(\nu-d+2)/2\}$$

to take expectations with respect to $\xi' T^{-1}\xi$, and multiplying by the density of $Y$:

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu-d+1}{2})\Gamma(\frac{d}{2})} y^{\frac{1}{2}d-1}(1+y)^{-\frac{\nu+1}{2}}$$

gives:

$$\begin{aligned}
&\mathrm{E}_{\xi' T^{-1}\xi}\left\{\mathrm{E}\left[|\det(\ddot{\mathbf{Y}}^-)|\,\middle|Y = y, \xi' T^{-1}\xi, \dot{\mathbf{Y}} = \mathbf{0}\right] f_N(\mathbf{0}|y, \xi' T^{-1}\xi)\right\} f(y) \\
= \quad & \pi^{-\frac{N}{2}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu-d-N+1}{2})\Gamma(\frac{d}{2})} y^{-\frac{\nu+1-d-N}{2}-1}\{1 + O(y^{-\frac{1}{2}})\}
\end{aligned}$$

after simplification using $x! = 2^x\Gamma\{(x+2)/2\}\Gamma\{(x+1)/2\}/\pi^{\frac{1}{2}}$. Integrating over $y$ gives (a).

Evaluating the expectations in Theorem 1.3.1 by conditioning on $e' T^{-1} e$ and then taking expectations over $e' T^{-1} e$ gives,

$$\begin{aligned}
\mathrm{E}(M_y^-(Y, C)) = \lambda(C) \int_y^\infty & \mathrm{E}_{e' T^{-1} e}\{\mathrm{E}[\det(\ddot{\mathbf{Y}}^+)\,\big|Y = y, e' T^{-1} e, \dot{\mathbf{Y}} = \mathbf{0}] \\
& f_N(\mathbf{0}|y, e' T^{-1} e)\} f(y) dy,
\end{aligned}$$

where $f_N(\mathbf{0}|y, e' T^{-1} e)$ is the density of $\dot{\mathbf{Y}}$ at $\mathbf{0}$ conditional on $Y = y$ and $e' T^{-1} e$. Following the same arguments as in the proof of Theorem 6.3.1 of Adler (1981), page 134, we can see that when $y \to 0$ then $\ddot{\mathbf{Y}}$ converges to $2(1+y)e' T^{-1} e$ and so

$$\mathrm{E}\left[\det(\ddot{\mathbf{Y}}^+)\,\middle|Y = y, e' T^{-1} e, \dot{\mathbf{Y}} = \mathbf{0}\right] = 2^N(1+y)^N \det(e' T^{-1} e)\{1 + O(y^{\frac{1}{2}})\}.$$

By Lemma 2.3.1(a) and Lemma 2.3.3,

$$\dot{\mathbf{Y}} = 2(1+y)y^{\frac{1}{2}}(e'T^{-1}e)^{\frac{1}{2}}v_2,$$

where $v_2$ is defined in Lemma 2.3.3 with the distribution in Lemma 2.3.3(b). Hence,

$$f_N(\mathbf{0}|y, e'T^{-1}e) = \frac{\Gamma(\frac{d}{2})}{\pi^{\frac{N}{2}}\Gamma(\frac{d-N}{2})}\{2(1+y)y^{\frac{1}{2}}\}^{-N}\det(e'T^{-1}e)^{-\frac{1}{2}}. \tag{5.21}$$

Note that by Lemma A.0.1 with $n = \nu+1$, $p = d$ and $k = N$,

$$\mathrm{E}(\det(e'T^{-1}e)^{\frac{1}{2}}) = \frac{\Gamma[(\nu+1-d)/2]\Gamma[(d+1)/2]}{\Gamma[(\nu+1+N-d)/2]\Gamma[(d-N+1)/2]}.$$

Therefore, putting these all together gives:

$$\begin{aligned} & \mathrm{E}_{e'T^{-1}e}\left\{\mathrm{E}\left[\det(\ddot{\mathbf{Y}}^{+})\,\middle|Y=y, e'T^{-1}e, \dot{\mathbf{Y}}=\mathbf{0}\right] f_N(\mathbf{0}|y, e'T^{-1}e)\right\} f(y) \\ = \quad & \frac{2^{d-N-1}\Gamma(\frac{\nu+1}{2})\Gamma(\frac{d+1}{2})}{\pi^{\frac{N+1}{2}}(d-N-1)!\Gamma(\frac{\nu+1+N-d}{2})} y^{\frac{d-N}{2}-1}\{1+O(y^{\frac{1}{2}})\}. \end{aligned}$$

Integrating over $y$ gives the second result. □

Suppose $Y(\mathbf{t})$ has a local maximum (minimum) at $\mathbf{0}$ with a height exceeding (below) $y$, then the above theorem could be used to find out the distribution of $Y(\mathbf{0})$ in the limit of $y \to \infty$ ($y \to 0$). Such a study is via the notion "conditioning in the ergodic sense", or "horizontal window conditioning" introduced by Kac and Slepian (1959) (see next chapter for further references). We will later refer this type of conditioning as "HW conditioning" and we shall adopt the notation $\|$ used by Aronowich and Adler (1988) to denote it.

**Theorem 2.5.2** *Given that the Hotelling's $T^2$ field $Y(\mathbf{t}), \mathbf{t} \in \mathrm{I\!R}^N$, has a local maximum (minimum) at $\mathbf{t} = \mathbf{0}$ with a height exceeding (below) $y$, then for $0 < v < 1$,*

$$\lim_{y\to\infty} \mathrm{P}\{Y(\mathbf{0}) > \frac{y}{1-v} \parallel Y(\mathbf{0}) > y; \text{maximum}\} = (1-v)^{\frac{\nu+1-d-N}{2}}, \quad \textit{for } \nu \geq d+N.$$
$$\lim_{y\to 0} \mathrm{P}\{Y(\mathbf{0}) < (1-v)y \parallel Y(\mathbf{0}) < y; \text{minimum}\} = (1-v)^{\frac{d-N}{2}}, \quad \textit{for } d > N.$$

*Proof.* Since $M_y^+(Y,C)$ is derived from an ergodic point process, we can apply the results of Sections 6.5 and 6.6 from Adler (1981) to conclude that

$$\mathrm{P}\{Y(\mathbf{0}) > \frac{y}{1-v} \parallel Y(\mathbf{0}) > y; \text{maximum}\} = \frac{\mathrm{E}\{M^+_{\frac{y}{1-v}}(Y,C)\}}{\mathrm{E}\{M_y^+(Y,C)\}}.$$

Then by Lemma 2.5.1,

$$\lim_{y\to\infty} \mathrm{P}\{Y(\mathbf{0}) > \frac{y}{1-v} \parallel Y(\mathbf{0}) > y; \text{maximum}\} = (1-v)^{\frac{\nu+1-d-N}{2}}.$$

Similar arguments will lead to the other result. □

As we pointed out before, if $d \geq 2$, the scaled Hotelling's $T^2$ field is not an $F$ random field although at each point, it obeys an $F_{d,\nu-d+1}$ distribution. However, the expected number of extremal points of the Hotelling's $T^2$ field and that of an $F$ field (Worsley (1994), Theorem 4.3) share a similar form. So it is no surprise that the above distribution would share the same form as that of the $F$ field (Chapter 3, Theorem 3.4.2).

## 2.6 The limiting conditional distribution of curvature at local extrema of the Hotelling's $T^2$ field

In this section, we shall determine the limiting conditional distribution of $\ddot{\mathbf{Y}}(\mathbf{t})$ given that $\mathbf{t} = \mathbf{0}$ is a local extremum with height $y$. Again, the appropriate approach is via HW conditioning.

**Theorem 2.6.1** *Given that the Hotelling's $T^2$ field $\nu Y(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$, has a local extremum at $\mathbf{0}$ with height $y$, then*

**(a)** *if $\mathbf{t} = \mathbf{0}$ is a local maximum, $y^{-1}(1+y)^{-1}\ddot{\mathbf{Y}} \to -2A^{-1}\mathbf{B}$ as $y \to \infty$, where* $A \sim \chi^2_{\nu-d-N+2}$, $\mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{I}, \nu-d+2)$.

**(b)** *if $\mathbf{t} = \mathbf{0}$ is a local minimum, $\ddot{\mathbf{Y}} \to 2\mathbf{B}$ as $y \to 0$, where $\mathbf{B} \sim$* Multi-$T^2(d+1, N, \nu+1)$.

*Proof.* Since $Y$ has a local extremum at $\mathbf{0}$ with height $y$, then by Lemma 2.3.1 we can write

$$\ddot{\mathbf{Y}} = 2(1+y)\{e'T^{-1}e - y^{\frac{1}{2}}(\xi'T^{-1}\xi)^{\frac{1}{2}}[e'T^{-\frac{1}{2}}f + f'T^{-\frac{1}{2}}e - H] - y(\xi'T^{-1}\xi)U\}$$

at $\mathbf{t} = \mathbf{0}$. If $\mathbf{t} = \mathbf{0}$ is a local maximum, then when $y \to \infty$, following the same arguments as in the proof of Theorem 6.3.1 of Adler (1981), page 134, we can see that $(1+y)^{-1}y^{-1}\ddot{\mathbf{Y}}$ converges to the negative definite matrix $-2(\xi'T^{-1}\xi)U$. So now we need to find the joint HW conditional distribution of $\xi'T^{-1}\xi$ and $U$ at $\mathbf{0}$ given that $\mathbf{0}$ is a local maximum of $Y(\mathbf{t})$ with height $y$. By applying the corollary in Section 5.1 and the results of Section 6.5 and 6.6 from Adler (1981), the joint HW conditional distribution of $\xi'T^{-1}\xi = v$ and $U = u$ at $\mathbf{t} = \mathbf{0}$ is:

$$\begin{aligned}
& f_{\xi'T^{-1}\xi,U}(v, u \| Y = y, \dot{\mathbf{Y}} = \mathbf{0}, \ddot{\mathbf{Y}} < \mathbf{0}) \\
= \quad & \mu^{-1}\mathrm{E}\left(|\det(\ddot{\mathbf{Y}}^{-})| \,\Big|\, Y = y, \dot{\mathbf{Y}} = \mathbf{0}, \xi'T^{-1}\xi = v, U = u\right) f(y, \mathbf{0}, v, u) \\
\to \quad & \mu^{-1}2^N y^N (1+y)^N v^N \det(u) f(y, \mathbf{0}, v, u),
\end{aligned}$$

where $f(y, \dot{\mathbf{y}}, v, u)$ is the joint distribution of $Y, \dot{\mathbf{Y}}, \xi'T^{-1}\xi, U$ and $\mu$ is the derivative at $y$ of the expected number of local maxima of the random field $Y$ above $s$ as a function of $s$ at a level $s = u$. By Lemma 2.3.1, $f(y, \mathbf{0}, v, u) = f_Y(y) f_{\xi'T^{-1}\xi}(v) f_U(u) f_N(\mathbf{0}|y, v)$ where $f_N(\mathbf{0}|y, v)$ is the density of $\dot{\mathbf{Y}}$ at $\mathbf{0}$ conditional on $Y = y, \xi'T^{-1}\xi = v$. From (5.20) in the proof of Theorem 2.5.1, $f_N(\mathbf{0}|y, v) \propto v^{-\frac{N}{2}}$. Hence as $y \to \infty$,

$$\begin{aligned}
& f_{\xi'T^{-1}\xi,U}(v, u \| Y = y, \dot{\mathbf{Y}} = \mathbf{0}, \ddot{\mathbf{Y}} < \mathbf{0}) \quad \propto \quad v^{\frac{N}{2}} \det(u) f_{\xi'T^{-1}\xi}(v) f(u) \\
\propto \quad & v^{-\frac{\nu-d-N+2}{2}-1} e^{\frac{1}{2v}} \det(u)^{\frac{\nu-d+2-N-1}{2}} \exp(-\frac{1}{2}\mathrm{tr}(u)),
\end{aligned}$$

where the constant of proportionality contains $y$. Hence (a).

If $\mathbf{0}$ is a local minimum, then when $y \to 0$, by the same kind of argument, we know that $\ddot{\mathbf{Y}}$ approaches $2e'T^{-1}e$ at $\mathbf{0}$. So now we need to find the HW conditional distribution of $e'T^{-1}e$ given that random field $Y$ has a local minimum at $\mathbf{0}$ with height

$y$. By applying the corollary in Section 5.1 and the results of Sections 6.5 and 6.6 from Adler (1981), the HW conditional distribution of $e'T^{-1}e = V_1$ at $\mathbf{t} = \mathbf{0}$ has the density

$$\begin{aligned}
& f_{e'T^{-1}e}(V_1 \| Y = y, \dot{\mathbf{Y}} = \mathbf{0}, \ddot{\mathbf{Y}} > \mathbf{0}) \\
= \ & \mu^{-1}\mathrm{E}\left(|\det(\ddot{\mathbf{Y}}^{+})|\, \middle| Y = y, \dot{\mathbf{Y}} = \mathbf{0}, e'T^{-1}e = V_1)f(y, \mathbf{0}, V_1\right) \\
\to \ & \mu^{-1}2^N(1+y)^N \det(V_1) f(y, \mathbf{0}, V_1),
\end{aligned}$$

where $f(y, \dot{\mathbf{y}}, V_1)$ is the joint density of $Y, \dot{\mathbf{Y}}$ and $e'T^{-1}e$, $\mu$ is the derivative at $y$ of the expected number of local minima of random field $Y$ below level $s$ as a function of $s$ at a level $s = u$. Since $f(y, \mathbf{0}, V_1) = f_Y(y) f_{e'T^{-1}e}(V_1) f(\mathbf{0}|y, V_1)$ and from (5.21),

$$f(\mathbf{0}|y, V_1) \propto \{2(1+y)y^{\frac{1}{2}}\}^{-N} \det(V_1)^{-\frac{1}{2}},$$

hence as $y \to 0$,

$$\begin{aligned}
f_{e'T^{-1}e}(V_1 \| Y = y, \dot{\mathbf{Y}} = \mathbf{0}, \ddot{\mathbf{Y}} > \mathbf{0}) \ & \propto \ \det(V_1)\phi(V_1)\det(V_1)^{-\frac{1}{2}} \\
& \propto \ \det(V_1)^{\frac{d-N}{2}} \det(I + V_1)^{-\frac{\nu+1+N}{2}},
\end{aligned}$$

where the proportionality contains $y$. Hence (b). □

Suppose $\Lambda$ is the covariance-variance matrix of the first derivative of the component Gaussian fields in the definition of the Hotelling's $T^2$ field $\nu Y^*(\mathbf{t})$. For simplicity, we have assumed $\Lambda = I$ in the above sections. The extension to the general $\Lambda$ is obvious under a simple change of coordinates $\mathbf{s} = \Lambda^{\frac{1}{2}}\mathbf{t}$. We can write $Y^*(\mathbf{t})$ as $Y(\mathbf{s})$ where the covariance-variance matrix of the first derivative of component Gaussian fields in $Y(\mathbf{s})$ is the identity. Denote $\rho_y(Y^*, C)$ and $\rho_y(Y, C)$ to be the EC density for $Y^*$ and $Y$ separately. We can prove,

$$\frac{\partial Y^*}{\partial \mathbf{t}} = \Lambda^{\frac{1}{2}}\frac{\partial Y}{\partial \mathbf{s}}, \qquad \frac{\partial^2 Y^*}{\partial \mathbf{t}^2} = \Lambda^{\frac{1}{2}}\frac{\partial^2 Y}{\partial \mathbf{s}^2}\Lambda^{\frac{1}{2}}$$

and so

$$\begin{aligned}
\rho_y(Y^*, C) &= \det(\Lambda)^{\frac{1}{2}} \rho_y(Y, C), \\
\mathrm{E}(M_y(Y^*, C)) &= \det(\Lambda)^{\frac{1}{2}} \mathrm{E}(M_y(Y, C))
\end{aligned}$$

and then we can adjust Theorem 2.4.1, 2.3.1, 2.6.1 accordingly.

## 2.7 Applications to brain deformation studies

An experiment was carried out to study brain structure differences between males and females. There are 52 subjects in the experiment which produces 52 3-D deformation fields. Among the subjects, 29 are males and 23 are females. This gives a Hotelling's $T^2$ field with $d = 3$ components and $\nu = 52 - 2 = 50$ degrees of freedom under the null hypothesis. The Hotelling's $T^2$ field is then scaled to an $F$ distribution at each point by multiplication by the factor $(\nu - d + 1)/d/\nu$ and the excursion set of the scaled Hotelling's $T^2$ field above level 4 is shown in Figure B-9.

The unitless factor $\lambda(C) \det(\Lambda)^{\frac{1}{2}}$ was estimated to be 913, using methods of Worsley (1996a) and the global maximum of the scaled Hotelling's $T^2$ field was found to be 11.89. Our result for P-value approximation was applied and no significant structural difference was found at 0.05 level. We have also plotted the expected EC under the null hypothesis against the observed EC which is shown in Figure B-10. Although the global maximum is not significant ($0.05 <$ P-value $< 0.10$), the observed EC appears to be somewhat larger than the expected EC (Figure B-10). Inspection of Figure B-10 shows that this can be attributed to large deformations in the frontal area, anterior cerebellum and the superior central region. These are all possibly due to deformations outside the brain: the frontal difference can be attributed to more pronounced eye-brows in males, the anterior cerebellum deformations appear to be due to thicker necks in male. The superior central deformations may be an artifact due to poor sampling near the top of the brain.

# Chapter 3

# The size of the connected components of the excursion set of $\chi^2$, $t$ and $F$ fields

## 3.1 Introduction

In typical activation studies of brain function with positron emission tomography (PET), a 3-D image representing the difference of the brain activity under baseline and stimulation conditions was obtained. The purpose of these studies is to see which areas of the brain show an increase in brain activity, or "activation", due to the stimulation condition. In most of the cases, the activation is so slight that the experiment must be repeated on several subjects and the differenced images are averaged to improve the signal-to-noise ratio. The averaged image is then standardized to have unit variance and then searched for significant activation.

If the variance of the differenced images is stationary across the brain, then the averaged image after standardization is a Gaussian-statistic image (see equation 2.7). Worsley et al. (1992) have modelled the Gaussian-statistic image as a stationary Gaussian random field under the hypothesis of no activation.

However, the stationarity of the variance of the differenced images is questioned by some workers. If the variance at each location is arbitrary, a $t$ statistic image could be used to test instead. Under the hypothesis of no activation, the $t$ statistic image can be modelled as a $t$ stationary random field (Worsley et al. 1992, Worsley 1994). Therefore, the assessment of significant activations can be done based on the P-value of the global maximum of the $t$ random field. In more general hypothesis testing problems, a $\chi^2$ or $F$ statistic image might arise from the test for coefficients of a linear model as explained in section 1.2.1. This can also be modelled as a $\chi^2$ or $F$ random field under the hypothesis of no brain functional change and hence the P-value of the global maximum of the corresponding field would be required.

While in typical activation studies, the global maximum is chosen as the test statistic for significant activations, some workers suggested using the size of the largest connected component of the excursion set above a relatively high threshold as an alternative test statistic, based on the assumption that areas of true neural activity will tend to stimulate signal changes over contiguous voxels. Based on the spatial extent of the activation foci, Friston et al. (1994) presented an approximate analysis for testing the activations. Although their analysis was principally intended for images modelled as a Gaussian random field, they suggested that for some non-Gaussian fields, like the $t$ field, their approach could still be applied after appropriate "Gaussianization", i.e. point-wise transformation to a Gaussian distribution. If $T(\mathbf{t})$ is a $t$ random field with $\nu$ degrees of freedom, then the "Gaussianized" $t$ field is defined as $\Phi^{-1}(F(T(\mathbf{t}))$, where $\Phi$ is the Gaussian cumulative density function and $F$ is the cumulative density function of $t_\nu$. Although this approach could serve as an approximation for $t$ field with reasonably high degrees of freedom, its major problem lies in the fact that the Gaussianized field is *not* a Gaussian field.

Nosko (1969) and Adler (1981) studied the local maxima and high-level excursions of a stationary Gaussian field extensively. A result from Nosko (1969) states that at high level local maxima, the stationary Gaussian random field has a deterministic quadratic shape or form. He then obtained an asymptotic distribution of the size of one connected component of a high level excursion set for the stationary Gaussian field, which is reported in Adler (1981). Some extensions to the $\chi^2$ field in 1-D and

2-D case were done by Aronowich and Adler (1986, 1988), where the structure of the field at local extrema was investigated.

This chapter aims to study the behavior of a $\chi^2, t$ or $F$ field at local extrema above high (or below low) levels, which will lead to a limiting distribution of the size of the connected components of the excursion set. At the end, the theory will be applied to detect significant activations in images from $\chi^2, t$ and $F$ fields.

The overall plan is the following. In section 3.2 we shall give the definition of $\chi^2, t$ and $F$ fields and the representations of their derivatives from Worsley (1994). In section 3.3, we shall find the limiting conditional distribution of the curvature at a local extremum of these fields. In section 3.4, we shall first state the expected number of extremal points for $\chi^2, t$ and $F$ fields from Worsley (1994) and then apply the results to find the limiting distribution of the height of the local extrema. In section 3.5, we will focus on the structure of the surface near a local extremum for $\chi^2, t$ and $F$ fields. In section 3.6, the limiting distribution of the size of one connected component of the excursion set will be found. In Section 3.7 and 3.8, the limiting distribution of the size of the largest connected component of the excursion set will be discussed and validated by simulations. And finally, in section 3.9 we will apply this work to some PET images from a study on pain perception.

## 3.2 Definitions of $\chi^2, t$ and $F$ fields and the representations of their derivatives

More notations are needed at this point besides what were introduced in the previous chapters. We shall use the notation Exponential($\mu$) to represent the exponential distribution with expectation $\mu$, Beta($\alpha, \beta$) to represent a beta distribution with parameter $\alpha$ and $\beta$. Recall that $t_\nu$ represents the Student $t$ distribution with $\nu$ degrees of freedom.

### 3.2.1 $\chi^2$ field

Let $X_1(\mathbf{t}), \ldots, X_n(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$, be independent, identically distributed, stationary, real-valued, zero mean, unit-variance Gaussian random fields and $\mathbf{\Lambda} = \mathrm{Var}(\dot{\mathbf{X}}_i),\ i = 1, \ldots, n$. Then Adler(1981), page 169, defines the $\chi^2$ field $U(\mathbf{t})$ as

$$U(\mathbf{t}) = \sum_{i=1}^{n} X_i(\mathbf{t})^2, \quad \mathbf{t} \in \mathbb{R}^N. \tag{2.1}$$

The definition of $\chi^2$ field is restricted to $n \geq N+1$ to avoid exact zeroes as discussed in Section 2.3 (Worsley, 1994). Worsley (1994) has the following representation of variates $\dot{\mathbf{U}}$ and $\ddot{\mathbf{U}}$ in terms of independent random variables.

**Lemma 3.2.1** *We can write the first two derivatives of $U = U(\mathbf{t})$ as follows, where $\stackrel{D}{=}$ is equality in law:*

*(a)* $\dot{\mathbf{U}} \stackrel{D}{=} 2U^{\frac{1}{2}}\mathbf{z}$,

*(b)* $\ddot{\mathbf{U}} \stackrel{D}{=} 2(\mathbf{P} + \mathbf{z}\mathbf{z}' - U\mathbf{\Lambda} + U^{\frac{1}{2}}\mathbf{H})$,

*where* $U \sim \chi^2_n$, $\mathbf{z} \sim \mathrm{Normal}_N(\mathbf{0}, \mathbf{\Lambda})$, $\mathbf{P} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, n-1)$ *and* $\mathbf{H} \sim \mathrm{Normal}_{N\times N}$ $(\mathbf{0}, \mathbf{M}(\mathbf{\Lambda}))$*, all independently.*

The definition of $\mathbf{M}(\mathbf{\Lambda})$ can be found in section 2.2  2.2 when we replace $\mathbf{I}$ by $\mathbf{\Lambda}$.

### 3.2.2 $t$ field

Let $X(\mathbf{t}), Y_1(\mathbf{t}), \ldots, Y_m(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^N$, be independent, identically distributed, stationary, real-valued Gaussian random fields with zero mean, unit variance, and $\mathbf{\Lambda} = \mathrm{Var}(\dot{\mathbf{X}}(\mathbf{t})) = \mathrm{Var}(\dot{\mathbf{Y}}_i(\mathbf{t})), i = 1, \ldots, m$. Worsley (1994) defines the $t$ field as

$$T(\mathbf{t}) = X(\mathbf{t}) \Big/ \left\{ \sum_{i=1}^{m} Y_i(\mathbf{t})^2/m \right\}^{\frac{1}{2}}. \tag{2.2}$$

The definition of $t$ field is restricted to $m \geq N$ to avoid exact zeroes as discussed in Section 2.3 (Worsley, 1994). Worsley (1994) has the following Lemma about representation of $\dot{\mathbf{T}}$ and $\ddot{\mathbf{T}}$ in terms of independent variables.

**Lemma 3.2.2** *We can express the first and second derivatives of $T = T(\mathbf{t})$ in terms of independent random variables as follows:*

(a) $\dot{\mathbf{T}} \stackrel{D}{=} m^{\frac{1}{2}}(1+T^2/m)S^{-\frac{1}{2}}\mathbf{z}_1,$

(b) $\ddot{\mathbf{T}} \stackrel{D}{=} m^{\frac{1}{2}}(1+T^2/m)S^{-1}\{-m^{-\frac{1}{2}}T(\mathbf{Q}-2\mathbf{z}_1\mathbf{z}'_1)-\mathbf{z}_1\mathbf{z}'_2-\mathbf{z}_2\mathbf{z}'_1+S^{\frac{1}{2}}\mathbf{H}\},$

*where* $T \sim \mathrm{t}_m$, $S \sim \chi^2_{m+1}$, $\mathbf{z}_1, \mathbf{z}_2 \sim \mathrm{Normal}_N(\mathbf{0}, \mathbf{\Lambda})$, $\mathbf{Q} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, m-1)$ *and* $\mathbf{H} \sim \mathrm{Normal}_{N\times N}(\mathbf{0}, \mathbf{M}(\mathbf{\Lambda}))$, *all independently.*

### 3.2.3 $F$ field

Let $X_1(\mathbf{t}), \cdots, X_n(\mathbf{t}), Y_1(\mathbf{t}), \cdots, Y_m(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$, be i.i.d. stationary, real-valued Gaussian random fields with zero mean, unit variance, and $\mathrm{Var}(\dot{\mathbf{X}}_i(\mathbf{t})) = \mathrm{Var}(\dot{\mathbf{Y}}_j(\mathbf{t})) = \mathbf{\Lambda}$, $i = 1, \cdots, n$, $j = 1, \cdots, m$. Worsley (1994) defines the $F$ field as

$$F(\mathbf{t}) = \left\{\sum_{i=1}^{n} X_i(\mathbf{t})^2/n\right\} \Big/ \left\{\sum_{i=1}^{m} Y_i(\mathbf{t})^2/m\right\}. \tag{2.3}$$

The definition is restricted to $m+n \geq N+1$ to avoid exact zeroes as discussed in Section 2.3 (Worsley, 1994). Define $G(\mathbf{t}) = (n/m)F(\mathbf{t})$, then Worsley (1994) has the following representation of $\dot{\mathbf{G}}$ and $\ddot{\mathbf{G}}$.

**Lemma 3.2.3** *We can express the first and second derivatives of $G$ in terms of independent random variables as follows:*

(a) $\dot{\mathbf{G}} \stackrel{D}{=} 2G^{\frac{1}{2}}(1+G)W^{-\frac{1}{2}}\mathbf{z}_1$

(b) $\ddot{\mathbf{G}} \stackrel{D}{=} 2(1+G)[W^{-1}\{\mathbf{P}-G\mathbf{Q}+(1+3G)\mathbf{z}_1\mathbf{z}'_1-G^{\frac{1}{2}}(\mathbf{z}_1\mathbf{z}'_2+\mathbf{z}_2\mathbf{z}'_1)\}+G^{\frac{1}{2}}W^{-\frac{1}{2}}\mathbf{H}]$

*where* $(m/n)G \sim F_{n,m}$, $W \sim \chi^2_{m+n}$, $\mathbf{z}_1, \mathbf{z}_2 \sim \mathrm{Normal}_N(\mathbf{0}, \mathbf{\Lambda})$, $\mathbf{P} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, n-1)$, $\mathbf{Q} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, m-1)$ *and* $\mathbf{H} \sim \mathrm{Normal}_{N\times N}(\mathbf{0}, \mathbf{M}(\mathbf{\Lambda}))$, *all independently.*

## 3.3 The limiting distribution of the curvature at a local extremum of the $\chi^2, t$ or $F$ field

Throughout the following sections we shall assume that we are dealing with $N$-dimensional, real-valued, stationary, ergodic Gaussian fields, with mean zero, unit variance and satisfying the condition

$$\max_{i,j} \mathrm{E}\{|X^{(ij)}(\mathbf{t}) - X^{(ij)}(\mathbf{0})|^2\} \leq a\|\mathbf{t}\|^2 \tag{3.4}$$

for some finite $a > 0$ and all $\mathbf{t}$ with $\mathbf{t}$ small enough. The reason for imposing this condition is to provide 'adequate' regularity conditions for $\chi^2, t$ and $F$ fields as are required in the future use of the Slepian model process and HW conditioning. For instance, one such regularity condition would be that the random field has continuous first- and second-order derivatives with probability one.

We shall determine, in this section, the limiting conditional distribution of the curvature of a $\chi^2, t$ or $F$ field, $\ddot{\mathbf{Y}}$, at $\mathbf{t} = \mathbf{0}$, given that $\mathbf{0}$ is a local extremum with height $u$. So we are conditioning on the event that $\dot{\mathbf{Y}}(\mathbf{0}) = \mathbf{0}, Y(\mathbf{0}) = u$, and $\ddot{\mathbf{Y}}(\mathbf{0}) < \mathbf{0}$ (local maximum) or $\ddot{\mathbf{Y}}(\mathbf{0}) > \mathbf{0}$ (local minimum). It is now well known that the appropriate framework for such a study is via the notion of 'horizontal window' (HW) conditioning introduced by Kac and Slepian (1959), and the so-called Slepian model process. A more detailed description about this approach will be given in the beginning of Section 5 when we consider the structure of the surface near the extremal points. We will adopt the notation $\|$ by Aronowich and Adler (1986, 1988) to denote conditioning in the ergodic (HW) sense and use $|$ to denote conditioning in the usual sense.

In what follows, we are going to adopt all the notations established in the previous section about the definition of $\chi^2, t$ and $F$ fields and their derivative representations unless otherwise stated.

**Theorem 3.3.1** *Given that a $\chi^2$ field $U(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$ has a local extremum at $\mathbf{0}$ with height $u$, if*

*(a) $\mathbf{t} = \mathbf{0}$ is a local maximum, then $u^{-1}\ddot{\mathbf{U}}$ converges to $-2\mathbf{\Lambda}$ as $u \to \infty$.*

*(b)* $\mathbf{t}=\mathbf{0}$ *is a local minimum, then* $\ddot{\mathbf{U}}$ *converges to* $2\times \mathrm{Wishart}_N(\mathbf{\Lambda}, n+1)$ *as* $u\to 0$.

*Proof.* Since the $\chi^2$ field $U(\mathbf{t})$ has a local extremum at $\mathbf{0}$ with height $u$, by Lemma 3.2.1, at $\mathbf{t}=\mathbf{0}$,

$$\ddot{\mathbf{U}} = 2(\mathbf{P} - u\mathbf{\Lambda} + u^{\frac{1}{2}}\mathbf{H}).$$

If $\mathbf{0}$ is a local maximum, then when $u\to\infty$, following the same arguments as in the proof of Theorem 6.3.1 of Adler (1981), page 134, we can see that $u^{-1}\ddot{\mathbf{U}}$ converges to the negative definite matrix $-2\mathbf{\Lambda}$.

If $\mathbf{t}=\mathbf{0}$ is a local minimum, then when $u\to 0$, by the same kind of arguments, we know $\ddot{\mathbf{U}}$ approaches to $2\mathbf{P}$. So we need to find the HW conditional distribution of $\mathbf{P}$ at $\mathbf{0}$ given that the $\chi^2$ field $U$ has a local minimum at $\mathbf{0}$ with height $u$. Under the regularity conditions, the HW conditional distribution of $\mathbf{P}$ has the density

$$\begin{aligned}
& f(\mathbf{p} \parallel U=u, \dot{\mathbf{U}}=\mathbf{0}, \ddot{\mathbf{U}}>\mathbf{0}) \\
= \ & \mu^{-1}\mathrm{E}(|\det(\ddot{\mathbf{U}}^{+})| \,\Big|\, U=u, \dot{\mathbf{U}}=\mathbf{0}, \mathbf{P}=\mathbf{p}) f(u,\mathbf{0},\mathbf{p}) \\
\to \ & \mu^{-1}\det(2\mathbf{p})\phi(u,\mathbf{0},\mathbf{p}),
\end{aligned}$$

where $f(u,\mathbf{0},\mathbf{p})$ is the joint density of $U, \dot{\mathbf{U}}$ and $\mathbf{P}$, and $\mu$ is the derivative of the expected number of local minima below $s$ as a function of $s$ at level $s=u$.

Since variate $\mathbf{P}$ is independent of $U, \dot{\mathbf{U}}$ by Lemma 3.2.1, therefore

$$f(u,\mathbf{0},\mathbf{p}) = f(u,\mathbf{0})f(\mathbf{p}).$$

Hence when $u\to 0$,

$$\begin{aligned}
f(\mathbf{p} \parallel U=u, \dot{\mathbf{U}}=\mathbf{0}, \ddot{\mathbf{U}}>\mathbf{0}) &\propto \det(\mathbf{p})\phi(\mathbf{p}) \\
&\propto \det(\mathbf{p})^{\frac{n+1-N-1}{2}}\exp(\mathrm{tr}(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{p})),
\end{aligned}$$

where the proportionality contains $u$. Thus at $\mathbf{t}=\mathbf{0}$, $\mathbf{P}\to\mathrm{Wishart}_N(\mathbf{\Lambda}, n+1)$, which proves (b). $\square$

**Theorem 3.3.2** *Given that a $t$ field $T(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$, has a local maximum at $\mathbf{0}$ with height $t$, then at $\mathbf{t} = \mathbf{0}$ when $t \to \infty$, the limiting HW conditional distribution of $\ddot{\mathbf{T}}$ has the property that:*

$$-t^{-1}(1 + t^2/m)^{-1}\ddot{\mathbf{T}} \text{ converges to } A^{-1}\mathbf{B},$$

*where $A \sim \chi^2_{m+1-N}, \mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, m+1)$ independently.*

*Proof.* Since the $t$ field $T(\mathbf{t})$ has a local maximum at $\mathbf{0}$ with height $u$, by lemma 3.2.2, we can write

$$\ddot{\mathbf{T}} = -t(1 + t^2/m)S^{-1}\mathbf{Q} + m^{\frac{1}{2}}(1 + t^2/m)S^{-\frac{1}{2}}\mathbf{H}.$$

Following the same arguments as in the proof of Theorem 6.3.1 of Adler (1981), page 134, we can see as $t \to \infty$, $\ddot{\mathbf{T}}$ converges to the negative definite matrix $-t(1 + t^2/m)S^{-1}\mathbf{Q}$. So we need to find the joint HW conditional distribution of of $S$ and $\mathbf{Q}$ given that the $t$ field $T(\mathbf{t})$ has a local maximum at $\mathbf{0}$ with height $t$. The joint HW conditional distribution of $S$ and $\mathbf{Q}$ is:

$$\begin{aligned}
& f(s, \mathbf{q} \| T = t, \dot{\mathbf{T}} = \mathbf{0}, \ddot{\mathbf{T}} < \mathbf{0}) \\
= \;& \mu^{-1}\mathrm{E}(|\det(\ddot{\mathbf{T}}^{-})| \,\big|\, T = t, \dot{\mathbf{T}} = \mathbf{0}, S = s, \mathbf{Q} = \mathbf{q}) f(t, \mathbf{0}, s, \mathbf{q}) \\
\to \;& \mu^{-1}t^N(1 + t^2/m)^N s^{-N} \det(\mathbf{q}) f(t, \mathbf{0}, s, \mathbf{q}) \text{ as } t \to \infty,
\end{aligned}$$

where $f(t, \mathbf{0}, s, \mathbf{q})$ is the joint density of $T, \dot{\mathbf{T}}, S, \mathbf{Q}$. By Lemma 3.2.2,

$$f(t, \mathbf{0}, s, \mathbf{q}) = f(t)f(s)f(\mathbf{q})f(\mathbf{0}|t, s)$$

and conditional on $T$ and $S$,

$$\dot{\mathbf{T}} \sim \mathrm{Normal}_N(\mathbf{0}, mS^{-1}(1 + T^2/m)^2\mathbf{\Lambda}),$$

so the conditional density of $\dot{\mathbf{T}} = \mathbf{0}$ conditional on $T = t, S = s$ is

$$f(\mathbf{0}|t,s) = \{2\pi m(1+t^2/m)^2 s^{-1}\}^{-\frac{1}{2}N} \det(\mathbf{\Lambda})^{-\frac{1}{2}} \propto s^{\frac{1}{2}N}.$$

Hence as $t \to \infty$,

$$\begin{aligned} & f(s,\mathbf{q}\|T=t,\dot{\mathbf{T}}=\mathbf{0},\ddot{\mathbf{T}}<\mathbf{0}) \propto s^{-N}\det(\mathbf{q})f(s)f(\mathbf{q})s^{\frac{1}{2}N} \\ & \propto s^{\frac{m+1-N}{2}-1}\exp(-\frac{s}{2})\det(\mathbf{q})^{\frac{m+1-N-1}{2}}\exp(\mathrm{tr}(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{q})). \end{aligned}$$

This proves the theorem. □

**Theorem 3.3.3** *Given that an $F$ field $F(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^N$, has a local extremum at $\mathbf{0}$ with height $f$, then*

*(a) if $\mathbf{t} = \mathbf{0}$ is a local maximum, $f^{-2}\ddot{\mathbf{F}}$ converges to $(-2n/m)A^{-1}\mathbf{B}$ as $f \to \infty$, where $A \sim \chi^2_{m+n-N}$ and $\mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, m+1)$ independently.*

*(b) if $\mathbf{t} = \mathbf{0}$ is a local minimum, $\ddot{\mathbf{F}}$ converges to $(2m/n)A^{-1}\mathbf{B}$, as $f \to 0$, where $A \sim \chi^2_{m+n-N}$ and $\mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, n+1)$ independently.*

*Proof.* Since the $F$ field $F(\mathbf{t})$ has a local extremum at $\mathbf{0}$, if we define $G(\mathbf{t}) = (n/m)F(\mathbf{t})$, then by the Lemma 3.2.3, at $\mathbf{t} = \mathbf{0}$,

$$\ddot{\mathbf{G}} = 2(1+g)[W^{-1}(\mathbf{P} - g\mathbf{Q}) + g^{\frac{1}{2}}W^{-\frac{1}{2}}\mathbf{H}],$$

where $g = (n/m)f$. Following the same arguments as in the proof of Theorem 6.3.1 of Adler (1981), page 134, we can see that as $f \to \infty$ then $g^{-2}\ddot{\mathbf{G}}$ converges to the negative definite matrix $-2W^{-1}\mathbf{Q}$. So we need to find the joint HW conditional distribution of $W$ and $\mathbf{Q}$ at $\mathbf{0}$ given that the $F$ field $F(\mathbf{t})$ has a local maximum at $\mathbf{0}$ with height $f$, which is the same as the field $G$ has a local maximum at $\mathbf{0}$ with height $g$. The joint HW conditional density of $W$ and $\mathbf{Q}$ at $\mathbf{0}$ is:

$$f(w,\mathbf{q}\|G=g,\dot{\mathbf{G}}=\mathbf{0},\ddot{\mathbf{G}}<\mathbf{0})$$

$$
\begin{aligned}
&\propto \ \mathrm{E}(|\det(\ddot{\mathbf{G}}^{-})|\,\Big|G=g,\dot{\mathbf{G}}=\mathbf{0},W=w,\mathbf{Q}=\mathbf{q})f(g,\mathbf{0},w,\mathbf{q})\\
&\propto \ \mathrm{E}(\det(w^{-1}\mathbf{q}))f(g,\mathbf{0},w,\mathbf{q}) \ \text{ as } g\to\infty\\
&\propto \ w^{-N}\det(\mathbf{q})f(g,\mathbf{0},w,\mathbf{q}) \ \text{ as } g\to\infty,
\end{aligned}
$$

where $f(g,\mathbf{0},w,\mathbf{q})$ is the joint density of $G,\dot{\mathbf{G}},W,\mathbf{Q}$. By Lemma 3.2.3,

$$
f(g,\mathbf{0},w,\mathbf{q})=f(g)f(w)f(\mathbf{q})f(\mathbf{0}|g,w)
$$

and conditional on $G$ and $W$,

$$
\dot{\mathbf{G}}\sim \mathrm{Normal}_N(\mathbf{0},4G(1+G)^2W^{-1}\mathbf{\Lambda}),
$$

so

$$
f(\mathbf{0}|g,w)=\{8\pi g(1+g)^2w^{-1}\}^{-\frac{1}{2}N}\det(\mathbf{\Lambda})^{-\frac{1}{2}}\propto w^{\frac{1}{2}N}.
$$

Hence as $f\to\infty$,

$$
\begin{aligned}
&f(w,\mathbf{q}\|G=g,\dot{\mathbf{G}}=\mathbf{0},\ddot{\mathbf{G}}<0)\propto w^{-N}\det(\mathbf{q})f(w)f(\mathbf{q})w^{\frac{1}{2}N}\\
&\propto w^{\frac{m+n-N}{2}-1}\exp(-\frac{w}{2})\det(\mathbf{q})^{\frac{m+1-N-1}{2}}\exp(\mathrm{tr}(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{q})).
\end{aligned}
$$

This proves (a). Notice that when $f\to 0$, $\ddot{\mathbf{G}}$ converges to $2W^{-1}\mathbf{P}$. Similar arguments will lead to (b). $\square$

## 3.4 The expected number of extremal points of $\chi^2, t$ and $F$ fields

Recall that $M_u^+(Y,C)$ is the number of local maxima of a random field $Y(\mathbf{t})$ greater than $u$ inside a compact subset $C$, and $M_u^-(Y,C)$ is the number of local minima less than $u$ inside $C$. Worsley (1994) gave us the following lemma about the asymptotic expectation of $M_u^+(Y,C)$ and $M_u^-(Y,C)$ of a $\chi^2$ field $U$, $t$ field $T$ and $F$ field $F$. From now on we shall assume that each of the component Gaussian fields satisfy the regularity conditions of Theorem 1.3.1.

**Lemma 3.4.1** *(a) for the $\chi^2$ field $U$,*

$$
\begin{aligned}
\mathrm{E}(M_u^+(U,C)) &= \frac{\lambda(C)\det(\mathbf{\Lambda})^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}N}2^{\frac{1}{2}(n-2)}\Gamma(\frac{1}{2}n)} u^{\frac{1}{2}(n+N)-1}\exp(-\frac{1}{2}u)\{1+O(u^{-\frac{1}{2}})\},\\
\mathrm{E}(M_u^-(U,C)) &= \frac{\lambda(C)\det(\mathbf{\Lambda})^{\frac{1}{2}}(n-1)!}{(2\pi)^{\frac{1}{2}N}2^{\frac{1}{2}(n-2)}\Gamma(\frac{1}{2}n)(n-N)!} u^{\frac{1}{2}(n-N)}\{1+O(u^{\frac{1}{2}})\},\ \textit{for } n>N;
\end{aligned}
$$

*(b) for the $t$ field $T$, $m>N$,*

$$
\mathrm{E}(M_t^+(T,C)) = \frac{\lambda(C)\det(\mathbf{\Lambda})^{\frac{1}{2}}\Gamma(\frac{1}{2}(m+1))}{2(\pi)^{\frac{1}{2}(N+1)}\Gamma(\frac{1}{2}(m+2-N))}(\frac{t^2}{m})^{-\frac{1}{2}(m-N)}\{1+O(t^{-1})\};
$$

*(c) for the $F$ field $F$, $m>N$,*

$$
\begin{aligned}
\mathrm{E}(M_f^+(F,C) &= \frac{\lambda(C)\det(\mathbf{\Lambda})^{\frac{1}{2}}\Gamma(\frac{1}{2}(m+n-N))(m-1)!}{(2\pi)^{\frac{1}{2}N}2^{\frac{1}{2}(N-2)}\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)(m-N)!}\\
&\quad\times(\frac{nf}{m})^{-\frac{1}{2}(m-N)}\{1+O(f^{-\frac{1}{2}})\};
\end{aligned}
$$

*where $\lambda(C)$ is the Lebesgue measure of $C$.*

The above results, aside from being of intrinsic interest, can also be applied to give the limiting distribution of the height of local maximum (or minimum) above high (or low) threshold.

**Theorem 3.4.2**

*(a) Given that a $\chi^2$ field $U(\mathbf{t})$, $\mathbf{t}\in\mathbb{R}^N$ has a local maximum (or minimum) at $\mathbf{t}=\mathbf{0}$ with a height exceeding (or below) $u$, then*

$$
\begin{aligned}
\lim_{u\to\infty}\mathrm{P}\{U(\mathbf{0})-u>v \parallel U(\mathbf{0})>u;\text{maximum}\} &= \exp(-\frac{1}{2}v),\quad v>0;\\
\lim_{u\to 0}\mathrm{P}\{u-U(\mathbf{0})>uv \parallel U(\mathbf{0})<u;\text{minimum}\} &= (1-v)^{\frac{1}{2}(n-N)},
\end{aligned}
$$

*for $0<v<1$ and $n>N$.*

(*b*) *Given that a t field* $T(\mathbf{t})$, $\mathbf{t} \in \mathrm{I\!R}^N$ *and* $m > N$, *has a local maximum at* $\mathbf{t} = \mathbf{0}$ *with a height exceeding* $t$, *then for* $0 < v < 1$

$$\lim_{t\to\infty} \mathrm{P}\{T(\mathbf{0}) > \frac{t}{\sqrt{1-v}} \parallel T(\mathbf{0}) > t; \text{maximum}\} = (1-v)^{\frac{1}{2}(m-N)}.$$

(*c*) *Given that an F field* $F(\mathbf{t})$, $\mathbf{t} \in \mathrm{I\!R}^N$ *and* $m > N$ *has a local maximum at* $\mathbf{t} = \mathbf{0}$ *with a height exceeding* $f$, *then for* $0 < v < 1$

$$\lim_{f\to\infty} \mathrm{P}\{F(\mathbf{0}) > \frac{f}{1-v} \parallel F(\mathbf{0}) > f; \text{maximum}\} = (1-v)^{\frac{1}{2}(m-N)}.$$

*Proof.* Since $M_u^+(U,C)$ and $M_u^-(U,C)$ is derived from an ergodic point process, we can apply the results of Sections 6.5 and 6.6 from Adler (1981) to conclude that

$$\mathrm{P}\{U(\mathbf{0}) - u > v \parallel U(\mathbf{0}) > u; \text{maximum}\} = \frac{\mathrm{E}\{M_{u+v}^+(U,C)\}}{\mathrm{E}\{M_u^+(U,C)\}}.$$

Then by Lemma 3.4.1,

$$\begin{aligned} & \lim_{u\to\infty} \mathrm{P}\{U(\mathbf{0}) - u > v \parallel U(\mathbf{0}) > u; \text{maximum}\} \\ = \quad & \lim_{u\to\infty} \frac{(u+v)^{\frac{1}{2}(n+N)-1}\exp(-\frac{1}{2}(u+v))}{u^{\frac{1}{2}(n+N)-1}\exp(-\frac{1}{2}u)} = \exp(-\frac{1}{2}v). \end{aligned}$$

By the same kind of arguments, we can prove (b),(c) by using Lemma 3.4.1(b),(c).

□

## 3.5 The structure of $\chi^2, t$ or $F$ field near local extrema

Suppose $Y(\mathbf{t})$ is a real-valued, stationary, ergodic, $N$-dimensional random field. If we are interested in the behavior of $Y(\mathbf{t})$ in the neighbourhood of an extremum with height $u$, then we need to investigate the *conditional field*

$$Y_u(\mathbf{t}) = Y(\mathbf{t})|A, \tag{5.5}$$

where $A$ denotes the event that $Y$ has a local extremum with height $u$ at $\mathbf{t} = \mathbf{0}$. The right tool is to use the ergodic-sense Slepian model, based on horizontal-vertical window conditioning (see for example Lindgren (1972)). That is, if we wish to investigate the properties of the 'conditioned surface' $Y(\mathbf{t})$, given an extremum at the origin of height $u$, we have to approach this conditioning event by the series of events

$$\begin{aligned} A(h, h') \quad = \quad & \{Y(\mathbf{t}) \text{ has an extremum of height in } (u, u+h) \\ & \text{at a point in the circle } \|\mathbf{t}\| < h'\}. \end{aligned}$$

Denote by $Y_u(\mathbf{t})$ the surface with the distribution

$$\mathrm{P}\{Y_u(\mathbf{t}) \in B\} = \lim_{h \downarrow 0} \lim_{h' \downarrow 0} \mathrm{P}\{Y(\mathbf{t}) \in B | A(h, h')\},$$

where $B$ is any set in the $\sigma$-field. Making use of ergodicity, we can rewrite the above probability as a ratio of expectations. In particular, Adler (1981) (page 152) gave the following lemma about *the finite-dimensional (fi-di)* distributions for the conditional field $Y_u(\mathbf{t})$.

**Lemma 3.5.1** *Let $\mathbf{t}^1, \ldots, \mathbf{t}^n$ be any set of points in $\mathbb{R}^N$. Under the appropriate regularity conditions, the conditional distribution of $\mathbf{v} = (Y(\mathbf{t}^1), \ldots, Y(\mathbf{t}^n))$, given that $Y$ has a local extremum at $\mathbf{0}$ with height $u$, has the density*

$$\beta(\mathbf{v}, u) = \frac{\int_D |\det \ddot{\mathbf{y}}| \phi(u, \mathbf{0}, \ddot{\mathbf{y}}, \mathbf{v}) d\ddot{\mathbf{y}}}{\int_D |\det \ddot{\mathbf{y}}| \phi(u, \mathbf{0}, \ddot{\mathbf{y}}) d\ddot{\mathbf{y}}}, \tag{5.6}$$

*where $D$ is the region such that $\{\ddot{\mathbf{y}} > \mathbf{0}\}$ or $\{\ddot{\mathbf{y}} < \mathbf{0}\}$ depending on whether conditioning is on a maximum or minimum, $\phi(u, \mathbf{0}, \ddot{\mathbf{y}}, \mathbf{v})$ is the joint density of $Y = u, \dot{\mathbf{Y}} = \mathbf{0}, \ddot{\mathbf{Y}} = \ddot{\mathbf{y}}$ and $\mathbf{v}$, $\phi(u, \mathbf{0}, \ddot{\mathbf{y}})$ is the joint density of $Y = u, \dot{\mathbf{Y}} = \mathbf{0}$ and $\ddot{\mathbf{Y}} = \ddot{\mathbf{y}}$.*

*Furthermore, (5.6) is equivalent to*

$$\beta(\mathbf{v}, u) = \int_D \psi_u(\ddot{\mathbf{y}})\phi(\mathbf{v}|u, \mathbf{0}, \ddot{\mathbf{y}})d\ddot{\mathbf{y}},$$

*where*

$$\psi_u(\ddot{\mathbf{y}}) = \begin{cases} \frac{|\det \ddot{\mathbf{y}}|\phi(\ddot{\mathbf{y}}|u,\mathbf{0})}{\int_D |\det \ddot{\mathbf{y}}|\phi(\ddot{\mathbf{y}}|u,\mathbf{0})d\ddot{\mathbf{y}}} & \text{if } \ddot{\mathbf{y}} \in D \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

*is the HW conditional density of* $\ddot{\mathbf{Y}}(\mathbf{0})$ *given that* $Y$ *has an extremum at* $\mathbf{0}$ *with height* $u$*,* $\phi(\mathbf{v}|u, \mathbf{0}, \ddot{\mathbf{y}})$ *is the density of* $Y(\mathbf{t}^1), \ldots, Y(\mathbf{t}^n)$ *conditional on* $Y(\mathbf{0}) = u$*,* $\dot{\mathbf{Y}}(\mathbf{0}) = \mathbf{0}$*,* $\ddot{\mathbf{Y}}(\mathbf{0}) = \ddot{\mathbf{y}}$*.*

Now we will restrict our attention on $\chi^2, t$ and $F$ fields. Suppose $Y(\mathbf{t})$ is one of these fields, and we impose the regularity conditions mentioned in the beginning of Section 3.3 on each of the component Gaussian fields in the definitions of $\chi^2, t$ and $F$ fields. If we can specify $\psi_u(\ddot{\mathbf{y}})$ and $\phi(\mathbf{v}|u, \mathbf{0}, \ddot{\mathbf{y}})$, then Lemma 3.5.1 can be used to find the fi-di distributions of the conditional field $Y_u$. Note that each of the component fields is a Gaussian random field, and Adler (1981) in Section 6.7 has worked out the form of $\phi(\mathbf{v}|u, \mathbf{0}, \ddot{\mathbf{y}})$. His result can be extended as follows after some complicated calculations.

**Lemma 3.5.2** *Let* $X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N$ *be one of the component Gaussian random fields in the definition of the* $\chi^2, t$ *or* $F$ *field; then conditional on* $X(\mathbf{0}), \dot{\mathbf{X}}(\mathbf{0})$ *and* $\ddot{\mathbf{X}}(\mathbf{0})$*,* $X(\mathbf{t})$ *has the same distribution as*

$$X(\mathbf{0}) + \dot{\mathbf{X}}'(\mathbf{0})\mathbf{t} + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{X}}(\mathbf{0})\mathbf{t} + \epsilon(\mathbf{t}^i) + K(\mathbf{t}),\ i = 1, \ldots, n, \tag{5.8}$$

*where* $\epsilon(\mathbf{t}) = m\ o(\|\mathbf{t}\|^2)$ *is deterministic with* $m = \max\{|X(\mathbf{0})|, \|\dot{\mathbf{X}}(\mathbf{0})\|, \|\ddot{\mathbf{X}}(\mathbf{0})\|\}$ *and* $K(\mathbf{t})$ *is normal with mean* $\mathbf{0}$ *and covariance matrix of* $o(\|\mathbf{t}\|^4)$*.*

*Proof.* For simplicity, we shall prove (5.8) for the case $N = 2$. The proof for any other dimensions is similar. Suppose $\mathbf{\Lambda} = \mathrm{Var}(\dot{\mathbf{X}})$ with elements $\lambda_{ij}$ and $\rho(\mathbf{t})$ is the autocorrelation function of $X(\mathbf{t})$. By standard results, we have $\ddot{\rho}(\mathbf{0}) = -\mathbf{\Lambda}$. It is easy to see that the joint distribution of $(X(\mathbf{t}), X(\mathbf{0}), \dot{X}(\mathbf{0}), \ddot{X}^{(11)}(\mathbf{0}) + \lambda_{11}X(\mathbf{0}), \ddot{X}^{(12)}(\mathbf{0}) +$

$\lambda_{12}X(\mathbf{0}), \ddot{X}^{(22)}(\mathbf{0}) + \lambda_{22}X(\mathbf{0}))$ is normal with mean $\mathbf{0}$ and covariance matrix

$$\begin{pmatrix} 1 & \rho(\mathbf{t}) & -\dot{\rho}(\mathbf{t})' & \ddot{\rho}_{11}(\mathbf{t}) + \lambda_{11}\rho(\mathbf{t}) & \ddot{\rho}_{12}(\mathbf{t}) + \lambda_{12}\rho(\mathbf{t}) & \ddot{\rho}_{22}(\mathbf{t}) + \lambda_{22}\rho(\mathbf{t}) \\ \rho(\mathbf{t}) & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\dot{\rho}(\mathbf{t}) & \mathbf{0} & \mathbf{\Lambda} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \ddot{\rho}_{11}(\mathbf{t}) + \lambda_{11}\rho(\mathbf{t}) & \mathbf{0} & \mathbf{0} & \rho^{(4)}_{1111}(\mathbf{0}) - \lambda_{11}^2 & \rho^{(4)}_{1112}(\mathbf{0}) - \lambda_{11}\lambda_{12} & \rho^{(4)}_{1122}(\mathbf{0}) - \lambda_{11}\lambda_{22} \\ \ddot{\rho}_{12}(\mathbf{t}) + \lambda_{12}\rho(\mathbf{t}) & \mathbf{0} & \mathbf{0} & \rho^{(4)}_{1112}(\mathbf{0}) - \lambda_{11}\lambda_{12} & \rho^{(4)}_{1212}(\mathbf{0}) - \lambda_{12}^2 & \rho^{(4)}_{1222}(\mathbf{0}) - \lambda_{12}\lambda_{22} \\ \ddot{\rho}_{22}(\mathbf{t}) + \lambda_{22}\rho(\mathbf{t}) & \mathbf{0} & \mathbf{0} & \rho^{(4)}_{1122}(\mathbf{0}) - \lambda_{11}\lambda_{22} & \rho^{(4)}_{1222}(\mathbf{0}) - \lambda_{12}\lambda_{22} & \rho^{(4)}_{2222}(\mathbf{0}) - \lambda_{22}^2 \end{pmatrix},$$

where $\rho^{(4)}_{ijkl}$ is $4th$ order derivative of $\rho$ with respect to $t_i, t_j, t_k, t_l$. We shall denote the $3 \times 3$ right-diagonal sub-matrix of the covariance matrix by $\mathbf{M}$. Hence $X(\mathbf{t})|X(\mathbf{0}), \dot{\mathbf{X}}(\mathbf{0}), \ddot{X}^{(11)}(\mathbf{0}) + \lambda_{11}X(\mathbf{0}), \ddot{X}^{(12)}(\mathbf{0}) + \lambda_{12}X(\mathbf{0}), \ddot{X}^{(22)}(\mathbf{0}) + \lambda_{22}X(\mathbf{0})$ is normal with mean

$$\Big(\rho(\mathbf{t}), -\dot{\rho}(\mathbf{t})'\mathbf{\Lambda}^{-1}, (\ddot{\rho}_{11}(\mathbf{t}) + \lambda_{11}\rho(\mathbf{t}), \ddot{\rho}_{12}(\mathbf{t}) + \lambda_{12}\rho(\mathbf{t}), \ddot{\rho}_{22}(\mathbf{t}) + \lambda_{22}\rho(\mathbf{t}))\,\mathbf{M}^{-1}\Big) \cdot \\ \Big(X(\mathbf{0}), \dot{\mathbf{X}}(\mathbf{0}), \ddot{X}_{11}(\mathbf{0}) + \lambda_{11}X(\mathbf{0}), \ddot{X}_{12}(\mathbf{0}) + \lambda_{12}X(\mathbf{0}), \ddot{X}_{22}(\mathbf{0}) + \lambda_{22}X(\mathbf{0})\Big)'$$

and variance

$$1 - \rho(\mathbf{t})^2 - \dot{\rho}(\mathbf{t})'\mathbf{\Lambda}^{-1}\dot{\rho}(\mathbf{t}) - [\ddot{\rho}_{11}(\mathbf{t}) + \lambda_{11}\rho(\mathbf{t}), \ddot{\rho}_{12}(\mathbf{t}) + \lambda_{12}\rho(\mathbf{t}), \ddot{\rho}_{22}(\mathbf{t}) + \lambda_{22}\rho(\mathbf{t})]\mathbf{M}^{-1} \\ \times(\ddot{\rho}_{11}(\mathbf{t}) + \lambda_{11}\rho(\mathbf{t}), \ddot{\rho}_{12}(\mathbf{t}) + \lambda_{12}\rho(\mathbf{t}), \ddot{\rho}_{22}(\mathbf{t}) + \lambda_{22}\rho(\mathbf{t}))'.$$

By a Taylor expansion,

$$\rho(\mathbf{t}) = 1 - \frac{1}{2}\mathbf{t}'\Lambda\mathbf{t} + o(\mathbf{t}^2), \quad \dot{\rho}(\mathbf{t}) = -\mathbf{\Lambda}\mathbf{t} + o(\mathbf{t}^2), \\ (\ddot{\rho}_{11}(\mathbf{t}) + \lambda_{11}\rho(\mathbf{t}), \ddot{\rho}_{12}(\mathbf{t}) + \lambda_{12}\rho(\mathbf{t}), \ddot{\rho}_{22}(\mathbf{t}) + \lambda_{22}\rho(\mathbf{t})) = \frac{1}{2}(t_1^2, 2t_1t_2, t_2^2)\mathbf{M} + o(\mathbf{t}^2).$$

It is easy to see, after some calculations, that the conditional mean of $X(\mathbf{t})$ is

$$X(\mathbf{0}) + \dot{\mathbf{X}}(\mathbf{0})'\mathbf{t} + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{X}}(\mathbf{0})\mathbf{t} + \epsilon(\mathbf{t}),$$

where $\epsilon(\mathbf{t}) = m\, o(\|\mathbf{t}\|^2)$ with $m = \max\{|X(\mathbf{0})|, \|\dot{\mathbf{X}}(\mathbf{0})\|, \|\ddot{\mathbf{X}}(\mathbf{0})\|\}$. By a Taylor expansion for the relevant terms in the variance and after simplifications, the conditional variance of $X(\mathbf{t})$ is $o(\|\mathbf{t}\|^4)$. This proves the Lemma. $\square$

It is now straightforward to conclude that the Gaussian surface $X(\mathbf{t})$ conditioned on the random variables $X(\mathbf{0}), \dot{\mathbf{X}}(\mathbf{0}), \ddot{\mathbf{X}}(\mathbf{0})$ has the same distribution as the random surface $X(\mathbf{0}) + \dot{\mathbf{X}}'(\mathbf{0})\mathbf{t} + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{X}}(\mathbf{0})\mathbf{t} + m\ o(\|\mathbf{t}\|^2) + K(\mathbf{t})$, where $K(\mathbf{t}) = o(\|\mathbf{t}\|^2)$ a.s., as easily follows from the fact that $K(\mathbf{t})$ is twice continuously differentiable.

We shall say that a sequence of surfaces $Y_u$ converges to a surface $Y^\star$ *uniformly* as $u \to a$ if, for each $h > 0$ and $\|\mathbf{t}\| < h$, we have

$$\mathrm{P}\left\{\lim_{u\to a} \sup_{\|\mathbf{t}\|<h} |Y_u(\mathbf{t}) - Y^\star(\mathbf{t})| = 0\right\} = 1.$$

Now Lemma 3.5.2 can be applied to study the structure of $\chi^2, t$ and $F$ random fields near the local extrema through their Gaussian components.

**Theorem 3.5.3** *Assume $\mathbf{0}$ is a local extremum of the $\chi^2$ field $U(\mathbf{t})$ with height $u$,*

*(a) if $\mathbf{0}$ is a local maximum and $U_u(\mathbf{t})$ is the conditional field as defined in (5.5), then when $u \to \infty$, the surface $U_u(\mathbf{t}/\sqrt{u}) - u$ converges uniformly to the elliptic paraboloid $-\mathbf{t}'\mathbf{\Lambda}\mathbf{t}$;*

*(b) if $\mathbf{0}$ is a local minimum and $\overline{U}_u$ is the corresponding conditional field, then as $u \to 0$, the surface $\overline{U}_u(\sqrt{u}\mathbf{t})/u$ converges uniformly to the random elliptic paraboloid $1 + \mathbf{t}'\mathbf{B}\mathbf{t}$, where $\mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, n+1)$.*

*Proof.* Suppose $X_i(\mathbf{t})$, $i = 1, \ldots, n$ are the Gaussian component fields in the definition of the $\chi^2$ field (2.1). By Lemma (3.5.2), the $\chi^2$ surface $U(\mathbf{t})$ is of the form

$$\sum_{i=1}^{n} \{X_i(\mathbf{0}) + \mathbf{t}'\dot{\mathbf{X}}_i(\mathbf{0}) + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{X}}_i(\mathbf{0})\mathbf{t} + \epsilon_i(\mathbf{t}) + K_i(\mathbf{t})\}^2, \tag{5.9}$$

where $\epsilon_i(\mathbf{t}) = m_i o(\|\mathbf{t}\|^2)$, $m_i = \max\{|X_i(\mathbf{0})|, |\dot{\mathbf{X}}_i(\mathbf{0})|, |\ddot{\mathbf{X}}_i(\mathbf{0})|\}$ and $K_i(\mathbf{t}) = o(\|\mathbf{t}\|^2)$. Given $\mathbf{0}$ is a local maximum, then when $u \to \infty$, $U_u(\mathbf{t}/\sqrt{u})$ can be represented as

$$\sum_{i=1}^{n} \left\{X_i(\mathbf{0}) + \frac{\mathbf{t}'}{\sqrt{u}}\dot{\mathbf{X}}_i(\mathbf{0}) + \frac{1}{2u}\mathbf{t}'\ddot{\mathbf{X}}_i(\mathbf{0})\mathbf{t} + \epsilon_i\left(\frac{\mathbf{t}}{\sqrt{u}}\right) + K_i\left(\frac{\mathbf{t}}{\sqrt{u}}\right)\right\}^2, \tag{5.10}$$

for each $h > 0$ and $\|\mathbf{t}\| < h$. Since the HW distribution of each component Gaussian field and its derivatives at $\mathbf{t} = \mathbf{0}$ has the property that $X_i \sim O(\sqrt{u})$, $\dot{\mathbf{X}}_i \sim O(1)$, $\ddot{\mathbf{X}}_i \sim O(\sqrt{u})$ and $K_i(\mathbf{t}) = o(\|\mathbf{t}^2\|)$ a.s., so $\epsilon_i(\frac{\mathbf{t}}{\sqrt{u}})$ and $K_i(\frac{\mathbf{t}}{\sqrt{u}})$ go to 0 uniformly on $\|\mathbf{t}\| < h$. By expanding the square in (5.10) and using the fact that at $\mathbf{t} = \mathbf{0}$,

$$U = \sum_{i=1}^{n} X_i^2 = u, \qquad \dot{\mathbf{U}} = \sum_{i=1}^{n} X_i \dot{\mathbf{X}}_i = 0,$$

we have the surface

$$U_u(\mathbf{t}/\sqrt{u}) - u - \frac{1}{2u}\mathbf{t}'\ddot{\mathbf{U}}(\mathbf{0})\mathbf{t} \to 0$$

uniformly as $u \to \infty$. Applying Theorem 3.3.1 (a), $u^{-1}\ddot{\mathbf{U}}(\mathbf{0}) \to -2\mathbf{\Lambda}$. Hence (a).

Given $\mathbf{0}$ is a local minimum, then by (5.9), when $u \to 0$, $\overline{U}_u(\sqrt{u}\mathbf{t})/u$ can be represented as

$$1 + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{U}}(\mathbf{0})\mathbf{t} + \text{error term};$$

Since the HW distribution of each component Gaussian field and its derivatives has the property that $X_i \sim O(\sqrt{u})$, $\dot{\mathbf{X}}_i \sim O(1)$ and $\ddot{\mathbf{X}}_i \sim O(1)$, we can control the error term and prove that the error term goes to 0 uniformly on $\|\mathbf{t}\| < h$. Then (b) follows by using Theorem 3.3.1 (b). □

**Theorem 3.5.4** *If $\mathbf{0}$ is a local maximum of the t field $T(\mathbf{t})$ with height $t$, let $T_t(\mathbf{t})$ be the conditional field as defined in (5.5), then as $t \to \infty$, the surface*

$$\frac{T_t(\mathbf{t}/t)}{t} \text{ converges uniformly to } \sqrt{\frac{mA}{mA + \mathbf{t}'\mathbf{B}\mathbf{t}}},$$

*where $A \sim \chi^2_{m+1-N}, \mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, m+1)$ independently.*

*Proof.* Suppose $X(\mathbf{t}), Y_i(\mathbf{t}), i = 1, \ldots, n$ are the Gaussian component fields given by (2.2). Following the same arguments, we can substitute each component field $X(\mathbf{t}), Y_i(\mathbf{t}), i = 1, \cdots, n$ by (5.8) in $T(\mathbf{t})^2$. Then $X(\mathbf{t})^2$ is of the form:

$$\{X(\mathbf{0}) + \mathbf{t}'\dot{\mathbf{X}}(\mathbf{0}) + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{X}}(\mathbf{0})\mathbf{t} + \epsilon(\mathbf{t}) + K(\mathbf{t})\}^2, \quad K(\mathbf{t}) = o(\|\mathbf{t}\|^2).$$

Define $U(\mathbf{t}) = \sum_{i=1}^{m} Y_i(\mathbf{t})^2$, then $U(\mathbf{t})$ is of the form:

$$\sum_{i=1}^{m}\{Y_i(\mathbf{0}) + \mathbf{t}'\dot{\mathbf{Y}}_i(\mathbf{0}) + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{Y}}_i(\mathbf{0})\mathbf{t} + \epsilon_i(\mathbf{t}) + K_i(\mathbf{t})\}^2, \quad K_i(\mathbf{t}) = o(\|\mathbf{t}\|^2).$$

Hence $X(\mathbf{t}/t)^2$ can be represented as:

$$X(0)^2 + \frac{2}{t}\mathbf{t}'X(0)\dot{\mathbf{X}}(\mathbf{0}) + \frac{1}{t^2}\mathbf{t}'[X(\mathbf{0})\ddot{\mathbf{X}}(\mathbf{0}) + \dot{\mathbf{X}}(\mathbf{0})'\dot{\mathbf{X}}(\mathbf{0}]\mathbf{t} + \text{error}_1 = \text{main}_1 + \text{error}_1 \quad (5.11)$$

and $t^2U(\mathbf{t}/t)$ can be represented as:

$$t^2U(\mathbf{0}) + t\mathbf{t}'\dot{\mathbf{U}}(\mathbf{0}) + \frac{1}{2}\mathbf{t}'\ddot{\mathbf{U}}(\mathbf{0})\mathbf{t} + \text{error}_2 = \text{main}_2 + \text{error}_2. \quad (5.12)$$

Define $R = T/\sqrt{m}$, by using the same notations as in the Lemma 3.2.2 and following the proof of Lemma 5.1 in Worsley (1994), we have:

$$\begin{aligned}
X^2 &= \frac{SR^2}{1+R^2}, \quad X\dot{\mathbf{X}} = \frac{S^{\frac{1}{2}}R}{1+R^2}(\mathbf{z}_1 + R\mathbf{z}_2), \\
X\ddot{\mathbf{X}} + \dot{\mathbf{X}}\dot{\mathbf{X}}' &= (1+R^2)^{-1}(\mathbf{z}_1 + R\mathbf{z}_2)(\mathbf{z}_1 + R\mathbf{z}_2)' - (1+R^2)^{-1}SR^2\mathbf{\Lambda} + \\
&\quad (1+R^2)^{-\frac{1}{2}}S^{\frac{1}{2}}R\mathbf{H}_x, \\
V &= \frac{S}{1+R^2}, \quad \dot{\mathbf{V}} = \frac{2S^{\frac{1}{2}}}{1+R^2}(-R\mathbf{z}_1 + \mathbf{z}_2), \\
\ddot{\mathbf{V}} &= 2\{\mathbf{Q} + (1+R^2)^{-1}(-R\mathbf{z}_1 + \mathbf{z}_2)(-R\mathbf{z}_1 + \mathbf{z}_2)' - (1+R^2)^{-1}S\mathbf{\Lambda} \\
&\quad + (1+R^2)^{-\frac{1}{2}}S^{\frac{1}{2}}\mathbf{H}_v\}.
\end{aligned}$$

Substitute $\text{main}_1$ and $\text{main}_2$ in (5.11), (5.12) by the above expression, we can see that when $T = t \to \infty$, which is the same as $R = t/\sqrt{m} \to \infty$,

$$\text{main}_1 \to S(\mathbf{0}), \quad \text{main}_2 \to mS(\mathbf{0}) + \mathbf{t}'\mathbf{Q}(\mathbf{0})\mathbf{t},$$

uniformly on $\|\mathbf{t}\| < h$. Furthermore, when $t \to \infty$, both of the error terms go to 0 uniformly on $\|\mathbf{t}\| < h$ by observing that the HW conditional distribution of the component fields and their derivatives have the property that

$$
\begin{aligned}
X(\mathbf{0}) \sim O(1), &\quad \dot{\mathbf{X}}(\mathbf{0}) \sim O(1), \quad \ddot{\mathbf{X}}(\mathbf{0}) \sim O(1), \\
Y_i(\mathbf{0}) \sim O(1/t), &\quad \dot{\mathbf{Y}}_i(\mathbf{0}) \sim O(1), \quad \ddot{\mathbf{Y}}_i(\mathbf{0}) \sim O(1).
\end{aligned}
$$

Now the theorem follows from the fact that the HW conditional distribution of $S(\mathbf{0})$ and $\mathbf{Q}(\mathbf{0})$ are distributed as $\chi^2_{m+1-N}$ and $\mathrm{Wishart}_N(\mathbf{\Lambda}, m+1)$ independently when $t \to \infty$ as proved in Theorem 3.3.2. $\square$

Now we will state the corresponding result about the $F$ field without proof.

**Theorem 3.5.5** *Assume $\mathbf{0}$ is a local extremum of $F$ field $F(\mathbf{t})$ with height $f$.*

*(a) If $\mathbf{0}$ is a local maximum and $F_f(\mathbf{t})$ is the corresponding conditional field, then as $f \to \infty$, the surface*

$$\frac{F_f(\mathbf{t}/\sqrt{f})}{f} \text{ converges uniformly to } \frac{mA}{mA + n\mathbf{t}'\mathbf{B}\mathbf{t}},$$

*where $A \sim \chi^2_{m+1-N}, \mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, m+1)$ independently.*

*(b) If $\mathbf{0}$ is a local minimum and $\overline{F}_f(\mathbf{t})$ is the corresponding conditional field, then as $f \to 0$, the surface*

$$\frac{\overline{F}_f(\sqrt{f}\mathbf{t})}{f} \text{ converges uniformly to } 1 + \frac{m}{n}\mathbf{t}'A^{-1}\mathbf{B}\mathbf{t},$$

*where $A \sim \chi^2_{m+1-N}, \mathbf{B} \sim \mathrm{Wishart}_N(\mathbf{\Lambda}, n+1)$ independently.*

Comparing these results to Nosko's result for stationary Gaussian fields (Theorem 1.3.2), it is interesting to see that not all the random fields have a quadratic shape or form at high (low) level excursions. The behavior of the $\chi^2$ surface at high level excursions shows a deterministic quadratic form, like that of the Gaussian field. However, at low local minima it has a random quadratic form. Unlike both the Gaussian and $\chi^2$ fields, at high level excursions, it is the inverse of the $F$ field which has a random quadratic form. Similar behavior of the square of the $t$ field can be expected since it can be viewed as a special case of the $F$ field with one numerator degree of freedom.

Nosko has results which go beyond Theorem 1.3.2. By applying the lemma, he obtained the limiting distribution of the size $S$ of one connected component of the excursion set for stationary Gaussian fields at a high level. His result states that $S^{\frac{2}{N}}$ has a limiting exponential distribution. The extension of his result to $\chi^2, t$ and $F$ fields is the topic of the next section.

## 3.6 The limiting distribution of the size of a connected component of the excursion set

In the last section, we have investigated the behavior of a $\chi^2, t$ or $F$ field near extremal points. As a consequence of that, we can obtain the limiting distribution of the size of a connected component of the excursion sets above high or below low levels.

We first need to introduce some notation. Suppose we are interested in the excursion set above (or below) level $u$ which contains $\mathbf{t} = \mathbf{0}$ as a local maximum (or minimum) above (or below) $u$. Define $B_u^+$ (or $B_u^-$) to be the connected component of the excursion set that contains $\mathbf{0}$, that is

$$\begin{aligned} B_u^+ &= \{\mathbf{t} : Y(\mathbf{t}) \geq u, \mathbf{t} \text{ is connected with } \mathbf{0}\}, \\ B_u^- &= \{\mathbf{t} : Y(\mathbf{t}) \leq u, \mathbf{t} \text{ is connected with } \mathbf{0}\}. \end{aligned}$$

Furthermore, denote the size of $B_u^+$ and $B_u^-$ by $S_u^+$, $S_u^-$ respectively:

$$S_u^+ = \lambda(B_u^+), \qquad S_u^- = \lambda(B_u^-),$$

where $\lambda$ stands for Lebesgue measure in $\mathbb{R}^N$. Define $a_N = \pi^{\frac{N}{2}}/\Gamma(\frac{N}{2}+1)$ to be the Lebesgue measure of the unit sphere in $\mathbb{R}^N$. Then we have the following theorems about the limiting distributions of $S_u^+$ and $S_u^-$ for $\chi^2, t$ and $F$ fields.

**Theorem 3.6.1** *For a $\chi^2$ field $U$,*

(a) *when $u \to \infty$, $u^{\frac{N}{2}} S_u^+ \to a_N \det(\mathbf{\Lambda})^{-\frac{1}{2}} E^{\frac{N}{2}}$, where $E \sim$ Exponential(2);*

(b) *when $u \to 0$, $u^{-\frac{N}{2}} S_u^- \to a_N \det(\mathbf{\Lambda})^{-\frac{1}{2}} D^{\frac{N}{2}} \det(\mathbf{B}^\star)^{-\frac{1}{2}}$, where $D \sim \mathrm{Beta}(1, \frac{1}{2}(n - N))$, $\mathbf{B}^\star \sim \mathrm{Wishart}_N(n+1, \mathbf{I})$ independently.*

*Proof.* If $\mathbf{0}$ is a local maximum above $u$, then by Theorem 3.5.3, with probability approaching one as $u \to \infty$, $U$ has the following representation over that component of the excursion set containing $\mathbf{0}$:

$$U(\mathbf{t}) = U(\mathbf{0}) - U(\mathbf{0})\mathbf{t}'\mathbf{\Lambda}\mathbf{t} + o(1). \tag{6.13}$$

Actually, the arguments that lead to the above representation have to go through a process to remove the conditioning, $U(\mathbf{0}) = u$, in Theorem 3.5.3. So,

$$\begin{aligned} B_u^+ &= \{\mathbf{t} : U(\mathbf{0}) - U(\mathbf{0})\mathbf{t}'\mathbf{\Lambda}\mathbf{t} + o(1) \geq u\} \\ &= \left\{\mathbf{t} : \frac{u[U(\mathbf{0}) - u]}{U(\mathbf{0})} + o(1)\frac{u}{U(\mathbf{0})} \geq u\mathbf{t}'\mathbf{\Lambda}\mathbf{t}\right\}. \end{aligned}$$

By the ergodicity of the $\chi^2$ field and Lemma 3.4.1, the HW conditional distribution of $u[U(\mathbf{0}) - u]/U(\mathbf{0})$ is:

$$\begin{aligned} &\lim_{u\to\infty} \mathrm{P}\{\frac{u[U(\mathbf{0}) - u]}{U(\mathbf{0})} > v \| U(\mathbf{0}) > u; \text{maximum}\} \\ = &\lim_{u\to\infty} \mathrm{P}\{U(\mathbf{0}) > \frac{u^2}{u-v} \| U(\mathbf{0}) > u; \text{maximum}\} = \lim_{\mathrm{u}\to\infty} \frac{\mathrm{E}\{\mathrm{M}^+_{\frac{\mathrm{u}^2}{\mathrm{u}-\mathrm{v}}}(\mathrm{U}, \mathrm{C})\}}{\mathrm{E}\{\mathrm{M}^+_{\mathrm{u}}(\mathrm{U}, \mathrm{C})\}} \\ = &\lim_{u\to\infty} \left(\frac{u}{u-v}\right)^{\frac{1}{2}(n+N-1)} \exp\left(-\frac{uv}{2(u-v)}\right) = \exp(-\frac{v}{2}), \end{aligned}$$

and as $u \to \infty$, $u/U(\mathbf{0}) \to 1$. Hence by using the Skorokhod Theorem and the fact that Lebesgue measure is continuous with respect to a.s. convergence, we have $u^{\frac{N}{2}} S_u^+ \to a_N \det(\mathbf{\Lambda})^{-\frac{1}{2}} E^{\frac{N}{2}}$, where $E \sim$ Exponential(2).

If $\mathbf{0}$ is a local minimum below $u$, then by Theorem 3.5.3, with probability approaching one as $u \to 0$, $U(\mathbf{t})$ has the following representation over that component of the excursion set containing $\mathbf{0}$:

$$U(\mathbf{t}) = U(\mathbf{0}) + \mathbf{t}'\mathbf{B}\mathbf{t} + o(u). \tag{6.14}$$

So,

$$B_u^- = \{\mathbf{t} : U(\mathbf{0}) + \mathbf{t}'\mathbf{B}\mathbf{t} + o(u) \leq u\} = \left\{\mathbf{t} : \frac{u - U(\mathbf{0})}{u} + o(1) \geq u^{-1}\mathbf{t}'\mathbf{B}\mathbf{t}\right\}.$$

By Theorem 3.4.2, the HW conditional distribution of $(u - U(\mathbf{0}))/u$ is $\text{Beta}(1, \frac{1}{2}(n - N))$. Define $\mathbf{B}^\star = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{B}\mathbf{\Lambda}^{-\frac{1}{2}}$, then (b) follows immediately by using the same arguments as in (a).

□

**Theorem 3.6.2** *For a $t$ field $T$,*

$$t^N S_t^+ \to a_N \det(\mathbf{\Lambda})^{-\frac{1}{2}} m^{\frac{N}{2}} D^{\frac{N}{2}} A^{\frac{N}{2}} \det(\mathbf{B}^\star)^{-\frac{1}{2}}$$

*as $t \to \infty$, where $D \sim \text{Beta}(1, \frac{1}{2}(m - N))$, $\mathbf{B}^\star \sim \text{Wishart}_N(m + 1, \mathbf{I})$, $A \sim \chi^2_{m+1-N}$ all independently.*

*Proof.* If $\mathbf{0}$ is a local maximum above $t$, then by Theorem 3.5.4, with probability approaching one as $t \to \infty$, $T(\mathbf{t})$ has the following representation over that component of the excursion set containing $\mathbf{0}$:

$$T(\mathbf{t})^2 = \frac{mAT(\mathbf{0})^2}{mA + T(\mathbf{0})^2\mathbf{t}'\mathbf{B}\mathbf{t}} + o(t^2), \tag{6.15}$$

So,

$$\begin{aligned} B_u^+ &= \{\mathbf{t} : \frac{mAT(\mathbf{0})^2}{mA + T(\mathbf{0})^2\mathbf{t}'\mathbf{B}\mathbf{t}} + o(t^2) \geq t^2\} \\ &= \left\{\mathbf{t} : mA\frac{T(\mathbf{0})^2 - t^2}{T(\mathbf{0})^2} + o(1) \geq t^2\mathbf{t}'\mathbf{B}\mathbf{t}\right\}. \end{aligned}$$

By ergodicity of the $t$ field and Lemma 3.4.1, the HW conditional distribution of

$[T(\mathbf{0})^2 - t^2]/T(\mathbf{0})^2$ is:

$$
\begin{aligned}
& \lim_{t\to\infty} \mathrm{P}\{\frac{T(\mathbf{0})^2 - t^2}{T(\mathbf{0})^2} > v \| T(0) > t; \text{maximum}\} \\
= & \lim_{t\to\infty} \mathrm{P}\{T(\mathbf{0})^2 > \frac{t^2}{1-v} \| T(0) > t; \text{maximum}\} = \lim_{t\to\infty} \frac{\mathrm{E}\{M^+_{\frac{t}{\sqrt{1-v}}}(T, C)\}}{\mathrm{E}\{M^+_t(T, C)\}} \\
= & \lim_{t\to\infty} \left(\frac{t^2}{1-v}\right)^{-\frac{1}{2}(m-N)} t^{m-N} = (1-v)^{-\frac{1}{2}(m-N)}.
\end{aligned}
$$

Define $\mathbf{B}^\star = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{B} \mathbf{\Lambda}^{-\frac{1}{2}}$, then the result follows immediately by following the same arguments in Theorem 3.6.1 (a). $\square$

Now we state the corresponding result about the distribution of the size of a connected component of an $F$ field without proof.

**Theorem 3.6.3** *For an $F$ field $F$,*

$$
f^{\frac{N}{2}} S_f^+ \to a_N \det(\mathbf{\Lambda})^{-\frac{1}{2}} n^{-\frac{N}{2}} m^{\frac{N}{2}} D^{\frac{N}{2}} W^{\frac{N}{2}} \det(\mathbf{B}^\star)^{-\frac{1}{2}},
$$

*as $f \to \infty$, where $D \sim \mathrm{Beta}(1, \frac{1}{2}(m-N))$, $\mathbf{B}^\star \sim \mathrm{Wishart}_N(m+1, \mathbf{I})$, $W \sim \chi^2_{m+n-N}$ independently.*

One can see the similarities between the limiting distribution of $S_u^-$ for the $\chi^2$ field and that of $S_t^+, S_f^+$ for the $t$ or $F$ field. It can be explained by the fact that when the $t$ or $F$ field goes to infinity, the $\chi^2$ field $\sum Y_i^2$, which is the denominator in the definition of the $t$ or $F$ field, goes to zero.

## 3.7 The limiting distribution of the size of the largest connected component of the excursion set

In the previous section, we have found out the limiting distribution of the size of one connected component of excursion set of $\chi^2, t$ and $F$ fields. In this section, the distribution of the size of the *largest* connected component of the excursion set will be discussed in an informal way.

We need some notations. Let $B_1, \ldots, B_L$ be all the connected components of the excursion set, where $L$ is the number of such connected components. Denote $S_i = \lambda(B_i)$ to be the Lebesgue measure of $B_i, i = 1, \ldots, L$. Then we are interested in $S_{\max} = \max_{1 \le i \le L} S_i$, which is the size of the largest connected component of the excursion set. In Section 6.9, Adler (1981) pointed out that $L$ has a limiting Poisson distribution. A Poisson distribution is intuitively sensible, in the sense that one can regard the centers of the connected components as a multidimensional point process with no 'memory' and hence $S_i, i = 1, \ldots, L$ are independent identically distributed in the limit. Define the size of one connected component as $S$, then

$$\mathrm{P}(L = l) \approx \frac{\mathrm{E}(L)^l}{l!} e^{-\mathrm{E}(L)}, \tag{7.16}$$

and

$$\begin{aligned} & \mathrm{P}(S_{\max} \le x | L \ge 1) \\ = & \sum_{l=1}^{\infty} \mathrm{P}(S_i \le x, 1 \le i \le l, L = l | L \ge 1) \\ = & \sum_{l=1}^{\infty} \mathrm{P}(L = l | L \ge 1) \mathrm{P}(S_i \le x, 1 \le i \le l | L = l) \\ \approx & \frac{1}{\mathrm{P}(L \ge 1)} \sum_{l=1}^{\infty} \frac{\mathrm{E}(L)^l}{l!} e^{-\mathrm{E}(L)} \mathrm{P}(S \le x)^l \\ = & \frac{\exp\{-\mathrm{E}(L)\mathrm{P}(S \ge x)\} - \exp\{-\mathrm{E}(L)\}}{1 - \exp\{-\mathrm{E}(L)\}}. \end{aligned} \tag{7.17}$$

In the extreme case when $\mathrm{E}(L)$ approaches 0, $\mathrm{P}(S_{\max} \le x) \to \mathrm{P}(S \le x)$. However, when $\mathrm{E}(L)$ is large enough (say, bigger than 5) so that we can ignore $\exp\{-\mathrm{E}(L)\}$ in (7.17) and obtain

$$\mathrm{P}(S_{\max} \le x | L \ge 1) \approx \exp\{-\mathrm{E}(L)\mathrm{P}(S \ge x)\}.$$

As argued in Chapter 6 of Adler (1981), $\mathrm{E}(L)$, $\mathrm{E}(\chi(A_u(Y, C)))$, $\mathrm{E}(M_u(Y, C))$ are asymptotically the same at high (or low) level $u$, that is,

$$\mathrm{E}(L) \approx \mathrm{E}(\chi(A_u(Y,C))) \approx \mathrm{E}(M_u(Y,C)). \tag{7.18}$$

In practice, we tend to choose $\mathrm{E}(\chi(A_u(Y,C)))$ as an approximation of $\mathrm{E}(L)$ more often since the exact formula of $\mathrm{E}(\chi(A_u(Y,C)))$ can usually be obtained. The expression of $\mathrm{E}(\chi(A_u(Y,C)))$ for the Gaussian field can be found in Adler (1981) and for $\chi^2, t$ and $F$ fields in Worsley (1994).

We shall now mention the approach adopted by Friston et al. (1994) to find the distribution of $S_{\max}$ for non-Gaussian random fields, especially for the $t$ field. They proposed to find the distribution of $S_{\max}$ by using the result of its Gaussian counterpart. The field is first *Gaussianized*, that is, at each point, the distribution is transformed to the Gaussian distribution as explained in section 3.1. Obviously, the excursion set of the field corresponds to the excursion set of the Gaussianized field at a different level. Then by treating the Gaussianized field as an approximation of a Gaussian random field, one can approximate the distribution of $S_{\max}$ for the non-Gaussian field by the corresponding distribution for the Gaussian field. However, the Gaussianized field is *not* a Gaussian random field since the fi-di distribution does not have a multi-dimensional Gaussian form. Furthermore, for the $t$ field, the resulting Gaussianized field should be slightly rougher than the original $t$ field, so Worsley et al. (1992) suggested a correction factor to account for this and improve the accuracy of the approximation. Nevertheless, this approach could serve only as an approximation when the degrees of freedom of the $t$ field is high enough.

Since only an asymptotic distribution of $S$ is available, a correction to the mean of $S$ might achieve a better approximation in actual applications. Friston et al. (1994) proposed a way to correct the mean of $S$ as follows. Denote the Lebesgue measure of the excursion set to be $\lambda(A_u(Y,C))$. Then,

$$\mathrm{E}(L)\mathrm{E}(S) = \mathrm{E}(\lambda(A_u(Y,C))) = \lambda(C)\Phi(-u) = \lambda(C)\int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2}dz. \tag{7.19}$$

Now the distribution of the size could be corrected by multiplication by a factor to ensure that $\mathrm{E}(S) = \lambda(C)\Phi(-u)/\mathrm{E}(L)$.

## 3.8 Validation

We present some $N = 2$ dimensional simulations to validate the distribution of $L$, $S$ and $S_{\max}$ from results in Section 3.6 and 3.7 for $\chi^2, t$ and $F$ fields. We first generated each of the Gaussian component fields and then generated $\chi^2, t$ and $F$ fields according to their definitions in (2.1), (2.2), (2.3). Each Gaussian field was independently simulated by convolving a Gaussian white noise with a Gaussian kernel. The 2-D simulations used random images with $256 \times 256$ pixels and a Gaussian kernel with FWHM $= 9.2$. FWHM, or *Full Width at Half Maximum*, the width of the kernel at half its maximum height, is commonly used in imaging to measure the spread of a kernel. For a Gaussian kernel, FWHM is $\sqrt{8 \log 2} \times$ standard deviation. In each realization, we recorded the number of connected components in the excursion set, $L$, the size of each connected component, $S$, and the size of the largest connected component, $S_{\max}$. The number of simulations was around 5000. Using this data, we compared the empirical estimates and theoretical results of $\mathrm{P}(L = x), \mathrm{P}(S = x), \mathrm{P}(S_{\max} = x)$. Note all the theoretical results are in the limiting sense.

Figure B-11 shows the simulation result for a $t$ field with $n = 7$ degrees of freedom and threshold $t = 4.75$. Figure B-12 is for the excursion set of a $\chi^2$ field with $n = 7$ degrees of freedom above the threshold $u = 22.04$, while Figure B-13 is for the excursion set of the same $\chi^2$ field below threshold $u = 0.80$. Figure B-14 shows the simulation result for an $F$ field with $n = 4$ numerator degrees of freedom, $m = 5$ denominator degrees of freedom and threshold $f = 21.05$. The thresholds were chosen so that the exceedance 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 the empirical distributions from simulations and the dotted lines are the theoretical distributions after correcting the mean size by (7.19). For the $t$ field in Figure B-11, the dashed lines represent the distribution from the Gaussianization approach. The agreement between empirical and theoretical distributions is evident. However, we can see some discrepancies between the solid lines and the dashed lines in Figure B-11, which suggest that the Gaussianization approach may not provide a good approximation when the degrees of freedom of the $t$ field are relatively small. Actually, even larger discrepancies are expected for higher dimensions.

As mentioned in Friston et al. (1994), the small discrepancies between the empirical and theoretical distributions may derive from a number of sources: (i) many of our approximations are only asymptotically true, in the limit of high or low thresholds; (ii) the effect of sampling the random field on a lattice of voxels; (iii) the edge effects of our simulated process.

## 3.9 Application to positron emission tomography images

The PET technique uses positron-emitting labelled carriers, created in an on-site cyclotron, to produce an image of brain activity such as glucose utilization, oxygen consumption or blood flow. An increasingly important application of PET involves the measurement of regional CBF (cerebral blood flow) during the performance of various sensorimotor or cognitive tasks. Regional changes in CBF between two tasks reflect the mobilization of functional units specific to the new task. The purpose of these studies is to identify the region activated by the stimulation condition. In many cognitive studies, the activation is so slight that the experiment must be repeated on several subjects and the subtracted images are averaged to improve the signal-to-noise ratio. Typically, the number of subjects in the experiment is small. A detailed description of such an experiment can be found in Worsley et al. (1992), Worsley (1994).

While typically in these studies, the global maximum is used as a test statistic, some investigators have suggested an alternative. They proposed the use of the size of the largest connected component of the excursion set above a relatively high threshold as the test statistic for activations based on the assumption that areas of true neural activity will tend to stimulate signal changes over contiguous voxels. If the underlying variance is stationary, so that a Gaussian random field in 2.7 is used for the hypothesis testing, then one can apply the work of Friston et al. (1994) to assess the significance of the activated regions by their spatial extent. However, if the variance is not stationary and the resulting statistic image is a $t$ random field, we would apply what we have established in the previous sections to assess the significance.

Experiments were carried out to investigate the neural activity in response to a painful heat stimulus. There were $p = 9$ subjects and two experiments: a control experiment in which the subjects were given the same baseline stimulus twice and a treatment experiment in which the subjects were exposed to a hot painful stimulus after the baseline stimulus. The null hypothesis is that there is no difference in brain activity between the two stimuli.

The stationarity of the variance was examined by the sample variance image in (2.5), which under the stationarity assumption, should be a $\chi^2$ random field with $p-1$ degrees of freedom. Evidence against the stationarity was found by using the global maximum as the test statistic, which was reported in Worsley (1994). Hence, the $t$ statistic image calculated by (2.6) was used to test regions of activation.

Figure B-15 shows the images of the same horizontal slice through two $t$ random fields generated from the control (left) and treatment (right) experiment respectively. The corresponding excursion sets in the same slice, above threshold 5.0 or below threshold $-5.0$, are also shown.

Throughout the analysis, $\lambda(C) = 1000\text{cm}^3$ was used as the volume of the whole brain, the search region $C$. Without smoothing ($\text{FWHM}_x = 6.1\text{mm}, \text{FWHM}_y = 7.0\text{mm}, \text{FWHM}_z = 7.302\text{mm}$), we found that there was no evidence against the null hypothesis both in control and treatment experiments. Smoothing the image to have $\text{FWHM}_x = \text{FWHM}_y = \text{FWHM}_z = 11.6\text{mm}$, we can see some significant areas of activation in the treatment experiment, while no significant activation was found in the control experiment as expected. For comparison, we also tried the Gaussianization approach, and its performance as well as ours is presented in Table 3.1. Listed in the table is the threshold of the excursion set, the expected number of the connected

Table 3.1: Comparison between the Gaussianization and the t-field approach in the pain experiment (painful heat stimulus applied)

| Method | Threshold | E($L$) | P($S \geq 4.74$) | P($S \geq 1.21$) | P($S \geq 0.88$) | P($S \geq 0.44$) | P($S \geq 1.06$) |
|---|---|---|---|---|---|---|---|
| t-field | 5 | 10.89 | 0.0000 | 0.0046 | 0.0145 | 0.0983 | 0.0077 |
| Gaussianized | 3.276 | 3.38 | 0.0000 | 0.0288 | 0.0708 | 0.2584 | 0.0433 |

components of the excursion set, E($L$), followed by the P-values of the observed sizes for four activated and one deactivated region. The unit of the sizes is $\mathrm{cm}^3$. The threshold for the $t$ field was chosen to be 5 with a corresponding threshold for the Gaussian field of 3.28. That is, $\mathrm{P}(T \geq 5) = \mathrm{P}(Z \geq 3.28)$, where $T$ is a Student-$t$ random variable with 8 degrees of freedom and $Z$ is a Gaussian random variable.

From the table, we can see that for the same activated area, our result gives a smaller P-value than the Gaussianization approach. Hence an improvement in the power of detection can be expected. It is also worth noting that if $T_{\max}$ was used as the test statistic, then no significant activation can be detected (reported in Worsley 1994). This may suggest that the global maximum as a test statistic may not be powerful enough in this case.

# Chapter 4

# Further work

In the detection of brain functional and structural change, computation of a statistic indicating activity at each voxel of the brain gives a statistical image $Y$. To identify significant activations, two test statistics from the statistical image are proposed. One is the global maximum, $Y_{\max}$, the other is the size of the largest connected component of the excursion set, $S_{\max}$.

For the statistical image which can be modelled as a Hotelling's $T^2$ field under the null hypothesis, the result in Chapter 2 gives an approximate P-value of $Y_{\max}$ at high levels. However, no null distribution of $S_{\max}$ is obtained. Some attempt has been made in Chapter 2, such as the discussions on the distribution of the curvature of Hotelling's $T^2$ field at high (or low) level extremal points. Yet, a full development has not been done.

The two test statistics emphasize different features of the activation foci in the statistical images: peak height $Y_{\max}$ and peak spatial extent $S_{\max}$. A combination of the two may result in an increase of power of detection. Poline et al. (1996) presented such an analysis for images from Gaussian random field. The extension to $\chi^2, t$ and $F$ fields can be done in the future.

Throughout the thesis, the images we discussed are those with a strong correlation between the adjacent voxels. This is caused by smoothing or the intrinsic low resolution of the scanner (for PET images). However, in functional MRI images, a much better resolution can be expected and hence only a slight correlation is present. In this case, the random field theory cannot be applied any longer because of the effect of discretizing the process on voxels. The P-value of $Y_{\max}$ can still be approximated by a Bonferroni correction, but the P-value of $S_{\max}$ is unknown. Although there are some simulation results in Forman et al. (1995), a rigorous theoretical development needs to be done.

# Appendix A

# Lemmas

Suppose $G \sim$ Multi-$T^2(p,k,n)$ as defined in Lemma 2.3.2, then the following lemma concerns the expectations of the determinant of the matrix $G$, $G^{\frac{1}{2}}$.

**Lemma A.0.1** *If $k \leq p$, the expectations of* $\det(G), \det(G)^{\frac{1}{2}}$ *are:*

$$\begin{aligned} \mathrm{E}[\det(G)] &= \frac{p!(n-p-2)!}{(p-k)!(n+k-p-2)!} \quad \textit{for} \quad n \geq p+2, \\ \mathrm{E}[\det(G)^{\frac{1}{2}}] &= \frac{\Gamma[\frac{1}{2}(n-p)]\Gamma[\frac{1}{2}(p+1)]}{\Gamma[\frac{1}{2}(n+k-p)]\Gamma[\frac{1}{2}(p-k+1)]} \quad \textit{for} \quad n \geq p+1. \end{aligned}$$

*If $k > p$, the lemma still holds if division by the factorial of a negative integer, and division by the $\Gamma$ function at a negative number is treated as multiplication by zero.*

*Proof.* By Lemma 2.3.2, we have

$$\begin{aligned} \mathrm{E}[\det(G)] &= \int \det(G) f(G) dG \\ &= c(p,k,n) \int \det(G)^{\frac{(p+2)-k-1}{2}} \det(I+G)^{-(n+k)/2} dG \\ &= \frac{c(p,k,n)}{c(p+2,k,n)}. \end{aligned}$$

Likewise,

$$\mathrm{E}[\det(G)^{\frac{1}{2}}] = \frac{c(p,k,n)}{c(p+1,k,n)}.$$

The result follows after we plug the formula for $c(p,k,n)$ in (3.12) and then simplify.

$\square$

In what follows, for a symmetric matrix $A$, we shall denote its principal minor consisting of $j$ rows and columns by $A_j$.

**Lemma A.0.2** *Let $U \sim \mathrm{Wishart}_N(I,\nu)$, $A$ be a fixed symmetric $N \times N$ matrix and let $b$ be a fixed scalar. Then*

$$\mathrm{E}(\det(A+bU)) = \sum_{j=0}^{N} \frac{b^j \nu!}{(\nu-j)!} \mathrm{detr}_{N-j}(A),$$

*Proof.* where $U^* = O'UO$. Let $O$ be an orthonormal matrix such that $O'AO = L = \mathrm{diag}(l_1,\ldots,l_N)$ is a diagonal matrix of eigenvalues of $A$. Then,

$$\mathrm{E}(\det(A+bU)) = \mathrm{E}(\det(L+bO'UO)) = \mathrm{E}(\det(bU^*+L)).$$

Since $L$ is a diagonal matrix then $\det(bU^*+L)$ can be expanded in terms of products of the determinant of each $k \times k$ principal minor of $U^*$ with $N-k$ members of $\{l_1,\ldots,l_N\}$ corresponding to the remaining rows and columns not included in the principal minor. Note that each $j \times j$ principal minor of $U^*$ has a $\mathrm{Wishart}_j(I_{j\times j},\nu)$ distribution and from standard multivariate statistics its expectation is $\nu!/(\nu-j)!$, we obtain the result. $\square$

**Lemma A.0.3** *Let $U \sim \mathrm{Wishart}_N(I,\nu)$, $H \sim \mathrm{Normal}_{N\times N}(0, M(I_N))$ independently, $A$ be a fixed symmetric $N \times N$ matrix and let $a$ and $b$ be fixed scalars. Then*

$$\mathrm{E}(\det(A+aH+bU)) = \sum_{j=0}^{\lfloor N/2 \rfloor} \frac{(-1)^j}{2^j j!} a^{2j} \sum_{k=0}^{N-2j} b^k \begin{pmatrix} \nu \\ k \end{pmatrix} (2j+k)!\mathrm{detr}_{N-2j-k}(A).$$

*Proof.* Holding $U$ fixed and applying Lemma A.2 in Worsley (1994) we have

$$\mathrm{E}(\det(A+aH+bU)) = \sum_{j=0}^{\lfloor N/2 \rfloor} \frac{(-1)^j (2j)!}{2^j j!} a^{2j} \mathrm{E}(\mathrm{detr}_{N-2j}(A+bU)). \qquad (0.1)$$

By Lemma A.0.2, we have

$$\begin{aligned}
\mathrm{E}(\mathrm{detr}_{N-2j}(A+bU)) &= \sum \sum_{k=0}^{N-2j} \frac{b^k \nu!}{(\nu-k)!} \mathrm{detr}_{N-2j-k}(A_{N-2j}) \\
&= \sum_{k=0}^{N-2j} \frac{b^k \nu!}{(\nu-k)!} \begin{pmatrix} 2j+k \\ k \end{pmatrix} \mathrm{detr}_{N-2j-k}(A), \qquad (0.2)
\end{aligned}$$

where the first summation is over all possible choices of $(N-2j)\times(N-2j)$ principal minors from $N \times N$ matrix $A$. Note that $U$ and $H$ are independent. Combining (0.1), (0.2) and after proper simplification, we can obtain the result. $\square$

**Lemma A.0.4** *Let $x \sim \mathrm{Normal}_{M\times N}(0, I)$, $V$ be a fixed symmetric $N \times N$ matrix, $W$ be a fixed $M \times N$ matrix with $W'W = V$ and $a$ be a fixed scalar. Then*

$$\mathrm{E}\{\det[V - a(W'x + x'W)]\} = \sum_{m=0}^{\lfloor N/2 \rfloor} (-1)^m a^{2m} \frac{(N-m)!}{(N-2m)!} \mathrm{detr}_{N-m}(V).$$

*Proof.* Let $W = (W_1, W_2, \ldots, W_N)$, $x = (x_1, x_2, \ldots, x_N)$ where $W_i, x_i$ are column vectors of $W$ and $x$ respectively. Let $e = (e_1, e_2, \ldots, e_j)$ be a sequence consisting of $j$ distinct positive integers, let $P[e]$ be the set of all permutations of $e$ with elements $p[e] = (i_{e_1}, i_{e_2}, \ldots, i_{e_j})$ and let $\eta(e; p[e])$ equal $+1$ or $-1$ depending on the order of the permutation from $e$ to $p[e]$. If we define $e_0 = (1, 2, \ldots, N)$, then

$$\begin{aligned}
&\det[V - a(W'x + x'W)] \qquad (0.3) \\
= &\sum_{P[e_0]} \eta(e_0; p[e_0])(V_{1i_1} - aW_1' x_{i_1} - aW_{i_1}' x_1) \ldots (V_{Ni_N} - aW_N' x_{i_N} - aW_{i_N}' x_N).
\end{aligned}$$

Suppose $l, m, n$ are integers satisfying $1 \leq l, m, n \leq N$ and $l + m + n = N$. Let $q_{l,m,n} = (r_1, \ldots, r_l; s_1, \ldots, s_m; t_1, \ldots, t_n) \in P[e_0]$ with order $r_1 < \ldots < r_l; s_1 < \ldots < s_m; t_1 < \ldots < t_n$. Such a choice of $q_{l,m,n}$ is uniquely determined by a choice of $\phi_{l,n} = (r_1, \ldots, r_l; t_1, \ldots, t_n)$. Therefore, if we expand the product inside the summation in (0.3), the terms would be of the form

$$\eta(q_{l,m,n}; p[q_{l,m,n}])(-a)^{l+m} W_{r_1}' x_{i_{r_1}} \ldots W_{r_l}' x_{i_{r_l}} \times W_{i_{s_1}}' x_{s_1} \ldots W_{i_{s_m}}' x_{s_m}$$

$$\times V_{t_1 i_{t_1}} \dots V_{t_n i_{t_n}}. \tag{0.4}$$

Since $x \sim \text{Normal}_{M\times N}(0, I)$ and $W'W = V$, we have

$$\mathrm{E}(W_i' x_j \times W_k x_l) = \delta_{jl} W_i' W_k = \delta_{jl} V_{ik}.$$

Note that if $(Y_1, Y_2, \dots, Y_{2j+1})$ joint normal with mean zero then

$$\mathrm{E}(Y_1 Y_2 \dots Y_{2j+1}) = 0,$$

therefore for the terms (0.4) to have nonzero expectation, we must have $l = m$ and $(i_{r_1}, \dots, i_{r_m})$ is a permutation of $(s_1, \dots, s_m)$. To have $(r_1, \dots, r_m)$ ordered and permute $i_{r_1}, \dots, i_{r_m}$ would have the same effect as ordering $i_{r_1}, \dots, i_{r_m}$ and let $r_1, \dots, r_m$ permute. Hence, form now on we could assume that $i_{r_1} = s_1, \dots, i_{r_m} = s_m$ and $r_1, \dots, r_m$ with no particular order. Let $\tilde{\Phi}_m$ be the set consisting of all choices of $\phi_{m,N-2m}$ but without the ordering on $(r_1, \dots, r_m)$ and define $\tilde{\phi}_m = \phi_{m,N-2m}$. Note that when $i_{r_1} = s_1, \dots, i_{r_m} = s_m$, $(i_{s_1}, \dots, i_{s_m}; i_{t_1}, \dots, i_{t_{N-2m}})$ is a permutation of $(r_1, \dots, r_m; t_1, \dots, t_{N-2m})$ and

$$\eta(q_{m,m,N-2m}; p[q_{m,m,N-2m}]) = (-1)^m \eta(\tilde{\phi}_m; p[\tilde{\phi}_m]),$$

then

$$\begin{aligned} &\mathrm{E}\{\det[V - a(W'x + x'W)]\} \\ = &\sum_{m=0}^{\lfloor (N/2) \rfloor} \sum_{\tilde{\Phi}_m} (-1)^m a^{2m} \sum_{P(\tilde{\phi}_m)} \eta(\tilde{\phi}_m; p[\tilde{\phi}_m]) V_{r_1 i_{s_1}} \dots V_{r_m i_{s_m}} V_{t_1 i_{t_1}} \dots V_{t_{N-2m} i_{t_{N-2m}}}. \end{aligned}$$

As we know, the last summation actually comes down to the determinant of the $(N-m) \times (N-m)$ principal minor of matrix $V$ with rows and columns from $r_1, \dots, r_m; t_1, \dots, t_{N-2m}$ and there are all together $\frac{N!}{(N-2m)!m!}$ such principal minors. Recall $\mathrm{detr}_{N-m}(V)$ is the summation of all the $\frac{N!}{m!(N-m)!}$ distinct principal minors of

$V$, then

$$\mathrm{E}\{\det[V - a(W'x + x'W)]\} = \sum_{m=0}^{\lfloor N/2 \rfloor} (-1)^m a^{2m} \frac{(N-m)!}{(N-2m)!} \mathrm{detr}_{N-m}(V).$$

$\square$

**Lemma A.0.5** *Let* $U \sim \mathrm{Wishart}_N(I, \nu)$, $H \sim \mathrm{Normal}_{N\times N}(0, M(I_N))$, $x \sim \mathrm{Normal}_{N\times N}(0, I)$ *and* $V$ *be a fixed symmetric* $N \times N$ *matrix and* $a, b, c$ *be fixed scalars. Then*

$$\begin{aligned} &\mathrm{E}\{\det[V - c(V^{\frac{1}{2}}x + x'V^{\frac{1}{2}}) + aH + bU]\} \\ = &\sum_{j=0}^{\lfloor N/2 \rfloor} \sum_{k=0}^{N-2j} \sum_{m=0}^{\lfloor (N-2j-k)/2 \rfloor} \begin{pmatrix} \nu \\ k \end{pmatrix} \frac{(-1)^{j+m}(2j+k+m)!(N-2j-k-m)!}{2^j j! m! (N-2j-k-2m)!} a^{2j} b^k c^{2m} \\ &\mathrm{detr}_{N-2j-k-m}(V). \end{aligned}$$

*Proof.* Define $A = V - c(V^{\frac{1}{2}}x + x'V^{\frac{1}{2}})$. Holding $x$ fixed, then by Lemma A.0.3 we have

$$\mathrm{E}\{\det[A + aH + bU]\} = \sum_{j=0}^{\lfloor N/2 \rfloor} \frac{(-1)^j}{2^j j!} a^{2j} \sum_{k=0}^{N-2j} b^k \begin{pmatrix} \nu \\ k \end{pmatrix} (2j+k)! \mathrm{detr}_{N-2j-k}(A).$$

For any $l \times l$ principal minor of $A$, we can show that it has the form $V^* + W^{*'}x^* + x^{*'}W^*$ where $V^*$ is a $l \times l$ principal minor of $V$, $W^*$ is a $N \times l$ matrix satisfying $W^{*'}W^* = V^*$ and $x^* \sim \mathrm{Normal}_{N\times l}(0, I)$. Hence by Lemma A.0.4 we have

$$\begin{aligned} &\mathrm{E}\{\det[V - c(V^{\frac{1}{2}}x + x'V^{\frac{1}{2}}) + aH + bU]\} \\ = &\sum_{j=0}^{\lfloor N/2 \rfloor} \frac{(-1)^j}{2^j j!} a^{2j} \sum_{k=0}^{N-2j} b^k \begin{pmatrix} \nu \\ k \end{pmatrix} (2j+k)! \sum \sum_{m=0}^{\lfloor (N-2j-k)/2 \rfloor} (-1)^m c^{2m} \\ &\frac{(N-2j-k-m)!}{(N-2j-k-2m)!} \mathrm{detr}_{N-2j-k-m}(V_{N-2j-k}), \end{aligned}$$

where the second to the last summation is over all possible choices of $(N-2j-k) \times (N-2j-k)$ principal minors from the $N \times N$ matrix $V$. Proper simplification leads to

$$
\begin{aligned}
& \mathrm{E}\{\det[V - c(V^{\frac{1}{2}}x + x'V^{\frac{1}{2}}) + aH + bU]\} \\
= & \sum_{j=0}^{\lfloor N/2 \rfloor} \sum_{k=0}^{N-2j} \sum_{m=0}^{\lfloor (N-2j-k)/2 \rfloor} \begin{pmatrix} \nu \\ k \end{pmatrix} \frac{(-1)^{j+m}(2j+k+m)!(N-2j-k-m)!}{2^j j! m! (N-2j-k-2m)!} a^{2j} b^k c^{2m} \\
& \qquad \mathrm{detr}_{N-2j-k-m}(V).
\end{aligned}
$$

□

**Theorem A.0.6** (Skorohod Theorem) *Let $(S, \rho)$ be a complete separable metric space and $P_n, n = 1, 2, \ldots,$ and $P$ be probabilities on $(S, \mathcal{B}(S))$ such that $P_n \xrightarrow{w} P$ as $n \to \infty$. Then, on a probability space $(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{P})$, we can construct $S$-valued random variables $X_n, n = 1, 2, \ldots,$ and $X$ such that*

**(i)** $P_n = \tilde{P}^{X_n}, n = 1, 2, \ldots$ *and* $P = \tilde{P}^X$.

**(ii)** *$X_n$ converges to $X$ almost everywhere.*

*Thus, in particular, convergence in law of random variables $X_n$ can be realized by an almost everywhere convergence without changing the law of any $X_n$.*

# References


Adler, R.J. (1981). *The Geometry of Random Fields.* Wiley, New York.

Adler, R.J. and Hasofer, A.M. (1976). Level crossings for random fields. *Annals of Probability,* **4**, 1-12.

Aronowich M. and Adler, R.J. (1986). Extrema and level crossings of $\chi^2$ processes. *Advances in Applied Probability*, **18**, 901-920.

Aronowich, M. and Adler, R.J. (1988). Sample path behaviour of $\chi^2$ surfaces at extrema. *Advances in Applied Probability*, **20**, 719-738.

Beljaev, Ju.K. (1967). Bursts and shines of random fields. *Soviet Math. Dokl.*, **8**, 1107-1109.

Collins, D.L., Holmes, C.J., Peters, T.M. and Evans, A.C. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, **3**, 190-208.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. and Noll, D.C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, **33**, 636-647.

Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S.J. (1991). Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism*, **11**, 690-699.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**, 214-220.

Hasofer, A.M. (1978). Upcrossings of random fields. *Supplement to Advances in Applied Probability,* **10**, 14-21.

Kac, M., Slepian, D. (1959). Large excursions of Gaussian processes. *Annals of Mathematical Statistics*, **30**, 1215-1228.

Lindgren, G. (1972). Local maxima of Gaussian fields. *Ark. Math.*, **10**, 195-218.

Lindgren, G. (1980). Model processes in nonlinear prediction with application to detection and alarm. *The Annals of Probability*, **8**, 775-792.

Lindgren, G. (1984). Use and structure of Slepian model processes for prediction and detection in crossing and extreme value theory. In *Statistical Extremes and Applications*, Ed. J. Tiago de Oliveira. Reidel, Dordrecht.

MacDonald, J.D., Avis, D. and Evans, A.C. (1994). Multiple surface identification and matching in magnetic resonance images. *Visualization in Biomedical Computing*, 160-169.

Nosko, V.P. (1969). Local structure of Gaussian random fields in the vicinity of high-level light sources. *Soviet Math. Dokl.*, **10**, 1481-1484.

Olkin, I., Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Annals of Mathematical Statistics*, **35**, 261-269.

Poline, J.B., Worsley, K.J., Evans, A.C. and Friston, K.J. (1996). Testing for activations using both signal peak intensity and spatial extent. *NeuroImage*, **3**, S254.

Siegmund, D.O. and Worsley, K.J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, **23**, 608-639.

Talbot, J.D., Marrett, S., Evans, A.C., Meyer, E., Bushnell, M.C. and Duncan, G.H. (1991). Multiple representations of pain in human cerebral cortex. *Science*, **251**, 1355-1358.

Worsley, K.J., Evans, A.C., Marrett, S. and Neelin, P. (1992). A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**, 900-918.

Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of $\chi^2$, $F$ and $t$ fields. *Advances in Applied Probability*, **26**, 13-42.

Worsley, K.J. (1995a). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Annals of Statistics*, **23**, 640-669.

Worsley, K.J. (1995b). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability*, **27**, 943-959.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. and Evans, A.C. (1995c). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, **4**, 58-73.

Worsley, K.J., Poline, J-B., Vandal, A.C., and Friston, K.J. (1995d). Tests for distributed, non-focal brain activations. *NeuroImage*, **2**, 183-194.

Worsley, K.J. (1996a). An unbiased estimator for the roughness of a multivariate Gaussian random field. *Technical Report, Department of Mathematics and Statistics, McGill University.*

Worsley, K.J., MacDonald, D., Cao, J., Shafie, Kh. and Evans, A.C. (1996b). Statistical analysis of cortical surfaces. *Second International Conference on Functional Mapping of the Human Brain*, S108.