P-VALUES IN PROJECTION PURSUIT

by

Jiayang Sun
Stanford University

TECHNICAL REPORT NO. 2
AUGUST 1989

Prepared Under Contract 89-0355
For The Air Force Office of Scientific Research

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
P-VALUES IN PROJECTION PURSUIT

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF STATISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Jiayang Sun
August 1989
Contents

1 Introduction ........................................... 1
   1.1 Why P-values of PP ......................... 1
   1.2 Friedman's Projection Pursuit Index .... 3
   1.3 The Statistical Problems .................. 5
   1.4 Results ...................................... 6

2 Approximate Formulae for the Tail Probabilities ... 9
   2.1 Limiting Process (Field) ................. 9
   2.2 The Maximum of the Related Gaussian Random Field 12
   2.3 Connection with Differential Geometry .... 16
   2.4 Approximation Formulae for the Tail Probabilities 22

3 Geometric Meanings and Applicable Formulas ........ 35
   3.1 Metric Tensor and Covariance Function .... 36
      3.1.1 Unosphered Case ..................... 39
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.2 Sphered Case</td>
<td>42</td>
</tr>
<tr>
<td>3.2 The First Coefficient $\kappa_0$ in the Approximation Formula</td>
<td>43</td>
</tr>
<tr>
<td>3.2.1 Unsphere Case</td>
<td>44</td>
</tr>
<tr>
<td>3.2.2 Sphered Case</td>
<td>46</td>
</tr>
<tr>
<td>3.3 The Second Coefficient $\kappa_2$ in the Approximation Formula</td>
<td>47</td>
</tr>
<tr>
<td>3.3.1 Unsphere Case</td>
<td>47</td>
</tr>
<tr>
<td>3.3.2 Sphered Case</td>
<td>48</td>
</tr>
<tr>
<td>3.4 Tables and Their Construction</td>
<td>49</td>
</tr>
<tr>
<td>4 Comparisons with Monte Carlo Results</td>
<td>58</td>
</tr>
<tr>
<td>4.1 Notations and Algorithms in Simulation</td>
<td>58</td>
</tr>
<tr>
<td>4.2 Unsphere Case</td>
<td>62</td>
</tr>
<tr>
<td>4.3 Sphered Case</td>
<td>69</td>
</tr>
<tr>
<td>5 More about the Sphered Case</td>
<td>76</td>
</tr>
<tr>
<td>5.1 Centered Case</td>
<td>78</td>
</tr>
<tr>
<td>5.2 Sphered Case</td>
<td>82</td>
</tr>
<tr>
<td>6 Discussions and Remarks</td>
<td>90</td>
</tr>
<tr>
<td>6.1 Comparisons for Hall and Friedman's PP indices</td>
<td>93</td>
</tr>
<tr>
<td>6.2 Remarks and Conclusions</td>
<td>98</td>
</tr>
<tr>
<td>A Supplementary Theorems and Proofs for Chapter 2</td>
<td>112</td>
</tr>
<tr>
<td>B Supplementary Theorems and Proofs for Chapter 3</td>
<td>149</td>
</tr>
<tr>
<td>C Supplementary Tables for Chapter 4</td>
<td>170</td>
</tr>
</tbody>
</table>
# List of Tables

3.1 $A_{p,J} (A^*_{p,J})$ ................................................. 53
3.2 $B_{p,J} = \frac{1}{2^{p-1}} \cdot \mathcal{E} [ C (\beta)^{(p-1)/2} ]$ ................................................. 54
3.3 $C_{p,J} = \frac{1}{2p} \cdot \mathcal{E} [ D(p, J, C (\beta)) ]$ ................................................. 55
3.4 $B^*_{p,J} = \frac{1}{2^{p-1}} \cdot \mathcal{E} [ C^* (\beta)^{(p-1)/2} ]$ ................................................. 56
3.5 $C^*_{p,J} = \frac{1}{2p} \cdot \mathcal{E} [ D(p, J, C^* (\beta)) ]$ ................................................. 57

4.1 Case: $p = 3, J = 4$ ................................................. 67
4.2 Case: $p = 4, J = 4$, part I ................................................. 68
4.3 $\hat{Y}_i(\alpha)$ in Sphered Case ................................................. 70
4.4 Sphered Case: $p = 4, J = 4$ ................................................. 75

5.1 Limiting Variances of $\hat{Y}_j(\alpha)$ ................................................. 89

6.1 The First Experiment ................................................. 105
6.2 The Second Experiment ................................................. 106

C.1 Case: $p = 2, J = 1$ ................................................. 173
C.2 Case: $p = 2, J = 2$ ................................................. 174
C.3 Case: $p = 2, J = 4$ ................................................. 174
C.4 Case: $p = 3, J = 1, 2$ ........................................ 175
C.5 Case: $p = 4, J = 4$, part II .................................. 175
C.6 Case: $p = 4, J = 4$, part III ................................ 176
C.7 Case: $p = 6, J = 4$, part I .................................. 176
C.8 Case: $p = 6, J = 4$, part II ................................ 177
C.9 Case: $p = 8, J = 4$ ............................................ 177
C.10 Sphered Case: $p = 4, 6, J = 4$, (bad) ................. 178
C.11 $\tilde{Y}(\alpha)$ in Centered Case .............................. 179
C.12 Sphered Case: $p = 4, J = 4$ ............................... 179
C.13 Sphered Case: $p = 6, J = 4$ ............................... 180
C.14 Sphered Case: $p = 6, 8, J = 4$ ........................... 180
List of Figures

2.1 Tube of the One Dimensional Manifold in $S^2$. .................. 21

4.1 Approximate $\Phi(x)$ by $ax + 0.5$ ............................. 71
4.2 The $Q-Q$ Plot for $\hat{Y}_i(\alpha)$, $\alpha = (1, 0, 0, 0)$. .............. 74

6.1 The Mixture of Two Normals $g_\mu$ .............................. 95
6.2 The First Experiment. ........................................ 107
6.3 The Second Experiment. ....................................... 108
6.4 The Third Experiment. ........................................ 109
6.5 The Fourth Experiment. ....................................... 110
6.6 The Fifth Experiment. ........................................ 111

B.1 The Ordering Diagram in the Bivariate Case ...................... 159
Chapter 1

Introduction

Projection pursuit (PP) is concerned with finding interesting "non-linear" structure of a high dimensional (say, p-dim.) data set by some low dimensional (say, one-dim.) projections of the data. To fulfill this goal, finding interesting projections, we use a projection pursuit index $I(\alpha)$ (PP index), a function in all possible directions $\alpha = (\alpha_1, \ldots, \alpha_p)$ with $\alpha^T\alpha = 1$, and then to design and (or) use some optimization algorithm (PP algorithm) to obtain $\alpha^*$ which maximizes $I(\alpha)$, so that the "non-linear" structure of the data can be seen clearly after they are projected in the direction $\alpha^*$.

1.1 Why P-values of PP

The first PP algorithm was proposed by Friedman and Tukey in 1974. Since
then, there has been some literature on the subject (e.g. Friedman (1987), Huber (1985), Jones and Sibson (1987), and Öhrvik (1988)). However, there has been a concern about the significance of the so called interesting projection, or interesting structure, as R. Miller said in the discussion of Huber (1985):

"With the reference to this new technology (PP), I feel that there is a gap developing between practice and theory. While interesting shapes, clusters, and separations are now more discoverable in high dimensional data, how do we evaluate whether they are real or merely random variation? The theory for separating real from random seems somewhat limited at this point in time."

Again in a later paragraph, he augmented this argument with an example:

"Suppose the statistician through his/her visualization of the multidimensional data on the computer thinks that he/she sees some sort of smooth contour which effectively separates the populations. How does he/she evaluate whether this separation is real or random? The theory and critical constants to measure the significance of the separation into a 2 × 2 table, or the significance of coefficients in an equation defining the boundary, are unavailable at the present. The multiple comparisons problem involved in this evaluation is enormous."

On the other hand, heuristically, a projection is uninteresting if it is random or unstructured.

- Any suitable projection pursuit index essentially amounts a test for nonnormality according to Huber (1985). John Öhrvik (1988) also emphasized this viewpoint.
The P-value associated with a PP index gives us a reasonable calibration, i.e. how large is large, or how much structure can be found by using a particular PP index.

Therefore, it is helpful to develop a "significance test" for the departure from the "null" or "uninteresting" model in the PP setting, or to obtain the P-value of the "test statistic" which is the maximum PP index in the projection pursuit set up (cf. 2.2 in Huber (1985)).

Remark 1.1: A good PP procedure is one which finds multiple views of a multivariate data set, i.e. it enables one to keep "running" its PP algorithm based on its PP index to discover as many informative directions as possible (cf. Friedman, 1987). One of by-products of such P-values is that it helps us to decide when reasonably to stop the algorithm. This idea will be made clear in Section 4.2.

1.2 Friedman's Projection Pursuit Index

Based on the principle of seeking the "least normal" projection as an interesting one, Friedman (1987) proposed an algorithm for his new PP index described below, which has many computational advantages.

Suppose $Y_1, Y_2, \ldots, Y_p$ are $p$-dim. i.i.d. random variables from a population $F$; $y_1, y_2, \ldots, y_p$ are the corresponding $p$-dim. observations from same population $F$. In the following, if there is no specification, big letters are for random variables and small letters are for the corresponding observations. Assume for the moment that
\( \mathcal{E}(Y_i) = 0 \) and \( \operatorname{Cov}(Y_i) = \text{identity} \). Set

\[
Z_i = Y_i, \quad i = 1, \ldots, N. \tag{1.1}
\]

Then a consistent statement of the above principle is that a direction \( \alpha \) is interesting, or shows an interesting view after data are projected on it, if the density function \( p_\alpha(z) \) for the projected random variable \( X = \alpha^T Z \) is "very far" away from the density function \( \phi(z) \) of the random variable from \( \mathcal{N}(0,1) \), the normal population with mean zero and variance one.

Under the transformation

\[
X \longrightarrow R = 2\Phi(X) - 1, \tag{1.2}
\]

the density function \( p_\alpha(z) \) of random variable \( X = \alpha^T Z \) is transformed to the density function \( q_\alpha(r) \) of the transformed random variable \( R = 2\Phi(\alpha^T Z) - 1 \), and the density function \( \phi(z) \) of normal random variable from \( \mathcal{N}(0,1) \) is transformed to the density function \( f(r) = \frac{1}{2} \) of an uniform random variable on \((-1,1)\).

The \( L_2 \) distance between these two density functions for transformed random variables is

\[
\int_{-1}^{1} (q_\alpha(r) - \frac{1}{2})^2 \, dr.
\]

Expanding this integral into infinite series in terms of Legendre polynomials, \textit{viz.}

\[
\sum_{j=1}^{\infty} \frac{2j+1}{2} \left[ \mathcal{E}(P_j(R)) \right]^2,
\]

Friedman formed his PP index

\[
I_J(\alpha) = \sum_{j=1}^{J} \frac{2j+1}{2} \left( \frac{1}{N} P_j (2\Phi(\alpha^T Z_i) - 1) \right)^2, \tag{1.3}
\]
and considered \( \alpha^* = \max_{\alpha \in S^{p-1}} I_J(\alpha) \) as an interesting projection direction. Here

\[
S^{p-1} \overset{\text{def}}{=} \{ \alpha \in \mathcal{R}^p, \, \alpha^T \alpha = 1 \},
\]

\( \mathcal{E} \) stands for the expectation with respect to \( \mathcal{F} \), \( P_j \)'s are Legendre polynomials (see Section 2.1), and \( J \) is usually picked from 2 to 6 depending on \( p \).

In the following, we call the PP index defined in (1.3) \textit{JHF PP index}.

### 1.3 The Statistical Problems

Now we can give a precise meaning of Miller's question: how shall we test the null hypothesis that the data are just consistent with a realization of structureless Gaussian noise if we use the index defined in (1.3)? The answer is to try to calculate P-value under the null assumption \( H_0 \): \( Y_i \)'s are independent, identically distributed random samples from \( \mathcal{N}(0, I_p) \)

\[
P_o = P\{ \max_{\alpha \in S^{p-1}} I_J(\alpha) \geq I_{\text{obs}}^* \}
\]

for \( I_{\text{obs}} = I_J(\alpha^*) = \max_{\alpha \in S^{p-1}} I_J(\alpha) \) of the observed data set.

In practice, however, \( \mathcal{E}(Y_i) \) and \( \text{Cov}(Y_i) \) are arbitrary. In this case, we sphere the data first before implementing the PP procedure. One version of spherred data is

\[
Z_i = D^{-\frac{1}{2}} U^T(Y_i - \bar{Y}), \quad i = 1, \ldots, N.
\]

Here \( \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i \) is the sample mean, and \( U D U^T = \hat{\Sigma} \) is an eigenvalue-eigenvector decomposition of the estimate

\[
\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})(Y_i - \bar{Y})^T
\]
of the covariance matrix of $\mathcal{F}$.

If $Z_i$'s are defined as (1.1), we call them unsphered. If $Z_i$'s are defined as (1.5), we call them sophered. In (1.3), if $Z_i$'s are sophered, the JHF index (1.3) is location and scale invariant.

Sphering is very important in application (cf. Huber(1985), Jones and Sibson(1987)). It also introduces a great amount of computational saving, but causes extra difficulty for one to make inferences. In our case, we need to calculate a more difficult P-value than the one given in (1.4), viz.

$$P'_o = P\{ \max_{\alpha \in \mathcal{S}} I_J(\alpha) \geq I^*_{obs} \}$$

(1.6)

under $H'_o$: Y_i's are independent, identically distributed random samples from $p$-dimensional population $\mathcal{N}(\mu, \Sigma)$ for some $p$-dimensional vector $\mu$, $p \times p$ nonsingular matrix $\Sigma$, for $I^*_{obs} = I_J(\alpha^*) = \max_{\alpha \in \mathcal{S}} I_J(\alpha)$ of the observed sophered data set.

**Question:** What are the P-values in (1.4) and (1.6)?

It is natural to try to use some Monte-Carlo methods for calculating P-values in (1.4) and (1.6). For example, one can try to compute $I^*_{obs}$ for many Gaussian replications and estimate $P_o$, $P'_o$ from the empirical distribution. Unfortunately, Monte Carlo methods are computationally intensive and difficult for non experienced programmer to use in this case. Hence, it is ideal and important to search some analytical approach for P-values in (1.4) and (1.6).

### 1.4 Results
Under the null hypothesis that the data are an independent, identically distributed, sample from a \( p \)-dimensional normal population, Chapter 2 derives the limiting process for \( I_j(\alpha) \). This leads to a **general problem** for the tail probability of the extreme of differentiable Gaussian random fields. We then show the connection with integral and differential geometry based on Karhunen-Loève expansion, and derive a two term approximation for the tail probability by using Weyl's formula (1939) for the volume of a tube about a manifold embedded in the unit sphere in Euclidean space.

In Chapter 3, we show that two coefficients in the two term approximation mentioned above are the volume and some kind of total scalar curvature of the manifold. When the Gaussian random field is the one related to JHF PP index, the matrix of the metric tensor turns out to be diagonal and applicable formulas for those constants are given. This indicates that those extremely complicated constants in differential geometry can be calculated for a real world problem: P-values in PP. Then simple approximate formulas for the P-values of the maximum JHF PP index are obtained. Finally the tables of the computed constants are presented.

In Chapter 4, our theoretical result is compared with Monte Carlo simulations. As mentioned earlier, Monte Carlo methods in PP are computationally intensive and difficult to use, since they involve repeated solution of a difficult *optimization* problem. After experimenting with some “modern” optimizers and algorithms, we also propose modified algorithms to evaluate the JHF index.

In chapter 5, we deal with the sphered data case in detail. If the data are sphered before implementing PP procedure, there is some reduction of the dimensions, i.e. the first two terms in JHF PP index are negligible. This loss of dimensionality phenomenon occurs for all the 'polynomial' based PP indices, for example, the
index proposed by P. Hall (1989) which is based on Hermite functions (i.e. Hermite polynomial multiplied by the kernel $\exp\{-\frac{x^2}{2}\}$).

In Chapter 6, different indices (including JHF two dimensional PP index) are discussed. Especially, the JHF PP index is compared with another index suggested recently by P. Hall (1989). There are also some concluding remarks.

In Appendix A, there are supplementary Theorems and Proofs for Chapter 2. Garsia's sufficient condition for a uniformly convergent Karhunen-Loève expansion of a random field is presented. Several lemmas, two propositions, Weyl's formula and the generalized one term approximation (to non-homogeneous Gaussian random fields) from one in Adler (1981, p160) are used to prove two term (upper) approximation formula in Theorem 2.3 (Theorem 2.2).

In Appendix B, we introduce some useful geometrical definitions and give theorems and proofs related to the work of Chapter 3. A connection between the Karhunen-Loève expansion of a random field and the Fourier expansion of its covariance function is presented.

In Appendix C, there are supplementary Tables on comparisons of our two term approximation formula and Monte Carlo result for P-values in PP in Chapter 4.
Chapter 2

Approximate Formulae for the Tail Probabilities

2.1 Limiting Process (Field)

In this section, we use the central limit theorem to derive the limiting process for JHF's index $I_J(\alpha)$ defined in (1.3) as $N \to \infty$. Again here $N$ is the sample size.

The recursive definition of Legendre polynomials is that for $r \in [-1,1],$

\[
P_0(r) = 1 \\
P_1(r) = r \\
P_j(r) = \frac{1}{j}((2j-1)rP_{j-1}(r) - (j-1)P_{j-2}(r)), \quad \text{for } j = 2, 3, \ldots
\]  

(2.1)

From this definition, we have the following well known lemma which is useful in
deriving the limiting process.

**Lemma 2.1** For Legendre Polynomials defined in (2.1),

\[
\int_{-1}^{1} P_i(r) \, dr = 0, \text{ for all } i \geq 1,
\]

\[
\int_{-1}^{1} P_i(r) \cdot P_j(r) \, dr = \begin{cases} 
\frac{2}{2i+1} & \text{if } i = j \\
0 & \text{otherwise}
\end{cases}
\]

\[
\square
\]

Define

\[
\tilde{Y}_j(\alpha) = \sqrt{\frac{2j+1}{N}} \sum_{i=1}^{N} P_j(2\Phi(\alpha^T Z_i) - 1),
\]

(2.2)

for \( j = 1, \ldots, J \). Then by (1.3), The JHF index can be rewritten as

\[
I_J(\alpha) = \frac{1}{2N} \sum_{j=1}^{J} (\tilde{Y}_j(\alpha))^2.
\]

(2.3)

We consider first the simpler case, where the data are unsphered.

In this case, \( Z_1, Z_2, \ldots \) are \( p \)-dimensional independent, identically distributed random variables from \( \mathcal{N}(0, I_p) \), so \( P_j(2\Phi(\alpha^T Z_1) - 1) \), \( P_j(2\Phi(\alpha^T Z_2) - 1) \), \ldots are also independent, identically distributed random variables for each fixed \( j = 1, \ldots, J \) and \( \alpha \in S^{p-1} \). Suppose \( R \) is a r.v. with uniform density \( \frac{1}{2} \) on \([-1, 1]\). By Lemma 2.1, we see that

\[
\mathcal{E}[\tilde{Y}_j(\alpha)] = \sqrt{\frac{2j+1}{N}} \mathcal{E}[\sum_{i=1}^{N} P_j(2\Phi(\alpha^T Z_i) - 1)]
\]

\[
= c \cdot \mathcal{E}[P_j(R)] = 0,
\]

(2.4)
\[
\text{Cov} [\tilde{Y}_j(\alpha), \tilde{Y}_i(\alpha)] = \sqrt{(2j + 1)(2i + 1)} \cdot \text{Cov} [P_j(R), P_i(R)] = \delta_{ij}, \tag{2.5}
\]

where \(\text{Cov}(X, Y)\) is the covariance for two random variables \(X\) and \(Y\), \(\delta_{ij}\) is 1 if \(i = j\), is 0 if \(i \neq j\).

Further by applying multivariate central limit theorem, we have as \(N \to \infty\),

\[
\begin{pmatrix}
\tilde{Y}_1(\alpha) \\
\tilde{Y}_2(\alpha) \\
\vdots \\
\tilde{Y}_J(\alpha)
\end{pmatrix} \overset{\mathcal{L}}{\longrightarrow} \begin{pmatrix}
\tilde{X}_1(\alpha) \\
\tilde{X}_2(\alpha) \\
\vdots \\
\tilde{X}_J(\alpha)
\end{pmatrix} \overset{\text{dis}}{\sim} N(0, I_J), \tag{2.6}
\]

where \(\overset{\text{dis}}{\sim}\) stands for "is distributed as". Hence, it is easy to prove that as \(N \to \infty\),

\[
2N \cdot I_J(\alpha) = \sum_{j=1}^{J}(\tilde{Y}_j(\alpha))^2 \overset{\mathcal{L}}{\longrightarrow} \sum_{j=1}^{J}(\tilde{X}_j(\alpha))^2 \overset{\text{dis}}{\sim} \chi^2_J(\alpha), \tag{2.7}
\]

i.e. the limiting process (field) is a central Chi-square process (field) with degree of freedom \(J\) and a covariance function related to the matrix \(A\) in Proposition 2.1 on page 14. We do not specify this covariance function explicitly as it is not needed for us to achieve the goal of this thesis.

Now, suppose that the data \(Z_i\)'s are taken as the spherered version of \(p\) dimensional random variables \(Y_i\)'s, which are from \(N(\mu, \Sigma)\) for some \(p\)-dimensional vector \(\mu, p \times p\) nonsingular matrix \(\Sigma\). One might think that we just have eliminated the effect of unknown scale and location, but otherwise the above argument on limiting behavior of \(2N I_J(\alpha)\) should also work in this spherered case. We shall examine this situation in a later chapter.
2.2 The Maximum of the Related Gaussian Random Field

Let \( \tilde{X}(\alpha) = (\tilde{X}_1(\alpha), \ldots, \tilde{X}_J(\alpha)) \) with \( \tilde{X}_j(\alpha) \) as the limiting random field defined in (2.6) for \( j = 1, \ldots, J \). Then \( \tilde{X}(\alpha) \) is a Gaussian random field indexed by point (or vector \( \alpha \)) in the unit sphere \( S^{p-1} \) with mean vector 0. Recall that we are to calculate the tail probability of \( \max_{\alpha \in S^{p-1}} I_J(\alpha) \), which in turn is approximately the tail probability of the maximum of a Chi-square field as the result of (2.7). Since a Gaussian random field is the easiest random field for one to deal with, in the following we convert the tail probability of the maximum of a Chi-square field into the tail probability of the maximum of a Gaussian random field by the fact (cf. almost any standard text book on Linear Algebra):

\[
\| \tilde{X} \| = \max_{\| \beta \|_1} \langle \beta, \tilde{X} \rangle = \max_{\beta \in S^{J-1}} \beta^T \tilde{X} \tag{2.8}
\]

(The usefulness of (2.8) was suggested by D. Siegmund).

Define \( Z(\alpha, \beta) \equiv \beta^T \tilde{X}(\alpha) \) and \( \frac{1}{2} S^{p-1} \equiv \{ \alpha : \alpha = (\alpha_1, \ldots, \alpha_p) \in S^{p-1}, \alpha_p \geq 0 \} \). Then from (1.3), (2.7),

\[
P\{ \max_{\alpha \in S^{p-1}} I_J(\alpha) \geq a \}
\]

\[
= P\{ \max_{\alpha \in \frac{1}{2} S^{p-1}} I_J(\alpha) \geq a \} \tag{2.9}
\]

\[
= P\{ \max_{\alpha \in \frac{1}{2} S^{p-1}} \sqrt{2N \cdot I_J(\alpha)} \geq \sqrt{2aN} \}
\]

\[
\approx P\{ \| \tilde{X}(\alpha) \| \geq b \}
\]

\[
= P\{ \max_{\alpha \in \frac{1}{2} S^{p-1}} \max_{\beta \in S^{J-1}} \beta^T \tilde{X}(\alpha) \geq b \} \tag{2.10}
\]

12
\[ = \mathcal{P}\{ \max_{\alpha \in \mathcal{S}^{p-1}, \beta \in \mathcal{S}^{j-1}} Z(\alpha, \beta) \geq b \} \quad (2.11) \]
\[ \equiv \mathcal{P}\{ \max_{t \in I} Z(t) \geq b \} \quad (2.12) \]

where \( b = \sqrt{2aN}, \| \vec{X} \| = \sqrt{\tilde{X}_1^2 + \tilde{X}_2^2 + \ldots + \tilde{X}_J^2} \). Here (2.9) is by the symmetry of JHF PP index in upper hemisphere and in lower hemisphere and (2.10) in the fourth equation follows from Equation (2.8). \( Z(\alpha, \beta) = \beta^T \hat{X}(\alpha) \) is a linear combination of the Gaussian random fields, and therefore is a Gaussian random field. The parameter space of this Gaussian random field is \( d = p + J - 2 \) dimensional, i.e. there is a reparametrization in terms of \( t \) so that the tail probability in (2.11) can be represented as the tail probability (2.12) of the maximum of a Gaussian random field \( Z(t) \) for \( t \in I \), a \( d \)-dimensional space.

One typical example of such a reparametrization is to use the ordinary polar coordinate transformation as follows:

\[
\begin{align*}
\alpha_1 &= r \cos \theta_1, \\
\alpha_2 &= r \sin \theta_1 \cos \theta_2, \\
\alpha_3 &= r \sin \theta_1 \sin \theta_2 \cos \theta_3, \\
\vdots \\
\alpha_{p-1} &= r \sin \theta_1 \sin \theta_2 \sin \theta_3 \ldots \sin \theta_{p-2} \cos \theta_{p-1}, \\
\alpha_p &= r \sin \theta_1 \sin \theta_2 \sin \theta_3 \ldots \sin \theta_{p-2} \sin \theta_{p-1},
\end{align*}
\]

where \( r = 1; \theta_1, \theta_2, \ldots, \theta_{p-1} \in [0, \pi] \); 

\[
\begin{align*}
\beta_1 &= \rho \cos \varphi_1, \\
\beta_2 &= \rho \sin \varphi_1 \cos \varphi_2, \\
\beta_3 &= \rho \sin \varphi_1 \sin \varphi_2 \cos \varphi_3,
\end{align*}
\]
\[
\beta_{J-1} = \rho \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \ldots \sin \varphi_{J-2} \cos \varphi_{J-1},
\]
\[
\beta_J = \rho \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \ldots \sin \varphi_{J-2} \sin \varphi_{J-1},
\]
where \( \rho = 1; \ \varphi_1, \varphi_2, \ldots, \varphi_{J-2} \in [0, \pi]; \ \varphi_{J-1} \in [0, 2\pi], \)
where \( \varphi_{J-1} = 0 \) and \( \varphi_{J-1} = 2\pi \) represent the same point. Under this reparametrization, \( Z(t) \) is a \( d = p + J - 2 \) dimensional Gaussian random field indexed by
\[
t = (t_1, t_2, \ldots, t_d) = (\theta, \varphi) = (\theta_1, \theta_2, \ldots, \theta_{p-1}, \varphi_1, \varphi_2, \ldots, \varphi_{J-1})
\]
on a \( d \)-dimensional space \( I \equiv [0, \pi] \times \ldots \times [0, \pi] \times [0, 2\pi] \).

**Proposition 2.1** The random field \( Z(t) \) induced from JHF PP index, is a \( d = p + J - 2 \) dimensional differentiable Gaussian random field on \( I \), with mean zero, and the following covariance function \( r(t, t') \) for \( t, t' \in I \),

\[
r(t, t') \\
= \text{Cov}(Z(t), Z(t')) \\
= \mathbb{E}[\beta^T \tilde{X}(\alpha) \cdot \beta^T \tilde{X}(\alpha')] \\
= \beta^T A \beta' \\
= 1 \quad \text{if} \ t = t'.
\]

Here \( A = \mathbb{E}[\tilde{X}(\alpha) \cdot \tilde{X}^T(\alpha')] \equiv (a_{ij}(\alpha^T \alpha'))_{J \times J} \) is a \( J \times J \) matrix, with element \( a_{ij}(\alpha^T \alpha') \) being \( C^\infty \) differentiable \( \text{i.e.} \ a_{ij} \) has \( k \)-th derivative with respect to \( \theta \) and \( \theta' \) for any positive integer \( k \), and depending only on \( \alpha^T \alpha' \).

In particular, as \( \|\theta - \theta'\|^2 + \|\varphi - \varphi'\|^2 \rightarrow 0 \),
\[
r(s, t) = 1 - a_1(\varphi, \varphi')\|\varphi - \varphi'\|^2 - a_2(\theta, \theta', \varphi)\|\theta - \theta'\|^2 + o_p(\|\theta - \theta'\|^2 + \|\varphi - \varphi'\|^2),
\]
(2.14)
where $a_1(\varphi, \varphi'), a_2(\theta, \theta', \varphi)$ are bounded on $I$.

**Proof:** (Idea of the proof. For details, see Appendix A)

Use equation (2.1)-(2.2), (2.4)-(2.6), and the fact that $X(\alpha)$ and $X(\alpha')$ are from $\mathcal{N}(0,1)$ with correlation coefficient $\rho = \alpha^r \alpha'$, as well as Laplace approximation (cf. (4.2.4) on p65 in Bruijin, 1961).

\[\square\]

**Remark 2.1:** If $p = 2, J = 1$, then $d = 1$. In this case, the Gaussian random field derived from JHF PP index is a stationary process, since $r(t, t') = a_{11}(\alpha^r \alpha') = b(\theta_1 - \theta_1')$ for some function $b(\cdot)$. This field is not homogeneous if its dimension $d$ is greater than or equal to 2.

Accordingly, we can introduce a general problem from this concrete problem in Projection Pursuit, which is to calculate the tail probability of the extreme of differentiable Gaussian random field

\[
P\{ \max_{t \in I} Z(t) \geq b \} \tag{2.15}
\]

under the assumption that $Z(t)$ is a differentiable Gaussian random field with mean zero, variance one. According to Adler (1981, p159), Leadbetter, Lindgren, and Rootzen (1983), if $d = 1$, there are only five special cases, i.e. for five kinds of covariance functions, for which we know the exact results for the tail probability in equation (2.15). For other cases, we can only obtain approximate results. If this Gaussian random field is homogeneous with dimension $d \geq 2$, there is one term approximation formula for (2.15) given by Adler (1981, p160). This one term approximation formula is not good enough for our problem. First, our $Z(t)$ is not homogeneous for $d \geq 2$ as mentioned in Remark 2.1. Second, as shown in a later
chapter, we need at least a two term approximation for those statistically interesting cases, \( i.e. \ p \geq 4, \)  which imply \( d = p + J - 2 \geq 3 \) for \( J \geq 1. \)

Fortunately, there is a neat connection between this kind of Gaussian random field and some differential geometry related to the Karhunen-Loève expansion. This connection helps us obtain an approximation to (2.15).

2.3 Connection with Differential Geometry

Even in the 1930’s, there were some statistical questions inspiring the calculation of the volume of a tube about a manifold. For example, suppose that we consider the following regression model

\[
y_i = \beta f_i(\theta) + \epsilon_i, \quad i = 1, 2, ..., n.
\]

Then the problem of finding the significance level in testing the hypothesis \( H_0 : \beta = 0 \) against \( H_1 : \beta \neq 0, \) based on the likelihood ratio statistic, is reduced to the problem of calculating the volume of a tube about some manifold. Here the \( f_i \)'s are known functions of an unknown parameter \( \theta, \) and \( \epsilon_1, ..., \epsilon_n \) are \( i.i.d. \ \mathcal{N}(0, \sigma^2) \) random variables (cf. Hotelling, 1939. For recent applications, see Knowles and Siegmund, 1988). In this section, we present the connection between a differentiable random field and a differentiable manifold, and then introduce some definitions related to the tube of a manifold embedded in a unit sphere, as well as Weyl's formula (1939) for volume of the tube. These preliminaries are useful for deriving approximate formulas in the next section.

Assume \( \{Z(t), t \in I\}, \) is a \( d \) dimensional Gaussian random field with \( \mathcal{E}[Z(t)] = \)
0, $\mathbb{E}[(Z(t))^2] = 1$ and a uniformly convergent Karhunen-Loève expansion in $t \in I$, viz.

$$Z(t) = \sum_{i=1}^{\infty} u_i(t)X_i = <u(t), X>,$$

(2.16)

where $u(t) = (u_1(t), u_2(t), \ldots)$ satisfies $\|u\| = [\sum_{i=1}^{\infty} u_i^2(t)]^{1/2} = 1$, and $X_1, X_2, \ldots$ are independent, identically distributed $\mathcal{N}(0, 1)$ random variables. Note the Karhunen-Loève expansion defined here does not require the orthogonality

$$\int_I u_i(t) u_j(t) \, dt = 0, \quad \text{for } i \neq j$$

(2.17)
as the standard Karhunen-Loève expansion does. The standard Karhunen-Loève expansion is a Karhunen-Loève expansion defined as above with additionally that $u_i(t)$'s satisfy the orthogonal relation in (2.17). (cf. Adler(1981), Yaglom(1987) and discussion in Remark 2.2.)

Set

$$\tilde{Z}_k(t) = \sum_{i=1}^{k} u_i(t)X_i,$$

(2.18)

$$Z_k(t) = \sum_{i=1}^{k} v_i(t)X_i$$

(2.19)

with $v_i(t) = \frac{u_i(t)}{\sigma_k(t)}$ for $i = 1, 2, \ldots, k$, $\sigma_k(t) = \sqrt{\sum_{i=1}^{k} (u_i(t))^2}$ for $k = 1, 2, \ldots$. Then

$$\mathcal{V}^k = \{v^k(t) : t \in I, v^k(t) = (v_1(t), \ldots, v_k(t))\}$$
is a $d$ dimensional manifold embedded in $S^{k-1}$, the unit sphere in $\mathcal{R}^k$; and

$$\mathcal{U} = \{u(t) : t \in I, u(t) = (u_1(t), u_2(t), \ldots)\}$$
is a $d$ dimensional manifold embedded in $S^{\infty-1}$, the unit sphere in $\mathcal{R}^\infty$. Here

$$S^{k-1} = \{x : x = (x_1, \ldots, x_k) \in \mathcal{R}^k, \sum_{i=1}^{k} x_i^2 = 1\},$$
\[ S^{\infty -1} = \{ x : x = (x_1, x_2, \ldots) \in \mathcal{R}^\infty, \sum_{i=1}^{\infty} x_i^2 = 1 \}. \]

Hence, there is an \textit{one-to-one} correspondence between a differentiable manifold embedded in the unit sphere, and a differentiable Gaussian random field through a uniformly convergent Karhunen-Loève expansion of the random field.

**Remark 2.2:** (Existence of a Karhunen-Loève expansion) Let \( r(s, t) \) be the covariance function of a \( d \) dimensional non singular differentiable random field \( Z(t) \) with mean zero, unit variance. In general, by Mercer’s theorem \( r(s, t) \) has an absolutely and uniformly convergent eigenvalue \( (\lambda_l) \) eigenfunction \( (\Lambda_l(t)) \) expansion:

\[ r(s, t) = \sum_{l=1}^{\infty} \lambda_l \Lambda_l(s) \Lambda_l(t), \]

where \( \lambda_l \)'s are in decreasing order \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \) and

\[ \int_I r(s, t) \Lambda_l(t) \, dt = \lambda_l \Lambda_l(s), \]

\[ \int_I \Lambda_l(t) \Lambda_m(t) \, dt = \delta_{lm} = \begin{cases} 1 & \text{if } l \neq m \\ 0 & \text{otherwise} \end{cases}. \]

This expansion is called the \textit{Mercer expansion} (cf. pp138-140, Courant and Hilbert (1953)). By Garsia (1972), under an additional mild condition on \( r(s, t) \) (cf. Lemma A.1), there exist independent, identically distributed \( \mathcal{N}(0, 1) \) random variables \( X_1, X_2, \ldots \) such that

\[ \sum_{l=1}^{k} u_l(t) X_l \to Z(t) \quad \text{uniformly in } t, \quad a.s. \]

where \( u_l(t) = \sqrt{\lambda_l} \Lambda_l(t) \) for \( l = 1, 2, \ldots \) Therefore, there exists a Karhunen-Loève expansion to a random field in most cases.
Definition 2.1 (Tube) The tube with radius $r$ of a manifold $U^k = \{u^k(t) \in (u_1(t), \ldots, u_k(t)), t \in I\}$ embedded in $S^{k-1}$ is

$$T(r) = \{y : y \in S^{k-1}, \inf_{t \in I} \|y - u^k(t)\| \leq r\}, \quad (2.20)$$

where $k$ is finite or infinite.

If $d = 1, k = 3$, the manifold embedded in $S^{k-1}$ can be represented as a line in the following picture (Figure 2.1) on page 21, the tube of this manifold is the dotted area around this line on the unit sphere $S^2$.

Definition 2.2 (Critical) The critical radius of the tube $T(r)$ is

$$r_{kc} = \inf\{r : r \geq 0, T(r) \text{ has selfoverlap}\},$$

the critical point of the tube is

$$d_k = \frac{r_{kc}^2}{2 - r_{kc}^2}.$$

The semi critical radius of the tube $T(r)$ is

$$\tilde{r}_{kc} = \inf\{r : r \geq 0, \exists y(t, \xi) = u^k(t) + \sum_{i=1}^{k-d-1} \xi_i n_i^k(t), \frac{y(t, \xi)}{\|y(t, \xi)\|} \in T(r), \text{ s.t. } J(y) = 0\}.$$

Here $\exists$ means "there exists", "s.t." means "such that", $n_i^k(t), i = 1, 2, \ldots, k - d - 1$ are mutually orthogonal unit normal vectors of $U^k$ at $t$ and are orthogonal to $u^k(t)$, and $J(y)$ is the volume element at $y = y(t, \xi) = y_1(t, \xi), \ldots, y_k(t, \xi))$:

$$J(y) \equiv \left\| \begin{array}{cccc}
\frac{\partial y_1}{\partial t_1} & \ldots & \frac{\partial y_1}{\partial t_d} & n_1^k(t), n_2^k(t), \ldots, n_{k-d-1}^k(t) \\
\vdots & \ddots & \vdots & \vdots \\
\frac{\partial y_k}{\partial t_1} & \ldots & \frac{\partial y_k}{\partial t_d} & \frac{\partial y_k}{\partial \xi_{k-d-1}}
\end{array} \right\|.$$

$$\equiv \left| \begin{array}{cccc}
y_1 & \frac{\partial y_1}{\partial t_1} & \ldots & \frac{\partial y_1}{\partial t_d} \\
\vdots & \ddots & \vdots & \vdots \\
y_k & \frac{\partial y_k}{\partial t_1} & \ldots & \frac{\partial y_k}{\partial t_d}
\end{array} \right|,$$

19
the semi critical point of the tube is

\[ \tilde{d}_k = \frac{\tilde{r}_{kc}^2}{2 - \tilde{r}_{kc}^2}, \]

where \( T(r) \) is defined in (2.20).

According to the inverse function theorem (cf. almost any standard text book on Calculus),

\[
\{ \exists \ y(t, \xi) = u^k(t) + \sum_{i=1}^{k-d-1} \xi_i n_i^k(t), \ \frac{y(t, \xi)}{\|y(t, \xi)\|} \in T(r), \ \text{s.t.} \ J(y) = 0 \}
\subset\{ \exists \ y(t, \xi) = u^k(t) + \sum_{i=1}^{k-d-1} \xi_i n_i^k(t), \ \frac{y(t, \xi)}{\|y(t, \xi)\|} \in T(r), \ \text{s.t.} \ \exists t' \neq t, \]

\[ t' \in \text{a small neighbourhood of } t, y(t, \xi) = y(t', \xi') = u^k(t') + \sum_{i=1}^{k-d-1} \xi'_i n_i^k(t') \}
\subset\{ r : r \geq 0, T(r) \text{ has selfoverlap} \},

hence \( r_{kc} \leq \tilde{r}_{kc} \).

**Lemma 2.2 (Weyl)** Suppose \( U^k \) is a smooth manifold (without boundary) imbedded in \( S^{k-1} \), \( k < \infty \). Then the volume of the tube \( T(r) \) defined in Definition 2.1 is

\[
V(r) \begin{cases} 
= \frac{2 \pi^{rac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} \sum_{\epsilon=0, \text{even}}^{d} \kappa_\epsilon \cdot J_\epsilon(\theta) & \text{if } r \leq r_{kc} \\
\leq \frac{2 \pi^{rac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} \sum_{\epsilon=0, \text{even}}^{d} \kappa_\epsilon \cdot J_\epsilon(\theta) & \text{if } r \leq \tilde{r}_{kc}
\end{cases}
\]

(2.21)

where \( r_{kc} \) and \( \tilde{r}_{kc} \) are defined in Definition 2.2, \( m = k - d - 1 \), \( \kappa_\epsilon \)'s are certain integral invariants of the manifold, \( \theta \) is the corresponding spherical radius related to
radius \( r \), i.e.

\[
\theta = \arccos(\omega), \quad \omega = 1 - \frac{r^2}{2},
\]

\( J_e(\theta) \)'s are certain incomplete beta functions

\[
J_0(\theta) = \int_0^\theta \sin^{m-1}(x) \cos^d(x) \, dx,
\]

\[
m(m + 2) \ldots (m + e - 2) \cdot J_e(\theta) = \int_0^\theta \sin^{m+e-1}(x) \cos^{d-e}(x) \, dx.
\]

for \( e = 2, 4, \ldots, \leq d \).

\[
\square
\]

In the simple case: \( d = 1, k = 3 \), the first term in equation (2.21) says that if the radius \( r \) is small enough (smaller than the critical radius) the dotted area in unit sphere \( S^2 \) (see Figure 2.1) equals to \( 2\kappa_0 \sin(\theta) \) if the curve is closed and
approximately equals to $2\kappa_0\theta$ if the curve is not closed, where $\kappa_0$ is the area of the manifold, which is the length of the line in this case.

2.4 Approximation Formulae for the Tail Probabilities

In this section, we give a $d$ terms approximation formula for the tail probability of the extreme of a Gaussian random field (Theorem 2.1) if the corresponding Karhunen-Loève expansion is finite, a two term (upper) approximation formula in Theorem 2.3 (Theorem 2.2) if the corresponding Karhunen-Loève expansion is infinite. Thereby, we obtain the approximate formulas (Corollary 2.1, Equation (2.30)) for the P-value of PP when data are sphered, and unsphered, respectively.

Lemma 2.3 Suppose a $d$-dimensional twice differentiable random field $Z(t)$ on $I$, a compact subset of Euclidian space $\mathcal{R}^d$, is non-singular and has mean zero and a finite Karhunen-Loève expansion

$$Z(t) = \sum_{i=1}^{k} u_i(t)X_i,$$

where $\parallel u(t) \parallel = 1$ for $u(t) = (u_1(t),...,u_k(t))$, and $X_1,X_2,...,X_k$ are independent, identically distributed from $\mathcal{N}(0,1)$. Then for the corresponding manifold $U^k = \{u^k(t): t \in I, u^k(t) = (u_1(t),...,u_k(t))\}$, the critical radius $r_{kc}$, as defined in definition 2.2, is positive.

Proof: Notice that $I$ is compact, $u^k(t)$ is twice differentiable on $I$. If $r_{kc} = 0$, 22
there exist \( t_1, t_2 \in I, t_1 \neq t_2 \), such that

\[
u^k(t_1) = u^k(t_2),
\]

which is equivalent to

\[
Z(t_1) = Z(t_2) \text{ a.s. ,}
\]

by the assumed finite representation for \( t \in I \).

However,

\[
P\{Z(t_1) = Z(t_2)\}
= \int_{-\infty}^\infty P\{Z(t_1) = r \mid Z(t_2) = r\} P\{Z(t_2) \in dr\}
= \int_{-\infty}^\infty 0 \cdot P\{Z(t_2) \in dr\}
= 0.
\]

Here, for (2.22) we used the fact that the random variable \( Z(t_1) \), conditional on \( Z(t_2) = r \), is still a continuous Gaussian random variable, since \( Z(t), t \in I \), is a non-singular random field.

This contradiction shows that \( r_{kc} > 0 \).

\[\square\]

**Theorem 2.1** Let \( Z(t) \) be a \( d \)-dimensional non singular differentiable random field on a bounded \( d \)-dimensional closed space \( I \), with mean zero, unit variance and covariance function \( r(s, t) \). If \( r(s, t) \) has finite expansion

\[
r(s, t) = \sum_{i=1}^k u_i(s)u_i(t), \quad k < \infty,
\]
and the manifold $U^k = \{u^k(t) : u^k(t) = (u_1(t), \ldots, u_k(t)), t \in I\}$ has no boundary, then as $z \to \infty$,

$$P\{\max_{t \in I} Z(t) \geq z\} = \kappa_0\psi_0(z) + \kappa_2\psi_2(z) + \ldots + \kappa_{\bar{d}}\psi_{\bar{d}}(z) \cdot (1 + o(1)). \quad (2.23)$$

Here $\kappa_0, \kappa_2, \ldots, \kappa_{\bar{d}}$ are the same constants as those in Weyl's formula for the manifold $U = \{u(t) : u(t) = (u_1, \ldots, u_k(t)), t \in I\}$ (see Lemma 2.2), and

$$\psi_0(z) = \frac{1}{2\pi^{(d+1)/2}} \int_{z^2/2}^{\infty} u^{(d+1)/2} \exp\{-u\} \, du \quad (2.24)$$

$$\psi_\varepsilon(z) = \frac{1}{2^{1+\varepsilon/2} \pi^{(d+1)/2}} \int_{z^2/2}^{\infty} u^{(d+1-\varepsilon)/2} \exp\{-u\} \, du \quad (2.25)$$

for $\varepsilon = 2, 4, \ldots, \bar{d}$, where $\bar{d} = d$ if $d$ is even, $\bar{d} = d - 1$ if $d$ is odd.

**Proof:** (Idea of the proof. For details, see Appendix A.)

First, define $\tilde{Z}(t) = \sum_{i=1}^{k} u_i(t) X_i$, where $X_1, X_2, \ldots, X_k$ are independent, identically distributed $\mathcal{N}(0,1)$ random variables. Then $\tilde{Z}(t)$ and $Z(t)$ are identically distributed, and hence

$$P\{\max_{t \in I} Z(t) \geq z\} = P\{\max_{t \in I} \tilde{Z}(t) \geq z\}.$$

Second, by Lemma 2.3 and Definition 2.2, the critical value $d_k$, of the tube of $U^k = \{u(t) : u(t) = (u_1(t), \ldots, u_k(t)), t \in I\}$, is positive.

Third, since

$$U = \left( \frac{X_1}{\|X\|}, \frac{X_2}{\|X\|}, \ldots, \frac{X_k}{\|X\|} \right)$$

is uniformly distributed on unit sphere $S^{k-1}$, and is independent of $\|X\|$, which is distributed as a Chi random variable with $k$ degrees of freedom, we can rewrite the
tail probability as follows:

\[
P\{\max_{t \in I} Z(t) \geq z\} \\
= P\{\max_{t \in I} \tilde{Z}(t) \geq z\} \\
= \int_{\substack{x}}^\infty P\{\max_{t \in I} < u^k(t), U > \geq \frac{z}{y} \} \, P\{\|X\| \in dy\} \\
= \int_{z}^{(1+d_k)z} + \int_{(1+d_k)z}^{\infty} \equiv A + B.
\]

Fourth, we use Weyl's formula for the integrand of A, and elementary probability inequalities for B and the remainder from A, we then conclude the theorem.

One key assumption in Theorem 2.1 is that \( r(s, t) \) has a finite term expansion, or there exists a \( \tilde{Z}(t) \), which is identically distributed as \( Z(t) \) and has a finite Karhunen-Loève expansion. However, in most cases, it is hard to prove that \( r(s, t) \) has a finite term expansion as in Theorem 2.1. The reasons are as follows

1). \( r(s, t) \) usually has infinitely many eigenfunctions, and hence the Mercer expansion of \( r(s, t) \) (or the standard Karhunen-Loève expansion of \( Z(t) \)) is an infinite term expansion.

2). There is no theory to guarantee the existence of a finite term expansion of \( r(s, t) \) as in Theorem 2.1, which is not eigenvalue eigenfunction expansion of \( r(s, t) \).

For the problems in PP, we know that the random field corresponding to the index based on Hermite polynomials, suggested by Johansen and Johnstone (1985) for Projection Pursuit Regression, has a finite Karhunen-Loève expansion; but we can
not prove and doubt that there is a finite Karhunen-Loève expansion of the random field corresponding to the index based on Legendre polynomials (see Equation (1.3)), proposed by Friedman (1987) for (Exploratory) Projection Pursuit. Therefore, it is safer if we have a result which holds even if the Karhunen-Loève expansion is infinite, which also contributes to extreme value theory, viz. giving a general two term (upper) approximation under assumptions on the covariance function of the random field $Z(t)$. Of course, if the expansion is infinite, it is a much harder problem, since we can not use Weyl's formula as in the finite expansion case. In the following Theorem 2.3 (Theorem 2.2), under some regularity conditions, we derive a two term (upper) approximation for the tail probability of the maximum of the Gaussian random field, which are the same first two terms in the approximation formula of Theorem 2.1. To get a higher order approximation, we have to have stronger conditions, and in principle it can be done.

Another key assumption in Theorem 2.1 is that the manifold $U^k$ formed from the expansion of $r(s,t)$ has no boundary. This assumption leads to the following questions:

1). How does one know $U^k$ has no boundary?

2). Is there an approximation similar to (2.23) if $U^k$ has boundary?

For the question 1), the answer is easy to find if we know $U^k$ explicitly, for example, in the case of Projection Pursuit Regression index suggested by Johansen and Johnstone (1988). Otherwise, one can try to apply the following two criterions.

a). If the image of $U^k$ is a closed and open set of some space, $U^k$ has no boundary.
b). If there is an one to one continuous mapping from the parameter space $I$ to $S^d = \{\alpha : \alpha = (\alpha_1, \ldots, \alpha_{d+1}), \|\alpha\| = 1\}$, and $u^k(t)$ is continuous, $U^k$ has no boundary.

The criterion b) is easy to use, but is not a necessary condition for a manifold to have no boundary. On the other hand, the related manifold is intrinsically determined by the covariance function of $Z(t)$. Hence some day one might also be able to tell if $U^k$ has boundary by checking the properties of $r(s, t)$. We leave this question open to readers. For the question 2), the answer is yes. If $U^k$ has boundary, Weyl's formula should be modified. In this case, the expansion of the volume of the tube has an extra term with order between $J_0(\theta)$ and $J_2(\theta)$, and the coefficient of $J_2(\theta)$ is $\kappa_2 + c$, $c$ is another intrinsic invariant of the manifold. (See the discussion at the end of Knowles and Siegmund, 1988, for the case $d = 2$.) When an explicit formula, for the volume of a tube about a general $d$ dimensional manifold embedded in $S^{k-1}$, is available, a reader can obtain an approximation to $P\{\max_{t \in I} Z(t) \geq z\}$ by using similar tactics as in the proof of Theorem 2.1.

In the rest of this section, we discuss and give Theorem 2.2, 2.3, where the expansion of $r(s, t)$ is not assumed to be a finite expansion, and present the applications of Theorem 2.2, 2.3.

Notations and Definitions

A real vector function $f = (f_1, \ldots, f_n)$ is called full rank, if there is no Borel measurable function $F$ such that

$$f_i = F(f_1, \ldots, \tilde{f}_i, \ldots, f_n)$$

for any $i = 1, \ldots, n$, where the $\tilde{}$ denotes missing. If a (vector) function $f$ has $l$-th
order mixed continuous partial derivatives for \( l = 1, 2, \ldots, n \) on its domain \( I \), we denote \( f \in C^n(I) \).

Let \( r(s,t) \) be the covariance function of a random manifold \( Z(t) \) on a \( d \) dimensional space with mean zero and unit variance. The four regularity conditions R.1-R.4 of \( r(s,t) \) are as follows:

**R.1.** One of the following is true for some positive integer \( m \).
1. There exist functions \( f \) and \( g \), such that
\[
r(s,t) = g(f(s) - f(t)),
\]
where \( f, g \in C^{md^2}(I) \), \( g \) is an even real function in each of its coordinate(s), \( f \) is a real vector function with full rank.

2. There exist integers \( d_3 < \infty \), \( d_1 < d \) and functions \( f, h_{ij}, h_i, i, j = 1, \ldots, d_3 \), such that
\[
r(s,t) = \sum_{i,j=1}^{d_3} h_i(s^{(1)})h_j(t^{(1)})h_{ij}(f(s^{(2)}) - f(t^{(2)}))
\]
where \( f, h_{ij} \in C^{md_3(d_2+1)}(I) \), \( d_2 = d - d_1 \), \( h_i \in C^4 \), \( h_{ij} \) are even functions in each of their coordinates, \( f \) is a real vector function with full rank, \( s^{(1)} = (s_1, \ldots, s_{d_1}) \), and \( s^{(2)} = (s_{d_1+1}, \ldots, s_d) \).

**R.2.** The \( d \times d \) matrix \( R(t) = (\frac{\partial r(s,t)}{\partial x_i \partial t_j} |_{s=t})_{d \times d} \) is nonsingular on \( I \) and for \( i = 1, 2, \ldots, d \), \( \frac{\partial r(s,t)}{\partial x_i} |_{s=t} = 0 \).

**R.3.** The manifolds \( \mathcal{V}^k = \{ v^k(t) : t \in I, v^k(t) = (v_1(t), \ldots, v_k(t)) \} \) derived from an expansion of \( r(s,t) \) have no boundary for all \( k > d \). Here \( r(s,t) = \sum_{l=1}^\infty u_l(s)u_l(t) \), \( v_l(t) = \frac{u_l(t)}{\sigma_l(t)} \) for \( l = 1, 2, \ldots, k \), and \( \sigma_k(t) = \sqrt{\sum_{l=1}^k (u_l(t))^2} \) for \( k = 1, 2, \ldots, \), for \( s, t \in I \).
R.4. For some \( c_0 > 0 \), the critical radius \( r_{kc} \) of the tube \( T(r) \) of the manifolds \( Y^k \) defined in Definition 2.2 satisfies:

\[
r_{kc} \geq c_0.
\]

Presentation

**Theorem 2.2** Suppose \( Z(t) \) is a \( d \)-dimensional non-singular Gaussian random field on a bounded \( d \) dimensional closed space \( I \), with mean zero, unit variance and covariance function \( r(s,t) \) which satisfies, for some \( \alpha \in (0,2] \),

\[
r(s,t) = 1 - \sum_{l=1}^{d} a_l(s,t) |s_i - t_i|^{\alpha} + o(|s - t|^{\alpha})
\]
as \( |s - t| \to 0 \), where \( a_l(s,t) \)'s are bounded and non-negative on \( I \). Then \( Z(t) \) has a uniformly convergent standard Karhunen-Loève expansion in \( t \in I \). (cf.(2.16).)

Further, assume the regularity conditions **R.1** (for \( m = 6 \), **R.2** and **R.3** on \( r(s,t) \) hold. Then as \( z \to \infty \),

\[
P\{\max_{t \in I} Z(t) \geq z\}
\leq \kappa_0 \psi_0(z) + \kappa_2 \psi_2(z) \cdot (1 + o(1)). \tag{2.26}
\]

where \( \kappa_0, \kappa_2 \) are the same constants as those in Weyl's formula for some manifold which depend only on the double mixed derivatives of \( r(s,t) \), and

\[
\psi_0(z) = \frac{1}{2\pi^{(d+1)/2}} \int_{z^2/2}^{\infty} u^{(d+1)/2-1} \exp\{-u\} \, du \tag{2.27}
\]

\[
\psi_2(z) = \frac{1}{4\pi^{(d+1)/2}} \int_{z^2/2}^{\infty} u^{(d+1-a)/2-1} \exp\{-u\} \, du \tag{2.28}
\]

29
Proof: (See Appendix A.)

Note **R.2** implies that the $\alpha$, in the asymptotic expression of $r(s,t)$ in Theorem 2.2, equals to 2. (cf. Remark A.2.)

**Theorem 2.3** Suppose $Z(t)$ satisfies the conditions in Theorem 2.2, and **R.4**. Then as $z \to \infty$,

$$P \left\{ \max_{t \in I} Z(t) \geq z \right\} = \kappa_0 \psi_0(z) + \kappa_2 \psi_2(z) \cdot (1 + o(1)). \quad (2.26')$$

where $\kappa_0, \kappa_2, \psi_0(z)$ and $\psi_2(z)$ are same as those in Theorem 2.2.

Proof: (See Appendix A.)

Remarks

The condition **R.1** is basically to include the interesting random fields and can be examined easily. The homogeneous random fields have representations like 1) in **R.1**. The random field derived from JHF PP index has the representation like 2) in **R.1**.

The condition **R.1** (for $m = 6$) and **R.2** on the covariance $r(s,t)$ are to ensure the following two properties of the manifolds related to the random field $Z(t)$ through an expansion of $r(s,t)$ (may not be Mercer expansion), which are used to prove the two term (upper) approximation formula.

The two properties are:

30
1). For semi critical point $\tilde{r}_{kc}$ defined as Definition 2.2 in Section 2.3, there exists $c_0 > 0$ such that $\tilde{r}_{kc} \geq c_0 > 0$. (cf. Proposition A.1 in Appendix A.)

2). The uniformly convergent Karhunen-Loève expansion of $Z(t)$ satisfies some regularity conditions for the rate of the convergence of

$$\sum_{l=k+1}^{\infty} (u_l(t))^2$$

to zero as $k \to \infty$. (cf. Condition A.1 and A.2 (or Condition A.3) and Proposition A.2 in Appendix A.).

The condition $\textbf{R.3}$ ensures that we can use Weyl's formula. As in the discussion of Theorem 2.1 on page 26, if the manifold related to the expansion of $r(s,t)$ has boundary, there should be a similar approximation formula as (2.26) or (2.26′).

Some numerical examples about these $r_{kc}$'s (and hence $d_k$'s) in Johansen and Johnstone (1985) confirm the condition $\textbf{R.4}$. In most applications, the $\lambda_l$'s decrease rapidly as $l \to \infty$, viz. the first few terms of the Karhunen-Loève expansion (cf. (2.16)) usually dominate the variability of $Z(t)$. Hitherto, people more often just neglect the tail terms of a Karhunen-Loève expansion: for example, in the application to Image Processing (cf. Yaglom (1987a, pp 450 ff)), Yaglom (1987b, note 118). We do not want to judge the suitability of this approach in practice at this point. However, we do see that $\textbf{R.4}$ and properties 2) are reasonable since the $\lambda_l$'s decrease rapidly as $l \to \infty$ in most cases.

\textbf{Remark 2.3}: Unfortunately, the condition $\textbf{R.4}$ is relatively hard to be checked. We believe that there are similar conditions as $\textbf{R.1}$ (for $m = 6$) and $\textbf{R.2}$ which are sufficient for $r_{kc} \geq c_0 > 0$, the condition $\textbf{R.4}$. We leave this problem open to
readers. Without R.4, we have proved a two term upper approximation formula in
Theorem 2.2.

Remark 2.4: A careful reading of the proof will convince the reader that these
regularity conditions, especially condition R.1 (for \( m = 6 \)), can be weakened.

On summery, R.1-R.4 are all reasonable, where R.1-R.3 can be examined easily.

Conjecture

Conditions in Theorem 2.2 are sufficient to (2.26'), i.e., R.1, R.2 and R.3 are
sufficient for the lower bound to hold.

Applications

For the P-value in PP of the unsphered data case (see Equation (1.4)), we have
the following corollary.

Corollary 2.1 Suppose \( Z(t) \) be the random field derived from JHF PP index \( I_j(\alpha) \)
defined in (1.3). Then as \( N \to \infty \)

\[
P\{\max_{t \in I} Z(t) \geq \sqrt{2aN} \} \\
\leq \frac{\kappa_0}{2\pi(p+J-1)/2} \int_{aN}^{\infty} u^{(p+J-1)/2 - 1} \exp\{-u\} \, du \\
+ \frac{\kappa_2}{4\pi(p+J-1)/2} \int_{aN}^{\infty} u^{(p+J-3)/2 - 1} \exp\{-u\} \, du \,(1 + o(1)).
\]  

(2.29)

where \( \kappa_0, \kappa_2 \) are the corresponding volume and total scalar curvature respectively,
of the manifold (see next chapter for their representations), and \( I \) is the \( p + J - 2 \)
dimensional space \([0, \pi] \times \ldots \times [0, \pi] \times [0, 2\pi] \).
**Proof:** Let $I' = \frac{1}{2}S^{p-1} \times S^{J-1}$, $I'' = S^{p-1} \times S^{J-1}$, where $\frac{1}{2}S^{p-1} = \{ \alpha, \alpha \in S^{p-1}, \alpha_p \geq 0 \}$. Then $I = I'$.

Proposition 2.1 says that \textbf{R.1-R.2} in Theorem 2.2 are satisfied.

As for \textbf{R.3}, note that the original random field derived from JHF PP index is $Z(\alpha, \beta) = \beta^T \hat{X}(\alpha)$ on $I''$ (cf. (2.9)-(2.11)). According to the second criterion for checking if the related manifold has no boundary on page 27, the related manifold $\mathcal{V}^k$ has no boundary. As (2.9), by the symmetry (through the origin) of JHF PP index in upper hemisphere and in lower hemisphere,

$$P\left\{ \max_{\alpha \in S^{p-1}, \beta \in S^{J-1}} Z(\alpha, \beta) \geq b \right\} = P\left\{ \max_{\alpha \in \frac{1}{2}S^{p-1}, \beta \in S^{J-1}} Z(\alpha, \beta) \geq b \right\} = P\left\{ \max_{t \in I} Z(t) \geq b \right\}.$$

Hence the related manifold derived from $Z(t)$ should be treated as no boundary in the role of obtaining approximation to $P\{ \max_{t \in I} Z(t) \geq b \}$. (Rigorously speaking, there is some subtle mathematics theory behind "should be treated", which is rather complicated and we do not present it here.) Hence, \textbf{R.3} is satisfied. Therefore this corollary holds.

\[ \square \]

This corollary and the above conjecture give the corresponding P-value in (1.4) for JHF index:

$$P\left\{ \max_{\alpha \in S^{p-1}} I_\beta(\alpha) \geq a \right\} \simeq (2.29).$$

For the P-value in PP of the sphered data case (see Equation 1.6), the situation is quite different, the above approximation formula is not good at all. In this case, the large sample theory we used to derive the limit process does not work. Based on a regression argument, we shall see in Chapter 4 that the first two terms in
JHF PP index are negligible if the data are sphered, viz. there is reduction of dimensionality due to sphering. Intuitively, we can treat JHF PP index with $J$ terms for $j = 1, \ldots, J$ as $J - 2$ term index for $j = 3, \ldots, J$ to derive a similar random field and then obtain a similar approximation formula (cf. Chapter 4 for its intuitive explanation and Chapter 5 for its theoretical justification).

For JHF PP index $I_J(\alpha)$ defined in (1.3), if we sphere the data first before implementing PP procedure, then as $N \to \infty$, the corresponding P-value in (1.6) for JHF index is

$$
P\{ \max_{\alpha \in S^{p-1}} I_J(\alpha) \geq a \}
\approx \frac{\kappa_0^2}{2\pi(p+J-3)/2} \int_{aN}^{\infty} u^{(p+J-5)/2-1} \exp\{-u\} \, du
+ \frac{\kappa_2^2}{4\pi(p+J-3)/2} \int_{aN}^{\infty} u^{(p+J-5)/2-1} \exp\{-u\} \, du.
$$

(2.30)

where $\kappa_0^2$ and $\kappa_2^2$ are the corresponding volume and total scalar curvature of the manifold, which are represented in next chapter. The proof is similar as that of Corollary 2.1.

**Remark 2.5:** We have seen some gap between theory and application, because in practice people do whatever is 'economic' or 'convenient' or please them rather than follow the 'strict rule'. For example, people usually sphere data first before implementing PP procedures, while the derivation of PP indices is based on the assumption that there is no sphering involved. Therefore, we have a task of closing the gap, which often introduce some more interesting problems (one solution here is in Equation (2.30), see chapters 4 and 5 for more details).
Chapter 3

Geometric Meanings and Applicable Formulas

In the last chapter, we applied Weyl’s formula to derive a two term approximation formula for the tail probability of the extreme of a class of differentiable Gaussian random fields. This class of random fields contains all homogeneous ones. However, in applications the two coefficients $\kappa_0$, $\kappa_2$ in this formula are difficult to evaluate from the original abstract expressions in Weyl’s paper. In this chapter, we shall explain the geometric meanings of $\kappa_0$, $\kappa_2$ introduced in last chapter in an easily understandable way, and express them in applicable formulas. The numerical values for related constants are calculated and reported in Table 3.1-3.5.

The main steps for deriving the so called applicable formulas are as follows:

1. Find a relation between a metric tensor and a covariance function (Proposition 3.1 and Equation (3.21) in Section 3.1).
2. Write the coefficients $\kappa_0$ and $\kappa_2$ in terms of the covariance function based on the relation in step 1 (Theorem 3.1, 3.2, Equation (3.31) and (3.36) in Section 3.2 and 3.3).

3. Substitute the representations for $\kappa_0$ and $\kappa_2$ from step 2 into (2.29) and (2.30) to obtain final simple applicable formulas (3.37) and (3.41) in Section 3.4.

Those people who only wish to know how to use the two term approximation formula for their problems should go directly to Section 3.4.

3.1 Metric Tensor and Covariance Function

Suppose $Z(t), t \in I$, is a differentiable Gaussian random field with the uniformly convergent Karhunen-Loève expansion

$$Z(t) = \sum_{i=1}^{k} u_i(t) X_i = < u(t), X >$$

(cf. (2.16)). Let $\mathcal{U} = \{u(t) : u(t) = (u_1(t), u_2(t), \ldots, u_k(t))\}$ be the corresponding $d$-dimensional manifold embedded in the unit sphere $S^{k-1}$. Here $d$ is the dimension of the parameter space $I$, $k$ is either finite or infinite. As shown in Section 2.3, there is a one-to-one correspondence between $Z(\cdot)$ and $\mathcal{U}$:

$$Z(t) \longleftrightarrow u(t), \ t \in I. \quad (3.1)$$

By applying this correspondence (3.1) and Weyl's formula, we proved a two term approximation formula (2.26) for the tail probability in (2.15). The two coefficients
$\kappa_0$ and $\kappa_2$ in this approximation formula are certain integral invariants of the manifold $\mathcal{U} = \{u(t) : u(t) = (u_1(t), u_2(t), \ldots, u_k(t))\}$ in Weyl’s formula. Through this correspondence (3.1), we shall be able to express $\kappa_0$ and $\kappa_2$ in terms of functions of the covariance function of $Z(t)$ for convenience of statistical applications.

The properties of the $d$-dimensional differentiable manifold $\mathcal{U} = \{u(t) : u(t) = (u_1(t), u_2(t), \ldots, u_k(t))\}$ are determined by its metric tensor, which has the components defined by the following inner product

$$g_{ij}(t) = \langle \frac{\partial u(t)}{\partial t_i}, \frac{\partial u(t)}{\partial t_j} \rangle = \sum_{i=1}^{k} \frac{\partial u_l(t)}{\partial t_i} \cdot \frac{\partial u_l(t)}{\partial t_j},$$

(3.2)

for $i, j = 1, 2, \ldots, d$, $t = (t_1, t_2, \ldots, t_d) \in I$. These functions $g_{ij}(\cdot)$ form a $d \times d$ symmetric matrix $R(\cdot)$. $R(t)$ is called metric tensor matrix for the convenience in this paper.

**Definition 3.1 (Metric Tensor Matrix)** The metric tensor matrix of a $d$ dimensional differentiable manifold $\mathcal{U} = \{u(t) : u(t) = (u_1(t), u_2(t), \ldots, u_k(t))\}$ is

$$R(t) = (g_{ij}(t))_{d \times d}, \ t \in I,$$

(3.3)

where $g_{ij}(t)$ is the $(i, j)$ coefficient of the metric tensor defined in (3.2).

More specifically, in principle everything we need to know about the manifold $\mathcal{U}$, like volume and scalar curvature, can be represented as some functions of $R(t), t \in I$.

Recall that we do not know the Karhunen-Loève expansion of the random field related to JHF PP index explicitly, not even whether the expansion is finite or infinite. The following Lemma 3.1 builds a connection between the metric tensor matrix of a manifold and the covariance function of a random field, provided that
there is a one to one correspondence as in (3.1). This connection enables one to calculate the metric tensor matrix without knowing the Karhunen-Loève expansion explicitly.

**Lemma 3.1** Let \( r(s, t) \) be the covariance function of a non-degenerate differentiable random field \( Z(t) \) on a compact set \( I \), which has an uniformly convergent expansion

\[
 r(s, t) = \sum_{i=1}^{k} u_{i}(s)u_{i}(t). 
\]  

(3.4)

where \( k \) is finite or infinite. Further, if

\[
 \sum_{i=1}^{k} \frac{\partial u_{i}(s)}{\partial s_i} \frac{\partial u_{i}(t)}{\partial t_j} 
\]

converges uniformly in \( I \), then

\[
 \frac{\partial^{2}r(s, t)}{\partial s_i \partial t_j} = \sum_{i=1}^{k} \frac{\partial u_{i}(s)}{\partial s_i} \frac{\partial u_{i}(t)}{\partial t_j} 
\]

(3.6)

and in particular,

\[
 g_{ij}(t) = \left. \frac{\partial^{2}r(s, t)}{\partial s_i \partial t_j} \right|_{s=t}. 
\]

(3.7)

**Proof:** By (3.4) and (3.5), we can exchange the summation and differentiation as follows (cf. Theorem (5.5) on p26 in Walker, 1988),

\[
 \frac{\partial^{2}r(s, t)}{\partial s_i \partial t_j} = \frac{\partial^{2}}{\partial s_i \partial t_j} \sum_{i=1}^{k} u_{i}(s)u_{i}(t) \quad \text{(by (3.4))}
\]

\[
 = \sum_{i=1}^{k} \frac{\partial^{2}}{\partial s_i \partial t_j} u_{i}(s)u_{i}(t)
\]

\[
 = \left. < \frac{\partial u}{\partial s_i}, \frac{\partial u}{\partial s_j} > \right|
\]

\[
 = g_{ij}(t) \quad \text{by the definition.}
\]

38
Therefore, (3.6) and hence (3.7) are valid.

Equation (3.7) says that these metric coefficients only depend on the mixed partial derivatives of the covariance function evaluated on the diagonal line $s = t$.

**Remark 3.1:** In Remark 2.2, we described the existence of an expansion as in (3.4) in a quite general class. In Appendix B, we give a sufficient condition for (3.4) and (3.5) for a wide class of the random fields (see Lemma B.1 and Proposition B.1), and prove that the Karhunen-Loève expansion corresponding to JHF PP index satisfy (3.4) and (3.5) (the first part of the proof for Proposition 3.1 in Appendix B).

Therefore, the expression of a metric tensor matrix in terms of the double mixed partial derivatives is valid in a quite general class.

In the following two subsections, we give explicit expressions for the corresponding tensor matrices, for the case where the random fields are related to JHF PP index with unsphered and sphered data.

### 3.1.1 Un sphered Case

**Proposition 3.1** Suppose $Z(t)$ is the random field derived from the JHF PP index with unsphered data. Then the metric tensor matrix defined in Equation (3.3) is

$$ R(t) = \begin{pmatrix} R_{11}(t) & R_{12}(t) \\ R_{21}(t) & R_{22}(t) \end{pmatrix} \quad (d \times d), \quad (3.8) $$

where

$$ R_{11}(t) = \left( \frac{\partial^2 r(s,t)}{\partial \theta_i \partial \theta'_j} \right)_{s=t}^{(p-1) \times (p-1)} \quad (3.9) $$
\[ R_{22}(t) = \left( \frac{\partial^2 r(s, t)}{\partial \varphi_i \partial \varphi_j'} \right) \bigg|_{s=t}^{(J-1) \times (J-1)} \]  
\[ R_{12}(t) = \left( \frac{\partial^2 r(s, t)}{\partial \theta_i \partial \varphi_j'} \right) \bigg|_{s=t}^{(p-1) \times (J-1)} = R_{21}(t)^T \]  
with \( d = p + J - 2 \), \( s = (\theta_1, \ldots, \theta_{p-1}, \varphi_1, \ldots, \varphi_{J-1}) \), \( t = (\theta_1', \ldots, \theta_{p-1}', \varphi_1', \ldots, \varphi_{J-1}') \) 
\( \in I = (0, \pi] \times (0, \pi] \times \cdots \times (0, \pi] \times (0, 2\pi] \). Further,

\[
R_{11}(t) = C(\beta) \begin{pmatrix}
\sum_{l=1}^p (\frac{\partial \beta_l}{\partial \theta_1})^2 & 0 & \ldots & 0 \\
0 & \sum_{l=1}^p (\frac{\partial \beta_l}{\partial \theta_2})^2 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & \sum_{l=1}^p (\frac{\partial \beta_l}{\partial \theta_{p-1}})^2
\end{pmatrix},
\]

\[
R_{22}(t) = \begin{pmatrix}
\sum_{l=1}^J (\frac{\partial \beta_l}{\partial \varphi_1})^2 & 0 & \ldots & 0 \\
0 & \sum_{l=1}^J (\frac{\partial \beta_l}{\partial \varphi_2})^2 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & \sum_{l=1}^J (\frac{\partial \beta_l}{\partial \varphi_{J-1}})^2
\end{pmatrix},
\]

\[
R_{12}(t) = 0_{(p-1) \times (J-1)} = R_{21}(t)^T.
\]

Here, \( 0_{(p-1) \times (J-1)} \) is a \((p-1) \times (J-1)\) zero matrix; \( C(\beta) \) is a quadratic function in \( \beta' \)s,

\[
C(\beta) = 4 \cdot \mathcal{E} \left\{ \left[ \sum_{j=1}^J \beta_j \sqrt{2j + 1} P_j'(2\Phi(X) - 1) \phi(X) \right]^2 \right\}
\]

\[
= 4 \sum_{i,j=1}^J c_{ij} \beta_i \beta_j,
\]

with

\[
c_{ij} = (2i + 1)(2j + 1) \cdot \mathcal{E} \left\{ (\phi(X))^2 P_i'(2\Phi(X) - 1) P_j'(2\Phi(X) - 1) \right\}
\]

for \( i, j = 1, \ldots J \); \( P_j' \) is the first derivative of the \( j \)-th Legendre polynomial, and \( X \) is a \( \mathcal{N}(0, 1) \) random variable.
Proof: (Idea of the proof. For details, see Appendix B.)

Since the condition (3.5) in Lemma 3.1 is satisfied for the random field derived from JHF index, Equation (3.9)-(3.11) hold.

For Equation (3.12)-(3.14), we apply Lemma 3.3 given below, and the following facts:

\[
\sum_{i=1}^{p} \frac{\partial \alpha_i}{\partial \theta_i} \alpha_i = \sum_{i=1}^{J} \frac{\partial \beta_i}{\partial \varphi_j} \beta_i = 0
\]  

(3.17)

for \( i = 1, 2, \ldots, p - 1; j = 1, 2, \ldots, J - 1 \).

\[
\sum_{i=1}^{p} \frac{\partial \alpha_i}{\partial \theta_i} \frac{\partial \alpha_i}{\partial \theta_{i'}} = \delta_{ii'};
\]

(3.18)

\[
\sum_{i=1}^{J} \frac{\partial \beta_i}{\partial \varphi_i} \frac{\partial \beta_i}{\partial \varphi_{j'}} = \delta_{jj'};
\]

(3.19)

for \( i, i' = 1, 2, \ldots, p - 1 \), \( j, j' = 1, 2, \ldots, J - 1 \). Here \( \delta_{kl} \) is the Kronecker \( \delta \).

\( \square \)

Lemma 3.2 Suppose the random vector \( Z = (Z_1, Z_2, \ldots, Z_p) \) is from the population \( \mathcal{N}(0, I_p) \), \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p) \) and \( \beta = (\beta_1, \beta_2, \ldots, \beta_p) \) are on \( S^{p-1} \). Then for \( i, j = 1, 2, \ldots, p \), we have

\[
\mathcal{E}(Z_i | \alpha^T Z = x) = \alpha_i x;
\]

\[
\mathcal{V}ar(Z_i | \alpha^T Z = x) = 1 - \alpha_i^2;
\]

\[
\mathcal{C}ov(Z_i, Z_j | \alpha^T Z = x) = -\alpha_i \alpha_j.
\]

and

\[
\mathcal{E}(Z_i | \alpha^T Z = x, \beta^T Z = y) = \alpha_i x + \beta_i y;
\]

\[
\mathcal{V}ar(Z_i | \alpha^T Z = x, \beta^T Z = y) = 1 - (\alpha_i^2 + \beta_i^2);
\]

\[
\mathcal{C}ov(Z_i, Z_j | \alpha^T Z = x, \beta^T Z = y) = -(\alpha_i \alpha_j + \beta_i \beta_j).
\]
Proof: It is easy to see that for \( i \neq j \),
\[
\begin{pmatrix}
Z_i \\
Z_j \\
\alpha^r Z \\
\beta^r Z
\end{pmatrix}
\overset{\text{dist}}{\sim}
\mathcal{N}
\begin{pmatrix}
0 \\
0 \\
\alpha_i \\
\beta_i
\end{pmatrix},
\begin{pmatrix}
1 & 0 & \alpha_i & \beta_i \\
0 & 1 & \alpha_j & \beta_j \\
\alpha_i & \alpha_j & 1 & 0 \\
\beta_i & \beta_j & 0 & 1
\end{pmatrix}
\tag{3.20}
\]
where \( \overset{\text{dist}}{\sim} \) means " is distributed from ". Then using Theorem 2.5.1 in Anderson (1984), we can present conditional means, variances, and covariance as above.

\[\square\]

3.1.2 Sphered Case

In the case of the sphered data (see Equation (1.5) and the successive paragraph), as indicated briefly at the end of Chapter 2, we treat the corresponding limiting process \( Z^*(t) \) from JHF PP index as from the index \( I^*_j(\alpha) \) summed up from \( j = 3, \ldots, J \):
\[
I^*_j(\alpha) = \frac{1}{2N} \sum_{j=3}^{J} (\tilde{Y}_j(\alpha))^2,
\]
with \( \tilde{Y}_3(\alpha), \ldots, \tilde{Y}_j(\alpha) \) as from unsphered version, which are defined in (2.2) Its intuitive explanation and theoretical justification are presented in Chapter 4 and 5 respectively.

In this sphered case, the parameter space has dimension \( d = p + J - 3 \), and the metric tensor matrix \( R^*(t) \) for \( Z^*(t) \) is similar to \( R(t) \) of unsphered data case,
\[
R^*(t) = \begin{pmatrix}
R^*_{11}(t) & R^*_{12}(t) \\
R^*_{21}(t) & R^*_{22}(t)
\end{pmatrix}_{d \times d}
\tag{3.21}
\]
Here
\[
R_{11}^s(t) = C^s(\beta) \cdot \begin{pmatrix}
\sum_{l=1}^p (\frac{\partial \alpha_l}{\partial \theta_1})^2 & 0 & \cdots & 0 \\
0 & \sum_{l=1}^p (\frac{\partial \alpha_l}{\partial \theta_2})^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sum_{l=1}^p (\frac{\partial \alpha_l}{\partial \theta_{p-1}})^2
\end{pmatrix}_{(p-1) \times (p-1)}
\] (3.22)

\[
R_{22}^s(t) = \begin{pmatrix}
\sum_{l=1}^{J-2} (\frac{\partial \beta_l}{\partial \varphi_1})^2 & 0 & \cdots & 0 \\
0 & \sum_{l=1}^{J-2} (\frac{\partial \beta_l}{\partial \varphi_2})^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sum_{l=1}^{J-2} (\frac{\partial \beta_l}{\partial \varphi_{J-3}})^2
\end{pmatrix}_{(J-3) \times (J-3)}
\] (3.23)

\[
R_{12}^s(t) = 0_{(p-1) \times (J-3)} = R_{21}^s(t)^T,
\] (3.24)

with
\[
C^s(\beta) = 4 \sum_{j=1}^{J-2} \tilde{c}_{ij} \beta_i \beta_j,
\] (3.25)

where
\[
\tilde{c}_{ij} = c_{i+2j+2},
\] (3.26)

c_{ij}'s are defined in Equation (3.16).

### 3.2 The First Coefficient \( \kappa_0 \) in the Approximation Formula

The first coefficient \( \kappa_0 \) in the approximation formula (2.23) is the *volume* (or area) of the manifold (Weyl, 1939). The volume of a manifold is the integral of the square
root of the determinant of the metric tensor matrix associated with this manifold, and is relatively easy to calculate compared to other terms in differential geometry. However, if the metric tensor matrix is not diagonal and has high order, i.e. \( d \) is large, it is often tough to get a clean formula. In the case of JHF PP index, our metric tensor matrix is diagonal, and therefore we can present a simple formula for \( \kappa_0 \) and give an example for \( \kappa_0 \) related to a real world problem: calculate a P-value in PP.

### 3.2.1 Unsphered Case

**Theorem 3.1** \( \kappa_0 \) in Weyl’s formula is the volume of a manifold. If the manifold is the one derived from JHF PP index with unsphered data, we have

\[
\kappa_0 = \int_I \| R(t) \|^{1/2} \ d\theta_1 \ldots d\theta_{p-1} d\varphi_1 \ldots d\varphi_J \quad (3.27)
\]

\[
= \frac{1}{2} \omega_{p-1} \omega_{J-1} \cdot \mathcal{E} \left[ C(\beta)^{(p-1)/2} \right]. \quad (3.28)
\]

Here the expectation \( \mathcal{E} \) is with respect to the probability under which \( \beta_1, \beta_2, \ldots, \beta_J \) are uniformly distributed on \( S^{J-1} = \{ \beta : \beta = (\beta_1, \ldots, \beta_J), \beta \beta^* = 1 \} \), and \( \omega_{r-1} \) is the surface area of \( S^{r-1} \), i.e.

\[
\omega_{r-1} = \frac{2 \pi^{r/2}}{\Gamma(r/2)}, \quad (3.29)
\]

where \( \Gamma \) is the Gamma function,

\[
\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} \, dx, \quad p > 0.
\]

**Proof:** Weyl had described \( \kappa_0 \) as the volume of the tube in his paper in 1939. (3.27) is by the intrinsic expression of the volume of a manifold (cf. Definition B.1 in Appendix B or Millman and Parker, p130, 1977). (3.28) then follows since the
determinant of $R_{22}(t)$ and $R_{11}(t)/C(\beta)$ are the Jacobian matrices of the Cartesian-polar coordinate transformation from $\beta$ to $\varphi$, $\alpha$ to $\theta$ respectively.

\[ \Box \]

**Remark 3.2:** To obtain the numerical value of a multiple integral can be a headache. For example, the direct calculation of a multiple integral might take an unrealistically long time for even the most powerful computers, even if we know the integral exists. The above theorem expresses the multiple integral as the expectation of certain power of the simple quadratic function in (3.28), so we can use the Monte Carlo method to simulate $\kappa_0$ and get a numerical estimate of its value very fast (within one minute).

**Remark 3.3:** Readers may wonder why we do not simulate the P-value in (1.4) (or (1.6)) directly, but perform a simulation to estimate some constant (e.g. $\kappa_0$) in our approximation formula for the P-value. The answer is as follows. To simulate the P-value with simulation size $M$, we have to maximize $M$ multivariate functions, *viz.* to obtain

$$\max_{\alpha \in \mathbb{S}^{p-1}} I_{f}(\alpha \mid \text{data set } i), \ i = 1, \ldots, M. \quad (3.30)$$

Each function $I_{f}(\alpha \mid \text{data set } i)$ has $N$ parameters from 'data set $i$': $Z_{1}^{i}, Z_{2}^{i}, \ldots, Z_{N}^{i}$ and is very bumpy under the null hypothesis: data are from normal population (for details, see the hypotheses associated with Equation (1.4) and (1.6)), $i = 1, \ldots, M$, where $N$ is sample size. They have a lot of local maxima. Therefore, there are two major difficulties in simulating the P-value. First, since there is no perfect fast optimizer to maximize a multivariate function, it is rather complicated for one to design a good algorithm for our problem. In fact, *optimization* is still a big subject in Operational Research. Second, in this case, the problem is very computationally
intensive as our experiments show. For example, we used a ‘good enough’ algorithm (see chapter 4 for the algorithm) in the following situation: the data dimension \( p = 4 \), the JHF PP index used \( J = 4 \) terms, the sample size \( N \) is 100, the simulation size \( M \) is 1000. This took about a week to run the program of simulating the P-value on Sun3/50 under the condition that no one else is using the machine. Needless to say, it will be very expensive when data dimension \( p \) gets higher - 6, 8, ... The simulation for \( \kappa_0 \) is like an evaluation of a simple quadratic function, its numerical result can be obtained very quickly and reliably. On the another hand, the function \( I_f(\alpha \mid \text{data set } i) \) which is to be maximized in simulating P-values, depends on the data, the quadratic function for \( \kappa_0 \) does not. Those useful constants (\( \kappa_0 \) in this section and \( \kappa_2 \) in next section etc.) in our approximation formula need to be calculated only once.

### 3.2.2 Sphered Case

If the data are sphered, our treatment in section 3.1.2 induces a random field \( Z^s(i) \) and a metric tensor matrix \( R^s(t) \) as in Equation (3.21). Therefore, in this case, the corresponding \( \kappa_0 \), which we denote as \( \kappa_0^s \), is the volume of the manifold associated with \( Z^s(t) \), and has the following representation

\[
\kappa_0^s = \frac{1}{2} \omega_{p-1} \omega_j \mathcal{E} \left[ C^s(\beta)^{p-1/2} \right].
\]  

(3.31)

where \( C^s(\beta) \) is defined as in Equation (3.25) and \( \omega_{r-1} \)'s are defined as in (3.29).
3.3 The Second Coefficient $\kappa_2$ in the Approximation Formula

In Weyl's paper, he did not present a geometric meaning for the second coefficient $\kappa_2$, but only vaguely described it as "certain integral invariant". In this section, we give the explicit geometric meaning for $\kappa_2$ and its applicable formula.

3.3.1 UnspHERed Case

Theorem 3.2 ($\kappa_2$) $\kappa_2$ in Weyl's formula is the total scalar curvature of the manifold as a submanifold embedded in the unit sphere $S^{d-1}$. If the manifold is one derived from JHF PP index with data as unspHERed version, we have

$$\kappa_2 = \int \left( \frac{S}{2} - \frac{d(d-1)}{2} \right) \| R(t) \|^{1/2} \, d\theta_1 \ldots d\theta_{p-1} d\varphi_1 \ldots d\varphi_{J-1}$$

$$= \frac{1}{2} \omega_{p-1} \omega_{J-1} \cdot \mathcal{E} \left[ D \left( p, J, C(\beta) \right) \right]$$

where for $p \geq 5$,

$$D \left( p, J, C(\beta) \right)$$

$$= \left( \frac{(J-1)(J-2)}{2} \right) \cdot \left( \frac{d(d-1)}{2} \right) \cdot \left[ C(\beta)^{(p-1)/2} \right]$$

$$+ \left( \frac{(p-1)(p-2)}{2} \right) \left[ C(\beta)^{(p-3)/2} \right]$$

$$+ \frac{(p-1)(p-2)}{8} \left[ C(\beta)^{(p-5)/2} \cdot \sum_{l=1}^{J-1} \left( \frac{\partial C(\beta)}{\partial \varphi_l} \right)^2 \sum_{m=1}^{J} \left( \frac{\partial \beta_m}{\partial \varphi_l} \right)^2 \right]$$

and for $p < 5$,

$$D \left( p, J, C(\beta) \right)$$
\[
\begin{align*}
&= \left( \frac{(J-1)(J-2)}{2} - \frac{d(d-1)}{2} \right) \cdot [C(\beta)^{(p-1)/2}] \\
&+ \frac{(p-1)(p-2)}{2} \left[ C(\beta)^{(p-3)/2} \right] \\
&- \frac{(p-1)(p/4 - 1)}{2} \left[ C(\beta)^{(p-5)/2} \cdot \sum_{l=1}^{J-1} \left( \frac{\partial C(\beta)}{\partial \varphi_l} \right)^2 \left( \sum_{m=1}^{J} \left( \frac{\partial \beta}{\partial \varphi_m} \right)^2 \right)^{-1} \right] \\
&- \frac{p-1}{2} \left[ C(\beta)^{(p-3)/2} \cdot \left( \sum_{l=1}^{J-1} \frac{\partial^2 C(\beta)}{\partial \varphi_l^2} \right) \left( \sum_{m=1}^{J} \left( \frac{\partial \beta}{\partial \varphi_m} \right)^2 \right)^{-1} \right. \\
&\left. + \sum_{l=1}^{J-2} (J-1-l) (\cot \varphi_l) \frac{\partial C(\beta)}{\partial \varphi_l} \left( \sum_{m=1}^{J} \left( \frac{\partial \beta}{\partial \varphi_m} \right)^2 \right)^{-1} \right]. \\
\end{align*}
\] (3.35)

Here $S$ is the intrinsic scalar curvature of the manifold (cf. Definition B.5, Appendix B); $R$ is the metric tensor matrix; $\omega_{p-1}$ and $\omega_{J-1}$ are as defined in (3.29).

**Proof:** See Appendix B.

\[\square\]

### 3.3.2 Sphered Case

For sphered data, similar to the first constant in the approximation formula, the second constant $\kappa_2^*$ is the scalar curvature of the manifold related to the random field $Z^*(t)$ (see section 3.2), with the following expression

\[
\kappa_2^* = \frac{1}{2} \omega_{p-1} \omega_{J-3} \cdot \mathcal{E} [D(p, J-2, C^*(\beta))],
\] (3.36)

where $C^*(\beta)$ is as in Equation (3.25), $\omega_{p-1}$ and $\omega_{J-3}$ are similarly defined as in Equation (3.29), $D(\cdot, \cdot, \cdot)$ is defined in Equation (3.34) if data dimension $p \geq 5$ or (3.35) if $p < 5$. 

48
3.4 Tables and Their Construction

Substituting the above expressions for $\kappa_0$ and $\kappa_2$ into the two term approximation formulas (2.29) and (2.30), we have the following directly applicable formula for P-values of JHF PP index.

If the data are unsphered,

$$P\{ \max_{\alpha \in \mathcal{S}^{p-1}} I_J(\alpha) \geq a \}$$

$$\simeq P\{ \max_{t \in \mathcal{I}} Z(t) \geq \sqrt{2aN} \}$$

$$= A_{p,J} \left[ B_{p,J} \int_{aN}^{\infty} u^{(p+J-1)/2-1} \exp\{-u\} du \right.$$  

$$+ C_{p,J} \int_{aN}^{\infty} u^{(p+J-3)/2-1} \exp\{-u\} du (1 + o(1)) \left. \right] \right. \) (3.37)

as $N \to \infty$, where

$$A_{p,J} = \frac{2^{p-1}}{4\pi^{(p+J-1)/2}} \omega_{p-1} \omega_{J-1} = \frac{2^{p-1}\sqrt{\pi}}{\Gamma(p/2)\Gamma(J/2)}$$ (3.38)

$$B_{p,J} = \frac{1}{2^{p-1}} \cdot \mathcal{E} \left[ C(\beta)^{(p-1)/2} \right],$$ \hspace{1cm} (3.39)

$$C_{p,J} = \frac{1}{2^p} \cdot \mathcal{E} \left[ D(p, J, C(\beta)) \right].$$ \hspace{1cm} (3.40)

with $D(\cdot, \cdot, \cdot)$ defined in Equation (3.34) if data dimension $p \geq 5$ or (3.35) if $p < 5$.

In the case of the sphered data,

$$P\{ \max_{\alpha \in \mathcal{S}^{p-1}} I_J(\alpha) \geq a \}$$

$$\simeq P\{ \max_{t \in \mathcal{I}} Z^s(t) \geq \sqrt{2aN} \}$$

$$= A_{p,J}^s \left[ B_{p,J}^s \int_{aN}^{\infty} u^{(p+J-3)/2-1} \exp\{-u\} du \right.$$  

$$+ C_{p,J}^s \int_{aN}^{\infty} u^{(p+J-5)/2-1} \exp\{-u\} du (1 + o(1)) \left. \right] \right. \) (3.41)

49
as \( N \to \infty \), where

\[
A_{pJ}^s = \frac{2^{p-1}}{4\pi^{(p+J-3)/2}} \omega_{p-1} \omega_{J-3} = \frac{2^{p-1}\sqrt{\pi}}{\Gamma(p/2)\Gamma((J-2)/2)},
\]

(3.42)

\[
B_{pJ}^s = \frac{1}{2^{p-1}} \cdot \mathcal{E} \left[ C^s(\beta)^{(p-1)/2} \right],
\]

(3.43)

\[
C_{pJ}^s = \frac{1}{2^p} \cdot \mathcal{E} \left[ D(p, J-2, C^s(\beta)) \right].
\]

(3.44)

with \( D(\cdot, \cdot, \cdot) \) defined in (3.34) or (3.35) depending on \( p \), \( \omega^s \)'s defined in Equation (3.29), and \( C^s(\beta) \) as before.

How do we choose \( J \) for different \( p \) and \( N \)? It depends on each individual. However, the following guidelines suggested by J. Friedman may be useful:

- \( 3 \leq J \leq 6 \) in most cases;
- the smaller \( N \) is, the smaller \( J \) should be taken.

Less experienced PP users may find the examples in Friedman (1987) useful.

We can evaluate \( A_{pJ} \) by a simple arithmetic calculation, so the table (Table 3.1) is easy to build.

To obtain numerical values for \( B_{pJ} \) (see Table 3.2), we need to obtain the numerical value for the coefficients \( c_{ij} \) in \( C(\beta) \) first, then use numerical integration to calculate \( \mathcal{E} \left[ C(\beta)^{(p-1)/2} \right] \) if \( J = 2 \), and apply

\[
\frac{1}{M} \sum_{l=1}^{M} \left[ C(\beta^{(l)})^{(p-1)/2} \right]
\]

(3.45)
as an estimate for \( \mathcal{E} \left[ C(\beta)^{(p-1)/2} \right] \) if \( J \geq 3 \). Here \( M \) is the simulation size, \( \beta^{(l)} \) is the \( l \)-th pseudo uniform random vector on the unit sphere \( S^{J-1} \) constructed as follows,

\[
\beta^{(l)} = \left( \frac{x_{1l}}{\|x_i\|}, \frac{x_{2l}}{\|x_i\|}, \ldots, \frac{x_{Jl}}{\|x_i\|} \right),
\]

(3.46)
where $x_{11}, x_{12}, \ldots, x_{1J}, x_{21}, \ldots, x_{MJ}$ are independent, identically distributed $\mathcal{N}(0, 1)$ observed (pseudo) random variables, $\|x_i\| = \sqrt{x_{i1}^2 + x_{i2}^2 + \ldots + x_{iJ}^2}$ for $l = 1, 2, \ldots, M$. Notice that $c_{ij}$'s are certain linear combinations of $e_i, i = 0, 2, 4, \ldots, 2J - 2$, with

$$e_i = \mathbb{E} \left\{ (\phi(X))^2 (2\Phi(X) - 1)^i \right\}, \quad X \overset{d}{\sim} \mathcal{N}(0, 1).$$

If $i$ is small we have an exact numerical expression for $e_i$; otherwise we have numerical values with accuracy to 5th digit.

In Table 3.2, when $J = 2$, we use cautious adaptive Romberg extrapolation with desired absolute error $10^{-5}$ and relative error $10^{-10}$ to get the numerical integration result, which is specified by "num" in the table. When $J = 4, 6$, we use the Monte Carlo method indicated by (3.45) with simulation size 20,000 to estimate $B_{pJ}$, which we call "est" in the table with estimated standard error "se".

For $C_{pJ}$, the numerical calculation involved is a bit more complicated than the calculation for $B_{pJ}$, but is similar. See Table 3.3 for corresponding numerical results, where * indicates combinations of $p$ and $J$ which may be inferior to other combinations for fixed $p$. All other notations have same meaning as ones in Table 3.2, only the simulation size is 50,000 rather than 20,000.

The computation and accuracy involved for constructing tables for $A_{pJ}^*, B_{pJ}^*, C_{pJ}^*$ are similar to ones for $A_{pJ}, B_{pJ}, C_{pJ}$. See Table 3.1, 3.4, 3.5.

In summary, the recommended procedure for calculating the P-value in (1.4) (or (1.6) if we use spherred data) in application, is as follows:

2. Do a very simple calculation for an incomplete Gamma function: substitute $a$ in the formula (3.37) (or (3.41) if we use sphered data) by the observed $\max_{\alpha \in S^p-1} I_J(\alpha)$, then try to obtain the numerical value for the simple incomplete $\Gamma$ function evaluated at $(N \cdot \max_{\alpha \in S^p-1} I_J(\alpha))$ (see (3.37) and (3.41)), either by using the standard statistical table or call a standard subroutine on the computer. Here $N$ is the sample size.

3. Use the simple algebra to combine values from step 1 and 2 for the final P-value.

Our approximation formula and tables make the P-values easy to compute.
<table>
<thead>
<tr>
<th>$A_{pJ}$</th>
<th>J=2</th>
<th>J=4</th>
<th>J=6</th>
<th>J=8</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=2</td>
<td>3.5449</td>
<td>3.5449</td>
<td>1.7725</td>
<td>0.59082</td>
</tr>
<tr>
<td>p=3</td>
<td>8.0000</td>
<td>8.0000</td>
<td>4.0000</td>
<td>1.3333</td>
</tr>
<tr>
<td>p=4</td>
<td>14.180</td>
<td>14.180</td>
<td>7.0900</td>
<td>2.3633</td>
</tr>
<tr>
<td>p=7</td>
<td>34.133</td>
<td>34.133</td>
<td>17.067</td>
<td>5.6888</td>
</tr>
<tr>
<td>p=8</td>
<td>37.812</td>
<td>37.812</td>
<td>18.906</td>
<td>6.3020</td>
</tr>
<tr>
<td>p=9</td>
<td>39.010</td>
<td>39.010</td>
<td>19.505</td>
<td>6.5017</td>
</tr>
<tr>
<td>p=10</td>
<td>37.812</td>
<td>37.812</td>
<td>18.906</td>
<td>6.3020</td>
</tr>
<tr>
<td>p=11</td>
<td>34.675</td>
<td>34.675</td>
<td>17.337</td>
<td>5.7792</td>
</tr>
<tr>
<td>p=12</td>
<td>30.250</td>
<td>30.250</td>
<td>15.125</td>
<td>5.0417</td>
</tr>
<tr>
<td>p=13</td>
<td>25.218</td>
<td>25.218</td>
<td>12.609</td>
<td>4.2030</td>
</tr>
<tr>
<td>p=15</td>
<td>15.519</td>
<td>15.519</td>
<td>7.7595</td>
<td>2.5865</td>
</tr>
<tr>
<td>p=16</td>
<td>11.524</td>
<td>11.524</td>
<td>5.7620</td>
<td>1.9207</td>
</tr>
<tr>
<td>p=17</td>
<td>8.2768</td>
<td>8.2768</td>
<td>4.1384</td>
<td>1.3795</td>
</tr>
<tr>
<td>p=18</td>
<td>5.7619</td>
<td>5.7619</td>
<td>2.8810</td>
<td>0.96032</td>
</tr>
<tr>
<td>p=19</td>
<td>3.8949</td>
<td>3.8949</td>
<td>1.9475</td>
<td>0.64915</td>
</tr>
<tr>
<td>p=20</td>
<td>2.5608</td>
<td>2.5608</td>
<td>1.2804</td>
<td>0.42680</td>
</tr>
</tbody>
</table>

Note: for $A_{pJ}^{s}$, use $A_{pJ}^{s} = A_{p(j-2)}$

Table 3.1: $A_{pJ} (A_{pJ}^{s})$
<table>
<thead>
<tr>
<th>$B_{pJ}$</th>
<th>$J=2$</th>
<th></th>
<th>$J=4$</th>
<th></th>
<th>$J=6$</th>
<th></th>
<th>$J=8$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>num</td>
<td>est</td>
<td>se</td>
<td>est</td>
<td>se</td>
<td>est</td>
<td>se</td>
<td>est</td>
</tr>
<tr>
<td>p=2</td>
<td>0.67818</td>
<td>0.99944</td>
<td>0.0070671</td>
<td>1.3359</td>
<td>0.0094461</td>
<td>1.6799</td>
<td>0.011879</td>
<td></td>
</tr>
<tr>
<td>p=3</td>
<td>0.47038</td>
<td>1.0412</td>
<td>0.0073627</td>
<td>1.8568</td>
<td>0.013130</td>
<td>2.9467</td>
<td>0.020836</td>
<td></td>
</tr>
<tr>
<td>p=4</td>
<td>0.33306</td>
<td>1.1204</td>
<td>0.0079224</td>
<td>2.6877</td>
<td>0.019005</td>
<td>5.3624</td>
<td>0.037918</td>
<td></td>
</tr>
<tr>
<td>p=5</td>
<td>0.24022</td>
<td>1.2456</td>
<td>0.0088076</td>
<td>4.0486</td>
<td>0.028628</td>
<td>9.9778</td>
<td>0.070554</td>
<td></td>
</tr>
<tr>
<td>p=6</td>
<td>0.17606</td>
<td>1.4101</td>
<td>0.0099708</td>
<td>6.2117</td>
<td>0.043923</td>
<td>19.609</td>
<td>0.13866</td>
<td></td>
</tr>
<tr>
<td>p=7</td>
<td>0.13083</td>
<td>1.6501</td>
<td>0.011668</td>
<td>9.5938</td>
<td>0.067839</td>
<td>39.036</td>
<td>0.27603</td>
<td></td>
</tr>
<tr>
<td>p=8</td>
<td>0.098352</td>
<td>1.9637</td>
<td>0.013886</td>
<td>15.862</td>
<td>0.11230</td>
<td>77.650</td>
<td>0.54908</td>
<td></td>
</tr>
<tr>
<td>p=9</td>
<td>0.074661</td>
<td>2.3365</td>
<td>0.016522</td>
<td>25.912</td>
<td>0.18323</td>
<td>163.75</td>
<td>1.1579</td>
<td></td>
</tr>
<tr>
<td>p=10</td>
<td>0.057139</td>
<td>2.8966</td>
<td>0.020483</td>
<td>43.718</td>
<td>0.30914</td>
<td>348.82</td>
<td>2.4666</td>
<td></td>
</tr>
<tr>
<td>p=11</td>
<td>0.044026</td>
<td>3.6027</td>
<td>0.025475</td>
<td>76.220</td>
<td>0.53897</td>
<td>759.75</td>
<td>5.3723</td>
<td></td>
</tr>
<tr>
<td>p=12</td>
<td>0.034114</td>
<td>4.5443</td>
<td>0.032134</td>
<td>127.90</td>
<td>0.90441</td>
<td>1655.9</td>
<td>11.709</td>
<td></td>
</tr>
<tr>
<td>p=13</td>
<td>0.026559</td>
<td>5.7577</td>
<td>0.040721</td>
<td>226.59</td>
<td>1.6023</td>
<td>3588.2</td>
<td>25.373</td>
<td></td>
</tr>
<tr>
<td>p=14</td>
<td>0.020760</td>
<td>7.2715</td>
<td>0.051418</td>
<td>397.53</td>
<td>2.8110</td>
<td>8267.4</td>
<td>58.461</td>
<td></td>
</tr>
<tr>
<td>p=15</td>
<td>0.016282</td>
<td>9.6237</td>
<td>0.068079</td>
<td>712.49</td>
<td>5.0382</td>
<td>18751.</td>
<td>132.59</td>
<td></td>
</tr>
<tr>
<td>p=16</td>
<td>0.012808</td>
<td>12.374</td>
<td>0.087502</td>
<td>1314.7</td>
<td>9.2967</td>
<td>42258.</td>
<td>298.82</td>
<td></td>
</tr>
<tr>
<td>p=17</td>
<td>0.010100</td>
<td>16.281</td>
<td>0.11513</td>
<td>2403.2</td>
<td>16.993</td>
<td>103100</td>
<td>729.04</td>
<td></td>
</tr>
<tr>
<td>p=18</td>
<td>0.0079818</td>
<td>21.183</td>
<td>0.14979</td>
<td>4435.8</td>
<td>31.366</td>
<td>228070</td>
<td>1612.7</td>
<td></td>
</tr>
<tr>
<td>p=19</td>
<td>0.0063200</td>
<td>28.724</td>
<td>0.20312</td>
<td>7991.3</td>
<td>56.910</td>
<td>542260</td>
<td>3834.4</td>
<td></td>
</tr>
<tr>
<td>p=20</td>
<td>0.0050127</td>
<td>35.363</td>
<td>0.25006</td>
<td>14885.</td>
<td>105.26</td>
<td>1437200</td>
<td>10163.</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: $B_{pJ} = \frac{1}{2^{p-1}} \cdot \mathcal{E}[C(\beta)^{(p-1)/2}]$
<table>
<thead>
<tr>
<th>$C_{p,J}$</th>
<th>$J=2$</th>
<th>$J=4$</th>
<th>$J=6$</th>
<th>$J=8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>num</td>
<td>est</td>
<td>se</td>
<td>est</td>
</tr>
<tr>
<td>p=2</td>
<td>-0.33910</td>
<td>1.1492</td>
<td>0.0029695</td>
<td>*</td>
</tr>
<tr>
<td>p=3</td>
<td>-0.55950</td>
<td>-3.3609</td>
<td>0.00089475</td>
<td>*</td>
</tr>
<tr>
<td>p=4</td>
<td>-0.70274</td>
<td>-5.8856</td>
<td>0.0019350</td>
<td>*</td>
</tr>
<tr>
<td>p=5</td>
<td>-0.79150</td>
<td>-9.4732</td>
<td>0.0081824</td>
<td>-46.475</td>
</tr>
<tr>
<td>p=6</td>
<td>-0.83959</td>
<td>-14.733</td>
<td>0.021303</td>
<td>-95.018</td>
</tr>
<tr>
<td>p=7</td>
<td>-0.85655</td>
<td>-22.366</td>
<td>0.045486</td>
<td>-191.20</td>
</tr>
<tr>
<td>p=8</td>
<td>-0.84969</td>
<td>-33.565</td>
<td>0.087522</td>
<td>-381.79</td>
</tr>
<tr>
<td>p=9</td>
<td>-0.82497</td>
<td>-49.993</td>
<td>0.15824</td>
<td>-745.97</td>
</tr>
<tr>
<td>p=10</td>
<td>-0.78732</td>
<td>-74.431</td>
<td>0.27435</td>
<td>-1512.8</td>
</tr>
<tr>
<td>p=11</td>
<td>-0.74086</td>
<td>-109.17</td>
<td>0.46083</td>
<td>-3039.3</td>
</tr>
<tr>
<td>p=12</td>
<td>-0.68893</td>
<td>-161.68</td>
<td>0.76320</td>
<td>-6062.0</td>
</tr>
<tr>
<td>p=13</td>
<td>-0.63421</td>
<td>-239.62</td>
<td>1.2486</td>
<td>-12219.</td>
</tr>
<tr>
<td>p=14</td>
<td>-0.57878</td>
<td>-351.59</td>
<td>1.9959</td>
<td>-24739.</td>
</tr>
<tr>
<td>p=15</td>
<td>-0.52420</td>
<td>-520.62</td>
<td>3.1798</td>
<td>-50355.</td>
</tr>
<tr>
<td>p=16</td>
<td>-0.47163</td>
<td>-758.38</td>
<td>5.0075</td>
<td>-100450.</td>
</tr>
<tr>
<td>p=17</td>
<td>-0.42183</td>
<td>-1115.1</td>
<td>7.8292</td>
<td>-205580.</td>
</tr>
<tr>
<td>p=18</td>
<td>-0.37532</td>
<td>-1649.6</td>
<td>12.248</td>
<td>-409450.</td>
</tr>
<tr>
<td>p=19</td>
<td>-0.33238</td>
<td>-2422.0</td>
<td>18.968</td>
<td>-849320.</td>
</tr>
<tr>
<td>p=20</td>
<td>-0.29311</td>
<td>-3526.0</td>
<td>29.178</td>
<td>-1768500.</td>
</tr>
</tbody>
</table>

Table 3.3: \[ C_{p,J} = \frac{1}{2^{p}} \cdot \mathbb{E} \left[ D(p, J, C(\beta)) \right] \]
<table>
<thead>
<tr>
<th>$B_{p,J}^*$</th>
<th>J=4</th>
<th>J=6</th>
<th>J=8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>num</td>
<td>est</td>
<td>se</td>
</tr>
<tr>
<td>p=2</td>
<td>1.2662</td>
<td>1.5817</td>
<td>0.011185</td>
</tr>
<tr>
<td>p=3</td>
<td>1.6146</td>
<td>2.5544</td>
<td>0.018062</td>
</tr>
<tr>
<td>p=4</td>
<td>2.0728</td>
<td>4.2442</td>
<td>0.030011</td>
</tr>
<tr>
<td>p=5</td>
<td>2.6785</td>
<td>7.1432</td>
<td>0.050510</td>
</tr>
<tr>
<td>p=6</td>
<td>3.4830</td>
<td>12.382</td>
<td>0.087556</td>
</tr>
<tr>
<td>p=7</td>
<td>4.5562</td>
<td>21.456</td>
<td>0.15171</td>
</tr>
<tr>
<td>p=8</td>
<td>5.9933</td>
<td>38.049</td>
<td>0.26905</td>
</tr>
<tr>
<td>p=9</td>
<td>7.9248</td>
<td>68.685</td>
<td>0.48568</td>
</tr>
<tr>
<td>p=10</td>
<td>10.529</td>
<td>126.06</td>
<td>0.89138</td>
</tr>
<tr>
<td>p=11</td>
<td>14.052</td>
<td>228.61</td>
<td>1.6166</td>
</tr>
<tr>
<td>p=12</td>
<td>18.830</td>
<td>428.47</td>
<td>3.0298</td>
</tr>
<tr>
<td>p=13</td>
<td>25.327</td>
<td>805.84</td>
<td>5.6981</td>
</tr>
<tr>
<td>p=14</td>
<td>34.181</td>
<td>1516.6</td>
<td>10.724</td>
</tr>
<tr>
<td>p=15</td>
<td>46.272</td>
<td>2860.9</td>
<td>20.230</td>
</tr>
<tr>
<td>p=16</td>
<td>62.815</td>
<td>5539.0</td>
<td>39.168</td>
</tr>
<tr>
<td>p=17</td>
<td>85.488</td>
<td>10534.</td>
<td>74.489</td>
</tr>
<tr>
<td>p=18</td>
<td>116.61</td>
<td>19916.</td>
<td>140.83</td>
</tr>
<tr>
<td>p=19</td>
<td>159.40</td>
<td>39272.</td>
<td>277.69</td>
</tr>
<tr>
<td>p=20</td>
<td>218.29</td>
<td>75130.</td>
<td>531.30</td>
</tr>
</tbody>
</table>

Table 3.4: $B_{p,J}^* = \frac{1}{2^{p-1}} \cdot \mathcal{E} [C^*(\beta)^{(p-1)/2}]$
<table>
<thead>
<tr>
<th>$C_{pJ}^s$</th>
<th>J=4</th>
<th>J=6</th>
<th>J=8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>num</td>
<td>est</td>
<td>se</td>
</tr>
<tr>
<td>p=2</td>
<td>-0.63311</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>p=3</td>
<td>-2.2743</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>p=4</td>
<td>-5.6583</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>p=5</td>
<td>-11.966</td>
<td>-58.994</td>
<td>0.033799</td>
</tr>
<tr>
<td>p=6</td>
<td>-23.077</td>
<td>-139.41</td>
<td>0.14022</td>
</tr>
<tr>
<td>p=7</td>
<td>-41.949</td>
<td>-319.10</td>
<td>0.46933</td>
</tr>
<tr>
<td>p=8</td>
<td>-73.210</td>
<td>-716.23</td>
<td>1.3888</td>
</tr>
<tr>
<td>p=9</td>
<td>-124.05</td>
<td>-1597.5</td>
<td>3.7861</td>
</tr>
<tr>
<td>p=10</td>
<td>-205.56</td>
<td>-3521.9</td>
<td>9.9240</td>
</tr>
<tr>
<td>p=11</td>
<td>-334.79</td>
<td>-7713.6</td>
<td>25.030</td>
</tr>
<tr>
<td>p=12</td>
<td>-537.83</td>
<td>-1672.8</td>
<td>61.324</td>
</tr>
<tr>
<td>p=13</td>
<td>-854.41</td>
<td>-3641.4</td>
<td>148.78</td>
</tr>
<tr>
<td>p=14</td>
<td>-1344.9</td>
<td>-7871.1</td>
<td>354.30</td>
</tr>
<tr>
<td>p=15</td>
<td>-2100.4</td>
<td>-17235.0</td>
<td>839.43</td>
</tr>
<tr>
<td>p=16</td>
<td>-3258.7</td>
<td>-36586.0</td>
<td>1937.2</td>
</tr>
<tr>
<td>p=17</td>
<td>-5026.7</td>
<td>-7892000</td>
<td>4485.0</td>
</tr>
<tr>
<td>p=18</td>
<td>-7715.2</td>
<td>-1718200</td>
<td>10369.</td>
</tr>
<tr>
<td>p=19</td>
<td>-11789.</td>
<td>-3672300</td>
<td>23657.</td>
</tr>
<tr>
<td>p=20</td>
<td>-17943.</td>
<td>-7990800</td>
<td>53986.</td>
</tr>
</tbody>
</table>

Table 3.5: $C_{pJ}^s = \frac{1}{2p} \cdot \mathcal{E} [ D(p, J, C^s(\beta)) ]$
Chapter 4

Comparisons with Monte Carlo Results

Our main goal of this thesis is to calculate

\[ P\{ \max_{\alpha \in S_{p-1}} I_f(\alpha) \geq a \}, \]  

(4.1)

which we denote by \( F(a) \). However, as mentioned in Section 2.2 (see the paragraph following Remark 2.1), it is not possible for us to find \( F(a) \) exactly. To check how our two term approximation formula (2.29) (or (2.30) for spheroid data) for (4.1) works, we can only compare this theoretical approximation (2.29) (or (2.30)), namely \( G(a) \), with a simulated value \( \hat{F}_M(a) \) of sample size \( M \).

4.1 Notations and Algorithms in Simulation
Suppose $z_i^1, z_i^2, \ldots, z_i^N$ is the observed $i$th data set with sample size $N$ from a
$p$ dimensional normal population $N(0, I_p)$, $a_i$ is the observed JHF index based on
the data set $i$: $a_i = \max_{\alpha \in S^{p-1}} I_J(\alpha | z_i^1, \ldots, z_i^N)$, $i = 1, 2, \ldots, M$. The standard
simulation estimator for (4.1) is

$$\hat{F}_M(a) = \frac{1}{M} \sum_{i=1}^{M} I_{\{a_i \geq a\}}, \quad (4.2)$$

where $I_A$ is an indicator function of $A$, which specifies $I_A(x) = 1$ if $x \in A$, or
$I_A(x) = 0$ otherwise.

Since "$\hat{F}_M(a) = G(a)$" iff "$a = G(\hat{F}_M^{-1}(\alpha))$", We shall compare the theoretical
formula $G$ in (2.29) and the simulation formula $\hat{F}_M$ in (4.2) in the following way:

$$\alpha \xrightarrow{\hat{F}_M} \hat{c}_M \equiv \hat{F}_M^{-1}(\alpha) \xrightarrow{G} G(\hat{c}_M) \quad (4.3)$$

viz. for a given $\alpha \in (0, 1)$, we calculate an $100\alpha\%$ quantile $\hat{c}_M$ of $\hat{F}_M$, then evaluate
$G$ at this $\hat{c}_M$. We interpret closeness of $\alpha$ and $G(\hat{c}_M)$ to indicate that $G$ is a
good approximation to $F$. Note, here $\alpha$ is a univariate in $(0, 1)$, rather than the $p$
dimensional index of a PP index.

Standard Error in Comparisons

The error in $G(\hat{c}_M)$ is due to simulation error in $\hat{c}_M$ and approximation error
from $G$, where simulation error is composed of imperfect optimizer as explained in
the next paragraph and randomness of the simulation formula $\hat{F}_M(a)$. Therefore, to
give a rigorous standard error formula associated with $G(\hat{c}_M)$ is difficult. However,
a Taylor series expansion shows that if $G \equiv F$, the standard error of $G(\hat{c}_M)$ is
asymptotically as $M \to \infty$,

$$se = \sqrt{\frac{\alpha(1-\alpha)}{M}}, \quad (4.4)$$

59
which thus is an indication for how good $G$ is as a theoretical approximation formula.

**PP Algorithm in Comparisons**

As Remark 3.3 in Section 3.2.1., Monte Carlo methods in PP are computationally intensive and difficult to use, since they involve repeated solution of a complicated optimization problem. If a PP algorithm is not good enough, the computer calculated value (output) for the observed $\max_{\alpha \in S^{p-1}} I_J(\alpha | z_1^i, \ldots, z_2^i)$ can be merely a local maximum of $I_J(\alpha | z_1^i, \ldots, z_2^i)$ which is far away from the global maximum. Hence the calculated quantile $\hat{\alpha}_M$ tends to be larger than the true quantile with the result that $G(\hat{\alpha}_M)$ would appear to be much smaller than $\alpha$ even if $G \equiv F$. Therefore, we should be cautious about choosing a PP algorithm (optimization algorithm). Generally speaking, when data dimension increases, the requirement for optimization procedure is higher and higher, and the number of the terms in the approximation formula (or order of approximation formula) becomes crucial.

Here is a description of JHF’s PP algorithm (cf. Friedman, 1987): for $p > 0$,

1. Preliminary Search: search along $p$ coordinates (or principle components in the sphered case), the one with largest PP index is called $\alpha^{(1)}$.

2. Coarse Search: large stepping search along $p$ orthogonal directions with origin at $\alpha^{(1)}$, the one with largest PP index is called $\alpha^{(2)}$.

3. Local search: starting from $\alpha^{(2)}$, search to a convergence point by using steepest ascent or quasi-Newton algorithm as the local optimizer, the one with largest PP index after convergence of the algorithm is called $\alpha^{(3)}$. 

60
4. Structure Removal: transform the data (or remove the structure), so that the PP index evaluated at $\alpha^{(3)}$ based on the new data set is a minimum of the PP index. Therefore, if we go through the above steps 1-3 again, we will not end at the same direction we found before.


After a PP algorithm is completed, we choose the direction which maximizes the PP index among all the directions examined in (above steps of) the algorithm as $\max_{\alpha \in I} I_J(\alpha)$.

The first modified version of JHF PP algorithm in our experiments is that the local optimizer in the third step of JHF PP algorithm is replaced by NPSOL(version 4.0).

NPSOL uses a sequential quadratic programming algorithm. It has been developed into several refined versions by Gill, Murray, Wright (cf. their book published in 1981 and User's Guide for NPSOL (Version 4.0), 1986). It is considered to be one of the most efficient algorithms among the existing local optimizers.

The second modified version of the PP algorithm in our experiments is that the step 1 - 3 in JHF PP algorithm are replaced by a so called global optimization subroutine from IMSL.

The so called global optimizer from IMSL is called ZXMWD in release 9.2. ZXMWD minimizes the objective function (in our problem, we like to minimize $-I_J(\alpha)$), and calls a modified subroutine using Newton-quasi method to do about 4 iterations with each of $n(= \min(2^{p-1} + 5, 100))$ randomly starting points, the five of which result in the lowest values of the functions are allowed to continue
to convergence. The local minimum found which, of these five, gives the lowest objective function value is taken to be the global minimum.

The third modified version of the PP algorithm in our experiments is that we add a rotation technique either to the end of JHF PP algorithm, or the above first modified version, or the second modified version.

The rotation technique is based on the idea that the maximum of the PP index is an invariant under any orthogonal transformation for the data. Hence, after we found a convergent point in the step 3 (or after step 4, if structure removal is used), we rotate data $180^\circ$, then repeat the step 2 and 3 (and step 4 if structure removal technique is used) and choose one direction with the biggest PP index. (Thanks to Professor I. Johnston for suggesting the rotation technique).

If the number of structure removals in the algorithm is zero in a particular experiment, it indicates that we do not use the step 4 (Structure removal) of JHF algorithm in this experiment.

In the most recent version of IMSL (release 10) and NAG (release 11), there is no subroutine claimed to be a global optimizer as ZXMWD does, there are DNCON as a local optimizer in IMSL and EO4HEF as a local optimizer in NAG, both of them use successive quadratic programming algorithms. We do not experiment with these two subroutines, as they use a similar algorithm to NPSOL, and the people who have used NPSOL at Stanford University find NPSOL (version 4.0) efficient and reliable.

4.2 Unshered Case
In this section, we discuss a few experiments in the unsphered case, and make some summaries. More experiments are presented in tables in Appendix C.

Table 4.1 on page 67 gives comparisons of \( G(\alpha) \) and \( \hat{F}_M(\alpha) \), when the data dimension \( p \) is 3 and the number \( J \) of the terms in JHF PP index is 4. In this table, the column labeled “p3j4” gives a one term approximation, \( G(\hat{e}_M) \), for various \( \hat{e}_M \) corresponding to different values of \( \alpha \) in column 1. The column labeled “p3j4.\( \kappa_2 \)” stands for \( G(\hat{e}_M) \), where the two term approximation formula rather than only one term approximation formula is applied. The column labeled “se” is the standard error associated with \( G(\hat{e}_M) \) defined in (4.4). Here, the PP algorithm is the second modified version, i.e. the optimizer is ZXMWD. We do see that the third and fifth column are not terribly far away from the first column, taking into account the given standard error (se). However, there is an improvement by adding the second term approximation. To conclude, \( G(\alpha) \) and the corresponding \( \hat{F}_M(\alpha) \) are close for the reasonable \( \alpha \) in this case.

Similarly, Table 4.2 on page 68 is about the comparisons of \( G(\alpha) \) and \( \hat{F}_M(\alpha) \), when the data dimension \( p \) is 4, the number \( J \) of the terms in JHF PP index is 4. In this table, we use sophisticated PP algorithms and two term approximation formula for four different experiments. The column labeled “nlp4j4.\( \kappa_2 \)” stands for \( G(\hat{e}_M) \), where the algorithm is the first modified version with zero number of the structure removals, i.e. the optimizer is a combination of step 1, 2 of JHF algorithm and the local optimizer NPSOL. The column labeled “rp4j4.\( \kappa_2 \)” stands for \( G(\hat{e}_M) \), where the PP algorithm is the third modified version with the rotation technique added to the second modified version, i.e. the optimizer is a combination of ZXMWD and rotation technique. The column labeled “nlp4j4m.\( \kappa_2 \)” stands for \( G(\hat{e}_M) \), where the algorithm is the first modified version with 4 structure removals, i.e. the optimizer
is a combination of step 1,2 of JHF PP algorithm, NPSOL and structure removal technique in step 4 of JHF PP algorithm. Columns 4 – 7 are all reasonably close to the first column for this case. We also see the improvement brought in by applying rotation technique and structure removals.

Other experiments for different sets of $p$ and $J$ ($p \in (2, 3, 4, 6, 8)$, $J \in (1, 2, 4, 6)$) are presented in Appendix C.

All our experiments, including ones in the next section, show the importance of introducing two terms approximation formula over one term approximation formula, especially, when $p + J \geq 4$. In fact, when data dimension $p$ is 2, or the number $J$ of the index used is 1 or 2, the one term approximation formula is good enough, but these cases are uninteresting statistically. When $p$ is 3, or $p$ is 2 with $J$ equal to 4, there is improvement by supplying the second term approximation. When $p \geq 4$, the second term approximation becomes crucial. When $p = 8$, the two term approximation is good enough for $\alpha \leq 0.10$. When $p = 10$, the two term approximation is good enough for $\alpha \leq 0.05$. For larger data dimension case, we may need a three term approximation formula. It is possible that we can strengthen the conditions in our Theorem 2.2 to get an extra term. However, this third term involves the calculation of its coefficient $\kappa_4$ which explicit geometric meaning and applicable formula are hard to get at this time. Therefore we are restricted to the two term approximation formula at this point.

We also see from these comparisons that

1. the bigger the data dimension is, the more complicated optimizer has to be used to achieve reliability;

2. the rotation technique and the local optimizer NPSOL are very useful for our
problem.

Our third modified version, with $p$ structure removals, and rotation technique added to the first modified version, turns out to be the most efficient and reliable one when data are from normal population (noise data) for data dimension $p$ is 4 or 5. This is our first recommended modified PP algorithm (RMA1 for short). Since the RMA1 is reliable in the null case (noise data case), it is also reliable in non-null case. As for the efficiency of the RMA1 in non-null case, a reader should read the third paragraph below.

The third modified version, with $p$ structure removals, rotation technique added to the second modified version, is better than the first recommended version in terms of the reliability of the optimizer if $p$ is greater or equal to 6. This is our second recommended modified version (RMA2 for short). Note that the RMA1 is satisfactory for $p = 6$.

In fact, the result from using the algorithm based on the RMA1 is comparable to the result for $p = 4$ for which grid search or exhaustive search are used on supercomputer CRAY. This RMA1, where NPSOL is the local optimizer, is faster than JHF PP algorithm, if an equal number of structure removals is used. The RMA2 is slower than the RMA1, so the reliability gained from the RMA2 over the RMA1 can be taken as the tradeoff of using a slower algorithm. We can imagine that if a few more structure removals are used in the RMA1, it might work as well as the RMA2 if $p > 6$. This needs further investigation.

If the data is really structured (or far away from normal), fewer structure removals than we specify as follows may be enough to achieve (or be close to) the true maximum. However, we always like to find as many informative views as possible, so
it is advised that we still do structure removals several times if the corresponding P-values encourage so. This is one of the so called by-products which helps us to decide when to reasonably stop the algorithm (cf. Remark 1.1.). Adding one rotation into the algorithm improves reliability and does not substantially increase computation compared to the case where we may have to use many structure removals to find the solution. One obvious example of such case is the null case, or the case close to null.

The first recommended JHF PP algorithm (RMA1) is

1. Preliminary Search: search along $p$ coordinates (or principle components in the sphered case), the one with largest PP index is called $\alpha^{(1)}$.

2. Coarse Search: large stepping search along $p$ orthogonal directions with origin at $\alpha^{(1)}$, the one with largest PP index is called $\alpha^{(2)}$.

3. Local search: starting from $\alpha^{(2)}$, search to a convergence point by NPSOL, the one with largest PP index after convergence of the algorithm is called $\alpha^{(3)}$.

4. Structure Removal: transform the data (or remove the structure), so that the PP index evaluated at $\alpha^{(3)}$ based on the new data set is a minimum of the PP index. Therefore, if we go through above step 1-3 again, we will not end at the same direction we found before.

5. Loop 1: repeat above step 2-4 no more than $p$ times.

6. Loop 2: rotate the original data, repeat step 2-5.

The second recommended JHF PP algorithm (RMA2) is to replace the above step 1 – 3 by ZXMWD.
In summary, in the unsphered data case, our theoretical formula $G$ in (2.29) matches the simulation formula $\hat{F}_M(\alpha)$, if the two terms approximation formula is used and our PP algorithm is reliable.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>p3j4</th>
<th>p3j4.$\kappa_2$</th>
<th>p3j4</th>
<th>p3j4.$\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.34323</td>
<td>0.20200</td>
<td>0.37090</td>
<td>0.21649</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.24493</td>
<td>0.14895</td>
<td>0.24760</td>
<td>0.15045</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.14821</td>
<td>0.093957</td>
<td>0.17029</td>
<td>0.10681</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.065848</td>
<td>0.044010</td>
<td>0.072232</td>
<td>0.048019</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.040901</td>
<td>0.028036</td>
<td>0.045975</td>
<td>0.031330</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.028074</td>
<td>0.019586</td>
<td>0.028144</td>
<td>0.019633</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.014332</td>
<td>0.010276</td>
<td>0.016171</td>
<td>0.011542</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0022305</td>
<td>0.0022676</td>
<td>0.0017212</td>
<td>0.012062</td>
<td>0.0087045</td>
</tr>
</tbody>
</table>

Table 4.1: Case: $p = 3, J = 4$
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$s$</th>
<th>nlp4j4.(\kappa_2)</th>
<th>rp4j4.(\kappa_2)</th>
<th>nlp4j4.(\kappa_2)</th>
<th>nlp4j4m.(\kappa_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.24151</td>
<td>0.22440</td>
<td>0.25946</td>
<td>0.21840</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.16866</td>
<td>0.15932</td>
<td>0.18252</td>
<td>0.14898</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.11372</td>
<td>0.10294</td>
<td>0.12377</td>
<td>0.094139</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.049910</td>
<td>0.052006</td>
<td>0.048690</td>
<td>0.045339</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.034796</td>
<td>0.026315</td>
<td>0.027539</td>
<td>0.027162</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.018910</td>
<td>0.012418</td>
<td>0.015391</td>
<td>0.021651</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.0088799</td>
<td>0.0065408</td>
<td>0.0055731</td>
<td>0.013646</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0022305</td>
<td>0.0035427</td>
<td>0.0036490</td>
<td>0.0034910</td>
<td>0.0070255</td>
</tr>
</tbody>
</table>

Table 4.2: Case: $p = 4, J = 4$, part I
4.3 Sphered Case

Sphering data before implementing PP procedures is for eliminating the effect of unknown scale and location. An improper hope is that the result for unsphered case should work for sphered case without any modification. Table C.10 in Appendix C shows if we still use (2.29) as $G$ for the sphered data case, $G(\hat{c}_M)$ is very bad compared to $\alpha$ when $\alpha$ is 0.20, 0.15, 0.10, 0.05, 0.03, 0.02, 0.01, 0.005. Some of $G(\hat{c}_M)$ are even bigger than 1. For example, when the data dimension $p$ is 4, the number $J$ of the terms in JHF PP index is 4, for $\alpha = 0.20$, the simulation result gives $G(\hat{c}_M)$ 2.1700 if one term approximation is used, and 0.83800 if two terms approximation formula is used.

For this sphered data, the large sample theory does not work in the same way as for the unsphered data which are independent, identically distributed.

Let us check the behavior of $\hat{Y}_l(\alpha)$, the $l$th term in the PP index, for $l = 1, 2, \ldots, J$ by simulations presented in Table 4.3 on page 70 and the following Figure 4.2 on page 74. In these simulations, the data are 4 dimensional pseudo normal random vectors from $\mathcal{N}(0, I_4)$, the sample size is $N$ and the simulation size $M$ is 1000. We first sphere the data and calculate $\hat{Y}_l(\alpha)$ for $l = 1, 2, 3, 4$ for each simulation, then compute the sample mean and variance of $\hat{Y}_l(\alpha)$. The sample mean and variance of the 1000 pseudo univariate standard normal random numbers (labeled st. normal) are also represented in Table 4.3. According to large sample theory, in the unsphered data case, these $\hat{Y}_l(\alpha), l = 1, 2, \ldots, J$ are independent and identically distributed as univariate standard normal random variable from $\mathcal{N}(0, 1)$.

In Figure 4.2, it is significant that $\hat{Y}_l(\alpha)$ are normally distributed by its $Q - Q$
plot, for $\alpha = (1, 0, 0, 0), i = 1, 2, 3, 4$.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha = (0, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$</th>
<th>$\alpha = (1, 0, 0, 0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N=100$</td>
<td>$N=1000$</td>
</tr>
<tr>
<td>$\bar{Y}_1(\alpha)$</td>
<td>0.0414</td>
<td>0.0147</td>
</tr>
<tr>
<td>$\bar{Y}_2(\alpha)$</td>
<td>0.22</td>
<td>-0.004</td>
</tr>
<tr>
<td>$\bar{Y}_3(\alpha)$</td>
<td>0.92</td>
<td>-0.080</td>
</tr>
<tr>
<td>$\bar{Y}_4(\alpha)$</td>
<td>0.853</td>
<td>0.0490</td>
</tr>
<tr>
<td>st. normal</td>
<td>0.98503</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Table 4.3: $\bar{Y}_l(\alpha)$ in Sphered Case

In this Table 4.3, the mean of these $\bar{Y}_l(\alpha)$'s are reasonably close to zero; the variances of $\bar{Y}_3(\alpha)$ and $\bar{Y}_4(\alpha)$ are close to one, but the variance of $\bar{Y}_1(\alpha)$ is only about 0.04, and the variance of $\bar{Y}_2(\alpha)$ is about 0.22 which are closer to 0 rather than 1. Hence, something must have changed dramatically for $\bar{Y}_1(\alpha)$ and $\bar{Y}_2(\alpha)$ from unsphered data case to sphered data case.

The following regression argument explains the above phenomenon.

When data $Z_1, Z_2, \ldots, Z_N$ are sphered, they satisfy

$$\sum_{i=1}^{N} Z_i = 0,$$  \hspace{1cm} (4.5)

$$\frac{1}{N} \sum_{i=1}^{N} Z_i Z_i^\top = I_p,$$ \hspace{1cm} (4.6)
and concentrate in a small neighborhood of 0, in which \( \Phi \) function can be approximate by a straight line with intercept 0.5 and slope \( a \). See Figure 4.1 on page 71.

Recall that

\[
I_J(\alpha) = \frac{1}{2N} \sum_{j=1}^{J} (\tilde{Y}_j(\alpha))^2,
\]

where

\[
\tilde{Y}_j(\alpha) = \sqrt{\frac{2j+1}{N}} \sum_{i=1}^{N} P_j(R_i), \quad R_i = 2\Phi(\alpha^T Z_i) - 1,
\]

for \( j = 1, \ldots, J \), \( P_j \) is the Legendre polynomial of order \( j \) (see section 2.1 for its representation). Henceforth, \( \tilde{Y}_l(\alpha) \) is like an average of the Legendre polynomial of order \( l \), and intuitively the first term \( \tilde{Y}_1(\alpha) \) of the index is affected by (4.5), the second term \( \tilde{Y}_2(\alpha) \) is affected by (4.6), the third, fourth, and higher terms are safe from the constraints (4.5) and (4.6). As \( \Phi \) can be approximated by a straight line in the small neighborhood of 0, in which the sphered data concentrate.
with a very high probability, we can approximately substitute \( \Phi(x) \) by \( ax + 0.5 \) in 
\[
R_i = 2\Phi(\alpha^T Z_i) - 1, \text{ i.e. } R_i \approx 2a\alpha^T Z_i. \text{ Then}
\]
\[
\hat{Y}_1(\alpha) = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} P_1(R_i)
\]
\[
\approx \sqrt{\frac{3}{N}} \sum_{i=1}^{N} 2a\alpha^T Z_i = 0, \text{ by (4.5)}.
\]
\[
\hat{Y}_2(\alpha) = \sqrt{\frac{5}{N}} \sum_{i=1}^{N} P_2(R_i),
\]
\[
\approx \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \frac{1}{2} (3 \cdot (2a\alpha^T Z_i)^2 - 1)
\]
\[
= 6 \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \frac{1}{2} (a^2 \alpha^T Z_i Z_i^\alpha \alpha - \frac{1}{12})
\]
\[
= 0, \text{ for } a = \frac{1}{\sqrt{12}}, \text{ by (4.6)}.
\]

Here \( a = 1/\sqrt{12} \) gives a good approximation \( ax + 1/2 \) to \( \Phi(x) \) in the sense of minimizing
\[
\mathcal{E} \left\{ (\Phi(X) - ax - 0.5)^2 \cdot I_{(|X| \leq 3)} \right\} + P \{ |X| > 3 \}
\]
and close to \( a = 1/\sqrt{4\pi} \) which minimizes
\[
\mathcal{E} \left\{ (\Phi(X) - ax - 0.5)^2 \right\},
\]
where \( X \) is a \( \mathcal{N}(0, 1) \) random variable. Intuitively speaking, our sphering reduces dimensionality of the corresponding manifold of the PP index by 2.

There is another way to check the above explanation for the effect of the sphering. If the data are centered rather than sphered, we only have one constraint \( \sum_{i=1}^{N} Z_i = 0 \), not (4.6). The first term \( \hat{Y}_1(\alpha) \) in PP index should be the only term affected by the analogy from the sphered data case. Table C.11 in Appendix C confirms this argument.
Table 4.3 and Table C.11 in Appendix C also reveals that the affected $\tilde{Y}_1(\alpha)$ has variance about 0.04, and the affected $\tilde{Y}_2(\alpha)$ has variance about 0.22 in all those simulations. In Chapter 5, we shall give a theoretical explanation for these two seemingly magical numbers.

Therefore, in the case of sphered data, it is reasonable that we treat the corresponding limiting process $Z^*(t)$ from JHF PP index as from the index $I_j^*(\alpha)$ summed up from $j = 3, \ldots, J$:

$$I_j^*(\alpha) = \frac{1}{2N} \sum_{j=3}^J (\tilde{Y}_j(\alpha))^2,$$

with $\tilde{Y}_3(\alpha), \ldots, \tilde{Y}_j(\alpha)$ as from unsphered version, which are defined in (2.2). Under this treatment, we have derived (2.30) in Chapter 2.

In Table 4.3, “srp4j4m.34.\kappa_2” has the same meaning as “rp4j4m.\kappa_2” in Table 4.2 except the data are sphered before implementing PP procedure, $G$ here is (2.30) not (2.29). This table demonstrates that our treatment, of deleting first two terms in sphered data case, is sensible. We also see that one term approximation formula is not good enough for the P-value of PP, and the rotation technique is very useful, when $p = 4$.

In summary, our two term theoretical formula (2.30) matches the simulation formula $G$. 
Figure 4.2: The $Q-Q$ Plot for $\hat{Y}_i(\alpha)$, $\alpha = (1, 0, 0, 0)$. 
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>srp4j4m.34</th>
<th>srp4j4m.34.$\kappa_2$</th>
<th>snlp4j4m.34</th>
<th>snlp4j4m.34.$\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.30994</td>
<td>0.21463</td>
<td>0.34585</td>
<td>0.23790</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.22917</td>
<td>0.16144</td>
<td>0.24161</td>
<td>0.16972</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.13661</td>
<td>0.098742</td>
<td>0.16664</td>
<td>0.11932</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.076385</td>
<td>0.056574</td>
<td>0.077261</td>
<td>0.057199</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.048656</td>
<td>0.036634</td>
<td>0.045052</td>
<td>0.034008</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.038613</td>
<td>0.029298</td>
<td>0.030433</td>
<td>0.023262</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.018110</td>
<td>0.014050</td>
<td>0.011606</td>
<td>0.091097</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0022305</td>
<td>0.010185</td>
<td>0.0080198</td>
<td>0.0044942</td>
<td>0.003600</td>
</tr>
</tbody>
</table>

Table 4.4: Sphered Case: $p = 4$, $J = 4$
Chapter 5

More about the Sphered Case

As before, suppose $Y_1, \ldots, Y_N$ are independent and identically distributed data from a $p$ dimensional normal population $\mathcal{N}(\mu, \Sigma)$, where $\mu$ is some $p$ dimensional vector $(\mu_1, \mu_2, \ldots, \mu_p)^\top$, and $\Sigma$ is a $p \times p$ nonsingular matrix. Assume $Z_1, \ldots, Z_N$ are the modified version of $Y_1, \ldots, Y_N$, either sphered, centered or unsphered as specified whenever it is necessary. Let $\tilde{Y}_j(\alpha)$ be the average of the $j$th Legendre polynomial $P_j$ evaluated at $N$ different points $\alpha^\top Z_i$, $i = 1, \ldots, N$:

$$
\tilde{Y}_j(\alpha) = \sqrt{\frac{2j + 1}{N}} \sum_{i=1}^{N} P_j(2\Phi(\alpha^\top Z_i) - 1)
$$

for $j = 1, \ldots, J$ (cf. (2.2)). Here for $i = 1, 2, \ldots, N$, $Z_i = (Z_{1i}, Z_{2i}, \ldots, Z_{pi})^\top$, and $Y_i = (Y_{1i}, Y_{2i}, \ldots, Y_{pi})^\top$ are $p$ dimensional vectors. For $j = 1, \ldots, J$, define $\tilde{Y}_{ju}(\alpha)$ as

$$
\tilde{Y}_{ju}(\alpha) = \sqrt{\frac{2j + 1}{N}} \sum_{i=1}^{N} P_j(2\Phi(\alpha^\top \tilde{Z}_i) - 1),
$$

(5.1)

where $\tilde{Z}_i = \tilde{D}^{-\frac{1}{2}} \tilde{U}^\top(Y_i - \mu)$, $\tilde{U} \tilde{D} \tilde{U}^\top = \Sigma$ is an eigenvalue-eigenvector decomposition of the covariance matrix of $\mathcal{N}(\mu, \Sigma)$. These $\tilde{Y}_{1u}(\alpha), \tilde{Y}_{2u}(\alpha), \ldots, \tilde{Y}_{Ju}(\alpha)$ are
asymptotically as $N \to \infty$ independent, identically distributed from $\mathcal{N}(0,1)$.

In this chapter, we show that in the sphered case, as $N \to \infty$ each $\tilde{Y}_j(\alpha)$ is distributed asymptotically according to an univariate normal population $\mathcal{N}(0, \sigma_j^2)$ for $j = 1, 2, \ldots, J$, with $\sigma_1^2, \sigma_2^2$ close to zero, $\sigma_3^2, \ldots, \sigma_J^2$ close to one; and prove that as $N \to \infty$,

$$
\tilde{Y}_j(\alpha) = \tilde{Y}_{ju}(\alpha) - R_{js} = O_p\left(\frac{1}{\sqrt{N}}\right)
$$

with $\text{Cov}(\tilde{Y}_{ju}, R_{js})$, $\text{Cov}(R_{js})$ close to zero for $j = 3, \ldots, J$. The summary of these $\sigma_j^2$'s is presented in Table 5.1 on page 89.

This limiting result of $\tilde{Y}_j(\alpha)$'s gives a theoretical justification for our treatment of the sphered case. Our treatment for sphered data is to approximate the PP index $I_J(\alpha) = \frac{1}{2N} \sum_{j=1}^{J} (\tilde{Y}_j(\alpha))^2$ by the index $I_J^*(\alpha)$ summed up from $j = 3, \ldots, J$:

$$
I_J^*(\alpha) = \frac{1}{2N} \sum_{j=3}^{J} (\tilde{Y}_j(\alpha))^2
$$

with $\tilde{Y}_3(\alpha), \ldots, \tilde{Y}_J(\alpha)$ same as those from unsphered data ($Z_i = Y_i, i = 1, \ldots, N$) when $\mu = 0, \Sigma = I_p$ (see Section 2.1). Then these $\tilde{Y}_3(\alpha), \ldots, \tilde{Y}_J(\alpha)$ are distributed as $\tilde{Y}_{3u}(\alpha), \ldots, \tilde{Y}_{Ju}(\alpha)$, and $2NI_J^*(\alpha)$ is asymptotically distributed as $\chi^2_{J-2}$. Therefore, the resulting two term approximation formula (2.30) for $Z^*(t)$ derived from $I^*(\alpha)$ is reasonable in the sphered case.

In following Section 5.1, we deal first with the simpler case of centered data, to give a simple prototype of our justification for sphered data which is studied in Section 5.2. In this prototype, Central Limit Theorem and Taylor Expansion are applied.
5.1 Centered Case

For this *centered* case, $Z_1, \ldots, Z_N$ are centered version of the data $Y_1, \ldots, Y_N$, i.e.,

$$Z_i = Y_i - \bar{Y}, \quad i = 1, \ldots, N$$

with $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$. Given $\Sigma = I_p$, a $p \times p$ identity matrix, we shall prove that $\sigma_1^2$ is close to zero, $\sigma_2^2, \sigma_3^2, \ldots, \sigma_J^2$ are close to one as illustrated by the intuitive argument in Section 4.3 and the simulation results presented in Table C.11.

Without loss of generality, let $\alpha$ be $(1, 0, \ldots, 0)^T$ in $\tilde{Y}_j(\alpha)$'s $(j = 1, \ldots, J)$. Then

$$\alpha^T Z_i = Z_{1i} = Y_{1i} - \bar{Y}_1$$

is a univariate random variable, where $\bar{Y}_1 = \frac{1}{N} \sum_{i=1}^{N} Y_{1i}$. In this section, Taylor series expansions around $Y_{1i} - \mu_1$ are used for $Z_{1i} = Y_{1i} - \bar{Y}_1$ at $Y_{1i} = \mu_1$ for $i = 1, \ldots, N$. Note that $Y_{1i} - \bar{Y}_1 = Y_{1i} - \mu_1 + \mu_1 - \bar{Y}_1, \bar{U} = \bar{D} = I_p$.

The first term

By (2.1),

$$\tilde{Y}_1(\alpha) = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} P_i(\alpha^T Z_i) = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} (2\Phi(Y_{1i} - \bar{Y}_1) - 1). \quad (5.3)$$

Since $(\mu_1 - \bar{Y}_1)^2 = Q_p(1/N)$ as $N \to \infty$, Taylor expansion of the above (5.3) is

$$\sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[ (2\Phi(Y_{1i} - \mu_1) - 1) + (\mu_1 - \bar{Y}_1) \cdot 2\phi(Y_{1i} - \mu_1) \right] + O_p\left( \frac{1}{\sqrt{N}} \right)$$

as $N \to \infty$, which is equivalent to

$$\tilde{Y}_{1u}(\alpha) + \sqrt{3N(\mu_1 - \bar{Y}_1)} \cdot \frac{1}{N} \sum_{i=1}^{N} 2\phi(Y_{1i} - \mu_1) + O_p\left( \frac{1}{\sqrt{N}} \right). \quad (5.4)$$

78
Since $\mathcal{E}\phi(Y_i - \mu_1) = 1/(2\sqrt{\pi})$, and when $N \to \infty$

$$
\frac{1}{N} \sum_{i=1}^{N} \phi(Y_{1i} - \mu_1) - \mathcal{E} (\phi(Y_{1i} - \mu_1)) = O_p (1/\sqrt{N}),
$$

$$
\sqrt{3N}(\mu_1 - \bar{Y}_1) = O_p (1),
$$

the above (5.4) is equal to

$$
\bar{Y}_{1u}(\alpha) + \sqrt{3N}(\mu_1 - \bar{Y}_1) \cdot \mathcal{E} (\phi(Y_{1i} - \mu_1)) + O_p \left( \frac{1}{\sqrt{N}} \right)
$$

$$
= \bar{Y}_{1u}(\alpha) - \sqrt{\frac{3N}{\pi}}(\bar{Y}_1 - \mu_1) + O_p \left( \frac{1}{\sqrt{N}} \right)
$$

as $N \to \infty$. Therefore,

$$
\bar{Y}_1(\alpha)
$$

$$
= \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[ (2\Phi(Y_{1i} - \mu_1) - 1) - \frac{Y_{1i} - \mu_1}{\sqrt{\pi}} \right] + O_p \left( \frac{1}{\sqrt{N}} \right)
$$

$$
= \bar{Y}_{1u}(\alpha) - R_{1c} + O_p \left( \frac{1}{\sqrt{N}} \right)
$$

$$
\overset{\text{dis}}{\sim} N(0, \sigma_1^2),
$$

(5.5)

as $N \to \infty$, where

$$
R_{1c} = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \frac{Y_{1i} - \mu_1}{\sqrt{\pi}},
$$

$$
\sigma_1^2 = 3 \cdot \mathcal{E} \left[ (2\Phi(Y_{1i} - \mu_1) - 1 - \frac{Y_{1i} - \mu_1}{\sqrt{\pi}})^2 \right]
$$

$$
= 3 \cdot \left( \frac{1}{3} - \frac{1}{\pi} \right) = 0.045070341.
$$

Here "\overset{\text{dis}}{\sim}" means "is asymptotically distributed from". The Central Limit Theorem is used for the independent, identically distributed random variables $2\Phi(Y_{11} - \mu_1) - 1 - \frac{Y_{11} - \mu_1}{\sqrt{\pi}}$, $2\Phi(Y_{12} - \mu_1) - 1 - \frac{Y_{12} - \mu_1}{\sqrt{\pi}}$ , ..., $2\Phi(Y_{1N} - \mu_1) - 1 - \frac{Y_{1N} - \mu_1}{\sqrt{\pi}}$ in (5.5).
The second term

Similarly as \( N \to \infty \)

\[
\tilde{Y}_2(\alpha) = \sqrt{\frac{5}{N}} \sum_{i=1}^{N} P_2(\alpha^T Z_i)
\]

\[
= \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} [(2\Phi(Y_{1i} - \bar{Y}_1) - 1)^2 - \frac{1}{3}]
\]

\[
= \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} [((2\Phi(Y_{1i} - \mu_1) - 1)^2 - \frac{1}{3})

+ (\mu_1 - \bar{Y}_1) \cdot 2(2\Phi(Y_{1i} - \mu_1) - 1)2\phi(Y_{1i} - \mu_1)] + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{5.6}
\]

\[
\tilde{Y}_{2u}(\alpha) + \frac{3}{2} \sqrt{5N} (\mu_1 - \bar{Y}_1) \cdot \frac{1}{N} \sum_{i=1}^{N} 4(2\Phi(Y_{1i} - \mu_1) - 1)\phi(Y_{1i} - \mu_1)

\]

\[
+ O_p\left(\frac{1}{\sqrt{N}}\right)
\]

\[
= Y_{2u}(\alpha) + \frac{3}{2} \sqrt{5N} (\mu_1 - \bar{Y}_1) \cdot 4\mathcal{E}[(2\Phi(Y_{1i} - \mu_1) - 1)\phi(Y_{1i} - \mu_1)] + O_p\left(\frac{1}{\sqrt{N}}\right)
\]

\[
= \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} [(2\Phi(Y_{1i} - \mu_1) - 1)^2 - \frac{1}{3}] + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{5.7}
\]

\[
= \tilde{Y}_{2u}(\alpha) + O_p\left(\frac{1}{\sqrt{N}}\right)
\]

\[
\overset{dist}{\sim} \mathcal{N}(0,1).
\]

Here (5.6) is by Taylor expansion and \((\bar{Y}_1 - \mu_1)^2 = O_p(1/N)\) as \( N \to \infty \). (5.7) is by

\[
\frac{1}{N} \sum_{i=1}^{N} [(2\Phi(Y_i) - 1)^2 - \frac{1}{3}] - \mathcal{E}[(2\Phi(Y_i - \mu_1) - 1)\phi(Y_i - \mu_1)] = O_p\left(\frac{1}{\sqrt{N}}\right),
\]

as \( N \to \infty \), and (5.8) is by

\[
\mathcal{E}[(2\Phi(Y_i - \mu_1) - 1)\phi(Y_i - \mu_1)] = 0.
\]
Therefore, $\sigma^2_x = 1$.

The third, fifth, and seventh terms

For $j = 3, 5, 7$, then as $N \to \infty$,

$$
\hat{Y}_j(\alpha) = \sqrt{\frac{2j + 1}{N}} \sum_{i=1}^{N} P_j(\alpha^T Z_i)
= \hat{Y}_{ju}(\alpha) - R_{jc} + O_p \left( \frac{1}{\sqrt{N}} \right) \overset{d}{\to} N(0, \sigma^2_j)
$$

with $R_{jc} = \sqrt{\frac{2j + 1}{N}} \sum_{i=1}^{N} c_j(Y_{1i} - \mu_1)$, where for $W = 2\Phi(Y) - 1, Y \overset{d}{\to} N(0, 1)$

$$
c_3 = \frac{3}{2} \mathcal{E} \left[ (5(2\Phi(Y) - 1)^2 - 1)2\phi(Y) \right] = 0.06917061,
\quad c_5 = \frac{15}{4} \mathcal{E} \left[ (21W^4 - 14W^2 + 1)\phi(Y) \right] = 0.024633,
\quad c_7 = \frac{1}{8} \mathcal{E} \left[ (3003W^6 - 3465W^4 + 945W^2 - 35)\phi(Y) \right] = 0.012324,
$$

and

$$
\text{Cov}(R_3) = 0.033492103, \quad \text{Cov}(Y_{ju}, R_3) = 0.033489596,
\quad \text{Cov}(R_3) = 0.00667463, \quad \text{Cov}(R_5, Y_{ju}) = 0.00667469,
\quad \text{Cov}(R_7) = 0.0022782, \quad \text{Cov}(R_7, Y_{ju}) = 0.0022782,
$$

which imply $\sigma^2_3 = 0.9665128$, $\sigma^2_5 = 0.993325244$ and $\sigma^2_7 = 0.9977218$.

The fourth, sixth, and eighth terms

For $j=4, 6, 8$, as $N \to \infty$,

$$
\hat{Y}_j(\alpha) = \sqrt{\frac{9}{N}} \sum_{i=1}^{N} P_4(\alpha^T Z_i)
= \hat{Y}_{ju}(\alpha) + O_p \left( \frac{1}{\sqrt{N}} \right) \overset{d}{\to} N(0, 1)
$$
i.e. $\sigma_4^2 = \sigma_6^2 = \sigma_8^2 = 1$.

We have examined all $\tilde{Y}_j(\alpha)$ for $j = 1, 2, \ldots, J, J = 8$, the maximum number of the terms in JHF index suggested in Chapter 3. We see that $\sigma_1^2$ is close to 0 and $\sigma_2^2, \ldots, \sigma_J^2$ are close to one as we expected.

## 5.2 Sphered Case

Now, we deal with the interesting case that the data are sphered, i.e.

$$Z_i = D^{-\frac{1}{2}} U^\tau(Y_i - \overline{Y}), \quad i = 1, \ldots, N.$$

Here data $Y_1, Y_2, \ldots, Y_N$ are i.i.d. from $\mathcal{N}(\mu, \Sigma)$, $\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$ is the sample mean, and $UDU^\tau = \hat{\Sigma}$ is an eigenvalue-eigenvector decomposition of the estimate

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \overline{Y})(Y_i - \overline{Y})^\tau$$

of the covariance matrix of $\mathcal{N}(\mu, \Sigma)$.

Set

$$D^{-\frac{1}{2}} U^\tau = (\tilde{\sigma}_{ij})_{p \times p}.$$

Without loss of generality, let $\Sigma = \sigma^2 I_p$, for a positive scalar $\sigma^2$ and a $p \times p$ identity matrix $I_p$. Then in the definition of $\tilde{Y}_{ju}(\alpha)$ in (5.1),

$$\tilde{D}^{-\frac{1}{2}} \tilde{U}^\tau(Y_i - \mu) = \frac{Y_i - \mu}{\sigma}.$$

Theorem 3.4.4 and its following paragraphs in Anderson (1984) give

$$(\tilde{\sigma}_{ij})_{p \times p} - \frac{1}{\sigma} I_p = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{as} N \to \infty. \quad (5.9)$$

82
Again, without loss generality, let \( \alpha \) be \((1, 0, \ldots, 0)^\tau \) in \( Y_j(\alpha) \) \((j = 1, \ldots, J)\). Then

\[
\alpha^\tau Z_i = Z_{i1} = \sum_{j=1}^{p} \hat{\sigma}_{1j}(Y_{ji} - \bar{Y}_j) = \frac{Y_{1i} - \bar{Y}_1}{\hat{\sigma}_{11}} + R_i
\]

(5.10)

is a univariate random variable with \( \hat{\sigma}_{11} = 1/\hat{\sigma}_{11} \), and

\[
R_i = \sum_{j=2}^{p} \hat{\sigma}_{1j}(Y_{ji} - \bar{Y}_j) .
\]

(5.11)

These \( Z_{1i} \)'s satisfy

\[
\sum_{i=1}^{N} Z_{1i} = 0 , \quad (5.12)
\]

\[
\frac{1}{N} \sum_{i=1}^{N} Z_{1i}^2 = 1 . \quad (5.13)
\]

We shall see that \( \alpha^\tau Z_i = Z_{1i} \) in Equation (5.10) can be treated as \( (Y_{1i} - \bar{Y}_1)/\hat{\sigma}_{11} \)
in the following derivations of the limiting processes (fields) for \( \tilde{Y}_j(\alpha) \), \( j = 1, \ldots, J \), where

\[
\hat{\sigma}_{11} = \frac{1}{N} \sum_{i=1}^{N} (Y_{1i} - \bar{Y}_1)^2.
\]

In fact, \( \max_{i=1, \ldots, p} |R_i| = O_p(1/\sqrt{N}) = o_p((Y_{1i} - \bar{Y}_1)/\hat{\sigma}_{11}) \) as \( N \to \infty \) by (5.9), and we have by (5.13) that

\[
\hat{\sigma}_{11}^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_{1i} - \bar{Y}_1 + \hat{\sigma}_{11} R_i)^2 \\
= \frac{1}{N} \sum_{i=1}^{N} (Y_{1i} - \bar{Y}_1)^2 + R = \hat{\sigma}_{11} + R \\
\]

(5.14)

with

\[
R = \frac{\hat{\sigma}_{11}^2}{N} \sum_{i=1}^{N} R_i^2 + \frac{2\hat{\sigma}_{11}}{N} \sum_{i=1}^{N} (Y_{1i} - \bar{Y}_1) \cdot R_i ,
\]

83
where $R_i$ is defined in (5.11) and hence $R = o_p(\hat{\sigma}_{11})$.

In this spherè case, a Taylor series expansion shall be used for

$$Z_{1i} = \sum_{j=1}^{p} \hat{\sigma}_{1j}(Y_{ji} - \bar{Y}_j) = \frac{Y_{1i} - \bar{Y}_1}{\hat{\sigma}_{11}} + R_i$$

at $(Y_{1i} - \mu_1)/\sigma$ for each summand, for $i = 1, \ldots, N$. Note

$$\left(\frac{Y_{1i} - \bar{Y}_1}{\hat{\sigma}_{11}} + R_i \right) - \left(\frac{Y_{1i} - \mu_1}{\sigma}\right) = \frac{(Y_{1i} - \mu_1)(\sigma - \hat{\sigma})}{\hat{\sigma}_{11}} - \frac{(Y_1 - \mu_1)}{\hat{\sigma}_{11}} + R_i,$$

and by $\max_{1 \leq i \leq p} |R_i| = O_p(\frac{1}{\sqrt{N}})$, we have that as $N \to \infty$

$$\left(\frac{Y_{1i} - \bar{Y}_1}{\hat{\sigma}_{11}} + R_i - \frac{Y_{1i} - \mu_1}{\sigma}\right)^2 = O_p\left(\frac{1}{N}\right).$$

The first term

By (2.1) and (5.10), as $N \to \infty$

$$\bar{Y}_1(\alpha) = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} P_1(\alpha^T Z_i) = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left(2\Phi\left(\frac{Y_{1i} - \bar{Y}_1}{\hat{\sigma}_{11}} + R_i\right) - 1\right),$$

which can be expanded into the following Taylor expansion by (5.15) and (5.16)

$$\sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[\left(2\Phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right) - 1\right)\right]$$

$$+ \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[\frac{(Y_{1i} - \mu_1)(\sigma - \hat{\sigma}_{11})}{\hat{\sigma}_{11}} 2\phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right)\right]$$

$$- \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[\frac{(Y_1 - \mu_1)}{\hat{\sigma}_{11}} 2\phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right)\right]$$

$$+ \sqrt{\frac{3}{N}} \sum_{i=1}^{N} [R_i 2\phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right) + O_p\left(\frac{1}{\sqrt{N}}\right)].$$

Since $\mathcal{E}(Y_{1i} - \mu_1)2\phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right) = 0$, and as $N \to \infty$

$$\frac{1}{N} \sum_{i=1}^{N} (Y_{1i} - \mu_1)2\phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right) - \mathcal{E}(Y_{1i} - \mu_1)2\phi\left(\frac{Y_{1i} - \mu_1}{\sigma}\right) = O_p\left(1/\sqrt{N}\right),$$

84
and \( \sqrt{N}(\sigma - \hat{\sigma}_{11}) = O_p(1) \) by (5.9), the second term in (5.17) is \( O_p(1/\sqrt{N}) \). By (5.11) and \( \mathcal{E}[(Y_{ji} - \mu_1)2\phi(Y_{ji} - \mu_1)] = 0, \) for \( j \neq 1 \), the fourth term in (5.17) is
\[
\sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[ R_i 2\phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) \right] = O_p(1/\sqrt{N}).
\]
Therefore,
\[
\bar{Y}_1(\alpha) = \bar{Y}_{1u}(\alpha) - \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \left[ \frac{(\bar{Y}_1 - \mu_1)}{\sigma} 2\phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) \right] + O_p(\frac{1}{\sqrt{N}})
\]
\[
= \bar{Y}_{1u}(\alpha) - R_{1s} + O_p(\frac{1}{\sqrt{N}}) \tag{5.18}
\]
\[
\sim N(0, \sigma_1^2) \tag{5.19}
\]
where in (5.18),
\[
R_{1s} = \sqrt{\frac{3}{N}} \sum_{i=1}^{N} \frac{Y_{ii} - \mu_1}{\sqrt{\pi}};
\]
in (5.19), \( \sigma_1^2 = 0.04507 \) and Central Limit Theorem is used.

The second term
As \( N \rightarrow \infty \)
\[
\bar{Y}_2(\alpha) = \sqrt{\frac{5}{N}} \sum_{i=1}^{N} P_2(\alpha^T Z_i) = \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \left[ (2\Phi(O_{1i} - \mu_1) + R_i - 1)^2 - \frac{1}{3} \right]. \tag{5.20}
\]
By (5.15), (5.16) and using Taylor expansion, (5.20) is equivalent to
\[
\frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \left[ (2\Phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) - 1)^2 - \frac{1}{3} \right]
\]
\[
+ \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \left[ (Y_{ii} - \mu_1)(\sigma - \hat{\sigma}) (2\Phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) - 1) 2\Phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) \right]
\]

85
\[
- \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \frac{(Y_i - \mu_1)}{\hat{\sigma}_1} \cdot 2(2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)2\Phi(\frac{Y_i - \mu_1}{\sigma}) \\
+ \frac{3}{2} \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \left[ R_i \cdot 2(2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)2\Phi(\frac{Y_i - \mu_1}{\sigma}) \right] + O_p \left( \frac{1}{\sqrt{N}} \right). \quad (5.21)
\]

Because of (5.15), (5.16), and \( \mathcal{E}[(2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)\phi(\frac{Y_i - \mu_1}{\sigma})] = 0 \), and as \( N \to \infty \)

\[
\frac{1}{N} \sum_{i=1}^{N} (2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)\phi(\frac{Y_i - \mu_1}{\sigma}) - \mathcal{E}[(2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)\phi(\frac{Y_i - \mu_1}{\sigma})] = O_p \left( \frac{1}{\sqrt{N}} \right),
\]

the third term in (5.21) is \( O_p \left( \frac{1}{\sqrt{N}} \right) \). As

\[
\mathcal{E}(Y_{ji} - \mu_j)\phi(\frac{Y_{ji} - \mu_1}{\sigma})(2\Phi(\frac{Y_{ji} - \mu_1}{\sigma}) - 1) = 0, \quad j \neq 1,
\]

applying (5.9), and (5.11), we have the fourth term in (5.21) is

\[
\sqrt{\frac{5}{N}} \sum_{i=1}^{N} \left[ R_i \cdot 2(2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)2\Phi(\frac{Y_i - \mu_1}{\sigma}) \right] = O_p \left( \frac{1}{\sqrt{N}} \right)
\]

as \( N \to \infty \). Therefore,

\[
\hat{Y}_2(\alpha)
\]

\[
= \hat{Y}_{2u}(\alpha) + \frac{3}{2} \sqrt{\frac{5}{N}} \left( \frac{\sigma - \hat{\sigma}_1}{\sigma \hat{\sigma}_1} \right) \frac{4}{\sqrt{N}} \sum_{i=1}^{N} \left( 2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1 \right)\phi(\frac{Y_i - \mu_1}{\sigma}) Y_i - \mu_1 \\
+ O_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
= \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \left[ \frac{1}{2}(3(2\Phi(\frac{Y_i - \mu_1}{\sigma}) - 1)^2 - 1) + \frac{3}{2} c_2 (1 - (\frac{Y_i - \mu_1}{\sigma})^2) \right] \\
+ O_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
= \hat{Y}_{2u}(\alpha) - R_{2s} + O_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
\hat{Y}_2(\alpha) \approx \mathcal{N}(0, \sigma_{\hat{Y}_2}^2)
\]

86
where
\[ R_{2s} = \sqrt{\frac{5}{N}} \sum_{i=1}^{N} \frac{3c_2}{2} \left( \frac{(Y_{ii} - \mu_1)}{\sigma} \right)^2 - 1 \]
with
\[ c_2 = 4E[\Phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) \frac{Y_{ii} - \mu_1}{\sigma} \phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right)] = \frac{1}{\sqrt{3\pi}} = 0.183776. \]
Hence
\[ \sigma_2^2 = \sqrt{5} E\left[\frac{1}{2} \left(3(2\Phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) - 1)^2 - 1\right) - \frac{3}{2}c_2 \left(\frac{(Y_{ii} - \mu_1)}{\sigma}\right)^2 - 1\right] = 0.2421419. \]
Here for (5.22), we used
\[
\frac{3}{2} \sqrt{5N} \frac{(\sigma - \hat{\sigma}_1)}{\hat{\sigma}} \frac{4}{N} \sum_{i=1}^{N} \left[2\Phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) - 1\right] \phi\left(\frac{Y_{ii} - \mu_1}{\sigma}\right) \frac{Y_{ii} - \mu_1}{\sigma} \right]
= \frac{3}{2} \sqrt{5N} \frac{\sigma^2 - \hat{\sigma}_1^2}{(\sigma + \hat{\sigma})\hat{\sigma}} \frac{2c_2 + O_p\left(\frac{1}{\sqrt{N}}\right)}{(\sigma + \hat{\sigma})\hat{\sigma}}
= \frac{3}{2} \sqrt{5N} (1 - \frac{1}{\sigma^2} \sum_{i=1}^{N} (Y_{ii} - \bar{Y}_1)^2 - \frac{R_i}{\sigma^2} c_2 + O_p\left(\frac{1}{\sqrt{N}}\right) \text{ (by (5.14), (5.9))}
= \frac{3}{2} \sqrt{5N} (1 - \frac{1}{\sigma^2} \sum_{i=1}^{N} (Y_{ii} - \mu_1)^2) c_2 + O_p\left(\frac{1}{\sqrt{N}}\right)
\]
as \( N \to \infty. \)

Again, from the above derivations for \( \hat{Y}_1(\alpha), \hat{Y}_2(\alpha) \), we see that we can just treat
\[ Z_{1i} = \frac{Y_{ii} - \bar{Y}_1}{\hat{\sigma}_{1i}} + R_i \text{ as } \frac{Y_{ii} - \bar{Y}_1}{\hat{\sigma}_{1i}} \text{ with} \]
\[ \hat{\sigma}_{1i} = \frac{1}{N} \sum_{i=1}^{N} (Y_{ii} - \bar{Y}_1)^2 \]
which really simplifies further derivations for \( \hat{Y}_j(\alpha), j = 3, \ldots, J. \)

The third, fifth, and seventh terms

87
For $j=3, 5, 7$, as $N \to \infty$

$$
\tilde{Y}_j(\alpha) = \sqrt{\frac{2j+1}{N}} \sum_{i=1}^{N} P_j(\alpha^T Z_i)
$$

$$
= Y_{ju}(\alpha) + R_{js} + O_p\left(\frac{1}{\sqrt{N}}\right) \overset{d}{\sim} \mathcal{N}(0, \sigma_j^2)
$$

with

$$
R_{js} = \sqrt{\frac{2j+1}{N}} \sum_{i=1}^{N} c_j \frac{Y_{1i} - \mu_1}{\sigma}
$$

and $c_j$ same as centered case (see page 81.). Therefore,

$$
\sigma_3^2 = 0.9665128; \\
\sigma_5^2 = 0.993325; \\
\sigma_7^2 = 0.9977218.
$$

The fourth, sixth, and eighth terms

For $j=4, 6, 8$, as $N \to \infty$

$$
\tilde{Y}_j(\alpha) = \sqrt{\frac{2j+1}{N}} \sum_{i=1}^{N} P_j(\alpha^T Z_i)
$$

$$
= Y_{ju}(\alpha) - R_{js} + O_p\left(\frac{1}{\sqrt{N}}\right) \overset{d}{\sim} \mathcal{N}(0, \sigma_j^2)
$$

with $R_{js} = \sqrt{\frac{2j+1}{N}} \sum_{i=1}^{N} c_j\left(\frac{Y_{1i} - \mu_1}{\sigma}\right)^2 - 1$, where $c_4 = 0.086855$, $c_6 = 0.042285$, $c_8 = 0.0249315$ and

$$
Cov(R_{4s}) = 0.135788238, \quad Cov(\tilde{Y}_4, R_{4s}) = 0.135783, \\
Cov(R_{6s}) = 0.046364395, \quad Cov(\tilde{Y}_6, R_{6s}) = 0.0463645, \\
Cov(R_{8s}) = 0.021133703, \quad Cov(\tilde{Y}_8, R_{8s}) = 0.0211344,
$$

88
which imply

\[
\begin{align*}
\sigma_4^2 &= 0.86422438, \\
\sigma_6^2 &= 0.953635396, \\
\sigma_8^2 &= 0.9788649.
\end{align*}
\]

Now we have showed that for spherred data each \( \tilde{Y}_j(\alpha) \) is asymptotically normal with mean zero, and variance \( \sigma_j^2 \) with \( \sigma_1^2, \sigma_2^2 \) close to zero, \( \sigma_3^2, \ldots, \sigma_j^2 \) close to one as in the following table and (5.2) hold. Our justification is done.

<table>
<thead>
<tr>
<th>( \sigma_j^2 )</th>
<th>Unsphered</th>
<th>Centered</th>
<th>Sphered</th>
</tr>
</thead>
<tbody>
<tr>
<td>j=1</td>
<td>1</td>
<td>0.04507</td>
<td>0.04507</td>
</tr>
<tr>
<td>j=2</td>
<td>1</td>
<td>1</td>
<td>0.2421419</td>
</tr>
<tr>
<td>j=3</td>
<td>1</td>
<td>0.9665128</td>
<td>0.9665128</td>
</tr>
<tr>
<td>j=4</td>
<td>1</td>
<td>1</td>
<td>0.8642244</td>
</tr>
<tr>
<td>j=5</td>
<td>1</td>
<td>0.9933252</td>
<td>0.9933252</td>
</tr>
<tr>
<td>j=6</td>
<td>1</td>
<td>1</td>
<td>0.9536364</td>
</tr>
<tr>
<td>j=7</td>
<td>1</td>
<td>0.9977218</td>
<td>0.9977218</td>
</tr>
<tr>
<td>j=8</td>
<td>1</td>
<td>1</td>
<td>0.9788649</td>
</tr>
</tbody>
</table>

Remark: \( \tilde{Y}_j(\alpha) \stackrel{d}{\sim} N(0, \sigma_j^2) \)

Table 5.1: Limiting Variances of \( \tilde{Y}_j(\alpha) \)
Chapter 6

Discussions and Remarks

In previous chapters, we have given a general two term approximation formula for the tail probability of the extreme of a class of differentiable Gaussian random fields, and its application to projection pursuit with JHF PP index. We have seen that sphering is important in applications, but introduces difficulties in making inferences about the data set. The higher the data dimension \( p \) is, the more complicated the data analysis and optimization procedure in PP are. That the metric tensor matrix \( R(t) \) corresponding to JHF PP index is diagonal played an important role in the derivation of clean formulas for two coefficients in the two term approximation formula.

Naturally, we would like to have similar results for some other indices. More specifically, we hope that tensor matrices are diagonal, if they exist, for other PP indices or for those indices which are competitive with JHF PP index.

Any Projection Pursuit index is essentially a measurement of the distance be-
tween the distribution of the data projected to a direction $\alpha$ and a normal distribution. The direction which maximizes this index reveals least normal (or most interesting) structure of the data (cf. Huber (1985)). The four most common measurements are based on the cumulants of the data, Shannon Entropy or Fisher information, Kolmogorov-Smirnov distance, or certain polynomial expansion of $L_2$ distance. For example, JHF PP index is from the Legendre polynomial expansion of the $L_2$ distance (see (1.3)).

- According to J. Friedman (1987), the PP indices based on cumulants of the data are inferior to JHF index proposed in his 1987 paper.

- The PP index based on Shannon Entropy or Fisher information does not have a corresponding practical PP algorithm so far, because of certain difficulties brought in by the way the index is defined. It is not applied in application at present.

- The PP index introduced by John Öhrvik (1988) based on Kolmogorov-Smirnov distance, for example, is much more expensive computationally than Friedman's polynomial based PP index, since to find the maximum of a Kolmogorov-Smirnov distance based PP index over all possible directions, we need to rank arguments for each computation of the index. Sorting can be very costly. Furthermore, it is hard to get a theoretical formula for the P-value associated to this Kolmogorov-Smirnov distance based PP index. There is only an empirical formula by John Öhrvik (1988), who used large simulations.

- The natural orthogonal polynomial basis for a $L_2$ distance of the distribution of the data and normal distribution is the set of Hermite Polynomial. This leads
to projection pursuit regression (PPR) index based on Hermite polynomial suggested by Johansen and Johnstone (1985), and PP index based on Hermite function introduced by Hall (1989). These two indices are closely related to the index introduced by Jones (1983), which was called “the index based on the high moments of the data” by Jones. If the data are first transformed in the way so that a standard normal random variable is converted into uniform random variable in \([-1, 1]\), the natural basis of orthogonal polynomial is Legendre Polynomial which leads to JHF PP index.

Hence, from practical point of view, we only wish to check if the metric tensor matrix $R(t)$ is also diagonal for the PPR index introduced by Johansen and Johnstone (1985) and the PP index introduced by P. Hall (1989), and if these two indices are competitive with JHF PP index. We shall call Hall’s PP index $PH$ index, and Johansen and Johnstone PPR index $JJ$ index.

JHF PP index aims at Exploratory Projection Pursuit, while JJ PPR index aims at Projection Pursuit Regression. They are not directly comparable. However, the metric tensor matrix for the index proposed by Johansen and Johnstone (1985) is diagonal, and the related Karhunen-Loève expansion of JJ index is finite, which implies that our theory presented in previous chapters works for this PPR index, viz. there is a $d$ term approximation for the P-value of JJ PPR index (cf. Theorem 2.1).

$PH$ index is for Exploratory Projection Pursuit, but its corresponding metric tensor matrix is not diagonal. Therefore, in the following we particularly compare PH PP index with JHF PP index in some interesting cases.
PH PP index is

\[ I_j^h(\alpha) = \sum_{j=1}^{J} \left( \frac{1}{N} \sum_{i=1}^{N} h_j(\alpha^\tau Z_i) \right)^2 \]

\[ + \left( \frac{1}{N} \sum_{i=1}^{N} h_0(\alpha^\tau Z_i) - \frac{1}{\sqrt{2\pi}^{1/4}} \right)^2 \]

\[ = \sum_{j=1}^{H} \frac{\sqrt{\pi}}{j! \cdot 2^{j-1}} \left( \frac{1}{N} \sum_{i=1}^{N} H_j(\alpha^\tau Z_i) \phi(\alpha^\tau Z_i) \right)^2 \]

\[ + 2\pi^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} H_0(\alpha^\tau Z_i) \phi(\alpha^\tau Z_i) - \frac{1}{\sqrt{2\pi}^{1/4}} \right)^2, \quad (6.1) \]

where \( h_j \) is the \( j \)th Hermite function based on \( j \)th Hermite polynomial \( H_j \):

\[ h_j(r) = \frac{\pi^{1/4}}{\sqrt{j! \cdot 2^{j-1}}} H_j(r) \phi(r), \quad (6.3) \]

with \( H_j \) defined as follows: for \( r \in (-\infty, +\infty) \),

\[ H_0(r) = 1 \]

\[ H_1(r) = 2r \]

\[ H_j'(r) = 2j H_{j-1}(r), \quad \int_{-\infty}^{\infty} H_j^2(r) \phi^2(r) \, dr = \frac{j! \cdot 2^{j-1}}{\sqrt{\pi}} \]

for \( j = 2, 3, \ldots. \) \quad (6.4)

Here \( \phi(r) = \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} \), and (6.4) & (6.3) yield

\[ h_j(r)' = rh_j(r) - \sqrt{2(j + 1)} h_{j+1}(r). \quad (6.5) \]

for \( j = 1, 2, 3, \ldots. \)

### 6.1 Comparisons for Hall and Friedman’s PP indices
The bimodal model is an interesting model in application. We shall compare the 4-term JHF index ($J = 4$) with the 4-term PH PP index under the following simple probability model. In PH index, there are five terms if we count the zero term. Here the comparison is to be of histograms:

\[ Y \xrightarrow{\text{sphere}} Z \xrightarrow{\alpha_{PH}} \alpha_{PH}^T Z \rightarrow \text{histogram}_{PH} \]

\[ \xrightarrow{\alpha_{JHF}} \alpha_{JHF}^T Z \rightarrow \text{histogram}_{JHF} \]  

(6.6)

viz. for each data $Y$ from a bimodal model, we sphere $Y$ to get the spherized version $Z$ of $Y$, then obtain two different histograms in two two different ways. In the first way (see the upper portion of (6.6)), based on the data $Z$, we run a PP algorithm with JHF index as the objective function to arrive at an interesting direction (or PP solution) $\alpha_{JHF}$, then project $Z$ onto $\alpha_{JHF}$ for having the univariate data $\alpha_{JHF}^T Z$. Finally use the histogram of $\alpha_{JHF}^T Z$ as an estimation of its density function. In the second way (see the lower portion of (6.6)), based on the same data $Z$ we run the same PP algorithm with PH index as the objective function to reach an interesting direction (or PP solution) $\alpha_{PH}$, then project $Z$ onto this $\alpha_{PH}$ for having the univariate data $\alpha_{PH}^T Z$. At last we use the histogram of $\alpha_{PH}^T Z$ as an estimation of its density function. We interpret the PP Index as performing well if the corresponding histogram is close to the original bimodal model.

Model

Suppose that the data are independent, identically distributed from the 4 dimensional population $\mathcal{G}_\mu$, which specifies that if $X = (X_1, X_2, X_3, X_4)^T \overset{\text{dis}}{\sim} \mathcal{G}_\mu$, then

- the first component $X_1 \overset{\text{dis}}{\sim}$ the mixture of two normals with the density function $g_\mu$ given in (6.7) below;
Figure 6.1: The Mixture of Two Normals $g_\mu$

- the second component $X_2 \overset{\text{dist}}{\sim} \mathcal{N}(0,1)$;
- the third component $X_3 \overset{\text{dist}}{\sim} \mathcal{N}(0,1)$;
- the fourth component $X_4 \overset{\text{dist}}{\sim} \mathcal{N}(0,1)$;

and $X_1, X_2, X_3, X_4$ are independent. Here $\overset{\text{dist}}{\sim}$ should be read as "is distributed from". The probability density function of $X_1$ is

$$g_\mu(x) = \frac{1}{2} \phi(x) + \frac{1}{2} \phi(x - \mu), \quad x \in (-\infty, \infty),$$

(6.7)

where $\phi(x)$ is the density function of the standard normal distribution $\mathcal{N}(0,1)$ and $\mu$ is a non zero location. The true interesting direction for this probability model is $\alpha^* = (1, 0, 0, 0)$.

Set-up of Experiments
In the reference of the comparison scheme in (6.6), the following is the concrete set-up of experiments for the above probability model.

1. Generate 4 independent, identically distributed pseudo random numbers \( N_1, N_2, N_3, N_4 \) from \( \mathcal{N}(0, 1) \) and an independent Uniform random number \( U \) from \( \mathcal{U}(0, 1) \), then form random vector \( X = (X_1, X_2, X_3, X_4)^\top \) as follows

\[
\begin{align*}
X_1 &= N_1 + \mu I_{(U > \frac{1}{2})}, \\
X_2 &= N_2, \\
X_3 &= N_3, \\
X_4 &= N_4.
\end{align*}
\]

This \( X \overset{\text{dis}}{\sim} \mathcal{G}_\mu \).

2. Repeat above procedure independently \( N \) times. Consequently, we have a sample of size \( N \) data from \( \mathcal{G}_\mu \). A typical \( N \) is 100.

3. Sphere the data as in (1.5).

4. Run two programs with the same algorithm suggested in Section 4.2, but two different objective functions. One objective function is JHF PP index, another one is PH PP index.

5. Project the original data onto the two directions from these two different programs. If 4-term JHF PP (PH PP) index is good, its direction from the above program should be close to \( \alpha^* = (1, 0, 0, 0) \), and hence there are two peaks from the histogram of the data projected to this direction, as Figure 6.1 shows on page 95.

6. When \( \mu \) is small, we may not be able to see the separation of two peaks from raw data even they are projected onto \( \alpha^* \). Therefore, for each set of random variables \( N_1, N_2, N_3, N_4, U \), we consider an increasing sequence of \( \mu \)'s. Here we chose \( \mu = 0.4, 0.8, \ldots, 7.6, 8.0 \).
Comparisons

Figure 6.2 on page 107 is about our first experiment. Those pictures are the histograms of the spherized data projected to PP solutions $\alpha_{JHF}$ and $\alpha_{PH}$ in the spherized space, for $\mu = 0.4, 0.8, \ldots, 8.0$. The only difference between the histogram of the spherized data projected to the PP solution in spherized space and the histogram of the original data projected to the corresponding PP solution in the original unspherized space is the scale, i.e. the separation is preserved. We see that with JHF PP index (see the left part of Figure 6.2), the separation starts from $\mu = 4.4$ (when $\mu = 4$, there is some evidence in seeing two peaks), while with PH index (see the right part of Figure 6.2), we do not see separation for any of these $\mu$'s: $0.4, 0.8, \ldots, 8.0$. Actually, if we check the directions $\alpha_{JHF}(\mu)$ found by using JHF index, we notice that $\alpha_{JHF}(\mu)$ are close to the line represented by the direction $\alpha^* = (1, 0, 0, 0)$ for $\mu = 4.0, 4.4, \ldots, 8.0$, while for all $\mu = 0.4, 0.8, \ldots, 8.0$, $\alpha_{PH}(\mu)$ are not. See Table 6.1 on page 105. Note that $(-1, 0, 0, 0)$ and $(1, 0, 0, 0)$ are on the same line.

Figure 6.3 on page 108 are about the second experiment, which is independent of the first one. We see the separation starting from $\mu = 3.6$ (there is some evidence in seeing the separation when $\mu = 3.2$) to 8.0 with JHF index, and the separation at $\mu = 4.8$, and 6.0 to 8.0 with PH index. Table 6.2 on page 106 reveals that the line represented by directions found by using JHF index are close to one represented by $\alpha^* = (1, 0, 0, 0)$ for $\mu = 3.2$ to 8.0, and those found by using PH index are close to $\alpha^* = (1, 0, 0, 0)$ for $\mu = 4.8$ and 6.0 to 8.0.

We have printed out 20 experiments of which the first five are included in this thesis. All these 20 experiments show the same pattern as described for the first two experiments. JHF PP index is superior to PH index for this bimodal model.
Explanations

By Cramér (p133, pp221-224, 1954), the statistics formed from the third Hermite polynomial gives the skewness of the data and the statistics formed from the fourth Hermite polynomial gives the kurtosis of the data. Recall the argument in Chapter 4 and 5: if the data are sphered, with high probability they are concentrated in a small neighborhood of zero. Therefore, similarly for the 4-term Hall’s index based on Hermite functions (i.e. Hermite polynomial multiplied by kernel \( \exp\left(\frac{-u^2}{2}\right) \)), the zero, first, and second terms do not change too much as they have rather small variances comparing to the third and fourth term, the third term is dominated by the skewness of the data, and the fourth term is dominated by the kurtosis of the data.

In our experiments, the 4-term PH index places too much weight on the skewness of the data and picks up views with skewed density rather than bimodal density in many cases. Histograms in the right portion of Figure 6.2 and 6.3 do reveal a lot of skewed pictures.

If more terms in PH PP index are used, one might get the same result as JHF PP index does for this model. However, it is not desirable in terms of computation efficiency, since the maximization of the higher order polynomial is much more expensive than the maximization of the lower order polynomial.

6.2 Remarks and Conclusions

In data analysis where PP should be used, it is advised that one first reduces the dimension of the search space for PP by Principal Components, then applies PP for
the "new" data which are the original data projected to lower dimensional space formed by the first several Principal Components, especially, if the data dimension \( p \) is large while sample size \( N \) is relatively small. The P-value of PP should be calculated for the new data set. For example, Friedman (1987) applied PP in the reduced 4-dimensional space formed by the first four principal components of the original data for the STATES DATA where the sample size \( N = 50 \), and the original data dimension \( p = 7 \).

In application, one may also like to apply two dimensional PP procedure to see the interesting views projected on a plane formed by two orthogonal directions. One dimensional PP is to provide with the interesting views of the data along the line formed by one direction which maximizes the PP index, two dimensional PP is to provide with the interesting views of the data along the plane formed by two orthogonal directions which maximizes the PP index. JHF 2-dimensional PP index

\[
I_2(\alpha, \beta) = \sum_{j=1}^{J} \frac{2j + 1}{4} \left( \frac{1}{N} P_j (2\Phi(\alpha^T Z_i) - 1) \right)^2 \\
+ \sum_{k=1}^{J} \frac{2k + 1}{4} \left( \frac{1}{N} P_k (2\Phi(\beta^T Z_i) - 1) \right)^2 \\
+ \sum_{j=1}^{J-1} \sum_{k=1}^{J-j} \frac{(2j + 1)(2k + 1)}{4} \left( \frac{1}{N} P_j (2\Phi(\alpha^T Z_i) - 1) P_k (2\Phi(\beta^T Z_i) - 1) \right)^2,
\]

for \( \alpha, \beta \in S^{p-1}, \alpha^T \beta = 0 \),

(6.8)

where \( P_j \) is the \( j \)th Legendre polynomial. We hope that our theory for one-dimensional JHF PP index works for two-dimensional JHF PP index in calculating the P-value of the 2-dimensional PP.
Let
\[ \hat{Y}_j(\alpha) = \sqrt{\frac{2j + 1}{N}} \sum_{i=1}^{N} P_j(2\Phi(\alpha^T Z_i) - 1), \quad j = 1, \ldots, J, \]
\[ \hat{Y}^k(\beta) = \sqrt{\frac{2k + 1}{N}} \sum_{i=1}^{N} P_k(2\Phi(\beta^T Z_i) - 1), \quad k = 1, \ldots, J, \]
\[ \hat{Y}_l^m(\alpha, \beta) = \sqrt{\frac{(2l + 1)(2m + 1)}{N}} \sum_{i=1}^{N} P_l(2\Phi(\alpha^T Z_i) - 1) P_m(2\Phi(\beta^T Z_i) - 1), \]
\[ l = 1, \ldots, J, \quad m = 1, \ldots, J - l. \]

It is easy to prove similarly as in one-dimensional JHF PP index case that as \( N \to \infty, \)
\[ \hat{Y}_j(\alpha) \xrightarrow{\mathcal{L}} \hat{X}_j(\alpha) \xrightarrow{\text{dist}} \mathcal{N}(0,1), \quad j = 1, \ldots, J, \]
\[ \hat{Y}^k(\beta) \xrightarrow{\mathcal{L}} \hat{X}^k(\beta) \xrightarrow{\text{dist}} \mathcal{N}(0,1), \quad k = 1, \ldots, J, \]
\[ \hat{Y}_l^m(\alpha, \beta) \xrightarrow{\mathcal{L}} \hat{X}_l^m(\alpha, \beta) \xrightarrow{\text{dist}} \mathcal{N}(0,1), \quad l = 1, \ldots, J, \quad m = 1, \ldots, J - l \]
and \( \{X_j, X^k, X_l^m\} \) themselves, and for all \( j = 1, \ldots, J, \quad k = 1, \ldots, J, \quad m = 1, \ldots, J - l, l = 1, \ldots, J, \) are mutually independent. Further, by (6.8)
\[ 4N \cdot I_J(\alpha, \beta) = \sum_{j=1}^{J} (\hat{Y}_j(\alpha))^2 + \sum_{k=1}^{J} (\hat{Y}^k(\beta))^2 + \sum_{l=1}^{J} \sum_{m=1}^{J-l} (\hat{Y}_l^m(\alpha, \beta))^2 \]
\[ \xrightarrow{\mathcal{L}} \chi^2_{m(J)}(\alpha, \beta), \quad \text{as } N \to \infty, \]
where \( \chi^2_{m(J)}(\alpha, \beta) \) is a Chi-square field (process) with degree of freedom
\[ m(J) = 2J + J^2 - \frac{J(J + 1)}{2}. \]

Similarly as one-dimensional projection pursuit, this limiting result implies
\[
P\left\{ \max_{\alpha, \beta \in \mathbb{S}^{p-1}, \alpha^\top \beta = 0} I_J(\alpha, \beta) \geq a \right\}
\]

100
\[ = \max_{\alpha \in \mathbb{S}^{p-1}, \beta \in \mathbb{S}^{p-1}, a^\gamma = 0} \frac{\sqrt{4N \cdot I_J(\alpha, \beta)}}{\sqrt{4aN}} \]
\[ \approx \max_{\alpha \in \mathbb{S}^{p-1}, \beta \in \mathbb{S}^{p-1}, a^\gamma = 0, \gamma \in S^{m(j)-1}} \gamma^T \hat{X}(\alpha, \beta) \geq b. \] (6.9)

Here \( b = \sqrt{4Na} \), \( \hat{X}(\alpha, \beta) = (\hat{X}_1(\alpha), \ldots, \hat{X}_J(\alpha), \hat{X}_1^1(\beta), \ldots, \hat{X}_J^J(\beta), \hat{X}_1^1(\alpha, \beta), \ldots, \hat{X}_1^{J-1}(\alpha, \beta), \hat{X}_2^1(\alpha, \beta), \ldots, \hat{X}_2^{J-2}(\alpha, \beta), \ldots, \hat{X}_J^1(\alpha, \beta)) \). Therefore, we can apply the general two term approximation formula in Theorem 2.2 to the P-value for the two dimensional JHF PP index. If there is some reparametrization such that the corresponding metric tensor matrix is diagonal, we should have easy applicable formulas for the two coefficients in this two term approximation formula as those in Chapter 3 for one-dimensional JHF PP index. Unfortunately, the calculation is very complicated and tedious. We can see that the tensor matrix is partially diagonal as represented in (6.11) under the following reparametrization,

\[ \alpha = Q_{p-1}Q_{p-2} \ldots Q_2Q_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta_1 \\ \sin \theta_1 \cos \theta_2 \\ \sin \theta_1 \sin \theta_2 \cos \theta_3 \\ \vdots \\ \sin \theta_1 \sin \theta_2 \ldots \sin \theta_{p-2} \cos \theta_{p-1} \\ \sin \theta_1 \sin \theta_2 \ldots \sin \theta_{p-2} \sin \theta_{p-1} \end{pmatrix}, \]

\[ \beta = Q_{p-1}Q_{p-2} \ldots Q_2Q_1 \begin{pmatrix} 0 \\ \cos \varphi_1 \\ \sin \varphi_1 \cos \varphi_2 \\ \vdots \\ \sin \varphi_1 \sin \varphi_2 \ldots \sin \varphi_{m(J)-2} \cos \varphi_{m(J)-1} \\ \sin \varphi_1 \sin \varphi_2 \ldots \sin \varphi_{m(J)-2} \sin \varphi_{m(J)-1} \end{pmatrix}, \]
\[
\gamma = \begin{pmatrix} 
\cos \psi_1 \\
\sin \psi_1 \cos \psi_2 \\
\vdots \\
\sin \psi_1 \sin \psi_2 \sin \psi_3 \ldots \sin \psi_{J-2} \cos \psi_{J-1} \\
\sin \psi_1 \sin \psi_2 \sin \psi_3 \ldots \sin \psi_{J-2} \sin \psi_{J-1} 
\end{pmatrix} 
\] 

\hspace{1cm} (6.10)

for

\[
\theta_1, \ldots, \theta_{p-1} \in (0, \pi] \times (0, \pi] \times \ldots \times (0, \pi], \\
\varphi_1, \ldots, \varphi_{p-2} \in (0, \pi] \times (0, \pi] \times \ldots \times (0, \pi], \\
\psi_1, \ldots, \psi_{m(J-1)} \in (0, \pi] \times (0, \pi] \times \ldots \times (0, \pi] \times (0, 2\pi].
\]

Here

\[
Q_1 = \begin{pmatrix} 
\cos \theta_1 & -\sin \theta_1 & 0 & \ldots & 0 & 0 \\
\sin \theta_1 & \cos \theta_1 & 0 & \ldots & 0 & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 1 
\end{pmatrix},
\]

\[
Q_i = \begin{pmatrix} 
1 \\
\vdots \\
1 \\
\cos \theta_i & -\sin \theta_i \\
\sin \theta_i & \cos \theta_i \\
1 & \vdots \\
\vdots \\
1 
\end{pmatrix},
\]

102
\[ Q_{p-1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \cos \theta_{p-1} & -\sin \theta_{p-1} \\ \sin \theta_{p-1} & \cos \theta_{p-1} \end{pmatrix}, \]

are the usual rotation matrices. The metric tensor matrix corresponding to the reparametrization in (6.10) is

\[
R(t) = \begin{pmatrix} R_{11}(t) & R_{12}(t) & R_{13}(t) \\ R_{21}(t) & R_{22}(t) & R_{23}(t) \\ R_{31}(t) & R_{32}(t) & R_{33}(t) \end{pmatrix}_{d \times d}, \tag{6.11}
\]

where

\[
R_{11}(t) = \left. \frac{\partial^2 r(s,t)}{\partial \theta_i \partial \theta_j^t} \right|_{s=t} \quad (p-1) \times (p-1),
\]

\[
R_{22}(t) = \left. \frac{\partial^2 r(s,t)}{\partial \phi_i \partial \phi_j^t} \right|_{s=t} = a \text{ diagonal matrix},
\]

\[
R_{33}(t) = \left. \frac{\partial^2 r(s,t)}{\partial \psi_i \partial \psi_j^t} \right|_{s=t} = a \text{ diagonal matrix},
\]

\[
R_{12}(t) = \left. \frac{\partial^2 r(s,t)}{\partial \theta_i \partial \phi_j^t} \right|_{s=t} = R_{21}(t)^{\top},
\]

\[
R_{13}(t) = \left. \frac{\partial^2 r(s,t)}{\partial \theta_i \partial \psi_j^t} \right|_{s=t} = R_{31}(t)^{\top} = 0,
\]

\[
R_{23}(t) = \left. \frac{\partial^2 r(s,t)}{\partial \phi_i \partial \psi_j^t} \right|_{s=t} = R_{32}(t)^{\top} = 0,
\]

with \( d = 2p - 3 + m(J) - 1 \), \( s = (\theta_1, \ldots, \theta_{p-1}, \phi_1, \ldots, \phi_{p-2}, \psi_1, \ldots, \psi_{m(J)-1}) \in I \), \( t = (\theta_1', \ldots, \theta_{p-1}', \phi_1', \ldots, \phi_{p-2}', \psi_1', \ldots, \psi_{m(J)-1}') \in I \), for \( I = (0, \pi) \times (0, \pi) \times \ldots \times \)
$(0, \pi) \times (0, 2\pi]. \text{viz.} \quad R(t) = \begin{pmatrix} R_{11}(t) & R_{12}(t) & 0 \\ R_{21}(t) & \ddots & 0 \\ 0 & 0 & \ddots \end{pmatrix}.

We hope that someone would carry on the calculation using the above parametrization or a new parametrization to show there is a corresponding diagonal metric tensor matrix.

Now we conclude that we have obtained an applicable two terms approximation formula for the P-value of PP, which makes inference about PP possible. We have made the gap between theory and application in using PP (a concern by R. Miller) much smaller.
<table>
<thead>
<tr>
<th>( \mu )</th>
<th>( \alpha_{JHF}(\mu) )</th>
<th>( \alpha_{PH}(\mu) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>(-0.083338, -0.85669, -0.40426, 0.30937)</td>
<td>(-0.16909, 0.78828, 0.50243, -0.31240)</td>
</tr>
<tr>
<td>0.8</td>
<td>(-0.10376, -0.83055, 0.52094, -0.16749)</td>
<td>(-0.15289, 0.78606, -0.58900, 0.10197)</td>
</tr>
<tr>
<td>1.2</td>
<td>(-0.10854, 0.73815, 0.66629, 0.04047)</td>
<td>(-0.11867, 0.71830, -0.67334, 0.12873)</td>
</tr>
<tr>
<td>1.6</td>
<td>(-0.10959, 0.58860, 0.79421, 0.10374)</td>
<td>(0.86626, 0.09006, -0.20291, 0.44755)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.18629, 0.51479, 0.81762, 0.17830)</td>
<td>(0.057124, 0.47739, 0.84848, 0.22116)</td>
</tr>
<tr>
<td>2.4</td>
<td>(0.15097, 0.41307, 0.88585, 0.14781)</td>
<td>(0.042300, 0.41027, 0.88120, 0.23104)</td>
</tr>
<tr>
<td>2.8</td>
<td>(0.17277, 0.38270, 0.89369, 0.15812)</td>
<td>(0.032780, 0.36910, 0.89836, 0.23589)</td>
</tr>
<tr>
<td>3.2</td>
<td>(0.19188, -0.36650, 0.89524, -0.16553)</td>
<td>(0.025925, -0.34154, 0.90874, -0.23848)</td>
</tr>
<tr>
<td>3.6</td>
<td>(-0.20821, 0.35711, -0.89435, 0.17105)</td>
<td>(0.020625, -0.32160, 0.91575, -0.23092)</td>
</tr>
<tr>
<td>4.0</td>
<td>(0.99628, 0.084509, 0.014131, -0.0085792)</td>
<td>(0.016368, -0.30634, 0.92083, -0.24073)</td>
</tr>
<tr>
<td>4.4</td>
<td>(0.99806, 0.060958, 0.0059161, 0.011181)</td>
<td>(0.65584, -0.56789, -0.12869, -0.48043)</td>
</tr>
<tr>
<td>4.8</td>
<td>(0.99874, -0.049130, -0.0049060, 0.0093570)</td>
<td>(0.64636, 0.58282, 0.11220, -0.47954)</td>
</tr>
<tr>
<td>5.2</td>
<td>(0.99908, -0.014659, -0.0061433, 0.0075456)</td>
<td>(0.63820, 0.59531, 0.007024, -0.47843)</td>
</tr>
<tr>
<td>5.6</td>
<td>(0.99928, -0.036557, -0.0081777, 0.0058907)</td>
<td>(-0.63101, -0.60597, -0.083045, 0.47721)</td>
</tr>
<tr>
<td>6.0</td>
<td>(0.99939, -0.033038, -0.010222, 0.0044798)</td>
<td>(0.0036457, 0.26268, -0.93417, -0.24149)</td>
</tr>
<tr>
<td>6.4</td>
<td>(0.99947, -0.030216, -0.011936, 0.0033291)</td>
<td>(0.61870, 0.62333, 0.058184, -0.47464)</td>
</tr>
<tr>
<td>6.8</td>
<td>(0.99952, -0.027891, -0.013236, 0.0024123)</td>
<td>(0.61333, 0.63051, 0.047100, -0.47334)</td>
</tr>
<tr>
<td>7.2</td>
<td>(0.99956, -0.025906, -0.014151, 0.0016884)</td>
<td>(-0.00039181, 0.24846, -0.93813, -0.24121)</td>
</tr>
<tr>
<td>7.6</td>
<td>(-0.99960, 0.021477, 0.011751, -0.0011175)</td>
<td>(-0.0014157, 0.24473, -0.93014, -0.24109)</td>
</tr>
<tr>
<td>8.0</td>
<td>(0.99963, -0.022652, -0.015105, 0.00066615)</td>
<td>(-0.0023232, 0.24138, -0.94003, -0.24097)</td>
</tr>
</tbody>
</table>

Table 6.1: The First Experiment
<table>
<thead>
<tr>
<th>( \mu )</th>
<th>( \alpha_{PH}(\mu) )</th>
<th>( \alpha_{PH}(\mu) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>(0.71285, 0.36443, -0.45505, -0.38984)</td>
<td>(0.084005, -0.20921, -0.56222, -0.79567)</td>
</tr>
<tr>
<td>0.8</td>
<td>(-0.25511, -0.006226, 0.69983, -0.66717)</td>
<td>(0.19501, -0.24838, -0.64503, 0.69585)</td>
</tr>
<tr>
<td>1.2</td>
<td>(-0.47785, -0.037650, 0.61193, 0.62912)</td>
<td>(0.43953, 0.091376, 0.66845, -0.59290)</td>
</tr>
<tr>
<td>1.6</td>
<td>(-0.51711, -0.001513, 0.64914, 0.55785)</td>
<td>(0.52236, 0.096520, -0.67104, -0.51724)</td>
</tr>
<tr>
<td>2.0</td>
<td>(-0.50760, 0.035192, 0.60462, -0.52543)</td>
<td>(0.53695, 0.17745, -0.67581, -0.47273)</td>
</tr>
<tr>
<td>2.4</td>
<td>(0.23240, -0.74695, 0.27143, -0.56070)</td>
<td>(0.12243, -0.21050, -0.14695, -0.55870)</td>
</tr>
<tr>
<td>2.8</td>
<td>(-0.30059, -0.74266, -0.21769, -0.55247)</td>
<td>(0.18994, 0.22677, -0.14095, 0.94479)</td>
</tr>
<tr>
<td>3.2</td>
<td>(0.98420, -0.016513, 0.053772, -0.16786)</td>
<td>(0.24278, -0.24563, -0.13253, -0.29067)</td>
</tr>
<tr>
<td>3.6</td>
<td>(0.98719, 0.0074661, 0.062255, -0.14674)</td>
<td>(-0.85279, -0.37562, 0.35878, 0.054219)</td>
</tr>
<tr>
<td>4.0</td>
<td>(-0.99076, 0.034163, -0.012933, 0.12402)</td>
<td>(0.41525, 0.40581, 0.67988, -0.47941)</td>
</tr>
<tr>
<td>4.4</td>
<td>(0.99263, -0.034992, 0.042527, -0.10798)</td>
<td>(0.31723, -0.30182, -0.10023, -0.78221)</td>
</tr>
<tr>
<td>4.8</td>
<td>(-0.99407, 0.034186, -0.041081, 0.091286)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
<tr>
<td>5.2</td>
<td>(0.99519, -0.032652, 0.011025, -0.082791)</td>
<td>(0.39119, -0.33279, -0.079071, -0.85438)</td>
</tr>
<tr>
<td>5.6</td>
<td>(0.99604, -0.030866, 0.030724, -0.073298)</td>
<td>(0.40862, -0.34610, -0.069336, -0.84068)</td>
</tr>
<tr>
<td>6.0</td>
<td>(0.99670, -0.029068, 0.038212, -0.065528)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
<tr>
<td>6.4</td>
<td>(0.99720, -0.027368, 0.036605, -0.059177)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
<tr>
<td>6.8</td>
<td>(0.99760, -0.025804, 0.034987, -0.053960)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
<tr>
<td>7.2</td>
<td>(0.99791, -0.024385, 0.033408, -0.049639)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
<tr>
<td>7.6</td>
<td>(0.99816, -0.023104, 0.031899, -0.046021)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
<tr>
<td>8.0</td>
<td>(0.99837, -0.021948, 0.030472, -0.042960)</td>
<td>(1.0000, 0, 0, 0)</td>
</tr>
</tbody>
</table>

Table 6.2: The Second Experiment
Figure 6.2: The First Experiment.
Figure 6.5: The Fourth Experiment.
Figure 6.6: The Fifth Experiment.
Appendix A

Supplementary Theorems and Proofs for Chapter 2

In this appendix, Garsia's sufficient condition for a uniformly convergent Karhunen-Loève expansion of a random field is presented in Lemma A.1. The proofs of Theorem 2.1, 2.2, and 2.3 start on page 119, 147, and 140 respectively. In Proposition A.1 and A.2, we conclude the two properties (cf. the discussion prior to Remark 2.3) of the related manifold of a random field under R.1 (for m = 6) and R.2, the conditions of Theorem 2.2. These two properties and the generalized one term approximation (Lemma A.3) from the approximation in Adler (1981, p160) are useful to prove Theorem 2.2 and 2.3.

Let \( r(s, t) \) be the covariance function \( Cov(Z(s), Z(t)) \) of a random field \( Z(t) \) on
$I$, a $d$ dimensional space. Define

$$\Delta r(s, t) = r(s, s) + r(t, t) - 2r(s, t) \quad (A.1)$$

and

$$p(u) = \max_{|s-t| \leq u \sqrt{d}} \sqrt[4]{\Delta r(s, t)}. \quad (A.2)$$

The following Lemma A.1 gives a sufficient condition for a $d$ dimensional random field $Z(t)$ to have a uniformly convergent Karhunen-Loève expansion.

**Lemma A.1 (Garsia)** Suppose $Z(t)$ is a $d$ dimensional non singular random field with mean zero and covariance function $r(s, t)$.

If

$$\int_0^1 \sqrt{-\log u} \ dp(u) < \infty, \quad (A.3)$$

then there exists a partial sum formed from the first $k$ eigenvalue-eigenfunctions of $r(s, t)$

$$\tilde{Z}_k(t) = \sum_{i=1}^k \sqrt{\lambda_i} \cdot \Lambda_i(t) \cdot X_i$$

which converges to $Z(t)$ uniformly in $t$ on $I$ as $k \to \infty$, with probability one. Here $X_1, X_2, \ldots$ are independent, identically distributed random variables from $N(0, 1)$, $\lambda_i$ is the $i$th eigenvalue of $r(s, t)$, $\Lambda_i(t)$ is the corresponding eigenfunction of $\lambda_i$ (cf. (2.18)), and $p(u)$ is as defined in (A.2).

$\square$

Existence of an orthogonal eigenvalue-eigenfunction expansion (Mercer expansion) of the covariance function $r(s, t)$ of a random field $Z(t)$ is given by Mercer (cf. pp138-140, Courant and Hilbert, 1953). The corresponding series $\sum_{i=1}^\infty u_i(t)X_i$ as in
Lemma A.1 is called the uniformly convergent Standard Karhunen-Loève expansion of $Z(t)$ (cf. Remark 2.2), where $u_l(t) = \sqrt{\lambda_l(t)}$ for $l = 1, 2, \ldots$

**Remark A.1:** If $\Delta r(s, t)$ satisfies Lipschitz-$\alpha$ condition:

$$ \Delta r(s, t) \leq c\|s - t\|^\alpha, \quad \text{for all } s, t \in I, \quad (A.4) $$

for some positive constant $c$ and $\alpha$, then $(A.3)$ in Lemma A.1 is valid.

**Remark A.2:** Suppose $Z(t)$ is a $d$ dimensional homogeneous random field and the covariance function $r(s, t) = r(s - t)$ of $Z(t)$ has two continuous derivatives with respect to $w = s - t$ on its bounded domain. Then $r(s, t)$ satisfies Lipschitz-2 condition, i.e. $(A.4)$ holds with $\alpha = 2$. Hence $Z(t)$ has a uniformly convergent standard Karhunen-Loève expansion on its bounded domain $I$.

**Lemma A.2** Suppose $Z(t)$ is a $d$ dimensional non-singular random field with mean zero and covariance function $r(s, t)$, and

$$ \int_0^1 \sqrt{-\log u} \, dp(u) < \infty. $$

If there is an expansion of $r(s, t)$ such that

$$ r(s, t) = \sum_{i=1}^{\infty} \lambda_i \Lambda_i(s) \Lambda_i(t) \quad (A.5) $$

and $\sum_{i=1}^{k} |\lambda_i \Lambda_i(t) \Lambda_i(s)|$ is dominated by $\sum_{i=1}^{k} c|\lambda_i|$ for some $c > 0$, which converges as $k \to \infty$, then the random field $\tilde{Z}(t)$:

$$ \tilde{Z}(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i \Lambda_i(t)} X_i, \quad (A.6) $$

is identically distributed as $Z(t)$ and has a uniformly convergent Karhunen-Loève expansion on its bounded domain $I$, where $X_1, X_2, \ldots$ are independent, identically distributed $N(0,1)$ random variables.
Proof: Just follow the proof in Adler (1981) from page 52 to page 57. Notice that our $\Lambda_i$ is $\theta_i$ in Adler’s proof. A few misprints in Adler’s proof should be corrected and there is a small change to match our assumption on expansion in (A.5), where $\Lambda_i$’s do not have to be orthogonal system.

The small change is to replace line 12 from the bottom of page 56 by

$$E \left\{ \sum_{k=1}^{\infty} \lambda_k \theta_k^2(\omega) \right\} \leq c \sum_{k=1}^{\infty} |\lambda_k| < \infty.$$  

Our corrections for misprints are:

- line 4 from the bottom of page 52 should be

$$E\{X_{m+1}|X_m,X_{m-1},\ldots,X_1\}\geq X_m \text{ with probability one}.$$  

- line 10 from the bottom of page 55 should be

$$\geq P_m \exp(E\{R_{m+1}|\theta_m,\ldots,\theta_1\})$$  

- line 7 from the bottom of page 55 should be

$$E\{P_{m+1}|P_m,\ldots,P_1\}\geq P_m.$$  

\[\square\]

Remark A.3: Lemma A.2 says that under certain conditions for the covariance function of a random field $Z(t)$, there exists a random field $\tilde{Z}(t)$ which is identically distributed as $Z(t)$ and has a uniformly convergent Karhunen-Loève expansion on its bounded domain $I$. Lemma A.1 says that if the expansion of $r(s,t)$ is orthogonal, there is a Karhunen-Loève expansion which converges to $Z(t)$ itself.

115
on its bounded domain $I$, viz. $\hat{Z}(t) \equiv Z(t)$. These two lemmas on existence of a Karhunen-Loève expansion, play the same role in finding an approximation formula for $P\{\max_{t \in I} Z(t) \geq z\}$ as $z \to \infty$, since

$$P\{\max_{t \in I} Z(t) \geq z\} = P\{\max_{t \in I} \hat{Z}(t) \geq z\}$$

and $\hat{Z}(t)$ has a uniformly convergent Karhunen-Loève expansion to itself as $Z(t)$ in Lemma A.1 does.

**Proof of Proposition 2.1:** From (2.1), (2.2) and (2.4)-(2.6), the random field $Z(t)$, derived from JHF index, has mean zero, unit variance and covariance function $r(s, t) = \beta^r A \beta'$, where $A = \mathcal{E} [\bar{X}(\alpha) \cdot \bar{X}^*(\alpha')] = (a_{ij}(\alpha^r \alpha'))_{J \times J}$ is a $J \times J$ matrix with the element

$$a_{ij}(\alpha^r \alpha') = \frac{\sqrt{(2i + 1)(2j + 1)}}{\sqrt{(2i + 1)(2j + 1)}} \mathcal{E} \left\{ P_i(2\Phi(\alpha^r X) - 1) \cdot P_j(2\Phi(\alpha'^r X) - 1) \right\}$$

$$= \begin{cases} \sqrt{(2i + 1)(2j + 1)} \int_{-\infty}^{\infty} P_i(2\Phi(x) - 1) \cdot P_j(2\Phi(y) - 1) f(x, y, \alpha^r \alpha') \, dx \, dy & \text{if } \alpha \neq \alpha' \\ \delta_{ij} & \text{if } \alpha = \alpha' \end{cases}$$

(2.7)

(2.8)

**viz.** (2.13) holds and $r(t, t) = 1$. Here $\delta_{kl}$ is the Kronecker $\delta$, $P_i$ is the $i$th Legendre polynomial, $X$ is a standard $p$ dimensional normal random variable, and $f(\cdot, \cdot, \rho)$ is a bivariate normal density function with mean zero, unit variance and correlation.
coefficient \( \rho \) not equal to 1:

\[
f(x,y,\rho) = \frac{1}{2\pi \sqrt{(1 - \rho^2)}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} (x^2 + y^2 - 2xy\rho) \right\}
\]

\[
= \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{x^2}{2} \right\} \cdot \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \exp\left\{ -\frac{(y - \rho x)^2}{2(1 - \rho^2)} \right\}.
\]

For any continuous function \( \tilde{f}(y) \) which is \( o(\exp\{y^2\}) \) as \( y \to \infty \),

\[
\int_{-\infty}^{\infty} \tilde{f}(y) \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \exp\left\{ -\frac{(y - \rho x)^2}{2(1 - \rho^2)} \right\} dy = \tilde{f}(\rho x)(1 + o(1)) \quad (A.9)
\]

as \( \rho \to 1 \) by the Laplace approximation (cf. (4.2.4) on p65 in Bruijn, 1961). From (A.7) and (A.9), it is not hard to prove that \( r(s, t) \) is \( C^\infty \) for

\[
(s, t) = (s_1, \ldots, s_d, t_1, \ldots, t_d) = (\theta, \varphi, \theta', \varphi')
\]

\[
= (\theta_1, \theta_2, \ldots, \theta_{p-1}, \varphi_1, \varphi_2, \ldots, \varphi_{j-1}, \theta'_1, \theta'_2, \ldots, \theta'_{p-1}, \varphi'_1, \varphi'_2, \ldots, \varphi'_{j-1})
\]
on its bounded domain

\[
I = [0, \pi] \times \ldots \times [0, \pi] \times [0, 2\pi] \times [0, \pi] \times \ldots \times [0, \pi] \times [0, 2\pi].
\]

Again, from (A.8) it is easy to see

\[
a_{ij}(\alpha' \alpha') = a_{ij}(\rho) = \sqrt{(2i + 1)(2j + 1)} \cdot \int_{-\infty}^{\infty} P_i(2\Phi(x) - 1) \cdot P_j(2\Phi(u\sqrt{1 - \rho^2} + \rho x) - 1) \frac{1}{2\pi} \exp\left\{ -\frac{x^2 + u^2}{2} \right\} dx du,
\]

under the transformation \( u = (y - x\rho)/\sqrt{1 - \rho^2} \). Using Taylor series expansion

\[
P_j(2\Phi(u\sqrt{1 - \rho^2} + \rho x) - 1) - P_j(2\Phi(x) - 1)
\]

\[
= P_j'(2\Phi(x) - 1)2\phi(x) \cdot (u\sqrt{1 - \rho^2} + \rho x - x)
\]

\[
+ (P_j'(2\Phi(x) - 1)2\phi(x))'_{|x=\xi} \cdot \frac{(u\sqrt{1 - \rho^2} + \rho x - x)^2}{2}
\]

117
with \( \rho x \leq \xi \leq \rho x + u\sqrt{1 - \rho^2} \), we have

\[
\lim_{\rho \to 1} \frac{a_{ij}(\rho) - a_{ij}(1)}{1 - \rho} = \lim_{\rho \to 1} \frac{1}{\sqrt{(2i + 1)(2j + 1)}} \cdot \left( \int_{-\infty}^{\infty} P_i(2\Phi(x) - 1) \cdot P'_j(2\Phi(x) - 1) \cdot 2\phi(x) \cdot (u\sqrt{1 + \rho \over 1 - \rho} - x) g(x, u) dx du \right. \\
+ \left. \int_{-\infty}^{\infty} P_i(2\Phi(x) - 1) \cdot (P'_j(2\Phi(x) - 1) \cdot 2\phi(x))' \right|_{x=\xi} \\
\cdot \frac{(u\sqrt{1 - \rho^2} - x(1 - \rho))^2}{2(1 - \rho)} g(x, u) dx du \right)
\]

\[
= \int_{-\infty}^{\infty} P_i(2\Phi(x) - 1) \cdot (P'_j(2\Phi(x) - 1) \cdot 2\phi(x))'u^2 g(x, u) dx du < \infty,
\]

where \( g(x, u) = \exp\{ -\frac{x^2 + u^2}{2} \}/2\pi \), viz. \( a_{ij}(\alpha^T \alpha') - a_{ij}(1) = O(1 - \rho) \), which is equivalent to \( O(\|\alpha - \alpha'\|^2) \) as \( \rho = \alpha^T \alpha' = 1 - \frac{\|\alpha - \alpha'\|^2}{2} \).

Therefore, by Proposition 2.1, we have as \( \|\theta - \theta'\|^2 + \|\varphi - \varphi'\|^2 \to 0 \),

\[
r(s, t) = \beta^T I \beta' + \beta^T (A - I) \beta' \\
= \beta^T \beta' + \beta^T (A - I) \beta + \beta^T (A - I) (\beta' - \beta) \\
= 1 - \frac{\|\beta - \beta'\|^2}{2} + \beta^T \left( -\frac{\|\alpha - \alpha'\|^2}{2} a_{ij}'(1) \right)_{JxJ} \beta + o_p(\|\alpha - \alpha'\|^2 + \|\beta - \beta'\|^2) \\
= 1 - \frac{\|\beta - \beta'\|^2}{2} - \frac{\|\alpha - \alpha'\|^2}{2} \beta^T \left( a_{ij}'(1) \right)_{JxJ} \beta + o_p(\|\alpha - \alpha'\|^2 + \|\beta - \beta'\|^2) \\
= 1 - a_1(\varphi, \varphi')\|\varphi - \varphi'\|^2 - a_2(\theta, \theta', \varphi')\|\theta - \theta'\|^2 + o_p(\|\theta - \theta'\|^2 + \|\varphi - \varphi'\|^2),
\]

where \( a_1(\varphi, \varphi') \), \( a_2(\theta, \theta', \varphi) \) are bounded on \( I \), \( P'_j \) is the first derivative of the \( j \)th Legendre polynomial and \( a_{ij}'(1) = -\lim_{\rho \to 1} (a_{ij}(\rho) - a_{ij}(1))/(1 - \rho) < \infty \).
This proves (2.14).

(A.10) implies that (A.4) is satisfied. The random field $Z(t)$ derived from JHF PP index has a uniformly convergent standard Karhunen-Loève expansion corresponding to Mercer expansion of $r(s,t)$. We shall also prove that there is an expansion for $r(s,t)$ whose corresponding manifold satisfies two properties listed in Remark 2.4, so Theorem 2.2 can be applied to the random field derived from JHF PP index.

**Proof of Theorem 2.1:**

(The idea of the proof is given in Section 2.4 following the statement of Theorem 2.2.)

Define $\hat{Z}(t) = \sum_{i=1}^{k} u_i(t) X_i$, where $X_1, X_2, \ldots, X_k$ are independent, identically distributed $N(0,1)$ random variables. Then $\hat{Z}(t)$ and $Z(t)$ are identically distributed, which implies

$$P \{ \max_{t \in I} Z(t) \geq z \} = P \{ \max_{t \in I} \hat{Z}(t) \geq z \}.$$

By Lemma 2.3 and Definition 2.2, the critical radius $r_{kc}$ and hence the critical point $d_k = r_{kc}^2/(2 - r_{kc}^2)$, of the manifold $U^k = \{(u_1(t), \ldots, u_k(t)) : t \in I\}$ of $\hat{Z}(t)$, is positive.

Since

$$U \equiv \left( \frac{X_1}{\|X\|}, \frac{X_2}{\|X\|}, \ldots, \frac{X_k}{\|X\|} \right)$$

119
is uniformly distributed on unit sphere $S^{k-1}$ and is independent of $\| X \|$, we can rewrite the tail probability as follows

\[
P\{\max_{t \in I} Z(t) \geq z\} \\
= P\{\max_{t \in I} \tilde{Z}(t) \geq z\} \\
= P\{\max_{t \in I} \sum_{i=1}^{k} u_i(t) X_i \geq z\} \\
= \int_{z}^{\infty} P\{\max_{t \in I} < u^k(t), U > \geq \frac{z}{x}\} \ P\{\| X \| \in dx\} \\
= \int_{z}^{(1+d_k)x} + \int_{(1+d_k)x}^{\infty} \equiv A + B. \tag{A.11}
\]

Here $u^k(t) = (u_1(t), \ldots, u_k(t))$, and $\| X \|$ is distributed as a Chi random variable with $k$ degree of freedom.

Applying Weyl's formula in Lemma 2.2, we have, for $r_{kz} \equiv \sqrt{2(1 - \frac{z}{x})} < r_{kz}$,

\[
P\{\max_{t \in I} < u^k(t), U > \geq \frac{z}{x}\} \\
= P\{ U : \inf_{t \in I} \| u^k(t) - U \| \leq \sqrt{2(1 - \frac{z}{x})}\} \\
= \frac{1}{\omega_{k-1}} \cdot \frac{2\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} \sum_{\epsilon = 0,\text{even}}^{d} \kappa_{\epsilon} J_{\epsilon}(\theta) \\
= \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \pi^{\frac{d+1}{2}}} \tag{A.12}
\]

Here $m = k - d - 1$, $\omega_{k-1} = \frac{2\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}$ is the surface area of $S^{k-1}$, $\kappa_{\epsilon}$'s are certain integral invariants of the manifold $U^k = \{ u^k(t) : t \in I \}$, $J_{\epsilon}(\theta)$'s are the incomplete beta functions as defined in Lemma 2.2, and

\[
\theta = \arccos\left(1 - \frac{r_{kz}^2}{2}\right) = \arccos\left(\frac{z}{x}\right) \tag{A.13}
\]
is the spherical radius of the tube

\[ T(r_{kz}) = \{ y : y \in S^{k-1}, \inf_{t \in I} \| u^k(t) - y \| \leq r_{kz} \} \]

(cf. Definition 2.1.) with radius \( r_{kz} \).

If \( x \in (z, (1 + d_k)x) \), then

\[
0 < r_{kz} \equiv \sqrt{2(1 - \frac{z}{x})} < \sqrt{\frac{2d_k}{1 + d_k}} = r_{kc}.
\]

Hence, combining it with (A.12), we get

\[
A = \int_z^{(1+d_k)x} P\{ \max_{t \in I} < u^k(t), U > \geq \frac{z}{x} \} \ P\{ \| X \| \in dx \}
\]

\[
= \frac{\Gamma(k/2)}{\Gamma(m/2)} \sum_{e=0, \text{even}}^{d} \kappa_e \int_z^{(1+d_k)x} f_k(x) J_e(\theta) dx
\]

\[
= \frac{\Gamma(k/2)}{\Gamma(m/2)} \sum_{e=0, \text{even}}^{d} \kappa_e \int_0^\infty f_k(x) J_e(\theta) dx
\]

\[
- \frac{\Gamma(k/2)}{\Gamma(m/2)} \sum_{e=0, \text{even}}^{d} \kappa_e \int_0^{(1+d_k)x} f_k(x) J_e(\theta) dx
\]

\[
\equiv A_1 + A_2,
\]

where \( f_k(y) \) is the density function of \( \| X \| \) (a Chi random variable with \( k \) degree of freedom):

\[
f_k(x) = \frac{x^{k-1} \exp\{-\frac{x^2}{2}\}}{2^{(k-1)/2} \Gamma(k/2)}.
\]

In the following, we shall prove that as \( z \to \infty \),

\[
\int_{z(1+d_k)}^\infty P\{ \| X \| \in dx \} = o(\psi_d(z)) \quad (A.14)
\]

121
\[
A_1 = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^{d+1}} \sum_{e=0,\text{even}}^{d} \kappa_e \int_{\mathbb{R}} f_k(x)J_e(\theta)dx, \\
= \kappa_0\psi_0(z) + \kappa_2\psi_2(z) + \ldots + \kappa_\tilde{d}\psi_\tilde{d}(z), \quad (A.15)
\]
\[
A_2 = -\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^{d+1}} \sum_{e=0,\text{even}}^{d} \kappa_e \int_{(1+d_k)\mathbb{R}} f_k(x)J_e(\theta)dx = o(\psi_\tilde{d}(z)), \quad (A.16)
\]
\[
B = \int_{(1+d_k)\mathbb{R}} P\{\max_{t \in I} < u^k(t), U > \geq \frac{z}{x} \} P\{\|X\| \in dx\} = o(\psi_\tilde{d}(z)) \quad (A.17)
\]
where \( \tilde{d} \) is \( d \) if \( d \) is even, and \( d-1 \) if \( d \) is odd. The theorem follows from (A.15)-(A.17).

For (A.14), applying transformations \( y = x^2/2 \) and \( u = y/(1 + d_k)^2 \), we obtain
\[
\int_{(1+d_k)\mathbb{R}} P\{\|X\| \in dx\},
\]
\[
= \int_{(1+d_k)\mathbb{R}} \frac{x^{k-1}\exp\{-\frac{x^2}{2}\}}{2^{\frac{k+2}{2}}\Gamma\left(\frac{k}{2}\right)} dx
\]
\[
= \int_{(1+d_k)\mathbb{R}} \frac{y^{\frac{k-2}{2}}\exp\{-\frac{1}{2}\}}{\Gamma\left(\frac{k}{2}\right)} dy
\]
\[
= \int_{\frac{2}{2}}^{\infty} \frac{1}{\Gamma\left(\frac{k}{2}\right)} (1 + d_k)^k u^{\frac{k-1}{2}} \exp\{-2d_k u - d_k^2 u\} \cdot u^{-\frac{1}{2}} \exp\{-u\} du
\]
\[
\leq R_1 \int_{\frac{2}{2}}^{\infty} u^{-\frac{1}{2}} \exp\{-u\} du \quad \text{(for } \frac{z^2}{2} \geq \frac{k-1}{2(2d_k + d_k^2)})
\]
\[
= o(\psi_\tilde{d}(z))
\]
as \( z \to \infty \). Here we used the fact
\[
R_1 = \frac{(1 + d_k)^k}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{z^2}{2}\right)^{\tilde{d} + \frac{d_k - 1}{2}} \exp\{-2d_k\left(\frac{z^2}{2}\right) - d_k^2\left(\frac{z^2}{2}\right)\} \to 0 \quad (A.18)
\]
as \( z \to \infty \). We finished the proof of (A.14).

For (A.15), we use a similar proof as one given by Knowles (pp49ff). Notice that the inequality "\( \leq \)" on page 49 of his paper should be "\( = \)" since \( \cos \theta = z/x \). This
proof gives

\[ A_1 = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^{\frac{d+1}{2}}} \sum_{e=0, even}^{d} \kappa_e \int_{0}^{\infty} f_k(y) J_e(\theta) dy \]

\[ = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^{\frac{d+1}{2}}} \sum_{e=0, even}^{d} \kappa_e \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)2^{\frac{d}{2}+1}} \int_{0}^{\infty} u^{\frac{(d+1)-e}{2}-1} \exp\{-u\} du \]

\[ = \kappa_0 \psi_0(z) + \kappa_2 \psi_2(z) + \ldots + \kappa_d \psi_d(z), \]

where \( \psi_e(z) \)'s are as defined in (2.24) and (2.25). (A.15) is proved.

For (A.16), it is enough to show that as \( z \to \infty \),

\[ \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^{\frac{d+1}{2}}} \int_{(1+d_k)z}^{\infty} f_k(x) J_0(\theta) dx = o(\psi_d(z)) \]  \hspace{1cm} (A.19)

for \( \cos \theta = z/x \). In fact,

\[ \int_{(1+d_k)z}^{\infty} f_k(x) J_0(\theta) dx \]

\[ = \int_{(1+d_k)z}^{\infty} f_k(x) dx \int_{0}^{\arccos \frac{z}{k}} \sin^{m-1} t \cos^d t dt \]

\[ = \frac{1}{2} \int_{(1+d_k)z}^{\infty} f_k(x) dx \int_{(\frac{z}{k})^2}^{1} (1 - u)^{\frac{m-2}{2}} u^{\frac{d-1}{2}} du \quad \text{(} u = \cos^2 t \text{)} \]

\[ \leq \frac{1}{2} \int_{(1+d_k)z}^{\infty} f_k(y) dy \int_{0}^{1} (1 - u)^{\frac{m}{2}-1} u^{\frac{d-1}{2}} du \]

\[ = \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{d+1}{2}\right)}{2\Gamma\left(\frac{k}{2}\right)} \int_{z(1+d_k)}^{\infty} P\{||X|| \in dx\}. \]

So (A.19) follows from (A.14). (A.16) is proved.

For (A.17),

\[ B = \int_{(1+d_k)z}^{\infty} P\{\max_{t \in I} u^k(t), U > \frac{z}{x} \} \{||X|| \in dx\} \]

\[ \leq \int_{z(1+d_k)}^{\infty} P\{||X|| \in dx\} = o(\psi_d(z)) \quad \text{(by (A.14)).} \]
The proof of (A.17) is completed.

Therefore, we conclude the theorem.

Suppose $Z(t), t \in I$, is a $d$ dimensional Gaussian random field with mean zero, unit variance and has a uniformly convergent Karhunen-Loève expansion in $t \in I$:

$$ Z(t) = \sum_{l=1}^{\infty} u_l(t)X_l, \quad (A.20) $$

where $\|u(t)\| = \left[ \sum_{l=1}^{\infty} u_l^2(t) \right]^{1/2} = 1$, $X_1, X_2, \ldots$ are independent, identically distributed random variables from $\mathcal{N}(0, 1)$ (cf (2.16)). Let

$$ \tilde{Z}_k(t) = \sum_{l=1}^{k} u_l(t)X_l $$

be the $k$th partial sum of the Karhunen-Loève expansion in (A.20),

$$ Z_k(t) = \sum_{l=1}^{k} \frac{u_l(t)}{\sigma_l(t)}X_l $$

be the normalized version of $\tilde{Z}_k(t)$ for $k = 1, 2, \ldots$, where $\sigma_k^2(t) = \sum_{l=1}^{k} u_l^2(t)$.

Define

$$ a_k^2 = \max_{t \in I} \sum_{l=k+1}^{\infty} u_l^2(t); \quad (A.21) $$

$$ b_k^2 = \min_{t \in I} \sum_{l=k+1}^{\infty} u_l^2(t). \quad (A.22) $$

Write $\kappa^k_e$ for the integral invariants in Weyl’s formula when the manifold is

$$ V^k = \{ v^k(t) : v^k(t) = (v_1(t), \ldots, v_k(t)) \} $$

124
with \( v_l(t) = \frac{u_l(t)}{\sigma_x(t)} \) for \( l = 1, 2, \ldots, k \) (cf. (2.19) and its following paragraph), \( \kappa_x \)'s corresponding to \( \kappa_x^k \)'s geometrically for the manifold

\[
\mathcal{U} = \{ u(t) : u(t) = (u_1(t), u_2(t), \ldots, ) \}.
\]

In the following, we shall use Lemma A.3, Proposition A.1 and A.2 to prove Theorem 2.2 and 2.3. Here Proposition A.1 and A.2 conclude the two properties (cf. the discussion prior to Remark 2.3) of the related manifolds to the random field \( Z(t) \) through an expansion of \( r(s, t) \), under \( \textbf{R.1} \) (for \( m = 6 \)) and \( \textbf{R.2} \), the conditions on the covariance function of \( Z(t) \) given in Theorem 2.2 and 2.3.

**Lemma A.3** Suppose \( Z(t) \) is a \( d \) dimensional random field (not necessarily homogeneous) with mean zero, unit variance and the covariance function \( r(s, t) \) which satisfies

\[
r(s, t) = 1 - \sum_{i=1}^{d} a_i(s, t)|s_i - t_i|^\alpha + o(|s - t|^\alpha)
\]

as \( |s - t| \to 0 \), where \( 0 < \alpha \leq 2 \) and \( a_i(s, t) \) are non-negative and bounded on a rectangle \( I \).

Then as \( z \to \infty \),

\[
P\{ \max_{t \in I} Z(t) \geq z \}
= c z^{2d/\alpha} (\sqrt{2\pi z})^{-1} \exp\{-\frac{z^2}{2}\} (1 + o(1)),
\]

where

\[
c = \lim_{a \to 0} \lim_{n \to \infty} \int_{t \in I} \frac{H_a(n, a, t)}{n^d a^d} dt.
\]

125
with
\begin{equation}
H_{\alpha}(n, a, t) = 1 + \int_{-\infty}^{0} e^{-x} P\{ \max_{\vec{j} \in (0, \vec{n})} Y_{\alpha}(\vec{j}, t) > -x \} \, dx,
\end{equation}

for \( \vec{j} = (j_1, j_2, \ldots, j_d) \). \( j_1, j_2, \ldots, j_d \) are integers and "\( \vec{j} \in (0, \vec{n}) \)" means that "\( 0 < j_1, j_2, \ldots, j_d < n \)". \( Y_{\alpha}(\vec{j}, t) \) is a normal random variable with mean and covariance as follows:

\begin{align*}
\mathbb{E}Y_{\alpha}(\vec{j}, t) &= -a^\alpha \sum_{i=1}^{d} a_i(\vec{j}, t) \mid j_i - t_i \mid^\alpha \\
\text{Cov}(Y_{\alpha}(\vec{i}, t), Y_{\alpha}(\vec{j}, t)) &= -a^\alpha \left( \sum_{i=1}^{d} a_i(\vec{i}, t) \mid i_i - t_i \mid^\alpha + \sum_{i=1}^{d} a_i(\vec{j}, t) \mid j_i - t_i \mid^\alpha \\
&\quad- \sum_{i=1}^{d} a_i(\vec{i}, \vec{j}) \mid i_i - j_i \mid^\alpha \right).
\end{align*}

\textbf{Proof:} The proof is very long and follows closely from Chapter 12 in Leadbetter, Lingren and Rootzen (1983). The main steps are as follows:

1). using conditional argument (on \( Z(t) \)) to approximate
\[ P\{ \max_{\vec{j} \in (0, \vec{n})} Z(t + \vec{q}) > z \} \]
for \( z \to \infty, q \to 0, z^{2/\alpha}q \to a > 0 \);

2). based on the approximation obtained in 1), using Bonferroni inequality to approximate \( P\{ \max_{\vec{j} \in (0, \vec{h})} Z(t + \vec{q}) > z \} \) for fixed \( \vec{h} = (h_1, \ldots, h_d) \), and then to \( P\{ \max_{t \in I} Z(t) > z \} \) for \( z \to \infty \).

\[ \square \]

\textbf{Remark A.4:} Lemma A.3 can be seen as a first order approximation to the tail probability \( P\{ \max_{t \in I} Z(t) > z \} \), as \( z \to \infty \). Indeed, it is easy to see that it has the
same order as \( \psi_0(z) \), the first term of the two term approximation formula, when the random field is differentiable \((\alpha = 2)\).

In Lemma A.3, it is not necessary to assume that the random field \( Z(t) \) is differentiable.

The following Proposition A.1 shows that the first property of the related manifold (cf. the discussion prior to Remark 2.3) is valid.

**Proposition A.1** Suppose \( Z(t) \) is a non singular random field on a \( d \) dimensional compact space \( I \) with mean zero, unit variance, covariance function \( r(s,t) \) which satisfies the regularity conditions \( R.1 \) (for \( m = 6 \)) and \( R.2 \). Let \( \tilde{r}_{kc} \) be the semi critical radius of the manifold

\[
\mathcal{V}^k = \{v^k(t) : \quad v^k(t) = \left( \frac{u_1(t)}{\sigma_k(t)}, \frac{u_2(t)}{\sigma_k(t)}, \ldots, \frac{u_k(t)}{\sigma_k(t)} \right) \quad t \in I \},
\]

(cf. Definition 2.2, ) where \( \sigma_k^2(t) = \sum_{i=1}^k u_i^2(t) \) and \( r(s,t) \) has a uniformly convergent expansion as the one in Lemma B.1:

\[
r(s,t) = \sum_{i=1}^\infty u_i(s)u_i(t).
\]

Then there exists a positive constant \( c_0 > 0 \) such that

\[
\tilde{r}_{kc} \geq c_0 > 0. \tag{A.27}
\]

**Proof:** By assumption, the manifold \( \mathcal{V}^k \) and

\[
\mathcal{U} = \{u(t) : \quad u(t) = (u_1(t), u_2(t), \ldots), t \in I \}
\]

are twice differentiable.

127
Assume there is no such $c_0$ for (A.27). Then for any $\epsilon_k$ which decreases to zero as $k \to \infty$, there is $n_k \to \infty$ such that

$$\tilde{r}_{n_k} < \epsilon_k,$$

(A.28)

i.e. $\lim_{k \to \infty} \tilde{r}_{n_k} = 0$. Without lose of generality, we denote $n_k$ by $k$.

According to Definition 2.2,

$$\tilde{r}_k = \inf \{ r : r \geq 0, \exists y(t, \xi) = v^k(t) + \sum_{i=1}^{k-d-1} \xi_i n_i^k(t), \frac{y(t, \xi)}{\|y(t, \xi)\|} \in T(r), \text{ s.t. } J(y) = 0 \},$$

where $n_i^k(t), i = 1, 2, \ldots, k-d-1$ are mutually orthogonal unit normal vectors of $\mathcal{V}^k$ at $t$ and are orthogonal to $v^k(t)$. By (A.28), for the tube of $\mathcal{V}^k$ with radius $r_k = 2r_{kc}$, there is $y = y(t, \xi) = v^k(t) + \sum_{i=1}^{k-d-1} \xi_i n_i^k(t), \frac{y(t, \xi)}{\|y(t, \xi)\|} \in T(r_k)$ such that

$$J(y) = \left\| y, \frac{\partial y}{\partial t_1}, \ldots, \frac{\partial y}{\partial t_d}, n_1^k(t), n_2^k(t), \ldots, n_{k-d-1}^k(t) \right\| = 0.$$  

(A.29)

In the following, we show that (A.29) introduces a contradiction. Therefore, (A.27) holds.

Expressing $\frac{\partial n_i^k(t)}{\partial t_i}$ by Weingarten Equation (cf. Definition B.5), we have

$$\frac{\partial n_i^k(t)}{\partial t_i} = -\sum_j L_i^j(l) \frac{\partial v^k(t)}{\partial t_j} + \ldots,$$

(A.30)

where $-L_i^j(l)$ is the coefficient of $\frac{\partial n_i^k(t)}{\partial t_i}$ in the direction $\frac{\partial v^k(t)}{\partial t_j}$, and "+..." are components orthogonal to the tangent space spanned by $\frac{\partial v^k(t)}{\partial t_j}, i = 1, \ldots d$. Hence,

$$\frac{\partial y}{\partial t_i} = \frac{\partial}{\partial t_i} \left( v^k(t) + \sum_{i=1}^{k-d-1} \xi_i n_i^k(t) \right)$$

$$= \frac{\partial v^k(t)}{\partial t_i} + \sum_{i=1}^{k-d-1} \xi_i \frac{\partial n_i^k(t)}{\partial t_i}.$$
\[
\frac{\partial v^k(t)}{\partial t_i} + \sum_{i=1}^{k-d-1} \xi_i^k \left( -\sum_j L_i^j(l) \frac{\partial v^k(t)}{\partial t_j} + \ldots \right),
\]

which implies that

\[
\left( \frac{\partial \hat{y}}{\partial t_1}, \ldots, \frac{\partial \hat{y}}{\partial t_d} \right) = \left( \frac{\partial v^k(t)}{\partial t_1}, \ldots, \frac{\partial v^k(t)}{\partial t_1} \right) \left( I_{d \times d} - \sum_{i=1}^{k-d-1} \xi_i^k \bar{L}(l) \right), \tag{A.31}
\]

where \( + \ldots \) are components orthogonal to the tangent space spanned by \( \frac{\partial v^k(t)}{\partial t_i} \), \( i = 1, \ldots, d \), \( \bar{L}(l) \) is the \( d \times d \) matrix \( \bar{L}(l) = (L_i^j(l))_{d \times d} \). Therefore,

\[
(\mathcal{J}(y))^2 = \left\| \frac{\partial y}{\partial t_1}, \ldots, \frac{\partial y}{\partial t_d}, n_1^k(t), n_2^k(t), \ldots, n_{k-d-1}^k(t) \right\|^2
\]

\[
= \left\| v^k(t), \frac{\partial \hat{y}}{\partial t_1}, \ldots, \frac{\partial \hat{y}}{\partial t_d}, n_1^k(t), n_2^k(t), \ldots, n_{k-d-1}^k(t) \right\|^2
\]

\[
= \left\| \left( \frac{\partial \hat{y}}{\partial t_1}, \ldots, \frac{\partial \hat{y}}{\partial t_d}, n_1^k(t), n_2^k(t), \ldots, n_{k-d-1}^k(t), v^k(t) \right)^T \cdot \left( \frac{\partial \hat{y}}{\partial t_1}, \ldots, \frac{\partial \hat{y}}{\partial t_d}, n_1^k(t), n_2^k(t), \ldots, n_{k-d-1}^k(t), v^k(t) \right) \right\|^2
\]

\[
= \left\| \left( \frac{\partial \hat{y}}{\partial t_1}, \ldots, \frac{\partial \hat{y}}{\partial t_d} \right)^T \left( \frac{\partial \hat{y}}{\partial t_1}, \ldots, \frac{\partial \hat{y}}{\partial t_d} \right) \right\|^2
\]

\[
= \left\| \left( \frac{\partial v^k(t)}{\partial t_1}, \ldots, \frac{\partial v^k(t)}{\partial t_d} \right)^T \cdot \left( \frac{\partial v^k(t)}{\partial t_1}, \ldots, \frac{\partial v^k(t)}{\partial t_d} \right) \right\|^2
\]

\[
\cdot \left\| I_{d \times d} - \sum_{i=1}^{k-d-1} \xi_i^k \bar{L}(l) \right\|^2
\]

\[
= \left\| \left( g^k_{ij}(t) \right)_{d \times d} \cdot I_{d \times d} - \sum_{i=1}^{k-d-1} \xi_i^k \bar{L}(l) \right\|^2, \tag{A.32}
\]

where \( \| \cdot \| \) represents the determinant of the matrix \( \cdot \cdot \cdot \), and \( (g^k_{ij}(t)) \) is the \( d \times d \) matrix.
matrix with elements
\[ g_{ij}^k(t) = \frac{\partial v^k(t)}{\partial t_i} \frac{\partial v^k(t)}{\partial t_j} = \sum_{i=1}^k \frac{\partial v_i(t)}{\partial t_i} \frac{\partial v_i(t)}{\partial t_j}. \]

In the following, we shall prove as \( k \to \infty \)
\[ \| (g_{ij}^k(t))_{d \times d} \| \to \| (g_{ij}(t))_{d \times d} \| \geq c_1 \]  \hspace{1cm} (A.33)
\[ \| I_{d \times d} - \sum_{i=1}^{k-d-1} \xi_i^k \tilde{L}(I) \| \to c_2. \]  \hspace{1cm} (A.34)

for some \( c_1, c_2 > 0 \), where \( g_{ij}(t) = \frac{\partial u(t)}{\partial t_i} \frac{\partial u(t)}{\partial t_j} = \sum_{i=1}^{\infty} \frac{\partial u_i(t)}{\partial t_i} \frac{\partial u_i(t)}{\partial t_j} \). By (A.32), (A.33) and (A.34) contradict to (A.29). This concludes (A.27).

By Lemma B.1, for some constant \( c > 0 \), and any \( t, s \in I \),
\[ |u_k(t)u_k(s)| \leq \frac{c}{k^6}; \]
\[ |u_k(t)| \leq \frac{c}{k^3}. \]
\[ |\frac{\partial u_k(t)}{\partial t_i}| \leq \frac{c}{k^2}. \]
\[ |\frac{\partial^2 u_k(t)}{\partial t_i \partial t_j}| \leq \frac{c}{k}. \]

This implies that as \( k \to \infty \),
\[ \sum_{i=1}^k u_i(s) u_i(t) \to r(s, t), \]
\[ \sum_{i=1}^k u_i(s) \frac{\partial u_i(t)}{\partial t_i} \to \frac{\partial r(s, t)}{\partial t_i}, \]
\[ \sum_{i=1}^k u_i(s) \frac{\partial^2 u_i(t)}{\partial t_i \partial t_j} \to \frac{\partial^2 r(s, t)}{\partial t_i \partial t_j}, \]
\[ \sum_{i=1}^k \frac{\partial u_i(s)}{\partial s_i} \frac{\partial u_i(t)}{\partial t_j} \to \frac{\partial^2 r(s, t)}{\partial s_i \partial t_j}, \]
\[ \sum_{i=1}^k \frac{\partial^2 u_i(s)}{\partial s_i \partial t_j} \frac{\partial u_i(t)}{\partial t_j} \to \frac{\partial^4 r(s, t)}{\partial s_i \partial s_j \partial t_i \partial t_j}. \]  \hspace{1cm} (A.35)
uniformly in \( s, t \). Thus as \( k \to \infty \),

\[
\frac{\partial \sigma_k}{\partial t_i} = \frac{\partial}{\partial t_i} \sqrt{\sum_{l=1}^{k} u_l(t)^2} = \frac{1}{\sigma_k} \sum_{l=1}^{k} \frac{\partial u_l(t)}{\partial t_i} \frac{u_l(t)}{u_l(t)} \\
\to \sum_{l=1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} u_l(t) = \frac{\partial r(s, t)}{\partial s_i} \bigg|_{s=t} = 0
\]

uniformly in \( t \) by \( \textbf{R.2} \), which gives as \( k \to \infty \)

\[
g^k_{ij}(t) = \left< \frac{\partial v^k}{\partial t_i}, \frac{\partial v^k}{\partial t_j} \right> \\
= \frac{1}{\sigma_k^2} \sum_{l=1}^{k} \left( \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right) - \frac{\partial \sigma_k}{\partial t_i} \frac{\partial \sigma_k}{\partial t_j} \frac{1}{\sigma_k^2} \\
\to \sum_{l=1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} = g_{ij}(t) \tag{A.36}
\]

uniformly in \( t \).

By Lemma 3.1, (A.35) gives \( g(t) \equiv (g_{ij}(t))_{d \times d} = \left( \frac{\partial r(s, t)}{\partial s_i, \partial t_j} \right)_{s=t} \). Hence (A.36), the nonsingularity and continuity of \( \left( \frac{\partial r(s, t)}{\partial s_i, \partial t_j} \right)_{s=t} \) on compact set \( I \) give the following two results:

1). (A.33) holds;

2). the inverse matrices \( g^{-1,k}(t) = (g^{k,ij}(t)) \) of \( g^k(t) \) and \( g^{-1}(t) = (g^{ij}(t)) \) of \( g(t) \) exist for \( k > k_0 \), some positive integer. The elements of these inverse matrices have a uniformly upper bound and \( \|g^{-1}(t)\| \) has a positive lower bound, i.e., for some \( M, M' > 0 \),

\[
g^{k,ij}(t) \leq M, \quad g^{ij}(t) \leq M, \quad \|g^{-1}(t)\| \geq M' \tag{A.37}
\]

Let \( L(l) \equiv (L_{ij})_{d \times d} = g^k \bar{L}(l), g^k = (g^k_{ij})_{d \times d} \), a \( d \times d \) symmetrical matrix, \( u^k(t) = (u_1(t), \ldots, u_k(t))^\tau \). By the Gauss Formula given in Definition B.5, \( \textbf{R.1} \) (for \( m = 6 \)
and (A.35)

\[
\sum_i L_{ij}^2(l) \leq \sum_i \left< \frac{\partial^2 u^k}{\partial t_i \partial t_j}, n_i^k \right> \leq \left\| \frac{\partial^2 u^k}{\partial t_i \partial t_j} \right\|^2 \\
= \left\| \frac{1}{\sigma_k(t)} \frac{\partial^2 u^k(t)}{\partial t_i \partial t_j} - \frac{1}{\sigma_k^2(t)} \left( \frac{\partial u^k(t)}{\partial t_i} \frac{\partial \sigma_k(t)}{\partial t_j} + \frac{\partial u^k(t)}{\partial t_j} \frac{\partial \sigma_k(t)}{\partial t_i} \right) \right\|^2 \\
+ \frac{u^k(t) \partial \sigma_k(t)}{2 \sigma_k^2(t)} \frac{\partial \sigma_k(t)}{\partial t_i} \left( \frac{\partial \sigma_k(t)}{\partial t_j} \right)^2 - \frac{u^k(t) \partial^2 \sigma_k(t)}{\sigma_k^2(t)} \frac{\partial \sigma_k(t)}{\partial t_i} \frac{\partial \sigma_k(t)}{\partial t_j} \right\|^2 \\
\leq 16 \left( \left\| \frac{\partial^2 u^k(t)}{\partial t_i \partial t_j} \right\|^2 + \left\| \frac{\partial u^k(t)}{\partial t_i} \right\|^2 \left( \frac{\partial \sigma_k(t)}{\partial t_j} \right)^2 \right. \\
\left. + \left( \frac{u^k(t) \partial^2 \sigma_k(t)}{\sigma_k^2(t)} \frac{\partial \sigma_k(t)}{\partial t_i} \frac{\partial \sigma_k(t)}{\partial t_j} \right)^2 \right)
\]

\leq M'' (A.38)

for some $M'' > 0$. (A.37) and (A.38) imply

\[
\sum_i (L_i^2(l))^2 \leq M''' ,
\]

for some $M''' > 0$. On the other hand, as $y(t, \xi)/\|y(t, \xi)\| \in T(r_k)$, we have

\[
\sum_i (\xi^k_i)^2/(1 + \sum_i (\xi^k_i)^2) \leq \bar{r}_k^2 ,
\]

which implies

\[
\sum_i (\xi^k_i)^2 \leq \frac{\bar{r}_k^2}{1 - \bar{r}_k^2} \to 0
\]

as $k \to \infty$. Consequently, (A.39) and (A.40) give

\[
\left\| I_{d \times d} - \sum_{i=1}^{k-d-1} \xi^k_i L(l) \right\| \to 1 > 0 ,
\]

as $k \to \infty$. This concludes (A.34).

Therefore, (A.33) and (A.34), which contradict to (A.29), hold. The (A.27) is proved.

\[\square\]
The following Condition A.1 and Condition A.2 are the regularity conditions for the rate of the convergence of
\[ \sum_{i=k+1}^{\infty} (u_i(t))^2 \]
to zero as \( k \to \infty \) (cf. the second property of the manifold of a random field \( Z(t) \)) (cf. teh discussion prior to Remark 2.3). The regularity conditions R.1 and R.2 of \( r(s,t) \) in Theorem 2.2 (2.3) imply Condition A.3 which is sufficient for Condition A.1 and A.2 to hold (cf. Proposition A.2).

**Condition A.1** Let \( Z(t) \) be a d dimensional random field with a uniformly convergent Karhunen-Loève expansion as in (A.20). There exist \( z_0 > 1, \epsilon_0 > 0 \) such that: for all \( z > z_0 \) there is a \( k \leq z^{2-\epsilon_0} \) and \( \epsilon_{kz} > 0 \) with
\[ a_k^2 \leq \frac{\epsilon_{kz}^2}{(1 + \delta)}, \]
where \( \delta \) satisfies that as \( z \to \infty \),
\[ \delta \frac{z^2}{2} - 2 \log z \to \infty, \]
and \( \epsilon_{kz} = o\left(\frac{1}{z^2}\right) \) as \( z \to \infty \) (for \( k \leq z^{2-\epsilon_0} \)). Here \( a_k^2 \) is defined in (A.21).

**Remark A.5**: Condition A.1 is satisfied if \( a_k^2 \leq c/k^{2+\epsilon} \) for some \( c \) and \( \epsilon > 0 \).

**Condition A.2** For the same random field \( Z(t) \) and \( k, \epsilon_{kz} \) in Condition A.1,
\[ \kappa_0^k \cdot \frac{\psi_0(z')}{\psi_0(z)} - \kappa_0 = o\left(\frac{1}{z^2}\right), \]
\[ \kappa_0^k \cdot \frac{\psi_0(z'')}{\psi_0(z)} - \kappa_0 = o\left(\frac{1}{z^2}\right), \]

133
as \( z \to \infty \), where \( \psi_0(z) \) is defined in (2.27), and

\[
z' = z \frac{1 - \epsilon_{kz}}{\sqrt{1 - b_k^2}}, \quad z'' = z \frac{1 - \epsilon_{kz}}{\sqrt{1 - a_k^2}}
\]

with \( a_k^2 \) and \( b_k^2 \) as defined in (A.21) and (A.22), \( \kappa_0 \) and \( \kappa_0^k \) defined in the paragraph following (A.22).

Condition A.2 can also be simplified.

For \( e = 0, 2, \ldots, \leq d \), as \( z, k \to \infty \), \( k \leq z^{2-\varepsilon_0} \),

\[
\frac{\psi_e(z')}{\psi_e(z)} = \frac{\int_{\frac{z}{2}}^{\infty} u \frac{e^{k-\frac{u}{2}}}{2} \exp \{-u\} \, du}{\int_{\frac{z}{2}}^{\infty} u \frac{e^{k-\frac{u}{2}}}{2} \exp \{-u\} \, du}
\]

\[
\sim \frac{z'^{d-\varepsilon-1}}{z^{d-\varepsilon-1}} \exp \left\{ -z'^2 \left( \frac{\epsilon_{kz}}{\sqrt{1 - b_k^2}} \right)^2 \right\}
\]

\[
= \left( \frac{1 - \epsilon_{kz}}{\sqrt{1 - b_k^2}} \right)^{d-1-e} \exp \left\{ -z'^2 \left( \frac{1 - \epsilon_{kz}}{\sqrt{1 - b_k^2}} \right)^2 - 1 \right\}
\]

\[
= \left( \frac{1 - \epsilon_{kz}}{\sqrt{1 - b_k^2}} \right)^{d-1-e} \exp \left\{ -z'^2 \left( \frac{1}{2} (\epsilon_{kz}^2 - 2 \epsilon_{kz} + b_k^2)(1 + o(1)) \right) \right\}
\]

\[
= \left( \frac{1 - \epsilon_{kz}}{\sqrt{1 - b_k^2}} \right)^{d-1-e} \left( 1 - \frac{z'^2}{2} (\epsilon_{kz}^2 - 2 \epsilon_{kz} + b_k^2)(1 + o(1)) \right)
\]

\[
= 1 - \frac{z'^2}{2} (\epsilon_{kz}^2 - 2 \epsilon_{kz} + b_k^2)(1 + o(1)).
\]

Similarly, as \( z, k \to \infty \), \( k \leq z^{2-\varepsilon_0} \),

\[
\frac{\psi_0(z'')}{\psi_0(z)} = 1 - \frac{z'^2}{2} (\epsilon_{kz}^2 - 2 \epsilon_{kz} + a_k^2)(1 + o(1)).
\]

Since

\[
\kappa_0^k \cdot \frac{\psi_0(z')}{\psi_0(z)} - \kappa_0 = (\kappa_0^k - \kappa_0) \frac{\psi_0(z')}{\psi_0(z)} - \kappa_0 \left( \frac{\psi_0(z')}{\psi_0(z)} - 1 \right),
\]

134
a sufficient condition for Condition A.2 is the following (A.42):

\[
\begin{align*}
\kappa_0^k - \kappa_0 &= o\left(\frac{1}{z^2}\right), \\
(b_k^2 - 2\epsilon_kz + \epsilon_k^2z^4) &= o(1), \\
(a_k^2 - 2\epsilon_kz + \epsilon_k^2z^4) &= o(1).
\end{align*}
\]

(A.42)

It is easy to see that the following Condition A.3 is a simplified and sufficient condition for Condition A.1 and (A.42) (and hence Condition A.2).

**Condition A.3** Let \(Z(t)\) be a \(d\) dimensional random field with a uniformly convergent Karhunen-Loève expansion as in (A.20). There exist \(\epsilon_0, \epsilon_1 > 0\), and \(K, c > 0\) such that:

\[
\begin{align*}
\kappa_0^k - \kappa_0 &= o\left(\frac{1}{k^{1+\epsilon_0}}\right), \\
\epsilon_k^2 \leq \frac{c}{k^{4+\epsilon_1}} & \quad \text{for } k > K.
\end{align*}
\]

(A.43) \hspace{1cm} (A.44)

**Lemma A.4** Condition A.3 implies Condition A.1 and (A.42) (and hence Condition A.2).

**Proof:** Without loss of generality, assume \(\epsilon_0, \epsilon_1 < 1\).

Choose \(k = z^{2-\epsilon_0\epsilon_1/4}, z_0 = 1.1, \delta = 1.1\epsilon_1 - \epsilon_2\epsilon_1^2/8 - 1, \epsilon_k^2 = c/k^{4+\epsilon_1}/2\). Then

\[
b_k^2 \leq \epsilon_k^2 \leq \frac{c}{k^{4+\epsilon_1}} = \frac{\epsilon_k^2}{k^{\epsilon_1/2}} \sim \frac{\epsilon_k^2}{z^{\epsilon_1-\epsilon_0\epsilon_1^2/8}} \leq \frac{\epsilon_k^2}{1 + \delta}
\]

for \(z > z_0\), and

\[
\epsilon_kz \sim \frac{1}{k^{2+\epsilon_1/4}} \sim o\left(\frac{1}{z^4}\right).
\]

Hence Condition A.1 holds.
On the other hand,

\[ b_k^2 \leq a_k^2 \leq \frac{\epsilon_k^2}{1 + \delta} \sim o\left(\frac{1}{z^8}\right), \]

and by the assumption,

\[ \kappa_0^k - \kappa_0 = o\left(\frac{1}{k^{1+\epsilon_0}}\right) = o\left(\frac{1}{z^{2-\epsilon_0\epsilon_1/4}(1+\epsilon_0)}\right) = o\left(\frac{1}{z^{2+2\epsilon_0-\epsilon_0\epsilon_1/4-\epsilon_1\epsilon_2^2/4}}\right). \]

Therefore, (A.42) and hence Condition A.2 hold.

\[ \square \]

**Lemma A.5** Suppose \( Z(t) \) is a \( d \) dimensional random field on a bounded \( d \) dimensional closed space \( I \) with mean zero, unit variance and covariance function \( r(s,t) \), which has uniformly convergent expansion \( r(s,t) = \sum_{l=1}^{\infty} u_l(s)u_l(t) \) in \( t \in I \).

Further, assume \( \frac{\partial^2 r(s,t)}{\partial s_i \partial t_j} \) is uniformly bounded, \( \frac{\partial r(s,t)}{\partial s_i} |_{s=t} = 0 \), the \( d \times d \) matrix \( (\frac{\partial^2 r(s,t)}{\partial s_i \partial t_j} |_{s=t})_{d \times d} \) is continuous and nonsingular on \( I \), \( \sum_{l=1}^{k} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \) and \( \sum_{l=1}^{k} \frac{\partial u_l(t)}{\partial t_i} u_l(t) \) converges uniformly in \( I \) as \( k \to \infty \), and for some \( \epsilon_0, \epsilon_1 > 0 \),

\[ \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} u_l(t) = o\left(\frac{1}{k^{1/2+\epsilon_0}}\right), \]

\[ | \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} | = o\left(\frac{1}{k^{1+\epsilon_1}}\right). \]

as \( k \to \infty \). Then "\( a_k^2 \leq \frac{\epsilon^2}{k^{1+\epsilon_2}} \)" for some \( \epsilon_2 > 0 \) gives

\[ \kappa_0^k - \kappa_0 = o\left(\frac{1}{k^{1+\epsilon_3}}\right) \]

for some \( \epsilon_3 > 0 \), i.e. (A.44), the second part of Condition A.3 implies its first part (A.43). Here \( \kappa_0^k \) and \( \kappa_0 \) are the corresponding manifold formed from \( u_l(t) \)'s (cf. the paragraph following (A.22)).
Proof: Suppose \( R(t) = (g_{ij}(t))_{d \times d} \) is the matrix formed by the metric tensor with components \( g_{ij}(t) \) of the manifold \( \mathcal{U} \) of \( Z(t) \), \( R^k(t) = (g_{ij}^k(t))_{d \times d} \), is the matrix formed by metric tensor of the manifold \( \mathcal{W}^k \) of \( Z(t) \) (cf. Definition 3.1).

In the following, we shall prove that

\[
\kappa_0^k - \kappa_0 = \int_{t \in I} \sqrt{\| R^k(t) \|} \, dt - \int_{t \in I} \sqrt{\| R(t) \|} \, dt \\
\leq \frac{c}{\max_{i,j=1,\ldots,d} \left| g_{ij}(t) - g_{ij}^k(t) \right|} \tag{A.45}
\]

for some positive constant \( c \), and as \( z \to \infty \),

\[
\max_{i,j=1,\ldots,d} \left| g_{ij}(t) - g_{ij}^k(t) \right| = o\left( \frac{1}{z^2} \right). \tag{A.46}
\]

(A.45) and (A.46) conclude the lemma.

That \( r(t, t) = 1 \) and \( r(s, t) = \sum_{i=1}^k u_i(s)u_i(t) \) uniformly in \( s, t \in I \) implies that

\[
\sigma_k^2(t) = \sum_{i=1}^k u_i(t)u_i(t) \to 1, \tag{A.47}
\]

uniformly as \( k \to \infty \).

As \( \frac{\partial r(s, t)}{\partial s_i} \big|_{s=t} = 0 \), and \( \sum_{i=1}^k \frac{\partial u_i(t)}{\partial t_i}u_i(t) \) converges uniformly in \( I \), we can differentiate the Karhunen-Loève expansion \( Z(t) \) term by term by the theorem on P26 of Walker(1988). Hence, as \( k \to \infty \),

\[
\frac{\partial \sigma_k}{\partial t_i} = \frac{\partial}{\partial t_i} \sqrt{\sum_{i=1}^k u_i(t)^2} = \frac{1}{\sigma_k} \sum_{i=1}^k \frac{\partial u_i(t)}{\partial t_i}u_i(t) \\
\to \sum_{i=1}^\infty \frac{\partial u_i(t)}{\partial t_i}u_i(t) = \frac{\partial r(s, t)}{\partial s_i} \big|_{s=t} = 0.
\]

As \( \sum_{i=1}^k \frac{\partial u_i(t)}{\partial t_i} \frac{\partial u_i(t)}{\partial t_j} \) converges uniformly, \( g_{ij}(t) = \frac{\partial^2 r(s, t)}{\partial s_i \partial s_j} \big|_{s=t} \) by Lemma 3.1. It is easy to see for \( u^k(t) = (u_1(t), \ldots, u_k(t)) \) and \( v^k(t) = u^k(t)/\sigma_k(t) \), as \( k \to \infty \),

\[
ge^k_{ij}(t) = \frac{\partial v^k}{\partial t_i} \frac{\partial v^k}{\partial t_j}.
\]

137
\[
\begin{align*}
&= \frac{1}{\sigma_k^2} \sum_{l=1}^{k} \left( \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right) - \frac{\partial \sigma_k}{\partial t_i} \frac{1}{\sigma_k^2} \frac{\partial \sigma_k}{\partial t_j} \\
&\to \sum_{l=1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} = g_{ij}(t)
\end{align*}
\]

Further,

\[
|g_{ij}^k(t) - g_{ij}(t)|
= \frac{1}{\sigma_k^2} \sum_{l=1}^{k} \left| \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} - \frac{\partial \sigma_k}{\partial t_i} \frac{1}{\sigma_k^2} \frac{\partial \sigma_k}{\partial t_j} - \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right|
\leq \left( \frac{1}{\sigma_k^2} - 1 \right) \sum_{l=1}^{k} \left| \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right|
+ \left| \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right|
+ \left| \frac{\partial \sigma_k}{\partial t_i} \frac{\partial \sigma_k}{\partial t_j} \right| .
\]

By (A.47) and assumptions, we see that as \(k \to \infty,\)

\[
\left( \frac{1}{\sigma_k^2} - 1 \right) \sum_{l=1}^{k} \left| \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right|
+ \left| \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right|
\leq \frac{a_k^2}{\sigma_k^2} \sum_{l=1}^{k} \left| \frac{\partial r(s,t)}{\partial s} \frac{\partial r(s,t)}{\partial t} \right|_{s=t} + (1 + \frac{a_k^2}{\sigma_k^2}) \left| \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} \frac{\partial u_l(t)}{\partial t_j} \right|

= o\left(\frac{1}{k^{1+\min\{\epsilon_1,\epsilon_2\}}}\right);
\]

and

\[
\left| \frac{\partial \sigma_k}{\partial t_i} \frac{\partial \sigma_k}{\partial t_j} \frac{1}{\sigma_k^2} \right|
= \left| \left( \sum_{l=1}^{k} \frac{\partial u_l(t)}{\partial t_i} u_l(t) \right) \left( \sum_{m=1}^{k} \frac{\partial u_m(t)}{\partial t_j} u_m(t) \right) \frac{1}{\sigma_k^4} \right|
\leq \left| \sum_{l=k+1}^{\infty} \frac{\partial u_l(t)}{\partial t_i} u_l(t) \right| \left| \sum_{m=k+1}^{\infty} \frac{\partial u_m(t)}{\partial t_j} u_m(t) \right|
\leq o\left(\frac{1}{k^{1+2\alpha}}\right).
\]
Hence from (A.48), we have as $k \to \infty$,

$$|g_{ij}(t) - g_{ij}^k(t)| = o\left(\frac{1}{k^{1+\epsilon_3}}\right)$$

uniformly in $t \in I$, where $\epsilon_3 = \min\{\epsilon_1, \epsilon_2, 2\epsilon_0\}$. (A.46) is proved.

The following representation

$$R(t) = \left(\frac{\partial^2 r(s, t)}{\partial s_i \partial t_j}\right)_{s=t} dx^d$$

and the assumption on $r(s, t)$ imply that the determinant $\|R(t)\|$ of $R(t)$ has a positive lower bound, say, $l_b$. Therefore, $\|R^k(t)\| > l_b/2$ for all $k > K$, some positive constant, by (A.46).

By Definition B.1 for $\kappa_0$ and $\kappa_0^k$, we obtain

$$\kappa_0^k - \kappa_0 = \int_{t \in I} \sqrt{\|R^k(t)\|} \, dt - \int_{t \in I} \sqrt{\|R(t)\|} \, dt$$

$$\leq c(d^2, l_b) \max_{i,j=1,\ldots,d} |g_{ij}(t) - g_{ij}^k(t)|$$

for some positive constant $c(d^2, l_b)$. Here the constant $c(d^2, l_b)$ only depends on the dimension $d^2$ of the domain of the continuous multivariate function $\| \cdot \|$ of $g_{ij}(t), i, j = 1, \ldots, d$ and the lower bound $l_b$ of $\|R(t)\|$. (A.45) is proved. Therefore, we proved the lemma.

$\square$

**Proposition A.2** Suppose $r(s, t)$ is a non-negative definite covariance function of a $d$ dimensional differentiable differentiable random field $Z(t), t \in I$, where $I$ is a bounded $d$ dimensional closed space. Further, assume regularity condition R.1 (for $m = 6$), in Section 2.4, hold. Then there is a uniformly convergent Karhunen-Loève expansion $\sum_{i=1}^{\infty} u_i(t) X_i$ for $\tilde{Z}(t)$ which is distributed identically as $Z(t)$, and Condition A.3 for the manifold formed from $u_i(t)$'s is satisfied.
Proof: By Lemma B.1, for some constant $c > 0$, and any $t, s \in I$,

\[
|u_k(t)u_k(s)| \leq \frac{c}{k^8},
\]

\[
\left| \frac{\partial u_k(t)}{\partial t_i} \right| \leq \frac{c}{k^4}.
\]

Hence, for some constant $c'$, as $k \to \infty$,

\[
b_k^2 \leq a_k^2 = \max_{t \in I} \sum_{i=k+1}^{\infty} u_i^2(t) \sim \frac{c'}{k^5},
\]

and

\[
\sum_{i=1}^{k} \frac{\partial u_i(t)}{\partial t_i} u_i(t),
\]

\[
\sum_{i=1}^{k} \frac{\partial u_i(t)}{\partial t_i} \frac{\partial u_i(t)}{\partial t_j}
\]

converges uniformly in $t$ as $k \to \infty$. On the another hand,

\[
\sum_{i=k+1}^{\infty} \frac{\partial u_i(t)}{\partial t_i} u_i(t) \sim \frac{c}{k^4},
\]

\[
\sum_{i=1}^{k} \frac{\partial u_i(t)}{\partial t_i} \frac{\partial u_i(t)}{\partial t_j} \sim \frac{c}{k^3}
\]

as $k \to \infty$.

Therefore, all the conditions in Lemma A.5 and the second part of Condition A.3 hold, which also implies (A.43), the first part of Condition A.3.

The proposition is proved.

\[ \square \]

Proof of Theorem 2.3:
By Lemma A.1, \( Z(t) \) has a uniformly convergent standard Karhunen-Loève expansion as in (A.20).

By Proposition A.2, the assumptions R.1 (for \( m = 6 \)) and R.2 imply Condition A.3 and hence Condition A.1 and A.2 by Lemma A.4, where \( \varepsilon_k^2 \) is as in the proof of Lemma A.4.

To prove (2.26'), we need to show that as \( z \to \infty \),

\[
\frac{P\{\max_{t \in I} Z(t) \geq z\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} = 1 + o(1). \quad \text{(A.49)}
\]

On one side, since

\[
P\{\max_{t \in I} Z(t) \geq z\} = P\{\max_{t \in I} (Z(t) - \tilde{Z}_k(t) + \tilde{Z}_k(t)) \geq z\} \leq P\{\max_{t \in I} \tilde{Z}_k(t) \geq z(1 - \epsilon_{kz})\} + P\{\max_{t \in I} (Z(t) - \tilde{Z}_k(t)) \geq z\epsilon_{kz}\},
\]

we have

\[
\frac{P\{\max_{t \in I} Z(t) \geq z\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} \leq \frac{P\{\max_{t \in I} \tilde{Z}_k(t) \geq z(1 - \epsilon_{kz})\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} + \frac{P\{\max_{t \in I} (Z(t) - \tilde{Z}_k(t)) \geq z\epsilon_{kz}\}}{\kappa_2 \psi_2(z)} \equiv A + B. \quad \text{(A.50)}
\]

On the another side, as

\[
P\{\max_{t \in I} Z(t) \geq z\} = P\{\max_{t \in I} (Z(t) - \tilde{Z}_k(t) + \tilde{Z}_k(t)) \geq z\}
\]

141
\[
\geq P\left\{ \max_{t \in I} \tilde{Z}_k(t) - \max_{t \in I} Z(t) - \tilde{Z}_k(t) \mid \geq z \right\} \\
\geq P\left\{ \max_{t \in I} \tilde{Z}_k(t) - \max_{t \in I} Z(t) - \tilde{Z}_k(t) \mid \geq z, \max_{t \in I} Z(t) - \tilde{Z}_k(t) \mid < z \epsilon_{kz} \right\} \\
\geq P\left\{ \max_{t \in I} \tilde{Z}_k(t) \geq z(1 + \epsilon_{kz}) \right\} - P\left\{ \max_{t \in I} Z(t) - \tilde{Z}_k(t) \mid \geq z \epsilon_{kz} \right\},
\]

we see that

\[
\frac{P\left\{ \max_{t \in I} Z(t) \geq z \right\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} \\
\geq \frac{P\left\{ \max_{t \in I} \tilde{Z}_k(t) \geq z(1 + \epsilon_{kz}) \right\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} \\
\quad - \frac{P\left\{ \max_{t \in I} Z(t) - \tilde{Z}_k(t) \mid \geq z \epsilon_{kz} \right\}}{\kappa_2 \psi_2(z)} \\
\equiv A' + B'.
\]

(A.51)

In the following, we shall prove that for \( k \) properly chosen as above (see Condition A.1 and A.2), as \( z \to \infty \)

\[
A = \frac{P\left\{ \max_{t \in I} \tilde{Z}_k(t) \geq z(1 - \epsilon_{kz}) \right\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} \leq 1 + o(1), \quad (A.52)
\]

\[
B = \frac{P\left\{ \max_{t \in I} \left( Z(t) - \tilde{Z}_k(t) \right) \geq z \epsilon_{kz} \right\}}{\kappa_2 \psi_2(z)} = o(1) \quad (A.53)
\]

\[
A' = \frac{P\left\{ \max_{t \in I} \tilde{Z}_k(t) \geq z(1 + \epsilon_{kz}) \right\} - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} \geq 1 + o(1) \quad (A.54)
\]

\[
B' = \frac{P\left\{ \max_{t \in I} \left( Z(t) - \tilde{Z}_k(t) \right) \mid \geq z \epsilon_{kz} \right\}}{\kappa_2 \psi_2(z)} = o(1) \quad (A.55)
\]

From (A.52)-(A.55) we conclude (2.26), and hence the theorem.

For (A.53), by R.2, we have for \( \alpha = 2 \),

\[
r(s, t) = 1 - \sum_{i=1}^{d} a_i(s, t) |s_i - t_i|^\alpha + o(|s - t|^\alpha)
\]

where \( a_i(s, t) \)'s are bounded and non-negative on \( I \).
Condition A.1 implies
\[
\frac{\varepsilon_k}{a_k^2} \geq 1 + \delta \tag{A.56}
\]
for some $\delta$ which satisfies $\delta \frac{\varepsilon_k^2}{2} - \log z^2 \to \infty$ as $z \to \infty$, and therefore,
\[
\frac{z \varepsilon_k}{a_k} \to \infty, \quad \text{as } z \to \infty. \tag{A.57}
\]
Here $a_k^2$ is defined in (A.21).

Denote
\[
\tilde{\sigma}_k^2(t) = \sum_{l=k+1}^{\infty} u_l^2(t), \quad \tilde{Z}(t) = \frac{Z(t) - \tilde{Z}_k(t)}{\tilde{\sigma}_k(t)}.
\]

By Lemma A.3, we obtain that as $z \to \infty$
\[
B = \frac{P\{\max_{t \in I} \tilde{Z}(t) \geq z \varepsilon_k \}}{\kappa_2 \psi_2(z)} \leq \frac{P\{\max_{t \in I} \tilde{Z}(t) \geq z \varepsilon_k / a_k \}}{\kappa_2 \psi_2(z)} \sim \frac{c(\varepsilon_k a_k)^d}{\kappa_2} \left( \frac{\varepsilon_k a_k}{a_k^2} \right)^{-1} \exp\left\{-\frac{1}{2} \left( \frac{\varepsilon_k a_k}{a_k^2} \right)^2 \right\} \frac{1}{4\pi} \int_{-\infty}^{\infty} u^{d-3} \exp\{-u\} du \tag{by (A.57)).
\]
\[
\propto c \left( \frac{\varepsilon_k}{a_k} \right)^{d-1} z^2 \cdot \exp\{-\frac{z^2}{2} \left( \frac{\varepsilon_k}{a_k} \right)^2 - 1 \right\} \tag{by (A.56)).
\]
\[
= o(1). \tag{by (A.56)).
\]

Here $c$ is some positive constant and "$\propto$" means "is proportional to". (A.53) is proved.

For (A.52),
\[
A = \frac{P\{\max_{t \in I} Z_k(t) \sigma_k(t) \geq z(1 - \varepsilon_k) \}}{\kappa_2 \psi_2(z)} - \kappa_0 \psi_0(z) \leq \frac{P\{\max_{t \in I} Z_k(t) \geq z' \}}{\kappa_2 \psi_2(z)} - \kappa_0 \psi_0(z). \tag{A.58}
\]
where $z' = z \frac{1 - c_{z,t}}{\sqrt{1 - b_k^2}}$, $b_k^2 = \min_{t \in I} \sum_{i=k+1}^{\infty} u_i^2(t)$ is defined in (A.22) and $\sigma_k^2 = \sum_{i=1}^{k} u_i^2(t)$.

Let

$$v^k(t) = \left( \frac{u_1(t)}{\sigma_k(t)}, \frac{u_2(t)}{\sigma_k(t)}, \ldots, \frac{u_k(t)}{\sigma_k(t)} \right).$$

By R.3, a similar proof as A in the proof of Theorem 2.1 gives that for $d_k = r_k^2/(2 - r_k^2)$

$$P\{\max_{t \in I} Z_k(t) \geq z'\}$$

$$= \int_{z'}^\infty P\{\max_{t \in I} U^k(t), U > z'/x\} P\{\|X\| \in dx\}$$

$$= \int_{z'}^{(1+d_k)z'} + \int_{(1+d_k)z'}^\infty$$

$$= \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{m}{2}) \pi \frac{d+1}{2}} \sum_{e=0, \text{even}}^d \kappa_e^k \int_{z'}^{(1+d_k)z'} f_k(x) J_e(\theta) dx$$

$$+ \int_{(1+d_k)z'}^\infty P\{\max_{t \in I} U^k(t), U > z'/x\} P\{\|X\| \in dx\}$$

$$= \kappa_0 \psi_0(z') + \kappa_2 \psi_2(z') + R_{kz'}, \quad (A.59)$$

where

$$R_{kz'} = -\frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{m}{2}) \pi \frac{d+1}{2}} \sum_{e=0,2}^d \kappa_e^k \int_{(1+d_k)z'}^\infty f_k(x) J_e(\theta) dx$$

$$+ \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{m}{2}) \pi \frac{d+1}{2}} \sum_{e=4, \text{even}}^d \kappa_e \int_{z'}^{(1+d_k)z'} f_k(x) J_e(\theta) dx$$

$$+ \int_{(1+d_k)z'}^\infty P\{\max_{t \in I} U^k(t), U > z'/x\} P\{\|X\| \in dx\}$$

$$\equiv R_{1kz'} + R_{2kz'} + R_{3kz'} \quad (A.60)$$

with $J_e(\cdot)$ as defined in Lemma 2.2 and the density function $f_k(y)$ of the Chi random variable with degree of freedom $k$. 

144
Therefore, from (A.58) and (A.59), we see that
\[
A \leq \frac{\kappa_0^k \psi_0(z') - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} + \frac{\kappa_2^k \psi_2(z')}{\kappa_2 \psi_2(z)} - \frac{R_{kz'}}{\kappa_2 \psi_2(z)}
\equiv A_1 + A_2 + A_3.
\]

To prove (A.52), it is enough to show that as \( z \to \infty \),
\[
A_1 = \frac{\kappa_0^k \psi_0(z') - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} \to 0,
\]
\[
A_2 = \frac{\kappa_2^k \psi_2(z')}{\kappa_2 \psi_2(z)} \to 1,
\]
\[
A_3 = -\frac{R_{kz'}}{\kappa_2 \psi_2(z)} \to 0
\]
\[(A.61)\]
\[(A.62)\]
\[(A.63)\]

For (A.61), by Condition A.2,
\[
A_1 = \frac{\kappa_0^k \psi_0(z') - \kappa_0 \psi_0(z)}{\kappa_2 \psi_2(z)} = \frac{(\kappa_0^k \psi_0(z') - \kappa_0) \psi_0(z)}{\kappa_2 \psi_2(z)}
\]
\[\propto (\kappa_0^k \psi_0(z') - \kappa_0) z^2 \]
\[= o(1)
\]
(A.61) is valid.

For (A.62), let \( \kappa_2 \) be the corresponding second integral invariant of the manifold \( U = \{ u(t) : i \in I \} \) in Weyl's formula. Lemma A.5 concludes that \( \kappa_2^k \to \kappa_2 \) as \( k \to \infty \).

Under Condition A.1, \( b_k^2 \leq a_k^2 \leq \frac{\tilde{c}_k}{1 + \tilde{c}_k} = o(\frac{1}{z^2}) \). Hence \( b_k^2 - 2 \epsilon_{kz} + \epsilon_{kz}^2 = o(\frac{1}{z^2}) \). By the same derivation as in (A.41), we have that \( \frac{\psi_2(z')}{\psi_2(z)} \to 1 \) as \( z \to \infty \). (A.62) is proved.

For (A.63), using Lemma A.5 again, we see that it is enough to prove
\[
\frac{R_{kz'}}{\psi_2(z)} = \frac{R_{1kz'} + R_{2kz'} + R_{3kz'}}{\psi_2(z)} \to 0
\]
145
as $z \to \infty$. (See (A.60).)

It is obvious that as $z \to \infty$,

$$R_{2kz'} = \frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^\frac{d+1}{2}} \sum_{e=4, \text{even}}^{d} \kappa_e \int_{z'}^{(1+d_k)z'} f_k(x)J_e(\theta)dx$$

$$= o(\psi_2(z'))$$

$$= o(\psi_2(z)) \quad \text{(by Condition A.1).}$$

Similarly to the proof for Theorem 2.1,

$$R_{3kz'} = \int_{z'(1+d_k)}^{\infty} P\{\max_{t \in I} u^k(t), U > \frac{z'}{x} \} P\{\|X\| \in dx\}$$

$$\leq \int_{(1+d_k)z'}^{\infty} P\{\|X\| \in dy\}$$

$$R_{1kz'} = -\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\pi^\frac{d+1}{2}} \sum_{e=0,2}^{d} \kappa_e \int_{(1+d_k)z'}^{\infty} f_k(x)J_e(\theta)dx$$

$$\leq c \frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^\frac{d+1}{2}} \int_{(1+d_k)z'}^{\infty} P\{\|X\| \in dy\},$$

for some $c > 0$. However, since $d_k = \frac{r_k^2}{2 - r_k^2} \geq \frac{r_k^2}{2} \geq \frac{c_0^2}{2}$ by R.4, using transformations $y = x^2/2$ and $u = y/(1 + c_0^2/2)^2$, we have

$$\int_{(1+d_k)z'}^{\infty} P\{\|X\| \in dx\}$$

$$\leq \int_{(1+\frac{c_0^2}{2})z'}^{\infty} f_k(x)dx$$

$$= \int_{z'(1+\frac{c_0^2}{2})}^{\infty} \frac{x^{k-1} \exp\{-\frac{x^2}{2}\}}{2^{\frac{k+1}{2}} \Gamma\left(\frac{k}{2}\right)}dx$$

$$= \int_{\frac{z'^2}{2}(1+\frac{c_0^2}{2})^2}^{\infty} \frac{y^{\frac{k-2}{2}} \exp\{-y\}}{2^{\frac{k-2}{2}} \Gamma\left(\frac{k}{2}\right)}dy$$

$$= \int_{\frac{z'^2}{2}}^{\infty} \frac{(1 + \frac{c_0^2}{2})^k}{\Gamma\left(\frac{k}{2}\right)} \int_{2u - u^2}^{\infty} \frac{\exp\{-\frac{k-4}{2}u - \frac{d_0}{2} - \frac{c_0^2}{4}\}}{2^{\frac{k-1}{2}} \Gamma\left(\frac{k}{2}\right)}du$$

146
\[
\leq R_1 \int_{\frac{k^2}{z^2}}^{\infty} u^{\frac{d-1}{2}} \exp\{-u\} du \\
= o(\psi_2(z')),
\]

where by Stirling's formula for \(\Gamma(\frac{d}{2})\), for \(k \leq cz^{2-t_0}\),

\[
R_1 = \frac{(1 + \frac{c_0^2}{2})^k}{\Gamma(\frac{d}{2})} \left(\frac{z'^2}{2}\right)^{k-\frac{d+1}{2}} \exp\left\{-c_0^2 \frac{z'^2}{2} - \frac{c_0^4 z'^2}{4 \cdot 2}\right\}
\]

\[
= \frac{1}{\sqrt{k}} \exp\left\{-\left(c_0^2 + \frac{c_0^4}{4}\right) \frac{z'^2}{2} + \frac{k - d + 1}{2} \log \frac{z'^2}{2} + k \log(1 + \frac{c_0^2}{2}) + \frac{k}{2} - \frac{k}{2} \log \frac{k}{2}\right\}
\]

\[
\to 0,
\]

as \(z \to \infty\). (A.63) is proved.

Hence, (A.52) is proved.

Similar proofs show that (A.54) and (A.55) hold.

As described in Chapter 3, \(\kappa_0, \kappa_2\) of a \(d\)-dimensional differentiable manifold \(U = \{u(t) : u(t) = (u_1(t), u_2(t), \ldots, u_k(t))\}\) are determined by the metric tensor \(g_{ij}\) of \(U\). By Lemma 3.1, \(g_{ij}(t) = \frac{\partial^2 r(s, t)}{\partial s_i \partial t_j} \big|_{s=t}\), which implies that \(\kappa_0\) and \(\kappa_2\) depend only on the double mixed derivatives of \(r(s, t)\).

Therefore, we conclude the theorem.

\[
\square
\]

**Proof of Theorem 2.2:**

The proof is almost same as the proof of Theorem 2.3, only some minor changes are needed.

147
By Proposition A.1,
\[ \tilde{r}_{kc} \geq c_0 > 0, \]  
(A.65)
for some \( c_0 > 0 \). (A.65) is similar to the regularity condition \( \mathbf{R.4} \), we call it \( \mathbf{R.4}' \).
In another words, \( \mathbf{R.1-R.3} \) and \( \mathbf{R.4}' \) are satisfied in this Theorem 2.2.

The minor changes from the proof of Theorem 2.3 are

1). \( \mathbf{R.4} \) in the proof of Theorem 2.3 is replaced by \( \mathbf{R.4} \).

2). The critical point \( d_k \) is replaced by the semi critical point \( \tilde{d}_k \): (cf. Definition 2.2.)

3). (A.59) on page 144 is changed into

\[ P(\max_{t \in I} Z_k(t) \geq z') \leq \kappa_0^k \psi_0(z') + \kappa_2^k \psi_2(z') + R_{ks'} \]

4). All the arguments about (A.54) and (A.55) are deleted.

Therefore, (A.52) and (A.53) still hold under the conditions of Theorem 2.2, which implies (2.26). We conclude the theorem.
Appendix B

Supplementary Theorems and Proofs for Chapter 3

This appendix introduces some useful geometrical definitions and gives theorems and proofs related to the work of Chapter 3. Notations in this appendix are the same as those used in Chapter 3.

Definitions

Suppose $U = \{u(t) : u(t) = (u_1(t), u_2(t), \ldots, u_k(t)) \in S^{k-1}, t \in I\}$ is a $d$-dimensional manifold, where $I$ is a d-dimensional metric space, $k \leq \infty$. Its metric tensor is the inner product

$$g_{ij}(t) = \langle \frac{\partial u(t)}{\partial t_i}, \frac{\partial u(t)}{\partial t_j} \rangle = \sum_{i=1}^{k} \frac{\partial u_i(t)}{\partial t_i} \cdot \frac{\partial u_i(t)}{\partial t_j} \quad (B.1)$$

of the partial derivatives of $u$. Here $i, j = 1, 2, \ldots, d$, $t = (t_1, t_2, \ldots, t_d) \in I$. The $d \times d$ symmetric matrix $R(\cdot) = (g_{ij}(\cdot))_{d \times d}$ is called the metric tensor matrix. The
inverse matrix $R^{-1}(t)$ of $R(t)$ is written as

$$R^{-1}(t) = (g^{ij}(t))_{dxd}. \quad (B.2)$$

The properties of this manifold $\mathcal{U}$ is determined by its metric tensor, which can be seen partially from the following definitions of some differential geometric terms. These definitions also indicate why the calculation involved for $\kappa_2$ is enormous.

**Definition B.1 (Volume (Area))** The volume, or area of a subset $M \subset \mathcal{U}$ is

$$V(M) = \int_{t \in u^{-1}(M)} \sqrt{||R(t)||} \ dt.$$ 

where $||R(t)||$ is the determinant of $R(t)$.

(cf. Millman and Parker, p130, 1977)

Hence, the volume of $\mathcal{U}$ is

$$V(\mathcal{U}) = \int_{t \in I} \sqrt{||R(t)||} dt. \quad (B.3)$$

**Definition B.2 (Riemannian Curvature Tensor)** The Riemannian Curvature Tensor of $\mathcal{U}$ is the tensor with components

$$R^l_{ijk}(t) = \frac{\partial \Gamma^l_{ik}(t)}{\partial t_j} - \frac{\partial \Gamma^l_{ij}(t)}{\partial t_k} + \sum_{p=1}^{d} (\Gamma^p_{ik}(t) \Gamma^l_{pj}(t) - \Gamma^p_{ij}(t) \Gamma^l_{pk}(t)), \quad (B.4)$$

where $\Gamma^k_{ij}$ is called the Christoffel Symbols defined on $\mathcal{U}$,

$$\Gamma^k_{ij} = \frac{1}{2} \sum_{l=1}^{d} g^{jk}(t) \left( \frac{\partial g_{lj}(t)}{\partial t_i} - \frac{\partial g_{lj}(t)}{\partial t_i} + \frac{\partial g_{li}(t)}{\partial t_j} \right) \quad (B.5)$$

with $g^{ij}$ as in (B.2).
(cf. (41.4) on p134 in Kreyszig, 1968.)

\[ R_{ij}(t) = \sum_k R^k_{ijk}(t). \]

(cf. Kreyszig, p309, 1968.)

**Definition B.3 (Ricc Curvature Tensor)** The Ricc Curvature Tensor of \( U \) is the tensor with components

\[ R_{ij}(t) = \sum_k R^k_{ijk}(t). \]

(cf. Kreyszig, p310, 1968.)

\[ S(t) = \sum_{i=1, j=1}^d g^{ij}(t) R_{ij}(t). \]

(cf. Kreyszig, p310, 1968.)

**Definition B.4 (Scalar Curvature)** The Scalar Curvature of \( U \) is

The Gauss formula of \( U \) gives linear expansion of \( \frac{\partial^2 u(t)}{\partial t_i \partial t_j} \) in terms of \( \frac{\partial u(t)}{\partial t_k} \), \( k = 1, \ldots, d \) and \( n_l, l = 1, 2, \ldots \):

\[ \frac{\partial^2 u(t)}{\partial t_i \partial t_j} = \sum_l \Gamma^l_{ij} \frac{\partial u(t)}{\partial t_l} + \sum_l L_{ij}(l) n_l, \]

where \( L_{ij}(l) = \langle \frac{\partial u(t)}{\partial t_i}, n_l \rangle = \sum_m g_{im} L^{m}_{ij}(l), \Gamma^l_{ij} \) is the Christoffel Symbols defined as in Definition B.2.
The Weingarten equation of $\mathcal{U}$ gives linear expansion of $\frac{\partial n_l(t)}{\partial t_i}$ in terms of $\frac{\partial u(t)}{\partial t_k}, k = 1, \ldots, d$ and $n_l, l = 1, 2, \ldots$:

$$\frac{\partial n_l^k(t)}{\partial t_i} = -\sum_j L^i_j (l) \frac{\partial u^k(t)}{\partial t_j} + \ldots,$$

where $-L^i_j (l)$ is the coefficient of $\frac{\partial u^k(t)}{\partial t_i}$ in the direction $\frac{\partial u^k(t)}{\partial t_j}$, and "+\ldots" are components orthogonal to the tangent space spanned by $\frac{\partial u^k(t)}{\partial t_j}, i = 1, \ldots d$.

(cf. Kreyszig, p126ff, 1968.)

\[\square\]

Definition B.6 (Dirichlet Condition): A function $f(x_1, \ldots, x_d)$ defined on a rectangular region $I = [a_1, b_1] \times \ldots \times [a_d, b_d]$ satisfies the Dirichlet Condition, if one of the following two conditions holds:

1). $f$ is bounded on $I$. For any fixed $x_j, j \neq i$ of $j \in (1, \ldots, d)$, the interval $(a_i, b_i)$ can be broken up into a finite number of open partial intervals, in each of which $f(x_1, \ldots, x_d)$, as a function of $x_i$, is monotonic for $i = 1, \ldots, d$;

2). $f(x_1, \ldots, x_d)$ has a finite number of points of infinite discontinuity in $I$. When the arbitrary small neighborhoods of these points are excluded, $f$ is bounded on $I$, and the remainder of $(a_i, b_i)$ can be broken up into a finite number of open partial intervals (when $x_j$ is fixed for $j \neq i$), in each of which $f$ is monotonic in terms of $x_i, i = 1, \ldots, d$. Further, the infinite integral $\int \int \int f(x_1, \ldots, x_d) dx_1 \ldots dx_d$ is absolutely convergent.

(cf. Carslaw, 1930, p226.)

\[\square\]
Definition B.7 (Fourier Kernel) A \{n_1, n_2, \ldots, n_d\}th Fourier Kernel \(f^{i}_{n_1, n_2, \ldots, n_d}(t)\), on a \(d\) dimensional rectangle \(I\), is one of the following \(d(d+1)/2\) combinations of sine and cosine functions:

\[
\begin{align*}
  f^1_{n_1, n_2, \ldots, n_d}(t) &= \sin(b_{n_1}t_1) \sin(b_{n_2}t_2) \sin(b_{n_3}t_3) \ldots \sin(b_{n_d}t_d), \\
  f^2_{n_1, n_2, \ldots, n_d}(t) &= \cos(b_{n_1}t_1) \sin(b_{n_2}t_2) \sin(b_{n_3}t_3) \ldots \sin(b_{n_d}t_d), \\
  f^3_{n_1, n_2, \ldots, n_d}(t) &= \sin(b_{n_1}t_1) \cos(b_{n_2}t_2) \sin(b_{n_3}t_3) \ldots \sin(b_{n_d}t_d), \\
  f^4_{n_1, n_2, \ldots, n_d}(t) &= \cos(b_{n_1}t_1) \cos(b_{n_2}t_2) \sin(b_{n_3}t_3) \ldots \sin(b_{n_d}t_d), \\
  &\vdots \\
  f^{d(d+1)}_{n_1, n_2, \ldots, n_d} &= \cos(b_{n_1}t_1) \cos(b_{n_2}t_2) \cos(b_{n_3}t_3) \ldots \cos(b_{n_d}t_d),
\end{align*}
\]

which are mutually orthogonal when certain relations among the \(b_n\) and the lengths of sides of \(I\) are satisfied:

\[
\int_{I} f^{i}_{n_1, n_2, \ldots, n_d}(t) f^{i'}_{n'_1, n'_2, \ldots, n'_d}(t) \, dt = \begin{cases} 
  a^{n_1, \ldots, n_d}_{i, i'} > 0 & \text{if } i = i', \ (n_1, n_2, \ldots, n_d) = (n'_1, n'_2, \ldots, n'_d), \\
  0 & \text{otherwise} \end{cases} \quad (B.6)
\]

for \(i, i' = 1, 2, \ldots, (d+1)d/2, \ n_i, n'_i = 1, 2, \ldots, \) where \(\frac{b_{n_i}}{n_i} \geq \frac{b_{n_0}}{n_0} > 0, \) for \(n_i \geq n_0\) for some positive number \(n_0\).

Assume the orthogonality condition (B.6) holds, and for \(i = 1, 2, \ldots, (d+1)d/2, \ n_i = 1, 2, \ldots\) Let

\[
c^{i}_{n_1, n_2, \ldots, n_d} = \frac{1}{a^{n_1, n_2, \ldots, n_d}_{i, i}} \int_{I} f(t) \cdot f^{i}_{n_1, n_2, \ldots, n_d}(t) \, dt.
\]

153
The following series with the real coefficients \(c_{n_1,n_2,\ldots,n_d}^i\) and the kernels \(f_{n_1,n_2,\ldots,n_d}^i(t)\) as in Definition B.7:

\[
\sum_{n_1=0,n_2=0,\ldots,n_d=0}^{\infty} \sum_{i=1}^{d(d+1)/2} c_{n_1,n_2,\ldots,n_d}^i \cdot f_{n_1,n_2,\ldots,n_d}^i(t)
\]

is called a Fourier Series.

**Remark B.1:** A routine generalization of the standard theorems in Walker (pp26ff, pp185-216, 1988), Carslaw (pp225ff, 1930), Sneddon (p34, p40, 1961) and Tolstov (pp125-180, 1962) shows that all the properties about the coefficients of the above Fourier series, and its convergence, integration, differentiation are the same as those for the regular Fourier series where \(b_{n_i} = cn_i\) for some positive \(c\), \(i = 1,\ldots,d(d+1)/2\).

For a regular Fourier series of a function, where \(b_{n_i} = cn_i\), \(i = 1,\ldots,d(d+1)/2\), that the function takes the same value at all its end points is a necessary condition for its Fourier series to converge to itself in its entire domain. Under our definition of a Fourier series, it is not necessary for a function to have the same values at the end points for such convergence. Here is one example.

Let \(r(s,t)\) be the covariance function of a Brownian Motion \(B(t)\) on \((0,1)\). Then \(r(s,t) = \min\{s,t\}\) for \(0 \leq s,t \leq 1\), and \(r(0,0) \neq r(1,1)\). However, the following Fourier series convergences uniformly to \(r(s,t)\):

\[
r(s,t) = 2 \sum_{n=1}^{\infty} \left(\frac{\sin(n - \frac{1}{2})\pi t}{(n - \frac{1}{2})\pi}\right) \left(\frac{\sin(n - \frac{1}{2})\pi s}{(n - \frac{1}{2})\pi}\right).
\]

Here \(d = 1\), \(b_{n_1} = (n_1 - \frac{1}{2})\pi\). This expansion also gives a uniformly convergent Karhunen-Loève expansion of \(B(t)\):

\[
B(t) = \sqrt{2} \sum_{n=1}^{\infty} \frac{\sin(n - \frac{1}{2})\pi t}{(n - \frac{1}{2})\pi} X_n,
\]

154
where $X_1, X_2, \ldots$ are independent, identically distributed $\mathcal{N}(0, 1)$ random variables. (cf. Yaglom, 1987, pp448-450.).

In fact, for any function $f(t)$ defined on some rectangular region $I = I_1 \times \ldots \times I_d$, which has the same value at its end points with the $j$th order mixed partial derivative $f(t)$ either continuous or satisfy the Dirichlet Condition (See Definition B.6) for $j = 1, 2, \ldots d$, there is a Fourier series for $f$ which converges uniformly to $f$ in any region which contains neither the interior nor an end point of the set of discontinuity point of the function and its derivatives (cf. p275, in Carslaw, 1930 and (3.6) on pp185ff in Walker, 1988). Here $I_i = [a_i, b_i]$ for $i = 1, \ldots, d$.

If $f(t)$ on $I$ does not have the same value at its end points, we can always transform $f(t)$, as indicated below, into a new function (cf. (B.7)), which has the same value at all end points. This new function has $j$th order bounded continuous mixed partial derivatives (or its $j$th order mixed partial derivatives satisfy Dirichlet Condition) if only if $f(t)$ does so. Without lose of generality, assume that $d = 2$, i.e. $t = (t_1, t_2)$ on $[a_1, b_1] \times [a_2, b_2]$. Then the new function

$$f_{\text{new}}(t) = f(t_1, t_2) - \frac{t_1 - a_1}{b_1 - a_1}(f(b_1, a_2) - f(a_1, a_2)) - \frac{t_2 - a_2}{b_2 - a_2}(f(a_1, b_2) - f(a_1, a_2))$$

$$- \frac{(t_1 - a_1)(t_2 - a_2)}{(b_1 - a_1)(b_2 - a_2)}(f(b_1, b_2) - f(b_1, a_2) - f(a_1, b_2) + f(a_1, a_2)) \quad (B.7)$$

has the same value $f(a_1, a_2)$ at the four end points $(a_1, a_2), (a_1, b_2), (b_1, a_2), (b_1, b_2)$, and the difference between the old and new functions, viz. the function

$$f_{\text{diff}}(t) = \frac{t_1 - a_1}{b_1 - a_1}(f(b_1, a_2) - f(a_1, a_2)) + \frac{t_2 - a_2}{b_2 - a_2}(f(a_1, b_2) - f(a_1, a_2))$$

$$+ \frac{(t_1 - a_1)(t_2 - a_2)}{(b_1 - a_1)(b_2 - a_2)}(f(b_1, b_2) - f(b_1, a_2) - f(a_1, b_2) + f(a_1, a_2))$$

155
has a uniformly convergent Fourier series by a similar trick used for \( r(s, t) \) of \( B(t) \) (cf. Yaglom, 1987, pp448-450.). Therefore, if \( f(t) \) has \( j \)th order mixed continuous bounded derivatives, for \( j = 1, \ldots, d \), \( f_{new} \) does so and hence has a uniformly convergent Fourier expansion, which implies \( f(t) \) has a uniformly convergent trigonometric series. This trigonometric series is a summation of two Fourier series. We call the summation of a finite number of Fourier series Semi Fourier Series. The properties of the coefficients of a semi Fourier series are similar as those of a Fourier series. Without loss of generality, we can assume any function under the consideration has the same value at all the end points when we cite a theorem from a standard textbook for a Fourier series. A Fourier series of \( f(t) \) is meant to be a semi Fourier series of \( f(t) \) if \( f(t) \) does not have the same value at its end points (cf. Remark B.1).

**Definition B.8 (Regular Ordering)** A regular ordering of a multidimensional indexing Fourier series to a univariate indexing Fourier series is as follows:

\[
\sum_{i=1}^{\infty} c_i f_i(t),
\]

where \( c_i \) is the coefficient of the \( i \)th Fourier Kernel in the following order:

\[
f^{\frac{d(d+1)}{2}}_{0,0,0,\ldots,0}(t), \ldots, f^{\frac{d(d+1)}{2}}_{0,0,0,\ldots,0}(t),
\]

\[
f^{\frac{d(d+1)}{2}}_{1,0,0,\ldots,0}(t), \ldots, f^{\frac{d(d+1)}{2}}_{1,0,0,\ldots,0}(t),
\]

\[
\ldots, f^{\frac{d(d+1)}{2}}_{0,0,0,\ldots,1}(t), \ldots, f^{\frac{d(d+1)}{2}}_{0,0,0,\ldots,1}(t),
\]

\[
f^{\frac{d(d+1)}{2}}_{1,1,0,\ldots,0}(t), \ldots, f^{\frac{d(d+1)}{2}}_{1,1,0,\ldots,0}(t),
\]

\[
\ldots, f^{\frac{d(d+1)}{2}}_{1,0,0,\ldots,1}(t), \ldots, f^{\frac{d(d+1)}{2}}_{1,0,0,\ldots,1}(t),
\]

156
(cf. Figure B.1 on page 159 for the ordering diagram in the bivariate case.) Under this ordering, \( f_{i_1 \ldots i_d}^i(t) \) is ordered as \( f_l \) for some \( l \leq \frac{d(d+1)}{2} \left( \max_{i_1, \ldots, i_d} n_i \right)^d \), where \( i = 1, \ldots, d(d + 1)/2 \).

**Theorems**

**Lemma B.1** Suppose that a non-negative definite covariance function \( r(s, t) \) of a \( d \) dimensional random field \( Z(t) \) on a rectangle \( I \) satisfies the regularity condition

157
R. 1 in Section 2.4 for \( m = k \). Then there is a uniformly convergent expansion for \( r(s, t) = \sum_{i=1}^{\infty} u_i(s)u_i(t) \), where \( |u_i(s)u_i(t)| \leq c/l^k \) for all \( s, t \in I \), for some positive constant \( c \).

**Proof:** Case 1: condition 1) holds.

Without loss of generality, assume \( f(s) \equiv s \), \( I = [-\pi/2, \pi/2] \times \ldots \times [-\pi/2, \pi/2] \), then \( r(s, t) = r(s-t) = r(w_1, w_2, \ldots, w_d) \) is symmetrical in terms of each \( w_i = s_i - t_i \), \( i = 1, \ldots, d \) on \( [-\pi, \pi] \times \ldots \times [-\pi, \pi] \) and is bounded, even, and have up to \( kd^2 \) mixed bounded continuous partial derivatives. Hence \( r \) has a uniformly convergent Fourier series. (cf. (3.6) in Walker, 1988.) Further, without loss of generality again, assume \( r(\cdot) \) in terms of \( w_i \)'s has the same value at its end points.

In the following, we expand \( r(w_1, w_2, \ldots, w_d) \) according to its coordinates one by one to show that its Fourier expansion is its eigenvalue-eigenvector expansion. Therefore the Fourier series corresponds to a Karhunen-Loève expansion of \( Z(t) \) when this expansion exists by Mercer's theorem (cf. pp138-140, Courant and Hilbert (1953)).

Expanding \( r(w_1, w_2, \ldots, w_d) \) into a Fourier series in terms of \( w_1 \), we see that this series is a cosine series as \( r(\cdot) \) is symmetric in \( w_1 \):

\[
 r(w_1, w_2, \ldots, w_d) = \sum_{n=0}^{\infty} c_n^1(w_2, \ldots, w_d) \cos(nw_1), \tag{B.9}
\]

where the coefficients \( c_n^1(w_2, \ldots, w_d) \) are symmetric in terms of \( w_2, \ldots, w_d \), respectively, as \( r(\cdot) \) does.

Similarly as (B.9), for \( m = 0, 1, 2, \ldots \), expanding \( c_m^1(w_2, \ldots, w_d) \) into Fourier series in terms of \( w_2 \), we have that this series is still a cosine series in \( w_2 \):

\[
c_m^1(w_2, \ldots, w_d) = \sum_{n=0}^{\infty} c_{n,m}^2(w_3, \ldots, w_d) \cos(nw_2),
\]

158
\[ f_{ij} = (f_{ij}^1 	o f_{ij}^2 	o \ldots \to f_{ij}^{\frac{d(d+1)}{2}}) \]

\[ (f_{ij} \text{ represents } \frac{d(d+1)}{2} \text{ components}) \]

Figure B.1: The Ordering Diagram in the Bivariate Case

where the coefficients \( c_{n,m}^2(w_2, \ldots, w_d) \) is still symmetric in terms of \( w_3, \ldots, w_d \) respectively.

Repeating the above procedures for \( w_3, \ldots, w_d \), we have

\[ r(w_1, w_2, \ldots, w_d) = \sum_{n_1, n_2, \ldots, n_d=0}^{\infty} c_{n_1, n_2, \ldots, n_d} \cos(n_1 w_1) \cos(n_2 w_2) \ldots \cos(n_d w_d) \]

\[ = \sum_{n_1, n_2, \ldots, n_d=0}^{\infty} c_{n_1, n_2, \ldots, n_d} \cdot [\cos(n_1 s_1) \cos(n_1 t_1) + \sin(n_1 s_1) \sin(n_1 t_1)] \]
\[ \cdot [\cos(n_2 s_2) \cos(n_2 t_2) + \sin(n_2 s_2) \sin(n_2 t_2)] \]
\[ \ldots [\cos(n_d s_d) \cos(n_d t_d) + \sin(n_d s_d) \sin(n_d t_d)] \]
\[
= \sum_{n_1, n_2, \ldots, n_d=0}^{\infty} c_{n_1, n_2, \ldots, n_d} \sum_{i=1}^{d(d+1)/2} \Lambda_{n_1, n_2, \ldots, n_d}^i(t) \Lambda_{n_1, n_2, \ldots, n_d}^i(s),
\]

where \(\Lambda_{n_1, n_2, \ldots, n_d}^i(t)\) is the \(i\)th configuration of the following \((p-1)p/2\) combination of sines and cosines

\[
\sin(n_1 t_1) \sin(n_2 t_2) \sin(n_3 t_3) \ldots \sin(n_d t_d),
\]
\[
\cos(n_1 t_1) \sin(n_2 t_2) \sin(n_3 t_3) \ldots \sin(n_d t_d),
\]
\[
\sin(n_1 t_1) \cos(n_2 t_2) \cos(n_3 t_3) \ldots \sin(n_d t_d),
\]
\[
\cos(n_1 t_1) \cos(n_2 t_2) \cos(n_3 t_3) \ldots \sin(n_d t_d),
\]
\[
\ldots,
\]
\[
\cos(n_1 t_1) \cos(n_2 t_2) \cos(n_3 t_3) \ldots \cos(n_d t_d).
\]

Since \(r(\cdot)\) is nonnegative definite, \(c_{n_1, n_2, \ldots, n_d}\) is nonnegative. By an argument similar to the theorem in Walker (p199, 1988) gives

\[
|c_{n_1, n_2, \ldots, n_d}| \leq \frac{c}{n_1^{kd} n_2^{kd} \ldots n_d^{kd}}
\]

for some positive constant \(c\). Therefore, using the regular ordering in Definition B.8 and the remark following the definition, the reordered Fourier series:

\[
\sum_{l=1}^{\infty} c_l \sum_{i=1}^{d(d+1)/2} \Lambda_{l}^i(t) \Lambda_{l}^i(s)
\]

has property \(c_l \leq c/l^k\) for some \(c > 0\). This Fourier series uniformly converges to \(r(s, t)\) for \(k > 2\). Therefore, the conclusion of Lemma B.1 is true in Case 1.

**Case 2: condition 2) holds.**

160
As in case 1, without loss of generality, assume the function under consideration has the same value at its end points when we try to apply Fourier series theory.

From (B.10), the result in Case 1, we have

\[
 r(s, t) = \sum_{i,j=1}^{d_3} h_i(s^{(1)})h_j(t^{(1)}) \cdot \sum_{n_1, n_2, \ldots, n_{d_2} = 0}^{\infty} c_{n_1,n_2,\ldots,n_{d_2}}^{ij} \Lambda_{n_1,n_2,\ldots,n_{d_2}}^i(t) \Lambda_{n_1,n_2,\ldots,n_{d_2}}^j(s)
\]

\[
 = \sum_{n_1, n_2, \ldots, n_{d_2} = 0}^{\infty} \sum_{i,j=1}^{d_3} h_i(s^{(1)})h_j(t^{(1)})c_{n_1,n_2,\ldots,n_{d_2}}^{ij} \Lambda_{n_1,n_2,\ldots,n_{d_2}}^i(t) \Lambda_{n_1,n_2,\ldots,n_{d_2}}^j(s)
\]

\[
 = \sum_{n_1, n_2, \ldots, n_{d_2} = 0}^{\infty} c_{n_1,n_2,\ldots,n_{d_2}}(s^{(1)}, t^{(1)})\Lambda_{n_1,n_2,\ldots,n_{d_2}}^i(t) \Lambda_{n_1,n_2,\ldots,n_{d_2}}^j(s) \tag{B.11}
\]

where

\[
 c_{n_1,n_2,\ldots,n_{d_2}}(s^{(1)}, t^{(1)}) = \sum_{i,j=1}^{d_3} h_i(s^{(1)})h_j(t^{(1)})c_{n_1,n_2,\ldots,n_{d_2}}^{ij}.
\]

Using integration by part, we see \(c_{n_1,n_2,\ldots,n_{d_2}}^{ij} = \frac{C_{i,j,n_1,\ldots,n_{d_2}}}{n_1^{k(d_2+1)}n_2^{k(d_2+1)}\ldots n_{d_2}^{k(d_2+1)}}\) for some \(C_{i,j,n_1,\ldots,n_{d_2}}\) which depends only on \(i, j; n_1, \ldots, n_{d_2}\) and \(l\)th mixed order partial derivatives of \(h_{ij}\) for \(l = 1, 2, \ldots, 6d_2(d_2 + 1)\). Define

\[
 \tilde{c}_{n_1,n_2,\ldots,n_{d}}(s^{(1)}, t^{(1)}) = n_1^{kd_2+1}\ldots n_{d_2}^{kd_2+1}c_{n_1,n_2,\ldots,n_{d}}(s^{(1)}, t^{(1)}),
\]

it is easy to see that \(\tilde{c}_{n_1,n_2,\ldots,n_{d}}(s^{(1)}, t^{(1)})\) is bounded uniformly in \(i, j; n_1, \ldots, n_{d_2}\) and is a continuous symmetrical kernel in \(s^{(1)}, t^{(1)}\) which has up to 4th mixed order partial derivatives. Hence, \(\tilde{c}_{n_1,n_2,\ldots,n_{d}}(s^{(1)}, t^{(1)})\) has a uniformly convergent Mercer expansion:

\[
 \sum_{l=1}^{\infty} \mu_l^{n_1,n_2,\ldots,n_{d_2}}(s^{(1)})\mu_l^{n_1,n_2,\ldots,n_{d_2}}(t^{(1)})
\]

which implies that

\[
 r(s, t)
\]

161
\[
= \sum_{n_1, n_2, \ldots, n_d = 0}^{\infty} \frac{1}{n_1^{kd_1+1} \cdots n_d^{kd_d+1}} \left( \sum_{i=1}^{\infty} u_i^{n_1, n_2, \ldots, n_d} (s^{(1)}) u_i^{n_1, n_2, \ldots, n_d} (t^{(1)}) \right) \Lambda_{n_1, n_2, \ldots, n_d}^i (t) \Lambda_{n_1, n_2, \ldots, n_d}^i (s).
\]

Therefore, by the regular ordering from multiple index \( \{n_1, \ldots, n_d, l\} \) to univariate index, the conclusion of Lemma B.1 is true in Case 2.

\[\square\]

**Proposition B.1 (Sufficient conditions for (3.4) and (3.5))** Suppose \( r(s, t) \) is a non-negative definite covariance function of a differentiable random field \( Z(t), t \in I \), where \( I \) is a bounded \( d \) dimensional closed space. If one of the conditions in Lemma B.1 holds for \( k = 6 \), then (3.4) and (3.5) in Lemma 3.1 converge uniformly on \( I \).

**Proof:** By Lemma B.1,

\[ r(s, t) = \sum_{i=1}^{\infty} u_i(s) u_i(t), \]

where \( u_i(s) u_i(t) \leq c/l^6 \) for some \( c > 0 \). Therefore the proof of Proposition B.1 is similar to the proof of Proposition A.2.

\[\square\]

**Proofs of Theorems in Chapter 3**

**Proof of Proposition 3.1:**

*First*, we prove that Equation (3.9)-(3.11) hold.
By Proposition 2.1, the covariance function of \( Z \) is

\[
r(t, t') = \beta^T A(\alpha^T \alpha') \beta,
\]

where \( A(\alpha^T \alpha') = \mathcal{E}[\tilde{X}(\alpha) \cdot \tilde{X}^T(\alpha')] = (a_{ij}(\alpha^T \alpha'))_{J \times J} \) is a \( J \times J \) matrix, with element \( a_{ij}(\alpha^T \alpha') = \text{Cov}(\tilde{X}_i(\alpha), \tilde{X}_j(\alpha)) \) being \( C^\infty \) differentiable in \( (t, t') \in I \times I \). Here

\[
t = (\theta_1, \ldots, \theta_{p-1}, \varphi_1, \ldots, \varphi_{J-1}) \in I = [0, \pi] \times \ldots \times [0, \pi] \times [0, 2\pi] = \frac{1}{2} S^{p-1} \times S^{J-1}
\]

for

\[
\frac{1}{2} S^{p-1} = \{ \alpha : \alpha \in S^{p-1}, \alpha_p \geq 0 \}.
\]

Equivalently, \( a_{ij}(\alpha^T \alpha') \) depends only on \( \| \alpha - \alpha' \|^2 \) as \( \alpha^T \alpha' = 1 - \frac{1}{2} \| \alpha - \alpha' \|^2 \). Without confusion, we also write \( a_{ij}(\| \alpha - \alpha' \|^2) \) for \( a_{ij}(\alpha^T \alpha') \).

Here is a different parametrization from the one in Section 2.2,

\[
\alpha_i = \alpha_i, \ i = 1, \ldots, p - 1,
\]

\[
\alpha_p = \sqrt{1 - \alpha_1^2 - \ldots - \alpha_{p-1}^2}.
\]

We see that \( a_{ii}(\| \alpha - \alpha' \|^2) \) is symmetric in \( \alpha_1, \ldots, \alpha_{p-1} \) for \( i = 1, \ldots, J \). Hence, an obvious generalization of Lemma B.1 shows that

\[
a_{ii}(\| \alpha - \alpha' \|^2) = \sum_{n_1, n_2, \ldots, n_{p-1} = 0}^{\infty} c_{n_1, n_2, \ldots, n_{p-1}} \cdot \left( \sum_{i=1}^{(p-1)p} h_{n_1, n_2, \ldots, n_{p-1}}^i(\alpha_1, \ldots, \alpha_{p-1}) h_{n_1, n_2, \ldots, n_{p-1}}^i(\alpha_1', \ldots, \alpha_{p-1}'), \right), \tag{B.12}
\]

on \((0, \pi), \ldots, (0, \pi)\), and the series converges uniformly in \([\epsilon, \pi - \epsilon], \ldots, [\epsilon, \pi - \epsilon]\) for any fixed \( \epsilon > 0 \). Here \( h_{n_1, n_2, \ldots, n_d}^i(\alpha_1, \ldots, \alpha_{p-1}) \) is the \( i \)th configuration of the
following $d(d + 1)/2$ combination of of sines and cosines

$$\sin(b_{n_1} \alpha_1) \sin(b_{n_2} \alpha_2) \ldots \sin(b_{n_{p-1}} \alpha_{p-1}),$$

$$\cos(b_{n_1} \alpha_1) \sin(b_{n_2} \alpha_2) \ldots \sin(b_{n_{p-1}} \alpha_{p-1}),$$

$$\sin(b_{n_1} \alpha_1) \cos(b_{n_2} \alpha_2) \ldots \sin(b_{n_{p-1}} \alpha_{p-1}),$$

$$\cos(b_{n_1} \alpha_1) \cos(b_{n_2} \alpha_2) \ldots \sin(b_{n_{p-1}} \alpha_{p-1}),$$

$$\ldots$$

$$\cos(b_{n_1} \alpha_1) \cos(b_{n_2} \alpha_2) \ldots \cos(b_{n_{p-1}} \alpha_{p-1}).$$

Since $A(\cdot)$ is nonnegative definite, $c_{n_1, n_2, \ldots, n_d}$ is nonnegative. Substituting the polar coordinate representation of $\alpha$, in terms of $\theta = (\theta_1, \theta_2, \ldots, \theta_{p-1})$ as in Section 2.2, into the above series, shows that this Fourier series (B.12) can be written as

$$\sum_{n_1, n_2, \ldots, n_{p-1}=0}^{\infty} c_{n_1, n_2, \ldots, n_{p-1}} \left( \sum_{i=1}^{(p-1)p} \hat{h}_{n_1, n_2, \ldots, n_{p-1}} \theta_1, \ldots, \theta_{p-1} \right) \hat{h}_{n_1, n_2, \ldots, n_{p-1}} \theta'_1, \ldots, \theta'_{p-1}).$$

(B.13)

Here $\hat{h}$ has the form

$$\cos(b_{n_1} \cos \theta_1) \cos(b_{n_2} \sin \theta_1 \cos \theta_2) \ldots \sin(b_{n_{p-1}} \sin \theta_1 \sin \theta_2 \ldots \cos \theta_{p-1}).$$

Note that the infinite differentiability of $r(s, t)$ implies that

$$\sum_{n_1, n_2, \ldots, n_{p-1}=0}^{\infty} | c_{n_1, n_2, \ldots, n_{p-1}} | < \infty.$$

Hence the series in (B.13) also converges for

$$\theta = (\theta_1, \ldots, \theta_{p-1}) \in [0, \pi] \times [0, \pi] - (\epsilon, \pi - \epsilon) \times (\epsilon, \pi - \epsilon).$$

164
and uniformly in terms of \( \theta \), viz. (B.12) converges uniformly to \( a_{\theta}^{\prime} \) in \( \theta, \theta' \in [0, \pi] \times \ldots \times [0, \pi] \).

Assume \( d_1 = J - 1, d_2 = p - 1, d = p + J - 2, d_3 = J, s^{(1)} = (\varphi_1, \ldots, \varphi_{J-1}), t^{(1)} = (\varphi'_1, \ldots, \varphi'_{J-1}), s^{(2)} = (\theta_1, \ldots, \theta_{J-1}), t^{(2)} = (\theta'_1, \ldots, \theta'_{J-1}), f(s^{(2)}) = \alpha^{(-1)} = \alpha^{(-1)}(\theta_1, \ldots, \theta_{J-1}), \alpha^{(-1)}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_{p-1}(\cdot))^T, h_i(s^{(1)}) = \beta_i = \beta_i(\varphi_1, \ldots, \varphi_{J-1}), h_{ij} = a_{ij}, \) where \( a_{ij} = a_{ij}(\alpha^\tau \alpha') \) which is a function of \( \|\alpha - \alpha'\| \) and hence also a function of \( \|\alpha^{(-1)} - \alpha'^{(-1)}\| \). Therefore, as in the proof of Lemma B.1, there is a uniformly convergent Karhunen-Loève expansion of \( \tilde{Z}(t) \), which has the same distribution as \( Z(t) \), the random field derived from JHF PP index. This expansion satisfies conditions in Lemma B.1. By Proposition B.1, Lemma B.1 implies that (3.4) and (3.5) in Lemma 3.1 holds, hence Equation (3.9)-(3.11) hold.

**Second**, we prove (3.12)-(3.14).

In the following, \( s = (\theta_1, \ldots, \theta_{p-1}, \varphi_1, \ldots, \varphi_{J-1}) \), \( t = (\theta'_1, \ldots, \theta'_{p-1}, \varphi'_1, \ldots, \varphi'_{J-1}) \) \( \in I = (0, \pi] \times (0, \pi] \times \ldots \times (0, \pi] \times (0, 2\pi] \).

**Proof of (3.14)**

For \( i = 1, \ldots, p - 1, \quad j = 1, \ldots, J - 1, \)

\[
\frac{\partial^2 r(s, t)}{\partial \theta_i \partial \varphi'_j} \bigg|_{\theta = \theta', \varphi_j = \varphi'_j} = \frac{\partial^2}{\partial \theta_i \partial \varphi'_j} (\beta^\tau \mathcal{E}X(\alpha)X'(\alpha')\beta') \bigg|_{\theta = \theta', \varphi_j = \varphi'_j} \quad \text{(by (2.13))}
\]

\[
= \sum_{l,m=1}^J \frac{\partial^2}{\partial \theta_i \partial \varphi'_j} (\beta_l \beta'_m \mathcal{E} \hat{Y}_i(\alpha)\hat{Y}_m'(\alpha')) \bigg|_{\theta = \theta', \varphi_j = \varphi'_j}
\]

\[
= \sum_{l,m=1}^J \beta_l \beta'_m \mathcal{E} \left( \frac{\partial}{\partial \theta_i} \left( \sum_{t=1}^N \frac{2l + 1}{N} P_t(2\Phi(\alpha^\tau \hat{Z}_t) - 1) \right) \cdot \hat{Y}_m'(\alpha) \right) \quad \text{(by (2.2))}
\]
\[
\sum_{l,m=1}^{J} \beta_i \frac{\partial \beta_m}{\partial \varphi_j} \mathcal{E}\{(\sum_{t=1}^{N} \sqrt{\frac{2l+1}{N}} P'_1(2\Phi(X_t) - 1)2\phi(X_t) \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st}) \cdot \tilde{Y}_m(\alpha)\},
\]

(B.14)

where \(X_t = \alpha^T Z_t, Z_t = (Z_{1t}, \ldots, Z_{pt})\), \(X = (X_1, X_2, \ldots, X_N)\) is distributed as \(\mathcal{N}(0, I_N)\). Since

\[
\mathcal{E}\left[ \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \mid X \right]
\]

\[
= \mathcal{E}\left[ \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \mid X_t \right]
\]

\[
= \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} \alpha_s X_t \quad \text{(by Lemma 3.2)}
\]

\[
= 0, \quad \text{(by (3.17))},
\]

from (B.14) we have

\[
\frac{\partial^2 r(s, t)}{\partial \theta_i \partial \varphi'_j} \bigg|_{\theta = \theta', \varphi_j = \varphi'_j}
\]

\[
= \sum_{l,m=1}^{J} \beta_i \frac{\partial \beta_m}{\partial \varphi_j} \mathcal{E}\{(\sum_{t=1}^{N} \sqrt{\frac{2l+1}{N}} P'_1(2\Phi(X_t) - 1)2\phi(X_t) \tilde{Y}_m(\alpha)\)}
\]

\[
\mathcal{E}\left( \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \mid X \right)
\]

\[
= 0.
\]

Here \(\tilde{Y}_m(\alpha)\) is a random variable depending only on \(X = (X_1, X_2, \ldots, X_N)\).

Therefore, \(R_{12} = R_{21} = 0\) and (3.14) is proved.

**Proof of (3.13)**

For \(i, j = 1, \ldots, J - 1,\)

\[
\frac{\partial^2 r(s, t)}{\partial \varphi_i \partial \varphi'_j} \bigg|_{\theta = \theta', \varphi_j = \varphi'_j}
\]

\(= 0.\)
\[
\begin{align*}
&= \sum_{i, m=1}^{J} \frac{\partial \beta_i}{\partial \varphi_i} \frac{\partial \beta_m}{\partial \varphi_j} \mathcal{E}(\bar{X}_i(\alpha) \bar{X}_m(\alpha)) \\
&= \sum_{i=1}^{J} \frac{\partial \beta_i}{\partial \varphi_i} \frac{\partial \beta_i}{\partial \varphi_i} \quad \text{(as } \bar{X}_1, \ldots, \bar{X}_N \text{ are i.i.d. from } \mathcal{N}(0, 1)) \\
&= \begin{cases} 
\sum_{i=1}^{J} (\frac{\partial \beta_i}{\partial \varphi_i})^2 & \text{if } i = j \\
0 & \text{otherwise}
\end{cases} \quad \text{(by (3.19)).}
\end{align*}
\]

Therefore, \(R_{22} = \text{right of (3.13), (3.13) is proved.}\)

**Proof of (3.12)**

For \(i, j = 1, \ldots, p - 1,\)

\[
\begin{align*}
\frac{\partial^2 r(s, t)}{\partial \theta_i \partial \theta_j} |_{\theta = \theta', \varphi_j = \varphi'}
&= \mathcal{E} \left\{ \sum_{i=1}^{J} (\beta_i \frac{\partial \bar{Y}_i(\alpha)}{\partial \theta_i}) \sum_{m=1}^{J} (\beta_m \frac{\partial \bar{Y}_m(\alpha)}{\partial \theta_m}) \right\} \\
&= \mathcal{E} \left\{ \left( \sum_{i=1}^{J} \beta_i \sum_{t=1}^{N} \sqrt{\frac{2l+1}{N}} P'_t(2\Phi(X_t) - 1)2\phi(X_t) \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \right) \right. \\
&\quad \cdot \left. \left( \sum_{m=1}^{J} \beta_m \sum_{t'=1}^{N} \sqrt{\frac{2m+1}{N}} P'_t(2\Phi(X_{t'}) - 1)2\phi(X_{t'}) \sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_i} Z_{st'} \right) \right\} \\
&= \frac{1}{N} \sum_{t=1}^{N} \mathcal{E} \left\{ \left( \sum_{i=1}^{J} \beta_i \sqrt{2l+1} P'_t(2\Phi(X_t) - 1)2\phi(X_t) \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \right) \left( \sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_i} Z_{st'} \right) \right\} \\
&\quad + \frac{1}{N} \sum_{t \neq u} \mathcal{E} \left\{ f(X_t, X_u) \left( \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \right) \left( \sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_i} Z_{st'} \right) \right\},
\end{align*}
\]

where \(X_t = \alpha^T Z_t, \ Z_t = (Z_{1t}, \ldots, Z_{pt}), \ f(\cdot, \cdot) \) is some bounded continuous function of \(X_t\) and \(X_u.\)
Note that for $t \neq v$, $X_t$ is independent of $X_v$, so we have

$$
\mathcal{E}\left\{ f(X_t, X_v) \left( \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \right) \left( \sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_j} Z_{s'v} \right) \right\}
= \mathcal{E}\left\{ f(X_t, X_v) \mathcal{E}\left\{ (\sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st}) (\sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_j} Z_{s'v}) \mid X_t, X_u \right\} \right\}
= \mathcal{E}\left\{ f(X_t, X_v) \mathcal{E}(\sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st} \mid X_t) \mathcal{E}(\sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_j} Z_{s'v} \mid X_v) \right\}
= \mathcal{E} f(X_t, X_v) X_t X_v \left( \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} \alpha_s \right) \left( \sum_{s'=1}^{p} \frac{\partial \alpha_{s'}}{\partial \theta_j} \alpha_{s'} \right)
= 0 \quad \text{(by (3.17)).} \quad (B.16)
$$

Similarly, for $t = 1, \ldots, N$,

$$
\mathcal{E}\left\{ (\sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} Z_{st}) (\sum_{v=1}^{p} \frac{\partial \alpha_v}{\partial \theta_j} Z_{vt}) \mid X_t \right\}
= \mathcal{E}\left\{ \sum_{s=1}^{p} \left( \frac{\partial \alpha_s}{\partial \theta_i} \frac{\partial \alpha_s}{\partial \theta_j} \right) Z_{st}^2 + \sum_{s \neq v} \frac{\partial \alpha_s}{\partial \theta_i} \frac{\partial \alpha_v}{\partial \theta_j} Z_{st} Z_{vt} \mid X_t \right\}
= \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} \frac{\partial \alpha_s}{\partial \theta_j} \left( 1 - \alpha_s^2 + \alpha_s^2 X_t^2 \right) + \sum_{s \neq v} \frac{\partial \alpha_s}{\partial \theta_i} \frac{\partial \alpha_v}{\partial \theta_j} \left( -\alpha_s \alpha_v + \alpha_s \alpha_v X_t^2 \right)
= \sum_{s=1}^{p} \frac{\partial \alpha_s}{\partial \theta_i} \frac{\partial \alpha_s}{\partial \theta_j} \left\{ \begin{array}{ll}
\sum_{s=1}^{p} \left( \frac{\partial \alpha_s}{\partial \theta_i} \right)^2 & \text{if } i = j \\
0 & \text{otherwise}
\end{array} \right. \quad \text{(by (3.17), (3.18)).} \quad (B.17)
$$

Therefore, combining (B.15)-(B.17), we have that

$$
\frac{\partial^2 r(s, t)}{\partial \theta_i \partial \theta_j} \bigg|_{\theta = \theta', \varphi_j = \varphi'_j} = C(\beta) \sum_{s=1}^{p} \left( \frac{\partial \alpha_s}{\partial \theta_i} \right)^2 \delta_{ij}
$$

where $C(\beta)$ is defined in (3.15).

Therefore, (3.12) is proved.

Proof of Theorem 3.2:
The calculation involved in this proof is enormous. The principle of the proof is as follows.

To get (3.32), just patiently manipulate the complicated expression for \( \kappa_2 \) in Weyl's paper and apply the definition of scalar curvature as in Definition B.5 (see Remark B.5).

To get (3.34) and (3.35), use the following fact which follows from integration by parts: for \( p \geq 5 \),

\[
\mathcal{E} \left\{ C(\beta)^{\frac{p+1}{2}} \left[ \sum_{m=1}^{J-1} \frac{\partial^2 C(\beta)}{\partial \varphi_m^2} \left( \sum_{s=1}^{p} (\frac{\partial \beta_s}{\partial \varphi_m})^2 \right)^{-1} \right] \right. \\
+ \sum_{m=1}^{J-2} \left( (J - 1 - m) \cot \varphi_m \frac{\partial C(\beta)}{\partial \varphi_m} \left( \sum_{s=1}^{p} (\frac{\partial \beta_s}{\partial \varphi_m})^2 \right)^{-1} \right) \right\} \\
= -\frac{p-3}{2} \mathcal{E} \left\{ C(\beta)^{\frac{p+1}{2}} \sum_{m=1}^{J-1} \left| \frac{\partial C(\beta)}{\partial \varphi_m} \left( \sum_{s=1}^{p} (\frac{\partial \beta_s}{\partial \varphi_m})^2 \right)^{-1} \right| \right\} .
\]

\[ \square \]

**Remark B.2:** There is also a relatively less complicated approach to prove (3.32) (the geometrical meaning of \( \kappa_2 \)) by generalizing one crucial step in Weyl's proof of his formula. We shall not introduce it here, since it involves more geometrical definitions than those introduced in this Appendix.
Appendix C

Supplementary Tables for Chapter 4

In this appendix, tables about more experiments on comparisons of $G$ and $\hat{F}_M$ in Chapter 4 are reported. The following notations are used.

- The column labeled "phji" gives $G(\hat{c}_M)$ for various $\hat{c}_M$'s corresponding to different $\alpha$ for the case $p = h, J = i$. Here the PP algorithm is the second modified version of JHF's with zero number of structure removals, i.e. optimizer is ZXMWD. One term approximation formula is used.

- The column labeled "phji.k" gives $G(\hat{c}_M)$ for the case $p = h, J = i$. Here the index is only composed of the term $k$: $\tilde{Y}_k(\alpha)$, i.e. the index $I_J(\alpha) = \frac{1}{2N} (\tilde{Y}_k(\alpha))^2$. The algorithm is the same as one for "phji". One term approximation formula is used.
- The column labeled “phji.kl” gives $G(\hat{c}_M)$ for the case $p = h, J = i$. Here the index is composed of $l - k + 1$ terms: $\hat{Y}_k(\alpha), \ldots, \hat{Y}_i(\alpha)$, i.e. the index $I_J(\alpha) = \frac{1}{2N} \sum_{j=k}^{l} (\hat{Y}_j(\alpha))^2$. The algorithm is the same as the one for “phji”. One term approximation formula is used.

- The column labeled “phji.kl.$\kappa_2$” gives $G(\hat{c}_M)$ for the case same as “phji.kl”, but two term approximation rather than only one term approximation is used. Here, the algorithm is the same as one for “phji”. Similar meanings are given to “phji.k.$\kappa_2$” and “phji.$\kappa_2$” etc.

- The column labeled “phjim.kl.$\kappa_2$” is the same as the column labeled by “phji.kl.$\kappa_2$”, but $h$ times structure removals are added. Similar meanings are given to “phjim”, “phjim.kl”, …, “nlphjim.$\kappa_2$”, “rphjim”, “spjim” etc. (See the definition for “nlphji.$\kappa_2$”, “rphji”, “spjhi” as follows.)

- The column labeled “nlphji.$\kappa_2$” is the same as the column labeled by “phji.$\kappa_2$”, but the PP algorithm is the first modified version of JHF’s with zero number of structure removals, i.e. it is a combination of step 1, 2, 5 of JHF algorithm and NPSOL.

- The column labeled “nlrphjim.$\kappa_2$” is the same as the column labeled by “nlphjim.$\kappa_2$”, but the rotation technique is added, i.e. the PP algorithm is the third modified version of JHF’s where we use the combination of the step 1, 2, 4, 5 of JHF PP algorithm and NPSOL and rotation technique. Similar meanings are given to “rphjim”, “rphjim.$\kappa_2$” etc.

- The column labeled “spjhi” gives $G(\hat{c}_M)$ as does “phji”, but we use sphered data for $\hat{F}_M$, (2.30) for $G$. Here the PP algorithm is the same as one used
in "phjii". Similar meanings are given to "snlphjim.kl.κ2", "snlphjim.kl.κ2", "snlphjii" etc.

(See the modified algorithms suggested in section 4.2)

In the following tables, sample size $N$ is 100 and the simulation size $M$ is 1000, if there is no specification.

In Table C.1, where $p = 2, J = 1$, we do see that the first and third column are very close, taking into account the given standard error (se), viz. $α$ and the corresponding $G(\hat{c}_M)$ are close for the reasonable $α$'s for the case of "p2j1.1". Columns 4 - 7 are all close to the first column, too. Notice that in this case, the optimizer ZXMD from IMSL is good enough and the grid search is possible and fast. In fact, when the data dimension is only 2, it is a optimization problem for one dimensional function, as there is only one constraint for the parameters.

In Table C.2, where $p = 2, J = 2$, the third to the sixth column are reasonably close to first column $α$. The fourth column "p2j2.12.κ2" and the sixth column "p2j2.12.κ2" also indicate that there is some improvement by using the second term with coefficient $κ_2$ in the approximation formula (2.29).

In Table C.3, where $p = 2, J = 3$, we see that the improvement is great by adding the second term approximation.

In Table C.4, where $p = 3, J = 1, 2$, the seventh column is closer to the first column than to the sixth column, i.e. some improvement is made by the second order approximation.

In Table C.5, where $p = 4, J = 4$(part II), we see that introducing rotation technique makes simulation result much better, and adding the second term approximation makes approximation formula better.
In Table C.6, where $p = 4, J = 4$ (part III), and Table C.7-C.9, where $p = 6, 8, J = 4, 6$, we experiment with NPSOL, rotation technique and structure removals.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>p2j1.1</th>
<th>p2j1.1</th>
<th>p2j1.2</th>
<th>p2j1.3</th>
<th>p2j1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.19817</td>
<td>0.21152</td>
<td>0.19799</td>
<td>0.19219</td>
<td>0.23301</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.15591</td>
<td>0.14666</td>
<td>0.15426</td>
<td>0.13568</td>
<td>0.17829</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.10020</td>
<td>0.094671</td>
<td>0.10390</td>
<td>0.087870</td>
<td>0.12724</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.046396</td>
<td>0.052505</td>
<td>0.049113</td>
<td>0.038821</td>
<td>0.053218</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.029973</td>
<td>0.037970</td>
<td>0.026718</td>
<td>0.022630</td>
<td>0.024764</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.019005</td>
<td>0.023320</td>
<td>0.019988</td>
<td>0.015875</td>
<td>0.014382</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.013273</td>
<td>0.010122</td>
<td>0.011071</td>
<td>0.0085336</td>
<td>0.0073942</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0022305</td>
<td>0.0088031</td>
<td>0.0059561</td>
<td>0.0057066</td>
<td>0.0040407</td>
<td>0.0046870</td>
</tr>
</tbody>
</table>

<p>| Table C.1: Case: $p = 2, J = 1$ |</p>
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>$p2j2.12$</th>
<th>$p2j2.12.\kappa_2$</th>
<th>$p2j2.12$</th>
<th>$p2j2.12.\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.22610</td>
<td>0.19755</td>
<td>0.23106</td>
<td>0.20170</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.15570</td>
<td>0.13797</td>
<td>0.17512</td>
<td>0.15455</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.10572</td>
<td>0.094779</td>
<td>0.11926</td>
<td>0.10656</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.050151</td>
<td>0.045726</td>
<td>0.053124</td>
<td>0.048384</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.037345</td>
<td>0.034229</td>
<td>0.036502</td>
<td>0.033470</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.025069</td>
<td>0.023119</td>
<td>0.019641</td>
<td>0.018175</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.013462</td>
<td>0.012515</td>
<td>0.017964</td>
<td>0.016642</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0047091</td>
<td>0.0044232</td>
<td>0.0037756</td>
<td>0.0035528</td>
</tr>
</tbody>
</table>

Table C.2: Case: $p = 2, J = 2$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>$p2j4$</th>
<th>$p2j4.\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.32416</td>
<td>0.24600</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.24163</td>
<td>0.18674</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.15672</td>
<td>0.12387</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.092905</td>
<td>0.075057</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.044546</td>
<td>0.036871</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.028245</td>
<td>0.023666</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.012911</td>
<td>0.011008</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0048872</td>
<td>0.0042390</td>
</tr>
</tbody>
</table>

Table C.3: Case: $p = 2, J = 4$
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>p3j3.1</th>
<th>p3j3.3</th>
<th>p3j3.3</th>
<th>p3j2.12</th>
<th>p3j2.12.$\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.19356</td>
<td>0.25588</td>
<td>0.22548</td>
<td>0.30982</td>
<td>0.23553</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.14171</td>
<td>0.17901</td>
<td>0.17152</td>
<td>0.21929</td>
<td>0.17073</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.099985</td>
<td>0.12309</td>
<td>0.10690</td>
<td>0.14808</td>
<td>0.11787</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.053159</td>
<td>0.062919</td>
<td>0.056579</td>
<td>0.075992</td>
<td>0.062284</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.034321</td>
<td>0.035080</td>
<td>0.034335</td>
<td>0.037345</td>
<td>0.031335</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.026126</td>
<td>0.016441</td>
<td>0.025829</td>
<td>0.025788</td>
<td>0.021856</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.012030</td>
<td>0.010409</td>
<td>0.013593</td>
<td>0.0092172</td>
<td>0.0079889</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.010893</td>
<td>0.0080194</td>
<td>0.0059822</td>
<td>0.0043352</td>
<td>0.0038058</td>
</tr>
</tbody>
</table>

Table C.4: Case: $p = 3, J = 1, 2$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>p4j4</th>
<th>p4j4.$\kappa_2$</th>
<th>p4j4</th>
<th>p4j4.$\kappa_2$</th>
<th>p4j4.$\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.55163</td>
<td>0.27317</td>
<td>0.58680</td>
<td>0.28826</td>
<td>0.22535</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.37524</td>
<td>0.19453</td>
<td>0.43114</td>
<td>0.22002</td>
<td>0.15877</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.23749</td>
<td>0.12897</td>
<td>0.24984</td>
<td>0.13502</td>
<td>0.093526</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.098676</td>
<td>0.057551</td>
<td>0.11576</td>
<td>0.066732</td>
<td>0.044025</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.050577</td>
<td>0.030796</td>
<td>0.061638</td>
<td>0.037083</td>
<td>0.022943</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.022395</td>
<td>0.014242</td>
<td>0.035391</td>
<td>0.021986</td>
<td>0.0011389</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.012869</td>
<td>0.0083923</td>
<td>0.016176</td>
<td>0.010439</td>
<td>0.0070589</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0036135</td>
<td>0.0024701</td>
<td>0.0099442</td>
<td>0.0065527</td>
<td>0.0019381</td>
</tr>
</tbody>
</table>

Table C.5: Case: $p = 4, J = 4$, part II

175
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>rp4j4m</th>
<th>rp4j4m, $\kappa_2$</th>
<th>nlp4j4m, $\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.41965</td>
<td>0.21483</td>
<td>0.22151</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.29061</td>
<td>0.15476</td>
<td>0.15193</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.18438</td>
<td>0.10242</td>
<td>0.085796</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.078323</td>
<td>0.046409</td>
<td>0.039975</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.039777</td>
<td>0.024557</td>
<td>0.026339</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.023127</td>
<td>0.014683</td>
<td>0.014441</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.011616</td>
<td>0.0076043</td>
<td>0.0063769</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0052513</td>
<td>0.0035458</td>
<td>0.0039892</td>
</tr>
</tbody>
</table>

Table C.6: Case: $p = 4, J = 4$, part III

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>p6j4</th>
<th>p6j4, $\kappa_2$</th>
<th>p6j4</th>
<th>p6j4, $\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>1.1391</td>
<td>0.29628</td>
<td>1.0505</td>
<td>0.27908</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.70983</td>
<td>0.20690</td>
<td>0.68338</td>
<td>0.20078</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.45047</td>
<td>0.14364</td>
<td>0.41260</td>
<td>0.13362</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.14270</td>
<td>0.054008</td>
<td>0.16099</td>
<td>0.060031</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.073923</td>
<td>0.030183</td>
<td>0.096319</td>
<td>0.038214</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.046380</td>
<td>0.019829</td>
<td>0.070063</td>
<td>0.028744</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.019302</td>
<td>0.0088917</td>
<td>0.046051</td>
<td>0.019662</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0089529</td>
<td>0.0043529</td>
<td>0.032170</td>
<td>0.014174</td>
</tr>
</tbody>
</table>

Table C.7: Case: $p = 6, J = 4$, part I
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>nlp6j4.(\kappa_2)</th>
<th>nlp6j4rm.(\kappa_2)</th>
<th>rp6j4m.(\kappa_2)</th>
<th>rp6j4.(\kappa_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.22714</td>
<td>0.22501</td>
<td>0.18567</td>
<td>0.20419</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.16645</td>
<td>0.16541</td>
<td>0.14104</td>
<td>0.14871</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.11307</td>
<td>0.10180</td>
<td>0.085677</td>
<td>0.095394</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.051253</td>
<td>0.042604</td>
<td>0.043456</td>
<td>0.046006</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.022819</td>
<td>0.019325</td>
<td>0.021126</td>
<td>0.021269</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.011194</td>
<td>0.014257</td>
<td>0.015945</td>
<td>0.016115</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.0047160</td>
<td>0.0073226</td>
<td>0.0033184</td>
<td>0.0033184</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0022565</td>
<td>0.0018948</td>
<td>0.0012810</td>
<td>0.0012810</td>
</tr>
</tbody>
</table>

Table C.8: Case: $p = 6, J = 4$, part II

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>p6j6.(\kappa_2)</th>
<th>p6j6.(\kappa_2)</th>
<th>p8j4</th>
<th>p8j4.(\kappa_2)</th>
<th>nlp8j4rm.(\kappa_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.33202</td>
<td>0.36780</td>
<td>5.5178</td>
<td>-0.21168</td>
<td>0.088340</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.24059</td>
<td>0.28524</td>
<td>3.6820</td>
<td>-0.018281</td>
<td>0.087899</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.15065</td>
<td>0.19666</td>
<td>1.9749</td>
<td>0.082641</td>
<td>0.068104</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.072096</td>
<td>0.072591</td>
<td>0.55025</td>
<td>0.068404</td>
<td>0.031751</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.041461</td>
<td>0.031892</td>
<td>0.20313</td>
<td>0.036367</td>
<td>0.020646</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.027382</td>
<td>0.024468</td>
<td>0.11571</td>
<td>0.023551</td>
<td>0.016513</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.016110</td>
<td>0.013415</td>
<td>0.034318</td>
<td>0.0088654</td>
<td>0.0077084</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0040131</td>
<td>0.0069271</td>
<td>0.020405</td>
<td>0.0054848</td>
<td>0.0039645</td>
</tr>
</tbody>
</table>

Table C.9: Case: $p = 8, J = 4$
In Table C.10, the two term approximation formula (2.29) for unsphered data is used for the sphered data. We see that even we used two terms approximation, the results of using (2.29) for sphered data case are so bad. These are examples of which the large sample theory for sphered data does not work in the same way as for the unsphered data.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>sp4j4</th>
<th>sp4j4.(\kappa_2)</th>
<th>sp6j4</th>
<th>sp6j4.(\kappa_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>2.1700</td>
<td>0.83800</td>
<td>4.8753</td>
<td>0.67954</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>1.4231</td>
<td>0.60520</td>
<td>3.4168</td>
<td>0.59113</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.99334</td>
<td>0.45129</td>
<td>2.1751</td>
<td>0.46016</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.51237</td>
<td>0.25611</td>
<td>1.1391</td>
<td>0.29628</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.28671</td>
<td>0.15289</td>
<td>0.70553</td>
<td>0.20591</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.16789</td>
<td>0.094004</td>
<td>0.49167</td>
<td>0.15426</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.11521</td>
<td>0.066438</td>
<td>0.30210</td>
<td>0.10309</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.065701</td>
<td>0.039370</td>
<td>0.17388</td>
<td>0.064210</td>
</tr>
</tbody>
</table>

Table C.10: Sphered Case: $p = 4, 6, J = 4, \text{(bad)}$

Table C.11 confirms that our treatment for sphered data case is reasonable. See Section 4.3.

Table C.12-C.14 show that our treatment, of deleting first two terms in sphered data case, is good.

Under this treatment and a reasonably good PP algorithm, the two term approximation formula and simulation formula match each other.
<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \bar{Y}_1(\alpha) )</th>
<th>( \bar{Y}_2(\alpha) )</th>
<th>( \bar{Y}_3(\alpha) )</th>
<th>( \bar{Y}_4(\alpha) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((1,0,0,0))</td>
<td>0.04688</td>
<td>-0.00466</td>
<td>0.0459</td>
<td>-0.0069</td>
</tr>
<tr>
<td>(\alpha = (0, \frac{1}{2}, \frac{\sqrt{3}}{2}, 0))</td>
<td>0.0459</td>
<td>-0.0069</td>
<td>0.044556</td>
<td>-0.003259</td>
</tr>
<tr>
<td>(\alpha = (0, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}))</td>
<td>0.1017</td>
<td>-0.028</td>
<td>1.0177</td>
<td>0.029</td>
</tr>
<tr>
<td>(\bar{Y}_3(\alpha))</td>
<td>0.95196</td>
<td>-0.026</td>
<td>0.964813</td>
<td>0.0339</td>
</tr>
<tr>
<td>(\bar{Y}_4(\alpha))</td>
<td>0.9696</td>
<td>0.0015</td>
<td>1.01349</td>
<td>-0.012</td>
</tr>
</tbody>
</table>

Table C.11: \( \bar{Y}_i(\alpha) \) in Centered Case

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>se</th>
<th>sp4j4.34</th>
<th>sp4j4.34.( \kappa_2 )</th>
<th>srp4j4.34.( \kappa_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.48013</td>
<td>0.32316</td>
<td>0.20281</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.29481</td>
<td>0.20476</td>
<td>0.16284</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.19539</td>
<td>0.13880</td>
<td>0.10969</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.092371</td>
<td>0.067902</td>
<td>0.057236</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.048237</td>
<td>0.036329</td>
<td>0.032236</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.026632</td>
<td>0.020441</td>
<td>0.026913</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.017579</td>
<td>0.013650</td>
<td>0.017020</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0094914</td>
<td>0.0074827</td>
<td>0.010924</td>
</tr>
</tbody>
</table>

Table C.12: Sphered Case: \( p = 4, J = 4 \)
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>$\text{sp6j4.34}$</th>
<th>$\text{sp6j4.34.} \kappa_2$</th>
<th>$\text{snlp6j4m.34}$</th>
<th>$\text{snlp6j4m.34.} \kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>1.0172</td>
<td>0.44994</td>
<td>0.62043</td>
<td>0.29149</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.68503</td>
<td>0.31825</td>
<td>0.40942</td>
<td>0.20089</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.41576</td>
<td>0.20370</td>
<td>0.25420</td>
<td>0.13028</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.20433</td>
<td>0.10659</td>
<td>0.12531</td>
<td>0.067809</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.12112</td>
<td>0.065696</td>
<td>0.077808</td>
<td>0.043483</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.081799</td>
<td>0.045556</td>
<td>0.043374</td>
<td>0.025071</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.048246</td>
<td>0.027713</td>
<td>0.021951</td>
<td>0.013137</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.026588</td>
<td>0.015752</td>
<td>0.016532</td>
<td>0.010051</td>
</tr>
</tbody>
</table>

Table C.13: Sphered Case: $p = 6, J = 4$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>se</th>
<th>$\text{srp6j4.34.} \kappa_2$</th>
<th>$\text{srp6j4m.34.} \kappa_2$</th>
<th>$\text{snlp8j4m.34.} \kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.012649</td>
<td>0.31739</td>
<td>0.22567</td>
<td>0.31636</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01129</td>
<td>0.21904</td>
<td>0.17038</td>
<td>0.23873</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0094868</td>
<td>0.15375</td>
<td>0.12671</td>
<td>0.15904</td>
</tr>
<tr>
<td>0.05</td>
<td>0.006892</td>
<td>0.065175</td>
<td>0.056836</td>
<td>0.084404</td>
</tr>
<tr>
<td>0.03</td>
<td>0.005394</td>
<td>0.040843</td>
<td>0.038122</td>
<td>0.057191</td>
</tr>
<tr>
<td>0.02</td>
<td>0.004427</td>
<td>0.023775</td>
<td>0.021173</td>
<td>0.048333</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0031464</td>
<td>0.012030</td>
<td>0.010255</td>
<td>0.029524</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0020305</td>
<td>0.0053158</td>
<td>0.0053152</td>
<td>0.015857</td>
</tr>
</tbody>
</table>

Table C.14: Sphered Case: $p = 6, 8, J = 4$
Bibliography


**P-VALUES IN PROJECTION PURSUIT**

**Jiayang Sun**

**Department of Statistics - Sequoia Hall**  
Stanford University  
Stanford, California 94305-4065

**Air Force Office of Scientific Research**  
Mathematical & Information Sciences Division  
Bolling Air Force Base  
Washington, DC 20352

**August 1989**

**Approved for public release; distribution unlimited.**

**NA**

**UNCLASSIFIED**

(See reverse side for abstract.)
20. ABSTRACT.

Projection Pursuit (PP) technology is a key for exploring the “non-linear” structure of high dimensional data. PP algorithms make it possible to find interesting directions so that the structure in data can be seen clearly after projection on these directions. It is useful to have a significance test to help us decide whether apparent structure is real or just the effect of noise. Monte Carlo methods can be helpful to achieve this goal; but unfortunately in this case, they are difficult to use and computationally expensive.

My work is to look for analytical methods for calculating the P-value associated with various projection pursuit indices, especially, the index suggested by J. Friedman (1987). In this thesis, under the null hypothesis that the data are an independent, identically distributed, sample from a p-dimensional normal population, a theoretical approximation formula for the P-value is derived using Weyl’s formula (1939) for the volume of a tube about a manifold imbedded in the unit sphere in Euclidean space. Weyl’s formula involves some complicated constants, for which applicable formulas and a table of numerical values are given. The result of Monte Carlo simulations is compared with our analytical result. The comparisons show that special care is needed when the number of dimension is large.

In addition to its contribution to the problem of projection pursuit, this work gives a general two term approximation for the tail probability of the extreme of a class of differentiable Gaussian random fields and illustrates some of the potential for using Weyl’s formula in Probability and Statistics. In the particular case of Friedman’s PP index, the matrix of the metric tensor turns out to be diagonal and hence it is possible to calculate Weyl’s second coefficient, which involves the total scalar curvature of the manifold, and in general is extremely complicated. The numerical results show that this term can improve the quality of the approximation enormously, and that use of only the first term, which involves the comparatively easily computed volume of the manifold, is inadequate in statistically interesting cases.