Karhunen-Loeve Expansions and their Applications

LIMIN WANG

Submitted for the Degree of
DOCTOR OF PHILOSOPHY
at
The London School of Economics and Political Science

March 16, 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Table of Contents

Declaration

I certify that this thesis presented for examination for a Ph.D degree of the London School of Economics and Political Science is solely my own work other than where it is clearly indicated that it is the work of others.

Acknowledgement

I would like to express my deep and sincere gratitude to my supervisor Professor Henry Wynn first. I would not imagine having finished my thesis without the support from him. In the last three years and a half, his knowledge in statistics and enthusiasm towards the research have left a very deep impression. It was the fruitful and frequent conversation with him that guided me in the right direction of my thesis. Besides his help in academia, his encouragement and unlimited support to me as an overseas student in the U.K. were also of great importance.

I would like to express my thanks to my department, the department of statistics at the London School of Economics and Political Science, which provided a conducive environment. It was a great pleasure and honour for me to stay in this world class department since my Masters. In addition to my supervisor, Professor Henry Wynn, I would also like to express my thanks to Dr Jeremy Penzer, my Masters tutor; Dr Pauline Barrieu, my supervisor of my Masters dissertation; and Dr Angelos Dassios, my second supervisor for my PhD. All of these individuals have provided support in my research. The research student scholarship I received from the department, and the Overseas Research Students Awards Scheme (ORSAS) set up by the Secretary of State for Education and Science, made it financially possible for me to complete my thesis.

My PhD and research fellows in the department have all given me the feeling of being at home when pursuing my research. Particularly, I owe my great gratitude to Billy Wu, who shared his valuable and constructive opinions with me; Pauline Sculli, who helped me with my language; Young Lee, who discussed with me the financial application; Noha Youssef, who provided me the code for the exchange algorithm in Chapter 3; and Hugo Maruri-Aguilar, who recommended several papers on my research.

My warmest thanks to my friends outside the department. In the U.K., I wish to thank Chin Yau Siah, who read and made detailed comments on the first draft of my thesis; and Michael Cheng, who encouraged me on my work. For my friends outside the U.K., I would like to express my heartfelt gratitude to Chor Ee Tan, who provides me the motivations and inspirations since my Master's course.

My special thanks are also to my beloved father, Dingxian Wang and mother Jinfen Luan. It is to them that I dedicate this work.

Abstract

The Karhunen-Loeve Expansion (K-L expansion) is a bi-orthogonal stochastic process expansion. In the field of stochastic process, the Karhunen-Loeve expansion decomposes the process into a series of orthogonal functions with the random coefficients. The essential idea of the expansion is to solve the Fredholm integral equation, associated with the covariance kernel of the process, which defines a Reproducing Kernel Hilbert Space (RKHS). This either has an analytical solution or special numerical methods are needed.

This thesis applies the Karhunen-Loeve expansion to some fields of statistics.

The first two chapters review the theoretical background of the Karhunen-Loeve expansion and introduce the numerical methods, including the integral method and the expansion method, when the analytical solution to the expansion is unavailable. Chapter 3 applies the theory of the Karhunen-Loeve expansion to the field of the design experiment using a criteria called "maximum entropy sampling". Under such setting, a type of duality is set up between maximum entropy sampling and the D-optimal design of the classical optimal design. Chapter 4 uses the Karhunen-Loeve expansion to calculate the conditional mean and variance for a given set of observations, with application to prediction. Chapter 5 extends the theory of the Karhunen-Loeve expansion from the univariate setting to the multivariate setting: multivariate space, univariate time. Adaptations of numerical methods of Chapter 2 are also provided for the multivariate setting, with a full matrix development. Chapter 6 applies the numerical method developed in Chapter 5 to the emerging area of multivariate functional data analysis with a detailed example on a trivariate autoregressive process.

Introduction

One of the fundamental methods with a wide range of scientific applications, such as solving partial differential equations, signal processing and option pricing, is Fourier analysis. It is a decomposition of a real function into an infinite linear combination of orthogonal basis terms, usually the trigonometric basis functions, together with the Fourier coefficients. In practical applications , it is usual to use only a finite number of terms, say p. Moreover, when p increases to infinity, it can also be shown that under certain conditions the mean squared error using the finite representation in the space of the deterministic function converges to zero.

The Karhunen-Loeve expansion can be regarded as an extension of Fourier analysis from deterministic functions to stochastic processes. The Karhunen-Loeve expansion is a representation, in which the process is decomposed into a series of orthogonal functions, analogous to Fourier analysis. The process of finding the coefficients is similar to that of Fourier analysis, which is to minimise the mean squared error of the finite representation. The minimisation process involves solving an integral equation, the Fredholm integral equation. Using a finite representation of the Karhunen-Loeve expansion (the truncated Karhunen-Loeve expansion), when the order, p, of the orthogonal functions increases to infinity, the mean squared error decreases to zero in the space of the stochastic process. Whereas, for Fourier analysis, the coefficients are

deterministic and the orthogonal functions are usually the trigonometric functions, the coefficients for the Karhunen-Loeve expansion are random and the orthogonal functions are derived from solving a Fredholm integral equation related to the covariance function of the process.

Early work on the Karhunen-Loeve expansion includes that of Karhunen (1947), Loeve (1948) and Trees (1968). More recent research includes (i) contributions to its numerical methods (see, for example, Boente and Fraiman (2000), Phoon et al. (2002b), Phoon et al. (2002a)), (ii) finite element methods (see, for example, Ghanem and Spanos (1991)), (iii) model reduction using the Karhunen-Loeve expansion (see, for example, Newman (1996b), Newman (1996a), Glavaski et al. (1998)), (iv) functional data analysis (see, for example, Ramsay and Silverman (1997), Ramsay and Silverman (2002)), (v) application in the field of finance (see, for example, Vargiolu (1998), Cont and Fonseca (2001), Schmidt (2004)), (vi) application in pattern recognition (see, for example, Kirby and Sirovich (1990)), and (vii) application in machine learning (see, for example, Rasmussen and Williams (2005)).

This thesis applies the Karhunen-Loeve expansions to some fields of statistics. Chapter 1 reviews the background of the Karhunen-Loeve expansion. It starts from a brief introduction to the Reproducing Kernel Hilbert Space (RKHS), which is isometrically isomorphic to the space our process lies in. It also provides a key theorem for the Karhunen-Loeve expansion and explains the optimality of the expansion in terms of the mean squared error. For examples such as the Browninan motion, the Brownian bridge, the Ornstein-Uhlenbeck process and the integrated Brownian motion, it is shown how to find the Karhunen-Loeve expansion analytically. The extension from the univariate Karhunen-Loeve expansion to the spatial Karhunen-Loeve expansion

is also discussed briefly.

Chapter 2 deals with the situation when the analytical solution to the Karhunen-Loeve expansion are unavailable. Numerical methods, instead, play an essential role in finding the expansion. The integral method and the expansion method are explained in detail with examples of both the Ornstein-Uhlenbeck process and the Gaussian process with squared exponential kernel. Numerical methods for the spatial Karhunen-Loeve expansion is explained briefly with an example provided on the Brownian sheet.

Chapter 3 applies the theory of the Karhunen-Loeve expansion to the field of the design experiment using a criteria called "maximum entropy sampling". Under such setting, a type of duality can be set up between maximum entropy sampling and the D-optimal design from the classical optimal design. Two kinds of algorithms, which are the "greedy exchange algorithm" and "DETMAX", are served to check the duality numerically.

Chapter 4 uses the Karhunen-Loeve expansion to calculate the conditional expectation and the conditional variance for the conditional on a set of observations. The behaviour of the generalised mean squared error for using the conditional expectation as the prediction is studied in detail. This chapter also develops an alternative way of treating conditional data, which is to calculate a "conditional Karhunen-Loeve expansion". Examples for both the Markovian process and the non-Markovian process are provided for the conditional Karhunen-Loeve expansion.

Chapter 5 extends the theory of the Karhunen-Loeve expansion from the univariate setting to the multivariate setting. "Multivariate", in this chapter, refers to multivariate state, but univariate time. For multivariate process satisfying certain

condition, the multivariate Karhunen-Loeve expansion is closely connected to the univariate Karhunen-Loeve expansion. Adaptations of multivariate numerical methods are also provided for this setting.

Chapter 6 applies the numerical method developed in chapter 5 in the field of multivariate functional data analysis. An example on trivariate $AR(1)$ process is studied in detail for the decomposition of the process, the reconstruction of the process, and the smoothing and the prediction of the process using the knowledge of both the covariance and the cross-covariance function under the multivariate Karhunen-Loeve expansion. The theory from chapter 5 proves useful in pointing to appropriate numerical methods.

Chapter 1

Introduction to the Karhunen-Loeve Expansion

The Karhunen-Loeve Expansion (K-L expansion) is a bi-orthogonal stochastic process expansion. It was derived and investigated by a number of researchers (see, for example, Karhunen (1947), Loeve (1948), Ghanem and Spanos (1991)). The essential idea is to solve the Fredholm integral equation, associated with the covariance kernel of the process, which defines a Reproducing Kernel Hilbert Space (RKHS). This gives either the analytical or the numerical expressions for the kernel's eigenvalue and eigenfunction. In this chapter, section 1 introduces the RKHS of the kernel and shows it is isometrically isomorphic to the space of our stochastic process. Section 2 presents the main theorem of the Karhunen-Loeve expansion and some properties of its eigenvalue and eigenfunction. Section 3 studies examples of some commonly used Gaussian processes. Section 4 extends the univariate Karhunen-Loeve expansion to the multivariate time, univariate state (spatial) Karhunen-Loeve expansion.

1.1 Reproducing Kernel Hilbert Space (RKHS)

A Reproducing Kernel Hilbert Space (RKHS) is a Hilbert Space of functions. It can be thought of as a space containing smoother function than the general Hilbert space. This section provides a brief introduction to it. See, for example, Aronszajn (1950), Kailath (1971), Wahba (1990), for details.

We start with a formal definition of the RKHS and present some main properties of this space. Then the RKHS is related to another commonly used subspace of the Hilbert Space, which is L_2.

Definition 1.1.1. A Hilbert Space is called a Reproducing Kernel Hilbert Space H, if for every function $f \in H$ defined on E, there exists a function $K : E \times E \rightarrow R$, such that

(1) For all y, as a function of x, $K(x, y) \in H$

(2) Reproducing property: for all $x \in E$

$$f(x) = (K(., x), f(.)) \tag{1.1.1}$$

Function $K(x, y)$ is called the reproducing kernel of the space H.

The reproducing kernel has the following properties, which make it possible for covariance functions to lie in this space.

Theorem 1.1.1. *If a reproducing kernel K exists, it is unique.*

Proof. Assume there exists another reproducing kernel K', then

$$\begin{aligned} 0 &\leq ||K(x,.) - K'(x,.)|| \\ &= (K(x,.) - K'(x,.), K(x,.) - K'(x,.)) \\ &= (K(x,.) - K'(x,.), K(x,.)) - (K(x,.) - K'(x,.), K'(x,.)) \\ &= (K(x,x) - K'(x,x)) - (K(x,x) - K'(x,x)) = 0 \end{aligned}$$

□

Theorem 1.1.2. *A reproducing kernel K is non-negative definite.*

Proof. For all $y_i \in E$, on which the kernel is defined, and $a_1, a_2, \cdots, a_n$

$$\begin{aligned} & \sum_i \sum_j K(y_i, y_j) a_i a_j \\ = & \sum_i \sum_j (K(., y_i), K(., y_j)) a_i a_j \\ = & (\sum_i K(., y_i) a_i, \sum_j K(., y_j) a_j) \\ = & ||\sum_i K(., y_i) a_i||^2 \geq 0 \end{aligned}$$

□

The following theorem is a key theorem in the RKHS theory. It states that the relationship between the RKHS and its corresponding kernel is one-to-one. See Aronszajn (1950) for detail.

Theorem 1.1.3. *(Moore-Aronszajn theorem): To every Reproducing Kernel Hilbert Space H, there exists a unique non-negative definite function $K(.,.)$. Conversely, for every non-negative definite function $K(.,.)$, there exists a unique Reproducing Kernel Hilbert Space H.*

Proof. In a Reproducing Kernel Hilbert Space H, uniqueness and non-negative definiteness of the kernel are guaranteed by theorem 1.1.1 and theorem 1.1.2 respectively. This proves the first part of the theorem. In order to prove the second part, the following two lemmas are required.

Lemma 1.1.4. *The reproducing kernel $K(.,x), x \in E$, on which the kernel is defined, spans a Reproducing Kernel Hilbert Space H, i.e.*

$$H = \text{Closure}(\text{lin}[K(.,x)]) = \text{Closure}\{\sum_i \lambda_i K(.,x_i), \lambda_i \in R\} \tag{1.1.2}$$

where $\text{lin}[K(.,x)]) = \{\sum_i \lambda_i K(.,x_i), \lambda_i \in R\}$

Proof. The zero vector is the only vector orthogonal to $K(.,x), \forall x \in E$, since

$$f(x) = (f(.), K(.,x)) = 0 \tag{1.1.3}$$

□

Lemma 1.1.5. *Norm convergence implies pointwise convergence in the Reproducing Kernel Hilbert Space H.*

Proof. Assume a Cauchy series $f_n \in H$ and $\forall x$

$$|f_n(x) - f(x)| = |(f_n(.) - f(.), K(.,x))| \leq ||f_n(.) - f(.)||||K(.,x)|| \tag{1.1.4}$$

The inequality above uses the Cauchy-Schwartz Inequality. Hence $||f_n(.) - f(.)|| \to 0$ implies $|f_n(x) - f(x)| \to 0$. □

Lemma 1.1.4 and 1.1.5 suggest a way of constructing the Reproducing Kernel Hilbert Space. This comprises all the linear combinations of the kernel function and the pointwise limit to all the sequences, to complete the space.

Assume $H_0 = (\text{lin})[K(.,x)]$ and $f(.) = \sum_i a_i K(.,x_i) \in H_0, g(.) = \sum_j b_j K(.,y_j) \in H_0$. The inner product is defined as

$$(f(.), g(.)) := \sum_i \sum_j a_i b_j K(x_i, y_j) \tag{1.1.5}$$

The above double summation is well defined since $\sum_i \sum_j a_i b_j K(x_i, y_j) \in H_0$. It meets the requirement of the inner product as well. Linearity and symmetry are satisfied because the double summation is a linear term. Hence it only needs to be shown that for all y, $(f(y), f(y)) = 0$, $f(y) = 0$. The reproducing property holds since,

$$(f(.), K(.y)) = \sum_i a_i K(x_i, y) = f(y) \tag{1.1.6}$$

Then, using Cauchy-Schwartz inequality, for all y, we obtain

$$0 \leq |f(y)| \leq ||f(.)||||K(.y)|| \tag{1.1.7}$$

Since $(f(.), f(.)) = 0$ implies $||f(.)|| = 0$, then $f(y) = 0$, for all y.

Finally, we complete this space by including all the limits of all the Cauchy sequences. Pointwise limit is enough, i.e. include all $f(y)$, for all y, such that for the Cauchy sequence $f_n(y) \in H_0$

$$|f_n(y) - f(y)| \to 0 \tag{1.1.8}$$

Hence the unique Reproducing Kernel Hilbert Space required is $H = \text{span}[K(.,x)]$.

□

The L_2 space is another typical subspace of the Hilbert space, but the L_2 space and the RKHS are not equal to each other. The L_2 space, in general, is a "rougher" space. It does not certainly contain functions that are pointwise convergent. However,

as is shown in lemma 1.1.5, norm convergent always implies pointwise convergent in the RKHS. The following example shows that norm convergent does not equal to pointwise convergent in L_2.

Example Consider the function in L_2

$$g_n(t) = \begin{cases} 1 & 0 < t < \frac{1}{n} \quad \text{if } n \text{ is odd} \\ -1 & 0 < t < \frac{1}{n} \quad \text{if } n \text{ is even} \\ 0 & \text{otherwise} \end{cases} \tag{1.1.9}$$

This is norm convergent to zero, since

$$||g_n - g|| = \int_0^{\frac{1}{n}} 1dt = \frac{1}{n} \to 0 \tag{1.1.10}$$

where $g(t) \equiv 0$, for all t. Hence g_n is norm convergent to g. However, when t approaches to zero, $g_n(t)$ oscillates between 1 and -1, which does not support pointwise convergent.

Another function, which belongs to the RKHS, but not the L_2 space is the Dirac delta function defined as

$$\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \tag{1.1.11}$$

The Dirac delta function has the property that $\int f(x)\delta(x)dx = f(0)$. From this property, it can be seen that $\delta(x)$ belongs to the RKHS, but not the L_2 space.

Under certain condition, the RKHS can be regarded as an embedded space of the L_2 space. The following theorem shows this result.

Theorem 1.1.6. *Assume $K(.,.)$ is an L_2 kernel, i.e. $\int\int K(s,t)^2 dsdt < \infty$. For all $f \in L_2$, assume that the Fourier coefficients for f is f_i, i.e.*

$$f_i = \int f(x)\phi_i(x)dx \tag{1.1.12}$$

where ϕ_i is defined in equation 1.1.15. Then $f \in RKHS$ if

$$||f||^2 = \sum_i \frac{f_i^2}{\lambda_i} < \infty \tag{1.1.13}$$

For all $f \in L_2$ and $g \in L_2$, the inner product in this RKHS is defined as

$$(f, g) = \sum_i \frac{f_i g_i}{\lambda_i} \ , \tag{1.1.14}$$

where f_i, g_i are the Fourier coefficients for f, g respectively. The series $\{\phi_i\}$ is an orthogonal series in L_2. Both λ_i and ϕ_i are from Mercer's theorem, i.e.

$$K(s,t) = \sum_i \lambda_i \phi_i(s) \phi_i(t) \tag{1.1.15}$$

(This equation for K can also be derived from the Karhunen-Loeve expansion, which will be introduced in the next section.)

Proof. The Fourier expansion for $f(.)$ and $K(x,.)$ can be written as

$$\begin{aligned} f(.) &= \sum_i f_i \phi_i(.) \\ K(x,.) &= \sum_i [\lambda_i \phi_i(x)] \phi_i(.) \end{aligned}$$

Since

$$\sum_i \frac{\lambda_i^2 \phi_i(x)^2}{\lambda_i} = \sum \lambda_i \phi_i(x)^2 = K(x,x) < \infty \tag{1.1.16}$$

the inner product $(f, K(x.))$ makes sense and its value is

$$(f, K(x,.)) = \sum_i \frac{f_i [\lambda_i \phi_i(x)]}{\lambda_i} = \sum_i f_i \phi_i(x) = f(x)$$

This means that $K(x.,)$ is a reproducing kernel and thus $f \in$ RKHS with norm $||f||^2 = \sum_i \frac{f_i^2}{\lambda_i}$. □

Remark 1.1.1. In the space of L_2, unless specified, the concept of orthogonal refers to orthonormal in this thesis. It means that if $\{f_i(t)\}$ is an orthogonal series in L_2,

$$\int f_i(t)f_j(t)dt = \delta_{ij} \tag{1.1.17}$$

Remark 1.1.2. For any function $f \in L_2$, only the condition $\sum_i f_i^2 < \infty$ is required. What the above theorem states is that if the function f also lies in the RKHS, an extra condition $\sum_i \frac{f_i^2}{\lambda_i} < \infty$ should be assumed, which might not be satisfied by all the L_2 functions.

1.2 The Karhunen-Loeve expansion

In the Karhunen-Loeve expansion, we represent a stochastic process $\{X(t), t \in \mathcal{T}\}$ via a sequence of independent simple random variables $\{\xi_i, i \in N\}$. Assume that the random process $X(t)$ is a zero mean second order process. Its corresponding kernel $K(s,t) = \text{cov}(X(t), X(s))$ is in the RKHS with its unique kernel function $K(.,.)$.

Assume that $\{\phi_i(t)\}$ is a series of orthogonal functions in L_2 derived from certain integral equations, the process can then be written down as

$$\begin{aligned} X(t) &= \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i\phi_i(t) \\ \xi_i &= \frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{T}} X(t)\phi_i(t)dt \end{aligned}$$

The basis function $\{\phi_i(t)\}$, together with the uncorrelated random coefficients $\{\xi_i\}$ constitute a bi-orthonormal system. That is, the basis functions are orthogonal as functions and ξ_i are independent in the sense that $E(\xi_i) = 0$ and $\text{cov}(\xi_i\xi_j) = \delta_{ij}$.

The following theorem provides a sufficient and necessary condition for the decomposition of the process.

Theorem 1.2.1. *Let $\{X(t), t \in \mathcal{T}\}$ be a zero mean second order process. Its covariance function is continuous and denoted as $K(s,t)$ for the covariance in between time s and time t.*

Let $\{\phi_i, i \in N\}$ be the set of orthogonal functions of L_2.

1. *Assume that λ_i and ϕ_i satisfies the following equation*

$$\int_{\mathcal{T}} K(s,t)\phi_i(t)dt = \lambda_i \phi_i(s) \tag{1.2.1}$$

where $\{\phi_i, i \in N\}$ and $\{\lambda_i, i \in N\}$ are called the eigenfunctions and eigenvalues respectively.

Furthermore, choose

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{T}} X(t)\phi_i(t)dt \tag{1.2.2}$$

Then,

$$X(t) = \lim_{p \to \infty} \sum_{i=1}^{p} \sqrt{\lambda_i}\xi_i\phi_i(t) \tag{1.2.3}$$

uniformly, in the sense that

$$E(X(t) - \sum_{i=1}^{p} \sqrt{\lambda_i}\xi_i\phi_i(t))^2 \to 0 \tag{1.2.4}$$

uniformly.

2. *Conversely, if $X(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i\phi_i(t)$, where $\{\xi_i\}$ is identically independent distributed (i.i.d.) with mean 0 and variance 1, then*

$$\int_{\mathcal{T}} K(s,t)\phi_i(t)dt = \lambda_i \phi_i(s) \tag{1.2.5}$$

Proof. (a) From the construction of ξ_i , it is easy to see that $E(\xi_i) = 0$ and $E(\xi_i\xi_j) = \delta_{ij}$. Since,

$$\begin{aligned} E(\xi_i) &= E(\frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{T}} X(t)\phi_i(t)dt) \\ &= \frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{T}} E(X(t))\phi_i(t)dt \\ &= 0 \end{aligned}$$

$$
\begin{aligned}
E(\xi_i\xi_j) &= E(\frac{1}{\sqrt{\lambda_i}}\frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{T}}\int_{\mathcal{T}} X(t)\phi_i(t)X(s)\phi_j(s)dtds) \\
&= \frac{1}{\sqrt{\lambda_i}}\frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{T}}\int_{\mathcal{T}} E(X(t)X(s))\phi_i(t)\phi_j(s)dtds \\
&= \frac{1}{\sqrt{\lambda_i}}\frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{T}}(\int_{\mathcal{T}} K(t,s)\phi_i(t)dt)\phi_j(s)ds \\
&= \frac{1}{\sqrt{\lambda_i}}\frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{T}} \lambda_i\phi_i(s)\phi_j(s)ds \\
&= \frac{\lambda_i}{\sqrt{\lambda_i}\sqrt{\lambda_j}}\delta_{ij} \\
&= \delta_{ij}
\end{aligned}
$$

Also, note that

$$
\begin{aligned}
E(X(t)\xi_i) &= \frac{1}{\sqrt{\lambda_i}}E(X(t)\int_{\mathcal{T}} X(s)\phi_i(s)ds) \\
&= \frac{1}{\sqrt{\lambda_i}}\int_{\mathcal{T}} E(X(t)X(s))\phi_i(s)ds \\
&= \frac{1}{\sqrt{\lambda_i}}\int_{\mathcal{T}} K(t,s)\phi_i(s)ds \\
&= \frac{1}{\sqrt{\lambda_i}}\lambda_i\phi_i(t) \\
&= \sqrt{\lambda_i}\phi_i(t)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& E(X(t) - \sum_{i=1}^{p}\sqrt{\lambda_i}\xi_i\phi_i(t))^2 \\
= \; & E(X(t)^2 - 2\sum_{i=1}^{p}\sqrt{\lambda_i}\xi_i\phi_i(t)X(t) + \sum_{i=1}^{p}\sum_{j=1}^{p}\sqrt{\lambda_i}\sqrt{\lambda_j}\xi_i\xi_j\phi_i(t)\phi_j(t)) \\
= \; & E(X(t)^2) - 2\sum_{i=1}^{p}\sqrt{\lambda_i}\phi_i(t)E(\xi_i X(t)) + \sum_{i=1}^{p}\sum_{j=1}^{p}\sqrt{\lambda_i}\sqrt{\lambda_j}\phi_i(t)\phi_j(t)E(\xi_i\xi_j) \\
= \; & E(X(t)^2) - 2\sum_{i=1}^{p}\sqrt{\lambda_i}\phi_i(t)\sqrt{\lambda_i}\phi_i(t) + \sum_{i=1}^{p}\lambda_i\phi_i(t)^2 \\
= \; & E(X(t)^2) - \sum_{i=1}^{p}\lambda_i\phi_i(t)^2 \to 0 \;,
\end{aligned}
$$

uniformly as $p \to \infty$. This convergence follows directly from Mercer's Theorem. (See, for example, Appendix 7.1, Mercer (1909), Porter and Stirling (1990), for details.)

(b) If $X(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\xi_i\phi_i(t)$, then

$$\begin{aligned} K(t,s) &= E(X(t)X(s)) \\ &= \sum_{i=1}^{\infty}\sum_{i=1}^{\infty} \sqrt{\lambda_i}\sqrt{\lambda_j}\phi_i(t)\phi_j(s)E(\xi_i\xi_j) \\ &= \sum_{i=1}^{\infty} \lambda_i\phi_i(t)\phi_i(s) \end{aligned}$$

Hence,

$$\int_{\mathcal{T}} K(s,t)\phi_i(t)dt = \int_{\mathcal{T}} \sum_{j=1}^{\infty} \lambda_j\phi_j(t)\phi_j(s)\phi_i(t)dt = \lambda_i\phi_i(s) \tag{1.2.6}$$

Thus, $\int_{\mathcal{T}} K(s,t)\phi_i(t)dt = \lambda_i\phi_i(s)$, using the fact that $\{\phi_i(t)\}$ is orthogonal. □

Theorem 1.2.1 provides a one-to-one relationship between equation 1.2.1 and the Karhunen-Loeve expansion. Equation 1.2.1 is well known in mathematics as the Fredholm integral equation. Only in a limited number of cases can the explicit solution to the integral equation be found. Numerical solutions play a major role in applications.

Assume that $X(t)$, $t \in \mathcal{T}$ is a family of zero mean Gaussian processes with $E[X(t)X(s)] = \text{cov}[X(t), X(s)] = K(t,s)$, where $t, s \in \mathcal{T}$ and E represents the expectation. Now a space, H_X, which is isometrically isomorphic to the RKHS with reproducing kernel K can be defined in the following way. We construct the space H_X spanned by $X(t)$, $t \in \mathcal{T}$, i.e.

$$H_X = \text{span}\{X(t), t \in \mathcal{T}\} \tag{1.2.7}$$

with the covariance as the inner product. It means that for $X_m, X_n \in H_X$, i.e.

$$X_m = \sum_i a_i X(t_i), \quad X_n = \sum_i b_i X(t_i) \tag{1.2.8}$$

where $t_i \in \mathcal{T}$, then the inner product $(X_m, X_n) = \text{cov}(X_m, X_n) = E[X_m X_n]$. This inner product is valid, since its properties, such as linearity and non-negative definiteness, can be checked using the properties of the expectation.

Meanwhile, because H_X is a closure, it should contain all the limits of all the Cauchy sequence, i.e. if X_m is a Cauchy sequence in H_X and X satisfies

$$||X - X_m||^2 = E[X - X_m]^2 \to 0 \tag{1.2.9}$$

then $X \in H_X$. The space of H_X, and the RKHS are isometrically isomorphic, since for $t, s \in \mathcal{T}$

$$\begin{aligned} (X(t), X(s)) &= E[X(t)X(s)] = \text{cov}[X(t), X(s)] = K(t,s) \\ &= (K(t,.), K(s,.)) \end{aligned}$$

In practice, only p terms is used in the Karhunen-Loeve expansion. The expression with a finite number of terms is called the *truncated* Karhunen-Loeve expansion. The use and the optimality of the truncated Karhunen-Loeve expansion is one of the main features of this thesis. The truncated Karhunen-Loeve expansion is optimal in the following sense:

Let $\{X(t), t \in \mathcal{T}\}$ be a zero mean second order process with covariance function $K(s,t)$ $s,t \in \mathcal{T}$. Let $\{\phi_i, i \in N\}$ be the set of orthogonal functions in L_2. We can then expand $X(t)$ as an infinite series of $\phi_i(t)$ as

$$X(t) = \sum_{i \geq 1} \sqrt{\lambda_i} \phi_i(t) \xi_i \tag{1.2.10}$$

where

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{T}} X(t) \phi_i(t) dt \tag{1.2.11}$$

The truncated version of $X(t)$ at order p is expressed as

$$X_p(t) = \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i(t)\xi_i \tag{1.2.12}$$

Then, the error of the truncated expansion can be defined as

$$e_p(t) = \sum_{i \geq p+1} \sqrt{\lambda_i}\phi_i\xi_i \tag{1.2.13}$$

The truncated Karhunen-Loeve expansion is optimal in the sense that its integrated mean squared error is minimised. This will be shown in theorem 1.2.2.

Theorem 1.2.2. *Among all the truncated expansion expressed as*

$$X_p(t) = \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i(t)\xi_i \ , \tag{1.2.14}$$

where $\phi_i(t)$ satisfies

$$\int_{\mathcal{T}} \phi_i(t)\phi_j(t)dt = \delta_{ij} \tag{1.2.15}$$

the Karhunen-Loeve expansion minimises the integrated mean squared error, i.e.

$$\int_{\mathcal{T}} E(e_p^2(t))dt \ , \tag{1.2.16}$$

where $E(.)$ represents the expectation and

$$e_p(t) = \sum_{m \geq p+1} \sqrt{\lambda_m}\phi_m(t)\xi_m \tag{1.2.17}$$

Proof. The square of the error can be written as

$$\begin{aligned} e_p^2(t) &= \sum_{m \geq p+1} \sum_{n \geq p+1} \sqrt{\lambda_m}\sqrt{\lambda_n}\phi_m(t)\phi_n(t)\xi_m\xi_n \\ &= \sum_{m \geq p+1} \sum_{n \geq p+1} \phi_m(t)\phi_n(t) \int_{\mathcal{T}}\int_{\mathcal{T}} X(t_1)X(s_1)\phi_m(t_1)\phi_n(s_1)dt_1ds_1 \end{aligned}$$

Then the mean squared error can be expressed as

$$E(e_p^2(t)) = \sum_{m \geq p+1} \sum_{n \geq p+1} \phi_m(t)\phi_n(t) \int_{\mathcal{T}}\int_{\mathcal{T}} K(t_1, s_1)\phi_m(t_1)\phi_n(s_1)dt_1ds_1 \tag{1.2.18}$$

Integrate both sides of the equation and use the orthogonality of eigenfunctions, then the integrated mean squared error can be obtained.

$$\begin{aligned}\int_{\mathcal{T}} E(e_p^2(t))dt &= \sum_{m\geq p+1}\sum_{n\geq p+1}\int_{\mathcal{T}}\phi_m(t)\phi_n(t)dt\int_{\mathcal{T}}\int_{\mathcal{T}}K(t_1,s_1)\phi_m(t_1)\phi_n(s_1)dt_1ds_1 \\ &= \sum_{m\geq p+1}\int_{\mathcal{T}}\int_{\mathcal{T}}K(t_1,s_1)\phi_m(t_1)\phi_m(s_1)dt_1ds_1\end{aligned}$$

Now the integrated mean squared error can be minimised given the orthogonality condition for the function $\phi_i(t)$, i.e.

$$\min\int_{\mathcal{T}}E(e_p^2(t))dt, \quad s.t.\int_{\mathcal{T}}\phi_m^2(t)=1 \tag{1.2.19}$$

Therefore, the target function for minimisation can now be re-written as

$$Q=\sum_{m\geq p+1}[\int_{\mathcal{T}}\int_{\mathcal{T}}K(t_1,s_1)\phi_m(t_1)\phi_m(s_1)dt_1ds_1-\lambda_m(\int_{\mathcal{T}}\phi_m^2(t_1)dt_1-1)] \tag{1.2.20}$$

Differentiate Q with respect to $\phi_i(t), i \geq p+1$. Notice that this is a functional derivative. Then we obtain

$$\frac{d}{d\phi_i(t)}Q=2\int_{\mathcal{T}}K(s,t)\phi_i(s)ds-2\lambda_i\phi_i(t) \tag{1.2.21}$$

Setting the above equation to zero results in the following equation

$$\int_{\mathcal{T}}K(s,t)\phi_i(s)ds=\lambda_i\phi_i(t) \tag{1.2.22}$$

This is the Fredholm integral equation. Since the Fredholm integral equation and the Karhunen-Loeve expansion have one-to-one relationship, the Karhunen-Loeve expansion minimises the integrated mean squared error among all the expansion expressed in equation 1.2.14. □

Remark 1.2.1. The functional derivative, used in deriving equation 1.2.21, is a generalisation of the directional derivative. Instead of a differentiation with respect to a variable, functional derivative is a differentiation of a function with respect to the function. It is defined as follows

$$\frac{d}{d\phi(t)}E[\phi(t')]=\lim_{\epsilon\to 0}\frac{E[\phi(t')+\epsilon\delta(t'-t)]-E[\phi(t')]}{\epsilon} \tag{1.2.23}$$

Functional derivative arises from calculus of variation, which can be regarded as an extension of calculus. For details of calculus of variation, see, for example, Sagan (1992).

Remark 1.2.2. If the stochastic system $\{X(t), t \in \mathcal{T}\}$ is a non-zero-mean process, we can subtract its mean first before applying the Karhunen-Loeve expansion, i.e.

$$X(t) = E(X(t)) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t)\xi_i \tag{1.2.24}$$

Remark 1.2.3. In some literature, especially that on the integral equations, compact integral operator theory is used. A compact integral operator $\mathcal{K}$ for the Fredholm integral equation is defined as

$$\mathcal{K}f(.) = \int_T K(s,.)f(s)ds \tag{1.2.25}$$

Hence the integral equation for the Karhunen-Loeve expansion can be written compactly as $\mathcal{K}\phi(t) = \lambda\phi(t)$. Using properties of compact operators, it can be shown that if there are infinitely many eigenvalues $\{\lambda_i, i \in N\}$ satisfying the Fredholm integral equation, then $\lambda_i \to 0$ when $i \to \infty$. See, for example, chapter 4, Porter and Stirling (1990), for details. Using an integral operator is helpful to solve certain integral equations, such as those related to the integrated Brownian motion. This will be presented in section 1.3.

1.3 Analytical examples

For certain processes related to the Brownian motion, the analytical solution to the Karhunen-Loeve expansion can be found. Four analytical examples will be discussed here, the Brownian motion, the Brownian bridge, the Ornstein-Uhlenbeck process (O-U process) and the integrated Brownian motion. For simplicity, the time interval for all the processes here are assumed to be at $[0, 1]$, i.e. $\mathcal{T} = [0, 1]$.

Example 1: The Brownian motion

The covariance function $K(s,t)$ for the Brownian motion is $\min(s,t)$. Hence, the integral equation is written as

$$\int_{\mathcal{T}} \min(s,t)\phi(s)ds = \lambda\phi(t) \tag{1.3.1}$$

Or equivalently

$$\int_0^t s\phi(s)ds + t\int_t^1 s\phi(s)ds = \lambda\phi(t) \tag{1.3.2}$$

Note that the above equation implies the boundary condition $\phi(0) = 0$. Evaluating the first derivative to both sides of the equation, we obtain

$$\int_t^1 s\phi(s)ds = \lambda\frac{d}{dt}\phi(t) \tag{1.3.3}$$

Take the derivative one more time,

$$-\phi(t) = \lambda\frac{d^2}{dt^2}\phi(t) \tag{1.3.4}$$

Solution to this ordinary differential equation is of the form

$$\phi(t) = A\sin(\frac{t}{\sqrt{\lambda}}) + B\cos(\frac{t}{\sqrt{\lambda}}) \ , \tag{1.3.5}$$

where A and B are constants. Since $\phi(0) = 0$, then $B = 0$ and $\phi(t) = A\sin(\frac{t}{\sqrt{\lambda}})$.

For eigenvalues, $\phi(t)$ is substituted into the first derivative equation.

$$\begin{aligned} A\int_t^1 \sin(\frac{s}{\sqrt{\lambda}})ds &= A\lambda\cos(\frac{t}{\sqrt{\lambda}})\frac{1}{\sqrt{\lambda}} \\ A\cos(\frac{t}{\sqrt{\lambda}}) - A\cos(\frac{1}{\sqrt{\lambda}}) &= A\cos(\frac{t}{\sqrt{\lambda}}) \end{aligned}$$

Hence, $\cos(\frac{1}{\sqrt{\lambda_i}}) = 0$, which implies

$$\lambda_i = \frac{4}{(2i-1)^2\pi^2}, \ i \geq 1 \tag{1.3.6}$$

The constant A can be found through the orthogonality condition for $\phi_i(t)$

$$1 = \int_0^1 \phi_i^2(t)dt = A^2\int_0^1 \sin^2(\frac{t}{\sqrt{\lambda_i}})dt = A^2\int_0^1 \sin^2[(i-\frac{1}{2})\pi t]dt = \frac{A^2}{2} \tag{1.3.7}$$

Thus, $A = \sqrt{2}$. Therefore, the Karhunen-Loeve expansion for the Brownian motion can be represented as

$$W(t) = \sqrt{2}\sum_{i\geq 1}\frac{2}{(2i-1)\pi}\sin[(i-\frac{1}{2})\pi t]\xi_i \tag{1.3.8}$$

where $\{\xi_i\}$ are i.i.d Gaussian process with mean zero and variance one.

Remark 1.3.1. The integral operator $\mathcal{K}$ for the Brownian motion can be expressed as a multiplication of two simplified operators defined as the following

$$T_0\phi(t) = \int_0^t \phi(s)ds, \quad T_1\phi(t) = \int_t^1 \phi(s)ds \tag{1.3.9}$$

Then the integral operator for the Brownian motion is T_0T_1, since

$$\begin{aligned}
\int_0^1 \mathrm{cov}(W(u), W(t))\phi(u)du &= \int_0^1 \min(u,t)\phi(u)du \\
&= \int_0^t u\phi(u)du + \int_t^1 \phi(u)du \\
&= \int_0^t \int_0^u \phi(u)dsdu + \int_t^1 \int_0^t \phi(u)dsdu \\
&= \int_0^t \int_s^1 \phi(u)duds + \int_0^t \int_t^1 \phi(u)duds \\
&= \int_0^t \int_s^1 \phi(u)duds = \int_0^t T_1\phi(s)ds \\
&= T_0T_1\phi(t)
\end{aligned}$$

This result will be shown to be useful for finding the eigenvalues of the integrated Brownian motion in example 4.

Example 2: The Brownian bridge

The Brownian bridge $X(t)$ can be derived from the Brownian motion $W(t)$ by conditioning on $W(1) = 0$. The analytical relationship can be written as

$$X(t) = W(t) - tW(1) \tag{1.3.10}$$

The corresponding covariance function for $X(s)$ and $X(t)$ can be expressed as

$$\min(s,t) - st$$

Hence, the integral equation is written as

$$\int_{\mathcal{T}} (\min(s,t) - st)\phi_i(s)ds = \lambda_i \phi_i(t) \tag{1.3.11}$$

Differentiate with respect to t for the first time,

$$\int_t^1 \phi(s)ds - \int_0^1 s\phi(s)ds = \lambda \frac{d}{dt}\phi(t) \tag{1.3.12}$$

Take the derivative one more time,

$$-\phi(t) = \lambda \frac{d^2}{dt^2}\phi(t) \tag{1.3.13}$$

This is exactly the same second derivative as that for Brownian motion. Hence, combining with the boundary condition $\phi(0) = 0$, we have

$$\phi(t) = A\sin(\frac{t}{\sqrt{\lambda}}) \tag{1.3.14}$$

Now $\phi(t)$ is substituted into the first derivative equation to obtain

$$\sqrt{\lambda}\cos(\frac{t}{\sqrt{\lambda}}) - \sqrt{\lambda}\cos(\frac{1}{\sqrt{\lambda}}) - \lambda\sin(\frac{1}{\sqrt{\lambda}}) + \sqrt{\lambda}\cos(\frac{1}{\sqrt{\lambda}}) = \lambda\frac{1}{\sqrt{\lambda}}\cos(\frac{1}{\sqrt{\lambda}}) \tag{1.3.15}$$

Hence, $\sin(\frac{1}{\sqrt{\lambda}}) = 0$, which implies

$$\lambda_i = \frac{1}{i^2\pi^2}, i \geq 1 \tag{1.3.16}$$

As before, the orthogonality condition of $\phi_i(t)$ is used to find A

$$A^2 \int_0^1 \sin^2(\frac{t}{\sqrt{\lambda_i}})dt = A^2 \int_0^1 \sin^2(i\pi t)dt = \frac{A^2}{2} = 1 \tag{1.3.17}$$

Again, $A = \sqrt{2}$. Therefore, the Karhunen-Loeve expansion for the Brownian bridge can be represented as

$$X(t) = \sqrt{2}\sum_{i\geq 1} \frac{1}{i\pi}\sin(i\pi t)\xi_i \tag{1.3.18}$$

where $\{\xi_i\}$ is a i.i.d Gaussian process with mean zero and variance one.

Remark 1.3.2. It can be seen that the Brownian bridge is the Brownian motion $W(t)$ conditional on $W(1) = 0$. Therefore, the Karhunen-Loeve expansion for the Brownian bridge can be regarded as an extension of the Karhuneen-Loeve expansion of the Brownian motion. This extension is related to a concept called the "conditional Karhunen-Loeve expansion". We refer to section 4.5, chapter 4, for a detailed discussion.

Remark 1.3.3. The Karhunen-Loeve expansion of the Brownian bridge can be applied to the goodness of fit test, since the asymptotic behavior of the test statistic involves a multivariate Brownian bridge. See, for example, Cheng and Jones (2004), for detail.

Example 3: The Ornstein Uhlenbeck process (O-U process)

For $\beta > 0$ and $\rho > 0$, We have a stochastic differential equation (SDE) for the stochastic process $Z(t)$

$$dZ(t) = -\beta Z(t)dt + \rho dW(t) \tag{1.3.19}$$

This SDE can be solved and the corresponding solution is

$$Z(t) = Z(0)e^{-\beta t} + \rho \int_0^t e^{-\beta(t-s)} dW(s) \tag{1.3.20}$$

If $Z(0) \sim N(0, c)$, where $c = \frac{\rho^2}{2\beta}$, Z is called an O-U process with parameters β and c. Note that

$$\begin{aligned} & \operatorname{cov}\left(\int_0^t e^{-\beta(t-u)} dW_u, \int_0^s e^{-\beta(s-u)} dW_u\right) = E\left(\int_0^t e^{-\beta(t-u)} dW_u, \int_0^s e^{-\beta(s-u)} dW_u\right) \\ = \; & e^{-\beta(t+s)} E\left[\left(\int_0^{\min(t,s)} e^{\beta u} dW_u\right)^2\right] = e^{-\beta(t+s)} \int_0^{\min(t,s)} e^{2\beta u} dW_u \\ = \; & \frac{e^{-\beta(t+s)}}{2\beta}\left(e^{2\beta \min(t,s)} - 1\right) = \frac{e^{-\beta|t-s|}}{2\beta} - \frac{e^{-\beta(t+s)}}{2\beta} \end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{cov}(Z(t),Z(s)) &= e^{-\beta(t+s)}\mathrm{Var}(Z(0)) + \rho^2\mathrm{cov}(\int_0^t e^{-\beta(t-u)}dW_u, \int_0^s e^{-\beta(s-u)}dW_u) \\ &= \rho^2\frac{e^{-\beta|t-s|}}{2\beta} = ce^{-\beta|t-s|}\end{aligned}$$

Once ρ and β are fixed, c is a constant. A constant term will not affect the calculation procedure of the Karhunen-Loeve expansion. The reason is as follows. Assume that there are two Fredholm integral equations.

$$\int K(s,t)\phi_1(s)ds = \lambda_1\phi_1(t) \quad \text{and} \quad \int cK(s,t)\phi_2(s)ds = \lambda_2\phi_2(t)\ , \tag{1.3.21}$$

where c is a constant. It can be seen that the relationship between λ_1 and λ_2, $\phi_1(t)$ and $\phi_2(t)$ is

$$\lambda_2 = c\lambda_1 \quad \text{and} \quad \phi_1(t) = \phi_2(t) \tag{1.3.22}$$

Hence the constant term c in the Fredholm integral equation will only affect the eigenvalues by multiplying c, while the eigenfunctions remain the same.

For simplicity, in this example, it is assumed that $c = 1$ and therefore, only the covariance function $\exp[-\beta|s-t|]$ is decomposed. This involves solving the following integral equation.

$$\int_0^1 \exp[-\beta|s-t|]\phi(s)ds = \lambda\phi(t) \tag{1.3.23}$$

Expand the left hand side of equation 1.3.23, we obtain,

$$\int_t^1 \exp[-\beta(s-t)]\phi(s)ds + \int_0^t \exp[-\beta(t-s)]\phi(s)ds = \lambda\phi(t) \tag{1.3.24}$$

Differentiate equation 1.3.24 with respect to t for the first time, we obtain

$$\beta[\int_t^1 \exp[-\beta(s-t)]\phi(s)ds - \int_0^t \exp[-\beta(t-s)]\phi(s)ds] = \lambda\frac{d\phi(t)}{dt} \tag{1.3.25}$$

Differentiate equation 1.3.24 with respect to t for the second time, we obtain

$$\beta[-2\phi(t)+\int_t^1 \beta\exp[-\beta(s-t)]\phi(s)ds+\int_0^t \beta\exp[-\beta(t-s)]\phi(s)ds] = \lambda\frac{d^2\phi(t)}{dt^2} \quad (1.3.26)$$

Simplifying equation 1.3.26 results in

$$\frac{d^2\phi(t)}{dt^2} + w^2\phi(t) = 0 \qquad \text{where} \quad w^2 = \frac{2\beta - \beta^2\lambda}{\lambda} \quad (1.3.27)$$

Solving the differential equation 1.3.27, we obtain

$$\phi(t) = A\cos(wt) + B\sin(wt) \quad (1.3.28)$$

Notice that the boundary conditions in the O-U process involve both the original eigenfunctions and their first derivative, since

$$\lambda\phi(0) = \int_0^1 \exp(-\beta s)\phi(s)ds, \lambda\frac{d}{dt}\phi(0) = \beta\int_0^1 \exp(-\beta s)\phi(s)ds$$
$$\lambda\phi(1) = \int_0^1 \exp(-\beta(1-s))\phi(s)ds, \lambda\frac{d}{dt}\phi(1) = -\beta\int_0^1 \exp(-\beta(1-s))\phi(s)ds$$

This gives the boundary conditions

$$\frac{d}{dt}\phi(0) - \beta\phi(0) = 0$$
$$\frac{d}{dt}\phi(1) + \beta\phi(1) = 0$$

These can be expressed as

$$A\beta - Bw = 0$$
$$A(\beta - w\tan(w)) + B(\beta\tan(w) + w) = 0$$

In order to obtain non-zero solutions for A and B, we need

$$\det\begin{pmatrix} \beta & -w \\ \beta - w\tan(w) & \beta\tan(w) + w \end{pmatrix} = 0 \quad (1.3.29)$$

which means that

$$\cot(w) = \frac{w^2 - \beta^2}{2\beta w} \Leftrightarrow w^2\sin(w) - \beta^2\sin(w) = 2\beta w\cos(w) \tag{1.3.30}$$

After solving for w, λ can be derived via $\lambda = \frac{2\beta}{w^2+\beta^2}$. For the unknown parameters A and B, using the fact that $B = \frac{\beta}{w}A$ and the orthogonal condition for the eigenfunction $\int_0^1 \phi^2(t)dx = 1$, we obtain

$$\begin{aligned} A &= \sqrt{\frac{2w^2}{2\beta + w^2 + \beta^2}} \\ B &= \frac{\beta}{w}A = \sqrt{\frac{2\beta^2}{2\beta + w^2 + \beta^2}} \end{aligned}$$

Then, when $k \geq 1$, the eigenvalue and the eigenfunction can be expressed as

$$\lambda_k = \frac{2\beta}{w_k^2 + \beta^2}, \tag{1.3.31}$$

$$\phi_k(t) = \sqrt{\frac{2w_k^2}{2\beta + w_k^2 + \beta^2}}\cos(w_k t) + \sqrt{\frac{2\beta^2}{2\beta + w_k^2 + \beta^2}}\sin(w_k t) \tag{1.3.32}$$

Now we can present the asymptotic behaviour for λ_k. Assume that w_k is an increasing function of k and $w_k \to \infty$ when $k \to \infty$. This assumption makes sense since, according to remark 1.2.3, λ_k is a decreasing function of k and $\lambda_k \to 0$ when $k \to \infty$. Then

$$\cot(w_k) = \frac{w_k^2 - \beta^2}{2\beta w_k} \to \infty \quad \text{when } k \to \infty$$

The asymptotic behaviour of w_k is therefore,

$$w_k \simeq k\pi \quad \text{when } k \to \infty$$

Replacing w_k in λ_k with $k\pi$, we obtain the asymptotic behaviour of λ_k

$$\lambda_k \simeq \frac{2\beta}{k^2\pi^2 + \beta^2} \simeq \frac{1}{k^2\pi^2} \quad \text{when } k \to \infty \tag{1.3.33}$$

Example 4: The m-integrated Brownian motion

The general m-integrated Brownian motion is defined as

$$X_m(t) = \int_{\mathcal{T}_m} \int_{\mathcal{T}_{m-1}} \cdots \int_{\mathcal{T}_1} W_{i_0}(t_0) dt_0 dt_1 dt_{m-1} \tag{1.3.34}$$

where i_0 is either 0 or 1, $W_0(t) = W(t)$, which is a Brownian motion, $W_1(t) = W(1-t)$, and

$$\mathcal{T}_i = [0, t_i] \text{ or } [t_i, 1], \ t_i \in (0,1), \ i = 1, \cdots, m-1$$
$$\mathcal{T}_m = [0, t] \text{ or } [t, 1], \ t \in (0,1)$$

The above multiple integral will be shown to be related to the Sturn-Liouville problem in this example.

We start by recalling operators T_0 and T_1, which have been introduced for the Brownian motion in example 1.

$$T_0\phi(t) = \int_0^t \phi(s)ds, \quad T_1\phi(t) = \int_t^1 \phi(s)ds \tag{1.3.35}$$

Using the operators T_0 and T_1, the m-integrated Brownian motion can be written as

$$X_m(t) = T_{i_m} T_{i_{m-1}} \cdots T_{i_1} W_{i_0}(t), \quad i_j \in \{0,1\}, \ j = 0, \cdots, m \tag{1.3.36}$$

Then the integral operator for $X_m(t)$ is expressed through the following two propositions.

Proposition 1.3.1. *Define $W_0(t) = W(t)$, which is a Brownian motion and $W_1(t) = W(1-t) = W_0(1-t)$. Then the integral operator for $W_i(t)$ is $T_i T_{1-i}$, $i \in \{0,1\}$.*

Proof. In remark 1.3.1, proposition 1.3.1 has been proved to be true for $i = 0$. For $i = 1$, the proof follows the same procedure. □

Proposition 1.3.2. *Let $\{X(t), t \in [0,1]\}$ be a zero mean second order process with the integral operator $\mathcal{K}$. Then the integral operator for $T_iX(t)$ is $T_i\mathcal{K}T_{1-i}$, $i \in \{0,1\}$*

Proof. Notice that

$$T_0X(t) = \int_0^t X(s)ds, \quad T_1X(t) = \int_t^1 X(s)ds \tag{1.3.37}$$

When $i = 0$, on one hand

$$\begin{aligned} T_0\mathcal{K}T_1\phi(t) &= \int_0^t \mathcal{K}T_1\phi(s)ds \\ &= \int_0^t \int_0^1 K(s,u)T_1\phi(u)duds \\ &= \int_0^t \int_0^1 K(s,u)\int_u^1 \phi(v)dvduds \\ &= \int_0^t \int_0^1 \int_u^1 K(s,u)\phi(v)dvduds \end{aligned}$$

On the other hand

$$\begin{aligned} \int_0^1 \mathrm{cov}(T_0(X(t)), T_0(X(v))\phi(v)dv &= \int_0^1 E(T_0(X(t))T_0(X(v))\phi(v)dv \\ &= \int_0^1 \int_0^v \int_0^t K(s,u)dsdu\phi(v)dv \\ &= \int_0^1 \int_0^t \int_0^v K(s,u)\phi(v)dudsdv \\ &= \int_0^t \int_0^1 \int_0^v K(s,u)\phi(v)dudvds \\ &= \int_0^t \int_0^1 \int_u^1 K(s,u)\phi(v)dvduds \end{aligned}$$

Hence, $T_0\mathcal{K}T_1\phi(t) = \int_0^1 \mathrm{cov}[T_0(X(t)), T_0(X(v))]\phi(v)dv$, i.e. the integral operator for $T_0(X(t))$ is $T_0\mathcal{K}T_1$. The proof for $i = 1$ follows similarly. □

Combining the above two propositions and using induction, corollary 1.3.3 can now be derived.

Corollary 1.3.3. *The integral operator for the m-integrated Brownian motion is*

$$\mathcal{K} = T_{i_m} T_{i_{m-1}} \cdots T_{i_0} T_{1-i_0} T_{1-i_2} \cdots T_{1-i_m} \tag{1.3.38}$$

Proof. This is a direct result from the above two propositions.

Using induction, when $m = 1$, since the integral operator for W_{i_0} is $\mathcal{K}' = T_{i_0} T_{1-i_0}$, the integral operator for $T_{i_1} W_{i_0}$ is $T_{i_1} \mathcal{K}' T_{1-i_1} = T_{i_1} T_{i_0} T_{1-i_0} T_{1-i_1}$

Assume the result holds when $m = n - 1$, i.e. the integral operator for $X_{n-1} = T_{i_{n-1}} \cdots T_{i_1} W_{i_0}$ is

$$\mathcal{K}'' = T_{i_{n-1}} \cdots T_{i_0} T_{1-i_0} \cdots T_{1-i_{n-1}} \tag{1.3.39}$$

Then when $m = n$, the integral operator for $X_n = T_{i_n} X_{n-1}$ is

$$T_{i_n} \mathcal{K}'' T_{1-i_n} = T_{i_n} T_{i_{n-1}} \cdots T_{i_0} T_{1-i_0} \cdots T_{1-i_{n-1}} T_{1-i_n} \tag{1.3.40}$$

as is required. □

Using corollary 1.3.3, in order to derive the Karhunen-Loeve expansion for the m-integrated Brownian motion, the following integral equation needs to be solved.

$$T_{i_m} T_{i_{m-1}} \cdots T_{i_0} T_{1-i_0} T_{1-i_2} \cdots T_{1-i_m} \phi(t) = \lambda \phi(t) \tag{1.3.41}$$

After differentiating equation 1.3.41 $(2m + 2)$ times, the following is obtained.

$$(-1)^{m+1} \phi(t) = \lambda \frac{d^{2m+2}}{dt^{2m+2}} \phi(t) \tag{1.3.42}$$

with boundary conditions

$$\underbrace{\phi(i_m) = \frac{d}{dt}\phi(i_{m-1}) = \cdots = \frac{d^m}{dt^m}\phi(i_0)}_{m+1 \text{ conditions}} = \frac{d^{m+1}}{dt^{m+1}}\phi(1 - i_0) = \cdots = \frac{d^{2m+1}}{dt^{2m+1}}\phi(1 - i_m) = 0 \tag{1.3.43}$$

This system of equations is known as the Sturm-Liouville problem. See, for example, Zill and Cullen (2001), Naimark (1968), for details.

There are two special cases worth discussing here. One is defined as

$$X_m^1(t) = \int_0^t \int_0^{t_{m-1}} \cdots \int_0^{t_1} W_0(t_0) dt_0 \cdots dt_{m-1} \tag{1.3.44}$$

with the corresponding boundary conditions

$$\phi(0) = \frac{d}{dt}\phi(0) = \cdots = \frac{d^m}{dt^m}\phi(0) = \frac{d^{m+1}}{dt^{m+1}}\phi(1) = \cdots = \frac{d^{2m+1}}{dt^{2m+1}}\phi(1) = 0 \tag{1.3.45}$$

We call X_m^1 the m-integrated Brownian motion of the *first type*. Its integral operator is

$$\underbrace{T_0 \cdots T_0}_{m+1}\underbrace{T_1 \cdots T_1}_{m+1} \tag{1.3.46}$$

$X_m^1(t)$ can also be written as

$$X_m^1(t) = \int_0^t X_{m-1}^1(s) ds \tag{1.3.47}$$

with $X_0(t) = W(t)$. In some literature (for example, Chen and Li (2003), Rue and Held (2005)), X_m^1 can also be expressed as

$$\begin{aligned} X_m^1(t) &= \int_0^t \frac{(t-s)^m}{m!} dW(s) \\ &= \frac{1}{m!} \int_0^t (t-s)^{m-1} W(s) ds \end{aligned}$$

The other special example of the integrated Brownian motion is defined as

$$X_m^2(t) = \begin{cases} \int_0^t \int_{t_{m-1}}^1 \cdots \int_0^{t_1} W_1(t_0) dt_0 \cdots dt_{m-1} & \text{if } m \text{ is odd} \\ \int_0^t \int_{t_{m-1}}^1 \cdots \int_{t_1}^1 W_0(t_0) dt_0 \cdots dt_{m-1} & \text{if } m \text{ is even} \end{cases} \tag{1.3.48}$$

with the corresponding boundary conditions

$$\phi(0) = \frac{d}{dt}\phi(1) = \cdots = \underbrace{\frac{d^n}{dt^n}\phi(0)}_{n \text{ is even}} = \frac{d^{n+1}}{dt^{n+1}}\phi(1) = \cdots = \frac{d^{2m+1}}{dt^{2m+1}}\phi(1) = 0 \tag{1.3.49}$$

We call X_m^2 the m-integrated Brownian motion of the *second type*. The covariance operator for X_m^2 is

$$\underbrace{T_0T_1T_0\cdots T_1T_0T_1}_{2m+2} \tag{1.3.50}$$

The covariance function for X_m^1 and X_m^2 can be expressed as follows. The covariance function for X_m^1 is

$$K_m^1(s,t) = \frac{1}{(m!)^2}\int_0^{\min(s,t)} (s-u)^m(t-u)^m du \tag{1.3.51}$$

and the covariance function for X_m^2 is

$$K_m^2(s,t) = \int_0^1 \cdots \int_0^1 (\min(s,s_1))(\min(s_1,s_2))\cdots(\min(s_m,t))ds_1ds_2\cdots ds_m \tag{1.3.52}$$

Normally, the Karhunen-Loeve expansion is derived from solving an integral equation involving a covariance function. However, in this example, due to the complicated structure of the covariance function, it is very difficult to solve the integral equation directly. Nevertheless, using the operators discussed above, the analytical solution to X_m^2 is straightforward to derive, while the analytical solution to X_m^1 remains intractable. After finding the eigenvalues and eigenfunctions for X_m^2, we provide a simplified example of X_m^1 when $m = 1$.

Theorem 1.3.4. *The eigenvalue and eigenfunction for* X_m^2 *are*

$$\lambda_i = \left[\frac{4}{(2i-1)^2\pi^2}\right]^{m+1}$$
$$\phi_i(t) = \sqrt{2}\sin[(i-\frac{1}{2})\pi t]$$
$$i \geq 1$$

respectively.

Proof. The integral operator for X_m^2 is

$$\underbrace{T_0T_1T_0\cdots T_1T_0T_1}_{2m+2} \tag{1.3.53}$$

Define $C = T_0T_1$, which is the covariance operator for the Brownian motion, $W(t)$. We denote the eigenvalue and the eigenfunction of $W(t)$ as λ^W and $\phi^W(t)$

$$C\phi^W(t) = \lambda^W\phi^W(t) \ ,$$

where, when $i \geq 1$

$$\lambda_i^W = \frac{4}{(2i-1)^2\pi^2}$$
$$\phi_i^W(t) = \sqrt{2}\sin[(i-\frac{1}{2})\pi t]$$

Now the integral operator for X_m^2 is C^{m+1}. Hence

$$\lambda\phi(t) = \mathcal{K}_m^2\phi(t) = C^{m+1}\phi(t) = (\lambda^W)^{m+1}\phi(t)$$

This means that when $i \geq 1$

$$\lambda_i = (\lambda_i^W)^{m+1} = \left[\frac{\cdot\ 4}{(2i-1)^2\pi^2}\right]^{m+1} \tag{1.3.54}$$
$$\phi_i(t) = \phi_i^W(t) = \sqrt{2}\sin[(i-\frac{1}{2})\pi t] \tag{1.3.55}$$

□

Although X_m^1 looks less complicated than X_m^2, no analytical solution to the Karhunen-Loeve expansion exists for X_m^1. In theorem 2, Gao et al. (2003), the asymptotic behavior for the eigenvalue of X_m^1 is derived, and can be seen as

$$\lambda_k \sim (\pi k)^{-(2m+2)} \qquad as \quad k \longrightarrow \infty \tag{1.3.56}$$

This is the same result as that provided by proposition 10, chapter IV, Ritter (2000).

For X_m^1, only $m = 1$ is considered. Using the formula provided above, it is easily shown that the covariance function for X_1^1 is

$$K_1^1(s,t) = \frac{1}{2}\min(s,t)^2\max(s,t) - \frac{1}{6}\min(s,t)^3 \tag{1.3.57}$$

Hence the corresponding integral equation $\int_0^1 K_1^1(s,t)\phi(s)ds = \lambda\phi(t)$ is expressed as

$$\int_0^t (\frac{1}{2}s^2t - \frac{1}{6}s^3)\phi(s)ds + \int_t^1 (\frac{1}{2}t^2s - \frac{1}{6}t^3)\phi(s)ds = \lambda\phi(t) \tag{1.3.58}$$

Differentiating equation 1.3.58 successively, we obtain

$$\phi(t) = \lambda\frac{d^4}{dt^4}\phi(t) \tag{1.3.59}$$

with the boundary conditions

$$\phi(0) = \frac{d}{dt}\phi(0) = \frac{d^2}{dt^2}\phi(1) = \frac{d^3}{dt^3}\phi(1) = 0 \tag{1.3.60}$$

The real solution to the differential equation 1.3.59 with the boundary condition 1.3.60 is

$$\phi(t) = c_1\sinh(\frac{t}{\lambda^{\frac{1}{4}}}) + c_2\cosh(\frac{t}{\lambda^{\frac{1}{4}}}) + c_3\sin(\frac{t}{\lambda^{\frac{1}{4}}}) + c_4\cos(\frac{t}{\lambda^{\frac{1}{4}}}) \tag{1.3.61}$$

The boundary conditions imply

$$\begin{aligned}
&\phi(0) = 0 \Rightarrow c_2 + c_4 = 0\\
&\frac{d}{dt}\phi(0) = 0 \Rightarrow c1 + c3 = 0\\
&\frac{d^2}{dt^2}\phi(1) = 0 \Rightarrow c_1\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) + c_2\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) - c_3\sin(\frac{1}{\lambda^{\frac{1}{4}}}) - c_4\cos(\frac{1}{\lambda^{\frac{1}{4}}}) = 0\\
&\frac{d^3}{dt^3}\phi(1) = 0 \Rightarrow c_1\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + c_2\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) - c_3\cos(\frac{1}{\lambda^{\frac{1}{4}}}) + c_4\sin(\frac{1}{\lambda^{\frac{1}{4}}}) = 0
\end{aligned}$$

Substitute $c_3 = -c_1$ and $c_4 = -c_2$ into $\frac{d^2}{dt^2}\phi(1) = \frac{d^3}{dt^3}\phi(1) = 0$.

$$c_1[\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) + \sin(\frac{1}{\lambda^{\frac{1}{4}}})] + c_2[\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + \cos(\frac{1}{\lambda^{\frac{1}{4}}})] = 0$$
$$c_1[\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + \cos(\frac{1}{\lambda^{\frac{1}{4}}})] + c_2[\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) - \sin(\frac{1}{\lambda^{\frac{1}{4}}})] = 0$$

Or equivalently,

$$\frac{c_1}{c_2} = -\frac{\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + \cos(\frac{1}{\lambda^{\frac{1}{4}}})}{\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) + \sin(\frac{1}{\lambda^{\frac{1}{4}}})}$$
$$\frac{c_1}{c_2} = -\frac{\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) - \sin(\frac{1}{\lambda^{\frac{1}{4}}})}{\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + \cos(\frac{1}{\lambda^{\frac{1}{4}}})}$$

Hence,

$$-\frac{\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + \cos(\frac{1}{\lambda^{\frac{1}{4}}})}{\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) + \sin(\frac{1}{\lambda^{\frac{1}{4}}})} = -\frac{\sinh(\frac{1}{\lambda^{\frac{1}{4}}}) - \sin(\frac{1}{\lambda^{\frac{1}{4}}})}{\cosh(\frac{1}{\lambda^{\frac{1}{4}}}) + \cos(\frac{1}{\lambda^{\frac{1}{4}}})} \tag{1.3.62}$$

Simplifying equation 1.3.62 results in

$$1 + \cosh(\frac{1}{\lambda^{\frac{1}{4}}})\cos(\frac{1}{\lambda^{\frac{1}{4}}}) = 0 \tag{1.3.63}$$

Hence, the eigenvalues for $X_1^1(t)$ are obtained as real solutions of the equation $1 + \cosh(\frac{1}{\lambda_k^{\frac{1}{4}}})\cos(\frac{1}{\lambda_k^{\frac{1}{4}}}) = 0$, or $1 + \cosh(x_k)\cos(x_k) = 0$, where $x_k = \frac{1}{\lambda_k^{\frac{1}{4}}}$. Using the Taylor expansion of $\cosh(x)$

$$\cosh(x) = \sum_{i=0}^{\infty} \frac{x^{2i}}{(2i)!} = 1 + h(x) \ , \tag{1.3.64}$$

where

$$h(x) = \sum_{i=1}^{\infty} \frac{x^{2i}}{(2i)!} \quad \text{and} \quad \lim_{x\to\infty} h(x) = \lim_{x\to\infty} \sum_{i=1}^{\infty} \frac{x^{2i}}{(2i)!} = \infty$$

According to remark 1.2.3, $\lim_{k\to\infty}\lambda_k = 0$. Since $x_k = \frac{1}{\lambda_k^{\frac{1}{4}}}$, $x_k \to \infty$ is equivalent to $\lambda_k \to 0$ when $k \to \infty$.

At the same time, according to equation 1.3.63,

$$\cos(x_k) = \frac{-1}{\cosh(x_k)} = \frac{1}{1+h(x_k)} \tag{1.3.65}$$

Therefore,

$$\cos(x_k) \to 0 \quad \Leftrightarrow \quad \cos\left(\frac{1}{\lambda_k^{\frac{1}{4}}}\right) \to 0 \quad \text{when } k \to \infty \tag{1.3.66}$$

This is equivalent to

$$\lambda_k \simeq \left(\frac{1}{k\pi}\right)^4 \qquad \text{when } k \to \infty \tag{1.3.67}$$

Remark 1.3.4. All the covariance functions which have been discussed as examples in this section satisfy the so-called Sacks-Ylvisaker Regularity conditions. These were introduced by Sacks and Ylvisaker in a series of papers (see, Sacks and Ylvisaker (1966), Sacks and Ylvisaker (1968), Sacks and Ylvisaker (1970a) and Sacks and Ylvisaker (1970b)). The essential idea of Sacks-Ylvisaker condition is to ensure that the process has no quadratic mean derivative. Typical examples on kernels of such processes are

$$\begin{aligned} &\min(s,t), \quad \min(s,t) - st, \quad 1 + \min(s,t), \\ &\frac{1}{2}(1 - |s-t|), \quad \frac{1}{2}\exp(-|s-t|) \end{aligned}$$

The above class of kernels was also called by Ritter (See Ritter (2000)), Sacks-Ylvisaker conditions of order $r = 0$. Ritter extended the idea of Sacks-Ylvisaker condition to order $r > 0$. Typical examples of processes with kernels satisfying Sacks-Ylvisaker condition of order $r > 0$ are the r-integrated Brownian motion of the first type and of the second type. Ritter further developed an important theorem for the behaviour of eigenvalues in the Sacks-Ylvisaker family. He proved that if a process with the covariance kernel K satisfies Sacks-Ylvisaker conditions of order $r \in N$, the

asymptotic behaviour of the eigenvalue λ_i is

$$\lambda_i \approx (\pi i)^{-(2r+2)} \tag{1.3.68}$$

The asymptotic behaviour of the eigenvalues from equation 1.3.68 for the Sacks-Ylvisaker family of order r is a general result. Ignoring the constant term, equation 1.3.56 and equation 1.3.54 for the eigenvalues of the r-integrated Brownian motion of the first type and of the second type respectively match equation 1.3.68 when $r > 0$. When $r = 0$, $\lambda_i \approx \frac{1}{(\pi i)^2}$. This matches our previous analysis on the eigenvalues of the Brownian motion, the Brownian bridge and the Ornstein-Uhlenbeck process ignoring the constant term.

1.4 The spatial Karhunen-Loeve expansion

In certain fields, such as geostatistics and weather forecasting, researchers are interested in the multivariate Karhunen-Loeve expansion. Multivariate in this section refers to multivariate time, and univariate state. This concept of multivariate is different from that in chapter 5, where multivariate refers to multivariate state, but univariate time. The multivariate time and univariate state Karhunen-Loeve expansion is also called the spatial Karhunen-Loeve expansion.

Let $X(\mathbf{t}) \in H_X$ be a spatial process, where multivariate time $\mathbf{t}$ is a vector time and $\mathbf{t} \in \mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2 \times \cdots \times \mathcal{T}_d$. To be more clear, $X(\mathbf{t})$ can be written down as $X(t_1, t_2, \cdots, t_d)$, where $t_i \in \mathcal{T}_i$, $1 \leq i \leq d$. Its corresponding covariance function between $\mathbf{s} \in \mathcal{T}$ and $\mathbf{t} \in \mathcal{T}$ can then be represented as

$$K(\mathbf{s}, \mathbf{t}) = \text{cov}[X(s_1, s_2, \cdots, s_d), X(t_1, t_2, \cdots, t_d)] \tag{1.4.1}$$

For the Karhunen-Loeve expansion of $X(\mathbf{t})$, it follows directly from the univariate Karhunen-Loeve expansion. We need to solve the integral equation $\int_{\mathcal{T}} K(\mathbf{s},\mathbf{t})\phi(\mathbf{s})d\mathbf{s} = \lambda\phi(\mathbf{t})$, which is equivalent to the following equation.

$$\begin{aligned} &\int_{\mathcal{T}_1}\int_{\mathcal{T}_2}\cdots\int_{\mathcal{T}_d} \mathrm{cov}[X(s_1,s_2,\cdots,s_d),X(t_1,t_2,\cdots,t_d)]\phi(s_1,s_2,\cdots,s_d)ds_1ds_2\cdots ds_d \\ = \;& \lambda\phi(t_1,t_2,\cdots,t_d) \end{aligned}$$

where λ and $\phi(t_1,t_2,\cdots,t_d)$ are the corresponding eigenvalue and eigenfunction respectively. Since the eigenfunction is orthogonal, it should satisfy the orthogonality condition $\int_{\mathcal{T}} \phi_i(\mathbf{t})\phi_j(\mathbf{t})d\mathbf{t} = \delta_{ij}$, which is equivalent to the following equation.

$$\int_{\mathcal{T}_1}\int_{\mathcal{T}_2}\cdots\int_{\mathcal{T}_d} \phi_i(t_1,t_2,\cdots,t_d)\phi_j(t_1,t_2,\cdots,t_d)dt_1dt_2\cdots dt_d = \delta_{ij} \tag{1.4.2}$$

Then the Karhunen-Loeve expansion for $X(\mathbf{t})$ is $X(\mathbf{t}) = \sum_{i\geq 1}\sqrt{\lambda_i}\phi_i(\mathbf{t})\xi_i$, which is equivalent to

$$X(t_1,t_2,\cdots,t_d) = \sum_{i\geq 1}\sqrt{\lambda_i}\phi_i(t_1,t_2,\cdots,t_d)\xi_i \;, \tag{1.4.3}$$

where $\{\xi_i\}$ is a series of independent process with mean 0 and variance 1.

In general, the analytical solution to the spatial process $X(\mathbf{t})$ is very difficult to obtain, since the Fredholm integral equation involves d-times integration. However, for a certain class of covariance function, which is separable, i.e. for $s_i, t_i \in \mathcal{T}_i$

$$K(\mathbf{s},\mathbf{t}) = \mathrm{cov}[X(s_1,s_2,\cdots,s_d),X(t_1,t_2,\cdots,t_d)] = \prod_{i=1}^{d}\mathrm{cov}[X(s_i),X(t_i)] = \prod_{i=1}^{d}K(s_i,t_i)$$

the eigenvalue and the eigenfunction for $X(\mathbf{t})$ are also separable. Both are the product of their univariate counterparts in $X(t_i), 1 \leq i \leq d$. This result is summarised in theorem 1.4.1.

Theorem 1.4.1. *Let $X(\mathbf{t}) \in H_X$ be a spatial process, where $\mathbf{t}$ is a vector time and $\mathbf{t} \in \mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2 \times \cdots \times \mathcal{T}_d$. Assume that the covariance function $K(\mathbf{t}, \mathbf{s})$ is separable, and the univariate Karhunen-Loeve expansion for $X(t_i)$, $t_i \in \mathcal{T}_i$ is*

$$X(t_i) = \sum_{k \geq 1} \sqrt{\lambda_k^{(i)}} \phi_k^{(i)}(t_i) \xi_k^{(i)} \quad t_i \in \mathcal{T}_i \tag{1.4.4}$$

Then the multivariate Karhunen-Loeve expansion for $X(\mathbf{t})$ is

$$X(\mathbf{t}) = X(t_1, t_2, \cdots, t_d) = \sum_{k \geq 1} \sqrt{\lambda_k} \phi_k(t_1, t_2, \cdots, t_d) \xi_k \tag{1.4.5}$$

where

$$\lambda_k = \prod_{j=1}^{d} \lambda_{i_j}^{(j)}, \quad \phi_k(t_1, t_2, \cdots, t_d) = \prod_{j=1}^{d} \phi_{i_j}^{(j)}(t_j) \quad i_j \in N, 1 \leq j \leq d$$

Proof. Assume the above eigenvalues and eigenfunctions are the solutions, then it only needs to be shown that the orthogonality of the eigenfunctions and that both satisfy the Fredholm integral equation. Due to the one-to-one relationship between the Fredholm integral equation and the Karhunen-Loeve expansion, the expansion listed above for $X(\mathbf{t})$ can then be proved. For $t_j, s_j \in \mathcal{T}_j$ and $i_j \in N, 1 \leq j \leq d$

$$\int_{\mathcal{T}_1} \int_{\mathcal{T}_2} \cdots \int_{\mathcal{T}_d} \mathrm{cov}[X(s_1, s_2, \cdots, s_d), X(t_1, t_2, \cdots, t_d)] \phi_k(s_1, s_2, \cdots, s_d) ds_1 ds_2 \cdots ds_d$$

$$= \int_{\mathcal{T}_1} \int_{\mathcal{T}_2} \cdots \int_{\mathcal{T}_d} \prod_{j=1}^{d} K(t_j, s_j) [\prod_{j=1}^{d} \phi_{i_j}^{(j)}(s_j)] ds_1 ds_2 \cdots ds_d$$

$$= \prod_{j=1}^{d} [\int_{\mathcal{T}_j} K(t_j, s_j) \phi_{i_j}^{(j)}(s_j) ds_j] = \prod_{j=1}^{d} [\lambda_{i_j}^{(j)} \phi_{i_j}^{(j)}(t_j)]$$

$$= [\prod_{j=1}^{d} \lambda_{i_j}^{(j)}][\prod_{j=1}^{d} \phi_{i_j}^{(j)}(t_j)]$$

Also, for $k_1, k_2, m_j, n_j \in N$, $1 \leq j \leq d$ and

$$\phi_{k_1}(s_1, s_2, \cdots, s_d) = \prod_{j=1}^{d} \phi_{m_j}^{(j)}(s_j), \quad \phi_{k_2}(s_1, s_2, \cdots, s_d) = \prod_{j=1}^{d} \phi_{n_j}^{(j)}(s_j) \ , \tag{1.4.6}$$

the orthogonality condition can be proved, since

$$\begin{aligned}
&\int_{\mathcal{T}_1}\int_{\mathcal{T}_2}\cdots\int_{\mathcal{T}_d}\phi_{k_1}(s_1,s_2,\cdots,s_d)\phi_{k_2}(s_1,s_2,\cdots,s_d)ds_1ds_2\cdots ds_d \\
= &\int_{\mathcal{T}_1}\int_{\mathcal{T}_2}\cdots\int_{\mathcal{T}_d}\prod_{j=1}^{d}\phi_{m_j}^{(j)}(s_j)\prod_{j=1}^{d}\phi_{n_j}^{(j)}(s_j)ds_1ds_2\cdots ds_d \\
= &\prod_{j=1}^{d}[\int_{\mathcal{T}_j}\phi_{m_j}^{(j)}(s_j)\phi_{n_j}^{(j)}(s_j)ds_j] = \prod_{j=1}^{d}\delta_{m_j,n_j} \\
= &\delta_{k_1,k_2}
\end{aligned}$$

Hence the eigenfunctions are orthogonal. Together with eigenvalues, they satisfy the Fredholm integral equation. □

Example: The Brownian sheet in $[0,1]$

The Brownian sheet $X(t,s), t,s\in[0,1]$, is a biparameter Gaussian random field with mean 0 and covariance function

$$\text{cov}[X(t_1,s_1),X(t_2,s_2)] = \min(s_1,s_2)\min(t_1,t_2) \tag{1.4.7}$$

It can be seen from the covariance function that in the two coordinate directions, a slice of the Brownian sheet looks like a Brownian motion multiplying by a constant. Since the covariance function is separable, we can use the result of the univariate Brownian motion to express the Karhunen-Loeve expansion for the Brownian sheet.

$$X(t,s) = \sum_{k\geq 1}\sqrt{\lambda_k}\phi_k(t,s)\xi_k \tag{1.4.8}$$

where $\xi_k \sim i.i.dN(0,1)$ and

$$\lambda_k = \frac{16}{[(2i-1)^2\pi^2][(2j-1)^2\pi^2]}, \quad \phi_k(t,s) = 2\sin[(i-\frac{1}{2})\pi t]\sin[(j-\frac{1}{2})\pi s] \quad i,j\geq 1 \tag{1.4.9}$$

Chapter 2

Computational Methods

As explained in section 1.2, chapter 1, in order to derive the Karhunen-Loeve expansion for the stochastic process $\{X(t), t \in \mathcal{T}\}$, it is often necessary to solve the Fredholm integral equation

$$\int_{\mathcal{T}} K(s,t)\phi_i(t)dt = \lambda_i \phi_i(s) \tag{2.0.1}$$

where $K(s,t)$ is the covariance function for $X(s)$ and $X(t)$

However, the analytical solution to the integral equation only exists for particular covariance functions. Numerical solutions often have to be sought in practice. This chapter deals with two commonly used numerical methods, the integral method and expansion method. The integral method proves to be computationally easy and quite fast (Ramsay and Silverman (1997)). However, it does not treat the eigenfunction as a real function, since it only discretises the time interval $\mathcal{T}$ into a finite number of small intervals and approximates the eigenfunction interval by interval. The expansion method, on the other hand, expands the eigenfunction as a function using certain bases. With suitable basis functions, the expansion methods are computationally efficient and provide quite good approximations.

2.1 Integral method

This method focuses on a direct approximation for the integral, i.e.

$$\int_0^1 f(t)dt \approx \sum_{i=0}^{n+1} \omega_i f(t_i) \tag{2.1.1}$$

To use the formula, we only need to know the discretised points t_i ($0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq t_{n+1} = 1$ and n is the total number of these points) and the weights ω_i.

A simple example is to discretise the integral range $[0, 1]$ into equally spaced intervals, i.e. $t_{j+1} - t_j = h = \frac{1}{n+1}$, $0 \leq j \leq n$. The length of the interval h is usually chosen to be very small and n very big, for accuracy. For each $f(t_i)$, the weights w_i are the same, i.e. $w_i = \frac{1}{n+2}$. This is called the *uniform* scheme. Under the uniform scheme,

$$\int_0^1 f(t)dt \approx \frac{1}{n+2} \sum_{i=0}^{n+1} f(t_i) \tag{2.1.2}$$

Another frequently used integral numerical scheme is the *trapezium* scheme, it is also based on n equally spaced points. Assume that $t_{j+1} - t_j = h = \frac{1}{n+1}$, $0 \leq j \leq n$, then

$$\begin{aligned} \int_0^1 f(t)dt &\approx \sum_{i=0}^{n} \frac{h}{2}(f(t_i) + f(t_{i+1})) \\ &= h(\frac{f(t_0)}{2} + f(t_1) + \cdots + f(t_{n-1}) + \frac{f(t_{n+1})}{2}) \end{aligned}$$

Other approaches, like the Gaussian quadrature approximation (see, for example, Cheney and Kincaid (2007), for detail on other numerical methods), greatly increase the accuracy of the numerical calculation. However a special placement of the discretised points needs to be considered, which is often not suitable in practice. In practical

applications, observation points are usually equally spaced, for example when data arise as a time series.

The Fredholm integral equation can be approximated as follows:

$$\int_0^1 K(s,t)\phi_i(t)dt \approx \sum_j K(s,t_j)\phi_i(t_j)\omega_j = \lambda_i\phi_i(s) \tag{2.1.3}$$

Furthermore, the orthogonality of the eigenfunctions ϕ_i means that

$$\int_0^1 \phi_i(t)\phi_j(t)dt = \delta_{ij} \approx \sum_k \phi_i(t_k)\phi_j(t_k)\omega_k \tag{2.1.4}$$

When written in the matrix form, the above integral equation can be written as

$$KW\phi_i = \lambda_i\phi_i \tag{2.1.5}$$

where

$$K = \begin{pmatrix} \mathrm{cov}(X(t_0),X(t_0)) & \mathrm{cov}(X(t_0),X(t_1)) & \cdots & \mathrm{cov}(X(t_0),X(t_{n+1})) \\ \mathrm{cov}(X(t_1),X(t_0)) & \mathrm{cov}(X(t_1),X(t_1)) & \cdots & \mathrm{cov}(X(t_1),X(t_{n+1})) \\ \cdots & & & \\ \mathrm{cov}(X(t_{n+1}),X(t_0)) & \mathrm{cov}(X(t_{n+1}),X(t_1)) & \cdots & \mathrm{cov}(X(t_{n+1}),X(t_{n+1})) \end{pmatrix}$$

$$W = \begin{pmatrix} w_0 & 0 & \cdots & 0 \\ 0 & w_1 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & w_{n+1} \end{pmatrix}$$

$$\phi_i^T = \begin{pmatrix} \phi_i(t_0), & \phi_i(t_1), & \cdots, & \phi_i(t_{n+1}) \end{pmatrix}$$

Defining $u_i = W^{1/2}\phi_i$, we obtain a symmetric eigenvalue problem of the form

$$W^{\frac{1}{2}}KW^{\frac{1}{2}}u_i = \lambda_i u_i \tag{2.1.6}$$

The vector ϕ_i is then recomputed via the inverse transformation $\phi_i = W^{-\frac{1}{2}}u_i$. Notice that using a uniform scheme, W is simply $\frac{1}{n+2}I$, where I is the identity matrix, while

using the slightly different trapezium scheme,

$$W = \begin{pmatrix} \frac{1}{2(n+1)} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{n+1} & \cdots & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & \cdots & \frac{1}{n+1} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{2(n+1)} \end{pmatrix} \tag{2.1.7}$$

2.2 Expansion method

Numerical methods, such as the Galerkin method (see, for example, Porter and Stirling (1990) for detail), usually employ certain basis function to expand the eigenfunctions up to M terms. The advantage of the Galerkin method over the integration method is that the eigenfunction is treated as a function and the accuracy of the approximation can be controlled by choosing sufficiently many basis terms. The commonly used basis functions are the trigonometrical functions or the wavelet functions.

Algorithm 1: the Fourier method

Using the Fourier basis to expand the eigenfunction is one of the first expansion methods that researchers have investigated. See, for example, Ghanem and Spanos (1991) and Huang et al. (2001). The calculation process is as follows.

Firstly, a set of M adequate basis functions $\{\theta_i(t), i = 1, 2, \cdots, M\}$ is chosen. For the Fourier basis defined on $[0, 1]$, M is chosen to be odd, so that the basis functions can be written down as

$$\theta_1(t) = 1, \theta_2(t) = \cos(2\pi t), \theta_3(t) = \sin(2\pi t), \cdots,$$
$$\theta_{2i}(t) = \cos(2\pi i t), \theta_{2i+1}(t) = \sin(2\pi i t), i = 1, 2, \cdots, \frac{M-1}{2}$$

Then, the eigenfunction is expanded as a linear combination of the basis functions up to term M

$$\phi_i(t) = \sum_{k=1}^{M} d_{ik}\theta_k(t) = \theta(t)^T D_i = D_i^T \theta(t) \ , \tag{2.2.1}$$

where

$$\begin{aligned} \theta(t)_{M\times 1}^T &= \Big(\theta_1(t),\ \theta_2(t),\ \cdots,\ \theta_M(t)\Big) \\ D_{i,M\times 1}^T &= \Big(d_{i1},\ d_{i2},\cdots,d_{iM}\Big) \ , \end{aligned}$$

while $\{d_{ik}\}$ are the unknown coefficients for this expansion.

Substitute (2.2.1) into the integral equation on the ith eigenvalue and the ith eigenfunction,

$$\int_0^1 K(s,t)\phi_i(s)ds = \lambda_i\phi_i(t) \Leftrightarrow D_i^T \int_0^1 K(s,t)\theta(s)ds = D_i^T\lambda_i\theta(t) \tag{2.2.2}$$

Integrate both sides of (2.2.2) with respect to t after multiplying both sides by $\theta(t)$:

$$D_i^T \int_0^1\int_0^1 K(s,t)\theta(s)\theta^T(t)dsdt = D_i^T\lambda_i \int_0^1 \theta(t)\theta^T(t)dt \tag{2.2.3}$$

Using the first p eigenvalues and eigenfunctions, equation 2.2.3 can be written in matrix form:

$$DA = \Lambda DB \Leftrightarrow AD^T = BD^T\Lambda \tag{2.2.4}$$

where

$$\begin{aligned} A &= \int_0^1 K(s,t)\theta(s)\theta^T(t)dsdt \qquad (A_{jk} = \int_0^1\int_0^1 K(s,t)\theta_k(s)\theta_j(t)dsdt) \\ B &= \int_0^1 \theta(t)\theta^T(t)dt \qquad (B_{jk} = \int_0^1 \theta_k(t)\theta_j(t)dt) \\ D_{p\times M} &= \begin{pmatrix} D_1^T \\ D_2^T \\ \cdots \\ D_p^T \end{pmatrix}, \qquad \Lambda_{p\times p} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \end{aligned}$$

This is an example of the so-called the generalised algebraic eigenvalue problem or the generalised eigenvalue problem. See, for example, chapter 12, Laub (2004) for details. Since A is a symmetric positive definite matrix and B is a diagonal positive matrix, the generalised eigenvalue problem can be solved in the following way. Express B as $B^{\frac{1}{2}}B^{\frac{1}{2}}$, where $B^{\frac{1}{2}}$ is the usual symmetric square root, then

$$AD^T = B^{\frac{1}{2}}B^{\frac{1}{2}}D^T\Lambda \tag{2.2.5}$$

Multiply both sides of equation 2.2.5.

$$B^{-\frac{1}{2}}AB^{-\frac{1}{2}}B^{\frac{1}{2}}D^T = B^{\frac{1}{2}}D^T\Lambda \tag{2.2.6}$$

Assume $E = B^{\frac{1}{2}}D^T$, then equation 2.2.6 can be simplified to

$$B^{-\frac{1}{2}}AB^{-\frac{1}{2}}E = E\Lambda \tag{2.2.7}$$

The eigenvalue Λ and the eigenfunction E for $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ can now be computed using equation 2.2.7. Then convert E to D via $D = E^TB^{-\frac{1}{2}}$.

Although using the Fourier basis provides more flexibility than the integration methods, the accuracy in estimating the eigenfunctions is worse (see chapter 2, Ghanem and Spanos (1991)). Moreover, in practice it takes a relatively longer computational time, even for small M.

After obtaining the ith eigenvalue λ_i and the ith eigenfunction $\phi_i(t)$, we can now construct the truncated covariance function at order p.

$$K(s,t) = \sum_{i=1}^{p} \lambda_i\phi_i(s)\phi_i(t) = \phi(s)^T\Lambda\phi(t) = \theta(s)^TD^T\Lambda D\theta(t) \ , \tag{2.2.8}$$

where

$$\phi(t)_{p\times 1}^T = \Big(\phi_1(t), \ \ \phi_2(t), \ \ \cdots, \ \ \phi_p(t)\Big) \tag{2.2.9}$$

Algorithm 2: the Haar wavelet method

Using the wavelet basis to expand the eigenfunctions was mentioned in Ramsay and Silverman (1997) and detailed calculations appear in Phoon et al. (2002b) and Phoon et al. (2002a). Here, the algorithm from Phoon et al. (2002b) is adapted to both the stochastic process satisfying Sacks-Ylvisaker condition and the stochastic process with the smooth covariance function, such as the squared exponential kernel. For simplicity, only the bounded interval $[0, 1]$ is considered again. Any function f between $[0, 1]$ can be expanded using the wavelet basis function $\psi_i(x)$

$$f(x) = a_0 \times 1 + \sum_{i=1}^{\infty} a_i \psi_i(x) \tag{2.2.10}$$

The Haar wavelet is the simplest wavelet basis function in the Daubechies's family. See, for example, Nievergelt (1999), for details of the Daubechies's family in wavelets. The Haar wavelet is defined as

$$\psi(x) = \begin{cases} 1 & 0 < x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.2.11}$$

A family of the orthogonal Haar wavelets can be constructed through shifting the above $\psi(x)$, i.e. for $j, k \in Z$

$$\psi_{j,k}(x) = \begin{cases} 1 & k2^{-j} < x < 2^{-j-1} + k2^{-j} \\ -1 & 2^{-j-1} + k2^{-j} \leq x < 2^{-j} + k2^{-j} \\ 0 & \text{otherwise} \end{cases} \tag{2.2.12}$$

They are orthogonal because

$$\int_0^1 \psi_{j,k}(x)\psi_{m,n}(x)dx = 2^{-j}\delta_{j,m}$$

Now, $M = 2^n$ orthogonal basis functions on $[0, 1]$ can be constructed in the following way

$$\psi_1 = 1; \qquad \psi_i = \psi_{j,k}(x) \quad i = 2^j + k + 1; \; j = 0, 1, \cdots, n-1; \; k = 0, 1, \cdots 2^j - 1 \tag{2.2.13}$$

Hence the eigenfunction and the covariance function can be expressed using the above basis function, i.e.

$$\phi_i(t) = \sum_{k=1}^{M} d_{ik}\psi_k(t) = \Psi^T(t)D_i \tag{2.2.14}$$

$$K(s,t) = \sum_{m=1}^{M}\sum_{n=1}^{M} a_{mn}\psi_m(s)\psi_n(t) = \Psi(s)^T A\Psi(t) \tag{2.2.15}$$

where

$$\Psi(t)^T_{M\times 1} = \Big(\psi_1(t), \;\; \psi_2(t), \;\; \cdots, \;\; \psi_M(t)\Big)$$
$$D^T_{i,M\times 1} = \Big(d_{i1}, \;\; d_{i2}, \;\; \cdots, \;\; d_{iM}\Big)$$
$$A_{M\times M} = \{a_{ij}\}$$

The matrix A can be derived through the 2D wavelet transform. In order to perform the wavelet transform, M time points need to be chosen at $t_i = \frac{2i-1}{2M}, 1 \le i \le M$, and M should be a number a power of two. In application, when the covariance function is known, M points satisfying the above conditions can always be chosen.

The integral equation involving the ith eigenvalue and the ith eigenfunction can now be written down as

$$\lambda_i\phi_i(t) = \int_0^1 K(s,t)\phi_i(s)ds \tag{2.2.16}$$

The expressions of $\phi_i(t)$ and $K(s,t)$ in equation 2.2.14 and equation 2.2.15 respectively are substituted into equation 2.2.16.

$$\lambda_i\Psi^T(t)D_i = \Psi^T(t)AHD_i \;\;, \tag{2.2.17}$$

where

$$H = \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & h_M \end{pmatrix}$$

$h_1 = 1, h_i = 2^{-j}$,

$i = 2^j + k + 1, j = 0, 1, \cdots, n-1$ and $k = 0, 1, \cdots 2^j - 1$

Combining the first p eigenvalues and eigenfunctions, we can express equation 2.2.9 in matrix form.

$$\Lambda_{p\times p} D_{p\times M} \Psi(t)_{M\times 1} = D_{p\times M} H_{M\times M} A_{M\times M} \Psi(t)_{M\times 1} \tag{2.2.18}$$

where

$$D_{p\times M} = \begin{pmatrix} D_1^T \\ D_2^T \\ \cdots \\ D_p^T \end{pmatrix}, \qquad \Lambda_{p\times p} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \tag{2.2.19}$$

Equating the coefficients for $\Psi(t)$, we obtain $\Lambda D = DHA$. Multiply by $H^{\frac{1}{2}}$ on both sides,

$$\Lambda D H^{\frac{1}{2}} = D H^{\frac{1}{2}} H^{\frac{1}{2}} A H^{\frac{1}{2}} \;\Leftrightarrow\; \Lambda \hat{D} = \hat{D} \hat{A} \ , \tag{2.2.20}$$

where $\hat{D} = DH^{\frac{1}{2}}$, $\hat{A} = H^{\frac{1}{2}} A H^{\frac{1}{2}}$ and $\hat{A}$ is a symmetric matrix. Solving equation 2.2.20 results in M eigenvalues and M orthogonal, linearly independent eigenvectors. The eigenfunction can now be written as

$$\phi(t) = D\Psi(t) = \hat{D} H^{-\frac{1}{2}} \Psi(t) \tag{2.2.21}$$

In terms of the truncated covariance function.

$$K(s,t) = \Psi(s)^T D^T \Lambda D \Psi(t) \tag{2.2.22}$$

2.3 Numerical examples

The Ornstein-Uhlenbeck process

The covariance function for the Ornstein-Uhlenbeck process is given in chapter 1, $\frac{\rho^2}{2\beta}\exp(-\beta|t-s|)$. For simplicity, it is further assumed that $\beta = 1$ and $\rho = \sqrt{2}$. So that the covariance function is $\exp(-|t-s|)$. Figure 2.1 plots the solution to the eigenvalues. From section 1.3, chapter 1, it is known that for the eigenvalues, $\cot(w) = \frac{w^2-1}{2w}$ is needed to be solved, and then we convert w to λ using the formula $\lambda = \frac{2}{w^2+1}$. Since the negative and the positive solution for w produce the same value of λ, figure 2.1 only covers the positive part of w. In the left plot of figure 2.1, the blue line plots $\cot(w)$, while the red line plots $\frac{w^2-1}{2w}$. It can been seen that at each interval $[k\pi, (k+1)\pi], k \in \mathbb{N}$, there exists an intersection, which represents one solution to w. Since k is an integer, the total number of intersections should be countable and the bigger the k is, the closer is the intersection to $k\pi$. The right plot of figure 2.1 plots $\lambda = \frac{2}{w^2+1}$, which reduces to zero very quickly. For positive w, the relationship between λ and w is one-to-one. One can see in this way how a countable number of eigenvalues is obtained.

Figures 2.2, 2.3 and 2.4 provide an idea of the performance of the analytical solution to the covariance function $\exp(-|t-s|)$ under different orders. This is helpful in suggesting a suitable order for the truncated Karhunen-Loeve expansion. Figure 2.2 shows the convergence in the Karhunen-Loeve expansion compared to the targeted covariance function. Define the time lag as $h = |t-s|$. Since the Ornstein-Uhlenbeck is a stationary process, we only need to plot the covariance for the targeted or the truncated covariance using the Karhunen-Loeve expansion versus h to study the convergence. It can be seen that the difference between order 10 and order 30 is

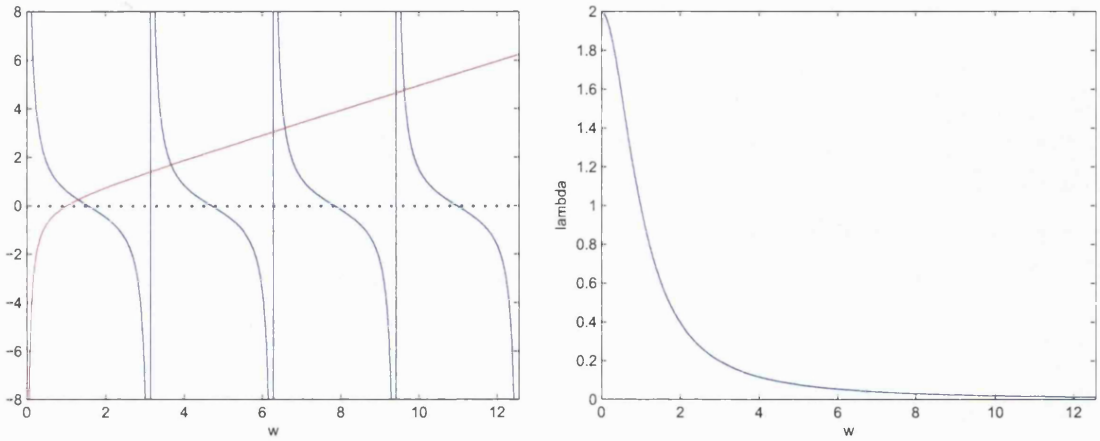

Figure 2.1: Graphical solution to $\cot(w) = \frac{w^2-1}{2w}$. (Left) Blue line $y = \cot(w)$; Red line $y = \frac{w^2-1}{2w}$. (Right) $\lambda = \frac{2}{w^2+1}$.

much bigger than that of order 30 and order 50.

Figure 2.3 serves to check stationarity for the different starting point t, $0 \leq t \leq 0.5$ under the same time lag $h = 0.5$. Theoretically, the targeted covariance function without the truncation should be a constant under the same h whatever the starting point t is. It should also be expected that the better the approximation using the truncated Karhunen-Loeve expansion is, the closer the truncated covariance is to a constant whatever the staring point t is. Again, it can be seen from the plot that the higher the order is, the better is the performance of the approximation. It is clearer from this plot that there is some improvement when the order is increased from 30 to 50, although not very big.

Figure 2.4 explains the cumulative expected variance preserved in the expansion. The cumulative expected variance is defined as

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \tag{2.3.1}$$

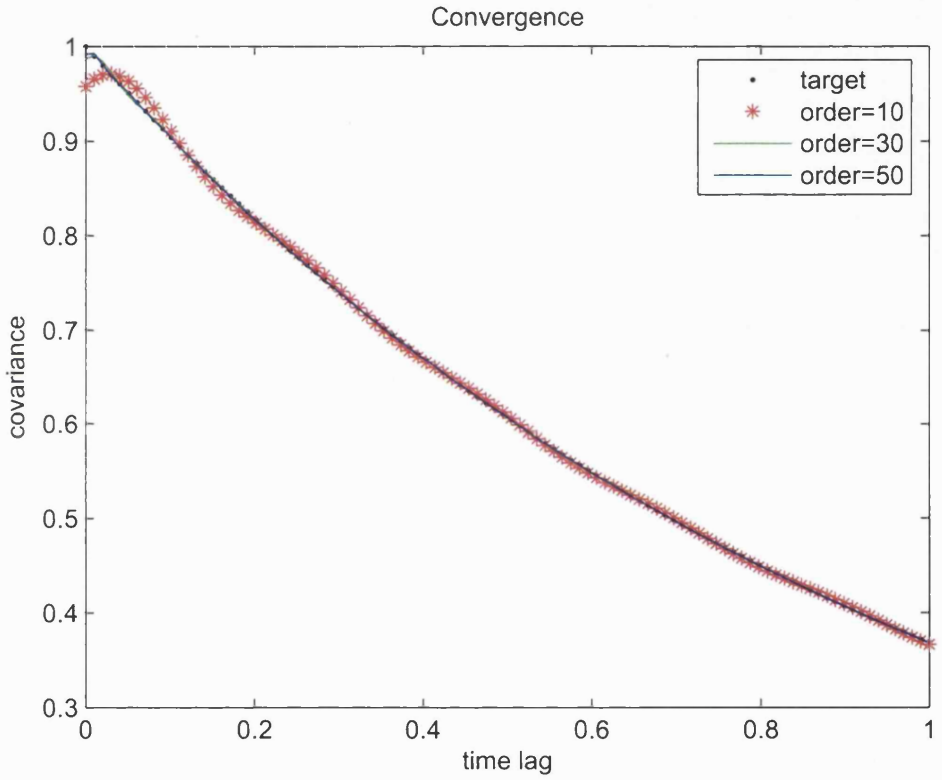

Figure 2.2: The analytical performance of the truncated Karhunen-Loeve expansion for the kernel $\exp(-|t-s|)$: covariance function versus time lag.

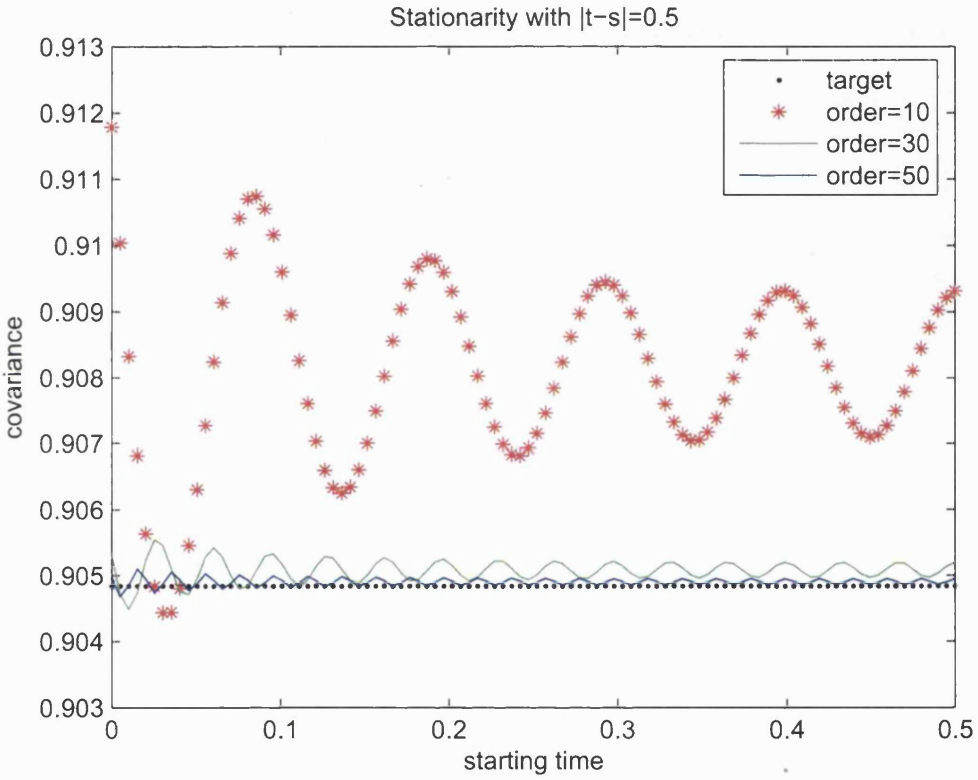

Figure 2.3: The analytical performance of the truncated Karhunen-Loeve expansion for the kernel $\exp(-|t-s|)$: stationarity checking when $|t-s| = 0.5$ for different starting points.

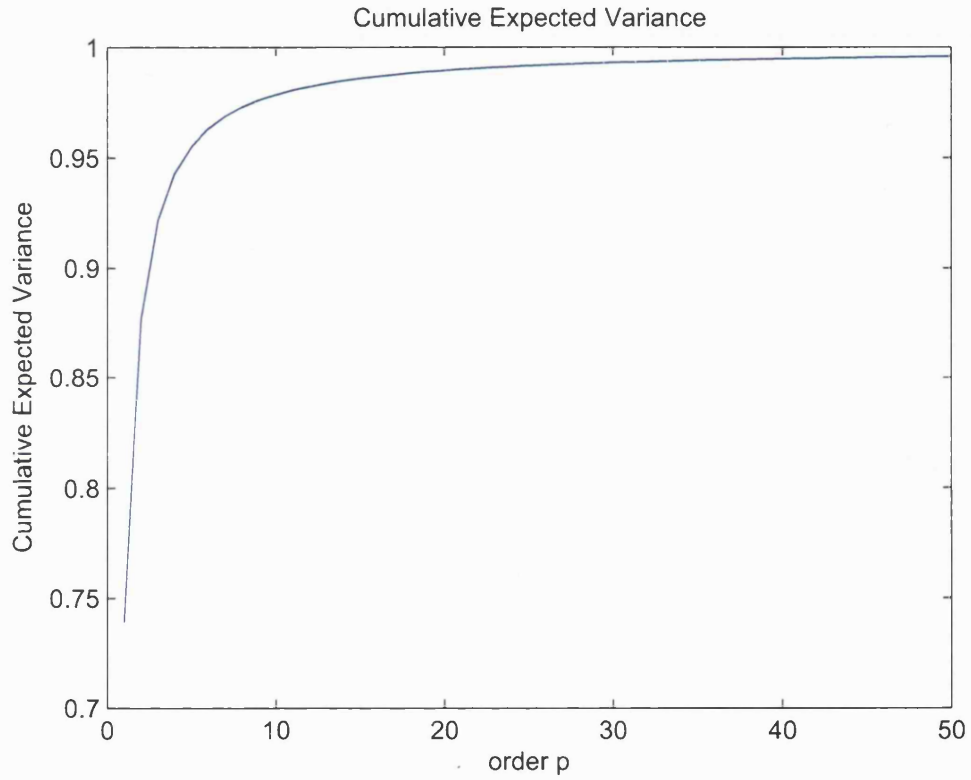

Figure 2.4: The analytical performance of the truncated Karhunen-Loeve expansion for the kernel $\exp(-|t-s|)$: cumulative expected variance versus order in the truncated Karhunen-Loeve expansion.

If the original stochastic process $X(t)$ and the truncated stochastic process $X_p(t)$ using the truncated Karhunen-Loeve expansion at order p are expressed respectively as

$$X(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i\xi_i \quad X_p(t) = \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i\xi_i \tag{2.3.2}$$

where ξ_i is i.i.d. with mean 0 and variance 1, the integrated variance using the untruncated expression and the truncated expression can be calculated as

$$\sum_{i=1}^{\infty} \lambda_i = \int_{\mathcal{T}} \mathrm{Var}(X(t))dt \quad \sum_{i=1}^{p} \lambda_i = \int_{\mathcal{T}} \mathrm{Var}(X_p(t))dt \tag{2.3.3}$$

Therefore, the cumulative expected variance can also be expressed as

$$\frac{\int_{\mathcal{T}} \mathrm{Var}(X_p(t))dt}{\int_{\mathcal{T}} \mathrm{Var}(X(t))dt} \tag{2.3.4}$$

For the Ornstein-Uhlenbeck process on $[0, 1]$, $\int_0^1 \mathrm{Var}(X(t))dt = 1$. Hence the cumulative expected variance is only the numerator of equation 2.3.4, $\int_0^1 \mathrm{Var}(X_p(t))dt$. It can be seen that the rate of the change of the cumulative expected variance decreases when the order p increases. When the order p approaches 50, the cumulative expected variance is quite close to 1.

Now, the analytical solution to the eigenvalues is compared with that derived from the integral method, the Fourier method and the Haar wavelet method respectively. The result is summarised in table 2.1. The error in the table is calculated as follows.

$$\text{Error} = \left| \frac{\text{Numerical Solution Using Certain Numerical Scheme} - \text{Analytical Solution}}{\text{Analytical Solution}} \right| \times 100\% \tag{2.3.5}$$

From table 2.1, it can be seen that in the integration methods, the trapezium scheme performs much better than the uniform scheme. The error dramatically decreases under the same number of points. The reason for using 11 basis functions in

	1st	2nd	3rd	4th	5th
Analytical	0.7388108	0.1380038	0.0450885	0.0213289	0.0122789
Uniform ($n = 200$)	0.7378086	0.1384447	0.0452880	0.0214326	0.0123420
Error	0.14%	0.32%	0.44%	0.49%	0.51%
Trapezium ($n = 200$)	0.7388105	0.1380060	0.0450919	0.0213327	0.0122828
Error	0.00%	0.00%	0.01%	0.02%	0.03%
Fourier ($M = 11$)	0.7388075	0.1299146	0.0451686	0.0198488	0.0122810
Error	0.00%	5.86%	0.18%	6.94%	0.02%
Haar ($M = 256$)	0.7388147	0.1380069	0.0450912	0.0213316	0.0122815
Error	0.00%	0.00%	0.01%	0.01%	0.02%

Table 2.1: The first five eigenvalues comparison among different numerical schemes for the kernel $\exp(-|t - s|)$.

the Fourier method is that it is computationally expensive for higher orders. It takes the computer a longer time than the time involved in any other method, even using 11 basis functions. Using the Fourier basis, the result is quite volatile, with the error as big as 5.86%, which is the biggest among all the methods. Using the Haar wavelet method, on the other hand, is computationally fast, although 256 basis functions are used. Its error is the smallest among all the methods. Although the results from the trapezium integral method and that from the Haar wavelet method are quite good and close to each other, it is the Haar method that regards the eigenfunction as a function. The accuracy of the Haar wavelet method can be controlled by the number of the basis functions involved.

Since the Haar wavelet method provides the result both computationally reliable and accurate, we take a further look at its result in terms of different number of basis functions involved. This result is shown in table 2.2.

It can be seen that the error dramatically decreases when the order needed for the expansion of the eigenfunction is doubled from 8 to 16. At the order of 16, the error is already quite close to that derived from the Fourier method. Although using

	1st	2nd	3rd	4th	5th
Analytical	0.7388108	0.1380038	0.0450885	0.0213289	0.0122789
Haar (M=8)	0.7428125	0.1412251	0.0480043	0.0242416	0.0153110
Error	0.54%	2.33%	6.47%	13.66%	24.69%
Haar (M=16)	0.7398109	0.1388036	0.0458001	0.0220190	0.0129669
Error	0.14%	0.58%	1.58%	3.24%	5.60%
Haar (M=32)	0.7390608	0.1382034	0.0452653	0.0214992	0.0124469
Error	0.03%	0.14%	0.39%	0.80%	1.37%
Haar (M=64)	0.7388733	0.1380537	0.0451326	0.0213713	0.0123207
Error	0.01%	0.04%	0.10%	0.20%	0.34%
Haar (M=128)	0.7388264	0.1380162	0.0450995	0.0213396	0.0122893
Error	0.00%	0.01%	0.02%	0.05%	0.08%

Table 2.2: The first five eigenvalues comparison of the Haar wavelet scheme using different number of basis functions for $\exp(-|t-s|)$.

the order 32, 64 and 128 provide slightly worse results than that from the integral method, especially the result using the trapezium scheme, it takes less computational time. Hence in practice, when the covariance function is known, it is reasonable to use lower orders when applying the Haar wavelet method.

Now consider the eigenfunctions. In the integration method, only the trapezium scheme is considered, since it is in general better than the uniform scheme.

The first two eigenfunctions in figure 2.5 and figure 2.6 correspond to the first two biggest eigenvalues. It can be seen that both the trapezium integration method and the Haar wavelet method perform quite well. The eigenfunctions from these two methods are almost the same as those from the analytical method. However, the eigenfunctions derived from the Fourier method are not smooth and exhibit oscillation. This oscillation might be caused by the cyclical variation in the trigonometric function. The oscillated eigenfunctions might cause problems in the reconstruction of the covariance function. Now the values of the first two eigenfunctions at the

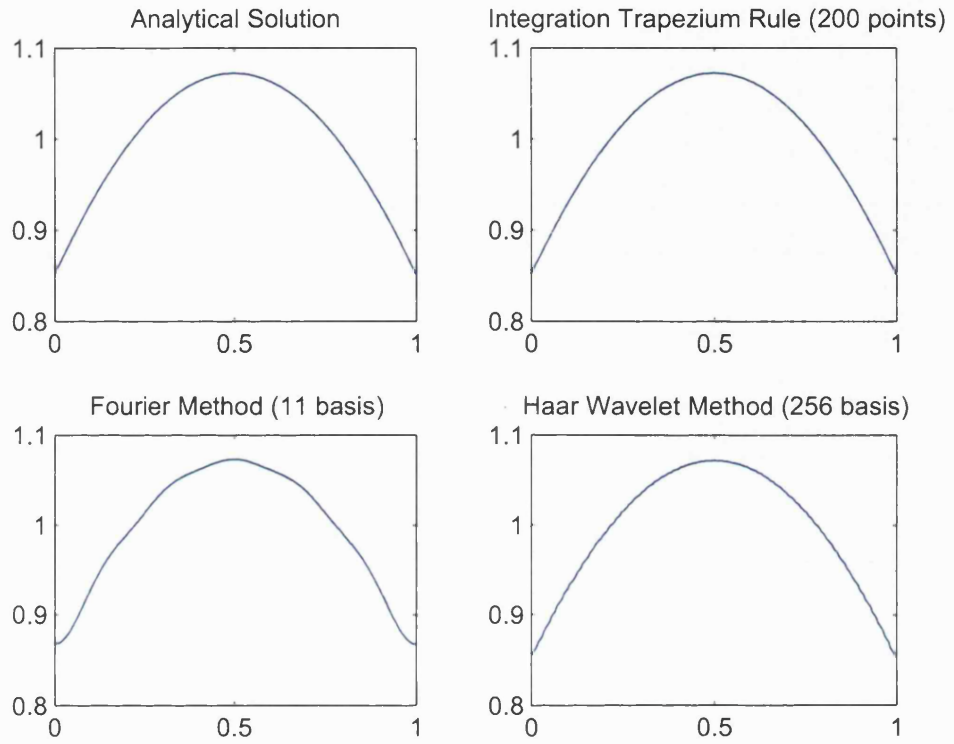

Figure 2.5: Comparison of the first eigenfunction between the analytical solution and three other numerical schemes for $\exp(-|t-s|)$.

beginning few time points are listed in table 2.3 and table 2.4.

Notice that although both the Haar wavelet method and the integral trapezium method provide similar results, the value provided by the integral trapezium method is more accurate. The difference between the analytical value and the value by the integral trapezium method is of the order 10^{-6}, while the difference between the analytical value and the value by the Haar wavelet appears at around 10^{-3}. This small difference 10^{-3} might affect our analysis when the accuracy needed is smaller

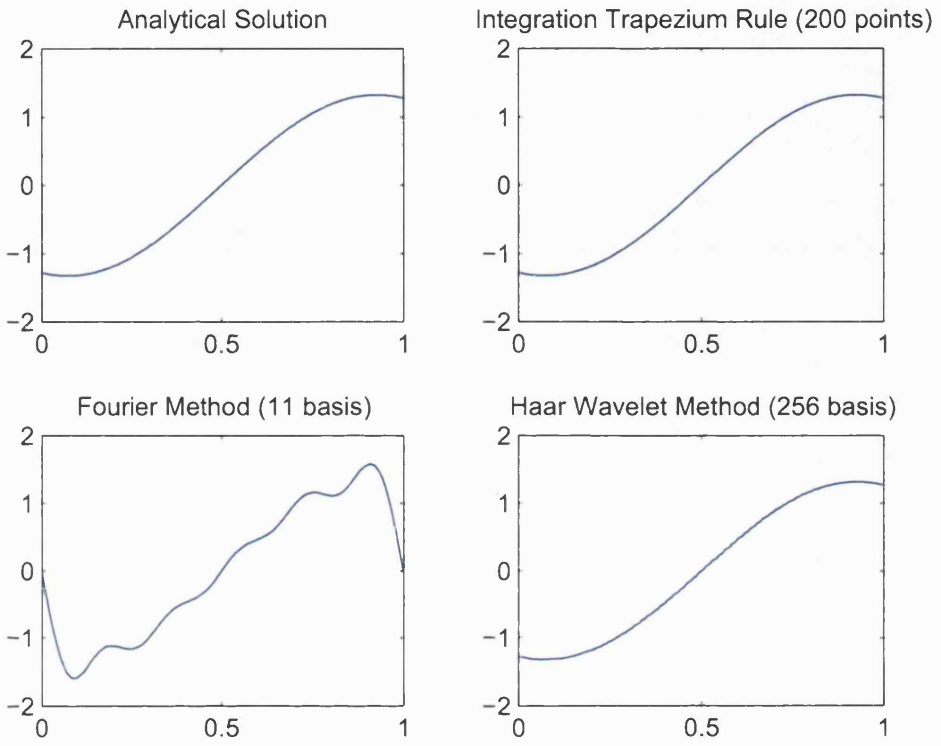

Figure 2.6: Comparison of the second eigenfunction between the analytical solution and three other numerical schemes for $\exp(-|t-s|)$.

than or equal to 10^{-3}. For the Fourier basis, the result is quite unstable. For the second eigenfunction listed in table 2.4, the error in the first few time points could be as big as above 100%, while the error using other methods are less than 1% for the same time points.

Further checking of the eigenfunctions can be implemented by introducing two

Time	0	$\frac{1}{202}$	$\frac{2}{202}$	$\frac{3}{202}$	$\frac{4}{202}$
Analytical	0.8516555	0.8558746	0.8600575	0.8642040	0.8683141
Trapezium ($n = 200$)	0.8516566	0.8558757	0.8600587	0.8642053	0.8683154
Error	0.00%	0.00%	0.00%	0.00%	0.00%
Fourier ($M = 11$)	0.8672837	0.8675013	0.8681515	0.8692274	0.8707174
Error	1.84%	1.36%	0.94%	0.58%	0.28%
Haar ($M = 256$)	0.8533158	0.8566203	0.8599025	0.8631624	0.8696145
Error	0.19%	0.09%	0.02%	0.12%	0.15%

Table 2.3: The first eigenfunction comparison for the first few time points for $\exp(-|t-s|)$.

Time	0	$\frac{1}{202}$	$\frac{2}{202}$	$\frac{3}{202}$	$\frac{4}{202}$
Analytical	-1.2791384	-1.2852884	-1.2910091	-1.2962986	-1.3011553
Trapezium ($n = 200$)	-1.2791395	-1.2852898	-1.2910109	-1.2963008	-1.3011579
Error	0.00%	0.00%	0.00%	0.00%	0.00%
Fourier ($M = 11$)	0	-0.147933	-0.2944067	-0.4379873	-0.577291
Error	150.19%	132.89%	115.88%	99.32%	83.36%
Haar ($M = 256$)	-1.2816094	-1.2863361	-1.2908039	-1.295006	-1.3026096
Error	0.29%	0.12%	0.02%	0.15%	0.17%

Table 2.4: The second eigenfunction comparison for the first few time points for $\exp(-|t-s|)$.

$l_{1,i}$	1	2	3	4	5	6
Trapezium	0.0000029	0.0000067	0.0000075	0.0000074	0.0000082	0.0000074
Fourier ($M = 11$)	0.0000827	0.0194733	0.0000411	0.0231734	0.0005216	0.0141584
Haar ($M = 256$)	0.0000360	0.0001514	0.0002740	0.0003003	0.0001275	0.0004956

Table 2.5: The difference of the eigenfunction between the analytical solution and the numerical solutions for the Ornstein-Uhlenbeck process in terms of l_1 measure.

$l_{2,i}$	1	2	3	4	5	6
Trapezium	0.0000031	0.0000073	0.0000144	0.0000219	0.0000295	0.0000370
Fourier ($M = 11$)	0.0029384	0.2640940	0.0048725	0.3074980	0.0052781	0.3409438
Haar ($M = 256$)	0.0004141	0.0004141	0.0025289	0.0025289	0.0049991	0.0049991

Table 2.6: The difference of the eigenfunction between the analytical solution and the numerical solutions for the Ornstein-Uhlenbeck process in terms of l_2 measure.

measures l_1 and l_2, which are defined as

$$l_{1,i} = \frac{1}{n}\left|\sum_{j=1}^{n}[f_i(t_j) - \hat{f}_i(t_j)]\right|$$

$$l_{2,i} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}[f_i(t_j) - \hat{f}_i(t_j)]^2} \ ,$$

where $f_i(t_j)$ and $\hat{f}_i(t_j)$ represent respectively the analytical solution to the ith eigenfunction and the approximation of the ith eigenfunction using the numerical method at time point t_j, $i \geq 1$ and $1 \leq j \leq n$. Using a statistical interpretation, l_1 can be treated as a measure for the bias, since its power in the summation is 1, while l_2 can be treated as a measure for the standard deviation, since its power in the summation is 2. In this example, $n = 202$ and $t_j = \frac{j-1}{n-1}$, $1 \leq j \leq n$.

Table 2.5 and table 2.6 show l_1 and l_2 of the first six eigenfunctions using three different numerical methods. In terms of l_1, the bias measure, the trapezium integral

method is the best, since all its values of l_1 are between 10^{-6} to 10^{-5}. The second best numerical scheme is the Haar wavelet method, since most of its values of l_1 are between 10^{-4} to 10^{-3}. The difference between these two methods is again around 10^{-3}, which matches the result from table 2.3 and table 2.4. The result from the Fourier method is not stable, since l_1 for its first eigenfunction is as small as 8×10^{-5}, while l_1 for its second eigenfunction is around 2×10^{-2}. l_2, the standard deviation measure, provides similar information to l_1. Although the trapezium integral method is better than the Haar wavelet method in terms of both l_1 and l_2, the Haar wavelet method takes less computational time. Again, the Haar wavelet method also treats the eigenfunction as a function expanded by the wavelets, while the eigenfunction from the integral based method depends on the discretisation of the interval. Hence, when the covariance function is known, the Haar wavelet method is still preferred in the following analysis.

Figure 2.7 shows the reconstruction of the covariance function under three different schemes with order 11. The truncated covariance function $\hat{K}(s,t)$ between time s and time t is expressed as

$$\hat{K}_{11}(s,t) = \sum_{i=1}^{11} \lambda_i \phi_i(s) \phi_i(t) \tag{2.3.6}$$

For the integral method, the Fourier method and the Haar wavelet method, the key equations for deriving the eigenvalues and the eigenfunctions are equation 2.1.5, equation 2.2.4 and equation 2.2.20 respectively. 11 is chosen since it is computationally doable for the Fourier basis and a universal order is preferred when comparing different numerical schemes. The analytical covariance structure is also provided for comparison. It can be seen that the Fourier basis method provides the worst result,

with variations almost everywhere. Again, this is due to the property of the trigonometric function itself. More order can be used to solve this problem. However, it will dramatically increase the computational time. The trapezium integration method and the Haar wavelet method provide the similar results. The difference between the covariance reconstruction using these two numerical methods and the analytical covariance function is further compared using the measure l_1^c and l_2^c, which are defined as

$$\begin{aligned} l_{1,p}^c &= \left| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{\mathrm{cov}[X(t_i), X(t_j)] - \widehat{\mathrm{cov}}_p[X(t_i), X(t_j)]\} \right| \\ l_{2,p}^c &= \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{\mathrm{cov}[X(t_i), X(t_j)] - \widehat{\mathrm{cov}}_p[X(t_i), X(t_j)]\}^2} \ , \end{aligned}$$

where $\mathrm{cov}[X(t_i), X(t_j)]$ represents the analytical covariance function between $X(t_i)$ and $X(t_j)$, while $\widehat{\mathrm{cov}}_p[X(t_i), X(t_j)]$ represents the covariance function using the numerical approximation between $X(t_i)$ and $X(t_j)$. The subscript p is the truncation order for the Karhunen-Loeve expansion when approximating $\widehat{\mathrm{cov}}_p[X(t_i), Y(t_j)]$. The construction of l_1^c and l_2^c is very similar to that of l_1 and l_2, hence the statistical interpretation of l_1^c and l_2^c is also the bias and the standard deviation respectively. In this example, again, $n = 202$ and $t_j = \frac{j-1}{n-1}$, $1 \leq j \leq n$.

Table 2.7 shows l_1^c and l_2^c using the trapezium integral method and the Haar wavelet method for the truncation order 10, 20, 30, 40 and 50. The reconstruction from both of these two methods are quite good, since l_1^c and l_2^c are generally very small. In terms of l_1^c, the bias measure, l_1^c using the trapezium integral method is roughly between 10^{-7} and 10^{-6}, while l_1^c using the Haar wavelet method is around 5×10^{-5}. It means that the bias using the integral trapezium method is smaller than the bias using the Haar wavelet method. In terms of l_2^c, the standard deviation

$l^c_{1,p}$	10	20	30	40	50
$l^c_{1,p}$ Trapezium	0.0000008	0.0000003	0.0000002	0.0000001	0.0000001
$l^c_{2,p}$ Trapezium	0.0040156	0.0013963	0.0007721	0.0005164	0.0003833
$l^c_{1,p}$ Haar	0.0000520	0.0000524	0.0000525	0.0000527	0.0000527
$l^c_{2,p}$ Haar	0.0041782	0.0018280	0.0014174	0.0012998	0.0012544

Table 2.7: The difference of the covariance between the analytical solution and the solutions from two numerical schemes for the Ornstein-Uhlenbeck process in terms of l^c_1 and l^c_2 measure.

measure, l^c_2 decreases when the order p increases, since the higher the order, the better the approximation, and therefore the less the variance. When $p = 10$, l^c_2 for the trapezium integral method and the Haar wavelet method is very close to each other. Both are around 4×10^{-3}. However when $p = 50$, l^c_2 using the Haar wavelet method, with value 1.25×10^{-3}, is about three times as big as l^c_2 using the trapezium integral method, with value 3.8×10^{-4}. It shows again that the trapezium integral method performs better in terms of the standard deviation in the reconstruction.

From figure 2.7, another issue worth mentioning is that although both the trapezium integral method and the Haar wavelet method produce similar results, they do not perform well in the diagonal part of the covariance when $p = 11$. Since the Ornstein-Uhlenbeck process satisfies Sacks-Ylvisaker condition, it does not have continuous first derivative in its covariance function, which is the reason for the peak in the analytical covariance. However, using numerical schemes with not enough order, this peak character can not be well-captured. One way of solving this problem is to introduce higher orders in the truncated Karhunen-Loeve expansion.

The Haar wavelet scheme is chosen as a representative numerical scheme to show how the higher order in the truncated Karhunen-Loeve expansion affects the diagonal

part in the covariance function. Figure 2.8 shows the reconstruction using the Haar wavelet with four different order p. It can be noticed that the higher the order is in the model, the better the performance of the numerical scheme is in the "peak" diagonal part. It becomes sharper when the order is increased from 11 to 50. From 50 to 100, the difference is not that obvious.

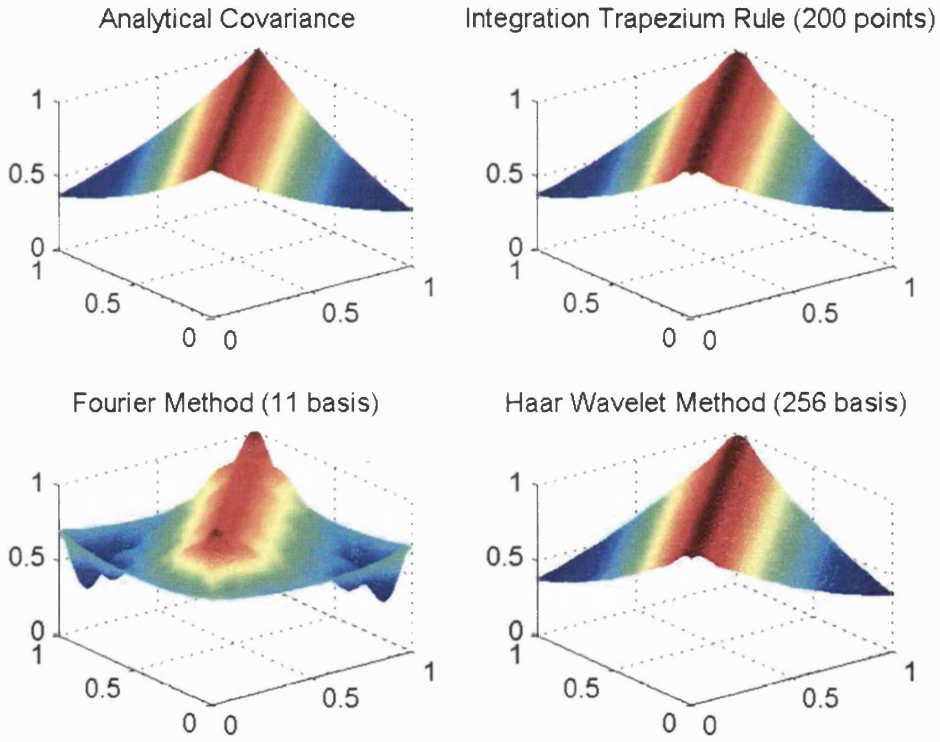

Figure 2.7: Comparison between the covariance of the analytical solution and the covariance reconstruction of three other numerical schemes with order 11 for $\exp(-|t-s|)$.

The squared exponential kernel $\exp(-\beta(t-s)^2)$

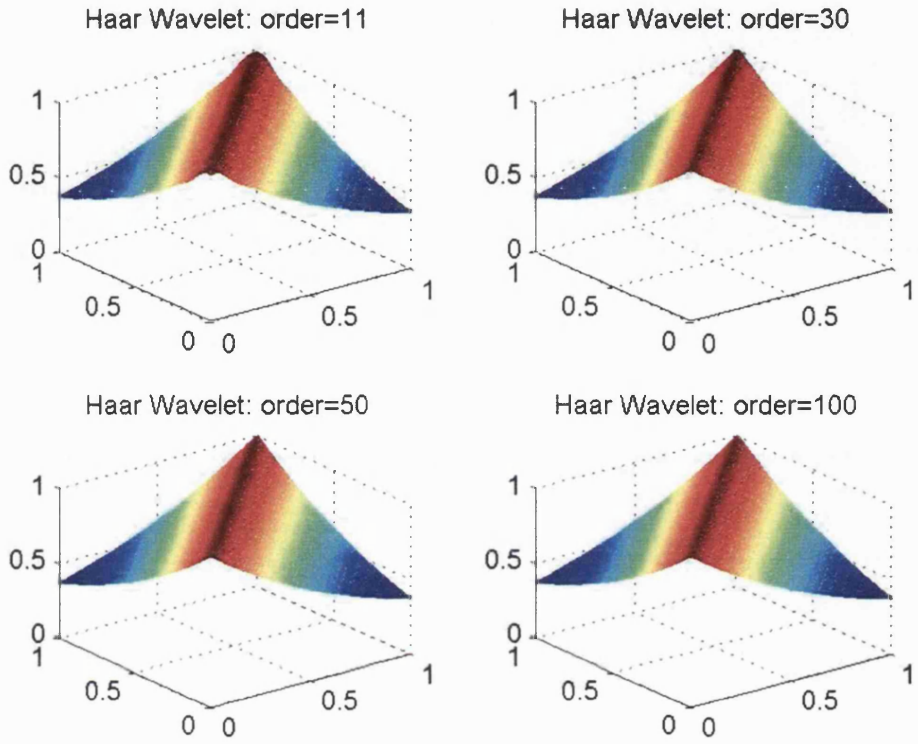

Figure 2.8: Comparison of the covariance reconstruction under different order, using the Haar wavelet with 256 basis, for $\exp(-|t-s|)$.

The most often used covariance kernel in practice is probably the squared exponential kernel, also called the Gaussian kernel, expressed as $\exp(-\beta(t-s)^2)$. For simplicity, β is assumed to be $\beta = 1$. This means that the covariance function is $\exp(-(t-s)^2)$. The analytical solution is very difficult to obtain for this covariance kernel. Some researchers assume Gaussian measure between 0 and 1, and then introduce Hermite polynomials to solve the problem. See, for example, Rasmussen and Williams (2005). Since only Lebesgue measure is used in this thesis, the analytical

	1st	2nd	3rd	4th	5th
Uniform ($n = 200$)	0.8637518	0.1271495	0.0087074	0.0003791	0.0000122
Diff	0.13%	0.74%	1.74%	2.76%	4.27%
Trapezium ($n = 200$)	0.8648371	0.1262215	0.0085601	0.0003692	0.0000118
Diff	0.00%	0.00%	0.02%	0.08%	0.85%
Fourier (11 basis)	0.8614099	0.1299146	0.0455937	0.0198488	0.0123127
Diff	0.08%	2.93%	432.75%	5280.54%	105136.75%
Standard					
Haar (256 basis)	0.8648431	0.1262178	0.0085582	0.0003689	0.0000117

Table 2.8: The first five eigenvalues comparison among different numerical schemes for $\exp[-(t-s)^2]$.

solution they derived is not really helpful to the current problem.

Table 2.8 provides the eigenvalue comparison among different numerical schemes. When comparing the eigenvalue, the result derived from the Haar wavelet method (256 basis function) is used as the standard. Since from the example in the Ornstein-Uhlenbeck process, the Haar wavelet method with $M = 256$ has shown its accuracy in deriving eigenvalues. Similar to the "Error" in the Ornstein-Uhlenbeck process, a measure called "Diff" used in this example is defined as

$$\text{Diff} = \left| \frac{\text{Other Numerical Solution} - \text{Numerical Solution by the Haar Wavelet}}{\text{Numerical Solution by the Haar Wavelet}} \right| \times 100\% \tag{2.3.7}$$

Like the solution to the Ornstein-Uhlenbeck process, the performance of the trapezium integral method is very close to that of the wavelet method. The uniform integral method provides a bigger difference in the eigenvalues compared to the wavelet method. In terms of the Fourier method, the approximation to the first two eigenvalues is not bad, since Diff is 0.08% and 2.93% for the first and the second eigenvalue respectively. But from the third eigenvalue and onwards, Diff for the Fourier

	1st	2nd	3rd	4th	5th
Standard: Haar (256 basis)	0.8648431	0.1262178	0.0085582	0.0003689	0.0000117
Haar (M=8)	0.8662842	0.1253201	0.0080818	0.0003062	0.0000075
Diff	0.17%	0.71%	5.57%	17.00%	35.90%
Haar (M=16)	0.8652013	0.1259949	0.0084399	0.000353	0.0000103
Diff	0.04%	0.18%	1.38%	4.31%	11.97%
Haar (M=32)	0.8649315	0.1261628	0.008529	0.000365	0.0000115
Diff	0.01%	0.04%	0.34%	1.06%	1.71%
Haar (M=64)	0.8648641	0.1262047	0.0085512	0.000368	0.0000117
Diff	0.00%	0.01%	0.08%	0.24%	0.00%
Haar (M=128)	0.8648473	0.1262151	0.0085568	0.0003688	0.0000117
Diff	0.00%	0.00%	0.02%	0.03%	0.00%

Table 2.9: The first five eigenvalues comparison of the Haar wavelet scheme using different orders for $\exp[-(t-s)^2]$.

method dramatically increases. In other words, using the Fourier method, the rate of the decrease of the eigenvalue is much slower than other methods. However, it must be noticed that the eigenvalues are already very small from the third eigenvalue and onwards using the Haar wavelet method and the first two eigenvalues explain more than 99% of the cumulative expected variance, i.e.

$$\frac{\lambda_1+\lambda_2}{\int_0^1 e^{-(t-t)^2}dt} = \lambda_1+\lambda_2 = 0.8648431+0.1262178 > 99\% \qquad (2.3.8)$$

where the values to λ_1 and λ_2 are derived from the Haar wavelet method. Hence the bad performance from the third eigenvalue and onwards using the Fourier method might not affect its overall performance.

Table 2.9 takes a further look at the eigenvalue result provided by the Haar wavelet method with respect to different orders M, with the result by order 256 as the comparison standard.

Again, for order $M = 64$ and onwards, the values almost converge, since the relative difference is very small. The result from lower orders of the Haar wavelet

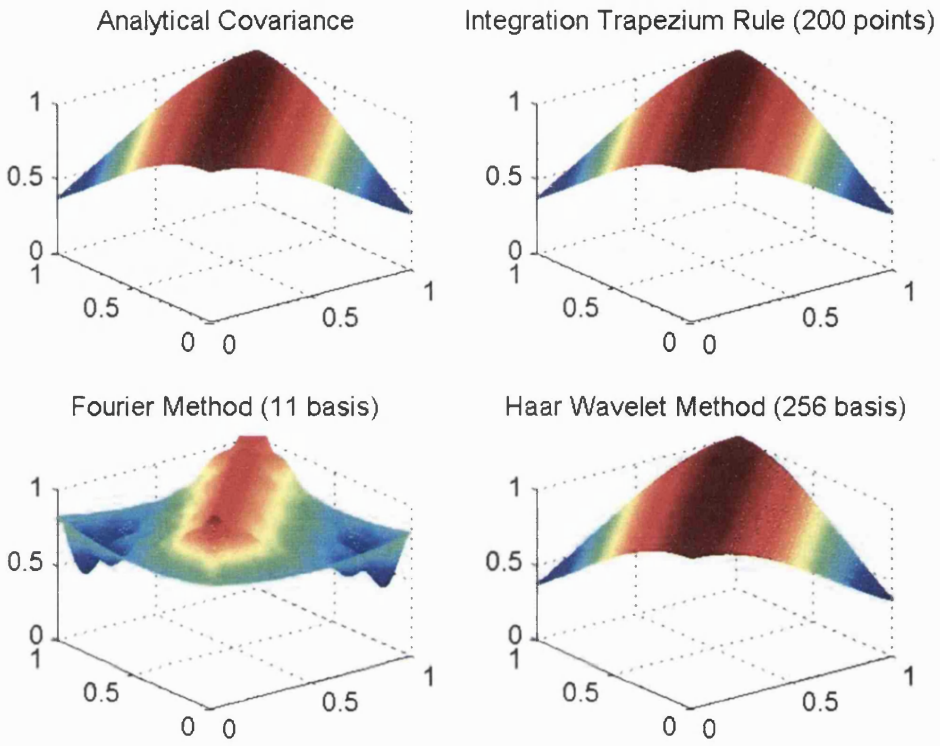

Figure 2.9: Comparison between the covariance of the analytical solution and the covariance reconstruction of three other numerical schemes with order 11 for $\exp[-(t-s)^2]$.

method is still more reliable than that from the Fourier basis method, and again, computationally fast. However, in practice, when data can be sampled at any point on $[0, 1]$, higher order of the wavelet basis will be preferred, since it provides the result both accurate and computationally efficient.

Now the covariance function is reconstructed using the order 11 in the truncated Karhunen-Loeve expansion. Figure 2.9 shows the reconstruction. Since the first two eigenvalues explains about 99% of the cumulative expected variance, 11 should be a

reasonable choice. As expected, the Fourier method does not perform very well, with oscillation everywhere. The covariance reconstructions from both the Haar wavelet and the trapezium integral method are very close to the analytical solution (without truncation). Another point worth mentioning here is that the squared exponential kernel is differentiable everywhere, which does not have any peak in the diagonal element of the covariance function. This is another reason that higher order is not required in the Karhunen-Loeve expansion to recapture the squared exponential kernel.

Table 2.10 provides a further check on the covariance reconstruction between the trapezium integral method and the Haar wavelet method using the measure l_1^c and the measure l_2^c again. Since the first two eigenvalues explains about 99% of the cumulative expected variance, table 2.10 only shows the reconstruction for order 1, 2, 4, 6 and 8. In terms of l_1^c, the bias measure, for the order 1 and 2, l_1^c using the trapezium integral method, with value 4×10^{-5}, is about four times as big as l_1^c using the Haar wavelet method, with value only about 10^{-5}. However, when p is larger than 4, l_1^c using the trapezium integral method reduces dramatically, with l_1^c resulting in about 5.8×10^{-16} when $p = 8$, while l_1^c remains between 10^{-5} and 10^{-4} using the Haar wavelet method. For l_2^c, the standard deviation measure, the result is still similar. When p is 1 and 2, l_2^c using both methods is very close to each other. But when p is bigger than four, l_2^c for the trapezium integral method reduces much faster than that for the Haar wavelet method.

Simulation

This thesis is not concerned with simulation but we now show briefly the application to simulation using the truncated Karhunen-Loeve expansion. For one sample

$l^c_{1,p}$	1	2	4	6	8
$l^c_{1,p}$ Trapezium	0.0000448	0.0000448	3.2×10^{-9}	1.9×10^{-12}	5.8×10^{-16}
$l^c_{2,p}$ Trapezium	0.1274479	0.0087160	0.0000122	6.6×10^{-9}	1.9×10^{-12}
$l^c_{1,p}$ Haar	0.0000108	0.0000108	0.0000551	0.0000551	0.0000551
$l^c_{2,p}$ Haar	0.1274000	0.0088000	0.0009272	0.0009272	0.0009272

Table 2.10: The difference of the covariance between the analytical solution and the numerical solution for the process with the squared exponential kernel $\exp[-(t-s)^2]$ in terms of l^c_1 measure and l^c_2 measure.

path of the Gaussian process, using the truncated Karhunen-Loeve expansion till order p, the truncated process can be expressed as

$$X_p(t) = \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i\xi_i$$

It can be observed that only p standard normal random variables $\{\xi_i\}$, $1 \leq i \leq p$ need to be simulated.

Figure 2.10 provides five simulated sample paths for the Ornstein-Uhlenbeck process with covariance function $\exp[-|t-s|]$ using the truncated Karhunen-Loeve expansion at order 50. The eigenvalues and the eigenfunctions follow the analytical expression in equation 1.3.31 and equation 1.3.32 respectively. The speed of the simulation is pretty fast.

2.4 Summary

This chapter contributes to the numerical solutions of the Fredholm integral equations. It provides a computational foundation for the later chapters, and the methods introduced here are applied throughout the thesis. The numerical methods discussed in the chapter are the integral methods, including the uniform integral method and

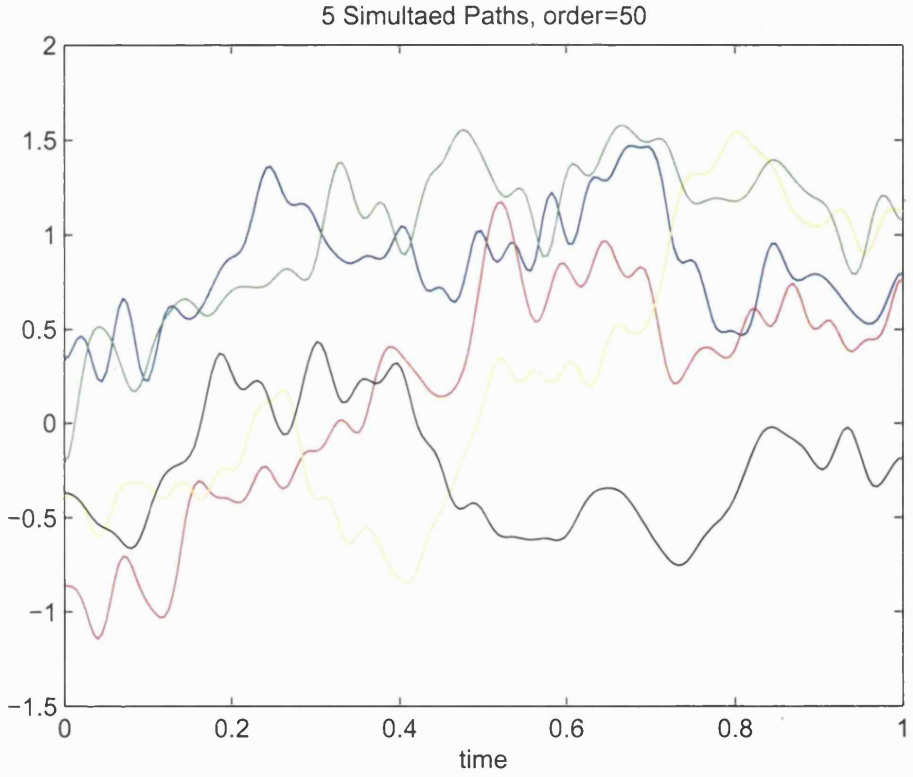

Figure 2.10: The simulation performance of the truncated Karhunen-Loeve expansion for the kernel $\exp(-|t-s|)$: five simulated paths using the truncated Karhunen-Loeve expansion when $p = 50$.

the trapezium integral method, and the expansion methods, including the Fourier basis method and the Haar wavelet basis method. The integral methods discretise the time interval $\mathcal{T}$ and transfer the Fredhom integral equation into a generalised eigen-equation problem, while the expansion methods employ certain basis functions to expand the eigenfunction. The examples being implemented are the Ornstein-Uhlenbeck process, whose analytical solution to the Karhunen-Loeve expansion is known, and the process with squared exponential kernel, whose analytical solution is not known.

Among all the four methods, the Fourier basis method is the worst performing method in terms of its computational speed and computational accuracy. In either of the examples, the Fourier method requires the longest computational time, although only 11 basis functions are involved, compared to 256 basis functions in the Haar wavelet method. This is due to the calculation of the integral involving both the kernel function and the basis function. The approximation of the eigenfunctions using the Fourier basis method result in non-smooth oscillated function. It might be caused by the cyclical variation existing in the trigonometric function. The eigenvalue approximation provides reasonable result, although the value is still not as good as that provided by other numerical methods.

According to the two examples provided in this chapter, the third best numerical method is the uniform integral method. The uniform integral method divides the time interval $\mathcal{T}$ into small intervals with equal length and then assigns equal weights to each interval. The speed of the integral methods can be controlled by the number of intervals involved. Using the uniform integral method, the error of the approximation of the eigenvalues is in general smaller than that provided by the Fourier basis method

and the computational speed is much faster.

The calculation procedure of the trapezium integral method is very similar to that of the uniform integral method, except for the weights. The weights assigned to the first and the last small intervals are only half of the weights assigned to the rest small intervals. Although only the weights are different compared with the uniform integral method, the result improves dramatically. In both examples, for the approximation of the eigenvalues, the error using the trapezium integral method is all below 1%, while the error using the uniform integral method sometimes goes beyond 4%. The approximation to the eigenfunction is very close to the analytical solution, in the example of the Ornstein-Uhlenbeck process, and very similar to that provided by the Haar wavelet method, in the example of the squared exponential kernel.

The Haar wavelet method utilises the easiest basis function in the Daubechies's wavelet family. In order to facilitate the calculation, both the number of the basis functions M and the number of sampled points are often chosen to be the power of two. The computational speed using the Haar wavelet method is very fast even with $M = 256$ in our example. Its approximation to the eigenvalues is the closest to the analytical solution, in the example of the Ornstein-Uhlenbeck process, and therefore is regarded as a comparison standard in the example of the squared exponential kernel. In terms of the approximation to the eigenfunctions, both the Haar wavelet method and the trapezium integral method provides similar results, and very close to the analytical solution in the example of the Ornstein-Uhlenbeck process. However, what is worth mentioning here is that if judging from both l_i measure and l_i^c measure, $i = 1$ and 2 depending on whether the bias or the standard deviation is concerned, the trapezium integral method outperforms the Haar wavelet method in

the approximation of the eigenfunctions and the covariance function reconstruction.

So far, the best two numerical methods are the trapezium integral method and the Haar wavelet method. Both of them are computationally accurate and fast. When the covariance function is known, the Haar wavelet method is preferred in the later chapters. Since it is the expansion methods which indeed regard eigenfunction as a function, which is one of the key ideas to some applications, such as the functional data analysis. The integral trapezium method can only obtain the value at the time points used in the approximation of the integral, although it is more accurate in terms of l_i measure and l_i^c measure, $i = 1$ and 2.

2.5 Numerical method for the spatial Karhunen-Loeve expansion

The spatial Karhunen-Loeve expansion has been discussed separately in section 1.4, chapter 1. The derivation of the spatial Karhunen-Loeve expansion follows directly from the univariate Karhunen-Loeve expansion, except for the time changing from the scalar to the vector.

In terms of the numerical methods, spatial processes with high dimensional time cause computational problems in practice. One one hand, the amount of calculation involved is huge. In d- dimensional time spatial processes, each time point $\mathbf{s} \in \mathcal{T}$ is a d-dimensional column vector, i.e. $\mathbf{s} = \left(s_1, \ s_2, \ \cdots, \ s_d\right)^T$. If n points are chosen in the direction of $s_i, 1 \leq i \leq d$, the total possible number of the points for $\mathbf{s}$ is n^d. Even if in the two dimensional case, if n is 100, the total possible number of the points involving in the calculation is 10^4, hence the covariance matrix

is with dimension $10^4 \times 10^4$. This is beyond the computing power of some popular mathematical package, for example, Matlab 7.0.4, in a normal computer, with CPU Pentium 4 3.6GHz. On the other hand, using the expansion method, for example, the Haar wavelet method, it requires $2d$-dimensional wavelet transform. This causes furthers delay and results in even longer time in practice. Therefore, the integral method is used as the numerical method here.

The result derived at the univariate case can still be applied. The major difference is that the vector time needs to be substituted for the scalar time. The calculation procedure for the two dimensional spatial process is discussed in detail in this section. It can be easily generalised to the $d, d \geq 3$ dimensional case.

Assume that $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \in \mathcal{T}$, $t_i \in [0, T_i], i = 1, 2$. For each direction of t_i, $i = 1, 2$, $n_i + 2$ points are chosen, including the boundary points, i.e. $0 = t_{i0} \leq t_{i1} \leq t_{i2} \leq \cdots \leq t_{i(n_i+1)} = T_i$. Hence the total possible number of points for $\mathbf{t}$ is $(n_1+2)\times(n_2+2)$ and can be expressed as

$$\mathbf{t}_0 = \begin{pmatrix} t_{10} \\ t_{20} \end{pmatrix}, \ \mathbf{t}_1 = \begin{pmatrix} t_{10} \\ t_{21} \end{pmatrix}, \ \cdots, \ \mathbf{t}_k = \begin{pmatrix} t_{1i} \\ t_{2j} \end{pmatrix}, \ \cdots, \ \mathbf{t}_{(n_1+1)(n_2+1)-1} = \begin{pmatrix} t_{1(n_1+1)} \\ t_{2(n_2+1)} \end{pmatrix} \tag{2.5.1}$$

where $0 \leq i \leq n_1+1$, $0 \leq j \leq n_2+1$ and $k = i(n_2+2)+j$. For simplicity, it is further assumed that $T_1 = T_2$, $n = n_1 = n_2$, $t'_j = t_{1j} = t_{2j}$, $0 \leq j \leq n+1$ and $t'_{k+1} - t'_k = \frac{1}{n+1}$, $0 \leq k \leq n$.

For $\mathbf{s}, \mathbf{t} \in \mathcal{T}$, the Fredholm integral equation on the ith eigenvalue and the ith eigenfunction can be approximated as follows.

$$\int_{\mathcal{T}} K(\mathbf{t}, \mathbf{s})\phi_i(\mathbf{s})d\mathbf{s} = \lambda_i\phi_i(\mathbf{t}) \Leftrightarrow \sum_{k=1}^{(n+2)^2} K(\mathbf{t}_j, \mathbf{s}_k)\phi_i(\mathbf{s}_k)w_k = \lambda_i\phi_i(\mathbf{t}_j) \tag{2.5.2}$$

Its corresponding orthogonal condition of the eigenfunction is

$$\int_{\mathcal{T}} \phi_i(\mathbf{s})\phi_j(\mathbf{s})d\mathbf{s} = \delta_{ij} \Leftrightarrow \sum_{k=1}^{(n+2)^2} \phi_i(\mathbf{s}_k)\phi_j(\mathbf{s}_k)w_k = \delta_{ij} \tag{2.5.3}$$

w_k is the weight and equals to

$$w_k = w_i' w_j' \quad 0 \leq i, j \leq n+1 \text{ and } k = i(n+2) + j \tag{2.5.4}$$

where w_i' is the weight in the univariate case. For the uniform scheme, $w_i' = \frac{1}{n+2}, 0 \leq i \leq n+1$; while for the trapezium scheme, $w_i' = \frac{1}{n+1}, 1 \leq i \leq n$ and $w_0' = w_{n+1}' = \frac{1}{2(n+1)}$.

Equation 2.5.2 can also be expressed into the matrix form.

$$KW\phi_i = \lambda_i \phi_i \ , \tag{2.5.5}$$

where

$$K = \begin{pmatrix} K(\mathbf{t}_0, \mathbf{t}_0) & K(\mathbf{t}_0, \mathbf{t}_1) & \cdots & K(\mathbf{t}_0, \mathbf{t}_{(n+2)^2-1}) \\ K(\mathbf{t}_1, \mathbf{t}_0) & K(\mathbf{t}_1, \mathbf{t}_1) & \cdots & K(\mathbf{t}_1, \mathbf{t}_{(n+2)^2-1}) \\ \cdots & & & \\ K(\mathbf{t}_0, \mathbf{t}_{(n+2)^2-1}) & K(\mathbf{t}_0, \mathbf{t}_{(n+2)^2-1}) & \cdots & K(\mathbf{t}_{(n+2)^2-1}, \mathbf{t}_{(n+2)^2-1}) \end{pmatrix}$$

$$W = \begin{pmatrix} w_0 & 0 & \cdots & 0 \\ 0 & w_1 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & w_{(n+2)^2-1} \end{pmatrix}$$

$$\phi_i^T = \begin{pmatrix} \phi_i(\mathbf{t}_0), & \phi_i(\mathbf{t}_1), & \cdots, & \phi_i(\mathbf{t}_{(n+2)^2-1}) \end{pmatrix}$$

Equation 2.5.5 is equivalent to equation 2.1.5 in the univariate case. The remaining numerical procedures are exactly the same as that in the univariate setting. See section 1 of this chapter for details.

	1st(1,1)	2nd(1,2)	3rd(2,1)	4th(3,1)	5th(1,3)	6th(1,4)
Analytical	0.1642557	0.0182506	0.0182506	0.0065702	0.0065702	0.0033522
Numerical	0.1642817	0.0182651	0.0182651	0.0065837	0.0065837	0.0033654
Error	0.02%	0.08%	0.08%	0.21%	0.21%	0.39%

Table 2.11: Eigenvalue comparison for the Brownian sheet. Numerical method is the integral trapezium method when $n = 50$. (i, j) in the first row represents the ith value and jth value used in equation 1.4.9.

Example: The Brownian sheet in $[0, 1]$ (numerical solution)

This example follows the analytical solution to the Brownian sheet derived in chapter 1. The analytical solution will be used to compare with the solution derived from the trapezium integral method. Due to the lack of computing power, we use $n = 50$. Table 2.7 compares the first six eigenvalues. Error is computed using equation 2.3.5.

The difference between the analytical solution and the numerical solution does not appear until 10^{-5}, which is relatively small. Although the error calculated in the table is bigger than that in the univariate case, the result is still very close considering only $n = 50$ points are used. More n could be considered to improve accuracy. In terms of the eigenfunctions, since the eigenfunction is now a function of two dimensional time, both the 3D plot and its corresponding contour for the first four eigenfunctions are plotted in figure 2.11 and figure 2.12 respectively. The straight line in the contour is where the eigenfunctions equal to zero. For example, for the second eigenfunction $\phi_2(t, s) = 2\sin(\frac{3}{2}\pi t)\sin(\frac{1}{2}\pi s)$, it equals to zero when $\sin(\frac{3}{2}\pi t) = 0$. It is equivalent to say that $t = \frac{2}{3}$.

In terms of the performance of the numerical method for the eigenfunctions, which

is shown in figure 2.13 and figure 2.14, the approximation gets worse when the eigenfunction's corresponding eigenvalue becomes smaller. Using the numerical method, the first eigenfunction is very like the analytical solution, while some deviation from the true value start to appear in the second and the third eigenfunction. For the fourth eigenfunction, the approximation is very unsatisfactory. It is quite different from the analytical solution. Again, one of the reasons is due to the limited number of points used in this example, which is $n = 50$.

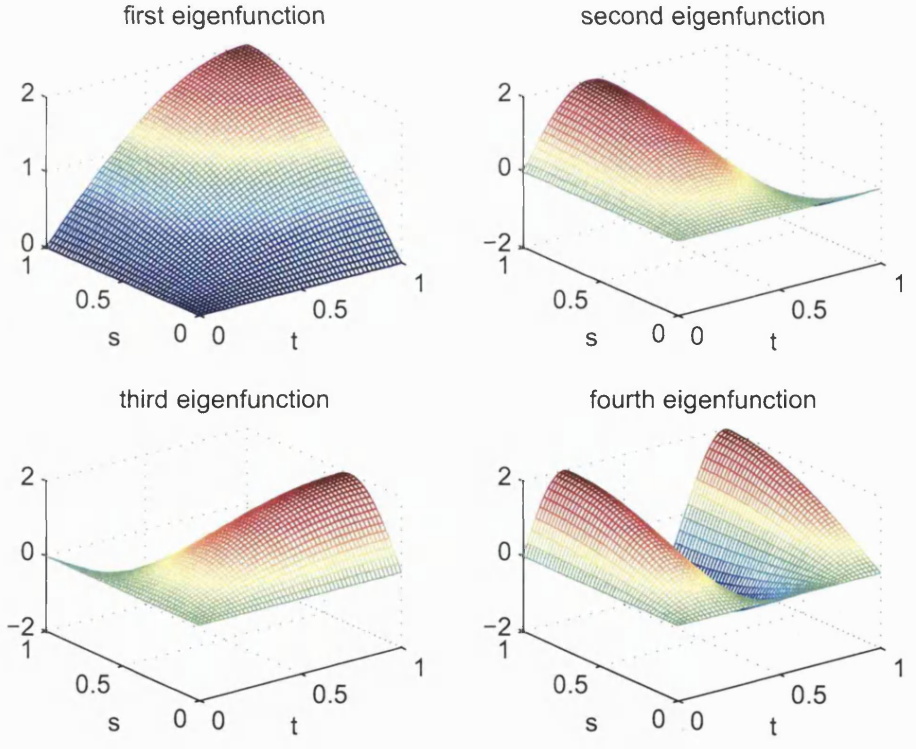

Figure 2.11: 3D plot for the first four analytical eigenfunctions of the Brownian sheet. 1st: $2\sin(\frac{1}{2}\pi t)\sin(\frac{1}{2}\pi s)$; 2nd: $2\sin(\frac{3}{2}\pi t)\sin(\frac{1}{2}\pi s)$; 3rd: $2\sin(\frac{1}{2}\pi t)\sin(\frac{3}{2}\pi s)$; 4th: $2\sin(\frac{3}{2}\pi t)\sin(\frac{3}{2}\pi s)$.

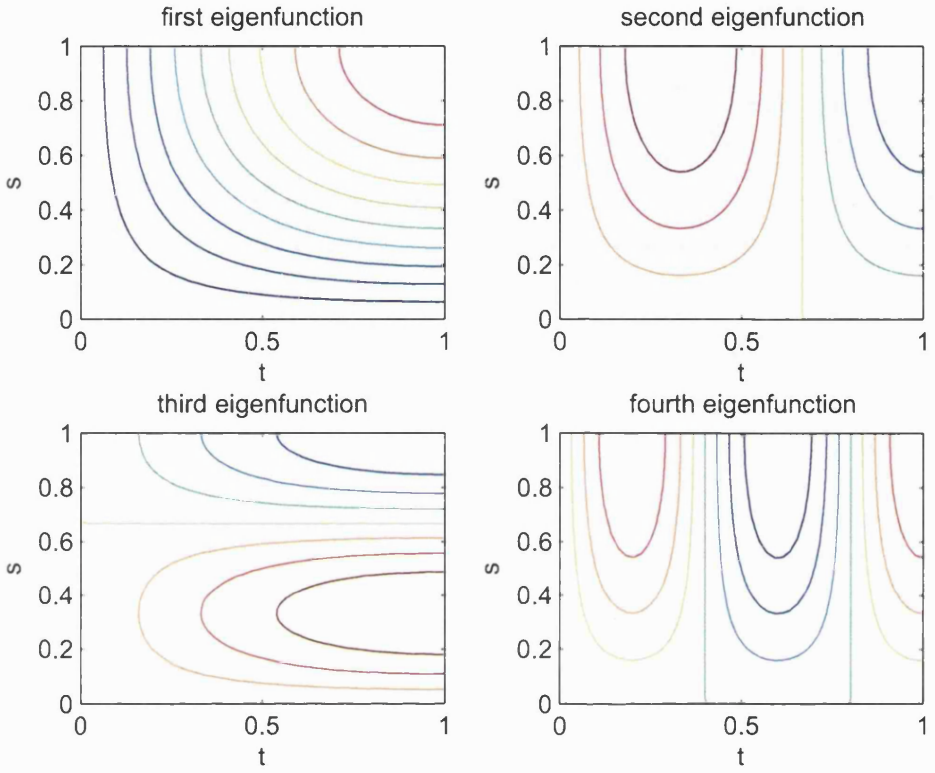

Figure 2.12: Contour for the first four analytical eigenfunctions of the Brownian sheet. 1st: $2\sin(\frac{1}{2}\pi t)\sin(\frac{1}{2}\pi s)$; 2nd: $2\sin(\frac{3}{2}\pi t)\sin(\frac{1}{2}\pi s)$; 3rd: $2\sin(\frac{1}{2}\pi t)\sin(\frac{3}{2}\pi s)$; 4th: $2\sin(\frac{3}{2}\pi t)\sin(\frac{3}{2}\pi s)$.

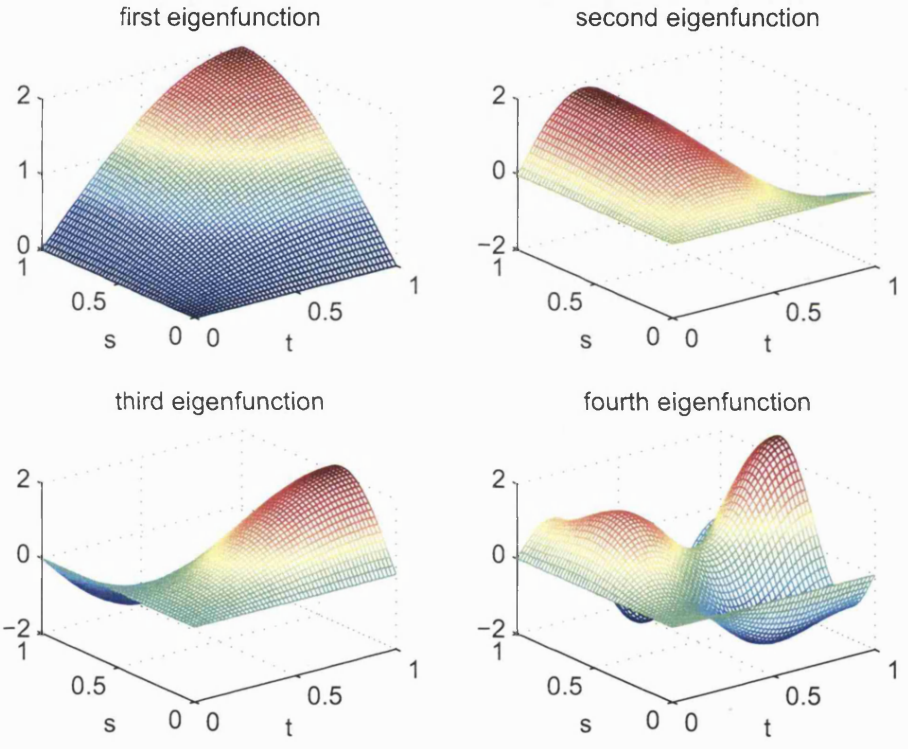

Figure 2.13: 3D plot for the first four eigenfunctions of the Brownian sheet using the integral trapezium scheme ($n = 50$).

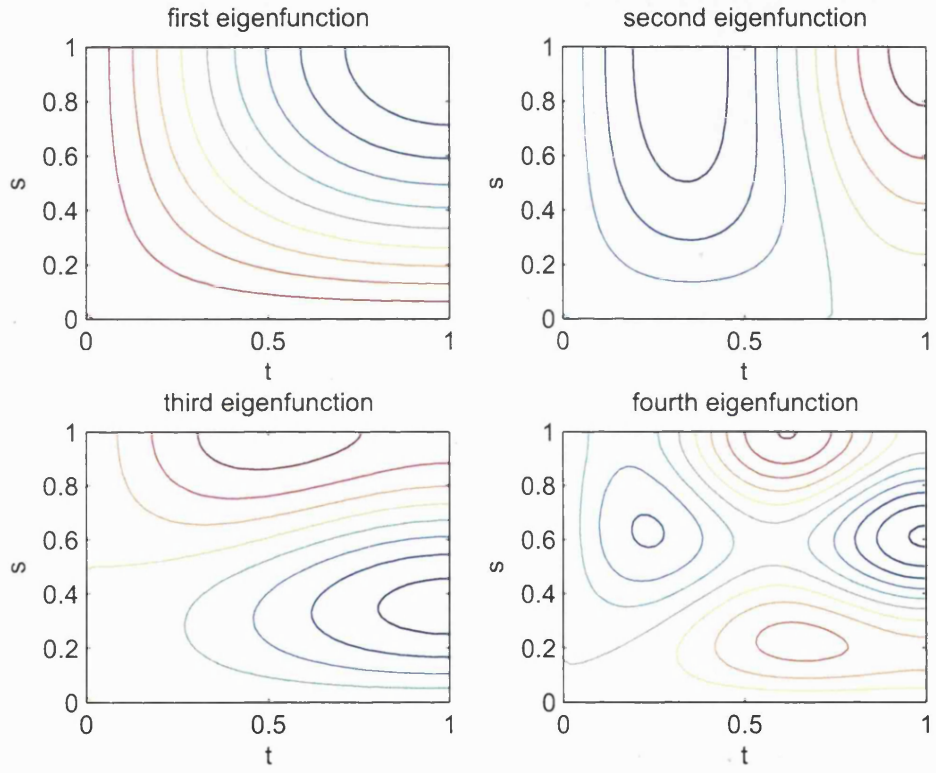

Figure 2.14: Contour for the first four eigenfunctions of the Brownian sheet using the integral trapezium scheme ($n = 50$).

Chapter 3

Maximum Entropy Sampling for Gaussian processes

In this chapter, the Karhunen-Loeve expansion is related to maximum entropy sampling in the Gaussian case. Entropy, or the Shannon entropy, is defined as

$$\text{Ent}(\Gamma) = E_{\Gamma}[-\log\{p(\Gamma)\}] \tag{3.0.1}$$

The following formula related to the entropy is used throughout this chapter. See, for example, Cover and Thomas (1991).

$$\text{Ent}(\Gamma) = \text{Ent}(\Gamma_s) + E_{\Gamma_s}\{\text{Ent}(\Gamma_{s^c}|\Gamma_s)\} \tag{3.0.2}$$

In our case, define $X_S = \{X_i\}_{i\in S}$, $X_s = \{X_i\}_{i\in s}$ and $X_{s^c} = \{X_i\}_{i\in s^c}$, where the index set $S = \{1, 2, \cdots, N\}$ represents the whole "population" and $s \subset S$, $|s| = n$ is a chosen sample. The the above formula can be rewritten as:

$$\text{Ent}(X_S) = \text{Ent}(X_s) + E_{X_s}\{\text{Ent}(X_{s^c}|X_s)\} \tag{3.0.3}$$

It means that the entropy for the "population" can be decomposed as the entropy for the sample set and the expected posterior entropy of the unsampled set.

Maximum entropy sampling claims that the expected posterior entropy of the unsampled population s^c should be minimized in terms of the entropy sampling. Since the left hand side of equation 3.0.3 does not depend on the experiment, minimizing the expected posterior entropy is equivalent to maximising the entropy of the chosen sample, which is the first term of the right hand side. The idea of maximum entropy sampling was first introduced by Shewry and Wynn (1987). Since then algorithms have been developed by Lee (1998) and Hoffman et al. (2004). Later discussions on this topic can be found at Sebastiani and Wynn (2000) and Wynn (2004).

In the Gaussian case, the entropy is related to the logorithm of the determinant of the covariance matrix. Assume X is a n-dimensional multivariate Gaussian distribution with mean μ and variance Σ, i.e. $X \sim N(\mu, \Sigma)$, then

$$\text{Ent}(X) = \frac{n}{2} + \frac{n}{2}\log(2\pi) + \frac{1}{2}\log(|\Sigma|) \tag{3.0.4}$$

Hence for a Gaussian process, given fixed N and n maximising the entropy is equivalent to maximise the determinant of the covariance matrix comprising of the sampling points.

3.1 Karhunen-Loeve expansion, maximum entropy sampling and D-optimality when $n = p$

In this section, using the truncated Karhunen-Loeve expansion, the true covariance matrix of the sample set is approximated by the truncated covariance matrix defined in equation 3.1.2. Then applying the General Equivalence Theorem (theorem 3.1.2), the maximum entropy sampling is connected with the D-optimal design.

Assume $\{Y(t)\}$ is a centered Gaussian process, i.e. $E[Y(t)] = 0$. Using the Karhunen-Loeve expansion, it can be expressed as

$$Y(t) = \sum_{i \geq 1} \sqrt{\lambda_i} \phi_i(t) \epsilon_i \tag{3.1.1}$$

where $\epsilon_i \sim i.i.d.N(0,1)$.

In practice, only the finite order of the expansion is used. In this section we further assume that the true stochastic model for $Y(t)$ is under truncation at order p and is written as $Y(t) = \sum_{i=1}^{p} \sqrt{\lambda_i} \phi_i(t) \epsilon_i$. Then with n design points $t_1, t_2, \cdots, t_n$, using the truncated Karhunen-Loeve expansion at order p, we can approximate the covariance matrix as K_p, which is

$$K_p = \Phi^T \Lambda \Phi \tag{3.1.2}$$

where

$$\Phi = \begin{pmatrix} \phi_1(t_1) & \phi_1(t_2) & \cdots & \phi_1(t_n) \\ \phi_2(t_1) & \phi_2(t_2) & \cdots & \phi_2(t_n) \\ \cdots & & & \\ \phi_p(t_1) & \phi_p(t_2) & \cdots & \phi_p(t_n) \end{pmatrix} \qquad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \tag{3.1.3}$$

According to chapter 1, each element of K_p is a truncated version of the true covariance function. The accuracy of the approximation can be controlled by choosing an adequate order p. Assuming a suitable choice of the order p, the rest of this chapter concentrates on maximising the determinant of the truncated covariance matrix K_p, rather than that of the true covariance matrix. From now on, the design maximising the determinant of K_p will be referred to maximum entropy sampling.

In order to set up a duality between maximum entropy sampling and the D-optimality, the condition that $n \geq p$ is required.

In this section we concentrate on $n = p$.

Proposition 3.1.1. *When $n = p$, $K_p = \Phi^T \Lambda \Phi$ with Φ and Λ defined in equation 3.1.3, then* $\det(K_p) = \det(\Phi\Phi^T)\det(\Lambda)$

Proof.

$$\det(K_p) = \det(\Phi^T \Lambda \Phi) = \det(\Phi\Phi^T \Lambda) = \det(\Phi\Phi^T)\det(\Lambda) \tag{3.1.4}$$

□

According to proposition 3.1.1, when $n = p$, the determinant of the covariance matrix can be expressed as $\det(\Phi\Phi^T)\det(\Lambda)$. Hence maximising the entropy is equivalent to maximising $\det(\Phi\Phi^T)\det(\Lambda)$. Since $\det(\Lambda)$ is fixed whatever design we choose, maximising $\det(K_p)$ for the maximum entropy is further equivalent to maximising $\det(\Phi\Phi^T)$, in this case.

Remark 3.1.1. When $n > p$, K_p is a singular matrix, i.e. $\det(K_p) = 0$, so that no equivalence is obvious. However, under a certain setting, we can still argue the equivalence between maximum entropy sampling and maximising $\det(\Phi\Phi^T)$, as long as p is big enough to capture most of the cumulative expected variance. This will be discussed in section 3.2.

We further denote $X = \Phi^T$ to match the standard notation of the D-optimality, so that

$$\Phi\Phi^T = X^T X = \sum_{i=1}^{n} f(t_i) f^T(t_i) \ , \tag{3.1.5}$$

where

$$\begin{aligned} f(t_i)^T &= \Big(\phi_1(t_i), \ \phi_1(t_i), \ \cdots, \ \phi_p(t_i)\Big) \\ X^T &= \Big(f(t_1), \ f(t_2), \ \cdots, \ f(t_n)\Big) \end{aligned}$$

Now define the information matrix F needed for the D-optimal design. In general, the information matrix F for the design ξ is expressed as

$$F(\xi) = \int f(t)f(t)^T p_c(t)dt = \sum_{i=1}^{n} f(t_i)f(t_i)p(t_i) \tag{3.1.6}$$

where $p_c(t)$ is a continuous measure and $p(t)$ is a discrete measure. The first equality is for the continuous design while the second equality is for the discrete design. We also need to define a sensitive function written as $d(x,\xi) = f(x)^T F^{-1}(\xi)f(x)$.

The following equivalence theorem plays a major role in the optimal design of this chapter. It builds a bridge between maximum entropy sampling and the D-optimal design. See, for example, Kiefer and Wolfowitz (1960), Kiefer (1974) and Fedorov (1972), for details.

Theorem 3.1.2. *(General Equivalence Theorem, GET) The following assertions are equivalent*

(i) the design ξ^* *maximizes* $\det(F(\xi))$

(ii) the design ξ^* *minimizes* $\max_t d(t,\xi)$

(iii) $\max_t d(t,\xi^*) = p$, *where* p *represents the number of unknown parameters.*

A design measure ξ maximising $\det(F(\xi))$ is called a D-optimal design measure. The GET provides a way of checking the D-optimality by plotting $d(t,\xi)$. When $n = p$, maximising K_p is equivalent to maximising $\det(X^T X)$. Under the discrete uniform design, the information matrix F is expressed as $F = \frac{X^T X}{n}$. Hence, given an adequate order p, maximum entropy sampling aims at maximising $\det(F)$. Or using the GET, given an adequate order p, a sampling ξ^* is a maximum entropy sampling if $\max_t d(t,\xi^*) = f(t)[\frac{X(\xi^*)^T X(\xi^*)}{n}]^{-1} f(t) = p$. For processes satisfying the Sacks-Ylvisaker conditions (see remark 1.3.3, chapter 1), there is a lot of "energy"

in the tail, i.e. p needs to be very big to make any connection between maximum entropy sampling and the D-optimal design. However, for the process with smooth covariance kernels, like $\exp[-(t-s)^2]$, p can be chosen as small as 2 to establish a strong connection between these two sampling schemes. Both types of processes will be discussed in more detail in the examples in the later sections of this chapter.

Now using the GET, we provide a simple example for the Brownian motion under a discrete uniform design. For simplicity, we only assume $n = p = 2$ in this case. However, using order 2 is not an accurate way of describing the covariance of a member of the Sacks-Ylvisaker family. The optimal points, although D-optimal, do not maximise the entropy. However this example goes some way towards explaining the connection and is analytically tractable. The optimal points will be used to compare with the result from numerical algorithm in section 3.4.

Example: The Brownian motion on $[0,1]$, when $n = p = 2$

As a prior knowledge, time point 1 is always chosen as one of the D-optimality points in the Brownian motion. The choice of 1 will be shown reasonable on the sensitivity function in figure 3.1 when $n = p = 2$. For bigger n and p, refer to theorem 3.3.1 on the equivalence of the D-optimality and maximum entropy for the Brownian motion.

In the $n = p = 2$ case, since 1 has been chosen as one of the design points, only one more point t $(0 \leq t < 1)$ needs to be chosen. Under the discrete uniform measure, $p(t) = p(1) = \frac{1}{2}$.

When $p = 2$, $f(t)$ gives two eigenfunctions.

$$f(t) = \begin{pmatrix} \sqrt{2}\sin(\frac{1}{2}\pi t) \\ \sqrt{2}\sin(\frac{3}{2}\pi t) \end{pmatrix} \tag{3.1.7}$$

Therefore, $F = \sum_{i=1}^{2} f(t_i)f^T(t_i)p(t_i)$, $(t_1 = t, t_2 = 1)$, i.e.

$$F = \begin{pmatrix} \sin^2(\frac{\pi t}{2}) + 1 & \sin(\frac{\pi t}{2})\sin(\frac{3\pi t}{2}) - 1 \\ \sin(\frac{\pi t}{2})\sin(\frac{3\pi t}{2}) - 1 & \sin^2(\frac{3\pi t}{2}) + 1 \end{pmatrix} \tag{3.1.8}$$

Then, $\det(F) = (\sin(\frac{\pi t}{2}) + \sin(\frac{3\pi t}{2}))^2$. Since $\sin(\frac{\pi t}{2}) + \sin(\frac{3\pi t}{2}) \geq 0$ when $0 \leq t \leq 1$, maximising $\det(F)$ is equivalent to maximising $\sin(\frac{\pi t}{2}) + \sin(\frac{3\pi t}{2})$.

Set $Q = \sin(\frac{\pi t}{2}) + \sin(\frac{3\pi t}{2})$, then

$$\frac{dQ}{dt} = \frac{1}{2}\cos(\frac{\pi t}{2}) + \frac{3}{2}\cos(\frac{3\pi t}{2}) \tag{3.1.9}$$

The optimal point can be found by setting $\frac{dQ}{dt} = 0$. Hence, $t = \frac{2\arccos(\frac{\sqrt{6}}{3})}{\pi} \simeq 0.3918$

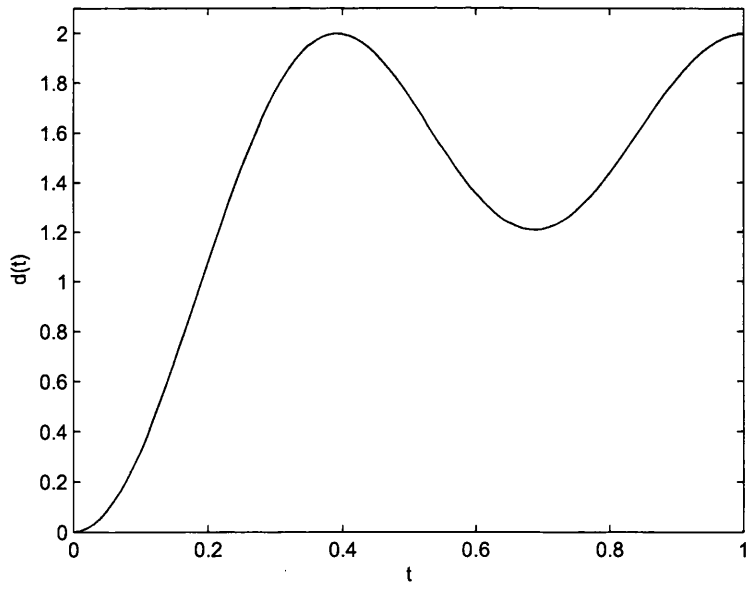

Figure 3.1: $d(t) = f(t)^T F^{-1} f(t)$ for the Brownian motion, when $n = p = 2$ and $0 \leq t \leq 1$.

Figure 3.1 plots $d(t) = f(t)^T F^{-1} f(t)$ under the optimal design points we just found. It shows that $\max_{0 \geq t \geq 1}(d(t)) = 2$, which is the number of the parameters.

It confirms that $\frac{2\arccos(\frac{\sqrt{6}}{3})}{\pi}$ and 1 are the D-optimal design points. However, as we have mentioned earlier, $p = 2$ is too small to approximate the covariance matrix for the Brownian motion. Although the two points we found are D-optimal, they are not the optimal points for the Brownian motion under the maximum entropy criteria.

3.2 Equivalence of D-optimality and maximum entropy sampling when $n > p$

In this section, we cope with the condition $n > p$. As is mentioned in remark 3.1.1, when $n > p$, the truncated covariance matrix K_p is singular if the true stochastic model is the truncated version of equation 3.1.1. Hence an alternative stochastic model should be considered under the current setting.

We assume that the true stochastic model is a regression model defined as follows.

$$Y_s = X_s V + \epsilon \ ,$$

where $|s| = n < N$, $V \sim N(0, W)$ and error term $\epsilon \sim N(0, \Sigma)$. Since the number of the sampling points is n and the number of the parameters is p, X_s is with dimension $n \times p$, while V is with dimension $p \times 1$. We assume that V and ϵ are independent and $\Sigma = \sigma^2 I$, so that the model for Y_s is a Gaussian model with the independent noise. In some literature, the Gaussian model with the independent noise is called the Gaussian regression model. See, for example, Rasmussen and Williams (2005) and MacKay (2003). A Gaussian regression model is a good stochastic model in dealing with the case when the number of the observations is bigger than the number of the parameters, i.e. $n > p$. Application of this model will be used in both chapter 3 and chapter 4.

For this chapter, it is further assumed that W is diagonal. When $n > p$, we will argue that if the first p parameters can explain almost all the cumulative expected variance, maximum entropy sampling for the process Y_s is equivalent to the D-optimal design.

We start from decomposing the entropy of Y_s.

$$\begin{aligned}
\det[\mathrm{Var}(Y_s)] &= \det(X_s W X_s^T + \sigma^2 I_{n\times n}) \\
&= \det(UDU^T + \sigma^2 UU^T) \\
&= \det[U(D+\sigma^2 I)U^T] \\
&= \sigma^{2(n-p)} \prod_{i=1}^{p} (\mu_i(D_1) + \sigma^2) \\
&= \sigma^{2(n-p)} \det(W^{\frac{1}{2}} X_s^T X_s W^{\frac{1}{2}} + \sigma^2 I_{p\times p}) \ ,
\end{aligned}$$

where UDU^T is the eigenvalue decomposition for $X_s W X_s^T$; $\mu_i(D_1)$ is the ith eigenvalue for D_1;

$$D = \begin{pmatrix} D_{1,p\times p} & 0_{p\times(n-p)} \\ 0_{(n-p)\times p} & 0_{(n-p)\times(n-p)} \end{pmatrix} \qquad \text{and} \qquad D_1 = \begin{pmatrix} \mu_1(D_1) & 0 & \cdots & 0 \\ 0 & \mu_2(D_1) & \cdots & 0, \\ \cdots & & & \\ 0 & 0 & \cdots & \mu_p(D_1) \end{pmatrix}$$

If n and p are fixed, σ^2 is fixed. Hence maximising the entropy of Y_s is equivalent to maximising $\det(W^{\frac{1}{2}} X_s^T X_s W^{\frac{1}{2}} + \sigma^2 I_{p\times p})$. In the case that $\sigma^2 \to 0$,

$$\begin{aligned}
& \det(W^{\frac{1}{2}} X_s^T X_s W^{\frac{1}{2}} + \sigma^2 I) \\
= & \prod_{i=1}^{p} (\mu_i(D_1) + \sigma^2) \\
= & \prod_{i=1}^{p} [\mu_i(D_1)] + o(\sigma^2) \\
= & \det(W^{\frac{1}{2}} X_s^T X_s W^{\frac{1}{2}}) + o(\sigma^2)
\end{aligned}$$

If we further assume that $X_s = \Phi^T = X$ and $W = \Lambda$,

$$\text{Var}(Y_s) = \Phi^T \Lambda \Phi + \sigma^2 I = X \Lambda X^T + \sigma^2 I \tag{3.2.1}$$

The first term on the right hand side is the same as equation 3.1.2, which is the covariance matrix expressed by the truncated Karhunen-Loeve expansion at order p. Since the true covariance function can be expressed by the un-truncated Karhunen-Loeve expansion, i.e. an infinite sum, the second term on the right hand side $\sigma^2 I$ only captures the tail $\sum_{i=p+1}^{\infty} \lambda_i \phi_i^2(t_i)$, $i = 1, 2, \cdots, n$ for the diagonal elements in the covariance matrix. When the first p eigenvalues can explain almost all the cumulative expected variance of the original process, i.e. $\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i}$ is close to 1, the majority of the variance of Y_s can be mainly captured by the truncated Karhunen-Loeve expansion with the remaining energy in the tail very small, implying very small σ^2. In the extreme case that $\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \to 1$, $\sigma^2 \to 0$. Therefore, maximising $\det[\text{cov}(Y_s)]$ is equivalent to maximising $\det[\Lambda^{\frac{1}{2}} X^T X \Lambda^{\frac{1}{2}}] = \det(\Lambda)\det(X^T X)$. Since Λ is not related to the design, maximising the entropy of the process is the same as maximising $\det(X^T X)$. This is the same result as what is derived in section 3.1.

Remark 3.2.1. In Wynn (2004), the model for Y_s is regarded as a Bayesian model, where the distribution for the parameter V needs to be updated. In order to have the maximum posterior information on V, we aim to minimize the posterior entropy of V, which is $\text{Ent}(V|Y_s)$. Using Shannon's theorem, there are two ways to express the joint entropy of V and Y_s.

$$\begin{aligned} \text{Ent}(Y_s, V) &= \text{Ent}(V) + E_V \text{Ent}(Y_s|V) \\ \text{Ent}(Y_s, V) &= \text{Ent}(Y_s) + E_{Y_s} \text{Ent}(V|Y_s) \end{aligned}$$

The first equation shows that $\text{Ent}(Y_s, V)$ is independent of the design, i.e. independent of s, since $\text{Ent}(Y_s|V) = \text{Ent}(\epsilon)$. Therefore minimizing the posterior entropy of the parameter V, (second term of the second equation) is equivalent to maximising $\text{Ent}(Y_s)$, which provides the same result as maximum entropy sampling.

In short, under the Bayesian setting, we quote the result by Wynn (2004)

Theorem 3.2.1. *Under a suitable Bayesian regression model, maximum entropy sampling for choice of a subsample s from a population S is equivalent to*
(i) minimisation of expected posterior entropy for the unsampled population units.
(ii) Bayes minimum entropy design for the regression parameters.

Remark 3.2.2. Another reason that eigenvalues will not change our result is as follows:

For the Gaussian regression model $Y_s = X_s V + \sigma^2 I$, we can also define $X_s = \Phi^T \Lambda^{\frac{1}{2}}$ and $W = I$. Then using the same arguments as those in the rest of this section, maximising the entropy of Y_s is the same as maximising $\det(X_s^T X_s) = \det[\sum_{i=1}^n f'(t_i) f'^T(t_i)]$. where

$$f'(t) = \begin{pmatrix} \sqrt{\lambda_1}\phi_1(t) \\ \sqrt{\lambda_2}\phi_2(t) \\ \cdots \\ \sqrt{\lambda_p}\phi_p(t) \end{pmatrix} = \Lambda^{\frac{1}{2}} f(t) \tag{3.2.2}$$

Under the uniform design, the information matrix F' is expressed as $F' = \frac{\sum_{i=1}^n [f'(t_i) f'(t_i)^T]}{n}$,

then the sensitivity function $d'(t)$ can be written down as

$$\begin{aligned} d'(t) &= f'(t)^T F'^{-1} f'(t) \\ &= nf^T(t)\Lambda^{\frac{1}{2}}[\Lambda^{-\frac{1}{2}}(X^TX)^{-1}\Lambda^{-\frac{1}{2}}]\Lambda^{\frac{1}{2}}f(t) \\ &= f^T(t)\left(\frac{X^TX}{n}\right)^{-1} f(t) = f^T(t)F^{-1}f(t) \\ &= d(t) \end{aligned}$$

Now following the GET, minmax $d'(t)$ is the same as minmax $d(t)$, which is equivalent to maximising $\det(X^TX)$. This is not related to the eigenvalues.

3.3 Brownian motion on $[0, 1]$

In this section, it is proved that, for the Brownian motion, the maximum entropy sampling is an approximate D-optimal design when conditions provided by theorem 3.3.1 are satisfied.

We have shown an example on the Brownian motion in section 3.1 when $n = p = 2$. However as we have mentioned, $p = 2$ is not big enough to capture the true covariance function, since the first two eigenvalues only explain about 90% of the cumulative expected variance.

We now show that analytically the optimal design points for maximising the entropy of the Brownian motion should be equally spaced. For the Brownian motion, the analytical covariance function $K(s,t) = \min(t,s)$. Hence the covariance matrix

for the discrete time points $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$ is

$$K = \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \cdots & & & & \\ t_1 & t_2 & t_3 & \cdots & t_n \end{pmatrix} \tag{3.3.1}$$

and $\det(K) = (t_1 - t_0)(t_2 - t_1)(t_3 - t_2)\cdots(t_n - t_{n-1}) = \prod_{i=1}^{n} u_i$, where $u_i = t_i - t_{i-1}$, $1 \leq i \leq n$ and $t_0 = 0$. Notice that $\sum_{i=1}^{n} u_i = t_n$. Maximum entropy sampling problem has now been transformed to finding the points t_i, $1 \leq i \leq n$, so that

$$\max(\det(K)) = \max \prod_{i=1}^{n} u_i = \max \prod_{i=1}^{n} (t_i - t_{i-1})$$

$$\text{subject to} \quad \sum_{i=1}^{n} u_i = t_n \text{ and } t_0 = 0$$

The solution to u_i is $u_i = \frac{t_n}{n}$, $1 \leq i \leq n$. Therefore, the solution to t_i is $t_i = i\frac{t_n}{n}$, $1 \leq i \leq n$. In our case we always choose $t_n = 1$, so that our sampling scheme is an equally spaced sampling on $[0, 1]$, with 1 as one of the design points, i.e. $t_i = \frac{i}{n}$, $1 \leq i \leq n$.

It is further assumed that the measure here is the uniform discrete measure ξ^*, i.e. for n design points on $[0, 1]$, $p(t_i) = \frac{1}{n}$. In the Brownian motion, the eigenfunction can be written as $\phi_i(t) = \sqrt{2}\sin[(i - \frac{1}{2})\pi t]$, $i \geq 1$, hence for $a, b \in \mathbb{N}$

$$\sum_{i=1}^{n} \phi_a(t_i)\phi_b(t_i) = 2\sum_{i=1}^{n} \sin[(a - \frac{1}{2})\pi\frac{i}{n}]\sin[(b - \frac{1}{2})\pi\frac{i}{n}]$$

$$= \sum_{i=1}^{n} \{\cos[(a-b)\pi\frac{i}{n}] - \cos[(a+b-1)\pi\frac{i}{n}]\}$$

It can be further calculated that

$$\sum_{i=1}^{n}\cos[(a-b)\pi\frac{i}{n}] = \begin{cases} n & a=b \\ 0 & |a-b| \text{ even} \\ -1 & |a-b| \text{ odd} \end{cases} \quad \sum_{i=1}^{n}\cos[(a+b-1)\pi\frac{i}{n}] = \begin{cases} -1 & a=b \\ -1 & |a-b| \text{ even} \\ 0 & |a-b| \text{ odd} \end{cases} \tag{3.3.2}$$

Hence

$$\sum_{i=1}^{n}\phi_a(t_i)\phi_b(t_i) = \begin{cases} n+1 & a=b \\ 1 & |a-b| \text{ even} \\ -1 & |a-b| \text{ odd} \end{cases} \tag{3.3.3}$$

Whatever p is even or odd, the calculation procedure for finding the D-optimality points is the same. Here, only the result for even p is demonstrated. When p is even

$$X^TX = \begin{pmatrix} n+1 & -1 & 1 & \cdots & 1 & -1 \\ -1 & n+1 & -1 & \cdots & -1 & 1 \\ \cdots & & & & & \\ 1 & -1 & 1 & \cdots & n+1 & -1 \\ -1 & 1 & -1 & \cdots & -1 & n+1 \end{pmatrix}$$

$$\det(X^TX) = n^{p-1}(n+p)$$

Therefore the inverse of the information matrix $F = \frac{X^TX}{n}$ is expressed as

$$F^{-1} = \frac{n}{n(n+p)} \begin{pmatrix} n+p-1 & 1 & -1 & \cdots & -1 & 1 \\ 1 & n+p-1 & 1 & \cdots & 1 & -1 \\ \cdots & & & & & \\ -1 & 1 & -1 & \cdots & n+p-1 & 1 \\ 1 & -1 & 1 & \cdots & 1 & n+p-1 \end{pmatrix} \tag{3.3.4}$$

with $f(t)^T = \left(\sqrt{2}\sin(\frac{1}{2}\pi t) \quad \sqrt{2}\sin(\frac{3}{2}\pi t) \quad \cdots \quad \sqrt{2}\sin((p-\frac{1}{2})\pi t)\right)$. The sensitivity

function under the uniform design ξ^* is

$$\begin{aligned} d(t,\xi^*) &= f(t)^T F^{-1} f(t) = 2\sum_{i=1}^{p} \frac{n(n+p-1)}{n(n+p)} \sin^2[(i-\frac{1}{2})\pi t] \\ &+ 2\sum_{i\neq j} (-1)^{i+j+1} \frac{n}{n(n+p)} \sin[(i-\frac{1}{2})\pi t]\sin[(j-\frac{1}{2})\pi t] \end{aligned}$$

Also notice that

$$2\sum_{i=1}^{p} \sin^2[(i-\frac{1}{2})\pi t] = \sum_{i=1}^{p}\{1-\cos[(2i-1)\pi t]\} = p - \frac{\sin(\pi t)\sin(2\pi tp)}{1-\cos(2\pi t)} \qquad (3.3.5)$$

Now the equally spaced uniform sampling can be related to the approximate D-optimal design for the Brownian motion. The result is summarised in theorem 3.3.1.

Theorem 3.3.1. *For the Brownian motion on* $[0,1]$, *the uniform sampling design* ξ^* *with the sampling points* $t_i = \frac{i}{n}$, $1 \leq i \leq n$ *is optimal in the following two senses*

(i) It is a maximum entropy sampling.

(ii) It is an approximate D-optimal design in the sense that: for all $\epsilon > 0$ *and any point* $t \in [\epsilon, 1-\epsilon]$, *if* $n \to \infty$ *and* $p \to \infty$ *in such a way that* $n \geq p$, $lim_{p\to\infty}\frac{n}{p} = c$, *where* c *is a constant and* $c \geq 1$, *then*

$$lim_{p\to\infty}\frac{d(t,\xi^*)}{p} = 1 \qquad (3.3.6)$$

Proof. The first part of the theorem on maximum entropy sampling is shown at the beginning of this section. The equally spaced design on $[0,1]$ maximises the entropy of the Brownian motion. We only need to show that the second part of the theorem

now.

$$
\begin{aligned}
& d(t,\xi^*) \\
= \; & 2\sum_{i=1}^{p}\frac{(n+p-1)}{(n+p)}\sin^2[(i-\frac{1}{2})\pi t]+2\sum_{i\neq j}(-1)^{i+j+1}\frac{1}{(n+p)}\sin[(i-\frac{1}{2})\pi t]\sin[(j-\frac{1}{2})\pi t] \\
= \; & \frac{n+p-1}{n+p}2\sum_{i=1}^{p}\sin^2[(i-\frac{1}{2})\pi t] \\
& +\frac{2}{n+p}\{\sum_{i=1}^{p}\sum_{j=1}^{p}(-1)^{i+j+1}\sin[(i-\frac{1}{2})\pi t]\sin[(j-\frac{1}{2})\pi t]-\sum_{i=1}^{p}(-1)^{2i-1}\sin^2[(i-\frac{1}{2})\pi t]\} \\
= \; & 2\sum_{i=1}^{p}\sin^2[(i-\frac{1}{2})\pi t]-\frac{2}{n+p}\{\sum_{i=1}^{p}(-1)^i\sin[(i-\frac{1}{2})\pi t]\}^2 \\
= \; & p-\frac{\sin(\pi t)\sin(2\pi tp)}{1-\cos(2\pi t)}-\frac{2}{(\frac{n}{p}+1)p}\{\sum_{i=1}^{p}(-1)^i\sin[(i-\frac{1}{2})\pi t]\}^2
\end{aligned}
$$

Then we obtain $\frac{d(t,\xi^*)}{p}$

$$
\frac{d(t,\xi^*)}{p}=1-\frac{1}{p}\frac{\sin(\pi t)\sin(2\pi tp)}{1-\cos(2\pi t)}-\frac{2}{(\frac{n}{p}+1)}\{\frac{\sum_{i=1}^{p}(-1)^i\sin[(i-\frac{1}{2})\pi t]}{p}\}^2 \qquad (3.3.7)
$$

For interval $[\epsilon, 1-\epsilon]$, there always exists $M_1 \geq \frac{1}{1-\cos(2\pi\epsilon)} > 0$ such that

$$
|\frac{\sin(\pi t)\sin(2\pi tp)}{1-\cos(2\pi t)}| \leq \frac{1}{1-\cos(2\pi\epsilon)} \leq M_1 \qquad (3.3.8)
$$

Hence

$$
\lim_{p\to\infty}\frac{1}{p}\frac{\sin(\pi t)\sin(2\pi tp)}{1-\cos(2\pi t)}=0 . \qquad (3.3.9)
$$

Also, for interval $[\epsilon, 1-\epsilon]$, there always exists $M_2 \geq \frac{9}{2(1-\cos(\pi\epsilon))} > 0$, such that

$$
\begin{aligned}
&|\sum_{i=1}^{p}(-1)^i \sin[(i-\frac{1}{2})\pi t]| = |\frac{1}{2[\cos(\pi t)+1]}\{(-1)^p \cos(\frac{1}{2}\pi t)\sin[\pi t(p+1)]\cos(\pi t) \\
+\ & (-1)^p \sin[\pi t(p+1)]\cos(\frac{1}{2}\pi t) - (-1)^p \cos(\frac{1}{2}\pi t)\cos[\pi t(p+1)]\sin(\pi t) \\
-\ & (-1)^p \sin(\frac{1}{2}\pi t)\sin[\pi t(p+1)]\sin(\pi t) - (-1)^p \sin(\frac{1}{2}\pi t)\cos[\pi t(p+1)]\cos(\pi t) \\
-\ & (-1)^p \cos[\pi t(p+1)]\sin(\frac{1}{2}\pi t) - \cos(\frac{1}{2}\pi t)\sin(\pi t) + \sin(\frac{1}{2}\pi t) + \cos(\pi t)\sin(\frac{1}{2}\pi t)\}| \\
\leq\ & \frac{9}{2(1-\cos(\pi\epsilon))} \leq M_2
\end{aligned}
$$

Therefore

$$\lim_{p\to\infty} \frac{\sum_{i=1}^{p}(-1)^i \sin[(i-\frac{1}{2})\pi t]}{p} = 0 \tag{3.3.10}$$

Since $\lim_{p\to\infty} \frac{n}{p} = c$, where $c \geq 1$,

$$\lim_{p\to\infty} \frac{2}{(\frac{n}{p}+1)} \left\{ \frac{\sum_{i=1}^{p}(-1)^i \sin[(i-\frac{1}{2})\pi t]}{p} \right\}^2 = 0 \tag{3.3.11}$$

Then we can conclude that for all $\epsilon > 0$ and any point $t \in [\epsilon, 1-\epsilon]$, if $n \to \infty$ and $p \to \infty$ in such a way that $n \geq p$, $\lim_{p\to\infty} \frac{n}{p} = c$, where c is a constant and $c \geq 1$, then, (3.3.9) and (3.3.11) hold. Following equation (3.3.7), we obtain,

$$\lim_{p\to\infty} \frac{d(t,\xi^*)}{p} = 1 \tag{3.3.12}$$

□

Theorem 3.3.1 shows that when both n and p are quite big in a certain relationship such that $\lim_{p\to\infty} \frac{n}{p} = c$, where $c \geq 1$, then we can almost get a straight line p for the sensitivity function across the interval except for the bounds. Figure 3.2 and figure 3.3 further confirm the result. In figure 3.2, $\lim_{p\to\infty} \frac{n}{p} = 1$, while in figure 3.3, $\lim_{p\to\infty} \frac{n}{p} = 5$. When n and p becomes bigger, the variation at the bounds becomes smaller. The value of $d(t,\xi^*)$ across the interval is almost p, when p goes to 1000.

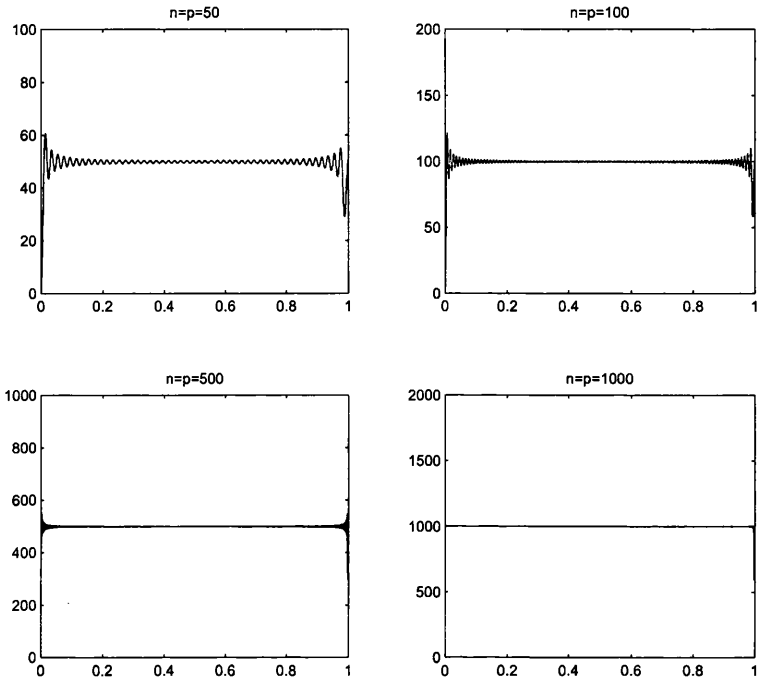

Figure 3.2: The sensitivity function $d(t)$ of the equally spaced uniform design for the Brownian motion, when choosing $n = p = 50, 100, 500$ and 1000 using the truncated Karhunen-Loeve expansion.

Remark 3.3.1. When $n \to \infty$, the discrete uniform design becomes a continuous uniform design. We denote the continuous uniform design as ξ^{**}

In the continuous uniform case, measure $p(t) = 1, 0 \leq t \leq 1$, hence the information matrix can be calculated as follows

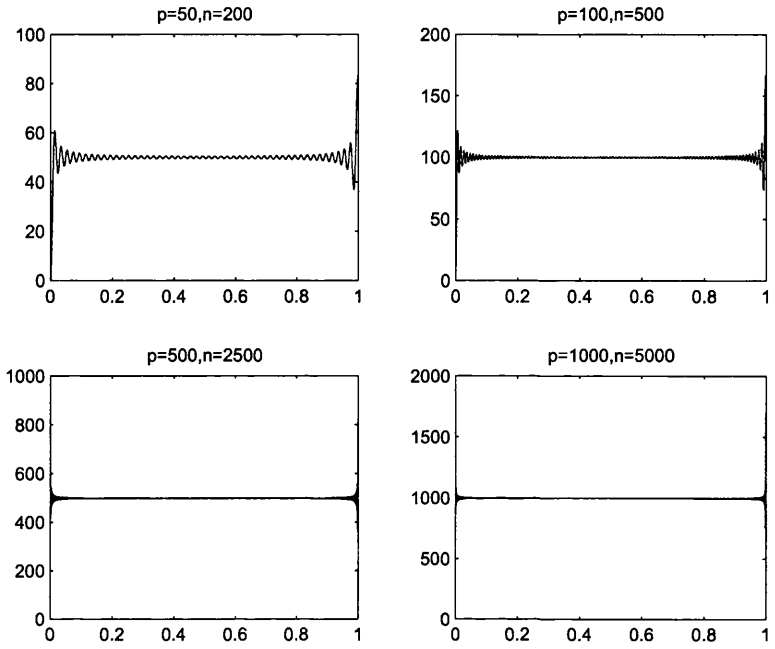

Figure 3.3: The sensitivity function $d(t)$ of the equally spaced uniform design for the Brownian motion, when choosing $p = 50, 100, 500$ and 1000 using the truncated Karhunen-Loeve expansion and $n = 5p$.

$$
\begin{aligned}
F &= \int_0^1 f(t)f(t)^T dt \\
&= \int_0^1 \begin{pmatrix} \phi_1(t) \\ \phi_2(t) \\ \cdots \\ \phi_p(t) \end{pmatrix} \begin{pmatrix} \phi_1(t) & \phi_2(t) & \cdots & \phi_p(t) \end{pmatrix} dt \\
&= \begin{pmatrix} \int_0^1 \phi_1^2(t)dt & \int_0^1 \phi_1(t)\phi_2(t)dt & \cdots & \int_0^1 \phi_1(t)\phi_p(t)dt \\ \int_0^1 \phi_2(t)\phi_1(t)dt & \int_0^1 \phi_2^2(t)dt & \cdots & \int_0^1 \phi_2(t)\phi_p(t)dt \\ & \cdots & & \\ \int_0^1 \phi_p(t)\phi_1(t)dt & \int_0^1 \phi_p(t)\phi_2(t)dt & \cdots & \int_0^1 \phi_p^2(t)dt \end{pmatrix}.
\end{aligned}
$$

Since the eigenfunctions are orthogonal, i.e. $\int_0^1 \phi_i(t)\phi_j(t)dt = \delta_{i,j}$, $F = I_{p\times p}$. Then

$$d(t,\xi^{**}) = f(t)^T F^{-1} f(t) = \sum_{i=1}^{p} \phi_i^2(t) = \sum_{i=1}^{p} 2\sin^2[(i-\frac{1}{2})\pi t] = p - \frac{\sin(\pi t)\sin(2\pi t p)}{1-\cos(2\pi t)}$$

Following the same calculation procedure as that for the proof of theorem 3.3.1, we can show that for all $\epsilon > 0$, there exists an interval $[\epsilon, 1-\epsilon]$, such that

$$\lim_{p\to\infty} \frac{d(t,\xi^{**})}{p} = 1 \tag{3.3.13}$$

It means that the continuous uniform design is also an approximate D-optimal design in the Brownian motion case on the interval $[\epsilon, 1-\epsilon]$.

Remark 3.3.2. The second type integrated Brownian motion is defined, in chapter 1, as

$$X_m^2(t) = \begin{cases} \int_0^{t_m}\int_{t_{m-1}}^1 \cdots \int_0^{t_1} W_1(t_0)dt_0\cdots dt_{m-1} & 0<t<\frac{1}{n} \quad \text{if } m \text{ is odd} \\ \int_0^{t_m}\int_{t_{m-1}}^1 \cdots \int_{t_1}^1 W_0(t_0)dt_0\cdots dt_{m-1} & 0<t<\frac{1}{n} \quad \text{if } m \text{ is even} \end{cases} \tag{3.3.14}$$

with covariance function

$$K_m^2(s,t) = \int_0^1 \cdots \int_0^1 (\min(s,s_1))(\min(s_1,s_2))\cdots(\min(s_m,t))ds_1 ds_2 \cdots ds_m \tag{3.3.15}$$

The eigenvalue and eigenfunction for X_m^2 are

$$\lambda_i = \left(\frac{4}{(2i-1)^2\pi^2}\right)^{m+1}, \quad \phi_i(t) = \sqrt{2}\sin[(i-\frac{1}{2})\pi t] \quad i \geq 1 \tag{3.3.16}$$

Its eigenfunction is the same as that of the Brownian motion. For the D-optimality, since it is the eigenfunction, not the eigenvalue, which relates to maximum entropy sampling. Hence the optimal sampling for the second type integrated Brownian motion is the same as that for the Brownian motion.

3.4 Numerical methods

In practice, the analytical solutions to the Karhunen-Loeve expansion are quite restricted to certain processes. The numerical methods have to be implemented instead to derive the Karhunen-Loeve expansion for many processes. In this section, in order to relate maximum entropy sampling to the D-optimality in the numerical setting, the expansion method is mainly considered. Since it is the expansion method which treats the eigenfunction as a function, so that the sensitivity function $d(x)$ can be calculated under the optimal design in the continuous interval. For the integration method, very fine intervals are required for $d(x)$. It involves loads of data, which might not be computationally efficient. For the sake of accuracy, we discuss the expansion method using the Haar wavelet here.

Although the uniform equally spaced design is one of the easiest designs to handle, it is not always D-optimal. We choose n equally spaced design points $t_i = \frac{2i-1}{2n}$, $1 \leq i \leq n$ to show this result. The condition required for these design points to be D-optimal is that the number of the eigenvalues used in the truncated Karhunen-Loeve expansion should be equal to the number of the basis functions, i.e. $M = p$. However, the condition $M = p$ does not allow p and M to go to infinity at different rates, which does not reflect the reality.

We assume that both M and n are specifically chosen, so that $n \geq M$ and both of them are the power of 2, i.e. $M = 2^x$, $n = 2^y$, $y \geq x$, $x, y \in N$. Using the same notation as that in chapter 2, matrix Φ for the eigenfunctions containing all the n design points of order p can be expressed as $\Phi = \Psi^T D^T$, where Ψ is a matrix involving

the Haar basis functions of order M, i.e.

$$\begin{aligned}
\Psi^T &= \left(\psi_1,\ \psi_2,\ \cdots,\ \psi_M\right) \\
\psi_i^T &= \left(\psi_i(t_1),\ \psi_i(t_2),\ \cdots,\ \psi_i(t_n)\right) \\
\psi_i(t) &= \psi_{j,k}(t) = \psi(2^j t - k), i = 2^j + k + 1, 0 \le j \le x-1, 0 \le k \le 2^j - 1
\end{aligned}$$

The equally spaced design points under consideration are $t_i = \frac{2i-1}{2n}$, $1 \le i \le n$, where $i = 2^j + k + 1$, $0 \le j \le x-1$, $0 \le k \le 2^j - 1$. In each ψ_i $(i \ge 2)$, from the $\frac{n}{2^j}k + 1$th element to the $\frac{n}{2^j}k + \frac{n}{2^{j+1}}$th element, the element value is 1. From the $\frac{n}{2^j}k + \frac{n}{2^{j+1}} + 1$th element to the $\frac{n}{2^j}(k+1)$th element, the element value is -1. The values for the rest elements are 0. It means that, when $i \ge 2$

$$\psi_i = \left(\underbrace{0,\ 0,\ \cdots,\ 0}_{\frac{n}{2^j}k},\ \underbrace{1,\ 1,\ \cdots,\ 1}_{\frac{n}{2^{j+1}}},\ \underbrace{-1,\ -1,\ \cdots,\ -1}_{\frac{n}{2^{j+1}}},\ \underbrace{0,\ 0,\ \cdots,\ 0}_{n-\frac{n}{2^j}(k+1)}\right) \tag{3.4.1}$$

To make this more clear,

$$\begin{aligned}
\psi_1 &= \left(\underbrace{1,\ 1,\ \cdots,\ 1}_{n}\right) \\
\psi_2 &= \left(\underbrace{1,\ 1,\ \cdots,\ 1}_{\frac{n}{2}},\ \underbrace{-1,\ -1,\ \cdots,\ -1}_{\frac{n}{2}}\right) \\
\psi_3 &= \left(\underbrace{1,\ 1,\ \cdots,\ 1}_{\frac{n}{4}},\ \underbrace{-1,\ -1,\ \cdots,\ -1}_{\frac{n}{4}},\ \underbrace{0,\ 0,\ \cdots,\ 0}_{\frac{n}{2}}\right) \\
\psi_4 &= \left(\underbrace{0,\ 0,\ \cdots,\ 0}_{\frac{n}{2}},\ \underbrace{1,\ 1,\ \cdots,\ 1}_{\frac{n}{4}},\ \underbrace{-1,\ -1,\ \cdots,\ -1}_{\frac{n}{4}}\right) \\
&\cdots \\
\psi_M &= \left(\underbrace{0,\ 0,\ \cdots,\ 0}_{n-\frac{2n}{M}},\ \underbrace{1,\ 1,\ \cdots,\ 1}_{\frac{n}{M}},\ \underbrace{-1,\ -1,\ \cdots,\ -1}_{\frac{n}{M}}\right)
\end{aligned}$$

Therefore,

$$\Psi\Psi^T = \begin{pmatrix} l_1 & 0 & \cdots & 0 \\ 0 & l_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & l_M \end{pmatrix} \tag{3.4.2}$$

where $l_1 = 1$ and $l_i = \frac{n}{2^j}$. It means that $\Psi\Psi^T = nH$.

In order to derive the coefficient matrix D, the eigenvalue problem $\Lambda(DH^{\frac{1}{2}}) = (DH^{\frac{1}{2}})(H^{\frac{1}{2}}AH^{\frac{1}{2}})$ needs to be solved. After solving for D, the matrix X in the D-optimality can be expressed as $X = \Phi^T = D\Psi_{M\times n}$. Then the information matrix F is simply an identity matrix, since,

$$F = \frac{1}{n}X^TX = \frac{1}{n}D\Psi\Psi^TD^T = \frac{1}{n}DnHD^T = I \tag{3.4.3}$$

The last equality comes from the fact that in the eigen-equation $\Lambda(DH^{\frac{1}{2}}) = (DH^{\frac{1}{2}})(H^{\frac{1}{2}}AH^{\frac{1}{2}})$, the eigenvectors $DH^{\frac{1}{2}}$ are orthogonal. Under the equally spaced uniform design ξ, if $f(t) = D\psi(t)$, the sensitivity function $d(t,\xi)$ is

$$\begin{aligned} d(t,\xi) &= f(t)^TF^{-1}f(t) \\ &= \psi(t)^TD^TF^{-1}D\psi(t) \\ &= \psi(t)^TD^TD\psi(t) \end{aligned}$$

Up till now, the order p for the truncated Karhunen-Loeve expansion can be chosen freely, as long as $n \geq p$ for the consideration of the inverse calculation of F. If the uniform equally spaced design is D-optimal, $d(t,\xi) = p$. However, $d(t,\xi) = p$ does not necessarily hold unless further assumptions on D^TD are made. One of the possibilities to support $d(t,\xi) = p$ is to assume $D^TD = H^{-1}$. This equality holds

when $M = p$. Hence under $M = 2^x = p$,

$$\begin{aligned} d(t,\xi) &= \psi(t)^T H^{-1} \psi(t) = 1 + \sum_{i=2}^{p} 2^j \psi_i^2(t) \\ &= 1 + \sum_{j=0}^{x-1} \sum_{k=0}^{2^j-1} 2^j \psi_{2^j+k+1}^2(t) = 1 + \sum_{j=0}^{x-1} 2^j \\ &= 2^x = M = p \end{aligned}$$

Since for $0 \leq t \leq 1$, and for each level j, there exists one and only one k, such that $\psi_{2^j+k+1}^2(t) = 1$, while the rest $\psi_{2^j+k+1}^2(t) = 0$.

However, as we have mentioned at the beginning of this section that $M = p$ is not a realistic assumption in practice. M and p should be able to be raised at different rates to serve different approximation purposes. M should be able to increase for the accuracy of the eigenfunction, while p should be able to increase for the better approximation of the original process. Hence $M = p$ lacks of flexibility in reality and therefore we cast doubt on treating the uniform equally spaced design as D optimal.

In practice, finding the D-optimality points depends on the eigenfunctions of the process, which is not always analytically tractable, hence we can not always derive the analytical solution to the optimal points. The rest of the chapter contributes to an algorithm looking for n design points to maximise the information matrix, i.e. maximise $nF = \det(X^T X)$, where $X = D\Psi$ using the Haar wavelet method. The algorithm we are using is called the "DETMAX". DETMAX was first introduced by Mitchell (1974) and then improved by Galil and Kiefer (1980). The DETMAX algorithm is an exchange algorithm. Each time it adds a point out of all the N candidate points or subtracts a point out of all the current chosen design points that can maximise $\det(X^T X)$. It also defines a failure set $\mathfrak{F}$ comprising of designs that does not improve determinant after each excursion. At the beginning of the algorithm, $\mathfrak{F}$ is

set to be an empty set and a random initial design with size n out of the N candidate points will be chosen. Each excursion starts from a set of design points with size n and then randomly decides whether to add or subtract a point first. With the size of the current design D denoted by n', we continue an excursion following the rule suggested by Mitchell (1974)

- If $n' > n$, we subtract a candidate point if D is not in $\mathfrak{F}$. Otherwise we add a candidate point.
- If $n' < n$, we add a candidate point if D is not in $\mathfrak{F}$. Otherwise we subtract a candidate point.

Here, "add"a point means that with the current $\det(X_\xi^T X_\xi)$, we choose a point t out of N candidate points to add to our design, so that we can maximise

$$\det[\begin{pmatrix} X_\xi^T, & f(t) \end{pmatrix} \begin{pmatrix} X_\xi \\ f(t)^T \end{pmatrix}] = \det(X_\xi^T X_\xi)[1 + f(t)^T (X_\xi^T X_\xi)^{-1} f(t)] \tag{3.4.4}$$

while "subtract" a point means that with the current $\det(X_\xi^T X_\xi)$, we choose a point t to remove out of the current design so that we can maximise

$$\det(X_\xi^T X_\xi)[1 - f(t)^T (X_\xi^T X_\xi)^{-1} f(t)] \tag{3.4.5}$$

Each excursion stops when $n' = n$. If there is an improvement in the information matrix, $\mathfrak{F}$ is emptied. Otherwise, all the designs during the excursion will be added to $\mathfrak{F}$.

The algorithm stops when $|n' - n| \leq 6$ to avoid the problem of longer algorithm and longer time. As is mentioned by Mitchell (1974), a longer algorithm does not seem to provide a better result. The other problem of the DETMAX algorithm is that it is

very easy to get stuck in a local maximum and does not guarantee a global optimum. In our case, the local maximum happens quite often and the algorithm continues with adding one point and then subtracting the same point. Once the excursion has run for more than a certain number of steps, say 200 in our case, and the algorithm still gets stuck in the local maximum and does not able to get out, we also stop the algorithm for the sake of computational efficiency. One way of improving the performance of the DETMAX algorithm is to run the algorithm several times with different starting points. We follow the method suggested by Galil and Kiefer (1980) to choose the initial design. It is found to be computationally faster than a completely random start.

Another drawback of the DETMAX algorithm is that it can not handle very big p due to the current computer power, although DETMAX is generally regarded as one of the fastest algorithms to maximise $\det(X^TX)$. For the process in the Sacks-Ylvisaker family, p is usually chosen to be big, say 1000, so that its first p eigenvalues can explain almost all the cumulative expected variance, i.e. $\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{\infty}\lambda_i}$ is almost 1. Since $n \geq p$ is required for the singularity of the information matrix, big p implies bigger n and even bigger N. In this situation, an extremely long computational time is usually expected and the algorithm is very likely to get stuck in a local maximum. Hence the most suitable process to implement the DETMAX algorithm is the smooth process whose cumulative expected variance can be almost captured by the first few eigenvalues, for example, the Gaussian process with the squared exponential kernel $\exp[(-(s-t)^2]$. Its first two eigenvalues already explain more than 99% of the cumulative expected variance. Therefore they allow big choice of n under small p, and the current computer power can handle the calculation.

For accuracy, when using the DETMAX algorithm, we always assume that the candidate points are $t_i = \frac{2i-1}{2N}$, $1 \leq i \leq N$, $N = 5000$, and the order for approximating the eigenfunction using the Haar wavelet is $M = 256$. The following figures on $\log[\det(X^T X)]$ and $\max[d(x)]$ at each iteration only show the results when each excursion finishes and succeeds, i.e. $n' = n$. When $n = p$, the algorithm stops at the last excursion that results in $n' = n$, while when $n > p$, the algorithm stops at the optimal design with size n' rather than n with $0 < |n' - n| \leq 6$.

We first consider the case when $n = p$. In order to show that the DETMAX algorithm can actually provide the optimal points in terms of D-optimality, we start from the example of the Brownian motion when $n = p = 2$. Although in this case, the optimal points we find is not optimal in terms of the maximum entropy criteria due to small p, the solution could be used to compare with the analytical solution we derived in section 3.1.

Figure 3.4 shows the result for $n = p = 2$ using the DETMAX. The result for the determinant converges in the first few iterations. The optimal points are 0.3907 and 0.9961, who are the 1954th and the 4981th point out of our 5000 candidate points respectively. They are very close to our analytical solution $\frac{2\arccos\frac{\sqrt{6}}{3}}{\pi} \approx 0.3918$ and 1. Both of them are equally chosen with probability $\frac{1}{2}$.

As we have mentioned, due to the limitation of the computational power, the Brownian motion is not a very good example to show the power of the DETMAX algorithm, since it has too much energy in the tail. The DETMAX algorithm is more useful in the process, whose cumulative expected variance can be explained by the first few eigenvalues. In order to show that the optimal points we find through the DETMAX algorithm under the D-optimality criteria is also the optimal points

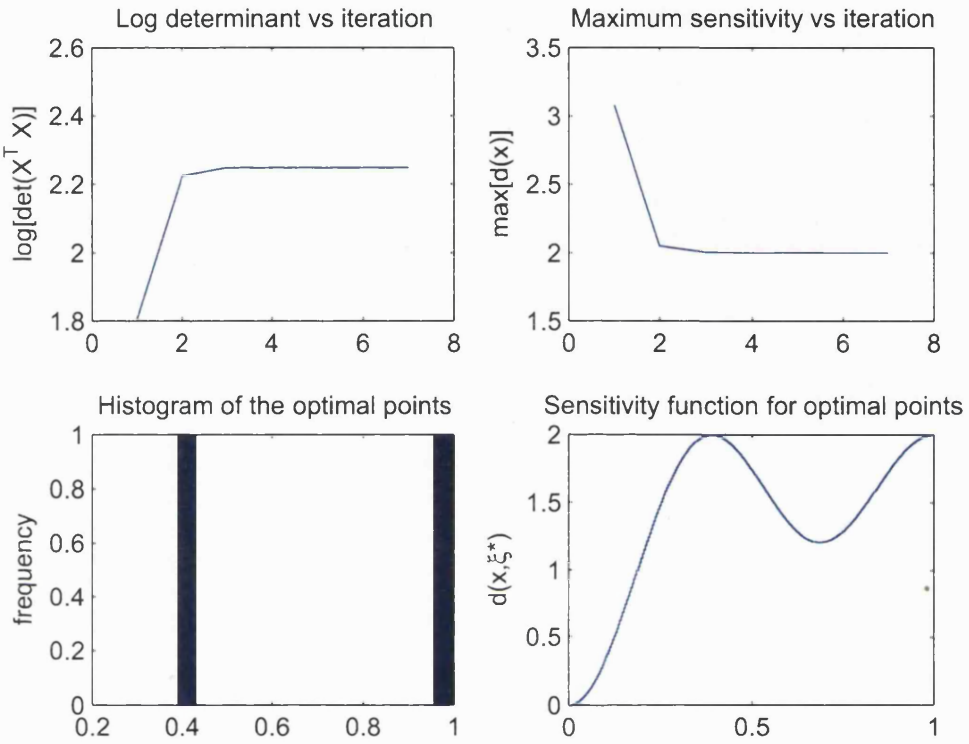

Figure 3.4: The Brownian motion when $n = p = 2$. (Top left) $\log[\det(X^T X)]$ vs iteration; (Top right) $\max(d(x))$ vs iteration; (Bottom left) Histogram for the optimal points; (Bottom right) Sensitivity function $d(x, \xi^*)$ under the optimal design.

under the maximum entropy criteria, we need to use another algorithm on finding the optimal points that maximise the entropy. The most common algorithms to maximise the entropy are the greedy exchange algorithm and the branch and bound algorithm. See, for example, Ko et al. (1995), for detail. The greedy exchange algorithm is fast in practice. The idea behind it is essentially the same as that behind the DETMAX. The algorithm starts from a random set with size $s < n$. Then at each step it increases the size of the design space by one so that the current design has the maximum

entropy. The greedy exchange algorithm also has the same drawback as that of the DETMAX, which is that the algorithm sometimes gets stuck in the local maximum. The branch and bound algorithm improves the greedy exchange algorithm, however, it takes extremely long time for the computation. We tried to choose 30 points out of 60 points using Matlab 7.0.4. It took the computer for more than 2 days' work, and it still did not return a result. Therefore, we only use the greedy exchange algorithm as a method for finding the optimal points under the maximum entropy. Due to the computing power, the size of the candidate points are set to be $N_g = 1024$ with each point $t_i = \frac{2i-1}{2N_g}, 1 \leq i \leq N_g$. We try to avoid the local maximum by running the algorithm several times.

We now focus on the Gaussian kernel with the covariance function $\exp[-(t-s)^2]$ when $n = p$, since its first two eigenvalues explain more than 99% of the cumulative expected variance. We expect that sampling using either the maximum entropy criteria or the D-optimality should result in more or less the same solution. We start again from the easiest case when $n = p = 2$. In fact, for $n = 2$, the analytical solution is still tractable. Assume we want to choose points t_1 and t_2 $(t_1 \leq t_2)$ on $[0, 1]$, so that we can maximise the following

$$\det \begin{pmatrix} \exp[-(t_1-t_1)^2] & \exp[-(t_2-t_1)^2] \\ \exp[-(t_2-t_1)^2] & \exp[-(t_2-t_2)^2] \end{pmatrix} = 1 - \{\exp[-(t_2-t_1)^2]\}^2 \qquad (3.4.6)$$

$1 - \{\exp[-(t_2-t_1)^2]\}^2$ can be maximized, if $|t_2 - t_1|$ can be maximized. In our case, the optimal points should be the bounds of the interval, i.e. $t_1 = 0$ and $t_2 = 1$. The result provided by the greedy exchange algorithm matches this. For $n = 2$, it ends up choosing 0.000488 and 0.99951, which are the first and the last points respectively in its candidate points set.

Figure 3.5 uses the DETMAX algorithm for the D-optimality points when $n =$

$p = 2$. The solution it provides is $t_1 = 0.0001$ and $t_2 = 0.9961$. They are the 1st and the 4981th point of its 5000 candidate points in on $[0, 1]$ respectively. They are also very close to our analytical solution.

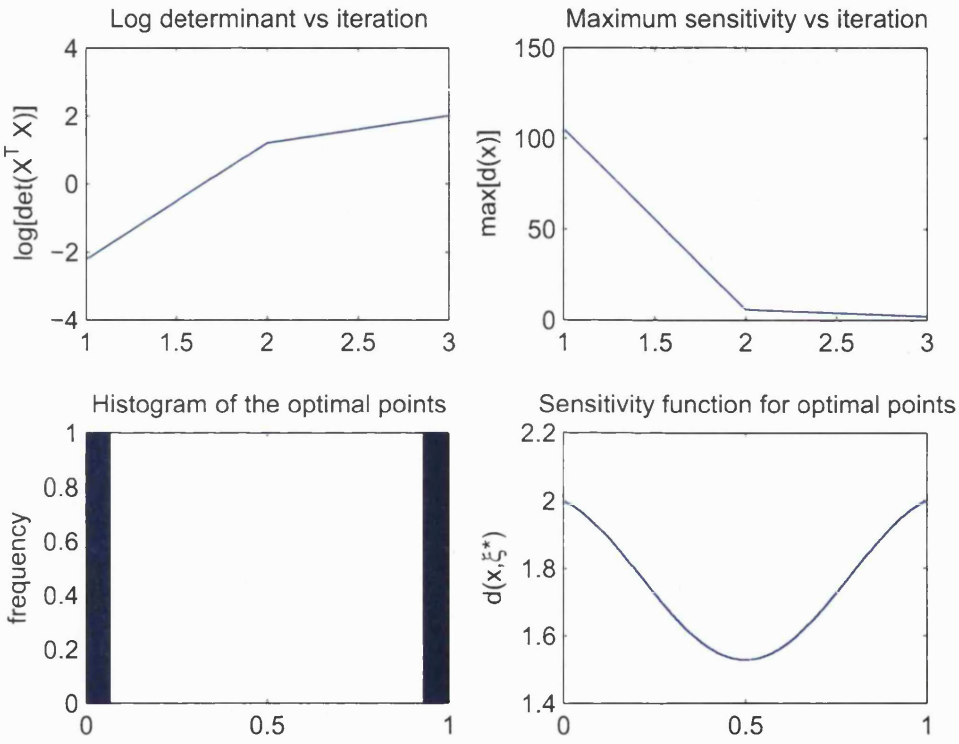

Figure 3.5: Kernel $\exp[-(t-s)^2]$ when $p = 2$. (Top left) $\log[\det(X^T X)]$ vs iteration; (Top right) $\max(d(x))$ vs iteration; (Bottom left) Histogram for the optimal points; (Bottom right) Sensitivity function $d(x, \xi^*)$ under the optimal design.

Figure 3.7 further investigates DETMAX algorithm on kernel $\exp[-(t-s)^2]$ from $n = p = 2$ to $n = p = 11$. It plots the sensitivity function $d(t, \xi^*)$ under the optimal design. For small n, the optimal points the DETMAX chooses satisfy D-optimality, since $\max d(t, \xi^*) = p$. When $n \geq 6$, there appears some strange behavior in the

sensitivity function, whose peak is bigger than p. If we run the algorithm for several times, the peak might move up and down due to the different choice of the optimal points. The peak bigger than p could be caused by three possible reasons. Firstly, as we have mentioned, the DETMAX algorithm is very likely to get stuck in a local maximum rather than a global maximum. Secondly, the rounding error is very easy to accumulate in the algorithm with big number of points and iterations. Last but not least, although our size of the candidate points is as big as $N = 5000$, it is still an approximation to the continuous interval. The exact location of the optimal points might not be included in the candidate points set. Hence strange behavior of the sensitivity function can be regarded as acceptable and is due to the numerical error.

The reasons mentioned above also implies that the sensitivity function itself is very sensitive to the choice of the optimal points. Even a small difference in the points could result in a big difference in the sensitivity function. Figure 3.6 and table 3.1 show four different tries of DETMAX when $n = p = 6$. The first try is the best, $\max d(t, \xi^*)$ is very close to 6, hence its $\det(X^T X)$ is also the highest among the four. The second try and the third try have one peak above 6, but the location of the peak is different. $\det(X^T X)$ of the second try is bigger than that of the third try. It is not only due to its optimal points much closer to the real solution, but also its sensitivity function more symmetric and more regular. The last try involves 2 irregular peaks above 6, which result in the worst $\det(X^T X)$ among all. Therefore, one way of getting out of local maximum using the DETMAX is to try the algorithm several times.

The DETMAX algorithm does provide reasonable optimal points after choosing the best results among all the tries. Figure 3.8 explains this by comparing the cumulative distribution function of the optimal points provided by the DETMAX with that

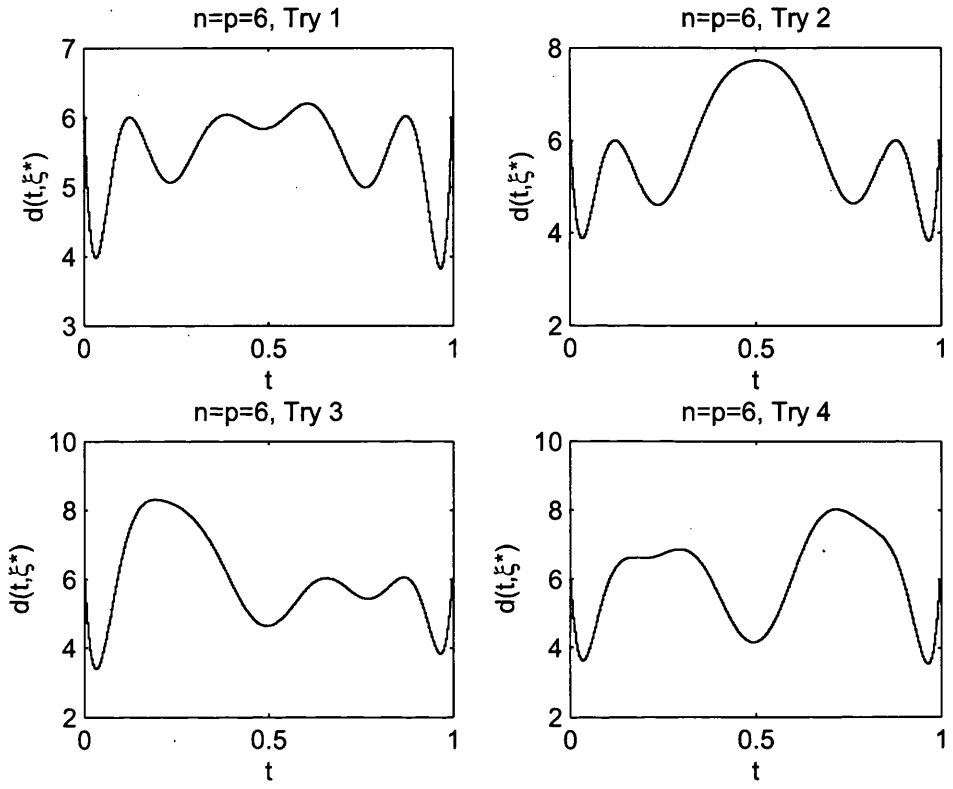

Figure 3.6: The sensitivity function $d(t, \xi^*)$ for four different tries using the DETMAX algorithm on the kernel $\exp[-(t-s)^2]$ when $n = p = 6$. (see, also, table 3.1)

provided by the greedy algorithm. Each jump in the y axis (cumulative distribution function $F(t)$) represents an optimal point in the x axis (time t) on $[0, 1]$. Since our optimal measure is uniform, the magnitudes of the jump are the same within each plot, which is $\frac{1}{n}$. With low n and p, the optimal points for both algorithms are almost the same. When n and p are increased, the difference of the optimal points between both algorithms is still relatively small. In fact, the closer the optimal points are to the bounds, which is 0 and 1 in our case, the smaller the difference is between both

	1st point	2nd point	3rd point	4th point	5th point	6th point	$\log[\det(X^TX)]$
Try 1	0.0001	0.1291	0.3663	0.6525	0.8791	0.9961	13.1954
Try 2	0.0001	0.1211	0.3351	0.6681	0.8791	0.9961	13.0932
Try 3	0.0001	0.0899	0.3947	0.6405	0.8791	0.9961	12.9989
Try 4	0.0001	0.1055	0.3791	0.5939	0.8947	0.9961	12.9795

Table 3.1: The optimal points of four different tries using the DETMAX algorithm on the kernel $\exp[-(t-s)^2]$ when $n = p = 6$. (see, also, figure 3.6)

algorithms. Also there are more optimal points that are closer to the bounds than the optimal points in the middle.

We now turn to the other situation when $n > p$. Using the DETMAX algorithm, we find out that out of the n design points found by the algorithm, quite a few points overlap. The number of the different optimal points is n_0, where $n > n_0$. Therefore, the frequency for each n_0 optimal point might be different depending on the number of points collapsing on it. The frequency for the optimal points can be treated as the design measure.

Theoretically, when measure is not uniform, and $n > p$, our argument for the equivalence of maximum entropy sampling and the D-optimality still holds after adjustment. The model we use is still the truncated Karhunen-Loeve expansion, plus the independent noise. However the variance of the noise should be defined as $\sigma^2 W^{-1}$ instead of $\sigma^2 I$, where W is a positive diagonal matrix not depending on the design points. Hence we need to maximise the entropy of the process

$$\det(\Phi^T \Lambda \Phi + \sigma^2 W^{-1}) \tag{3.4.7}$$

We can multiply the determinant of W to the above formula. Since W has no relationship with the design points, maximising the entropy of the process is equivalent

to maximising the following

$$\begin{aligned} & \det(W)\det(\Phi^T\Lambda\Phi + \sigma^2 W^{-1}) \\ = \; & \det(W^{\frac{1}{2}}\Phi^T\Lambda\Phi W^{\frac{1}{2}} + \sigma^2 I_{n\times n}) \\ = \; & \sigma^{2(n-p)}\det(\Lambda^{\frac{1}{2}}\Phi W\Phi^T\Lambda^{\frac{1}{2}} + \sigma^2 I_{p\times p}) \\ = \; & \sigma^{2(n-p)}[\det(\Lambda)\det(X^TWX)] + o(\sigma^2)] \quad \text{if } \sigma^2 \to 0 \end{aligned}$$

When the order p for the truncated Karhunen-Loeve expansion can explain almost all the cumulative expected variance, $\sigma^2 \to 0$. Hence maximising the entropy is equivalent to maximising $\det(X^TWX)$, where $X = \Phi^T$ and W can now be regarded as the weighting matrix comprising of the design measure. Notice that when $n = p$, the optimal measure for the D-optimal design is simply uniform. Since $\det(X^TWX) = \det(X^TX)\det(W)$, when $n = p$. W can then be maximized if all its diagonal elements are equal, which implies a uniform measure. Hence maximising the entropy is equivalent to maximising $\det(X^TX)$, which is the old problem we meet in section 3.1.

However, in practice, using the DETMAX algorithm, not only n points collapse to n_0 points, but also n_0 points have tendency to collapse to p points. Since there are a couple of points out of the n_0 points, that are very close to each other. They could be considered to be combined into one optimal point. If further comparing n_0 optimal points provided by the DETMAX and p optimal points provided by the greedy algorithm, they are quite close to each other, except for the measure. This will be gone through in detail in the following example on the Gaussian kernel $\exp[-(t-s)^2]$ still, but with $n = 500$ and $p = 11$. The reason we use such big n is that we hope to capture the right measure for each optimal point the DETMAX chooses.

Point	0.0001	0.0353	0.1173	0.2111	0.2267	0.3555	0.3673
Measure	0.0918	0.1796	0.0060	0.0040	0.0878	0.0758	0.0160
Point	0.4923	0.5001	0.5041	0.6291	0.6329	0.6407	0.6603
Measure	0.0200	0.0040	0.0619	0.0020	0.0020	0.0838	0.0020
Point	0.7697	0.7853	0.8829	0.8947	0.9611	0.9649	0.9961
Measure	0.0778	0.0140	0.0858	0.0040	0.0878	0.0020	0.0918

Table 3.2: The optimal points and their corresponding measure using the DETMAX for the kernel $\exp[-(t-s)^2]$, when $n = 500$ and $p = 11$.

Figure 3.9 shows the best result (in terms of maximising $\det(X^TX)$) after several tries. $n = 500$ initial points collapse to $n_0 = 21$ points (see table 3.2). Since $\max[d(t, \xi^*)]$ is very close to 11, where ξ^* is the optimal design, the algorithm provides reasonable optimal points meeting the D-optimality criteria. Within several tries of the algorithm, most of the points remain the same, while several points still change. It might be these minority points caused by the local maximum of the DETMAX, that will affect our further analysis when we compare the points with that provided by the greedy exchange algorithm. Nevertheless, most tries result in the points that make the sensitivity function close to p.

Another phenomenon worth mentioning here is that, although 500 points already collapse to 21 points, some points among 21 points are very close to each other, for example the $4th$ and the $5th$ point (0.2111 and 0.2267 respectively), or the 19th and the 20th point (0.9611 and 0.9649 respectively). If we combine these close points, the total number of optimal points could be further reduced to 11, which is the number of the parameters $p = 11$. Hence the number of the optimal points found by the DETMAX could be believed to be p, even when $n > p$. This can be further confirmed by comparing the results from the DETMAX and that from the greedy exchange, which is shown in figure 3.10.

In figure 3.10, although the optimal points from the DETMAX collapse to 21 and are D-optimal, these 21 points are quite different from the 21 points chosen by the greedy algorithm. In fact, after combining the close optimal points from the DETMAX and their corresponding measure, the new points almost match the 11 points chosen by the greedy algorithm to maximise the entropy. The only difference is on the measure. While after combination, most new points have the measure very close to $\frac{1}{11} \approx 0.0909$, there are 2 exceptions: the 2nd point 0.0353 with the measure as big as 0.1796 and the 3rd point 0.1173 with the measure as little as 0.006. This might be caused by DETMAX stuck in the local maximum. Although several other tries show that we can assign the right measure to these two points, there are other points whose measure might be affected.

In short, when $n > p$, the support of the optimal points found by the DETMAX are almost p. Except for the measure, which might be caused by the local maximum of DETMAX, the points (after combination) are also very close to the p optimal points found by the greedy algorithm. Hence the D-optimality and the maximum entropy criteria are almost equivalent even when $n > p$. Further research could be focused on the improvement of the DETMAX so that global optimal points could be found to maximise $\det(X^T X)$.

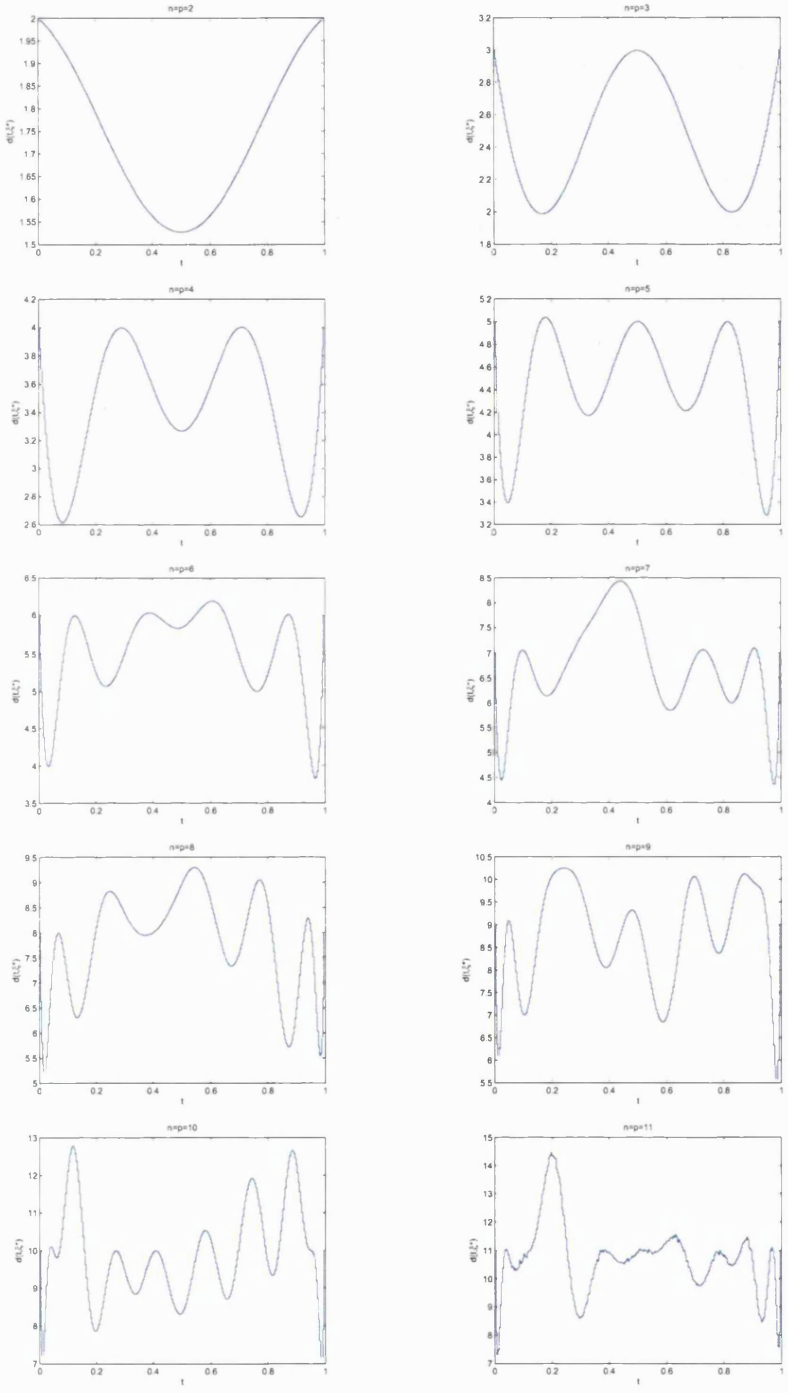

Figure 3.7: The sensitivity function $d(t, \xi^*)$ for the kernel $\exp[-(t-s)^2]$ under the optimal design, $n = p$ ranging from 2 to 11.

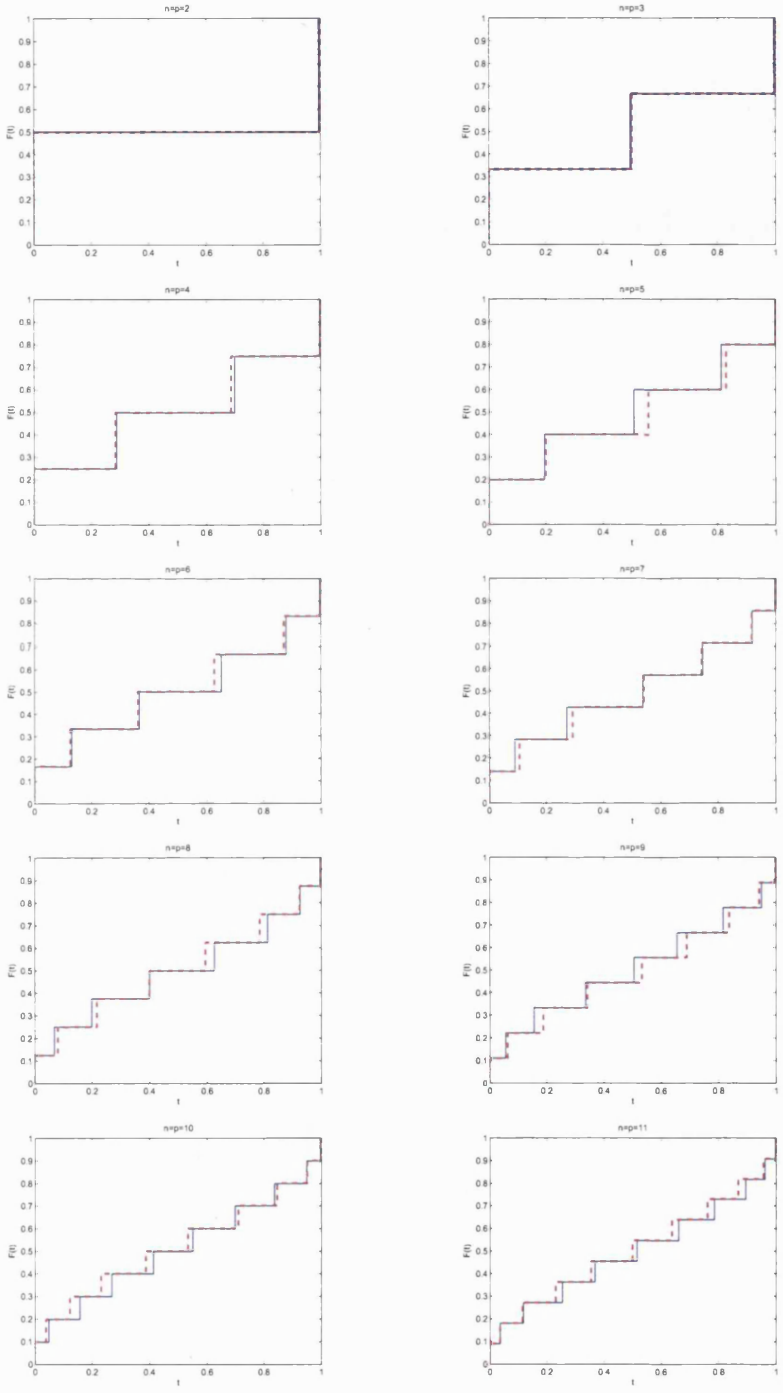

Figure 3.8: The cumulative distribution function for the optimal design points of the kernel $\exp[-(t-s)^2]$ using the DETMAX algorithm(blue solid line) and the greedy exchange algorithm(red dash line), $n = p$ ranging from 2 to 11.

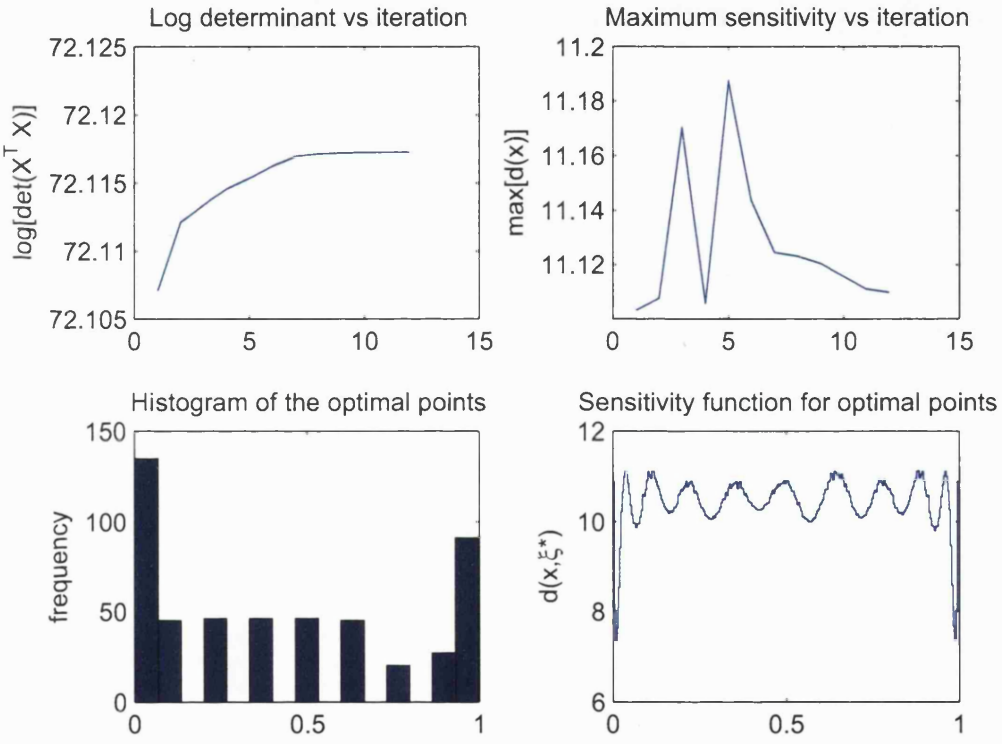

Figure 3.9: Kernel $\exp[-(t-s)^2]$ when $p = 11$ and $n = 500$. (Top left) $\log[\det(X^T X)]$ vs iteration; (Top right) $\max(d(x))$ vs iteration; (Bottom left) Histogram for the optimal points; (Bottom right) Sensitivity function $d(x, \xi^*)$ under the optimal design.

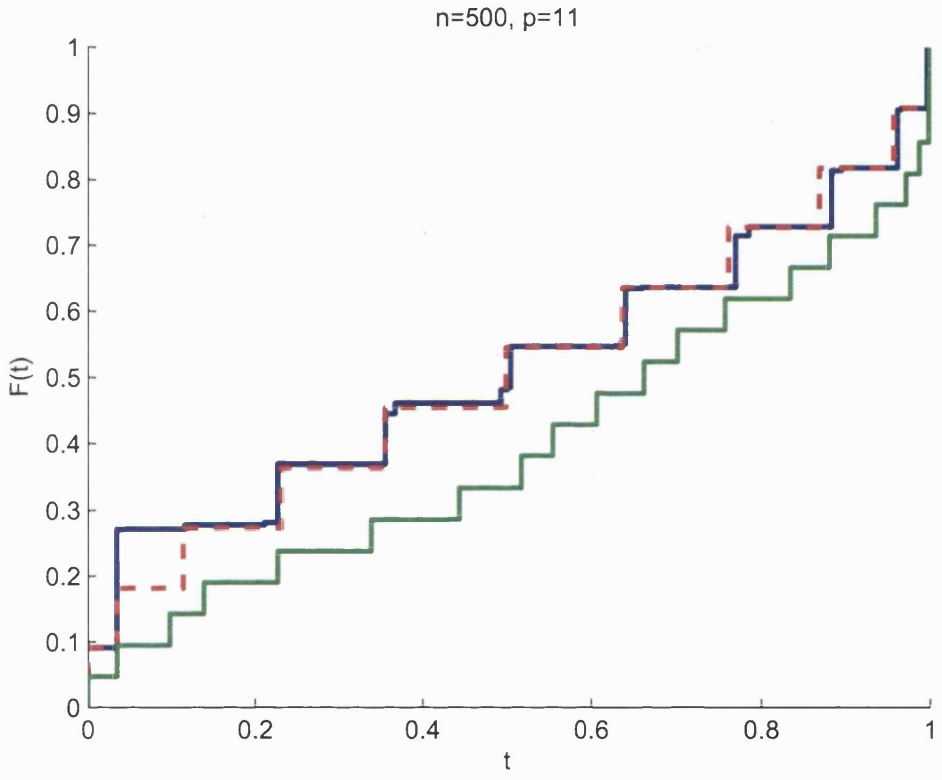

Figure 3.10: Cumulative distribution function for optimal design points of the kernel $\exp[-(t-s)^2]$ using the DETMAX algorithm when $n = 500$ and $p = 11$ (blue solid line), the greedy exchange algorithm when $n = 21$ (green solid line), and the greedy exchange algorithm when $n = 11$ (red dash line).

Chapter 4

Prediction and Mean Squared Error

Assume that there is a set of the sampling points (design) for the centered Gaussian process $D = \{t_i, y(t_i), t_i \in \mathcal{T}, 1 \leq i \leq n\}$, where $y(t_i)$ are assumed to be sampled from a "true" stochastic process, i.e. a process with an infinite order of the Karhunen-Loeve expansion. Then based on the sampling points, the value of y at $\{t | t \in \mathcal{T}, t \neq t_i, 1 \leq i \leq n\}$ can be predicted at any other point in the interval $\mathcal{T}$. In this chapter, the prediction for $y(t)$ is calculated from the conditional expectation based on the sampling points. This can be interpreted as the Bayes estimator under quadratic loss. Then the truncated Karhunen-Loeve expansion is used for the approximation of the covariance function. Section 1 provides a brief introduction to the calculation of the conditional expectation from the perspective of the Bayesian analysis and the perspective of the functional analysis. Section 2 calculates the mean squared error (MSE) for the prediction derived from the truncated Karhunen-Loeve expansion and the generalised MSE. Section 3 concentrates on solving the inverse problem for the truncation version of the covariance matrix and puts forward two possible solutions. Section 4 calculates the generalised MSE numerically using the

Haar wavelet method, described in chapter 2. Section 5 provides an alternative way of modelling the conditional data using the conditional Karhunen-Loeve expansion.

4.1 Prediction

Given the sampling points $y_s = \{y(t_i), t_i \in \mathcal{T}, 1 \leq i \leq n\}$, we try to predict $y(t)$ at any other point $t \neq t_i$. One way of computing this is through the Bayesian analysis. The distribution for y_s, before sampling, can be regarded as a prior distribution. After being given a new time point t, we can update the distribution for $y(t)$ and obtain its posterior distribution. The other way of predicting $y(t)$ is to derive a minimum norm interpolant through simple kriging. Refer to remark 4.1.2 for a brief on kriging. We assume throughout that all the covariances are known, and where necessary, distributions are the univariate or multivariate normal.

Assume that K_n is the covariance matrix for n sampling points, i.e. $(K_n)_{i,j} = \text{cov}(y(t_i), y(t_j))$. With a new time point t, a new covariance matrix K_{n+1} can be written as

$$K_{n+1} = \begin{pmatrix} K_n & k \\ k & \kappa \end{pmatrix} , \tag{4.1.1}$$

where $(k^T) = \begin{pmatrix} \text{cov}(y(t), y(t_1)) & \cdots & \text{cov}(y(t), y(t_n)) \end{pmatrix}$ and $\kappa = \text{cov}(y(t), y(t))$.

Using the partitioned inverse equation (see, for example, Press et al. (1992)), K_{n+1}^{-1} can be expressed in terms of K_n.

$$K_{n+1}^{-1} = \begin{pmatrix} M & m \\ m^T & u \end{pmatrix} , \tag{4.1.2}$$

where

$$\begin{aligned} u &= (\kappa - k^T K_n^{-1} k)^{-1} \\ m &= -u K_n^{-1} k \\ M &= K_n^{-1} + \frac{1}{u} m m^T \end{aligned}$$

Then since $y(t)$ and y_s follow the normal distribution, the posterior distribution for $y(t)$ given the sampling points y_s can be calculated as

$$\begin{aligned} p(y(t)|y_s) &\propto \exp[-\frac{1}{2}\begin{pmatrix} y_s^T, & y(t) \end{pmatrix} K_{n+1}^{-1} \begin{pmatrix} y_s \\ y(t) \end{pmatrix}] \\ &\propto \exp[-\frac{1}{2}\begin{pmatrix} y_s^T, & y(t) \end{pmatrix} \begin{pmatrix} My_s + y(t) \\ m^T y_s + uy(t) \end{pmatrix}] \\ &\propto \exp[-\frac{1}{2}(uy(t)^2 + 2y_s^T m y(t) + y_s^T M y_s)] \\ &\propto \exp[-\frac{u}{2}(y(t)^2 + \frac{2}{u} y_s^T m y(t))] \end{aligned}$$

Hence the posterior mean and the posterior variance are as follows

$$E(y(t)|y_s) = -\frac{1}{u} y_s^T m = (y_s^T u K_n^{-1} k)\frac{1}{u} = k^T K_n^{-1} y_s \tag{4.1.3}$$

$$\mathrm{Var}(y(t)|y_s) = \frac{1}{u} = \kappa - k^T K_n^{-1} k \tag{4.1.4}$$

Denote $\hat{y}(t) = E(y(t)|y_s)$, i.e. $\hat{y}(t)$ is the posterior mean, since the posterior mean minimises the posterior mean squared error.

In fact, from the perspective of simple kriging, together with the functional analysis, $\hat{y}(t)$ can also be regarded as the minimum norm interpolant of $y(t)$ in H_Y onto span(y_s). Under this setting, a set of coefficients $\{a_i\}$ needs to be found, so that the following can be minimised

$$||y(t) - \sum_i a_i y(t_i)||^2_{H_Y} \tag{4.1.5}$$

Since the norm in H_Y and RKHS are isometrically isomorphic, it is equivalent to minimise $||K(t,.) - \sum_i a_i K(t_i,.)||^2_{RKHS}$. Using the properties of RKHS (see, section 1.1, chapter 1, for detail) and using the notation $a^T = \Big(a_1, \ a_2, \ \cdots, \ a_n\Big)$, equation 4.1.5 can be further simplified as

$$\begin{aligned} & < K(t,.), K(t,.) > -2 < \sum_i a_i K(t_i,.), K(t,.) > + < \sum_i a_i K(t_i,.) > \\ = \quad & K(t,t) - 2\sum_i a_i K(t_i,t) + \sum_i \sum_j a_i a_j K(t_i,t_j) \\ = \quad & \kappa - 2a^T k + a^T K_n a \end{aligned}$$

The last equation can be minimised through the differentiation with respect to the vector a. It results in $a = K_n^{-1}k$. Hence $\hat{y}(t) = k^T K_n^{-1} y_s$ as in equation 4.1.3, which is an orthogonal projection.

Using a truncated Karhunen-Loeve expansion of order p, and with notation

$$\begin{aligned} \Phi(t)^T &= \Big(\phi_1(t), \ \phi_2(t), \ \cdots, \ \phi_p(t)\Big) \\ \Phi &= \Big(\Phi(t_1), \ \Phi(t_2), \ \cdots, \Phi(t_n)\Big) \\ \Lambda &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \end{aligned}$$

The prediction $\hat{y}(t)$ defined as the conditional expectation $E(y(t)|y_s)$ can be expressed as

$$\hat{y}(t) = \Phi(t)^T \Lambda \Phi (\Phi^T \Lambda \Phi)^{-1} y_s \tag{4.1.6}$$

Remark 4.1.1. we shall use the term "prediction", although when the time point is in the region of interest, it can be considered as interpolation or smoothing.

Remark 4.1.2. In general, the statistical technique on interpolating the value $y(t)$, using the minimum norm interpolant, from the prior observations $y_s = \{y(t_i), t_i \in \mathcal{T}, 1 \leq i \leq n\}$, where $t_i \neq t$, is often called kriging. The kriging model for $y(t_i)$ is usually expressed as

$$y(t_i) = \mu(t_i) + \xi(t_i) \tag{4.1.7}$$

so that $E(y(t_i)) = \mu(t_i)$ and $\text{cov}[y(t_i), y(t_j)] = \text{cov}[\xi(t_i), \xi(t_j)]$. In this chapter, $\mu(t_i) = 0$. This is called a simple kriging. Under simple kriging, the minimum norm interpolant is expressed in equation 4.1.3. In practice, there are more complicated kriging models, such as "ordinary" kriging, which assumes $\mu(t_i)$ to be a constant, or "universal" kriging, which assumes a general linear trend model on $\mu(t_i)$, i.e. $\mu(t_i) = \sum_{k=1}^{p} \beta_k f_k(t_i)$, where $f_k(t)$ can be treated as a basis function. For details on kriging, see, for example, Sacks et al. (1989), Cressie (1993) and Wahba (1990). One can apply the current methods by making the β_k random and incorporating these random effect into the covariance but we omit these calculations.

4.2 Mean squared error (MSE)

In this section, the sampling points are assumed to be from what is called the "full" model, which is a true model of the process expressed by the infinite sum of the Karhunen-Loeve expansion. At the same time, the model for the prediction of $y(t)$ is assumed to be from what is called the "reduced" model, where only p terms are used for the approximation of the covariance function. The reduced model is also called the truncated model in the previous analysis. The main task of this section is to calculate the mean squared error and the generalised mean squared error for the

prediction *under the reduced model when the true model is the full model.* Analytical expression will be derived. It can also be shown that under certain conditions, the generalised mean squared error decreases when the order p increases in the Karhunen-Loeve expansion.

To make the notation clear, since the number of testing points is always assumed to be n, we omit the subscript n from the covariance matrix K. Furthermore, since both the full and the reduced version of the covariance matrix/vector will be involved in the calculation, we distinguish them by using the subscript f. Thus K, k and K_f, k_f represent the covariance matrix/vector from the reduced model and from the full model, respectively.

Under the reduced model, $\hat{y}(t) = a^T y_s, a^T = k^T K$ as explained in equation 4.1.6. Under the full model, using the same calculation, the prediction at time point t can be written down as $\hat{\hat{y}}(t) = b^T y_s$, where

$$
\begin{aligned}
b^T &= k_f^T K_f^{-1} \\
k_f &= \left(\textstyle\sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t)\phi_i(t_1),\ \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t)\phi_i(t_2),\ \cdots,\ \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t)\phi_i(t_n)\right) \\
K_f &= \begin{pmatrix}
\sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_1)\phi_i(t_1) & \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_1)\phi_i(t_2) & \cdots & \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_1)\phi_i(t_n) \\
\sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_2)\phi_i(t_1) & \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_2)\phi_i(t_2) & \cdots & \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_2)\phi_i(t_n) \\
 & \cdots & & \\
\sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_n)\phi_i(t_1) & \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_n)\phi_i(t_2) & \cdots & \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t_n)\phi_i(t_n)
\end{pmatrix}
\end{aligned}
$$

Both $\hat{y}(t)$ and $\hat{\hat{y}}(t)$ are the linear combination of the centered Gaussian process y_s. Hence

$$E(y(t)) = E(\hat{y}(t)) = E(\hat{\hat{y}}(t)) = 0 \tag{4.2.1}$$

Now the mean squared error for the reduced model can be calculated as

$$
\begin{aligned}
& \mathrm{MSE}(\hat{y}(t)) \\
= \; & E[(\hat{y}(t) - y(t))^2] = E[(\hat{y}(t) - E(\hat{y}(t)) + E(\hat{y}(t)) - y(t))^2] \\
= \; & E[(\hat{y}(t) - E(\hat{y}(t)) + E(y(t)) - y(t))^2] = \mathrm{Var}(\hat{y}((t))) + \mathrm{Var}(y(t)) - 2\mathrm{cov}(\hat{y}(t), y(t)) \\
= \; & \kappa + \mathrm{Var}(a^T y_S) - 2\mathrm{cov}(a^T y_S, y(t)) = \kappa + a^T K_f a - 2a^T k_f \\
= \; & \kappa + k^T K^{-1} K_f K^{-1} k - 2k^T K^{-1} k_f
\end{aligned}
$$

Remark 4.2.1. The mean squared error can also be calculated from the MSE conditioning on the sampling points (conditional MSE). The conditional MSE can be decomposed as the conditional variance and the square of the conditional bias, since

$$
\begin{aligned}
\mathrm{MSE}(\hat{y}(t)|y_s) &= E[(\hat{y}(t) - y(t))^2|y_s] = \mathrm{Var}((\hat{y}(t) - y(t))|y_s) + \{E[\hat{y}(t) - y(t)|y_s]\}^2 \\
&= \mathrm{Var}(y(t)|y_s) + \{E[\hat{y}(t)|y_s] - E[y(t)|y_s]\}^2 \\
&= \mathrm{Var}(y(t)|y_s) + [\hat{y}(t) - \hat{\hat{y}}(t)]^2
\end{aligned}
$$

Then the unconditional MSE is the expectation of the conditional MSE.

$$
\begin{aligned}
& E[\mathrm{MSE}(\hat{y}(t)|y_s)] \\
= \; & E[\mathrm{Var}(y(t)|y_s)] + E[\hat{y}(t) - \hat{\hat{y}}(t)]^2 = E[\mathrm{Var}(y(t)|y_s)] + E[\hat{y}(t)^2 + \hat{\hat{y}}(t)^2 - 2\hat{y}(t)\hat{\hat{y}}(t)] \\
= \; & \mathrm{Var}(y(t)) - \mathrm{Var}[E(y(t)|y_s)] + \mathrm{Var}(\hat{y}(t)) - (E[\hat{y}(t)])^2 + \mathrm{Var}(\hat{\hat{y}}(t)) - (E[\hat{\hat{y}}(t)])^2 \\
& -2\mathrm{cov}(\hat{y}(t), \hat{\hat{y}}(t)) + 2E[\hat{y}(t)]E[\hat{\hat{y}}(t)] \\
= \; & \mathrm{Var}(y(t)) - \mathrm{Var}[E(y(t)|y_s)] + \mathrm{Var}(\hat{y}(t)) + \mathrm{Var}(\hat{\hat{y}}(t)) - 2\mathrm{cov}(\hat{y}(t), \hat{\hat{y}}(t)) \\
= \; & \mathrm{Var}(y(t)) - \mathrm{Var}(\hat{\hat{y}}(t)) + \mathrm{Var}(\hat{y}(t)) + \mathrm{Var}(\hat{\hat{y}}(t)) - 2\mathrm{cov}(\hat{y}(t), \hat{\hat{y}}(t)) \\
= \; & \mathrm{Var}(y(t)) + \mathrm{Var}(\hat{y}(t)) - 2\mathrm{cov}(\hat{y}(t), \hat{\hat{y}}(t)) = \kappa + \mathrm{Var}(a^T y_s) - 2\mathrm{cov}(a^T y_S, b^T y_s) \\
= \; & \kappa + a^T K_f a - 2a^T K_f b = \kappa + k^T K^{-1} K_f K^{-1} k - 2k^T K^{-1} K_f K_f^{-1} k_f \\
= \; & \kappa + k^T K^{-1} K_f K^{-1} k - 2k^T K^{-1} k_f
\end{aligned}
$$

Notice that the expectation of the conditional variance $E[\text{Var}(y(t)|y_s)]$ is expressed as $E[\text{Var}(y(t)|y_s)] = \text{Var}(y(t)) - \text{Var}(\hat{\hat{y}}(t))$, which does not depend on the truncation at all. It is the expectation of the conditional bias that plays a role in affecting the MSE when the order p in the truncated Karhunen-Loeve expansion is changed.

Remark 4.2.2. Using the same method, the MSE for $\hat{\hat{y}}(t)$ can also be calculated, i.e. the mean squared error for the full model when the true model is the full model. Notice that under this setting, the unconditional MSE and the conditional MSE are the same.

$$
\begin{aligned}
& \text{MSE}(\hat{\hat{y}}(t)) = E[(\hat{\hat{y}}(t) - y(t))^2] \\
= \; & \text{Var}(\hat{\hat{y}}((t))) + \text{Var}(y(t)) - 2\text{cov}(\hat{\hat{y}}(t), y(t)) \\
= \; & \kappa + \text{Var}(b^T y_S) - 2\text{cov}(b^T y_S, y(t)) \\
= \; & \kappa + b^T K_f b - 2b^T k_f \\
= \; & \kappa + k_f^T K_f^{-1} K_f K_f^{-1} k_f - 2k_f^T K_f^{-1} k_f \\
= \; & \kappa - k_f^T K_f^{-1} k_f \\
& E[\text{MSE}(\hat{\hat{y}}(t)|y_S)] = \text{Var}(y(t)) - \text{Var}(\hat{\hat{y}}(t)) \\
= \; & \kappa - \text{Var}(b^T y_S) \\
= \; & \kappa - b^T K_f b \\
= \; & \kappa - k_f^T K_f^{-1} k_f
\end{aligned}
$$

In the experimental design literature, e.g. Sacks and Ylvisaker (1978), Muller-Gronbach (1996) and Mukherjee (2006), and machine learning literature, e.g. Opper and Vivarelli (1999), Sollich (1999) and Rasmussen and Williams (2005), are interested in the generalised mean squared error, which is defined as the integral of the

MSE of the prediction with respect to t. For the experimental design, the optimal design points can be chosen so that the generalized mean squared error is minimised.

The generalized MSE for prediction using the truncated Karhunen-Loeve expansion of order p can now be calculated. We will first give some formula for the integration of some quadratic forms. With notation

$$\begin{aligned}
\phi_u^T &= \Big(\phi_u(t_1),\ \phi_u(t_2),\ \cdots,\ \phi_u(t_n)\Big) \\
\Phi(t)^T &= \Big(\phi_1(t),\ \phi_2(t),\ \cdots,\ \phi_p(t)\Big) \\
\Phi &= \Big(\Phi(t_1),\ \Phi(t_2),\ \cdots,\ \Phi(t_n)\Big) \\
k &= \Phi^T\Lambda\Phi(t) \\
B &= \{b_{ij}\} = \Lambda\Phi A\Phi^T\Lambda
\end{aligned}$$

A quadratic form based on a fixed matrix $A = \{a_{ij}\}$ can be calculated as

$$\begin{aligned}
&\int_{\mathcal{T}} k^T A k dt = \int_{\mathcal{T}} \Phi(t)^T\Lambda\Phi A\Phi^T\Lambda\Phi(t)dt = \int_{\mathcal{T}} \sum_{i\in S}\sum_{j\in S} b_{ij}\phi_i(t)\phi_j(t)dt \\
= &\sum_{i\in S}\sum_{j\in S} b_{ij}\int_{\mathcal{T}} \phi_i(t)\phi_j(t)dt = \sum_{i\in S} b_{ii} = \text{trace}(B) \\
= &\ \text{trace}(\Lambda\Phi A\Phi^T\Lambda)
\end{aligned}$$

Another quadratic form based on the fixed matrix A can be expressed as

$$\begin{aligned}
&\int_{\mathcal{T}} k^T A k_f dt \\
= &\int_{\mathcal{T}} \sum_{i\in S}\sum_{j\in S} a_{ij}k_i k_{f,j}dt = \sum_{i\in s}\sum_{j\in s} a_{ij}\sum_{u=1}^{p}\lambda_u\phi_u(t_i)\int_{\mathcal{T}}\phi_u(t)\text{cov}(y(t),y(t_j))dt \\
= &\sum_{i\in s}\sum_{j\in s} a_{ij}\sum_{u=1}^{p}\lambda_u\phi_u(t_i)\lambda_u\phi_u(t_j) = \sum_{i\in s}\sum_{j\in s} a_{ij}\sum_{u=1}^{p}\lambda_u^2\phi_u(t_i)\phi_u(t_j) \\
= &\sum_{u=1}^{p}\lambda_u^2\sum_{i\in s}\sum_{j\in s} a_{ij}\phi_u(t_i)\phi_u(t_j) = \sum_{u=1}^{p}\lambda_u^2\phi_u^T A\phi_u = \sum_{u=1}^{p}(\lambda_u\phi_u)^T A(\lambda_u\phi_u) \\
= &\ \text{trace}(\Lambda\Phi A(\Lambda\Phi)^T)
\end{aligned}$$

For the integration involving the variance κ, we obtain

$$\int_{\mathcal{T}} \kappa dt = \int_{\mathcal{T}} \sum_{u=1}^{\infty} \lambda_u \phi_u^2(t) dt = \sum_{u=1}^{\infty} \lambda_u \tag{4.2.2}$$

Therefore, the generalised MSE, using the truncated Karhunen-Loeve expansion, can now be expressed as

$$\begin{aligned} \int_{\mathcal{T}} \mathrm{MSE}(\hat{y}(t)) dt &= \sum_{i=1}^{\infty} \lambda_i - \mathrm{trace}(\Lambda\Phi(2K^{-1} - K^{-1}K_f K^{-1})\Phi^T\Lambda) \\ K &= \Phi^T\Lambda\Phi \end{aligned}$$

As is mentioned in remark 4.2.1, the MSE can be decomposed into two parts: the expectation of the conditional variance and the square of the expectation of the conditional bias. Since the expectation of the conditional variance does not depend on the order p in the truncated Karhunen-Loeve expansion, the change in the MSE due to the different order p is caused by the change in the expectation of the conditional bias. After integrating the MSE to derive the generalised MSE, the change in the generalised MSE for the different order p should also be caused by the change in the conditional bias. Therefore it might be expected that when the order p increases, the bias decreases, and therefore so does the generalised MSE. The conditions for this to hold are not obvious and the rest of this section is devoted to finding a tractable version. To make the notation clear, the truncated covariance matrix/vector of order p will be denoted as K_p, k_p, while the true covariance matrix/vector is still denoted as K_f, k_f.

Denote the generalised MSE for the order p as $\int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t)) dt$ and

$$\Phi_{p+1} = \begin{pmatrix} \Phi_p \\ \phi_{p+1}^T \end{pmatrix} \qquad \Lambda_{p+1} = \begin{pmatrix} \Lambda_p & 0 \\ 0 & \lambda_{p+1} \end{pmatrix} \qquad \Sigma_p = K_f - K_p \tag{4.2.3}$$

where Φ_p is a $p \times n$ matrix comprising of the first p eigenfunctions, and ϕ_{p+1} is a $n \times 1$ column vector comprising of the $(p+1)$th eigenfunction. For the simplicity of theorem 4.2.1, the following notations are further defined.

$$\begin{aligned} A &= \phi_{p+1}^T K_p^{-1} \Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1} \\ C &= \phi_{p+1}^T K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1} \\ b &= \phi_{p+1}^T K_p^{-1} \phi_{p+1} \\ d &= \phi_{p+1}^T K_p^{-1} \Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1} \phi_{p+1} \\ S &= 2Ab - Cd - C + b - bd \\ \Delta &= S^2 - 16Ab^2 \end{aligned}$$

Then, following the calculation in appendix 7.2, the difference of the generalised MSE between the order p and the order $p+1$ is,

$$\begin{aligned} & \int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t))dt - \int_{\mathcal{T}} \mathrm{MSE}_{p+1}(\hat{y}(t)) \\ = \quad & \frac{\lambda_{p+1}}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2}[\lambda_{p+1}^2(2b^2) + \lambda_{p+1}S + 2A] \end{aligned}$$

Since b is a quadratic form of a positive definite matrix K_p^{-1}, $b > 0$ if ϕ_{p+1} is a non-zero vector. Except for the constant term $\frac{\lambda_{p+1}}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2}$, which is positive, the remaining term is a quadratic form of λ_{p+1}, when $b > 0$.

It is expected that the difference of the generalised MSE between the order p and the order $p+1$ is non-negative, so that the generalised MSE is a non-increasing function of the order p. The conditions required for the above statement is summarised in theorem 4.2.1.

Theorem 4.2.1. *If one of the following statements is true, the change of the generalised MSE from the order p to the order $p+1$ is non-negative, i.e.*

$$\int_{\mathcal{T}} MSE_p(\hat{y}(t))dt - \int_{\mathcal{T}} MSE_{p+1}(\hat{y}(t))dt \geq 0 \tag{4.2.4}$$

(i) If $A \geq 0$, $\Delta \leq 0$, $S \geq 0$, or

(ii) If $A \geq 0$, $\Delta \leq 0$, $S < 0$, or

(iii) If $A \geq 0$, $\Delta > 0$, $S \geq 0$, or

(iv) If $A \geq 0$, $\Delta > 0$, $S < 0$,$\lambda_{p+1} \leq x_1$, or

(v) If $A \geq 0$, $\Delta > 0$, $S < 0$,$\lambda_{p+1} \geq x_2$, or

(vi) If $A < 0$, $\lambda_{p+1} \geq x_2$

where x_1 and x_2 are the roots for the quadratic equation $(2b^2)x^2+Sx+2A$ and $x_1 \leq x_2$.

Proof. Since all the eigenvalues of the Fredholm integral equation are positive, as is mentioned in appendix 7.1, $\lambda_{p+1} > 0$, and therefore $\frac{\lambda_{p+1}}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} > 0$. Notice that $b = 0$ if and only if ϕ_{p+1} is a zero vector, which also implies that $A = C = d = S = 0$. Hence

$$\int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t))dt - \int_{\mathcal{T}} \mathrm{MSE}_{p+1}(\hat{y}(t))dt = 0 \tag{4.2.5}$$

When $b > 0$, it is enough to show that when $x \geq 0$, the quadratic form $(2b^2)x^2 + Sx + 2A \geq 0$. Since $2b^2 > 0$, the quadratic form is a convex parabola. Geometrically, when $\Delta \leq 0$ as stated in the condition (i) and (ii), there is no intersection between the parabola and the horizontal axis. This implies that the quadratic form is always non-negative. Using the properties of the convex parabola, the other conditions are shown similarly. □

Remark 4.2.3. When ϕ_{p+1} is a zero vector, $b = 0$. It means that there is no increase or decrease to the generalised MSE when the order p changes to the order $p+1$. The

rest of this section contributes mainly to the other case when $b > 0$, i.e. not all the elements of ϕ_{p+1} is zero, since this is the more confusing case and worths more effort. Therefore, in the rest of the section, unless specified, we always assume that $b > 0$.

Remark 4.2.4. The above conditions are written down as the disjoint sets. The first three conditions can also be expressed as $\Delta \leq 0$ (implying $A \geq 0$) and $A \geq 0$, $S \geq 0$, which have some overlapping.

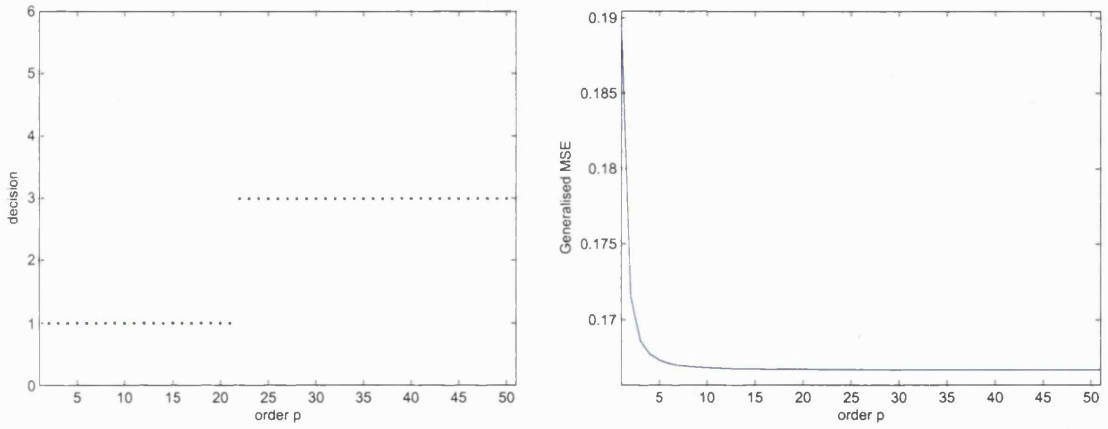

Figure 4.1: (Left) Reason checking for the difference in the generalised MSE for the Brownian motion due to the change of the order from p to $p+1$, when $n = 1$, $t = \frac{1}{2}$ and $1 \leq p \leq 51$. 1: $A \geq 0, \Delta \leq 0, S \geq 0$; 2: $A \geq 0, \Delta \leq 0, S < 0$; 3: $A \geq 0, \Delta > 0, S \geq 0$; 4: $A \geq 0, \Delta > 0, S < 0, \lambda_{p+1} \leq x_1$; 5: $A \geq 0, \Delta > 0, S < 0, \lambda_{p+1} \geq x_2$; 6: $A < 0, \lambda_{p+1} \geq x_2$. (Right) the generalised MSE for the posterior prediction when $n = 1$, $t = \frac{1}{2}$ and $1 \leq p \leq 51$.

From figure 4.1, figure 4.2 and figure 4.3, it can be seen that in the Brownian motion on $[0, 1]$, it is the first three conditions that decides the change of the generalised MSE when the order in the truncated Karhunen-Loeve expansion changes. When there is only one condition ($n = 1$), the change in the generalised MSE is mainly due to $S \geq 0$. When the number of the prior observations increases to $n = 30$ or even $n = 100$, the reason for the change moves to $\Delta \leq 0$.

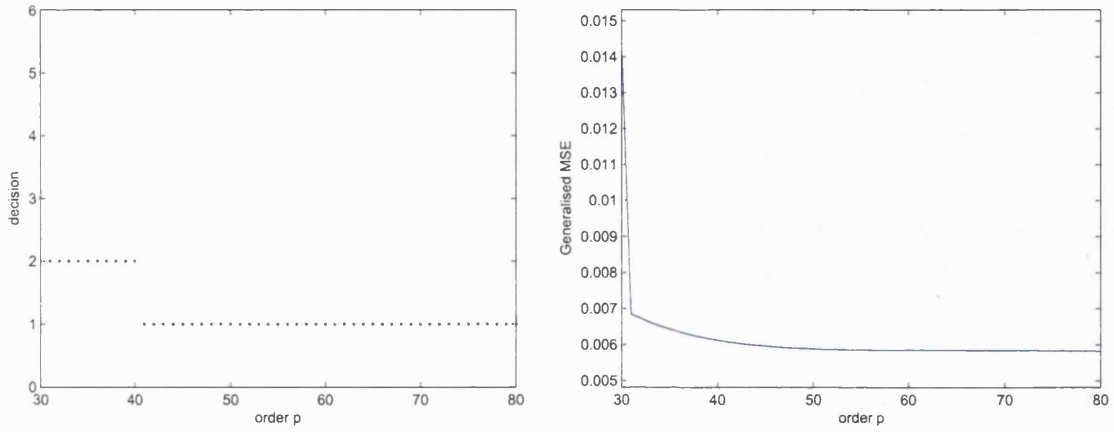

Figure 4.2: (Left) Reason checking for the difference in the generalised MSE for the Brownian motion due to the change of the order from p to $p+1$, when $n = 30$, $t_i = \frac{i}{31}, 1 \leq i \leq 30$ and $30 \leq p \leq 80$. 1: $A \geq 0, \Delta \leq 0, S \geq 0$; 2: $A \geq 0, \Delta \leq 0, S < 0$; 3: $A \geq 0, \Delta > 0, S \geq 0$; 4: $A \geq 0, \Delta > 0, S < 0, \lambda_{p+1} \leq x_1$; 5: $A \geq 0, \Delta > 0, S < 0, \lambda_{p+1} \geq x_2$; 6: $A < 0, \lambda_{p+1} \geq x_2$. (Right) the generalised MSE for the posterior prediction when $n = 30$, $t_i = \frac{i}{31}, 1 \leq i \leq 30$ and $30 \leq p \leq 80$.

However, theorem 4.2.1 is somewhat difficult to use in application, since there are six conditions in total to check whether the generalised MSE increases or not. Theorem 4.2.2 provides a simplified version of theorem 4.2.1. If two additional constraints are placed on $\frac{A}{C}$ and d, the generalised MSE does not increase when p increases.

Theorem 4.2.2. *If $\frac{A}{C} \geq \lambda_{p+1}$ and $d \leq 1$, the change of the generalised MSE from the order p to the order $p+1$ is non-negative, i.e.*

$$\int_{\mathcal{T}} MSE_p(\hat{y}(t))dt - \int_{\mathcal{T}} MSE_{p+1}(\hat{y}(t))dt \geq 0 \tag{4.2.6}$$

Proof. Since $K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}$ is non-negative definite, $d \geq 0$. If d further assumes to be $d \leq 1$,

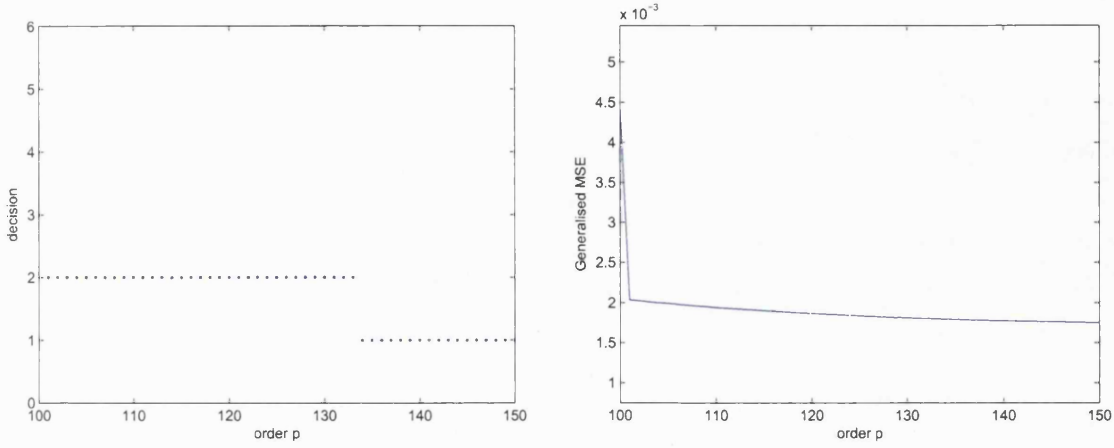

Figure 4.3: (Left) Reason checking for the difference in the generalised MSE for the Brownian motion due to the change of the order from p to $p+1$, when $n = 100$, $t_i = \frac{i}{101}, 1 \leq i \leq 100$ and $100 \leq p \leq 150$. 1: $A \geq 0, \Delta \leq 0, S \geq 0$; 2: $A \geq 0, \Delta \leq 0, S < 0$; 3: $A \geq 0, \Delta > 0, S \geq 0$; 4: $A \geq 0, \Delta > 0, S < 0, \lambda_{p+1} \leq x_1$; 5: $A \geq 0, \Delta > 0, S < 0, \lambda_{p+1} \geq x_2$; 6: $A < 0, \lambda_{p+1} \geq x_2$. (Right) the generalised MSE for the posterior prediction when $n = 100$, $t_i = \frac{i}{101}, 1 \leq i \leq 100$ and $100 \leq p \leq 150$.

$$
\begin{aligned}
& \int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t))dt - \int_{\mathcal{T}} \mathrm{MSE}_{p+1}(\hat{y}(t))dt \\
\geq\ & \frac{1}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} \\
& [2\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& +2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& -2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1})] \\
& +[\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1}) + 2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 - \lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})] \\
\geq\ & \frac{1}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} \\
& [2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& +[2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2] \geq 0
\end{aligned}
$$

The second inequality uses the condition that $\frac{A}{C} \geq \lambda_{p+1}$. Since $C > 0$ and $\lambda_{p+1} > 0$, $A > 0$. □

Remark 4.2.5. In theorem 4.2.2, the condition $\frac{A}{C} \geq \lambda_{p+1}$ can be explained as follows,

$$\begin{aligned} & A = \phi_{p+1}^T K_p^{-1} \Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1} \\ = \quad & \lambda_{p+1} (\phi_{p+1}^T K_p^{-1} \Phi_p^T \frac{\Lambda_p}{\lambda_{p+1}} \Lambda_p \Phi_p K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1}) \end{aligned}$$

Since, $\lambda_i > \lambda_{p+1}$, $1 \leq i \leq p$, $\frac{\Lambda_p}{\lambda_{p+1}} > I$. If $\Phi_p^T (\frac{\Lambda_p}{\lambda_{p+1}} \Lambda_p - \Lambda_p) \Phi_p$ is positive definite, and its multiplication with the other two positive definite matrix K_p^{-1} and $K_p^{-1} \Sigma_p K_p^{-1}$ is still positive definite,

$$\begin{aligned} & \lambda_{p+1} (\phi_{p+1}^T K_p^{-1} \Phi_p^T \frac{\Lambda_p}{\lambda_{p+1}} \Lambda_p \Phi_p K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1}) \\ \geq \quad & \lambda_{p+1} (\phi_{p+1}^T K_p^{-1} \Phi_p^T \Lambda_p \Phi_p K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1}) \\ = \quad & \lambda_{p+1} (\phi_{p+1}^T K_p^{-1} \Sigma_p K_p^{-1} \phi_{p+1}) = \lambda_{p+1} C \end{aligned}$$

When $n = 1$, all the key terms in the difference of the generalised MSE are the scalars, including A, C, b, d, S and Δ. Theorem 4.2.2 can be further simplified to equation 4.2.7 exhibited in theorem 4.2.3.

Theorem 4.2.3. *When the number of the design points $n = 1$, if*

$$\phi_{p+1}(t)^2 \sum_{i=1}^{p} \lambda_i^2 \phi_i(t)^2 \leq [\sum_{i=1}^{p} \lambda_i \phi_i(t)^2]^2 \tag{4.2.7}$$

the generalised MSE expressed by the truncated Karhunen-Loeve expansion is a non-increasing function of p, i.e.

$$\int_{\mathcal{T}} MSE_p(\hat{y}(t))dt - \int_{\mathcal{T}} MSE_{p+1}(\hat{y}(t))dt \geq 0, \ \ p \geq 1 \tag{4.2.8}$$

Proof. When $n = 1$, i.e. we only have one prior observation at time t. Using theorem 4.2.2, if it can be shown that $d \leq 1$ and $\frac{A}{C} \geq \lambda_{p+1}$, the generalised MSE is a non-increasing function of p.

$d \leq 1$ can be simplified as

$$\frac{\phi_{p+1}(t)^2 \sum_{i=1}^{p} \lambda_i^2 \phi_i(t)^2}{[\sum_{i=1}^{p} \lambda_i \phi_i(t)^2]^2} \leq 1 \tag{4.2.9}$$

It is equivalent to

$$\phi_{p+1}(t)^2 \sum_{i=1}^{p} \lambda_i^2 \phi_i(t)^2 \leq [\sum_{i=1}^{p} \lambda_i \phi_i(t)^2]^2 \tag{4.2.10}$$

As to $\frac{A}{C} \geq \lambda_{p+1}$, it always holds for $n = 1$, since both K_p^{-1} and $K_p^{-1}\Sigma_p K_p^{-1}$ are the positive scalars. Then $\frac{\sum_{i=1}^{p} \lambda_i^2 \phi_i(t)^2}{\lambda_{p+1}} > \sum_{i=1}^{p} \lambda_i \phi_i(t)^2$, because $\lambda_{p+1} < \lambda_i$, $1 \leq i \leq p$. □

The upper bound of the generalised MSE can also be calculated, when its expression using the truncated Karhunen-Loeve expansion is a decreasing function of the order p. The upper bound is obtained when the order p equals to the number of the design points n. The result is being summarised in theorem 4.2.4.

Theorem 4.2.4. *When the generalised MSE expressed by the truncated Karhunen-Loeve expansion is a decreasing function of the order p, i.e.*

$$\int_{\mathcal{T}} MSE_p(\hat{y}(t))dt - \int_{\mathcal{T}} MSE_{p+1}(\hat{y}(t))dt \geq 0, \;\; p \geq n \;\; , \tag{4.2.11}$$

where n represents the number of the design points, there exists an upper bound for the generalised MSE.

$$\int_{\mathcal{T}} MSE_p(\hat{y}(t))dt \leq \int_{\mathcal{T}} MSE_n(\hat{y}(t))dt = \sum_{i=p+1}^{\infty} \lambda_i + trace[\Sigma_p(\Phi_p^T \Phi_p)^{-1}], \;\; p \geq n \;\; , \tag{4.2.12}$$

where $\Sigma_p = K_f - K_p$ *and* Φ_p *is a* $p \times n$ *matrix comprising of the first p eigenfunctions.*

Proof. The condition $p \geq n$ guarantees the existence of the inverse of the covariance matrix using the truncated Karhunen-Loeve expansion. With fixed n, when $\int_{\mathcal{T}} \mathrm{MSE}(\hat{y}(t))dt$ is a decreasing function of p ($p \geq n$), an upper bound for the generalised MSE can be obtained. The upper bound is obtained when the order of the Karhunen-Loeve expansion p achieves its minimum possible value, which is the number of the prior observations n, i.e.

$$\int_{\mathcal{T}} \mathrm{MSE}_n(\hat{y}(t))dt \tag{4.2.13}$$

When $n = p$, the generalized MSE can be further simplified, since the inverse of K can be calculated matrix by matrix now, i.e.

$$K_p = \Phi_p^T \Lambda_p \Phi_p \,, \quad K_p^{-1} = (\Phi_p)^{-1}(\Lambda_p)^{-1}(\Phi_p^T)^{-1}$$

Therefore, the integration of a quadratic form based on the symmetric fixed matrix K^{-1} can be calculated as

$$\begin{aligned} & \int_{\mathcal{T}} k_p^T K_p^{-1} k_f dt \\ = \; & \mathrm{trace}[\Lambda_p \Phi_p K_p^{-1} (\Lambda_p \Phi_p)^T] = \mathrm{trace}[\Lambda_p \Phi_p (\Phi_p)^{-1}(\Lambda_p)^{-1}(\Phi_p^T)^{-1}(\Phi_p)^T \Lambda_p^T] \\ = \; & \mathrm{trace}(\Lambda_p) \end{aligned}$$

Meanwhile, when $n = p$, using the same expression for K_p^{-1}, the integration of the quadratic form based on the fixed matrix $K_p^{-1} K_f K_p^{-1}$ can be calculated as

$$\begin{aligned} & \int_{\mathcal{T}} k_p^T K_p^{-1} K_f K_p^{-1} k_p dt \\ = \; & \mathrm{trace}[\Lambda_p \Phi_p K_p^{-1} K_f K_p^{-1} (\Lambda_p \Phi_p)^T] \\ = \; & \mathrm{trace}[\Lambda_p \Phi_p (\Phi_p)^{-1}(\Lambda_p)^{-1}(\Phi_p^T)^{-1} K_f (\Phi_p)^{-1}(\Lambda_p)^{-1}(\Phi_p^T)^{-1}(\Phi_p)^T \Lambda_p^T] \\ = \; & \mathrm{trace}[(\Phi_p^{-1})^T K_f \Phi_p^{-1}] = \mathrm{trace}[K_f \Phi_p^{-1}(\Phi_p^{-1})^T] \\ = \; & \mathrm{trace}[K_f (\Phi_p^T \Phi_p)^{-1}] \end{aligned}$$

Hence, we can conclude that in the case of $n = p$,

$$\begin{aligned}
\int_{\mathcal{T}} \mathrm{MSE}_n(\hat{y}(t))dt &= \sum_{i=p+1}^{\infty} \lambda_i + \mathrm{trace}[K_f(\Phi_p^T\Phi_p)^{-1}] - \mathrm{trace}(\Lambda_p) \\
&= \sum_{i=p+1}^{\infty} \lambda_i - \mathrm{trace}(\Lambda_p) + \mathrm{trace}[K_f(\Phi_p^T\Phi_p)^{-1}] \\
&= \sum_{i=p+1}^{\infty} \lambda_i - \mathrm{trace}(\Lambda_p) + \mathrm{trace}[K_p(\Phi_p^T\Phi_p)^{-1}] + \mathrm{trace}[\Sigma_p(\Phi_p^T\Phi_p)^{-1}] \\
&= \sum_{i=p+1}^{\infty} \lambda_i - \mathrm{trace}(\Lambda_p) + \mathrm{trace}[(\Phi_p^T)^{-1}\Phi_p^T\Lambda_p\Phi_p\Phi_p^{-1}] + \mathrm{trace}[\Sigma_p(\Phi_p^T\Phi_p)^{-1}] \\
&= \sum_{i=p+1}^{\infty} \lambda_i + \mathrm{trace}[\Sigma_p(\Phi_p^T\Phi_p)^{-1}]
\end{aligned}$$

□

Example: The conditional Brownian motion on $[0,1]$ with $y(1) = a$

There is one observation here at time 1, i.e. $n = 1$ and $t = 1$. Using theorem 4.2.3, it only needs to be shown that

$$\phi_{p+1}(t)^2 \sum_{i=1}^{p} \lambda_i^2 \phi_i(t)^2 \leq [\sum_{i=1}^{p} \lambda_i \phi_i(t)^2]^2 \ , \tag{4.2.14}$$

where eigenfunction $\phi_i(t) = \sqrt{2}\sin[(i - \frac{1}{2})\pi t]$, $i \geq 1$ in the Brownian motion.

When $t = 1$, $\phi_i(1)^2 = 2$, $i \geq 1$.

$$\begin{aligned}
\phi_{p+1}(t)^2 \sum_{i=1}^{p} \lambda_i^2 \phi_i(t)^2 &= 4 \sum_{i=1}^{p} \lambda_i^2 \\
(\sum_{i=1}^{p} \lambda_i \phi_i(t)^2)^2 &= 4(\sum_{i=1}^{p} \lambda_i)^2
\end{aligned}$$

Since $(\sum_{i=1}^{p} \lambda_i)^2 = \sum_{i=1}^{p} \lambda_i^2 + \sum_{i \neq j} \lambda_i \lambda_j \geq \sum_{i=1}^{p} \lambda_i^2$,

$$\phi_{p+1}(1)^2 \sum_{i=1}^{p} \lambda_i^2 \phi_i(1)^2 \leq [\sum_{i=1}^{p} \lambda_i \phi_i(1)^2]^2 \tag{4.2.15}$$

Therefore, the generalised MSE is a non-increasing function of p based on one prior observation at time 1. Furthermore, the upper bound for the generalised MSE can be calculated using theorem 4.2.4. When $n = 1$ and $p = 1$

$$\begin{aligned}
\Sigma_1 &= K_f(1) - K_1(1) = 1 - \lambda_1\phi_1^2(1) = 1 - 2\lambda_1 \\
\Phi_1^T\Phi_1 &= \phi_1^2(1) = 2 \\
\sum_{i=2}^{\infty}\lambda_i &= \sum_{i=1}^{\infty}\lambda_i - \lambda_1 = \int_0^1 \mathrm{Var}(W(t))dt - \lambda_1 = \frac{1}{2} - \lambda_1
\end{aligned}$$

Hence, for the Brownian motion with $y(1) = a$, the upper bound for its corresponding generalised MSE is

$$\int_0^1 \mathrm{MSE}(\hat{y}(t))dt \leq \frac{1}{2} - \lambda_1 + \frac{1 - 2\lambda_1}{2} = 1 - 2\lambda_1 \qquad (4.2.16)$$

This seems to be new.

4.3 Markovian processes and the Gaussian regression model

Although the analytical form for the generalised MSE can be derived based on the truncated Karhunen-Loeve expansion, the inverse for the truncated covariance matrix K requires the order of the Karhunen-Loeve expansion p at least not less than the number of the observations n, i.e. $p \geq n$. However, in practice, we sometimes hope that $n > p$. Moreover, when $p \geq n$, with large n and even larger p, it is computationally expensive to calculate the inverse matrix. If p can be chosen so that $p < n$ for fixed n, computational time could be saved. In order to solve this inverse matrix problem, two suggestions are put forwarded in this section. The first method

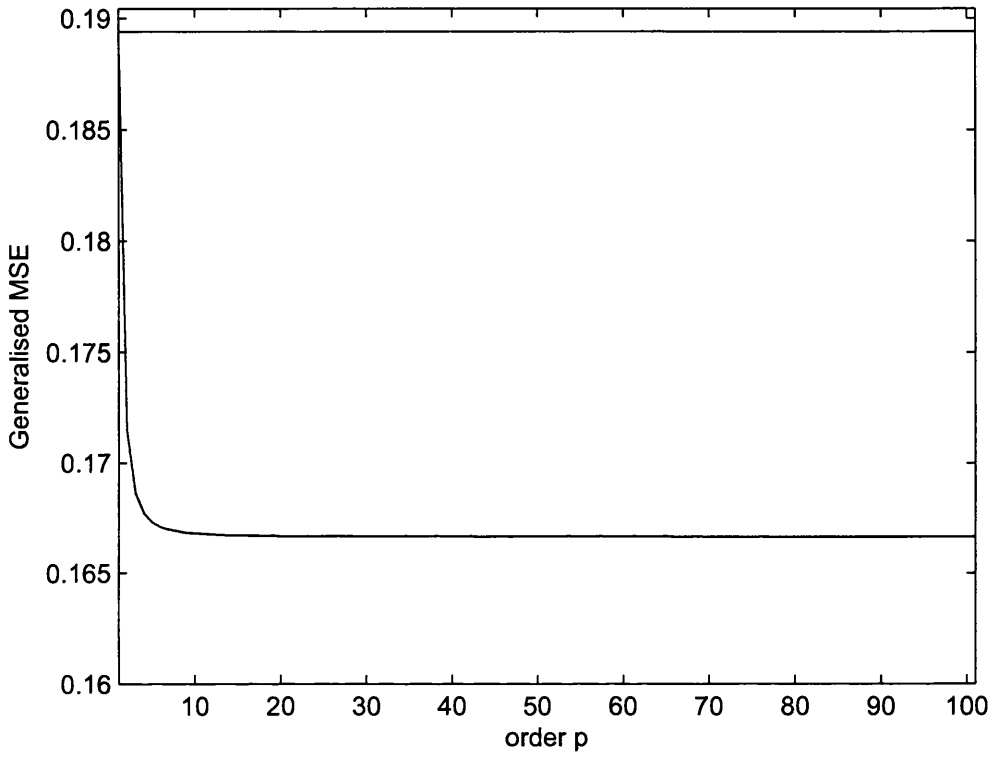

Figure 4.4: The generalised MSE for the conditional Brownian motion based on $y(1) = a$. Blue line: the change of the generalised MSE with the order p; Red line: the upper bound for the generalised MSE.

can be used when the process is Markovian. Under the Markovian model it can be shown that p can be chosen to be any number as long as $p \geq 2$. The second method involves the extension of the reduced model from the truncated Karhunen-Loeve expansion to the Gaussian regression model, which has been discussed in chapter 3. The added independent noise in the Gaussian regression model provides flexibility for the inverse. In the machine learning literature (See, for example, Rasmussen and Williams (2005)), researchers further assume that the true model is indeed the

Gaussian regression model, as well as the model for the prediction. Their results will also be stated in equation 4.3.8.

A process $y(t)$ is called a Markovian process, if the process only depends on the latest information, i.e.

$$p(y(t)|y(s), y(r)) = p(y(t)|y(s)) \qquad r \leq s \leq t \tag{4.3.1}$$

Processes, like the Brownian motion or the Ornstein-Uhlenbeck, satisfy the Markovian property.

In order to get the posterior prediction $\hat{y}(t)$ for $y(t)$, the inverse truncated covariance matrix K^{-1} comprising of the sampling points needs to be calculated. If the number of the sampling points n is too big, the inverse matrix is usually not analytically tractable. However, for the processes satisfying the Markovian property, the calculation for $\hat{y}(t)$ can be simplified.

Assume that the process lies in the interval $\mathcal{T} = [t_0, t_{n+1}]$ with $t_0 \leq t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_n \leq t_{n+1}$ where $t_1, t_2, \cdots, t_n$ are the time points for the sampling data. Now we try to calculate the posterior distribution $p(y(t)|y_s) = p(y(t)|y(t_1), y(t_2), \cdots, y(t_n))$, where $t_i < t < t_{i+1}, 1 \leq i \leq n-1$. It can be shown that $p(y(t)|y(t_1), y(t_2), \cdots, y(t_n)) = p(y(t)|y(t_i), y(t_{i+1}))$, since

$$\begin{aligned}
& p(y(t)|y(t_1), y(t_2), \cdots, y(t_n)) \\
= \; & \frac{p[y(t), y(t_{i+1}), y(t_{i+2}), \cdots, y(t_n)|y(t_1), y(t_2), \cdots, y(t_i)]}{p[y(t_{i+1}), y(t_{i+2}), \cdots, y(t_n)|y(t_1), y(t_2), \cdots, y(t_i)]} \\
= \; & \frac{p[y(t), y(t_{i+1}), y(t_{i+2}), \cdots, y(t_n)|y(t_i)]}{p[y(t_{i+1}), y(t_{i+2}), \cdots, y(t_n)|y(t_i)]} = p[y(t)|y(t_i), y(t_{i+1}), y(t_{i+2}), \cdots, y(t_n)] \\
= \; & \frac{p[y(t_{i+2}), y(t_{i+3}), \cdots, y(t_n)|y(t_i), y(t), y(t_{i+1})]p[y(t_i), y(t), y(t_{i+1})]}{p[y(t_{i+2}), y(t_{i+3}), \cdots, y(t_n)|y(t_i), y(t_{i+1})]p[y(t_i), y(t_{i+1})]} \\
= \; & \frac{p[y(t_{i+2}), y(t_{i+3}), \cdots, y(t_n)|y(t_{i+1})]p[y(t_i), y(t), y(t_{i+1})]}{p[y(t_{i+2}), y(t_{i+3}), \cdots, y(t_n)|y(t_{i+1})]p[y(t_i), y(t_{i+1})]} = p[y(t)|y(t_i), y(t_{i+1})]
\end{aligned}$$

If one of the bounds of t is one of the boundary points, i.e. $t_0 \leq t \leq t_1$ and $t_n \leq t \leq t_{n+1}$, using the same argument, it can be derived that $p(y(t)|y(t_1), y(t_2), \cdots, y(t_n)) = p(y(t)|y(t_1))$ and $p(y(t)|y(t_1), y(t_2), \cdots, y(t_n)) = p(y(t)|y(t_n))$ respectively. In short, the density function of $y(t)$ conditional on all the sampling data is the conditional density function depending on the nearest sampling points of t.

Now the posterior mean can be calculated. For $t_i \leq t \leq t_{i+1}, 1 \leq i \leq n-1$,

$$\begin{aligned}
\hat{y}(t) &= E[y(t)|y_s] \\
&= E[y(t)|y(t_i), y(t_{i+1})] \\
&= \begin{pmatrix} k_{t,i}, & k_{t,i+1} \end{pmatrix} \begin{pmatrix} k_{i,i} & k_{i,i+1} \\ k_{i,i+1} & k_{i+1,i+1} \end{pmatrix}^{-1} \begin{pmatrix} y_i \\ y_{i+1} \end{pmatrix} \\
&= \frac{1}{\Delta_{i,i+1}} \begin{pmatrix} k_{t,i}, & k_{t,i+1} \end{pmatrix} \begin{pmatrix} k_{i+1,i+1} & -k_{i,i+1} \\ -k_{i,i+1} & k_{i,i} \end{pmatrix} \begin{pmatrix} y_i \\ y_{i+1} \end{pmatrix} \\
&= \frac{1}{\Delta_{i,i+1}} (k_{t,i}k_{i+1,i+1}y_i - k_{t,i+1}k_{i,i+1}y_i - k_{t,i}k_{i,i+1}y_{i+1} + k_{t,i+1}k_{i,i}y_{i+1}) \ ,
\end{aligned}$$

where $k_{t,i} = \text{cov}(y(t), y(t_i))$,$k_{i,j} = \text{cov}(y(t_i), y(t_j))$ and $\Delta_{i,i+1} = k_{i,i}k_{i+1,i+1} - k_{i,i+1}^2$
For $t_0 \leq t \leq t_1$ and $t_n \leq t \leq t_{n+1}$, $\hat{y}(t) = E[y(t)|y_s] = E[y(t)|y(t_1)] = \frac{\text{COV}(y(t),y(t_1))}{\text{Var}(y(t_1))} y(t_1)$ and $\hat{y}(t) = E[y(t)|y_{1:n}] = E[y(t)|y(t_n)] = \frac{\text{COV}(y(t),y(t_n))}{\text{Var}(y(t_n))} y(t_n)$ respectively. Since only the nearest points are required in the Markovian model, K is at most 2×2 ($n = 2$). Then any $p \geq n = 2$ satisfies the condition for the inverse of the truncated covariance matrix K.

For simplicity, $\mathcal{T} = [0, 1]$, i.e. $t_0 = 0$ and $t_{n+1} = 1$. Under the Markovian

assumption, for $t_i < t < t_{i+1}$ $(1 \leq i \leq n-1)$, denote

$$\begin{aligned}
\Phi(t)^T &= \Big(\phi_1(t),\ \phi_2(t),\ \cdots,\ \phi_p(t)\Big) \quad \Phi(t_i, t_{i+1}) = \Big(\Phi(t_i),\ \Phi(t_{i+1})\Big) \\
k(t_i, t_{i+1}) &= [\Phi(t_i, t_{i+1})]^T \Lambda \Phi(t) \quad k_f(t_i, t_{i+1})^T = \Big(k_{t,i},\ k_{t,i+1}\Big) \\
K(t_i, t_{i+1}) &= [\Phi(t_i, t_{i+1})]^T \Lambda [\Phi(t_i, t_{i+1})] \quad K_f(t_i, t_{i+1}) = \begin{pmatrix} k_{i,i} & k_{i,i+1} \\ k_{i,i+1} & k_{i+1,i+1} \end{pmatrix}
\end{aligned}$$

At the boundary points, $\Phi(t_0, t_1) = \Phi(t_1)$, $k_f(t_0, t_1) = k_{t,1}$ and $K_f(t_0, t_1) = k_{1,1}$ for $0 < t < t_1$ and $\Phi(t_n, t_{n+1}) = \Phi(t_n)$, $k_f(t_n, t_{n+1}) = k_{t,n}$ and $K_f(t_n, t_{n+1}) = k_{n,n}$ for $t_n < t < 1$.

Therefore, for $t_i < t < t_{i+1}$

$$\begin{aligned}
\text{MSE}(\hat{y}(t)) &= \kappa + [k(t_i, t_{i+1})]^T [K(t_i, t_{i+1})]^{-1} K_f(t_i, t_{i+1}) [K(t_i, t_{i+1})]^{-1} k(t_i, t_{i+1}) \\
&\quad -2[k(t_i, t_{i+1})]^T [K(t_i, t_{i+1})]^{-1} k_f(t_i, t_{i+1}) \\
&= \kappa + \Phi(t)^T \Lambda \Phi(t_i, t_{i+1}) [K(t_i, t_{i+1})]^{-1} K_f(t_i, t_{i+1}) [K(t_i, t_{i+1})]^{-1} [\Phi(t_i, t_{i+1})]^T \Lambda \Phi(t) \\
&\quad -2\Phi(t)^T \Lambda \Phi(t_i, t_{i+1}) [K(t_i, t_{i+1})]^{-1} k_f(t_i, t_{i+1})
\end{aligned}$$

Now assume that $A(t_i, t_{i+1}) = \{a_{uv}\}$, which is a p by p matrix, then

$$\begin{aligned}
&\int_{t_i}^{t_{i+1}} \Phi(t)^T A(t_i, t_{i+1}) \Phi(t) dt = \int_{t_i}^{t_{i+1}} \sum_u \sum_v a_{uv} \phi_u(t) \phi_v(t) dt \\
= \quad &\sum_u \sum_v a_{uv} \int_{t_i}^{t_{i+1}} \phi_u(t) \phi_v(t) dt = \mathbf{1}^T [A(t_i, t_{i+1}) \circ B(t_i, t_{i+1})] \mathbf{1} \ ,
\end{aligned}$$

where

$$\begin{aligned}
&\mathbf{1} = \Big(1,\ 1,\ 1,\ \cdots,\ 1\Big)^T_{n \times 1} \\
&B(t_i, t_{i+1}) = \{\int_{t_i}^{t_{i+1}} \phi_u(t) \phi_v(t) dt\}_{uv} \\
&(A \circ B)_{ij} = A_{ij} B_{ij}, \circ \text{ represents Hadamard (or Schur) product}
\end{aligned}$$

In the current case,

$$A(t_i, t_{i+1}) = \{a_{uv}\} = \Lambda\Phi(t_i, t_{i+1})[K(t_i, t_{i+1})]^{-1}K_f(t_i, t_{i+1})[K(t_i, t_{i+1})]^{-1}[\Phi(t_i, t_{i+1})]^T\Lambda \tag{4.3.2}$$

On the other hand, if $E(t_i, t_{i+1}) = \{e_{uv}\} = [K(t_i, t_{i+1})]^{-1}$,

$$\begin{aligned}
& \int_{t_i}^{t_{i+1}} [k(t_i, t_{i+1})]^T E(t_i, t_{i+1}) k_f(t_i, t_{i+1}) dt \\
= & \int_{t_i}^{t_{i+1}} \sum_u \sum_v e_{uv} \sum_m \lambda_m \phi_m(t) \phi_m(t_u) \mathrm{cov}(t_v, t) dt \\
= & \sum_u \sum_v e_{uv} \sum_m \lambda_m \phi_m(t_u) \int_{t_i}^{t_{i+1}} \phi_m(t) \mathrm{cov}(t_v, t) dt \\
= & \sum_m \lambda_m \sum_u \sum_v e_{uv} \phi_m(t_u) D_m(t_v; t_i, t_{i+1}) \\
= & \sum_m \lambda_m \phi_m^T(t_i, t_{i+1}) E(t_i, t_{i+1}) D_m(t_i, t_{i+1}) \\
= & \mathrm{trace}(\Phi(t_i, t_{i+1}) E(t_i, t_{i+1}) D(t_i, t_{i+1}) \Lambda)
\end{aligned}$$

where for $1 \leq i \leq n-1$

$$\begin{aligned}
D_m(t_v; t_i, t_{i+1}) &= \int_{t_i}^{t_{i+1}} \phi_m(t) \mathrm{cov}(t_v, t) dt \\
D_m(t_i, t_{i+1})^T &= \Big(D_m(t_i; t_i, t_{i+1}), \quad D_m(t_{i+1}; t_i, t_{i+1}) \Big) \\
\phi_m(t_i, t_{i+1})^T &= \Big(\phi_m(t_i), \quad \phi_m(t_{i+1}) \Big) \\
D(t_i, t_{i+1}) &= \Big(D_1(t_i, t_{i+1}), \quad D_2(t_i, t_{i+1}), \quad \cdots, \quad D_p(t_i, t_{i+1}) \Big)
\end{aligned}$$

At the boundary point, $0 \leq t \leq t_1$

$$\begin{aligned}
\phi_m(t_i, t_{i+1}) &= \phi_m(t_1) \\
D(t_i, t_{i+1}) &= \Big(D_1(t_1; 0, t_1), \quad D_2(t_1; 0, t_1), \quad \cdots, \quad D_p(t_1; 0, t_1) \Big)
\end{aligned}$$

while $t_n \leq t \leq 1$

$$\begin{aligned}
\phi_m(t_i, t_{i+1}) &= \phi_m(t_n) \\
D(t_i, t_{i+1}) &= \Big(D_1(t_n; t_n, 1),\ D_2(t_n; t_n, 1),\ \cdots,\ D_p(t_n; t_n, 1)\Big)
\end{aligned}$$

Hence the generalised MSE can be expressed as

$$\begin{aligned}
\int_{\mathcal{T}} \mathrm{MSE}(\hat{y}(t))dt &= \kappa + \sum_{i=0}^{n} \mathbf{1}^T [A(t_i, t_{i+1}) \circ B(t_i, t_{i+1})]\mathbf{1} \\
&\quad -2\sum_{i=0}^{n} \mathrm{trace}(\Phi(t_i, t_{i+1})E(t_i, t_{i+1})D(t_i, t_{i+1})\Lambda)
\end{aligned}$$

where

$$\begin{aligned}
A(t_i, t_{i+1}) &= \Lambda\Phi(t_i, t_{i+1})K^{-1}(t_i, t_{i+1})K_f K^{-1}(t_i, t_{i+1})\Phi^T(t_i, t_{i+1})\Lambda \\
E(t_i, t_{i+1}) &= K^{-1}(t_i, t_{i+1})
\end{aligned}$$

The other way of solving the inverse matrix problem is to use the Gaussian regression model, which has been mentioned in chapter 3. Now the reduced model contains the noise term, as well as the truncated Karhunen-Loeve expansion, i.e.

$$y(t) = \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i(t)\xi_i + \tau_t \tag{4.3.3}$$

where $\{\xi_i\} \sim i.i.d.N(0,1)$, $\{\tau_t\} \sim i.i.d.N(0, \sigma^2)$ and ξ_i, τ_t independent. Hence $K = \Phi^T\Lambda\Phi + \sigma^2 I$. The formula for the generalised MSE remains the same form except for the change in K, i.e.

$$\begin{aligned}
\int_{\mathcal{T}} \mathrm{MSE}(\hat{y}(t))dt &= \sum_{i=1}^{\infty} \lambda_i - \mathrm{trace}(\Lambda\Phi(2K^{-1} - K^{-1}K_f K^{-1})\Phi^T\Lambda) \\
K &= \Phi^T\Lambda\Phi + \sigma^2 I
\end{aligned}$$

There are two ways to estimate the hyperparameter σ^2. Assume that the length of the interval $\mathcal{T}$ is T. If we know the covariance function

$$\begin{aligned}
\sigma^2 T &= \int_{\mathcal{T}} \sigma^2 dt = \int_{\mathcal{T}} \mathrm{Var}(\tau_t) dt \\
&= \int_{\mathcal{T}} \mathrm{Var}(\sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(t)\xi_i - \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i(t)\xi_i) dt \\
&= \sum_{i=p+1}^{\infty} \lambda_i \\
&= \int_{\mathcal{T}} \mathrm{Var}(y(t)) dt - \mathrm{trace}(\Lambda)
\end{aligned}$$

It is equivalent to

$$\sigma^2 = \frac{\int_{\mathcal{T}} \mathrm{Var}(y(t)) dt - \mathrm{trace}(\Lambda)}{T} \tag{4.3.4}$$

The approach to estimate σ^2 using equation 4.3.4 is called the eigenvalue approach. If the covariance function is not known, the other approach, called the maximum likelihood approach can be used. The likelihood function for y_s is expressed as

$$p(y_s) = \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\Phi^T \Lambda \Phi + \sigma^2 I)|^{\frac{1}{2}}} \exp[-\frac{1}{2} y_s^T (\Phi^T \Lambda \Phi + \sigma^2 I) y_s] \tag{4.3.5}$$

σ^2 is chosen so that $p(y_s)$ is maximised. Or equivalently, minimize $-\log(p(y))$

$$-\log(p(y)) = \frac{1}{2}\log[|\det(\Phi^T \Lambda \Phi + \sigma^2 I)^{-1}|] + \frac{1}{2}[y_s^T (\Phi^T \Lambda \Phi + \sigma^2 I)^{-1} y_s] \tag{4.3.6}$$

However, the estimator of σ^2 roughly decreases with the increase of the order p. When the first p eigenvalues can explain almost all the cumulative expected variance, i.e. $\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \to 1$, the estimator of σ^2 becomes very small, i.e. $\sigma^2 \to 0$. This can be seen from equation 4.3.4. When the covariance function is known, equation 4.3.4 can also be written as

$$\sigma^2 = \frac{\sum_{i=1}^{\infty} \lambda_i - \sum_{i=1}^{p} \lambda_i}{T} = \frac{1 - \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i}}{\frac{T}{\sum_{i=1}^{\infty} \lambda_i}} \tag{4.3.7}$$

Therefore, when $\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{\infty}\lambda_i} \rightarrow 1$, $\sigma^2 \rightarrow 0$. When the covariance function is not known, which means that the likelihood method will be applied to find the estimator of σ^2, the trend in the change of the estimator of σ^2 with respect to the order p remains more or less the same. Again, the higher the p is in the model, the smaller the estimator of σ^2 is. Refer to figure 4.6 for an example.

For the Gaussian regression model, the covariance matrix K is expressed as

$$K = \Phi^T \Lambda \Phi + \sigma^2 I$$

It can be seen that the noise is only added to the diagonal element of the truncated covariance matrix K, but not the off-diagonal element. When $\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{\infty}\lambda_i}$ is not very close to 1, to the diagonal element of K, σ^2 tends to capture the true covariance function in addition to the truncated Karhunen-Loeve expansion, but to the off-diagonal element, only the truncated expansion is used without the added noise σ^2. Under relatively small $\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{\infty}\lambda_i}$, the quality of the approximation using only the truncated Karhunen-Loeve expansion could be poor. This unbalance covariance matrix structure between the diagonal element and the off-diagonal element might result in a big generalised MSE. When p is increased so that $\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{\infty}\lambda_i}$ approaches 1, it is expected that the unbalanced structure due to σ^2 on the covariance matrix becomes smaller, since the estimator for σ^2 decreases. Thus the generalised MSE also becomes smaller.

When both the true model and the reduced model are the Gaussian regression model, i.e. $K_f = K = \Phi^T \Lambda \Phi + \sigma^2 I$ and $k_f = k = \Phi^T \Lambda \Phi(t)$,

$$\begin{aligned} \mathrm{MSE}(\hat{y}(t)) &= \kappa - k^T K^{-1} k \\ \int_{\mathcal{T}} \mathrm{MSE}(\hat{y}(t)) dt &= \mathrm{trace}\Lambda - \mathrm{trace}[\Lambda\Phi(\Phi^T\Lambda\Phi + \sigma^2 I)^{-1}\Phi^T\Lambda] \\ &= \mathrm{trace}[\Lambda - \Lambda\Phi(\Phi^T\Lambda\Phi + \sigma^2 I)^{-1}\Phi^T\Lambda] \end{aligned}$$

Hence,

$$\int_{\mathcal{T}} \text{MSE}(\hat{y}(t))dt = \text{trace}(\Lambda^{-1} + \sigma^{-2}\Phi\Phi^T)^{-1} \tag{4.3.8}$$

Equation 4.3.8 uses the matrix inversion formula

$$(A + UWV^T)^{-1} = A^{-1} - A^{-1}U(W^{-1} + V^T A^{-1}U)^{-1}V^T A^{-1}$$

The following example analyses the generalised MSE for the Brownian motion, in terms of the analytical solution as well as the Karhunen-Loeve approximation. Notice that the Brownian motion is a Markovian process. The computational time will be dramatically reduced, since the dimension of the truncated covariance matrix K is at most 2×2. In the second step of the analysis, the Gaussian regression model will also be used as the reduced model to investigate the change in the generalised MSE.

Example: The Brownian motion on $[0, 1]$

The analytical solution to the generalised MSE and the optimal design points in terms of the generalised MSE

The analytical solution to the generalised MSE of the Brownian motion is derived in the same way as that of the Ornstein-Uhlenbeck process in Rasmussen and Williams (2005). The covariance function of the Brownian motion is expressed as $\text{cov}(y(t_i), y(t_j)) = \min(t_i, t_j)$. Assume $0 = t_0 \leq t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_n \leq t_{n+1} = 1$ where $t_1, t_2, \cdots, t_n$ are the sampling points, then for $t_i \leq t \leq t_{i+1}$, $1 \leq i \leq n-1$

$$\Delta_{i,i+1} = t_i t_{i+1} - t_i^2 = t_i(t_{i+1} - t_i) = t_i\delta_i$$

$$k_{t,i}k_{i+1,i+1}y_i - k_{t,i+1}k_{i,i+1}y_i - k_{t,i}k_{i,i+1}y_{i+1} + k_{t,i+1}k_{i,i}y_{i+1}$$

$$= t_i[t_{i+1} - t]y_i + t_i(t - t_i)y_{i+1} = t_i\delta_i' y_i + t_i(\delta_i - \delta_i')y_{i+1}$$

$$\hat{y}(t) = \frac{\delta_i'}{\delta_i}y(t_i) + (1 - \frac{\delta_i'}{\delta_i})y(t_{i+1}) = r_i y(t_i) + (1 - r_i)y(t_{i+1})$$

where $\delta_i = t_{i+1} - t_i$, $\delta_i' = t_{i+1} - t$ and $r_i = \frac{\delta_i'}{\delta_i}$.

For the boundary points $0 \leq t \leq t_1$ and $t_n \leq t \leq 1$, $\hat{y}(t) = \frac{t}{t_1} y(t_1)$ and $\hat{y}(t) = \frac{t_n}{t_n} y(t_n) = y(t_n)$ respectively.

Now the mean squared error can be calculated.

$$\text{MSE}(\hat{y}(t)) = \text{Var}(\hat{y}((t))) + \text{Var}(y(t)) - 2\text{cov}(\hat{y}(t), y(t)) \tag{4.3.9}$$

For $t_i \leq t \leq t_{i+1}$, $1 \leq i \leq n-1$

$$\begin{aligned}
\text{MSE}(\hat{y}(t)) &= t + r_i^2 t_i + (1-r_i)^2 t_{i+1} + 2r_i(1-r_i)t_i - 2r_i t_i - 2(1-r_i)t \\
&= -t - r_i^2 t_i + t_{i+1} - 2r_i t_{i+1} + t_{i+1} r_i^2 + 2r_i t \\
&= \frac{(t-t_i)(t_{i+1}-t)}{t_{i+1}-t_i}
\end{aligned}$$

For $0 \leq t \leq t_1$, $\text{MSE}(\hat{y}(t)) = t + \frac{t^2}{t_1^2} t_1 - 2\frac{t^2}{t_1} = \frac{t(t_1-t)}{t_1}$. For $t_n \leq t \leq 1$, $\text{MSE}(\hat{y}(t)) = t + t_n - 2t_n = t - t_n$.

Hence the generalised MSE can be expressed as

$$\begin{aligned}
&\int_0^1 \text{MSE}(\hat{y}(t))dt = \sum_{i=0}^{n} \int_{t_i}^{t_{i+1}} \text{MSE}(\hat{y}(t))dt \\
= &\int_0^{t_1} \text{MSE}(\hat{y}(t))dt + \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} \text{MSE}(\hat{y}(t))dt + \int_{t_n}^{1} \text{MSE}(\hat{y}(t))dt \\
= &\frac{1}{6}\sum_{i=1}^{n-1}(t_{i+1}-t_i)^2 + \frac{t_1^2}{6} + \frac{(1-t_n)^2}{2} \\
= &\frac{1}{6}\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2 + \frac{(1-t_n)^2}{2}
\end{aligned}$$

The sampling points $\{t_i\}$, $1 \leq i \leq n$ can be chosen, so that the generalised MSE is minimised. This can be performed through the differentiation of the generalised MSE

with respect to each time point t_i

$$\frac{\partial}{\partial t_i}\int_0^1 \text{MSE}(\hat{y}(t))dt = 0 \Rightarrow 2t_i = t_{i-1} + t_{i+1} \qquad 1 \leq i \leq n-1$$
$$\frac{\partial}{\partial t_n}\int_0^1 \text{MSE}(\hat{y}(t))dt = 0 \Rightarrow 4t_n = t_{n-1} + 3$$

After the rearrangement, $t_i = \frac{3i}{3n+1}$ and the $\min[\int_0^1 \text{MSE}(\hat{y}(t))dt] = \frac{1}{2(3n+1)}$

The generalised MSE for the Brownian motion under the Markov assumption

Under the Markov assumption, the covariance matrix K and K_f are either 2×2 matrix or are scalars depending on whether the nearest sampling point is a boundary point or not. The generalised MSE is expressed as

$$\begin{aligned}\int_0^1 \text{MSE}(\hat{y}(t))dt &= \kappa + \sum_{i=0}^{n} \mathbf{1}^T[A(t_i, t_{i+1}) \circ B(t_i, t_{i+1})]\mathbf{1} \\ &\quad -2\sum_{i=0}^{n} \text{trace}(\Phi(t_i, t_{i+1})E(t_i, t_{i+1})D(t_i, t_{i+1})\Lambda) \ ,\end{aligned}$$

where

$$\begin{aligned}A(t_i, t_{i+1}) &= \Lambda\Phi(t_i, t_{i+1})[K(t_i, t_{i+1})]^{-1}K_f(t_i, t_{i+1})[K(t_i, t_{i+1})]^{-1}\Phi^T(t_i, t_{i+1})\Lambda \\ E(t_i, t_{i+1}) &= [K(t_i, t_{i+1})]^{-1}\end{aligned}$$

For the Brownian motion, the main term $B(t_i, t_{i+1})$ and $D(t_i, t_{i+1})$ can be calculated as follows

$$\begin{aligned}D_m(t_i; t_i, t_{i+1}) &= \int_{t_i}^{t_{i+1}} \phi_m(t)\min(t_i, t)dt = \int_{t_i}^{t_{i+1}} \sqrt{2}\sin[(m-\frac{1}{2})\pi t]t_i dt \\ &= \sqrt{2}\frac{t_i}{(m-\frac{1}{2})\pi}[\cos((m-\frac{1}{2})\pi t_i) - \cos((m-\frac{1}{2})\pi t_{i+1})]\end{aligned}$$

$$
\begin{aligned}
D_m(t_{i+1};t_i,t_{i+1}) &= \int_{t_i}^{t_{i+1}} \phi_m(t)\min(t_{i+1},t)dt = \int_{t_i}^{t_{i+1}} \sqrt{2}\sin[(m-\frac{1}{2})\pi t]tdt \\
&= -\frac{\sqrt{2}}{(m-\frac{1}{2})\pi}\int_{t_i}^{t_{i+1}} td[\cos((m-\frac{1}{2})\pi t)] \\
&= -\frac{\sqrt{2}}{(m-\frac{1}{2})\pi}[t_{i+1}\cos((m-\frac{1}{2})\pi t_{i+1}) - t_i\cos((m-\frac{1}{2})\pi t_i) \\
&\quad -\frac{1}{(m-\frac{1}{2})\pi}(\sin((m-\frac{1}{2})\pi t_{i+1}) - \sin((m-\frac{1}{2})\pi t_i))]
\end{aligned}
$$

$$
\begin{aligned}
B_{uv}(t_i,t_{i+1}) &= \int_{t_i}^{t_{i+1}} \phi_u(t)\phi_v(t)dt = \int_{t_i}^{t_{i+1}} 2\sin[(u-\frac{1}{2})\pi t]\sin[(v-\frac{1}{2})\pi t]dt \\
&= \int_{t_i}^{t_{i+1}} \cos[(u-v)\pi t] - \cos[(u+v-1)\pi t]dt \\
(\text{If } u\neq v) &= \frac{1}{(u-v)\pi}[\sin[(u-v)\pi t_{i+1}] - \sin[(u-v)\pi t_i]] \\
&\quad -\frac{1}{(u+v-1)\pi}[\sin[(u+v-1)\pi t_{i+1}] - \sin[(u+v-1)\pi t_i]] \\
(\text{If } u=v) &= \int_{t_i}^{t_{i+1}} 1-\cos[(2u-1)\pi t]dt \\
&= t_{i+1}-t_i - \frac{1}{(2u-1)\pi}[\sin[(2u-1)\pi t_{i+1}] - \sin[(2u-1)\pi t_i]]
\end{aligned}
$$

Figure 4.5 shows that assuming the optimal sampling points, the generalised MSE of the truncated Karhunen-Loeve expansion ($p = 55$) under the Markovian assumption and the analytical generalised MSE are quite close to each other. If we further check the value for this difference, it is around $10^{-5} - 10^{-6}$ (see, table 4.1). Also notice that under the Markovian assumption, we can choose any n as long as $n \geq 2$ instead of $p \geq n$, which provides flexibility and eases the computation.

Both figure 4.5 and table 4.1 imply that it is reasonable to use the truncated Karhunen-Loeve expansion for the prediction in the Brownian motion.

The reduced model is the Gaussian regression model

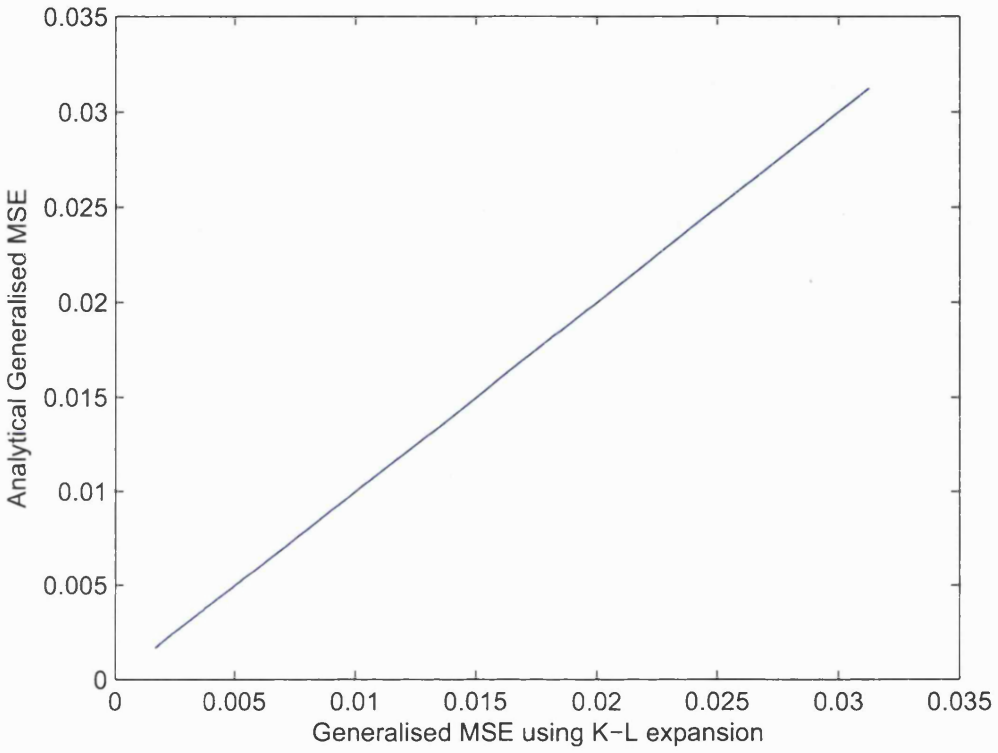

Figure 4.5: Under the optimal scheme, i.e. $t_i = \frac{3i}{3n+1}$, the generalised MSE using the Karhunen-Loeve expansion and the Markovian assumption vs the analytical solution in the Brownian motion, $p = 55$ and $5 \leq n \leq 100$.

We now consider the case, when the reduced model is the Gaussian regression model expressed in equation 4.3.3, which contains the added noise in addition to the truncated Karhunen-Loeve expansion, while the full model is the original process.

In order to apply the Gaussian regression model, the hyperparameter σ^2 needs to be estimated first. As is mentioned previously there are two approaches, the eigenvalue approach and the maximum likelihood approach, to estimate the hyperparamters depending on whether the covariance function is known or not. Both

n	5	25	45	65	85	100
Markovian	0.0312594	0.0066426	0.0036987	0.0025588	0.0019574	0.0016646
Analytical	0.0312500	0.0065789	0.0036765	0.0025510	0.0019531	0.0016611

Table 4.1: Comparison between the solution using the Markovian assumption and the analytical solution to the generalised MSE in the Brownian motion when $p = 55$.

approaches will be used to estimate σ^2 first since the covariance function for the Brownian motion is known to be $\min(s, t)$. In the maximum likelihood approach, the data is simulated at the sampling points.

It can be seen from figure 4.6 that σ^2 derived from both the eigenvalue approach and the maximum likelihood approach follows the same pattern. When p increases, σ^2 roughly decreases. However, σ^2 from the likelihood approach appears to be more volatile, which reflects the unpredictable characteristic of the stochastic process in reality. In the following analysis, for simplicity, the result from the eigenvalue approach is used, i.e. σ^2 derived from equation 4.3.4.

Figure 4.7 shows the performance of the generalised MSE under the Gaussian regression model with different p and n. As is expected, it does not perform very well under small p. For $n = 50$, and $p = 5$, $\frac{\sum_{i=1}^{5}\lambda_i}{\sum_{i=1}^{\infty}\lambda_i} = 0.9596$, i.e. there are still about 4% of the cumulative expected variance that can not be explained. The difference of the generalised MSE between the Gaussian regression model and that of the analytical solution is almost 10^{-2}. The reason appears to be that the balance is not well kept between the diagonal and the off-diagonal elements of the truncated covariance matrix. With lower p, the approximation of the covariance function at the diagonal elements relies more on bigger σ^2 to capture the true covariance function, while the approximation of the covariance function for the off-diagonal elements remains using merely

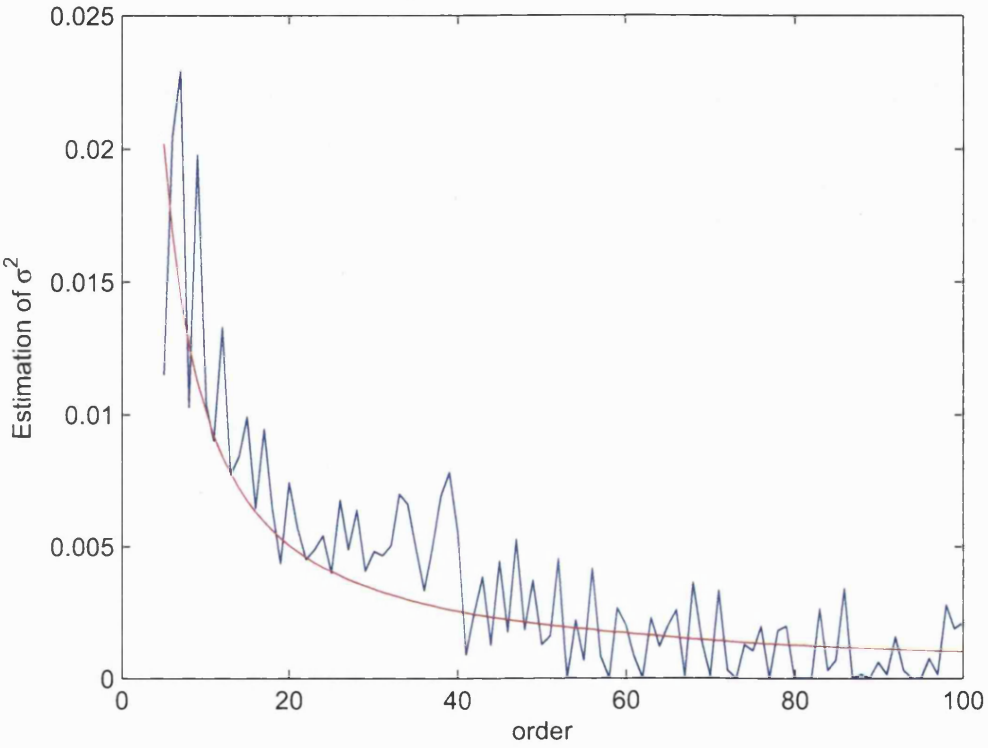

Figure 4.6: The estimator of σ^2 using both approaches, when order increases from 5 to 100. Red line: the eigenvalue approach; Blue line: the maximum likelihood approach.

the truncated Karhunen-Loeve expansion. The bigger the p, the less the diagonal elements rely on σ^2, and thereafter the more improvement there is in the generalised MSE. Furthermore, when $p \geq n$, the result under the Gaussian regression model (with σ^2) can be compared with the result under the truncated Karhunen-Loeve expansion (without σ^2). It is shown in the plot that the result with σ^2 actually performs better, although the difference between them are only $10^{-6} - 10^{-5}$. When n is increased to 200, and the p is used at a relatively big value, which lies in between 150 to 250,

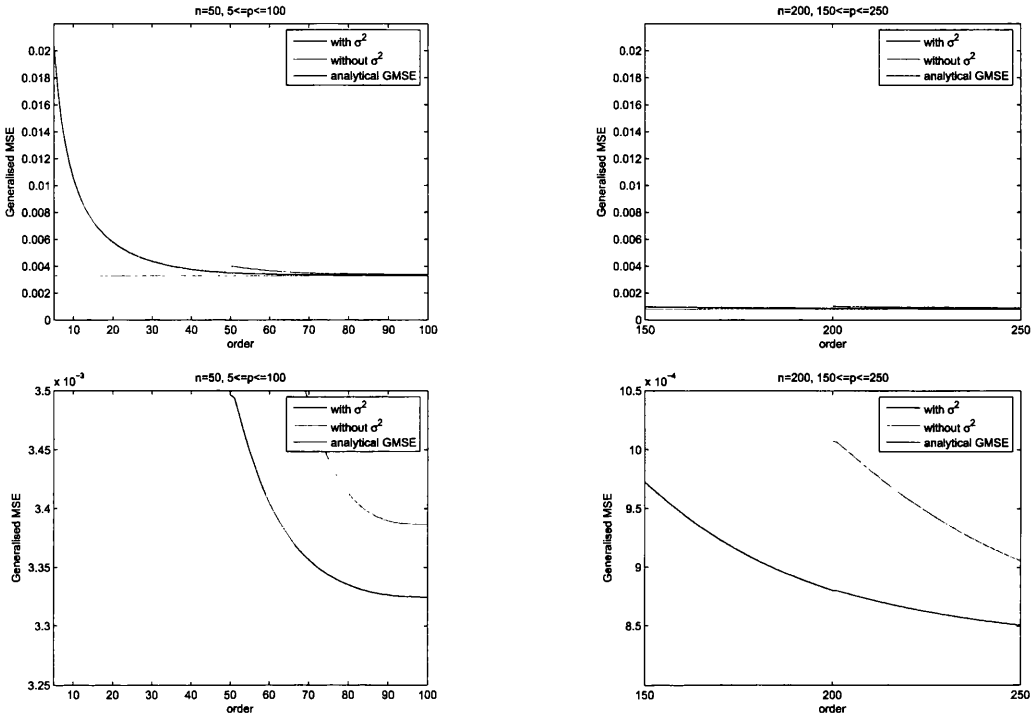

Figure 4.7: The generalised MSE using the Gaussian regression model, the truncated Karhunen-Loeve expansion and the analytical generalised MSE. Blue: The generalised MSE using the Gaussian regression model; Green: The generalised MSE using the truncated Karhunen-Loeve expansion; Red: The analytical generalised MSE. Left: $n = 50$ and $5 \leq p \leq 100$; Right: $n = 200$ and $150 \leq p \leq 250$.

the difference between the generalised MSE of the Gaussian regression model and that of the analytical solution reduces to only 10^{-4}. In fact, $\frac{\sum_{i=1}^{150} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \approx 0.999$. This means that the majority of the cumulative expected variance has been explained by this higher order p. The result derived by the Gaussian regression model continues to outperform the result derived by merely the truncated Karhunen-Loeve expansion. Their difference remains around $10^{-6} - 10^{-5}$.

4.4 Numerical methods for prediction and generalised MSE

For most processes as mentioned, where analytical solutions to the Karhunen-Loeve expansion are not applicable, numerical solutions play an important role. The numerical method under here is the Haar wavelet expansion method described in chapter 2, which has shown to be both computationally fast and accurate and treats the eigenfunction as a function, when the covariance function is known.

Use the same notation as chapter 2, when the reduced model is the truncated Karhunen-Loeve expansion and $p \geq n$

$$\begin{aligned}
\Phi(t) &= D\Psi(t) \\
\Phi &= D\Psi \\
K &= \Phi^T \Lambda \Phi = \Psi^T D^T \Lambda D \Phi \\
k &= \Phi^T \Lambda \Phi(t) = \Psi^T D^T \Lambda D \Psi(t)
\end{aligned}$$

Hence the prediction and its corresponding generalised MSE can be expressed as follows

$$\begin{aligned}
\hat{y}(t) &= \Psi(t)^T D^T \Lambda D \Psi (\Psi^T D^T \Lambda D \Phi)^{-1} y_s \\
\int_{\mathcal{T}} \mathrm{MSE}[\hat{y}(t)] dt &= \sum_{i=1}^{\infty} \lambda_i - \mathrm{trace}[\Lambda \Phi (2K^{-1} - K^{-1} K_f K^{-1}) \Phi^T \Lambda] \\
&= \sum_{i=1}^{\infty} \lambda_i - \mathrm{trace}\{\Lambda D \Psi [2(\Psi^T D^T \Lambda D \Psi)^{-1} \\
&\quad -(\Psi^T D^T \Lambda D \Psi)^{-1} K_f (\Psi^T D^T \Lambda D \Psi)^{-1}] \Psi^T D^T \Lambda\}
\end{aligned}$$

When the reduced model is the Gaussian regression model with σ^2 added as the variance of the independent noise for the prediction, the only change is that $K =$

$\Psi^T D^T \Lambda D \Psi + \sigma^2$. When the covariance function is known, and the time interval is on $[0,1]$, σ^2 can be estimated as

$$\sigma^2 = \int_0^1 \mathrm{Var}(y(t))dt - \mathrm{trace}(\Lambda) \tag{4.4.1}$$

We still consider the Brownian motion first. Since the analytical solution of its generalised MSE is known under the optimal sampling points, a comparison can be made between the analytical solution and the solution from the numerical scheme.

In figure 4.8, 50 optimal design points at $t_i = \frac{3i}{3\times 50+1}$, $1 \leq i \leq 50$, together with 50 testing points at $t_i = \frac{3i-2}{3\times 50+1}$ are simulated. They are displayed on the top left plot. For the prediction performance shown on the top right plot and the bottem left plot, both the truncated Karhunen-Loeve expansion and the Gaussian regression model perform quite well, since most of the predictions are within the 95% confidence interval. The way of constructing the confidence interval is quite like that used in the regression analysis. Denote $\hat{y}(t) = a^T y_s$, where $a^T = k^T K^{-1}$, then

$$\begin{aligned}
\mathrm{Var}(y(t) - \hat{y}(t)) &= \mathrm{Var}(y(t) - a^T y_s) \\
&= \mathrm{Var}(y(t)) + a^T \mathrm{Var}(y_s) a - 2a^T \mathrm{cov}(y(t), y_s) \\
&= \kappa + k^T K^{-1} K_f K^{-1} k - 2k^T K^{-1} k_f
\end{aligned}$$

Since $E[y(t) - \hat{y}(t)] = 0$, then

$$\frac{y(t) - \hat{y}(t)}{\sqrt{\mathrm{Var}(y(t) - \hat{y}(t))}} \sim N(0,1) \tag{4.4.2}$$

Therefore, the 95% confidence interval for $y(t)$ is calculated as

$$[\hat{y}(t) - 1.96\sqrt{\mathrm{Var}(y(t) - \hat{y}(t))}, \hat{y}(t) + 1.96\sqrt{\mathrm{Var}(y(t) - \hat{y}(t))}] \tag{4.4.3}$$

For the generalised MSE with the change of p shown on the bottom right plot, the results derived from the numerical schemes using the Haar wavelet method are slightly

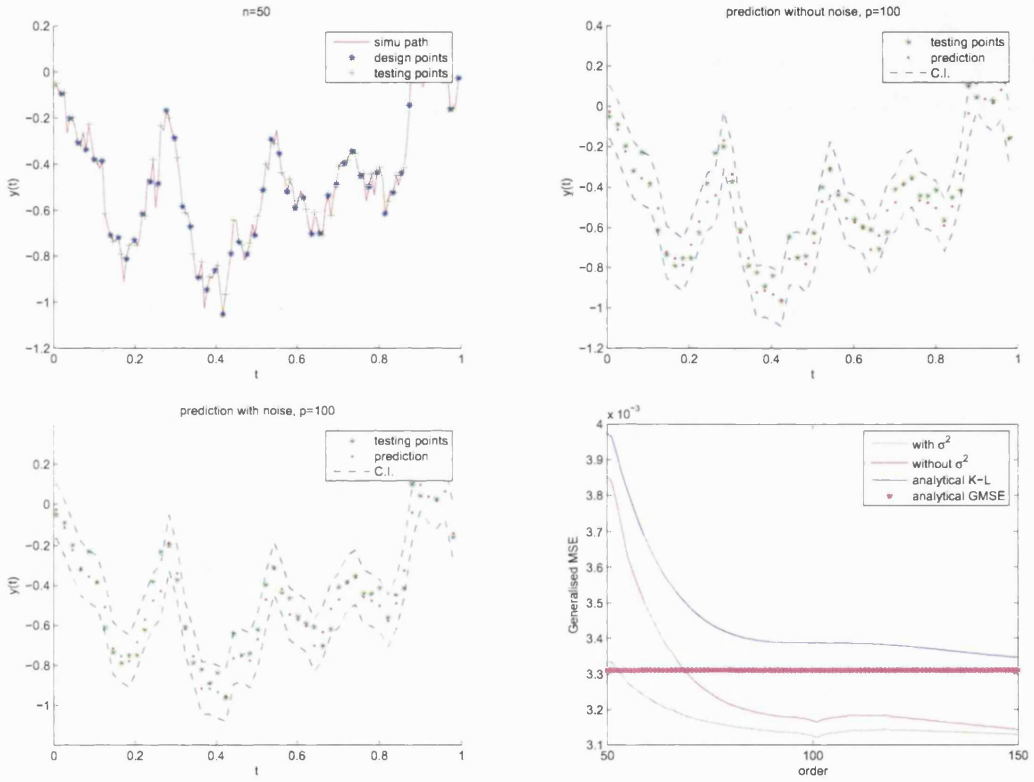

Figure 4.8: The numerical scheme for calculating the generalised MSE using the Haar wavelet when $M = 256$ for the Brownian motion. Top left: the simulation path of the Brownian motion (red), 50 sampling points (blue), together with 50 testing points (green); Top right: the prediction of the testing points using the truncated Karhunen-Loeve expansion ($p = 100$), with the corresponding 95% confidence interval; Bottom left: the prediction of the testing points using the Gaussian regression model ($p = 100$), with the corresponding 95% confidence interval; Bottom right: the generalised MSE comparison between the analytical Karhunen-Loeve expansion and the numerical scheme.

different from the one derived analytically using the Karhunen-Loeve expansion. The difference is about 10^{-3}, which is also the difference between the approximation of the eigenfunction using the Haar wavelet ($M = 256$) and the analytical solution (see, section 2.3, chapter 2). This value of 10^{-3} could also be the reason why the generalised MSE derived from the Haar wavelet is below the analytical solution $\frac{1}{2(3\times50+1)}$ when p approaches around 60. Moreover, the generalised MSE derived from the Gaussian regression model has generally even smaller error. This matches our previous analysis.

The reason that the Brownian motion is used as the first example is because its analytical solution can be used to compare with the numerical result. Next, another example will be shown, whose analytical solution is untractable. It is the stochastic process with the kernel function $(1 + |t - s|)\exp[-|s - t|]$. This time the sampling points are chosen to be $t_i = \frac{2i}{2n+1}$, $1 \leq i \leq n$, while the testing points are $t_i = \frac{2i-1}{2n+1}$, $1 \leq i \leq n$, when n equals to 50. The eigenvalues for this covariance function share the same character as that for the squared exponential kernel in the sense that the first two eigenvalues counts for 99% of the sum of all the eigenvalues. The first five eigenvalues (from highest to lowest) are 0.9435229, 0.0508775, 0.0043546, 0.0008119 and 0.0002382 respectively.

Figure 4.9 provides the same information for the kernel $(1 + |t - s|)\exp[-|s - t|]$ as that for the Brownian motion in figure 4.8. The variance for this prediction is relatively small, only about 10^{-4}. That is why the confidence interval is so narrow. In terms of the prediction shown on the top right plot and the bottom left plot, both the truncated Karhunen-Loeve expansion (without noise) and the Gaussian regression model (with noise) perform quite well again with most of the predictions lying within the 95% confidence interval. The generalised MSE displayed on the bottom right

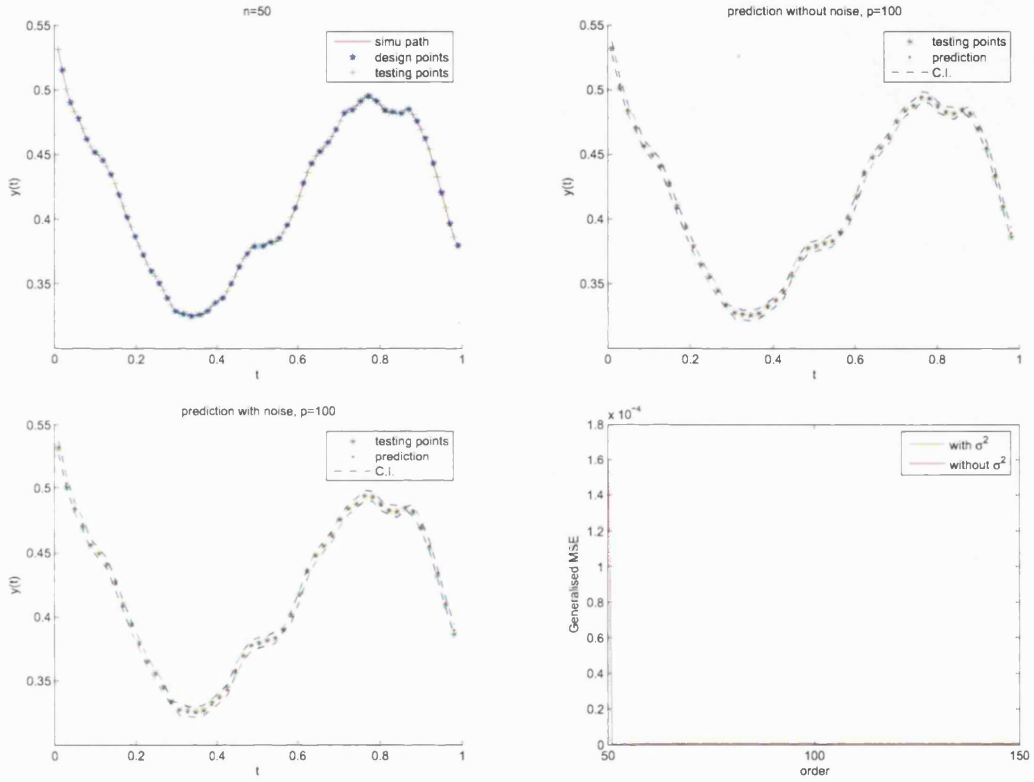

Figure 4.9: The numerical scheme for calculating the generalised MSE using the Haar wavelet when $M = 256$ for the process with the kernel $(1 + |t - s|)\exp[-|s - t|]$. Top left: the simulation path of the process (red), 50 sampling points (blue), together with 50 testing points (green); Top right: the prediction of the testing points using the truncated Karhunen-Loeve expansion ($p = 100$), with the corresponding 95% confidence interval; Bottom left: the prediction of the testing points using the Gaussian regression model ($p = 100$), with the corresponding 95% confidence interval; Bottom right: the generalised MSE comparison between the numerical scheme with σ^2 and that without σ^2.

order	50	70	130	150
with σ^2	4.8613191×10^{-6}	8.9756746×10^{-7}	8.9761501×10^{-7}	8.9640381×10^{-7}
without σ^2	1.7192874×10^{-4}	8.9676508×10^{-7}	9.022133×10^{-7}	8.9157505×10^{-7}

Table 4.2: The difference of the generalised MSE using the truncated Karhunen-Loeve expansion (without σ^2) and using the Gaussian regression model (with σ^2), when $n = 50$ for the kernel $(1 + |t - s|)\exp[-|s - t|]$.

plot is very small using both of these two models. Table 4.2 lists the value of the generalised MSE for the order 50, 70, 130 and 150, when $n = 50$.

Except for the result at order 50, which has relatively bigger difference (10^{-4}), the majority of the orders bigger than 50 result in the difference as small as 10^{-7}. This is a quite different result compared with that of the Brownian motion. One of the reasons is that we do not need such high orders in the model, since the first two eigenvalues already explain more than 99% of the cumulative expected variance. However for the truncated Karhunen-Loeve expansion (without noise), the calculation has to meet the assumption $p \geq n$, which means that p could be too big in this case if n is already 50. Hence the Gaussian regression model would be preferred as the reduced model in this problem to break the above assumption.

From figure 4.10, it shows that the generalised MSE dramatically decreases in the first few orders. When the order is 10, the generalised MSE is already quite close to 0 and remains stable for the later orders. This matches our expectation, since it is the first few terms in the truncated Karhunen-Loeve expansion which play a major role in this expansion.

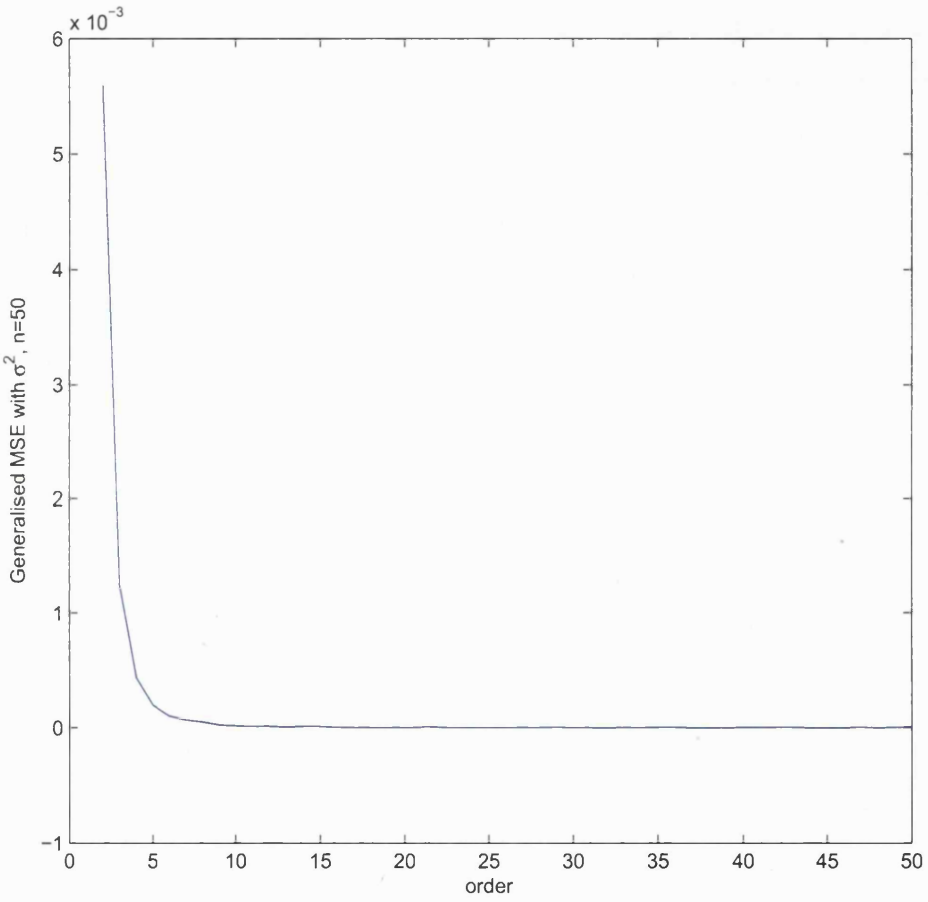

Figure 4.10: The numerical generalised MSE under the Gaussian regression model for the kernel $(1 + |t - s|)\exp[-|s - t|]$, when $n = 50$.

4.5 Conditional Karhunen-Loeve expansion

We assign value a to the sampling points $y_s = \{y(t_i), 1 \leq i \leq n\}$, where

$$a^T = \Big(a_1, \quad a_2, \quad \cdots, \quad a_n\Big); \quad a_i \in \mathbb{R} \ , 1 \leq i \leq n \tag{4.5.1}$$

Then the conditional expression $y(t)|\{y_s = a\}$ is a new stochastic process, which can be approximated using a Karhunen-Loeve expansion. In this section, this expansion under the conditional setting is called the conditional Karhunen-Loeve expansion.

In this section, the conditional Karhunen-Loeve expansion is derived for two types of processes, which are the Markovian process and the non-Markovian process. It is shown that the solution to the Markovian process can be calculated at each interval, whose boundary points are the sampling points, while the analytical solution to the non-Markovian process is in general more difficult to obtain than the solution to the Markovian process.

Using equation 1.2.24, chapter 1, we can now expand $y(t)|\{y_s = a\}$

$$y(t)|\{y_s = a\} = E(y(t)|\{y_s = a\}) + \sum_{i \geq 1} \sqrt{\lambda_i}\phi_i(t)\xi_i \tag{4.5.2}$$

where λ_i and $\phi_i(t)$ solves the Fredholm integral equation

$$\int_{\mathcal{T}} \text{cov}(y(t), y(s)|y_s = a)\phi_i(s)ds = \lambda_i\phi_i(t) \tag{4.5.3}$$

and ξ_i is the independent random variable with mean 0 and variance 1. If each $y(t_i)$ is a Gaussian process, $y(t)|\{y_s = a\}$ is also a Gaussian process, and thereafter $\xi_i \sim i.i.d.N(0, 1)$.

Using equation 4.1.3, the conditional expectation can be modelled as

$$E(y(t)|\{y_s = a\}) = k^T K_n^{-1} a \ , \tag{4.5.4}$$

where

$$\begin{aligned}(K_n)_{i,j} &= \text{cov}(y(t_i), y(t_j)), \quad 1 \leq i, j \leq n \\ k^T &= \Big(\text{cov}(y(t), y(t_1)), \ \text{cov}(y(t), y(t_2)), \ \cdots, \ \text{cov}(y(t), y(t_n))\Big)\end{aligned}$$

Section 4.2, chapter 4 provides a detailed explanation on the approximation of the conditional expectation using the truncated Karhunen-Loeve expansion and its corresponding generalised mean squared error. In this section, we try to capture the whole stochastic part of $y(t)|\{y_s = a\}$ using a biorthogonal system as defined in equation 4.5.2. For simplicity, it is further assumed that for each sampling point, $y(t_i) = 0$, $1 \leq i \leq n$, i.e. $a = 0$. Then $E(y(t)|\{y_s = 0\}) = 0$ and

$$y(t)|\{y_s = 0\} = \sum_{i \geq 1} \sqrt{\lambda_i}\phi_i(t)\xi_i \tag{4.5.5}$$

In terms of the conditional covariance, the partition inverse equation can be applied twice. For time s and time t, the conditional covariance is

$$\text{cov}(y(s), y(t)|y_s = a) = \text{cov}(y(s), y(t)) - k_s^T K_n^{-1} k_t \tag{4.5.6}$$

When there is only one prior observation, i.e. $n = 1$

$$\text{cov}(y(s), y(t)|y(t_1) = a_1) = \text{cov}(y(t), y(s)) - \frac{\text{cov}(y(s), y(t_1))\text{cov}(y(t), y(t_1))}{\text{Var}(y(t_1))} \tag{4.5.7}$$

Equation 4.5.7 can be applied to derive the covariance function of the Brownian bridge, since the Brownian bridge is constructed using the Brownian motion $W(t)$ conditional on $W(1) = 0$. Then the conditional covariance function for the Brownian bridge is

$$\begin{aligned}\text{cov}(W(s), W(t)|W(1) = 0) &= \text{cov}(W(t), W(s)) - \frac{\text{cov}(W(s), W(1))\text{cov}(W(t), W(1))}{\text{Var}(W(1))} \\ &= \min(t, s) - st\end{aligned}$$

which is exactly the covariance function we expect.

After obtaining the conditional covariance function, the corresponding Fredholm integral equation can now be solved.

$$\int_{\mathcal{T}} \text{cov}(y(s), y(t)|y_s = a)\phi(s)ds = \lambda\phi(t) \tag{4.5.8}$$

Applying equation 4.5.6, we obtain

$$\int_{\mathcal{T}} \text{cov}(y(s), y(t))\phi(s)ds - \int_{\mathcal{T}} (k_s^T K_n^{-1} k_t)\phi(s)ds = \lambda\phi(t) \tag{4.5.9}$$

If denoting $K_n^{-1} = \{a_{ij}\}$, equation 4.5.9 can be written down as

$$\int_{\mathcal{T}} \text{cov}(y(s), y(t))\phi(s)ds - \int_{\mathcal{T}} \sum_i \sum_j a_{ij}\text{cov}(y(s), y(t_i))\text{cov}(y(t), y(t_j))\phi(s)ds = \lambda\phi(t) \tag{4.5.10}$$

Notice further that if one of the points (t or s) is one of the sampling points, i.e. $t = t_i$ or $s = t_i$, $1 \leq i \leq n$, the conditional covariance equals to zero. This could be seen from equation 4.5.6 directly. Therefore, when solving the integral equation

$$\int_{\mathcal{T}} \text{cov}(y(t_i), y(s)|y_s = a)\phi(s)ds = \lambda\phi(t_i) \tag{4.5.11}$$

with $\text{cov}(y(t_i), y(s)|y_s = a) = 0$, a boundary condition for the eigenfunction can be obtained, which is $\phi(t_i) = 0$, $1 \leq i \leq n$.

The rest of this section presents some examples on deriving the conditional Karhunen-Loeve expansion.

Example 1: The Brownian motion on $[0, 1]$ with two prior observations $y(1) = 0$ and $y(\frac{1}{3}) = 0$.

We need to solve the following Fredholm integral equation.

$$\int_0^1 \{\min(t, s) - \frac{1}{\frac{1}{3} - \frac{1}{9}}[\min(t, \frac{1}{3})\min(s, \frac{1}{3}) + \frac{1}{3}st - \frac{1}{3}\min(t, \frac{1}{3})s - \frac{1}{3}\min(s, \frac{1}{3})t]\}\phi(s)ds = \lambda\phi(t)$$

Differentiating it twice results in

$$-\phi(t) = \lambda \frac{d^2}{dt^2}\phi(t) \tag{4.5.12}$$

It is the same differential equation as that of the Brownian motion. Since the boundary condition at $\frac{1}{3}$ divides the whole interval into two parts, the eigenfunction can then be solved in these two intervals separately. Denote

$$\phi(t) = \begin{cases} f(t) & 0 < t < \frac{1}{3} \\ g(t) & \frac{1}{3} < t < 1 \end{cases} \tag{4.5.13}$$

On one hand, for $0 < t < \frac{1}{3}$, the boundary conditions are $f(0) = f(\frac{1}{3}) = 0$. $f(0) = 0$ provides the functional form for the eigenfunction

$$f(t) = A\sin(\frac{t}{\sqrt{\lambda}}) \tag{4.5.14}$$

Since $f(\frac{1}{3}) = 0$,

$$\lambda_i = \frac{1}{9i^2\pi^2} \tag{4.5.15}$$

Using the orthogonality condition, $A = \sqrt{6}$. Then $f_i(t) = \sqrt{6}\sin(3\pi it)$

On the other hand, for $\frac{1}{3} < t < 1$, the functional form of the solution to equation 4.5.12 is

$$g(t) = A\sin(\frac{t}{\sqrt{\lambda}}) + B\cos(\frac{t}{\sqrt{\lambda}}) \tag{4.5.16}$$

The boundary conditions $g(\frac{1}{3}) = g(1) = 0$ suggest that

$$A\sin(\frac{1}{3\sqrt{\lambda}}) + B\cos(\frac{1}{3\sqrt{\lambda}}) = 0 \quad \text{and} \quad A\sin(\frac{1}{\sqrt{\lambda}}) + B\cos(\frac{1}{\sqrt{\lambda}}) = 0 \tag{4.5.17}$$

Further simplification of equation 4.5.17 results in

$$\sin(\frac{2}{3\sqrt{\lambda}}) = 0 \tag{4.5.18}$$

which implies that $\lambda_j = \frac{4}{9j^2\pi^2}$

Therefore the functional form of $g(t)$ is $g(t) = A\sin(\frac{3j\pi}{2}t) + B\cos(\frac{3j\pi}{2}t)$. Applying the boundary condition $\phi(\frac{1}{3}) = 0$ again, (Or alternatively we can also use $\phi(1) = 0$. Both of these conditions result in the same solution.)

$$A\sin(\frac{j\pi}{2}) + B\cos(\frac{j\pi}{2}) = 0 \tag{4.5.19}$$

Therefore, the solution to the eigenfunction is

$$j \text{ odd} \rightarrow A = 0 \rightarrow \phi_j(t) = B\cos(\frac{3j\pi t}{2})$$
$$j \text{ even} \rightarrow B = 0 \rightarrow \phi_j(t) = A\sin(\frac{3j\pi t}{2})$$

Using the orthogonality condition, $A = B = \sqrt{3}$.

The eigenvalues for this problem are a combination of both $\frac{1}{9i^2\pi^2}$, $i \geq 1$ for $1 < t < \frac{1}{3}$ and $\frac{4}{9j^2\pi^2}$, $j \geq 1$ for $\frac{1}{3} < t < 1$. Since the eigenvalue is calculated interval by interval, it implies that the eigenfunction can also be expressed interval by interval, i.e. the ith eigenfunction is $\phi_i(t) = f_i(t)I_{(0<t<\frac{1}{3})} + g_i(t)I_{(\frac{1}{3}<t<1)}$. When $\lambda_i = \frac{1}{9i^2\pi^2}$, $\phi_i(t) = f_i(t)I_{(0<t<\frac{1}{3})}$; when $\lambda_i = \frac{4}{9j^2\pi^2}$, $\phi_i(t) = g_i(t)I_{(\frac{1}{3}<t<1)}$.

Hence the conditional Karhunen-Loeve expansion for this process can be expressed as

$$\begin{aligned} y(t)|\{y(\frac{1}{3}) = 0, y(1) = 0\} = & \sum_{i\geq 1} \frac{1}{3\pi i}\sqrt{6}\sin(3\pi it)I_{(0<t<\frac{1}{3})}\xi_i \\ & + \sum_{j\geq 1} \frac{2}{3\pi(2j-1)}\sqrt{3}\cos[\frac{3\pi(2j-1)}{2}t]I_{(\frac{1}{3}<t<1)}\xi'_{2j-1} \\ & + \sum_{j\geq 1} \frac{2}{3\pi(2j)}\sqrt{3}\sin[\frac{3\pi(2j)}{2}t]I_{(\frac{1}{3}<t<1)}\xi'_{2j} \end{aligned}$$

where $\xi_i, \xi'_j \sim i.i.dN(0,1)$

	1st	2nd	3rd	4th	5th	6th
Analytical	$\frac{4}{9\pi^2}$	$\frac{1}{9\pi^2}$	$\frac{4}{36\pi^2}$	$\frac{4}{81\pi^2}$	$\frac{4}{144\pi^2}$	$\frac{1}{36\pi^2}$
Analytical value	0.045032	0.011258	0.011258	0.005004	0.002814	0.002814
Numerical(Haar)	0.045033	0.011259	0.011259	0.005005	0.002816	0.002816

Table 4.3: The analytical and the numerical solution to the first six eigenvalues of the conditional Brownian motion on $y(\frac{1}{3}) = y(1) = 0$.

Now the performance of the numerical solution using the Haar wavelet method is briefly compared with the analytical solution. For the accuracy of the Haar wavelet, 256 basis functions are chosen. (see table 4.3 and figure 4.11)

Whatever the eigenvalue or the eigenfunction, solutions from both the Haar wavelet numerical scheme and the analytical solution are very close to each other. It can reassure us again that the eigenvalues for the conditional Brownian motion are a combination of the eigenvalues from different intervals, while the eigenfunctions only have non-zero values in the interval from which the corresponding eigenvalues calculate.

Remark 4.5.1. Now take a further look at the conditional covariance function for the Brownian motion in this problem,

$$\text{cov}(y(s), y(t)|y(\frac{1}{3}) = 0, y(1) = 0) = 0 \quad \min(s,t) < \frac{1}{3} < \max(s,t) \tag{4.5.20}$$

Hence the Fredholm integral equation we need to solve can be separated into two parts.

$$\int_0^{\frac{1}{3}} \text{cov}(y(s), y(t)|y(\frac{1}{3}) = 0, y(1) = 0) f(s) ds = \lambda f(t) \quad f(0) = f(\frac{1}{3}) = 0$$
$$\int_{\frac{1}{3}}^1 \text{cov}(y(s), y(t)|y(\frac{1}{3}) = 0, y(1) = 0) g(s) ds = \lambda' g(t) \quad g(\frac{1}{3}) = g(1) = 0$$

Differentiating either part results in the same differential equation as that in the Brownian motion case, except for different boundary conditions. The above argument

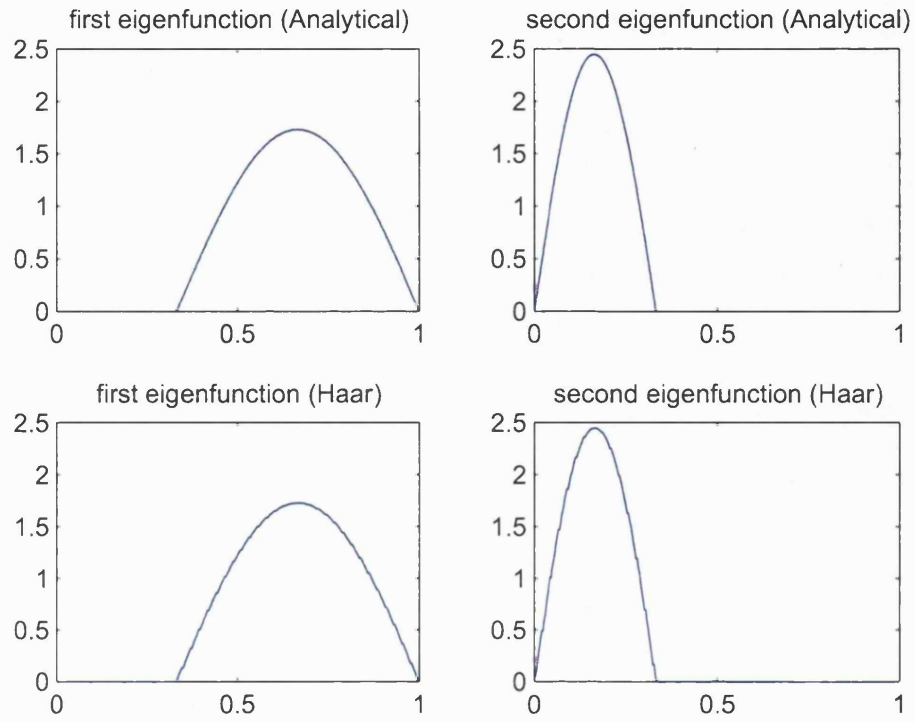

Figure 4.11: Comparison of the first two eigenfunctions between the analytical solution and the solution derived from the Haar wavelet (M=256) of the conditional Brownian motion on $y(\frac{1}{3}) = y(1) = 0$.

provides an alternative way to understand why the final solution to the eigenvalues is a combination of the eigenvalues of two different intervals and why the eigenfunctions only have non-zero values in one of the intervals.

Remark 4.5.2. Remark 4.5.1 also shows a fact that the conditional covariance function for the Brownian motion has a non-zero value if there is no sampling point in between s and t.

This is actually true for any Markovian process. We only show the case if neither of the closest sampling points of t and s is the boundary point. It can be easily

generalised to the situation involving the boundary points.

Assume $y(t)$ is a Markovian process. It has n sampling points t_i, $1 \leq i \leq n$ with $t_1 < t_2 < \cdots < t_n$. It needs to be shown that for $j+1 \leq i$ and $t_j < s < t_{j+1} \leq t_i < t < t_{i+1}$, $\text{cov}(y(s), y(t)|y_s = a) = \text{cov}(y(s), y(t)|y(t_1) = a_1, \cdots, y(t_n) = a_n) = 0$. This is true for the Markovian process because

$$\begin{aligned} p(y(t), y(s)|y_s) &= p(y(t)|y(s), y_s)p(y(s)|y_s) \\ &= p(y(t)|y(t_i), y(t_{i+1}))p(y(s)|y(t_j), y(t_{j+1})) \\ &= p(y(t)|y_s)p(y(s)|y_s) \end{aligned}$$

Hence $\text{cov}(y(s), y(t)|y_s = a) = 0$ if there is sampling point(s) in between $y(t)$ and $y(s)$.

Example 2: The Brownian motion on $[0, 1]$ when $y(t_i) = 0$, $1 \leq i \leq n$ is observed, where $0 = t_0 < t_1 < t_2 < \cdots < t_n = 1$

Example 2 is a generalisation of example 1.

Since $\text{cov}(y(t_i), y(t_j)) = \min(t_i, t_j)$, after denoting $a_{ij} = \min(t_i, t_j)$, the Fredholm integral equation can be written down as

$$\int_0^1 \min(s,t)\phi(s)ds - \sum_i \sum_j a_{ij} \min(t, t_j) \int_0^1 \min(s, t_i)\phi(s)ds = \lambda\phi(t) \quad (4.5.21)$$

It is equivalent to

$$\int_0^1 \min(s,t)f(s)ds - \sum_i \sum_{t \leq t_j} a_{ij} t \int_0^1 \min(s, t_i)\phi(s)ds$$
$$- \sum_i \sum_{t > t_j} a_{ij} t_j \int_0^1 \min(s, t_i)\phi(s)ds = \lambda\phi(t)$$

Differentiating it once, we obtain

$$\int_t^1 \phi(s)ds - \sum_i \sum_{t \leq t_j} a_{ij} \int_0^1 \min(s, t_i)\phi(s)ds = \lambda \frac{d}{dt}\phi(t) \quad (4.5.22)$$

Differentiating it a second time, we obtain

$$-\phi(t) = \lambda \frac{d^2}{dt^2}\phi(t) \tag{4.5.23}$$

Since the Brownian motion is a Markovian process, we can solve the above differential equation at each individual interval $[t_i, t_{i+1}]$, $0 \leq i \leq n-1$. Assume at the interval $[t_i, t_{i+1}]$, $0 \leq i \leq n-1$, we have the eigenvalue $\lambda^{(i+1)}$ and the corresponding eigenfunction $\phi^{(i+1)}$. Then the differential equation, together with its boundary conditions, are

$$-\phi^{(i+1)}(t) = \lambda^{(i+1)} \frac{d^2}{dt^2}\phi^{(i+1)}(t) \quad \phi^{(i+1)}(t_i) = \phi^{(i+1)}(t_{i+1}) = 0, 0 \leq i \leq n-1 \tag{4.5.24}$$

Equation 4.5.24 results in the functional form of $\phi^{(i+1)}(t)$,

$$\phi^{(i+1)}(t_{i+1}) = A^{(i+1)}\sin(\frac{t}{\sqrt{\lambda^{(i+1)}}}) + B^{(i+1)}\cos(\frac{t}{\sqrt{\lambda^{(i+1)}}}) \tag{4.5.25}$$

For the first interval $0 < t < t_1$, using the boundary condition $\phi^{(1)}(0) = \phi^{(1)}(t_1) = 0$, the functional form of the eigenfunction is $\phi_k^{(1)}(t) = A^{(1)}\sin(\frac{\pi k}{t_1}t)$ with the eigenvalue $\lambda_k^{(i+1)} = \frac{t_1^2}{k^2\pi^2}$. Combining with the orthogonality condition, $\int_0^{t_1}[\phi_k^{(1)}(s)]^2 ds = 1$, it can be obtained that

$$\phi_k^{(1)}(t) = \sqrt{\frac{2}{t_1}}\sin(\frac{\pi k}{t_1}t) \tag{4.5.26}$$

For the rest intervals $t_i < t < t_{i+1}$, $1 \leq i \leq n-1$, we have the boundary conditions $\phi^{(i+1)}(t_i) = \phi^{(i+1)}(t_{i+1}) = 0$. It means that

$$A^{(i+1)}\sin(\frac{t_i}{\sqrt{\lambda^{(i+1)}}}) + B^{(i+1)}\cos(\frac{t_i}{\sqrt{\lambda^{(i+1)}}}) = 0$$
$$A^{(i+1)}\sin(\frac{t_{i+1}}{\sqrt{\lambda^{(i+1)}}}) + B^{(i+1)}\cos(\frac{t_{i+1}}{\sqrt{\lambda^{(i+1)}}}) = 0$$

Simplifying the above equations, we obtain,

$$\sin(\frac{t_{i+1} - t_i}{\lambda^{(i+1)}}) = 0 \tag{4.5.27}$$

Hence the eigenvalue $\lambda^{(i+1)}$ is

$$\lambda_k^{(i+1)} = \frac{(t_{i+1} - t_i)^2}{k^2\pi^2} \tag{4.5.28}$$

$A^{(i+1)}$ and $B^{(i+1)}$ can be decided by looking at the boundary condition $\phi^{(i+1)}(t_i) = \phi^{(i+1)}(t_{i+1}) = 0$ again, as well as the orthogonality condition $\int_{t_i}^{t_{i+1}} [\phi^{(i+1)}(s)]^2 ds = 1$.

Hence, after solving for $A^{(i+1)}$ and $B^{(i+1)}$, the conditional Karhunen-Loeve expansion can be expressed as

$$y(t)|\{y(t_i) = 0, 0 < t_1 < t_2 < \cdots < t_n = 1\} = \sum_{k\geq 1} \frac{t_1}{k\pi}\sqrt{\frac{2}{t_1}}\sin(\frac{\pi k}{t_1}t) I_{(0<t<t_1)}\xi_{1,k}$$
$$+ \sum_{i=1}^{n-1}\sum_{k\geq 1} \frac{t_{i+1} - t_i}{k\pi}[A^{(i+1)}\sin(\frac{\pi k}{t_{i+1} - t_i}t) + B^{(i+1)}\cos(\frac{\pi k}{t_{i+1} - t_i}t)] I_{(t_i<t<t_{i+1})}\xi_{(i+1),k}$$

where $\xi_{i,k} \sim i.i.d.N(0,1)$, $1 \leq i - 1 \leq n, k \geq 1$.

The conditional Karhunen-Loeve expansion changes frequency in the eigenfunction, while the functional form of the expansion remains as a linear combination of the trigonometric functions.

For the Ornstein-Uhlenbeck process, a general solution given n sampling points will also be provided. Notice that the Ornstein-Uhlenbeck process is again a Markovian process, hence the integral equation can be solved interval by interval.

Example 3 The Ornstein-Ullenbeck process on $[0, 1]$ when $y(t_i) = 0$, $1 \leq i \leq n$ is observed, where $0 = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = 1$.

For the Onstein-Uhlenbeck process, the covariance function is $\exp(-\beta|t - s|)$ up to a scalar multiplier. Given n prior observations, the conditional covariance function can then be expressed as

$$\exp(-\beta|t - s|) - \sum_k \sum_h a_{kh}\exp(-\beta|t - t_h|)\exp(-\beta|s - t_k|) \tag{4.5.29}$$

where $a_{kh} = \exp(-\beta|t_k - t_h|)$.

Since the Ornstein-Uhlenbeck process is a Markovian process, it is assumed that for each interval $[t_i, t_{i+1}]$, $0 \leq i \leq n$, we have the eigenvalue $\lambda^{(i+1)}$, together with the eigenfunction $\phi^{(i+1)}(t)$. Then the integral equation at each interval $[t_i, t_{i+1}]$, $0 \leq i \leq n$, is

$$\begin{aligned} & \int_{t_i}^{t_{i+1}} \exp(-\beta|t-s|)\phi^{(i+1)}(s)ds \\ - & \sum_k \sum_h a_{kh}\exp(-\beta|t-t_h|) \int_{t_i}^{t_{i+1}} \exp(-\beta|s-t_k|)\phi^{(i+1)}(s)ds \\ = & \lambda\phi^{(i+1)}(t) \end{aligned}$$

Differentiating the integral equation twice, we obtain

$$-2\beta\phi^{(i+1)}(t) + \beta^2\lambda\phi^{(i+1)}(t) = \lambda\frac{d^2\phi^{(i+1)}(t)}{dt^2} \tag{4.5.30}$$

Simplifying equation 4.5.30, we obtain

$$\frac{d^2\phi^{(i+1)}(t)}{dt^2} + (w^{(i+1)})^2\phi^{(i+1)}(t) = 0 \quad \text{when setting} \quad (w^{(i+1)})^2 = \frac{2\beta - \beta^2\lambda^{(i+1)}}{\lambda^{(i+1)}} \tag{4.5.31}$$

Again this is the same differential equation as that of the unconditional case. Hence the functional form of the eigenfunction at interval $[t_i, t_{i+1}]$ is

$$\phi^{(i+1)}(t) = A^{(i+1)}\cos(w^{(i+1)}t) + B^{(i+1)}\sin(w^{(i+1)}t) \quad \phi^{(i+1)}(t_{i+1}) = \phi^{(i+1)}(t_i) = 0 \tag{4.5.32}$$

For any interval whose both bounds are the sampling points, i.e. $[t_i, t_{i+1}]$, $1 \leq i \leq n-1$, the boundary conditions are $\phi^{(i+1)}(t_i) = \phi^{(i+1)}(t_{i+1}) = 0$. It implies that

$$\begin{aligned} A^{(i+1)}\cos(w^{(i+1)}t_i) + B^{(i+1)}\sin(w^{(i+1)}t_i) &= 0 \\ A^{(i+1)}\cos(w^{(i+1)}t_{i+1}) + B^{(i+1)}\sin(w^{(i+1)}t_{i+1}) &= 0 \end{aligned}$$

Solving the above equations is equivalent to solving

$$\sin(w^{(i+1)}(t_{i+1} - t_i)) = 0 \tag{4.5.33}$$

Therefore the solution to $w_k^{(i+1)}$ is

$$w_k^{(i+1)}(t_{i+1} - t_i) = k\pi \quad w_k^{(i+1)} = \frac{k\pi}{t_{i+1} - t_i} \tag{4.5.34}$$

Hence the solution to the eigenvalue $\lambda_k^{(i+1)}$ is

$$\lambda_k^{(i+1)} = \frac{2\beta}{(w_k^{(i+1)})^2 + \beta^2} = \frac{2\beta(t_{i+1} - t_i)^2}{k^2\pi^2 + \beta^2(t_{i+1} - t_i)^2} \tag{4.5.35}$$

For the boundary points, the equations for deriving the eigenvalues are slightly different. In the interval $[0, t_1]$, the boundary conditions are as follows

$$\frac{d}{dt}\phi^{(1)}(0) - \beta\phi^{(1)}(0) = 0$$
$$\phi^{(1)}(t_1) = 0$$

Replacing the eigenfunction with its functional form, we have

$$A^{(1)}\beta - B^{(1)}w^{(1)} = 0$$
$$A^{(1)}\cos(w^{(1)}t_1) + B^{(1)}\sin(w^{(1)}t_1) = 0$$

Solving the above equations is equivalent to solving

$$\cot(w^{(1)}t_1) = -\frac{\beta}{w^{(1)}} \tag{4.5.36}$$

For $[t_n, 1]$, the boundary conditions are

$$\frac{d}{dt}\phi^{(n+1)}(1) + \beta\phi^{(n+1)}(1) = 0$$
$$\phi^{(n+1)}(t_n) = 0$$

Replacing the eigenfunction with its functional form again, we have

$$A^{(n+1)}(\beta - w^{(n+1)}\tan(w^{(n+1)})) + B^{(n+1)}(\beta\tan(w^{(n+1)}) + w^{(n+1)}) = 0$$

$$A^{(n+1)}\cos(w^{(n+1)}t_n) + B^{(n+1)}\sin(w^{(n+1)}t_n) = 0$$

Then the above boundary conditions can be simplified as

$$\cot(w^{(n+1)}t_n) = \frac{\beta - w^{(n+1)}\tan w^{(n+1)}}{\beta\tan w^{(n+1)} + w^{(n+1)}} \tag{4.5.37}$$

Combining the above results, the eigenvalues for the conditional Onstein-Uhlenbeck process can be obtained. $A^{(i+1)}$ and $B^{(i+1)}$ can be further solved using the orthogonality condition of the eigenfunction.

The above three examples calculate the conditional Karhunen-Loeve expansion for the Brownian motion and the Ornstein-Uhlenbeck process. However, both these two processes are Markovian. In practice, when the process is not Markovian, the analytical solution is usually very difficult to obtain, or at least not as simple as the solution for the Markovian process. Now the process with the kernel function $1-|t-s|$ based on one prior observation at time S will be used to demonstrate the calculation procedure of deriving the conditional Karhunen-Loeve expansion in general.

Example 4: Process with the kernel $1 - |t - s|$ on $[0, 1]$, conditional on one observation at time S

After observing at time S, we can express the conditional covariance function.

$$1 - |t - s| - (1 - |t - S|)(1 - |s - S|) \tag{4.5.38}$$

It is further assumed that the eigenfunction $\phi(t)$ is

$$\phi(t) = \begin{cases} f(t) & 0 < t < T \\ g(t) & T < t < 1 \end{cases} \tag{4.5.39}$$

Hence the following integral equations need to be solved

$$\begin{cases} \int_0^t [1-t+s-(1-S+t)(1-S+s)]f(s)ds \\ +\int_t^S [1-s+t-(1-S+t)(1-S+s)]f(s)ds \\ +\int_S^1 [1-s+t-(1-S+t)(1-s+S)]g(s)ds = \lambda f(t) \quad 0<t<S \\ \int_0^T [1-t+s-(1-t+S)(1-S+s)]f(s)ds \\ +\int_T^t [1-t+s-(1-t+S)(1-s+S)]g(s)ds \\ +\int_t^1 [1-s+t-(1-t+S)(1-s+S)]g(s)ds = \lambda g(t) \quad S<t<1 \end{cases}$$

Differentiating the equation twice with respect to t, the following differential equation is obtained

$$\begin{aligned} & 2f(t)+\lambda f''(t)=0 \quad 2g(t)+\lambda g''(t)=0 \\ \text{or} \quad & f''(t)+w^2 f(t)=0 \quad g''(t)+w^2 g(t)=0 \quad \text{where } w^2=\frac{2}{\lambda} \end{aligned}$$

Therefore, the solutions to $f(t)$ and $g(t)$ can be expressed as

$$f(t)=c_1\sin(wt)+c_2\cos(wt) \quad g(t)=c_3\sin(wt)+c_4\cos(wt) \tag{4.5.40}$$

The boundary conditions of the process can be found as follows

$$f(S)=g(S)=0 \quad f'(0)=-g'(1)=f(0)+g(1) \tag{4.5.41}$$

With the functional form of the eigenfunction, the boundary conditions can be simplified as

$$\begin{aligned} & c_2=-\tan(wS)c_1 \quad c_1=-c_3\cos(w)+c_4\sin(w) \\ & c_4=-\tan(wS)c_3 \quad c_1 w=c_2+c_3\sin(w)+c_4\cos(w) \end{aligned}$$

After combining these boundary conditions, we can solve for the eigenvalues out of the following equation

$$\tan(w-wS)+\tan(wS)+w=0 \quad \lambda=\frac{2}{w^2} \tag{4.5.42}$$

As an example, we assume that an observation is observed at one time point $S = \frac{1}{3}$. The analytical solution can now be plotted in figure 4.12.

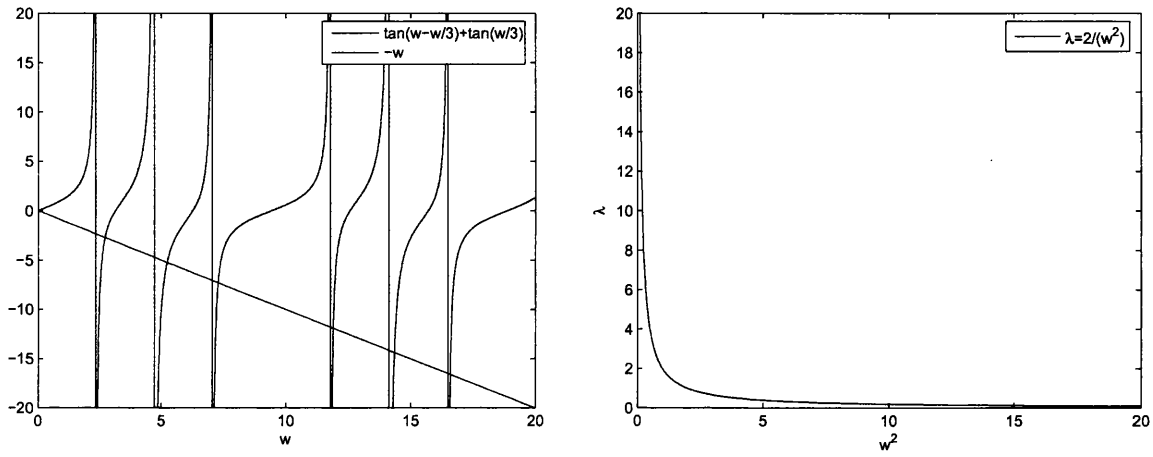

Figure 4.12: The analytical solution to the eigenvalues for the process with the kernel function $1 - |t - s|$. (Left) $\tan(w - \frac{w}{3}) + \tan(\frac{w}{3})$ vs $-w$; (Right) $\lambda = \frac{2}{w^2}$.

The left plot of figure 4.12 plots $\tan(w - \frac{1}{3}w) + \tan(\frac{1}{3}w)$ and $-w$ respectively. Each intersection in the plot represents one solution to w. After obtaining the solution to w from the left plot, the solution to the eigenvalue λ can be obtained from the right plot, which shows the inverse quadratic relationship between λ and w.

The eigenvalues from the analytical solution can now be further compared with the solution from the Haar wavelet method with 256 basis functions, i.e. $M = 256$. Table 4.4 lists the first six eigenvalues. It can be seen that they are very close to each other. The difference does not appear until 10^{-3}. The numerical solution confirms again the validity of the calculation procedure in deriving the analytical solution.

The above process with 1 sampling point can be generalised to n sampling points. The calculation procedure should still be the same, although it is usually difficult to derive the solution analytically. It should be noted that deriving the conditional

	Analytical	Numerical Haar ($M = 256$)
1st eigenvalue	0.2696687	0.2696716
2nd eigenvalue	0.0727949	0.0727979
3rd eigenvalue	0.0375350	0.0375377
4th eigenvalue	0.0141320	0.0141346
5th eigenvalue	0.0097205	0.0097232
6th eigenvalue	0.0072675	0.0072701

Table 4.4: Comparison of the analytical solution and the numerical solution to the first six eigenvalues of the process with the kernel function $1 - |t - s|$.

Karhunen-Loeve expansion is closely related to deriving the univariate time, multivariate state Karhunen-Loeve expansion using the "lining-up" method, which will be discussed in detail in the next chapter.

Chapter 5

Multivariate Karhunen-Loeve Expansion

In the previous chapters, we mainly deal with univariate stochastic processes. However, in practice, researchers and practitioners are also interested in the multivariate setting, either multivariate in time or multivariate in state. Multivariate in state and univariate in time have an important practical impact on the fields like finance, while multivariate in time, and univariate in state has been discussed in both chapter 1 and chapter 2 and is used for example in imaging and spatial methods. This chapter concentrates on the former multivariate setting. Section 1 starts from a univariate time, bivariate state Karhunen-Loeve expansion. Examples, like the Brownian motion and the Ornstein-Uhlenbeck process will be provided. Section 2 extends from the univariate time, bivariate state to the univariate time, multivariate state and presents the main theorem of the multivariate Karhunen-Loeve expansion. Section 3 explains the numerical methods in the multivariate setting. Section 4 applies a theorem in section 3 and further relates it to the linear stochastic differential equation (SDE) in the narrow sense.

5.1 Bivariate Karhunen-Loeve expansion

Assume that there are two correlated time series $X(t)$ and $Y(t)$. Although the analysis is harder than the analysis for one dimension, the problem of finding the Karhunen-Loeve expansion can be reduced to a one-dimensional problem under certain special condition (equation 5.2.21 and equation 5.2.22). We refer to this method of reducing the problem to the univariate setting as the "lining-up" method. For convenience, we assume the special condition is satisfied and only consider the time interval $\mathcal{T}$ starting from 0 with length T. It can be easily generalised to any interval with any starting point.

Define a new series $U(t)$ with $X(t)$ as its first part and $Y(t)$ its second part, i.e.

$$U(t) = \begin{cases} X(t) & 0 \leq t < T \\ Y(t-T) & T \leq t < 2T \end{cases} \tag{5.1.1}$$

Hence the covariance function for $\{U(t), t \in [0, 2T]\}$ can be written down as

$$K(t,s) = \text{cov}(U(t), U(s)) = \begin{cases} \text{cov}(X(t), X(s)) & 0 \leq t, s < T \\ \text{cov}(X(t), Y(s-T)) & 0 \leq t < T \leq s < 2T \\ \text{cov}(X(s), Y(t-T)) & 0 \leq s < T \leq t < 2T \\ \text{cov}(Y(t-T), Y(s-T)) & T \leq t, s < 2T \end{cases} \tag{5.1.2}$$

The Fredholm integral equation is $\int_0^{2T} K(t,s)\phi(s)ds = \lambda\phi(t)$ with eigenfunction $\phi(s)$ defined over two intervals $[0, T]$ and $[T, 2T]$ separately.

$$\phi(s) = \begin{cases} f(s) & 0 \leq s < T \\ g(s-T) & T \leq s < 2T \end{cases}$$

Depending on the location of t, there are two integral equations which need to be solved. When $0 \leq t < T$,

$$\int_0^{2T} K(t,s)\phi(s)ds = \lambda\phi(t) \tag{5.1.3}$$

Using the expression of $K(t,s)$ in equation 5.1.2, equation 5.1.3 can be reduced to

$$\int_0^T \text{cov}(X(t),X(s))f(s)ds + \int_0^T \text{cov}(X(t),Y(s))g(s)ds = \lambda f(t) \tag{5.1.4}$$

When $T \leq t < 2T$,

$$\int_0^{2T} K(t,s)\phi(s)ds = \lambda\phi(t) \tag{5.1.5}$$

Using the expression of $K(t,s)$ in equation 5.1.2 again, equation 5.1.5 can be simplified as

$$\int_0^T \text{cov}(X(s),Y(t-T))f(s)ds + \int_0^T \text{cov}(Y(s),Y(t-T))g(s)ds = \lambda g(t-T) \tag{5.1.6}$$

Since $T \leq t < 2T$, $0 \leq t-T < T$. Hence equation 5.1.6 is equivalent to

$$\int_0^T \text{cov}(X(s),Y(t))f(s)ds + \int_0^T \text{cov}(Y(s),Y(t))g(s)ds = \lambda g(t) \tag{5.1.7}$$

for $0 \leq t < T$. The orthogonal condition in the bivariate setting is

$$\int_0^{2T} \phi_i(s)\phi_j(s)ds = \delta_{ij} \tag{5.1.8}$$

Equation 5.1.8 is equivalent to

$$\int_0^T f_i(s)f_j(s)ds + \int_0^T g_i(s)g_j(s)ds = \delta_{ij} \tag{5.1.9}$$

After solving the integral equation 5.1.4 and 5.1.7, we can use Mercer's theorem to decompose the covariance function. Assuming that the covariance function is continuous after lining up the process, we have

$$\begin{aligned} K(t,s) &= \sum_i \lambda_i\phi_i(t)\phi_i(s) \\ &= \begin{cases} \text{cov}(X(t),X(s)) = \sum_i \lambda_i f_i(t)f_i(s) & 0 \leq t,s < T \\ \text{cov}(X(t),Y(s-T)) = \sum_i \lambda_i g_i(t)f_i(s-T) & 0 \leq t < T \leq s < 2T \\ \text{cov}(X(s),Y(t-T)) = \sum_i \lambda_i g_i(t-T)f_i(s) & 0 \leq s < T \leq t < 2T \\ \text{cov}(Y(t-T),Y(s-T)) = \sum_i \lambda_i g_i(t-T)g_i(s-T) & T \leq t,s < 2T \end{cases} \end{aligned}$$

Or for time $0 \leq s, t < T$

$$\text{cov}(X(t), X(s)) = \sum_i \lambda_i f_i(t) f_i(s) \quad \text{cov}(X(t), Y(s)) = \sum_i \lambda_i f_i(t) g_i(s)$$
$$\text{cov}(X(s), Y(t)) = \sum_i \lambda_i g_i(t) f_i(s) \quad \text{cov}(Y(t), Y(s)) = \sum_i \lambda_i g_i(t) g_i(s)$$

In summary, for the bivariate state, univariate time process $X(t)$ and $Y(t)$, $t \in [0, T]$, the bivariate Karhunen-Loeve expansion for them can be written down as

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \sum_{i \geq 1} \sqrt{\lambda_i} \begin{pmatrix} f_i(t) \\ g_i(t) \end{pmatrix} \xi_i$$

where $\xi_i \sim i.i.d.N(0, 1)$, λ_i, f_i and g_i are derived from the matrix form of the Fredholm integral equation

$$\int_0^T \begin{pmatrix} \text{cov}(X(t), X(s)) & \text{cov}(X(t), Y(s)) \\ \text{cov}(X(s), Y(t)) & \text{cov}(Y(t), Y(s)) \end{pmatrix} \begin{pmatrix} f_i(s) \\ g_i(s) \end{pmatrix} ds = \lambda_i \begin{pmatrix} f_i(t) \\ g_i(t) \end{pmatrix}$$

with the orthogonality condition

$$\int_0^T \begin{pmatrix} f_i(s), & g_i(s) \end{pmatrix} \begin{pmatrix} f_j(s) \\ g_j(s) \end{pmatrix} ds = \delta_{ij} \tag{5.1.10}$$

Mercer's theorem can now be expressed as

$$\begin{pmatrix} \text{cov}(X(t), X(s)) & \text{cov}(X(t), Y(s)) \\ \text{cov}(X(s), Y(t)) & \text{cov}(Y(t), Y(s)) \end{pmatrix} = \sum_i \lambda_i \begin{pmatrix} f_i(t) \\ g_i(t) \end{pmatrix} \begin{pmatrix} f_i(s) & g_i(s) \end{pmatrix} \tag{5.1.11}$$

Notice that in the bivariate case, the eigenfunctions are a two dimensional column vector but the eigenvalues are still scalars.

Example 1: The bivariate Brownian motion

Assume that there are two correlated Brownian motion $\{X(t), t \in [0, 1]\}$ and $\{Y(t), t \in [0, 1]\}$ with the covariance matrix defined as follows. For $\rho > 0$,

$$\begin{pmatrix} \text{cov}(X(t), X(s)) & \text{cov}(X(t), Y(s)) \\ \text{cov}(X(s), Y(t)) & \text{cov}(Y(t), Y(s)) \end{pmatrix} = \begin{pmatrix} \min(t, s) & \rho \min(t, s) \\ \rho \min(s, t) & \min(t, s) \end{pmatrix} \tag{5.1.12}$$

The bivariate Fredholm integration equation is

$$\int_0^1 \begin{pmatrix} \min(t,s) & \rho\min(t,s) \\ \rho\min(s,t) & \min(t,s) \end{pmatrix} \begin{pmatrix} f(s) \\ g(s) \end{pmatrix} ds = \lambda \begin{pmatrix} f(t) \\ g(t) \end{pmatrix} \tag{5.1.13}$$

Decomposing $\min(t,s)$, equation 5.1.13 can be simplified as

$$\begin{cases} \int_0^t sf(s) + \rho sg(s)ds + t\int_t^1 f(s) + \rho g(s)ds = \lambda f(t) \\ \int_0^t \rho sf(s) + sg(s)ds + t\int_t^1 \rho f(s) + g(s)ds = \lambda g(t) \end{cases} \tag{5.1.14}$$

with the boundary condition

$$f(0) = g(0) = 0 \tag{5.1.15}$$

Differentiate equation 5.1.14 once,

$$\begin{cases} \int_t^1 f(s) + \rho g(s)ds = \lambda f'(t) \\ \int_t^1 \rho f(s) + g(s)ds = \lambda g'(t) \end{cases}$$

Differentiating equation 5.1.14 one more time results in the following differential equations.

$$\begin{cases} -f(t) - \rho g(t) = \lambda f''(t) \\ -\rho f(t) - g(t) = \lambda g''(t) \end{cases} \tag{5.1.16}$$

Solving these differential equations provides the functional form of $f(t)$ and $g(t)$

$$\begin{cases} f(t) = c_1\cos(\sqrt{\frac{1-\rho}{\lambda}}t) + c_2\sin(\sqrt{\frac{1-\rho}{\lambda}}t) + c_3\cos(\sqrt{\frac{1+\rho}{\lambda}}t) + c_4\sin(\sqrt{\frac{1+\rho}{\lambda}}t) \\ g(t) = -c_1\cos(\sqrt{\frac{1-\rho}{\lambda}}t) - c_2\sin(\sqrt{\frac{1-\rho}{\lambda}}t) + c_3\cos(\sqrt{\frac{1+\rho}{\lambda}}t) + c_4\sin(\sqrt{\frac{1+\rho}{\lambda}}t) \end{cases} \tag{5.1.17}$$

Considering the boundary conditions in equation 5.1.15,

$$\left.\begin{array}{l} f(0) = 0 \Rightarrow c_1 + c_3 = 0 \\ g(0) = 0 \Rightarrow -c_1 + c_3 = 0 \end{array}\right\} \Longrightarrow c_1 = c_3 = 0 \tag{5.1.18}$$

Hence the solution to $f(t)$ and $g(t)$ can be further simplified to

$$\begin{cases} f(t) = c_2\sin(\sqrt{\frac{1-\rho}{\lambda}}t) + c_4\sin(\sqrt{\frac{1+\rho}{\lambda}}t) \\ g(t) = -c_2\sin(\sqrt{\frac{1-\rho}{\lambda}}t) + c_4\sin(\sqrt{\frac{1+\rho}{\lambda}}t) \end{cases} \tag{5.1.19}$$

Substitute the above solutions into the first differential equations to find eigenvalues

$$\begin{cases} c_2(\rho-1)\sqrt{\frac{1+\rho}{\lambda}}\cos(\sqrt{\frac{1-\rho}{\lambda}}) + c_4(\rho+1)\sqrt{\frac{1-\rho}{\lambda}}\cos(\sqrt{\frac{1+\rho}{\lambda}}) = 0 \\ c_2(\rho-1)\sqrt{\frac{1+\rho}{\lambda}}\cos(\sqrt{\frac{1-\rho}{\lambda}}) - c_4(\rho+1)\sqrt{\frac{1-\rho}{\lambda}}\cos(\sqrt{\frac{1+\rho}{\lambda}}) = 0 \end{cases} \tag{5.1.20}$$

Solving equation 5.1.20 results in two possible solutions.

$$\begin{cases} c_2 = 0, \cos(\sqrt{\frac{1+\rho}{\lambda}}) = 0 \quad \text{or} \\ c_4 = 0, \cos(\sqrt{\frac{1-\rho}{\lambda}}) = 0 \end{cases} \tag{5.1.21}$$

Thus there are also two possible solutions to the eigenvalue λ and the eigenfunction $\begin{pmatrix} f(t) \\ g(t) \end{pmatrix}$. Denote the jth possible solution to the ith eigenvalue and the $i$the eigenfunction as $\lambda_{j,i}$ and $\begin{pmatrix} f_{j,i}(t) \\ g_{j,i}(t) \end{pmatrix}$, where $j = 1, 2$ and $i \geq 1$, then equation 5.1.21 and equation 5.1.19 are equivalent to

$$\begin{cases} c_2 = 0, \lambda_{1,i} = \frac{4(1+\rho)}{(2i-1)^2\pi^2}, f_{1,i}(t) = c_4 \sin\frac{(2i-1)\pi t}{2}, g_{1,i}(t) = c_4 \sin\frac{(2i-1)\pi t}{2} \quad \text{or} \\ c_4 = 0, \lambda_{2,i} = \frac{4(1-\rho)}{(2i-1)^2\pi^2}, f_{2,i}(t) = c_2 \sin\frac{(2i-1)\pi t}{2}, g_{2,i}(t) = -c_2 \sin\frac{(2i-1)\pi t}{2} \end{cases} \tag{5.1.22}$$

Using the orthogonality condition

$$\int_0^1 f_i^2(t) + g_i^2(t)dt = 1 \tag{5.1.23}$$

and the result that

$$\int_0^1 [\sin\frac{(2i-1)\pi t}{2}]^2 dt = \frac{1}{2} \ , \tag{5.1.24}$$

The value for the other coefficient can be obtained, i.e. when $c_2 = 0$, $c_4 = 1$ or when $c_2 = 1$, $c_4 = 0$.

Hence, the complete solution to the original integral equation is

$$\begin{cases} \lambda_{1,i} = \frac{4(1+\rho)}{(2i-1)^2\pi^2}, f_{1,i}(t) = \sin\frac{(2i-1)\pi t}{2}, g_{1,i}(t) = \sin\frac{(2i-1)\pi t}{2} \quad \text{or} \\ \lambda_{2,i} = \frac{4(1-\rho)}{(2i-1)^2\pi^2}, f_{2,i}(t) = \sin\frac{(2i-1)\pi t}{2}, g_{2,i}(t) = -\sin\frac{(2i-1)\pi t}{2} \end{cases} \tag{5.1.25}$$

In short, the bivariate Karhunen-Loeve expansion to the bivariate Brownian motion can be summarised in lemma 5.1.1.

Lemma 5.1.1. *The bivariate Karhunen-Loeve expansion to the bivariate Brownian motion* $\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}$ *with the covariance matrix*

$$\begin{pmatrix} cov(X(t),X(s)) & cov(X(t),Y(s)) \\ cov(X(s),Y(t)) & cov(Y(t),Y(s)) \end{pmatrix} = \begin{pmatrix} min(t,s) & \rho min(t,s) \\ \rho min(s,t) & min(t,s) \end{pmatrix} \tag{5.1.26}$$

can be expressed as

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \sum_{i\geq 1} \frac{2\sqrt{1+\rho}}{(2i-1)\pi} \begin{pmatrix} sin\frac{(2i-1)\pi t}{2} \\ sin\frac{(2i-1)\pi t}{2} \end{pmatrix} \xi_i + \sum_{i\geq 1} \frac{2\sqrt{1-\rho}}{(2i-1)\pi} \begin{pmatrix} sin\frac{(2i-1)\pi t}{2} \\ -sin\frac{(2i-1)\pi t}{2} \end{pmatrix} \xi_i' ,$$

or

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \sum_{i\geq 1} \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} \xi_i \\ \xi_i' \end{pmatrix} \frac{2\sqrt{2}}{(2i-1)\pi} sin\frac{(2i-1)\pi t}{2} , \tag{5.1.27}$$

where ξ_i *and* ξ_i' *are mutually independent standard normal random variables.*

Also, $\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}$ *can be written in terms of two independent univariate Brownian motion* $B_1(t)$ *and* $B_2(t)$

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} B_1(t) \\ B_2(t) \end{pmatrix} \tag{5.1.28}$$

Example 2: The bivariate Ornstein-Uhlenbeck process

Assume that there are two correlated Ornstein-Uhlenbeck process $\{X(t), t \in [0,1]\}$ and $\{Y(t), t \in [0,1]\}$, i.e. for $\beta > 0$

$$dX(t) = -\beta X(t)dt + rdW_1(t)$$

$$dY(t) = -\beta Y(t)dt + rdW_2(t)$$

with

$$\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} B_1(t) \\ B_2(t) \end{pmatrix} \tag{5.1.29}$$

where $B_1(t)$ and $B_2(t)$ are two independent Brownian motion.

Assume further that the initial conditions for $X(t)$ and $Y(t)$ are: $X(0) \sim N(0, \frac{r^2}{2\beta})$, $Y(t) \sim N(0, \frac{r^2}{2\beta})$ and $\text{cov}(X(0), Y(0)) = \rho\frac{r^2}{2\beta}$ $(\rho > 0)$. The covariance matrix for $X(t)$ and $Y(t)$ is expressed as

$$\begin{pmatrix} \text{cov}(X(t), X(s)) & \text{cov}(X(t), Y(s)) \\ \text{cov}(X(s), Y(t)) & \text{cov}(Y(t), Y(s)) \end{pmatrix} = \frac{r^2}{2\beta} \begin{pmatrix} \exp(-\beta|t-s|) & \rho\exp(-\beta|t-s|) \\ \rho\exp(-\beta|s-t|) & \exp(-\beta|t-s|) \end{pmatrix} \tag{5.1.30}$$

Except for a constant term $\frac{r^2}{2\beta}$, the Fredholm integral equation is

$$\int_0^1 \begin{pmatrix} \exp(-\beta|t-s|) & \rho\exp(-\beta|t-s|) \\ \rho\exp(-\beta|s-t|) & \exp(-\beta|t-s|) \end{pmatrix} ds = \lambda \begin{pmatrix} f(t) \\ g(t) \end{pmatrix} \tag{5.1.31}$$

This is equivalent to solving

$$\begin{cases} \int_0^t \exp[-\beta(t-s)]f(s) + \rho\exp[-\beta(t-s)]g(s)ds \\ + \int_t^1 \exp[-\beta(s-t)]f(s) + \rho\exp[-\beta(s-t)]g(s)ds = \lambda f(t) \\ \int_0^t \rho\exp[-\beta(t-s)]f(s) + \exp[-\beta(t-s)]g(s)ds \\ + \int_t^1 \rho\exp[-\beta(s-t)]f(s) + \exp[-\beta(s-t)]g(s)ds = \lambda g(t) \end{cases}$$

Differentiating the above equations once, we obtain

$$\begin{cases} -\beta \int_0^t \exp[-\beta(t-s)]f(s) + \rho\exp[-\beta(t-s)]g(s)ds \\ +\beta \int_t^1 \exp[-\beta(s-t)]f(s) + \rho\exp[-\beta(s-t)]g(s) = \lambda f'(t) \\ -\beta \int_0^t \rho\exp[-\beta(t-s)]f(s) + \exp[-\beta(t-s)]g(s)ds \\ +\beta \int_t^1 \rho\exp[-\beta(s-t)]f(s) + \exp[-\beta(s-t)]g(s)ds = \lambda g'(t) \end{cases} \tag{5.1.32}$$

Differentiating one more time results in

$$\begin{cases} (\lambda\beta^2 - 2\beta)f(t) - 2\rho\beta g(t) = \lambda f''(t) \Longrightarrow \frac{2\beta - \lambda\beta^2}{\lambda} f(t) + \frac{2\rho\beta}{\lambda} g(t) + f''(t) = 0 \\ (\lambda\beta^2 - 2\beta)g(t) - 2\rho\beta f(t) = \lambda g''(t) \Longrightarrow \frac{2\beta - \lambda\beta^2}{\lambda} g(t) + \frac{2\rho\beta}{\lambda} f(t) + g''(t) = 0 \end{cases} \tag{5.1.33}$$

Define

$$w_1^2 = \frac{2\beta - \lambda\beta^2 + 2\rho\beta}{\lambda} \tag{5.1.34}$$

$$w_2^2 = \frac{2\beta - \lambda\beta^2 - 2\rho\beta}{\lambda} \tag{5.1.35}$$

Then equation 5.1.33 is equivalent to

$$\begin{cases} \frac{w_1^2+w_2^2}{2} f(t) + \frac{w_1^2-w_2^2}{2} g(t) + f''(t) = 0 \\ \frac{w_1^2+w_2^2}{2} g(t) + \frac{w_1^2-w_2^2}{2} f(t) + g''(t) = 0 \end{cases} \tag{5.1.36}$$

Hence solutions to $f(t)$ and $g(t)$ are

$$\begin{cases} f(t) = c_1\sin(w_2 t) + c_2\cos(w_2 t) + c_3\sin(w_1 t) + c_4\cos(w_1 t) \\ g(t) = -c_1\sin(w_2 t) - c_2\cos(w_2 t) + c_3\sin(w_1 t) + c_4\cos(w_1 t) \end{cases} \tag{5.1.37}$$

with the boundary conditions,

$$\begin{cases} f'(0) - \beta f(0) = 0, \;\; g'(0) - \beta g(0) = 0 \\ f'(1) + \beta f(1) = 0, \;\; g'(1) + \beta g(1) = 0 \end{cases}$$

Substitute the functional form of $f(t)$ and $g(t)$ into the boundary conditions.

$$\begin{cases} c_1(w_2\cos(w_2) + \beta\sin(w_2)) + c_2(-w_2\sin(w_2) + \beta\cos(w_2)) \\ +c_3(w_1\cos(w_1) + \beta\sin(w_1)) + c_4(-w_1\sin(w_1) + \beta\cos(w_1)) = 0 \\ -c_1(w_2\cos(w_2) + \beta\sin(w_2)) - c_2(-w_2\sin(w_2) + \beta\cos(w_2)) \\ +c_3(w_1\cos(w_1) + \beta\sin(w_1)) + c_4(-w_1\sin(w_1) + \beta\cos(w_1)) = 0 \\ -c_1 w_2 + c_2\beta + c_3 w_1 - c_4\beta = 0 \\ c_1 w_2 - c_2\beta + c_3 w_1 - c_4\beta = 0 \end{cases} \tag{5.1.38}$$

In order to have non-zero solutions, the determinant of the following matrix should be equal to zero.

$$\begin{vmatrix} w_2\cos(w_2) + \beta\sin(w_2) & -w_2\sin(w_2) + \beta\cos(w_2) & w_1\cos(w_1) + \beta\sin(w_1) & -w_1\sin(w_1) + \beta\cos(w_1) \\ -w_2\cos(w_2) - \beta\sin(w_2) & w_2\sin(w_2) - \beta\cos(w_2) & w_1\cos(w_1) + \beta\sin(w_1) & -w_1\sin(w_1) + \beta\cos(w_1) \\ -w_2 & \beta & w_1 & -\beta \\ w_2 & -\beta & w_1 & -\beta \end{vmatrix} = 0$$

Calculating the above determinant results in the following equation

$$4(2\beta w_2 \cos(w_2) + \beta^2 \sin(w_2) - w_2^2 \sin(w_2))(\beta^2 \sin(w_1) + 2\beta w_1 \cos(w_1) - w_1^2 \sin(w_1)) = 0 \tag{5.1.39}$$

This means that

$$\beta^2 \sin(w_1) + 2\beta w_1 \cos(w_1) - w_1^2 \sin(w_1) = 0 \quad \text{or,} \tag{5.1.40}$$

$$2\beta w_2 \cos(w_2) + \beta^2 \sin(w_2) - w_2^2 \sin(w_2) = 0 \tag{5.1.41}$$

Solving equation 5.1.40 is the same as setting $c_1 = c_2 = 0$. The reason is as follows. When $c_1 = c_2 = 0$, equation 5.1.38 can be reduced to

$$\begin{cases} c_3(w_1 \cos(w_1) + \beta \sin(w_1)) + c_4(-w_1 \sin(w_1) + \beta \cos(w_1)) = 0 \\ c_3 w_1 - c_4 \beta = 0 \end{cases} \tag{5.1.42}$$

For non-zero solutions, the determinant of equation 5.1.42 is zero, i.e.

$$\begin{vmatrix} w_1 \cos(w_1) + \beta \sin(w_1) & -w_1 \sin(w_1) + \beta \cos(w_1) \\ w_1 & -\beta \end{vmatrix} = \beta^2 \sin(w_1) + 2\beta w_1 \cos(w_1) - w_1^2 \sin(w_1) = 0 \tag{5.1.43}$$

Equation 5.1.43 is equivalent to equation 5.1.40.

Similarly, it can also be shown that solving equation 5.1.41 is the same as setting $c_3 = c_4 = 0$.

Since equation 5.1.40 and equation 5.1.41 are essentially the same and both of them are equivalent to equation 1.3.30 in the univariate case, the solution to w_1 and w_2 should also be the same as that of equation 1.3.30. Notice also that if the ith smallest solution to w_1 and w_2 is denoted as $w_i = w_{1,i} = w_{2,i}$, $i \geq 1$, w_i can be derived through a simplified version of either equation 5.1.40 or equation 5.1.41,

$$\cot(w_i) = \frac{w_i^2 - \beta^2}{2\beta w_i} \tag{5.1.44}$$

Please refer to section2.3, chapter 2 for the detailed illustration of finding w_i.

In terms of eigenvalues, after obtaining w_i and applying equation 5.1.34 and equation 5.1.35, the ith eigenvalue is

$$\lambda_{1,i} = \frac{2\beta(1+\rho)}{w_i^2+\beta^2} \tag{5.1.45}$$

$$\lambda_{2,i} = \frac{2\beta(1-\rho)}{w_i^2+\beta^2} \tag{5.1.46}$$

As to eigenfunctions, the ith eigenfunction related to the ith eigenvalue $\lambda_{1,i}$ is denoted as $\begin{pmatrix} f_{1,i}(t) \\ g_{1,i}(t) \end{pmatrix}$. Using equation 5.1.37 and the condition that $c_1 = c_2 = 0$ and $c_3 = \frac{\beta}{w_1}c_4$,

$$\begin{cases} f_{1,i}(t) = \frac{\beta}{w_i}c_4\sin(w_i t) + c_4\cos(w_i t) \\ g_{1,i}(t) = \frac{\beta}{w_i}c_4\sin(w_i t) + c_4\cos(w_{1,i} t) \end{cases} \tag{5.1.47}$$

The unknown coefficient c_4 can be solved using the orthogonality condition, as well as the fact that $f_{1,i}^2(t) = g_{1,i}^2(t)$. Since $\int_0^1 f_{1,i}^2(t) + g_{1,i}^2(t)dt = 1$

$$\int_0^1 f_{1,i}^2(t)dt = \frac{1}{2} \tag{5.1.48}$$

Hence, the solution to c_4, $f_{1,i}(t)$ and $g_{1,i}(t)$ are

$$c_4 = \sqrt{\frac{w_i^2}{2\beta + w_i^2 + \beta^2}} \tag{5.1.49}$$

$$f_{1,i}(t) = g_{1,i}(t) = \sqrt{\frac{w_i^2}{2\beta + w_i^2 + \beta^2}}\cos(w_i t) + \sqrt{\frac{\beta^2}{2\beta + w_i^2 + \beta^2}}\sin(w_i t) \ , \tag{5.1.50}$$

where w_i solves the equation $\cot(w_i) = \frac{w_i^2-\beta^2}{2\beta w_i}$

Similarly, the ith eigenfunction related to the ith eigenvalue $\lambda_{2,i}$ is denoted as $\begin{pmatrix} f_{2,i}(t) \\ g_{2,i}(t) \end{pmatrix}$. Following the same calculation procedure for finding $f_{1,i}(t)$ and $g_{1,i}(t)$, the value for $f_{2,i}(t)$ and $g_{2,i}(t)$ are

$$f_{2,i}(t) = -g_{2,i}(t) = \sqrt{\frac{w_i^2}{2\beta + w_i^2 + \beta^2}}\cos(w_i t) + \sqrt{\frac{\beta^2}{2\beta + w_i^2 + \beta^2}}\sin(w_i t) \ , \tag{5.1.51}$$

where w_i solves the equation $\cot(w_i) = \frac{w_i^2-\beta^2}{2\beta w_i}$

In short, the bivariate Karhunen-Loeve expansion to the bivariate Ornstein-Uhlenbeck process can be summarised in lemma 5.1.2.

Lemma 5.1.2. *The bivariate Karhunen-Loeve expansion to the bivariate Ornstein-Uhlenbeck process* $\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}$ *with the covariance matrix*

$$\begin{pmatrix} cov(X(t),X(s)) & cov(X(t),Y(s)) \\ cov(X(s),Y(t)) & cov(Y(t),Y(s)) \end{pmatrix} = \frac{r^2}{2\beta} \begin{pmatrix} exp(-\beta|t-s|) & \rho exp(-\beta|t-s|) \\ \rho exp(-\beta|s-t|) & exp(-\beta|t-s|) \end{pmatrix} \tag{5.1.52}$$

can be expressed as

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \sum_{i\geq 1} \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} \xi_i \\ \xi_i' \end{pmatrix} r\sqrt{\frac{1}{w_i^2+\beta_i^2}} \left[\sqrt{\frac{2w_i^2}{2\beta+w_i^2+\beta^2}}cos(w_i t) + \sqrt{\frac{2\beta^2}{2\beta+w_i^2+\beta^2}}sin(w_i t)\right] ,$$

where ξ_i *and* ξ_i' *are mutually independent standard normal random variables and* w_i *solves the equation*

$$cot(w_i) = \frac{w_i^2-\beta^2}{2\beta w_i} \tag{5.1.53}$$

The bivariate process $\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}$ *can also be expressed as a linear combination of two independent Ornstein-Uhlenbeck processes* $X'(t)$ *and* $Y'(t)$*. That is*

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} X'(t) \\ Y'(t) \end{pmatrix} , \tag{5.1.54}$$

where

$$\begin{aligned} dX^{'}(t) &= -\beta X^{'}(t)dt + rdB_1(t) \\ dY^{'}(t) &= -\beta Y^{'}(t)dt + rdB_2(t) \ , \end{aligned}$$

with initial condition $X^{'}(0), Y^{'}(0)$ *i.i.d.* $\sim N(0, \frac{r^2}{2\beta})$ *and* $B_1(t)$ *and* $B_2(t)$ *are two independent Brownian motion.*

In the above two examples, the bivariate Brownian motion and the bivariate Ornstein-Uhlenbeck process share almost the same covariance structure, which is

$$\mathrm{cov}(X_1(t), X_1(s)) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{5.1.55}$$

This covariance structure results in the same Karhunen-Loeve expansion for both of the processes, which is

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} X^{'}(t) \\ Y^{'}(t) \end{pmatrix} \ , \tag{5.1.56}$$

where $X(t)$ and $Y(t)$ are the correlated processes, while $X^{'}(t)$ and $Y^{'}(t)$ are the corresponding independent processes. This is not a coincidence. The following theorem guarantees that equation 5.1.56 is the right Karhunen-Loeve expansion for the bivariate process with the covariance structure mentioned in equation 5.1.55.

Theorem 5.1.3. *Assume that* $X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}$ *is a bivariate process with the covariance matrix* $cov(X_1(t), X_1(s)) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, *where* $t \in \mathcal{T}$ *and* $0 < \rho < 1$. *If the univariate Karhunen-Loeve expansion for* $X_1(t)$ *and* $X_2(t)$ *is*

$$X_1(t) = X_2(t) = \sum_{i \geq 1} \sqrt{\lambda_i}\phi_i(t)\tau_i \quad \tau_i \sim i.i.d. N(0,1)$$

there exists a bivariate Karhunen-Loeve expansion for $X(t)$ *expressed as*

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \sum_{i\geq 1} \sqrt{(1+\rho)\lambda_i} \begin{pmatrix} \frac{\phi_i(t)}{\sqrt{2}} \\ \frac{\phi_i(t)}{\sqrt{2}} \end{pmatrix} \xi_i + \sum_{i\geq 1} \sqrt{(1-\rho)\lambda_i} \begin{pmatrix} \frac{\phi_i(t)}{\sqrt{2}} \\ -\frac{\phi_i(t)}{\sqrt{2}} \end{pmatrix} \xi_i' \ ,$$

where ξ_i *and* ξ_i' *are the mutually independent standard normal random variables.*

If λ_i^{new}, $f_i(t)$ *and* $g_i(t)$ *represent the* i*th eigenvalues for the bivariate expansion, the* i*th eigenfunction for* $X_1(t)$ *and the* i*th eigenfunction for* $X_2(t)$ *respectively, the above expansion means that*

$$\lambda_i^{new} = (1+\rho)\lambda_i \qquad f_i(t) = \frac{\phi_i(t)}{\sqrt{2}} \qquad g_i(t) = \frac{\phi_i(t)}{\sqrt{2}} \qquad or$$
$$\lambda_i^{new} = (1-\rho)\lambda_i \qquad f_i(t) = \frac{\phi_i(t)}{\sqrt{2}} \qquad g_i(t) = -\frac{\phi_i(t)}{\sqrt{2}}$$

Proof. The bivariate Fredholm integral equation in this case is

$$\int_{\mathcal{T}} \mathrm{cov}(X_1(t), X_1(s)) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} f_i(s) \\ g_i(s) \end{pmatrix} ds = \lambda_i^{\mathrm{new}} \begin{pmatrix} f_i(t) \\ g_i(t) \end{pmatrix} \tag{5.1.57}$$

This is equivalent to

$$\int_{\mathcal{T}} K(t,s) f_i(s) ds + \rho \int_{\mathcal{T}} K(t,s) g_i(s) ds = \lambda_i^{\mathrm{new}} f_i(t) \tag{5.1.58}$$
$$\rho \int_{\mathcal{T}} K(t,s) f_i(s) ds + \int_{\mathcal{T}} K(t,s) g_i(s) ds = \lambda_i^{\mathrm{new}} g_i(t) \ , \tag{5.1.59}$$

where $K(t,s) = \mathrm{cov}(X_1(t), X_1(s))$.

Subtract equation 5.1.58 multiplied by ρ from equation 5.1.59, we obtain

$$(1-\rho^2) \int_{\mathcal{T}} K(t,s) g_i(s) ds = \lambda_i^{\mathrm{new}} g_i(t) - \lambda_i^{\mathrm{new}} \rho f_i(t) \tag{5.1.60}$$

Similarly, subtract equation 5.1.59 multiplied by ρ from equation 5.1.58, we obtain

$$(1-\rho^2) \int_{\mathcal{T}} K(t,s) f_i(s) ds = \lambda_i^{\mathrm{new}} f_i(t) - \lambda_i^{\mathrm{new}} \rho g_i(t) \tag{5.1.61}$$

Add equation 5.1.60 and equation 5.1.61.

$$(1+\rho)\int_{\mathcal{T}} K(t,s)[f_i(s)+g_i(s)]ds = \lambda_i^{\text{new}}[f_i(t)+g_i(t)] \tag{5.1.62}$$

Subtract equation 5.1.60 from equation 5.1.61.

$$(1-\rho)\int_{\mathcal{T}} K(t,s)[f_i(s)-g_i(s)]ds = \lambda_i^{\text{new}}[f_i(t)-g_i(t)] \tag{5.1.63}$$

Both equation 5.1.62 and equation 5.1.63 can be now treated as the Fredholm integral equation for the univariate process with kernel function $K(t,s)$. Therefore, the eigenvalues and the eigenfunctions can now be derived using the result of the univariate process. In terms of the eigenvalues, from equation 5.1.62, $\lambda_i^{\text{new}} = (1+\rho)\lambda_i$, while from equation 5.1.63, $\lambda_i^{\text{new}} = (1-\rho)\lambda_i$. Since both of the eigenvalues should solve the integral equation 5.1.57, the possible solutions to the eigenfunctions, as well as the eigenvalues, should be

$$\lambda_i^{\text{new}} = (1+\rho)\lambda_i, \quad f_i(t)+g_i(t) = c_1\phi_i(t), \quad f_i(t)-g_i(t) = 0 \quad \text{or}$$
$$\lambda_i^{\text{new}} = (1-\rho)\lambda_i, \quad f_i(t)-g_i(t) = c_2\phi_i(t), \quad f_i(t)+g_i(t) = 0 \ ,$$

where c_1 and c_2 are constants. Both c_1 and c_2 can be derived using the orthogonality condition

$$\int_{\mathcal{T}} f_i(t)f_j(t)+g_i(t)g_j(t)dt = \delta_{ij} \tag{5.1.64}$$

When $\lambda_i^{\text{new}} = (1+\rho)\lambda_i$, $f_i(t) = g_i(t) = \frac{c_1}{2}\phi_i(t)$, hence

$$1 = \int_{\mathcal{T}} f_i^2(t)+g_i^2(t)dt = \frac{c_1^2}{2} \tag{5.1.65}$$

Hence $c_1 = \sqrt{2}$. Similarly, $c_2 = \sqrt{2}$. Therefore, the ith eigenvalue and the ith eigenfunction are

$$\lambda_i^{\text{new}} = (1+\rho)\lambda_i \qquad f_i(t) = \frac{\phi_i(t)}{\sqrt{2}} \qquad g_i(t) = \frac{\phi_i(t)}{\sqrt{2}} \qquad \text{or}$$
$$\lambda_i^{\text{new}} = (1-\rho)\lambda_i \qquad f_i(t) = \frac{\phi_i(t)}{\sqrt{2}} \qquad g_i(t) = -\frac{\phi_i(t)}{\sqrt{2}}$$

□

In the next section, theorem 5.1.3 will be extended to a more complicated multivariate setting.

5.2 Multivariate Karhunen-Loeve expansions

In this section, the results from the univariate time, bivariate state, Karhunen-Loeve expansion are extended to the univariate time, multivariate state, Karhunen-Loeve expansion. The idea behind both is the same. Assume $\mathbf{X}(t)^T = \Big(X_1(t),\ X_2(t),\ \cdots,\ X_d(t)\Big)$ is a d-dimensional stochastic process. Each $X_i(t)$ is defined on the interval $[0, T]$. $\mathbf{X}(t)$ is further assumed to lie in the multivariate $H_{\mathbf{X}}$ space equipped with the finite energy $\sum_{i=1}^{d} E(X_i^2(t)) < \infty$. Its inner product is defined as

$$< \mathbf{X}(t), \mathbf{Y}(t) >= E(\mathbf{X}(t)^T \mathbf{Y}(t)) = \sum_{i=1}^{d} E[X_i(t) Y_i(t)] \tag{5.2.1}$$

given $\mathbf{X}(t), \mathbf{Y}(t) \in H_{\mathbf{X}}$. Using properties of the expectation, the above inner product can be shown as a valid inner product. In the multivariate setting, assume that the eigenfunction for each $X_i(t)$ is $f^{(i)}(t)$.

Under certain condition (equation 5.2.21 and equation 5.2.22), which assumes continuity of the covariance function after lining up the process, the lining-up method is used again to this d-dimensional process and a new stochastic process $U(t)$ is formed and defined on the interval $[0, dT]$. The corresponding eigenfunction for $U(t)$ can be

defined in the same way as that in the bivariate case.

$$U(t) = \begin{cases} X_1(t) & 0 \le t < T \\ \cdots & \\ X_i(t-(i-1)T) & (i-1)T \le t < iT \\ \cdots & \\ X_d(t-(d-1)T) & (d-1)T \le t < dT \end{cases} \tag{5.2.2}$$

$$\phi(t) = \begin{cases} f^{(1)}(t) & 0 \le t < T \\ \cdots & \\ f^{(i)}(t-(i-1)T) & (i-1)T \le t < iT \\ \cdots & \\ f^{(d)}(t-(d-1)T) & (d-1)T \le t < dT \end{cases} \tag{5.2.3}$$

Then for $(i-1)T \le t < iT$, $1 \le i \le d$ and $(j-1)T \le s < jT$, $1 \le j \le d$, the covariance function $K_{i,j}(t,s)$ between $U(t)$ and $U(s)$ can be expressed as

$$K_{i,j}(t,s) = \mathrm{cov}(U(t),U(s)) = \mathrm{cov}[X_i(t-(i-1)T), X_j(t-(j-1)T)] \tag{5.2.4}$$

For d-dimensional processes, the Fredholm integral equation is equivalent to the d simultaneous integral equations. Each of them corresponds to a specific time interval of $U(t)$. For $(i-1)T \le t < iT$, $1 \le i \le d$, the Fredholm integral equation yields

$$\int_0^{dT} \mathrm{cov}(U(t),U(s))\phi(s)ds = \lambda\phi(t) \qquad (i-1)T \le t < iT, 1 \le i \le d \tag{5.2.5}$$

Replace $U(t)$ with the corresponding $X_i(t)$,

$$\begin{aligned} &\int_0^T \mathrm{cov}[X_i(t-(i-1)T), X_1(s)]f^{(1)}(s)ds \\ &+ \int_T^{2T} \mathrm{cov}[X_i(t-(i-1)T), X_2(s-T)]f^{(2)}(s)ds \\ &+ \cdots + \int_{(d-1)T}^{dT} \mathrm{cov}[X_i(t-(i-1)T), X_d(s-(d-1)T)]f^{(d)}(s)ds \\ &= \lambda f^{(i)}(t-(i-1)T) \qquad (i-1)T \le t < iT, 1 \le i \le d \end{aligned}$$

After some rearrangements, the above equation can be simplified to

$$\sum_{j=1}^{d} \int_0^T \text{cov}[X_i(t-(i-1)T), X_j(s)] f^{(j)}(s)ds = \lambda f^{(i)}(t-(i-1)T)$$
$$(i-1)T \leq t < iT, 1 \leq i \leq d$$

This is equivalent to the statement that when $0 \leq t < T$, the Fredholm integral equation is expressed as

$$\sum_{j=1}^{d} \int_0^T \text{cov}[X_i(t), X_j(s)] f^{(j)}(s)ds = \lambda f^{(i)}(t) \quad 0 \leq t < T \tag{5.2.6}$$

In the multivariate setting, the orthogonality condition is

$$\int_0^{dT} \phi_i(s)\phi_j(s)ds = \delta_{ij} \Leftrightarrow \sum_{k=1}^{d} \int_{(k-1)T}^{kT} f_i^{(k)}(s-(k-1)T) f_j^{(k)}(s-(k-1)T)ds = \delta_{ij} \tag{5.2.7}$$

This is equivalent to

$$\sum_{k=1}^{d} \int_0^T f_i^{(k)}(s) f_j^{(k)}(s)ds = \delta_{ij} \tag{5.2.8}$$

A version of Mercer's theorem in the multivariate setting has been rigorously proved in Mahram et al. (2002). See, appendix 7.1, for the statement of the generalised Mercer's theorem. Here the result for the covariance function is briefly shown using the lining-up method.

Assume that the covariance function is continuous after lining up the process. For $(i-1)T \leq t < iT$, $1 \leq i \leq d$ and $(j-1)T \leq s < jT$, $1 \leq j \leq d$, the covariance function $K_{ij}(t,s)$ between $X_i(t)$ and $X_j(s)$ can then be expressed as

$$K_{ij}(t,s) = \text{cov}(U(t), U(s)) = \text{cov}(X_i(t), X_j(s))$$
$$= \sum_{k \geq 1} \lambda_k \phi_k(t)\phi_k(s) = \sum_{k \geq 1} \lambda_k f_k^{(i)}(t-(i-1)T) f_k^{(j)}(s-(j-1)T)$$

This is equivalent to the statement that for $0 \leq t, s \leq T$,

$$K_{ij}(t,s) = \sum_{k \geq 1} \lambda_k f_k^{(i)}(t) f_k^{(j)}(s) \tag{5.2.9}$$

The above results could be further expressed in the matrix form. Defining the covariance matrix $\mathbf{K}(t,s) = \text{cov}(\mathbf{X}(t), \mathbf{X}(s))$ and the eigenfunction $f(t)$ as follows

$$\mathbf{K}(t,s) = \begin{pmatrix} \text{cov}(X_1(t), X_1(s)) & \text{cov}(X_1(t), X_2(s)) & \cdots & \text{cov}(X_1(t), X_d(s)) \\ \text{cov}(X_2(t), X_1(s)) & \text{cov}(X_2(t), X_2(s)) & \cdots & \text{cov}(X_2(t), X_d(s)) \\ \text{cov}(X_d(t), X_1(s)) & \text{cov}(X_d(t), X_2(s)) & \cdots & \text{cov}(X_d(t), X_d(s)) \end{pmatrix}$$

$$f(t)^T = \begin{pmatrix} f^{(1)}(t), & f^{(2)}(t), & \cdots, & f^{(d)}(t) \end{pmatrix}$$

Notice that $f(t)$ defined above is a column vector, while $f(t)$ in the previous section represents the univariate eigenfunction for a univariate process. For the multivariate process $\mathbf{X}(t)$, $t \in [0,T]$, its multivariate Karhunen-Loeve expansion can be expressed as

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \cdots \\ X_d(t) \end{pmatrix} = \sum_{i \geq 1} \sqrt{\lambda_i} f_i(t) \xi_i = \sum_{i \geq 1} \sqrt{\lambda_i} \begin{pmatrix} f_i^{(1)}(t) \\ f_i^{(2)}(t) \\ \cdots \\ f_i^{(d)}(t) \end{pmatrix} \xi_i \ ,$$

where $\xi_i \sim i.i.dN(0,1)$. Then, the Fredholm integral equation is

$$\int_0^T \mathbf{K}(s,t) f(s) ds = \lambda f(t) \tag{5.2.10}$$

with the orthogonality condition

$$\int_0^T f_i(s)^T f_j(s) ds = \delta_{ij} \tag{5.2.11}$$

Mercer's theorem is

$$\mathbf{K}(t,s) = \sum_{i \geq 1} \lambda_i f_i(t) f_i^T(s) \tag{5.2.12}$$

In the multivariate setting, the eigenfunctions are vectors, while the eigenvalues are still scalars.

The multivariate results for the Karhunen-Loeve expansion will now be expressed using the generalised Mercer's theorem. A theorem, similar to theorem 1.2.1 in chapter 1, will be proposed and proved here.

Theorem 5.2.1. *Let* $\mathbf{X}(t) \in H_{\mathbf{X}}$, $t \in \mathcal{T}$ *be a zero mean vector process with* $\sum_{i=1}^{d} E(X_i^2(t)) < \infty$. *Its covariance matrix is denoted as* $\mathbf{K}(s,t)$ *for the covariance between time* s *and time* t.

Let $\{f_i\}$ *be the orthogonal eigenfunction vector, i.e.*

$$\int_{\mathcal{T}} f_i(t)^T f_j(t) = \delta_{ij} \tag{5.2.13}$$

1. *Assume that* λ_i *and* $\phi_i(t)$ *satisfy the following equation ,*

$$\int_{\mathcal{T}} \mathbf{K}(t,s) f_i(s) ds = \lambda_i f_i(t) \tag{5.2.14}$$

where $\{f_i, i \in N\}$ *and* $\{\lambda_i, i \in N\}$ *are called the eigenfunctions and eigenvalues respectively in the multivariate setting.*

Furthermore, choose

$$\rho_i = \int_{\mathcal{T}} f_i(t)^T \mathbf{X}(t) dt \tag{5.2.15}$$

Then

$$\mathbf{X}(t) = \lim_{p \to \infty} \sum_{i=1}^{p} \rho_i f_i(t) \tag{5.2.16}$$

The limit is defined in the sense of mean square convergence.

2. *Conversely, if* $\mathbf{X}(t) = \sum_{i=1}^{\infty} \rho_i f_i(t)$, *where* $\{\rho_i, i \in N\}$ *is i.i.d. with mean* 0 *and variance* λ_i,

$$\int_{\mathcal{T}} \mathbf{K}(s,t) f_i(s) ds = \lambda_i f_i(t) \tag{5.2.17}$$

Proof. The proof is also similar to that in the univariate case.

(a) We first show that $E(\rho_i) = 0$ and $\text{cov}(\rho_i, \rho_j) = \lambda_i \delta_{ij}$

$$\begin{aligned} E(\rho_i) &= E(\int_{\mathcal{T}} f_i(t)\mathbf{X}(t)dt) = \int_{\mathcal{T}} f_i(t)E(\mathbf{X}(t))dt = 0 \\ \text{cov}(\rho_i, \rho_j) &= E(\rho_i \rho_j) = \int_{\mathcal{T}}\int_{\mathcal{T}} f_i(t)^T E(\mathbf{X}(t)\mathbf{X}(s)^T) f_j(s) ds dt \\ &= \int_{\mathcal{T}} f_i(t)^T \int_{\mathcal{T}} \mathbf{K}(s,t) f_j(s) ds dt = \int_{\mathcal{T}} f_i(t)^T \lambda_j f_j(t) dt \\ &= \lambda_i \delta_{ij} \end{aligned}$$

Moreover,

$$\begin{aligned} E(\rho_i \mathbf{X}(t)^T) &= E(\int_{\mathcal{T}} f_i(s)^T \mathbf{X}(s)\mathbf{X}(t)^T ds) \\ &= \int_{\mathcal{T}} f_i(s)^T E(\mathbf{X}(s)\mathbf{X}(t)^T) ds \\ &= \lambda_i f_i(t)^T \end{aligned}$$

Therefore,

$$\begin{aligned} & ||\mathbf{X}(t) - \sum_{i=1}^{p} \rho_i f_i(t)||^2 \\ = & < \mathbf{X}(t) - \sum_{i=1}^{p} \rho_i f_i(t), \mathbf{X}(t) - \sum_{i}^{p} \rho_i f_i(t) > \\ = & < \mathbf{X}(t), \mathbf{X}(t) > -2 < \mathbf{X}(t), \sum_{i=1}^{p} \rho_i f_i(t) > + < \sum_{i=1}^{p} \rho_i f_i(t), \sum_{i=1}^{p} \rho_i f_i(t) > \\ = & E(\mathbf{X}(t)^T \mathbf{X}(t)) - 2\sum_{i=1}^{p} E(\mathbf{X}(t)^T \rho_i) f_i(t) + \sum_{i=1}^{p}\sum_{j=1}^{p} E(\rho_i f_i(t)^T \rho_j f_j(t)) \\ = & E(\mathbf{X}(t)^T \mathbf{X}(t)) - 2\sum_{i=1}^{p} \lambda_i f_i(t)^T f_i(t) + \sum_{i=1}^{p}\sum_{j=1}^{p} E(\rho_i \rho_j) f_i(t)^T f_j(t) \\ = & E(\mathbf{X}(t)^T \mathbf{X}(t)) - \sum_{i=1}^{p} \lambda_i f_i(t)^T f_i(t) = \sum_{i=1}^{d} E(X_i(t)^2) - \sum_{i=1}^{p} \lambda_i \text{trace}(f_i(t)^T f_i(t)) \\ = & \text{trace}(\mathbf{K}(t,t)) - \sum_{i=1}^{p} \lambda_i \text{trace}(f_i(t) f_i(t)^T) = \text{trace}[\mathbf{K}(t,t) - \sum_{i=1}^{p} \lambda_i f_i(t) f_i(t)^T] \to 0 \end{aligned}$$

The limit follows from the generalised Mercer's theorem. Appendix 7.1 states the generalised Mercer's theorem. See, for example, Mahram et al. (2002), for proof.

(b) Conversely, if $\mathbf{X}(t) = \sum_{i=1}^{\infty} \rho_i f_i(t)$,

$$\begin{aligned}\mathbf{K}(t,s) &= E(\mathbf{X}(t)\mathbf{X}(s)^T) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E(\rho_i f_i(t)\rho_j f_j(s)^T) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E(\rho_i\rho_j) f_i(t) f_j(s)^T = \sum_{i=1}^{\infty} \lambda_i f_i(t) f_i(t)^T\end{aligned}$$

and the integral equation is

$$\begin{aligned}\int_{\mathcal{T}} \mathbf{K}(t,s) f_i(s) ds &= \int_{\mathcal{T}} \sum_j \lambda_j f_j(t) f_j(s)^T f_s(s) ds \\ &= \sum_j \lambda_j f_j(t) \int_{\mathcal{T}} f_j(s)^T f_i(s) ds \\ &= \lambda_i f_i(t)\end{aligned}$$

□

We use two approaches above to derive the multivariate Karhunen-Loeve expansion. One is to use the lining-up method to transform the multivariate problem to a univariate problem, when the covariance function, after lining up, is continuous. The other approach relies on the generalised Mercer's theorem and derives the multivariate Karhunen-Loeve expansion to all $\mathbf{X}(t)$ in the space of $H_{\mathbf{X}}$. The latter is a more general approach, while the former might cast doubt on applying the univariate Mercer's theorem after lining up, since the corresponding covariance function might not be continuous. For example, to the d dimensional vector Brownian motion $\mathbf{X}(t)$ with cross covariance function $\mathrm{cov}(X_{i-1}(t), X_i(s)) = \rho \min(t,s)$, $0 \le t, s \le T$, if the covariance function is continuous after lining up,

$$\lim_{t \nearrow T} \mathrm{cov}(X_{i-1}(t), X_i(0)) = \lim_{t \nearrow T} \mathrm{cov}(X_{i-1}(t), X_{i-1}(t)) \ , \qquad (5.2.18)$$

where the limit in the equation is a limit from the left. However

$$\lim_{t\nearrow T}\text{cov}(X_{i-1}(t), X_i(0)) = \rho \times 0 = 0 \tag{5.2.19}$$

$$\lim_{t\nearrow T}\text{cov}(X_{i-1}(t), X_{i-1}(t)) = \lim_{t\nearrow T}\min(t) = T \tag{5.2.20}$$

Since equation 5.2.19 and equation 5.2.20 are not equivalent, the continuity assumption in the univariate Mercer's theorem, after lining up the vector Brownian motion, is not satisfied. Hence the lining-up approach is not suitable in this example.

Remark 5.2.1. As mentioned in Mahram et al. (2002), checking the continuity in the covariance function is the same as checking the mean square continuity of the process after lining up. In the multivariate setting, this is equivalent to checking the mean square continuous at the transition point in the process $\mathbf{X}(t)$, i.e.

$$\lim_{t\searrow T}E[|X_i(t-T) - X_i(0)|^2] = 0 \quad 1 \le i \le d \tag{5.2.21}$$

$$\lim_{t\nearrow T}E[|X_{i-1}(t) - X_i(0)|^2] = 0 \quad 2 \le i \le d \tag{5.2.22}$$

The first equation is a limit from the right, which is always right, since it involves only one process. However, the second equation, which is a limit from the left, goes through two processes and does not equal to zero to some often used processes in this thesis. Again use the example of the vector Brownian motion with the cross covariance $\text{cov}(X_{i-1}(t), X_i(s)) = \rho\min(t,s)$, $0 \le t, s \le T$

$$\begin{aligned} & \lim_{t\nearrow T}E[|X_{i-1}(t) - X_i(0)|^2] = \text{Var}[X_{i-1}(t) - X_i(0)] \\ = \ & \lim_{t\nearrow T}\text{Var}[X_{i-1}(t)] + \text{Var}[X_i(0)] - 2\text{cov}[X_{i-1}(t), X_i(0)] = \lim_{t\nearrow T} t = T \neq 0 \end{aligned}$$

If the vector process $\mathbf{X}(t)$ is mean square continuous, both approaches are equivalent, since the covariance function is continuous after lining up the process.

Theorem 5.1.3 in the last section is extended to a more complicated multivariate setting. It can be proved that if the vector process $\mathbf{X}(t) = A\mathbf{Y}(t)$, where $\mathbf{Y}(t)$ is independent and A is orthogonal, the multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ can be expressed in terms of the univariate Karhunen-Loeve expansion of $\mathbf{Y}(t)$ of each its component.

Theorem 5.2.2. *For a vector process* $\mathbf{X}(t)^T = \Big(X_1(t),\ X_2(t),\ \cdots,\ X_d(t)\Big)$, $t \in \mathcal{T}$, *assume* $\mathbf{X}(t) = A\mathbf{Y}(t)$, *where* A *is an orthogonal matrix, i.e.* $A^TA = I$, *and* $\mathbf{Y}(t)^T = \Big(Y_1(t),\ Y_2(t),\ \cdots,\ Y_d(t)\Big)$ *is an independent vector process, i.e.*

$$\mathrm{cov}(Y_i(t), Y_j(s)) = 0, \ \textit{if } i \neq j$$

If the univariate Karhunen-Loeve expansion for $Y_i(t)$ *is*

$$Y_i(t) = \sum_{k\geq 1} \sqrt{\lambda_{k,i}^Y}\phi_{k,i}(t)\epsilon_{k,i} \tag{5.2.23}$$

and $\lambda_{k,i}^Y \neq \lambda_{k,j}^Y$ *when* $i \neq j$, *the multivariate Karhunen-Loeve expansion for* $\mathbf{X}(t)$ *is a linear combination of the univariate Karhunen-Loeve expansion of* $Y_i(t)$, *i.e.*

$$\mathbf{X}(t) = \sum_{k\geq 1} A \begin{pmatrix} \sqrt{\lambda_{k,1}^Y} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_{k,2}^Y} & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \sqrt{\lambda_{k,d}^Y} \end{pmatrix} \begin{pmatrix} \phi_{k,1}(t) & 0 & \cdots & 0 \\ 0 & \phi_{k,2}(t) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \phi_{k,d}(t) \end{pmatrix} \begin{pmatrix} \epsilon_{k,1} \\ \epsilon_{k,2} \\ \cdots \\ \epsilon_{k,d} \end{pmatrix} \tag{5.2.24}$$

where $\{\epsilon_{k,i}\}$ *is a series of independent standard normal random variables.*

Proof. Since $\mathbf{X}(t) = A\mathbf{Y}(t)$, its covariance matrix is $\mathrm{cov}(\mathbf{X}(t), \mathbf{X}(s)) = A\mathrm{cov}(\mathbf{Y}(t), \mathbf{Y}(s))A^T$.

Hence the integral equation we need to solve is

$$\int_{\mathcal{T}} A\mathrm{cov}(\mathbf{Y}(t), \mathbf{Y}(s))A^T f(s)ds = \lambda f(t)\ , \tag{5.2.25}$$

where $f(t)^T = \left(f^{(1)}(t), \quad f^{(2)}(t), \quad \cdots, \quad f^{(d)}(t) \right)$ and each $f^{(i)}(t)$ is the eigenfunction for $X_i(t)$ in the multivariate setting.

Since matrix A is orthogonal, equation 5.2.25 is equivalent to

$$\int_{\mathcal{T}} \text{cov}(\mathbf{Y}(t), \mathbf{Y}(s)) A^T f(s) ds = \lambda A^T f(t) \tag{5.2.26}$$

We can further assume that

$$f'(t) = A^T f(t) = \left(f'^{(1)}(t), \quad f'^{(2)}(t), \quad \cdots, \quad f'^{(d)}(t) \right)^T \tag{5.2.27}$$

The orthogonality condition of $f'(t)$ follows from $f(t)$.

$$\int_{\mathcal{T}} f_i'(t)^T f_j'(t) dt = \int_{\mathcal{T}} f_i(t)^T A A^T f_j(t) dt = \int_{\mathcal{T}} f_i(t)^T f_j(t) dt = \delta_{ij} \tag{5.2.28}$$

Therefore the Fredholm integral equation can now be expressed as

$$\int_{\mathcal{T}} \begin{pmatrix} \text{cov}(Y_1(s), Y_1(t)) & 0 & \cdots & 0 \\ 0 & \text{cov}(Y_2(s), Y_2(t)) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \text{cov}(Y_d(s), Y_d(t)) \end{pmatrix} \begin{pmatrix} f'^{(1)}(s) \\ f'^{(2)}(s) \\ \cdots \\ f'^{(d)}(s) \end{pmatrix} ds = \lambda \begin{pmatrix} f'^{(1)}(t) \\ f'^{(2)}(t) \\ \cdots \\ f'^{(d)}(t) \end{pmatrix} \tag{5.2.29}$$

This is the same as the following d simultaneous equations.

$$\int_{\mathcal{T}} \text{cov}(Y_i(s), Y_i(t)) f'^{(i)}(s) ds = \lambda f'^{(i)}(t) \quad 1 \leq i \leq d \tag{5.2.30}$$

Now, the multivariate Fredholm integral equation of $\mathbf{X}(t)$ has been transformed to the univariate Fredholm integral equation of $\mathbf{Y}(t)$ of each its component. For the ith equation

$$\int_{\mathcal{T}} \text{cov}(Y_i(s), Y_i(t)) f'^{(i)}_k(s) ds = \lambda_{k,i} f'^{(i)}_k(t) \ , \tag{5.2.31}$$

where $\lambda_{k,i}$ is the kth eigenvalue and $f_k^{\prime(i)}(t)$ is the kth eigenfunction,

$$\lambda_{k,i} = \lambda_{k,i}^Y \qquad f_k^{\prime(i)}(t) = \phi_{k,i}(t) \qquad 1 \leq i \leq d, k \geq 1$$

Since $\lambda_{k,i}^Y \neq \lambda_{k,j}^Y$ when $i \neq j$, $\lambda_{k,i} \neq \lambda_{k,j}$ when $i \neq j$. However, $\lambda_{k,i}$ should be the common eigenvalue for all the components of the process $\mathbf{X}(t)$. Hence the solution to the above d simultaneous equations is

$$\begin{aligned}
&\lambda_{k,1} = \lambda_{k,1}^Y, f_{k,1}^{\prime(1)}(t) = \phi_{k,1}(t), f_{k,1}^{\prime(2)}(t) = f_{k,1}^{\prime(3)}(t) = \cdots = f_{k,1}^{\prime(d)}(t) = 0, \quad \text{or} \\
&\lambda_{k,2} = \lambda_{k,2}^Y, f_{k,2}^{\prime(1)}(t) = 0, f_{k,2}^{\prime(2)}(t) = \phi_{k,2}(t), f_{k,2}^{\prime(3)}(t) = \cdots = f_{k,2}^{\prime(d)}(t) = 0, \quad \text{or} \\
&\cdots \\
&\lambda_{k,d} = \lambda_{k,d}^Y, f_{k,d}^{\prime(1)}(t) = f_{k,d}^{\prime(2)}(t) = \cdots = f_{k,d}^{\prime(d-1)}(t) = 0, f_{k,d}^{\prime(d)}(t) = \phi_{k,d}(t) \ ,
\end{aligned}$$

where $f_{k,j}^{\prime(i)}$ is the jth possible solution to the kth eigenfunction of the process $Y_i(t)$. Notice that $f_{k,j}^{\prime(i)}$ follows the orthogonality condition in equation 5.2.28.

Alternatively all the possible solutions to the kth eigenfunction can be written down in the following matrix form

$$\begin{pmatrix} f_{k,1}^{\prime(1)}(t) & f_{k,2}^{\prime(1)}(t) & \cdots & f_{k,d}^{\prime(1)}(t) \\ f_{k,1}^{\prime(2)}(t) & f_{k,2}^{\prime(2)}(t) & \cdots & f_{k,d}^{\prime(2)}(t) \\ \cdots & & & \\ f_{k,1}^{\prime(d)}(t) & f_{k,2}^{\prime(d)}(t) & \cdots & f_{k,d}^{\prime(d)}(t) \end{pmatrix} = \begin{pmatrix} \phi_{k,1}(t) & 0 & \cdots & 0 \\ 0 & \phi_{k,2}(t) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \phi_{k,d}(t) \end{pmatrix} \tag{5.2.32}$$

Since $f(t) = Af'(t)$, all the possible solution to the $f_k(t)$ can now be expressed as

$$\begin{pmatrix} f_{k,1}^{(1)}(t) & f_{k,2}^{(1)}(t) & \cdots & f_{k,d}^{(1)}(t) \\ f_{k,1}^{(2)}(t) & f_{k,2}^{(2)}(t) & \cdots & f_{k,d}^{(2)}(t) \\ \cdots & & & \\ f_{k,1}^{(d)}(t) & f_{k,2}^{(d)}(t) & \cdots & f_{k,d}^{(d)}(t) \end{pmatrix} = A \begin{pmatrix} \phi_{k,1}(t) & 0 & \cdots & 0 \\ 0 & \phi_{k,2}(t) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \phi_{k,d}(t) \end{pmatrix} \tag{5.2.33}$$

Therefore, the multivariate Karhunen-Loeve expansion for $\mathbf{X}(t)$ is

$$\begin{aligned}
\mathbf{X}(t) &= \sum_{k\geq 1} \begin{pmatrix} f_{k,1}^{(1)}(t) & f_{k,2}^{(1)}(t) & \cdots & f_{k,d}^{(1)}(t) \\ f_{k,1}^{(2)}(t) & f_{k,2}^{(2)}(t) & \cdots & f_{k,d}^{(2)}(t) \\ \cdots & & & \\ f_{k,1}^{(d)}(t) & f_{k,2}^{(d)}(t) & \cdots & f_{k,d}^{(d)}(t) \end{pmatrix} \begin{pmatrix} \lambda_{k,1} & 0 & \cdots & 0 \\ 0 & \lambda_{k,2} & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_{k,d} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} \epsilon_{k,1} \\ \epsilon_{k,2} \\ \cdots \\ \epsilon_{k,d} \end{pmatrix} \\
&= \sum_{k\geq 1} A \begin{pmatrix} \sqrt{\lambda_{k,1}} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_{k,2}} & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \sqrt{\lambda_{k,d}} \end{pmatrix} \begin{pmatrix} \phi_{k,1}(t) & 0 & \cdots & 0 \\ 0 & \phi_{k,2}(t) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \phi_{k,d}(t) \end{pmatrix} \begin{pmatrix} \epsilon_{k,1} \\ \epsilon_{k,2} \\ \cdots \\ \epsilon_{k,d} \end{pmatrix}
\end{aligned}$$

□

Remark 5.2.2. Sometimes, although the correlated vector process $\mathbf{X}(t)$ can be written down as a linear combination of the independent vector process $\mathbf{Y}(t)$, i.e. $\mathbf{X}(t) = A\mathbf{Y}(t)$, A may not be orthogonal. However, it is still possible to write down the multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ in terms of the univariate Karhunen-Loeve expansion, if A satisfies the following three conditions.

(i) $A = UD$

(ii) $UU^T = I$, i.e. U is an orthogonal matrix

(iii) $\mathbf{Y}'(t) = D\mathbf{Y}(t)$, where $\mathbf{Y}'(t)^T = \left(Y_1'(t),\ Y_2'(t),\ \cdots,\ Y_d'(t)\right)$, is still an independent vector process, with the different eigenvalue for each new process $Y_i'(t)$ after transformation. Then $\mathbf{X}(t) = A\mathbf{Y}(t) = UD\mathbf{Y}(t) = U\mathbf{Y}'(t)$, where $UU^T = I$.

If AA^T is a symmetric positive definite matrix, i.e. $AA^T = U\Gamma U^T$, the above conditions satisfy if the new independent process has different eigenvalues for each of its component. Since A can be written down as $A = U\Gamma^{\frac{1}{2}}$, $\mathbf{Y}'(t) = \Gamma^{\frac{1}{2}}\mathbf{Y}(t)$ is the new independent vector process. If each $Y_i'(t)$ has different eigenvalues, $\mathbf{X}(t) = U\mathbf{Y}'(t)$,

with orthogonal matrix U. The multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ can now be expressed using theorem 5.2.2.

Remark 5.2.3. If there is no decomposition of A of this kind, the multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ might not be expressible as a linear combination of the univariate Karhunen-Loeve expansion of $\mathbf{Y}(t)$ following theorem 5.2.2. Since this linear combination might not solve the multivariate Fredholm integral equation. However, this linear combination can still be defined as so-called the Karhunen-Loeve-like expansion. Although it might lack some attractive properties of the Karhunen-Loeve expansion, like the minimal mean squared error, analysis like simulation and covariance reconstruction can still be implemented.

Theorem 5.2.2 can now be applied to prove the following corollary, which is an extension of theorem 5.1.3 in section 5.1.

Corollary 5.2.3. *For a vector process* $\mathbf{X}(t)^T = \Big(X_1(t),\ \ X_2(t),\ \ \cdots,\ \ X_d(t)\Big)$, *assume its covariance matrix* $K(s,t)$ *can be expressed as* $K(s,t) = \mathrm{cov}(X_1(s), X_1(t))B$, *where* B *is a symmetric positive definite matrix.*

Then the multivariate Karhunen-Loeve expansion for $\mathbf{X}(t)$ *can be written down as*

$$\mathbf{X}(t) = \sum_{i\geq 1} U\Gamma^{\frac{1}{2}} \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \cdots \\ \epsilon_{i,d} \end{pmatrix} \sqrt{\lambda_i^{\mathrm{old}}}\phi_i(t)\ , \tag{5.2.34}$$

where λ_i^{old} *and* $\phi_i(t)$ *is the ith eigenvalue and the ith eigenfunction for* $X_1(t)$ *respectively.* $\{\epsilon_{i,j}, i \geq 1, 1 \leq j \leq d\}$ *is a series of independent standard normal random*

variables. U and Γ come from the eigenvalue decomposition of the matrix B, i.e. $B = U\Gamma U^T$.

Proof. Since B is a symmetric positive definite matrix, it can be decomposed as

$$B = U\Gamma U^T \ , \tag{5.2.35}$$

where $U^TU = I$ and $\Gamma = \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ 0 & 0 & \cdots & \gamma_d \end{pmatrix}$, $\gamma_i > 0, 1 \leq i \leq d$, $\gamma_i \neq \gamma_j$ when $i \neq j$.
Hence the covariance matrix for $\mathbf{X}(t)$ can also be expressed as

$$\text{cov}(\mathbf{X}(t), \mathbf{X}(s)) = U[\text{cov}(X_1(t), X_1(s))\Gamma]U^T \tag{5.2.36}$$

An independent vector process $\mathbf{Y}(t)$ can now be defined with $\text{cov}[Y_i(t), Y_j(s)] = \gamma_i \text{cov}[X_1(t), X_1(s)]\delta_{ij}$, such that $\mathbf{X}(t) = U\mathbf{Y}(t)$. Therefore the kth eigenvalue and the kth eigenfunction for $Y_i(t)$ are $\gamma_i\lambda_k^{\text{old}}, k \geq 1$ and $\phi_k(t), k \geq 1$ respectively. Since the eigenvalue for $Y_i(t)$ is different from that for $Y_j(t)$, i.e. $\gamma_i\lambda_k^{\text{old}} \neq \gamma_j\lambda_k^{\text{old}}$ when $i \neq j$, and U is an orthogonal matrix, the result in the corollary is a direct application of theorem 5.2.2. □

Remark 5.2.4. Corollary 5.2.3 can now be applied to prove theorem 5.1.3 in the bivariate case in section 5.1. Since $B = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$,

$$\Gamma = \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix} \quad U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.2.37}$$

Hence using equation 5.2.34, the bivariate Karhunen-Loeve expansion for the bivariate process $\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}$ is

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \sum_{i \geq 1} \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} & \sqrt{\frac{1-\rho}{2}} \\ \sqrt{\frac{1+\rho}{2}} & -\sqrt{\frac{1-\rho}{2}} \end{pmatrix} \begin{pmatrix} \xi_{i,1} \\ \xi_{i,2} \end{pmatrix} \sqrt{\lambda_i^{\text{old}}} \phi_i(t) \ , \tag{5.2.38}$$

where $\xi_{i,1}$ and $\xi_{i,2}$ are the mutually independent standard normal random variables. This is the same result as what is derived in section 5.1.

5.3 Numerical methods

There are two approaches to implement the multivariate Karhunen-Loeve expansion in the multivariate setting. When the process is mean square continuous, the lining-up method can be applied to transform the multivariate problem into the univariate problem. The details of the numerical methods with respect to the univariate problem have been discussed in chapter 2. However, in practice, not all the processes are mean square continuous. A more general approach using the generalised Mercer's theorem has to be applied, as discussed in section 5.2. Correspondingly, some adjustments to the numerical methods are required. This section can be treated as a generalisation of section 2.1 and section 2.2, chapter 2. The main numerical methods discussed here are still the integral method and the expansion method using the Fourier basis and the Haar wavelet basis. For simplicity, the interval $\mathcal{T} = [0, 1]$. Other intervals could be easily generalised.

Integral Method

Suppose that we have a vector process $\mathbf{X}(t)^T = \Big(X_1(t), \ X_2(t), \ \cdots, \ X_d(t)\Big)$.

λ and $f^{(m)}(t)$ are used to denote the eigenvalue for the multivariate Karhunen-Loeve expansion and the eigenfunction for X_m at time t respectively. The Fredholm integral equation here is

$$\int_0^1 \begin{pmatrix} \mathrm{cov}(X_1(t),X_1(s)) & \mathrm{cov}(X_1(t),X_2(s)) & \cdots & \mathrm{cov}(X_1(t),X_d(s)) \\ \mathrm{cov}(X_2(t),X_1(s)) & \mathrm{cov}(X_2(t),X_2(s)) & \cdots & \mathrm{cov}(X_2(t),X_d(s)) \\ \cdots & & & \\ \mathrm{cov}(X_d(t),X_1(s)) & \mathrm{cov}(X_d(t),X_2(s)) & \cdots & \mathrm{cov}(X_d(t),X_d(s)) \end{pmatrix} \begin{pmatrix} f^{(1)}(s) \\ f^{(2)}(s) \\ \cdots \\ f^{(d)}(s) \end{pmatrix} ds$$
$$= \lambda \begin{pmatrix} f^{(1)}(t) \\ f^{(2)}(t) \\ \cdots \\ f^{(d)}(t) \end{pmatrix}$$

Assume that n points t_i, $1 \le i \le n$ on $[0,1]$ are used to approximate the integral, with $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_n \le t_{n+1} = 1$. Then the approximation using the integral method can be expressed as

$$\sum_{i=0}^{n+1} \begin{pmatrix} \mathrm{cov}(X_1(t_j),X_1(t_i)) & \mathrm{cov}(X_1(t_j),X_2(t_i)) & \cdots & \mathrm{cov}(X_1(t_j),X_d(t_i)) \\ \mathrm{cov}(X_2(t_j),X_1(t_i)) & \mathrm{cov}(X_2(t_j),X_2(t_i)) & \cdots & \mathrm{cov}(X_2(t_j),X_d(t_i)) \\ \cdots & & & \\ \mathrm{cov}(X_d(t_j),X_1(t_i)) & \mathrm{cov}(X_d(t_j),X_2(t_i)) & \cdots & \mathrm{cov}(X_d(t_j),X_d(t_i)) \end{pmatrix}$$
$$\begin{pmatrix} w_i^{(1)} & 0 & \cdots & 0 \\ 0 & w_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & w_i^{(d)} \end{pmatrix} \begin{pmatrix} f^{(1)}(t_i) \\ f^{(2)}(t_i) \\ \cdots \\ f^{(d)}(t_i) \end{pmatrix} = \lambda \begin{pmatrix} f^{(1)}(t_j) \\ f^{(2)}(t_j) \\ \cdots \\ f^{(d)}(t_j) \end{pmatrix}$$

When the uniform integral method is used, $w_i^{(j)} = \frac{1}{n+2}$, $0 \le i \le n+2$, $1 \le j \le d$, while when the trapezium method is used, $w_0^{(j)} = w_{n+1}^{(j)} = \frac{1}{2(n+1)}$, $1 \le j \le d$ and $w_i^{(j)} = \frac{1}{n+1}$, $1 \le i \le n$, $1 \le j \le d$.

Or after re-arrangement with the notation

$$X_m^T = \Big(X_m(t_1),\ \ X_m(t_2),\ \ \cdots,\ \ X_m(t_n)\Big), \tag{5.3.1}$$

and the jth eigenfunction for the process X_m,

$$(f_j^{(m)})^T = \Big(f_j^{(m)}(t_1),\ \ f_j^{(m)}(t_2),\ \ \cdots,\ \ f_j^{(m)}(t_{n+2})\Big), \tag{5.3.2}$$

the approximation to the integral equation is

$$\begin{pmatrix} \mathrm{cov}(X_1, X_1^T) & \mathrm{cov}(X_1, X_2^T) & \cdots \mathrm{cov}(X_1, X_d^T) \\ \mathrm{cov}(X_2, X_1^T) & \mathrm{cov}(X_2, X_2^T) & \cdots \mathrm{cov}(X_2, X_d^T) \\ \cdots & & \\ \mathrm{cov}(X_d, X_1^T) & \mathrm{cov}(X_d, X_2^T) & \cdots \mathrm{cov}(X_d, X_d^T) \end{pmatrix} \begin{pmatrix} W^{(1)} & 0 & \cdots & 0 \\ 0 & W^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & W^{(d)} \end{pmatrix}$$

$$\begin{pmatrix} f_1^{(1)} & f_2^{(1)} & \cdots & f_{(n+2)d}^{(1)} \\ f_1^{(2)} & f_2^{(2)} & \cdots & f_{(n+2)d}^{(2)} \\ \cdots & & & \\ f_1^{(d)} & f_2^{(d)} & \cdots & f_{(n+2)d}^{(d)} \end{pmatrix} = \begin{pmatrix} f_1^{(1)} & f_2^{(1)} & \cdots & f_{(n+2)d}^{(1)} \\ f_1^{(2)} & f_2^{(2)} & \cdots & f_{(n+2)d}^{(2)} \\ \cdots & & & \\ f_1^{(d)} & f_2^{(d)} & \cdots & f_{(n+2)d}^{(d)} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_{(n+2)d} \end{pmatrix}$$

The following notation is introduced so that the notation used here is consistent with that in the univariate case.

$$K = \begin{pmatrix} \mathrm{cov}(X_1, X_1^T) & \mathrm{cov}(X_1, X_2^T) & \cdots \mathrm{cov}(X_1, X_d^T) \\ \mathrm{cov}(X_2, X_1^T) & \mathrm{cov}(X_2, X_2^T) & \cdots \mathrm{cov}(X_2, X_d^T) \\ \cdots & & \\ \mathrm{cov}(X_d, X_1^T) & \mathrm{cov}(X_d, X_2^T) & \cdots \mathrm{cov}(X_d, X_d^T) \end{pmatrix} \tag{5.3.3}$$

$$f = \begin{pmatrix} f_1^{(1)} & f_2^{(1)} & \cdots & f_{(n+2)d}^{(1)} \\ f_1^{(2)} & f_2^{(2)} & \cdots & f_{(n+2)d}^{(2)} \\ \cdots & & & \\ f_1^{(d)} & f_2^{(d)} & \cdots & f_{(n+2)d}^{(d)} \end{pmatrix} \qquad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_{(n+2)d} \end{pmatrix}$$

$$W_{(n+2)d\times(n+2)d} = \begin{pmatrix} W^{(1)}_{(n+2)\times(n+2)} & 0 & \cdots & 0 \\ 0 & W^{(2)}_{(n+2)\times(n+2)} & \cdots & 0 \\ 0 & 0 & \cdots & W^{(d)}_{(n+2)\times(n+2)} \end{pmatrix}$$

W is a weighted matrix for the approximation. When the uniform integral method is used, $W = \frac{1}{n+2} I_{(n+2)d\times(n+2)d}$. When the trapezium method is used, each submatrix $W^{(i)}$ is identical and equals to

$$W^{(i)} = \begin{pmatrix} \frac{1}{2(n+1)} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{n+1} & \cdots & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & \cdots & \frac{1}{n+1} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{2(n+1)} \end{pmatrix} \tag{5.3.4}$$

Thus, the Fredholm integral equation has been changed to the eigenvalue problem

$$KWf = f\Lambda \tag{5.3.5}$$

under the orthogonality condition $f^T W f = I$. With $\phi = W^{1/2} f$, we need to solve a symmetric eigenvalue problem $W^{\frac{1}{2}} K W^{\frac{1}{2}} \phi = \phi\Lambda$ with the orthogonality condition $\phi^T\phi = I$. After obtaining ϕ and Λ, the inverse transformation can be computed, $f = W^{-\frac{1}{2}}\phi$.

The covariance matrix between X_{m_1} and X_{m_2} can now be expressed as

$$\mathrm{cov}(X_{m_1}, X_{m_2}^T) = \begin{pmatrix} f_1^{(m_1)}, & f_2^{(m_1)}, & \cdots, & f_{(n+2)d}^{(m_1)} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_{(n+2)d} \end{pmatrix} \begin{pmatrix} (f_1^{(m_2)})^T \\ (f_2^{(m_2)})^T \\ \cdots \\ (f_{(n+2)d}^{(m_2)})^T \end{pmatrix} \tag{5.3.6}$$

If the first $p \leq (n+2)d$ eigenvalues can explain most of the cumulative expected variance, the approximation of the covariance matrix at order p can be used.

$$\mathrm{cov}(X_{m_1}, X_{m_2}^T) \approx \begin{pmatrix} f_1^{(m_1)}, & f_2^{(m_1)}, & \cdots, & f_p^{(m_1)} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \begin{pmatrix} (f_1^{(m_2)})^T \\ (f_2^{(m_2)})^T \\ \cdots \\ (f_p^{(m_2)})^T \end{pmatrix} \tag{5.3.7}$$

Expansion Method 1: the Fourier method

Let $\mathbf{X}(t)^T = \begin{pmatrix} X_1(t) & X_2(t) & \cdots & X_d(t) \end{pmatrix}$ be a d-dimensional vector process and $\{\theta_i, 1 \leq i \leq M\}$ be a series of M adequate Fourier basis functions. Denote $f_i^{(m)}(t)$ the ith eigenfunction for the process $X_m(t)$ and $f_i^{(m)}(t)$ can be expanded as

$$f_i^{(m)}(t) = \sum_{k=1}^{M} d_{ik}^{(m)} \theta_k(t) = (D_i^{(m)})^T \theta(t) \ , \tag{5.3.8}$$

where

$$\begin{aligned} \theta(t)_{M\times 1}^T &= \begin{pmatrix} \theta_1(t), & \theta_2(t), & \cdots, & \theta_M(t) \end{pmatrix} \\ (D_{i,M\times 1}^{(m)})^T &= \begin{pmatrix} d_{i1}^{(m)}, & d_{i2}^{(m)}, \cdots, & d_{iM}^{(m)} \end{pmatrix} \end{aligned}$$

while $d_{ik}^{(m)}$ are the unknown coefficients for the expansion of the eigenfunction $f_i^{(m)}(t)$.

As in the univariate case, both sides of the integral equation are multiplied by each basis function and are then integrated one more time. Then the following equation can be obtained on the process X_m

$$\begin{aligned} &\sum_{k=1}^{M} d_{ik}^{(1)} \int_0^1 \int_0^1 K^{(m,1)}(s,t)\theta_k(s)\theta_j(t)dsdt + \sum_{k=1}^{M} d_{ik}^{(2)} \int_0^1 \int_0^1 K^{(m,2)}(s,t)\theta_k(s)\theta_j(t)dsdt + \cdots \\ + \quad &\sum_{k=1}^{M} d_{ik}^{(d)} \int_0^1 \int_0^1 K^{(m,d)}(s,t)\theta_k(s)\theta_j(t)dsdt - \lambda_i \sum_{k=1}^{M} d_{ik}^{(m)} \int_0^1 \theta_k(t)\theta_j(t)dt = 0 \ , \end{aligned}$$

where $K^{(a,b)}(t,s) = \mathrm{cov}(X_a(t), X_b(s))$, $a, b = 1, 2, \cdots, d$.

Denote

$$\begin{aligned} A^{(a,b)} &= \int_0^1 \int_0^1 K^{(a,b)}(s,t)\theta(s)\theta(t)^T dsdt \\ B &= \int_0^1 \theta(t)\theta(t)^T dt \\ D_{p\times M}^{(m)} &= \begin{pmatrix} (D_1^{(m)})^T \\ (D_2^{(m)})^T \\ \cdots \\ (D_p^{(m)})^T \end{pmatrix} \qquad \Lambda_{p\times p} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \end{aligned}$$

Then the equation on the process X_m can be written down in the matrix form

$$A^{(m,1)}(D^{(1)})^T + A^{(m,2)}(D^{(2)})^T + \cdots + A^{(m,d)}(D^{(d)})^T = B(D^{(m)})^T\Lambda$$
$$D^{(1)}A^{(m,1)} + D^{(2)}A^{(m,2)} + \cdots + D^{(d)}A^{(m,d)} = \Lambda D^{(m)}B$$

The matrix form can be extended to the whole vector process $\mathbf{X}(t)$ by further denoting

$$A_{Md\times Md} = \begin{pmatrix} A^{(1,1)} & A^{(2,1)} & \cdots & A^{(d,1)} \\ A^{(1,2)} & A^{(2,2)} & \cdots & A^{(d,2)} \\ \cdots & & & \\ A^{(1,d)} & A^{(2,d)} & \cdots & A^{(d,d)} \end{pmatrix}, \quad \hat{B}_{Md\times Md} = \begin{pmatrix} B & 0 & \cdots & 0 \\ 0 & B & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & B \end{pmatrix}$$
$$D_{p\times Md} = \left(D^{(1)},\ D^{(2)},\ \cdots,\ D^{(d)}\right)$$

Finding the eigenvalue and the eigenfunction in the Fredholm integral equation is reduced to solving the equation

$$DA = \Lambda D\hat{B} \tag{5.3.9}$$

This equation matches the one in the univariate case and is also a generalised algebraic eigenvalue equation.

Expansion Method 2: the Haar Wavelet method

Under the same notation as chapter 2, the eigenfunction, using the Haar wavelet basis function $\{\psi_k(t)\}$, can be expanded as

$$f_i^{(m)}(t) = \sum_{k=1}^{M} d_{ik}^{(m)}\psi_k(t) = \Psi^T(t)D_i^{(m)} \tag{5.3.10}$$

where

$$\Psi^T(t) = \left(\psi_1(t),\ \psi_2(t),\ \cdots,\ \psi_M(t)\right)$$
$$(D_i^{(m)})^T = \left(d_{i1}^{(m)},\ d_{i2}^{(m)},\ \cdots,\ d_{iM}^{(m)}\right)$$

Using the Haar wavelet basis function, the covariance function can be expressed as

$$K^{(a,b)}(s,t) = \sum_{i=1}^{M}\sum_{j=1}^{M} a_{mn}^{(a,b)}\psi_i(s)\psi_j(t) = \Psi^T(s)A^{(a,b)}\Psi(t) \tag{5.3.11}$$

Given data and the basis function, 2-D wavelet transform can be performed to compute the matrix $A^{(a,b)}$, where $A^{(a,b)} = \{a_{ij}^{(a,b)}\}_{M\times M}$. As in the univariate case, M time points need to be sampled at $t_i = \frac{2i-1}{2M}$, $1 \leq i \leq M$ and M should be a number which is the power of two.

The integral equation on X^m for the ith eigenvalue and the ith eigenfunction is

$$\int_0^1 K^{(m,1)}(s,t)f_i^{(1)}ds + \int_0^1 K^{(m,2)}(s,t)f_i^{(2)}ds + \int_0^1 K^{(m,d)}(s,t)f_i^{(d)}ds = \lambda_i f_i^{(m)} \tag{5.3.12}$$

This is equivalent to

$$\Psi^T(t)A^{(m,1)}HD_i^{(1)} + \Psi^T(t)A^{(m,2)}HD_i^{(2)} + \cdots + \Psi^T(t)A^{(m,d)}HD_i^{(d)} = \lambda_i\Psi^T(t)D_i^{(m)} \ , \tag{5.3.13}$$

where $H = \int \Psi(s)\Psi^T(s)ds$ is a diagonal matrix. Or written in a matrix form for the first p eigenvalues and the first p eigenfunctions, equation 5.3.13 can be expressed as

$$D^{(1)}HA^{(m,1)}\Psi(t) + D^{(2)}HA^{(m,2)}\Psi(t) + \cdots + D^{(d)}HA^{(m,d)}\Psi(t) = \Lambda D^{(m)}\Psi(t) \tag{5.3.14}$$

The coefficients of $\Psi(t)$ should be equal to both sides of equation 5.3.14.

$$D^{(1)}HA^{(m,1)} + D^{(2)}HA^{(m,2)} + \cdots + D^{(d)}HA^{(m,d)} = \Lambda D^{(m)} \tag{5.3.15}$$

Multiply both sides of equation 5.3.15 by $H^{\frac{1}{2}}$

$$D^{(1)}H^{\frac{1}{2}}H^{\frac{1}{2}}A^{(m,1)}H^{\frac{1}{2}} + D^{(2)}H^{\frac{1}{2}}H^{\frac{1}{2}}A^{(m,2)}H^{\frac{1}{2}} + \cdots + D^{(d)}H^{\frac{1}{2}}H^{\frac{1}{2}}A^{(m,d)}H^{\frac{1}{2}} = \Lambda D^{(m)}H^{\frac{1}{2}} \tag{5.3.16}$$

Furthermore, for $1 \leq i, j \leq d$, define

$$\begin{aligned} \hat{D}^{(i)} &= D^{(i)} H^{\frac{1}{2}} \\ \hat{A}^{(i,j)} &= H^{\frac{1}{2}} A^{(i,j)} H^{\frac{1}{2}} \end{aligned}$$

Then equation 5.3.16 is also equivalent to

$$\hat{D}^{(1)}\hat{A}^{(m,1)} + \hat{D}^{(2)}\hat{A}^{(m,2)} + \cdots + \hat{D}^{(d)}\hat{A}^{(m,d)} = \Lambda \hat{D}^{(m)} \tag{5.3.17}$$

Denoting

$$\begin{aligned} \hat{A}_{Md \times Md} &= \begin{pmatrix} \hat{A}^{(1,1)} & \hat{A}^{(2,1)} & \cdots & \hat{A}^{(N,1)} \\ \hat{A}^{(1,2)} & \hat{A}^{(2,2)} & \cdots & \hat{A}^{(N,2)} \\ \cdots & & & \\ \hat{A}^{(1,d)} & \hat{A}^{(2,d)} & \cdots & \hat{A}^{(d,d)} \end{pmatrix} \\ \hat{D}_{p \times Md} &= \left(\hat{D}^{(1)}, \ \hat{D}^{(2)}, \ \cdots, \ \hat{D}^{(d)} \right) \end{aligned}$$

Finding the eigenvalue and the eigenfunction of the Fredholm integra equation is reduced to solving

$$\hat{D}\hat{A} = \Lambda \hat{D} \tag{5.3.18}$$

The eigenfunction can now be written as

$$f^{(m)}(t) = D^{(m)}\Psi^T(t) = \hat{D}^{(m)} H^{-\frac{1}{2}} \Psi^T(t) \ , \tag{5.3.19}$$

where $f^{(m)}(t)^T = \left(f_1^{(m)}(t), \ f_2^{(m)}(t), \ \cdots, \ f_p^{(m)}(t) \right)$

Example: The bivariate Ornstein-Uhlenbeck process

Assume that there is a bivariate Ornstein-Uhlenbeck process $\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}$ with the covariance matrix $\exp(-|t-s|) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. ρ is chosen to be 0.3 here. Since the analytical solution has been calculated in section 5.1, the analytical solution can be used

λ_i	1	2	3	4	5	6
Analytical	0.9604541	0.5171676	0.1794049	0.0966026	0.058615	0.0315619
Trapezium($n = 200$)	0.9618451	0.5179166	0.1794054	0.0966029	0.058835	0.0316804
Error	0.14%	0.14%	0.00%	0.00%	0.38%	0.38%
Haar (M=256)	0.9604591	0.5171703	0.179409	0.0966048	0.0586186	0.0315639
Error	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%

Table 5.1: The eigenvalue comparison between the analytical solution and that from the numerical methods for the bivariate Ornstein-Uhlenbeck process.

to compare with the solution from the numerical methods. We mainly concentrate on the performance of two numerical methods here. One is the integration method using the trapezium rule, while the other is the Haar wavelet method, because both of them provide good result in the univariate case.

The performance of the eigenvalues λ_i, $i \geq 1$ is analysed first. Table 5.1 provides a comparison for the eigenvalues between solutions from the numerical methods and the analytical solutions.

The error is defined in the same way as that in chapter 2

$$\text{Error} = \left| \frac{\text{Numerical Solution} - \text{Analytical Solution}}{\text{Analytical Solution}} \right| \times 100\% \tag{5.3.20}$$

It can be seen that both these methods perform consistently well again with the difference to the analytical solution smaller than 1%. The Haar wavelet method outperforms the trapezium method in this case, and its difference with the analytical solution does not appear until 5th decimal and onwards.

For eigenfunctions, the first two eigenfunctions are plotted for both $X(t)$ and $Y(t)$ in three different figures. Figure 5.1, 5.2 and 5.3 show the eigenfunctions using the analytical method, the trapezium method and the Haar wavelet method respectively. Notice that in all these three plots, the first eigenfunctions are the same for both $X(t)$ and $Y(t)$, while the second eigenfunctions are different in sign, but same in modula.

$l_{1,i}$	1	2	3	4	5	6
Trapezium	0.0014685	0.0014685	0.0000047	0.0000047	0.0037380	0.0037380
Haar	0.0000635	0.0000635	0.0001779	0.0001779	0.0000517	0.0000517

Table 5.2: The difference of the eigenfunction between the analytical solution and the numerical solution for the bivariate Ornstein-Uhlenbeck process in terms of l_1 measure.

This matches our analytical result.

The results by all these three methods are very close to each other. The difference between the analytical method and the numerical method can be measure by the l_1 and l_2, which has been mentioned in chapter 2 and defined as follows

$$l_{1,i}^{X^{(m)}} = \frac{1}{n}\left|\sum_{j=1}^{n}[f_i^{(m)}(t_j) - \hat{f}_i^{(m)}(t_j)]\right|$$

$$l_{2,i}^{X^{(m)}} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}[f_i^{(m)}(t_j) - \hat{f}_i^{(m)}(t_j)]^2}\ ,$$

where both $l_{1,i}^{X^{(m)}}$ and $l_{2,i}^{X^{(m)}}$ measure the difference of the ith eigenfunction between its analytical solution, $f_i^{(m)}(t_j)$ $1 \leq j \leq n$, and its numerical approximation, $\hat{f}_i^{(m)}(t_j)$ $1 \leq j \leq n$, for the process $X^{(m)}(t)$. In this example $t_j = \frac{j-1}{n-1}$, where $1 \leq j \leq n$ and $n = 202$.

The powers in the summation of $l_{1,i}^{X^{(m)}}$ and $l_{2,i}^{X^{(m)}}$ are 1 and 2 respectively, hence the information they provide is similar to the information provided by the bias and the standard deviation respectively. According to theorem 5.1.3, in this example the eigenfunction for $X(t)$ and $Y(t)$ should be the same except for the sign, hence $l_{k,i}^{X}$ and $l_{k,i}^{Y}$ ($k = 1, 2$ and $i \geq 1$) should be equal to each other and they can be further denoted as $l_{k,i}$, where $l_{k,i} = l_{k,i}^{X} = l_{k,i}^{Y}$, $k = 1, 2$ and $i \geq 1$.

Table 5.2 and 5.3 measure the difference of the first six eigenfunctions between

$l_{2,i}$	1	2	3	4	5	6
Trapezium	0.0017583	0.0017583	0.0000052	0.0000052	0.0126944	0.0126944
Haar	0.0004141	0.0004141	0.0025289	0.0025289	0.0049991	0.0049991

Table 5.3: The difference of the eigenfunction between the analytical solution and the numerical solution for the bivariate Ornstein-Uhlenbeck process in terms of l_2 measure.

the analytical solutions and solutions from the integral method and the Haar wavelet method, in terms of l_1 and l_2. In general, both l_1 and l_2 are quite small, which implies good approximations using both of the numerical methods. For the first two eigenfunctions and the fifth and the sixth eigenfunction, both l_1 and l_2 using the Haar wavelet method is smaller than that using the trapezium integral method. It means that the bias and the standard deviation for the eigenfunction approximation is smaller using the Haar wavelet method. For the third and the fourth eigenfunction, on the contrary, the trapezium integral method outperforms the Haar wavelet method, since both l_1 and l_2 is smaller using the trapezium integral method.

Figure 5.4 shows the reconstruction of the covariance matrix for $X(t)$ and the cross covariance matrix for $X(t)$ and $Y(t)$ using the trapezium method and the Haar wavelet method. The difference between two methods is still very small. As what is expected, the cross covariance matrix is actually squashed by the covariance matrix by $\rho = 0.3$.

At the same time, the difference between the analytical covariance matrix and the numerical approximation can be further checked using $l_1^c(\mathbf{X}, \mathbf{Y})$ and $l_2^c(\mathbf{X}, \mathbf{Y})$, where both the process $\mathbf{X}(t)$ and the process $\mathbf{Y}(t)$ takes the value at time point t_i, $1 \leq i \leq n$ for the covariance function if the analytical solution is known, or the covariance reconstruction if the numerical methods are used. As is mentioned in

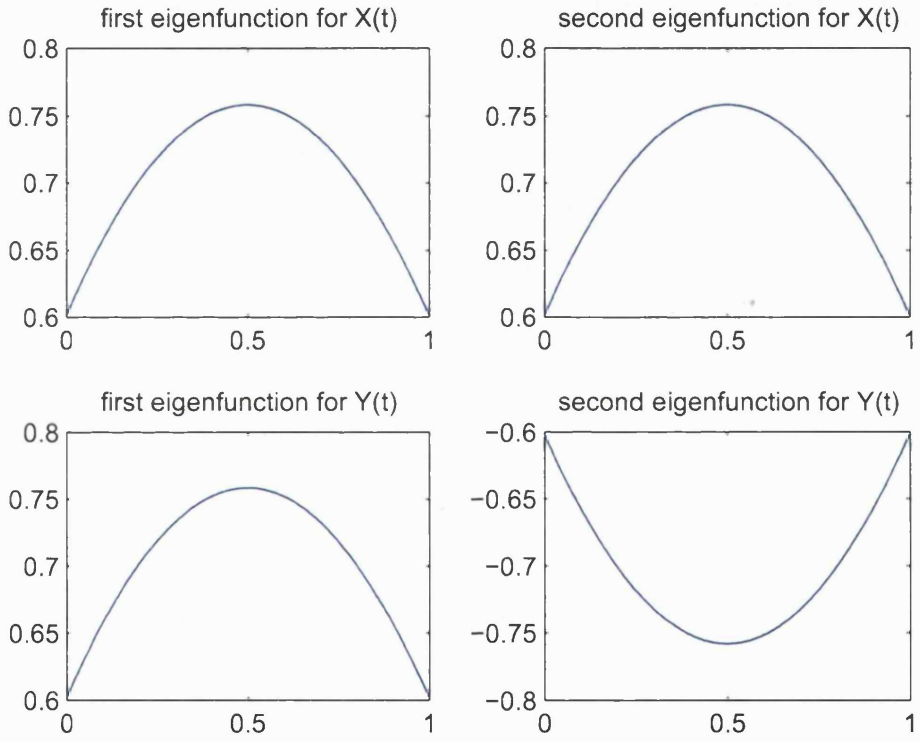

Figure 5.1: Comparison of the first two eigenfunctions of the bivariate Ornstein-Uhlenbeck process, the analytical method.

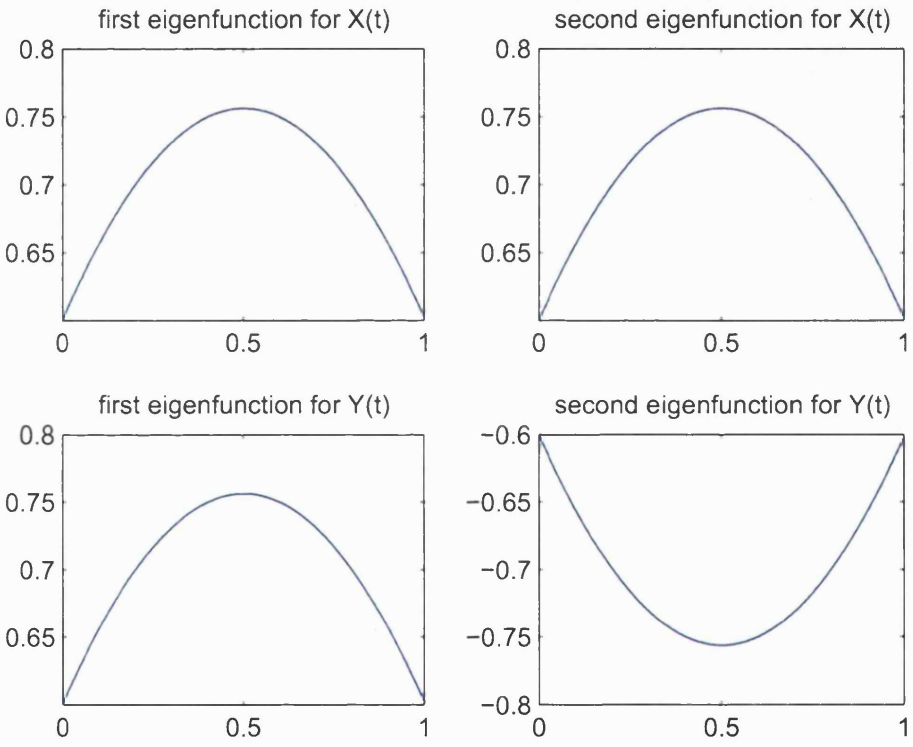

Figure 5.2: Comparison of the first two eigenfunctions of the bivariate Ornstein-Uhlenbeck process, the trapezium integral method ($n = 200$).

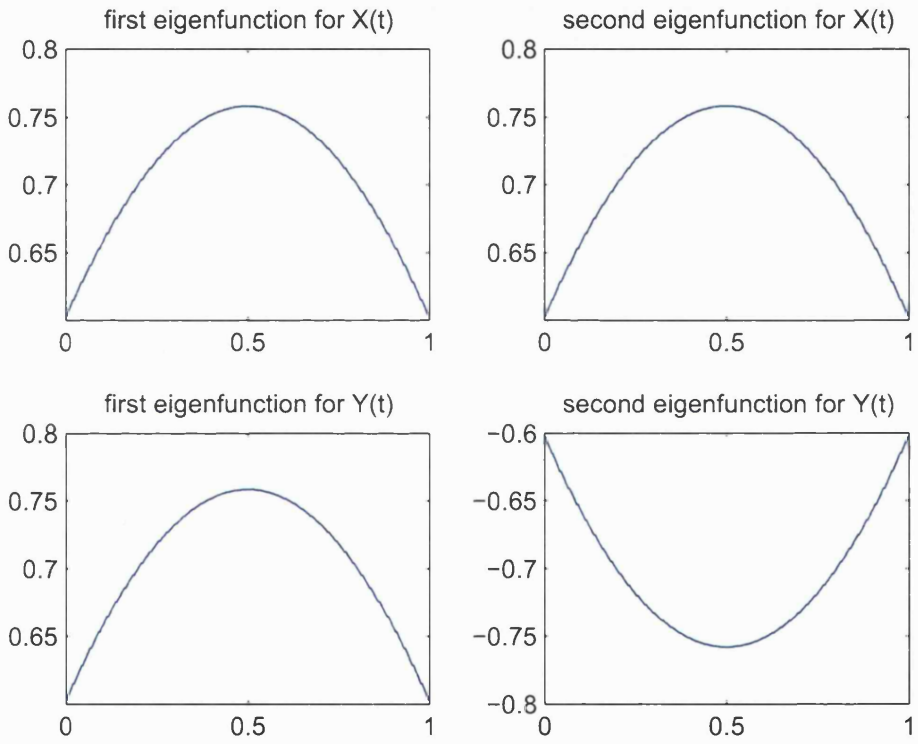

Figure 5.3: Comparison of the first two eigenfunctions of the bivariate Ornstein-Uhlenbeck process, the Haar wavelet($M = 256$).

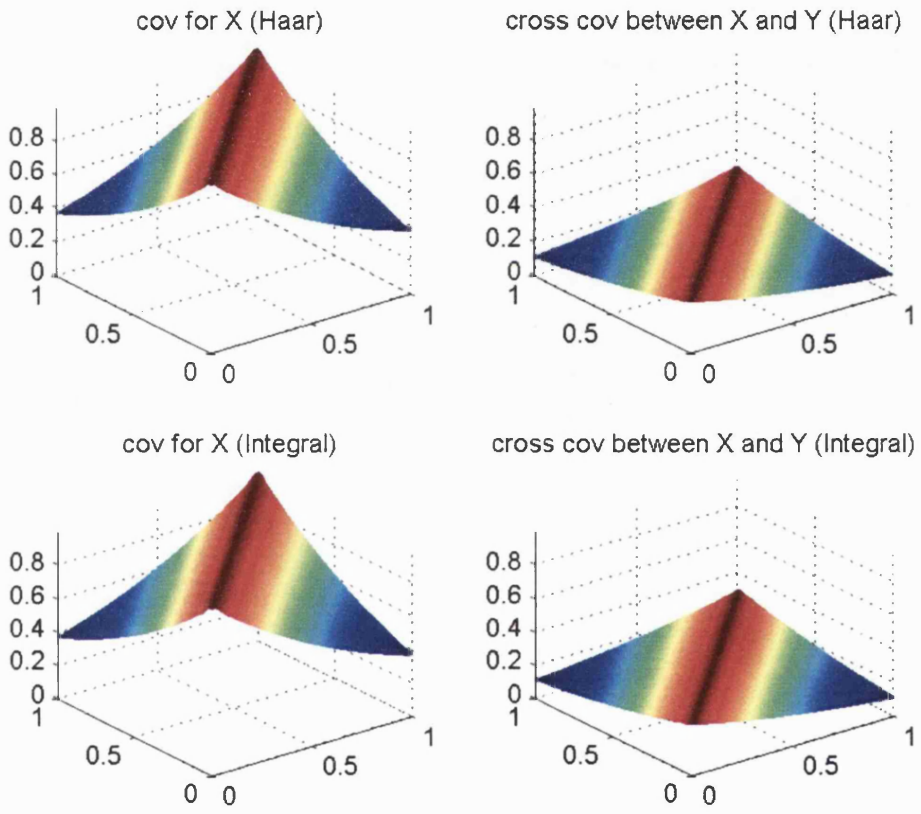

Figure 5.4: Covariance matrix reconstruction of the bivariate Ornstein-Uhlenbeck process between the trapezium integral method ($n = 200$) and the Haar wavelet method ($M = 256$), when the order for the Karhunen-Loeve expansion is $p = 50$.

chapter 2, $l_1^c(\mathbf{X},\mathbf{Y})$ and $l_2^c(\mathbf{X},\mathbf{Y})$ are defined similar to l_1 and l_2.

$$l_{1,p}^c(\mathbf{X},\mathbf{Y}) = \left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\mathrm{cov}[X(t_i),Y(t_j)] - \widehat{\mathrm{cov}}_p[X(t_i),Y(t_j)]\}\right|$$

$$l_{2,p}^c(\mathbf{X},\mathbf{Y}) = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\mathrm{cov}[X(t_i),Y(t_j)] - \widehat{\mathrm{cov}}_p[X(t_i),Y(t_j)]\}^2}\ ,$$

where $\mathrm{cov}[X(t_i),Y(t_j)]$ represents the analytical covariance function between $X(t_i)$ and $Y(t_j)$, while $\widehat{\mathrm{cov}}_p[X(t_i),Y(t_j)]$ represents the covariance function using the numerical approximation between $X(t_i)$ and $Y(t_j)$. The subscript p is the order for the Karhunen-Loeve expansion when approximating $\widehat{\mathrm{cov}}_p[X(t_i),Y(t_j)]$. In this example $t_j = \frac{j-1}{n-1}$, where $1 \leq j \leq n$ and $n = 202$. Again, the information provided by $l_{1,p}^c(\mathbf{X},\mathbf{Y})$ and $l_{2,p}^c(\mathbf{X},\mathbf{Y})$ are similar to the information provided by the bias and the standard deviation respectively.

Table 5.4 and table 5.5 show $l_{1,p}^c$ and $l_{2,p}^c$ for the covariance of $\mathbf{X}(t)$ and the cross covariance between $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, versus the order p in the Karhunen-Loeve expansion using the trapezium method and the Haar wavelet method. For $l_{1,p}^c$, in both the covariance and the cross covariance, it almost remains the same whatever the order p is. Since $l_{1,p}^c$ provides the information on bias, the bias using the Haar wavelet method is smaller than that using the integral method. For $l_{2,p}^c$, in both the covariance and the cross covariance, it decreases when the order p increases. The rate of the decrease is not linear with p. For example, to $l_{2,p}^c(\mathbf{X},\mathbf{X})$ using the integral method, it reduces from 0.01 to 0.004 when p increases from 10 and 20, while the further reduction, when p increases to 30, is only about 0.0013, ending at 0.0027 when p is 30. Since $l_{2,p}^c$ to both numerical methods are quite close to each other, the performance of both methods should also be quite similar in terms of the standard deviation.

$l^c_{1,i}$	10	20	30	40	50
$l^c_{1,p}(\mathbf{X},\mathbf{X})$, Trapezium	0.0004625	0.0004607	0.0004607	0.0004607	0.0004607
$l^c_{1,p}(\mathbf{X},\mathbf{Y})$, Trapezium	0.0001370	0.0001382	0.0001382	0.0001382	0.0001382
$l^c_{1,p}(\mathbf{X},\mathbf{X})$, Haar	0.0001226	0.0001265	0.0001267	0.0001268	0.0001269
$l^c_{1,p}(\mathbf{X},\mathbf{Y})$, Haar	0.0000398	0.0000380	0.0000381	0.0000381	0.0000381

Table 5.4: The difference of the covariance between the analytical solution and the numerical solution for the bivariate Ornstein-Uhlenbeck process in terms of l^c_1 measure.

$l^c_{2,i}$	10	20	30	40	50
$l^c_{2,p}(\mathbf{X},\mathbf{X})$, Trapezium	0.0104989	0.0040907	0.0026866	0.0022628	0.0021003
$l^c_{2,p}(\mathbf{X},\mathbf{Y})$, Trapezium	0.0058256	0.0016355	0.0010530	0.0008267	0.0007232
$l^c_{2,p}(\mathbf{X},\mathbf{X})$, Haar	0.0104070	0.0038221	0.0022435	0.0017045	0.0014753
$l^c_{2,p}(\mathbf{X},\mathbf{Y})$, Haar	0.0058122	0.0015714	0.0009473	0.0006863	0.0005589

Table 5.5: The difference of the covariance between the analytical solution and the numerical solution for the bivariate Ornstein-Uhlenbeck process in terms of l^c_2 measure.

	1st	2nd	3rd	4th	5th	6th
Trapezium ($n = 200$)	1.7517087	0.6256485	0.2547128	0.2267902	0.0909744	0.0329771
Haar (M=256)	1.745239	0.6233377	0.2547053	0.2259525	0.0909717	0.0329618
Diff	0.37%	0.37%	0.00%	0.37%	0.00%	0.05%

Table 5.6: The eigenvalue comparison for the tri-variate process with Gaussian kernel between the trapezium method and the Haar wavelet method.

Example: The tri-variate process with Gaussian kernel

When the dimension of the process is more than 2, i.e. $d > 2$, it is harder to derive the analytical solution. An example, whose analytical solution is untractable, is presented now. The example analyses a tri-variate process $\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix}$ with the covariance matrix $\exp[-(s-t)^2] \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}$. Assume that $\rho_1 = 0.3$, $\rho_2 = 0.5$ and $\rho_3 = 0.7$. Notice that the covariance kernel here is a generalisation of the squared exponential kernel $\exp(-(t-s)^2)$ in chapter 2. The performance of the numerical solutions using the trapezium integral method and the Haar wavelet method will be briefly discussed.

For eigenvalues, table 5.6 looks at the difference between two numerical methods in this case, since no analytical solution can be found and therefore be regarded as a standard. Diff in the table is defined in the same way as that in chapter 2.

$$\text{Diff} = \left| \frac{\text{Trapezium method} - \text{Haar wavelet method}}{\text{Haar wavelet method}} \right| \times 100\% \qquad (5.3.21)$$

Again eigenvalues calculated from both methods are close to each other. However the difference between both methods, which could be observed from the 2nd decimal and onwards, is slightly larger than that in the univariate case.

Also notice that the covariance structure in this example satisfies the condition

λ^{old}	$\gamma_1 = 2.0179834$	$\gamma_2 = 0.7207524$	$\gamma_3 = 0.2612642$
0.8648431	1.7452390	0.6233377	0.2259525
0.1262178	0.2547054	0.0909718	0.0329762

Table 5.7: The first two eigenvalues from the univariate squared exponential kernel $\exp[-(t-s)^2]$ using the Haar wavelet ($M = 256$) and a further calculation for the eigenvalues of the corresponding tri-variate process with Gaussian kernel.

in corollary 5.2.3. Hence the eigenvalues can be alternatively derived from those in the univariate squared exponential kernel $\exp(-(s-t)^2)$. Using corollary 5.2.3, the matrix $\begin{pmatrix} 1 & 0.3 & 0.5 \\ 0.3 & 1 & 0.7 \\ 0.5 & 0.7 & 1 \end{pmatrix}$ needs to be decomposed into $U\Gamma U^T$, where

$$\Gamma = \begin{pmatrix} 2.0179834 & 0 & 0 \\ 0 & 0.7207524 & 0 \\ 0 & 0 & 0.2612642 \end{pmatrix} \quad U = \begin{pmatrix} 0.4896118 & 0.8314071 & 0.2627595 \\ 0.5873968 & -0.5372107 & 0.6052848 \\ 0.6443953 & -0.1420105 & -0.7513906 \end{pmatrix} \tag{5.3.22}$$

Then the eigenvalues can be calculated from $2.0179834\lambda^{\text{old}}$, $0.7207524\lambda^{\text{old}}$ and $0.2612642\lambda^{\text{old}}$, where λ^{old} is the eigenvalue from the univariate squared exponential kernel $\exp(-(s-t)^2)$ (see section 2.3, chapter 2). Table 5.7 shows the first six eigenvalues for this tri-variate process with Gaussian kernel derived from its univariate counterpart. Comparing the result between table 5.7 and table 5.6, the eigenvalues agree very closely using both approaches.

Since no analytical solution is provided to this tri-variate Gaussian process, the reconstruction of the covariance matrix can be implemented to assure that the numerical methods for the truncated Karhunen-Loeve expansion can actually approximate the true Gaussian kernel. Only the Haar wavelet method is used for the reconstruction, since in general it provides consistent result and regards the eigenfunction as a

$l^c_{1,p}$	1	2	4	6	8
$l^c_{1,p}(\mathbf{X},\mathbf{X})$	0.44412172	0.0154236	0.0000980	0.0000980	0.000141
$l^c_{1,p}(\mathbf{X},\mathbf{Y})$	0.2412762	0.0357256	0.0000294	0.0000294	0.0000406
$l^c_{1,p}(\mathbf{X},\mathbf{Z})$	0.1176594	0.0444346	0.0000490	0.0000490	0.0000730
$l^c_{1,p}(\mathbf{Y},\mathbf{Z})$	0.0550000	0.1023139	0.0000686	0.0000686	0.00010416
$l^c_{1,p}$	10	20	30	40	50
$l^c_{1,p}(\mathbf{X},\mathbf{X})$	0.0001414	0.0001414	0.0001414	0.0001414	0.0001414
$l^c_{1,p}(\mathbf{X},\mathbf{Y})$	0.0000424	0.0000424	0.0000424	0.0000424	0.0000424
$l^c_{1,p}(\mathbf{X},\mathbf{Z})$	0.0000707	0.0000707	0.0000707	0.0000707	0.0000707
$l^c_{1,p}(\mathbf{Y},\mathbf{Z})$	0.0000990	0.0000990	0.0000990	0.0000990	0.0000990

Table 5.8: The difference of the covariance reconstruction between the analytical solution and the numerical solution using the Haar wavelet method, for the tri-variate process with Gaussian kernel in terms of l^c_1 measure.

function indeed.

Figure 5.5 shows the reconstruction. Again as what is expected, the plot for the cross covariance between $X(t)$ and $Y(t)$ (top right plot), the plot for the cross covariance between $X(t)$ and $Z(t)$ (bottom left plot) and the plot for the cross covariance between $Y(t)$ and $Z(t)$ (bottom right plot), are squashed compared with the plot of the covariance of $X(t)$ (top left plot) by the corresponding proportion, $\rho_1 = 0.3$, $\rho_2 = 0.5$ and $\rho_3 = 0.7$ respectively.

Further checking of the covariance reconstruction using the Haar wavelet method can be performed by l^c_1 measure and l^c_2 measure. In terms of l^c_1, which can be regarded as the bias, whatever the covariance function or the cross covariance function, l^c_1 decreases dramatically in the first few orders. The reduction of $l^c_{1,p}(\mathbf{X},\mathbf{X})$ is about 0.429 when the order p increases from 1 to 2, while the reduction of $l^c_{1,p}(\mathbf{X},\mathbf{X})$ is only about 0.015 when the order p increases from 2 to 4. From the order $p = 2$ to the order $p = 8$, $l^c_{1,p}(\mathbf{X},\mathbf{X})$ decreases first, and then increases. When p is bigger than

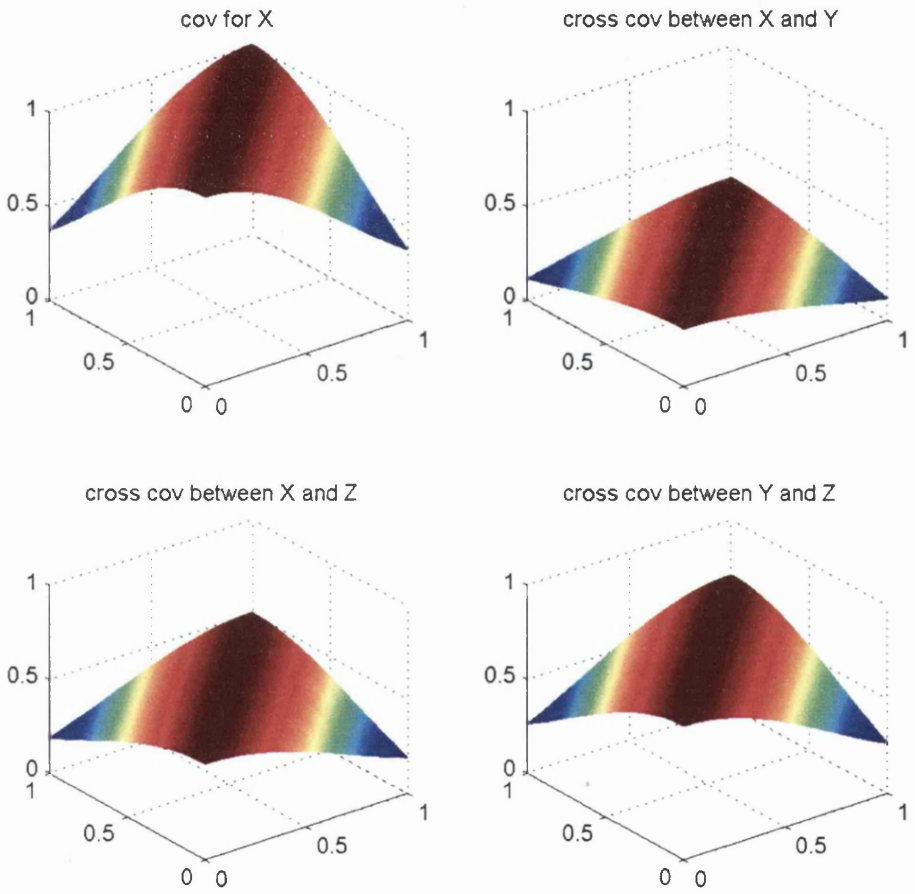

Figure 5.5: Covariance reconstruction for the tri-vairate process with Gaussian kernel using the Haar wavelets ($M = 256$), while the order for the Karhunen-Loeve expansion is $p = 50$.

$l^c_{2,p}$	1	2	4	6	8
$l^c_{2,p}(\mathbf{X},\mathbf{X})$	0.4637025	0.1283918	0.0662803	0.0087658	0.0010238
$l^c_{2,p}(\mathbf{X},\mathbf{Y})$	0.2452395	0.0524192	0.0356934	0.0026297	0.0004755
$l^c_{2,p}(\mathbf{X},\mathbf{Z})$	0.1342348	0.0778016	0.0178559	0.0043829	0.0006684
$l^c_{2,p}(\mathbf{Y},\mathbf{Z})$	0.1049386	0.1360755	0.0101084	0.0061360	0.0012397
$l^c_{2,p}$	10	20	30	40	50
$l^c_{2,p}(\mathbf{X},\mathbf{X})$	0.0009552	0.0009347	0.0009347	0.0009347	0.0009347
$l^c_{2,p}(\mathbf{X},\mathbf{Y})$	0.0002996	0.0002804	0.0002804	0.0002804	0.0002804
$l^c_{2,p}(\mathbf{X},\mathbf{Z})$	0.0004702	0.0004674	0.0004674	0.0004674	0.0004674
$l^c_{2,p}(\mathbf{Y},\mathbf{Z})$	0.0006548	0.0006543	0.0006543	0.0061360	0.0006543

Table 5.9: The difference of the covariance reconstruction between the analytical solution and the numerical solution using the Haar wavelet method, for the tri-variate process with Gaussian kernel in terms of l^c_2 measure.

10, $l^c_{1,p}(\mathbf{X},\mathbf{X})$ almost becomes a constant. Similar reduction trends exist in other $l^c_{1,p}$ measure with respect to the cross covariance functions.

The reason for the dramatic reduction of l^c_1 in the first few orders can be explained as follows. Assume that the ith eigenfunction for the vector process $\mathbf{X}(t)$, $t \in \mathcal{T}$ is $f(t)$. Then $\mathbf{K}_1(t,s)$ and $\mathbf{K}^p_2(t,s)$, the full covariance matrix and the truncated covariance matrix at order p between time t and time s respectively, are expressed as

$$[\mathbf{K}_1(t,s)]_{i,j} = \mathrm{cov}(X_i(t), X_j(s)) \quad [\mathbf{K}^p_2(t,s)]_{i,j} = \sum_{i=1}^{p} \lambda_i f_i(t) f_i(s)^T \tag{5.3.23}$$

Therefore,

$$\mathbf{K}_1(t,s) = \sum_{i=1}^{\infty} \lambda_i f_i(t) f_i(s)^T \quad \mathbf{K}^p_2(t,s) = \sum_{i=1}^{p} \lambda_i f_i(t) f_i(s)^T \tag{5.3.24}$$

Integrate $\mathbf{K}_1(t,t)$ and apply the trace,

$$\begin{aligned} \mathrm{trace}[\int_{\mathcal{T}} \mathbf{K}_1(t,t)dt] &= \sum_{i=1}^{\infty} \lambda_i \int_{\mathcal{T}} \mathrm{trace}[f_i(t) f_i^T(t)]dt = \sum_{i=1}^{\infty} \lambda_i \mathrm{trace}[\int_{\mathcal{T}} f_i(t)^T f_i(t)dt] \\ &= \sum_{i=1}^{\infty} \lambda_i \end{aligned}$$

Similarly, integrate $\mathbf{K}_2^p(t,s)$ and apply the trace,

$$\text{trace}[\int_{\mathcal{T}} \mathbf{K}_2^p(t,t)dt] = \sum_{i=1}^{p} \lambda_i \tag{5.3.25}$$

Combine the results for $\mathbf{K}_2^p(t,s)$ and $\mathbf{K}_1(t,s)$ together,

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} = \frac{\text{trace}[\int_{\mathcal{T}} \mathbf{K}_2^p(t,t)dt]}{\text{trace}[\int_{\mathcal{T}} \mathbf{K}_1(t,t)dt]} \tag{5.3.26}$$

Equation 5.3.26 can be regarded as an generalisation of equation 2.3.1 in the univariate setting. It can also be interpreted as the cumulative expected variance explained by the first p eigenvalues. In this example, $\mathcal{T} = [0,1]$. Since the diagonal elements of $\mathbf{K}_1(t,s)$ are all equal to $\exp[-(t-s)^2]$, then

$$\text{trace}[\int_{\mathcal{T}} \mathbf{K}_1(t,t)dt] = 3 \times \int_0^1 \exp[-(t-t)^2]dt = 3 \tag{5.3.27}$$

Use the first 6 eigenvalues listed in table 5.6.

$$\frac{\sum_{i=1}^{6} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} = 0.9911 \tag{5.3.28}$$

It means that the first six eigenvalues already explains more than 99% of the cumulative expected variance. This explains why $l_{1,p}^c$, which can be treated as bias, reduces dramatically for the first few orders and then remains very small.

Wherever for the covariance function or for the cross covariance function, the other measure $l_{2,p}^c$, which can be regarded as a measure for the standard deviation, decreases when p increases from 1 to 20. When p is bigger than 20, $l_{2,p}^c$ varies little and performs almost like constants. Again, $l_{2,p}^c$ decreases dramatically in the first few orders. This could also be explained by the cumulative expected variance, which is more than 99% using the first six eigenvalues.

5.4 The analytical multivariate Karhunen-Loeve expansion for the stationary linear stochastic differential equation (SDE)

In the above section, theorem 5.2.2 is derived for the analytical solution to the multivariate state univariate time Karhunen-Loeve expansion from its corresponding solution to the univariate Karhunen-Loeve expansion. If a correlated vector process $\mathbf{X}(t)$ can be written down as a linear combination of the independent vector process $\mathbf{Y}(t)$, i.e. $\mathbf{X}(t) = A\mathbf{Y}(t)$ and A is orthogonal, the multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ can be expressed in terms of a linear combination of the univariate Karhunen-Loeve expansion of each component of $\mathbf{Y}(t)$. In this section, theorem 5.2.2 will be related to the setting of the stationary linear stochastic differential equation (in the narrow sense). The stationary linear stochastic differential equation (in the narrow sense), according to Arnold (1974), is defined in the following theorem.

Theorem 5.4.1. *Denote* $\mathbf{W}(t)$ *as an independent d dimensional Brownian motion. A stochastic differential equation*

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + Fd\mathbf{W}(t) \tag{5.4.1}$$

is a linear stationary Gaussian stochastic differential equation (in the narrow sense) for a d dimensional vector process $\mathbf{X}(t)$*, if the eigenvalues of A have negative real parts, with initial condition* $\mathbf{X}(0) \sim N(0, K)$*. K solves the following equation.*

$$AK + KA^T = -FF^T \tag{5.4.2}$$

or written as $K = \int_0^\infty \exp(At)FF^T\exp(A^Tt)dt$

Then for the process $\mathbf{X}(t)$,

$$E(\mathbf{X}(t)) = 0 \tag{5.4.3}$$

$$Var(\mathbf{X}(s), \mathbf{X}(t)) = \begin{cases} K\exp[A^T(t-s)] & t > s \\ \exp[A(s-t)]K & s > t \end{cases} \tag{5.4.4}$$

The proof of theorem 5.4.1 provided in this section is shorter than that provided by Arnold (1974), since we only concern with the special case when the linear term A can be diagonalised and all the eigenvalues of A are real. The proof using is through a transform of $\mathbf{X}(t)$ to another process $\mathbf{Y}(t) = U^{-1}\mathbf{X}(t)$, where U is the eigenvector matrix of A with each column of U corresponding to a eigenvector. The stationarity of $\mathbf{Y}(t)$ will be proven first. We shall use the following lemma

Lemma 5.4.2. *Consider a stationary Gaussian process*

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + Fd\mathbf{W}(t) \tag{5.4.5}$$

If $\mathbf{Y}(t)$ *is a linear combination of* $\mathbf{X}(t)$, *i.e.* $\mathbf{Y}(t) = L\mathbf{X}(t)$, $\mathbf{Y}(t)$ *is still a stationary Gaussian process when* L *is invertible.*

Proof.

$$\begin{aligned} d\mathbf{Y}(t) &= Ld\mathbf{X}(t) = LA\mathbf{X}(t)dt + LFd\mathbf{W}(t) \\ &= LAL^{-1}L\mathbf{X}(t)dt + LFd\mathbf{W}(t) \\ &= A_2\mathbf{Y}(t)dt + F_2d\mathbf{W}(t) \end{aligned}$$

where $A_2 = LAL^{-1}$ and $F_2 = LF$. If the initial condition for $\mathbf{X}(t)$ is $\mathbf{X}(0) \sim N(0, K)$, the initial condition for $\mathbf{Y}(t)$ is $\mathbf{Y}(0) \sim N(0, K_2)$,where $K_2 = LKL^T$. Since

$E(\mathbf{Y}(t)) = LE(\mathbf{X}(t)) = 0$ and $\text{Var}(\mathbf{Y}(0)) = \text{Var}(LX(0)) = L\text{Var}(\mathbf{X}(0))L^T = LKL^T$. Then K_2 satisfies the following equation.

$$A_2K_2 + K_2A_2^T = LAL^{-1}LKL^T + LKL^T(L^{-1})^TA^TL^T \tag{5.4.6}$$

$$= L(AK + KA^T)L^T = -LFF^TL^T = -F_2F_2^T \tag{5.4.7}$$

Moreover, A and A_2 share the same eigenvalue. Let γ be an eigenvalue of A_2. Then γ solves the following equation

$$0 = |A_2 - \gamma I| = |LAL^{-1} - \gamma LL^{-1}| = |L(A - \gamma I)L^{-1}| \tag{5.4.8}$$

Hence when A has negative eigenvalues for the stationarity of $\mathbf{X}(t)$, $A_2 = LAL^{-1}$ also has the same negative eigenvalues. Since K_2 satisfies equation 5.4.7 and A has negative eigenvalue, applying theorem 5.4.1, $\mathbf{Y}(t)$ is a stationary Gaussian process.

□

Theorem 5.4.1 can be proved when A can be diagonalised and all the eigenvalues of A are real.

Proof. (Proof of theorem 5.4.1)

We have the SDE for the d dimensional multivariate process $\mathbf{X}(t)$

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + Fd\mathbf{W}(t) \tag{5.4.9}$$

with initial condition $\mathbf{X}(0) \sim N(0, K)$.

Assume that matrix A has negative eigenvalues and can be diagonalised , i.e.

$U^{-1}AU = \Gamma$, where

$$\Gamma = \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \gamma_d \end{pmatrix} \tag{5.4.10}$$

In order to calculate matrix exponential, a new random vector $\mathbf{Y}(t) = U^{-1}\mathbf{X}(t)$ is defined. We only need to prove that $\mathbf{Y}(t)$ is a stationary Gaussian process, since $\mathbf{X}(t)$ is a linear combination of $\mathbf{Y}(t)$ and $\mathbf{X}(t) = U\mathbf{Y}(t)$. Using Ito's lemma, the SDE for $\mathbf{Y}(t)$ is

$$\begin{aligned} d\mathbf{Y}(t) &= U^{-1}d\mathbf{X}(t) = U^{-1}A\mathbf{X}(t)dt + U^{-1}Fd\mathbf{W}(t) \\ &= U^{-1}AUU^{-1}\mathbf{X}(t)dt + U^{-1}Fd\mathbf{W}(t) \\ &= \Gamma\mathbf{Y}(t)dt + U^{-1}Fd\mathbf{W}(t) \end{aligned}$$

The solution to the above SDE is given by Arnold (1974).

$$\mathbf{Y}(t) = \Phi(t)[\mathbf{Y}(0) + \int_0^t \Phi(s)^{-1}U^{-1}Fd\mathbf{W}(s)] \ , \tag{5.4.11}$$

where $\Phi(t)$ is the solution to $\frac{d\Phi(t)}{dt} = \Gamma\Phi(t)$ with $\Phi(0) = I$, i.e. $\Phi(t) = \exp(\Gamma t)$. Since Γ is a diagonal matrix,

$$\Phi(t) = \begin{pmatrix} \exp(\gamma_1 t) & 0 & \cdots & 0 \\ 0 & \exp(\gamma_2 t) & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \exp(\gamma_d t) \end{pmatrix} \tag{5.4.12}$$

Notice that $E(\mathbf{Y}(t)) = \Phi(t)E[\mathbf{Y}(0)]$. If the expectation of $E(\mathbf{Y}(t))$ is independent of time t, $E[\mathbf{Y}(0)] = 0$. For covariance function between t and s,

$$\text{cov}(\mathbf{Y}(s), \mathbf{Y}(t)) = \Phi(s)[\text{Var}(\mathbf{Y}(0)) + \int_0^{\min(s,t)} \Phi(u)^{-1}U^{-1}FF^T(U^T)^{-1}\Phi(u)^{-1}du]\Phi(t)^T$$

Assume $V = U^{-1}FF^T(U^T)^{-1} = \{v_{ij}\}$, then

$$\Phi(u)^{-1}U^{-1}FF^TU^{-1}\Phi(u)^{-1} = \Phi(u)^{-1}V\Phi(u)^{-1} = \{v_{ij}\exp[-(\gamma_i+\gamma_j)u]\} \quad (5.4.13)$$

Hence the integration of equation 5.4.13 is

$$\begin{aligned}\int_0^{\min(s,t)} \Phi(u)^{-1}V\Phi(u)^{-1}du &= \{-\frac{v_{ij}\exp[-(\gamma_i+\gamma_j)u]}{\gamma_i+\gamma_j}|_0^{\min(s,t)}\} \\ &= \{-\frac{v_{ij}(\exp[-(\gamma_i+\gamma_j)\min(s,t)]-1)}{\gamma_i+\gamma_j}\}\end{aligned}$$

Therefore,

$$\begin{aligned}&\Phi(s)\int_0^{\min(s,t)} \Phi(u)^{-1}U^{-1}FF^TU^{-1}\Phi(u)^{-1}du\Phi(t)^T \\ =\ &\Phi(s)\{-\frac{v_{ij}(\exp[-(\gamma_i+\gamma_j)\min(s,t)]-1)}{\gamma_i+\gamma_j}\}\Phi(t)^T \\ =\ &\{-\frac{v_{ij}}{\gamma_i+\gamma_j}\exp(\gamma_i s)\exp(\gamma_j t)(\exp[-(\gamma_i+\gamma_j)\min(s,t)]-1)\} \\ =\ &\begin{cases}\{\frac{v_{ij}}{\gamma_i+\gamma_j}[\exp(\gamma_i s+\gamma_j t)-\exp(\gamma_j(t-s))]\} & t>s \\ \{\frac{v_{ij}}{\gamma_i+\gamma_j}[\exp(\gamma_i s+\gamma_j t)-\exp(\gamma_i(s-t))]\} & s>t\end{cases}\end{aligned}$$

If $\mathbf{Y}(t)$ is stationary, $\mathrm{cov}(\mathbf{Y}(s),\mathbf{Y}(t))$ is only dependent on the difference of $s-t$, hence the initial condition for $\mathrm{Var}(\mathbf{Y}(0))$ should be chosen to cancel the term $\{\frac{v_{ij}}{\gamma_i+\gamma_j}\exp(\gamma_i s+\gamma_j t)\}$. Set $K_{\mathbf{Y}} = \mathrm{Var}(\mathbf{Y}(0)) = \{-\frac{v_{ij}}{\gamma_i+\gamma_j}\}$, then $\Phi(s)\mathrm{Var}(\mathbf{Y}(0))\Phi(t)^T = \{-\frac{v_{ij}}{\gamma_i+\gamma_j}\exp(\gamma_i s+\gamma_j t)\}$. Then

$$\mathrm{Var}(\mathbf{Y}(s),\mathbf{Y}(t)) = \begin{cases}\{-\frac{v_{ij}}{\gamma_i+\gamma_j}\exp[\gamma_j(t-s)]\} = K_{\mathbf{Y}}\exp[\Gamma(t-s)] & t>s \\ \{-\frac{v_{ij}}{\gamma_i+\gamma_j}\exp[\gamma_i(s-t)]\} = \exp[\Gamma(s-t)]K_{\mathbf{Y}} & s>t\end{cases} \quad (5.4.14)$$

Notice that $V = U^{-1}FF^T(U^T)^{-1}$ is a symmetric matrix, it means that $v_{ii} > 0$. Since $K_{\mathbf{Y}}$ is a covariance matrix, the diagonal element of it should be non-negative.

This implies that $\gamma_i < 0$. Actually in the matrix form, the initial variance $K_{\mathbf{Y}}$ is also the solution to the equation

$$\Gamma K_{\mathbf{Y}} + K_{\mathbf{Y}}\Gamma = -V \tag{5.4.15}$$

Equation 5.4.15 matches equation 5.4.2 in theorem 5.4.1. The matrix $K_{\mathbf{Y}}$ is chosen so that $\mathrm{Var}(\mathbf{Y}(s), \mathbf{Y}(t))$ is a function of $s - t$. Together with $E(\mathbf{Y}(t)) = 0$, $\mathbf{Y}(t)$ is stationary. Therefore $\mathbf{X}(t)$ is also stationary.

After obtaining $E(\mathbf{Y}(t))$ and $\mathrm{Var}(\mathbf{Y}(s), \mathbf{Y}(t))$, we can calculate the corresponding expectation and variance for $\mathbf{X}(t)$. With the initial variance matrix $K = UK_{\mathbf{Y}}U^T$

$$\begin{aligned} E(\mathbf{X}(t)) &= UE(\mathbf{Y}(t)) = 0 \\ \mathrm{Var}(\mathbf{X}(s), \mathbf{X}(t)) &= U\mathrm{Var}(\mathbf{Y}(s), \mathbf{Y}(t))U^T \end{aligned}$$

When $t > s$,

$$\begin{aligned} \mathrm{Var}(\mathbf{X}(s), \mathbf{X}(t)) &= U\mathrm{Var}(\mathbf{Y}(s), \mathbf{Y}(t))U^T = UK_{\mathbf{Y}}\exp[\Gamma(t-s)]U^T \\ &= UK_{\mathbf{Y}}U^T(U^T)^{-1}\exp[\Gamma(t-s)]U^T = K\exp[A^T(t-s)] \end{aligned}$$

When $t < s$

$$\begin{aligned} \mathrm{Var}(\mathbf{X}(s), \mathbf{X}(t)) &= U\mathrm{Var}(\mathbf{Y}(s), \mathbf{Y}(t))U^T = U\exp[\Gamma(t-s)]K_{\mathbf{Y}}U^T \\ &= U\exp[\Gamma(t-s)]U^{-1}UK_{\mathbf{Y}}U^T = \exp[A(t-s)]K \end{aligned}$$

□

Thus, the process $\mathbf{X}(t)$ has been represented as a linear combination of an analytically tractable process $\mathbf{Y}(t)$. If $\mathbf{Y}(t)$ is an independent process, we can write down either the multivariate Karhunen-Loeve expansion or the Karhunen-Loeve-like expansion depending on whether U satisfies the condition mentioned in remark 5.2.2. The following theorem provides a way of checking whether $\mathbf{Y}(t)$ can be independent.

Theorem 5.4.3. *Assume the SDE for the vector process* $\mathbf{X}(t)$ *is*

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + Fd\mathbf{W}(t) \ , \tag{5.4.16}$$

where A *can be diagonalised and all the eigenvalues of* A *are real and negative.* $\mathbf{Y}(t) = U^{-1}\mathbf{X}(t)$ *is independent if and only if* $FF^T = UDU^T$*, where* U *is the eigenvector matrix for* A *and* D *is any diagonal matrix.*

Proof. $\mathbf{Y}(t)$ is independent if and only if matrix $V = U^{-1}FF^T(U^T)^{-1}$ is diagonal. Assume this diagonal matrix is D, then

$$D = U^{-1}FF^T(U^T)^{-1} \tag{5.4.17}$$

This is equivalent to

$$FF^T = UDU^T \tag{5.4.18}$$

□

Combining the above theorems, under certain condition, we can write down the multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ as a linear combination of the univariate Karhunen-Loeve expansion of $\mathbf{Y}(t)$ of each its components.

Theorem 5.4.4. *Assume that the SDE for the vector process* $\mathbf{X}(t)$ *is*

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + Fd\mathbf{W}(t) \tag{5.4.19}$$

If A *can be diagonalised, i.e.* $U^{-1}AU = \Gamma$*, where* Γ *is diagonal matrix, the eigenvalues of* A *are real and negative and* $FF^T = UDU^T$*, where* D *is any diagonal matrix. The process* $\mathbf{X}(t)$ *can be written down as a linear combination of an independent vector process* $\mathbf{Y}'(t)$ *and* $\mathbf{X}(t) = U_2\mathbf{Y}'(t)$*, where*

(i) $U = U_2 D_2$

(ii) $U_2 U_2^T = I$, *i.e.* U_2 *is an orthogonal matrix*

(iii) $\mathbf{Y}'(t) = D_2 \mathbf{Y}(t)$ *is still an independent vector process, with different eigenvalue for each new process* $\mathbf{Y}_i'(t)$ *after transformation.*

Then the multivariate state univariate time Karhunen-Loeve expansion of $\mathbf{X}(t)$ *is a linear combination of the univariate Karhunen-Loeve expansion of* $\mathbf{Y}'(t)$ *of each its component.*

Example: A **is negative symmetric matrix and** $F = I$

The SDE for the stationary Gaussian process $\mathbf{X}(t)$ can now be expressed as

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + Id\mathbf{W}(t) \tag{5.4.20}$$

Since A is negative symmetric, A can be diagonalised as $U^T A U = \Gamma$, where $U^T U = I$. We choose $D = I$, so that $FF^T = I$ satisfies the condition that $FF^T = UIU^T = I$. The process $\mathbf{X}(t)$ can be expressed as $\mathbf{X}(t) = U\mathbf{Y}(t)$. Since U is orthogonal, the multivariate Karhunen-Loeve expansion of $\mathbf{X}(t)$ can be derived from the univariate Karhunen-Loeve expansion of $\mathbf{Y}(t)$ of each its component, where $\mathbf{Y}(t)$ satisfies SDE

$$d\mathbf{Y}(t) = \Gamma\mathbf{Y}(t)dt + U^T d\mathbf{W}(t) \tag{5.4.21}$$

Example: Conditional independence structure in F **and** $A = -I$

The linear SDE (in the narrow sense) can sometimes be regarded as a continuous version of $AR(1)$ process. For $AR(1)$ process, research on the conditional independence has been performed by Caines and Wynn (2007). This example extends the discrete case to its continuous counterpart in a special situation.

Assume that there is a tri-variate process $\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix}$. The SDE for $\mathbf{X}(t)$

here is written down as

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} = -I \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} dt + F d\mathbf{W}(t) \ , \tag{5.4.22}$$

where I is an identity matrix and

$$F = \begin{pmatrix} f_{11} & 0 & f_{13} \\ 0 & f_{22} & f_{23} \\ 0 & 0 & f_{33} \end{pmatrix} \tag{5.4.23}$$

Since, $A = -I$, any inverse matrix can be its eigenvector matrix. We therefore choose $U = F$, so that $FF^T = UIU^T = FF^T$. The process $\mathbf{X}(t)$ can be expressed as $\mathbf{X}(t) = F\mathbf{Y}(t)$, where $\mathbf{Y}(t)$ satisfies SDE

$$d\mathbf{Y}(t) = -I\mathbf{Y}(t)dt + Id\mathbf{W}(t) \tag{5.4.24}$$

Depending on whether F satisfies the condition in theorem 5.4.4 or not, we can decide whether the multivariate Karhunen-Loeve expansion for $\mathbf{X}(t)$ can be written down in terms of the univariate Karhunen-Loeve expansion.

For the conditional independence structure, only the inverse of the initial covariance matrix K for the process $\mathbf{X}(t)$ needs to be calculated. See, for example, Rue and Held (2005). K can be derived through solving

$$AK + KA^T = -FF^T \ , \tag{5.4.25}$$

where $A = -I$. Hence

$$K = \frac{1}{2}FF^T = \frac{1}{2}\begin{pmatrix} f_{11}^2 + f_{13}^2 & f_{13}f_{23} & f_{13}f_{33} \\ f_{13}f_{23} & f_{22}^2 f_{23}^2 & f_{23}f_{33} \\ f_{13}f_{33} & f_{23}f_{33} & f_{33}^2 \end{pmatrix} \tag{5.4.26}$$

and

$$K^{-1} = \begin{pmatrix} \frac{1}{f_{11}^2} & 0 & -\frac{f_{13}}{f_{33}f_{11}^2} \\ 0 & \frac{1}{f_{22}^2} & -\frac{f_{23}}{f_{33}f_{22}^2} \\ -\frac{f_{13}}{f_{33}f_{11}^2} & -\frac{f_{23}}{f_{33}f_{22}^2} & \frac{f_{11}^2 f_{22}^2 + f_{11}^2 f_{23}^2 + f_{13}^2 f_{22}^2}{f_{11}^2 f_{22}^2 f_{33}^2} \end{pmatrix} \tag{5.4.27}$$

Equation 5.4.27 contains the conditional independent structure and it matches the information in F. Hence one possible conjecture is that the conditional independence structure lies in the deterministic coefficient term of the Karhunen-Loeve expansion, since in this example $\mathbf{X}(t) = F\mathbf{Y}(t)$.

The original research of the conditional independence structure related to the multivariate Karhunen-Loeve expansion is to put the structure into matrix A, rather than matrix F. However, it is very difficult to derive an analytical expression for K^{-1} using the truncated Karhunen-Loeve expansion, and the numerical calculation of K^{-1} does not provide satisfactory results. Further research is needed in this direction.

Chapter 6

Multivariate Functional Data Analysis

The idea of functional data analysis starts from Ramsay and Silverman (1997). Its property and application have drawn the interest from a wide range of researchers, see, for example, Boente and Fraiman (2000), James et al. (2000), Ramsay and Silverman (2002), Yao et al. (2005) and Hall and Hosseini-Nassab (2006). Functional data analysis can be regarded as an extension to a functional version of the principal component analysis. The canonical principal component analysis treats each data point in the data set as an individual discrete point and try to find a direction in which the variance of the data set is best explained, while the functional data analysis regards each data point from the data set as an observation from a continuous process and tries to find certain deterministic functions out of the process so that the variance of the process can be captured mostly by these deterministic functions. In this sense, the main idea of the functional data analysis is essentially the same as the Karhunen-Loeve expansion. In most of the literatures, the applications of functional data analysis have been restricted to the univariate case, where both the state and the time are one-dimension. In the time series applications, however, interest lies

more in finding the relationship among different series. Since most time series in practice only have one-dimensional time, what is called "multivariate" in this chapter restricts to the process with the multivariate state and the univariate time. In summary, this chapter deals with multivariate functional data analysis (MFDA). It draws for intuition on chapter 5, in which a theoretical background of the multivariate Karhunen-Loeve expansion is investigated. Section 1 introduces the data set we use. Section 2 explains the decomposition using the numerical methods and the reconstruction of the covariance and the cross covariance function. Section 3 provides a method for the smoothing and the prediction using the truncated Karhunen-Loeve expansion.

6.1 The introduction to the data

The method discussed in this chapter applies mainly to the weakly stationary time series. Since for the weakly stationary time series, the empirical version of the covariance matrix (see, equation 6.2.1) needed for the numerical methods can be calculated using the autocovariance function, while for the non-stationary time series, the empirical version of the covariance matrix might not be easy to derive. Once the covariance matrix is available, the method can be generalised to any other time series.

In time series analysis, a considerable number of statistical methods are based on the assumption that the series can be transformed approximately to weakly stationarity, so that the statistical analysis, such as the prediction can be performed relatively easily. However, most economics and finance data in reality are far from stationary. They can be observed non-stationary characteristics, such as trend and seasonality. In the case of non-stationarity, transformations, such as de-trending and differencing,

can be made to produce a new series which is approximately stationary.

In the multivariate setting, a d-dimensional series

$$\mathbf{X}(t)^T = \Big(X_1(t),\ \ X_2(t),\ \ \cdots,\ \ X_d(t)\Big), \quad t \in \mathbb{N} \tag{6.1.1}$$

is weakly stationary, if

(i) $E[\mathbf{X}(t)]$ is independent of t.

(ii) $\Gamma(h) = \text{cov}[\mathbf{X}(t+h), \mathbf{X}(t)]$ is independent of t for each h.

The data set used throughout this chapter is a tri-variate $TAR(1)$ process. As a representative process of the multivariate $ARMA$ process, the multivariate $AR(1)$ process has been proved to be stationary and its structure is analytically tractable (see, for example, Brockwell and Davis (2002)). A tri-variate $TAR(1)$ process is defined in the following way.

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} = A \begin{pmatrix} X_1(t-1) \\ X_2(t-1) \\ X_3(t-1) \end{pmatrix} + \epsilon(t), \ , \tag{6.1.2}$$

where $\epsilon(t) \sim N(0, I_{3\times 3})$. For the existence of the unique stationary solution to $TAR(1)$, A is chosen so that all its eigenvalues are less than 1 in absolute value. We further assume that we have $n = 300$ observations, i.e. $t = 1, 2, \cdots, n = 300$. Then the data set can be regarded as a simulated version of the daily observation for about a year time.

As an example of $TAR(1)$, A is chosen as the following matrix,

$$A = \begin{pmatrix} 0.4 & 0 & -0.9 \\ 0 & 0.5 & -0.89 \\ -0.54 & 0 & 0.1 \end{pmatrix} \tag{6.1.3}$$

Its corresponding eigenvalues are 0.5, 0.96309 and -0.46309, which are less than 1 in absolute value.

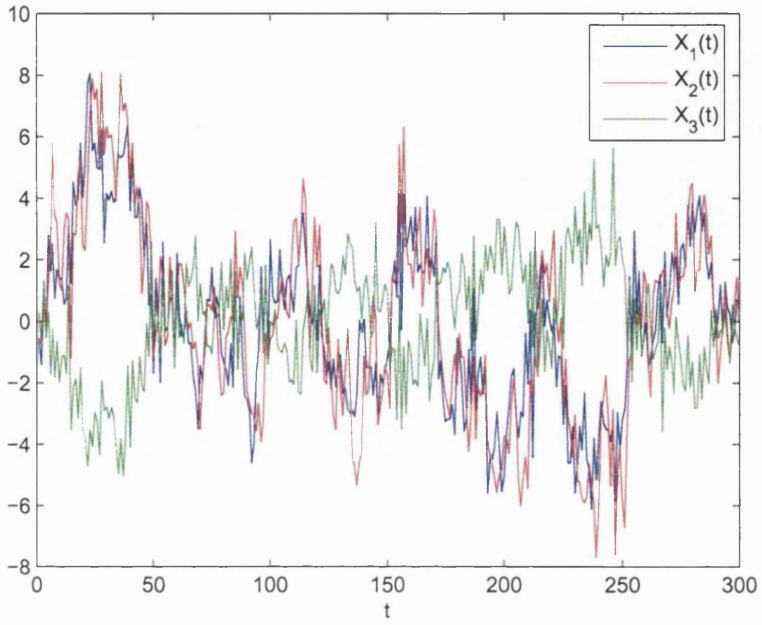

Figure 6.1: Three simulated paths for the series $X_1(t)$, $X_2(t)$ and $X_3(t)$ respectively from $TAR(1)$.

Three correlated paths for the series $X_1(t)$, $X_2(t)$ and $X_3(t)$ $(t = 1, 2, \cdots, 300)$ are simulated from the above $TAR(1)$ with the initial value $X_1(0) = X_2(0) = X_3(0) = 0$. For simplicity of the future work, notation $X_i(t), i = 1, 2, 3$ is further adjusted to denote the series after removing its mean (centerised series), which is calculated as $X_i(t) - \frac{1}{n}\sum_{j=1}^{n} X_i(j)$; $t = 1, 2, \cdots, n$, $i = 1, 2, 3$.

Figure 6.1 shows one sample paths for $X_1(t)$, $X_2(t)$ and $X_3(t)$. It can be seen that $X_1(t)$ and $X_2(t)$ are quite close to each other, while $X_3(t)$ has a negative relationship with the rest two series. This relationship can be roughly checked through the correlation coefficient, which is listed in table 6.1. Although the correlation coefficient is

	$X_1(t)$	$X_2(t)$	$X_3(t)$
$X_1(t)$	1	0.8813226	-0.8022594
$X_2(t)$	0.8813226	1	-0.7926202
$X_3(t)$	-0.8022594	-0.7926202	1

Table 6.1: The correlation coefficients for $X_1(t)$, $X_2(t)$ and $X_3(t)$.

mainly used to check the linear relationship among different series in statistics, it can still provide us a brief idea of how the multivariate series is correlated. From table 6.1, the correlation coefficient between $X_1(t)$ and $X_2(t)$ is about 0.9, while the correlation coefficient between $X_3(t)$ and the rest series is about -0.8. The sample paths simulated in figure 6.1 will be used to demonstrate the truncated Karhunen-Loeve expansion in the covariance reconstruction, the smoothing and the prediction in this chapter.

6.2 Empirical MFDA

In practice, numerical methods play an important role when little is known about the time series data. In this chapter, the numerical method is the trapezium integral method. Although we have shown in the previous chapters that the Haar wavelet method is also numerically accurate and efficient, the number of the points involved to perform the wavelet transform is a power of two. However, in practice, the number of the data points varies, which might make the implementation of the Haar wavelet method inefficient. The trapezium integral method, on the contrary, does not have any restriction on the number of the data points as long as the number is suitable for approximating the integral. The main condition that the trapezium integral method

has is that the data points are equally spaced, which satisfies most time series data sets. Therefore, the trapezium integral method is more practical to use and more flexible when dealing with the real data.

For the multivariate series, the trapezium integral method has been discussed in detail in chapter 5. Although $\mathbf{X}(t)$ are only sampled at the integer points in this chapter, i.e. $t = 1, 2, \cdots, n$, when approximating the Fredholm integral equation, the time interval involved is still assumed to be $[0, 1]$ for accuracy. Therefore, $\mathbf{X}(t)$ can actually be regarded as an observation at time point $\frac{t}{n}$ in the approximation.

In order to approximate the Fredholm integral equation, we need to calculate the covariance matrix K defined in equation 5.3.3 for multivariate series, i.e.

$$K = \begin{pmatrix} \text{cov}(X_1, X_1^T) & \text{cov}(X_1, X_2^T) & \cdots \text{cov}(X_1, X_d^T) \\ \text{cov}(X_2, X_1^T) & \text{cov}(X_2, X_2^T) & \cdots \text{cov}(X_2, X_d^T) \\ \cdots & & \\ \text{cov}(X_d, X_1^T) & \text{cov}(X_d, X_2^T) & \cdots \text{cov}(X_d, X_d^T) \end{pmatrix}, \tag{6.2.1}$$

where

$$X_i = \begin{pmatrix} X_i(1) \\ X_i(2) \\ \cdots \\ X_i(n) \end{pmatrix} \tag{6.2.2}$$

Since $TAR(1)$ is a stationary series, K can be derived through the sample auto-covariance function $\hat{\mathbf{\Gamma}}(h)$ defined as follows

$$\hat{\mathbf{\Gamma}}(h) = \frac{1}{n} \sum_{t=1}^{n-h} [\mathbf{X}(t+h) - \bar{\mathbf{X}}][\mathbf{X}(t) - \bar{\mathbf{X}}]^T \quad \text{for } 0 \leq |h| \leq n-1 \tag{6.2.3}$$

Since all the series have been centerised, $\bar{\mathbf{X}} = 0$. Properties of the sample autocorrelation function can refer to, for example, Brockwell and Davis (2002) and Brockwell and Davis (1991).

Figure 6.2 shows the covariance and the cross covariance function for $X_1(t)$, $X_2(t)$ and $X_3(t)$ using the sample autocovariance function. Since the number of the pairs in estimating the autocovariance function $\hat{\Gamma}(h)$ for lag h is $n-h$, the higher the lag, the less pairs are used in the estimation, hence the more variation there are in the result. It can also be clearly observed from the figure that $X_3(t)$ has negative correlation with the rest two series, while the correlation between $X_1(t)$ and $X_2(t)$ is positive.

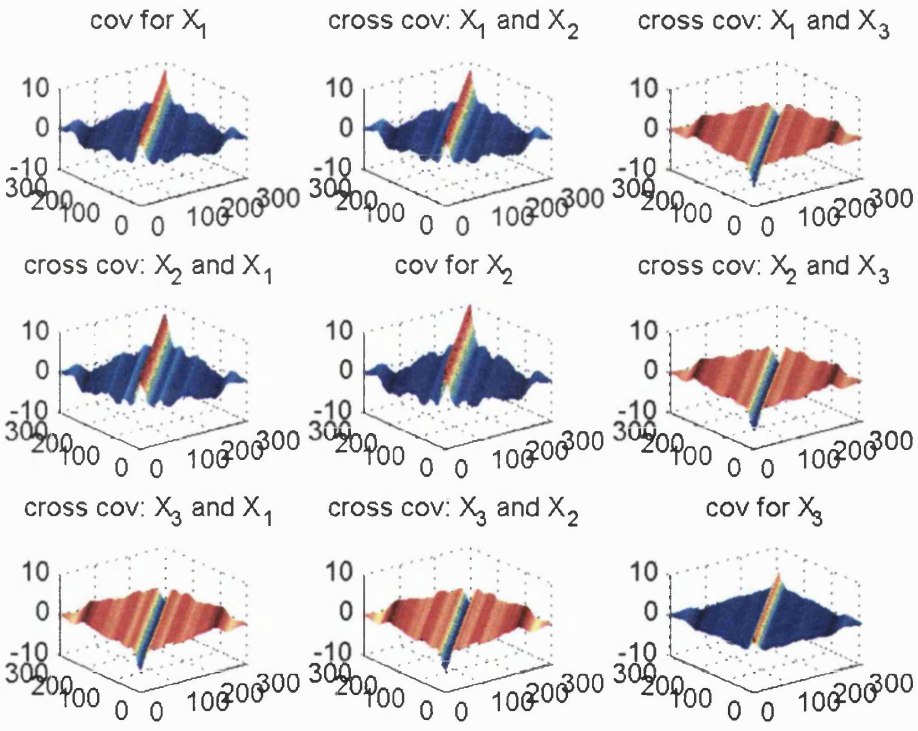

Figure 6.2: The covariance and the cross covariance for $X_1(t)$, $X_2(t)$ and $X_3(t)$ using the autocovariance function.

Using $\hat{\Gamma}(h)$, $0 \leq |h| \leq n-1$, we can estimate the covariance matrix K and denote

the estimator as $\hat{\Gamma}$

$$\hat{\Gamma} = \begin{pmatrix} \hat{\Gamma}_{11} & \hat{\Gamma}_{12} & \cdots & \hat{\Gamma}_{1d} \\ \hat{\Gamma}_{21} & \hat{\Gamma}_{22} & \cdots & \hat{\Gamma}_{2d} \\ \cdots & & & \\ \hat{\Gamma}_{d1} & \hat{\Gamma}_{d2} & \cdots & \hat{\Gamma}_{dd} \end{pmatrix}, \tag{6.2.4}$$

where $\hat{\Gamma}_{ij}$ is

$$\hat{\Gamma}_{ij} = \begin{pmatrix} \hat{\Gamma}_{ij}(0) & \hat{\Gamma}_{ij}(1) & \cdots & \hat{\Gamma}_{ij}(n-1) \\ \hat{\Gamma}_{ij}(1) & \hat{\Gamma}_{ij}(0) & \cdots & \hat{\Gamma}_{ij}(n-2) \\ \cdots & & & \\ \hat{\Gamma}_{ij}(n-1) & \hat{\Gamma}_{ij}(n-2) & \cdots & \hat{\Gamma}_{ij}(0) \end{pmatrix} \quad 1 \leq i, j \leq d \tag{6.2.5}$$

and $\hat{\Gamma}_{ij}(h) = \{\hat{\Gamma}(h)\}_{ij}$. Based on the covariance matrix estimator $\hat{\Gamma}$, the eigen-equation for the eigenvalues and eigenfunctions can be solved. As shown in section 5.3, chapter 5, the eigen-equation is expressed as

$$\hat{\Gamma} W f = \Lambda f \tag{6.2.6}$$

For the trapezium integral method, the notation in equation 6.2.6 is the same as the notation in section 5.3, chapter 5, i.e.

$$f = \begin{pmatrix} f_1^{(1)} & f_2^{(1)} & \cdots & f_{(n+2)d}^{(1)} \\ f_1^{(2)} & f_2^{(2)} & \cdots & f_{(n+2)d}^{(2)} \\ \cdots & & & \\ f_1^{(d)} & f_2^{(d)} & \cdots & f_{(n+2)d}^{(d)} \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \lambda_{(n+2)d} \end{pmatrix}$$

$$W = \begin{pmatrix} W^{(1)} & 0 & \cdots & 0 \\ 0 & W^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & W^{(d)} \end{pmatrix}$$

$$W^{(i)} = \begin{pmatrix} \frac{1}{2(n+1)} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{n+1} & \cdots & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & \cdots & \frac{1}{n+1} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{2(n+1)} \end{pmatrix} \quad 1 \leq i \leq d$$

Figure 6.3 shows the performance of the first 40 eigenvalues after solving the eigen-equation using the $TAR(1)$ data. The left plot of figure 6.3 displays their values. It can be seen that the eigenvalues decay relatively slowly. The value of the first 6 eigenvalues are above 1, with the biggest eigenvalue at about 2.2527. For the next 8 eigenvalues, i.e. from the 7th eigenvalue to the 14th eigenvalues, they appear to be clustering, with value in between 0.7857 and 0.4062. For the rest eigenvalues, i.e. from the 15th eigenvalue to the 40th eigenvalues, the decay rate is even slower than that of the first 14s, with the value for the 40th eigenvalue at around 0.0408.

The right plot of figure 6.3 shows $\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{40} \lambda_i}$, which is roughly the cumulative expected variance, discussed in section 5.3, chapter 5. Theoretically, when the analytical solution to the covariance function of the process is known, the cumulative expected variance should be expressed as

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \tag{6.2.7}$$

where the number of the eigenvalues in the denominator of equation 6.2.7 is infinity. However, in practice, only finite number of eigenvalues can be derived. In this example, the denominator of equation 6.2.7 is roughly approximated using $\sum_{i=1}^{40} \lambda_i$, since

for the order higher than 40, the numerical calculation sometimes results in complex eigenvalues, due to the limitation of the computer package we use (Matlab 7.0.4). From the plot, it can be seen that the first 6 eigenvalues explain around 56.26% out of the first 40 eigenvalues, while the first 14 eigenvalues explain around 86.55% out of the first 40 eigenvalues. Again this means that the decay of the eigenvalues is quite slow. This is due to the complexity and the variation in the real data.

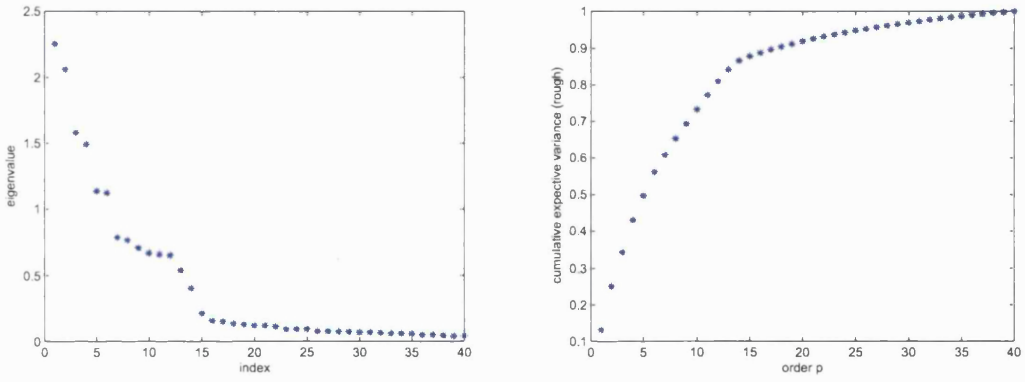

Figure 6.3: (Left): The first 40 eigenvalues of $TAR(1)$. (Right) The cumulative expected variance (rough) using $\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{40}\lambda_i}$.

Figure 6.4 shows the first three eigenfunctions. Although the original three series $X_1(t)$, $X_2(t)$ and $X_3(t)$ are very volatile due to the existence of the noise, the first three eigenfunctions are quite smooth. The first eigenfunction corresponds to the biggest eigenvalue. The direction of the first eigenfunction more or less follows the real series, except for the variation. Hence the first eigenfunction may be interpreted as a very rough "trend" factor. More accurate trend could be modelled using more eigenfunctions. The second and the third, possibly later eigenfunctions, start to

capture the variation in the series. The figure also shows other three facts. Firstly, since the original noisy series can be decomposed into a brunch of smooth series, it is possible to use the truncated Karhunen-Loeve expansion to smooth the data in practice. Secondly, if the original series is reconstructed using the Karhunen-Loeve expansion, higher orders are preferred, since the first few eigenfunctions are too smooth to capture all the variance. In other words, it means that there are a proportion of the "energy" containing in the lower tail of the Karhunen-Loeve expansion. Last but not least, the correlation among three series is kept in the eigenfunctions. It can be seen that the eigenfunctions for $X_1(t)$ and $X_2(t)$ are highly positively correlated, while the eigenfunction for $X_3(t)$ is highly negatively correlated with that for the rest two.

The covariance function can now be reconstructed using the truncated Karhunen-Loeve expansion. The order p is chosen to be $p = 40$. On one hand, the 40th eigenvalue is quite small, compared with the first few eigenvalues. Its value is actually 0.0408. On the other hand, as is mentioned, order higher than 40 sometimes results in complex eigenvalue due to the limitation of the computer package we use (Matlab 7.0.4). However, as we have examined previously, there is a lot of energy in the tail terms of the process. Truncation at order 40 might affect the overall performance of the reconstruction. This problem might be able to be resolved using higher order through more stable and more powerful software, such as C or C^{++}.

The other problem for reconstructing the covariance function using the truncated Karhunen-Loeve expansion is that the covariance matrix is no longer stationary after truncation. We now denote the truncated covariance function between $\mathbf{X}(t)$ and $\mathbf{X}(s)$ straightly after solving eigen-equation 6.2.6 as $\text{cov}_p[\mathbf{X}(t), \mathbf{X}(s)]$. To the $TAR(1)$

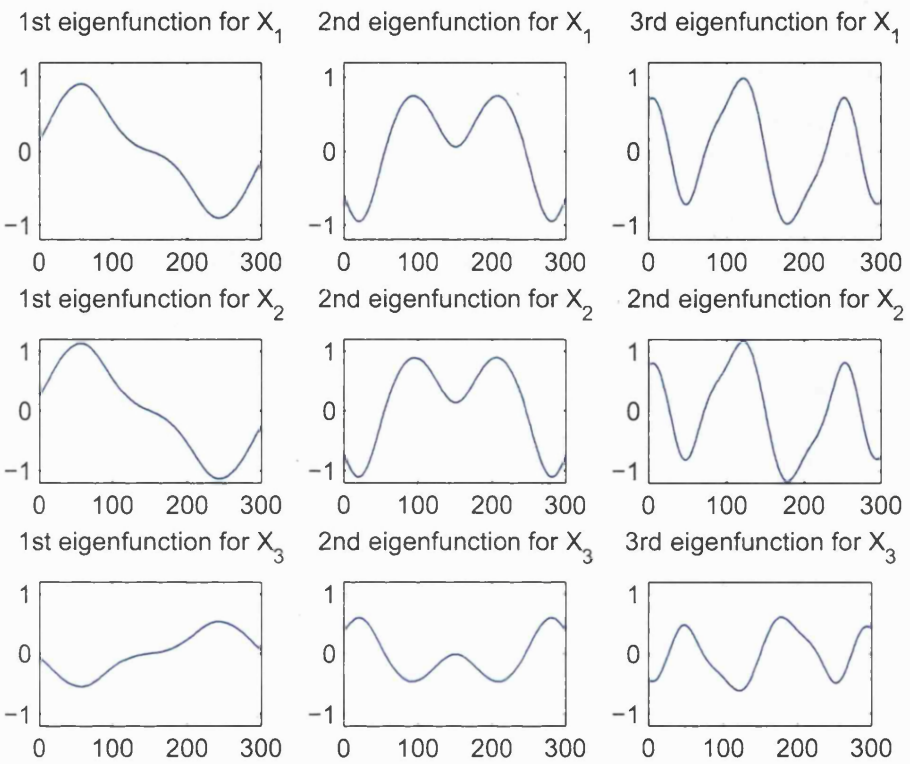

Figure 6.4: The first three eigenfunctions for $X_1(t)$, $X_2(t)$ and $X_3(t)$.

process plotted in figure 6.1, although both $\text{cov}[\mathbf{X}(t_1), \mathbf{X}(t_1)]$ and $\text{cov}[\mathbf{X}(t_2), \mathbf{X}(t_2)]$ represent the covariance function for lag 0 and are supposed to be equal to each other, their value after truncation at $p = 40$ is different.

$$\text{cov}_p[\mathbf{X}(t_1), \mathbf{X}(t_1)] = \begin{pmatrix} 5.3076 & 6.4255 & -3.2874 \\ 6.4255 & 8.0213 & -4.0437 \\ -3.2874 & -4.0437 & 2.1216 \end{pmatrix}$$

$$\mathrm{cov}_p[\mathbf{X}(t_2),\mathbf{X}(t_2)] = \begin{pmatrix} 5.8108 & 7.0219 & -3.5476 \\ 7.0219 & 8.7409 & -4.3615 \\ -3.5476 & -4.3615 & 2.2742 \end{pmatrix}$$

The way of resolving this stationary problem in the current chapter is to replace each $\mathrm{cov}_p[\mathbf{X}(t),\mathbf{X}(s)]$ by the average of the truncated covariance matrix with the same lag. We further denote the truncated sample autocovariance function with order p and lag h as $\hat{\Gamma}_p^A(h)$ after adjustment through averaging, the relationship between $\mathrm{cov}_p[\mathbf{X}(t),\mathbf{X}(s)]$ and $\hat{\Gamma}_p^A(h)$ is as follows. When $1 \le h \le n-1$

$$\hat{\Gamma}_p^A(0) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{cov}_p[\mathbf{X}(i),\mathbf{X}(i)] \tag{6.2.8}$$

$$\hat{\Gamma}_p^A(h) = \frac{1}{2(n-h)}\{\sum_{i=1}^{n-h} \mathrm{cov}_p[\mathbf{X}(i+h),\mathbf{X}(i)] + \sum_{i=1}^{n-h} \mathrm{cov}_p[\mathbf{X}(i),\mathbf{X}(i+h)]\} \tag{6.2.9}$$

Correspondingly, the covariance matrix estimator $\hat{\Gamma}$, after averaging, is denoted as $\hat{\Gamma}_p^A$.

$$\hat{\Gamma}_p^A = \begin{pmatrix} \hat{\Gamma}_{p,11}^A & \hat{\Gamma}_{p,12}^A & \cdots & \hat{\Gamma}_{p,1d}^A \\ \hat{\Gamma}_{p,21}^A & \hat{\Gamma}_{p,22}^A & \cdots & \hat{\Gamma}_{p,2d}^A \\ \cdots & & & \\ \hat{\Gamma}_{p,d1}^A & \hat{\Gamma}_{p,d2}^A & \cdots & \hat{\Gamma}_{p,dd}^A \end{pmatrix}, \tag{6.2.10}$$

where $\hat{\Gamma}_{p,ij}^A$ is

$$\hat{\Gamma}_{p,ij}^A = \begin{pmatrix} \hat{\Gamma}_{p,ij}^A(0) & \hat{\Gamma}_{p,ij}^A(1) & \cdots & \hat{\Gamma}_{p,ij}^A(n-1) \\ \hat{\Gamma}_{p,ij}^A(1) & \hat{\Gamma}_{p,ij}^A(0) & \cdots & \hat{\Gamma}_{p,ij}^A(n-2) \\ \cdots & & & \\ \hat{\Gamma}_{p,ij}^A(n-1) & \hat{\Gamma}_{p,ij}^A(n-2) & \cdots & \hat{\Gamma}_{p,ij}^A(0) \end{pmatrix} \quad 1 \le i,j \le d \tag{6.2.11}$$

and $\hat{\Gamma}_{p,ij}^A(h) = \{\hat{\Gamma}_p^A(h)\}_{ij}$.

True covariance function				Truncated covariance function		
	$X_1(1)$	$X_2(1)$	$X_3(1)$	$X_1(1)$	$X_2(1)$	$X_3(1)$
$X_1(1)$	7.5631	7.5493	-3.7207	4.2102/5.8543	5.2752/7.0777	-2.6317/-3.5970
$X_2(1)$	7.5493	10.1721	-4.2506	5.2752/7.0777	6.6754/8.8820	-3.2866/-4.4367
$X_3(1)$	-3.7207	-4.2506	3.6355	-2.6317/-3.5970	-3.2866/-4.4367	1.6575/2.3775

Table 6.2: The covariance and the cross covariance function for $X_1(1)$, $X_2(1)$ and $X_3(1)$ using the untruncated (true) version and the truncated approximation. From the second to the fourth column: the true version; From the fifth to the seventh column: the truncated version. Value above "/", $p = 10$; Value below "/", $p = 40$.

Figure 6.5 and figure 6.6 show the covariance and the cross covariance function reconstruction for $X_1(t)$, $X_2(t)$ and $X_3(t)$ for the above $TAR(1)$ process, using the truncated Karhunen Loeve expansion after truncation and averaging. The truncation order is 10 and 40 respectively. Since the first 10 eigenvalues only explain about 73% of the first 40 eigenvalues as mentioned above, the covariance reconstruction is not very good. Compared with the true covariance function, some values in the covariance function using $p = 10$ are much smaller. When the truncation order p increases from 10 to 40, the performance of the approximation using the truncated Karhunen-Loeve expansion improves a lot. The shape of the true covariance function and that of the truncated version are quite similar to each other. However, some values using the truncated approximation are still smaller than that from the true function, even when using $p = 40$.

Table 6.2 further displays the value comparison of the covariance and the cross covariance function between the untruncated version (true version) and the truncated version for the lag $h = 0$. For both the covariance and the cross covariance function, when p increases from 10 to 40, the difference of the value between the true version and the truncated version decreases. It implies that higher order p can be introduced for better approximation. However, as is mentioned, current computer package used

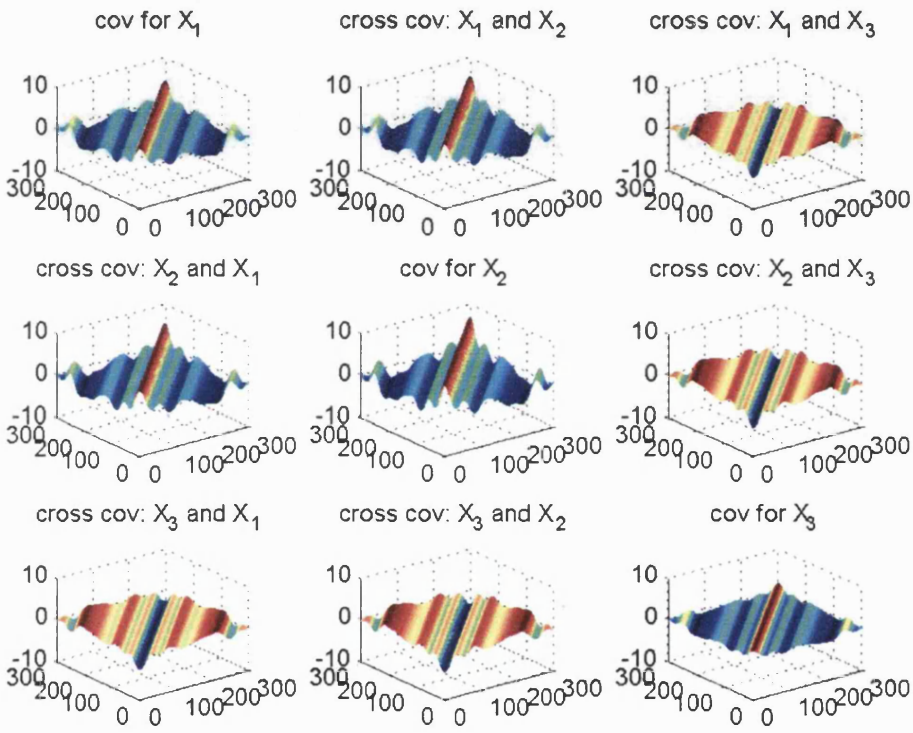

Figure 6.5: Reconstruction of the covariance and the cross covariance function for $X_1(t)$, $X_2(t)$ and $X_3(t)$, using the truncated Karhunen Loeve expansion when $p = 10$.

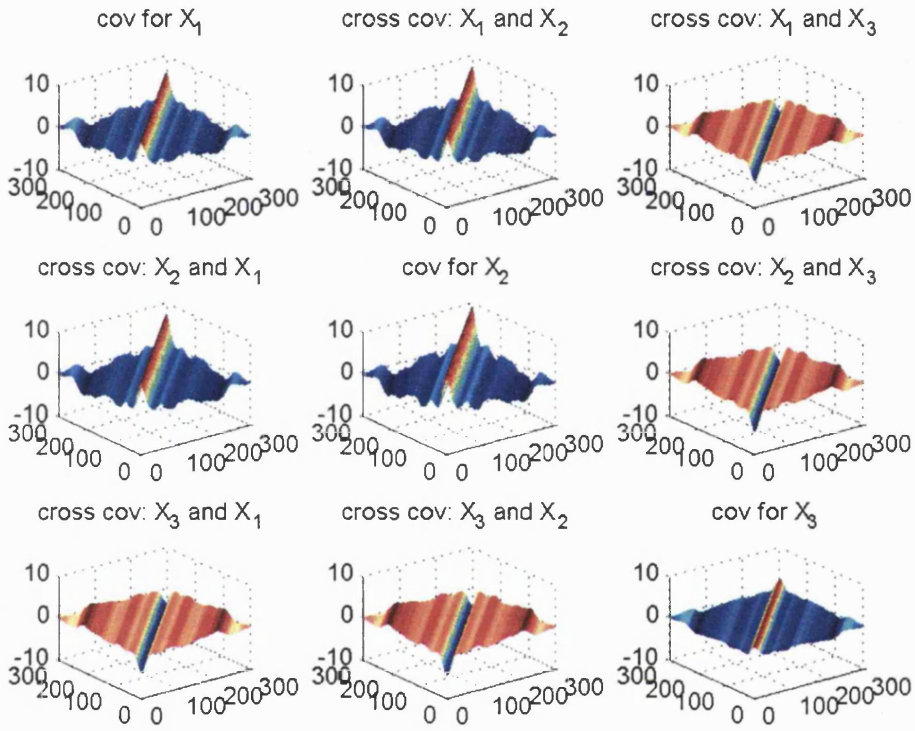

Figure 6.6: Reconstruction of the covariance and the cross covariance function for $X_1(t)$, $X_2(t)$ and $X_3(t)$, using the truncated Karhunen Loeve expansion when $p = 40$.

here (Matlab 7.0.4) is not powerful enough to handle large p.

In general, we can summarise the procedure of dealing with the multivariate data when using the multivariate Karhunen-Loeve expansion. The multivariate data can be the data either from the simulation or from the real database.

Step 1: Stationarise the data using certain transformation techniques, such as detrending and differencing, so that the data after transformation is weakly stationary.

Step 2: Obtain an estimator $\hat{\Gamma}$ for the covariance matrix K, as shown in equation 6.2.3.

Step 3: Perform the multivariate Karhunen-Loeve expansion using certain numerical scheme and decide the order p for the truncation. In this section, the numerical scheme we use is the trapezium integral method. The key equation to solve is the eigen-equation 6.2.6.

Step 4: Obtain $\hat{\Gamma}_p^A$, which is the truncated covariance matrix after averaging. The average procedure is defined in equation 6.2.8 and equation 6.2.9. $\hat{\Gamma}_p^A$ is also the covariance matrix used for the reconstruction.

6.3 Smoothing and prediction

For the noisy process, it can be decomposed into an infinite sum of relatively smooth functions using the Karhunen-Loeve expansion. If the expansion is truncated and only part of these smooth functions are used, we should be able to capture some characteristic, such as the patterns of the original series, while leaving out the substantial random noise. This reconstruction of the original series using the finite smooth functions is called "smoothing" in this section. In the multivariate setting, this smoothing process uses not only the knowledge of one series, it utilises the knowledge from the

other correlated series as well.

The method for smoothing in this chapter is an extension of the method introduced in chapter 4 from a univariate setting to a multivariate setting. The smoother is the conditional expectation. Assume that there is a d-dimensional correlated series $\mathbf{X}(t) = \Big(X_1(t),\ X_2(t),\ \cdots,\ X_d(t)\Big)^T$, $t = 1, 2, \cdots n$, we try to calculate $E(\mathbf{X}(t)|X_s)$, where

$$X_s = \begin{pmatrix} \mathbf{X}(1) \\ \mathbf{X}(2) \\ \cdots \\ \mathbf{X}(n) \end{pmatrix}$$

Using the partition inverse equation and following the same procedure as that described in chapter 4, we can derive that

$$E(\mathbf{X}(t)|X_s) \ = \ k^T K_n^{-1} X_s \tag{6.3.1}$$

where

$$k_{nd\times d} \ = \ \Big(\mathrm{cov}[\mathbf{X}(t), \mathbf{X}(1)],\ \mathrm{cov}[\mathbf{X}(t), \mathbf{X}(2)],\ \cdots,\ \mathrm{cov}[\mathbf{X}(t), \mathbf{X}(n)]\Big)^T$$

$$K_{n,nd\times nd} \ = \ \begin{pmatrix} \mathrm{cov}[\mathbf{X}(1), \mathbf{X}(1)] & \mathrm{cov}[\mathbf{X}(1), \mathbf{X}(2)] & \cdots & \mathrm{cov}[\mathbf{X}(1), \mathbf{X}(n)] \\ \mathrm{cov}[\mathbf{X}(2), \mathbf{X}(1)] & \mathrm{cov}[\mathbf{X}(2), \mathbf{X}(2)] & \cdots & \mathrm{cov}[\mathbf{X}(2), \mathbf{X}(n)] \\ \cdots & & & \\ \mathrm{cov}[\mathbf{X}(n), \mathbf{X}(1)] & \mathrm{cov}[\mathbf{X}(n), \mathbf{X}(2)] & \cdots & \mathrm{cov}[\mathbf{X}(n), \mathbf{X}(n)] \end{pmatrix}$$

When smoothing the data, we will use the truncated version of $E(\mathbf{X}(t)|X_s)$, $t = 1, 2, \cdots n$ to represent the original data point at time t. When using the original covariance function without truncation, $E(\mathbf{X}(t)|X_s) = \mathbf{X}(t)$, $t = 1, 2, \cdots n$. Hence, under this case, we are only able to obtain the original value at time point t rather than the smoothing value. Using the truncated Karhunen-Loeve expansion, on the

other hand, approximates the covariance function in equation 6.3.1, so that an estimator of $E(\mathbf{X}(t)|X_s)$ can be calculated, which can be regarded as the smoothing value. Nevertheless, using the truncated Karhunen-Loeve expansion only, one of the problems we meet in chapter 4 arises again. When n is big, such as 300 in this chapter, and $n > p$, the truncated version of the matrix K_n is singular, hence not invertible. One way to solve this problem is to add an extra noise term σ^2 to the diagonal element. This can be regarded as a regularization. It is equivalent to treating the original process as coming from the Gaussian regression model, which is the truncated Karhunen-Loeve expansion with added independent noise. However, since K_n after the truncation has been adjusted to reflet stationarity and the analytical expression of the covariance is unknown, the maximum likelihood approach in equation 4.3.5 can not be simply applied, to find σ^2. The adjusted method used here to find σ^2 is to minimise the mean squared error instead. For the process, $X_i(t), 1 \leq i \leq d$, denote the value at time t after smoothing as $\hat{X}_i(t)$, then we need to minimise the following.

$$MSE = \frac{1}{nd} \sum_{i=1}^{d} \sum_{t=1}^{n} [X_i(t) - \hat{X}_i(t)]^2 \tag{6.3.2}$$

We continue with the example of $TAR(1)$ in the first section. In order to smooth the data, a suitable σ^2 needs to be sought after first to minimise the mean squared error. Figure 6.7 lists σ^2 for four different orders, $p = 1$, $p = 10$, $p = 20$ and $p = 40$. Due to the limitation of the computing power, σ^2 is chosen to be the integer in between 1 and 40. When p is chosen to be low, such as 1 and 10. There exists some variation in the mean squared error. Moreover, under small p, since not much cumulative expected variance can be explained by the first p eigenvalues, σ^2 needs to be relatively big to recover the "energy" in the tail. For $p = 1$ and $p = 10$, $\sigma^2 = 29$

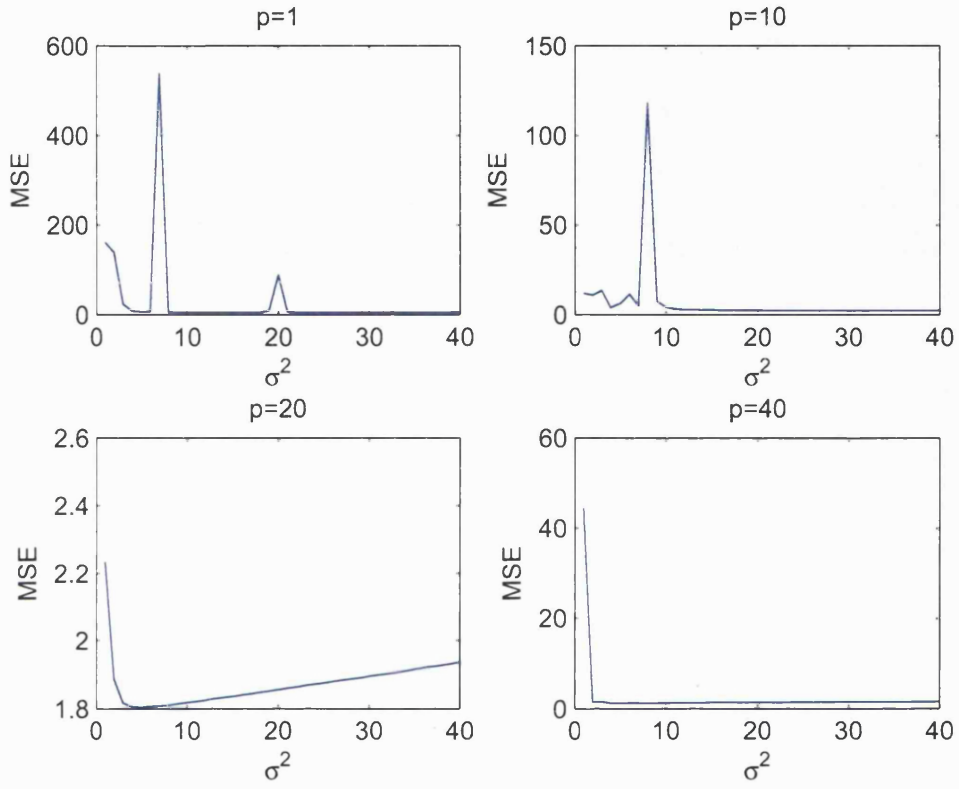

Figure 6.7: The mean squared error for choosing σ^2 using equation 6.3.2 for TAR(1) using the truncated Karhunen-Loeve expansion when p is 1, 10, 20 and 40 respectively.

and $\sigma^2 = 38$ respectively. When the order p increases, the plot for the mean squared error gets smoother. Not only the mean squared error exhibits less variation, σ^2 dramatically decreases as well. For $p = 20$ and $p = 40$, σ^2 has been reduced to 5 in both cases.

Figure 6.8 plots $X_1(t)$, $X_2(t)$ and $X_3(t)$ after smoothing using the truncated orders $p = 1$, $p = 10$, $p = 20$ and $p = 40$ respectively. When only the first eigenfunction corresponding to the biggest eigenvalue is used in the smoothing, the shape of the

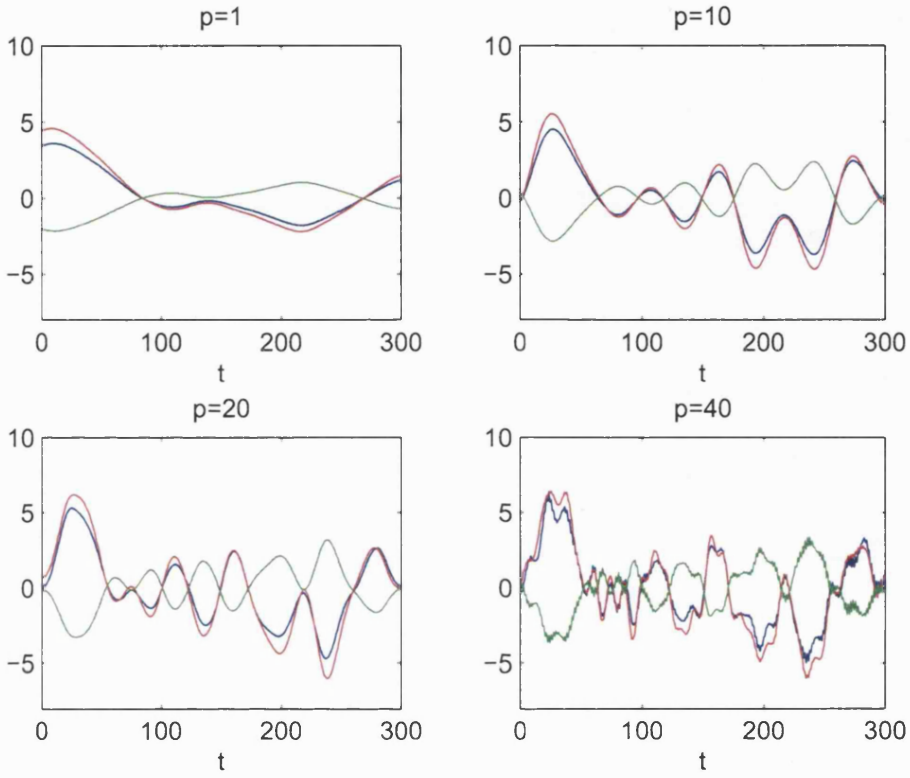

Figure 6.8: Smoothing for TAR(1) using the truncated Karhunen-Loeve expansion when p is 1, 10, 20 and 40 respectively. Blue line: smoothing for $X_1(t)$; Red line: smoothing for $X_2(t)$; Green line: smoothing for $X_3(t)$.

series after smoothing is more or less the same as that of the first eigenfunction itself. It is too smooth to capture any variation in the original series. When p is increased to 10 and 20, more variation of the original series has been captured by the local minimum or local maximum appearing in the smoothing series. It is still too smooth though. Using the biggest p in this example, i.e. $p = 40$, some zigzag patterns exhibit in the series after smoothing. It is a reflection of the noise lying in the original series. In short the smoothing performance using the truncated Karhunen-Loeve expansion

can be controlled by the truncation order p. The bigger the p, the more variation there is after smoothing.

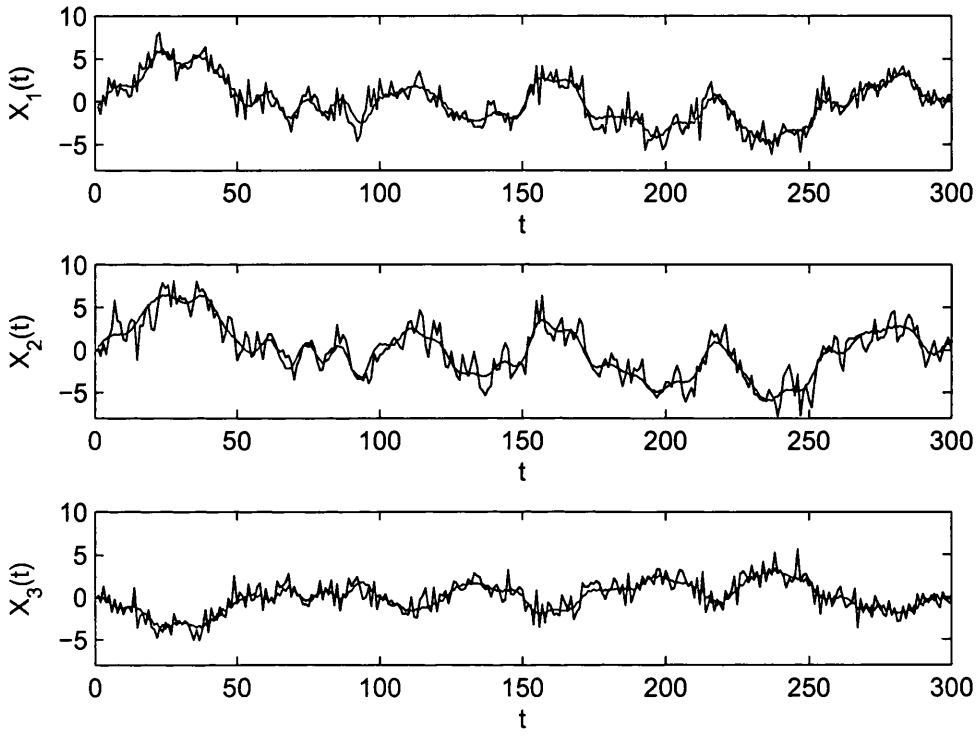

Figure 6.9: The original series of $TAR(1)$ and its smoothing version using the truncated Karhunen-Loeve expansion when $p = 40$. Blue line: original series; Red line: smoothing series.

Figure 6.9 further plots the original series and the smoothing series under $p = 40$ altogether. It can be seen that the smoothing series performs quite well. The smoothing series more or less follows the patterns of the original series without sudden change, which is caused by the big noise in the original series. If more noise is intended to be captured after smoothing, even higher order of p can be used.

Another smoothing technique, which has been proved to be successful and similar to the smoothing using the Karhunen-Loeve expansion, is called singular-spectrum analysis (SSA). SSA has been widely used in the field of climate, meteorology and geophysics, see for example, Vautard et al. (1992), Allen and Smith (1996) and Yiou et al. (2000). The core step of SSA is also to find eigenvalues and eigenfunctions, but through singular value decomposition. Detail of SSA can refer to Golyandina et al. (2001).

For time series $\mathbf{X}(t)$, $t = 1, 2, \cdots, n$, smoothing discussed above is a process to reduce random fluctuations and identify patterns to its existing data $t = 1, 2, \cdots, n$ using certain smoother. In this chapter, the smoother is the conditional expectation, i.e. $E(\mathbf{X}(t)|X_s)$, $t = 1, 2, \cdots, n$. The conditional expectation can also be used to forecast the future value for the series at $t \geq n+1$ using the same calculation procedure. This time the conditional expectation is called a predictor rather than a smoother.

The performance of the prediction using the truncated Karhunen-Loeve expansion is analysed through one-step ahead prediction using the above $TAR(1)$ process. The one-step ahead prediction is calculated at each point from $t = 251$ to $t = 300$, while the information which is conditional on is from the original series and kept in the window $[t - 249, t - 1]$ with window length 249. Each time, when a prediction point moves from t to $t + 1$, the information window moves correspondingly from $[t - 249, t - 1]$ to $[t - 248, t]$. Such setting allows us to compare the prediction using the truncated Karhunen-Loeve expansion with the original value from the series. σ^2 used to adjust the covariance matrix is calculated using the data points from $t = 1$ to $t = 250$ through minimising the mean squared error using equation 6.3.2.

Figure 6.10 shows the original series, the prediction using the truncated Karhunen-Loeve expansion when $p = 40$ and the prediction using the untruncated Karhunen-Loeve expansion (full covariance function without the truncation and without σ^2 adjustment) from $t = 251$ to $t = 300$. It can be seen that using the truncated Karhunen-Loeve expansion when $p = 40$, the prediction does not strictly follow the original series, although it is performing quite well in capturing the trend of the series. The prediction under the current truncation version $p = 40$, in some sense, is still like smoothing. It mainly goes through the main pattern without considering too much on the sudden movement due to the big noise. However, when using the untruncated Karhunen-Loeve expansion, which is to use the true covariance function without any truncation, the one-step ahead prediction performs quite well. Its difference from the value in the original series is relatively small. This implies that better prediction can be conducted when p increases so that variability in the tail can be retrieved and be contributed to the calculation when dealing with the real data.

Table 6.3 further lists the value using the truncated Karhunen-Loeve expansion when $p = 40$, the prediction using the true covariance function and the real value of $TAR(1)$ for the first three points and the last three points. The error in the table is defined as $\left|\frac{\text{prediction–real value}}{\text{real value}}\right| \times 100\%$. It confirms again that the prediction performance using the truncation version is not very satisfactory. Although the error for the prediction using the untruncated true covariance function is still about 5% or even more, it generally performs much better than the truncation version. Hence, we can increase the order p for the better prediction if computing power allows.

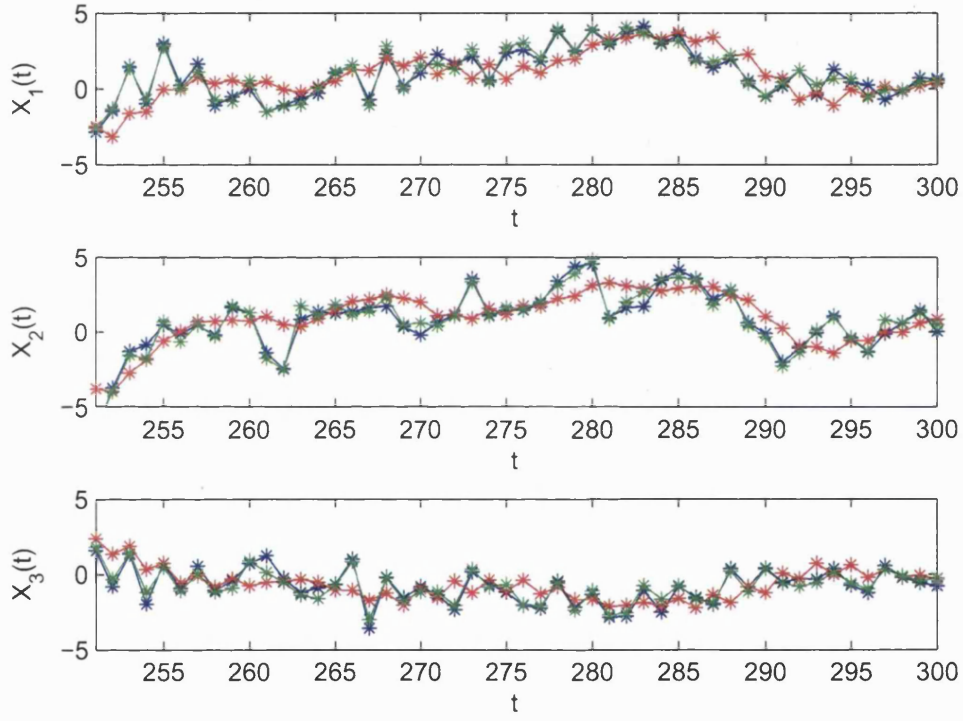

Figure 6.10: The original series of $TAR(1)$ (blue), the prediction using the truncated Karhunen-Loeve expansion with $p = 40$ (red) and the prediction using the true covariance function (green) from $t = 251$ to $t = 300$.

t	251	252	253	$\cdots$	298	299	300
X_1 (real)	-2.8461	-6.7093	1.579	$\cdots$	-0.1876	0.617	-0.2296
X_1 (untruncated)	-2.5866	-6.5303	1.7899	$\cdots$	-0.1141	0.581	-0.1305
Error	9.12%	2.67%	13.36%	$\cdots$	39.18%	5.83%	43.16%
X_1 (truncated)	-2.4464	-3.7849	2.3913	$\cdots$	-0.1746	0.0074	-0.1615
Error	14.04%	43.59%	51.44%	$\cdots$	6.93%	98.80%	29.66%
X_2 (real)	-1.4543	-3.7284	-0.8105	$\cdots$	0.7089	1.4439	-0.5675
X_2 (untruncated)	-1.3289	-3.934	-0.3366	$\cdots$	0.4209	1.3277	-0.4675
Error	8.62%	5.51%	58.47%	$\cdots$	40.63%	8.05%	17.62%
X_2 (truncated)	-3.1262	-3.9827	1.3701	$\cdots$	0.1948	0.6285	-0.0691
Error	114.96%	6.82%	269.04%	$\cdots$	72.52%	56.47%	87.82%
X_3 (real)	1.4771	-1.2949	1.3572	$\cdots$	0.6538	0.0623	-0.7967
X_3 (untruncated)	1.3776	-1.5471	1.4226	$\cdots$	0.514	0.4961	-0.2797
Error	6.74%	19.48%	4.82%	$\cdots$	21.38%	696.31%	64.89%
X_3 (truncated)	-1.599	-2.7318	1.9004	$\cdots$	0.4005	0.881	-0.3161
Error	208.25%	110.97%	40.02%	$\cdots$	38.74%	1314.13%	60.32%

Table 6.3: The original series of $TAR(1)$, the prediction using the truncated Karhunen-Loeve expansion with $p = 40$ and the prediction using the true covariance function (untruncated) for $t = 251$, 252, 253, 298, 299 and 300.

Chapter 7

Appendix

7.1 Mercer's theorem

As is introduced in chapter 1, one of the key formulae of this thesis is the Fredholm integral equation. Using the compact integral operator defined in remark 1.2.3, chapter 1, the Fredholm integral equation is expressed as

$$\mathcal{K}\phi(s) = \int_{\mathcal{T}} K(s,t)\phi(s)ds = \lambda\phi(t) \tag{7.1.1}$$

where $\mathcal{T}$ is often a compact interval and $K(s,t)$, the kernel of a certain stochastic process $X(t)$, is an L_2 kernel satisfying the condition,

$$\int_{\mathcal{T}}\int_{\mathcal{T}} K(s,t)^2 dsdt < \infty \tag{7.1.2}$$

If $X(t)$ is a zero mean stochastic process, $K(s,t) = E(X(s)X(t))$, where E represents the expectation. According to theorem 1.1.6, chapter 1, when the L_2 kernel $K(s,s) < \infty$, for all $s \in \mathcal{T}$, $K(.,.)$ also belongs to RKHS.

In general, there are infinite number of eigenvalues λ_i, $i \in N$. We assume that all the eigenvalues have been ordered so that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots$. Correspondingly,

there are infinite number of eigenfunctions as well, $\phi_i(.)$, $i \in N$. The eigenfunctions are orthogonal in the sense that $\int_{\mathcal{T}} \phi_i(s)\phi_j(s)ds = \delta_{ij}$.

Mercer's theorem (see, for example, Porter and Stirling (1990)) provides a way of expressing the kernel $K(s,t)$ in terms of the eigenvalues and the eigenfunctions.

Theorem 7.1.1. *Assume that $K(s,t)$ is a continuous, symmetric, positive definite L_2 kernel of the stochastic process $X(t)$. $K(s,t)$ is defined on $\mathcal{T} \times \mathcal{T}$, where $\mathcal{T}$ is a compact interval. Then there exists a L_2 orthogonal basis consisting of $\{\phi_i\}$, which are the eigenfunctions of $\mathcal{K}$, together with the corresponding eigenvalues $\{\lambda_i > 0\}$. The kernel $K(s,t)$ has the representation*

$$K(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi(t) \tag{7.1.3}$$

where the series converges absolutely and uniformly in the sense that

$$\sum_{i=1}^{\infty} |\lambda_i \phi_i(s)\phi(t)| \tag{7.1.4}$$

converges uniformly with respect to both variables simultaneously.

Theorem 7.1.1 states the expansion of the kernel $K(s,t)$ in the univariate setting. It can be generalised to the multivariate setting, which is useful in chapter 5. Multivariate, as is described in chapter 5, refers to the multivariate state, and the univariate time.

Assume that we have a zero mean d-dimensional stochastic process $\mathbf{X}(t)^T = \Big(X_1(t),\ X_2(t),\ \cdots,\ X_d(t)\Big)$ defined on a compact interval $\mathcal{T}$. The process $\mathbf{X}(t)$ is with finite energy, i.e. $\sum_{i=1}^{d} E(X_i^2(t)) < \infty$. The kernel $\mathbf{K}(t,s)$ of the process $\mathbf{X}(t)$ defined as $\mathbf{K}(t,s) = E(\mathbf{X}(t)\mathbf{X}(t)^T)$, is usually a multivariate L_2 kernel satisfying

$$\sum_{i=1}^{d}\sum_{j=1}^{d} \int_{\mathcal{T}}\int_{\mathcal{T}} [E(X_i(s)X_j(t))]^2 dsdt < \infty \tag{7.1.5}$$

Then the Fredholm integral equation on the kernel $\mathbf{K}(t,s)$ and its relevant integral operator $\mathcal{K}_d$ can be expressed as

$$\mathcal{K}_d f(s) = \int_{\mathcal{T}} \mathbf{K}(s,t) f(s) ds = \lambda f(t) \tag{7.1.6}$$

where λ, the eigenvalue, is a scalar, while, $f(t)$, the eigenfunction, is a d-dimensional vector. The same as that in the univariate setting, there are infinite number of eigenvalues λ_i, $i \in N$, with the infinite number of eigenfunctions $f_i(t)$, $i \in N$, in the multivariate L_2 space. The orthogonality of the eigenfunctions in the multivariate L_2 space refers to

$$\int_{\mathcal{T}} f_i(t)^T f_j(t) = \delta_{ij} \tag{7.1.7}$$

Generalised Mercer's theorem (see, for example, Mahram et al. (2002) is an extension of the Mercer's theorem, which provides a way of expanding the kernel $\mathbf{K}(t,s)$.

Theorem 7.1.2. *Assume that* $\mathbf{K}(t,s)$ *is a continuous, positive definite, multivariate* L_2 *kernel of the d-dimensional stochastic process* $\mathbf{X}(t)$ *with finite energy.* $\mathbf{K}(t,s)$ *is defined on* $\mathcal{T} \times \mathcal{T}$*, where* $\mathcal{T}$ *is a compact interval. Then there exists a multivariate* L_2 *orthogonal basis consisting of* $\{f_i\}$*, which are the eigenfunctions of* $\mathcal{K}_d$*, together with the corresponding eigenvalues* $\{\lambda_i > 0\}$*. The kernel* $\mathbf{K}(s,t)$ *has the representation*

$$\mathbf{K}(s,t) = \sum_{i=1}^{\infty} \lambda_i f_i(s) f_i(t)^T \tag{7.1.8}$$

and the converge is uniform in both s *and* t*.*

7.2 Change in the generalised mean squared error

This section provides the detailed calculation for the difference of the generalised MSE between order p and the order $p+1$. To make notation clear, the truncated

covariance matrix/vector of order p is denoted as K_p/k_p, while the true covariance matrix/vector is denoted as K_f/k_f.

Define

$$\Phi_{p+1} = \begin{pmatrix} \Phi_p \\ \phi_{p+1}^T \end{pmatrix} \qquad \Lambda_{p+1} = \begin{pmatrix} \Lambda_p & 0 \\ 0 & \lambda_{p+1} \end{pmatrix} \qquad \Sigma_p = K_f - K_p \tag{7.2.1}$$

where Φ_p is a $p \times n$ matrix comprising of the first p eigenfunctions, and ϕ_{p+1} is a $n \times 1$ column vector comprising of the $p+1$ eigenfunction. Then

$$\begin{aligned} K_{p+1} &= \Phi_{p+1}^T \Lambda_{p+1} \Phi_{p+1} \\ &= \Phi_p^T \Lambda_p \Phi_p + \lambda_{p+1} \phi_{p+1} \phi_{p+1}^T \\ &= K_p + \lambda_{p+1} \phi_{p+1} \phi_{p+1}^T \end{aligned}$$

Using the inverse matrix formula, we obtain

$$K_{p+1}^{-1} = K_p^{-1} - \frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}} K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1}$$

$$K_{p+1}^{-1} K_f K_{p+1}^{-1} = K_p^{-1} K_f K_p^{-1} - \frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}} K_p^{-1} K_f K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1}$$

$$- \frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}} K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1} K_f K_p^{-1}$$

$$+ \left(\frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}}\right)^2 K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1} K_f K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1}$$

$$2K_{p+1}^{-1} - K_{p+1}^{-1} K_f K_{p+1}^{-1}$$

$$= 2K_p^{-1} - 2\frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}} K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1}$$

$$- K_p^{-1} K_f K_p^{-1} + \frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}} K_p^{-1} K_f K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1}$$

$$+ \frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}} K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1} K_f K_p^{-1}$$

$$- \left(\frac{\lambda_{p+1}}{1 + \lambda_{p+1} \phi_{p+1}^T K_p^{-1} \phi_{p+1}}\right)^2 K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1} K_f K_p^{-1} \phi_{p+1} \phi_{p+1}^T K_p^{-1}$$

Therefore, denoting the generalised MSE of order p $\int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t))dt$, the difference in the generalised MSE between the order $p+1$ and the order p is

$$
\begin{aligned}
& \int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t))dt - \int_{\mathcal{T}} \mathrm{MSE}_{p+1}(\hat{y}(t))dt \\
= \quad & \mathrm{trace}[(2K_{p+1}^{-1} - K_{p+1}^{-1}K_f K_{p+1}^{-1})(\Phi_{p+1}^T \Lambda_{p+1}^2 \Phi_{p+1})] - \mathrm{trace}[(2K_p^{-1} - K_p^{-1}K_f K_p^{-1})(\Phi_p^T \Lambda_p^2 \Phi_p)] \\
= \quad & \mathrm{trace}[(2K_{p+1}^{-1} - K_{p+1}^{-1}K_f K_{p+1}^{-1})(\Phi_p^T \Lambda_p^2 \Phi_p + \lambda_{p+1}^2 \phi_{p+1}\phi_{p+1}^T)] \\
& -\mathrm{trace}[(2K_p^{-1} - K_p^{-1}K_f K_p^{-1})(\Phi_p^T \Lambda_p^2 \Phi_p)] \\
= \quad & 2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1}) - 2\frac{\lambda_{p+1}}{1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1}}\phi_{p+1}^T K_p^{-1}\Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1}\phi_{p+1} \\
& -2\frac{\lambda_{p+1}^3}{1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1}}(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 - \lambda_{p+1}^2 \phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1} \\
& +\frac{\lambda_{p+1}}{1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1}}\phi_{p+1}^T K_p^{-1}\Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1}K_f K_p^{-1}\phi_{p+1} \\
& +\frac{\lambda_{p+1}^3}{1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1}}(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& +\frac{\lambda_{p+1}}{1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1}}\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1}\phi_{p+1} \\
& +\frac{\lambda_{p+1}^3}{1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1}}(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& -\frac{\lambda_{p+1}^2}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2}(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T \Lambda_p^2 \Phi_p K_p^{-1}\phi_{p+1}) \\
& -\frac{\lambda_{p+1}^4}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2}(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1})
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{1}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} \\
& [2\lambda_{p+1}^2\phi_{p+1}^T K_p^{-1}\phi_{p+1} + 4\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 + 2\lambda_{p+1}^4(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^3 \\
& -2\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) - 2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 - 2\lambda_{p+1}^4(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^3 \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) - 2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^4(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) + \lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& +2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& +2\lambda_{p+1}^4(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) + \lambda_{p+1}(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^4(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1})
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{1}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} \\
& [2\lambda_{p+1}^2\phi_{p+1}^T K_p^{-1}\phi_{p+1} + 2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 \\
& -2\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) - 2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1}) + \lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}K_f K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}K_f K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1})] \\
= \quad & \frac{1}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} \\
& [\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& +\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1})] + [\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1}) \\
& +2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 - \lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1})]
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{1}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2} \\
& [2\lambda_{p+1}(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& +2\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1}) \\
& -\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1})] + [\lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1}) \\
& +2\lambda_{p+1}^3(\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2 - \lambda_{p+1}^2(\phi_{p+1}^T K_p^{-1}\phi_{p+1})(\phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1})]
\end{aligned}
$$

For simplicity, we define,

$$
\begin{aligned}
A &= \phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1} \\
C &= \phi_{p+1}^T K_p^{-1}\Sigma_p K_p^{-1}\phi_{p+1} \\
b &= \phi_{p+1}^T K_p^{-1}\phi_{p+1} \\
d &= \phi_{p+1}^T K_p^{-1}\Phi_p^T\Lambda_p^2\Phi_p K_p^{-1}\phi_{p+1} \\
S &= 2Ab - Cd - C + b - bd \\
\Delta &= S^2 - 16Ab^2
\end{aligned}
$$

Then the difference of the generalised MSE between the order p and the order $p+1$ is,

$$
\begin{aligned}
& \int_{\mathcal{T}} \mathrm{MSE}_p(\hat{y}(t))dt - \int_{\mathcal{T}} \mathrm{MSE}_{p+1}(\hat{y}(t))dt \\
= \quad & \frac{\lambda_{p+1}}{(1+\lambda_{p+1}\phi_{p+1}^T K_p^{-1}\phi_{p+1})^2}[\lambda_{p+1}^2(2b^2) + \lambda_{p+1}S + 2A]
\end{aligned}
$$

Bibliography

M.R. Allen and L.A. Smith. Monte carlo ssa: Detecting irregular oscillation in the presence of coloured noise. *Journal of Climate*, 9:3373–3404, 1996.

L. Arnold. *Stochastic Differential Equations: Theory and Applications.* Wiley-Interscience, New York, 1974.

N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

G. Boente and R. Fraiman. Kernel-based functional principal components. *Statistics and Probability Letters*, 48:335–345, 2000.

P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting, Second Edition.* Springer-Verlag, New York, 2002.

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods, Second Edition.* Springer-Verlag, Berlin, 1991.

P.E. Caines and H.P. Wynn. An algebraic framework for bayes nets of time series. *Modeling, Estimation and Control. Festschrift in Honor of Giorgio Picci on the Occasion of his Sixty-Fifth Birthday*, Lecture Notes in Control and Information Sciences 364:45–57, 2007.

X. Chen and W.V. Li. Quadratic functionals and small ball probabilities for the m-fold integrated brownian motion. *The Annals of Probability*, 31, No.2:1052–1077, 2003.

E.W. Cheney and D.R. Kincaid. *Numerical Mathematics and Computing.* Thomson Brooks, 2007.

R.C.H. Cheng and O.D. Jones. Analysis of distributions in factorial experiments. *Statistica Sinica*, 14, No. 4:1085–1103, 2004.

R. Cont and J.da Fonseca. Dynamics of implied volatility surfaces. *Working paper*, 2001.

M.V. Cover and J.V. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

N. Cressie. *Statistics for Spatial Data.* Wiley, New York, 1993.

V.V. Fedorov. *Theory of Optimal Experiments.* Academic Press, New York, 1972.

Z. Galil and J. Kiefer. Time- and space-saving computer methods, related to mitchell's detmax, for finding d-optimum designs. *Technometrics*, 22, No.3:301–313, 1980.

F. Gao, J. Hannig, and F. Torcaso. Integrated brownian motions and exact l_2-small balls. *The Annals of Probability*, 31, No.3:1320–1337, 2003.

R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: a Spectral Approach.* Springer, Berlin, 1991.

S. Glavaski, J.E. Marsden, and R.M. Murray. Model reduction, centering, and the karhunen-loeve expansion. *Decision and Control, 1998. Proceedings of the 37th IEEE Conference on*, 2:2071–2076, 1998.

N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques.* Chapman and Hall/CRC, 2001.

P. Hall and M. Hosseini-Nassab. On properties of functional principal components analysis. *Journal of the Royal Statistical Society: series B*, 68, Part. 1:109–126, 2006.

A. Hoffman, J. Lee, and J. Williams. New upper bounds for maximum entropy sampling. *mODA6- Advances in Model-Oriented Design and Analysis*, 6:143–153, 2004.

S.P. Huang, S.T. Quek, and K.K. Phoon. Convergence study of the truncated karhunen-loeve expansion for simulation of stochastic processes. *International Journal for Numerical Methods in Engineering*, 52:1029–1043, 2001.

G.M. James, T.J. Hastie, and C.A. Sugar. Principal component models for sparse functional data. *Biometrika*, 87:587-602, 2000.

T. Kailath. An rkhs approach to detection and estimation problems. part i. deterministic signals in gaussian noise. *IEEE Transactions on Information Theory*, 17: 530–549, 1971.

K. Karhunen. Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientarum Fennicae*, 37:3–79, 1947.

J. Kiefer. General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2:849–879, 1974.

J. Kiefer and J. Wolfowitz. The equivalence of two extremal problems. *Canadian Journal of Mathematics*, 14:363–366, 1960.

M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12, Issue 1:103–108, 1990.

C.W. Ko, J. Lee, and M. Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43, No.4:684–691, 1995.

A.J. Laub. *Matrix Analysis for Scientists and Engineers.* SIAM, Society for Industrial and Applied Mathematics, 2004.

J. Lee. Constrained maximum entropy sampling. *Operations Research*, 43:684–691, 1998.

M. Loeve. Fonctions aleatoires du second ordre. *Processus Stochastiques et Mouvement Brownien, P. Levy (ed.),*, 1948.

D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, Cambridge, UK, 2003.

H.H. Mahram, D. Dahlhaus, and D. Blomker. Karhunen-loeve expansion of vector random processes. *Technical Report, CTL, Swiss Federal Institute of Technology*, No. IKT-NT 1019, 2002.

J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society London (A)*, 209:415–446, 1909.

T.J. Mitchell. An algorithm for the construction of "d-optimal" experimental designs. *Technometrics*, 16, No.2:203–210, 1974.

B. Mukherjee. A note on sampling designs for random processes with no quadratic mean derivative. *Australian and New Zealand Journal of Statistics*, 48 (3):305–319, 2006.

T. Muller-Gronbach. Optimal designs for approximating the path of a stochastic process. *Journal of Statistical Planning and Inference*, 49:371–385, 1996.

M.A. Naimark. *Linear Differential Operators / translated from the Russian! by E.R. Dawson, English translation edited by W.N. Everitt Part 1, Elementary theory of linear differential operators, with additional material by the author.* Harrap, London, 1968.

A.J. Newman. Model reduction via the karhunen-loeve expansion part ii: Some elementary examples. *Technical Report, University of Maryland, College Park, MD*, T.R. 96-33, 1996a.

A.J. Newman. Model reduction via the karhunen-loeve expansion part i: An exposition. *Technical Report, University of Maryland, College Park, MD*, T.R. 96-32, 1996b.

Y. Nievergelt. *Wavelets Made Easy.* Birkhauster, 1999.

M. Opper and F. Vivarelli. General bounds on bayes errors for regression with gaussian processes. *Advances in Neural Information Processing Systems*, 11, 1999.

K.K. Phoon, H.W. Huang, and S.T. Quek. Comparison between karhunen-loeve and wavelet expansions for simulation of gaussian processes. *Probabilistic Engineering Mechanics*, 17:293–303, 2002a.

K.K. Phoon, S.P. Huang, and S.T. Quek. Implementation of karhunen-loeve expansion for simulation using wavelet-galerkin scheme. *Probabilistic Engineering Mechanics*, 17:293–303, 2002b.

D. Porter and D.G. Stirling. *Integral Equations: a Practical Treatment from Spectral Theory to Applications, Cambridge Texts in Applied Mathematics.* Cambridge University Press, Cambridge, 1990.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C, Second edition.* Cambridge University Press, Cambridge, UK, 1992.

J.O. Ramsay and B.W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies.* Springer, New York, 2002.

J.O. Ramsay and B.W. Silverman. *Functional Data Analysis.* Springer-Verlag, 1997.

C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning.* The MIT Press, Cambridge, Massachusetts, 2005.

K. Ritter. *Average-Case Analysis of Numerical Problems, Lecture Notes in Mathematics, 1733.* Springer-Verlag, Berlin Heidelberg, 2000.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications.* Chapman and Hall/CRC, 2005.

J. Sacks and D. Ylvisaker. Designs for regression problems with correlated error. *the Annals of Mathematical Statistics*, 37:66–89, 1966.

J. Sacks and D. Ylvisaker. Designs for regression problems with correlated errors; many parameters. *the Annals of Mathematical Statistics*, 39:49–69, 1968.

J. Sacks and D. Ylvisaker. Design for regression problems with correlated errors iii. *the Annals of Mathematical Statistics*, 41:2057–2074, 1970a.

J. Sacks and D. Ylvisaker. Statistical design and integral approximation. *Proceeding of the 12th Biennial Seminary of the Canadian Mathematical Congress*, pages 115–136, 1970b.

J. Sacks and D. Ylvisaker. Linear estimation for approximately linear models. *Annals of Statistics*, 6:1122–1137, 1978.

J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, pages 409–435, 1989.

H. Sagan. *Introduction to Calculus of Variation.* Dover, 1992.

T. Schmidt. Credit risk modeling with gaussian random fields. *Working paper*, 2004.

P. Sebastiani and H.P. Wynn. Maximum entropy sampling and bayesian experimental design. *Journal of the Royal Statistical Society: series B*, 62:145–157, 2000.

P. Shewry and H.P. Wynn. Maximum entropy sampling. *Applied Statistics*, 14: 165–170, 1987.

P. Sollich. Learning curves for gaussian processes. *Neural Information Processing Systems*, 11, 1999.

H.L.Van Trees. *Detection, Estimation, and Modulation Theory.* John Wiley and Sons, Inc., New York, 1968.

T. Vargiolu. Calibration of the gaussian musiela model using the karhunen-loeve expansion. *Atti del XXII Convegno AMASES*, 1998.

R. Vautard, P. Yiou, and M. Ghil. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, 58:95–126, 1992.

G. Wahba. *Spline Models for Observational Data, volume 59 of CBMS-NSF Regional Conference Series in Applied Mathematics.* SIAM, Philadelphia, 1990.

H.P. Wynn. Maximum entropy sampling and general equivalence theory. *mODA7-Advances in Model-Oriented Design and Analysis*, 7:211–218, 2004.

F. Yao, H.G. Muller, and J.L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590, 2005.

P. Yiou, D. Sornette, and M. Ghil. Data-adaptive wavelets and multi-scale singular-spectrum analysis. *Physica D*, 142:254–290, 2000.

D.G. Zill and M.R. Cullen. *Differential Equations with Boundary-Value Problems.* Brooks/Cole, 2001.