IOWA STATE UNIVERSITY
Digital Repository

1994

Kriging and splines, and their application to the characterization of near-circular features

Hae-Il Ahn
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Industrial Engineering Commons

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Kriging and splines, and their application to the characterization of near-circular features

Ahn, Hae-Il, Ph.D.

Iowa State University, 1994

Kriging and splines, and their application to the characterization of

near-circular features

by

Hae-Il Ahn

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Industrial and Manufacturing Systems Engineering
Major: Industrial Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Members of the Committee:

Signature was redacted for privacy.

Signature was redacted for privacy.

Signature was redacted for privacy.

Signature was redacted for privacy.

Iowa State University
Ames, Iowa
1994

TABLE OF CONTENTS

v

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1.   INTRODUCTION

Traditionally, there have been two major streams of research in spatial interpolation and prediction: kriging and splines. The two methods have developed almost independently, and each provides a variety of models. Regarding splines, an extensive comparative study was carried out by Franke [25]. He reported that radial basis function methods rank high in predictive power, when compared to other spline methods, such as tensor product, blending or finite element methods. Further, within the group of radial basis function methods, Hardy's multiquadric harmonic (MQ-H) and biharmonic (MQ-B) [29] and Duchon's thin plate spline (TPS) [21] methods performed especially well. It is worthy of note that Franke's study excluded kriging and that other studies such as Laslett *et.al.* [40], have reported kriging to be superior to splines, without considering MQ method.

The two approaches generally are thought to be distinct. In point of fact, Cressie [15] and Wahba [73] adopt somewhat partisan positions in the matter: the former for kriging and the latter for splines. On the other hand, the point of view also has been expressed that kriging and splines share essentially the same structure as in Kimeldorf and Wahba [38]. Further, Matheron [51] showed that thin plate spline (TPS) modeling is equivalent to a certain form of kriging in the sense of convergence in a Hilbert space setting. Also, Salkaukas [65] has demonstrated that kriging and TPS method are algebraically equivalent in terms of a certain generalized function concept. A third sense in which kriging and

splines are similar is espoused by Dubrule [20] and Watson [74]. The notion is extended by Myers [56] to vector valued functions viz., cokriging.

Comparisons of splines with kriging typically are performed with a view to TPS. Such comparisons rarely are done with MQ-H or MQ-B in mind. The reason for this may be that the origins of MQ lie in the physical sciences rather than in mathematics.

This work reviews the equivalence of kriging and splines, from a variety of points of view. Their common basis in such matters as model building and validity checking of kernels are reinvestigated, with a view to an application to the geometric-shape rendering of manufactured parts. Our main illustrative emphasis is on the engineering problem of characterizationing the attained shape of manufactured objects intended to be "round".

Chapter 2 is reserved for an exposition of the author's perception of the common ground of splines and kriging. Chapter 3 is devoted to a description of kriging with a view to comparison with splines, while section 4 deals reciprocally with splines.

Chapter 5 examines some details of unified model building and explores the possibility of validity checking in relation to MQ kernels. The author's perception of the common aspects of splines and kriging is exposed in chapter 6. In chapter 7, the author's interpretation of MQ model in kriging terms is discussed. In chapter 8, the links between existing spatial prediction methods including least square method were examined with a view to interpreting the spatial prediction method as a generalization of least square method.

In the remaining chapters 9 and 10, an application to a manufacturing problem is studied, particularly for the estimation of geometric objects intended to exhibit some form of roundness, such as circles, spheres, cylinders and toruses, in an attempt to develop an approach to geometric-shape rendering based on the material discussed in the previous chapters.

CHAPTER 2.   COMMON GROUND OF SPLINES AND KRIGING

Kriging and spline models have at least one point of view in common : irregular surfaces $z(x)$ viewed as realizations of stochastic process $Z(x)$, these viewed in turn as superpositions of large and small-scale components. In some cases, such a superposition model will include a micro-scale component, typically interpreted as measurement error; thus $Z(x)$ is viewed as composed of large (μ), small (ε) and micro(ϑ) scale variations:

$$Z(x) = \mu(x) + \varepsilon(x) + \vartheta(x) \tag{2.1}$$

Here, $\mu(x)$ is deterministic but $\varepsilon(x)$ and $\vartheta(x)$ are random with statistical expectation zero. It is also usually assumed that $\varepsilon(x)$ and $\varepsilon(x')$, $x \neq x'$, are correlated, while $\vartheta(x)$ and $\vartheta(x')$ are not. Further, both krigers and spliners assume that $z(x)$ is available on a sub-domain $\mathcal{D}'$ of a domain $\mathcal{D} \subset \mathcal{R}^n$ of interest. Throughout this paper, $\mathcal{D}'$ is assumed to be a finite set of distinct points: $\mathcal{D}' \equiv \{x_1, \cdots, x_N\}$. The observations $z(\mathcal{D}') \equiv \{z(x_i)\}_{i=1}^{N}$ is used to deduce an estimate:

$$\hat{z}(x) \equiv m(x) + e(x) \tag{2.2}$$

of $z(x)$ on $\mathcal{D}$. When $\mu(x)$ is the multivariable polynomial of degree $k-1$, the model for $Z(x)$ is said to be of order k. If the large scale part is omitted, it is of order 0.

Both krigers and spliners avail themselves of the option of excluding the micro-scale variation ϑ from the model; in that case, krigers call the problem of developing an $\hat{z}(\cdot)$

one of predictions by interpolation, the latter term also being common among spliners. When micro-scale variation is included, both disciplines speak of "measurement error" and "smoothing". Krigers deal with measurement error through introducing a "nugget effect", while spliners invoke a certain variational principle.

2.1 Large-scale Variation

The large-scale variation $\mu(x)$ of $z(x)$ is usually represented as a multivariable polynomial $p_k(x)$ of degree $k - 1$. Let $\mathcal{Q}_k$ be the set $\{q(x)|q(x) = x_1^{\xi_1} \cdots x_n^{\xi_n} : 0 \leq \xi_1 + \cdots + \xi_n \leq k - 1\}$. The cardinality $|\mathcal{Q}_k|$ of this set is $\binom{n+k-1}{n} = m$. A polynomial of degree, say $k - 1 = 2$, in, say $n = 2$, dimensions will then consist of $m = 6$ monomials; $q_1(x) = 1$, $q_2(x) = x$, $q_3(x) = y$, $q_4(x) = x^2$, $q_5(x) = xy$ and $q_6(x) = y^2$. Let $\mathcal{P}_k$ be the class of multivariate polynomials of degree $k - 1$. A polynomial $p_k(x)$ of degree $k - 1$ can be expressed as $\sum_{j=1}^{m} \beta_j q_j(x)$, where the β_j's are real constants. For the discussion in later sections, it is useful to introduce the following constructs related to large-scale variation.

$$
X \equiv \begin{bmatrix} q_1(x_1) & q_2(x_1) & \cdots & q_m(x_1) \\ q_1(x_2) & q_2(x_2) & \cdots & q_m(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ q_1(x_N) & q_2(x_N) & \cdots & q_m(x_N) \end{bmatrix} \quad \mathbf{x} \equiv \begin{bmatrix} q_1(x) \\ q_2(x) \\ \vdots \\ q_m(x) \end{bmatrix} \quad \mathbf{z} \equiv \begin{bmatrix} z(x_1) \\ z(x_2) \\ \vdots \\ z(x_N) \end{bmatrix} \quad (2.3)
$$

In the context of known z and k, a key task for both krigers and spliners is to estimate $\beta \equiv (\beta_1, \cdots, \beta_m)'$ and therefore also $p_k(x) = \sum_{j=1}^{m} \beta_j q_j(x) = \beta' \mathbf{x}$.

Example 2.0.1 *In the case of multivariable polynomial regression, the model can be identified as $Z(x) = \mu(x) + \vartheta(x)$, where $\vartheta(x)$ consists of independent errors. The associated normal equation is given by $\hat{\beta} = \mathbf{b} = (X'X)^{-1} X' \mathbf{z}$. The corresponding estimates $\hat{\mu}(x)$ of $\mu(x)$ can be expressed as $\hat{\mu}(x) = \mathbf{b}' \cdot \mathbf{x} = \mathbf{z}'X \cdot (X'X)^{-1} \cdot \mathbf{x}$.*

2.2 Small-scale Variation

When large and small-scale variations are considered, the model for $z(\cdot)$ becomes:

$$z(x) = \sum_{j=1}^{m} \beta_j q_j(x) + \epsilon(x) = \beta' \cdot x + \epsilon(x) \qquad (2.4)$$

where $\epsilon(x)$ is a mutually correlated residual, with estimated interpolant: $\hat{z}(x) = b'x + e(x)$, where $b \equiv \hat{\beta}$ and $e(x) = \hat{\epsilon}(x)$. The correlation structure of $e(x)$, and the interpolation and treatment of that correlation structure, is the key part of the model and the major source of difference among interpolation methods such as splines, kriging and possibly other types of techniques. The structure described below generalizes and unifies the above viewpoints. Given $z(x_i), i = 1, \cdots, N$, the interpolant $\hat{z}(x)$ must satisfy

$$\hat{z}(x_i) = \sum_{j=1}^{m} b_j q_j(x_i) + e(x_i) = \sum_{j=1}^{m} \beta_j q_j(x_i) + \epsilon(x_i) = z(x_i) \qquad (2.5)$$

The problem common to krigers and spliners is to choose an interpolant (or approximant)

$$\hat{z}(x) = \sum_{j=1}^{m} b_j q_j(x) + e(x) \qquad (2.6)$$

which amounts to choosing $\{b_j\}$ and $\{e(x_i)\}$ subject to (2.5). To analyze this type of *general order* model, a so-called "detrending" technique is employed. Detrending is applicable to "smoothing" models that incorporate independent errors $\vartheta(x)$, as well as to interpolation models such as (2.4) that do not. For maximal generality, the next section discusses detrending in the former case.

2.3 Detrending

This section on detrending emphasizes conceptual approaches to eliminate the trend rather than actually implementable data analytic detrending methodologies. The problem

faced by both krigers and spliners is how to treat the nuisance parameter β. This problem is in effect approached by both schools in the following way. The whole idea is to *filter*, linearly, the data z such that $Gz = G\epsilon$, where G is a *linear operator* (matrix). There are several known filtering methods depending on the matrix G used. In this study, only the subtraction and partial differential operator ∇ is being considered. The easiest way of detrending is to use the subtraction,

$$Z_0(x) = Z(x) - \beta' x = \varepsilon(x) + \vartheta(x) \tag{2.7}$$

in which case the nature of the stochastic processes $\varepsilon(x)$ and $\vartheta(x)$ remain intact. The more sophisticated one would be to use the *partial differential operator* ∇,

$$Z_0(x) = \nabla^k Z(x) = \nabla^k \mu(x) + \nabla^k \varepsilon(x) + \nabla^k \vartheta(x) = \varepsilon_0(x) + \vartheta_0(x) \tag{2.8}$$

The large scale part $\mu(x)$ vanishes and the nature of the stochastic process changes. As noted, the detrended model embraces small-scale and micro-scale variations only.

$$Z_0(x) = \varepsilon_0(x) + \vartheta_0(x) \tag{2.9}$$

Note that after being detrended, the nature of the stochastic process changes. Basically, it is assumed that $\varepsilon_0(x)$ and $\vartheta_0(x)$ are covariance stationary white noise processes, the kernels of which must be of particular type, as will be discussed in later sections.

2.4 Whitening Operation and Positive Transformation

Now we consider the simple algebra involved in the calculation of an interpolant, which corresponds to the above mentioned reproducing kernel hilbert space approach. Consider the simplest form of a spatial prediction model that reflects only the small

scale variation $Z(\boldsymbol{x}) = \varepsilon(\boldsymbol{x})$ taken to be the detrended model. Let us define a vector $\mathbf{E} = [\varepsilon(\boldsymbol{x}_1), \cdots, \varepsilon(\boldsymbol{x}_N)]'$ of correlated random variables $\{\varepsilon(\boldsymbol{x}_i)\}_{i=1}^{N}$ with mean zero and finite variance. Suppose $\mathbf{E} \sim N(\mathbf{0}, \Sigma)$, where Σ is a positive definite variance covariance matrix. If we define $\mathbf{N} = T^{-1}\mathbf{E} = [v(\boldsymbol{x}_1), \cdots, v(\boldsymbol{x}_N)]'$, where T is a symmetric and invertible square root of Σ as in appendix 12.3, then $\mathbf{N} \sim N(\mathbf{0}, I)$, where $T^{-1} \equiv \Sigma^{-1/2}$ is called *Whitening operator* that changes correlated random variables $\{\varepsilon(\boldsymbol{x}_i)\}_{i=1}^{N}$ into independent random variables $\{v(\boldsymbol{x}_i)\}_{i=1}^{N}$. Again, let us define $\mathbf{A} = \Sigma^{-1/2}\mathbf{N} = [\alpha(\boldsymbol{x}_1), \cdots, \alpha(\boldsymbol{x}_N)]'$, then $\mathbf{A} \sim N(\mathbf{0}, \Sigma^{-1})$. Further define $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ by $\mathbf{E} = I\mathbf{E} = T\mathbf{N} = T^2\mathbf{A}$. Let the observation vectors of $\mathbf{E}$, $\mathbf{N}$ and $\mathbf{A}$ be $\boldsymbol{\epsilon}$, $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$, respectively. Then the detrended observation vector $\mathbf{z}$ can be expressed as

$$\mathbf{z} = I\boldsymbol{\epsilon} = T\boldsymbol{\nu} = T^2\boldsymbol{\alpha} = \Sigma\boldsymbol{\alpha} \qquad (2.10)$$

In the language of vector spaces, $\boldsymbol{\epsilon}$ and $\boldsymbol{\alpha}$ are mutually *covariant* and *contravariant*. Both for kriging and splines, $\boldsymbol{\nu}$ is regarded as a vector of independent residuals. For simplicity, we introduce the following vectors of kernel functions functions:

$$\boldsymbol{\lambda} \equiv \begin{bmatrix} \lambda(\boldsymbol{x}, \boldsymbol{x}_1) \\ \lambda(\boldsymbol{x}, \boldsymbol{x}_2) \\ \vdots \\ \lambda(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix}, \quad \boldsymbol{\tau} \equiv \begin{bmatrix} \tau(\boldsymbol{x}, \boldsymbol{x}_1) \\ \tau(\boldsymbol{x}, \boldsymbol{x}_2) \\ \vdots \\ \tau(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix} \text{ and } \boldsymbol{\kappa} \equiv \begin{bmatrix} \kappa(\boldsymbol{x}, \boldsymbol{x}_1) \\ \kappa(\boldsymbol{x}, \boldsymbol{x}_2) \\ \vdots \\ \kappa(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix} \qquad (2.11)$$

An interpolant can be obtained by specifying a kernel $\{\lambda(\cdot)\}$ such that $I \equiv (\lambda_{ij})_{i,j=1}^{N}$ with $\lambda_{ij} = \lambda(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Alternatively, one can obtain an interpolant by specifying $\{\tau(\cdot)\}$ such that $T \equiv (\tau_{ij})_{i,j=1}^{N}$ with $\tau_{ij} = \tau(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Still another method common to kriging and splines for obtaining an interpolant is to specify a function $\{\kappa(\cdot)\}$ such that $\Sigma \equiv T^2 \equiv (\kappa_{ij})_{i,j=1}^{N}$ with $\kappa_{ij} \equiv \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where Σ is *positive definite* because it is a multiplication

of two identical symmetric invertible matrices. The interpolant corresponding to (2.10) is an extension of observations beyond the data points using the relation $\nu = T\alpha = T^{-1}\epsilon$ and $\tau = T^{-1}\kappa = T\lambda$.

$$\hat{z}(x) = \hat{\epsilon}(x) \;=\; \epsilon' \cdot \lambda \;=\; \epsilon' \cdot T^{-1} \cdot \tau \;=\; \epsilon' \cdot \Sigma^{-1} \cdot \kappa \qquad (2.12)$$

$$= \; \epsilon' \cdot \lambda \;=\; \nu' \cdot \tau \;=\; \alpha' \cdot \kappa \qquad (2.13)$$

Note that equations in (2.12) are standard form of Hilbert space approach. Usually, krigers try to estimate the covariance kernel $\kappa(\cdot)$ from the given observational data, while spliners choose and specify splines $k(\cdot)$ on their own criterion.

The purpose of this chapter has been to introduce the reader to the notation involved in the subsequent chapters, and to models common to both kriging and splines in the context of linear spatial prediction models. The rest of this work is based on this notation and these models.

CHAPTER 3. KRIGING WITH REGIONAL DATA

The spatial prediction method named "kriging" is an analogue for spatial stochastic processes of Wiener-Kolmogorov prediction theory that has been developed and used mainly by Matheron and his school in mining industry. In this method, an irregular surface $z(x)$ is regarded as a realization of multi-dimensional random variable $Z(x)$, where $\{Z(x) : x \in \mathcal{D} \subset \mathcal{R}^n\}$ is thought of as a real-valued stochastic process defined on a domain $\mathcal{D}$ of $\mathcal{R}^n$. Conceptually, we divide the random process $Z(x)$ into a drift or trend $\mu(x)$ and a small-scale random process $\varepsilon(x)$: $Z(x) = \mu(x) + \varepsilon(x)$. The distinction between the drift (trend) and fluctuation is not clear cut. Usually, $\mu(x)$ is a multivariate polynomial. Generally, we think of the drift as a large scale variation, regarded as fixed, and the fluctuation as a small-scale variation. The model could encompass a micro-scale process for pure "nugget effect" or measurement error of the process.

The kriging model for spatial prediction is based on the multivariate normal distribution, which is assumed throughout. The model for interpolation takes the form of $\hat{Z}(x) = \hat{\mu}(x) + \hat{\varepsilon}(x)$. The model type depends on the way we regard $\varepsilon(x)$. In the literature, there are many types of models for kriging. In some cases, the small-scale variation is viewed as a white noise process. Under this assumption, the estimation lends itself to *Ordinary Least Square*. In some other cases, $\varepsilon(x)$ is regarded as a Brownian motion or a higher order random process. The prediction of kriging is much the same as Generalized

Linear Regression analysis and *logged likelihood analysis.*

3.1 Stationarity

The most common assumption of kriging is that the stochastic process is *stationary* in the domain of interest ($\mathcal{D} \subset \mathcal{R}^n$). The term *stationary*, however, can have different meanings and different degrees. A stochastic process $\{Z(x) : x \in \mathcal{D} \subset \mathcal{R}^n\}$ is called

1. *Strictly Stationary* : if the joint distribution of $\{Z(x_1), \cdots, Z(x_N)\}$ is the same as the joint distribution of $\{Z(x'_1), \cdots, Z(x'_N)\}$ for any set of points $\{x_i\}_{i=1}^N$ and $\{x'_i\}_{i=1}^N \equiv \{x_i + h\}_{i=1}^N \in \mathcal{R}^n$.

2. *Covariance Stationary* : if it has finite *second moments* and satisfies 1) $E[Z(x)]$ exists for all $x \in \mathcal{R}^n$, and 2) $Cov[Z(x'), Z(x)]$ is a function of $\|x' - x\|$ for all $(x', x) \in \mathcal{R}^n \times \mathcal{R}^n$, where $\| \cdot \|$ is the Euclidean norm. A covariance stationary stochastic process is also called *Wide Sense Stationary* or *Second Order Stationary*.

3. *Intrinsically Stationary* : if it satisfies 1) $E[Z(x)]$ exists for all $x \in R^n$, and 2) for any $(x', x) \in \mathcal{R}^n \times \mathcal{R}^n$, the increment variable $[Z(x') - Z(x)]$ has finite variance, which depends only on $\|x' - x\|$. An intrinsic stationary stochastic process is also called *Variogram Stationary* or *Wider Sense Stationary*, since 3. does in fact generalizes 2.

3.2 Covariogram and Variogram

The covariogram and variogram are functions that characterize the second moment dependence properties of a stochastic process defined on $\mathcal{R}^n$, and underlie intrinsic stationarity or variogram stationarity a further sort of stationarity. They play an important

role in stochastic spatial prediction or *kriging*. In particular, with $x' - x = h$, suppose that for the increment variable $[Z(x') - Z(x)]$,

$$E[Z(x') - Z(x)] = 0, \quad \text{for all } x', x \in \mathcal{D} \tag{3.1}$$

$$Var[Z(x') - Z(x)] = 2\gamma(h), \quad \text{for all } x', x \in \mathcal{D} \tag{3.2}$$

The quantity $2\gamma(\cdot)$, which is a function only of the *difference* between the spatial locations x and x', has been called the *variogram* by Matheron [49]. Earlier appearances in the scientific literature can be found and have been called *structure functions* in physics and meteorology, and a *serial variation function* in time series. Replacing equation (3.2) by

$$Cov\left[Z(x'), Z(x)\right] = \kappa(h) \text{ for all } x, x' \in \mathcal{D} \tag{3.3}$$

defines the class of *covariance stationary* processes. Here $\kappa(\cdot)$ is called the covariograms or covariance kernels. Usually, the semi-variograms, one half of the variograms, have the properties of $\gamma(-h) = \gamma(h)$ and $\gamma(0) = 0$. When the semi-variogram $\gamma(h) = \gamma^0(\|h\|)$, it is said to be *isotropic* or *rotationally symmetric*. If $Var[Z(x)] < +\infty$ is not a function of x and we define $h = \|h\|$, the *semi-variogram* $\gamma^0(h)$ is related very simply to the covariance kernel $\kappa^0(\cdot)$ if $\kappa^0(h)$ exists. This can be easily seen from the followings

$$\gamma^0(h) = \frac{1}{2}Var\left[Z(x') - Z(x)\right] \tag{3.4}$$

$$= \frac{1}{2}\left\{Var[Z(x')] + Var[Z(x)] - 2Cov[Z(x'), Z(x)]\right\} \tag{3.5}$$

$$= \frac{1}{2}\{\kappa^0(0) + \kappa^0(0) - 2\kappa^0(h)\} = \kappa^0(0) - \kappa^0(h) \tag{3.6}$$

Further, it is often seen that $\lim_{h\to\infty} \gamma^0(h) = \kappa^0(0)$, called the *sill*, by the the near-independence of distant quantities.

While conventional models envisage $\gamma^0(h)$ with finite asymptote $\kappa^0(0)$, one can extend the class of covariance stationary processes to include cases where the variance

$\gamma^0(h) = Var[Z(x') - Z(x)]$ can become arbitrarily large as $h \to \infty$. Thus, one obtains the class of the *intrinsically stationary* processes, defined by (3.1) and (3.2), that includes covariance stationary processes with finite sill as well as stationary-increment processes like *Brownian motion*. Thus the class Ω_1 of variogram-stationary processes strictly contains the class Ω_2 of second-order (covariance) stationary processes, and define the class Ω_3 of variogram-only-stationary processes as $\Omega_3 = \Omega_1 \backslash \Omega_2$, we have that

$$\Omega_1 = \Omega_2 + \Omega_3 \tag{3.7}$$

The major distinction between *variogram-only-stationary* processes and *covariance stationary* processes is that the latter posesses a covariogram depending on $h = \|x' - x\|$, in which case the variogram will have a finite sill, while the former does not. Various variogram models are presented in Journel and Huijbregts [35] and Salkaukas [65]. Variogram models are more versatile than covariogram models in the sense that (3.2) holds more often than does (3.3). In this connection, Matheron [50] has invented a *generalized covariance* concept. The functional relationship between the (generalized) covariance $\kappa(\cdot)$ and the semi-variogram $\gamma(\cdot)$ is the same as that given by (3.1), (3.2) and (3.3). Kitanidis [39] discusses on the estimation of generalized covariance function from the data.

Example 3.0.2 *Suppose* $\{Z(x)|Z(x) = \mu + \varepsilon(x), x \in \mathcal{D}\}$ *is a Brownian motion process, where μ is a constant. The increment variable* $[Z(x') - Z(x)] = [\varepsilon(x') - \varepsilon(x)]$ *is normally distributed with mean zero and variance* $Var[\varepsilon(x') - \varepsilon(x)] = \gamma^0(h) = \sigma^2 h$, *where σ^2 a fixed parameter and $h \equiv \|x' - x\| \geq 0$. Then, $Z(x)$ is an intrinsically stationary, but not covariance stationary, stochastic process. When $\sigma^2 = 1$, the process is often called standard Brownian motion.*

3.3 Choices of Covariogram and Variogram

In this work, $\kappa^0(h)$ represents the generalized covariance kernels, unless stated otherwise. Examples of the covariance-stationary and variogram stationary covariograms are given in Fig. 3.1 and Fig. 3.2, respectively. For the former, the sill is set equal to c_1.
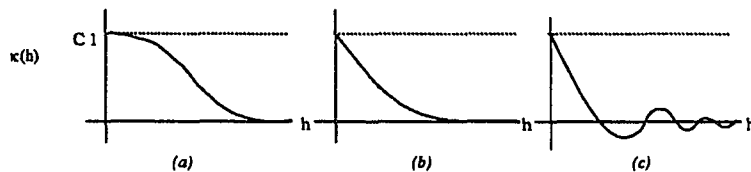
Figure 3.1: Covariance-Stationary Semi-variograms
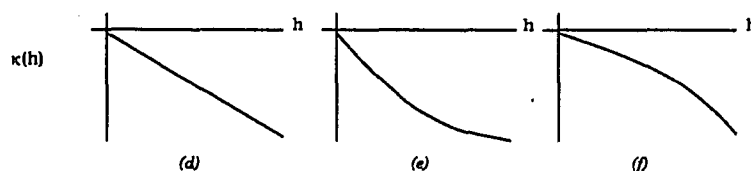
Figure 3.2: Variogram-Only-Stationary Semi-variograms

Among the covariance-stationary functions $\kappa^0(h)$ are the (a) Gaussian, (b) Exponential, and (c) Hole Effect, while the (d) Linear, (e) Logarithmic, and (f) Power functions are examples of Variogram-only-stationary functions.

We now consider a family of (generalized) covariance functions called *Gaussian Family* by Whittle [75] and largely used by krigers that include most existing (generalized) covariance functions as special cases.

$$\kappa_\nu^0(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{a}\right)^\nu \mathcal{K}_\nu \left(\frac{h}{a}\right) \quad , \quad \nu > 0 \tag{3.8}$$

where $\mathcal{K}_\nu(\cdot)$ is the modified Bessell function of order ν. The *range* parameter a works in the same way as *scale* parameters do in the *exponential* and *Gaussian* functions. The parameter ν is the order and controls the smoothness of the process. This family can be easily generalized to higher dimensions. For smoothing, krigers construct the following function using the kernel $\kappa_\nu^0(h)$

$$C_\nu(h) = c_0\delta(h) + c_1\kappa_\nu^0(h) \tag{3.9}$$

where $\delta(\cdot)$ is the indicator function (or white noise kernel), c_0 the nugget, and $(c_0 + c_1)$ the sill.

Now let us discuss how the kernel is tied to the order of the model. When one assumes the model $Z(x) = \mu(x) + \varepsilon(x)$, $\varepsilon(x)$ is the process that has the associated (generalized) covariance function $\kappa^0(h)$ and semi-variogram $\gamma^0(h)$. The partial derivative $\nabla\varepsilon(x)$ is also a stochastic process, whose associated covariogram or semi-variogram becomes negative Laplacians of the original ones, provided they are isotropic or radially symmetric. Suppose that $Z_k(x)$ is a k times "differentiable" random function on $\mathcal{D} \subset \mathcal{R}^n$ with a semi-variogram $\gamma_k^0(h)$ or a covariogram $\kappa_k^0(h)$. Assume the model to be: $Z_k(x) = \mu_k(x) + \varepsilon_k(x)$, where $\mu_k(x)$ is a polynomial of degree not exceeding $k-1$ and $\varepsilon_k(x)$ is a k times differentiable small-scale stochastic process. Let $\gamma_{k-1}^0(h)$ and $\kappa_{k-1}^0(h)$ be the associated semi-variogram and covariogram of $\nabla Z_k(x)$. As shown in Matheron [49] and Kent and Mardia [37],

$$\gamma_{k-1}^0(h) = -\Delta\gamma_k^0(h) \tag{3.10}$$

$$\kappa_{k-1}^0(h) = -\Delta\kappa_k^0(h) \tag{3.11}$$

where Δ is the Laplacian operator ∇^2. In the same fashion, if we let $\nabla^\ell Z_k(x)$ have associated semi-variogram and covariogram $\gamma_{k-\ell}^0(h)$ or $\kappa_{k-\ell}^0(h)$, then they can be represented

as follows

$$\gamma^0_{k-\ell}(h) = (-1)^\ell \Delta^\ell \gamma^0_k(h) \tag{3.12}$$

$$\kappa^0_{k-\ell}(h) = (-1)^\ell \Delta^\ell \kappa^0_k(h) \tag{3.13}$$

These facts are enough to give one a hint that certain random functions can be expanded in a finite Taylor-like series with the remainder expressed in terms of a white noise process $\varepsilon_0(x)$, which is the main topic of chapter 5.

Example 3.0.3 *White noise is defined by analogy to the continuous energy distribution in white light from an incandescent body. A covariance stationary stochastic process which has equal power in all frequency intervals over a wide frequency range is called white noise. For most purposes, it suffices to regard white noise as a normal process with covariogram*

$$\kappa^0(h) = \rho e^{-\rho h} \quad 0 < h < \infty \tag{3.14}$$

where ρ is so large as to be considered infinite. If ρ is thought of as a variance of the process, the semi-variogram becomes

$$\gamma^0(h) = \kappa^0(0) - \kappa^0(h) = \rho - \rho e^{-\rho h} \quad 0 < h < \infty \tag{3.15}$$

With finite ρ, (3.14) becomes the semi-variogram of covariance stationary processes. There could exist many possible ways of giving a precise mathematical definition of white noise.

Typical covariograms (or semi-variograms) of the white noise and Brownian motion are shown in Fig 3.3. The covariogram (or semi-variogram) of the white noise process works like a *Dirac Delta Function* encountered frequently in differential equation problems and regarded as a *Generalized Function*. A white noise kernel is characterized by complete monotonicity, which plays a key role in validity checking of kernels, as will be discussed in
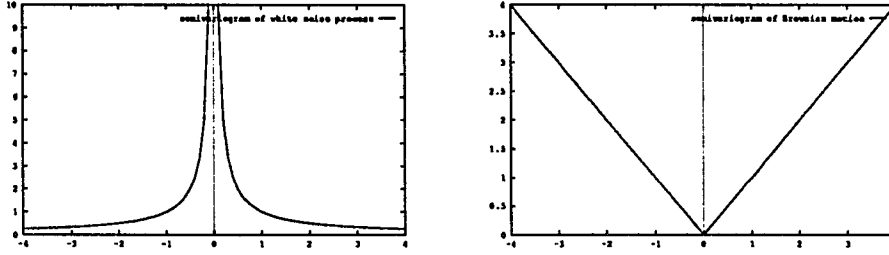
Figure 3.3: White Noise (a) and Standard Brownian Motion (b)

chapter 5. It would be of help in later chapters to note that the derivative of a white noise kernel is again a white noise kernel, and one can take the Laplacian an infinite number of times. The sign of the Laplacian alternates, whenever the Laplacian is taken, and hence the white noise kernel is completely monotonic.

3.4 Kriging Models of General Order

The use of kernels in the model is interrelated with the order of the model. The kriging predictor defined as $\hat{Z}(x) = \sum_{i=1}^{N} \lambda_i(x) Z(x_i)$ with $\lambda_i \equiv \lambda_i(x) \equiv \lambda(x, x_i)$ is known as the *Best Linear Unbiased Predictor* (*BLUP*), where $\lambda_j(x_i) = \lambda(x_i, x_j) = \lambda_{ij} = \delta_{ij}$. Let $z = [z(x_1), \cdots, z(x_N)]'$ be an observation vector of the random vector $Z = [Z(x_1), \cdots, Z(x_N)]'$ in the context of known $\{x_i, z(x_i)\}_{i=1}^{N}$, then the kriging predictor for $\lambda = [\lambda_1(x), \lambda_2(x), \cdots, \lambda_N(x)]'$ becomes

$$\hat{Z}(x) = \sum_{i=1}^{N} \lambda_i(x) \cdot z(x_i) = \lambda' \cdot z \qquad (3.16)$$

The *Unbiasedness* requires the following condition: $E[\hat{Z}(x)] = E\left[\sum_{i=1}^{N} \lambda_i(x) Z(x_i)\right]$ which reduces to $x'\beta = \lambda' X \beta$ or $x = X'\lambda$. A key task for krigers is to determine the order k: $p_k(x) = \sum_{j=1}^{m} \beta_j q_j(x) = \beta' \cdot x$ as in section 2. Depending on the order k, there

exist three types of *Covariance Stationary* models:

1. Simple Kriging ($k=0$): $\mu(\boldsymbol{x})$ is known or zero.

2. Ordinary Kriging ($k=1$): $\mu(\boldsymbol{x})$ is an unknown constant.

3. Universal Kriging ($k > 1$): $\mu(\boldsymbol{x})$ is an unknown multivariate polynomial.

In the same manner, the *variogram-only-stationary* models can be classified in accordance with the order ranging from 0 to $k > 1$.

1. Order 0 ($k=0$): $\mu(\boldsymbol{x})$ is known or zero.

2. Order 1 ($k=1$): $\mu(\boldsymbol{x})$ is an unknown constant.

3. Order k ($k > 1$): $\mu(\boldsymbol{x})$ is an unknown multivariate polynomial.

According to the classification by Matheron [50] and Chritakos [11], a 0-IRF model is of order 1, a 1-IRF model order 2 and so on. Based on stochastic process theory, order 0 model corresponds to white noise and order 1 to Brownian motion. If $k > 1$, then it is called Brownian motion of higher order.

3.5 Objective of Kriging

Kriging predictors are devised with a view to an *Error Variance Reduction* at prediction points $\boldsymbol{x}$, which is often referred to as a *Conditional Variance* reduction. The variance of $Z(\boldsymbol{x})$ can be decomposed into two parts; *Conditional Variance (CV)* and *Variance of Conditional Expectation (VCE)*:

$$Var\{Z(\boldsymbol{x})\} = Var\{Z(\boldsymbol{x}) - \hat{Z}(\boldsymbol{x})\} + Var\{\hat{Z}(\boldsymbol{x})\} \tag{3.17}$$

$$= Var\left[Z(\boldsymbol{x}) - \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{x}_i)\right] + Var\left[\sum_{i=1}^{N} \lambda_i Z(\boldsymbol{x}_i)\right] \tag{3.18}$$

The second term of right hand side of (3.17) is the variance of conditional expectation, which is less than $Var\{Z(x)\}$ itself by virtue of having observed data. Actually, use of $\hat{Z}(x)$ minimize the first term, which is one way of achieving what kriger's hoped-for error variance reduction.

3.6 Covariogram Modeling

Together with the unbiasedness of the estimates, the objective can be given by

$$\min_{\boldsymbol{\lambda}} \; Var\left[Z(\boldsymbol{x}) - \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{x}_i) \right] = \min_{\boldsymbol{\lambda}} \; E\left[\left(Z(\boldsymbol{x}) - \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{x}_i) \right)^2 \right] \quad (3.19)$$

$$\text{subject to} \quad \sum_{i=1}^{N} \lambda_i q_j(\boldsymbol{x}_i) = q_j(\boldsymbol{x}) \; , \; j = 1, \cdots, m \quad (3.20)$$

Note that in case of simple kriging ($k=0$), no such constraint is needed. In general, however, $Z(x)$ is supposed to be k-th order stochastic process. After removing the large-scale part from the model, the MSE function can be rearranged as follows:

$$E\left[\left(Z(\boldsymbol{x}) - \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{x}_i) \right)^2 \right] = E\left[\left(\mathbf{x}'\boldsymbol{\beta} + \varepsilon(\boldsymbol{x}) - \boldsymbol{\lambda}'X\boldsymbol{\beta} - \sum_{i=1}^{N} \lambda_i \varepsilon(\boldsymbol{x}_i) \right)^2 \right] \quad (3.21)$$

$$= E\left[\left(\varepsilon(\boldsymbol{x}) - \sum_{i=1}^{N} \lambda_i \varepsilon(\boldsymbol{x}_i) \right)^2 \right] \quad (3.22)$$

since the unbiasedness condition $\mathbf{x}'\boldsymbol{\beta} = \boldsymbol{\lambda}'X\boldsymbol{\beta}$ holds. Thus, the optimization becomes

$$\min_{\boldsymbol{\lambda}} \; Var[\varepsilon(\boldsymbol{x})] + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \; Cov[\varepsilon(\boldsymbol{x}_i), \varepsilon(\boldsymbol{x}_j)] - 2 \sum_{i=1}^{N} \lambda_i \; Cov[\varepsilon(\boldsymbol{x}), \varepsilon(\boldsymbol{x}_i)] \quad (3.23)$$

$$\text{subject to} \quad \sum_{i=1}^{N} \lambda_i q_j(\boldsymbol{x}_i) = q_j(\boldsymbol{x}) \; , \; j = 1, \cdots, m \quad (3.24)$$

The problem is identified as an *Equality Constrained Quadratic Optimization (ECQO)*:

$$\min_{\boldsymbol{\lambda}} \; \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\boldsymbol{\kappa} + \kappa(0) \quad \text{subject to} \quad X'\boldsymbol{\lambda} = \mathbf{x} \quad (3.25)$$

where $\kappa(0) = Var[\varepsilon(\boldsymbol{x})]$. Let the *Lagrangian Multiplier* be $2\boldsymbol{\omega} = 2(\omega_1, \omega_2, \cdots, \omega_m)'$. Upon taking derivatives with respect to $(\boldsymbol{\lambda}', \boldsymbol{\omega}')$, this optimization problem reduces to a linear system called *First Order Necessary Condition (FONC)*

$$\Sigma\boldsymbol{\lambda} + X\boldsymbol{\omega} = \boldsymbol{\kappa} \tag{3.26}$$

$$X'\boldsymbol{\lambda} = \mathbf{x} \tag{3.27}$$

The system of equation solves uniquely, if the matrix Σ is *Positive Definite* on the *Subspace* defined by $\mathcal{M} = \{\boldsymbol{\lambda} : X'\boldsymbol{\lambda} = \mathbf{0}\}$.

3.7 Variogram Modeling

The *Covariogram Model* changes to the *Variogram Model* through the relations

$$\kappa(\boldsymbol{x}, \boldsymbol{x}_j) = Cov[\varepsilon(\boldsymbol{x}), \varepsilon(\boldsymbol{x}_j)] = Var[\varepsilon(\boldsymbol{x})] - \gamma(\boldsymbol{x}, \boldsymbol{x}_j) \tag{3.28}$$

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = Cov[\varepsilon(\boldsymbol{x}_i), \varepsilon(\boldsymbol{x}_j)] = Var[\varepsilon(\boldsymbol{x}_i)] - \gamma(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{3.29}$$

On condition that $\sum_{i=1}^{N} \lambda_i = 1$, the equation (3.23)-(3.24) can be rearranged as

$$\min_{\boldsymbol{\lambda}} \; 2\sum_{i=1}^{N} \lambda_i \; \gamma(\boldsymbol{x}, \boldsymbol{x}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \; \gamma(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{3.30}$$

$$\text{subject to} \; \sum_{i=1}^{N} \lambda_i q_j(\boldsymbol{x}_i) = q_j(\boldsymbol{x}) \tag{3.31}$$

Let us define the matrix $V = (\gamma_{ij})_{i,j=1}^{N}$ with $\gamma_{ij} = \gamma(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{\gamma} = [\gamma(\boldsymbol{x}, \boldsymbol{x}_1), \cdots, \gamma(\boldsymbol{x}, \boldsymbol{x}_N)]'$. In turn, the objective function can be given by

$$\min_{\boldsymbol{\lambda}} \; 2\boldsymbol{\lambda}'\boldsymbol{\gamma} - \boldsymbol{\lambda}'V\boldsymbol{\lambda} \quad \text{subject to} \quad X'\boldsymbol{\lambda} = \mathbf{x} \tag{3.32}$$

This is also an *ECQO*, the *FONC* of which is the following

$$V\lambda + X\omega = \gamma \tag{3.33}$$

$$X'\lambda = \mathbf{x} \tag{3.34}$$

where 2ω is the *Lagrangian Multiplier*. Just as in covariogram modeling, the solution in λ and ω are uniquely determined if $V_A = \begin{bmatrix} V & X \\ X' & \mathbf{0} \end{bmatrix}$ is invertible. The system of equations solves uniquely, if the matrix V is *positive (negative) definite* on the tangent subspace $\mathcal{M} = \{\lambda : X'\lambda = \mathbf{0}\}$. From equation (3.33),

$$\lambda = V^{-1}(\gamma - X\omega) = V^{-1}\gamma - V^{-1}X\omega \tag{3.35}$$

Due to the unbiasedness condition (3.34), it follows that $X'\lambda = X'V^{-1}(\gamma - X\omega) = \mathbf{x}$. Thus, the estimates of λ and ω can be obtained as follows:

$$\omega = (X'V^{-1}X)^{-1} \cdot \left[X'V^{-1}\gamma - \mathbf{x}\right] \tag{3.36}$$

By substuting equation (3.36) into (3.35), one obtains

$$\lambda = V^{-1}\gamma - V^{-1}X(X'V^{-1}X)^{-1}\left[X'V^{-1}\gamma - \mathbf{x}\right] \tag{3.37}$$

$$= \left[V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\right]\gamma + \left[V^{-1}X(X'V^{-1}X)^{-1}\right]\mathbf{x} \tag{3.38}$$

For convenience, let $\lambda = A \cdot \gamma + B \cdot \mathbf{x}$. The interpolant can be expressed as

$$\hat{z}(\boldsymbol{x}) = \mathbf{z}' \cdot \lambda = \mathbf{z}' \cdot B \cdot \mathbf{x} + \mathbf{z}' \cdot A \cdot \gamma = \mathbf{b}' \cdot \mathbf{x} + \mathbf{a}' \cdot \gamma \tag{3.39}$$

where

$$\mathbf{a}' = \mathbf{z}' \cdot \left[V^{-1} - V^{-1}X \cdot (X'V^{-1}X)^{-1} \cdot X'V^{-1}\right] \tag{3.40}$$

$$\mathbf{b}' = \mathbf{z}' \cdot V^{-1}X \cdot (X'V^{-1}X)^{-1} \tag{3.41}$$

Note that $X'\mathbf{a} = \mathbf{0}$, which indicate the orthogonal decomposition. (3.39) is called the dual form for kriging.

3.8 Confidence Interval of Prediction

Since the kriging objective is to minimize MSE, the optimized value is the variance estimate of kriging prediction. Assuming covariance stationarity, the variance estimate $s^2(x)$ at a prediction point x is given by

$$s^2(x) = \hat{\sigma}^2(x) = \lambda' \Sigma \lambda - 2\lambda \kappa + \kappa(0) \tag{3.42}$$

where $\kappa(0)$ is $Var\left[\varepsilon(x)\right]$. Assuming variogram only stationarity, one can infer that

$$s^2(x) = \hat{\sigma}^2(x) = 2\lambda\gamma - \lambda'V\lambda \tag{3.43}$$

The 95% confidence interval (CI) can be evaluated using standard deviation $s(x)$

$$CI \equiv \left(\hat{Z}(x) - 1.96\ s(x)\ ,\ \hat{Z}(x) + 1.96\ s(x)\right) \tag{3.44}$$

Under the assumption that $Z(x)$ is *Gaussian*, the probability that the actual surface $z(x)$ is within the interval is 0.95 : $Pr\left[z(x) \in CI\right] = 0.95$, where the $Pr[\cdot]$ is calculated from the joint distribution of $Z(x)$, $Z(x_1)$, $\cdots$, $Z(x_N)$.

3.8.1 Smoothing of Data

In most cases of real world problems, the sampled data are subject to random effect errors. In the absence of formal fitting procedures, the best way to assess the smoothing parameter or nugget c_0 seems to be *Cross Validation*. Each data point is deleted in turn and its value predicted from the rest of the data, using the fitted or specified kernel function. The predicted error are then assessed. It is tempting to form the sum of squares of these errors. Based on this idea, Cressie [16] and Wahba [72] presented the formal fitting procedures, such as *Ordinary Cross Validation* and *Generalized Cross Validation*.

CHAPTER 4.   SPLINES WITH SCATTERED DATA

We now consider a deterministic approach to multi-dimensional interpolation and smoothing problem for "scattered" data using spline functions. Formally and computationally, the analysis of deterministic splines follows the Hilbert space approach.

The deviation of an interpolant $\hat{z}(x)$ from the actual surface $z(x)$ is called the error of the interpolant. The error $\mathcal{E}(x)$ at a point x committed by interpolation is defined as $\mathcal{E}(x) = z(x) - \hat{z}(x)$. The objective function is min $|\mathcal{E}(x)|^2$ or

$$\min \ \|z(x) - \hat{z}(x)\|^2 \tag{4.1}$$

The optimal estimate $\hat{z}(x) \in \mathcal{M} = \text{span}\{\xi_i(x)\}_{i=1}^{m}$, where $\xi_i(x)$'s form a basis, becomes an interpolant when $m \geq N$ and the least square approximant when $m < N$.

4.1 Representation of an Interpolant

We first describe a way of representating an interpolant at the prediction point $x \in \mathcal{D}$. Consider a scalar-valued function $\lambda(x, x_j)$ defined on $x \in \mathcal{D}$ such that

$$\lambda(x_i, x_j) \ = \ \delta_{ij} = \begin{cases} 1 & , x_i = x_j \\ 0 & , x_i \neq x_j \end{cases} \tag{4.2}$$

where δ_{ij} $(i, j = 1, \cdots, N)$ is *Kronecker Delta*. These are what spliners call *Cardinal Functions* $\lambda_i \equiv \lambda(x, x_i)$. When x is fixed, the interpolant $\hat{z}(x)$ is just a linear combination

of the cardinal functions with different $z(x_i)$. The interpolant $\hat{z}(x)$ based on these basis functions has a *Representation* in terms of a linear combination with the observations $z_i \equiv z(x_i)$, which is often called *Lagrange Formula*

$$\hat{z}(x) = \sum_{i=1}^{N} \lambda(x, x_i) z(x_i) = \sum_{i=1}^{N} \lambda_i z_i \tag{4.3}$$

Among famous examples of the cardinal function found in the literature is the kernel of *Lagrangian Polynomial*. This type of representation leaves something to be desired, for the interpolant is not unique and in some cases the interpolant might behave erratically. The measure against this would be to secure the boundedness and uniqueness of the interpolant.

4.1.1 Optimization at Prediction Point

For our purpose, let us define $z_a = [z_0, z_1, \cdots, z_N]' \equiv [z(x), z(x_1), \cdots, z(x_N)]'$ and $\lambda_a = [-1, \lambda_1, \cdots, \lambda_N]' \equiv [-1, \lambda(x, x_1), \cdots, \lambda(x, x_N)]'$. The error $|\mathcal{E}(x)|$ can be written as a dot product of the two vectors

$$\mathcal{E}(x) = z(x) - \sum_{i=1}^{N} \lambda(x, x_i) z(x_i) \tag{4.4}$$

$$= -(z_0, z_1, \cdots, z_N) \cdot (-1, \lambda_1, \cdots, \lambda_N)' = -z_a' \cdot \lambda_a \tag{4.5}$$

Further, let us define a $(N + 1) \times (N + 1)$ matrix $K_a = T_a^2$ with T_a and K_a are:

$$T_a = \begin{bmatrix} c & \tau' \\ \tau & T \end{bmatrix} \quad \text{and} \quad K_a = \begin{bmatrix} c & \kappa' \\ \kappa & K \end{bmatrix} \tag{4.6}$$

where c is a certain constant (possibly 1). For our purposes, we shall insist that T_a be symmetric. Now it is possible to show that the magnitude of $\mathcal{E}(x)$ in equation (4.4) can

be bounded as follows:

$$|\mathcal{E}(x)| \; = \; |z'_a \cdot \lambda_a| \le |z'_a K_a^{-1} z_a|^{1/2} \cdot |\lambda'_a K_a \lambda_a|^{1/2} \tag{4.7}$$

This is the consequence of the well known *Cauchy–Schwarz* inequality. The upper bound can take several forms and can be expressed in terms of weighted norms:

$$|z'_a K_a^{-1} z_a|^{1/2} \cdot |\lambda'_a K_a \lambda_a|^{1/2} \; = \; C |\lambda'_a K_a \lambda_a|^{1/2} \; = \; C |\kappa'_a K_a^{-1} \kappa_a|^{1/2} \tag{4.8}$$

where the constant C depends on K_a and the values $z_0, z_1, \cdots, z_N$.

4.1.2 Zero Order Model ($k = 0$)

We come now to the optimization of $\sum_{i=1}^{N} \lambda(x, x_i) z(x_i)$. Ideally, we would like to minimize the absolute error $|\mathcal{E}(x)|$, but since $z(x)$ is unknown, this is not possible. Instead, we choose a suitable K and minimize the upper bound of $|\mathcal{E}(x)|$ in (4.8):

$$\min_{\lambda_a} \lambda'_a K_a \lambda_a \; = \; \min_{\lambda_a} \begin{bmatrix} -1, \lambda_1, \cdots, \lambda_N \end{bmatrix} \begin{bmatrix} c & \kappa \\ \kappa & K \end{bmatrix} \begin{bmatrix} -1 \\ \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} \tag{4.9}$$

$$= \; \min_{\lambda} \left\{ c - 2\lambda'\kappa + \lambda' K \lambda \right\} \tag{4.10}$$

where $\kappa = [\kappa(x, x_1), \cdots, \kappa(x, x_N)]'$ and $K = (\kappa_{ij})_{i,j=1}^{N}$ with $\kappa_{ij} = \kappa(x_i, x_j)$. Note that equation (4.10) is an unconstrained quadratic optimization problem. In order for this problem to solve uniquely, K must be *Positive Definite* and hence T invertible. The interpolant can be given as

$$\hat{z}(x) = z' \cdot \lambda = z' \cdot T^{-1} \cdot \tau = z' \cdot K^{-1} \cdot \kappa \tag{4.11}$$

Note that this is the standard form of Hilbert space approach in chapter 2.

4.1.3 General Order Model $(k \geq 1)$

In order to make large-scale variation independent of small-scale one, we assume that m associated with the order k be smaller than the number N of data points. In addition, the constraint $X'\lambda = \mathbf{x}$ is to be imposed on the permissible values of $\lambda_1, \cdots, \lambda_N$. Then, the optimization becomes

$$\min_{\lambda} \; \lambda'K\lambda - 2\lambda'\kappa + c \quad \text{subject to} \quad X'\lambda = \mathbf{x} \tag{4.12}$$

Note that this is a typical *Quadratic Programming* of optimization problem with equality constraints as in Luenberger [43]. The use of Lagrangian multipliers (2ω) leads to the following *First Order Necessary Condition (FONC)*

$$K\lambda + X\omega \;=\; \kappa \tag{4.13}$$

$$X'\lambda \;=\; \mathbf{x} \tag{4.14}$$

whose solution in λ and ω are uniquely obtainable if $K_A = \begin{bmatrix} K & X \\ X' & \mathbf{0} \end{bmatrix}$ is invertible. The system of equations solves uniquely, if and only if V is *Positive (Negative) definite* on the tangent subspace $\mathcal{M} = \{\lambda : X'\lambda = \mathbf{0}\}$.

4.2 Standard Form of Splines

The standard representation of interpolant in splines are based on *Reproducing Kernel Hilbert Space* theory and can be given by

$$z(\mathbf{x}) = \beta' \cdot \mathbf{x} + \sum_{j=1}^{N} \alpha_j \kappa(\mathbf{x}, \mathbf{x}_j) \tag{4.15}$$

where $z(\mathbf{x})$ is a polynomial in $\mathbf{x}$ of degree $k-1$ plus a linear combination of N copies of the kernel function centered at data sites $\{\mathbf{x}_j\}_{j=1}^{N}$. This form can be derived directly

from (4.13) and (4.14) as will be shown in chapter 6. The parameter estimates are

$$\mathbf{a} = \hat{\alpha} = \left(K^{-1} - K^{-1}X \cdot (X'K^{-1}X)^{-1} \cdot X'K^{-1} \right) \cdot \mathbf{z} \qquad (4.16)$$

$$\mathbf{b} = \hat{\beta} = (X'K^{-1}X)^{-1} \cdot X'K^{-1} \cdot \mathbf{z} \qquad (4.17)$$

Note that $X'\mathbf{a} = X'(K^{-1} - K^{-1}X \cdot (X'K^{-1}X)^{-1} \cdot X'K^{-1}) = \mathbf{0}$, where $\mathbf{a}$ represents any vector residing in the subspace orthogonal to the subspace spanned by column vector of matrix X.

4.3 Choices of Spline Functions

Usually, spliners adopt the class of *Radially Symmetric Basis Function* for kernels $\kappa(\cdot)$, mainly because of the *Rotation Invariance* of the interpolants. According to Powell and his associates [63], radially symmetric basis function models have many desirable properties and hence they provide many opportunities for application. For this reason, radial function method is the major part of the analysis of splines.

Radial symmetry is analogous to the concept of *Isotropy* of covariogram or semi-variogram kernels of kriging. Usually, the functions $\boldsymbol{x} \rightarrow \phi^0(\|\boldsymbol{x} - \boldsymbol{x}_j\|_2^2)$, $\boldsymbol{x} \in \mathcal{R}^n$, $j = 1, \cdots, N$ are referred to as radially symmetric functions "centered" at $\boldsymbol{x}_j$. If a radial function is centered at $\mathbf{0}$, then $\phi(\boldsymbol{x}) = \phi(-\boldsymbol{x})$ and hence $\phi(\boldsymbol{x}) = \phi^0(\|\boldsymbol{x}\|_2^2)$. A function $\phi(\boldsymbol{x})$ is called *Harmonic* in $\mathcal{D} \subset \mathcal{R}^n$, where the Laplacian of the function vanishes.

Example 4.0.4 *Let $\boldsymbol{x} = (x, y, z)'$: n=3, then the following condition, called Laplace equation, must be satisfied*

$$\Delta\phi(\boldsymbol{x}) = \nabla^2\phi(\boldsymbol{x}) = \frac{\partial^2}{\partial x^2}\phi(\boldsymbol{x}) + \frac{\partial^2}{\partial y^2}\phi(\boldsymbol{x}) + \frac{\partial^2}{\partial z^2}\phi(\boldsymbol{x}) = 0 \qquad (4.18)$$

for the $\phi(\boldsymbol{x})$ to be harmonic. The theory of the solutions of Laplace's equation is called Potential Theory. In the same way, the Biharmonic and Polyharmonic conditions can be expressed as follows:

$$(\text{ biharmonic }) \quad \Delta^2\phi(\boldsymbol{x}) \;=\; \nabla^4\phi(\boldsymbol{x}) = \nabla^2\left(\nabla^2\phi(\boldsymbol{x})\right) = 0 \tag{4.19}$$

$$(\text{ polyharmonic }) \quad \Delta^k\phi(\boldsymbol{x}) \;=\; \nabla^{2k}\phi(\boldsymbol{x}) = \nabla^2\left(\cdots\left(\nabla^2\phi(\boldsymbol{x})\right)\cdots\right) = 0 \tag{4.20}$$

The kernel function that satisfies these conditions renders very smooth surface in $\mathcal{D} \subset \mathcal{R}^n$ in terms of curvature.

Example 4.0.5 *Both $\phi_1(\boldsymbol{x}) = \log(x^2 + y^2)$ and $\phi_2(\boldsymbol{x}) = \sqrt{x^2 + y^2 + z^2}$ are harmonic:*

$$\nabla^2\phi_1(x,y) \;=\; \nabla^2\log(x^2 + y^2) \equiv 0 \;,\quad x^2 + y^2 \neq 0 \tag{4.21}$$

$$\nabla^2\phi_2(x,y,z) \;=\; \nabla^2\sqrt{x^2 + y^2 + z^2} \equiv 0 \;,\quad x^2 + y^2 + z^2 \neq 0 \tag{4.22}$$

Note that the Laplacians of the two functions are not defined when $x^2 + y^2 = 0$ or $x^2 + y^2 + z^2 = 0$. This fact gives the notion that the Laplacian is proportional to $\delta(\cdot)$.

In this work, we only take into account the multiquadrics and Thin Plate Spline (Surface Spline), while we do not exclude other spline models. Our selection is based on Franke's [25] test of various methods for interpolation over planar domain, in which the performance of these three methods ranks high.

4.3.1 Thin Plate Spline (Surface Spline)

The *Thin Plate Spline* (TPS) was introduced by Duchon [21] and later technically advanced by Duchon and Meinguet in a series of papers [52, 53, 54]. The interpolant is known to be a generalization of univariate *Natural Spline* or *Surface Spline*. The objective

of TPS is to find a function $g^*(x)$ that minimizes the *Roughness Measure* or *Penalty*,

$$J_k^n(g) = \int_{\mathcal{R}^n} |\nabla^k g(x)|^2 dx \qquad (4.23)$$

among all functions $g(x)$ interpolating the data $\{g(x_i) = z(x_i), i = 1, \cdots, N\}$, where k is the order of the model. In the case of $n=2$, $J_2^2(g)$ measures the *Bending Energy* of a thin plate of infinite extent, and the solution models such a plate clamped at the datapoints $\{(x_i, z(x_i))\}_{i=1}^N$. Suppose any interpolant $g(x)$ takes the form of

$$g(x) = \beta' \cdot \mathbf{x} + \sum_{j=1}^N \alpha_j \kappa(x, x_j) \qquad (4.24)$$

and for which $\nabla^k g(x)$ is square integrable, let $g(x) = g^*(x) + g_1(x)$, where $g_1(x) = g(x) - g^*(x)$ satisfies $\{g_1(x_i) = 0\}_{i=1}^N$. The roughness measure can be expressed as

$$J_k^n(g) = J_k^n(g^*) + J_k^n(g_1) + 2\int (\nabla^k g^*) \cdot (\nabla^k g_1) dx \qquad (4.25)$$

With the last cross term vanishing, the following always holds.

$$J_k^n(g) \geq J_k^n(g^*) \qquad (4.26)$$

The equality holds if and only if $g(x) = g^*(x)$. Taking integration by parts k times of the last term of (4.25) and using the facts that $g^*(x) = \beta' \cdot \mathbf{x} + \sum_j \alpha_j \kappa^*(x, x_j)$ and

$$\nabla^{2k} g^*(x) = \nabla^{2k} \sum_{j=1}^N \alpha_j \kappa^*(x, x_j) = (-1)^k (2\pi)^n \sum_{j=1}^N \alpha_j \delta(x - x_j) \qquad (4.27)$$

one can make vanish the last term of (4.25) $2\int (\nabla^k g^*) \cdot (\nabla^k g_1) dx$

$$= 2(-1)^k \int \{\nabla^{2k} g^*(x)\} g_1(x) dx = 2(2\pi)^n \int \left\{ \sum_{j=1}^N \alpha_j \delta(x - x_j) \right\} g_1(x) dx$$

$$= 2(2\pi)^n \sum_{j=1}^N \alpha_j \left\{ \int \delta(x - x_j) g_1(x) \right\} dx = 2(2\pi)^n \sum_{j=1}^N \alpha_j g_1(x_j) = 0 \qquad (4.28)$$

Using the fact that $-\Delta \kappa_\ell(x) = \kappa_{\ell-1}(x)$, the kernel of the optimal interpolating surface can be obtained by solving *Partial Differential Equations* :

$$\nabla^4 \kappa^*(x) = \nabla^2 \nabla^2 \kappa^*(x) \propto \delta(x) \qquad (4.29)$$

As discussed earlier, for $n = 2$, $\nabla^2 \log(x^2 + y^2) \propto \delta(x)$ and for $n = 3$, $\nabla^2 \sqrt{x^2 + y^2 + z^2} \propto \delta(x)$. For the order k, the optimal kernel $\kappa^*(t)$ is given as follows:

$$\kappa^*(x, x_j) = \begin{cases} t^{k-1} \log t^{1/2} & \text{if } n = 2, 4, \cdots \\ t^{k-1/2} & \text{if } n = 1, 3, \cdots \end{cases} \qquad (4.30)$$

where t, in (4.30) and henceforth is set equal to $\|x - x_j\|_2^2$. For $k = 2$, the model can be constructed as follows

$$(\text{ odd dim }) \quad g^*(x) = \beta' \cdot \mathbf{x} + \sum_{j=1}^{N} \alpha_j \|x - x_j\|_2^3 \qquad (4.31)$$

$$(\text{even dim }) \quad g^*(x) = \beta' \cdot \mathbf{x} + \sum_{j=1}^{N} \alpha_j \|x - x_j\|_2^2 \log \|x - x_j\|_2 \qquad (4.32)$$

After estimating the parameters β and α by $\mathbf{b}$ and $\mathbf{a}$, respectively, one gets, with $X'\mathbf{a} = 0$, as in section 4.2,

$$(\text{ odd dim }) \quad \hat{z}(x) = \hat{g}^*(x) = \mathbf{b}' \cdot \mathbf{x} + \sum_{j=1}^{N} a_j \|x - x_j\|_2^3 \qquad (4.33)$$

$$(\text{even dim }) \quad \hat{z}(x) = \hat{g}^*(x) = \mathbf{b}' \cdot \mathbf{x} + \sum_{j=1}^{N} a_j \|x - x_j\|_2^2 \log \|x - x_j\|_2 \qquad (4.34)$$

For more detailed justification, refer to Duchon [21] and Kent and Mardia [37].

4.3.2 Radial Basis Functions and Matern Spline

The choice of splines is not restricted to thin plate splines. In fact, many types of spline functions are being used in practice. Among these spline functions, one can name

the following, of which the first three are derivable from a *Variational Principle*, now is to be discussed in section 4.4.

1) Linear	$\kappa(x, x_j)$	$=$	$\|x - x_j\|$
2) Cubic	$\kappa(x, x_j)$	$=$	$\|x - x_j\|^3$
3) Thin Plate	$\kappa(x, x_j)$	$=$	$\|x - x_j\|^2 \log \|x - x_j\|$
4) Exponential	$\kappa(x, x_j)$	$=$	$\exp\left\{-\|x - x_j\|\right\}$
5) Gaussian	$\kappa(x, x_j)$	$=$	$\exp\left\{-\|x - x_j\|^2\right\}$
6) MQ-H	$\kappa(x, x_j)$	$=$	$1/\sqrt{\|x - x_j\|^2 + c^2}$
7) MQ-B	$\kappa(x, x_j)$	$=$	$\sqrt{\|x - x_j\|^2 + c^2}$

Note that some spliners regard the multiquadric kernels 6) and 7) as spline functions. We now consider a family of splines that includes most existing models as special cases. Consider the kernel function $C_\nu(h)$ depending on four parameters c_0, c_1, a and ν, already mentioned in chapter 3

$$C_\nu(h) = c_0 \delta(h) + c_1 \kappa_\nu^0(h) \tag{4.35}$$

where κ_ν is the function $c(\nu)(\sqrt{\nu}h/2)^\nu \mathcal{K}_\nu(\sqrt{\nu}h/2)$ where $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of order ν and $\delta(\cdot)$ is the indicator (dirac delta) function. Note that the scale parameter a works in the same way as in the *Exponential* and *Gaussian* functionss. The parameter ν is the order and controls the smoothness of the spline. This family, first suggested in kriging by Whittle [75] and later extended in the area of splines by Matern [48], can be easily generalized to higher dimensions.

The most interesting special cases are that with $\nu = \frac{1}{2}$, where $\kappa^0(h)$ will follow the exponential model, and that with $\nu = 1.5$ and a near infinity, where $\kappa^0(h)$ will approximate the thin plate spline 3) associated with a cubic smoothing (thin plate) spline.

For this reason, we consider this family as defining a larger class of splines : Matern or M spline.

4.4 Smoothing with Splines

In the case of smoothing with splines, we are assuming the model to be

$$z(x) = \mu(x) + \epsilon(x) + \theta(x) \tag{4.36}$$

where $\epsilon(x)$ and $\theta(x)$ are the realizations of $\varepsilon(x)$ and $\vartheta(x)$, respectively. Let $g(x) = \mu(x) + \epsilon(x)$ be any smoothing function. Because of the presence of $\theta(x)$,

$$z(x_i) \neq g(x_i) = \mu(x_i) + \epsilon(x_i) \ , \ \text{for } i = 1, \cdots, N \tag{4.37}$$

To provide a setting to incorporate the θ into the model, a *Variational Principle* is invoked as in Wahba [72] and Watson [74]. A minimizing problem is defined as follows:

$$\min_g \left[\frac{\sum_{i=1}^{N} [z(x_i) - g(x_i)]^2}{N\sigma^2} + \phi \int_{\mathcal{R}^n} |\nabla^k g(x)|^2 dx \right] \tag{4.38}$$

If we let $\theta N \sigma^2 \phi = 1$, then equation (4.38) reduces to

$$\min_g \left[\theta \sum_{i=1}^{N} [z(x_i) - g(x_i)]^2 + \int_{\mathcal{R}^n} |\nabla^k g(x)|^2 dx \right] \tag{4.39}$$

The choice of θ, $0 < \theta < \infty$ determines the extent to which smoothing is allowed. Let $\mathbf{g} = [g^*(x_1), \cdots, g^*(x_N)]'$; then $\mathbf{g} = X\beta + K\alpha$. Based on the same argument in (4.28),

$$J_k^n(g^*) = \int_{\mathcal{R}^n} |\nabla^k g^*(x)|^2 dx = (2\pi)^n \sum_{i=1}^{n} \alpha_i g^*(x_i) = (2\pi)^n \alpha' \mathbf{g} \tag{4.40}$$

$$= (2\pi)^n \alpha'[X\beta + K\alpha] = (2\pi)^n \alpha' K\alpha = (2\pi)^n \mathbf{g}' A\mathbf{g} \tag{4.41}$$

Here, one uses the facts that $X'\alpha = 0$ and $\alpha' = z'A$, as in section 4.2. Using this fact, the minimization problem (4.39) can be restated as

$$\min_{\mathbf{g}} \ \theta \ |\mathbf{z} - \mathbf{g}|^2 + \mathbf{g}' \cdot A \cdot \mathbf{g} = \min_{\mathbf{g}} \ \theta \ |\mathbf{z} - X\beta - K\alpha|^2 + \alpha' \cdot K \cdot \alpha \qquad (4.42)$$

The Minimization is straightforward and yields $\mathbf{g}^* = (\theta I + A)^{-1}\mathbf{z}$. Thus,

$$\hat{g}^*(x) = \mathbf{g}' \cdot B \cdot \mathbf{x} + \mathbf{g}' \cdot A \cdot \kappa \qquad (4.43)$$

$$= \mathbf{z}' \cdot (\theta I + A)^{-1} \cdot B \cdot \mathbf{x} + \mathbf{z}' \cdot (\theta I + A)^{-1} \cdot A \cdot \kappa \qquad (4.44)$$

The formulas (4.42)-(4.44) turn out to be the same as those of kriging. To determine the smoothing parameter θ, *Ordinary* and *Generalized Cross Validation* techniques are employed as in kriging.

CHAPTER 5.   THE ROLE OF KERNELS AND CONSTRAINTS

So far, we have briefly reviewed kriging and spline methods, and that the kernels of kriging and the spline functions play equivalent roles in a spatial prediction model, even though the objective functions of the two method look different.

As is to be discussed below, the isotrophy of a semi-variogram , or radial symmetry of a spline function lends the property of *Rotation Invariance* to the interpolant, while the constrained models features *Translation Invariance (splines)* and *Annihilation (kriging)*. The radially symmetric splines or isotrophic kernels have very interesting properties that facilitate the application of spatial prediction. It is convenient and even customary to express the norm of radius vector as $t = h^2 = \|x\|_2^2$, as in chapter 4. For later use, let us define further that $t_{0j} = h_{0j}^2 = \|x - x_j\|^2$ and $t_{ij} = h_{ij}^2 = \|x_i - x_j\|^2$.

As noted earlier and also to be further discussed below, the optimization problem demands that the matrices V or K be *Positive Definite* on the *Tangent Subspace* given by $\mathcal{M} = \{u : Xu = 0\}$. The kernels that yield V or K and satisfy these conditions are called *Valid* and *Conditionally Positive Definite*. By employing conditionally positive definite kernels of order k, the boundedness and uniqueness of the interpolant can be attained at the same time. Conversely, one can say that the positive definiteness condition on matrices V or K is ralaxable by increasing the order of the model. The main objective of this chapter is to discuss how to relax the restrictions on the matrices K or V.

5.1 Effect of Constraints on Splines and Kriging

Now, we investigate the effect of *Constraints* imposed on the objective function of spline model, the proof of which is also available in Micchelli [55]. Recall that the objective of a spline model is to minimize the roughness penalty $\alpha' K \alpha$ accompanied by $X' \alpha = 0$. Since the splines are functions of distance, the matrix K is a function of t_{ij} or h_{ij}.

Lemma 5.0.1 *The conditions of splines $\sum_{i=1}^{N} a_i q_j(x_i) = 0$, for $j = 1, 2, \cdots, m$ imply that $\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \|x_i - x_j\|^{2\ell} = 0$, for $\ell = 0, 1, \cdots, k - 1$.*

Proof : By expanding the kernel, one gets $\|x_i - x_j\|^{2\ell} = \left(\|x_i\|^2 + \|x_j\|^2 - 2(x_i, x_j) \right)^{\ell}$. The objective function is a sum of multiples of polynomial terms with respect to x_i or x_j

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \sum_{u+v+w=\ell} \|x_i\|^{2u} \|x_j\|^{2v} (-2)^w (x_i, x_j)^w \qquad (5.1)$$

where u, v and w are nonnegative integers that sum to ℓ. For $u \leq v$, let us define

$$Q(x) = \sum_{j=1}^{N} a_j \sum_{u+v+w=\ell} \|x\|^{2u} \|x_j\|^{2v} (-2)^w (x, x_j)^w \qquad (5.2)$$

Note that $Q(x)$ is the summation of polynomials of degree at most $k - 1$. Therefore, $\sum_{i=1}^{N} a_i Q(x_i) = 0$. Likewise, the same thing is true of the case with $u < v$. Thus, the lemma holds.

Corollary 5.0.1 *Due to the condition of $X' \mathbf{a} = \mathbf{0}$, adding to the kernel function the term $p_k(h^2) = \sum_{i=0}^{k-1} c_i h^{2i} = \sum_{i=0}^{k-1} c_i t^i$ has no effect at all on the associated spline model.*

Example 5.0.6 *In the case of splines, $\sum_{i=1}^{N} \sum_{j=1}^{N} c_1 \|x_i - x_j\|_2^{2\ell} a_i a_j = 0$, for $\ell = 0, 1, 2, \cdots, k - 1$, on condition that $\sum_{i=1}^{N} a_i q_j(x_i) = 0$, for $j = 1, \cdots, m$.*

In the case of kriging, it suffices to investigate intrinsically stationary models, because the unbiasedness constraint $X'\lambda = \mathbf{x}$ can be used without any restriction in the case of covariance stationary models. In the intrinsically stationary cases, one can derive the *MSE* function.

$$\min MSE = \min \left[2 \sum_{j=1}^{N} \lambda_j \gamma(h_{0j}) - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(h_{ij}) \right] \tag{5.3}$$

Lemma 5.0.2 *The unbiasedness conditions $\sum_{i=1}^{N} \lambda_i q_j(\mathbf{x}_i) = q_j(\mathbf{x})$, for $j = 1, 2, \cdots, m$ or $X'\lambda = \mathbf{x}$ imply the equation*

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \|\mathbf{x}_i - \mathbf{x}_j\|^{2\ell} = 2 \sum_{j=1}^{N} \lambda_j \|\mathbf{x} - \mathbf{x}_j\|^{2\ell} , \quad \ell = 0, 1, 2, \cdots, k-1 \tag{5.4}$$

Proof : Due to the unbiasedness constraints, the *MSE* is the same whether one uses $\gamma(h)$ or $\gamma(h) + \sum_{i=0}^{k-1} c_i h^{2i}$. By expanding the functions in (5.4), one obtains

$$\|\mathbf{x}_i - \mathbf{x}_j\|^{2\ell} = \left(\|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - 2(\mathbf{x}_i, \mathbf{x}_j) \right)^{\ell} , \quad \ell = 1, 2, \cdots, k-1 \tag{5.5}$$

$$\|\mathbf{x} - \mathbf{x}_j\|^{2\ell} = \left(\|\mathbf{x}\|^2 + \|\mathbf{x}_j\|^2 - 2(\mathbf{x}, \mathbf{x}_j) \right)^{\ell} , \quad \ell = 1, 2, \cdots, k-1 \tag{5.6}$$

The left hand side of equation (5.4) is a sum of multiples of polynomial terms in $\mathbf{x}_i$ or $\mathbf{x}_j$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \sum_{u+v+w=\ell} \|\mathbf{x}_i\|^{2u} \|\mathbf{x}_j\|^{2v} (-2)^w (\mathbf{x}_i, \mathbf{x}_j)^w \tag{5.7}$$

where $u, v,$ and w are nonnegative integers that sum to ℓ. Let us assume $u \leq v$, let

$$Q(\mathbf{x}) = \sum_{j=1}^{N} \lambda_j \sum_{u+v+w=\ell} \|\mathbf{x}\|^{2u} \|\mathbf{x}_j\|^{2v} (-2)^w (\mathbf{x}, \mathbf{x}_j)^w \tag{5.8}$$

which is a polynomial of degree at most $k-1$ in $\mathbf{x}$. Due to the unbiasedness constraint,

$$\sum_{i=1}^{N} \lambda_i Q(\mathbf{x}_i) = Q(\mathbf{x}) = \sum_{j=1}^{N} \lambda_j \|\mathbf{x}_i - \mathbf{x}_j\|^{2\ell} \tag{5.9}$$

Likewise, the expression is the same when $u > v$. Hence, the following holds

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\|x_i - x_j\|^{2\ell} = 2\sum_{j=1}^{N}\lambda_j\|x - x_j\|^{2\ell} \tag{5.10}$$

Corollary 5.0.2 *Due to the unbiasedness conditions, adding to the variogram function the term $p_k(h^2) = \sum_{i=0}^{k-1} c_i h^{2i} = \sum_{i=0}^{k-1} c_i t^i$ has no effect at all on the MSE or on the associated kriging objective function.*

These phenomena are often called *Annihilation* and play a key role in constructing a general order model of kriging and splines. For more details, refer to Kitanidis [39]

5.2 Laplacian of Radially Symmetric Functions

As discussed in previous sections, there exists the link between models of different order. Now let us review a known technique of obtaining the Laplacian of radially symmetric functions. If we let $f(x)$ be a radially symmetric function centered at $x = 0$ and $g(x)$ be the Laplacian of $f(x)$, then $g(x)$ is also radially symmetric and with slight abuse of notation : $g(x) = g^0(\|x\|_2^2) = g^0(t)$, $t > 0$.

Example 5.0.7 *The Laplacian of a radially symmetric function $f(\cdot)$ is expressible as a combination of the first and second order derivatives of $f(\cdot)$, expressed as a function of t.*

Let $f(x)$ be radially symmetric, then with a slight abuse of notation, $f(x) = f^0(\|x\|_2^2) = f^0(t)$.

$$\nabla^2 f(x) = \sum_{i=1}^{n}\frac{\partial^2 f(x)}{\partial x_i^2} = \sum_{i=1}^{n}\frac{\partial^2 f^0(\|x\|_2^2)}{\partial x_i^2} = \sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(\frac{\partial f^0(t)}{\partial t}\frac{\partial t}{\partial x_i}\right) \tag{5.11}$$

$$= \sum_{i=1}^{n}\left[\frac{\partial^2 f^0(t)}{\partial x_i \partial t}\frac{\partial t}{\partial x_i} + \frac{\partial f^0(t)}{\partial t}\frac{\partial^2 t}{\partial x_i^2}\right] = \sum_{i=1}^{n}\left[\frac{\partial^2 f^0(t)}{\partial t^2}\left(\frac{\partial t}{\partial x_i}\right)^2 + \frac{\partial f^0(t)}{\partial t}\frac{\partial^2 t}{\partial x_i^2}\right] \tag{5.12}$$

$$= \frac{d^2 f^0(t)}{dt^2}\cdot 4t + \frac{df^0(t)}{dt}\cdot 2n , \quad t > 0 \tag{5.13}$$

Note that the second order derivative term of equation (5.13) is multiplied by t. There are some cases for which this feature allows $\Delta f(x)$ to be expressed simply as a constant multiple of $\frac{df^0(t)}{dt}$.

Example 5.0.8 *Using this relationship, it is easy to verify that $f_1^0(t) = \|x\|_2^{-1}$ is harmonic only for $n = 3$ and $f_2^0(t) = \log\|x\|$ is harmonic only for $n = 2$ and hence $f_3^0(t) = \|x\|^2 \log\|x\|$ is biharmonic for $n = 2$, but not for $n = 3$. In general, a harmonic function is biharmonic and biharmonic function is triharmonic, and so on.*

Example 5.0.9 *If $f^0(t) = a_0 + a_1 t^\alpha$, $\alpha \neq 0$, then $g(t) = c_1 t^{\alpha-1}$, where a_0, a_1 and c_1 are arbitrary constants. Likewise, if $f^0(t) = t^\alpha \log t, \alpha \neq 0$, then $g^0(t) = t^{\alpha-1}\log t + t^{\alpha-1} = t^{\alpha-1}(\log t + 1)$.*

Let $f(x) \equiv \Delta f(x)$ and $g(x)$ be radially symmetric. For known $g(x)$, $f(x)$ can be obtained by solving

$$4t \cdot \frac{d^2 f^0(t)}{dt^2} + 2n \cdot \frac{df^0(t)}{dt} = g^0(t), \quad t > 0 \tag{5.14}$$

Example 5.0.10 *If one has $g^0(t) = c_0 + c_1 \log t$, then $f^0(t) = d_0 + d_1 t + d_2 t \log t$. Likewise, if one has $g^0(t) = c_0 + c_1 t^{-\alpha}$, then the indefinite integral of $g^0(t)$, providing a candidate $f^0(t)$ to be checked by (5.14), is given by*

$$d_0 + d_1 t + d_2 t^{-\alpha+1}, \quad \alpha \neq 1 \tag{5.15}$$

$$d_0 + d_1 t + d_2 \log t, \quad \alpha = 1 \tag{5.16}$$

For example, the semi-variogram of a white noise process centered at x_j is $\gamma_0(x, x_j) = \|x - x_j\|_2^{-1}$. The interpolant in this case can be expressed as $z_0(x) = \sum_{j=1}^N a_j \cdot \gamma_0(x, x_j)$. By integrating $\gamma_0(x, x_j) = \|x - x_j\|_2^{-1} \equiv t^{-1/2}$, one obtains a semi-variogram

for Brownian motion with corresponding interpolant given by $z_1(x) = \beta_0 + \sum_{j=1}^{N} a_j \cdot \gamma_1(x, x_j)$.

5.3 Validity of Kernels

Our next task is to investigate the admissibility (i.e. everywhere– invertiblity) of the kernels of the model. A thorough study of validity checking is available in the literature as in Matheron [49], Christakos [11], Micchelli [55], Powell [61] and Sun [70]. Their approaches are based in part on Fourier transforms, so complex variables are involved. On the other hand, Christakos [11] dispenses with complex variables by introducing a *Generalized Function* concept to prove the validity of certain kernels. As discussed in Micchelli [55], Powell [61] and Sun [70], validity checking in real domain also is possible for monotonic kernels. The relationship between non-negative definite and monotonic functions plays a key role in checking the admissibility of wider sense stationary semi-variograms or spline functions.

5.3.1 Positive Definite Functions

A set $\mathcal{T}$ of elements, or points $x, x', x_1, x_2, \cdots$, is said to be a metric space if it is provided with a distance function on $d(x, x')$ with the following

1. $d(x, x') = d(x', x)$

2. $d(x, x') = 0$ if and only if $x = x'$

3. $d(x, x'') \geq d(x, x') + d(x', x'')$

A real or complex valued function $\phi(d(x_i, x_j))$, a metric transform of pairs of points of $\mathcal{T}$ is called positive definite (Hermitian), if it enjoys the following two properties

1. Hermitian symmetry : $\phi(\boldsymbol{x}_i, \boldsymbol{x}_j) = \overline{\phi}(\boldsymbol{x}_j, \boldsymbol{x}_i)$, $i, j = 1, \cdots, N$

2. For any N points $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ of $\mathcal{T}$: $\sum_{i=1}^{N} \sum_{j=1}^{N} \phi(\boldsymbol{x}_i, \boldsymbol{x}_j) \xi_i \overline{\xi}_j \geq 0$

where $\overline{\phi}$ and $\overline{\xi}_j$ are complex conjugates of ϕ and ξ_j, respectively.

Now, let us assume that $\mathcal{T} \equiv \mathcal{R}^n$ is a linear vector space with the metric or norm $d(\boldsymbol{x}, \boldsymbol{x}') = \|\boldsymbol{x} - \boldsymbol{x}'\|$. If the function ϕ is a function of the length $\|\boldsymbol{x} - \boldsymbol{x}'\|$ of the vector only, then it is necessarily real, because of the Hermitian symmetry, and will be denoted by $g(\cdot)$.

A real valued function $G(\cdot)$, a metric transform, defined in $\mathcal{T}$, is said to be positive definite over R^N, if for any N points $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ of $\mathcal{T}$,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} G(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) a_i a_j \geq 0 \tag{5.17}$$

for arbitrary real $\{a_i\}_{i=1}^{N}$ and any N points $\{\boldsymbol{x}_i\}_{i=1}^{N}$, $N = 2, 3, \cdots$ of $\mathcal{T}$. Usually, the theory goes with saying that if a function is positive definite over $\mathcal{H}$, then it is also positive definite over $\mathcal{R}^N$, while the converse is not true. Refer to Shoenberg [66, 67].

Bochner [8] establishes the identity of the class of positive definite functions $g(\cdot)$ with the class of characteristic functions of distribution functions in $\mathcal{R}^N$.

$$g(x) = \int_{\mathcal{R}^N} e^{-ixu} d\psi(u) , \quad -\infty < x < \infty \tag{5.18}$$

where $\psi(u)$ is a non-negative and bounded measure on $[0, \infty)$. On the other hand, Shoenberg [66, 67] showed that the class of positive definite functions $g(\cdot)$ is expressible as

$$g(h) = \int_0^\infty e^{-h^2 u^2} d\alpha(u) , \quad 0 < h < \infty \tag{5.19}$$

where $\alpha(u)$ is a non-negative and bounded measure on $[0, \infty)$. The positive definite functions are closely related to the complete monotonicity of a function. A real function

$f(h)$ is said to be completely monotonic for $h \geq 0$, if

$$(-1)^k f^{(k)}(h) \geq 0 \quad \text{for} \ \ 0 < h < \infty \ \ \text{and} \ \ k = 0, 1, 2, \cdots \tag{5.20}$$

where $f(0) = f(0^+)$, which expresses the continuity of $f(h)$ at the origin. A fundamental theorem of Shoenberg [66] states the identity of this class of completely monotonic functions with the class of functions representable as a Laplace-Stieltjes integral

$$f(h) = \int_0^\infty e^{-hu} d\beta(u) \ , \quad 0 \leq h < \infty \tag{5.21}$$

where $d\beta(u)$ is a non-negative and bounded measure on $[0, \infty)$. If $\phi(h)$ is positive definite in Hilbert space $(\mathcal{H})$ of real functions, then one may write, in view of equations (5.19) and (5.21),

$$\phi(h) = \int_0^\infty e^{-h^2 u^2} d\alpha(u) = \int_0^\infty e^{-h^2 u} d\alpha(\sqrt{u}) = \int_0^\infty e^{-h^2 u} d\beta(u) = \phi(h^2) \tag{5.22}$$

This relationship states that a function $f(h)$ is completely monotonic for $h \geq 0$, if and only if $f(h^2)$ is positive definite in $\mathcal{H}$, as shown in Shoenberg [66].

Recall that for convenience, we defined $t = \|x - x_j\|_2^2$ and $h = \|x - x_j\|_2 : t = h^2$.

Example 5.0.11 *Consider a white noise semi-variogram* $\gamma(t) = t^{-\alpha}$, $\alpha > 0$. *It is easy to see that the function is completely monotonic. Since* $\gamma(t) = \int_0^\infty e^{-tu} d\beta(u)$ *by (5.21)*

$$\sum_{i=1}^N \sum_{j=1}^N \gamma(x_i, x_j) a_i a_j = \int_0^\infty \sum_{i=1}^N \sum_{j=1}^N \left[e^{-u\|x_i - x_j\|^2} \right] a_i a_j d\beta(u) \geq 0 \tag{5.23}$$

where $\beta(u)$ *can be identified as in Powell [61]. In general, the white noise kernels (of which* $\gamma(\cdot)$ *is an example) are positive definite.*

5.3.2 Validity Checking of Splines and Intrinsic Kriging

In the case of kriging, wide-sense stationary semi-variograms are said to be valid if they are positive definite. Thus, if one can prove the positive definiteness of a certain kernel, then that kernel is valid.

In the case of splines or intrinsic kriging, the operative property of a semi-variogram $\gamma(t)$ (or a spline function $\kappa(t)$) becomes conditional positive definiteness. In other words, one needs check whether a given kernel is positive definite under the relevant linear conditions

$$\sum_{i=1}^{N} a_i q_j(x_i) = 0 , \quad \text{for} \quad j = 1, \cdots, m \tag{5.24}$$

The following procedure is available for intrinsically stationary processes. For such processes, the Taylor series allows writing, as in Micchelli [55] and Powell [61],

$$\gamma(t) = \sum_{\ell=0}^{k-1} \frac{\gamma^{(\ell)}(0)}{\ell!} t^\ell + \frac{1}{k!} \int_{0+}^{t} (t-\theta)^{k-1} \gamma^{(k)}(\theta) d\theta, \quad t \geq 0 \tag{5.25}$$

where $\gamma(\cdot)$ is the *semi-variogram* of *white noise* process and hence *completely monotonic*, so that $\gamma^{(k)}(\theta) = \int_0^\infty e^{-u\theta} d\psi(u)$, $\theta > 0$. Hence,

$$\gamma(t) = \sum_{\ell=0}^{k-1} \frac{\gamma^{(\ell)}(0)}{\ell!} t^\ell + \frac{1}{(k-1)!} \int_0^\infty \int_{0+}^{t} (t-\theta)^{k-1} e^{-u\theta} d\theta d\psi(u) \tag{5.26}$$

$$= \sum_{\ell=0}^{k-1} \frac{\gamma^{(\ell)}(0)}{\ell!} t^\ell + (-1)^{k-1} \int_0^\infty \left[\sum_{\ell=0}^{k-1} \frac{(-ut)^\ell}{\ell!} - e^{-ut} \right] u^{-k} d\psi(u) \tag{5.27}$$

where $\psi(u)$ is a non-negative measure on $[0, \infty)$ such that

$$\int_0^\infty u^{-k} d\psi(u) < \infty \quad \text{and} \quad \int_0^\infty d\psi(u) > 0 \tag{5.28}$$

It follows in turn that

$$\sum_{i,j=1}^{N} \gamma(t_{ij}) a_i a_j = \sum_{i,j=1}^{N} \gamma(x_i, x_j) a_i a_j = \sum_{i,j=1}^{N} \left\{ \sum_{\ell=0}^{k-1} \frac{\gamma^{(\ell)}(0)}{\ell!} t_{ij}^\ell \right.$$

$$+ \quad (-1)^{k-1} \int_0^\infty \left[\sum_{\ell=0}^{k-1} \frac{(-ut_{ij})^\ell}{\ell!} - e^{-ut_{ij}} \right] u^{-k} d\psi(u) \Bigg\} a_i a_j \quad (5.29)$$

Now it has been shown in section 5.1 that

$$\sum_{i,j=1}^{N} \left\{ \sum_{\ell=0}^{k-1} \frac{\gamma^{(\ell)}(0)}{\ell!} t_{ij}^\ell \right\} a_i a_j = \sum_{i,j=1}^{N} \left\{ \sum_{\ell=0}^{k-1} \frac{\gamma^{(\ell)}(0)}{\ell!} \|x_i - x_j\|^{2\ell} \right\} a_i a_j = 0 \quad (5.30)$$

and

$$\sum_{i,j=1}^{N} \left\{ \sum_{\ell=0}^{k-1} \frac{(-ut_{ij})^\ell}{\ell!} \right\} a_i a_j = \sum_{i,j=1}^{N} \left\{ \sum_{\ell=0}^{k-1} \frac{(-u)^\ell \|x_i - x_j\|^{2\ell}}{\ell!} \right\} a_i a_j = 0 \quad (5.31)$$

under conditions (5.24). If the second term of (5.29) is *positive (negative) definite*, then the function $\gamma(\cdot)$ is positive (negative) definite on the tangent subspace $\mathcal{M} = \{\alpha | X' \alpha = 0\}$; i.e., *conditionally positive (negative) definite of order k.*

Example 5.0.12 *Let us suppose that the kernel $\gamma(t) = t^{1/2}$ is given. Then one can infer from Taylor series expansion that $\gamma(t) = \gamma(0) + \int_{0+}^{t} \gamma^{(1)}(\theta) d\theta$ where $\gamma^{(1)}(\theta)$ is a white noise kernel and hence completely monotonic, and*

$$\gamma(t) = \gamma(0) + \int_{0+}^{t} \int_0^\infty e^{-u\theta} d\theta d\psi(u) \quad (5.32)$$

$$= \gamma(0) + \int_0^\infty (1 - e^{-ut}) u^{-1} d\psi(u) \quad (5.33)$$

$$= \gamma(0) + \int_0^\infty u^{-1} d\psi(u) - \int_0^\infty e^{-ut} u^{-1} d\psi(u) \quad (5.34)$$

The first and second terms are just constants. Also the third term is negative definite in view of the general argument just given. Thus, it can be seen that $\sum_{i=1}^{N} \sum_{j=1}^{N} \gamma(\|x_i - x_j\|^2) a_i a_j \leq 0$ on condition that $\sum_{j=1}^{N} a_j = 0$; i.e., conditionally negative definite of order $k = 1$. In the same manner, it can be shown that the TPS kernel is conditionally positive definite of order $k = 2$.

5.4 Decomposition of Distance Matrix

Based on the arguments in the previous sections, if $\gamma(\cdot)$ is a valid kernel, the matrix V can be decomposed into several matrices based on a Taylor-like expansion:

$$V = V_0 + V_1 + \cdots + V_{k-1} + V_k \ , \quad k \geq 0 \tag{5.35}$$

It should be noted that $\mathbf{u}'V\mathbf{u} = \mathbf{u}'(V_0 + V_1 + \cdots + V_{k-1} + V_k)\mathbf{u} = \mathbf{u}'V_k\mathbf{u}$, on condition that $X'\mathbf{u} = \mathbf{0}$, since $V_0, V_1, \cdots, V_{k-1}$ is annihilated by the condition. The matrix V_k is positive (or negative) definite over the whole space, hence V must be positive (or negative) definite on the subspace $\mathcal{M} = \{\mathbf{u}|X'\mathbf{u} = \mathbf{0}\}$. Note that $\mathbf{u}$ is a dummy variable and hence the above fact holds for $\mathbf{u} = \boldsymbol{\lambda}$ or $\mathbf{u} = \boldsymbol{\alpha}$. As will be discussed in some detail in chapter 6, if the Vandermondian (distance) matrix is positive definite on $\mathcal{M}$ then a unique solution is obtainable. Consequently, one can say that if a kernel is valid, then there exists a unique solution. If the predicted value at a prediction point $\boldsymbol{x}$ is unique, then the entire interpolant is uniquely determined.

5.5 Relaxation of Restrictions on Kernels

To discuss the idea of relaxation, consider the following kernel function

$$\gamma(\boldsymbol{x}, \boldsymbol{x}_j) \ = \ \|\boldsymbol{x} - \boldsymbol{x}_j\|_2^\eta \ , \quad -\infty < \eta < \infty \tag{5.36}$$

By applying section 5.3, it can be shown that, the kernel is valid, if $\eta < 2k - 1$, $k \geq 0$, where k is the order of the model, in other words, if $k \geq \lceil \frac{\eta+1}{2} \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Here, η can be an arbitrary real number. For example, one can infer that

$$\gamma(\boldsymbol{x}, \boldsymbol{x}_j) = \|\boldsymbol{x} - \boldsymbol{x}_j\|^{1.3} \tag{5.37}$$

can be used for a model of order $k = 2$ or higher. Indeed, it can be shown that it is not valid for $k = 1$. In the case of TPS in 2-D,

$$\gamma(x, x_j) = \|x - x_j\|^{2k-2} \log \|x - x_j\| \tag{5.38}$$

is valid for $k \geq 2$, In general, the higher the order of a model, the more freedom one has in choosing the kernel functions. In other words, restrictions on valid kernels can be relaxed by increasing the order k of the model.

CHAPTER 6. COMMON ASPECTS OF SPLINES AND KRIGING

As discussed in previous chapters, the actual surface $z(x) = \mu(x) + \epsilon(x)$ in arbitrary dimension n is the realization of random process $Z(x) = \mu(x) + \varepsilon(x), x \in \mathcal{D} \subset \mathcal{R}^n$. Thus, $\epsilon(x)$ is a realization of $\varepsilon(x)$. Behind the notion that kriging and splines are essentially the same methodology, one is backed by the definition of

$$(\epsilon(s), \epsilon(t)) = \kappa(s, t) = E[\varepsilon(s)\varepsilon(t)] = (\varepsilon(s), \varepsilon(t)) \qquad s, t \in \mathcal{D} \subset \mathcal{R}^n \qquad (6.1)$$

where $(\cdot, \cdot)$ is the inner product, and $\kappa(\cdot, \cdot)$ the *(generalized) covariance kernel* or *reproducing kernel*. From this fact, it is not difficult to see that the objective functions of splines and kriging of general order k are equivalent to each other

$$\min \ \|z(x) - \hat{z}(x)\|^2 \ \equiv \ \min \ E[Z(x) - \hat{Z}(x)]^2 \qquad x \in \mathcal{D} \subset \mathcal{R}^n \qquad (6.2)$$

while $X'\lambda = \mathrm{x}$, provided that the order k of the model and the basis functions are equivalent. If we impose the rotation invariance on kernel $\kappa(s, t)$, $\kappa(s, t) = \kappa(\|s - t\|_2)$.

The interpolation $\sum_{i=1}^{N} \lambda_i z_i$ is best with respect to Ka, if $\lambda = [\lambda_1, \cdots, \lambda_N]'$ minimizes $\|(-1, \lambda')\|_K$, while $X'\lambda = \mathrm{x}$. Seen from kriging side, there exists two types of modeling: covariogram and variogram modeling. In covariogram modeling case, the objective function of kriging can be expressed as

$$\min \ \lambda'\Sigma\lambda - 2\lambda'\gamma + \kappa(0) \text{ subject to } X'\lambda = \mathrm{x} \qquad (6.3)$$

In the case of variogram modeling, (6.3) reduces to

$$\min \; 2\lambda'\gamma - \lambda'V\lambda \quad \text{subject to} \quad X'\lambda = x \tag{6.4}$$

The relation between (generalized) covariance $\kappa(x)$ and the semi-variogram $\gamma(x)$ can be expressed as $\kappa(x) = -\gamma(x) + \kappa(0)$. In view of the result in chapter 4, the splines correspond to (generalized) covariance kernels. The use of Lagrangian multipliers (ω) leads to the *First Order Necessary Condition (FONC)*: from (6.3),

$$\Sigma\lambda + X\omega = \kappa \tag{6.5}$$

$$X'\lambda = x \tag{6.6}$$

Similarly, one obtains similar *FONC* from (6.4),

$$V\lambda + X\omega = \gamma \tag{6.7}$$

$$X'\lambda = x \tag{6.8}$$

whose solution in λ and ω are uniquely determined, if $\Sigma_A = \begin{bmatrix} \Sigma & X \\ X' & 0 \end{bmatrix}$ is invertible.

6.1 Invariance Properties

As discussed in detail in chapter 5, the polynomial terms added to the kernel function has no effect at all on the prediction due to the model constraint, which property is often called *Anninilation*. For simplicity, consider the simple case of $\kappa'(\cdot) = d_2\kappa(\cdot) + d_1$, $d_2 \neq 0$. The use of $\kappa'(\cdot)$ instead of $\kappa(\cdot)$ makes no difference to the interpolant. The ineffectiveness of d_1 is referred to as *Shift Invariance*. The fact that nonzero d_2 always yields the same prediction leads to *Dilation Invariance*. These two invariance properties are termed *Translation Invariance*.

In some cases, the change of scale $\kappa^0(c\|x - x_j\|)$ for $c > 0$ ends up with the form of $\kappa'(\cdot)$, in which case the kernel $\kappa(\cdot)$ is referred to as being *Self-Similar*, since $\kappa^0(c\|x - x_j\|$ and $\kappa^0(\|x - x_j\|)$ yield the same prediction. A formal definition of Self-Similarity can be given as follows: A random field $\{Z(x) : x \in \mathcal{R}^n\}$ is called *self-similar* of index $\alpha \in \mathcal{R}$ if for each $c > 0$, $\{c^{-\alpha}Z(cx)\}$ has the same *distribution* as $\{Z(t)\}$. [Kent and Mardia 1994, Taqqu 1988]

Example 6.0.13 *For a scaling factor $c > 0$, suppose that $\kappa(x, x_j) = \kappa^0(\|x - x_j\|) = \|x - x_j\|^2 \log \|x - x_j\|$*

$$\kappa^0(c\|x - x_j\|) = c\|x - x_j\| \log c\|x - x_j\| \tag{6.9}$$

$$= d_0 + d_1\|x - x_j\|^2 + d_2\kappa^0(\|x - x_j\|) , \quad d_2 \neq 0 \tag{6.10}$$

From translation invariance and annihilation, one knows that

$$\kappa^0(c\|x - x_j\|) \sim \kappa^0(\|x - x_j\|) \tag{6.11}$$

which means $\kappa^0(c\|x - x_j\|)$ and $\kappa^0(\|x - x_j\|)$ yield the same prediction. The kernels $\kappa(\cdot)$ of this property is usually called self-similar. Basically, the family of functions belonging to the class of Matern spline or Whittle's Gaussian family is characterized by this property. MQ kernels weakly satisfy this property, because of tension parameter c^2.

6.2 Unisolvence of the Interpolant

In order to verify that the solution with respect to (λ', ω') of the above problems are globally optimal, we had better examine the second order necessary condition $(SONC)$. For more details and second order sufficiency condition $(SOSC)$, see Luenberger [43]. The theory goes with saying that if the matrix K of quadratic optimization is positive definite

on the *Tangent Subspace* formed at *Regular Point* of the constraints, then the solution is globally optimal and thus unique. The key point is that K need not be positive definite over the whole space represented by $\boldsymbol{\lambda}$, but just for $\boldsymbol{\lambda}$ satisfying $X'\boldsymbol{\lambda} = 0$, the concept being what they call *Conditional Positive Definiteness of Order k*.

Proposition 6.0.1 (Luenberger [43]) *Let K and X be $N \times N$ and $N \times m$ matrices, respectively. Suppose that X has rank m and that K is positive definite on the tangent subspace $\mathcal{M} = \{\mathbf{u} : X'\mathbf{u} = 0\}$. Then the matrix* $\begin{bmatrix} K & X \\ X' & 0 \end{bmatrix}$ *is nonsingular.*

Proof : Suppose $(\mathbf{u}, \mathbf{v}) \in \mathcal{R}^{N+m}$ is such that

$$K\mathbf{u} + X\mathbf{v} = 0 \tag{6.12}$$

$$X'\mathbf{u} = 0 \tag{6.13}$$

Multiplication of the first equation by $\mathbf{u}'$ yields $\mathbf{u}'K\mathbf{u} + \mathbf{u}'X\mathbf{v} = 0$ and substitute of $K\mathbf{u} = 0$ yields $\mathbf{u}'K\mathbf{u} = 0$. However, clearly $\mathbf{u} \in \mathcal{M}$, and thus the hypothesis on K together with $\mathbf{u}'K\mathbf{u} = 0$ implies that $\mathbf{u} = 0$. It then follows form the first equation that $X\mathbf{v} = 0$. The full-rank condition on X then implies that $\mathbf{v} = 0$. Thus the only solution to (6.12)-(6.13) is $\mathbf{u} = 0$, $\mathbf{v} = 0$. *Q.E.D.*

Note that K need not be always positive definite and hence $\kappa(\cdot)$ need not be a positive definite function, since the condition can be relaxed by increasing the order of the model as discussed in section 5 in connection with relaxation of kernels.

6.3 Estimation of Interpolant

Due to the invariance properties, it is easy to see that $K \sim \Sigma \sim V$, if one use the same type of kernels. The interpolant estimated from (6.5)-(6.6) can be rearranged as

$$\hat{z}(x) = [z', 0'] \begin{bmatrix} \lambda \\ \omega \end{bmatrix} = [z', 0'] \begin{bmatrix} \Sigma & X \\ X' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \kappa \\ x \end{bmatrix} \tag{6.14}$$

$$= z_A \cdot \Sigma_A^{-1} \cdot \kappa_A = z_A \cdot \lambda_A \tag{6.15}$$

where $\lambda_A = [\lambda', \omega']'$, $\kappa_A = [\kappa', x']'$, and $z_A = [z', 0']'$. This is also reminiscent of the standard form of Hilbert space approach.

6.4 Parametric Expression of Models

An alternative way of obtaining λ, ω is based on the assumption that the inverse of V exists, which is known to be true in most cases. From equation (6.7),

$$\lambda = K^{-1}(\kappa - X\omega) = K^{-1}\kappa - K^{-1}X\omega \tag{6.16}$$

Due to the unbiasedness condition (6.8), it follows that $X'\lambda = X'V^{-1}(\kappa - X\omega) = x$. Thus, the estimates of λ and ω can be obtained as follows:

$$\omega = (X'K^{-1}X)^{-1} \cdot [X'K^{-1}\kappa - x] \tag{6.17}$$

By substuting equation (6.17) into (6.16), one obtains

$$\lambda = K^{-1}\kappa - K^{-1}X(X'K^{-1}X)^{-1}[X'K^{-1}\kappa - x] \tag{6.18}$$

$$= [K^{-1} - K^{-1}X(X'K^{-1}X)^{-1}X'K^{-1}]\kappa + [K^{-1}X(X'K^{-1}X)^{-1}]x \tag{6.19}$$

For convenience, let $\lambda = P_\kappa \cdot \kappa + P_x \cdot x$. The interpolant can be expressed as

$$\hat{z}(x) = z' \cdot \lambda = z' \cdot B \cdot x + z' \cdot A \cdot \kappa = b' \cdot x + a' \cdot \kappa \tag{6.20}$$

which can be identified as a Boolean sum of two projections A and B, where

$$a' = z' \cdot \left[K^{-1} - K^{-1}X \cdot (X'K^{-1}X)^{-1} \cdot X'K^{-1} \right] \tag{6.21}$$

$$b' = z' \cdot K^{-1}X \cdot (X'K^{-1}X)^{-1} \tag{6.22}$$

Note that $X'a = 0$, which indicate the orthogonal decomposition. This form is standard for splines and dual for kriging.

The small-scale variation is the key part of the interpolation model and the major source of difference among interpolation methods such as splines, kriging and possibly other types of techniques. When large and small-scale is considered, the model becomes:

$$z(x) = \sum_{j=1}^{m} \beta_j q_j(x) + \epsilon(x) = \beta' \cdot x + \epsilon(x) \tag{6.23}$$

with estimated interpolant: $\hat{z}(x) = b' \cdot x + e(x)$, where $b \equiv \hat{\beta}$ and $e(x) = \hat{\epsilon}(x)$. The structure described below generalizes and unifies the above viewpoints. Given z, the interpolant $\hat{z}(x)$ must satisfy

$$\hat{z}(x_i) = \sum_{j=1}^{m} b_j q_j(x_i) + e(x_i) = \sum_{j=1}^{m} \beta_j q_j(x_i) + \epsilon(x_i) = z(x_i) \tag{6.24}$$

for $i = 1, 2, \cdots, N$. Choosing an interpolant $\hat{z}(x) = \sum_{j=1}^{m} b_j q_j(x) + e(x)$ amounts to choosing $\{b_j\}$ and $\{e(x_i)\}$. This is the problem common to krigers and spliners and is in effect approached by both schools in the following way : consider a symmetric real matrix K, which can be identified as an operator of a *Positive Transformation*, and define α by $\epsilon = I\epsilon = K\alpha$. Then we have:

$$z = X\beta + \epsilon = X\beta + I\epsilon = X\beta + K\alpha \tag{6.25}$$

Note that K is positive definite on condition that $X'\alpha = 0$. An interpolant can be obtained by specifying $\lambda(\cdot)$ such that $I \equiv (\lambda_{ij})_{i,j=1}^N$ with $\lambda_{ij} = \lambda(x_i - x_j)$. Another way of obtaining an interpolant is to specify a function $\kappa(\cdot)$ such that $K \equiv (\kappa_{ij})_{i,j=1}^N$ with $\kappa_{ij} \equiv \kappa(\|x_i - x_j\|)$ is *Positive Definite* on the tangent subspace $\mathcal{M} = \{u : X'u = 0\}$. Note that the system of equations in (6.25) is equivalent to:

$$\hat{z}(x_i) = z(x_i) = \sum_{j=1}^m \beta_j q_j(x_i) + \sum_{j=1}^N \epsilon(x_j)\lambda(x_i, x_j) \tag{6.26}$$

$$= \sum_{j=1}^m \beta_j q_j(x_i) + \sum_{j=1}^N \alpha(x_j)\kappa(x_i, x_j) \tag{6.27}$$

Relations (6.26) and (6.27) respectively yield the following models:

$$z(x) = \beta \cdot \mathbf{x} + \epsilon(x) = \beta \cdot \mathbf{x} + \sum_{j=1}^N \epsilon(x_j)\lambda(x, x_j) \tag{6.28}$$

$$= \beta \cdot \mathbf{x} + \sum_{j=1}^N \alpha(x_j)\kappa(x, x_j) \tag{6.29}$$

Using vector notation, interpolation models corresponding to (6.28) and (6.29) becomes

$$z(x) = \mathbf{z}' \cdot \boldsymbol{\lambda} = \beta' \cdot \mathbf{x} + \epsilon' \cdot \boldsymbol{\lambda} = \beta' \cdot \mathbf{x} + \alpha' \cdot \kappa \tag{6.30}$$

where $\alpha = K^{-1}\epsilon$ and $\kappa = K\lambda$, provided $\mathbf{x} = X'\lambda$. The λ is a vector of functions of x, since λ is a function of κ. The parameters $\alpha_1, \cdots, \alpha_N$ and $\beta_1, \cdots, \beta_N$ of polynomial $p_k \in \mathcal{P}_k$ are to be determined such that for $i = 1, \cdots, N$,

$$\hat{z}(x_i) = z(x_i) = \sum_{j=1}^m \beta_j q_j(x_i) + \sum_{j=1}^N \alpha(x_j)\kappa(x_i, x_j) \tag{6.31}$$

$$\sum_{j=1}^N \alpha(x_j)q_i(x_j) = 0 \ , \ i = 1, \cdots, m \tag{6.32}$$

Equivalently, in matrix and vector notation, equations (6.31)-(6.32) become

$$K\alpha + X\beta = z \qquad (6.33)$$

$$X'\alpha = 0 \qquad (6.34)$$

As a matter of fact, this is the *Standard Form* of splines and *Dual Form* of kriging.

6.5 The Link to Generalized Least Square

The parameters involved in kriging and spline models are ϵ, α and β. They are estimated based on N observations. In the language of vector spaces, ϵ and α are mutually *Covariant* and *Contravariant*. By the Hermitian character of the transformation K,

$$\min \epsilon' K^{-1}\epsilon = \min (z - X\beta)' K^{-1}(z - X\beta) = (z - Xb)' K^{-1}(z - Xb) \qquad (6.35)$$

and thus conforms to *General Least Squares (GLS)*. By symmetry, we can extend this notion to obtain another type of error vector α orthogonal to $\mathcal{B}_X$:

$$\alpha = K^{-1}(z - X\beta) = w - R\beta \qquad (6.36)$$

Thus, (6.35) can be rewritten as

$$\min \alpha' K\alpha = \min (w - R\beta)' K(w - R\beta) = (w - Rb)' K(w - Rb) \qquad (6.37)$$

Based on the criterion (6.35), $\hat{\beta}$ is given by $b = (X' K^{-1}X)^{-1}X' K^{-1}z$. Also, letting e and a be, respectively, the elements of the residuals $z - Xb$ and $w - Rb$, we have as well

$$e = z - Xb = K (w - Rb) \qquad (6.38)$$

$$a = w - Rb = K^{-1}(z - Xb) \qquad (6.39)$$

6.6 Dual Property

Alternatively, we can obtain an interpolant as well by estimating paramaters involved and inventing function $\kappa(\cdot)$. Note that now we are working with N data points, while we worked with $N + 1$ data points in the previous subsection. Let us get started with the fact that that our K is indeed restricted to be positive (semi)definite follows from

$$0 \leq \alpha' K \alpha \tag{6.40}$$

The existence of T^{-1} implies the existence of K^{-1}, so that K is in fact restricted to be positive definite. Indeed, the structure of K is more refined : K is conditionally positive definite of order k, a concept that can be detailed as follows.

In order to discuss conditional positive definiteness of the matrix K, it will be useful to identify the two mutually orthogonal subspaces $\mathcal{B}_X \equiv \{X\beta : \beta \in \mathcal{R}^m\}$ and $\mathcal{A}_X \equiv \{\alpha : X'\alpha = 0, \alpha \in \mathcal{R}^N\}$ of respective dimensions m and $N - m$ in an N dimensional vector spaces. In the same manner, the vector space can be partitioned into two mutually orthogonal subspaces $\mathcal{B}_R \equiv \{R\beta : \beta \in \mathcal{R}^m\}$ and $\mathcal{E}_R \equiv \{\epsilon : R'\epsilon = 0, \epsilon \in \mathcal{R}^N\}$ of respective dimensions m and $N - m$. In other words, the vector space can be expressed as a *Boolean (Direct)* sum of subspaces as

$$\mathcal{R}^N = \mathcal{A}_X \oplus \mathcal{B}_X = \mathcal{E}_R \oplus \mathcal{B}_R \tag{6.41}$$

It is interesting to observe that the conditions $R'e = X'a = 0$ always hold. Thus, the inequalities in (6.40) need not hold for all α values , but just for the values of α satisfying

$$R'\epsilon = X'\alpha = 0 \tag{6.42}$$

which actually weakens the positive definiteness condition of matrix K and K^{-1} so that the condition become less restrictive. The corresponding theory for the matrix K

involves the notion of validity of the kernel functions. The functions that satisfy the above-mentioned conditions are possible candidate kernel functions for both kriging and splines. See Huijbregts and Matheron [33], Olea [57] and Michelli [55].

All the properties discussed so far are attributable to the fact that both kriging and spline models defined in a *Finite Dimensional* vector space naturally or inevitably enjoy the dual structure, and both krigers and spliners are exploiting the *Dual Structure* of vector space. The following is the common way of estimating the parameters involved in the model. Recall that

$$\mathbf{b}' = (X'K^{-1}X)^{-1} \cdot XK^{-1} \cdot \mathbf{z} \tag{6.43}$$

$$\mathbf{a}' = K^{-1}(\mathbf{z} - X\mathbf{b}) = \left[K^{-1} - K^{-1}X \cdot (X'K^{-1}X)^{-1} \cdot X'K^{-1} \right] \cdot \mathbf{z} \tag{6.44}$$

From the fact that $K\mathbf{a} = \mathbf{z} - X\mathbf{b}$ and $X'\mathbf{a} = \mathbf{0}$, it is easy to see that the estimates in (6.44) and (6.43) can be obtained from the system of equations given in (I) or (II).

$$(\text{ I }) \quad K\boldsymbol{\alpha} + X\boldsymbol{\beta} = \mathbf{z} \qquad (\text{ II }) \quad K^{-1}\boldsymbol{\epsilon} + X\boldsymbol{\beta} = \mathbf{w} \tag{6.45}$$

$$X'\boldsymbol{\alpha} \qquad = \mathbf{0} \qquad\qquad\qquad X'\boldsymbol{\epsilon} \qquad = \mathbf{0}$$

The system of equations in (I) can be solved for arbitrary $\mathbf{z}$, if and only if for every $\boldsymbol{\alpha} \in \mathcal{A}_X \backslash \{0\}$, $\boldsymbol{\beta} \in \mathcal{B}_X$, the augmented matrix $\begin{bmatrix} K & X \\ X' & 0 \end{bmatrix}$ are nonsingular, which can be easily verified using the *SONC* provided K is positive definite on condition $X'\boldsymbol{\alpha} = \mathbf{0}$. The same is true of (II). We will discuss later that the formulation for kriging model takes the primal form (II), while splines take the dual (I) as in chapter 3 and 4, respectively.

6.7 Computational Aspects

Both in kriging and splines, positive definiteness of the matrix K on the tangent subspace is required for boundedness of the interpolant at the prediction point x. However, it is not trivial to verify the conditional positive definiteness of the function $\kappa(\cdot)$. In fact, the existing analytic methods guarantee only the conditional positive semi-definiteness of the function $\kappa(\cdot)$, as discussed in chapter 5. Besides, the problem is more serious than is thought to be. Even though the matrix K is conditionally positive definite, it could be numerically singular especially when the condition number of the matrix K is very small, where the *Condition Number* is defined as

$$\text{Condition Number} = \left| \frac{\text{the smallest eigenvalue}}{\text{the largest eigenvalue}} \right| \qquad (6.46)$$

Even this problem is also common to both kriging and splines. As a general rule, it is most efficient to use *Factorization Methods* such as *LU* decomposition that exploit the structure of the symmetric matrix, but we shall not review this here.

6.8 Smoothing Techniques

As mentioned earlier, the measurement error of observations are regarded as being independent. How to incorporate these errors into the prediction model is more like a trick. The changes are made only on the diagonal elements of the matrices.

If we define $K(\theta)$ to be the matrix K plus a diagonal matrix which has all elements θ, then the optimization problem can be identified as follows

$$\min_{\epsilon} \theta |z - X\beta - K\epsilon|^2 + \epsilon' K^{-1}\epsilon = \min_{\alpha} \theta |z - X\beta - K\alpha|^2 + \alpha' K\alpha \qquad (6.47)$$

The same result from another approach is available in Kimeldorf and Wahba [38] and

Watson [74]. Both krigers and spliners employ *Ordinary or General Cross Validation* techniques to estimate the parameter (θ).

6.9 Difference between Kriging and Splines

The major difference lies in the objectives of each group and hence the criterion for the choice of kernels. Usually, krigers try to estimate the kernels from the observational data, as in Cressie [16] and Kitanidis [39], while spliners invent splines on their own criterion, say, *Roughness Penalty*, as in chapter 4.

CHAPTER 7. MULTIQUADRIC METHOD

The multiquadric models were originally devised by Hardy [28, 29] as a global method to interpolate N observational data. The initial study of multiquadrics dates as far back as early 1970's. The origin and the development of the multiquadric method is well presented in Hardy [30] with extensive bibliographies. Many successful applications of the MQ method in the case of multidimensional data are also reported in many recent publications as in Pottmann and Eck [59], Carson and Foley [10] and Kansa [36]. Unlike other spatial prediction methods, the multiquadrics has its ground in *Potential Theory*.

7.1 Potential Theory

Multiquadric models are constructed on the basis of the afore-mentioned potential theory, a good example of which occurs in connection with gravitational forces. If a particle A of mass M is fixed at a point (x_j, y_j, z_j) and another particle B of mass m is at point (x, y, z) then A attracts (or repulses B), the gravitational force being the gradient of the scalar function

$$g(x, y, z) = \frac{\ell}{r} , \quad \ell = GMm \equiv const \tag{7.1}$$

$$r = \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \tag{7.2}$$

where G is coefficient of gravitation. This function is called the potential of the gravitational field, and it satisfies Laplace's equation. On the other hand, the function

$$f(x, y, z) = r = \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \tag{7.3}$$

represents the potential energy. These two types of functions are employed by Hardy and are called MQ harmonic (MQ-H) and MQ biharmonic (MQ-B), respectively. The functions can be expressed in terms of norms:

$$(\text{MQ-H}) \quad \kappa_h(x, x_j) = 1/\|x - x_j\|_2 \, , \; x \neq x_j \tag{7.4}$$

$$(\text{MQ-B}) \quad \kappa_b(x, x_j) = \|x - x_j\|_2 \quad , \; x \neq x_j \tag{7.5}$$

for $j = 1, 2, \cdots, N$, where $\|x - x_j\|_2$ is the distance between the two points x and x_j. Of course, the Laplacian of $\kappa_b(x - x_j)$ is supposed to be $\kappa_h(x, x_j)$;

$$\nabla^2 \kappa_b(x, x_j) = \kappa_h(x, x_j) \, , \; x \neq x_j \tag{7.6}$$

$$\nabla^2 \kappa_h(x, x_j) = 0 \qquad , \; x \neq x_j \tag{7.7}$$

Note that the original potential functions are defined only on 3 dimensional space.

7.2 Tension Parameter

As discussed in chapter 6, it is known and easy to verify that MQ-H and MQ-B kernels are harmonic and biharmonic, respectively, only for $n = 3$. Note that $\nabla^2 \kappa_h(x - x_j) \neq 0$ when $n=2$ and hence $\nabla^4 \kappa_b(x - x_j) \neq 0$ when $n = 2$. To use the models in $\mathcal{D} \subset \mathcal{R}^2$, a slight modification is needed and Hardy proposed the following by substuting c^2 for the term $(z - z_j)^2$.

$$(\text{MQ-H}) \quad \kappa_h(x, x_j) = 1/\sqrt{(x - x_j)^2 + (y - y_j)^2 + c^2} \tag{7.8}$$

$$(\text{MQ-B}) \quad \kappa_b(x, x_j) = \sqrt{(x - x_j)^2 + (y - y_j)^2 + c^2} \tag{7.9}$$

According to Hardy's recommendation, a small value of c would be enough to round off the discontinuity of the vertex at each center. In spite of the modification, the harmonic and biharmonic property is not maintained in $\mathcal{R}^2$. Some investigators such as Carson and Foley [10] have tried to find the "optimal value" of c^2 in some sense. According to their studies, the existence of an optimal c^2 is not apparent. Some other spliners as Eck [24] regard this constant as a "*Tension Parameter*" and thus a sort of *Design Parameter* in surface design.

7.3 Zero Order Model ($k = 0$)

Originally, Hardy [28, 29] utilized the form of order 0 without the large scale part of polynomial and chose $\kappa(\cdot)$ to be a harmonic or biharmonic function.

$$(\text{MQ-H}) \qquad \hat{z}_h(x) = \sum_{j=1}^{N} a_j \kappa_h(x, x_j) = \sum_{j=1}^{N} a_j \frac{1}{\sqrt{\|x - x_j\| + c^2}} \qquad (7.10)$$

$$(\text{MQ-B}) \qquad \hat{z}_b(x) = \sum_{j=1}^{N} a_j \kappa_b(x, x_j) = \sum_{j=1}^{N} a_j \sqrt{\|x - x_j\| + c^2} \qquad (7.11)$$

These models are valid in that the *Grammian* matrices T_h and T_b are invertible without any conditions so that the interpolant is always obtainable. As discussed earlier, however, the boundedness of the interpolant is not guaranteed within this framework, as discussed in chapter 4. As seen in Figure 7.1, the MQ harmonic kernel can be related to the semi-variogram of a white noise process. With positive constant c, it becomes much like the *Covariance Stationary* kernel of kriging, viz. *Simple Kriging*.

The MQ biharmonic kernel is also valid within this frame work, since it can be proven that the *Grammian* of MQ-B kernel is always invertible. The drawback of the zero order model framework is that it does not guarantee the boundedness of the interpolant.
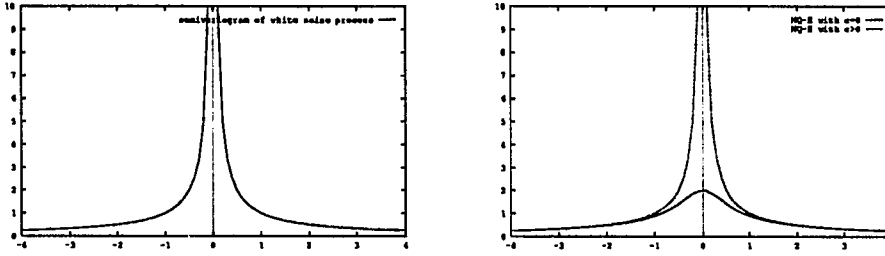
Figure 7.1: Covariogram of White Noise and MQ Harmonic Kernel

7.4 First Order Model ($k = 1$)

Later, Hardy [32] switched to what they call multiquadric biharmonic model. The model can be identified as of order 1 and falls into the following framework.

$$\text{(MQ-H)} \quad \hat{z}_h(x) = b_0 + \sum_{j=1}^{N} a_j \frac{1}{\sqrt{\|x - x_j\| + c^2}} \tag{7.12}$$

$$\text{(MQ-B)} \quad \hat{z}_b(x) = b_0 + \sum_{j=1}^{N} a_j \sqrt{\|x - x_j\| + c^2} \tag{7.13}$$

The MQ *harmonic* model of order 1 is valid in that the matrices κ_h is *Positive Definite* without any condition so that the interpolant is bounded and uniquely determined. With positive constant c, it becomes the *Covariance Stationary* kernel of kriging, viz. *Ordinary Kriging*. The MQ biharmonic model of order 1 is also valid within this frame work since the kernel is *Conditionally Positive Definite of Order 1*. See Figure 7.2. Note that the biharmonic kernel is much the same as the semi-variogram of a Brownian motion porcess, which corresponds to Matheron's 0-IRF.
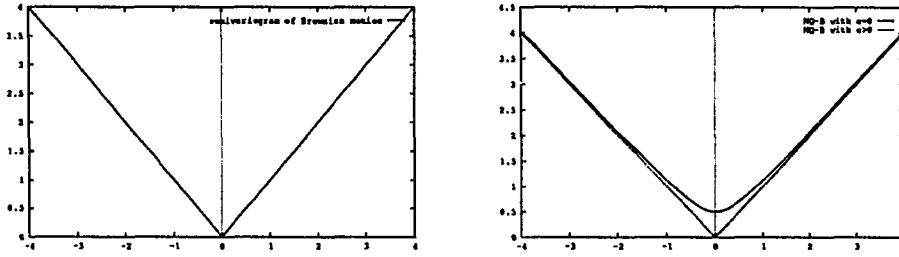
Figure 7.2:   Standard Brownian Motion and MQ-B Kernel

7.5 General Order Model ($k > 1$)

More general types of models are also available. MQ models of general order belong to what Hardy call *Osculating Mode*. In fact, the model embraces the polynomial term for large-scale variation and the kernel of Hardy's osculating mode remains the same as the order of the model changes. On the other hand, Riesz *Potential Model* provides more general framework for this type of multiquadric kernels

$$(\text{ Riesz }) \quad \kappa(\boldsymbol{x}, \boldsymbol{x}_j) \;=\; \left(\|\boldsymbol{x} - \boldsymbol{x}_j\|_2^2 + c^2 \right)^{k-\eta} \,, \; 0 < \eta < 1 \,, \; k = 0, 1, \cdots \quad (7.14)$$

Note that harmonic and biharmonic models are characterized by $k = 0$ and $k = 1$, respectively, with $\eta = 1/2$.

Example 7.0.14 *A MQ biharmonic model of order 1 takes the spline form*

$$\hat{z}(\boldsymbol{x}) \;=\; \beta_0 + \sum_{j=1}^{N} a_j \kappa(\boldsymbol{x}, \boldsymbol{x}_j) \,, \quad (7.15)$$

and also the parameters involved must satisfy the condition that $\sum_{j=1}^{N} a_j = 0$, *and*

$$z(\boldsymbol{x}_i) \;=\; \beta_0 + \sum_{j=1}^{N} a_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \,, \quad i = 1, \cdots, N \quad (7.16)$$

The MQ potential model of order k is also valid within this framework, since the kernel is *conditionally positive definite of order k*.
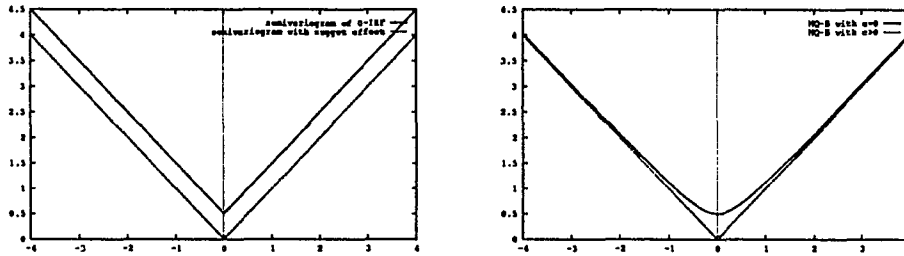
Figure 7.3: Smoothing with 0-IRF and MQ-B

7.6 Smoothing

A main difference between kriging and MQ lies in the use of the term "smoothing constant" (in the case of splines) and "nugget effect" (in the case of kriging) for essentially the same quantities.

Unlike other spatial prediction methods, in case of MQ or potential models, the objective function is not defined. For this reason, it is not clear how to incorporate measurement errors into MQ models. As a matter of fact, Hardy's multiquadric models have been used extensively for interpolation only. It would be possible, however, to incorporate the smoothing into MQ potential model just as is done in kriging, as in Figure 7.3.

CHAPTER 8. THE LINKS AMONG EXISTING MODELS

The spatial prediction models are the most general type of linear model. Accordingly, the generalized regression models or time series models are special cases of spatial prediction models. Some of typical spatial prediction methods can be understood in a unified context. Defined on a finite dimensional vector space or in a Hilbert space, both kriging and splines naturally and inevitably enjoy the same structure. The kriging predictor can be made equivalent to splines and multiquadric predictor with a particular choice of *Generalized Covariance or Semi-variogram*. To facilitate the comparison of existing prediction models, the order of the model and the type of kernel function must be considered.

8.1 The Links Among Models of Different Order

For both kriging and splines, the interpolation models of order k can be constructed by solving the *Partial Differential Equations*:

$$\Delta^k \kappa(\cdot) \equiv white\ noise \tag{8.1}$$

$$\Delta^{k-1} \kappa(\cdot) \equiv Brownian\ motion \tag{8.2}$$

If the kernels are radially symmetric, the differential equation can be easily solved and thus the model building is performed more systematically, as discussed in section 6. An analogue between *Covariance Stationary* kriging and *Variogram Stationary* kriging are

Table 8.1:   Classification by Order of Models

Stationarity	Order 0	Order 1	Order 2	Order k
Wide Sense	Simple	Ordinary	Universal	Universal
Condition on $(-1)^k \kappa(\cdot)$	Positive Definite	Positive Definite	Positive Definite	Positive Definite
Wider Sense	White Noise	0-IRF	1-IRF	$(k-1)$-IRF
Admissibility Condition on $(-1)^k \gamma(\cdot)$	Pos. Def.	Conditionally Pos. Def. of order 1	Conditionally Pos. Def. of order 2	Conditionally Pos. Def. of order k

tabulated as in Table 8.1.   As a matter of fact, such a kriger as Matheron [50] views the lowest order process as a Brownian motion, which is of order 1, whereas most spliners as Duchon [21] and Micchelli [55] equate the semi-variogram of the lowest order process as that of *Dirac Delta (or white noise)*, which is completely monotonic. In this context, MQ-H belongs to a semi-variogram of white noise model, just as MQ-B to that of a Brownian motion. In the same fashion, TPS can be identified as belonging to a triharmonic or polyharmonic model.

The functions for white noise models are featured by being completely monotonic. The white noise kernels play a key role in identifying the validity of a kernel. To ease the comparison, the following Table 8.2 might be helpful. Note that the kernel functions of lower order model can be used as the kernel of higher order models.

8.1.1 The Order of Models

Multiquadric users would relish the zero order model, where their distance matrix (or Vandermondian) is just invertible. Seen from spliner's stand point, the zero order model is not the bounded framework. Generally, the zero order model is not used by krigers',

Table 8.2: Harmonic Series of Models

Models	Order 0	Order 1	Order 2	Order k
MQ-2D	$t^{-1/2}(t^{1/2})$	$t^{1/2}$	$t^{1/2}\;[t^{3/2}]$	$t^{1/2}\;[t^{k-1/2}]$
MQ-3D	$t^{-1/2}(t^{1/2})$	$t^{1/2}$	$t^{1/2}\;[t^{3/2}]$	$t^{1/2}\;[t^{k-1/2}]$
TPS-2D	$[\delta(t)]$	$[\log t]$	$t\log t$	$t^k\log t$
TPS-3D	$[t^{-1/2}]$	$[t^{1/2}]$	$t^{3/2}$	$t^{k-1/2}$
Krg-2D	$[\delta(t)(t^{-1/2})]$	$[\log t(t^{1/2})]$	$t\log t(t^{3/2})$	$t^k\log t(t^{k-1/2})$
Krg-3D	$[t^{-1/2}]$	$t^{1/2}$	$t^{3/2}$	$t^{k-1/2}$

MQ-2D	: MQ in (x,y)	MQ-3D	: MQ in (x,y,z)
TPS-2D	: TPS in (x,y)	TPS-3D	: TPS in (x,y,z)
Krg-2D	: Kriging in (x,y)	Krg-2D	: Kriging in (x,y,z)
$[\;\cdot\;]$	: models not in use		

for the it does not have the large scale part and hence the interpolant is seriously biased in the statistical sense.

For the same reason, multiquadric users seem to prefer the MQ-B, which is order 1 and hence bounded. Multiquadric users made scarce use of the second order model. Usually, the second order $k = 2$ models are preferred by spliners, while krigers do not seem to have preference on the order of the model, is so far as the model is valid.

A question that can arise is what the use of higher order $(k > 2)$ models is. A sort of answer is available in Mardia $et.al.$ [47]. According to their study, one need to employ higher order model to incorporate into the model the information on slopes, curvatures and the higher order derivatives, which can be available when measured by such measuring devices as laser interferometer.

8.2 The Link between Least Square and Spatial Prediction

The estimation of parameters involved in spatial prediction models are essentially the same in that kriging estimates are based on maximum likelihood function and splines have its ground on generalized least square. It is famous that MLE and LS criterions are basically the same. In the sense, parameter estimates of spatial prediction are all MLE's. It is interesting and important to notice that spatial prediction method can be recognized as a natural extension of least squares or polynomial regression. For example, given a set of data $\{(y_i, x_i)\}_{i=1}^N$, the least square criterion is given by

$$\min \sum_{i=1}^N (y_i - \mu)^2 \quad \text{and} \quad \hat{\mu} = \frac{\sum_{i=1}^N y_i}{N} \tag{8.3}$$

where μ is a constant. Note that we are assuming the residuals $\{\nu_i = z_i - \mu\}_{i=1}^N$ are independent. To compare the least square method with spatial prediction, consider the first order model

$$z(x) = \beta_0 + \sum_{j=1}^N \alpha_j \kappa(x, x_j), \quad x, x_j \in \mathcal{D} \tag{8.4}$$

which is the standard form for splines and the dual form for kriging. If one adopts a white noise kernel for $\kappa(x, x_j)$, then one gets $\hat{\mu} = \hat{\beta}_0$. The white noise kernel provides a connection between the least squares and spatial prediction method. Recall that the spatial prediction model satisfies the following condition, for $i = 1, \cdots, N$,

$$z(x_i) = \beta_0 + \sum_{j=1}^N \epsilon_j \lambda(x_i, x_j) = \beta_0 + \sum_{j=1}^N \nu_j \tau(x_i, x_j) = \beta_0 + \sum_{j=1}^N \alpha_j \kappa(x_i, x_j) \tag{8.5}$$

In case that $\kappa(\cdot)$ is a white noise kernel, it is easy to see that $\epsilon_j = \nu_j = \alpha_j$ and hence $K = I$. It is interesting to note that the least square criterion provides a method for obtaining a the center of gravity in terms of physics.

8.2.1 Ordinary Least Square Method

Let's consider a general order model. The least squares criterion is equivalent to the *Polynomial Regression*. The objective function can be identified as

$$\min_{\beta} \sum_{i=1}^{N} [y(x_i) - \mu(x_i)]^2 = \min_{\beta} (y - U\beta)' \cdot (y - U\beta) = \min_{\beta} \nu' \cdot \nu \quad (8.6)$$

where $x_i \in \mathcal{D}$. Note that ν is the vector of independent residuals $\{\nu_i = y(x_i) - \mu(x_i) = y_i - \mu(\xi)\}_{i=1}^{N}$. The estimate $\hat{\beta} = (U'U)^{-1} \cdot U'y$ and $\hat{\mu}(x) = \hat{\beta}' \cdot x$.

8.2.2 Generalized Least Squares

The *Least Square* idea can be expanded further to *Generalized Least Squares* usually in connection with Covariance Stationarity. The *GLS* can be expressed as

$$\min_{\beta} \sum_{i=1}^{N} \sum_{j=1}^{N} \kappa^{ij} [z(x_i) - \mu(x_i)][z(x_j) - \mu(x_j)] \quad (8.7)$$

$$= \min_{\beta} (z - X\beta)' \cdot \Sigma^{-1} \cdot (z - X\beta) = \min_{\beta} \epsilon' \cdot \Sigma^{-1} \cdot \epsilon \quad (8.8)$$

where $[\kappa^{ij}] = \Sigma^{-1}$ and $[\kappa_{ij}] = \Sigma \equiv$ variance-covariance matrix, which is given by

$$\Sigma = [\kappa_{ij}] = \begin{bmatrix} \kappa(x_1, x_1) & \kappa(x_1, x_2) & \cdots & \kappa(x_1, x_N) \\ \kappa(x_2, x_1) & \kappa(x_2, x_2) & \cdots & \kappa(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(x_N, x_1) & \kappa(x_N, x_2) & \cdots & \kappa(x_N, x_N) \end{bmatrix} \quad (8.9)$$

In *GLS*, the positive definiteness of matrices Σ and Σ^{-1} is needed for the *Unisolvence* and *Boundedness* of the interpolant, which in turn requires the covariance kernel function to be positive definite. If we insist on the positive definiteness of the kernel, the model is

analogous to the universal kriging. If we use the white noise kernel, the model reduces to the polynomial least square model.

The *GLS* idea can be expanded further to *Constrained GLS (CGLS)* on the basis of *Wider Sense (Variogram) Stationarity* assumption. The objective function of *CGLS* can be identified as

$$\min_{\beta} \sum_{i=1}^{N} \sum_{j=1}^{N} \kappa^{ij} [z(x_i) - \mu(x_i)][z(x_j) - \mu(x_j)] \tag{8.10}$$

The restriction on the matrix K can be relaxed by adding constraints $X'\alpha = 0$ or $R'\epsilon = 0$.

$$\min_{\beta} \alpha' \cdot K \cdot \alpha = \min_{\beta} \epsilon' \cdot K^{-1} \cdot \epsilon \tag{8.11}$$

After estimating parameters as before, the prediction of interpolant can be performed by

$$\hat{f}(x) = b' \cdot x + \sum_{j=1}^{N} a_j \cdot \kappa(x, x_j) = b' \cdot x + \sum_{j=1}^{N} e_j \cdot \lambda(x, x_j) \tag{8.12}$$

The choice of $\kappa(\cdot)$ and the order k of the model determines the effectiveness of interpolant estimation, while the *Conditional Positive Definiteness of Order k* of matrices K and K^{-1} is needed for the *Unisolvence* and *Boundedness* of the problem.

The main point of this section is that any least square problem can be a special case of spatial prediction. Conversely, it can be said that any least square problem can be naturally extended to spatial prediction problems.

CHAPTER 9. SPATIAL PREDICTION OF ROUND FEATURES

In this chapter, we discuss what can be done with geometric objects which have circular features such as circles, spheres and cylinders. As discussed in the previous chapters, spatial prediction models can be unified into the *Generalized Least Squares* framework by extending the *Least Square Criterion.* When it comes to applying the spatial prediction theory to near-circular features, one may recall that the *Least Squares Method* provides a framework for circular profile approximation known as *Best Fitting* of circular features, the statistical behavior of which is well studied and illustrated in Berman [6], Berman and Culpin [7] and Coope [12]. Our primary emphasis will be on assessing departures from circularity. This will be done by fixing in an initial least squares estimate the "center of the profile, followed by detailed examination of the "radius" of the profile, applying methods of the previous sections, in addition to developing suitably periodic kernels. While it seems intuitively clear that pre-selection of the center will not subtantially affect the final estimated near-circular contour, precise analysis of this phenomenon awaits research beyond this work.

9.1 Circular Profile Estimation

According to ISO 1101 or ANSI Y14.5M, a profile is a outline of the projected shadow of an object. If the outline is "approximately" round, it is called circular. The problem

of determining the circle of *Best Fit* to a set of points of circular profile in the plane (or the obvious generalization to n-dimensions) is easily formulated as a *Nonlinear Total Least Squares* problem which may be solved using a nonlinear optimization algorithm. The distance r_i from the center (a, b) to a point $x_i = (x_i, y_i)$ can be given by

$$r_i = \sqrt{(u_i - a)^2 + (v_i - b)^2} \tag{9.1}$$

The *Ordinary Least Square Criterion* for circular feature with unknown center and radus can be expressed as

$$\min_{a,b,r} \sum_{i=1}^{N} \left[r_i - r \right]^2 = \min_{a,b,r} \sum_{i=1}^{N} \left[\sqrt{(u_i - a)^2 + (v_i - b)^2} - r \right]^2$$
$$= \sum_{i=1}^{N} \left[\sqrt{(u_i - \hat{a})^2 + (v_i - \hat{b})^2} - r \right]^2 \tag{9.2}$$

For fixed center (a, b), say $(\hat{a}, \hat{b})$, and unknown radius r, the criterion becomes

$$\min_{r} \sum_{i=1}^{N} [r_i - r]^2 \tag{9.3}$$

The estimates $\hat{a}, \hat{b}$ and $\hat{r}$ are *Maximum Likelihood Estimates (MLE)* under the assumption that the sample distances of (u_i, v_i) from the true center are independent and approximately normally distributed about the true radius.

9.1.1 Reference Circles

Just as the cartesian doordinate system is used for a reference in planar surface estimation, a circular reference with a center is needed for a set of near-circular data points imbedded in the plane. The spatial prediction models for circular features depends on the choice of setting or circular reference; also the use of kernel type is closely related

to the domain of interest $\mathcal{D}$; in particular, peroidicity is a nature attribute of kernels to be used for spatial analysis of circular features.

There are many ways to designate a reference circle and center, such as the best fit circle, the minimum radial separation circle, the largest inscribed circle, the smallest circumscribed circle and so on. Among these, the best fit circle would be the most compatible one with spatial prediction methodology.
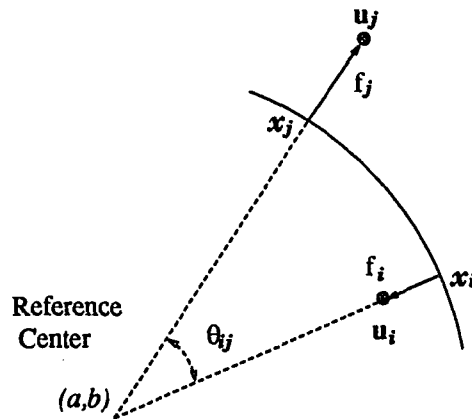
Figure 9.1: Observations of Circular Features

In general, the center of the reference circle is of particular importance. Indeed distances from the reference circles are regarded as observations.

For later use, let U^m denote the unit sphere in $\mathcal{R}^{m+1}$ so that U^1 can define a unit circle in $\mathcal{R}^2$ and let B^m be the best fit sphere in $\mathcal{R}^{m+1}$ so that B^1 denotes the best fit circle in $\mathcal{R}^2$. Throughout this work, the center of a reference circle lying in the plane, is assumed, without loss of generality, to be at the origin.

9.1.2 Domain of Interest

The most popular way of dealing with the circular problem would be to consider it in a univariate setting tied to a single angle θ. In the case that the observation points $x_i, x_j \in U^1$, since $(\mathbf{a}, \mathbf{b}) = \|\mathbf{a}\|\|\mathbf{b}\| \cos \theta$ for arbitrary vector $\mathbf{a}$ and $\mathbf{b}$, the intervening angle $\theta_{ij} \leq \pi$ between x_i and x_j can be given by $\theta_{ij} = \arccos(x_i, x_j)$, where $(\cdot)$ denotes the inner product. In case that x_i and x_j are not necessarily on U^1 but on an arbitrary circle,

$$\theta_{ij} = \arccos \left(\frac{x_i}{\|x_i\|}, \frac{x_j}{\|x_j\|} \right) \tag{9.4}$$

An alternative way of identifying the domain would be to resort to chords, rather than angles as in Light and Cheney [42]. The metric involved is the Euclidean distance $\|x_i - x_j\|_2$, which is the length of the chord between the two points x_i and x_j, given by

$$\|x_i - x_j\|_2 = \sqrt{2\tilde{r} - 2\tilde{r} \cos \left\{ \arccos(\frac{x_i}{\|x_i\|}, \frac{x_j}{\|x_j\|}) \right\}} = \sqrt{2\tilde{r} - 2\tilde{r} \cos \theta_{ij}} \tag{9.5}$$

Figure 9.2 illustrates the metric of each system. A closer look reveals that there exists
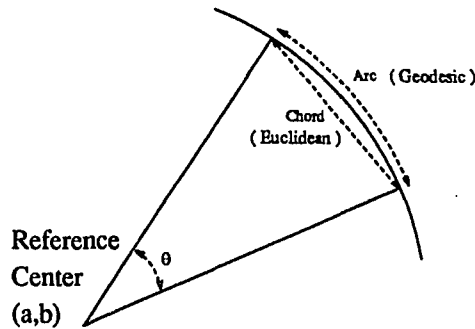
Figure 9.2: Geodesic vs. Euclidean Distance

an algebraic link between the two metric systems. A relation alternative to (9.5) is

$$\|\theta - \theta_j\| = \arccos\left(1 - \frac{\|x - x_j\|^2}{2\tilde{r}}\right) \tag{9.6}$$

It is also possible to exploit the theory of *Completely Monotone Functions* and *Conditionally Positive Definite Functions* to obtain information about interpolation on any reference circle, as discussed in previous chapters.

9.1.3 Criterion for the Choice of Kernels

Self-similarity, already defined in section 6.1, is especially useful in the spherical case, where it produces invariance of the interpolant to the radius of the spherical domain. Beyond this, kernel selection in spherical surface analysis will depend on the domain $\mathcal{D}$. If $\mathcal{D}$ is a circle, it is required for the kernel to be symmetric and periodic:

$$\kappa(\theta_1, \theta_2) = \kappa^0(\|\theta_1 - \theta_2\|) \text{ and } \kappa^0(\theta) = \kappa^0(\theta \pm 2\ell\pi) , \quad \ell = 0, 1, 2, 3, \cdots \tag{9.7}$$

This type of kernels lend itself to Fourier series analysis. In the case of a sphere, the kernel should satisfy properties generalizing (9.7), as discussed below.

9.1.4 Univariate Circular Approach

In keeping with *Fourier Analysis*, as in Wahba [72], kernels on the circle can be obtained by imposing periodic boundary conditions on kernel functions, but it is more convenient to begin in the first instance with periodic kernel functions: *Eigenfunctions* and *Eigenvalues* have a particularly simple form

$$\kappa(\theta) \sim \sqrt{2} \sum_{\nu=1}^{\infty} a_\nu \cos 2\pi\nu\theta + \sqrt{2} \sum_{\nu=1}^{\infty} b_\nu \sin 2\pi\nu\theta \quad \theta \in [0, 1] \equiv U^1 \subset \mathcal{R}^1 \tag{9.8}$$

with $\sum_{\nu=1}^{\infty} (a_\nu^2 + b_\nu^2)(2\pi\nu)^{2k} < \infty$, where the domain of interest is the unit circle. Since

$$\frac{d^k}{d\theta^k} \left\{ \begin{array}{c} \cos 2\pi\nu\theta \\ \sin 2\pi\nu\theta \end{array} \right\} = (2\pi\nu)^k \times \left\{ \begin{array}{c} \pm\cos 2\pi\nu\theta \\ \pm\sin 2\pi\nu\theta \end{array} \right\} \tag{9.9}$$

then if $\sum_{\nu=1}^{\infty} (a_\nu^2 + b_\nu^2)(2\pi\nu)^{2k} < \infty$ holds, we have

$$\sum_{\nu=1}^{\infty} (a_\nu^2 + b_\nu^2)(2\pi\nu)^{2k} = \int_0^1 \left| \kappa^{(k)}(u) \right|^2 du \tag{9.10}$$

A natural way of obtaining a valid kernel is to derive the *Reproducing Kernel*. From the famous *Parseval Identity*, the reproducing kernel function (inner product) is given by

$$\kappa(\theta_1, \theta_2) = \sum_{\nu=1}^{\infty} \frac{2}{(2\pi\nu)^{2k}} [\cos 2\pi\nu\theta_1 \cos 2\pi\nu\theta_2 + \sin 2\pi\nu\theta_1 \sin 2\pi\nu\theta_2] \tag{9.11}$$

$$= \sum_{\nu=1}^{\infty} \frac{2}{(2\pi\nu)^{2k}} \cos 2\pi\nu(\theta_1 - \theta_2) \tag{9.12}$$

which is a member of *Covariance Kernel*. The eigenvalues of the reproducing kernel are all of multiplicity 2 and are $\lambda_\nu = (2\pi\nu)^{-2k}$, and the eigenfunctions are $\sqrt{2}\sin 2\pi\nu\theta$ and $\sqrt{2}\cos 2\pi\nu\theta$. A closed form expression for $\kappa(\theta_1, \theta_2)$ using Bernoulli polynomials was given by Craven and Wahba [13]. The Bernoulli polynomials $B_\ell(\theta), \ell = 0, 1, \cdots, \theta \in [0, 1]$ are known to satisfy the recursive relations

$$B_0(\theta) = 1 \tag{9.13}$$

$$\frac{1}{\ell} \frac{d}{d\theta} B_\ell(\theta) = B_{\ell-1}(\theta) \tag{9.14}$$

$$\int_0^1 B_\ell(u) du = 0, \quad \ell = 1, 2, \cdots \tag{9.15}$$

The following formula is available in Abramowitz and Stegun [1].

$$B_{2k}(\theta) = (-1)^{k-1} 2(2k)! \sum_{\nu=1}^{\infty} \frac{\cos 2\pi\nu\theta}{(2\pi\nu)^{2k}}, \quad \theta \in [0, 1] \tag{9.16}$$

so that $\kappa(\cdot)$ of (9.12) is given by

$$\kappa(\theta_1, \theta_2) = \frac{(-1)^{k-1}}{(2k)!} B_{2k}([\theta_1 - \theta_2]) \tag{9.17}$$

where $[\theta_1 - \theta_2]$ is the fractional part of $\theta_1 - \theta_2$. $\kappa(\theta_1, \theta_2)$ is a stationary covariance on the circle, whose associated stochastic process $Z(\theta), \theta \in [0,1]$ possess exactly $k-1$ quadratic-mean derivatives and satifies the periodic boundary conditions $Z^{(\nu)}(0) = Z^{(\nu)}(1), \nu = 0, 1, 2, \cdots, k-1$. The first few such functions are given by

$$k = 1, \quad B_2(|\theta|) = \frac{1}{2}\{\frac{1}{6} - |\theta|(1 - |\theta|)\} \tag{9.18}$$

$$k = 2, \quad B_4(|\theta|) = -\frac{1}{24}\{-\frac{1}{30} + |\theta|^2(1 - |\theta|)^2\} \tag{9.19}$$

$$k = 3, \quad B_6(|\theta|) = \frac{1}{720}\{\frac{1}{42} - \frac{1}{2}|\theta|^2(1 - |\theta|)^2 - |\theta|^3(1 - |\theta|)^3\} \tag{9.20}$$

Due to *Translation Invariance*, the constant term and scaling factor are removable so a bit of simplification is possible here. With $\theta, \theta_j \in [0, 2\pi] \in \mathcal{R}^1$,

$$\kappa_1(\theta, \theta_j) = \|\theta - \theta_j\| \left(2\pi - \|\theta - \theta_j\|\right) \tag{9.21}$$

$$\kappa_2(\theta, \theta_j) = \|\theta - \theta_j\|^2 \left(2\pi - \|\theta - \theta_j\|\right)^2 \tag{9.22}$$

$$\kappa_3(\theta, \theta_j) = \frac{1}{2}\|\theta - \theta_j\|^2 \left(2\pi - \|\theta - \theta_j\|\right)^2 + \|\theta - \theta_j\|^3(2\pi - \|\theta - \theta_j\|)^3 \tag{9.23}$$

This enable us to construct different types of models for circular profile interpolation.

$$k = 1, \quad f(\theta) = \beta_0 + \sum_{j=1}^{N} \alpha_j \kappa_1(\theta, \theta_j), \quad \sum_{j=1}^{N} \alpha_j = 0 \tag{9.24}$$

$$k = 2, \quad f(\theta) = \beta_0 + \sum_{j=1}^{N} \alpha_j \kappa_2(\theta, \theta_j), \quad \sum_{j=1}^{N} \alpha_j = 0 \tag{9.25}$$

$$k = 3, \quad f(\theta) = \beta_0 + \sum_{j=1}^{N} \alpha_j \kappa_3(\theta, \theta_j), \quad \sum_{j=1}^{N} \alpha_j = 0 \tag{9.26}$$

which varies depending on k. In general, any periodic function on the circle $[0,1]$ which is continuous, even and positive-definite has a *Fourier Representation*

$$\kappa(\theta_1 - \theta_2) = \sum_{\nu=1}^{\infty} a_j \cos 2\pi\nu(\theta_1 - \theta_2) \qquad (9.27)$$

with $a_j \geq 0$, $\sum_{\nu=1}^{\infty} a_\nu < \infty$. Schoenberg [68] proved that the most general nonnegative definite functions on the circle take the above form. In other words, such a function $\kappa(\cdot)$ has the property that for arbitrary N, $x_i \in S^1$, and $\alpha_i \in \mathcal{R}$,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \kappa(d(x_i, x_j)) \geq 0 \qquad (9.28)$$

where $d(x_i, x_j)$ is the geodesic distance between the two points on the circle. Note that the models (9.24)-(9.26) are analogous to that of *Ordinary Kriging*, in the sense that the kernel $\kappa(\theta_1 - \theta_2)$ is the covariance function of the *Gaussian Brownian Bridge*. In chapter 10, the models of this category will be denoted by BPk. It is known that the models belonging to this category performs well especially for circular profiles, thus deserve to be used as benchmarks in teh circular case.

9.1.5 Bivariate Circular Approach

To discuss the estimation of circular profile in 2-D space, we need to define the domain of interest $\mathcal{D} \equiv \{(u,v)|(u-\hat{a})^2 + (v-\hat{b})^2 - \tilde{r}^2 = 0\}$, where $\tilde{r}$ is an unspoecified radius. Let us recall the *Variational Principle* in chapter 4. The following kernel, known as Thin Plate Spline, useful here. For $k = 2, 3, \cdots$, and $x, x_j \in \mathcal{D}$,

$$\kappa_k(x, x_j) = \|x - x_j\|_2^{2k-2} \log \|x - x_j\|_2 \qquad (9.29)$$

where $\|x - x_j\|_2 = \sqrt{(x - x_j)^2 + (y - y_j)^2}$. It is important to see that since this function is *Radially Symmetric* and *Self-Similar*, one can restrict the domain to the perimeter of

arbitrary radius. Using this kernel, one can construct the following prediction model

$$k = 2\,, \quad f(x) = \beta_0 + \beta_{10}x + \beta_{01}y + \sum_{j=1}^{N} \alpha_j \kappa_2(x, x_j) \tag{9.30}$$

where $x = (\tilde{r}\cos\theta, \tilde{r}\sin\theta) \subset \mathcal{D}$ and the models can be of any order other than 2. According to the *Variational Principle*, the order $k=2$ typically is preferred, as discussed in Chapter 4. Note that the above discussion can be extended to embedding spaces of any even dimensional circular approach. In chapter 10, the models of this category will be denoted by TPk.

9.1.6 Trivariate Circular Approach

Here it is assumed that the reference circle is imbedded in 3 dimensional space. The domain of interest can be designated by

$$\mathcal{D} \equiv \left\{(u,v,w)|(u-\hat{a})^2 + (v-\hat{b})^2 - \tilde{r}^2 = 0 \text{ and } w = 0\right\} \subset \mathcal{R}^3 \tag{9.31}$$

According to kriging and spline theory, the following kernels, depending on the order k, are valid in 3 dimensional space.

$$\kappa_k(x, x_j) = \|x - x_j\|^{2k-1} = \left[\sqrt{(x-x_j)^2 + (y-y_j)^2 + (z-z_j)^2}\right]^{2k-1} \tag{9.32}$$

for $k = 0, 1, 2, 3, 4, \cdots$. Note that these kernels are *Self-similar*. For this reason, one may consider that the circle is imbedded on X-Y plane with the best fit center $(\hat{a}, \hat{b})$ located at the origin. To make these kernels work in 2 dimensional space, one needs to assume $z = z_j = 0$, in which case (9.32) becomes

$$\begin{aligned}
\kappa_k(x, x_j) &= \left[\sqrt{(x-x_j)^2 + (y-y_j)^2}\right]^{2k-1} = \|x - x_j\|_2^{2k-1} \tag{9.33} \\
&= \left[2 - 2\cos\{\arccos(\frac{x}{\|x\|}, \frac{x_j}{\|x_j\|})\}\right]^{2k-1} = \left[2 - 2\cos\theta_{0j}\right]^{2k-1} \tag{9.34}
\end{aligned}$$

Let the radius of a reference circle be denoted by $\tilde{r}$, then on $B^1 \equiv \mathcal{D} \in \mathcal{R}^3$,

$$\kappa_k(x, x_j) = \left[2\tilde{r} - 2\tilde{r}\cos\theta_{0j}\right]^{2k-1} \tag{9.35}$$

From this fact, one can consider the following prediction models, which can vary depending on the order k of the models.

$$k = 0, \quad f(x) = \sum_{j=1}^{N} \alpha_j \kappa_0(x, x_j) \tag{9.36}$$

$$k = 1, \quad f(x) = \beta_0 + \sum_{j=1}^{N} \alpha_j \kappa_1(x, x_j) \tag{9.37}$$

$$k = 2, \quad f(x) = \beta_0 + \beta_{10}x + \beta_{01}y + \sum_{j=1}^{N} \alpha_j \kappa_2(x, x_j) \tag{9.38}$$

Note that the model (9.36) corresponds to *Spherical Multiquadric* of Hardy and Goepfert [31], which is always of order zero. According to spliners' criterion,, the order $k=2$ is always preferred, while in kriging there seems to be no preference as long as the kernel is valid. In chapter 10, the models of this category will be denoted by KRk.

9.1.7 Smoothing of Circular Profile Data

In most cases of real world problems, the sampled data are subject to random effect errors. In the absence of formal fitting procedures, the best way to assess the fit of a covariance seems to be *cross-validation*. Each data point is deleted in turn and its value predicted from the rest of the data, using the fitted or specified kernel function. The predicted error are then assessed. It is tempting to form the sum of squares of these errors. Based on this idea, Wahba [72] and Cressie [14] presented the formal fitting techniques, such as *Ordinary Cross Validation* and *Generalized Cross Validation*.

9.2 Spherical Surface Estimation

Now, consider a sphere imbedded in 3D space. There are some previous studies, such as Wahba [72], Hardy and Goepfert [31], and Pottmann and Eck [59]. The following approach suggests analogous, but somewhat different point of view.

Again, let us get started with *Least Square Method*, or "best fitting" for sphere. The distance r_i from the center (a, b, c) to a point $\mathbf{u}_i = (u_i, v_i, w_i)$ can be given by $r_i = \sqrt{(u_i - a)^2 + (v_i - b)^2 + (w_i - c)^2}$. Hence, the *Least Square Criterion* for spherical feature can be expressed as

$$\min_{\mu} \sum_{i=1}^{N} \left[\sqrt{(u_i - a)^2 + (v_i - b)^2 + (w_i - c)^2} - r \right]^2 \equiv \min_{\mu} \sum_{i=1}^{N} \left[r_i - r \right]^2 \quad (9.39)$$

For convenience, assume that the reference center $(\hat{a}, \hat{b}, \hat{c})$ is situated at the origin $(0, 0, 0)$, where $(\hat{a}, \hat{b}, \hat{c})$ is the *Best Fit Center*.

9.2.1 Bivariate Spherical Approach

A sphere can be looked upon as a natural extension of a circle or a surface of revolution of a circle. On bivariate setting, let us suppose that the domain of interest

$$\mathcal{D} = \{(\theta, \phi) | 0 \leq \theta < 2\pi \text{ and } 0 \leq \phi < \pi\}$$

To discuss spherical kernels from a bivariate point of view, we concentrate on the diatance $\|\cdot, \cdot\|$ between two points on the sphere. The geodesic distance between any two points $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_j$ is given by $\|\boldsymbol{\theta}, \boldsymbol{\theta}_j\| = \min\left\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|, 2\pi - \|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|\right\}$. In terms of that distance, the following two kernels are of special interest

$$\kappa_1(\boldsymbol{\theta}, \boldsymbol{\theta}_j) = \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|(2\pi - \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|) \quad (9.40)$$

$$\kappa_2(\boldsymbol{\theta}, \boldsymbol{\theta}_j) = \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|^2(2\pi - \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|)^2 \quad (9.41)$$

The reader will notice that these kernels, as in the geodesic distance itself, are periodic on a great circle of the spaere considered. Using these kernels, the prediction can be performed by

$$\hat{f}(\boldsymbol{\theta}) = b_0 + \sum_{j=1}^{N} a_j \cdot \kappa(\boldsymbol{\theta}, \boldsymbol{\theta}_j) = b_0 + \sum_{j=1}^{N} e_j \cdot \lambda(\boldsymbol{\theta}, \boldsymbol{\theta}_j) \qquad (9.42)$$

Note that the models are always of order 1 regardless of k, just as in the case of the univariate circular approach. BPk denotes the models of this category in chapter 10.

9.2.2 Trivariate Setting

Now consider a sphere imbedded in 3 dimensional space. The domain is given by

$$\mathcal{D} = \left\{ (u, v, w) | (u - \hat{a})^2 + (v - \hat{b})^2 + (w - \hat{c})^2 - \tilde{r}^2 = 0 \right\} \subset \mathcal{R}^3 \qquad (9.43)$$

According to kriging and spline theory, the following kernels, depending on $k = 0, 1, 2, \cdot$, are valid on three dimensional setting. For $\boldsymbol{x}, \boldsymbol{x}_j \in \mathcal{D} \subset \mathcal{R}^3$,

$$\kappa_k(\boldsymbol{x}, \boldsymbol{x}_j) = \|\boldsymbol{x} - \boldsymbol{x}_j\|^{2k-1} = \left\{ \sqrt{2\tilde{r} - 2\tilde{r}\cos\theta_{0j}} \right\}^{2k-1} \qquad (9.44)$$

$$= \left\{ \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \right\}^{2k-1} \qquad (9.45)$$

where $\theta_{0j} = \arccos\left(\dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \dfrac{\boldsymbol{x}_j}{\|\boldsymbol{x}_j\|} \right)$. The prediction can be conducted based on

$$\hat{f}(\boldsymbol{x}) = \mathbf{b}' \cdot \mathbf{x} + \sum_{j=1}^{N} a_j \cdot \kappa(\boldsymbol{x}, \boldsymbol{x}_j) = \mathbf{b}' \cdot \mathbf{x} + \sum_{j=1}^{N} e_j \cdot \lambda(\boldsymbol{x}, \boldsymbol{x}_j) \qquad (9.46)$$

The smoothing technique is also applicable to these models based on cross validation. In chapter 10, KRk denotes the models of this category.

9.3 Cylindrical Surface Estimation

Given a set of points sampled from an actual cylindrical surface, one can consider the *Least Squares Criterion* for the estimation of cylindrical surface. The distance r_i from a point on the best fit center line or axis $(\hat{a}, \hat{b})$ to a point $\mathbf{u}_i = (u_i, v_i, w_i)$ is given by $r_i = \sqrt{(u_i - a)^2 + (v_i - b)^2}$. The *Least Squares Criterion* for a cylinder is expressed as

$$\min_{\mu} \sum_{i=1}^{N} \left[\sqrt{(u_i - a)^2 + (v_i - b)^2} - \mu(\boldsymbol{x}) \right]^2 \equiv \min_{\mu} \sum_{i=1}^{N} \left[r_i - \mu(\boldsymbol{x}_i) \right]^2, \boldsymbol{x} \in \mathcal{D} \quad (9.47)$$

The reference axis $(\hat{a}, \hat{b})$ is set to $(0, 0)$, where $(\hat{a}, \hat{b})$ is the *Best Fit Axis*.

9.3.1 Trivariate Cylindrical Approach

Consider a cylinder imbeded in 3 dimensional space. The domain of interest can be identified as

$$\mathcal{D} \equiv \{(x, y, z) | (x - \hat{a})^2 + (y - \hat{b})^2 - \tilde{r}^2 = 0 \text{ and } c \leq z \leq d\} \subset \mathcal{R}^3 \quad (9.48)$$

where $\tilde{r}$ is the radius of a cylinder of any magnitude. In view of the spatial prediction theory, only the kernels on three dimensional setting is considered. For $\boldsymbol{x}, \boldsymbol{x}_j \in \mathcal{D}$, the kernel can be given depending on $k = 0, 1, 2, \cdots$,

$$\kappa_k(\boldsymbol{x}, \boldsymbol{x}_j) = \|\boldsymbol{x} - \boldsymbol{x}_j\|^{2k-1} = \left\{ \sqrt{2\tilde{r} - 2\tilde{r}\cos\theta_{0j} + (z - z_j)^2} \right\}^{2k-1} \quad (9.49)$$

$$= \left\{ \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \right\}^{2k-1} \quad (9.50)$$

where $\tilde{r}$ is the radius of a cylinder of any magnitude. The prediction is performed by

$$\hat{f}(\boldsymbol{x}) = \mathbf{b}' \cdot \mathbf{x} + \sum_{j=1}^{N} a_j \cdot \kappa(\boldsymbol{x}, \boldsymbol{x}_j) = \mathbf{b}' \cdot \mathbf{x} + \sum_{j=1}^{N} e_j \cdot \lambda(\boldsymbol{x}, \boldsymbol{x}_j) \quad (9.51)$$

Note that smoothing and variance calculation are also possible by usual procedure. In chapter 10, the models of this category will be denoted by KRk.

9.4 Toroidal Surface Estimation

A torus can be regarded as a natural extension of a circle: a surface of revolution of a circle around w axis. Now consider a torus imbedded in 3 dimensional space. Again, let us get started with *Least Square Method* for torus. In this case, the crosection is the circle $(u - R)^2 + w^2 = r^2$ with two kinds of radii: the Large scale radius R and small scale radius r. In parametric terms the equation of a torus can given as

$$\mathbf{r} = \mathbf{r}(\theta, \phi) = (R + a\cos\theta)\cos\phi \cdot \boldsymbol{i} + (R + a\cos\theta)\sin\phi \cdot \boldsymbol{j} + a\sin\theta \cdot \boldsymbol{k}$$

where i, j and k are standard orthonormal vectors in 3 dimensional space. For convenience, suppose that the reference center $(\hat{u}_0, \hat{v}_0, \hat{w}_0)$ is situated at the origin $(0, 0, 0)$, where $(\hat{u}_0, \hat{v}_0, \hat{w}_0)$ is the *Best Fit Center*. The center of the cross-sectional circle can be given by $(R\cos\phi, R\sin\phi)$.

9.4.1 Bivariate Toroidal Approach

In a bivariate approach, a torus is viewed as a surface of two paramaters $\boldsymbol{\theta} = (\theta, \phi)$. On bivariate setting, the domain of interest can be identified as

$$\mathcal{D} = \{(\theta, \phi) | 0 \le \theta < 2\pi \text{ and } 0 \le \phi < 2\pi\}$$

To discuss toroidal kernels from a bivariate point of view, we concentrate on the *Geodesic Distance* $\|\cdot, \cdot\|$ between two points on the reference torus. The geodesic distance between any two points $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_j$ is given by $\|\boldsymbol{\theta}, \boldsymbol{\theta}_j\| = \min\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|, 2\pi - \|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|\right)$. In terms of that distance, the following two kernels are of special interest

$$\kappa_1(\boldsymbol{\theta}, \boldsymbol{\theta}_j) = \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|(2\pi - \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|) \tag{9.52}$$

$$\kappa_2(\boldsymbol{\theta}, \boldsymbol{\theta}_j) = \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|^2(2\pi - \|\boldsymbol{\theta}, \boldsymbol{\theta}_j\|)^2 \tag{9.53}$$

Note that these kernels, as in the geodesic distance itself, are periodic both with respect to the revolution and on a cross-sectional circle of the torus considered. Using these kernels, the prediction can be performed by

$$\hat{f}(\boldsymbol{\theta}) = b_0 + \sum_{j=1}^{N} a_j \cdot \kappa(\boldsymbol{\theta}, \boldsymbol{\theta}_j) = b_0 + \sum_{j=1}^{N} e_j \cdot \lambda(\boldsymbol{\theta}, \boldsymbol{\theta}_j) \tag{9.54}$$

Note that the models are always of order 1 regardless of k, just as in the case of the bivariate spherical approach. BPk denotes the models of this category in chapter 10. Trivariate toroidal approach based on Euclidean distance is also possible, but not shown here.

CHAPTER 10.   EXAMPLES OF ROUND FEATURE ESTIMATION

This chapter is to demonstrate how the previous studies can be applied to the estimation of surfaces of machined parts that have round features. The set of data points is assumed to be obtained in three-coordinate form, as by three-coordinate measuring machines (3-CMM).

There are some reasons for measuring manufactured parts using three dimensional CMM: tolerance checking of the manufactured part, testing for compensation of machine tools and testing the accuracy of three-coordinate measuring machine itself.

In the case of tolerance checking of circular profiles, one may fit a circle into the data and check whether the specification is met. In case of testing for compensation of machine tools, the residuals are of more concern. In practice, the circular test is in wide use to test the accuracy of CMM itself. The circular test is a fast method of testing the geometrical accuracies of three-axis machines positioning, straightness, roll, pitch, yaw and perpendicularity. In the circular test, a standard disc is used. The disc is measured in different positions in the working area of the machine.

The calculated mean square fit diameter, standard deviation and Fourier analysis result in an analysis of the error sources and of the geometric error components. However, the calculation of the two or three dimensional uncertainty from measured geometric error components is generally not easy because of the random error effect.

In relation to this, we discuss how the three methods — Bernoulli Polynomial, Thin Plate Spline and IRF kriging — in the previous chapter can be used for round feature estimation such as circular profile, spherical and cylindrical surfaces.

10.1 Circular Profile Estimation

The easiest way of getting a glimpse of how those models work is to designate some test functions. Since the test function is perfectly known, one can define a measure of the performance in one way or another.

10.1.1 Test of Reproducibility

Among the performance measures of predictors would be the reproducibility of pre-specified circular feature. To test the performance of the prediction models, a set of test functions are designated: The criterion would be the sum of squared errors (SSQ) at prediction points. Candidate test functions are $f(\theta) = a + b\cos\ell\theta$, $a > 0$, $|b| < a$, $\ell = 0, 1, 2, 3, \cdots$. Examples of test functions are shown in Figure 10.1.

Let us define the reproducibility to be the sum of squared deviation from prespecified near-circular shape at each $\theta_i, i = 1, \cdots, 360$, which is given by

$$SSQ = \sum_{i=1}^{360} \left[\hat{f}(\theta_i) - f(\theta_i)\right]^2 \tag{10.1}$$

where $f(\theta_i)$ is the known test function and $\hat{f}(\theta_i)$ is the prediction at each $\theta_i, i = 1, \cdots, 360$. Table 10.1 and 10.2 condensed from Figure 10.2 and Figure 10.3, respectively, shows that the three methods do not differ in any substantial way. It is quite inspiring that one can detect 3 lobed round features with as small a sample size as 9. It seems to be almost
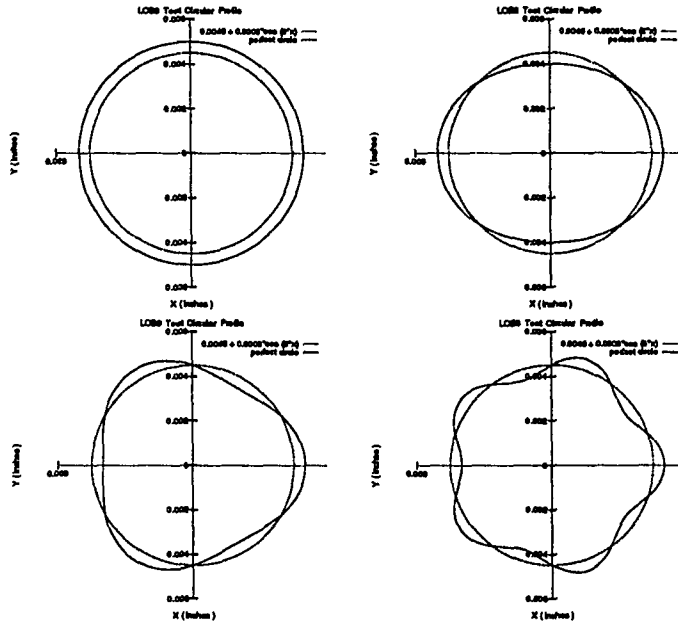
Figure 10.1:   Test Functions

Table 10.1:   Comparison of Test Results k=1

	LOB0	LOB2	LOB3	LOB5
MQ0	.1141476E-05	.2103417E-5	.6685560E-5	.3329501E-4
BP1	.2520260E-33	.1308544E-5	.5947378E-5	.3225025E-4
KR1	.2279519E-33	.1178821E-5	.5780964E-5	.3237041E-4

MQ0 : Multiquadric of $k = 0$ on 2-D Setting
BP1 : Bernoulli Polynomial of $k = 1$ on 1-D Setting
KR1 : IRF Kriging (MQ-B) of $k = 1$ on 3-D Setting

Table 10.2:   Comparison of Test Results $k=2$

	LOB0	LOB2	LOB3	LOB5
BP2	.1148965E-29	.5000040E-8	.3418271E-6	.4495495E-4
TP2	.1035142E-29	.4337514E-7	.1165846E-5	.3919316E-4
KR2	.1086565E-29	.9343919E-9	.2145103E-6	.4662230E-4

BP2 : Bernoulli Polynomial of $k = 2$ on 1-D Setting
TP2 : Thin Plate Spline of $k = 2$ on 2-D Setting
KR2 : IRF Kriging of $k = 2$ on 3-D Setting

impossible to detect 5 lobed objects. However, if the sample size is doubled, it would be possible to make a good estimation of 5 lobed round features.

10.1.2 Interpolation of Circular Profile .

The interpolation of circular profile is characterized by the fact that the interpolant passes through all the observation points. Let's suppose 18 points of circular data are given as in Fig 10.4. The calculated interpolant of each method is plotted as in Fig 10.4, which indicates that the two models of order $k=1$ do not make big differences.

The shortcoming of this type of interpolation is that the interpolant has cusps at each data point. Note that if one employ tension parameter c^2 to round off the vertices, then it reduces to Hardy's multiquadric biharmonic model. Instead of employing *Tension Parameter*, one may consider an increase of the order of the model. Figure 10.5 shows that the performance of the three models of order $k=2$ do not make any substantial differences, either. A drawback of this method might be that the interpolant looks too wavy and hence is not naturally smooth. A plausible cause of this would be that the interpolant is being forced to pass through all the observation points.
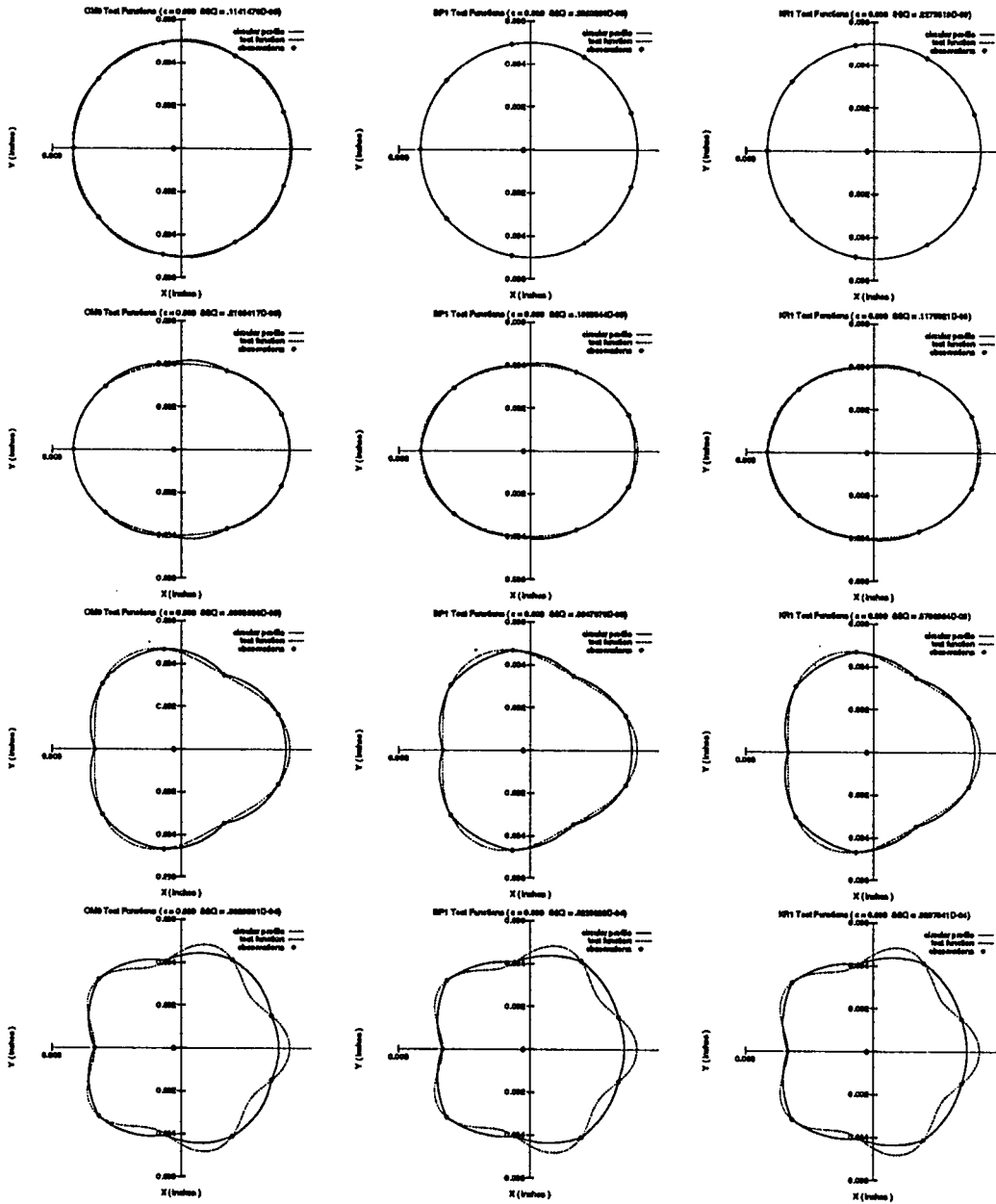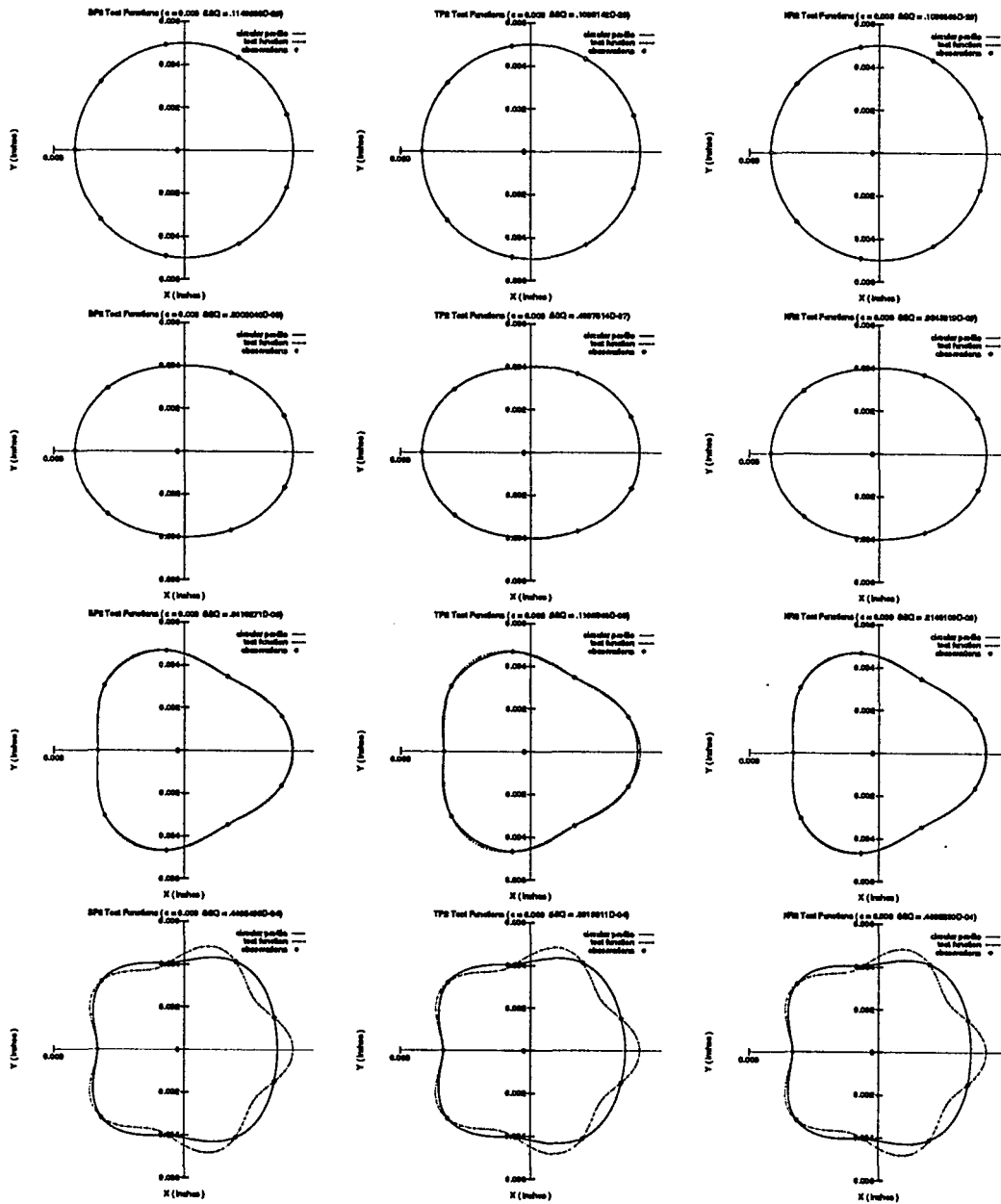
Figure 10.2: Test Results – MQ0, BP1 and KR1
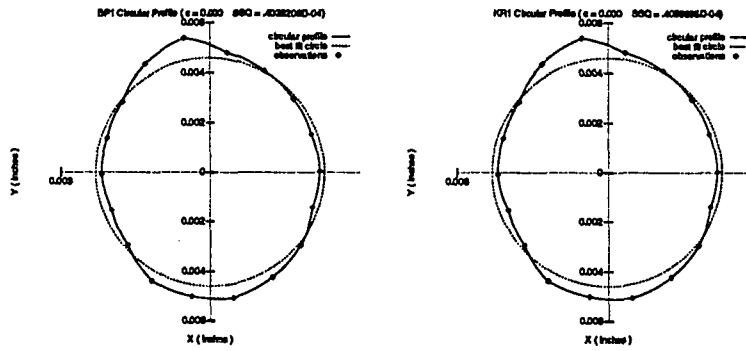
Figure 10.3: Test Results – BP2, TP2 and KR2

Figure 10.4:   Interpolated Circular Profile – BP1, KR1

10.1.3 Smoothing of Circular Profile Data

One may suspect that the measurements by coordinate measuring machine (CMM) are subject to some random errors, just like sampling errors in statistics. Smoothed circular profiles based on *Ordinary Cross Validation* are shown in Figure 10.6. Note that the three methods yields very similar results. For practical purposes, all of them provide excellent estimates pr prediction.
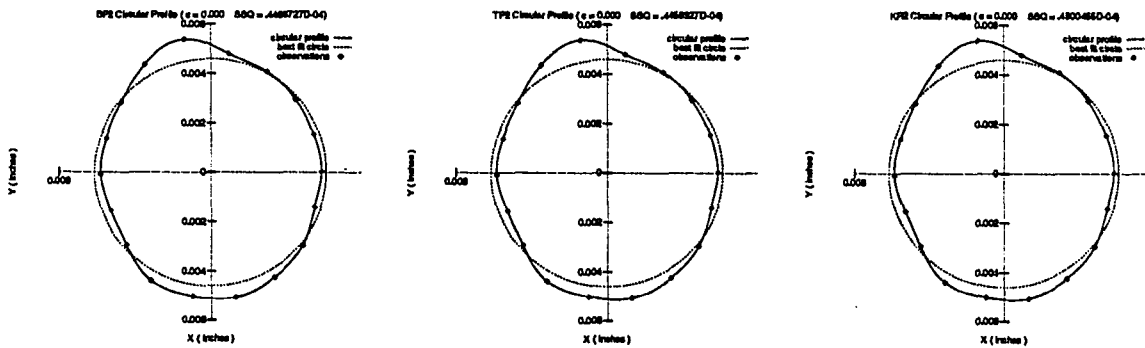

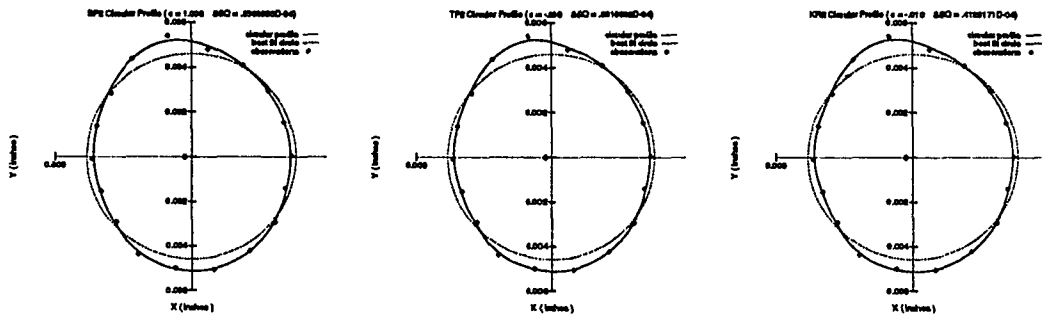Figure 10.5:   Interpolated Circular Profile – BP2, TP2, and KR2

Figure 10.6: Smoothed Circular Profile – BP2, TP2, and KR2

10.1.4 Lessons of Circular Profile Fitting

It is interesting to see that circular profiles can be estimated in 1,2 or 3 dimensional setting, owing to the invariance properties of kernel functions. The three methods perform almost equally well. There would be no preference among the three methods. It might be because of the fact that they belong to the same family of functions, say M spline or Gaussian family. The strong point of KR model is that it can be easily extended to the estimation of arbitrary round objects in 3 dimensional settings.

10.2 Spherical Surface Estimation

Now, consider a sphere as a natural extension of a circle. An example data is shown in Figure 10.7. In this example, 9×5 grid data are used, while it is also applicable to non-grid data. Three types of models are considered. It is expected that the prediction model should work good as long as the circular profile is concerned. The actual data and reproduced surface is plotted in Fig 10.7 and Fig 10.8. Note that BPk and KRk work much the same way. Figure 10.7 and Figure 10.8 indicate that the sphere is a little bit distorted.
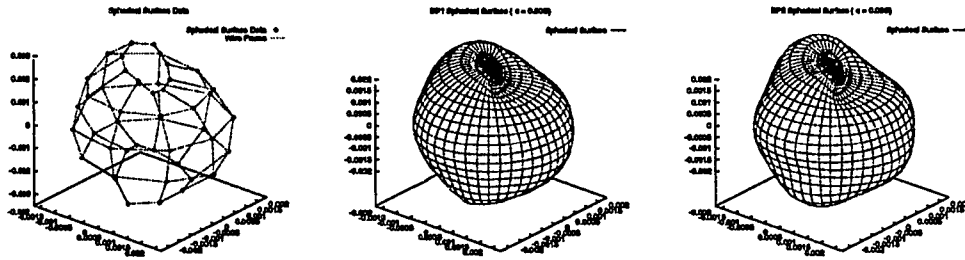
Figure 10.7: Actual Data and Interpolated Surfaces - BP1 and BP2
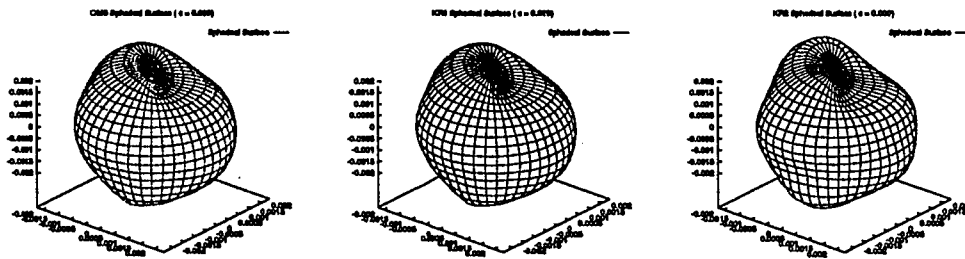
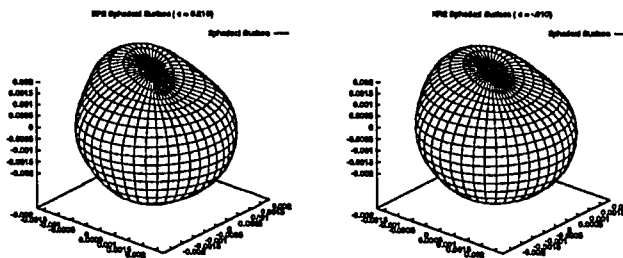Figure 10.8: Interpolated Surfaces - MQ0, KR1 and KR2

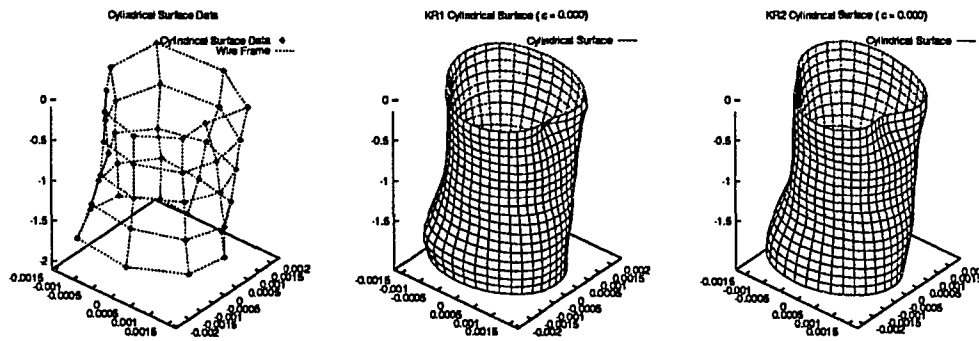Figure 10.9: Smoothed Surfaces - BP2 and KR2

Figure 10.10:   Actual Data and Interpolated Surfaces - KR1 and KR2

10.3 Cylindrical Surface Estimation

Figure 10.10 shows the actual data and the estimated surface. The reproduced surfaces look more realistic. The KR1 has some cusps on the surface, while the KR2 reproduces no cusps at all, but too wavy surface. By smoothing, one might be able to reduce the unnecessary willowing of the surfaces as shown in Fig 10.11. An interesting fact is that the abstract of the cylinder is a little bit tilted. In such a case, it would be better to try to fit a tilted cylinder by least square methods. According to ANSI standard, there are many types of distorted version of cylinders; Taper, Barrel, and so on, most of which are detectable by this simple approach. Such a study can find its wide applications in manufacturing area: pilot study of geometric modeling and wear pattern study of cylindricval surfaces such as bushings and inner surface of a cylinder.

10.4 Toroidal Surface Estimation

An example data is plotted in Figure 10.12. As in the previous sections, the BP1 reproduces some cusps on the surface, while the BP2 does not. Here, only the result of
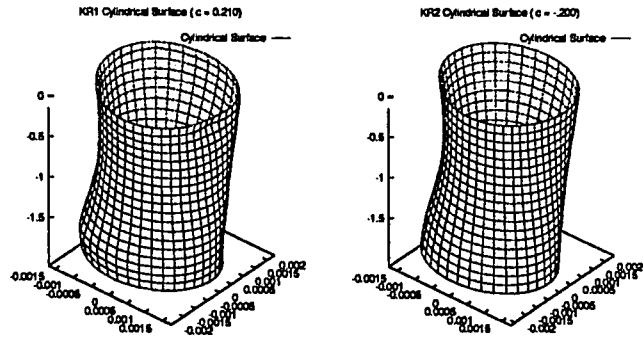
Figure 10.11:   Smoothed Surfaces - KR1 and KR2
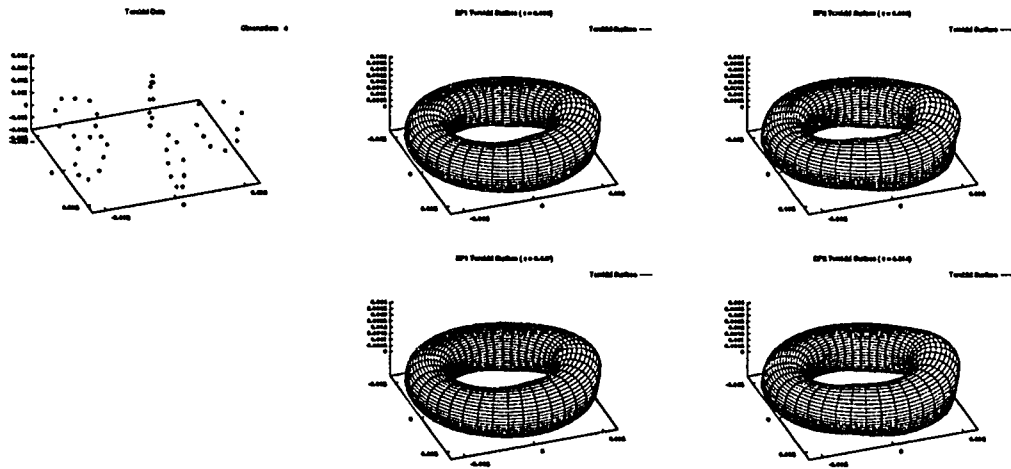
Figure 10.12:   Actual Data, Interpolated and Smoothed Surfaces – BP1 and BP2

geodesic approach is shown, while Euclidean approach is expected to produce as good a result as shown here. This method applies basically to any ring-shaped objects such as tires, rims and so on. For this type of study to be applied to the real world problems, it is desirable to use the larger sample size with some modifications of the model.

CHAPTER 11.   CONCLUSION

There exist several spatial prediction methods : kriging, splines and some other methods such as multiquadrics. Kriging is based on stochastic process theory and the most common assumption is the normality of the underlying process, which ease the analysis of the process. Among many aspects of kriging is the fact that the kriging model need not be tied to conventional *Covariance (Second Order, Wide Sense) Stationary* models. Indeed, one can extend *Covariance Stationarity* to *Intrinsically (Variogram, Wider Sense)* Stationarity. The objective of deterministic spline modeling is to minimize the error committed by interpolation. It is worthy of note that this minimized error coincides with the minimized mean squared (prediction) error (MSE) advocated by krigers. Even though spatial prediction methods differ in their objectives, they enjoy the same structure identifiable within a unified framework.

Kriging and spline prediction models take several forms, but these models are simultaneously interpretable in both disciplines. Given the order k of the model, spline function kernels or semi-variograms may be subject to certain conditions such as *Conditional Positive Definiteness of Order k*, which guarantees the boundedness and the uniqueness of the interpolant, and the optimization process common to both features of *Translation Invariance* and *Annihilation*.

The major difference between kriging and spline methods lies in the criterion for

the choice of kernels. Seen from the krigers' point of view, the kernel type depends on assumptions concerning the underlying stochastic process, while spliners choose spline function on the basis of physical criteria, such as roughness minimization.

According to the author's short experience, the multiquadric method is amazingly good for practical purposes, while it leaves something to be desired in some theoretical sense as discussed in chapter 7. Interpolants derived from the MQ biharmonic model of order zero may assume infinite values, while the MQ biharmonic model of order 1 does not suffer from this defect. MQ harmonic and biharmonic models are interpretable in the *Wider Sense Stationary* kriging context: 1) The MQ *harmonic* kernel can be interpreted as a covariogram of a white noise process, 2) The MQ *biharmonic* kernel can be interpreted as a covariogram of Brownian motion.

There is a close link between least squares and spatial prediction. The least square methods can be viewed as a special case of spatial prediction method. Accordingly, any least square model can be naturally extended to the spatial prediction model.

The model framework for circular profile interpolation and smoothing can be constructed by extending the *Least Squares Model* for round features to a *Generalized Least Squares* model. One is able to construct various models depending on domain, kernel type and order of the model. One can formulate three types of models depending on the dimension of domain: one, two or three dimensional setting. For circles only, those three types of models perform almost equally well. The three dimensional setting or trivariate approach is easy and versitile, especially when dealing with three dimensional geometric objects like spheres or cylinders or possibly any other type of objects.

Spatial prediction methods can find its wide applicability in the area of manufacturing, especially for geometric rendering of form-errors of machined parts or wear-pattern

of used parts.

The major conclusion of this dissertation is that interpolation should proceed in a three-dimensional setting, by imbedding if necessary, when the domain of interest is less than three-dimensional. This produces smooth interpolants on which the three major schools essentially agree. It would be of great help to the engineers engaged in manufacturing to have the corresponding software incorporated in coordinate measuring machines.

BIBLIOGRAPHY

[1] Abramowitz, M. and Stegun, I., *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, U.S. Government Printing Office, Washington D.C., 1968.

[2] ANSI B89.31, *Measurement of Out-of-Roundness*, The American Society of Mechanical Engineers, New York, 1972.

[3] ANSI Y14.5M, *Dimensioning and Tolerancing*, The American Society of Mechanical Engineers, New York, 1982.

[4] Armstrong, M. and Diamond, P., "Testing Variograms for Positive Definiteness," *Mathematical Geology*, Vol. 16, No. 4, 1984, pp. 407-421.

[5] Aronszajn, N., "Theory of Reproducing Kernel," *Trans. Amer. Math. Soc.*, Vol. 68, 1950, pp. 337-404.

[6] Berman, M., "Estimating the Parameter of a Circle When Angular Differences are Known," *Applied Statistics*, Vol. 32, No. 1, 1983, pp. 1-6.

[7] Berman, M. and Culpin, D., "The Statistical Behavior of Some Least Squares Estimators of the Centre and Radius of a Circle," *J. R. Statist. Soc.*, Series B, Vol. 48, No. 2, 1986, pp. 183-196.

[8] Bochner, S., "Hilbert Distances and Positive Definite Functions," *Annals of Mathematics*, Vol. 42, 1941, pp. 647-656.

[9] Buhman, M.D., "Multivariate Interpolation with Radial Basis Functions," *Report DAMTP 1988/NA8*, University of Cambridge, Cambridge, U.K., 1988.

[10] Carson, R.E. and Foley, T.A., " The Parameter R^2 in Multiquadric Interpolation," *Computers for Mathematics with Application*, Vol. 21, No. 9, 1991, pp. 29-42.

[11] Christakos, G., "On the Problem of Permissible Covariance and Variogram Models," *Water Resources Research*, Vol. 20, No. 2, 1985, pp. 251-265.

[12] Coope, I.D., "Circle Fitting by Linear and Nonlinear Least Squares," *J. of Optimization Theory and Applications*, Vol. 76, No. 2, 1993, pp. 381-388.

[13] Craven, P. and G. Wahba, "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numer. Math.*, Vol. 31, 1979, pp. 377-403.

[14] Cressie, N., "Geostatistics," *The American Statistician*, Vol. 43, No. 4, 1989, pp. 197-202.

[15] Cressie, N., "Reply to Wahba's letter," *The American Statistician*, Vol. 44, No. 3, 1990, pp. 256-258.

[16] Cressie, N., *Statistics for Spatial Data*, John Wiley and Sons, Inc., New York, 1991.

[17] David, H.T. and Yoo, S., "Where Next ? Adaptive Measurement Site Selection for Area Remediation," *Environmental Statistics, Assessment and Forecasting*, Lewis Publishers Inc., Ann Arbor, Michigan, 1993.

[18] David, H.T. and Yoo, S., "The Best of Both Worlds: Integrating Statistical and Deterministic Approaches to Area Remediation," *Unpublished Preprint*, 1993.

[19] Davis, P.J., *Interpolation and Approximation*, Dover Publications Inc., New York, 1975, pp. 316-326.

[20] Dubrule, O., "Two Methods with Different Objectives: Splines and Kriging," *Journal of Mathematical Geology*, Vol. 15, 1983, pp. 245-257.

[21] Duchon, J., "Interpolation Des Functions De Deux Variables Suivant Le Principe De La Flesion Des Plaques Minces", *R.A.I.R.O. Analyse numerique*, Vol. 10, No. 12, 1976, pp. 5-12.

[22] Dyn, N., Levin, D. and Rippa, S., "Numerical Procedures for Surface Fitting of Scattered Data by Radial Functions," *SIAM J. Sci. Stat. Comput.*, Vol. 7, NO. 2, April 1986, pp. 639-659.

[23] Dyn, N., "Interpolation and Approximation by Radial and Related Functions," *Approximation Theory VI*, C. Chui, L. Schumaker and J. Ward, eds., Academic Press, Orlando, 1986, pp. 211-234.

[24] Eck, M., "MQ-Curves are Curves in Tension," *Math. Methods in CAGD II*, Ton Lyche and Larry L. Schumaker (eds.), Amsterdam, North-Holland, 1992, pp. 217-228.

[25] Franke, R., "Scattered Data Interpolation: Tests of Some Methods," *Mathematics of Computation*, Vol. 38, 1982, pp. 181-199.

[26] Franke, R., "Thin Plate Splines with Tension," *CAGD*, Vol. 2, 1985, pp. 181-199.

[27] Harder, R.L., "Interpolation Using Surface Splines," *Journal of Aircraft*, Vol. 9, No. 2, 1972, pp. 189-191.

[28] Hardy, R.L., "Multiquadric Equations of Topography and Other Irregular Surfaces," *Journal of Geophysical Research*, Vol. 76, 1971, pp. 1905-1915.

[29] Hardy, R.L., "Geodetic Applications of Multiquadric Equations ," *Final Report, Engineering Research Institute*, Iowa State University, Ames, Iowa. ISU-ERI-Ames-76245, 1976.

[30] Hardy, R.L., "Theory and Applications of the Multiquadric Biharmonic Method," *Computers for Mathematics with Application*, Vol. 19, 1990, pp. 163-208.

[31] Hardy, R.L. and Goepfert, W.M. "Least Squares Prediction of Gravity Anomalies, Geoidal Undulations, and Deflections of the Vertical with Multiquadric Harmonic Functions," *Geophysics Research Letters*, Vol. 2, 1975, pp. 163-208.

[32] Hardy, R.L. and Nelson, S.A. "A Multiquadric Biharmonic Representation and Approximation of Disturbing Potential," *Geophysics Research Letters*, Vol. 13, 1986, pp. 18-21.

[33] Huijbregts, C. and Matheron, G., "Universal Kriging: An Optimal Method for Estimating and Contouring in Trend Surface Analysis," *Can. Inst. Mining Metall*, Vol. 12 (C.I.M.M. Special Volume), 1970, pp. 159-169.

[34] Jackson, I.R.H., "An Order of Convergence for Radial Basis Functions," *IMA J. Numer. Anal*, Vol. 9, 1989, pp. 567-587.

[35] Journel, A.G. and Huijbregts, C.J., *Mining Geostatistics*, Academic Press, London, U.K., 1978.

[36] Kansa, E.J., "Multiquadrics – a Scattered Data Approximation Scheme with Applications to Computational Fluid Dynamics," *Computers for Mathematics with Application*, Vol. 19, 1990, pp. 127-145.

[37] Kent, J.T. and Mardia, K.V., "The Link between Kriging and Thin-plate Splines," *Probability, Statistics and Optimization: Chapter 24*, Edited by F.P. Kelly, John Wiley and Sons, Ltd., New York, 1994.

[38] Kimeldorf, G.S. and Wahba, G., "A Correspondence between Bayesian Estimation on Stochastic Processes and Smoothing by Splines," *The Annals of Mathematical Statistics*, Vol. 41, No. 2, 1970, pp. 495-502.

[39] Kitanidis, P.K., "Generalized Covariance Functions in Estimation," *Mathematical Geology*, Vol. 25, No. 5, 1993, pp. 525-540.

[40] Laslett, G.M., McBratney, A.B., Pahl, P.J., and Hutchinson, M.F., "Comparison of Several Spatial Prediction Methods for Soil pH," *Journal of Soil Science*, Vol. 38, 1987, pp. 325-341.

[41] Laslett, G.M., "Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications," *JASA*, Vol. 89, No. 426, 1994, pp. 391-409.

[42] Light, W.A. and Cheney, E.W., "Interpolation by Periodic Radial Basis Functions," *Journal of Mathematical Analysis and Applications*, Vol. 168, 1992, pp. 111-130.

[43] Luenberger, D.G., *Linear and Nonlinear Programming*, Second Ed., Addison Wesley, Melo Park, California, 1989, pp. 423-425.

[44] Madych, W. R., "Miscellaneous Error Bounds for Multiquadric and Related Interpolators," *Computer Math. Applic.*, Vol. 24. No. 12, 1992, pp. 121-138.

[45] Madych, W.R. and Nelson, S.A., "Multivariate Interpolation and Conditionally Positive Definite Functions," *Approx. Theory and its Appl.*, Vol. 4, No. 4, 1988, pp. 77-89.

[46] Mandelbrot, B.B. and Van Ness, J.W., "Fractional Brownian Motion, Fractional Noises and Applications," *SIAM Review*, Vol. 10, 1968, pp. 77-95.

[47] Mardia, K.V., Kent, J.T., Goodall, C.R. and Little, J.A., "Kriging and Splines with Derivative Information," *Unpublished Preprint*, University of Leeds, 1994.

[48] Matern, B., "Spatial Variation," *Lecture Notes in Statistics*, Vol. 36, Splinger - Verlag, Berlin, 1986.

[49] Matheron, G., *La Theorie Des Functions Aleatoires Intrinseques Generalisees*, Internal Report N-252: CGMM, Fontainebleau, France, 1971.

[50] Matheron, G., "The Intrinsic Random Functions and Their Applications," *Adv. Appl. Prob.*, Vol. 5, 1973, pp. 439-468.

[51] Matheron, G., "Splines and Kriging: Their Formal Equivalence," *Down-to-Earth Statistics: Solutions Looking for Geological Problems*, D.F. Merriam, ed. Syracuse University Geological Contributions, Syracuse, 1981, pp. 77-95.

[52] Meinguet, J., "Multivariate Interpolation at Arbitrary Points Made Simple," *Z. Angrew. Math. Phys.*, Vol. 30, 1979, pp. 292-304.

[53] Meinguet, J., "An Intrinsic Appraoch to Multivariate Spline Interpolation at Arbitrary Points," *Polynomial and Spline Approximations*, N.B. Sahney (e.d.), Reidel, Dordrecht, Germany, 1979, pp. 163-190.

[54] Meinguet, J., "Surface Spline Interpolation: Basic Theory and Computational Aspects," *Approximation Theory and Spline Functions*, S.P. Singh, J.H.Burry, and B. Watson (eds.), Reidel, Dordrecht, Germany, 1984, pp. 127-142.

[55] Micchelli, C.A., "Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions," *Constructive Approximation*, Vol. 2, 1986, pp. 11-22.

[56] Myers, D.E., "Kriging, Cokriging, Radial Basis Functions and the Role of Positive Definiteness," *Computers for Mathematics with Application*, Vol. 24, No. 12, 1992, pp. 139-148.

[57] Olea, R.A., "Optimum Mapping Techniques Using Regionalized Variable Theory," *Kansas Geological Survey*, Lawrence, Kansas, 1975.

[58] Parzen, E., "An Approach to Time Series Analysis," *Annals of Mathematical Statistics*, Vol. 32, 1961, pp. 951-989.

[59] Pottman, H. and Eck, M., "Modified Multiquadric Methods for Scattered Data Interpolation over a Sphere," *Computer-Aided Geom. Design*, Vol. 7, 1990, pp. 313-321.

[60] Powell, M.J.D., *Approximation Theory and Methods*, Cambridge University Press, Cambridge, U. K., 1981.

[61] Powell, M.J.D., "Radial Basis Functions for Multivariable Interpolation: A Review," *Proceedings of the IMA Conference on Algorithms for the Approximation of Functions and Data*, M.G. Cox and J.C. Mason (eds.), Oxford University Press, Oxford, U.K., 1990, pp. 143-167.

[62] Powell, M.J.D., "The Theory of Radial Basis Function Approximation," *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms and Radial Functions*, (Edited by W. Light), Oxford University Press, Oxford, U.K., 1992, pp. 105-210.

[63] Powell, M.J.D., *The Theory of Radial Basis Function Approximation in 1990*, Advances in Numerical Analysis, Oxford Science Publications, Vol. 2, Oxford, U.K., 1993, pp. 105-210.

[64] Pruess, S., "Altermnatives to the Exponential Spline in Tension," *Math. Comp.*, Vol. 33, 1979, pp. 1273-1281.

[65] Salkaukas, K., "Some Relationship between Surface Splines and Kriging," *Multivariate Approximation Theory II*, 1982, pp. 313-325.

[66] Schoenberg, I.J., "Metric Spaces and Positive Definite Functions," *Trans. Amer. Math. Soc.*, Vol. 44, 1938, pp. 522-536.

[67] Schoenberg, I.J., "Metric Spaces and Completely Monotone Functions," *Annals of Mathematics*, Vol. 39, No.4, 1938, pp. 811-841.

[68] Schoenberg, I.J., "Positive Definite Functions on Spheres," *Duke Math. J.*, Vol. 9, 1942, pp. 96-108.

[69] Sirayanone, S. and Hardy, R.L. "Comparative Studies of Kriging, Multiquadric Biharmonic and other methods for solving Mineral Resource Problems," *Unpublished Preprint*, 1994.

[70] Sun, X., "On the Solvability of Radial Function Interpolation," *Approximation Theory IV*, Vol. 2, 1989, pp. 643-646.

[71] Wahba, G., "Comment on Cressie," *The American Statistician*, Vol. 44, No. 3, 1990, pp. 255.

[72] Wahba, G., *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

[73] Wahba, G., "Reply to Comment on Cressie," *The American Statistician*, Vol. 44, No. 3, 1990, pp. 255-256.

[74] Watson, G.S., "Smoothing and Interpolation by Kriging and with Splines," *Journal of Mathematical Geology*, Vol. 16, No. 6, 1984, pp. 601-615.

[75] Whittle, P., "Topographic Correlation, Power-Law Covariance Functions, and Diffusion," *Biometrica*, Vol. 49, 1962, pp. 305-314.

[76] Williamson, R.E., "Multiply Monotone functions and Their Laplace Transforms," *Duke Math. J.*, Vol. 23, 1956, pp. 189-207.

[77] Xu, Y. and Cheney, E. W., "Interpolation by Periodic Radial Functions," *Computers Math. Applic.*, Vol. 24, No. 12, 1992, pp. 201-215.

[78] Yaglom, A.M., "Some Classes of Random Fields in n-Dimensional Space, Related to Stationary Random Process," *Theory of Probability and its Applications*, Vol. 2, No. 3, 1957, pp. 273-320.

ACKNOWLEDGEMENTS

APPENDIX A.   SPATIAL PREDICTION

The whole idea of spatial prediction methodology might be well illustrated by Figure A.1. In spatial prediction, sample data are collected from the actual surface $z(x), x \in \mathcal{D}$ and the surface are reproduced based on the observational data $\{(x_i, z(x_i))\}_{i=1}^{N}$. The
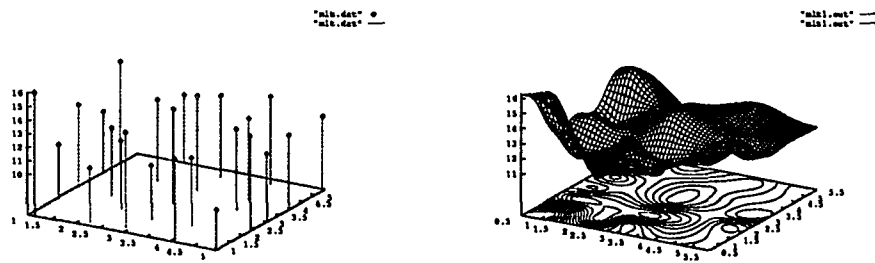

Figure A.1:   2-D Observational Data and Reproduced Surface

performance of spatial predictor can be measured by how close the reproduced surface is to the actual one. Since the actual surface is generally not known, they minimize the mean squared error in kriging and the upper bound of the error committed by the prediction in splines.

APPENDIX B. CARDINAL FUNCTIONS (HAT FUNCTIONS)

As discussed earlier, an interpolant has a representation in terms of the scalar valued function $\lambda(\cdot)$, which has a special property and shape. One can get an interpolant by specifying $\lambda(\cdot)$. In spatial prediction, the usual way to get the function $\lambda(\cdot)$ is to specify $\kappa(\cdot)$ or $\tau(\cdot)$ first and derive $\lambda(\cdot)$ from the following algebraic relations. Define

$$\lambda_i = \lambda_{0i} = \lambda(\boldsymbol{x}, \boldsymbol{x}_i) \ , \ \tau_i = \tau_{0i} = \tau(\boldsymbol{x}, \boldsymbol{x}_i) \text{ and } \kappa_i = \kappa_{0i} = \kappa(\boldsymbol{x}, \boldsymbol{x}_i) \ , \ i = 1, \cdots, N \text{ (B.1)}$$

and let $\boldsymbol{\lambda} = [\lambda_{01}, \lambda_{02}, \cdots, \lambda_{0N}]'$, $\boldsymbol{\tau} = [\tau_{01}, \tau_{02}, \cdots, \tau_{0N}]'$ and $\boldsymbol{\kappa} = [\kappa_{01}, \kappa_{02}, \cdots, \kappa_{0N}]'$, as before. Now let K and T consist of column vectors of $\{T_1, \cdots, T_N\}$ and $\{K_1, \cdots, K_N\}$, respectively, i.e.,

$$T = [T_1, T_2, \cdots, T_N] = \begin{bmatrix} \tau_{11} & \cdots & \tau_{1N} \\ \vdots & \cdots & \vdots \\ \tau_{N1} & \cdots & \tau_{NN} \end{bmatrix} \tag{B.2}$$

$$K = [K_1, K_2, \cdots, K_N] = \begin{bmatrix} \kappa_{11} & \cdots & \kappa_{1N} \\ \vdots & \cdots & \vdots \\ \kappa_{N1} & \cdots & \kappa_{NN} \end{bmatrix} \tag{B.3}$$

If onee let $K\boldsymbol{\lambda} = \boldsymbol{\kappa}$, then by Cramer's Rule,

$$\lambda_{0i} = \frac{det(K_1, \cdots, K_{i-1}, \boldsymbol{\kappa}, K_{i+1}, \cdots, K_N)}{det(K_1, \cdots, K_{i-1}, K_i, K_{i+1}, \cdots, K_N)} = \begin{cases} 1 & , \ \boldsymbol{\kappa} = K_i \\ 0 & , \ \boldsymbol{\kappa} \neq K_j, j \neq i \end{cases} \tag{B.4}$$

for $j = 1, \cdots, N$. Consequently, it follows that

$$
\lambda_{ij} = \lambda(x_i, x_j) = \delta_{ij} = \begin{cases} 1 & , \; x_i = x_j \\ 0 & , \; x_i \neq x_j \end{cases} \tag{B.5}
$$

where $i, j = 1, \cdots, N$. In the same way, given $\tau(\cdot)$, $\boldsymbol{\lambda}$ is obtainable from the relation $T\boldsymbol{\lambda} = \boldsymbol{\tau}$.
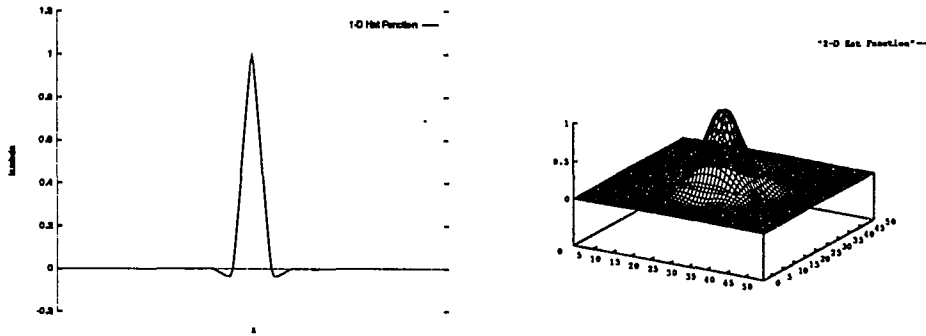

Figure B.1: 1-D and 2-D Hat Function

The $\boldsymbol{\lambda}$ is a vector of functions of x, since $\boldsymbol{\kappa}$ is a vector of functions of x. The functions $\lambda(\cdot)$ are called *Cardinal (Hat)* function in splines and *Weignt Function* in kriging.

APPENDIX C.   EIGENVALUES AND QUADRATIC FORMS

Let Q be an $N \times N$ square matrix of full rank. A scalar λ and a non zero vector x satisfying the equation $Qx = \lambda x$ are said to be, respectively, an eigenvalue and eigenvector of Q. In order that λ be an eigenvalue it is clear that it is necessary and sufficient for $Q - \lambda I$ to be singular, and hence $det(Q - \lambda I) = 0$. This last result, when expanded, yields an N-th order polynomial equation which can be solved for N (possibly nondistinct) complex roots λ which are the eigenvalues of Q. Assume that Q is symmetric. Then the following holds: 1) the eigenvalues of Q are real, 2) Eigenvectors associated with distinct eigenvalues are orthogonal and hence form an orthogonal basis.

Let $\{v_1, v_2, \cdots, v_N\}$ be an orthogonal basis as in 2). By normalization, we get an orthonormal basis $\{u_1, u_2, \cdots, u_N\}$. Defining matrix $B = [u_1, u_2, \cdots, u_N]$, we have that $B'B = I$ and hence $B' = B^{-1}$. Observe that

$$B^{-1}QB = B'QB = B'[Qu_1, \cdots, Qu_N] = B'[\lambda_1 u_1, \cdots, \lambda_N u_N] \qquad (C.1)$$

and hence the diagonalized matrix can be represented by

$$B^{-1}QB = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} \qquad (C.2)$$

A symmetric matrix Q is said to be positive definite if the quadratic form $x'Qx$ is positive for all nonzero vector x. For any x let $y = Q^{-1}x$ where Q is defined as above. Then

$\mathbf{x}'Q\mathbf{x} = \mathbf{y}'B'QB\mathbf{y} = \sum_{i=1}^{N} \lambda_i y_i^2$. Since the $y_i's$ are arbitrary, Q is positive definite (or positive semidefinite) if and only if all eigenvalues of Q are positive (or nonnegative).

It can be easily shown that a positive definite matrix Σ has positive definite square roots $\Sigma^{1/2}$ satisfying $\Sigma^{1/2} \cdot \Sigma^{1/2} = \Sigma$. For this, we use B as above and define

$$\Sigma^{1/2} = B^{-1} \cdot \begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_N^{1/2} \end{bmatrix} \cdot B \equiv T \qquad (C.3)$$

Note that, in (C.3), $\lambda_i^{1/2}$ may be chosen to be of either sign, so that T may be chosen as positive definite or negative definite or root of mixed signs, but in any event T will be invertible.