ESTIMATION OF RANDOM FIELDS
FROM NETWORK OBSERVATIONS

BY

ANDRÉ FRANÇOIS CABANNES

TECHNICAL REPORT NO. 26

JUNE 1979

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH (SIMS)

PREPARED UNDER SUPPORT TO SIMS FROM

DEPARTMENT OF ENERGY (DOE)

ROCKEFELLER FOUNDATION

SLOAN FOUNDATION

ENVIRONMENTAL PROTECTION AGENCY (EPA)

NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

zp

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ACKNOWLEDGEMENTS

I wish to express my gratitude to my adviser, Professor Paul Switzer, for proposing the problem to me and for his efficient guidance during the past year. My work benefited from discussions with Charles Stein and Peter Bloomfield. Further thanks are due to Ingram Olkin and Ray Faith who read the manuscript and made useful suggestions.

I would also like to take this opportunity to acknowledge my indebtedness to my teachers Persi Diaconis and David Siegmund.

The credit for fast and expert typing of the dissertation goes to Judi Davis and Wanda Edminster.

TABLE OF CONTENTS

	Page
Acknowledgements	iii
1 INTRODUCTION	1
1.1 Definition and Examples of Random Fields	1
1.2 The Problem of Estimation of Random Fields and the Problem of Network Design	2
1.3 Review of the Literature on Related Problems	3
1.4 Description of the Dissertation	4
2 ELEMENTARY PROPERTIES OF UNBIASED ESTIMATION OF RANDOM FUNCTIONS	10
2.1 Preliminaries	10
2.2 Interpretation of the B.L.U.E. $\hat{Z}(x)$ and Its Risk	13
2.3 A Class of Covariance Functions Which Yield the Same Estimator $\hat{Z}(x)$	15
3 FURTHER PROPERTIES OF THE B.L.U.E. $\hat{Z}(x)$	20
3.1 Bayes Estimation with Respect to the Parameter a	21
3.2 Admissibility of $\hat{Z}$	24
3.3 Uniformly Minimum Variance Unbiased Estimation	27
4 A NONLINEAR ESTIMATOR OF $Z(x)$	28
4.1 Biased Linear Estimator $\hat{Z}_a$	28
4.2 Nonlinear Estimator Z^*	30
4.3 A Numerical Example Comparing $\hat{Z}$ and Z^*	43

	Page
5 ESTIMATION OF A MAP OF Z USING THE B.L.U.E. AND THE PROBLEM OF NETWORK DESIGN	46
5.1 Preliminaries	46
5.2 An Updating Theorem	48
5.3 Possible Criteria for Assessing the Quality of a Network	54
5.4 Study of Criterion $e^{K^{-1}}$ when the Random Field Z is Weakly Stationary	56
5.5 An Estimation Paradox	61
5.6 Study of Criterion $\int_A \text{Var}(\hat{Z}(x) - Z(x))dx$	66
5.7 A Suggestion in Network Design for Estimation of Smooth Random Fields	79
5.8 A Numerical Example of Estimation of a Smooth Random Field on Artificial Data	80
6 STEIN-LIKE ESTIMATOR OF A MAP OF Z	96
6.1 A Theorem on Estimation of $C\theta$ with $X \sim N(\theta, I_p)$. .	96
6.2 Application to the Estimation of a Map	101
6.3 Further Comments and an Example	105
7 APPENDIX	110
7.1 Singularity of the Matrix $[\ x_i - x_j\ ^2]$	110
7.2 A Class of Covariance Function in R^2 Vanishing at a Finite Distance	114
REFERENCES	121

CHAPTER 1

INTRODUCTION

1.1. Definition and Examples of Random Fields

Many applied sciences study phenomena which present a random character in their variations both in time and in space.

Random phenomena varying only in time are the subject of the theory of time series, understood in the broad sense of being either discrete or continuous in time. The variation of the value of a stock, or the variation of the temperature at a given place over time are such examples.

Random phenomena varying in space, or in time and space, are generalizations represented by random fields, which are stochastic processes, i.e. collections of random variables, whose indexing set has more than one dimension. Examples are ore grade at each three dimensional point of an ore body, or the variation of temperature or pollutant concentrations both in time and space over a geographic area.

Several definitions of a random field are possible according to the level of generality one wants. We adopt the following

Definition: A random field is a collection of real-valued random variables indexed in the m -dimensional real vector space $\mathbb{R}^m$. We denote it

$$\{Z(x), x \in \mathbb{R}^m\}.$$

Further examples are:

- the thickness of a layer of coal over a two-dimensional area (dimension $m=2$),
- the diameter of a cotton thread at each point along the thread ($m=1$),
- the local density of trees over a geographic area ($m=2$),
- the density of particles in a liquid, and its evolution in time if the liquid is turbulent ($m=4$),
- the rainfall at each point of a region, after a storm ($m=2$).

Note that the definition considers real valued random variables: the wind vector measured at each point over a region falls outside its range. However, the modulus of the wind vector, that is the wind speed, fits the definition.

1.2. The Problem of Estimation of Random Fields and the Problem of Network Design

When one has observed a random field Z at some points $x_1, x_2, \dots, x_n$ and has recorded the n values

$$Z(x_1), Z(x_2), \dots, Z(x_n),$$

called the network observations, a natural problem is to estimate Z at points where we do not have observations. (Another term is to predict Z .)

In this dissertation a first part deals with this problem in an abstract setting, with the dimension m arbitrary; in a second part, more applied, we consider the estimation of a spatial two-dimensional

random field. The problem is then one of constructing an estimated map of Z over a geographic area A . The points $x_1, x_2, \dots, x_n$ are called the stations. For a given network of stations the quality of a map depends on the method of estimation. But for a given method of estimation the quality of a map depends on the choice of location for the stations. This is the problem of network design.

We shall be concerned with both the study of methods of estimation and the problem of network design.

1.3. Review of the Literature on Related Problems

The problem of estimation of a random field has been considered by Mathéron (1965), (1973). He and his colleagues developed a method that they dubbed "Kriging" to do estimation from a single realization, with very weak distributional hypotheses on the random field. For expository papers see Delfiner and Delhomme (1973), Delfiner (1975), and Delhomme (1978). See also Starks et al. (1978) for a detailed application to mining data.

The most studied problem on random fields has been the estimation of mean areal quantities, see Matern (1960), Bras and Rodriguez-Iturbe (1976a), (1976b), Bras and Colon (1978), Jones et al. (1978a), (1978b). Important examples of application of this problem are:

- in forestry, the evaluation of the quantity of wood in a forest,
- in hydrology, river flood forecast and control of dam level from rainfalls.

The problem of network design, which is intimately related to the previous ones, has been considered in the works cited above; but see also the more specific references Gribik, Kortanek, and Sweigart (1976), Rodriguez-Iturbe and Mejia (1973), Veneziano et al. (1977).

Finally, let us mention that Dalenius, Hajek, and Zubrzycki (1961) solved in particular cases the problem of optimal pattern for a regular network (that is, grid, or vertices of equilateral triangles, or vertices of hexagons, ...). For the general case, referring to the work of Matern (1960), they write (p. 142): "in all cases [of correlation functions] the triangular net proved to be somewhat better than the other ones."

1.4. Description of the Dissertation

The dissertation is divided into seven chapters. The first chapter is the introduction. Chapters 2, 3, and 4 form a first part in which we study the estimation of Z at a fixed point. In Chapter 3 we show some optimal properties of a usual estimator. In Chapter 4 we propose a new estimator that, in some cases, would be preferred. Chapters 5 and 6 form a second part in which we consider the estimation of the map of Z over an area A , and the problem of network design. In Chapter 5 we prove, in the second section, a theorem that is useful for speeding up the computation of estimated maps. We also obtain a result which leads to a new suggestion in the design of networks for estimation of smooth random fields. This suggestion should find practical applications in monitoring atmospheric phenomena. Chapter 6 shows how the properties of the point-estimator studied in Chapter 3 carry over to the case of map-estimation. Finally, Chapter 7 is an appendix containing a matrix theorem and a study of covariance functions, both related to Chapter 5.

We now give a detailed account of the contents of the dissertation and set some generalities that will not be repeated later.

The criterion we use to measure the quality of an estimator $\hat{Z}(x)$ of $Z(x)$ at a fixed point x , based on the network observations, is the mean squared error

$$E(\hat{Z}(x) - Z(x))^2.$$

The best estimator then is the conditional expectation $E\{Z(x) | Z(x_1), Z(x_2), \dots, Z(x_n)\}$; its computation requires the knowledge of the entire distribution of Z . However, it is often more reasonable to only partially specify the distribution of the random field, or its first two moments.

Our assumptions are that the random field Z has a mean function $m(x) = EZ(x)$, and a covariance function $K(x,y) = \text{Cov}(Z(x), Z(y))$ which are finite for all points x and y . Throughout the covariance function is completely specified, while the mean function is specified up to a vector parameter a entering its definition linearly.

Under these assumptions the mean squared error becomes also a function of the unspecified vector a . When there is no other distributional indeterminacy (and x is fixed) we can write

$$E(\hat{Z}(x) - Z(x))^2 = R(a, \hat{Z}).$$

Two estimators $\hat{Z}$ and Z^* are compared through their risk functions $R(a, \hat{Z})$ and $R(a, Z^*)$.

Chapter 2 summarizes some basic known results. A preliminary section sets the notation, reviews some elementary properties of mean squared estimation, and sketches the derivation of the best linear unbiased estimator $\hat{Z}(x)$ of the form

$$(1) \quad \hat{Z}(x) = \lambda_1 Z_1(x) + \lambda_2 Z(x_2) + \dots + \lambda_n Z(x_n) .$$

For notational convenience, we do not indicate the dependence of the λ_i 's on x . Section 2.2 shows some alternative forms of $\hat{Z}(x)$ and of its risk, and interprets them. Section 2.3 gives a matrix proof of a theorem showing how much $\hat{Z}(x)$ depends on the covariance function $K(x,y)$. In this chapter we need not specify anything more than the two first moments of Z .

In Chapters 3 and 4 to study further the properties of $\hat{Z}$ and to compare it to another estimator we need a more complete specification of the distribution of the random field Z : to the above assumptions we add the assumption that Z is Gaussian.

Chapter 3 contains a section on Bayes estimation when the parameter a has a multivariate normal prior distribution; it is shown that $\hat{Z}(x)$ is extended Bayes and minimax. Then we prove that $\hat{Z}(x)$ is admissible and has uniformly minimum variance among all unbiased estimators of $Z(x)$. The proofs of these results are essentially exercises in estimation of random variables, and are done by eventual reduction to classical results in estimation of parameters.

In Chapter 4 we study what happens when we release the unbiasedness requirement on the estimator given by (1). It leads to a natural nonlinear estimator $Z^*(x)$ of $Z(x)$. We study $Z^*(x)$ in detail and show that its risk beats that of $Z(x)$ when the parameter a lies in a certain region. An example is given, illustrating when Z^* should be preferred to $\hat{Z}$. This completes the first part of the dissertation.

In Chapter 5 the best linear unbiased estimator is chosen to draw estimated maps of Z . The variance of $\hat{Z}(x) - Z(x)$ as a function of x and $x_1, x_2, \dots, x_n$ is then the function of interest - it does not depend on a . Figure 1a shows $\text{Var}(\hat{Z}(x) - Z(x))$ as a function of x for a given network of stations. In Section 5.2 we prove a basic theorem updating $\hat{Z}(x)$ and $\text{Var}(\hat{Z}(x) - Z(x))$ when a station is added or deleted. Considerable computation can be saved using this theorem to update the estimated map of Z when a new station is added. Theorem 5.2.1 is also useful theoretically to establish certain properties of networks presented in Section 5.6. In Sections 5.3 through 5.6 we study several criteria to assess the quality of a network. Section 5.5 presents an estimation paradox about unbiased estimation. The main result of Section 5.6 is that if the random field Z is differentiable in quadratic mean, then the variance of the estimation error is not continuous in its arguments $x_1, x_2, \dots, x_n$. We give a substantial interpretation to Theorem 5.6.1 (which includes that result) and make further conjectures. In Section 5.7 we show the implication of Theorem 5.6.1 in the design of networks for estimation of smooth random fields, see Figure 1 below. This figure is taken from Section 5.8 which presents a thorough illustration of the results of Section 5.6 and the discussion of Section 5.7.

In Chapter 6 we show that the B.L.U.E. (which was shown in Section 3.2 to be admissible when used to estimate $Z(x)$ at a fixed point) is no longer admissible in general when used to estimate point-wise the whole map of Z over an area A . We prove an abstract

1a

1b

Figure 1. Example of improvement of a network of six stations for estimation of a smooth random field. Figure 1a represents the level curves of the function $\text{Var}(\hat{Z}(x) - Z(x))$ in the standard use of the network; figure 1b shows the improvement realized by taking three measurements close to each other instead of one at a single station.

theorem pertaining to the theory of Stein estimation, and then apply it to construct a Stein-like estimator of a whole map of Z that is uniformly better than the B.L.U.E. A numerical example is given in Section 6.3.

In the appendix, Chapter 7, the first section contains a proof of the singularity of the $n \times n$ matrix $[\|x_i - x_j\|^2]$ constructed from n points $x_1, x_2, \dots, x_n$ in R^2 provided $n \geq 5$. This result puts a limitation on the use of the network criterion considered in Section 5.4. The second section of the appendix presents an incomplete study of a class of isotropic covariance functions vanishing at a finite distance. This study was developed in an attempt to settle a conjecture related to the estimation paradox discussed in Section 5.5.

It should be indicated that the computations in the first part of the dissertation are not essentially related to the fact that Z is a random field. In particular, the calculations involved in studying the estimator $\hat{Z}(x)$ for a fixed x truly belong to the subject of generalized least squares prediction, see for example, Goldberger (1962) for a concise introduction.

Convention for numbering: In the following chapters the formulas, lemmas, theorems, and figures are each numbered from 1 to n in each section. When a formula of one section is referred to in another section, its number is preceded by the number of its section: for instance, formula (1) is the first numbered formula of the present section, whereas formula 2.1.7 is the seventh numbered formula of Section 2.1. Because they are less numerous and have more importance, the theorems and lemmas are directly given numbers preceded by the number of their sections.

CHAPTER 2

ELEMENTARY PROPERTIES OF UNBIASED ESTIMATION OF RANDOM FUNCTIONS

Before considering the estimation of a map of Z over an area A , from the network observations, we need first to consider the problem of estimating $Z(x)$ at a single point x . This chapter deals with the best linear unbiased estimator of $Z(x)$.

2.1. Preliminaries

We specify the dependence of the mean function $m(x)$ of the random field to be linear in the unknown vector parameter a :

$$(1) \quad m(x) = a_1 f_1(x) + a_2 f_2(x) + \dots + a_p f_p(x) .$$

It is abbreviated $m(x) = f'_x a$. This parametric mean function represents in a uniform manner many cases. For example

1) $m(x) = m$, unknown constant; when the covariance function $K(x,y)$ is only a function of $y-x$, this corresponds to weak stationarity of Z .

2) When $x \in R^2$ and its coordinates are $(u(x), v(x))$,

$$m(x) = a_1 + a_2 u(x) + a_3 v(x) ,$$

that is the surface representing the mean function is a plane.

3) Examples 1 and 2 generalize obviously to mean functions of higher degree in the coordinates of x .

4) When $x \in R$,

$$m(x) = a_1 \sin(\omega x) + a_2 \cos(\omega x) ,$$

that is the mean function is a sinusoid of known frequency, but unknown amplitude and unknown phase.

As said, the covariance function $K(x,y)$ is entirely specified.

In the sequel we shall use the following condensed notation

$$\underline{Z} = \begin{pmatrix} Z(x_1) \\ Z(x_2) \\ \vdots \\ Z(x_n) \end{pmatrix} = \text{the network observations}$$

K = covariance matrix of $\underline{Z}$

k_x = vector of covariances between $Z(x)$ and $Z(x_i)$, $i=1, 2, \dots, n$

$$E^Z Z(x) = E\{Z(x) | Z(x_1), Z(x_2), \dots, Z(x_n)\}$$

$$F' = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \dots & f_p(x_1) \\ \vdots & & & \\ f_1(x_n) & f_2(x_n) & \dots & f_p(x_n) \end{bmatrix}$$

thus $E\underline{Z} = f'a$ (in harmony with $EZ(x) = f'_x a$). Finally, let

$$(2) \quad \lambda' \underline{Z} = \lambda_1 Z(x_1) + \lambda_2 Z(x_2) + \dots + \lambda_n Z(x_n) .$$

Before deriving the best linear unbiased estimator of $Z(x)$, in order to avoid confusion later let us recall two basic properties of mean squared estimation.

Only when the random field Z is Gaussian the conditional expectation of $Z(x)$ given the observations has the form

$$(3) \quad E^Z Z(x) = m(x) + k'_x K^{-1}[\underline{Z} - E\underline{Z}] .$$

This formula is unusually simple; in general $E^Z Z(x)$ is complicated or hard to find. However, provided only that the distribution of Z has its two first moments finite, the estimator $\tilde{Z}(x) = \lambda_0 + \lambda' \underline{Z}$ which minimizes the mean squared error $E(\tilde{Z}(x) - Z(x))^2$ has the same form

$$(4) \quad \tilde{Z}(x) = m(x) + k'_x K^{-1}[\underline{Z} - E\underline{Z}] .$$

Formula (4) is true for an arbitrary mean function. When $m(x)$ has the parametric form given by (1), in particular formula (4) depends on a .

Let us now derive the best linear unbiased estimator

$\hat{Z}(x) = \lambda_0 + \lambda' \underline{Z}$ of $Z(x)$. It satisfies

$$(5a) \quad E(\hat{Z}(x) - Z(x))^2 \text{ minimum}$$

$$(5b) \quad E\hat{Z}(x) = f'_x a, \text{ for all } a .$$

Unbiasedness condition (5b) implies

$$\lambda_0 + \lambda' F' a = f'_x a, \text{ for all } a ,$$

hence

$$\lambda_0 = 0 \text{ and } F\lambda = f_x .$$

Therefore the best linear unbiased estimator of the form $\lambda_0 + \lambda' \underline{Z}$ is also the best linear unbiased estimator of the form $\lambda' \underline{Z}$.

To transform condition (5a), dropping the variable x which is fixed, we write

$$E(\hat{Z}-Z)^2 = \text{Var}(\hat{Z}-Z) = \text{Var}(\lambda'Z-Z) = \lambda'K\lambda - 2\lambda'k + K(x,x) .$$

So conditions (5a) and (5b) become

$$(6a) \quad \lambda'K\lambda - 2\lambda'k \text{ minimum}$$

$$(6b) \quad F\lambda = f .$$

This system is classical; its solution can be found, for example, in Goldberger (1962). The optimal λ is

$$(7) \quad \lambda = K^{-1}k_x - K^{-1}F'(FK^{-1}F')^{-1} FK^{-1}k_x + K^{-1}F'(FK^{-1}F')^{-1} f_x$$

and the mean squared error using the B.L.U.E. is

$$(8) \quad \begin{aligned} \text{Var}(\hat{Z}(x) - Z(x)) &= K(x,x) - k_x'K^{-1}k_x \\ &+ (FK^{-1}k_x - f_x)' (FK^{-1}F')^{-1} (FK^{-1}k_x - f_x) . \end{aligned}$$

Because the following quantities have an important role and will appear often in the subsequent chapters, we adopt the notation

$$\begin{aligned} G &= FK^{-1}F' , \\ \phi_x &= FK^{-1}k_x - f_x . \end{aligned}$$

2.2. Interpretation of the B.L.U.E. $\hat{Z}(x)$ and its Risk

Since $\underline{Z} = F'a + \varepsilon$, where $E\varepsilon = 0$, $\text{Cov } \varepsilon = K$, the generalized least squares estimator of a based on $\underline{Z}$ is

$$(1) \quad \hat{a} = (FK^{-1}F')^{-1} FK^{-1} \underline{Z} \\ = G^{-1} FK^{-1} \underline{Z} .$$

It is also the maximum likelihood estimator of a under the normality assumption $\varepsilon \sim N(0, K)$.

Goldberger made the observation that the B.L.U.E. obtained from formula 2.1.7 can be rewritten

$$(2) \quad \hat{Z}(x) = f'_x \hat{a} + k'_x K^{-1} [\underline{Z} - F' \hat{a}] .$$

If a were specified in advance then, from 2.1.4, the best linear estimate $\tilde{Z}(x)$ would have been

$$(3) \quad \tilde{Z}(x) = f'_x a + k'_x K^{-1} [\underline{Z} - F' a] .$$

So the B.L.U.E. $\hat{Z}(x)$ has the same form as $\tilde{Z}(x)$ with a replaced by the G.L.S. estimator $\hat{a}$.

Notes:

1) The expectation, and covariance matrix, of $\hat{a}$ given by (1) are $E\hat{a} = a$ and $\text{Cov } \hat{a} = G^{-1}$.

2) $f'_x \hat{a}$ is the B.L.U.E. of $f'_x a = m(x)$.

3) We see that $\hat{Z}(x)$ is the sum of two terms, one estimating $EZ(x)$ and the other estimating the deviation of $Z(x)$ from its mean.

4) In least square estimation with no unknown parameter it is a classical fact that

$$\text{Cov}(\underline{Z}, Z(x) - \tilde{Z}(x)) = 0 \quad (\text{zero vector})$$

whereas, with the unknown parameter a , and using $\hat{Z}(x)$, we have

$$\text{Cov}(\underline{Z}, Z(x) - \hat{Z}(x)) = F'G^{-1} \phi_x.$$

Let us now interpret the risk of $\hat{Z}(x)$. With the notations $\sigma^2 = K(x,x) = \text{Var } Z(x)$, $R = (k'K^{-1}k/\sigma^2)^{1/2}$ the multiple correlation coefficient between $Z(x)$ and the observations, and ϕ for ϕ_x , formula 2.1.8 on page 13 becomes

$$(4) \quad \text{Var}(\hat{Z}(x) - Z(x)) = \sigma^2(1-R^2) + \phi'G^{-1}\phi.$$

On the other hand, it is known that $\text{Var}(\tilde{Z}(x) - Z(x)) = \sigma^2(1-R^2)$. Hence we see that the term $\phi'G^{-1}\phi$ represents the price we pay for not knowing a , while requiring unbiasedness of $\hat{Z}(x)$.

The term $\phi'G^{-1}\phi$ is very interesting. In Chapters 4 and 6 we show how this added term in the mean squared error may be reduced.

2.3. A Class of Covariance Functions Which Yield the Same Estimator

$\hat{Z}(x)$

We have assumed that the covariance function $K(x,y)$ was entirely specified; that is, in applications, entirely known. Therefore, it is a matter of importance to know how dependent is the estimator $\hat{Z}(x)$ on the function $K(x,y)$.

If $K(x,y)$ is multiplied by a positive constant it is easy to see that $\hat{Z}(x)$ is unchanged.

Suppose that $K(x,y) = \sigma^2(1-\alpha N(x,y))$ and that it is a covariance function for all α 's in an interval (see the example below). If, in the mean function

$$m(x) = a_1 f_1(x) + a_2 f_2(x) + \dots + a_p f_p(x) ,$$

the function $f_1(x)$ is identically 1, then it is elementary to show, using equations 2.1.6a and 2.1.6b, that $\hat{Z}(x)$ does not depend on α .

Example: $K(x,y) = \sigma^2(1-\alpha(1-e^{-\|x-y\|^2})) = \sigma^2(1-\alpha + \alpha e^{-\|x-y\|^2})$ is a covariance function for all $\alpha \in [0,1]$, and all $\sigma^2 > 0$.

We shall demonstrate what is the most general condition on the $f_i(x)$'s for which the B.L.U.E. $\hat{Z}(x)$ does not depend on α .

Theorem 2.3.1. When $K(x,y) = \sigma^2(1-\alpha N(x,y))$ the estimator $\hat{Z}(x)$ is independent of α for all x if and only if the vector

$$e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

belongs to the vector space spanned by the columns of F' .

Theorem 2.3.1 will be proved as a consequence of the lemmas below; its proof comes on page 19 after their proofs.

Corresponding to K and k , let

$$N = [N(x_i, x_j)] \text{ and } n = [N(x_i, x)] .$$

Lemma 2.3.1. When $K(x,y) = \sigma^2(1-\alpha N(x,y))$ the general form of the coefficients of the estimator $\hat{Z}(x)$ is

$$(1) \quad \lambda = N^{-1}n - N^{-1}F'(FN^{-1}F')^{-1} [FN^{-1}n - f] + r .$$

To express the remainder term r and prove lemma 1 we need the following notation

$$\begin{aligned}
\text{Let } H &= (FN^{-1}F')^{-1} \\
L &= FN^{-1}e \\
\psi &= FN^{-1}n - f \\
\xi &= 1 - e'N^{-1}n + L'H\psi \\
\eta &= \alpha - e'N^{-1}e + L'HL,
\end{aligned}$$

then

$$r = \frac{\xi}{\eta} N^{-1} [F'HFN^{-1}e - e].$$

$$\text{Also, let } I = e'N^{-1}e \text{ and } \rho = \frac{e'N^{-1}n - 1}{e'N^{-1}e - \alpha}.$$

Lemma 2.3.2. If B is a symmetric nonsingular matrix and x is a column vector of corresponding size, then

$$[B + xx']^{-1} = B^{-1} - \frac{B^{-1}xx'B^{-1}}{1 + x'B^{-1}x},$$

provided $x'B^{-1}x \neq -1$.

A proof is obtained by direct multiplication.

Proof of Lemma 2.3.1. We start from the expression for λ given by formula 2.1.7. It can be written

$$(2) \quad \lambda = K^{-1}k - K^{-1}F'G^{-1}\phi.$$

We shall substitute the new form of K , k , F , G and ϕ in formula (2).

Since $K = \sigma^2(ee' - \alpha N)$, by lemma 2.3.2 we get

$$K^{-1} = -\frac{1}{\alpha\sigma^2} \left(N^{-1} + \frac{N^{-1}ee'N^{-1}}{\alpha - e'N^{-1}e} \right)$$

and

$$FK^{-1}F' = -\frac{1}{\alpha\sigma^2} \left(FN^{-1}F' + \frac{FN^{-1}ee'N^{-1}F'}{\alpha - e'N^{-1}e} \right).$$

This last quantity is still of the form $B + xx'$, therefore

$$(FK^{-1}F')^{-1} = -\alpha\sigma^2 \left\{ (FN^{-1}F')^{-1} - \frac{(FN^{-1}F')^{-1} FN^{-1}ee'N^{-1}F' (FN^{-1}F')^{-1}}{\alpha - e'N^{-1}e + e'N^{-1}F' (FN^{-1}F')^{-1} FN^{-1}e} \right\}.$$

With the notations defined above it can be rewritten

$$(FK^{-1}F')^{-1} = -\alpha\sigma^2 \left(H - \frac{HLL'H}{\eta} \right).$$

On the other hand,

$$FK^{-1}k = -\frac{1}{\alpha\sigma^2} F \left(N^{-1} + \frac{N^{-1}ee'N^{-1}}{\alpha - I} \right) \sigma^2 (1 - \alpha n).$$

After calculations this becomes

$$FN^{-1}n - FN^{-1}e \left(\frac{e'N^{-1}n - 1}{e'N^{-1}e - \alpha} \right).$$

Therefore $\phi = FK^{-1}k - f = \psi - \rho L$. Then

$$(FK^{-1}F')^{-1}\phi = -\alpha\sigma^2 \left(H - \frac{HLL'H}{\eta} \right) (\psi - \rho L).$$

After calculations we get

$$= -\alpha\sigma^2 \left[H\psi - HL \frac{\xi}{\eta} \right].$$

Now we are ready to compute λ . Using formula (2) λ is rewritten

$$\lambda = -\frac{1}{\alpha\sigma^2} \left(N^{-1} + \frac{N^{-1}ee'N^{-1}}{\alpha - I} \right) \left(\sigma^2 (1 - \alpha n) + \alpha\sigma^2 \left(F'H\psi - F'HL \frac{\xi}{\eta} \right) \right)$$

$$= N^{-1} \left(n - F'H\psi + F'HL \frac{\xi}{\eta} \right) + \frac{1}{I-\alpha} N^{-1} e \left(1 - e'N^{-1}n + e'N^{-1}F'H\psi - e'N^{-1}F'HL \frac{\xi}{\eta} \right)$$

$$= N^{-1}(n - F'H\psi) + r$$

where

$$\begin{aligned} r &= N^{-1}F'HL \frac{\xi}{\eta} + \frac{1}{I-\alpha} N^{-1} e \left(1 - e'N^{-1}n + e'N^{-1}F'H\psi - e'N^{-1}F'HL \frac{\xi}{\eta} \right) \\ &= N^{-1}F'HL \frac{\xi}{\eta} + \frac{N^{-1}e}{I-\alpha} \left(\xi - L'HL \frac{\xi}{\eta} \right) \\ &= \frac{\xi}{\eta} N^{-1}[F'HL - e] . \end{aligned}$$

In conclusion

$$\lambda = N^{-1}[n - F'(FN^{-1}F')^{-1} (FN^{-1}n - f)] + r$$

where

$$r = \frac{\xi}{\eta} N^{-1}[F'(FN^{-1}F')^{-1} FN^{-1}e - e] . \quad \text{Q.E.D.}$$

Proof of Theorem 2.3.1. $\hat{Z}(x)$ is independent of α for all x if and only if the remainder term r in the expression (1) for λ is independent of α for all x . This holds when

$$F'(FN^{-1}F')^{-1} FN^{-1}e - e = 0 .$$

The matrix $F'(FN^{-1}F')^{-1} FN^{-1}$ is the N -orthogonal projection onto $\mathcal{A}(F')$, the vector space spanned by the columns of F' . Therefore the eigenvectors of $F'(FN^{-1}F')^{-1} FN^{-1}$ are exactly the vectors belonging to $\mathcal{A}(F')$. Q.E.D.

Note: Theorem 2.3.1 generalizes in a straightforward fashion as follows: the estimator $\hat{Z}(x)$ is unaffected by a change in K of the form $K \rightarrow K + uu'$ if and only if u is in $\mathcal{A}(F')$.

CHAPTER 3

FURTHER PROPERTIES OF THE B.L.U.E. $\hat{Z}(x)$

In this chapter we assume that the random field $\{Z(x), x \in \mathbb{R}^m\}$ is Gaussian.

Since $x \in \mathbb{R}^m$ is fixed, and we work only with the $n+1$ random variables $Z(x_1), Z(x_2), \dots, Z(x_n)$ and $Z(x)$, we can simplify the notation and write

$$\underline{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} \sim N(F'a, K) ,$$

$$Z(x) = Z \sim N(f'a, \sigma^2) ,$$

$$\text{Cov}(\underline{Z}, Z) = k ,$$

F, f, K, k and σ^2 are known,

$a \in \mathbb{R}^p$ is unknown .

This simplified setting shows clearly that $\{Z(x), x \in \mathbb{R}^m\}$ being a random field is not yet essential.

In the preceding chapter we presented the best linear unbiased estimator $\hat{Z}$ of Z based on $\underline{Z}$. One of its expressions is

$$\hat{Z} = f'\hat{a} + k'K^{-1}[\underline{Z} - F'\hat{a}] .$$

In this chapter we shall prove that $\hat{Z}$ is extended Bayes, minimax, admissible, and UMVUE.

3.1. Bayes Estimation with Respect to the Parameter a

Suppose that the vector parameter a has the following multivariate normal prior distribution

$$a \sim N(\alpha, \Gamma) .$$

Let us compute the posterior distribution of a given $\underline{z}$ because it will be useful for subsequent computations. This is obtained as an application of Bayes theorem for densities which says

$$(1) \quad f(a|\underline{z}) = \frac{f(\underline{z}|a) f(a)}{\int_{R^p} f(\underline{z}|a) f(a) da} .$$

We know that

$$f(\underline{z}|a) = \frac{1}{(2\pi)^{n/2}} |K|^{-1/2} \exp[-\frac{1}{2} (\underline{z}-F'a)' K^{-1}(\underline{z}-F'a)]$$

and

$$f(a) = \frac{1}{(2\pi)^{p/2}} |\Gamma|^{-1/2} \exp[-\frac{1}{2} (a-\alpha)' \Gamma^{-1}(a-\alpha)] .$$

Since $a|\underline{z}$ must be normally distributed, all we need to examine is the quadratic form in a in the exponential on the numerator in (1).

This is

$$\begin{aligned} & (\underline{z}-F'a)' K^{-1}(\underline{z}-F'a) + (a-\alpha)' \Gamma^{-1}(a-\alpha) \\ &= \underline{z}' K^{-1} \underline{z} - 2\underline{z}' K^{-1} F'a + a' G a + a' \Gamma^{-1} a - 2\alpha' \Gamma^{-1} a + \alpha' \Gamma^{-1} \alpha \end{aligned}$$

(where $G = FK^{-1}F'$)

$$(2) \quad = a'[G+\Gamma^{-1}]a - 2[\underline{z}'K^{-1}F' + \alpha'\Gamma^{-1}]a + \text{term independent of } a.$$

This must be of the form

$$(3) \quad (a-u)' A^{-1}(a-u) + \text{cst}$$

where $u = E(a|\underline{z})$ and $A = \text{Var}(a|\underline{z})$.

Comparing (2) and (3) we conclude that

$$(4) \quad E(a|\underline{z}) = (G+\Gamma^{-1})^{-1} [\Gamma^{-1}\alpha + FK^{-1}\underline{z}]$$

$$(5) \quad \text{Var}(a|\underline{z}) = (G+\Gamma^{-1})^{-1}.$$

Notes: 1) If Γ tends to zero, in the sense that $\max_i(i^{\text{th}} \text{ eigenvalue of } \Gamma) \rightarrow 0$, then $E(a|\underline{z}) \rightarrow \alpha$ and $\text{Var}(a|\underline{z}) \rightarrow 0$.

2) If $\Gamma \rightarrow +\infty$, in the sense that $\min_i(i^{\text{th}} \text{ eigenvalue of } \Gamma) \rightarrow +\infty$, then $E(a|\underline{z}) \rightarrow \hat{a}$ and $\text{Var}(a|\underline{z}) \rightarrow G^{-1}$.

Remember that the conditional distribution of $\hat{a}$ given a is $N(a, G^{-1})$.

Now we are ready to compute the Bayes rule for Z when $a \sim N(\alpha, \Gamma)$.

Theorem 3.1.1. Let $\hat{Z}_{\alpha, \Gamma}$ be the Bayes rule using mean squared error to estimate Z , when we observe $\underline{z}$, and when $a \sim N(\alpha, \Gamma)$. It has the form

$$\hat{Z}_{\alpha, \Gamma} = k'K^{-1}\underline{z} - \phi' E(a|\underline{z})$$

and its Bayes risk is

$$(6) \quad r(\hat{Z}_{\alpha, \Gamma}) = \sigma^2 - k'K^{-1}k + \phi'(G+\Gamma^{-1})^{-1} G(G+\Gamma^{-1})^{-1}\phi.$$

Proof. The Bayes rule $\hat{Z}_{\alpha, \Gamma}$ is the mean of the posterior distribution of Z given $\underline{z}$ after averaging over a .

$$\begin{aligned}\hat{Z}_{\alpha, \Gamma} &= E\{E(Z|\underline{Z}, a)|\underline{Z}\} = f'E(a|\underline{Z}) + k'K^{-1}\{\underline{Z} - F'E(a|\underline{Z})\} \\ &= k'K^{-1}\underline{Z} - \phi' E(a|\underline{Z})\end{aligned}$$

(where $\phi = FK^{-1}k - f$). Its risk is

$$r(\hat{Z}_{\alpha, \Gamma}) = E(\hat{Z}_{\alpha, \Gamma} - Z)^2 = \text{Var}(\hat{Z}_{\alpha, \Gamma} - Z)^2 + \{E(\hat{Z}_{\alpha, \Gamma} - Z)\}^2$$

(the term within curly brackets vanishes because $\hat{Z}_{\alpha, \Gamma}$ is unbiased)

$$\begin{aligned}(7) \quad &= \text{Var}(\hat{Z}_{\alpha, \Gamma} - Z)^2 \\ &= k'K^{-1}k + \phi' \text{Cov}\{E(a|\underline{Z})\}\phi - \\ &\quad 2 k'K^{-1} \text{Cov}\{\underline{Z}, E(a|\underline{Z})\}\phi - \\ &\quad 2 \text{Cov}(\hat{Z}_{\alpha, \Gamma}, Z) + \sigma^2.\end{aligned}$$

Without going into the details of the calculations, we can show using formulas (4) and (5) that

$$E(a|\underline{Z}) = (G+\Gamma^{-1})^{-1} \Gamma^{-1} \alpha + (G+\Gamma^{-1})^{-1} FK^{-1}\underline{Z}$$

$$\text{Cov}\{E(a|\underline{Z})\} = (G+\Gamma^{-1})^{-1} G(G+\Gamma^{-1})^{-1}$$

$$\text{Cov}\{E(a|\underline{Z}), \underline{Z}\} = (G+\Gamma^{-1})^{-1} F$$

$$\text{Cov}(\hat{Z}_{\alpha, \Gamma}, Z) = k'K^{-1}k - \phi' \text{Cov}\{E(a|\underline{Z}), Z\}$$

$$\text{Cov}\{E(a|\underline{Z}), Z\} = (G+\Gamma^{-1})^{-1} FK^{-1}k.$$

Substitution of these covariance expressions into formula (7) above yields

$$E(\hat{Z}_{\alpha, \Gamma} - Z)^2 = \sigma^2 - k'K^{-1}k + \phi' (G+\Gamma^{-1})^{-1} G(G+\Gamma^{-1})^{-1} \phi. \quad \text{Q.E.D.}$$

Corollary 3.1.1. The B.L.U.E. $\hat{Z}$ of Z is extended Bayes and minimax.

Proof. The risk of $\hat{Z}$ as a function of a is

$$R(a, \hat{Z}) = \sigma^2 - k'K^{-1}k + \phi'G^{-1}\phi.$$

This expression does not depend on a , therefore $\hat{Z}$ is an equalizer. When $\min(i^{\text{th}} \text{ eigenvalue of } \Gamma) \rightarrow +\infty$, $R(\hat{Z}_{\alpha, \Gamma})$, given by (6), tends to $R(a, \hat{Z})$. Therefore $\hat{Z}$ is extended Bayes, and since it is an equalizer it must be minimax.

3.2. Admissibility of $\hat{Z}$

The problem is to show that there is no estimator Z^* such that for all a

$$(1) \quad E(Z^* - Z)^2 \leq E(\hat{Z} - Z)^2$$

with strict inequality for at least one a .

Consider an estimator Z^* of Z based on $\underline{Z}$. We shall split the risk of Z^* into two parts using the following elementary fact:

Lemma 3.2.1. Any two random variables X and Y , possibly vector valued, with second moments satisfy

$$\text{Cov}\{Y, X - E(X|Y)\} = 0.$$

$$\begin{aligned} \text{Proof. } \text{Cov}\{Y, X - E(X|Y)\} &= E\{YX - YE(X|Y)\} - EY E\{X - E(X|Y)\} \\ &= E\{YX - E(YX|Y)\} = 0. \quad \text{Q.E.D.} \end{aligned}$$

Consider $E(Z^* - Z)^2 = E\{Z - E(Z|\underline{Z}) + E(Z|\underline{Z}) - Z^*\}^2$. (In order to avoid confusion, let us make it clear that the vector-parameter a is

no longer a random variable with some prior distribution; it is a fixed unknown quantity.) By Lemma 3.2.1 $Z - E(Z|\underline{Z})$ is uncorrelated with $\underline{Z}$; hence, by normality it is independent of $\underline{Z}$, and independent also of $E(Z|\underline{Z}) - Z^*$ because this is a function of $\underline{Z}$. Therefore the mean squared error splits into two parts:

$$(2) \quad E(Z^* - Z)^2 = E\{Z - E(Z|\underline{Z})\}^2 + E\{E(Z|\underline{Z}) - Z^*\}^2.$$

Next we observe that $E\{Z - E(Z|\underline{Z})\}^2 = \sigma^2 - k'K^{-1}k$. This does not depend on a . So the choice of Z^* affects only the second term of the right hand side of (2). Therefore we can modify (1): The problem is now to show that there is no $Z^* = g(\underline{Z})$ such that for all a

$$(3) \quad E\{E(Z|\underline{Z}) - Z^*\}^2 \leq E\{E(Z|\underline{Z}) - \hat{Z}\}^2$$

with strict inequality for at least one a .

$$E(Z|\underline{Z}) = f'a + k'K^{-1}[\underline{Z} - F'a],$$

hence

$$\begin{aligned} E\{E(Z|\underline{Z}) - g(\underline{Z})\}^2 &= E\{f'a + k'K^{-1}\underline{Z} - k'K^{-1}F'a - g(\underline{Z})\}^2 \\ &= E\{k'K^{-1}\underline{Z} - g(\underline{Z}) - \phi'a\}^2. \end{aligned}$$

The problem reduces to that of estimating a linear combination of the components of a using $\underline{Z}$.

Lemma 3.2.2. $\hat{a} = G^{-1}FK^{-1}\underline{Z}$ is sufficient for a .

Proof. It will be shown in Theorem 4.2.1 of Chapter 4 that

$$E(\underline{Z}|\hat{a}) = F'\hat{a}$$

and

$$\text{Cov}(\underline{Z}|\hat{a}) = K - F'G^{-1}F . .$$

Hence $\underline{Z}|\hat{a}$, which is normally distributed, has a distribution which does not depend on a . Q.E.D.

Therefore any admissible estimator of $\phi'a$ which is a function of $\underline{Z}$ is a function of $\hat{a}$. We saw that $\hat{a} \sim N(a, G^{-1})$; then it is a classical result that $\phi'\hat{a}$ is admissible for $\phi'a$, see Cohen (1965) for the whole class of admissible estimators of $\phi'a$.

The estimator $g(\underline{Z})$ such that

$$k'K^{-1}\underline{Z} - g(\underline{Z}) = \phi'\hat{a}$$

is

$$\begin{aligned} g(\underline{Z}) &= k'K^{-1}\underline{Z} - \phi'\hat{a} \\ &= k'K^{-1}\underline{Z} - k'K^{-1}F'\hat{a} + f'\hat{a} \\ &= f'\hat{a} + k'K^{-1}[\underline{Z} - F'\hat{a}] \end{aligned}$$

and this is the B.L.U.E. $\hat{Z}$ of Z . This completes the proof of the admissibility of $\hat{Z}(x)$ to estimate $Z(x)$, for x fixed.

In Chapter 6 we will show that when we want to estimate the random field $\{Z(x), x \in R^m\}$ at N grid points $y_1, y_2, \dots, y_N$, in order to construct an approximate map, the procedure which uses the individual (pointwise admissible) B.L.U.E.'s $\hat{Z}(y_1), \hat{Z}(y_2), \dots, \hat{Z}(y_N)$ is not admissible in general when we use the risk

$$\sum_{i=1}^N E(\hat{Z}(y_i) - Z(y_i))^2 .$$

3.3. Uniformly Minimum Variance Unbiased Estimation

In this section we show that $\hat{Z}$ is UMVUE. To prove it, from the formula 3.2.2 and the subsequent comment we see that it is sufficient to show that the term $E\{E(Z|\underline{Z}) - \hat{Z}\}^2$ is uniformly minimum, in a , among unbiased estimators of Z .

An unbiased estimator $g(\underline{Z})$ of Z must satisfy

$$Eg(\underline{Z}) = EZ = E\{E(Z|\underline{Z})\}.$$

Then $E\{f'a + k'K^{-1}[\underline{Z} - F'a] - g(\underline{Z})\} = 0$. Therefore it is such that $k'K^{-1}\underline{Z} - g(\underline{Z})$ is an unbiased estimator of $\phi'a$.

By the generalized version of Cramer-Rao inequality any unbiased estimator of $\phi'a$ has its variance bounded from below by $\phi' I(a)^{-1} \phi$ where $I(a)$ is the Fisher information matrix of $\underline{Z}$ on a . Its (i,j) element is

$$\begin{aligned} I_{ij}(a) &= E\left\{\left[\frac{\partial}{\partial a_i} \log f_a(\underline{Z})\right] \left[\frac{\partial}{\partial a_j} \log f_a(\underline{Z})\right]\right\} \\ &= E\left\{-\frac{\partial^2}{\partial a_i \partial a_j} \log f_a(\underline{Z})\right\} \end{aligned}$$

with

$$f_a(\underline{z}) = \frac{1}{(2\pi)^{n/2}} |K|^{-1/2} \exp\left\{-\frac{1}{2} (\underline{z} - F'a)' K^{-1} (\underline{z} - F'a)\right\}.$$

Then $I_{ij}(a)$ turns out to be the (i,j) element of $FK^{-1}F'$.

Hence the Cramer-Rao lower bound is $\phi'G^{-1}\phi$. This is attained by $\phi'\hat{a}$, because $E(\phi'\hat{a} - \phi'a)^2 = \phi' \text{Cov}(\hat{a})\phi = \phi'G^{-1}\phi$. Therefore $g(\underline{Z}) = k'K^{-1}\underline{Z} - \phi'\hat{a} = \hat{Z}$ makes $E\{E(Z|\underline{Z}) - \hat{Z}\}^2$, and hence $E(Z - \hat{Z})^2$, uniformly minimum among unbiased estimators of Z .

CHAPTER 4

A NONLINEAR ESTIMATOR OF $Z(x)$

Since we are still concerned with estimation of $Z(x)$ at a fixed point x , we keep the simplified notations introduced at the beginning of Chapter 3. In Section 4.1 we don't need normality assumption; in Sections 4.2 and 4.3 we do.

We saw in Section 2 of Chapter 2 that the price we pay for unbiased estimation, when we do not know the vector parameter a , is represented by the extra term $\phi'G^{-1}\phi$ in the mean squared error of the estimator $\hat{Z}$. A natural idea is to try to dispense with unbiasedness. This will lead to an estimator $\hat{Z}_a$, which depends on the unspecified vector parameter a . It is therefore of no use itself but the estimator denoted symbolically by $\hat{Z}_{\hat{a}}$, obtained by substituting $\hat{a}$ for a in $\hat{Z}_a$, turns out to be interesting; we will denote it Z^* .

4.1. Biased Linear Estimator $\hat{Z}_a$

Let us find the estimator $\hat{Z}_a$ of the form

$$\hat{Z}_a = \lambda_1 Z_1 + \lambda_2 Z_2 + \dots + \lambda_n Z_n = \lambda' \underline{Z}$$

which minimizes $E(\hat{Z}_a - Z)^2$ with no unbiasedness restriction with respect to the parameter a .

$$\begin{aligned} E(\hat{Z}_a - Z)^2 &= \text{Var}(\hat{Z}_a - Z) + \{E(\hat{Z}_a - Z)\}^2 \\ &= \sigma^2 - 2\lambda'k + \lambda'K\lambda + (f'a)^2 - 2(\lambda'F'a)(f'a) + (\lambda'F'a)^2. \end{aligned}$$

This mean squared error is minimum at

$$\lambda = \lambda_a = (K + F'aa'F)^{-1} (F'aa'f + k)$$

and its value then is

$$E(\hat{Z}_a - Z)^2 = \sigma^2 + (f'a)^2 - (k + F'aa'f)' (K + F'aa'F)^{-1} (k + F'aa'f) .$$

By lemma 2.3.2 we have

$$(K + F'aa'F)^{-1} = K^{-1} - \frac{K^{-1}F'aa'FK^{-1}}{1 + a'FK^{-1}F'a} .$$

After some algebra using this formula we obtain

$$(1) \quad \lambda_a = K^{-1}k - \frac{a'\phi K^{-1}F'a}{1 + a'Ga}$$

and, for the risk of $\hat{Z}_a = \lambda_a'Z$,

$$(2) \quad E(\hat{Z}_a - Z)^2 = \sigma^2 - k'K^{-1}k + \frac{(\phi'a)^2}{1 + a'Ga} .$$

Comments: 1) When we did not know a and derived the B.L.U.E.

$\hat{Z}$ we obtained $E(\hat{Z} - Z)^2 = \sigma^2 - k'K^{-1}k + \phi'G^{-1}\phi$.

Since $\hat{Z}_a$ is not subject to unbiasedness we must have

$\frac{(\phi'a)^2}{1+a'Ga} \leq \phi'G^{-1}\phi$, all a . It is easy and instructive to verify this directly: Let $a = G^{-1}\alpha$, then $(a'\phi)^2 = (\alpha'G^{-1}\phi)^2$. By the inequality of Cauchy-Schwarz

$$\begin{aligned} (\alpha'G^{-1}\phi)^2 &\leq (\alpha'G^{-1}\alpha) (\phi'G^{-1}\phi) \\ &= (a'Ga) (\phi'G^{-1}\phi) \\ &\leq (1 + a'Ga) (\phi'G^{-1}\phi) \dots \quad \text{Q.E.D.} \end{aligned}$$

2) When $a = 0$, $\hat{Z}_a = E(Z|Z)|_{a=0}$, whereas when $\|a\|^2 = a_1^2 + a_2^2 + \dots + a_p^2 \rightarrow +\infty$,

$$\frac{(\phi'a)^2}{1 + a'Ga} \sim \frac{(\phi'a)^2}{a'Ga} = \frac{a'G^{\frac{1}{2}} G^{-\frac{1}{2}} \phi \phi' G^{-\frac{1}{2}} G^{\frac{1}{2}} a}{\|G^{\frac{1}{2}} a\|^2}.$$

The maximum of this quantity, when $G^{\frac{1}{2}} a$ varies, is max eigenvalue of $G^{-\frac{1}{2}} \phi \phi' G^{-\frac{1}{2}} = \phi' G^{-1} \phi$.

Therefore the constant risk of the B.L.U.E. is actually the least upper bound on the risk of $\hat{Z}_a$.

3) Bias of $\hat{Z}_a$:

$$\begin{aligned} E\hat{Z}_a &= E\lambda'_a Z = \lambda'_a F'a = k'K^{-1}F'a - \frac{\phi'a a'FK^{-1}F'a}{1 + a'Ga} \\ &= \frac{1}{1 + a'Ga} [k'K^{-1}F'a + k'K^{-1}F'a a'Ga - k'K^{-1}F'a a'Ga + f'a a'Ga] \\ &= f'a + \frac{k'K^{-1}F'a - f'a}{1 + a'Ga}. \end{aligned}$$

So $\frac{\phi'a}{1+a'Ga}$ is the bias of $\hat{Z}_a$.

4) The reason why we did not start Section 4.1 with the formula $\hat{Z}_a = \lambda_0 + \lambda'_a Z$ is because we would have obtained the best linear estimator $f'a + k'K^{-1}[Z - F'a]$; and then substituting $\hat{a}$ instead of a , which is the next step, would just get us back to the B.L.U.E. $\hat{Z}$.

4.2. Nonlinear Estimator Z^*

The purpose of this section is to study the risk of the estimator Z^* obtained by using $\hat{a} = G^{-1}FK^{-1}Z$ instead of a in the expression for $\hat{Z}_a$ derived from formula 4.1.1. So let

$$\lambda^* = K^{-1} - \frac{\phi' \hat{a} K^{-1} F' \hat{a}}{1 + \hat{a}' G \hat{a}}$$

and define $Z^* = \lambda^* Z$. Thus Z^* is a nonlinear estimator of Z .

In order to compute $E(Z^* - Z)^2$ we shall use lemma 4.2.1 below, and Theorem 4.2.1 that follows it giving the expressions for $E(Z|\hat{a})$, $E(Z|\hat{a})$, and for the covariance matrix of $\begin{pmatrix} Z \\ Z \end{pmatrix}$ conditional on $\hat{a}$. The lemma is stated in a general form for more flexibility.

Lemma 4.2.1. Consider $X \sim N(\mu, \Sigma)$ of size $n \times 1$; suppose Σ is positive definite. Let M be a $p \times n$ matrix ($p \leq n$) of full rank, and v a $p \times 1$ vector. Then

$$E(X|MX = v) = \mu - \Sigma M'(M\Sigma M')^{-1}M\mu + \Sigma M'(M\Sigma M')^{-1}v$$

$$\text{Cov}(X|MX = v) = \Sigma(I - M'(M\Sigma M')^{-1}M\Sigma)$$

Proof. Define $Y = \Sigma^{-\frac{1}{2}}X$ (so $X = \Sigma^{\frac{1}{2}}Y$). Then $Y \sim N(\Sigma^{-\frac{1}{2}}\mu, I)$, and we have $MX = v \Leftrightarrow M\Sigma^{\frac{1}{2}}Y = v \Leftrightarrow M\Sigma^{\frac{1}{2}}P^{-1}PY = v$. The last equality holds for any choice of a nonsingular matrix P .

Define $Z = PY$; then $Z \sim N(P\Sigma^{-\frac{1}{2}}\mu, PP')$. Let us pick P such that $PP' = I$ and $M\Sigma^{\frac{1}{2}}P^{-1} = [A, 0]$ where A is $p \times p$ nonsingular. There are many such matrices P . A matrix P will satisfy these conditions if and only if it can be written $P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$, where P_1 is $p \times n$, P_2 is $(n-p) \times n$, and they are such that

$$\begin{cases} \mathcal{A}(P_1') = \mathcal{A}(\Sigma^{\frac{1}{2}}M') , \\ \mathcal{A}(P_2') \perp \mathcal{A}(P_1') , \\ PP' = I_n \end{cases}$$

where $\mathcal{R}$ (of a matrix) represents the vector space spanned by the columns of the matrix.

We choose

$$P_1 = (M \Sigma M')^{-\frac{1}{2}} M \Sigma^{\frac{1}{2}}$$

and P_2 being any completing orthogonal set of rows. Then we have $Z = P \Sigma^{-\frac{1}{2}} X \sim N(P \Sigma^{-\frac{1}{2}} \mu, I)$; hence $Z_1 = P_1 \Sigma^{-\frac{1}{2}} X$ and $Z_2 = P_2 \Sigma^{-\frac{1}{2}} X$ are independent. With this the conditioning event $MX = v$ can be simplified:

$$MX = v \Leftrightarrow M \Sigma^{\frac{1}{2}} P^{-1} Z = v \Leftrightarrow A Z_1 = v \Leftrightarrow Z_1 = A^{-1} v,$$

and the conditional expectation of X given $MX = v$ becomes easy to compute:

$$\begin{aligned} E(X|MX = v) &= E(\Sigma^{\frac{1}{2}} P^{-1} Z | Z_1 = A^{-1} v) = \Sigma^{\frac{1}{2}} P^{-1} E(Z | Z_1 = A^{-1} v) \\ &= \Sigma^{\frac{1}{2}} P^{-1} \begin{bmatrix} A^{-1} v \\ P_2 \Sigma^{-\frac{1}{2}} \mu \end{bmatrix}. \end{aligned}$$

Since P is orthogonal $P^{-1} = [P_1', P_2']$. $E(X|MX = v) = \Sigma^{\frac{1}{2}} P_1' A^{-1} v + \Sigma^{\frac{1}{2}} P_2' P_2 \Sigma^{-\frac{1}{2}} \mu$, $P'P = I = P_1'P_1 + P_2'P_2$. Hence $P_2'P_2 = I - P_1'P_1$; so $E(X|MX = v) = \Sigma^{\frac{1}{2}} P_1' A^{-1} v + \mu - \Sigma^{\frac{1}{2}} P_1' P_1 \Sigma^{-\frac{1}{2}} \mu$.

Finally, from the formulas $P_1 = (M \Sigma M')^{-\frac{1}{2}} M \Sigma^{\frac{1}{2}}$ and $A = M \Sigma^{\frac{1}{2}} P_1'$ we get, after some algebra,

$$E(X|MX = v) = \mu - \Sigma M' (M \Sigma M')^{-1} M \mu + \Sigma M' (M \Sigma M')^{-1} v,$$

which was the first formula to prove. Furthermore, given a fixed vector λ , $\text{Var}(\lambda' X | MX = v) = \text{Var}(\lambda' \Sigma^{\frac{1}{2}} P' Z | Z_1 = A^{-1} v)$. We have

$\lambda' \Sigma^{\frac{1}{2}} P' Z = \lambda' \Sigma^{\frac{1}{2}} P_1' Z_1 + \lambda' \Sigma^{\frac{1}{2}} P_2' Z_2$. Hence $\text{Var}(\lambda' \Sigma^{\frac{1}{2}} P' Z | Z_1 = A^{-1} v)$
 $= \text{Var}(\lambda' \Sigma^{\frac{1}{2}} P_2' Z_2) = \lambda' \Sigma^{\frac{1}{2}} P_2' P_2 \Sigma^{\frac{1}{2}} \lambda = \lambda' \Sigma \lambda - \lambda \Sigma^{\frac{1}{2}} P_1' P_1 \Sigma^{\frac{1}{2}} \lambda$; and $\Sigma^{\frac{1}{2}} P_1' P_1 \Sigma^{\frac{1}{2}}$
 $= \Sigma M' (M \Sigma M')^{-1} M \Sigma$. Since this is true for any λ , we conclude that

$$\text{Cov}(X | MX = v) = \Sigma(I - M' (M \Sigma M')^{-1} M \Sigma) \quad \text{Q.E.D.}$$

Note: Lemma 4.2.1 is a generalization of Theorem 2.5.1 on page 29 of Anderson (1958), and can also be proved using it.

Theorem 4.2.1.

$$E\left(\begin{pmatrix} Z \\ Z \end{pmatrix} \middle| \hat{a}\right) = \begin{pmatrix} F' \hat{a} \\ -\phi' a + k' K^{-1} F' \hat{a} \end{pmatrix}$$

$$\text{Cov}\left(\begin{pmatrix} Z \\ Z \end{pmatrix} \middle| \hat{a}\right) = \begin{bmatrix} K - F' G^{-1} F & k - K H k \\ k' - k' H K & \sigma^2 - k' H k \end{bmatrix}$$

where $H = K^{-1} F' G^{-1} F K^{-1}$.

Proof. Let us prove the expectation part as an example of application of lemma 4.2.1, and leave the covariance part which proof is not essentially different.

Let $\xi = \begin{pmatrix} Z \\ Z \end{pmatrix}$. Since $\hat{a} = G^{-1} F K^{-1} \underline{Z}$, let $M = G^{-1} F K^{-1}$ and $M^* = [M; 0]$. Then $\hat{a} = M \underline{Z} = M^* \xi$.

Let $K^* = \text{Cov}(\xi) = \begin{bmatrix} K & k \\ k' & \sigma^2 \end{bmatrix}$. By lemma 4.2.1

$$E(\xi | M^* \xi = \hat{a}) = E\xi - K^* M^{*'} (M^* K^* M^{*'})^{-1} M^* E\xi + K^* M^{*'} (M^* K^* M^{*'})^{-1} \hat{a}.$$

To simplify this, note first

$$M^* K^* M^{*'} = [M; 0] \begin{bmatrix} K & k \\ k' & \sigma^2 \end{bmatrix} \begin{bmatrix} M' \\ 0 \end{bmatrix} = M K M' = G^{-1} F K^{-1} K K^{-1} F' G^{-1} = G^{-1};$$

Secondly $M^{*'} G = \begin{bmatrix} M'G \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} K^{-1}F' \\ \vdots \\ 0 \end{bmatrix}$. Hence

$$M^{*'} GM^* = \begin{bmatrix} K^{-1}F' \\ \vdots \\ 0 \end{bmatrix} [M:0] = \begin{bmatrix} K^{-1}F'M & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

and $K^{-1}F'M = K^{-1}F'G^{-1}FK^{-1} \triangleq H$.

Then

$$K^{**'} (M^{*'} K^{*'} M^{*'})^{-1} M^* = K^* \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} KH & 0 \\ k'H & 0 \end{bmatrix}$$

while

$$K^{**'} (M^{*'} K^{*'} M^{*'})^{-1} = \begin{bmatrix} K & k \\ k' & 0 \end{bmatrix} \begin{bmatrix} M'G \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} KM'G \\ k'M'G \end{bmatrix} = \begin{bmatrix} F' \\ k'K^{-1}F' \end{bmatrix}.$$

Then we can write

$$E(\xi|\hat{a}) = \begin{pmatrix} F'a \\ f'a \end{pmatrix} - \begin{bmatrix} KH & 0 \\ k'H & 0 \end{bmatrix} \begin{pmatrix} F'a \\ f'a \end{pmatrix} + \begin{pmatrix} F' \\ k'K^{-1}F' \end{pmatrix} \hat{a}.$$

Moreover, we have the following simplifications:

$$KHF' = KK^{-1}F'G^{-1}F K^{-1}F' = F'$$

$$k'HF' = k'K^{-1}F'G^{-1}F K^{-1}F' = k'K^{-1}F'.$$

So the conditional expectation reduces to

$$E\left\{\begin{pmatrix} Z \\ \underline{Z} \end{pmatrix} \middle| \hat{a}\right\} = \begin{pmatrix} F'\hat{a} \\ f'a - k'K^{-1}F'a + k'K^{-1}F'\hat{a} \end{pmatrix},$$

that is $E(\underline{Z}|\hat{a}) = F'\hat{a}$, and $E(Z|\hat{a}) = -\phi'a + k'K^{-1}F'\hat{a}$. Q.E.D.

Theorem 4.2.2. The estimator Z^* can be reexpressed as

$$(1) \quad Z^* = \hat{Z} + \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}$$

where $\hat{Z}$ is the B.L.U.E. of Z . And its risk is

$$(2) \quad E(Z^* - Z)^2 = \sigma^2 - k'K^{-1}k + \text{Var}\left(\frac{\phi'\hat{a} \hat{a}'G\hat{a}}{1 + \hat{a}'G\hat{a}}\right) + \left\{E\left(\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}\right)\right\}^2.$$

Let us make two remarks before the proof:

Remarks. 1) The estimator $\hat{Z}_a$ that we derived in Section 4.1 satisfied

$$E\hat{Z}_a = EZ + \frac{\phi'a}{1 + a'Ga}.$$

This sheds some light on the nature of the estimator Z^* .

However $\hat{Z}_a$ itself was not a simple modification of $\hat{Z}$; some algebra shows that

$$\hat{Z}_a = \hat{Z} + \frac{\phi'\hat{a}}{1 + a'Ga} + \frac{\phi'\hat{a} a'Ga - \phi'a a'G\hat{a}}{1 + a'Ga}.$$

2) The risk of $\hat{Z}_a$ can be written

$$E(\hat{Z}_a - Z)^2 = \sigma^2 - k'K^{-1}k + a'Ga \left(\frac{\phi'a}{1 + a'Ga}\right)^2 + \left(\frac{\phi'a}{1 + a'Ga}\right)^2.$$

This shows an analogy between the risk of Z^* and the risk of $\hat{Z}_a$.

Proof of Theorem 4.2.2. Consider $Z^* - \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}$; it is equal to

$$k'K^{-1}\underline{Z} - \frac{\phi'\hat{a} \hat{a}'FK^{-1}\underline{Z}}{1 + \hat{a}'G\hat{a}} - \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}} = k'K^{-1}\underline{Z} - \frac{\phi'\hat{a} \hat{a}'G\hat{a}}{1 + \hat{a}'G\hat{a}} - \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}$$

because $\hat{a}'FK^{-1}\underline{Z} = \hat{a}'GG^{-1}FK^{-1}\underline{Z} = \hat{a}'G\hat{a}$. After reduction we get

$$Z^* - \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}} = k'K^{-1}\underline{Z} - \phi'\hat{a} = \hat{Z}, \text{ thus establishing formula (1).}$$

Next the risk of Z^* is

$$(3) \quad E(Z^* - Z)^2 = E\left(\hat{Z} - Z + \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}\right)^2 \\ = E(\hat{Z} - Z)^2 + E\left[\left(\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}\right)^2\right] + 2E\left[(\hat{Z} - Z) \left(\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}\right)\right].$$

To compute the cross-product term in (3) we condition on $\hat{a}$:

$$E \left[(\hat{Z}-Z) \left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) \middle| \hat{a} \right] = \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} E(\hat{Z}-Z) \middle| \hat{a}$$

$$\begin{aligned} E(\hat{Z}-Z \mid \hat{a}) &= \lambda' E(Z \mid \hat{a}) - E(Z \mid \hat{a}) = (k' K^{-1} - \phi' G^{-1} F K^{-1}) F' \hat{a} - (k' K^{-1} F' \hat{a} - \phi' a) \\ &= k' K^{-1} F' \hat{a} - \phi' \hat{a} - k' K^{-1} F' \hat{a} + \phi' a = \phi' (a - \hat{a}) . \end{aligned}$$

So

$$\begin{aligned} E \left[(\hat{Z}-Z) \left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) \right] &= E \left[\phi' (a - \hat{a}) \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right] \\ &= \phi' a E \left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) - E \left[\frac{(\phi' \hat{a})^2}{1 + \hat{a}' G \hat{a}} \right] . \end{aligned}$$

It is useful to notice that this is also

$$- \text{Cov} \left(\phi' \hat{a}, \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) .$$

Remembering that $E(\hat{Z}-Z)^2 = \sigma^2 - k' K^{-1} k + \phi' G^{-1} \phi$ we get

$$E(Z^* - Z)^2 = \sigma^2 - k' K^{-1} k + \phi' G^{-1} \phi + E \left[\left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right)^2 \right] - 2 \text{Cov} \left(\phi' \hat{a}, \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) .$$

Then observing that $\phi' G^{-1} \phi = \text{Var}(\phi' \hat{a})$, the risk of Z^* is reexpressed as

$$\begin{aligned} \sigma^2 - k' K^{-1} k + \text{Var}(\phi' \hat{a}) - 2 \text{Cov} \left(\phi' \hat{a}, \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) + E \left[\left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right)^2 \right] \\ = \sigma^2 - k' K^{-1} k + \text{Var} \left(\phi' \hat{a} - \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) + \left[E \left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) \right]^2 . \end{aligned}$$

And finally, since

$$\phi' \hat{a} - \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} = \phi' \hat{a} \left(\frac{\hat{a}' G \hat{a}}{1 + \hat{a}' G \hat{a}} \right)$$

we obtain the risk of Z^* in the form (2) given in the statement of the theorem. Q.E.D.

The next problem is to compare $R(a, Z^*)$ to $R(a, \hat{Z})$. Since we showed in Section 2 of Chapter 3 the admissibility of $\hat{Z}$, Z^* cannot be uniformly better than $\hat{Z}$ when a varies, although $\hat{Z}_a$ was.

However, when $a=0$, we have

$$E\left(\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}\right) = 0$$

because $\hat{a} \sim N(0, G^{-1})$ and $\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}$ is symmetric about zero. Therefore

$$\begin{aligned} R(0, Z^*) &= \sigma^2 - k'K^{-1}k + \text{Var}\left(\frac{\phi'\hat{a} \hat{a}'G\hat{a}}{1 + \hat{a}'G\hat{a}}\right) < \sigma^2 - k'K^{-1}k + \text{Var}(\phi'\hat{a}) \\ &= \sigma^2 - k'K^{-1}k + \phi'G^{-1}\phi = R(0, \hat{Z}) . \end{aligned}$$

Thus we see that in turn $\hat{Z}$ is not uniformly better than Z^* .

Next note that when $\|a\| = (a_1^2 + a_2^2 + \dots + a_p^2)^{1/2} \rightarrow +\infty$ the term $\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}}$ tends to zero in probability: to prove this fact we observe that $\forall \epsilon > 0, \exists A_1$ such that, for any vector $b \in \mathbb{R}^p$,

$$\|b\| > A_1 \Rightarrow \left| \frac{\phi'b}{1 + b'Gb} \right| < \epsilon ;$$

also $\forall \eta > 0, \exists A_2$ such that

$$\|a\| > A_2 \Rightarrow \Pr\{\|\hat{a}\| \leq A_1\} < \eta .$$

Hence $\forall \epsilon > 0$ and $\forall \eta > 0, \exists A_2$ such that

$$\|a\| > A_2 \Rightarrow \Pr\left\{\left| \frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}} \right| > \epsilon\right\} < \eta . \quad \text{Q.E.D.}$$

Consequently we have also $\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \xrightarrow{\mathcal{D}} 0$ where $\xrightarrow{\mathcal{D}}$ means tends in distribution to. Finally, since $\exists C$ such that $\forall b \in \mathbb{R}^p \quad \left| \frac{\phi' b}{1 + b' G b} \right| < C$, by classical convergence theorems (see Chung (1974), Chapter 4, Section 5) we infer that all the moments of $\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}}$ tend to zero, and since $\text{Var } \phi' \hat{a} = \text{constant}$, also that

$$\text{Var} \left(\phi' \hat{a} - \frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right) + E \left[\left(\frac{\phi' \hat{a}}{1 + \hat{a}' G \hat{a}} \right)^2 \right]$$

tends to $\text{Var}(\phi' \hat{a}) = \phi' G^{-1} \phi$. Therefore we have shown the following:

$$\lim_{\|a\| \rightarrow +\infty} R(a, Z^*) = \text{risk of the equalizer } \hat{Z}.$$

Section 4.3 contains a numerical example comparing $\hat{Z}$ and Z^* .

Before that we prove the next theorem, which gives $R(a, Z^*)$ in a form which lends itself readily to numerical approximations. We also prove a series of lemmas which will be used again in Chapter 6.

Observe that $\hat{a}' G \hat{a} \sim \chi_p^2 \left(\frac{a' G a}{2} \right)$; that is $\hat{a}' G \hat{a}$ has a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter $\frac{a' G a}{2}$. A property of this distribution is that it can be expressed as a Poisson mixture of central chi-squared distributions. Symbolically $\chi_p^2 \left(\frac{a' G a}{2} \right) = \chi_{p+2K}^2$ where $K \sim \text{Poisson} \left(\frac{a' G a}{2} \right)$.

Theorem 4.2.3. $R(a, Z^*) = R(a, \hat{Z}) + \xi(a)$ where

$$(4) \quad \xi(a) = \frac{1}{2} \phi' G^{-1} \phi \left[E \frac{1}{1 + \chi_{p+2K}^2} - E \frac{1}{1 + \chi_{p+2K+2}^2} \right] + \frac{5}{2} (\phi' a)^2 \left[E \frac{1}{1 + \chi_{p+2K+2}^2} - E \frac{1}{1 + \chi_{p+2K+4}^2} \right],$$

with K as above.

Proof of Theorem 4.2.3. We saw in the proof of Theorem 4.2.2 that $R(a, Z^*)$ can be written

$$\sigma^2 - k'K^{-1}k + \phi'G^{-1}\phi + E \left[\left(\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}} \right)^2 \right] \\ + 2\phi'a E \left(\frac{\phi'\hat{a}}{1 + \hat{a}'G\hat{a}} \right) - 2 E \left[\frac{(\phi'\hat{a})^2}{1 + \hat{a}'G\hat{a}} \right] .$$

Let $X = G^{\frac{1}{2}}\hat{a} \sim N(G^{\frac{1}{2}}a, I_p)$; also let $\mu = G^{\frac{1}{2}}a$ and $c = \phi'G^{-\frac{1}{2}}$, so that $\hat{a}'G\hat{a} = \|X\|^2$ and $\phi'\hat{a} = cX$. In order to reexpress $\xi(a) = R(a, Z^*) - R(a, \hat{Z})$ we need to compute the following quantities:

$$E \left[\left(\frac{cX}{1 + \|X\|^2} \right)^2 \right] , \quad E \frac{(cX)^2}{1 + \|X\|^2}$$

and

$$E \left(\frac{cX}{1 + \|X\|^2} \right) .$$

This is done using the lemmas that we state next and prove after the proof of Theorem 4.2.3.

Lemma 4.2.2. If $X \sim N(\mu, I_p)$, for any row vector c and any function g such that the following expectations exist, we have

$$E(cX g(\|X\|^2)) = c\mu E g(\chi_{p+2K+2}^2)$$

where $K \sim \text{Poisson} \left(\frac{\|\mu\|^2}{2} \right)$. (This lemma is essentially proved in Stein (1966).)

Lemma 4.2.3. Under the assumptions of the above lemma

$$E\{XX' g(\|X\|^2)\} = I_p E g(\chi_{p+2K+2}^2) + \mu\mu' E g(\chi_{p+2K+4}^2) .$$

Lemma 4.2.4. Provided $n \geq 3$ the following identity holds

$$E\left[\left(\frac{1}{1 + \chi_n^2}\right)^2\right] = \frac{1}{2} E \frac{1}{1 + \chi_{n-2}^2} - \frac{1}{2} E \frac{1}{1 + \chi_n^2}.$$

Proof of Theorem 4.2.3 (continued). As an application of lemma 4.2.3 we have

$$E\{(cX)^2 g(\|X\|^2)\} = \|c\|^2 E g(\chi_{p+2K+2}^2) + (c\mu)^2 E g(\chi_{p+2K+4}^2).$$

Now consider the function $g(x) = \frac{1}{1+x}$; we can reexpress $\xi(a)$ as

$$\xi(a) = E\{(cX)^2 g^2(\|X\|^2)\} + 2 \phi'a E\{cX g(\|X\|^2)\} - 2 E\{(cX)^2 g(\|X\|^2)\}.$$

To simplify the writing define $\psi(K) = g(\chi_{p+2K}^2)$. By lemma 4.2.2 and the application of lemma 4.2.3 we have

$$\begin{aligned} \xi(a) &= \|c\|^2 E \psi^2(K+1) + (c\mu)^2 E \psi^2(K+2) + 2(\phi'a)(c\mu) E \psi(K+1) - \\ &\quad 2 \|c\|^2 E \psi(K+1) - 2 (c\mu)^2 E \psi(K+2). \end{aligned}$$

Note that $\|c\|^2 = \phi'G^{-1}\phi$ and $c\mu = \phi'a$. By lemma 4.2.4 $E \psi^2(K+1) = \frac{1}{2} E \psi(K) - \frac{1}{2} E \psi(K+1)$ and $E \psi^2(K+2) = \frac{1}{2} E \psi(K+1) - \frac{1}{2} E \psi(K+2)$. This yields

$$\begin{aligned} \xi(a) &= \phi'G^{-1}\phi\left[\frac{1}{2} E\psi(K) - \frac{1}{2} E\psi(K+1) - 2 E\psi(K+1)\right] \\ &\quad + (\phi'a)^2 \left[\frac{1}{2} E\psi(K+1) - \frac{1}{2} E\psi(K+2) + 2 E\psi(K+1) - 2 E\psi(K+2)\right] \\ &= \frac{1}{2} \phi'G^{-1}\phi[E\psi(K) - 5 E\psi(K+1)] + \frac{5}{2} (\phi'a)^2 [E\psi(K+1) - E\psi(K+2)] . \text{ Q.E.D.} \end{aligned}$$

Proof of Lemma 4.2.2. (Stein (1966))

$$\begin{aligned}
 E(X_1 g(\|X\|^2)) &= \frac{1}{(2\pi)^{p/2}} \int x_1 g(\|x\|^2) \exp\left\{-\frac{1}{2} \|x-\mu\|^2\right\} dx \\
 &= \frac{\exp\left(-\frac{\|\mu\|^2}{2}\right)}{(2\pi)^{p/2}} \int x_1 g(\|x\|^2) \exp\left\{-\frac{\|x\|^2}{2} + \mu'x\right\} dx \\
 &= e^{-\|\mu\|^2/2} \frac{\partial}{\partial \mu_1} \{e^{\|\mu\|^2/2} E g(\|x\|^2)\} \\
 &= e^{-\|\mu\|^2/2} \frac{\partial}{\partial \mu_1} \sum_{k=0}^{\infty} \frac{\left(\frac{\|\mu\|^2}{2}\right)^k}{k!} E g(\chi_{p+2k}^2) \\
 &= \mu_1 E g(\chi_{p+2(K+1)}^2) .
 \end{aligned}$$

The result of lemma 4.2.2 follows by rotation.

Sketch of proof of lemma 4.2.3. The technique used in the proof of lemma 4.2.2 can be extended to compute $E(X_1^2 g(\|X\|^2))$ and $E(X_i X_j g(\|X\|^2))$ for $i \neq j$: one has to differentiate twice instead of once, and the result of lemma 4.2.3 follows after collecting the terms.

Proof of lemma 4.2.4.

$$E\left[\left(\frac{1}{1 + \chi_n^2}\right)^2\right] = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{+\infty} \frac{s^{n/2-1}}{(1+s)^2} e^{-s/2} ds .$$

Let $du = \frac{ds}{(1+s)^2}$ and $v = s^{n/2-1} e^{-s/2}$. Then integrate by parts: we obtain

$$E\left[\left(\frac{1}{1 + \chi_n^2}\right)^2\right] = \frac{1}{2^{n/2} \Gamma(n/2)} \left\{ \left(-\frac{1}{1+s}\right) s^{n/2-1} e^{-s/2} \Big|_0^{+\infty} - \right.$$

$$\begin{aligned}
& \int_0^{+\infty} \left(-\frac{1}{1+s} \right) \left(\left(\frac{n}{2} - 1 \right) s^{n/2-1} e^{-s/2} - \frac{1}{2} s^{n/2-1} e^{-s/2} \right) ds \Big\} \\
&= \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{+\infty} \frac{(n/2 - 1)}{1+s} s^{(n/2-1)-1} e^{-s/2} ds - \\
&\quad \frac{1}{2^{n/2} \Gamma(n/2)} \frac{1}{2} \int_0^{+\infty} \frac{s^{n/2-1}}{1+s} e^{-s/2} ds \\
&= \frac{1}{2} E \frac{1}{1 + \chi_{n-2}^2} - \frac{1}{2} E \frac{1}{1 + \chi_n^2} . \quad \text{Q.E.D.}
\end{aligned}$$

Theorem 4.2.3 can be used as follows to approximate $R(a, Z^*)$, the risk function of the nonlinear estimator Z^* : first we take $E \frac{1}{\chi_{p+2K}^2}$ as an approximation for $E \frac{1}{1 + \chi_{p+2K}^2}$. Then we can use the method given by Stein (1966) to approximate $E \frac{1}{\chi_{p+2K}^2}$:

$$E \frac{1}{\chi_{p+2K}^2} = E \left\{ E \left[\frac{1}{\chi_{p+2K}^2} \middle| K \right] \right\} = E \frac{1}{p-2+2K} = \frac{1}{p-2+t} E \left\{ \frac{1}{1 + \frac{2K-t}{p-2+t}} \right\} .$$

Following Stein who checks the quality of the next approximation in the case $p=4$ (because then the exact calculations are also easy to carry out) we write this last equality

$$(5) \quad \approx \frac{1}{p-2+t} E \left\{ 1 - \frac{2K-t}{p-2+t} + \frac{(2K-t)^2}{(p-2+t)^2} \right\} .$$

When $t/2$ is the parameter of the Poisson variable K , (5) becomes

$$\frac{1}{p-2+t} \left\{ 1 - \frac{2t}{(p-2+t)^2} \right\} . \text{ Therefore we obtain}$$

$$(6) \quad E \frac{1}{1 + \chi_{p+2K}^2} \approx \frac{1}{p-2+t} \left\{ 1 - \frac{2t}{(p-2+t)^2} \right\} .$$

Finally, approximation (6) applied in the expression for $\xi(a)$ in (4), with $t = a'Ga$, yields an approximation for the risk function of Z^* .

4.3. A Numerical Example Comparing $\hat{Z}$ and Z^*

Consider the case $m(x) = m$, i.e. $a = m \in R^1$, and let e as before be a column vector of n ones. The risk function of the B.L.U.E. $\hat{Z}$ can be written

$$(1) \quad R(m, \hat{Z}) = \sigma^2 - k'K^{-1}k + \frac{(e'K^{-1}k - 1)^2}{e'K^{-1}e}.$$

Note: This expression is also obtained by Faith (1975).

On the other hand, the function ξ , introduced in Theorem 4.2.3 of the preceding section, can be simplified if we note that

$$\phi = (e'K^{-1}k - 1), \quad \hat{a} = \hat{m} = \frac{e'K^{-1}Z}{e'K^{-1}e},$$

$$\hat{a}'G\hat{a} = \frac{(e'K^{-1}Z)^2}{e'K^{-1}e}, \quad \text{and} \quad \hat{m} \sim N\left(m, \frac{1}{e'K^{-1}e}\right).$$

Consider a random variable $X \sim N(m, 1)$ and define

$$\xi_0(m) = E\left[\left(\frac{X}{1+X^2}\right)^2\right] - 2 \operatorname{Cov}\left(X, \frac{X}{1+X^2}\right).$$

Then

$$\xi(m) = \frac{(e'K^{-1}k - 1)^2}{e'K^{-1}e} \xi_0(m \sqrt{e'K^{-1}e}).$$

Let us suppose further that we have three observation points
(n=3) and that the covariance function has the simple form

$$K(x,y) = \begin{cases} 1 & \text{if } x=y \\ 0 & \text{if } x \neq y \end{cases}.$$

Then $\phi = -1$, $G = 3$,

$$\hat{Z} = \frac{Z_1 + Z_2 + Z_3}{3}, \quad Z^* = \hat{Z} \left(1 - \frac{1}{1 + 3\hat{m}^2} \right),$$

$$R(m, \hat{Z}) = 1 + \frac{1}{3} = \frac{4}{3},$$

$$R(m, Z^*) = R(m, \hat{Z}) + \frac{1}{3} \xi_0(m \sqrt{3}).$$

Actually, in this example we are truly comparing two estimators of the mean and adding to the risk of each one of them the constant one. Indeed $\hat{Z} = f'\hat{a} + k'K^{-1}[\underline{Z} - F'\hat{a}]$ reduces to $\hat{Z} = \hat{m}$. Thus $E(\hat{Z} - Z)^2$ may be written

$$E(\hat{m} - Z)^2 = E(\hat{m} - m - (Z - m))^2 = E(\hat{m} - m)^2 + \sigma^2$$

by independence of Z and $\hat{m}$.

Similarly, Z^* is an estimator of m . Let $m^* = Z^*$, then as above

$$E(m^* - Z)^2 = E(m^* - m)^2 + \sigma^2.$$

Thus we may write $R(m, \hat{m}) = \frac{1}{3}$ and $R(m, m^*) = \frac{1}{3} + \frac{1}{3} \xi_0(m \sqrt{3})$. Figure 1 represents these two functions. To plot $R(m, m^*)$ I calculated

several values of the function ξ_0 by numerical integration on a computer using Simpson's method.

We observe that if we have the prior information that the unknown parameter m lies in the interval $[m_0 - .8\sigma, m_0 + .8\sigma]$ then we should subtract the constant m_0 to all our observations and use the estimator m^* instead of $\hat{m}$, because its risk may be much better, and may not be worse. Such a situation is not uncommon in practice.

Figure 1. Comparison of the risk functions of $\hat{m}$ and m^* .

CHAPTER 5

ESTIMATION OF A MAP OF Z USING THE B.L.U.E. AND THE PROBLEM OF NETWORK DESIGN

We see in this chapter how the best linear unbiased estimator $\hat{Z}(x)$ performs when used to pointwise estimate a two-dimensional random field $\{Z(x), x \in \mathbb{R}^2\}$ over an area A .

As said in Section 1.2 of the introduction, once an estimation method is chosen the next problem is the design of a network which has good properties in terms of quality of estimation, and which is reasonable in practice in the sense that it does not cost too much money to build. This question is examined in detail in Sections 5.3 through 5.8. We come up with a new suggestion for the design of networks used to monitor smooth random phenomena; it is the subject of Section 5.7 and is further illustrated in Section 5.8.

5.1. Preliminaries

From the measurements $Z(x_1), Z(x_2), \dots, Z(x_n)$ taken at the station-points (usually located in A , although not necessarily) the estimated map of Z is the function $x \in A \rightarrow \hat{Z}(x)$.

To study the quality of this map the object of interest is the mean squared error as a function of x . Since it is also a function of $x_1, x_2, \dots, x_n$ we shall write it

$$E(\hat{Z}(x) - Z(x))^2 = s(x; x_1, x_2, \dots, x_n) \quad .$$

Its expression is familiar to us; we now write it emphasizing the dependence on x :

$$s(x; x_1, x_2, \dots, x_n) = \sigma_x^2 - k_x' K^{-1} k_x + \phi_x' G^{-1} \phi_x.$$

It is worthwhile noting that it does not depend on the values $Z(x_1), Z(x_2), \dots, Z(x_n)$, only on the position of the points $x_1, x_2, \dots, x_n$. Therefore, inasmuch as the covariance function $K(x,y)$ depends only on the geometry of the spatial area, the function s depends only on the geometry of the stations.

Also, let us observe that the estimated map is exact at the stations; that is, for $x_i \in \{x_1, x_2, \dots, x_n\}$ we have

$$\hat{Z}(x_i) = Z(x_i)$$

and then $s(x_i; x_1, x_2, \dots, x_n) = 0$. This can be seen from the equation

$$\lambda(x) = K^{-1} k_x - K^{-1} F' G^{-1} (F K^{-1} k_x - f_x);$$

when $x = x_i$, $K^{-1} k_{x_i} = e_i = i^{\text{th}}$ canonical vector of R^n and $F K^{-1} k_{x_i}$ is exactly f_{x_i} ; therefore $\lambda(x_i) = e_i$ and $\hat{Z}(x_i) = \lambda'(x_i) \underline{Z} = Z(x_i)$.

However, the function $x \in A \rightarrow \hat{Z}(x)$ need not be continuous. The smoothness properties of the map depend on the covariance function $K(x,y)$. Covariance functions used in practice are differentiable both in x and y at each pair (x,y) , except perhaps pairs such that $x = y$. In this case the map is differentiable at each point which is not a station.

These smoothness properties of the real valued function $x \rightarrow \hat{Z}(x)$, given $\underline{Z}$, should not be confused with the smoothness properties, in a stochastic sense, of the random field Z which also depend on the covariance function. This will be studied in detail in Section 5.6.

For the problem of network design we shall also consider $s(x; x_1, x_2, \dots, x_n)$ as a function of $x_1, x_2, \dots, x_n$. In particular, we shall study its behavior when we remove one of the stations, x_n say, or when we add a new station x_{n+1} .

To this question the theorem which is the object of the next section is basic.

5.2. An Updating Theorem

Theorem 5.2.1. Let $\hat{Z}(x)$ be the B.L.U.E. of $Z(x)$ based on the n stations $x_1, x_2, \dots, x_n$, and $\hat{Z}_1(x)$ that based on the $n-1$ stations $x_1, x_2, \dots, x_{n-1}$. Then

$$(1) \quad \hat{Z}(x) = \hat{Z}_1(x) - \frac{\text{cov}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n))}{\text{Var}(\hat{Z}_1(x_n) - Z(x_n))} [\hat{Z}_1(x_n) - Z(x_n)]$$

and

$$(2) \quad \frac{s(x; x_1, x_2, \dots, x_n)}{s(x; x_1, x_2, \dots, x_{n-1})} = 1 - \rho^2(x, x_n)$$

where $\rho(x, x_n) = \text{corr}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n))$

Lemma 5.2.1. $\text{cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y)) = K(x, y) - k'_x K^{-1} k_y + \phi'_x G^{-1} \phi_y$.

Proof of Lemma 5.2.1. We use for $\hat{Z}(x)$ the form

$$\hat{Z}(x) = k'_x K^{-1} \underline{Z} - \phi'_x G^{-1} F K^{-1} \underline{Z} = \lambda'(x) \underline{Z}$$

$$\text{then } \text{cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y)) = \text{cov}(\hat{Z}(x), \hat{Z}(y)) - \text{cov}(Z(x), \hat{Z}(y)) -$$

$$\text{cov}(\hat{Z}(x), Z(y)) + \text{cov}(Z(x), Z(y))$$

$$= (k'_x K^{-1} - \phi'_x G^{-1} F K^{-1}) K (K^{-1} k_y - K^{-1} F' G^{-1} \phi_y) -$$

$$\lambda'(y) k_x - \lambda'(x) k_y + K(x, y)$$

$$= k'_x K^{-1} k_y - k'_x K^{-1} F' G^{-1} \phi_y - \phi'_x G^{-1} F K^{-1} k_y +$$

$$\phi'_x G^{-1} F K^{-1} F' G^{-1} \phi_y - k'_y K^{-1} k_x + \phi'_y G^{-1} F K^{-1} k_x -$$

$$k'_x K^{-1} k_y + \phi'_x G^{-1} F K^{-1} k_y + K(x, y)$$

after cancellation, and reduction noting that $FK^{-1}F' = G$, this becomes

$$K(x, y) - k'_x K^{-1} k_y + \phi'_x G^{-1} \phi_y \quad . \quad \text{Q.E.D.}$$

Proof of Theorem 5.2.1. On every quantity an index 1 will indicate that the quantity has been computed on basis of the $(n-1)^{\text{st}}$ station-points. Thus,

$$K_1 = \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_{n-1}) \\ \vdots & & \vdots \\ K(x_{n-1}, x_1) & \dots & K(x_{n-1}, x_{n-1}) \end{bmatrix}, \quad k_{1x} = \begin{bmatrix} K(x_1, x) \\ \vdots \\ K(x_{n-1}, x) \end{bmatrix}$$

$$F_1 = \begin{bmatrix} f_1(x_1) & f_1(x_{n-1}) \\ \vdots & \vdots \\ f_k(x_1) & \dots & f_k(x_{n-1}) \end{bmatrix}, \quad \underline{Z}_1 = \begin{bmatrix} Z(x_1) \\ \vdots \\ Z(x_{n-1}) \end{bmatrix}$$

$$G_1 = F_1 K_1^{-1} F_1', \quad \text{and so on.}$$

To reexpress $\hat{Z}(x)$ in terms of $\hat{Z}_1(x)$ and $\hat{Z}_1(x_n)$ we will use the formula $\hat{Z}(x) = \lambda'(x)\underline{Z}$, where $\lambda(x)$ is given by formula 2.1.7, and will partition the various quantities in the expression of $\lambda(x)$.

$$\text{Let} \quad K = \left[\begin{array}{c|c} K_1 & k_{1n} \\ \hline k'_{1n} & \sigma_n^2 \end{array} \right],$$

$$\text{then} \quad K^{-1} = \left[\begin{array}{c|c} K_1^{-1} + \frac{K_1^{-1} k_{1n} k'_{1n} K_1^{-1}}{\delta} & -\frac{K_1^{-1} k_{1n}}{\delta} \\ \hline -\frac{k'_{1n} K_1^{-1}}{\delta} & \frac{1}{\delta} \end{array} \right]$$

with

$$(3) \quad \delta = \sigma_n^2 - k'_{1n} K_1^{-1} k_{1n}.$$

$G = F K^{-1} F'$ may be reexpressed using this partitioning of K^{-1} together with $F' = [F_1' : f_n']$. We obtain after some algebra

$$G = F_1 K_1^{-1} F_1' + \frac{1}{\delta} (F_1 K_1^{-1} k_{1n} - f_n') (F_1 K_1^{-1} k_{1n} - f_n')',$$

that is,

$$G = G_1 + \frac{\phi_{1n} \phi_{1n}'}{\delta}.$$

This is of the form $B + xx'$ (the assumption that K is positive definite ensures $\delta > 0$); therefore, by Lemma 2.3.2

$$G^{-1} = G_1^{-1} - \frac{G_1^{-1} \phi_{1n} \phi_{1n}' G_1^{-1}}{\delta + \phi_{1n}' G_1^{-1} \phi_{1n}}.$$

We shall only sketch the next calculations because they involve a great deal of matrix manipulation.

By analogy with δ given in formula (3) let

$$(4) \quad \gamma = \text{cov}(Z(x), Z(x_n)) - k_{1n}' K_1^{-1} k_{1x}.$$

Then from $FK^{-1} = \left[F_1 K_1^{-1} + \frac{\phi_{1n} k_{1n}' K_1^{-1}}{\delta} \mid -\frac{\phi_{1n}}{\delta} \right]$, we get $FK^{-1} k_x =$

$F_1 K_1^{-1} k_{1x} - \frac{\gamma}{\delta} (F_1 K_1^{-1} k_{1n} - f_n)$, and hence $\phi_x = \phi_{1x} - \frac{\gamma}{\delta} \phi_{1n}$. We also

derive

$$K^{-1} F' G^{-1} = \left[\begin{array}{c} K_1^{-1} F_1' G_1^{-1} - \frac{K_1^{-1} F_1' G_1^{-1} \phi_{1n} \phi_{1n}' G_1^{-1}}{V} \\ + \frac{K_1^{-1} k_{1n} \phi_{1n}' G_1^{-1}}{\delta} - \frac{K_1^{-1} k_{1n} \phi_{1n}' G_1^{-1} \phi_{1n} \phi_{1n}' G_1^{-1}}{\delta V} \\ \hline - \frac{\phi_{1n}' G_1^{-1}}{\delta} + \frac{\phi_{1n}' G_1^{-1} \phi_{1n} \phi_{1n}' G_1^{-1}}{\delta V} \end{array} \right]$$

where $V = \delta + \phi_{1n}' G_1^{-1} \phi_{1n} = \text{Var}(\hat{Z}_1(x_n) - Z(x_n))$. Then we can compute

$\lambda(x) = K^{-1} k_x - K^{-1} F' G^{-1} \phi_x$. To express it let

$$T(n,x) = \frac{\frac{\gamma}{\delta} \phi'_{1n} G_1^{-1} \phi_{1n} - \phi'_{1n} G_1^{-1} \phi_{1x}}{\delta + \phi'_{1n} G_1^{-1} \phi_{1n}}$$

After some calculations it turns out that

$$\lambda(x) = \left[\frac{\begin{aligned} &K_1^{-1} k_{1x} - K_1^{-1} F_1' G_1^{-1} \phi_{1x} - \frac{\gamma}{\delta} K_1^{-1} k_{1n} \\ &+ \frac{\gamma}{\delta} K_1^{-1} F_1' G_1^{-1} \phi_{1n} + (K_1^{-1} k_{1n} - K_1^{-1} F_1' G_1^{-1} \phi_{1n}) T(n,x) \end{aligned}}{\frac{\gamma}{\delta} - T(n,x)} \right]$$

Finally,

$$\begin{aligned} \frac{\gamma}{\delta} - T(n,x) &= \frac{\phi'_{1n} G_1^{-1} \phi_{1x} - \frac{\gamma}{\delta} \phi'_{1n} G_1^{-1} \phi_{1n}}{\delta + \phi'_{1n} G_1^{-1} \phi_{1n}} + \frac{\gamma}{\delta} \\ &= \frac{\delta \phi'_{1n} G_1^{-1} \phi_{1x} - \gamma \phi'_{1n} G_1^{-1} \phi_{1n} + \gamma \delta + \gamma \phi'_{1n} G_1^{-1} \phi_{1n}}{\delta(\delta + \phi'_{1n} G_1^{-1} \phi_{1n})}; \end{aligned}$$

two terms cancel in the numerator, and a factor δ simplifies; we get

$$\frac{\gamma + \phi'_{1n} G_1^{-1} \phi_{1x}}{\delta + \phi'_{1n} G_1^{-1} \phi_{1n}} = \frac{K(x, x_n) - k'_{1n} G_1^{-1} k_{1x} + \phi'_{1n} G_1^{-1} \phi_{1x}}{K(x_n, x_n) - k'_{1n} G_1^{-1} k_{1n} + \phi'_{1n} G_1^{-1} \phi_{1n}}$$

From Lemma 5.2.1 we recognize this term: it is

$$\frac{\text{cov}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n))}{\text{Var}(\hat{Z}_1(x_n) - Z(x_n))}$$

Let us call it S . Then we obtain

$$\lambda(x) = \left[\frac{\lambda_1(x) - S \lambda_1(x_n)}{S} \right]$$

and

$$\begin{aligned} \hat{Z}(x) &= \lambda'(x)Z = \lambda_1'(x)Z_1 - S \lambda_1'(x_n)Z_1 + S Z(x_n) = \hat{Z}_1(x) - S \hat{Z}_1(x_n) + S Z(x_n) \\ &= \hat{Z}_1(x) - \frac{\text{cov}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n))}{\text{Var}(\hat{Z}_1(x_n) - Z(x_n))} [\hat{Z}_1(x_n) - Z(x_n)] \end{aligned}$$

This completes the first part of the proof of Theorem 5.2.1.

To prove the second part note that

$$\hat{Z}(x) - Z(x) = \hat{Z}_1(x) - Z(x) - S [\hat{Z}_1(x_n) - Z(x_n)]$$

and therefore

$$\begin{aligned} \text{Var}(\hat{Z}(x) - Z(x)) &= \text{Var}(\hat{Z}_1(x) - Z(x)) - \\ &\quad 2S \text{cov}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n)) + \\ &\quad S^2 \text{Var}(\hat{Z}_1(x_n) - Z(x_n)) \end{aligned}$$

which simplifies to

$$\text{Var}(\hat{Z}_1(x) - Z(x)) = \frac{[\text{cov}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n))]^2}{\text{Var}(\hat{Z}_1(x_n) - Z(x_n))}$$

We divide by $\text{Var}(\hat{Z}_1(x) - Z(x))$ and get the formula for the ratio of variances:

$$\frac{\text{Var}(\hat{Z}(x) - Z(x))}{\text{Var}(\hat{Z}_1(x) - Z(x))} = 1 - \rho^2$$

where

$$\rho = \text{corr}(\hat{Z}_1(x) - Z(x), \hat{Z}_1(x_n) - Z(x_n)) \quad . \quad \text{Q.E.D.}$$

The decomposition (1) for $\hat{Z}(x)$ looks like a regression line equation, although it is not one. It can be rewritten as

$$\hat{Z}(x) = \hat{Z}_1(x) - \sqrt{\text{Var}(\hat{Z}_1(x) - Z(x))} \rho(x, x_n) \frac{\hat{Z}_1(x_n) - Z(x_n)}{\sqrt{\text{Var}(\hat{Z}_1(x_n) - Z(x_n))}}$$

and has the following interpretation: The extra information that $\hat{Z}(x)$ has over $\hat{Z}_1(x)$ is $Z(x_n)$; the way it uses it is by computing the discrepancy between $\hat{Z}_1(x_n)$ and $Z(x_n)$, normalizing it, getting from it the best linear estimator of the error of $\hat{Z}_1(x)$, and then correcting $\hat{Z}_1(x)$ by subtracting this estimated error.

Besides its theoretical interest for studying the function $s(x; x_1, x_2, \dots, x_n)$, Theorem 5.2.1 is of great value in numerical computations on a computer because it allows tremendous reduction of computation times. This will be illustrated in Section 5.8 which deals with a numerical example.

5.3. Possible Criteria for Assessing the Quality of a Network

In order to design a network or to improve an already existing one or to compare the merits of two different networks, we need a criterion measuring the quality of a network.

There are several candidates. For instance, we may choose to use the maximum mean squared error

$$(1) \quad \max_{x \in A} s(x; x_1, x_2, \dots, x_n)$$

or the average mean squared error

$$(2) \quad \int_A s(x; x_1, x_2, \dots, x_n) d\mu(x)$$

where μ is a measure of our choice; the simplest is the ordinary Lebesgue measure dx in $\mathbb{R}^2$.

Of special interest is the case when the random field $\{Z(x), x \in \mathbb{R}^2\}$ is second-order-stationary, because then a simple criterion suggests itself. In this case $m(x) = m$, an unknown constant, and $K(x,y)$ is a function only of $y-x$; in particular, $K(x,x) = \sigma^2$ does not depend on x . Under the natural hypothesis that

$$\lim_{\|y-x\| \rightarrow +\infty} K(x,y) = 0,$$

we obtain, from formula 4.3.1,

$$(3) \quad \lim_{\|x\| \rightarrow +\infty} \text{Var}(\hat{Z}(x) - Z(x)) = \sigma^2 + \frac{1}{e' K^{-1} e}$$

which expresses the mean squared error at a distant point.

Therefore, $e' K^{-1} e$ is a possible criterion (to be maximized) to measure the quality of a network. And it is appealing because it is simple to compute.

We first examine this criterion, in Section 5.4. Its relation to criterion (1) is studied in Section 5.5. Then, in Section 5.6, the criterion (2) is examined in detail.

5.4. Study of Criterion $e' K^{-1} e$ When the Random Field Z is Weakly Stationary

Suppose we have a network of $n-1$ stations $x_1, x_2, \dots, x_{n-1}$ and we want to add a new station x_n .

Consider the $n \times n$ matrix $K = [K(x_i, x_j)]$ partitioned with light notation as follows

$$K = \begin{bmatrix} K_1 & k \\ k & \sigma^2 \end{bmatrix} ;$$

an expression for K^{-1} has been given in the proof of Theorem 5.2.1, and may be used to show that

$$(1) \quad e' K^{-1} e = e_1' K_1^{-1} e_1 + \frac{(e_1' K_1^{-1} k - 1)^2}{\sigma^2 - k' K_1^{-1} k} .$$

So, in general, the question is to locate x_n so as to maximize

$$\frac{(e_1' K_1^{-1} k - 1)^2}{\sigma^2 - k' K_1^{-1} k} .$$

When $K(x, y)$ has more structure, we can say more:

Theorem 5.4.1. Suppose, as in Section 2.3, that the covariance function may be written

$$(2) \quad K(x, y) = \sigma^2 (1 - \alpha H(x, y))$$

then

$$e' K^{-1} e = \frac{e' H^{-1} e}{\sigma^2 (e' H^{-1} e - \alpha)} ;$$

and when α is small enough so that $e' H^{-1} e$ stays greater than α for all x_n 's in some bounded area, the local maximization of $e' K^{-1} e$ is equivalent to the local minimization of $e' H^{-1} e$.

Proof. $K = B + x x'$ with $B = -\alpha \sigma^2 H$ and $x = \sigma e$. Then by lemma 2.3.2

$$K^{-1} = -\frac{1}{\alpha \sigma^2} \left(H^{-1} + \frac{H^{-1} e e' H^{-1}}{\alpha - e' H^{-1} e} \right) .$$

Therefore,

$$\begin{aligned} e' K^{-1} e &= -\frac{1}{\alpha \sigma^2} \left(e' H^{-1} e + \frac{(e' H^{-1} e)^2}{\alpha - e' H^{-1} e} \right) \\ &= \frac{e' H^{-1} e}{\sigma^2 (e' H^{-1} e - \alpha)} . \end{aligned}$$

The second part of the theorem is proved by noting that

$$\frac{d}{dx} \left[\frac{x}{x - \alpha} \right] = -\frac{\alpha}{(x - \alpha)^2} < 0 . \quad \text{Q.E.D.}$$

When Theorem 5.4.1 applies and the problem is to maximize $e' H^{-1} e$, then again formula (1) is valid substituting H for K ; so the quantity to be maximized is

$$\frac{(e_1' H_1^{-1} h - 1)^2}{\tau^2 - h' H_1^{-1} h}$$

(with the obvious similar partitioning of H).

On the other hand, in the problem of shutting down one monitoring station in a network (i.e., removing one of the station-point), the use of the criterion $e'K^{-1}e$ becomes straightforward: we have to compute the n quantities $e_1'K_{(i)}^{-1}e_1$, where $K_{(i)}$ is the matrix obtained from K by deleting the i th row and the i th column, and we remove the station i such that $e_1'K_{(i)}^{-1}e_1$ is maximum.

Before leaving these elementary considerations, let us see an example which shows how sensible the criterion $e'K^{-1}e$ is.

Consider six stations forming the network represented in Figure 1. Suppose $K(x,y) = \exp(-\frac{1}{10} \|x-y\|^2)$; then $e'K^{-1}e = 2.7057$, and when we remove station i , $i = 1, 2, \dots, 6$ we obtain

$$\begin{aligned} e_1'K_{(i)}^{-1}e_1 &= 2.3330 \\ &2.7055 \\ &2.5612 \\ &2.3615 \\ &2.6770 \\ &1.9538 \end{aligned}$$

Therefore, the best station to remove is x_2 , and the ranking of the stations is

$$x_2, x_5, x_3, x_4, x_1, x_6$$

Secondly, instead of the preceding covariance function, suppose that

$$K(x,y) = \sigma^2(1 - \alpha \|x-y\| + o(\|x-y\|)) ,$$

and α is small enough so that with our data the approximation

Figure 1. Example of network. If we want to remove one station, which one should we choose?

$$(3) \quad K(x,y) \approx \sigma^2(1 - \alpha \|x-y\|)$$

is valid. Then we can apply Theorem 5.4.1: With N denoting the $n \times n$ matrix $[\|x_i - x_j\|]$, the problem is to minimize $e_1' N_{(i)}^{-1} e_1$. We have $e' N^{-1} e = .2980$, and

$$\begin{aligned} e_1' N_{(i)}^{-1} e_1 = & .3394 \\ & .2989 \\ & .3062 \\ & .3229 \\ & .3006 \\ & .3825 \end{aligned}$$

Therefore, the ranking of the stations is

$$x_2, x_5, x_3, x_4, x_1, x_6$$

That in both cases we obtain the same ranking of the stations is a very satisfactory result on the behavior of the criterion $e' K^{-1} e$. Moreover, geometrically this ranking makes sense.

In future work it would be of interest to investigate further this stability property, and more generally to study $e' K^{-1} e$ from the point of view of a generalized distance between several points.

Remark: Although certain covariance functions are such that

$$K(x,y) = \sigma^2(1 - \alpha \|x-y\|^2 + o(\|x-y\|^2))$$

(for instance the first one used above), the local approximation

$$(4) \quad K(x,y) \approx \sigma^2(1 - \alpha \|x-y\|^2)$$

cannot be used as we did above with the representation (3). Indeed, in general, the matrix K corresponding to (4) is singular. This

is because for any n points $x_1, x_2, \dots, x_n$ in R^2 , with $n \geq 5$, the $n \times n$ matrix $[\|x_i - x_j\|^2]$ is singular. This result and a generalization are proved in the first section of the Appendix.

5.5. An Estimation Paradox

When the random field $\{Z(x), x \in R^2\}$ has constant mean m (unknown), and its covariance function $K(x,y)$ satisfies

$$(1) \quad \begin{cases} K(x,x) = \sigma^2 & \text{for all } x \in R^2 \\ K(x,y) \geq 0 & \text{for all } x,y \\ K(x,y) \rightarrow 0 & \text{as } \|x-y\| \rightarrow +\infty, \end{cases}$$

one would think that the maximum mean squared estimation error occurs for estimation of very distant points; i.e., one would suppose that

$$\lim_{\|x\| \rightarrow +\infty} \text{Var}(\hat{Z}(x) - Z(x)) ,$$

which is equal to

$$\sigma^2 + \frac{1}{e' K^{-1} e} ,$$

is also equal to

$$\max_{x \in R^2} \text{Var}(\hat{Z}(x) - Z(x)) .$$

This would further support the use of the criterion $e' K^{-1} e$.

But this is not true. The crux of the matter is that the inequality

$$\text{Var}(\hat{Z}(x) - Z(x)) \leq \lim_{\|y\| \rightarrow +\infty} \text{Var}(\hat{Z}(y) - Z(y))$$

is equivalent to

$$\sigma^2 - k' K^{-1} k + \frac{(e' K^{-1} k - 1)^2}{e' K^{-1} e} \leq \sigma^2 + \frac{1}{e' K^{-1} e}$$

which in turn is equivalent to

$$(2) \quad (e' K^{-1} k)^2 - 2 e' K^{-1} k \leq (k' K^{-1} k)(e' K^{-1} e) .$$

Formula (2) is close to Cauchy-Schwarz inequality, but it is possible to find K and k , satisfying the set of conditions (1) on $K(x,y)$, such that the reverse of (2) holds.

For example, consider three station-points x_1, x_2, x_3 with covariance matrix

$$K = \begin{pmatrix} 1 & \alpha & 0 \\ \alpha & 1 & \alpha \\ 0 & \alpha & 1 \end{pmatrix} .$$

Let x be a point where we want to estimate $Z(x)$, and y another one farther away, such that

$$\text{Var}(Z(x)) = \text{Var}(Z(y)) = 1$$

$$\text{cov} \begin{pmatrix} Z(x_1) \\ Z(x_2) \\ Z(x_3) \end{pmatrix}, Z(x) = \begin{pmatrix} 0 \\ \beta \\ 0 \end{pmatrix} = k$$

$$\text{Cov}(Z(x_i), Z(y)) = 0 \quad i = 1, 2, 3 .$$

Let us find α and β satisfying

$$(3a) \quad \alpha > 0, \quad \beta > 0,$$

$$(3b) \quad \begin{bmatrix} K & k \\ k' & 1 \end{bmatrix} \text{ positive definite},$$

$$(3c) \quad \text{Var}(\hat{Z}(x) - Z(x)) > \text{Var}(\hat{Z}(y) - Z(y)).$$

With some calculations we obtain

$$e' K^{-1} e = \frac{3 - 4\alpha}{1 - 2\alpha^2}, \quad e' K^{-1} k = \frac{(1 - 2\alpha)\beta}{1 - 2\alpha^2},$$

$$k' K^{-1} k = \frac{\beta^2}{1 - 2\alpha^2}.$$

$$\text{Condition (3b) is equivalent to } \begin{cases} 1 - 2\alpha^2 > 0 \\ 1 - \frac{\beta^2}{1 - 2\alpha^2} > 0 \end{cases}.$$

Condition (3c) can be rewritten

$$\left(\frac{(1 - 2\alpha)\beta}{1 - 2\alpha^2} \right)^2 - 2 \frac{(1 - 2\alpha)\beta}{1 - 2\alpha^2} > \left(\frac{3 - 4\alpha}{1 - 2\alpha^2} \right) \left(\frac{\beta^2}{1 - 2\alpha^2} \right)$$

and simplifies to $\beta < 2\alpha - 1$.

Finally, the three conditions (3) are equivalent to

$$(4a) \quad \alpha > 0, \quad \beta > 0,$$

$$(4b) \quad 2\alpha^2 + \beta^2 < 1,$$

$$(4c) \quad \beta < 2\alpha - 1.$$

In the (α, β) -plane the solution region is represented in Figure 1.

Figure 1. Solution region to system (4).

For instance, if

$$(5) \quad K = \begin{bmatrix} 1 & \frac{2}{3} & 0 \\ \frac{2}{3} & 1 & \frac{2}{3} \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \quad \text{and} \quad k = \begin{bmatrix} 0 \\ \frac{1}{4} \\ 0 \end{bmatrix},$$

to fix ideas let us suppose x_1, x_2, x_3, x , and y are as represented in Figure 2.

Figure 2. Example of five points in the plane illustrating the estimation paradox.

Then

$$\text{Var}(\hat{Z}(x) - Z(x)) = \frac{35}{24}$$

and

$$\text{Var}(\hat{Z}(y) - Z(y)) = \frac{32}{24} .$$

Unsolved question: Does there exist a stationary (or even stationary isotropic) covariance function $K(x,y)$ consistent with K and k given by (5) and consistent with the geometry of Figure 2?

A study of a class of isotropic covariance functions vanishing at a finite distance, which was developed in an attempt to settle this question, is contained in the second section of the Appendix.

5.6. Study of Criterion $\int_A \text{Var}(\hat{Z}(x) - Z(x))dx$

Recall the notation $s(x; x_1, x_2, \dots, x_n) = \text{Var}(\hat{Z}(x) - Z(x))$ where $\hat{Z}(x)$ is based on the measurements taken at $x_1, x_2, \dots, x_n$. And let

$$t(x_1, x_2, \dots, x_n) = \int_A s(x; x_1, x_2, \dots, x_n)dx$$

be the network criterion introduced in formula 5.3.2 with the ordinary Lebesgue measure in $\mathbb{R}^2$.

It does not seem possible in general to give a closed algebraic form to the function t , because $K(x,y)$ enters the function s in too intricate a way, and because integration over an area A with irregular boundary would be hard. However, numerical approximations on a computer with any specific data are very easy. We will see an example in Section 5.8.

Actually, the criterion t derives its local properties from those of the function s itself. A most interesting one holds when the random field is in a certain sense smooth. Before stating our result we need to recall some facts about the smoothness of stationary random processes according to the form of their covariance functions.

A second-order-stationary random process $\{Z(x), x \in \mathbb{R}\}$ is continuous in quadratic mean if and only if its covariance function $K(x,y)$ (which can then be written $K(y-x)$) is continuous at the origin. It is differentiable in quadratic mean if and only if

$$K(h) = \sigma^2(1 - \frac{a^2}{2} h^2 + o(h^2)) \text{ as } h \rightarrow 0.$$

Of course these results still hold if the process instead of being second-order-stationary has a differentiable mean function and a stationary covariance function.

These results extend to a random field $\{Z(x), x \in \mathbb{R}^2\}$ with differentiable mean function and isotropic covariance function (that is $K(x,y)$ function only of $\|y - x\|$).

Theorem 5.6.1. If a random field $\{Z(x), x \in \mathbb{R}^2\}$, with differentiable mean function and isotropic covariance function, is differentiable in quadratic mean, then the function $s(x; x_1, x_2, \dots, x_n, y)$ (that is $\text{Var}(\hat{Z}(x) - Z(x))$ where $\hat{Z}$ is based on $x_1, x_2, \dots, x_n, y$) is not continuous in y , in the sense that

$$\lim_{y \rightarrow x_n} s(x; x_1, x_2, \dots, x_n, y) \neq s(x; x_1, x_2, \dots, x_n)$$

Indeed, we have

$$(1) \quad \lim_{y \rightarrow x_n} \frac{s(x; x_1, x_2, \dots, x_n, y)}{s(x; x_1, x_2, \dots, x_n)} = 1 - \rho^{*2}$$

(with fixed direction)

with

$$(2) \quad \rho^* = \kappa \frac{w'(x_n - y)}{[(x_n - y)' A (x_n - y)]^{1/2}}$$

where κ is a constant, w is a vector, and A is a positive definite quadratic form, each depending only on $x, x_1, x_2, \dots, x_n$.

In the case $n=1$, formula (2) simplifies to

$$(3) \quad \rho^* = \kappa \cos(\widehat{y x_n x})$$

where $\widehat{y x_n x}$ is the angle between the directions of $y - x_n$ and $x - x_n$.

If the random field is not differentiable but continuous in quadratic mean and $K(h) = \sigma^2(1 - ah + o(h))$, $a > 0$, then $s(x; x_1, x_2, \dots, x_n, y)$ is continuous in y in the sense indicated above.

Finally, if the random field is not continuous in q.m., then again the function $s(x; x_1, x_2, \dots, x_n, y)$ is not continuous in y .

Proof. If the random field, with differentiable mean function and isotropic covariance function, is differentiable then

$$K(x, y) = \sigma^2(1 - \frac{a^2}{2} \|x - y\|^2 + o(\|x - y\|^2))$$

With no loss of generality, let's take $\sigma = 1$.

In theorem 5.2.1 we proved that

$$(4) \quad \frac{s(x; x_1, x_2, \dots, x_n, y)}{s(x; x_1, x_2, \dots, x_n)} = 1 - \rho^2(x, y)$$

where

$$(5) \quad \rho(x, y) = \frac{\text{cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y))}{[\text{Var}(\hat{Z}(x) - Z(x))\text{Var}(\hat{Z}(y) - Z(y))]^{\frac{1}{2}}},$$

$\hat{Z}$ being computed on the basis of $x_1, x_2, \dots, x_n$. To prove theorem 5.6.1 we shall derive the expansion of $\rho(x, y)$, when $y \rightarrow x_n$,

by computing those of

$$(6) \quad \text{Var}(\hat{Z}(y) - Z(y)) = 1 - k_y' K^{-1} k_y + \phi_y' G^{-1} \phi_y$$

and

$$(7) \quad \text{Cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y)) = K(x, y) - k_x' K^{-1} k_y + \phi_x' G^{-1} \phi_y.$$

We need a Taylor expansion of $K(x, y)$ when $y \rightarrow x_n$. To do this it will be convenient to use for $K(x, y)$ the form

$$(8) \quad K(x, y) = R\left(\frac{a^2}{2} \|x - y\|^2\right).$$

So we have $R(t) = 1 - t + o(t)$ as $t \downarrow 0$. Next note that

$$(9) \quad \|x - y\|^2 = \|x - x_n\|^2 + 2(x - x_n)'(x_n - y) + \|x_n - y\|^2,$$

hence

$$\begin{aligned} R\left(\frac{a^2}{2} \|x - y\|^2\right) &= R\left(\frac{a^2}{2} \|x - x_n\|^2\right) + \\ &\quad \frac{a^2}{2} [2(x - x_n)'(x_n - y) + \|x_n - y\|^2] R^{[1]}\left(\frac{a^2}{2} \|x - x_n\|^2\right) + \\ &\quad o(\|x_n - y\|), \end{aligned}$$

($R^{[1]}$ denotes the first derivative of the function R). From (8)

we conclude that for all x

$$(10) \quad K(x, y) = K(x, x_n) + a^2(x - x_n)'(x_n - y) R^{[1]}\left(\frac{a^2}{2} \|x - x_n\|^2\right) + o(\|x_n - y\|),$$

while for x_n

$$(11) \quad K(x_n, y) = 1 - \frac{a^2}{2} \|x_n - y\|^2 + o(\|x_n - y\|^2) .$$

First we apply (10) and (11) to

$$k_y = \begin{pmatrix} K(x_1, y) \\ \vdots \\ K(x_{n-1}, y) \\ K(x_n, y) \end{pmatrix} .$$

With the following notation

$$(12) \quad \Omega = a^2 \begin{bmatrix} (x_1 - x_n)', R^{[1]} \left(\frac{a^2}{2} \|x_1 - x_n\|^2 \right) \\ \vdots \\ (x_{n-1} - x_n)', R^{[1]} \left(\frac{a^2}{2} \|x_{n-1} - x_n\|^2 \right) \\ - \frac{1}{2} (x_n - y)' \end{bmatrix}$$

and

$$(13) \quad \epsilon = \begin{pmatrix} o(\|x_n - y\|) \\ \vdots \\ o(\|x_n - y\|) \\ o(\|x_n - y\|^2) \end{pmatrix} ,$$

we can write

$$(14) \quad k_y = k_{x_n} + \Omega(x_n - y) + \epsilon .$$

Secondly, applying (14) to $\phi_y = FK^{-1}k_y - f_y$ we obtain

$$\begin{aligned}
 \phi_y &= FK^{-1}k_{x_n} + FK^{-1}\Omega(x_n - y) + FK^{-1}\epsilon - f_y \\
 &= f_{x_n} + FK^{-1}\Omega(x_n - y) + FK^{-1}\epsilon - f_y \\
 (15) \quad &= (FK^{-1}\Omega + f_{x_n}^{[1]})(x_n - y) + \eta
 \end{aligned}$$

where $f_{x_n}^{[1]} = [\text{grad } f_y]_{y=x_n}$, and η is a p -dimensional vector, each component of which is $o(\|x_n - y\|)$.

With these preliminaries, now we can compute the variance and covariance given in (6) and (7). Using (14) and (15) we obtain, on the first hand,

$$\begin{aligned}
 \text{Var}(\hat{Z}(y) - Z(y)) &= 1 - (k_{x_n}' + \Omega(x_n - y) + \epsilon)' K^{-1} (k_{x_n} + \Omega(x_n - y) + \epsilon) \\
 (16) \quad &+ [(FK^{-1}\Omega + f_{x_n}^{[1]})(x_n - y) + \eta]' G^{-1} [(FK^{-1}\Omega + f_{x_n}^{[1]})(x_n - y) + \eta].
 \end{aligned}$$

Since $k_{x_n}' K^{-1} k_{x_n} = 1$, $k_{x_n}' K^{-1} \Omega = -\frac{a^2}{2}(x_n - y)'$ and

$k_{x_n}' K^{-1} \epsilon = o(\|x_n - y\|^2)$, the R.H.S. of (16) becomes

$$\begin{aligned}
 &a^2 \|x_n - y\|^2 - (x_n - y)' \Omega' K^{-1} \Omega (x_n - y) + o(\|x_n - y\|^2) + \\
 (17) \quad &(x_n - y)' [FK^{-1}\Omega + f_{x_n}^{[1]}]' G^{-1} [FK^{-1}\Omega + f_{x_n}^{[1]}] (x_n - y) + o(\|x_n - y\|^2)
 \end{aligned}$$

or in condensed notation

$$(18) \quad (x_n - y)' A (x_n - y) + o(\|x_n - y\|^2),$$

where A (which depends only on $x_1, x_2, \dots, x_n$ and x) is a positive definite quadratic form since (18) is the expansion of a variance. On the other hand, we have

$$(19) \quad \text{Cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y)) = K(x, y) - k'_x K^{-1} (k_{x_n} + \Omega(x_n - y) + \epsilon) + \phi'_x G^{-1} [(FK^{-1} \Omega + f_{x_n}^{[1]})(x_n - y) + \eta].$$

Using (10) and $k'_x K^{-1} k_{x_n} = K(x, x_n)$, the R.H.S. of (19) simplifies to

$$(20) \quad a^2 (x - x_n)' (x_n - y) R^{[1]} \left(\frac{a^2}{2} \|x - x_n\|^2 \right) + o(\|x_n - y\|) - k'_x K^{-1} \Omega(x_n - y) - k'_x K^{-1} \epsilon + \phi'_x G^{-1} [FK^{-1} \Omega + f_{x_n}^{[1]}](x_n - y) + \phi'_x G^{-1} \eta$$

or in condensed notation

$$(21) \quad w'(x_n - y) + o(\|x_n - y\|),$$

where w is a vector depending only on $x_1, x_2, \dots, x_n$ and x .

Now, substituting (18) and (21) in formula (5) we obtain

$$(22) \quad \rho(x, y) = \frac{w'(x_n - y) + o(\|x_n - y\|)}{\{\text{Var}(\hat{Z}(x) - Z(x)) [(x_n - y)' A(x_n - y) + o(\|x_n - y\|^2)]\}^{\frac{1}{2}}}.$$

The first part of the theorem, i.e. formula (2), follows from (22) when we let y tend to x_n .

In the case $n=1$, to establish the more explicit formula (3), we must go back to formulas (17) and (20). Observe that in this case

$$a) \quad \Omega = -a^2(x_n - y)'$$

$$b) \quad F = 1 \quad \text{and} \quad f_{x_n}^{[1]} = 0$$

$$c) \quad K, G, K^{-1}, \text{ and } G^{-1} \text{ are equal to } 1,$$

hence (17) becomes

$$(23) \quad \text{Var}(\hat{Z}(y) - Z(y)) = a^2 \|x_n - y\|^2 + o(\|x_n - y\|^2)$$

and (20) becomes

$$(24) \quad \text{Cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y)) =$$

$$a^2(x - x_n)'(x_n - y)R^{[1]}(\frac{a^2}{2} \|x - x_n\|^2) + o(\|x_n - y\|).$$

Since we have

$$[a^2 \|x_n - y\|^2 + o(\|x_n - y\|^2)]^{\frac{1}{2}} = a \|x_n - y\| + o(\|x_n - y\|),$$

from (23) and (24), we conclude that

$$\rho(x, y) = \kappa \frac{(x - x_n)'(y - x_n)}{\|x - x_n\| \|y - x_n\|} + o(\|x_n - y\|),$$

where κ is a scalar independent of y . This completes the proof of formula (3).

Some more calculations, that we omit, show that in the case $n=1$ the constant κ is actually equal to

$$- \frac{a \|x - x_n\| R^{[1]} \left(\frac{a^2}{2} \|x - x_n\|^2 \right)}{\left[2 \left(1 - R \left(\frac{a^2}{2} \|x - x_n\|^2 \right) \right) \right]^{\frac{1}{2}}} = \frac{d}{du} \left[(1 - R(u^2))^{\frac{1}{2}} \right],$$

where

$$u = \frac{a}{\sqrt{2}} \|x - x_n\|.$$

We now consider the case when the random field is not differentiable but continuous in q.m. and $K(h) = 1 - ah + o(h)$, with $a > 0$. To establish the continuity of $s(x; x_1, x_2, \dots, x_n, y)$ we use the same scheme as before. The important difference is that $\|x - y\|$ plays the role of $\|x - y\|^2$ and formula (9) is replaced by

$$\begin{aligned} \|x - y\| &= [\|x - x_n\|^2 + 2(x - x_n)'(x_n - y) + \|x_n - y\|^2]^{\frac{1}{2}} \\ &= \|x - x_n\| \left[1 + \frac{2(x - x_n)'(x_n - y)}{\|x - x_n\|^2} + \frac{\|x_n - y\|^2}{\|x - x_n\|^2} \right]^{\frac{1}{2}} \\ (25) \quad &= \|x - x_n\| \left[1 + \frac{(x - x_n)'(x_n - y)}{\|x - x_n\|^2} + o(\|x_n - y\|) \right] \\ &= \|x - x_n\| + \frac{(x - x_n)'(x_n - y)}{\|x - x_n\|} + o(\|x_n - y\|). \end{aligned}$$

Formulas (10) and (11) become the following: For all x

$$(26) \quad K(x, y) = K(x, x_n) + a \frac{(x - x_n)'(x_n - y)}{\|x - x_n\|} R^{[1]}(a\|x - x_n\|) + o(\|x_n - y\|)$$

while for x_n

$$(27) \quad K(x_n, y) = 1 - a\|x_n - y\| + o(\|x_n - y\|).$$

To express k_y we use the new notation

$$(28) \quad \Omega_1 = a \begin{bmatrix} (x_1 - x_n)' \frac{R^{[1]}(a \|x_1 - x_n\|)}{\|x_1 - x_n\|} \\ \vdots \\ (x_{n-1} - x_n)' \frac{R^{[1]}(a \|x_{n-1} - x_n\|)}{\|x_{n-1} - x_n\|} \\ - \frac{(x_n - y)'}{\|x_n - y\|} \end{bmatrix}$$

and

$$(29) \quad \epsilon_1 = \begin{pmatrix} o(\|x_n - y\|) \\ \vdots \\ o(\|x_n - y\|) \\ o(\|x_n - y\|) \end{pmatrix}.$$

The difference between ϵ given by (13) and ϵ_1 given by (29) is that the last component of ϵ_1 is $o(\|x_n - y\|)$ instead of being $o(\|x_n - y\|^2)$.

Formulas (14) and (15) still hold with Ω_1 and ϵ_1 substituting for Ω and ϵ . But the calculations made in (16), (17), and (18) are replaced by

$$(30) \quad \text{Var}(\hat{Z}(y) - Z(y)) = a \|x_n - y\| + o(\|x_n - y\|).$$

On the other hand, formula (21), giving the development of $\text{Cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y))$ is still true. Therefore

$$\text{Cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y)) = O(\|x_n - y\|) ,$$

whereas, from (30),

$$[\text{Var}(\hat{Z}(y) - Z(y))]^{\frac{1}{2}} = O(\|x_n - y\|^{\frac{1}{2}}) .$$

We conclude that

$$\lim_{\|y - x_n\| \rightarrow 0} \rho(x, y) = 0 ,$$

and this, in view of (4), establishes the continuity of

$s(x; x_1, x_2, \dots, x_n, y)$.

Lastly, when the random field is not continuous in quadratic mean, the covariance function $K(h)$ is not continuous at the origin. Then neither $\text{Cov}(\hat{Z}(x) - Z(x), \hat{Z}(y) - Z(y))$ nor $\text{Var}(\hat{Z}(y) - Z(y))$ tend to zero when $\|y - x_n\|$ tends to zero. Hence $\rho(x, y) \not\rightarrow 0$ and s is not continuous. Q.E.D.

Before interpreting the theorem just proved, we observe that expressions (17) and (20) are often well approximated by expressions (23) and (24). As a consequence, the formula

$$\rho^* = \kappa \cos(\widehat{y x_n x})$$

is an approximation, even when $n > 1$, to the more complicated formula (2) which enters expression (1), giving the reduction of uncertainty at x when a station y is added in the vicinity of x_n . This will be the case in the example of Section 5.8.

The interpretation of Theorem 5.6.1 is that when station y is very close to x_n three possibilities can arise:

1. $K(h)$ has a jump at the origin and Z can be decomposed into two parts

$$Z = \zeta + \epsilon ,$$

where ζ is a random field with isotropic covariance function continuous at the origin, and ϵ is an independent stationary white noise. Then the couple $\{Z(y), Z(x_n)\}$ provides information on the variance of ϵ that $Z(x_n)$ alone does not provide.

2. $K(h) = 1 - ah + o(h)$, Z is continuous in q.m. but not differentiable. Then $Z(y)$ and $Z(x_n)$, in the limit, measure the same thing and do not provide more information than $Z(x_n)$ alone.

3. However, if $K(h) = 1 - \frac{a}{2} h^2 + o(h^2)$, Z is differentiable, and then $\{Z(y), Z(x_n)\}$ provides information on the partial derivative of Z at x_n in the direction of $y - x_n$. In the case $n=1$ the reduction of uncertainty about $Z(x)$ at any x is proportional to the square of the cosine of the angle between $y - x_1$ and $x - x_1$.

Note: The three cases considered above do not exhaust the possible forms of isotropic covariance functions in R^2 . However, the cases left over are pathological cases such as, for instance, a function $K(h)$ for which

$$\lim_{h \rightarrow 0} \sup K(h) \neq \lim_{h \rightarrow 0} \inf K(h) ,$$

or, an isotropic covariance function continuous but not differentiable at the origin. Such covariance functions are not used in practice. To insist on the pathological character of the above examples, we observe that in R^n , $n \geq 3$, isotropic covariance

functions, which are continuous at the origin, are then differentiable everywhere on $[0, +\infty)$, at the origin this being a right derivative (see Schoenberg (1938), Lemma 4, page 822).

When $s(x; x_1, x_2, \dots, x_n, y)$ is continuous in y , by the dominated convergence theorem, this carries over to its average $t(x_1, x_2, \dots, x_n, y)$. When s is discontinuous, since for all x

$$s(x; x_1, x_2, \dots, x_n, y) \leq s(x; x_1, x_2, \dots, x_n) ,$$

$t(x_1, x_2, \dots, x_n, y)$ is discontinuous too.

The preceding interpretation can be given the following extension, that I have not proved: When the random field Z is differentiable in quadratic mean, three measurements $\{Z(x_n), Z(y_1), Z(y_2)\}$, where y_1 and y_2 are very close to x_n , and the directions of $y_1 - x_n$ and $y_2 - x_n$ are different, for instance perpendicular, will provide all possible information on the first-order partial derivatives of Z at x_n . They will determine the tangent plane. Moreover, in the limit the reduction of uncertainty on $Z(x)$ at any point x will be independent of the directions $y_1 - x_n$, $y_2 - x_n$, provided they are not the same. This effect is indeed observed in the example of Section 5.8.

If the random field Z does not have curvature in quadratic mean, then four measurements around x_n will be redundant, in the same way as two are if Z is continuous but not differentiable in quadratic mean.

It should be noted that to assume that our measurements come from a smooth random field Z rules out models including measurement

errors, because they would entail a discontinuous covariance function; for instance, an isotropic covariance function $K(h)$ would have a jump at the origin.

Finally, Theorem 5.6.1 generalizes to covariance functions which instead of being isotropic with respect to the Euclidean norm, are isotropic with respect to another norm defined by a positive definite quadratic form. The expression of the results and the proof remain unchanged as long as everything is understood with respect to the new norm. Theorem 5.6.1 generalizes furthermore to covariance functions which are not isotropic at all. In particular, to generalize the result relative to $K(h) = 1 - \frac{a^2}{2} h^2 + o(h^2)$, the only important property they must have is to be differentiable in the neighborhood of every pair (x, x) and flat at every such pair; but then with such general conditions on $K(x, y)$ the discontinuity of $s(x; x_1, x_2, \dots, x_n, y)$ and the reduction of uncertainty on $Z(x)$ have no longer simple forms.

5.7. A Suggestion in Network Design for Estimation of Smooth Random Fields

In the design of networks used to monitor smooth random phenomena like, for instance, barometric pressure, or variation of temperature over an area in high altitude, the discontinuity of $s(x; x_1, x_2, \dots, x_n, y)$ ($= \text{Var}(\hat{Z}(x) - Z(x))$), with $\hat{Z}(x)$ based on the $n+1$ stations $x_1, x_2, \dots, x_n, y$ and of its average $t(x_1, x_2, \dots, x_n, y) = \int_A s(x; x_1, x_2, \dots, x_n, y) dx$, when $y \rightarrow x_n$, is of practical interest.

If we want to improve an existing network in a given area the first idea, if we are to use the criterion t , is to find the point y in A minimizing $t(x_1, x_2, \dots, x_n, y)$ and build a new station there.

However, we can generally do better in terms of the same criterion by taking, at each already existing stations, three measurements close to each other and forming an angle (optimally of 90 degrees). This would require only some minor modifications in the routine of taking measurements at stations, and would altogether cost much less money than to build a new station.

5.8. A Numerical Example of Estimation of a Smooth Random Field on Artificial Data

In this section we give an illustration of estimation of a random field over an area using pointwise the best linear unbiased estimator $\hat{Z}(x)$ studied in Chapters 2 and 3, and in the preceding sections of Chapter 5. We shall develop the example introduced in Section 5.4, when studying criterion $e' K^{-1} e$.

So consider a random field $\{Z(x), x \in \mathbb{R}^2\}$ with constant mean function m (unknown) and stationary isotropic covariance function

$$K(x, y) = e^{-\frac{1}{10} \|x-y\|^2};$$

suppose we are interested in estimating the map of Z over the square area A , with six monitoring stations, represented in Figure 1 on the next page.

We first illustrate the function $s(x; x_1, x_2, \dots, x_n) =$ variance of the estimation error $\hat{Z}(x) - Z(x)$, and the function

$t(x_1, x_2, \dots, x_n) = \int_A s(x; x_1, x_2, \dots, x_n) dx$. Then we illustrate the principal result of Theorem 5.6.1 about the discontinuity of $s(x; x_1, x_2, \dots, x_n, y)$ and $t(x_1, x_2, \dots, x_n, y)$ when $y \rightarrow x_n$. And finally we give support to the discussion of the effect of taking three measurements close to each other at monitoring stations.

Figure 1. Square area A over which we want to construct an estimated map of Z from measurements taken at the stations $x_1, x_2, \dots, x_6$.

Figure 2 on page 84 shows the function

$$s(x; x_1, x_2, \dots, x_6) ,$$

represented in two-dimension by its level-curves of level 0.01, 0.06, 0.11, 0.16, 0.21 The level-curve of level α , also called isopleth of level α , is defined as the set

$$C_{\alpha} = \{x \in \mathbb{R}^2 : s(x; x_1, x_2, \dots, x_6) = \alpha\} .$$

The value of the criterion $\int_A s(x; x_1, x_2, \dots, x_6) dx$ associated to the six-station-network and the area A is

$$t(x_1, x_2, \dots, x_6) = 49.15 .$$

Since the surface area of A is 100 (see Figure 1), the average value of s over A is .4915; this is in accordance with the shape of s shown in Figure 2.

Notes: 1. The numerical integration of the surface $s(x; x_1, x_2, \dots, x_6)$ to compute $t(x_1, x_2, \dots, x_6)$ was done by dividing the area A into a regular grid of 400 squares, computing the value of s at the top left corner of each square of the grid, summing up the values, and multiplying by $\frac{1}{4}$ which is the surface area of each square.

2. To draw the level curves of s , shown in Figure 2, I took only the 400 values already computed at the grid points, and designed a plotting program which from the values at these points extrapolates the curves in a simple manner, by barycentric

interpolation of s in each square of the grid. (This actually left over the 39 squares which form the bottom row and the right column of the grid of 400 squares, since I did not compute the value of s at each of their corners; but it is irrelevant for the purpose of the illustration.)

3. The method of interpolation accounts for the artificially jagged look of some of the curves. In particular, the concavities of the isopleth .01 near stations x_3, x_4 are not accurate, although they are the natural effect of barycentric interpolation. Indeed, a surface which is in reality like (a) will be interpreted as (b).

(a)

(b)

Figure 2. Function $s(x; x_1, x_2, \dots, x_6) = \text{Var}(\hat{Z}(x) - Z(x))$, represented by its level curves in the square area A .

Note: $\lim_{\|x\| \rightarrow +\infty} s(x; x_1, x_2, \dots, x_6) = 1.3696$.

Figure 3 on page 87 shows the function

$$t(x_1, x_2, \dots, x_6, y)$$

as a function of y . That is, for a given point y , the quantity $t(x_1, x_2, \dots, x_6, y)$ is the value of the criterion $\int_A s dx$ for the network of seven stations $x_1, x_2, \dots, x_n, y$.

To plot Figure 3 I computed $t(x_1, x_2, \dots, x_6, y)$ for y at each of the 400 points which I had already used to plot Figure 2. And for each point y , to compute $t(x_1, x_2, \dots, x_6, y)$, in turn, I used the method explained in Note 1 on page 82. Altogether, this involved the computation of $s(x; x_1, x_2, \dots, x_6, y)$ 160 000 times. To do this one can

- either, for each y , compute the covariance matrix of $(Z(x_1), Z(x_2), \dots, Z(x_6), Z(y))$ corresponding to the matrix K , and do all the subsequent computations with seven stations instead of six.

- or use the updating Theorem 5.2.1 proved in Section 5.2.

I used both methods to compare their speeds and also to check my results: With the first method the computations for Figure 3 take 30 minutes on a PDP 11/34 computer, whereas with the second method they take only 10 minutes. This shows the practical interest of Theorem 5.2.1.

Figure 3 gives the location of the point y which minimizes $t(x_1, x_2, \dots, x_n, y)$. It is near the bottom left corner of the map, where the criterion reaches the value 38.0.

However, the most interesting feature of Figure 3 is the display of the discontinuity of $t(x_1, x_2, \dots, x_6, y)$ when y tends to one of the already existing stations. This phenomenon is best illustrated at station x_6 , and requires comment.

The plotting program had difficulty representing the discontinuities of t on a large scale map. In order to improve on the representation of the discontinuity of t at x_6 , we computed its values on a finer grid around this station.

Figure 4 shows the corresponding, more accurate, map of t around station x_6 . Disregarding the plotting oddity arising again at the center, this figure shows nicely the discontinuity of t .

The variation of the improvement of $t(x_1, x_2, \dots, x_6, y)$ on $t(x_1, x_2, \dots, x_6)$ ($= 49.15$), according to the direction of $y - x_6$, is very clear. The local symmetry of $t(x_1, x_2, \dots, x_6, y)$, when y is close to x_6 , is particularly striking: This comes from the fact that when y is very close to x_6 the two cases

$$\arg(y - x_6) = \theta$$

and

$$\arg(y - x_6) = \theta + \pi$$

are identical because the position of the unordered pair $\{y, x_6\}$ (which is all that matters) is the same.

Figure 5 shows the function $s(x; x_1, x_2, \dots, x_6, x_7)$ when the seventh station is added at its optimal location, obtained from Figure 3.

Figure 3. Function $t(x_1, x_2, \dots, x_6, y)$ as a function of y .

$$t(x_1, x_2, \dots, x_6, y) = \int_A \text{Var}(\hat{Z}(x) - Z(x)) dx$$

where $\hat{Z}$ is calculated from the measurements taken at the seven stations $x_1, x_2, \dots, x_6, y$. The smaller square shows the scale of Figure 4. The level curves are at every .25 units.

Figure 4. Enlargement of Figure 3 around station x_6 showing the function $t(x_1, x_2, \dots, x_6, y)$ with more accuracy. Note the local central symmetry. The level curves are at every .25 units.

Figure 5. Function $s(x; x_1, x_2, \dots, x_6, x_7)$ when the seventh station is added at its optimal location to minimize the criterion $t(x_1, x_2, \dots, x_6, y)$.

Note: $\lim_{\|x\| \rightarrow +\infty} s(x; x_1, x_2, \dots, x_6, x_7) = 1.3118$

Figures 6 and 7 on pages 92 - 93 show the variation of the map of $s(x; x_1, x_2, \dots, x_6, y)$ when y approaches x_6 from the West direction on a straight line.

In Figure 6, $\|y - x_6\| = 1$.

In Figure 7, $\|y - x_6\| = .05$.

The point of these two figures is to show that not only is there discontinuity when $y \rightarrow x_6$, but it is not necessarily a local phenomenon: In our example the function s remains almost unchanged when $\|y - x_6\|$ goes from 1 to .05 with fixed direction.

In practical terms, if the size of the square A is 100 miles by 100 miles, we have no substantial gain in building a new station 10 miles away from x_6 instead of taking another measurement half a mile away, or actually even closer.

Figures 7 and 8 give a conspicuous illustration when y is close to x_6 ($\|y - x_6\| = .05$) of the effect of the direction of $y - x_6$ on the map of $s(x; x_1, x_2, \dots, x_6, y)$.

Note: The covariance function $K(x, y) = \exp(-\frac{1}{10} \|x - y\|^2)$, that we chose, is such that Z has derivatives in quadratic mean of any order. This comes from a simple adaptation of the proof of the following result: A strictly stationary random process $\{Z(x), x \in \mathbb{R}\}$ is differentiable in q.m. up to order n if and only if its correlation function is of the form

$$c(h) = 1 - a_1^2 \frac{h^2}{2!} + a_2^2 \frac{h^4}{4!} + \dots + (-1)^n a_n^2 \frac{h^{2n}}{(2n)!} + o(h^{2n}) \quad .$$

(See Lévy (1948), page 113).

However, the directional effect illustrated by Figures 7 and 8 is not due to the extreme smoothness of our covariance function. It is only due to the fact that $K(h) = \sigma^2(1 - \frac{a^2}{2} h^2 + o(h^2))$; that is, $K(h)$ is twice continuously differentiable at the origin, and its first derivative is zero.

Finally, Figure 9, which should be compared to Figure 2, shows the improvement of taking three measurements close to each other instead of one at station x_6 .

To plot Figure 9, the two extra stations put in the vicinity of $x_6 = (4, -3)$ (see Figure 1 on page 81) were

$$y_1 = (3.93, -3.00)$$

$$y_2 = (4.00, -3.07) .$$

To check the independence of the function s on the orientation of the triplet (x_6, y_1, y_2) we also tried

$$z_1 = (3.95, -2.95)$$

$$z_2 = (3.95, -3.05) ,$$

and obtained an identical map.

In addition, the numerical integration yielded

$$t(x_1, x_2, \dots, x_6, y_1, y_2) = 41.4572$$

$$t(x_1, x_2, \dots, x_6, z_1, z_2) = 41.4747 .$$

Figure 6. Function $s(x; x_1, x_2, \dots, x_6, y)$ when y is one unit away from station x_6 in the West direction.

Figure 7. Function $s(x; x_1, x_2, \dots, x_6, y)$ when y is .05 units away from station x_6 in the West direction. Note that there is no substantial decrease in quality of estimation from Figure 6.

Figure 8. Function $s(x; x_1, x_2, \dots, x_6, y)$ when y is .05 units away from station x_6 in the South direction. Figure 8 together with Figure 7 illustrate the effect of the direction of $y - x_6$, when y is close to x_6 , on the function $s(x; x_1, x_2, \dots, x_6, y)$.

Figure 9. Function $s(x; x_1, x_2, \dots, x_6, y_1, y_2)$ when y_1 and y_2 are .07 units away from station x_6 in the West and South direction respectively. Then $\{Z(x_6), Z(y_1), Z(y_2)\}$ measure the tangent plane, in quadratic mean, to the random field Z at x_6 . This figure should be compared to Figure 2 (page 84).

CHAPTER 6

STEIN-LIKE ESTIMATOR OF A MAP OF Z

The best linear unbiased estimation is admissible when used to estimate Z at a single point, but not in general when used to simultaneously estimate a whole map of Z . In other words, $\hat{Z}(x)$ is admissible for $Z(x)$, but $(\hat{Z}(y_1), \hat{Z}(y_2), \dots, \hat{Z}(y_N))$ is not admissible for $(Z(y_1), Z(y_2), \dots, Z(y_N))$. The object of this chapter is to exhibit an estimator $\tilde{Z}$, of the type introduced by James and Stein (1961), such that

$$\sum_{i=1}^N E(\tilde{Z}(y_i) - Z(y_i))^2 < \sum_{i=1}^N E(\hat{Z}(y_i) - Z(y_i))^2 .$$

In Section 3.2 we saw that the admissibility of $\hat{Z}(x)$ ultimately reduced to the admissibility of $\phi'_x \hat{a}$ to estimate $\phi'_x a$ when $\hat{a} \sim N(a, G^{-1})$. So it is natural that in the present chapter the key result will be the inadmissibility of $\Phi \hat{a}$ to estimate Φa where Φ is an $N \times p$ matrix satisfying certain conditions. This is stated in a general form in the first section.

6.1. A Theorem on Estimation of $C\theta$ with $X \sim N(\theta, I_p)$

Theorem 6.1.1. Consider $X \sim N(\theta, I_p)$, a p -dimensional normal random vector; let C be an $N \times p$ matrix. Then $\exists \eta > 0$ such that for all $\theta \in R^p$

$$(1) \quad E \left\| CX \left(1 - \frac{\eta}{\|X\|^2} \right) - C\theta \right\|^2 < E \|CX - C\theta\|^2 ,$$

if and only if

$$(2) \quad \max \text{ eigenvalue of } CC' < \frac{1}{2} \text{tr } CC'.$$

Remark: Condition (2) implies $\text{rank } C \geq 3$ because if $\text{rank } C \leq 2$ then CC' has at most two nonzero eigenvalues and the greater of the two is certainly at least equal to their mean. In particular, if C is the identity, condition (2) is satisfied if and only if $p \geq 3$, which is a classical fact in Stein estimation (see James and Stein (1961).)

Theorem 6.1.1 will be proved as a consequence of

Theorem 6.1.2. Under the conditions of Theorem 6.1.1, if the random variable K is Poisson $\left(\frac{\|\theta\|^2}{2}\right)$, then

$$(3) \quad E\left\|CX\left(1 - \frac{\eta}{\|X\|^2}\right) - C\theta\right\|^2 = \text{tr } CC' \left[1 - 2\eta E \frac{1}{p+2K} + \eta^2 E \frac{1}{(p+2K)(p-2+2K)}\right] + \frac{\|C\theta\|^2}{\|\theta\|^2} (4\eta + \eta^2) E \frac{2K}{(p+2K)(p-2+2K)}.$$

Proof of Theorem 6.1.2. The proof is computational and uses lemmas 4.2.2 and 4.2.3:

$$(4) \quad \begin{aligned} E\left\|CX\left(1 - \frac{\eta}{\|X\|^2}\right) - C\theta\right\|^2 &= E\left\|C(X-\theta) - \eta \frac{CX}{\|X\|^2}\right\|^2 \\ &= E(X-\theta)' C'C(X-\theta) - 2 E(X-\theta)' C'C \frac{X\eta}{\|X\|^2} + \eta^2 E \frac{X'C'CX}{\|X\|^4}. \end{aligned}$$

Since for two vectors in R^p $x'y = \text{tr } yx'$, expression (4) can be rewritten

$$(5) \quad \text{tr } C E(X-\theta)(X-\theta)' C' - 2\eta \text{tr } C E \frac{X(X-\theta)'}{\|X\|^2} C' + \eta^2 \text{tr } C E \frac{XX'}{\|X\|^4} C' .$$

By lemmas 4.2.2 and 4.2.3 and

$$E \frac{1}{\chi_n^2} = \frac{1}{n-2} , \quad E \left[\frac{1}{\chi_n^2} \right]^2 = \frac{1}{(n-2)(n-4)} ,$$

expression (5) is equal to

$$\begin{aligned} & \text{tr } CC' - 2\eta \text{tr } C \left\{ I_p E \frac{1}{\chi_{p+2K+2}^2} + \theta\theta' E \frac{1}{\chi_{p+2K+4}^2} - \theta\theta' E \frac{1}{\chi_{p+2K+2}^2} \right\} C' + \\ & \quad \eta^2 \text{tr } C \left\{ I_p E \left[\frac{1}{\chi_{p+2K+2}^2} \right]^2 + \theta\theta' E \left[\frac{1}{\chi_{p+2K+4}^2} \right]^2 \right\} C' \\ &= \text{tr } CC' - 2\eta \text{tr } CC' E \frac{1}{p+2K} - 2\eta \|C\theta\|^2 E \left\{ \frac{1}{p+2K+2} - \frac{1}{p+2K} \right\} + \\ & \quad \eta^2 \text{tr } CC' E \frac{1}{(p+2K)(p+2K-2)} + \eta^2 \|C\theta\|^2 E \frac{1}{(p+2K+2)(p+2K)} \\ &= \text{tr } CC' \left[1 - 2\eta E \frac{1}{p+2K} + \eta^2 E \frac{1}{(p+2K)(p+2K-2)} \right] + \\ & \quad \|C\theta\|^2 (4\eta + \eta^2) E \frac{1}{(p+2K+2)(p+2K)} . \end{aligned}$$

Finally, to get expression (3) given in the statement of Theorem 6.1.2 we have to show

$$E \frac{1}{(p+2K+2)(p+2K)} = \frac{1}{\|\theta\|^2} E \frac{2K}{(p+2K)(p+2K-2)} .$$

$$\begin{aligned}
E \frac{1}{(p+2K+2)(p+2K)} &= \sum_{k=0}^{\infty} e^{-\|\theta\|^2/2} \frac{\left(\|\theta\|^2/2\right)^k}{k!} \frac{1}{(p+2k+2)(p+2k)} \\
&= \frac{2}{\|\theta\|^2} \sum_{k=0}^{\infty} e^{-\|\theta\|^2/2} \frac{\left(\|\theta\|^2/2\right)^{k+1}}{(k+1)!} \frac{k+1}{(p+2(k+1))(p+2(k+1)-2)} \\
&= \frac{2}{\|\theta\|^2} \sum_{k=1}^{\infty} e^{-\|\theta\|^2/2} \frac{\left(\|\theta\|^2/2\right)^k}{k!} \frac{k}{(p+2k)(p+2k-2)} \\
&= \frac{1}{\|\theta\|^2} E \frac{2K}{(p+2K)(p+2K-2)},
\end{aligned}$$

which completes the proof of Theorem 6.1.2.

From Theorem 6.1.2 one can easily derive the basic formula of James-Stein estimation, shown below:

$$E\|X\left(1 - \frac{\eta}{\|X\|^2}\right) - \theta\|^2 = p - 2\eta(p-2) E \frac{1}{p-2+2K} + \eta^2 E \frac{1}{p-2+2K}.$$

We can also now prove Theorem 6.1.1.

Proof of Theorem 6.1.1. The key to the proof is to observe that

$$\max_{\theta} \frac{\|C\theta\|^2}{\|\theta\|^2} = \text{max eigenvalue of } CC'.$$

First if

$$(6) \quad \text{max eigenvalue of } CC' \leq \left(\frac{1-\epsilon}{2}\right) \text{tr } CC',$$

then $E\|CX\left(1 - \frac{\eta}{\|X\|^2}\right) - C\theta\|^2$ is bounded by

$$(7) \quad \text{tr } CC' \left\{ 1 - 2\eta \mathbb{E} \frac{1}{p+2K} + \eta^2 \mathbb{E} \frac{1}{(p+2K)(p-2+2K)} + \right. \\ \left. 2(1-\varepsilon)\eta \mathbb{E} \frac{2K}{(p+2K)(p-2+2K)} + (1-\varepsilon)\eta^2 \mathbb{E} \frac{K}{(p+2K)(p-2+2K)} \right\}.$$

In turn to bound expression (7) we need $p \geq 3$; but, as observed in the remark on page 97, this is implied by (6). Then expression (7) is bounded by

$$\text{tr } CC' \left\{ 1 - 2\varepsilon\eta \mathbb{E} \frac{1}{p+2K} + \eta^2 \mathbb{E} \frac{1}{p+2K} \right\}.$$

Therefore if we pick $\eta = \varepsilon$, we get

$$\mathbb{E} \|CX \left(1 - \frac{\varepsilon}{\|X\|^2}\right) - C\theta\|^2 \leq \text{tr } CC' \left(1 - \varepsilon^2 \mathbb{E} \frac{1}{p+2K}\right) < \text{tr } CC' = \mathbb{E} \|CX - C\theta\|^2.$$

Conversely if max eigenvalue of $CC' \geq \frac{1}{2} \text{tr } CC'$ we can choose the direction of θ such that

$$\frac{\|C\theta\|^2}{\|\theta\|^2} = \max \text{ eigenvalue of } CC'$$

and therefore such that

$$\mathbb{E} \|CX \left(1 - \frac{\eta}{\|X\|^2}\right) - C\theta\|^2 \geq \text{tr } CC' \left[1 - 2\eta \mathbb{E} \frac{1}{p+2K} + \right. \\ \left. 2\eta \mathbb{E} \frac{2K}{(p+2K)(p-2+2K)} + \eta^2 \mathbb{E} \frac{K+1}{(p+2K)(p-2+2K)} \right] \\ = \text{tr } CC' \left[1 - 2\eta \mathbb{E} \frac{p-2}{(p+2K)(p-2+2K)} + \eta^2 \mathbb{E} \frac{K+1}{(p+2K)(p-2+2K)} \right].$$

Now, for any $\eta > 0$, there exists θ large enough such that the above quantity is strictly greater than $\text{tr } CC'$, because when $\|\theta\| \rightarrow +\infty$

$$E \frac{p-2}{(p+2K)(p-2+2K)} = o\left(E \frac{K+1}{(p+2K)(p-2+2K)}\right).$$

This completes the proof of Theorem 6.1.1.

6.2. Application to the Estimation of a Map

Consider the problem of estimating Z at N points $y_1, y_2, \dots, y_N$. In practice they will usually form, as in Section 5.8, a regular grid of points over the area A , and the corresponding set of estimated values of Z is what we call the estimated map of Z over A . However, the next theorem does not require that the points be on a grid.

Theorem 6.2.1. With the notation of Chapters 2 and 3, let

$$(1) \quad \tilde{Z}(y_i) = k'_{y_i} K^{-1} \underline{Z} - \phi'_{y_i} \hat{a} \left(1 - \frac{n}{\hat{a}' G \hat{a}}\right)$$

for $i=1, 2, \dots, N$. Then $\exists \eta > 0$ such that, for all a

$$(2) \quad \sum_{i=1}^N E(\tilde{Z}(y_i) - Z(y_i))^2 < \sum_{i=1}^N E(\hat{Z}(y_i) - Z(y_i))^2$$

if and only if

$$(3) \quad \max \text{ eigenvalue of } [\phi'_{y_i} G^{-1} \phi_{y_j}] < \frac{1}{2} \sum_{i=1}^N \phi'_{y_i} G^{-1} \phi_{y_i}$$

where $[\phi'_{y_i} G^{-1} \phi_{y_j}]$ is the $N \times N$ matrix whose (i, j) element is $\phi'_{y_i} G^{-1} \phi_{y_j}$.

Proof. Since $\underline{Z} \sim N(F'a, K)$ we have $K^{-\frac{1}{2}} \underline{Z} \sim N(K^{-\frac{1}{2}} F'a, I_p)$.

Consider $H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$ such that $\mathcal{V}(H_1') = \mathcal{V}(K^{-\frac{1}{2}} F')$ and $HH' = I_p$.

Then

$$HK^{-\frac{1}{2}} \underline{Z} \sim N \begin{pmatrix} H_1 K^{-\frac{1}{2}} F'a & \\ 0 & I_p \end{pmatrix}.$$

Let $M = H_1 K^{-\frac{1}{2}} F'$: it is $p \times p$ and nonsingular since we assumed implicitly in Chapter 2 that $FK^{-1}F'$ is of full rank.

Let $\alpha = Ma$ and $\hat{\alpha} = M\hat{a}$, then $\hat{\alpha} \sim N(M\hat{a}, MG^{-1}M')$. To compute $MG^{-1}M'$ let us start with $M'M$:

$$\begin{aligned} M'M &= FK^{-\frac{1}{2}} H_1' H_1 K^{-\frac{1}{2}} F' \\ &= FK^{-\frac{1}{2}} (I - H_2' H_2) K^{-\frac{1}{2}} F' \\ &= FK^{-1} F' - FK^{-\frac{1}{2}} H_2' H_2 K^{-\frac{1}{2}} F'. \end{aligned}$$

But $H_2 K^{-\frac{1}{2}} F' = 0$; therefore $M'M = G$. Then $G^{-\frac{1}{2}} M'MG^{-\frac{1}{2}} = I_p$. This means that $MG^{-\frac{1}{2}}$ is orthogonal, hence

$$MG^{-1}M' = I_p.$$

In conclusion, we have $\hat{\alpha} \sim N(\alpha, I_p)$.

Following the decomposition of $E(\hat{Z}(x) - Z(x))^2$ given in formula 3.2.2, we can write

$$\begin{aligned} (4) \quad \sum_{i=1}^N E(\tilde{Z}(y_i) - Z(y_i))^2 &= \sum_{i=1}^N E(Z(y_i) - E^Z Z(y_i))^2 + \\ &\quad \sum_{i=1}^N E(E^Z Z(y_i) - \tilde{Z}(y_i))^2. \end{aligned}$$

The first term of the right hand side of (4) is $\sum_{i=1}^N (\sigma_{y_i}^2 - k'_{y_i} K^{-1} k_{y_i})$.

To reexpress the second term of the right hand side of (4), let us remember that

$$E^Z Z(y_i) = f'_{y_i} a + k'_{y_i} K^{-1} [Z - F'a] = k'_{y_i} K^{-1} Z - \phi'_{y_i} a$$

while $\tilde{Z}(y_i) = k'_{y_i} K^{-1} Z - \phi'_{y_i} \hat{a} \left(1 - \frac{\eta}{\hat{a}' G \hat{a}}\right)$. So the second term can be rewritten

$$(5) \quad \sum_{i=1}^N E \left(\phi'_{y_i} \hat{a} \left(1 - \frac{\eta}{\hat{a}' G \hat{a}}\right) - \phi'_{y_i} a \right)^2.$$

Note that $\|\hat{\alpha}\|^2 = \hat{\alpha}' \hat{\alpha} = \hat{a}' M' M \hat{a} = \hat{a}' G \hat{a}$, and let C be the $N \times p$ matrix whose i^{th} row is $\phi'_{y_i} M$. Then the expression (5) becomes

$$\sum_{i=1}^N E \left(\phi'_{y_i} M \hat{\alpha} \left(1 - \frac{\eta}{\|\hat{\alpha}\|^2}\right) - \phi'_{y_i} M \alpha \right)^2 = E \left\| C \hat{\alpha} \left(1 - \frac{\eta}{\|\hat{\alpha}\|^2}\right) - C \alpha \right\|^2.$$

Since $\hat{\alpha} \sim N(\alpha, I_p)$, by Theorem 6.1.1 we know that $\exists \eta > 0$ such that for all α

$$E \left\| C \hat{\alpha} \left(1 - \frac{\eta}{\|\hat{\alpha}\|^2}\right) - C \alpha \right\|^2 < E \|C \hat{\alpha} - C \alpha\|^2$$

if and only if max eigenvalue of $CC' < \frac{1}{2} \text{tr } CC'$.

To finish the proof let us observe on the one hand that $CC' = [\phi'_{y_i} G^{-1} \phi_{y_j}]$, and on the other hand (by going backward in the computations) that

$$E \|C \hat{\alpha} - C \alpha\|^2 = \sum_{i=1}^N E (E^Z Z(y_i) - \hat{Z}(y_i))^2. \quad \text{Q.E.D.}$$

Remark: The $N \times N$ matrix $[\phi'_{y_i} G^{-1} \phi_{y_j}] = CC'$ may be quite cumbersome for computing max eigenvalue of $[\phi'_{y_i} G^{-1} \phi_{y_j}]$. In practice it is much more convenient to first compute $C'C$, which is of size $p \times p$ and has the same nonzero eigenvalues as CC' , and then find max eigenvalue of $C'C$.

We should note that Theorem 6.2.1 gives equal weight to the mean squared errors at each point y_i . The consequence of this is that the absolute lower bound on $\sum_{i=1}^N E(\hat{Z}(y_i) - Z(y_i))^2$, where $\hat{Z}$ represents any estimator, is

$$(6) \quad \sum_{i=1}^N E(E^Z Z(y_i) - Z(y_i))^2 = \sum_{i=1}^N (\sigma_{y_i}^2 - k'_{y_i} K^{-1} k_{y_i})^2.$$

The difference between $\sum_{i=1}^N E(\hat{Z}(y_i) - Z(y_i))^2$ and the lower bound given by (6) is the term

$$\sum_{i=1}^N \phi'_{y_i} G^{-1} \phi_{y_i}.$$

The next section contains an example which gives an idea of the relative sizes of these three terms.

We also note that equal weights prevent us from using our knowledge of the correlations between the $Z(y_i)$'s.

One may instead want to minimize

$$\sum_{i=1}^N w_i E(\hat{Z}(y_i) - Z(y_i))^2$$

with different weights for different points, but this idea was not pursued because it is hard to find a convincing justification for these unequal weights in practice.

6.3. Further Comments and an Example

In the preceding section we saw that the estimation of a whole map using the B.L.U.E. is inadmissible if

$$(1) \quad \text{max eigenvalue of } [\phi'_{y_i} G^{-1} \phi_{y_j}] < \frac{1}{2} \sum_{i=1}^N \phi'_{y_i} G^{-1} \phi_{y_i}.$$

For reasons of rank, already mentioned, this condition requires that the vector of unknown parameters a be of dimension at least 3. This is the case if we assume, for instance, that the mean function of Z is a plane.

It would be of interest to obtain results saying under which general conditions of a geometric character on the points y_i 's relation (1) holds.

Also, some further admissibility questions are left unsolved by Theorem 6.2.1:

- if the dimension of the vector parameter a is 2, does this imply that the estimation of a map using the B.L.U.E. is admissible?
- or, more generally, does

$$\text{max eigenvalue of } [\phi'_{y_i} G^{-1} \phi_{y_i}] \geq \frac{1}{2} \sum_{i=1}^N \phi'_{y_i} G^{-1} \phi_{y_i}$$

imply this admissibility?

The interest of Theorem 6.2.1 is to show that admissibility is not to be expected in general. Let us now see a simple numerical example. Consider a random field $\{Z(x), x \in \mathbb{R}^2\}$ with mean function

$$m(x) = a_1 + a_2 u + a_3 v$$

and covariance function

$$K(x,y) = \begin{cases} 1 & \text{if } x=y \\ 0 & \text{if } x \neq y \end{cases}$$

Suppose we have three stations x_1, x_2, x_3 and we want to estimate Z at three points y_1, y_2, y_3 forming the pattern shown in Figure 1.

Figure 1: Example of simultaneous estimation of Z at three points from a network of three stations.

Then

$$F' = \begin{bmatrix} 1 & 0 & \frac{\sqrt{3}}{2} \\ 1 & \frac{1}{2} & 0 \\ 1 & -\frac{1}{2} & 0 \end{bmatrix}$$

and

$$[\phi_{y_1}, \phi_{y_2}, \phi_{y_3}] = - \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}.$$

With some matrix calculations one shows that

$$G^{-1} = (FK^{-1}F')^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{\sqrt{3}} \\ 0 & 2 & 0 \\ -\frac{1}{\sqrt{3}} & 0 & 2 \end{bmatrix}$$

and

$$[\phi_{y_i}' G^{-1} \phi_{y_j}] = CC' = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}.$$

Then max eigenvalue of $CC' = 4$ and $\text{tr } CC' = 9$.

Hence in this example relation (1) holds. More precisely,

$$\text{max eigenvalue of } CC' = \frac{1 - \frac{1}{9}}{2} \text{tr } CC';$$

therefore, from the proof of Theorem 6.1.1, we see that we can take $\eta = \frac{1}{9}$ in the definition of the estimator $\tilde{Z}(y_i)$ given by formula 6.2.1.

If we let

$$\Lambda = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1' \\ \lambda_2' \\ \lambda_3' \end{bmatrix}$$

after some more computations, observing that $\hat{a}'G\hat{a}$ turns out to be $\|Z\|^2$, one gets

$$\tilde{Z}(y_i) = \lambda_i' Z \left(1 - \frac{1/9}{\|Z\|^2} \right),$$

and $\hat{Z}(y_i) = \lambda_i' Z$. The vectors λ_i can actually be obtained from geometrical considerations.

Finally, the risks of the two types of estimators compare as follows:

$$\begin{aligned} \sum_{i=1}^3 E(\hat{Z}(y_i) - Z(y_i))^2 &= \sum_{i=1}^3 (\sigma_{y_i}^2 - k_{y_i}' K^{-1} k_{y_i}) + \text{tr } CC' \\ &= 3 + 9 \\ &= 12 \end{aligned}$$

while

$$\begin{aligned} \sum_{i=1}^3 E(\tilde{Z}(y_i) - Z(y_i))^2 &< \sum_{i=1}^3 (\sigma_{y_i}^2 - k_{y_i}' K^{-1} k_{y_i}) + \text{tr } CC' \left(1 - \eta^2 E \frac{1}{p+2K} \right) \\ &= 3 + 9 \left(1 - \frac{1}{81} E \frac{1}{3+2K} \right) \end{aligned}$$

where K is Poisson $\left(\frac{a'Ga}{2}\right)$.

In this example the reduction in risk produced by the Stein-like estimator is not tremendous, but the point is that there is a reduction. Note that the absolute lower bound given by formula 6.2.6 here has the value 3.

CHAPTER 7

APPENDIX

A thorough discussion of covariance functions, stationary covariance functions, and stationary isotropic covariance functions of random fields can be found in Matern (1960), Chapter 2. There can also be found a thorough list of references up to the date of the book. In addition to that, a short list of the works related to the subject, from 1960 to 1978, would contain Berman (1978), Cramer and Leadbetter (1966), Dalenius, Hajek, and Zubrzycki (1961), Matheron (1965), Mittal (1976), and Yaglom (1961), (1962).

In this chapter we give two side results on covariance functions: one related to the criterion $e'K^{-1}e$ studied in Section 5.4, the other one related to finding an isotropic covariance function supporting the estimation paradox subject of Section 5.5.

7.1. Singularity of the Matrix $[\|x_i - x_j\|^2]$

In Section 5.4 we showed that when the covariance function is represented locally by

$$K(x,y) = \sigma^2(1 - \alpha\|x - y\|)$$

then the maximization of $e'K^{-1}e$ is equivalent to the minimization of $e'N^{-1}e$ where

$$N = [n_{ij}]_{n \times n}, \quad n_{ij} = \|x_i - x_j\|.$$

Here we show that the representation

$$K(x,y) = \sigma^2(1 - \alpha \|x - y\|^2)$$

used for the same purpose is inadequate because, in general, the corresponding matrix K then is singular.

Theorem 7.1.1. Given n points $x_1, x_2, \dots, x_n$ in R^2 , let $M(x_1, x_2, \dots, x_n)$ be the $n \times n$ matrix whose (i,j) element is $\|x_i - x_j\|^2$, and let $K(x_1, x_2, \dots, x_n) = \sigma^2(ee' - \alpha M(x_1, x_2, \dots, x_n))$. Then

$$n \geq 5 \text{ implies } \det M(x_1, x_2, \dots, x_n) = 0$$

and

$$n > 5 \text{ implies } \det K(x_1, x_2, \dots, x_n) = 0.$$

This result generalizes to $x_1, x_2, \dots, x_n \in R^p$ and $n \geq p + 3$.

The proof of Theorem 7.1.1 uses two elementary lemmas.

Lemma 7.1.1. Given four points x_1, x_2, x_3, x_4 in R^2 , $\det M(x_1, x_2, x_3, x_4) = 0$ if and only if x_1, x_2, x_3, x_4 are on a circle (possibly a line).

Proof. Consider $\det M(x_1, x_2, x_3, x_4) = 0$ as an equation in the coordinates of x_4 . It is the equation of a circle which goes through x_1, x_2 , and x_3 (possibly a line). Conversely, if x_1, x_2, x_3 , and x_4 are on a circle, then x_4 satisfies

$$\det M(x_1, x_2, x_3, x_4) = 0.$$

Lemma 7.1.2. Let x_1, x_2, x_3, x_4 be four points in R^2 , not on a circle, let $M = M(x_1, x_2, x_3, x_4)$ and for any x define

$$m(x) = \begin{bmatrix} \|x - x_1\|^2 \\ \|x - x_2\|^2 \\ \|x - x_3\|^2 \\ \|x - x_4\|^2 \end{bmatrix}$$

then for any x and $y \in \mathbb{R}^2$

$$m(x)' M^{-1} m(y) = \|x - y\|^2 .$$

Proof. Let us fix y , and define

$$\phi(x) = m(x)' M^{-1} m(y) - \|x - y\|^2 .$$

It is a function of the coordinates (u, v) of x of the form

$$\alpha(u^2 + v^2) + \beta u + \gamma v + \delta .$$

Moreover, $\phi(x_1) = \phi(x_2) = \phi(x_3) = \phi(x_4) = 0$. Then

$\alpha = \beta = \gamma = \delta = 0$ unless x_1, x_2, x_3, x_4 are on a circle (or a straight line); but this was eliminated by hypothesis. Therefore,

$$\phi(x) \equiv 0 .$$

Proof of Theorem 7.1.1. We now consider $x_1, x_2, \dots, x_n$, $n \geq 5$. We can suppose that there are four points among them which are not on a circle because, if not, then $x_1, x_2, \dots, x_n$ are all on the same circle and $\det M = 0$ as an easy consequence of Lemma 7.1.1. Hence, suppose x_1, x_2, x_3, x_4 are not on a circle, and decompose M as follows

$$M = \left[\begin{array}{c|c} M_1 & m \\ \hline m' & \Omega \end{array} \right]$$

where M_1 is 4×4 , m is $4 \times (n-4)$ and Ω is $(n-4) \times (n-4)$.

Then

$$\det M = \det M_1 \det(\Omega - m' M_1^{-1} m) .$$

But, by Lemma 7.1.2, one sees that

$$\Omega = m' M_1^{-1} m ,$$

hence $\det M(x_1, x_2, \dots, x_n) = 0$.

Finally, to prove $\det K = 0$ when $n \geq 6$, we use the following decomposition of the determinant of a sum: if

$A = [A_1, A_2, \dots, A_n]$ and $B = [B_1, B_2, \dots, B_n]$ where the A_i 's and B_i 's are columns vectors, then

$$\begin{aligned} \det(A+B) &= \det A + \det(A_1, A_2, \dots, A_{n-1}, B_n) + \\ &\quad \det(A_1, A_2, \dots, B_{n-1}, A_n) + \dots + \det(A_1, B_2, \dots, B_n) + \det B . \end{aligned}$$

Here $A = \sigma^2 ee'$ and $B = -\sigma^2 \alpha M$, hence all the determinants with two or more columns of A are zero.

On the other hand, those having one or zero column of A have five or more columns of B , and by an adaptation of the first part of the theorem they must be zero too.

To generalize the result to R^p , one can go through the same sequence of lemmas generalized to R^p . The key fact is that a sphere in R^p is determined uniquely by $p+1$ distinct points chosen on it (not on a sphere of dimension $p-1$). Q.E.D.

Note: Lemma 2 has the following curious corollary.

Corollary. $\forall x_1, x_2, x_3, x_4 \in \mathbb{R}^2$ not on a circle or a line
and $\forall x \in \mathbb{R}^2$ we have

$$e' M^{-1} m(x) = 1 .$$

Proof. Write

$$\frac{m(y)'}{\|x-y\|^2} M^{-1} m(x) = 1$$

and let $\|y\| \rightarrow +\infty$, we get

$$e' M^{-1} m(x) = 1 .$$

This also generalizes to $\mathbb{R}^p$.

7.2. A Class of Covariance Functions in $\mathbb{R}^2$ Vanishing at a Finite Distance

In Dalenius, Hajek, and Zubrzycki (1961) it is mentioned that we can construct a stationary isotropic random field in $\mathbb{R}^2$ whose covariance function is

$$(1) \quad K_0(h) = \begin{cases} \frac{2}{\pi} \left[\arccos\left(\frac{h}{a}\right) - \frac{h}{a} \sqrt{1 - \left(\frac{h}{a}\right)^2} \right] & 0 \leq h \leq a , \\ 0 & \text{otherwise} \end{cases}$$

where a is a positive constant. The method is very geometric; the quantity $K_0(h)$ is proportional to the overlapping area of two disks of radius $\frac{a}{2}$ whose centers are at a distance h from each other.

Here we show that the function K_0 is only the simplest case of a whole class of stationary isotropic covariance functions vanishing at a finite distance. Let W be the two-dimensional Gaussian

white noise that is, following Mittal (1976), for any Borel sets B and C in $\mathbb{R}^2$, $W(B)$ and $W(C)$ are normal random variables with mean zero. Variance of $W(B) = \text{Volume of } B$, and $E\{W(B) W(C)\} = \text{volume of } B \cap C$.

Consider a function $f : [0,1] \rightarrow \mathbb{R}$ and define the random field $\{Z(x), x \in \mathbb{R}^2\}$ as follows:

$$(2) \quad Z(x) = c \int_{\{y: \|y-x\| \leq 1\}} f(\|y-x\|) W(dy) .$$

Then Z is a stationary isotropic Gaussian field whose covariance function is, for $h \leq 2$,

$$(3) \quad K(h) = 4c^2 \int_{r=\frac{h}{2}}^1 \int_{\theta=0}^{\arccos(h/2r)} f(r)f(s)r dr d\theta ,$$

with r , s , and θ represented on Figure 1 below:

Figure 1. Computation of $K(h)$.

In Figure 1 $\|x-y\| = h$ and $s^2 = h^2 + r^2 - 2hr \cos \theta$. Let us pick c such that $K(0) = 1$: This is satisfied when

$$1 = 4c^2 \int_0^1 \int_0^{\pi/2} f^2(r) r dr d\theta ,$$

then

$$4c^2 = \frac{2}{\pi \int_0^1 r f^2(r) dr} .$$

The function $K_0(h)$, with $a = 2$, corresponds to the case $f(r) = 1$, $0 \leq r \leq 1$. Let us show it by direct calculations:

$$(4) \quad \int_{r=\frac{h}{2}}^1 \int_{\theta=0}^{\arccos \frac{h}{2r}} r dr d\theta = \int_{\frac{h}{2}}^1 \arccos\left(\frac{h}{2r}\right) r dr .$$

Let $u = \arccos \frac{h}{2r}$, then $r dr = \frac{2r^3}{h} \sin u du$, $r^3 = \frac{h^3}{8 \cos^3 u}$, and the R.H.S. of (4) becomes

$$(5) \quad \int_0^{\arccos \frac{h}{2}} u \frac{h^2}{4} \cdot \frac{\sin u}{\cos^3 u} du \\ \propto \int_0^{\arccos \frac{h}{2}} u \tan u d(\tan u) ,$$

the sign $\propto$ means proportional to. Let $v = \tan u$; then the integral (5) is

$$(6) \quad \int_0^{\tan(\arccos \frac{h}{2})} v \arctan v dv .$$

Let $s = \tan(\arccos \frac{h}{2})$. By integration by parts one shows that
(6) is proportional to

$$(7) \quad [v^2 \tan v]_0^s - \int_0^s \frac{v^2}{1+v^2} dv = s^2 \arctan s - \int_0^s \left(1 - \frac{1}{1+v^2}\right) dv.$$

After some final reduction (7) is shown to be proportional to

$$\arccos \frac{h}{2} - \frac{h}{2} \sqrt{1 - \frac{h^2}{4}} \quad . \quad \text{Q.E.D.}$$

The incomplete result stated below gives the derivative of
 $K(h)$ for any function f sufficiently regular:

Theorem 7.2.1.

$$(8) \quad \frac{d}{dh} K(h) = k \int_{r=\frac{h}{2}}^1 \left[- \frac{f^2(r)}{2\sqrt{1 - \frac{h^2}{4r^2}}} + rf(r) \int_0^{\arccos \frac{h}{2r}} \frac{1}{f'(s)} \frac{h - r \cos \theta}{s} d\theta \right] dr$$

where $k = \frac{2}{\pi \int_0^1 r f^2(r) dr}$ and $s = \sqrt{r^2 + h^2 - 2rh \cos \theta}$.

Proof. Formula (3) can be rewritten

$$K(h) = k \int_{r=\frac{h}{2}}^1 \left[rf(r) \int_{\theta=0}^{\arccos \frac{h}{2r}} f(s) d\theta \right] dr.$$

Let us denote $g(r)$ the term within square brackets. Then

$$(9) \quad \frac{d}{dh} K(h) = - \frac{k}{2} g\left(\frac{h}{2}\right) + k \int_{\frac{h}{2}}^1 \frac{d}{dh} g(r) dr,$$

$$(10) \quad g(r) = r f(r) \int_0^{\arccos \frac{h}{2r}} f\left(\sqrt{r^2 + h^2 - 2rh \cos \theta}\right) d\theta ,$$

$$(11) \quad \frac{d}{dh} g(r) = r f(r) \left[\frac{d}{dh} \left(\arccos \frac{h}{2r} \right) f\left(\sqrt{r^2 + h^2 - 2rh \cos \theta}\right) + \int_0^{\arccos \frac{h}{2r}} \frac{d}{dh} f\left(\sqrt{r^2 + h^2 - 2rh \cos \theta}\right) d\theta \right] .$$

Note that $d \arccos x = -\frac{dx}{\sqrt{1-x^2}}$, and

$$\frac{d}{dh} f(s) = f'(s) \frac{ds}{dh} = f'(s) \frac{h - r \cos \theta}{s} .$$

On the other hand,

$$(12) \quad g\left(\frac{h}{2}\right) = \frac{h}{2} f\left(\frac{h}{2}\right) \int_{\theta=0}^{\arccos 1} f(s) d\theta = 0 .$$

Then modifying (11) and substituting (11) and (12) in (9) yields formula (8) as stated. Q.E.D.

Applications:

1. Case $h = 0$. Then $s = r$, therefore

$$\begin{aligned} K'(0) &= k \int_{r=0}^1 \left[-\frac{1}{2} f^2(r) + r f(r) \int_0^{\pi/2} f'(r) (-\cos \theta) d\theta \right] dr \\ &= k \int_0^1 \left[-\frac{1}{2} f^2(r) - r f(r) f'(r) \right] dr , \end{aligned}$$

hence
$$K'(0) = -\frac{\int_0^1 [f^2(r) + 2r f'(r) f(r)] dr}{\pi \int_0^1 r f^2(r) dr} ,$$

but $[f^2(r) + 2r f'(r) f(r)]dr = d(r f^2(r))$,

therefore
$$K'(0) = - \frac{\int_0^1 d(r f^2(r))}{\pi \int_0^1 r f^2(r) dr}$$

In particular, for $f = 1_{[0,1]}$ this yields $K'_0(0) = -\frac{2}{\pi}$,
which can be checked directly.

2. Case $h = 2$.

Theorem 7.2.2. If the left derivative of f at the point 1 is bounded, then

$$\lim_{h \rightarrow 2} \frac{d}{dh} K(h) = 0 \quad .$$

Proof. When $h \rightarrow 2$, the first term in (8) is

$$\int_{r=\frac{h}{2}}^1 - \frac{f^2(r)}{2\sqrt{1-\frac{h^2}{4r^2}}} dr \sim -f^2(1) \int_{\frac{h}{2}}^1 \frac{dr}{2\sqrt{1-\frac{h^2}{4r^2}}} .$$

Next
$$\begin{aligned} \int_{\frac{h}{2}}^1 \frac{dr}{\sqrt{1-\frac{h^2}{4r^2}}} &= \int_{\frac{h}{2}}^1 \frac{2r dr}{\sqrt{4r^2 - h^2}} = \frac{h}{4} \int_{\frac{h}{2}}^1 \frac{d \frac{4r^2}{h^2}}{\sqrt{\frac{4r^2}{h^2} - 1}} = \frac{h}{2} \int_{u=1}^{\frac{4}{h^2}} \frac{du}{2\sqrt{u-1}} \\ &= \frac{h}{2} [\sqrt{u-1}]_1^{\frac{4}{h^2}} = \frac{h}{2} \sqrt{\frac{4}{h^2} - 1} = \sqrt{1 - \frac{h^2}{4}} , \end{aligned}$$

and this tends to zero as $h \rightarrow 2$.

Hence, we only have to show that the second term in (8) satisfies

$$\lim_{h \rightarrow 2} \int_{r=\frac{h}{2}}^1 r f(r) \int_0^{\arccos \frac{h}{2r}} f'(s) \frac{h - r \cos \theta}{s} d\theta dr = 0 .$$

The term on the right of the limit is equivalent to

$$f(1) \int_{r=\frac{h}{2}}^1 \arccos \frac{h}{2r} f'(1-) dr .$$

We assumed $f'(1-)$ bounded. Then, since

$$\lim_{h \rightarrow 2} \int_{r=\frac{h}{2}}^1 \arccos \frac{h}{2r} dr = 0 ,$$

the result follows.

REFERENCES

- Anderson, T.W. (1958): An Introduction to Multivariate Statistical Analysis. John Wiley.
- Berman, S.M. (1978): A class of isotropic distributions in $\mathbb{R}^n$ and their characteristic functions. Pacific Journal of Mathematics, 78, No. 1.
- Bras, R.L. and Rodriguez-Iturbe, I. (1976a): Rainfall generation: A nonstationary time-varying multidimensional model. Water Resources Research, 12, No. 3.
- Bras, R.L. and Rodriguez-Iturbe, I. (1976b): Network design for the estimation of areal mean of rainfall events. Water Resources Research, 12, No. 6.
- Bras, R.L. and Colon, R. (1978): Time-averaged areal mean of precipitation: Estimation and network design. Water Resources Research, 14, No. 5.
- Chung, K.L. (1974): A Course in Probability Theory. Second edition. Academic Press.
- Cohen, A. (1965): Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution. Annals of Mathematical Statistics, 36, 78-87.
- Cramer, H. and Leadbetter, M.R. (1966): Stationary and Related Stochastic Processes. John Wiley: New York.
- Dalenius, T., Hajek, J., and Zubrzycki, S. (1961): On plane sampling and related geometrical problems. Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, 125-150.
- Delfiner, P. (1975): Linear estimation of nonstationary spatial phenomena. In Advanced Geostatistics in the Mining Industry. M. Guarascio et al. (Eds.), 49-68. D. Reidel: Dordrecht.
- Delfiner, P. and Delhomme, J.P. (1973): Optimum interpolation by kriging. In Proceedings of NATO A.S.I. for Display and Analysis of Spatial Data. 96-114. John Wiley: London.
- Delhomme, J.P. (1978): Kriging in the hydrosiences. Advances in Water Resources, 1, No. 5.
- Efron, B. and Morris, C. (1975): Data analysis using Stein's estimator and its generalizations. J.A.S.A., 70, 311-319.

- Faith, R. (1976): The interpolation of air pollution: A stochastic approach. Stanford University, SIMS, Working Paper, No. 5.
- Goldgerger, A.S. (1962): Best linear unbiased prediction in the generalized linear regression model. J.A.S.A., 57, 369-375.
- Gribik, P.R., Kortanek, K.O., and Sweigart, J.R. (1976): Designing a regional air pollution monitoring network: An appraisal of a regression experimental design approach. Environmental Modeling and Simulation, 86-91.
- Hodges, J.L. Jr. and Lehmann, E.L. (1951): Some applications of the Cramer-Rao inequality. Proc. Second Berkeley Symposium on Mathematical Statistics and Probability, 13-22.
- James, W. and Stein, C. (1960): Estimation with quadratic loss. Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, 361-379.
- Jones, D.A., Gurney, R.J., and O'Connell, P.E. (1978a): Network design using optimal estimation procedures. Institute of Hydrology, Wallingford, Oxon, United Kingdom. Paper presented at the Chapman Conferences "Design of Hydrologic Data Network," Tucson, Arizona, U.S.A., December 11-14, 1978.
- Jones, D.A., Gurney, R.J., O'Connell, P.E., et al. (1978b): A case study of rationalization of a raingauge network in South-West England. Institute of Hydrology, Wallingford, Oxon, United Kingdom. Paper presented at the Chapman Conferences "Design of Hydrologic Data Network," Tucson, Arizona, U.S.A., December 11-14, 1978.
- Lévy, P. (1948): Processus Stochastiques et Mouvement Brownien. Gauthier-Villars: Paris.
- Matern, B. (1960): Spatial variation (stochastic models and their application to some problems in forest surveys and other sampling investigations). Meddelanden Fran Statens Skogsforskningsinstitut Ban 49. No. 5 (reviewed 1970).
- Mathéron, G. (1965): Les Variables Régionalisées et leur Estimation. Masson et Cie., Editeurs. Paris.
- Matheron, G. (1973): The intrinsic random functions and their applications. Adv. Appl. Prob., 5, 439-468.
- Mittal, Y. (1976): A class of isotropic covariance functions. Pacific Journal of Mathematics, 64, 517-538.
- Rodriguez-Iturbe, I. and Mejia, J.M. (1974): The design of rainfall networks in time and space. Water Resources Research, 10, No. 4, 713-728.

Shoenberg, I.J. (1938): Metric spaces and completely monotone functions. Annals of Mathematics, 39, No. 4.

Stark, T., Fang, J., and Chiu, C. (1978): Intrinsic random functions: A discussion which includes computer programs and an application to coal data. Geostatistics report No. 1. Coal Extraction and Utilization Research Center, Southern Illinois University, Carbondale, Illinois.

Stein, C. (1960): Multiple Regression. In Contributions to Probability and Statistics (Essays in honor of Harold Hotelling), I. Olkin, ed. Stanford University Press.

Stein, C. (1966): An approach to the recovery of inter-block information in balanced incomplete designs. In Research Papers in Statistics (Festschrift for J. Neyman), F. N. David, ed. John Wiley.

Switzer, P. (1976): Probability models for spatial variability. Unpublished lecture notes.

Tubilla, A. (1975): Error Convergence Rates for Estimates of Multidimensional Integrals of Random Functions. Ph.D. dissertation. Statistics Department, Stanford University.

Veneziano, D., Kitanidis, P.K., and Queiroz, C.S. (1977): Two problems of spatial sampling. Technical Report. Department of Civil Engineering, M.I.T.

Yaglom, A.M. (1961): Second-order homogeneous random fields. Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, 2, 593-622.

Yaglom, A.M. (1962): An Introduction to the Theory of Stationary Random Functions. Prentice-Hall, Inc.