On Experimental Designs for Derivative Random Fields

Von der Fakultät für Mathematik und Informatik
der Technischen Universität Bergakademie Freiberg
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
Dr. rer. nat.,

vorgelegt
von Dipl.-Ing. Jaroslav Šimák

geboren am 2. Oktober 1974 in Příbram, Tschechien

Gutachter: Prof. Dr. rer. nat. Wolfgang Näther, Freiberg
Prof. Dr.-Ing. Joachim Menz, Freiberg
Prof. RNDr. Viktor Beneš, Prag

Tag der Verleihung: 5. Juli 2002

Acknowledgement

At this place, I would like to take the opportunity to express my sincere thanks to Prof. Dr. Wolfgang Näther. Under his supervisory, I have enjoyed cooperation in a friendly and informal atmosphere. I appreciate a lot his help which has sometimes gone far beyond the scope of our work. I am grateful to Prof. Dr. Joachim Menz and Dr. Markus Holtmann for some useful comments. Finally, I am indebted to Bärbel Winterholler who helped me very much to improve the readability of this manuscript.

Contents

1. Introduction

The design of an experiment aims at an efficient choice of the experimental conditions in order to optimize the performance of the statistical inference which will be based on the observations to be obtained. Schwabe (1996, page 1)

In practice, it is seldom possible to observe every potential datum of some study variable associated with a phenomenon that varies over a continuous (or even a large finite) spatial domain. Typically, an inference about the study variable may be made from a very small subset of potential data. Thus, an important task of statistics is an optimization of a particular data-collecting procedure. This procedure is called experimental design or sampling plan (network).

In the spatial context, the interest is typically in the prediction of (some function of) the study variable at multiple unsampled sites. An optimal experimental design usually refers to the choice of locations at which to sample in order to achieve optimality according to a given criterion (e.g., minimize average mean square prediction error, where the average is taken over multiple prediction locations). In practice, optimal experimental designs may be extremely difficult to achieve. On the other hand, good, although suboptimal, designs may be relatively easy to obtain and these designs, at least, should be sought.

Two directions in the theoretical statistics related to these problems can be found in literature, namely, the so-called sampling theory (see e.g. Aldworth and Cressie (1999)) and the theory of optimal experimental design (Pukelsheim, 1993). They have been developed separately, though with a considerable theoretical overlap. For problems which are motivated by geostatistics (Chilès and Delfiner, 1999) this overlap is more apparent.

Given an observation site, it is rather usual that there are several information sources available. For instance, in environmental monitoring, it is possible to obtain information about the amount of lead, zinc, sulphur or other pollutants from one soil sample. In this sense, it is a multivariate problem. In such a situation, the question formulation is not only where to observe, but also which component(s) should be observed, since not all observed components contribute equally to the predictor. Another issue is that any observation (and the following analysis) is associated with costs (financial, material or environmental). This thesis is an attempt to provide a solution to a particular **multivariate design problem** which is motivated by the following application areas.

In science and engineering, there are situations where gradient information on a stochastic quantity is directly measurable. Sometimes, direct measurements of a study variable are associated with very big costs or it is even impossible, but derivatives can be obtained relatively easily. One example is the geophysical determination of gravity anomalies using the Eötvös torsion balance (Jordan et al., 1967). These measurements are important for exploration of deposits. A detailed listing of applications in earth science, where derivatives are available is given in Jäkel (1997). As a more general and modern notion the

so-called satellite gradiometry should be mentioned (Freeden, 1999). Differentiable models and gradients are of certain importance in meteorology (Daley, 1996).

Finally, another differently motivated area are so-called computer experiments (see Koehler and Owen (1996) and Sacks et al. (1989) for overview). Computer modeling of physical or behavioral systems has become an important tool in many fields of research and also in manufacturing. Many of considered systems are so complex, that physical experimentation is too time consuming or too expensive. Consequently, experimenters have turned to simulations of these systems. Typically, simulators help to describe the relationship between the model, the production process and the real system (or final product). Equally important are the effects of uncontrollable variation in the production process parameters to the end products. Here, computer experiments can be seen as a kind of sensitivity analysis.

Many simulators need to solve differential equations and can provide the gradient of the response at a design point with little or no additional cost. For a listing of applications, which includes problems from semiconductor industry, nuclear research and physics, see Koehler and Owen (1996); Sacks et al. (1989) and Currin et al. (1991).

There are only a few references considering optimal design for derivatives. Some of them are motivated by a series of papers (Sacks and Ylvisaker, 1966, 1968, 1970) and they investigate large samples in an asymptotic optimal sense (Benhenni and Cambanis, 1992). In this context, the inclusion of derivative information means that a stochastic process is observed at a certain point simultaneously with all mean square derivatives of the process up to a certain order. This is the case for example of papers related to computer experiments (see Koehler and Owen (1996); Mitchell et al. (1994) and Morris et al. (1993) inter alia), where this limitation arises from problem specifics. This restriction is not necessary in the earth science and it is left in Holtmann (2001) where links to spline approximation theory, regression and prediction problems are established. Some numerical results are presented in Albrecht (1998).

For a good understanding of design problems in general, it is necessary to conceive what their essential properties are. They are a **space of design parameters**, which should be set optimally according to an **objective function**, and a **model** of described reality.

The concept adopted in this work arises from stochastics. The observations are supposed to be realizations of a **mean square differentiable random field** or realizations of its derivative fields. A focus is put to second order characteristics of the field, which is described by a known covariance function. The first order characteristic, i.e. the mean, is supposed to be known or equivalently equal to zero. This setup is sufficient for using the so-called **simple kriging predictor**. Throughout this thesis, the prediction is regarded with respect to the underlying (non-derivative) random field. The space of design parameters to be set can typically be observation coordinates in Euclidian space.

The objective function exclusively considered here is the so-called **imse-update**. This criterion is derived from the well known integrated mean square error (imse), which has been widely used (see for example Sacks et al. (1989) and Barnes (1989)) because it is intuitively appealing. Basically, one seeks a design that does 'best' on the average, where 'best' is measured by the mean square error of a linear predictor. Let us also note that the imse criterion often appears in a rather different context (cf. e.g. Khuri and Cornell (1996) and Biedermann and Dette (2001)).

In geostatistics, the covariance function is never a priori known. Typically, after some data are collected, the so-called model fitting is done. This means that some parameter-

ized covariance function (family) is chosen and its parameters are estimated from data. The present work is an attempt to explain **how the choice of a covariance function influences the quality of information obtained from derivatives of diverse order.**

The problem can be also formulated in another way. Can a careless choice of a covariance function lead (in the context of derivative observations) to an erroneous or over-optimistic inference? This question formulation is, in particular, inspired by the work of Stein (1999). However, the way it is solved here, is rather far from that of Stein (1999). Finally, the question is, what are simple rules of thumb which lead to an efficient design.

This thesis is organized as follows. **Chapter 2** provides a basic introduction to the terminology. For good understanding, the notion of mean square differentiability of a random field is essential (Section 2.1.4). Lemma 2.2, formulated in this section, is of a certain importance too, since it is used later on many times. Section 2.2 gives an orientation in a hierarchy of prediction problems. One of them is the topic of this work. Section 2.3 provides very briefly some links to other design problems.

Chapter 3 summarizes some properties of covariance functions. A focus is put to mean square differentiability and spectral properties of covariance functions. The chapter follows with a detailed overview of some examples of (families of) covariance functions. Attention is paid to the Matérn class, which has recently been recommended (Stein, 1999) for its flexibility and other reasonable properties. Some formulas, which are used later, are stated here. In particular, the formulas for the variances of derivative random fields under the Matérn, the Gaussian and the J-Bessel model are relevant to the further considerations. Lemmata 3.1 and 3.3 provide simplifications of formulas for differentiation of the Matérn and the J-Bessel class. Proofs are more or less technical and they are postponed to Appendix A.

Chapters 4 and **5** contain the main results of this thesis. Unlike to Chapter 4, where the reasoning is carried out theoretically, in Chapter 5, an empirical investigation is mainly made, however, with a lot of important links to the previous results

In **Chapter 4**, a certain class of designs is studied, namely, **designs with vanishing inter-site correlations**. The aim of the first part of this chapter is purposely not to construct an optimal design (it can be seen that these designs are in most cases not good), but to obtain some notion about the information, which a particular derivative observation can serve in some general sense. As already said, as a measure of this information, the imse-update criterion is used and this criterion should be maximized. Further, covariance functions and their L^2-properties are a very essence of the discussed approach.

The imse-update depends both on a particular design and on a covariance function of the study random field. This dependence is typically very complex. Due to an asymptotic, it disappears dependence on the design, i.e. all limiting desing (given number and derivative grades) are mutually equivalent.

Thus, can the (limiting) imse-update be bigger for derivative observations of higher order than for lower ones? **Theorem 5** and connected claims are a survey of this problem. Basically, both positive and negative answers can turn out and the solution is closely connected to a particular covariance function of the random field under study. Although the case, where the J-Bessel class is considered, is likely not of practical importance, the opposite is hopefully true for the Matérn class. For the later class and for the Gaussian covariance function, it is shown that the (limiting) imse-update decreases rapidly to zero as the order of derivative increases. These results can be used for establishing the efficiency of designs, which consist of diverse derivative observations or as **a comparison between**

several designs. In this context, the J-Bessel class can be seen as a counterpart of the Matérn class. It is an example for the fact that covariance functions allow a large variety of possibilities and their proper choice is of great importance in modeling.

The main focus of the second part of Chapter 4 is an investigation of the problem whether the (limiting) imse-update can be improved by a suitable combination of observations of distinct derivative random fields in one site. It is shown that **correlated pairs of derivatives** outperform the other discussed possibilities.

In **Chapter 5**, the asymptotic setup is left and designs are examined empirically. At first, an alternative analytical formula for imse-update is given, which is further substantiated for the case of two observations. Attention is paid to the Gaussian covariance function and some numerical case studies are investigated.

On the basis of some evidence, well constructed **regular grids** are proposed to be an efficient solution of the design problem for derivatives (at least for Gaussian-like or Matérn-like covariance functions). Finally, an analogous comparison of efficiency as in Theorem 5 is given in this chapter but for optimally constructed regular grids. This chapter is supplemented with many plots and tables which support the argumentation.

The **Appendix** consists of three parts. In the first part, the postponed **proofs** can be found. In the second part, the problem of **conditional simulation** using derivatives is solved. This section is not directly related to design problems and therefore it was postponed to Appendix B. Note that, although simulations are of a great interest in geostatistics, simultaneous (conditional) simulation of multiple derivative fields was not explicitely formulated in the literature. The third part of the appendix supplements Chapter 5 with **figures** related to the Matérn and the J-Bessel class.

The presented work offers a rather specialized view to an optimal design problem and it attempts to contribute to the current discussion of the geostatistical modeling. It remains, of course, a lot of open problems and many questions with unsatisfactory answers. Chapter 5 points out some effects, which can be verified on an empirical basis but a rigorous mathematical proof is missing. A further investigations in this area is surely a great challenge for the future research.

2. Basic definitions

2.1. Random processes and fields

Suppose an observer has collected certain data, for example, pollution levels in an area. The problem is now, how to model the observed reality. There are certainly many approaches to the problem. The approach adopted here assumes the observations to be a realization of a parent random function.

In this chapter, random processes and random fields will be introduced and a simple classification of random fields will be summarized. The goal of this chapter is to define mean square properties of random fields, above all mean square differentiability. The chapter continues with a hierarchy of prediction problems. The last section provides a short overview of other design problems.

Let us recall some definitions from the theory of stochastic functions. For more details see Yaglom (1987a); Skorokhod (1996); Todorovic (1992) and Gikhman and Skorokhod (1996).

2.1.1. Definitions

A probability space is a triple $(\Omega, \mathcal{A}, \mathrm{P})$, where Ω is a certain nonempty set called *set of elementary events*, $\mathcal{A}$ is a collection of subsets of Ω forming σ-algebra and P is a probability measure on $\mathcal{A}$. Sets from $\mathcal{A}$ are referred to as random events and $\mathrm{P}(A)$ is the probability of $A \in \mathcal{A}$.

Each measurable function $\zeta : \Omega \to \mathbb{R}$ is said to be *a random variable*. This means that $\{\omega : \zeta(\omega) < t\} \in \mathcal{A}$ for all $t \in \mathbb{R}$. A measurable vector acting from Ω and having values in $\mathbb{R}^m$ is said to be *a random vector*.

A random function is a mapping

$$Z : \Omega \times D \to \mathbb{R}, \qquad D \subset \mathbb{R}^d,$$

such that for every $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ from the domain set D the function $Z(\cdot, \boldsymbol{x})$ is a random variable. On the other hand, for a given $\omega \in \Omega$ we refer to $Z(\omega, \cdot)$ as *a realization of the random function* Z. It is customary to denote a random function by $Z(\boldsymbol{x})$ and a realization by lowercase $z(\boldsymbol{x})$.

It is common to distinguish the following cases. If $d = 1$ and $d > 1$, we speak about *a random process* or about *a random field* respectively. The one-dimensional case corresponds typically to temporal phenomena. The cases of higher dimensions are frequent in geostatistics or environmental science and they correspond to spatial (or spatial-temporal) observations.

Sometimes, it can be profitable to consider complex valued random functions, which are defined as

$$Z(\boldsymbol{x}) = Z_1(\boldsymbol{x}) + iZ_2(\boldsymbol{x}), \qquad \boldsymbol{x} \in D,$$

where Z_1 and Z_2 are real valued random functions.

The basic characteristics of random processes are *finite dimensional distribution functions*, i.e. probabilities

$$F_{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n}(t_1,t_2,\ldots,t_n) = \mathrm{P}(Z(\boldsymbol{x}_1) < t_1, Z(\boldsymbol{x}_2) < t_2, \ldots, Z(\boldsymbol{x}_n) < t_n), \tag{2.1}$$

which are defined for all $n \in \mathbb{N}$, and for all $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n \in D$.

If an n-dimensional distribution function $F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(t_1,\ldots,t_n)$ can be written as

$$F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(t_1,\ldots,t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} p_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(t'_1,\ldots,t'_n)dt'_1,\ldots,dt'_n$$

for all $n \in \mathbb{N}$, $p_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(t'_1,\ldots,t'_n)$ is referred to as *an n-dimensional probability density.*

In natural sciences as well as in finance we usually have several information sources which can be observed simultaneously. For instance in meteorology an observer can obtain information about temperature, atmospheric pressure, wind velocity, concentration of ozone etc. All the observations could be realized at the same time or time interval at a certain place. Similarly, a broker possesses information about prices of different shares, contracts or their composite indices. Thus, it is useful to generalize the above concept and we can define *a multivariate random function*, that is a mapping

$$Z : \Omega \times D \to \mathbb{R}^m, \qquad D \subset \mathbb{R}^d,$$

such that for every $\boldsymbol{x} \in D$ the mapping $Z(\cdot, \boldsymbol{x})$ is a random vector.

2.1.2. Moments of random processes and fields

In practice, it is usually very complicated to get a complete description of a (multivariate) random function by means of probability distribution functions. A theoretical determination of a probability distribution function is possible only in rare cases as for instance for Gaussian fields. For this reasons, it is common to restrict oneself to the relatively simple characteristics of probability distribution functions. The most important numerical characteristics of distribution functions are their moments.

Mean function *Mean function* $m(\boldsymbol{x})$ (trend or drift) is a basic characteristic, which describes a systematic variation of the random function Z. It is simply the expected value of the random variable $Z(\boldsymbol{x})$

$$m(\boldsymbol{x}) := \mathrm{E}\, Z(\boldsymbol{x}) = \int_{\mathbb{R}} t \; dF_{\boldsymbol{x}}(t).$$

Covariance function *Covariance function* $C(\boldsymbol{x}, \boldsymbol{y})$ describes a stochastic variation or local deviations from the mean. It is defined as

$$\begin{aligned} C(\boldsymbol{x},\boldsymbol{y}) = \mathrm{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) &:= \mathrm{E}\{(Z(\boldsymbol{x}) - m(\boldsymbol{x}))(Z(\boldsymbol{y}) - m(\boldsymbol{y}))\} \\ &= \mathrm{E}\, Z(\boldsymbol{x})Z(\boldsymbol{y}) - m(\boldsymbol{x})m(\boldsymbol{y}). \end{aligned} \tag{2.2}$$

It follows immediately from (2.2) that covariance functions are symmetric, i.e.

$$C(\boldsymbol{x},\boldsymbol{y}) = C(\boldsymbol{y},\boldsymbol{x}). \tag{2.3}$$

A covariance function normalized by the standard deviations $C(\boldsymbol{x},\boldsymbol{x})^{1/2}$ and $C(\boldsymbol{y},\boldsymbol{y})^{1/2}$

$$c(\boldsymbol{x},\boldsymbol{y}) := \frac{C(\boldsymbol{x},\boldsymbol{y})}{C(\boldsymbol{x},\boldsymbol{x})^{1/2}C(\boldsymbol{y},\boldsymbol{y})^{1/2}} \tag{2.4}$$

is called *correlation function.* The definition (2.2) implies that the *variance function* $C(\boldsymbol{x},\boldsymbol{x}) = \mathrm{var}(Z(\boldsymbol{x}))$ is everywhere non-negative. In the following, it will be assumed that $C(\boldsymbol{x},\boldsymbol{x})$ is in fact strictly positive, in order to assure the correlation function (2.4) corresponding to a given covariance function is well defined.

Covariance and variance matrix, positive definiteness Mean function and covariance function are defined using appropriate random variables. Analogously, for random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we define *covariance matrix of* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$

$$\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y}^T) = \mathrm{E}\{(\boldsymbol{X}-\mathrm{E}\,\boldsymbol{X})(\boldsymbol{Y}-\mathrm{E}\,\boldsymbol{Y})^T\},$$

where the expected value of a random matrix (vector) is just a matrix (vector) of expected values. If $\boldsymbol{X}=\boldsymbol{Y}$ we refer $\mathrm{cov}(\boldsymbol{X},\boldsymbol{X}^T)$ to as *covariance matrix of* $\boldsymbol{X}$, $\mathrm{var}(\boldsymbol{X})$.

A real symmetric $(n\times n)$-matrix $\boldsymbol{A}$ is said to be *positive semi-definite* if $\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{\lambda}\geq 0$ for all real $(n\times 1)$-vectors $\boldsymbol{\lambda}$. This matrix is termed as *positive definite* if $\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{\lambda}>0$ for all real $(n\times 1)$-vectors $\boldsymbol{\lambda}\neq\boldsymbol{0}$. Note that if matrix $\boldsymbol{A}$ is positive definite, it follows immediately that $\boldsymbol{A}$ is invertible and $\boldsymbol{A}^{-1}$ is also positive definite. The family of all positive definite and positive semi-definite $(n\times n)$-matrices will be denoted by PD_n and PSD_n, respectively.

Note that, since

$$0\leq \mathrm{var}(\boldsymbol{\lambda}^T\boldsymbol{X}) = \boldsymbol{\lambda}^T\,\mathrm{var}(\boldsymbol{X})\boldsymbol{\lambda}$$

for any random vector $\boldsymbol{X}=(X_1,\ldots,X_n)^T$, such that $\mathrm{cov}(X_i,X_j)<\infty$, $i,j\in\{1,\ldots,n\}$ and for all $(n\times 1)$-real vectors $\boldsymbol{\lambda}$, covariance matrix $\mathrm{var}(\boldsymbol{X})$ is at least positive semi-definite.

General moments *General moment functions* are given for the sake of completeness by the following equation

$$\mu^{j_1 j_2\ldots j_n}(\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots\boldsymbol{x}_n) := \mathrm{E}\,Z(\boldsymbol{x}_1)^{j_1}Z(\boldsymbol{x}_2)^{j_2}\ldots Z(\boldsymbol{x}_n)^{j_n}.$$

Therefore, a mean function is the first moment of $Z(\cdot)$ and a covariance function is the centered second moment of $Z(\cdot)$. The exclusive topic of the thesis will be properties of random fields considering only the first two moments.

Gaussian random fields There are a couple of arguments that this restriction is reasonable. An important class of random fields are *Gaussian random fields*, which are characterized by the property that all finite dimensional probability densities are multivariate Gaussian, i.e., it holds for all $n\in\mathbb{N}$

$$p_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(\boldsymbol{t}) = \frac{1}{(2\pi)^{n/2}(\det\boldsymbol{\Sigma})^{1/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{t}-m(\boldsymbol{x}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{t}-m(\boldsymbol{x}))\Big),$$

where $\boldsymbol{t}$ and $\boldsymbol{x}$ denote $(t_1,\ldots,t_n)^T$ and $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^T$ for short. $\boldsymbol{\Sigma}$ is $n\times n$-matrix with elements $\boldsymbol{\Sigma}_{ij}=C(\boldsymbol{x}_i,\boldsymbol{x}_j)$, $i,j\in\{1,\ldots,n\}$. Therefore, a Gaussian random field is completely determined by its mean and its covariance function.

Random fields of the second order A random variable ζ is termed to be of the *second order* if $\mathrm{E}\,|\zeta|^2 < \infty$. A random field Z is said to be of the *second order* if $Z(\boldsymbol{x})$ is of the second order for all $\boldsymbol{x} \in D$.

All random processes and fields considered in the following text are of the second order.

2.1.3. Stationarity and isotropy

Even if we confine ourselves only to the simplest quantities like a mean and a covariance function, a statistical inference from data about the quantities can be very difficult. A common simplifying assumption is that the underlying probabilistic structure remains the same or similar in some sense in different parts of the domain set D.

Stationarity

A random field is said to be *stationary in the strong sense* if all its finite dimensional distribution functions are translation invariant, i.e.,

$$F_{\boldsymbol{x}_1+\boldsymbol{h},\boldsymbol{x}_2+\boldsymbol{h},\ldots,\boldsymbol{x}_n+\boldsymbol{h}}(t_1,t_2,\ldots,t_n) = F_{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n}(t_1,t_2,\ldots,t_n)$$

for all $\boldsymbol{h},\boldsymbol{x}_1,\ldots\boldsymbol{x}_n \in D$, such that $\boldsymbol{h}+\boldsymbol{x}_i \in D$ $(i=1,\ldots,n)$, for all $t_1,t_2\ldots,t_n \in \mathbb{R}$ and for all $n \in \mathbb{N}$. As a rule, stationary random fields are always considered on the whole $\mathbb{R}^d$.

A random field Z of the second order is called *stationary (in the wide sense)*, if the mean function $\mathrm{E}\,Z(\boldsymbol{x})$ is identically constant for all $\boldsymbol{x} \in D$ and if the covariance function C is translation invariant, i.e.,

$$C(\boldsymbol{x},\boldsymbol{y}) = C(\boldsymbol{x}+\boldsymbol{h},\boldsymbol{x}+\boldsymbol{h}), \quad \text{for all } \boldsymbol{x},\boldsymbol{y},\boldsymbol{h} \in D.$$

Translation invariant functions defined on $\mathbb{R}^d \times \mathbb{R}^d$ can be *represented* by a function $K(\cdot)$, which is defined on $\mathbb{R}^d$ as follows. Given the translation invariant function $C(\boldsymbol{x},\boldsymbol{y})$, define $K(\boldsymbol{x}) := C(\boldsymbol{0},\boldsymbol{x})$. Then

$$K(\boldsymbol{y}-\boldsymbol{x}) = C(\boldsymbol{0},\boldsymbol{y}-\boldsymbol{x}) = C(\boldsymbol{x},\boldsymbol{y}), \quad \text{for all } \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d.$$

If the function $C(\cdot,\cdot)$ is moreover covariance function, its representative $K(\cdot)$ is usually called covariance function, too. Observe that

$$K(\boldsymbol{0}) = \mathrm{var}(Z(\boldsymbol{x})), \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d,$$

whenever $K(\cdot)$ is a covariance function of a stationary random field Z. Further, in this case, $K(\cdot)$ is also an even function with respect to each of its arguments by virtue of the symmetry property (2.3).

Isotropy

A second order random field Z is called *isotropic (in the wide sense)*, if the mean function $\mathrm{E}\,Z(\boldsymbol{x})$ is identically constant for all $\boldsymbol{x} \in D$ and the covariance function are rotation invariant, i.e.,

$$C(\boldsymbol{x},\boldsymbol{y}) = C(\boldsymbol{H}\boldsymbol{x},\boldsymbol{H}\boldsymbol{y}), \quad \text{for all } \boldsymbol{x},\boldsymbol{y} \in D,$$

where $\boldsymbol{H}$ is an orthogonal $d \times d$-matrix, i.e. $\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I}$.

Isotropy means that there is no reason to distinguish between one direction from another for a random field under consideration.

Covariance function $C(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, of a stationary and isotropic random field depends on $\boldsymbol{x}$ and $\boldsymbol{y}$ only through the Euclidean norm $||\boldsymbol{y}-\boldsymbol{x}|| := \sqrt{\sum_{i=1}^{d}(y_i - x_i)^2}$. Therefore, a function $\varphi : [0, \infty) \to \mathbb{R}$ exists such that

$$C(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{y} - \boldsymbol{x}) = \varphi(||\boldsymbol{y} - \boldsymbol{x}||), \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$$

(see Gaspari and Cohn (1999)). However, it is usually convenient to define φ on the whole $\mathbb{R}$ as its even extension, i.e. $\varphi(-t) := \varphi(t), t \in [0, \infty)$.

The class of all functions $\varphi : [0, \infty) \to \mathbb{R}$, which are covariance functions of a stationary and isotropic random field on $\mathbb{R}^d$ will be denoted as Φ_d.

Anisotropy

If a random field is not isotropic, it is said to be *anisotropic*. In nature, isotropic random fields can be quite rare and a detection of a potential anisotropy from data is an important statistical task. Nevertheless, anisotropy is in many cases modeled as a linear transformation of coordinates. Precisely, Z exhibits a *geometric anisotropy*, if there exist an invertible matrix $\boldsymbol{V}$ such that $Z(\boldsymbol{V}\boldsymbol{x})$ is isotropic. For instance, in $\mathbb{R}^2$ the matrix $\boldsymbol{V}$ transforms any ellipse to a circle (cf. Chilès and Delfiner (1999) or Journel and Huijbregts (1978) for other types of anisotropy). In this text, only isotropic random fields are considered.

2.1.4. Mean square continuity and differentiability

In this section some important properties of random processes will be summarized. The section provides simple classification of processes based only on a mean and a covariance function. The major goal is to introduce mean square differentiable processes, which are of the main interest throughout the thesis. Here, we restrict ourselves only to the definition and basic properties, examples are widely discussed in Chapter 3.

Convergence in the mean square sense

A sequence $\{\zeta_n\}_{n=1}^{\infty}$ of random variables of the second order is said to *converge in the mean square sense* to a random variable ζ if

$$\lim_{n\to\infty} \mathrm{E}(\zeta_n - \zeta)^2 = 0. \tag{2.5}$$

We use the notation $\underset{\mathrm{n}\to\infty}{\text{l.i.m.}}\, \zeta_n = \zeta$.

A sequence of random fields $\{Z_n\}_{n=1}^{\infty}$ of second order is said to converge to a random field Z if

$$\underset{\mathrm{n}\to\infty}{\text{l.i.m.}}\, Z_n(\boldsymbol{x}) = Z(\boldsymbol{x}), \quad \text{for all } \boldsymbol{x} \in D.$$

It means that a convergence of a process in the mean square sense is defined as a convergence of corresponding random variables in the same sense. It holds, (cf. Yaglom (1987a, p.64)) that $\underset{\mathrm{n}\to\infty}{\text{l.i.m.}}\, Z_n(\boldsymbol{x}) = Z(\boldsymbol{x})$ if and only if

$$\lim_{n,m\to\infty} \mathrm{E}(Z_n(\boldsymbol{x}) - Z_m(\boldsymbol{x}))^2 = 0, \tag{2.6}$$

which is further equivalent to $\lim_{n,m\to\infty} \mathrm{E}\, Z_n(\boldsymbol{x}) Z_m(\boldsymbol{x}) < \infty$.

Mean square continuity

A random field $Z(\boldsymbol{x})$ is said to be *mean square continuous* if its mean $m(\boldsymbol{x})$ and its covariance function $C(\boldsymbol{x}, \boldsymbol{y})$ are continuous with respect to all components of $\boldsymbol{x}$ and $\boldsymbol{x}, \boldsymbol{y}$ respectively. Note that if a stationary random field is mean square continuous in the origin, it is mean square continuous everywhere. If a stationary random field is not mean square continuous in the origin, it is not mean square continuous in any point of $\mathbb{R}^d$.

Note that mean square continuity of the field Z does not mean that all its realizations $z(\cdot)$ are continuous. It can easily be found examples of mean square continuous random processes whose realizations have discontinuities. On the other hand, for processes, which are not mean square continuous, it is possible to construct their infinitely differentiable realizations (cf. Yaglom (1987a)).

Mean square differentiability

Definition 1 *A process $Z(\cdot)$ is said to be* mean square differentiable *at a point x if a random variable $Z^{(1)}(x)$ exists, such that $\mathrm{E}\,|Z^{(1)}(x)|^2 < \infty$ and*

$$\lim_{h \to 0} \mathrm{E} \left(\frac{Z(x+h) - Z(x)}{h} - Z^{(1)}(x) \right)^2 = 0. \tag{2.7}$$

By the virtue of (2.5), (2.7) can be rewritten as

$$\operatorname*{l.i.m.}_{\mathrm{h} \to 0} \frac{Z(x+h) - Z(x)}{h} = Z^{(1)}(x).$$

Lemma 2.1 *In order that a process Z be mean square differentiable on an interval $[a, b]$, it is necessary and sufficient that the derivatives*

$$\frac{d}{dx} m(x), \quad \frac{\partial}{\partial x} C(x, y), \quad \frac{\partial C(x, y)}{\partial x \partial y} = \frac{\partial C(x, y)}{\partial y \partial x}$$

exist for all $x, y \in [a, b]$. It is sufficient if the derivative of the mean function $\frac{d}{dx} m(x)$ and the limit

$$\lim_{h, \delta \to 0} \frac{C(x+h, x+\delta) - C(x, x+\delta) - C(x+h, x) + C(x, x)}{h\delta}$$

exist for all $x \in [a, b]$.

Proof : Let $Z^{(1)}(x)$ exist for all $x \in [a, b]$. Then

$$\mathrm{E}\, Z^{(1)}(x) = \lim_{h \to 0} \mathrm{E}\, \frac{Z(x+h) - Z(x)}{h} = \lim_{h \to 0} \frac{m(x+h) - m(x)}{h} = \frac{d}{dx} m(x). \tag{2.8}$$

In (2.8), the first equation is a consequence of the Schwarz inequality. Hence, the derivative $m^{(1)}(x) := \frac{d}{dx} m(x)$ exists. Therefore, the zero mean process

$$Z_0(x) = Z(x) - m(x)$$

is differentiable. In the same manner, it can be shown that

$$\begin{aligned} \operatorname{cov}(Z^{(1)}(x), Z(y)) &= \mathrm{E}\, Z_0^{(1)}(x) Z_0(y) = \frac{\partial}{\partial x} C(x, y) \\ \operatorname{cov}(Z(x), Z^{(1)}(y)) &= \mathrm{E}\, Z_0(x) Z_0^{(1)}(y) = \frac{\partial}{\partial y} C(x, y) \\ \operatorname{cov}(Z^{(1)}(x), Z^{(1)}(y)) &= \mathrm{E}\, Z_0^{(1)}(x) Z_0^{(1)}(y) = \frac{\partial C(x, y)}{\partial x \partial y} = \frac{\partial C(x, y)}{\partial y \partial x} \end{aligned} \tag{2.9}$$

On the other hand, to prove sufficiency, it suffices (by means of claim (2.6)) to show that the limit

$$\lim_{\delta,h\to 0} \mathrm{E}\{\frac{Z(t+h)-Z(t)}{h}\frac{Z(t+\delta)-Z(t)}{\delta}\}$$

exists. This limit can be re-written as

$$\lim_{\delta,h\to 0} \mathrm{E}\left\{\frac{Z_0(t+h)-Z_0(t)+[m(t+h)-m(t)]}{h}\frac{Z_0(t+\delta)-Z_0(t)+[m(t+\delta)-m(t)]}{\delta}\right\},$$

where $Z_0(t) = Z(t) - m(t)$. This is further equal to

$$\lim_{\delta,h\to 0} \frac{C(t+h,t+\delta)-C(t+h,t)-C(t,t+\delta)+C(t,t)}{h\delta} + \left[\frac{d}{dt}m(t)\right]^2,$$

and this limit exists by the assumed conditions. □

The derivatives of higher order are defined simply by induction: $Z^{(2)}(x)$ is the mean square derivative of $Z^{(1)}(x)$ and $Z^{(k)}(x)$ is the derivative of $Z^{(k-1)}(x)$, $k \geq 2$. In order that a process $Z(\cdot)$ has the nth mean square derivative, it is necessary and sufficient that the derivatives

$$\frac{d^k}{dx^k}m(x), \quad \frac{\partial^{k+l}}{\partial x^k \partial y^l}C(x,y),$$

exist for all $k, l \leq n$. Analogously to (2.8) and (2.9) in the proof of Lemma 2.1, the mean function of the kth derivative process $Z^{(k)}$ is given by

$$\mathrm{E}\, Z^{(k)}(x) = m^{(k)}(x) = \frac{d^k}{dx^k}m(x).$$

The cross-covariance function between $Z^{(k)}(x)$ and $Z^{(l)}(y)$ is determined by

$$\mathrm{cov}(Z^{(k)}(x), Z^{(l)}(y)) = C_{Z^{(k)}Z^{(l)}}(x,y) = \frac{\partial^{k+l}C(x,y)}{\partial x^k \partial y^l}.$$

Differentiability in $\mathbb{R}^d$

The concept of the mean square differentiability of processes generalizes straightforward to the higher dimensions. Nevertheless, unlike the univariate case, where only one derivative process $Z^{(1)}(t)$ exists for $Z^{(0)}(t)$, in $\mathbb{R}^d$ there are d random fields, which play the role of the first derivatives. Precisely, we have to consider the following random fields or the d-variate random field (see. definition p. 6)

$$\frac{\partial}{\partial x_1}Z(\boldsymbol{x}),\ \frac{\partial}{\partial x_2}Z(\boldsymbol{x}),\ \ldots,\ \frac{\partial}{\partial x_d}Z(\boldsymbol{x}), \qquad \boldsymbol{x} = (x_1,\ldots,x_d)^T, \tag{2.10}$$

which are the mean square derivatives in the direction of axes. Sometimes, we use to refer to (2.10) as *gradient random field*. Further, there are d^2 second mean square derivative fields in $\mathbb{R}^d$, d^3 third derivative fields etc. Generally, for a multi-index $\boldsymbol{n} = (n_1,\ldots,n_d) \in \mathbb{N}_0^d$, we can define the mean square $\boldsymbol{n}$th derivative $Z^{(\boldsymbol{n})}$ of Z

$$Z^{(\boldsymbol{n})}(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{n}|}}{\partial x_1^{n_1}\ldots\partial x_d^{n_d}}Z(\boldsymbol{x}) \tag{2.11}$$

where $|\boldsymbol{n}| = n_1 + n_2 + \ldots + n_d$. If $Z^{(\boldsymbol{n})}$ exists, the field Z is called $\boldsymbol{n}$-times mean square differentiable. Similarly as in $\mathbb{R}^1$, the mean function of the derivative process $Z^{(\boldsymbol{n})}(\boldsymbol{x})$ is given by

$$\mathrm{E}\, Z^{(\boldsymbol{n})}(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{n}|}}{\partial x_1^{n_1} \ldots \partial x_d^{n_d}} m(\boldsymbol{x}) \tag{2.12}$$

and the covariance function of the derivative random fields $Z^{(\boldsymbol{k})}(\boldsymbol{x})$ and $Z^{(\boldsymbol{l})}(\boldsymbol{y})$ is given by

$$\mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{y})) = \frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d} \partial y_1^{l_1} \ldots \partial y_d^{l_d}} C(\boldsymbol{x}, \boldsymbol{y}). \tag{2.13}$$

The sum $\boldsymbol{k}+\boldsymbol{l}$ is defined elementwise.

Clearly, a random field $Z(\boldsymbol{x})$ is $\boldsymbol{n}$-times differentiable, if and only if the derivatives

$$\frac{\partial^{|\boldsymbol{k}|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}} m(\boldsymbol{x}), \qquad \frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d} \partial y_1^{l_1} \ldots \partial y_d^{l_d}} C(\boldsymbol{x}, \boldsymbol{y}),$$

exist for all multi-indices $\boldsymbol{k} = (k_1, \ldots, k_d)$ and $\boldsymbol{l} = (l_1, \ldots, l_d)$, such that $k_i \leq n_i, l_i \leq n_i$ for all $i \in \{1, \ldots, d\}$.

If the random field Z is stationary and $K(\cdot)$ is the stationary representative of $C(\cdot, \cdot)$, the equation (2.13) obviously shorten to

$$\mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{y})) = K_{Z^{(\boldsymbol{k})}, Z^{(l)}}(\boldsymbol{h}) = (-1)^{|\boldsymbol{k}|} \frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial h_1^{k_1+l_1} \ldots \partial h_d^{k_d+l_d}} K(\boldsymbol{h}), \tag{2.14}$$

where $\boldsymbol{h} = \boldsymbol{y} - \boldsymbol{x}$. Particularly, it holds

$$\mathrm{var}(Z^{(\boldsymbol{k})}(\boldsymbol{x})) = \mathrm{var}(Z^{(\boldsymbol{k})}(\boldsymbol{0})) = (-1)^{|\boldsymbol{k}|} \frac{\partial^{2|\boldsymbol{k}|}}{\partial h_1^{2k_1} \ldots \partial h_d^{2k_d}} K(\boldsymbol{h}) \bigg|_{||\boldsymbol{h}||=\boldsymbol{0}}. \tag{2.15}$$

Let us note that under the stationarity of Z, it follows from (2.14) that

$$\mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{y})) = (-1)^{|\boldsymbol{k}+\boldsymbol{l}|} \mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{y}), Z^{(\boldsymbol{l})}(\boldsymbol{x})).$$

If the $\boldsymbol{n}$-times differentiable random field Z is moreover isotropic, $Z^{(\boldsymbol{n})}$ exists for any permutation of $\boldsymbol{n}$. In this case, Z is termed to be an $|\boldsymbol{n}|$-times differentiable random field.

From the text above it clear that we must be careful when speaking about observations of derivative random fields. For instance, we have to distinguish if all the first partial derivatives are to be observed or only one of them.

Definition 2 *A multi-index $\boldsymbol{k} = (k_1, \ldots, k_n) \in \mathbb{N}_0^d$ is called* even *if* **all** *$k_1, \ldots, k_n$ are even. The multi-index $\boldsymbol{k}$ is called* odd *if* **any** *of the integers $k_1, \ldots, k_n$ is odd.*

Lemma 2.2 *Let Z be a stationary, isotropic, n-times differentiable random field on $\mathbb{R}^d$ and let $\boldsymbol{k} = (k_1, \ldots, k_d)$ and $\boldsymbol{l} = (l_1 \ldots, l_d)$ be multi-indices, such that $|\boldsymbol{k}|, |\boldsymbol{l}| \leq n$. Then*

$$\mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x})) = 0, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

if $\boldsymbol{k} + \boldsymbol{l}$ is odd.

The proof of Lemma 2.2 can be found in Appendix A on the page 83. Note, that the proof formalizes the very fact that an odd derivative of an even function is an odd function in $\mathbb{R}$ and it extends this claim to higher dimensions. Note that the claim can not be conversed under given prerequisites. This can be seen from the fact that any function, which is constant in $\mathbb{R}$ is a covariance function. Note that Lemma 2.2 also holds under the geometric anisotropy (see page 9).

2.2. Linear prediction and design

In the opening of Chapter 2, we have considered that an observer (meteorologist or mining engineer) has collected some spatial distributed data (temperature levels resp. pollutions levels in the soil in different parts of a region). His or her main task is typically to suggest some assessment of a quantity in the locations, where the observations of the quantity were not realized. We speak about a prediction of the quantity. Here, the prediction does not mean a prediction of a behavior in the future, but a reconstruction of the quantity in different parts of the space, where it is not known precisely.

Design

Let us denote by

$$V_n = \{\boldsymbol{x}_i : i = 1 \dots n\} \in \mathcal{V}_n \subset X \subset D \tag{2.16}$$

a set of locations, where a realization $z(\boldsymbol{x})$ of a random field $Z(\boldsymbol{x})$ has been observed. The set X is referred to as *an observation area*. In the context of Section 2.3, the set V_n is called *(discrete) n-points experimental design* and it is an element of the set $\mathcal{V}_n$ of all discrete n-points experimental designs. The integer $n \in \mathbb{N}$ in (2.16) is called *size of the design* V_n.

There are some other issues considered in geostatistics. Typically, data are sampled not only at single points of X, but they always involve a support of a positive volume. The reader can take, for instance, a core drill or a soil sample of a certain diameter into consideration. This aspects are always application dependent and they are not considered here, i.e. observation locations will always be in the idealized form (2.16).

In the following, we will not distinguish the (uppercase) parent random field and its particular (lowercase) realization. It is clear that probabilistic calculations involve the random field and all numerical estimates involve the realization. Thus, let us denote by

$$\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \dots, Z(\boldsymbol{x}_n))^T$$

the observed particular data, or precisely, the random vector we have observed at the points given by the design V_n.

Later in Chapter 4, observations of multivariate random fields will be regarded. Specifically, we will consider derivative random fields. In this case, we call also the set

$$V_n = \{(\boldsymbol{x}_i, \boldsymbol{k}_i) : i = 1 \dots n\} \in \mathcal{V}_n \subset X \subset D \tag{2.17}$$

n-points experimental design. The notations (2.17) means that the random field $Z_{\boldsymbol{k}_i}$ should be observed at the location $\boldsymbol{x}_i$. In the case of derivative random fields, (2.17) means that the random field $Z^{(\boldsymbol{k}_i)}$ has to be observed at the location $\boldsymbol{x}_i$. The observed random vector

$$\boldsymbol{Z} = (Z_{\boldsymbol{k}_1}(\boldsymbol{x}_i), \dots, Z_{\boldsymbol{k}_n}(\boldsymbol{x}_n))^T \quad \text{resp.} \quad \boldsymbol{Z} = (Z^{(\boldsymbol{k}_1)}(\boldsymbol{x}_i), \dots, Z^{(\boldsymbol{k}_n)}(\boldsymbol{x}_n))^T \tag{2.18}$$

is called data. A design is called *regular* if $\mathrm{cov}(\boldsymbol{Z}, \boldsymbol{Z}^T)$ is regular.

Without going to details, let us note that the aim of an experiment planning is to find a design that performs efficiently or optimally in some pre-defined sense. To preconceive what will follow, the reader can observe that different settings of the optimization problem are possible under the definition (2.17). Basically, if a set of multi-indices $\{\boldsymbol{k}_1, \dots, \boldsymbol{k}_n\}$ is given, we can look for the optimal locations $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$. On the other hand, it is possible that the locations $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ are given and the task is to find corresponding an optimal set of multi-indices $\{\boldsymbol{k}_1, \dots, \boldsymbol{k}_n\}$. Of course, design conditions can be formulated so, that both the locations $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ and the multi-indices $\{\boldsymbol{k}_1, \dots, \boldsymbol{k}_n\}$ should be chosen optimally.

Structure of random fields

It is often reasonable to split the observed random field into several components. The first part is deterministic and it is given in a form of a functional relationship. Typically, it represents a macro-structure of the observed phenomenon, i.e. it is a natural law, which is known in some sense.

The second part represents variations from this law, i.e. it is a micro-structure of the observed random fields. There could be also a natural law behind this component, but this law is unknown. Therefore, the second part is modeled as a zero mean random field. Since the kind of randomness is closely related to the nature of the study quantity, the modeling of its stochastic structure is of the main interest.

Nevertheless, sometimes it is necessary to consider an additional source of uncertainty which arises from a particular observation mechanism. It is called *measurement error* or *(white) noise* and it is supposed to be an independent identically distributed (i.i.d.) zero mean random variable $\varepsilon(\boldsymbol{x})$ of the second order. The noise $\varepsilon(\boldsymbol{x})$, as emphasized, can depend on the location $\boldsymbol{x}$. Thus, the random field $Z(\boldsymbol{x})$ can be split into a sum of three components, the mean function $m(\boldsymbol{x})$, the zero mean random field $S(\boldsymbol{x})$ representing the smooth or noiseless version of $Z(\boldsymbol{x})$ and the noise $\varepsilon(\boldsymbol{x})$:

$$Z(\boldsymbol{x}) = m(\boldsymbol{x}) + S(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}), \quad \text{for } \boldsymbol{x} \in D. \tag{2.19}$$

The stochastic components $S(\boldsymbol{x})$ and $\varepsilon(\boldsymbol{x})$ are supposed to be independent.

Apart from that, since the main focus of the thesis is an investigation of differentiable random fields, a direct incorporation of the measurement error $\varepsilon(\cdot)$ into the model is not correct. The random field Z would no longer be mean square continuous and therefore not mean square differentiable.

Prediction error

Let us denote by $p(\boldsymbol{x}_0; \boldsymbol{Z})$ an arbitrary predictor of Z at the point $\boldsymbol{x}_0 \in D$ from data $\boldsymbol{Z}$. The difference

$$Z(\boldsymbol{x}_0) - p(\boldsymbol{x}_0; \boldsymbol{Z}) \tag{2.20}$$

is referred to as the *prediction error* of $p(\boldsymbol{x}_0; \boldsymbol{Z})$ and

$$\mathrm{E}\{Z(\boldsymbol{x}_0) - p(\boldsymbol{x}_0; \boldsymbol{Z})\}^2$$

is called *mean square error* (mse) of $p(\boldsymbol{x}_0; \boldsymbol{Z})$. We say that the predictor $p(\boldsymbol{x}_0; \boldsymbol{Z})$ is *unbiased* if

$$\mathrm{E}\, p(\boldsymbol{x}_0; \boldsymbol{Z}) = \mathrm{E}\, Z(\boldsymbol{x}_0) = m(\boldsymbol{x}_0). \tag{2.21}$$

Note that (2.21) can be rewritten as $\mathrm{E}\{p(\boldsymbol{x}_0; \boldsymbol{Z}) - Z(\boldsymbol{x}_0)\} = 0$, so $p(\boldsymbol{x}_0; \boldsymbol{Z})$ is unbiased if and only if its expected prediction error equals zero.

The main focus of the presented work is a survey of mean square error properties, considering not only prediction at a single location but rather prediction over an (possibly very large) *prediction area* denoted by W. We define the *integrated mean square error* over the area W (imse_W) as

$$\mathrm{imse}_W = \int_W \mathrm{mse}(\boldsymbol{x})\, \boldsymbol{dx}, \tag{2.22}$$

where $\boldsymbol{dx} = dx_1 \dots dx_d$ denotes a volume element of $\mathbb{R}^d$. Similarly, we can also consider the *maximal mean square error* over the area W (mmse_W), which is defined as

$$\text{mmse}_W = \sup_{\boldsymbol{x} \in W} \text{mse}(\boldsymbol{x}). \tag{2.23}$$

Note that Cressie (1991) modifies (2.22) and (2.23) by including weight function in the integral and behind the sup operator. He gives as an example of weights which focus attention on sub-regions whose mean and variance exceed some given threshold.

2.2.1. Simple hierarchy of prediction problems

In prediction problems, the following alternative states of knowledge can be considered.

1. All finite dimensional distribution functions (2.1) of $Z(\cdot)$ are known.

2. Only the mean $m(\boldsymbol{x})$ and covariance function $C(\boldsymbol{x}, \boldsymbol{y})$ of $Z(\boldsymbol{x})$ are known.

3. The covariance function $C(\boldsymbol{x}, \boldsymbol{y})$ of $Z(\boldsymbol{x})$ is known, but its mean function $m(\boldsymbol{x})$ is supposed to be in the linear form

$$m(\boldsymbol{x}) = \sum_{i=1}^{p} \beta_i f_i(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta}, \tag{2.24}$$

   where $\boldsymbol{f} = (f_1, \dots, f_p)^T$ is a known vector of the so-called *regression functions* and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ are unknown parameters.

4. The only knowledge is that $m(\boldsymbol{x})$ is assumed to be in the same form as in the third state, but the covariance function $C(\boldsymbol{x}, \boldsymbol{y})$ of $Z(\boldsymbol{x})$ is supposed to be a known function of an unknown parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)^T$, whose values are restricted to a known set Θ.

5. A nonparametric setup for $m(\boldsymbol{x})$ and $C(\boldsymbol{x}, \boldsymbol{y})$ can be assumed.

2.2.2. Best prediction

Suppose that we know all finite dimensional distribution functions of an unobservable random variable $Z(\boldsymbol{x}_0)$, whose realization is to predict. The predictor will be based on the observable random vector $\boldsymbol{Z}$. A possible predictor is the conditional expectation

$$p(\boldsymbol{x}_0; \boldsymbol{Z}) = \text{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z}). \tag{2.25}$$

It is well known that $\text{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z})$ has the minimal mean square error among all predictors. A predictor which minimizes mse is referred to as *the best predictor.* Note that the predictor (2.25) is unbiased. The mse of predictor (2.25) is

$$\text{mse}\,\text{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z}) = \text{var}\left\{Z(\boldsymbol{x}_0) - \text{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z})\right\} = \text{E}(\text{var}(Z(\boldsymbol{x}_0)|\boldsymbol{Z})).$$

2.2.3. Best linear prediction

Suppose that Z has mean function $m(\boldsymbol{x})$ and covariance function $C(\boldsymbol{x}, \boldsymbol{y})$, which are known, but we have no further knowledge of the finite dimensional distribution functions. Therefore, the conditional mean $\mathrm{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z})$ is generally not determinable from the available information. One possibility and a very frequent approach used in geostatistics is to consider a predictor which is a linear function of the observations

$$p_L(\boldsymbol{x}_0; \boldsymbol{Z}) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i) + \lambda_0 = \boldsymbol{\lambda}^T \boldsymbol{Z} + \lambda_0, \tag{2.26}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$. The coefficients λ_i are chosen in some optimal way to obtain desirable properties of the predictor (2.26). The mean square prediction error is equal

$$\mathrm{E}(Z(\boldsymbol{x}_0) - p(\boldsymbol{x}_0; \boldsymbol{Z}))^2 = \{m(\boldsymbol{x}_0) - \lambda_0 - \boldsymbol{\lambda}^T \boldsymbol{m}\}^2 + C(\boldsymbol{x}_0, \boldsymbol{x}_0) - 2\boldsymbol{\lambda}^T \boldsymbol{c} + \boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda}, \tag{2.27}$$

where $\boldsymbol{m} = \mathrm{E}\, \boldsymbol{Z}$ and $\boldsymbol{c} = \mathrm{cov}(\boldsymbol{Z}, Z(\boldsymbol{x}_0))$ and $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{Z}, \boldsymbol{Z}^T)$. The equation (2.27) means that the mse of the linear predictor (2.26) is just the squared mean of the prediction error plus its variance. It can be seen that, for any choice of $\boldsymbol{\lambda}$, the square mean term can be set to zero by taking $\lambda_0 = m(\boldsymbol{x}_0) - \boldsymbol{\lambda}^T \boldsymbol{m}$. Suppose further that $\boldsymbol{\Sigma}$ is invertible. It is well known (cf. e.g. Stein (1999); Cressie (1991) or Rao (1973)) that the second term reaches its minimum if

$$\boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1} \boldsymbol{c}. \tag{2.28}$$

The condition on λ_0 goes to

$$\lambda_0 = m(\boldsymbol{x}_0) - \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{m}. \tag{2.29}$$

The best linear predictor (BLP) is therefore equal to

$$p_L^*(\boldsymbol{x}_0; \boldsymbol{Z}) = m(\boldsymbol{x}_0) + \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Z} - \boldsymbol{m}) \tag{2.30}$$

and its mean square error equals

$$\mathrm{mse}\, p_L^*(\boldsymbol{x}_0; \boldsymbol{Z}) = C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c}. \tag{2.31}$$

Let us emphasize that both $\boldsymbol{c}$ and $\boldsymbol{\Sigma}$ directly depend on the design V_n. Moreover, inverting of $\boldsymbol{\Sigma}$ usually leads to a very complex dependence, where every entry in $\boldsymbol{\Sigma}^{-1}$ is a function of the whole design V_n.

Note that the best linear predictor is usually called *simple kriging* in literature related to geostatistics or to spatial prediction (Chilès and Delfiner, 1999; Cressie, 1991; Journel and Huijbregts, 1978; Deutsch and Journel, 1992; Stein, 1999). The name origins from G. Matheron, who named this method after South African mining engineer D. G. Krige, who proposed the method for estimation of ore deposits. The system of linear equations (2.28) and (2.29) is called *kriging system.*

It is very significant property of the mean square prediction error (2.31) that it does not depend on the particular collected data. The mse (2.31) is merely function of the location of the predicted variable and the design V_n. This observation is a starting point for a formulation of a design problem as discussed later in Section 2.3.

2.2.4. Best linear unbiased prediction

Suppose the third state of knowledge (see page 15), i.e.

$$\mathrm{E}\, Z(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}.$$

Let us denote $\boldsymbol{F} = (\boldsymbol{f}(\boldsymbol{x}_1), \ldots, \boldsymbol{f}(\boldsymbol{x}_n))^T$ for $\boldsymbol{x}_i \in V_n$ (i.e. $\boldsymbol{F}$ is $(n \times p)$-matrix). Further, let $\boldsymbol{\Sigma}$ be a covariance matrix and $\boldsymbol{c}$ vector of covariances defined the same way as in Section 2.2.3 and let $\boldsymbol{\Sigma}$ be regular. If $\boldsymbol{\beta}$ were known, we could use the BLP

$$p_L^*(\boldsymbol{x}_0; \boldsymbol{Z}) = \boldsymbol{f}(\boldsymbol{x}_0)^T\boldsymbol{\beta} + \boldsymbol{c}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{F}\boldsymbol{\beta}) \tag{2.32}$$

for estimating of $Z(\boldsymbol{x}_0)$. Nevertheless, $\boldsymbol{\beta}$ is supposed to be unknown, then it is natural to substitute $\boldsymbol{\beta}$ in (2.32) by an estimator. One possibility is to use generalized least squares estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}, \tag{2.33}$$

where $\boldsymbol{F}$ is supposed to be of full rank. It is well known (see for instance Rao (1973) or Rao and Toutenburg (1995)) that $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$, i.e. covariance matrix of $\hat{\boldsymbol{\beta}}$

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1} \tag{2.34}$$

is minimal (in sense of partial ordering of positive definite matrices). Note that the predictor (2.32) with $\boldsymbol{\beta}$ replaced by its estimate (2.33) is linear and unbiased. We denote this predictor as $p_{LU}^*(\boldsymbol{x}_0; \boldsymbol{Z})$. The predictor is referred to as *the best linear unbiased predictor* (BLUP).

An alternative but equivalent approach how to obtain the BLUP, is to minimize the mean square prediction error among all predictors of the linear form (2.26) subject to the unbiasedness constraint, i.e.

$$\mathrm{E}(\lambda_0 + \boldsymbol{\lambda}^T\boldsymbol{Z}) = \mathrm{E}\, Z(\boldsymbol{x}_0) = \boldsymbol{f}(\boldsymbol{x}_0)^T\boldsymbol{\beta}, \qquad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p.$$

This explains the used terminology. The minimization of the mse leads to the same result and it can be seen that the mse of the BLUP is

$$\mathrm{mse}\, p_{LU}^*(\boldsymbol{x}_0; \boldsymbol{Z}) = C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{c}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{c} + ||\boldsymbol{m}(\boldsymbol{x}_0) - \boldsymbol{F}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{c}||^2_{\mathrm{var}(\hat{\boldsymbol{\beta}})},$$

where $||\boldsymbol{v}||_A$ denotes norm defined by $||\boldsymbol{v}||^2_{\boldsymbol{A}} = \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}$ for an $(n \times n)$-positive definite matrix $\boldsymbol{A}$ and $\boldsymbol{v} \in \mathbb{R}^n, n \in \mathbb{N}$. The variance matrix $\mathrm{var}\,\hat{\boldsymbol{\beta}}$ is given in (2.34) and since $\boldsymbol{\Sigma}$ is supposed to be regular and $\boldsymbol{F}$ of the full rank, $\mathrm{var}\,\hat{\boldsymbol{\beta}}$ is positive definite.

2.2.5. Components of prediction error

Now, let us consider the fourth state of knowledge. Let $p(\boldsymbol{x}_0; \boldsymbol{Z})$ be an arbitrary predictor of $Z(\boldsymbol{x}_0)$. As noted by Harville (1985), the prediction error (2.20) can be decomposed into four components in accordance with the hierarchy given on the page 15:

$$\begin{aligned} Z(\boldsymbol{x}_0) - p(\boldsymbol{x}_0; \boldsymbol{Z}) = &\underbrace{\Big[Z(\boldsymbol{x}_0) - \mathrm{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z})\Big]}_{\text{(inherent term)}} + \underbrace{\Big[\mathrm{E}(Z(\boldsymbol{x}_0)|\boldsymbol{Z}) - p_L^*(\boldsymbol{x}_0; \boldsymbol{Z})\Big]}_{\text{(unknown distributions)}} \\ &+ \underbrace{\Big[p_L^*(\boldsymbol{x}_0; \boldsymbol{Z}) - p_{LU}^*(\boldsymbol{x}_0; \boldsymbol{Z})\Big]}_{\text{(unknown mean)}} + \underbrace{\Big[p_{LU}^*(\boldsymbol{x}_0; \boldsymbol{Z}) - p(\boldsymbol{x}_0; \boldsymbol{Z})\Big]}_{\text{(unknown covariance function)}}. \end{aligned}$$

The first component represent an error that can not be avoided even if we would know all finite dimensional distribution functions of Z. The second, third and fourth components represents errors that can be attributed to losses of information incurred in passing successively from the first state of knowledge to second, from the second state to the third and from the third state to the fourth, under consideration that linear predictors have been chosen.

In Harville (1985), it is shown in more general setting that the first three components are mutually uncorrelated and conditions under that the fourth component is uncorrelated with the other are given in this paper. Further, he discussed an approximation of the fourth component when the unknown parameters of covariance function model are estimated. This approach was later revisited in Harville and Jeske (1992) and Zimmerman and Cressie (1992) and recently further developed by Abt (1999).

The previous section gives a basic view to the problem of spatial prediction. The practical problems are usually more complicated since they are liable to many application specific constraints. In the cited geostatistical literature, many additional conditions are considered, as for instance solution of the kriging problem under inequalities, problem of support and many other.

Let us note to the terminology that the best linear unbiased prediction is also called *ordinary kriging* if $p = 1$ and $f_1(\boldsymbol{x}) \equiv 1$ and *universal kriging* in the general case (2.24).

Note that all kriging predictors are unbiased and moreover they interpolate given data. Note that for construction of kriging predictors the assumption of stationarity and isotropy is not necessary. As long as the first and second moment is known, kriging can be used.

2.2.6. Cokriging and using gradient information

As already discussed, the information available on the study natural phenomenon is rarely limited to values of a single variable over a set of sample points. The concept of kriging can be straightforward generalized to handle multivariate problems. The multivariate best (linear, unbiased) predictor are sometimes referred to as *cokriging*. Despite an increased complexity, which is due to the notation involving multiple indices over covariables, there is only a little novelty from the theoretical point of view. Formally, it suffices to pool all the random variables representing data into a single Hilbert space and project the predicted variables onto that space.

A sufficient condition for construction of simple cokriging predictor is knowledge of covariances between all observed random variables and covariances between observed and predicted variables (Chilès and Delfiner, 1999; Cressie, 1991).

In case of universal cokriging, i.e. when trends of covariables under study need to be considered, the situation is not further more complicated. We must just consider particular functional dependencies between variables.

From the previous, it clear that there is no problem when incorporating mean square derivatives of a random field into prediction. The mean and covariance function of mean square derivatives are just appropriate ordinary derivatives of mean or covariance function (Section 2.1.4) and therefore the functional dependencies are easily available.

Consider an observable vector of random variables $\boldsymbol{Z} = (Z^{(\boldsymbol{k}_1)}(\boldsymbol{x}_1), \ldots, Z^{(\boldsymbol{k}_n)}(\boldsymbol{x}_n))^T$, where $Z^{(\boldsymbol{k}_i)}(\boldsymbol{x}_i)$ is assumed to be the $\boldsymbol{k}_i$th (partial) mean square derivative of the random field Z (see (2.11)). Let us denote $\boldsymbol{m} = \mathrm{E}\,\boldsymbol{Z}$, $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{Z}, \boldsymbol{Z}^T)$ and let $m(\boldsymbol{x}_0) = \mathrm{E}\, Z^{(\boldsymbol{k}_0)}(\boldsymbol{x}_0)$. Let us recall that $\boldsymbol{m}$, $m(\boldsymbol{x}_0)$ and $\boldsymbol{\Sigma}$ are defined as appropriate partial derivatives of mean and

covariance function of Z (see (2.12) and (2.13)). The BLP of the value of the random field $Z^{(\boldsymbol{k}_0)}$ at location $\boldsymbol{x}_0$ remains formally the same as in case with non-derivate observations and it is given by (2.30).

An analogous situation is when considering the BLUP (for prediction $Z^{(\boldsymbol{k}_0)}$ at the location $\boldsymbol{x}_0$). Also the BLUE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is given by (2.33), where

$$\boldsymbol{F} = (\boldsymbol{f}^{(\boldsymbol{k}_1)}(\boldsymbol{x}_1), \ldots, \boldsymbol{f}^{(\boldsymbol{k}_n)}(\boldsymbol{x}_n))^T \in \mathbb{R}^{n \times p}.$$

Examples of an explicite formulation of kriging equations for derivative observations can be found in Röttig (1996) and Albrecht (1998).

2.3. Optimal design

In the previous section, we have discussed a few very basic possibilities for a prediction of an unknown value (depending on different states of knowledge about the observed response Z) available to a practitioner after he or she collected data. On the other hand, the objective of an experiment is usually known before the data are sampled. Therefore, it is natural to do an active step already before collecting data and adapt the data-collecting procedure to a particular problem if possible. The aim is loosely speaking to obtain "optimal" data for possibly (but not necessary) "optimal" procedure.

A rather general characterization of design problems is summarized according to Fedorov (1996) in the next five assumptions.

1. There are *design parameters*, i.e. variables $\boldsymbol{x} \in X \subset \mathbb{R}^d$, which may be controlled. Usually, $d = 2$, and in the observing station problem, $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ are coordinates of stations and X is a region where these stations may be allocated.

2. There exists a model describing the observed response(s) or dependent variables(s) Z. More specifically, $\boldsymbol{x}$ and Z are linked together by a model, which may contain some stochastic components.

3. An experimenter or a practitioner can formulate the quantitative objective function.

4. Once a station or sensor is allocated, the response Z can be observed either continuously or according to any given time schedule without any additional significant expense.

5. Observations made at different sites may be correlated.

If Assumptions 4 and 5 are not considered, such a design problem is referred to as *standard design problem*. This problem was extensively discussed (see for instance Pukelsheim (1993); Pázman (1986) and Bandemer (1977)) and the theory is very compact and well developed.

Assumption 4 is related to quite usual case in that the aim is to ensure continuous monitoring of a data generating process or to allow for point prediction of present or future states of nature. Environmental monitoring, meteorology and seismology are most typical areas where this assumption is addressed. This assumption related more to time dependent observations is, however, not considered here at all and we concentrate only to spatial meaning of prediction.

Assumption 3 is the point, where the design theory splits in many directions. The following section gives a very rough overview about possible objective functions.

2.3.1. Design criteria

Following Müller (2001), the main construction principle for a well designed experimental plan is as follows. Asserting hypothesis-testing or parameter estimation as the aim of the experiment, the task remains to take a (possibly given) number of measurements in a way, that either the power of the test or precision of the estimate is maximized. Additional requirements such as the unbiasedness of the statistical method, restrictions on the experimental conditions (e.g. on a certain experimental region, etc.) usually enter as side-conditions of the optimization problem.

Design criteria for parameter estimation

The main stream of the design theory literature assumes the third state of knowledge (see page 15) to be the model mentioned in Assumption 2. The aim is to suggest a design which leads to an effective estimation of the unknown vector of parameters $\boldsymbol{\beta}$.

Suppose the standard design problem, i.e.,

$$Z(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta} + \varepsilon$$

where ε is a zero-mean i.i.d. variable with $\mathrm{var}(\varepsilon) = \sigma^2$. How can the goodness of an estimator of $\boldsymbol{\beta}$ be measured and compared? How can the information a particular design serves be described? The key to the problem are the so-called information matrices.

As the observation are uncorrelated, the least squares estimator of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}\boldsymbol{Z}$$

coincides with the best linear unbiased estimator of $\boldsymbol{\beta}$. The covariance matrix of this estimator is

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{F}^T\boldsymbol{F})^{-1}. \tag{2.35}$$

The problem is how to choose an discrete design (2.16) (or (2.17)) which minimizes (a suitable functional of) $\mathrm{cov}\,\hat{\boldsymbol{\beta}}$. As ε is i.i.d., repeated observations of Z can be considered at a particular design point. Thus design (2.16) can be re-written as

$$V_n = \left\{ \begin{array}{c} \boldsymbol{x}_i \\ r_i \end{array} \right\}_{i=1}^{N} \quad \text{or more often as} \quad V_n = \left\{ \begin{array}{c} \boldsymbol{x}_i \\ p_i \end{array} \right\}_{i=1}^{N},$$

where r_i (p_i) stands for (relative) number of repetitions of measurements at the location $\boldsymbol{x}_i$. It means that $\sum_{i=1}^{N} r_i = n$ and $p_i = r_i/n$. Instead of (2.35), it is more convenient, nevertheless equivalent, to consider the *information matrix*

$$\boldsymbol{M}(V_n) = \frac{1}{n}\boldsymbol{F}\boldsymbol{F}^T = \sum_{i=1}^{N} p_i \boldsymbol{f}(\boldsymbol{x_i})\boldsymbol{f}(\boldsymbol{x_i})^T,$$

where heteroscedascity due to the repeated observations is taken into account. Note that a design defined this way is a probability measure on the set X.

Ordering of designs

The aim is to maximize the matrix $\boldsymbol{M}(V_n)$ in some sense. We can use the so-called *Loewner ordering* $<_L$ of positive definite matrices. We write $\boldsymbol{A} <_L \boldsymbol{B}$ if $\boldsymbol{B} - \boldsymbol{A}$ is positive definite, for $\boldsymbol{A}, \boldsymbol{B} \in PD_n$. It means, the Loewner ordering is a semi-ordering with respect to the definition of positive definiteness. Note that another type of ordering (Schur ordering) is discussed in Pázman (1986, Chapter 3). The so-called Kiefer ordering is introduced by Pukelsheim (1993).

The design V_n^1 is said to be *(uniformly) better* than the design V_n^2 (write $V_n^2 < V_n^1$) if $\boldsymbol{M}(V_n^2) <_L \boldsymbol{M}(V_n^1)$. However, since the Loewner ordering is unfortunately only a semi-ordering, non-comparable matrices (and therefore non-comparable designs) can be easily found. Thus, we have to introduce a suitable scalar functional Ψ on the space of positive definite matrices to arrive at a complete ordering of candidate designs.

The reasonable condition such a functional Ψ must fulfill is monotony with the Loewner ordering, i.e.,

$$\text{if } \boldsymbol{A} <_L \boldsymbol{B} \quad \text{then} \quad \Psi(\boldsymbol{A}) < \Psi(\boldsymbol{B}). \tag{2.36}$$

The functional with the property (2.36) is referred to as *optimality criterion.* Subsequently, the design

$$V_n^* = \arg \min_{V_n \in \mathcal{V}_n} \Psi(\boldsymbol{M}(V_n)) \tag{2.37}$$

is called $\Psi-$*optimal.*

Let us mention the most important optimality criteria. In the following, $W \subset \mathbb{R}^d$ denotes the so-called *prediction area*, i.e. the set for which the estimation of trend is required. Furthermore, $\lambda_{\max}(\boldsymbol{A})$ stands for the maximal eigenvalue of the matrix $\boldsymbol{A}$.

$$\begin{aligned}
&D\text{-}optimality, \text{ i.e.,} && \Psi(\boldsymbol{M}(V_n)) = \det \boldsymbol{M}^{-1}(V_n),\\
&A\text{-}optimality, \text{ i.e.,} && \Psi(\boldsymbol{M}(V_n)) = \operatorname{tr} \boldsymbol{M}^{-1}(V_n),\\
&E\text{-}optimality, \text{ i.e.,} && \Psi(\boldsymbol{M}(V_n)) = \lambda_{\max}(\boldsymbol{M}^{-1}(V_n))\\
&G\text{-}optimality, \text{ i.e.,} && \Psi(\boldsymbol{M}(V_n)) = \sup_{\boldsymbol{x} \in W} \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{M}^{-1}(V_n) \boldsymbol{f}(\boldsymbol{x}),\\
&I\text{-}optimality, \text{ i.e.,} && \Psi(\boldsymbol{M}(V_n)) = \int_W \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{M}^{-1}(V_n) \boldsymbol{f}(\boldsymbol{x})\, d\boldsymbol{x},\\
&C\text{-}optimality, \text{ i.e.,} && \Psi(\boldsymbol{M}(V_n)) = \boldsymbol{c}^T \boldsymbol{M}^{-1}(V_n) \boldsymbol{c}, \quad \boldsymbol{c} \in \mathbb{R}^p.
\end{aligned}$$

The criteria above have an intuitive interpretation. For instance, if ε is Gaussian and given a confidence level α, a D-optimal design minimizes the volume of the confidence ellipsoid for $\boldsymbol{\beta}$ (see e.g. Pázman (1986) and Müller (2001) for further interpretation details on the other criteria).

Approximative design

The previous paragraph provides a basic view to the standard optimal design problem. However, the approach has many lacks. One of them is that discrete designs are not closed to some simple operations. Clearly, the sum of two particular $n-$points discrete designs can, for instance, be a $2n$-points discrete design. Therefore, the concept has been further generalized and the so-called *approximative design* has been introduced. An approximative design is a probability measure ξ with supporting set belonging to X. The information

matrix is defined as

$$M(\xi) = \int_X \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T \xi(\boldsymbol{dx}).$$

This formulation has many advantages. It can be seen that the space of information matrices forms a convex set and information matrices depend linearly on design points. If the optimality criteria are convex (and fortunately many criteria are convex), some general theorems from the theory of optimization can be used. In the design theory, the equivalence theorem of Whittle and the theorem of Kiefer and Wolfowitz became famous (see e.g. Näther (1985); Pukelsheim (1993) and Fedorov (1996)).

In practice, construction of an optimal design is usually done iteratively. Starting from an initial design, this design is improved in every iteration. There is a huge amount of literature, where the convergence of iterative procedures is discussed and a lot of modifications supposed. Let us cite Pázman (1986) and Bandemer (1977) inter alia. The whole concept of the standard design problem was generalized by Pukelsheim (1993).

Design criteria for parameter estimation - correlated observations

Suppose the following linear model (cf. (2.19))

$$Z(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta} + S(\boldsymbol{x}), \quad \boldsymbol{x} \in D.$$

The covariance function $C(\cdot,\cdot)$ of $S(\cdot)$ is supposed to be mean square continuous. The aim of a candidate design is the best possible estimation of the unknown vector of parameters $\boldsymbol{\beta}$.

Similarly to the previous section, we can take the best linear unbiased estimator of $\boldsymbol{\beta}$, which is given by the equation (2.33). The variance (2.34) of the estimator (2.33) is a reasonable criterion of optimality. Note that the continuity of the covariance function $C(\cdot,\cdot)$ immediately avoids repeated observations, since $\boldsymbol{\Sigma}$ in (2.34) is assumed to be regular.

We can of course compare designs by means of the Loewner ordering. Unfortunately, information matrices $\boldsymbol{M}(V_n) = (\operatorname{var}\hat{\boldsymbol{\beta}})^{-1}$ are no longer linear in the design points. Moreover, the dependence on the particular design is usually very complex due to the inverse of the covariance matrix in (2.34). We can formally define optimality criteria as in the previous section and we can consider the following optimization problem

$$V_n^* = \arg\min_{V_n \in \mathcal{V}_n} \Psi(\boldsymbol{M}(V_n)),$$

which has not too much common with (2.37) besides the notation. For instance, convexity of Ψ is not very helpful anymore. In the most studies, authors try to imitate the iterative construction methods mentioned in the previous section. An example of this heuristic procedures was proposed by Brimkulov (see Näther (1985); Müller (2001) and Fedorov (1996)). Unfortunately, there is no proof of convergence of this algorithm.

Another approach to the optimal design for the BLUE (the so-called method of perturbed covariance matrices) was developed by Müller (2001) and recently further adapted for prediction by Holtmann (2001).

Design criteria for prediction

If the goal is a prediction of a random field at a given number of points, it is trivial that the optimal design (without restriction on the number of support points) is realized when

observations of the field are taken exactly at the sites of interests. However, the number of observations is typically much less than number of locations where is to predict. The prediction area can be even a continuous set. In the case of prediction, design criteria can be based on the mean square error of a particular predictor. The optimality of a design can be taken in regard to the value of (2.22) or (2.23).

It can be seen that an exhaustive search over all potential designs is unfeasible. If the number of points of interest is not large some heuristic exchange algorithms (Barnes, 1989; Gao et al., 1996) or fast cooling can be used (Sacks and Schiller, 1988). Relative performance of the kriging predictor on regular triangular, hexagonal and square grids is investigated by Yfantis et al. (1987).

2.3.2. Analytical form of optimal designs for some covariance functions

One of the rare cases, where the explicite form of the optimal design is known, is due to Abt (1992). He studied Gaussian processes in the interval $W = [0, 1]$ with three possible covariance functions, namely, *the linear covariance function*

$$C_l(x, y) = 1 - (1 - \rho)|x - y|,\ 0 < \rho < 1,$$

covariance function of the Brownian motion

$$C_m(x, y) = \min(x, y)$$

and *the Brownian bridge*

$$C_b(x, y) = \min(x, y) - xy.$$

He considered ordinary and simple kriging model and found the exact optimal designs regarding to the D-criterion, integrated mean square error and maximum mean square error. These optimality criteria are denoted here as mmse_{ok}, mmse_{sk}, imse_{ok}, and imse_{sk} with an obvious interpretation. Note that only the linear covariance function is the stationary one and none of them is differentiable.

The proofs are purely analytical and they are enabled because of the unique structure of the covariance matrix, which inverse is in a tridiagonal form with a possible entry in the lower left and upper right corner.

For $n \geq 2$, let V_n be a discrete n-points design (2.16) with $0 \leq x_1 < x_2 < \ldots < x_n \leq 1$. The optimal choice of x_i is summarized in the following table.

-	C_l	C_m	C_b
D	$(i-1)/(n-1)$	i/n	$i/(n+1)$
mmse_{ok}	$(4i-3)/(4n-2)$	$(4i-3)/(4n-2)$	$*$
imse_{ok}	$(3i-2)/(3n-1)$	$(3i-2)/(3n-1)$	$*$
mmse_{sk}	-	$4i/(4n+1)$	$i/(n+1)$
imse_{sk}	-	$3i/(3i+1)$	$i/(n+1)$

Table 2.1.: Discrete optimal designs for some covariance functions.

Considering the Brownian bridge, it can be shown that any design satisfying $0 < x_1 < \ldots < x_n < 1$ and $x_i - x_{i-1} = 1/n$ for $i = 2, \ldots, n$ is optimal with respect to the imse_{ok} as well as to the mmse_{ok} criterion. This is denoted as " $*$ " in Table 2.1.

For the linear covariance function C_l, Abt (1992) states that optimal designs with respect to the both imse_{ok} and mmse_{ok} criterion are equispaced in the interval $[0, 1]$, but the difference between design points depends basically on t_1, which cannot be found explicitely. This is denoted as " $-$ " in Table 2.1.

3. More about covariance functions

Covariance functions play a crucial role in the analysis of random fields. For purposes of modeling, it is important that a variety of covariance functions with different additional properties is available. Unfortunately, especially many books written mainly for an applied audience discuss only a few covariance function models even if a lot of other ones are known for many years.

Being far from complete, let us first recall some necessary and sufficient conditions, which help to decide whether some candidate function is or is not a covariance function. The chapter follows with a more detailed listing of properties of some examples of differentiable covariance functions.

3.1. Properties and construction of covariance functions

3.1.1. Characterization of covariance functions

Let Z be a real second order random field. Its covariance function C has the property that

$$0 \leq \sum_{i=1}^{n} \sum_{j=1}^{n} a_i C(\boldsymbol{x}_i, \boldsymbol{x}_j) a_j \tag{3.1}$$

for all $n \in \mathbb{N}$ and for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $a_1, \ldots, a_n \in \mathbb{R}$. This can easily be seen from $\mathrm{var}\{\sum_{i=1}^{n} a_i Z(\boldsymbol{x}_i)\}$ which is equal to the right side of (3.1). A function satisfying (3.1) is said to be *positive semidefinite*. The condition (3.1) is necessary and sufficient for the existence of a second order random field with the covariance function C. The necessity is already noted above. The sufficiency follows by showing that for a given C a Gaussian random field exists with this covariance function.

Suppose Z is a second order stationary random field on $\mathbb{R}^d$ with a covariance function C. Then C must satisfy

$$\begin{aligned} C(\mathbf{0}) &\geq 0, \\ C(\boldsymbol{x}) &= C(-\boldsymbol{x}) \quad \text{and} \\ |C(\boldsymbol{x})| &\leq C(\mathbf{0}). \end{aligned}$$

3.1.2. Spectral representation

In this section, a representation of second order stationary and isotropic random fields by means of the Stieltjes-Fourier integral will be given. Spectral methods are a powerful tool for studying random fields and they simplify the derivation of many formulae as it will be seen later. In Fourier analysis, it is more natural to consider complex-valued functions rather than restricting oneself to real-valued functions. Thus, let us suppose that Z is a complex valued second order stationary random field on $\mathbb{R}^d$ with a covariance function C.

Bochner's theorem

Theorem 1 (Bochner) *A continuous function $C : \mathbb{R}^d \to \mathbb{C}$ is positive definite if, and only if it can be expressed in the form*

$$C(\boldsymbol{x}) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T\boldsymbol{x})\mathcal{F}(\boldsymbol{d\omega}), \tag{3.2}$$

where $\mathcal{F}$ is a non-negative finite measure (cf. e.g. Stein (1999) or Gikhman and Skorokhod (1996)).

In the equation (3.2), $\boldsymbol{d\omega} = d\omega_1 \dots d\omega_d$ denotes a volume element of the d-dimensional space $\mathbb{R}^d$. The prerequisite that $\mathcal{F}$ is a non-negative finite measure means that $\mathcal{F}(\boldsymbol{\Delta})$ is a non-negative functional on the d-dimensional set $\boldsymbol{\Delta}$ determined, in particular, for all d-dimensional intervals $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_d)$. Since $\mathcal{F}$ is a finite measure, it must hold

$$\mathcal{F}(\boldsymbol{\Delta}_1 \cup \boldsymbol{\Delta}_2) = \mathcal{F}(\boldsymbol{\Delta}_1) + \mathcal{F}(\boldsymbol{\Delta}_2), \quad \text{for all disjoint Borel sets } \boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2,$$

and

$$\mathcal{F}(\mathbb{R}^d) < \infty.$$

Following Yaglom (1987a, p.329), the equation (3.2) can be re-written as

$$C(\boldsymbol{x}) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T\boldsymbol{x})d\mathcal{F}(\boldsymbol{\omega}), \tag{3.3}$$

where $\mathcal{F}(\boldsymbol{\omega})$ is a non-negative bounded nondecreasing function, which is continuous everywhere on the left. The function $\mathcal{F}(\boldsymbol{\omega})$ in (3.3) is called *spectral distribution function* of the stationary random field Z. If spectral distribution function $\mathcal{F}(\cdot)$ is absolutely continuous, the function $f(\boldsymbol{\omega})$ defined by

$$\mathcal{F}(\boldsymbol{u}) = \int_{-\boldsymbol{\infty}}^{\boldsymbol{u}} f(\boldsymbol{\omega})\boldsymbol{d\omega}$$

is called the *spectral density*. We can write

$$C(\boldsymbol{x}) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T\boldsymbol{x})f(\boldsymbol{\omega})\boldsymbol{d\omega}.$$

This means that $C(\boldsymbol{x})$ is the *Fourier transform* F of $f(\boldsymbol{\omega})$, $Ff = C$. If the covariance function C is absolutely integrable on the $\mathbb{R}^d$, i.e. $\int_{\mathbb{R}^d} |C(\boldsymbol{x})|\boldsymbol{dx} < \infty$, we have the inversion formula

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-i\boldsymbol{\omega}^T\boldsymbol{x})C(\boldsymbol{x})\boldsymbol{dx}.$$

It means, $f(\boldsymbol{\omega})$ is the *inverse Fourier transform* F^{-1} of the covariance function $C(\boldsymbol{x})$, $F^{-1}C = f$.

Bochner's theorem for isotropic random fields

If a second order stationary random fields is also isotropic then the equation (3.3) has a special form. Note that the isotropy of a covariance function implies the isotropy of its spectral density, i.e. $f(\boldsymbol{\omega}) = f(\omega)$, for $\omega = ||\boldsymbol{\omega}||$. It is shown e.g. in Gikhman and Skorokhod (1996) by a straightforward computation that the following theorem holds.

Theorem 2 *A real function $C(r)$, $(0 \leq r < \infty)$ is a continuous isotropic covariance function of a random field on $\mathbb{R}^d$ $(d \geq 2)$, if and only if it can be represented as*

$$C(r) = 2^{(d-2)/2}\,\Gamma(d/2)\int_0^\infty (ur)^{-(d-2)/2}\,J_{(d-2)/2}(ur)d\mathcal{G}(u), \quad \textit{for all } r \in [0,\infty), \tag{3.4}$$

where $\mathcal{G}(u)$ is a nondecreasing function bounded on $[0,\infty)$ and $\mathcal{G}(0) = 0$. Here, $J_\nu(\cdot)$ stands for the Bessel function of the first kind of the order ν.

The right side of equation (3.4) is also known as the Hankel transform of the order $(d-2)/2$ of $\mathcal{G}$. Let us note that the integrand in (3.4) is itself a covariance function (cf. Section 3.2.3). From Theorem 2, it also follows that any isotropic covariance function can be expressed as a scale mixture of certain (the so-called J-Bessel) covariance functions. Particularly, for $d = 2$, the formula (3.4) has the following simple form

$$C(r) = \int_0^\infty J_0(ur)d\mathcal{G}(u),$$

and for $d = 3$

$$C(r) = 2\int_0^\infty \frac{\sin(ur)}{ur}d\mathcal{G}(u).$$

Lower bounds

As a consequence, Theorem 2 yields the following statements about lower bounds of a covariance function (see Yaglom (1987a)). If $C(\cdot)$ is a continuous isotropic correlation function on $\mathbb{R}^d$ then

$$\begin{aligned} d = 1 \quad &: \quad C \geq \inf_{s\geq 0}\cos(s) = -1, \\ d = 2 \quad &: \quad C \geq \inf_{s\geq 0} J_0(s) \approx -0.40276, \\ d = 3 \quad &: \quad C \geq \inf_{s\geq 0}\frac{\sin(s)}{s} \approx -0.21723. \end{aligned}$$

Plancherel theorem

For a function $f : \mathbb{R}^d \to \mathbb{R}$, let us denote by $f \in L^p(\mathbb{R}^d)$ the fact that

$$||f||_p := \Big(\int_{\mathbb{R}^d} |f(\boldsymbol{x})|^p \boldsymbol{dx}\Big)^{1/p} < \infty.$$

We will mainly use absolutely integrable functions, i.e. functions from $L^1(\mathbb{R}^d)$, and square integrable functions, i.e. functions that are elements of $L^2(\mathbb{R}^d)$. Note that since $\mathbb{R}^d$ is not finite, $L^1(\mathbb{R}^d) \not\subseteq L^2(\mathbb{R}^d)$ and also $L^2(\mathbb{R}^d) \not\subseteq L^1(\mathbb{R}^d)$.

In the next chapter, the following classical result from the spectral analysis will be used (cf. e.g. Kuttler (1998)).

Theorem 3 (Plancherel) *For $f \in L^2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, it holds*

$$||f||_2^2 = (2\pi)^{-d}\,||Ff||_2^2 = (2\pi)^d\,||F^{-1}f||_2^2.$$

Convolution theorem

Let $f, g \in L^1(\mathbb{R}^d)$. Define *the convolution* $f * g : \mathbb{R}^d \to \mathbb{R}$ as

$$f * g(\boldsymbol{x}) = \int_{\mathbb{R}^d} f(\boldsymbol{x} - \boldsymbol{y}) g(\boldsymbol{y}) d\boldsymbol{y}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

The function $f * g$ is an element of $L^1(\mathbb{R}^d)$ and $||f * g||_1 \leq ||f||_1 ||g||_1$.

Theorem 4 *Let $f, g \in L^1(\mathbb{R}^d)$. Then*

$$\begin{aligned} F(f * g)(\boldsymbol{x}) &= Ff(\boldsymbol{x})\, Fg(\boldsymbol{x}) \quad \textit{and} \\ F^{-1}(f * g)(\boldsymbol{x}) &= (2\pi)^d\, F^{-1} f(\boldsymbol{x})\, F^{-1} g(\boldsymbol{x}) \end{aligned}$$

for almost every $\boldsymbol{x} \in \mathbb{R}^d$. If moreover $F^{-1}(f * g) \in L^1(\mathbb{R}^d)$ then

$$f * g = (2\pi)^d F(F^{-1} f F^{-1} g) \textit{ almost everywhere.} \tag{3.5}$$

Analogous claims and its generalizations can be found in e.g. Kuttler (1998).

3.1.3. Construction of covariance functions

Fourier transform of a non-negative function From the Bochner's theorem, it follows immediately that the Fourier transform (3.3) of any non-negative function is positive definite. The Bochner's theorem is therefore often used as a criterion for positive definiteness.

Rescaling Rescaling of the domain $C(\boldsymbol{x}) = C(\alpha \boldsymbol{x})$ is a trivial transformation on the space of covariance functions. The parameter α controls the so-called range of the covariance function. Anyway, if changing the scale and considering diverse derivative fields, some care is needed. A change of the scale of a underlying fields means also a change of the variance for a particular derivative field. In Section 3.2, exact dependencies of the variance of derivative fields on the scale parameter are given for some particular examples of covariance functions.

Adding and multiplying Any linear linear combination of positive definite functions with *non-negative* scalars a_i yields a positive definite function. The product of two real-valued positive definite functions is positive definite.

Reduction of dimensionality A valid model C for a stationary isotropic random field in $\mathbb{R}^d$ is valid also in $\mathbb{R}^{d-1}$.

Pointwise limit If $C_1, C_2, \ldots$ are positive definite and if the sequence C_n converges pointwise to a function C as $n \to \infty$, then C is positive definite. This follows directly from the definition of positive definite functions (3.1).

Turning band The turning band operator transforms a stationary isotropic correlation function C_d valid in $\mathbb{R}^d$, $d \geq 3$, into

$$C_{d-2}(t) = C_d(t) + (d-2)^{-1} h \frac{\partial}{\partial t} C_d(t),$$

which is valid in $\mathbb{R}^{d-2}$ (Gneiting, 2000). Further, $C_d \in \Phi_d$, if and only if $C_{d-2} \in \Phi_{d-2}$. Recall from Section 2.1.3, that Φ_d is a class of covariance functions of a stationary and isotropic random field on $\mathbb{R}^d$.

3.1.4. Random fields and differentiability

For a covariance functions C of a sufficiently differentiable stationary random field with a spectral density f, it is easy to see that

$$\begin{aligned} \frac{\partial^{|\boldsymbol{k}|}}{\partial x_1^{k_1} \dots \partial x_d^{k_d}} C(\boldsymbol{x}) &= \int_{\mathbb{R}^d} (i\omega_1)^{k_1} \dots (i\omega_d)^{k_d} \exp(i\boldsymbol{\omega}^T \boldsymbol{x}) f(\boldsymbol{\omega})\, d\boldsymbol{\omega} \\ &=: \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T \boldsymbol{x})\, f^{(\boldsymbol{k})}(\boldsymbol{\omega})\, d\boldsymbol{\omega}. \end{aligned} \tag{3.6}$$

Note that from this spectral formulation, it can be seen that the operator of differentiation can be interpreted as a high-pass filter. The frequencies near zero get small weights in comparison to high frequencies.

If Z is moreover isotropic, it follows immediately from (2.15) and (3.6) that

$$\operatorname{var}(Z^{(\boldsymbol{k})}) = \int_{\mathbb{R}^d} \omega_1^{2k_1} \dots \omega_d^{2k_d} f(\boldsymbol{\omega})\, d\boldsymbol{\omega}. \tag{3.7}$$

The last equation will be used several times in this section and it will be applied on some examples of covariance functions. Although the equation (2.15) can be directly used for establishing the variance of $Z^{(\boldsymbol{k})}$, such a calculation is sometimes very tedious. This calculation is typically straightforward using equation (3.7), at least for the given examples.

Differentiability and the property to be a covariance function on a space of higher dimension has an interesting consequence. If $C(\cdot)$ is a covariance function of a stationary and isotropic random field on $\mathbb{R}^d$, which is n times mean square differentiable (it means $C(\cdot)$ is $2n$ times differentiable at origin), then C has $2n + \lfloor (d-1)/2 \rfloor$ continuous derivatives away from the origin. Here, "$\lfloor x \rfloor$" stands for the greatest integer less or equal to x (see Yaglom (1987a) and Gneiting (1999b)).

3.2. Examples of covariance functions

In the following paragraphs, several examples of (classes of) covariance functions will be introduced. The significant properties, especially differentiability will be discussed. Explicite formulas of derivatives of the covariance functions are given in a short form. Some formulas are later used in Chapter 4. The listing of the covariance functions presented here is of course not complete. Here, we restrict ourselves only to some differentiable models. Some other covariance functions can be found e.g. in Yaglom (1987a); Chilès and Delfiner (1999).

3.2.1. Matérn class

Perhaps the most important family of covariance functions is the class introduced by Matérn, whose work appeared in 1960 even though some particular cases of this class were suggested before. Until now, it is known under various names. The class is called *the $\mathcal{K}$-Bessel model* (Chilès and Delfiner, 1999), *the autoregressive model* (Gaspari and Cohn, 1999), *the von Kármán* or *the Whittle-Matérn model* (Gneiting, 1999a). The class was successfully used in many fields including meteorology (Handcock and Wallis, 1994), oceanography, hydrology etc. (see Chilès and Delfiner (1999); Stein (1999) and Gneiting (1999a) for further references). Recently, properties of the Matérn class were investigated by Stein (1999) who also recommended it for use.

Definition The Matérn covariance function $K_\nu(r)$ in the basic form can be defined by the following equation

$$K_\nu(r) = \phi(\alpha r)^\nu \mathcal{K}_\nu(\alpha r), \qquad r \geq 0, \phi > 0, \alpha > 0, \nu > 0, \tag{3.8}$$

where $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of the order ν.

Spectral density The Matérn class can be equivalently introduced by its d-dimensional spectral density

$$f(\omega) = \frac{\phi\ 2^{\nu-1}\Gamma(\nu + \frac{d}{2})\alpha^{2\nu}}{\pi^{d/2}(\alpha^2 + \omega^2)^{\nu+d/2}}, \qquad \omega = ||\boldsymbol{\omega}||. \tag{3.9}$$

Since the function $f(\omega)$ is positive for all $\omega \in \mathbb{R}$, its Fourier transform given by (3.8) (cf. e.g. Yaglom (1987a)) is positive definite. In (3.9), $\Gamma(\cdot)$ denotes the gamma function and $||\cdot||$ is the Eucleidean norm in $\mathbb{R}^d$..

Differentiability From the spectral representation follows that Z is m times mean square differentiable if and only if $\nu > m$, since $\int_{-\infty}^{\infty} \omega^{2m} f(\omega) d\omega < \infty$ if and only if $\nu > m$.

Further important properties The Matérn covariance function owes its popularity to its flexibility, since its behavior near the origin is like $r^{2\nu}$ if ν is not an integer and and like $r^{2\nu}\log(r)$ if ν is an integer (Stein, 1999).

Special cases If $\nu = n + \frac{1}{2}$, $n \in \mathbb{N}$, then $K_\nu(r)$ can be written by means of elementary functions (see Gradstein and Ryshik (1981a) f. 8.468).

$$K_{n+\frac{1}{2}}(r) = \phi\sqrt{\frac{\pi}{2}}e^{-\alpha r}\sum_{k=0}^{n}\frac{(n+k)!}{k!(n-k)!2^k}(\alpha r)^{n-k}.$$

Moreover, if $n = 0$, we have the well known *exponential covariance function*

$$K_{\frac{1}{2}}(r) = Ce^{-\alpha r}, \qquad C > 0,$$

which is often used in geostatistical applications (Chilès and Delfiner, 1999; Cressie, 1991).

Suitable parameterization The form of the Matérn covariance function given in (3.8) seems sometimes not suitable because the function becomes flatter as ν increases. In Handcock and Wallis (1994), the authors recommend an alternative parameterization, which is very useful because of its clear and intuitive interpretation.

For $\boldsymbol{\eta} = (\sigma^2, \nu, \rho)$, $\nu > 0$, $\rho > 0$, they set

$$K_{\boldsymbol{\eta}}(r) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}\left(\frac{2\nu^{1/2}r}{\rho}\right)^\nu \mathcal{K}_\nu\left(\frac{2\nu^{1/2}r}{\rho}\right), \qquad r \geq 0. \tag{3.10}$$

The parameter ν controls differentiability or local behavior of the underlying process Z as before, σ^2 is simply the variance of $Z(\boldsymbol{x})$ and ρ can be interpreted as effective or practical range of an isotropic covariance function considered in geostatistical literature (Chilès and Delfiner, 1999; Cressie, 1991). Another remarkable fact is that the parameterization (3.10) does not depend on the dimension d.

Derivatives of Matérn class covariance functions in $\mathbb{R}$ Derivatives of the Matérn class covariance functions can be calculated directly using formulas for the nth derivatives of product

$$\frac{d^u}{dz^u}\left(z^\nu \mathcal{K}_\nu(z)\right) = \sum_{i=0}^{u} \binom{u}{i} \frac{d^{u-i}}{dz^{u-i}} z^\nu \cdot \frac{d^i}{dz^i} \mathcal{K}_\nu(z),$$

rules for derivatives of composite functions (see Gradstein and Ryshik (1981b) f. 0.42 and 0.43) and applying an iterative formula for differentiation of the Bessel $\mathcal{K}_\nu$ function

$$\frac{d^i}{dz^i}\mathcal{K}_\nu(z) = \frac{(-1)^i}{2^i} \sum_{j=0}^{i} \binom{i}{j} \mathcal{K}_{\nu-i+2j}(z)$$

given here for convenience (see Prudnikov et al. (1990)). Thus, for $\lfloor \nu \rfloor \geq k, \lfloor \nu \rfloor \geq l$, where $k, l \in \mathbb{N}$ and for $|t_2 - t_1| = r > 0$, using formulas above, can be shown that

$$\begin{aligned}
\operatorname{cov}(Z^{(k)}(t_1), Z^{(l)}(t_2)) &= (-1)^k \frac{d^{k+l}}{dr^{k+l}} K_\nu(r) \\
&= (-1)^k \phi\, \alpha^{k+l} \frac{d^{k+l}}{dz^{k+l}} \left(z^\nu \mathcal{K}_\nu(z)\right)[\alpha r] \\
&= (-1)^k \phi\, \alpha^{k+l} \left(\sum_{i=0}^{k+l} \sum_{j=0}^{i} \binom{k+l}{i} \binom{i}{j} \frac{(-1)^i}{2^i} \frac{d^{k+l-i}}{dz^{k+l-i}} z^\nu \cdot \mathcal{K}_{\nu-i+2j}(z) \right)[\alpha r] \\
&= (-1)^k \phi \sum_{i=0}^{k+l} \sum_{j=0}^{i} \binom{k+l}{i} \binom{i}{j} \frac{(-1)^i}{2^i} \alpha^{\nu+i} \frac{\Gamma(\nu+1)}{\Gamma(\nu-k-l+i+1)}\, r^{\nu-k-l+i} \mathcal{K}_{\nu-i+2j}(\alpha r) \qquad (3.11)
\end{aligned}$$

Another formula For some calculations or for the implementation in computer routines, the formula (3.11) does not need be a good one. From the equation

$$\frac{d}{dr}\left[(\alpha r)^\nu \mathcal{K}_{\nu-k}(\alpha r)\right] = -\alpha\left[(\alpha r)^\nu \mathcal{K}_{\nu-k-1}(\alpha r) - k(\alpha r)^{\nu-1} \mathcal{K}_{\nu-k}(\alpha r)\right],$$

which is a trivial consequence of differential formulae for the Bessel functions, (see Gradstein and Ryshik (1981a), formula 8.486.14) and by its recursive application, the following more comprehensive formula can be found

$$\frac{d^u}{dr^u}\left[(\alpha r)^\nu \mathcal{K}_\nu(\alpha r)\right] = \alpha^u \sum_{i=0}^{\lfloor \frac{u}{2} \rfloor} C_i (\alpha r)^{\nu-i} \mathcal{K}_{\nu-u+i}(\alpha r).$$

The coefficient C_i does not depend on ν for all $i \in \{0, \ldots, \lfloor \frac{u}{2} \rfloor\}$ and, in particular, $C_0 = (-1)^u$. Their exact form is given by the following lemma.

Lemma 3.1

$$\frac{d^u}{dr^u}\left[r^\nu \mathcal{K}_\nu(r)\right] = \sum_{i=0}^{\lfloor \frac{u}{2} \rfloor} (-1)^{u+i} \binom{\lfloor \frac{u}{2} \rfloor}{i} \frac{(2\lfloor \frac{u+1}{2} \rfloor - 1)!!}{(2\lfloor \frac{u+1}{2} \rfloor - 1 - 2i)!!} r^{\nu-i} \mathcal{K}_{\nu-u+i}(r) \qquad (3.12)$$

or equivalently

$$\frac{d^u}{dr^u}[r^\nu \mathcal{K}_\nu(r)] = \sum_{i=0}^{\frac{u}{2}} (-1)^{u+i} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-1-2i)!!} r^{\nu-i} \mathcal{K}_{\nu-u+i}(r), \quad \textit{for } u \textit{ even},$$

$$\frac{d^u}{dr^u}[r^\nu \mathcal{K}_\nu(r)] = \sum_{i=0}^{\frac{u-1}{2}} (-1)^{u+i} \binom{\frac{u-1}{2}}{i} \frac{u!!}{(u-2i)!!} r^{\nu-i} \mathcal{K}_{\nu-u+i}(r), \quad \textit{for } u \textit{ odd}.$$

Proof : By induction over u. For the exact implementation see Appendix A.2. □

Thus, for $\lfloor \nu \rfloor \geq k, \lfloor \nu \rfloor \geq l$, where $k, l \in \mathbb{N}$ and for $|t_2 - t_1| = r > 0$, using the formulas above, we have an alternative formula for the computation of covariances of derivatives under the Matérn model

$$\begin{aligned} \operatorname{cov}(Z^{(k)}(t_1), Z^{(l)}(t_2)) &= (-1)^k \frac{d^{k+l}}{dr^{k+l}} K_\nu(r) \\ &= (-1)^l \phi \alpha^{k+l} \sum_{i=0}^{\lfloor \frac{k+l}{2} \rfloor} (-1)^i \binom{\lfloor \frac{k+l}{2} \rfloor}{i} \frac{(2\lfloor \frac{k+l+1}{2} \rfloor - 1)!!}{(2\lfloor \frac{k+l+1}{2} \rfloor - 1 - 2i)!!} (\alpha r)^{\nu-i} \mathcal{K}_{\nu-k-l+i}(\alpha r) \end{aligned}$$

Covariance of derivatives in the origin An important question is, what is the covariance of two derivative random fields, which are observed at the same location. In particular, what is the variance of a derivative random field. To calculate easy values of a derivative of the Matérn covariance function in the origin the equation (3.6) can be used. From the spectral representation of the Matérn class covariance function and using formula 3.251.11 in Gradstein and Ryshik (1981b), we obtain after some algebraic operations (see Appendix A.4, page 90)

$$\begin{aligned} \operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x})) &= (-1)^{|\boldsymbol{k}|} \frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial x_1^{k_1+l_1} \dots \partial x_d^{k_d+l_d}} K_\nu(\boldsymbol{x}) \Big|_{\boldsymbol{x}=\boldsymbol{0}} = \\ &= \begin{cases} (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} \phi 2^{\nu-1} \pi^{-\frac{d}{2}} \alpha^{|\boldsymbol{k}+\boldsymbol{l}|} \Gamma(\nu - \frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}) \prod_{j=1}^d \Gamma(\frac{k_j+l_j}{2} + \frac{1}{2}), & \text{if } \boldsymbol{k}+\boldsymbol{l} \text{ is even}, \\ 0, & \text{if } \boldsymbol{k}+\boldsymbol{l} \text{ is odd}, \end{cases} \end{aligned} \tag{3.13}$$

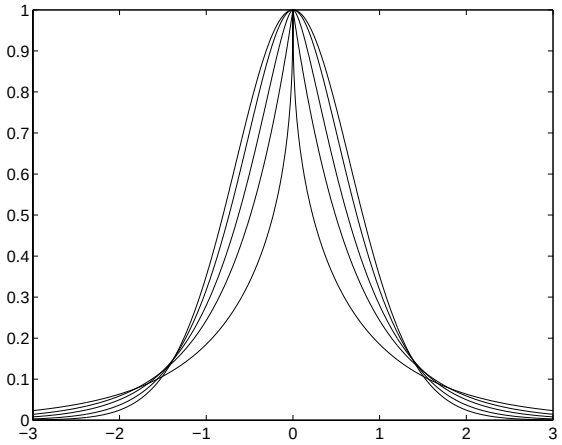
for $\boldsymbol{k} = (k_1, \dots, k_d)$ and $\boldsymbol{l} = (l_1, \dots, l_d)$ multi-indices, such that $\lfloor \nu \rfloor \geq |\boldsymbol{k}|, \lfloor \nu \rfloor \geq |\boldsymbol{l}|$.

Variance of derivative random field in $\mathbb{R}^d$ By setting $\boldsymbol{k} = \boldsymbol{l}$ in (3.13) we get the variance of $Z^{(\boldsymbol{k})}$

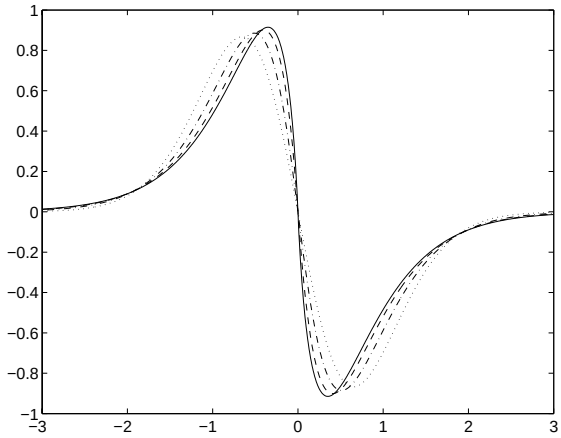
$$\operatorname{var}(Z^{(\boldsymbol{k})}(\boldsymbol{x})) = \phi \, \frac{2^{\nu-1} \alpha^{2|\boldsymbol{k}|}}{\pi^{d/2}} \, \Gamma(\nu - |\boldsymbol{k}|) \, \prod_{i=1}^d \Gamma(k_i + \frac{1}{2}). \tag{3.14}$$

Notice from (3.14) the dependence of $\operatorname{var}(Z^{(\boldsymbol{k})}(\boldsymbol{x}))$ on the scale parameter α and also

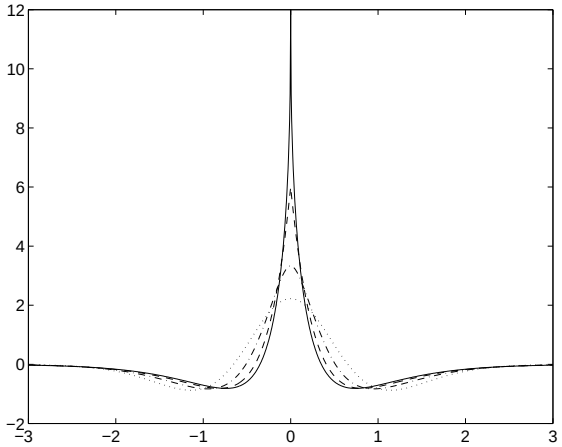
$$\lim_{\nu \to |\boldsymbol{k}|} \operatorname{var}(Z^{(\boldsymbol{k})}(\boldsymbol{x})) = \infty.$$



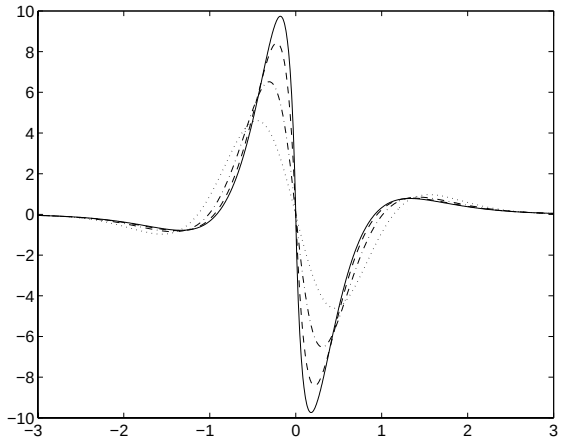
(a) The covariance function of the underlying process. The parameters values ν are 10, 2.5, 1, 0.5 and 0.2 (from top).



(b) The cross-covariance function of the original process and its 1st derivative for $\nu = 1.2$ (solid), $\nu = 1.5$ (dashed), $\nu = 2.5$ (dash-dot) and $\nu = 10$ (dotted line).



(c) The covariance function of the 1st derivative process for $\nu = 1.2$ (solid), $\nu = 1.5$ (dashed), $\nu = 2.5$ (dash-dot) and $\nu = 10$ (dotted line).



(d) The cross-covariance function of the 1st and the 2nd derivative process for $\nu = 2.2$ (solid), $\nu = 2.5$ (dashed), $\nu = 3.5$ (dash-dot) and $\nu = 10$ (dotted line).

Figure 3.1.: Plots of Matérn class covariance functions and its derivatives under the parameterization (3.10). In all cases, $\rho = 1$ and $\sigma^2 = 1$.

Implementation notes Although effective subroutines, which calculate all kinds of the Bessel functions are available and they are a standard part of many mathematical packages, some numerical problems could appear, when computing the Matérn class covariance function near origin. The Bessel $\mathcal{K}-$function diverges in the origin. Thus, for small values of r, the formulas like (3.11) and (3.13) could be numerically very unstable when using this standard subroutines. This could happen more often for higher derivatives. I would recommend to check the behavior of the functions for a particular data set before used (or to define a trash-hold value, under which the distances are set equal zero). Another possibility is to approximate the Matérn model near origin by another stable formula. In particular,

it explains also why the equation (3.13) is given separately from (3.11) and (3.12).

3.2.2. Gaussian covariance function

An infinitely differentiable covariance function is the *Gaussian* covariance function

$$K(h) = \sigma^2 e^{-\alpha^2 ||\boldsymbol{h}||^2}. \tag{3.15}$$

Some authors call the model *squared exponential* (Handcock and Wallis, 1994), maybe to weaken its seemingly importance associated with the personality of C. F. Gauss. Actually, the model plays no central role in geostatistics. The parameter σ^2 is simply the variance of the underlying process, the parameter α is linked with geostatistical (or practical) range. Following Chilès and Delfiner (1999), the value of the range is about $1.73/\sqrt{\alpha}$.

The Gaussian covariance function can be seen as the limiting case of the Matérn class for $\nu \to \infty$. The Gaussian model can be defined in the space $\mathbb{R}^d$ for arbitrary $d \in \mathbb{N}$.

The Gaussian model has been used in some studies in geostatistics (e.g. Röttig (1996); Albrecht (1998)), but today it is already considered as not suitable because of its big regularity. The predictor based on this covariance function typically underestimates the prediction error as shown in Stein (1999) and it is usually more tolerant to a misspecification of covariance function model than other models. Moreover, as the model is infinitely differentiable, it is not suitable for modeling of physical processes (Stein, 1999).

However, since the Gaussian covariance function belongs to the Matérn class, some attention is paid to this model in Chapter 5. The reason is not to recommend the Gaussian model for data analysis, but to seek for analogies to the Matérn class which can be used for a construction of an optimal design.

Spectral density The spectral representation can also derived from the spectral representation of the Matérn class as $\nu \to \infty$. The limit yields (see Stein (1999))

$$f(\omega) = \frac{\sigma^2}{(2\sqrt{\pi}\alpha)^d} \exp\big(-\frac{\omega^2}{4\alpha^2}\big), \qquad \omega = ||\boldsymbol{\omega}||. \tag{3.16}$$

Derivatives of the Gaussian covariance function in $\mathbb{R}^1$

Lemma 3.2 *It holds*

$$\frac{d^n}{dh^n} e^{-\alpha^2 h^2} = (-1)^n \; e^{-\alpha^2 h^2} \alpha^n n! \; \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \frac{1}{(n-2k)!k!} \; (2\alpha h)^{n-2k} \tag{3.17}$$

$$= (-1)^n \; e^{-\alpha^2 h^2} \alpha^n H_n(\alpha h), \tag{3.18}$$

where $H_n(\cdot)$ stands for the Hermite polynomial of the nth order.

Proof : The equation (3.17) can be obtained by induction (cf. formula 0.432.2 in Gradstein and Ryshik (1981b)). The following equation (3.18) can be directly derived from the definition of the Hermite polynomial of the nth order (see e.g. Gradstein and Ryshik (1981a) formula 8.950). □

Derivatives of the Gaussian covariance function in $\mathbb{R}^d$ The generalization of Lemma 3.2 to higher dimensions is trivial, since the Gaussian model is *separable*. A covariance function $C : \mathbb{R}^d \to \mathbb{R}$ is called separable if

$$C(\boldsymbol{h}) = \prod_{i=1}^{d} C_i(h_i), \tag{3.19}$$

where h_i is the ith component of the vector $\boldsymbol{h} \in \mathbb{R}^d$ and $C_i(\cdot)$ is a covariance functions in $\mathbb{R}^1$ for all $i \in \{1, \ldots, d\}$.

From (3.19) it follows that

$$\begin{aligned}\frac{\partial^{|\boldsymbol{n}|}}{\partial h_1^{n_1} \partial h_2^{n_2} \ldots \partial h_d^{n_d}} \exp(-\alpha^2 ||\boldsymbol{h}||^2) &= \prod_{i=1}^{d} \frac{d^{n_i}}{dh_i^{n_i}} \exp(-\alpha^2 h_i^2) \\ &= (-1)^{|\boldsymbol{n}|} \ \exp(-\alpha^2 ||\boldsymbol{h}||^2) \ \alpha^{|\boldsymbol{n}|} \prod_{i=1}^{d} H_{n_i}(\alpha h_i), \end{aligned} \tag{3.20}$$

where $|\boldsymbol{n}| \equiv n_1 + n_2 + \ldots + n_d$ and $||\cdot||$ is the Euclidean norm on $\mathbb{R}^d$.

Variance of derivative random field in $\mathbb{R}^d$ It follows from (3.20) and from the properties of the Hermite polynomials that

$$\operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x})) = (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} \sigma^2 \ \alpha^{|\boldsymbol{k}+\boldsymbol{l}|} \prod_{i=1}^{d} \frac{(k_i + l_i)!}{((k_i + l_i)/2)!} \tag{3.21}$$

and therefore

$$\operatorname{var}(Z^{(\boldsymbol{k})}(\boldsymbol{x})) = \sigma^2 \ \alpha^{2|\boldsymbol{k}|} \prod_{i=1}^{d} \frac{(2k_i)!}{k_i!}. \tag{3.22}$$

3.2.3. J-Bessel class

Definition Previous examples of covariance functions are valid on Euclidean space of any dimension and therefore they must be positive everywhere (cf. inequalities in Yaglom (1987a, p. 356)). Covariance functions, which exist only on spaces of certain dimension, can generally change their sign. This is the case of the following isotropic and differentiable class. The *J-Bessel* class (Chilès and Delfiner, 1999, p. 93) or simply the *Bessel* class on $\mathbb{R}^d$ is defined as follows

$$K_\nu(r) = \phi \ (\alpha r)^{-\nu} J_\nu(\alpha r) \qquad \nu \geq (d-2)/2, \ \phi > 0, \ \alpha > 0, \ r \geq 0. \tag{3.23}$$

Here, J_ν stands for the Bessel function of the first kind of the order ν. This model is very regular and probably only of theoretical interest. The differentiability is shown later in the text. If $\nu = (d-2)/2$, the model is positive definite since it is the Hankel transform (3.4) of the spectral distribution function $\mathcal{G}$ with the unit mass concentrated on one point $u = \alpha$. For an arbitrary $\nu \geq (d-2)/2$, the positive definiteness is shown in Yaglom (1987a, p. 367). Since the spectral density (3.24) of the model is everywhere nonnegative, the function (3.23) is positive definite.

Spectral density The d-dimensional spectral density of the J-Bessel class is

$$f(\omega) = \begin{cases} \frac{\phi(\alpha^2-\omega^2)^{\nu-d/2}}{2^\nu \pi^{d/2}\alpha^{2\nu}\Gamma(\nu+1-(d/2))} & \text{for } \omega = ||\boldsymbol{\omega}|| < \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

Special cases For some particular values of parameter ν, we have covariance functions that are known also under another names.

For $\nu = -\frac{1}{2}$ (i.e. the model is valid in $\mathbb{R}^1$ only), the equation (3.23) reduces to the *cosine* covariance function

$$K_{-\frac{1}{2}}(r) = \phi \cos(\alpha r).$$

For $\nu = \frac{1}{2}$, we obtain the *cardinal-sine* model

$$K_{\frac{1}{2}}(r) = \phi(\alpha r)^{-1} \sin(\alpha r).$$

The cardinal sine model is a valid covariance function in $\mathbb{R}^3$, whence in all spaces of lower dimension, but not in $\mathbb{R}^4$.

Covariance of the derivative random field $Z^{(\boldsymbol{k})}$ Analogously to the Matérn class, the covariance of the two derivative random field, which are observed at the same location can be calculated using spectral representation. This demonstrates that the J-Bessel class is infinitely differentiable simultaneously. Using for instance the formula 2.2.4.8 in Prudnikov et al. (1996), one gets after some algebraic operations

$$\begin{aligned} \operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x})) &= (-1)^{|\boldsymbol{k}|} \frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial x_1^{k_1+l_1} \dots \partial x_d^{k_d+l_d}} K_\nu(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{0}} = \\ &= \begin{cases} (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} \phi\, \pi^{-\frac{d}{2}} \frac{\alpha^{|\boldsymbol{k}+\boldsymbol{l}|}}{2^\nu \Gamma(\nu+\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}+1)} \prod_{i=1}^d \Gamma\big(\frac{k_i+l_i}{2}+\frac{1}{2}\big) & \text{if } \boldsymbol{k}+\boldsymbol{l} \text{ is even} \\ 0 & \text{if } \boldsymbol{k}+\boldsymbol{l} \text{ is odd} \end{cases} \end{aligned} \tag{3.25}$$

for arbitrary multi-indices $\boldsymbol{k}, \boldsymbol{l} \in \mathbb{N}_0^d$ and $\nu \geq -\frac{1}{2}$. Thus, the $\boldsymbol{k}$th derivative random field $Z^{(\boldsymbol{k})}$ always exists for all $\boldsymbol{k}$ and its variance equals

$$\operatorname{var}(Z^{(\boldsymbol{k})}(\boldsymbol{x})) = \phi\, \alpha^{2|\boldsymbol{k}|}\, 2^{-\nu}\, \pi^{-\frac{d}{2}} \frac{1}{\Gamma\big(\nu+|\boldsymbol{k}|+1\big)} \prod_{i=1}^d \Gamma\big(k_i+\frac{1}{2}\big) \tag{3.26}$$

Derivatives of J-Bessel class covariance function in $\mathbb{R}$ Similarly to the Matérn class, we have two equivalent forms for calculation of derivatives of the J-Bessel model. At first, the direct way uses formulas for the nth derivatives of a product

$$\frac{d^u}{dz^u}\big(z^{-\nu} J_\nu(z)\big) = \sum_{i=0}^u \binom{u}{i} \frac{d^{u-i}}{dz^{u-i}} z^{-\nu} \cdot \frac{d^i}{dz^i} J_\nu(z),$$

rules for derivatives of a composite function (see Gradstein and Ryshik (1981b) f. 0.42 and 0.43) and an iterative formula for differentiation of the J_ν-function (see e.g. Prudnikov et al. (1990))

$$\frac{d^i}{dz^i} J_\nu(z) = \frac{1}{2^i} \sum_{j=0}^i (-1)^j \binom{i}{j} J_{\nu-i+2j}(z).$$

Thus, for $|t_2 - t_1| = r > 0$, $u = k + l$, from formulas above follows that

$$\begin{aligned}
\operatorname{cov}(Z^{(k)}(t_1), Z^{(l)}(t_2)) &= (-1)^k \frac{d^{k+l}}{dr^{k+l}} K_\nu(r) \\
&= (-1)^k \phi\, \alpha^{k+l} \, \frac{d^{k+l}}{dz^{k+l}} \big(z^{-\nu} J_\nu(z)\big)[\alpha r] \\
&= (-1)^k \phi\, \alpha^{k+l} \left(\sum_{i=0}^{u} \sum_{j=0}^{i} \binom{k+l}{i} \binom{i}{j} \frac{(-1)^j}{2^i} \frac{d^{k+l-i}}{dz^{k+l-i}} z^{-\nu} \cdot J_{\nu-i+2j}(z) \right) [\alpha r] \\
&= (-1)^k \phi \sum_{i=0}^{k+l} \sum_{j=0}^{i} \binom{k+l}{i} \binom{i}{j} \frac{(-1)^j}{2^i} \alpha^{-(\nu-i)}\, O_i\, r^{-(\nu+k+l-i)}\, J_{\nu-i+2j}(\alpha r) \quad (3.27)
\end{aligned}$$

where

$$O_i = \prod_{m=0}^{k+l-i-1} -(\nu+m), \text{ for } \nu \geq -\frac{1}{2} \quad \text{or} \quad O_i = \frac{\Gamma(-\nu+1)}{\Gamma(-\nu-k-l+i+1)}, \text{ for } \nu \geq -\frac{1}{2},\ \nu \in \mathbb{R} \setminus \mathbb{N}_0.$$

A more elegant formulation for derivatives of the J-Bessel covariance function is given by the following lemma (cf. also Lemma 3.1 with an analogous statement about the Matérn class).

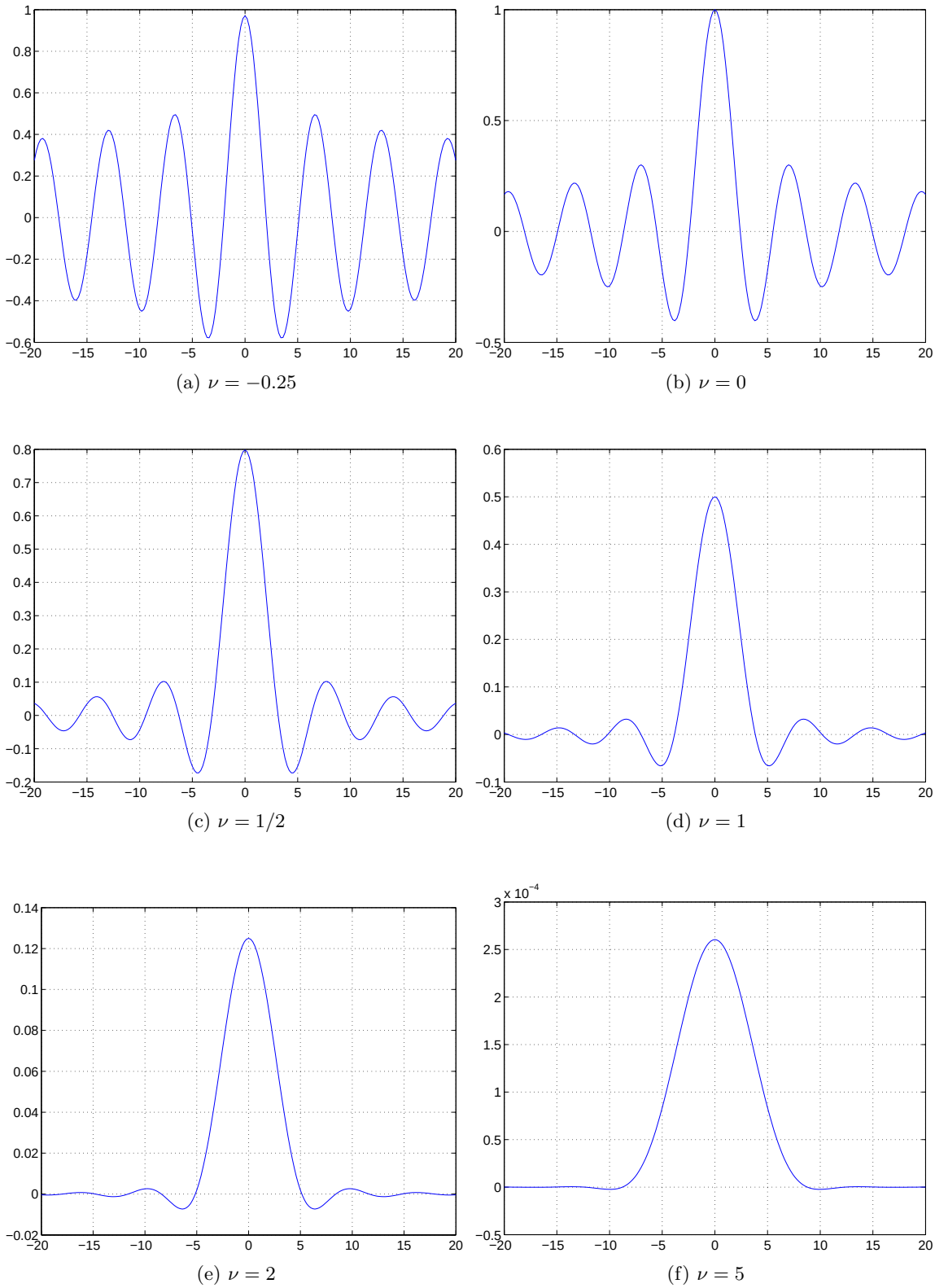
Lemma 3.3

$$\frac{d^u}{dr^u}\big[r^{-\nu} J_\nu(r)\big] = \sum_{i=0}^{\lfloor \frac{u}{2} \rfloor} (-1)^{u+i} \binom{\lfloor \frac{u}{2} \rfloor}{i} \frac{(2\lfloor \frac{u+1}{2} \rfloor - 1)!!}{(2\lfloor \frac{u+1}{2} \rfloor - 1 - 2i)!!} r^{-(\nu+i)} J_{\nu+u-i}(r)$$

or equivalently

$$\begin{aligned}
\frac{d^u}{dr^u}\big[r^{-\nu} J_\nu(r)\big] &= \sum_{i=0}^{\frac{u}{2}} (-1)^{u+i} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-2i-1)!!} r^{-(\nu+i)} J_{\nu+u-i}(r), \quad \textit{for } u \textit{ even} \\
\frac{d^u}{dr^u}\big[r^{-\nu} J_\nu(r)\big] &= \sum_{i=0}^{\frac{u-1}{2}} (-1)^{u+i} \binom{\frac{u-1}{2}}{i} \frac{u!!}{(u-2i)!!} r^{-(\nu+i)} J_{\nu+u-i}(r), \quad \textit{for } u \textit{ odd.}
\end{aligned}$$

Proof : By induction for u even and odd (see Appendix A.3). □



(a) $\nu = -0.25$ (b) $\nu = 0$

(c) $\nu = 1/2$ (d) $\nu = 1$

(e) $\nu = 2$ (f) $\nu = 5$

Figure 3.2.: Plots of the J-Bessel class covariance function (3.23) for $\phi = 1$ and $\alpha = 1$.

3.2.4. Differentiable covariance functions with compact support

For many applications, a compact support can be a reasonable property of a covariance function, for mainly computational reasons. Typically, vanishing correlations beyond a cut-off distance can reduce computational burden of kriging. It also enable the use of efficient sparse matrix techniques. Finally, the setup discussed in the next chapter can be used in a non-limiting sense considering covariance fuctions with compact support.

Exchange based algorithms are used in optimal design problems, where the number of observations is given (Aspie and Barnes, 1990; Gao et al., 1996). These algorithms can be the only way how to explore the behavior of complex (costs related) criteria. Assuming compact support of a covariance function, the kriging predictor changes only locally when one observation is picked up or put in a design. Thus, compact support is highly desirable in this case. Another point is that stochastic simulation of random fields can be proceeded faster.

Recently, big effort was put in a survey of compactly supported covariance functions (see Wendland (1995); Gaspari and Cohn (1999); Gneiting (1999a) and Gneiting (2000) among others). It was shown that these covariance functions have a similar degree of flexibility as for instance the Matérn class.

Transformations on the class of compactly supported covariance functions

The subsequent constructions rely on transformations between classes Φ_d. Recall from Section 2.1.3 (page 9) that Φ_d is the class of covariance functions of a stationary and isotropic random field. The first operator, the so-called *descente*, is defined as

$$I\varphi(t) = \int_t^\infty u\varphi(u)du / \int_0^\infty u\varphi(u)du \qquad t \geq 0. \tag{3.28}$$

The second one, the so-called *montée* is given by

$$D\varphi(t) = \begin{cases} 1, & t = 0 \\ \varphi'(t)/(t\varphi''(0)), & t > 0. \end{cases}$$

Both operators obviously preserve compact support. It was shown by Gneiting (2000) that if φ is an element of Φ_d, $d \geq 3$ and $u\varphi(u) \in L^1([0,\infty))$, then $I\varphi$ is an element of Φ_{d-2}. Further, if $\varphi \in \Phi_d$, $d \geq 3$ and $\varphi''(0)$ exists, then $D\varphi$ is an element of Φ_{d+2}. Under mild conditions, I and D are the inverse operators (see Wendland (1995), where also links to the Fourier transform and to the convolution are established).

Wendland's construction

Now, let us review the construction of covariance functions, which was investigated in Wendland (1995) and introduced in statistical literature by Gneiting (1999a). The starting point was the result that the truncated power function

$$\psi_{\nu,0}(t) = (1-t)_+^\nu = \begin{cases} (1-t)^\nu, & 0 \leq t \leq 1 \\ 0, & t \geq 1 \end{cases}$$

is an element of Φ_d if and only if $\nu \geq \frac{d+1}{2}$ (see Wendland (1995) and references therein). In the next step, he defines

$$\psi_{\nu,k}(t) = I^k\psi_{\nu,0}(t), \qquad k = 0, 1, 2, \ldots$$

by repeated application of (3.28) to the truncated power function. By the claim above, $\psi_{\nu,k}(t)$ belongs to Φ_d if and only if $\nu \geq \frac{d+1}{2} + k$. Furthermore, $\psi_{\nu,k}(t)$ is $2k$-times differentiable at zero, positive and strictly decreasing on its support. In Wendland (1995), the main focus was put on the case when ν is an integer, it means $\psi_{\nu,k}$ is a polynomial on its support. It shown there that this polynomial is of the minimal degree for a given order of differentiability.

Particularly, the first two correlation functions are

$$\psi_{\nu,1}(t) = \Big(1 + (\nu+1)t\Big)(1-t)_+^{\nu+1},$$

which is an element of Φ_d if and only if $\nu \geq \frac{d+3}{2}$ and

$$\psi_{\nu,2}(t) = \Big(1 + (\nu+2)t + \frac{1}{3}((\nu+2)^2 - 1)t^2\Big)(1-t)_+^{\nu+2},$$

which is an element of Φ_d if and only if $\nu \geq \frac{d+5}{2}$. These functions are twice and 4-times differentiable, respectively.

4. A comparison of designs with derivatives

The aim of this chapter is to obtain a notion about the quality of information that observations of derivatives can serve. The main focus lies on demonstrating the influence of the choice of a covariance function on the efficiency of designs when using derivatives, rather than on finding an optimal solution of a design problem. A theorem and some consequences are stated in a d-dimensional setting and discussed. The given results are closely related to the imse-criterion for the simple kriging predictor (2.30).

Before we begin with the survey related to the imse, let us mention the following motivation example. The example is just a very simple evidence for the fact that observations of derivatives are sometimes of more interest than direct observations of the field to predict. Some other examples can be found in Näther and Šimák (2001).

Example 4.1 Let a zero mean process $Z(x)$ has the Gaussian covariance function (3.15). Consider the simple kriging predictor (2.30) of $Z(x_0)$ based on only one observation of Z at x, where x has to be chosen from the interval $X = (-\infty, T]$, $x_0 > T$. Let us denote this predictor as $p_L^*(Z(x_0); Z(x))$. Using the notation in (2.30), $\operatorname{mse} p_L^*(Z(x_0); Z(x))$ becomes minimal if and only if

$$\boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} = (C_Z(x, x_0))^2 = \sigma^2 \, \exp(-2\alpha^2 (x_0 - x)^2)$$

is maximized, i.e. if and only if $x = T$. Hence, $V_1 = \{T\}$ is the best one-point design for prediction of $Z(x_0)$ regarding to the mean square error.

Now, for comparison, let us consider the predictor (2.30) of $Z(x_0)$ based on one observation of $Z^{(1)}$ at $x' \in X$. Let us denote this predictor as $p_L^*(Z(x_0); Z^{(1)}(x'))$. Now, $\operatorname{mse} p_L^*(Z(x_0); Z^{(1)}(x'))$ is minimized if and only if

$$\boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} = \frac{(C_{Z^{(1)},Z}(x', x_0))^2}{\operatorname{var}(Z^{(1)}(x'))} = \sigma^2 \, 2\alpha^2 (x_0 - x')^2 \exp(-2\alpha^2 (x_0 - x')^2) \tag{4.1}$$

is maximized, which is satisfied for $x_0 - x' = 1/(\alpha\sqrt{2})$. Hence, the best location for observing the first derivative of $Z(x)$ is at $x' = x_0 - 1/(\alpha\sqrt{2})$.

Compare both predictors at the same observation point $x = x'$ by the following ratio

$$\frac{\operatorname{mse} p_L^*(Z(x_0); Z(x))}{\operatorname{mse} p_L^*(Z(x_0); Z^{(1)}(x))} = \frac{1 - \exp(-2\alpha^2 (x_0 - x)^2)}{1 - 2\alpha^2 (x_0 - x)^2 \exp(-2\alpha^2 (x_0 - x)^2)} = \frac{1 - e^{-t}}{1 - t e^{-t}} =: r(t) \tag{4.2}$$

where $t := 2\alpha^2 (x_0 - x)^2$. For $t > 1$, we have $r(t) > 1$, i.e. observing the derivative leads to a smaller mean square error than observing the original process. Assume long range prediction in the sense that $1/(\alpha\sqrt{2}) < x_0 - T$. Then $t^* := 2\alpha^2 (x_0 - T)^2 > 1$ and $r(t^*) > 1$. Thus, optimal observing of Z at $x = T$ leads to larger mean square error than observing the first derivative $Z^{(1)}$ at the same point.

Note that in the case of short range prediction, i.e. $1/(\sqrt{2}\alpha) > x_0 - T$, we have $t^* < 1$ and observing Z at $x = T$ is always better than observing $Z^{(1)}$. Note that if $T = x_0$ and x' is chosen such that $x' = T$, then (4.1) is equal to zero. It means that (unlike to the observation of Z) the observation of $Z^{(1)}$ is totally uninformative at the prediction point.

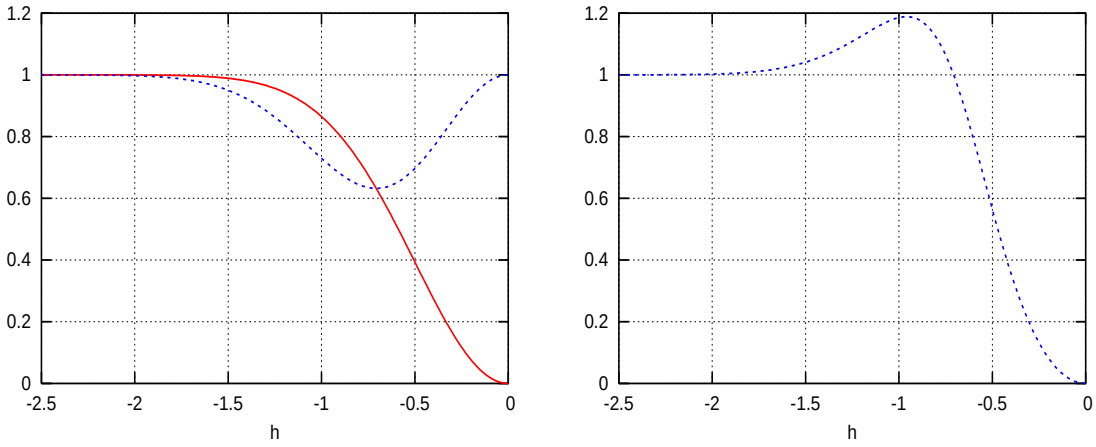


Figure 4.1.: On the left: Plots of $\operatorname{mse} p_L^*(Z(x_0); Z(x))$ (solid) and $\operatorname{mse} p_L^*(Z(x_0); Z^{(1)}(x'))$ (dashed line). On the right: Plot of (4.2) for $\sigma^2 = 1$, $\alpha = 1$ and $h = x - x_0$.

4.1. Imse-update and update ratio

Imse-update

Let us recall the definition of the integrated mean square error for the simple kriging predictor. This means, we automatically assume the second state of knowledge (see page 15). Using the notation (2.31) the imse-criterion is given as

$$\operatorname{imse}_W = \int_W \left(C(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} \right) \boldsymbol{d}\boldsymbol{x}_0, \quad W \subset \mathbb{R}^d. \tag{4.3}$$

The aim of optimal planning is to find a design V_n^*, so that V_n^* minimizes (4.3). It is obvious that minimizing (4.3) is equivalent to maximizing the *imse-update for prediction area* W, i.e.

$$U_W(V_n) = \int_W \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} \, \boldsymbol{d}\boldsymbol{x}_0. \tag{4.4}$$

Let us assume that the prediction area W is very large with respect to the range of covariance function and we approximate (4.4) as integral over the whole $\mathbb{R}^d$

$$U(V_n) = \int_{\mathbb{R}^d} \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} \, \boldsymbol{d}\boldsymbol{x}_0, \tag{4.5}$$

if the integral exists. An intuitive justification for the approximation is as follows. If the prediction region W is large enough, the given observations will be uncorrelated (as in case of covariance function with compact support) or effectively uncorrelated with the points on the border of W and then the approximation can be accepted. The approximation $U(V_n)$ will be simply referred to as *(total) imse-update*. Note that the imse-update is just the variance of the simple kriging predictor integrated over $\mathbb{R}^d$.

Update ratio

Let us consider two experimental designs (2.17) of the same size, i.e. with the same number of design points. The question is, which of these designs should be preferred and, particularly, what is the efficiency or performance of the better one comparing to the worse one.

Consider for instance the situation that an observer has several different measurement instruments which can be used at a particular site, but an application of one of them excludes the further use of the other measurements at the site. Suppose that the locations for the future measurement are already pre-defined. The observer faces the problem, which of the particular measurements should be preferred, to obtain the maximal information from the experiment.

As a criterion of efficiency we can take the *update ratio*, i.e. the fraction

$$R = \frac{U(V_n^1)}{U(V_n^2)},$$

where $U(V_n^1)$ and $U(V_n^2)$ are imse-updates for some designs V_n^1 and V_n^2, respectively. It is clear that if $R > 1$ (resp. $R < 1$) the design V_n^1 (resp. V_n^2) should be preferred. If $R = 1$ the designs are equivalent with respect to the imse-update.

It is also clear that using the update ratio does not lead to an optimal design in the usual sense. The update ratio is just a natural criterion for a comparison of designs. They can, however, be non-optimal. Applying the update ratio to some given set of designs, we get an answer to the question which of these designs performs best. Nevertheless, we do not get an answer to the question whether there exists a design outside from this set, which is better than the designs we have restricted to.

In the following, we compare regular designs (see (2.18)) based on observations of derivative random fields of diverse order. For purposes which will later be clear, let us refer to the correlations between observations, which were realized at different locations, as *inter-site* correlations. We will consider sequences of designs for which absolute values of correlations between observations become small. These designs will be referred to as *designs with vanishing inter-site correlations.*

Note that the strategy, i.e. to place observations so that they are correlated as little as possible, can be rather reasonable. Let us assume a positive covariance function as for instance a covariance function from the Matérn class. Let us suppose designs that consists of observations of the non-derivative random field only. Minimizing correlations between observations leads to the optimal design with respect to imse-update. However, differentiation of random fields introduces negative values of correlation and, therefore, for designs based on derivative information, this strategy may not be the best one. Nevertheless, for the first impression, it is quite instructive to consider designs with vanishing inter-site correlations.

Maximin design

Assuming positive covariance functions only, minimizing correlations between observations means to *max*imize the (*min*imal) distance between observations in a design. This procedure is also known as *maximin design.* Asymptotic properties of maximin designs are studied e.g. in Johnson et al. (1990) and Mitchell et al. (1994) where the authors also assume weak correlated designs. In the later paper, even observations of derivatives up to some order were considered for so-called Hermitian data. This approach was recently generalized

by Holtmann (2001) for non-Hermitian data. Note that in the cited literature a finite observation area is presumed. The bounds of the prediction area influences the form of the optimal design (cf. also Abt (1992) for non-asymptotic designs). Maximin distance designs, even for small sample problems, may be very difficult to construct, especially if the design region is irregular. It is known, that several structurally different solutions (see e.g. Müller (2001, page 75)) may exist.

4.2. Design with vanishing inter-site correlations

Let Z be a second order stationary isotropic random field on $\mathbb{R}^d$ and let $\boldsymbol{k} = (k_1, \ldots, k_d)$ and $\boldsymbol{l} = (l_1, \ldots, l_d)$ be some multi-indices. Now, let us consider fields $Z^{(\boldsymbol{k})}$ and $Z^{(\boldsymbol{l})}$ where $Z^{(\boldsymbol{k})}$ and $Z^{(\boldsymbol{l})}$ are $\boldsymbol{k}$th and $\boldsymbol{l}$th partial mean square derivative random fields of the underlying random field Z, which is supposed to be differentiable sufficiently.

Let C be the covariance function of Z. We will compare n-points discrete designs $V^{(\boldsymbol{k})}$ and $V^{(\boldsymbol{l})}$ where the former is based on n observations of $Z^{(\boldsymbol{k})}$ and the later consist of n observations of $Z^{(\boldsymbol{l})}$. The subindex n was omitted for convenience since all what follows is related to n-points designs.

Let us consider a sequence of designs $\{V_m^{(\boldsymbol{k})}\}_{m=1}^{\infty}$ and their corresponding covariance matrices $\{\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}\}_{m=1}^{\infty}$, $\boldsymbol{\Sigma}_m^{(\boldsymbol{k})} \in PD_n$. For a particular design $V_m^{(\boldsymbol{k})} = \{(\boldsymbol{x}_1^m, \boldsymbol{k}), \ldots, (\boldsymbol{x}_n^m, \boldsymbol{k})\}$ the elements of the matrix $\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}$ are defined as

$$(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})})_{ij} = \mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}_i^m), Z^{(\boldsymbol{k})}(\boldsymbol{x}_j^m)), \qquad i, j \in \{1, \ldots, n\}. \tag{4.6}$$

Since we assume that the matrices $\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}$ are all positive definite, they are regular for all m. Let us denote $v_{\boldsymbol{k}} = \mathrm{var}(Z^{(\boldsymbol{k})})$ and assume vanishing inter-site correlations for $m \to \infty$. It means that $\boldsymbol{\Sigma}_m^{(\boldsymbol{k})} \to v_{\boldsymbol{k}} \boldsymbol{I}_n$ as $m \to \infty$, where $\boldsymbol{I}_n$ denotes the identity matrix and the convergence is defined element-wise.

It is clear that $\left(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}\right)^{-1} \to 1/v_{\boldsymbol{k}} \boldsymbol{I}_n$ as $m \to \infty$. Since the eigenvalues of a matrix are continuous functions of the matrix elements (cf. e.g. Schott (1997, theorem 3.12)), particularly, the sequence $\lambda_{\min}(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})})$ of the smallest eigenvalues of $\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}$ converges to $v_{\boldsymbol{k}}$.

From this follows that $\lambda_{\min}(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})})$ is bounded by a positive constant K whenever m is sufficient large:

$$\boldsymbol{c}^T \left(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}\right)^{-1} \boldsymbol{c} \leq \lambda_{\min}(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}) \boldsymbol{c}^T \boldsymbol{c} < K \boldsymbol{c}^T \boldsymbol{c}. \tag{4.7}$$

Let us suppose that $C \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ as well as all derivatives of C. In this case, the function on the right side of (4.7) is in $L^1(\mathbb{R}^d)$. Then using the dominated convergence theorem (see e.g. Theorem 8.19 in Kuttler (1998)), Plancherel theorem (Theorem 3) and (3.6), we obtain the following equality:

$$\begin{aligned} L_{\boldsymbol{k}} := \lim_{\boldsymbol{\Sigma}_m^{(\boldsymbol{k})} \to v_{\boldsymbol{k}} \boldsymbol{I}_n} U(V_m^{(\boldsymbol{k})}) &= \int_{\mathbb{R}^d} \lim_{\boldsymbol{\Sigma}_m^{(\boldsymbol{k})} \to v_{\boldsymbol{k}} \boldsymbol{I}_n} \boldsymbol{c}^T \left(\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}\right)^{-1} \boldsymbol{c} \, d\boldsymbol{x}_0 && (4.8) \\ &= \frac{n}{v_{\boldsymbol{k}}} ||C_{Z^{(\boldsymbol{k})} Z^{(0)}}||_2^2 && (4.9) \\ &= (2\pi)^d \frac{n}{v_{\boldsymbol{k}}} ||f^{(\boldsymbol{k})}||_2^2. && (4.10) \end{aligned}$$

The previous schema is a general basis of the proof of the following theorem.

Theorem 5 *Let Z be a second order stationary isotropic random field on $\mathbb{R}^d$ with a covariance function C. Let us denote $\boldsymbol{k} = (k_1, \ldots, k_d)$ and $\boldsymbol{l} = (l_1, \ldots, l_d)$ multi-indices, such that $|\boldsymbol{k}| \leq |\boldsymbol{l}|$. Let $\{V_m^{(\boldsymbol{k})}\}$ and $\{V_m^{(\boldsymbol{l})}\}$ be sequences of n-points designs based on observations of $Z^{(\boldsymbol{k})}$ and $Z^{(\boldsymbol{l})}$, respectively. Let $\{\boldsymbol{\Sigma}_m^{(\boldsymbol{k})}\}$ and $\{\boldsymbol{\Sigma}_m^{(\boldsymbol{l})}\}$ be sequences of covariance matrices, defined by (4.6).*

Then

$$\frac{L_{\boldsymbol{k}}}{L_{\boldsymbol{l}}} = 2^{|\boldsymbol{l}|-|\boldsymbol{k}|} \tag{4.11}$$

$$\frac{L_{\boldsymbol{k}}}{L_{\boldsymbol{l}}} = \prod_{i=|\boldsymbol{k}|+1}^{|\boldsymbol{l}|} \left(2 + \frac{2i+d}{2(\nu-i)}\right) \quad \text{for } \nu > |\boldsymbol{l}| \geq |\boldsymbol{k}| \tag{4.12}$$

$$\frac{L_{\boldsymbol{k}}}{L_{\boldsymbol{l}}} = \prod_{i=|\boldsymbol{k}|+1}^{|\boldsymbol{l}|} \left(2 - \frac{2i+d}{2(\nu+i)}\right) \quad \text{for } |\boldsymbol{l}| \geq |\boldsymbol{k}| \text{ and } \nu > \frac{d-1}{2} \tag{4.13}$$

if the random field $Z(\boldsymbol{x})$ has the Gaussian, Matérn or J-Bessel covariance function, respectively.

Proof : We calculate directly the limit $L_{\boldsymbol{k}}$ using (4.10). This spectral form significantly simplifies calculation. In the Gaussian case, using formula 3.461.2 in Gradstein and Ryshik (1981b) and (3.22) it can be shown by direct calculation (see Appendix A, page 93), that

$$L_{\boldsymbol{k}} = n\sigma^2 \left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d 2^{-|\boldsymbol{k}|}. \tag{4.14}$$

In the Matérn class case, the formula 3.251.11 in Gradstein and Ryshik (1981b) and (3.14) gives

$$L_{\boldsymbol{k}} = n\ \phi\ 2^{\nu-1}\ \left(\frac{2\sqrt{\pi}}{\alpha}\right)^d\ B\Big(\nu + \frac{d}{2}, \nu + \frac{d}{2}\Big)\ \frac{\Gamma(2\nu - |\boldsymbol{k}| + \frac{d}{2})}{\Gamma(\nu - |\boldsymbol{k}|)}. \tag{4.15}$$

See Appendix A (page 93) for the detailed calculation. Finally, for the J-Bessel class (where condition $\nu > \frac{d-1}{2}$ assures $C^{(\boldsymbol{k})} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$) using formula 3.251.1 in Gradstein and Ryshik (1981b) and (3.26) we get

$$L_{\boldsymbol{k}} = n\ \phi\ 2^{-\nu} \left(\frac{2\sqrt{\pi}}{\alpha}\right)^d \frac{\Gamma(2\nu+1-d)}{\Gamma^2(\nu+1-\frac{d}{2})} \frac{\Gamma(\nu+|\boldsymbol{k}|+1)}{\Gamma(2\nu+1+|\boldsymbol{k}|-\frac{d}{2})}. \tag{4.16}$$

The detailed calculation is given in Appendix A on the page 94. The claim in Theorem 5 immediately follows from the integrals above. □

Note that the condition $\nu > |\boldsymbol{l}| \geq |\boldsymbol{k}|$ in (4.12) assures that the random field Z is sufficiently differentiable. On the other hand, the condition $\nu > \frac{d-1}{2}$ in (4.13) assures that the appropriate cross-covariance function is in $L^2(\mathbb{R}^d)$ and in $L^1(\mathbb{R}^d)$.

4.2.1. Gaussian covariance function vs. Matérn class

In his monograph, Stein (1999) criticizes the use of the Gaussian covariance function. He demonstrates that using this covariance function typically leads to an underestimation

of the prediction error comparing with the Matérn class, in the case that the covariance function model is misspecified. In this context, the Gaussian covariance function is said to be optimistic.

Here, we have shown another aspect of the problem. Since

$$2^{|\boldsymbol{l}|-|\boldsymbol{k}|} < \prod_{i=|\boldsymbol{k}|+1}^{|\boldsymbol{l}|} \left(2 + \frac{2i+d}{2(\nu-i)}\right)$$

for finite $\nu > |\boldsymbol{l}| \geq |\boldsymbol{k}|$, it can be seen that the Gaussian covariance function remains optimistic regarding to the Matérn class also if derivative random fields are considered. More precisely, observing high derivatives comparing to lower ones of a field with the Gaussian covariance function gives us more information than analogous observations of a random field driven by a covariance function from the Matérn class. The problem is that in practice the covariance function model is rarely known with great certainity. Therefore, assuming the Gaussian covariance function to be the proper model (although the Matérn class is the right one) leads to an over-weighting of the seeming importance of high derivative observations. On the other hand, geostatisticians decide in some stage of statistical data analysis for one parameterizable model and they proceed the so-called model fitting, i.e. model parameter estimation. Theorem 5 points out the potential danger, which is hidden in the prior choice of the Gaussian covariance function.

Note that the Gaussian model is the limiting case of the Matérn model as $\nu \to \infty$ and also the fraction (4.12) converges to (4.11). In this light, the result of Theorem 5 is not surprising. The Gaussian covariance function is the smoothest covariance function in the Matérn class and therefore observation of derivative fields driven by the Gaussian covariance function must be much more informative than observations of derivative fields under the general Matérn class.

4.2.2. The critical role of the parameter ν in the Matérn model for an estimation of the efficiency

Moreover, it can readily be seen from (4.15) that the parameter ν plays a critical role for estimating the efficiency of designs, that use derivatives. Say, for instance, we want to suggest a sufficiently large size $n_{\boldsymbol{k}}$ of a design $V_{n_{\boldsymbol{k}}}^{(\boldsymbol{k})}$ based on observations of a $\boldsymbol{k}$th derivative random field such that imse-update $U(V_{n_{\boldsymbol{k}}}^{(\boldsymbol{k})})$ is at least the same as the imse-update $U(V_{n_{\boldsymbol{0}}}^{(\boldsymbol{0})})$ of a design $V_{v_{\boldsymbol{0}}}^{(\boldsymbol{0})}$ of a given size $n_{\boldsymbol{0}}$ based on field values. But as soon as ν gets close to $|\boldsymbol{k}|$, $n_{\boldsymbol{k}}$ rapidly increases. It means $Z^{(\boldsymbol{k})}$ is close to the uninformative white noise.

This shows a possible danger even in the decision to use derivative observations. For instance, *it does not suffice to know that the observed field is k-times differentiable, if we plan to observe a* $\boldsymbol{k}$*th derivative field,* $k = |\boldsymbol{k}|$. If the field is driven by the Matérn class (and this can be considered as a good assumption as discussed in Stein (1999)) and ν is very close to $|\boldsymbol{k}|$, observations of the $\boldsymbol{k}$th derivative field are nearly uninformative. Clearly, $\Gamma(\nu - |\boldsymbol{k}|) \to \infty$ as $\nu \to |\boldsymbol{k}|$ in (4.15). Therefore, some lower bound for the smoothness parameter ν have to be required for estimation of the efficiency of a particular design. In the previous case, it suffices for instance that the observed field is $(|\boldsymbol{k}|+1)$-times differentiable.

4.2.3. Discussion of the J-Bessel class

It can be immediately seen from Theorem 5 that the following statement holds.

Consequence 4.1 *Let $Z(\boldsymbol{x})$ be a random field on $\mathbb{R}^d$. If $Z(\boldsymbol{x})$ has the J-Bessel covariance function with parameter $\nu = \frac{d}{2}$ then*

$$\frac{L_{\boldsymbol{k}}}{L_{\boldsymbol{l}}} = 1 \quad \text{for arbitrary } \boldsymbol{k}, \boldsymbol{l} \in \mathbb{N}_0^d. \tag{4.17}$$

Moreover, if $Z(\boldsymbol{x})$ has the J-Bessel covariance function and $\frac{d-1}{2} < \nu < \frac{d}{2}$

$$\frac{L_{\boldsymbol{k}}}{L_{\boldsymbol{l}_1}} < \frac{L_{\boldsymbol{k}}}{L_{\boldsymbol{l}_2}} < 1 \quad \text{for } |\boldsymbol{l}_1| > |\boldsymbol{l}_2| \geq |\boldsymbol{k}|.$$

The result for the J-Bessel class seems a little bit surprising. Since

$$\prod_{i=|\boldsymbol{k}|+1}^{|\boldsymbol{l}|} \left(2 - \frac{2i+d}{2(\nu+i)}\right) < 2^{|\boldsymbol{l}|-|\boldsymbol{k}|}$$

for finite $|\boldsymbol{l}| \geq |\boldsymbol{k}|$ and $\nu > 0$, it can be seen that observations of high derivatives of a random field with the J-Bessel class covariance function (for any ν) comparing to lower ones, are even better than an analogous situation with the Gaussian covariance function. Further, in any dimension, there exists a covariance function for which observations of derivatives of any finite order are equivalent in the imse-update sense (see (4.17)). Moreover, if $\frac{d-1}{2} < \nu < \frac{d}{2}$ then $L_{\boldsymbol{k}}/L_{\boldsymbol{l}} < 1$. In this case observation of a derivative random field of any order gives us a bigger imse-update than direct observations of the underlying random field.

Similarly to Section 4.2.1, it can be claimed that using the J-Bessel class leads likely to an over-estimating of the imse-update for higher derivative observations. Thus, if it is definitely not sure that the study random field is driven by the J-Bessel class, (and we can never be sure of it making inference from a data set) due to its extreme regularity, an application of this class is really questionable.

What is the significant property, which makes the J-Bessel class so different from the Matérn class or from the Gaussian covariance function? It can be conjectured that the key attribute is the compact support of the spectral density (3.24) of the J-Bessel class covariance functions[1]. Recall (3.6) which suggests to interpret the operator of differentiation as a high pass filter. Clearly, if the spectral density of a random field is positive everywhere, the high (less informative) frequencies obtain bigger weights after differentiation. However, for the interpolation, lower frequencies are of the main interest. On the other hand, a compact support of the spectral density helps to keep the over-weighting of high frequencies under control. Similarly, differentiable random fields with a narrow spectral density (although not restricted to a finite domain) could possess similar properties.

Construction problems

A problem is, however, the construction of a design whose covariance matrix is (close to) a diagonal matrix. In the Gaussian case as well as for the Matérn class, the covariance function vanishes quite fast. Thus, we can easily construct a design for that the actual (non-limiting) fraction approximates the limiting fraction in Theorem 5 rather well. Simultaneously, distances between observations do not increase much (cf. Chapter 5). Roughly speaking, the distances could be a little bit bigger than the practical range of the correlation function.

[1] I am grateful to Prof. Joachim Menz for this idea.

The J-Bessel class covariance functions as well as their derivatives are alternating functions, but the runs of these functions as the whole tend to zero very slowly (see Figure 3.2) at least for small values of ν. The zeros of the functions are generally not a periodic sets. It means that we can easily find a $d+1$-points designs (in $\mathbb{R}^d$) with uncorrelated observations as follows. The design points are vertexes of a regular polyhedron, i.e. join of any two vertexes is equal to the edges in the polyhedron. For instance, regular triangle or regular tetrahedron are examples for that in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. The length of the edges have to be equal to the distance between the origin and one of the zero points. Then, the designs points will obviously be uncorrelated.

However, when constructing a general n-points design ($n > d+1$), we have to set very large distances between possibly $(d+1)$-observations-blocks to obtain "nearly" uncorrelated observations. For larger values of ν, the correlations under the J-Bessel model remain close to zero provided that the distance from the origin is large enough.

4.2.4. Notes to Theorem 5

Let us give some comments on the meaning and on the applications of Theorem 5. First, let us note that Theorem 5 gives no answer to the question whether the plans with vanishing inter-site correlations are optimal or even good ones. The main focus lies rather on the demonstration of possible effects, which are caused by the prior choice of a covariance function, than on the efficiency of the plans. Theorem 5 can be seen as a discussion of different covariance function models and it can be useful in the stage of modeling. If we concentrate on the informations derivatives can serve, the Matérn class is here seen as the most pessimistic one. It is in some sense the counterpart to the J-Bessel class, which is the most optimistic one.

Another point could be that by means of the results of Theorem 5 a sequential planning procedure could be accepted as reasonable approach. For instance, if observations of $Z^{(\boldsymbol{k})}$ are more informative than $Z^{(\boldsymbol{l})}$, we should concentrate primarily on finding an optimal design for the observations of $Z^{(\boldsymbol{k})}$. Once this smaller problem is solved, we can start to think about how to improve the existing design with additional observations of less informative $Z^{(\boldsymbol{l})}$. As a result, we could reduce the complexity of a particular design problem this way.

4.2.5. Sensitivity of the imse-update on covariance function parameters

We can also ask, how the imse-update behaves when the parameters of a particular covariance function are misspecified. Misspecification of covariance model parameters and its consequences was one of the central themes in Stein (1999). Here, we present just a simple observation based on the formulae given in the proof of Theorem 5.

First, let us look at the scale parameter α. Given α, we denote the limiting imse-update by $L_{\boldsymbol{k}}(\alpha)$. For all mentioned covariance function models, it holds

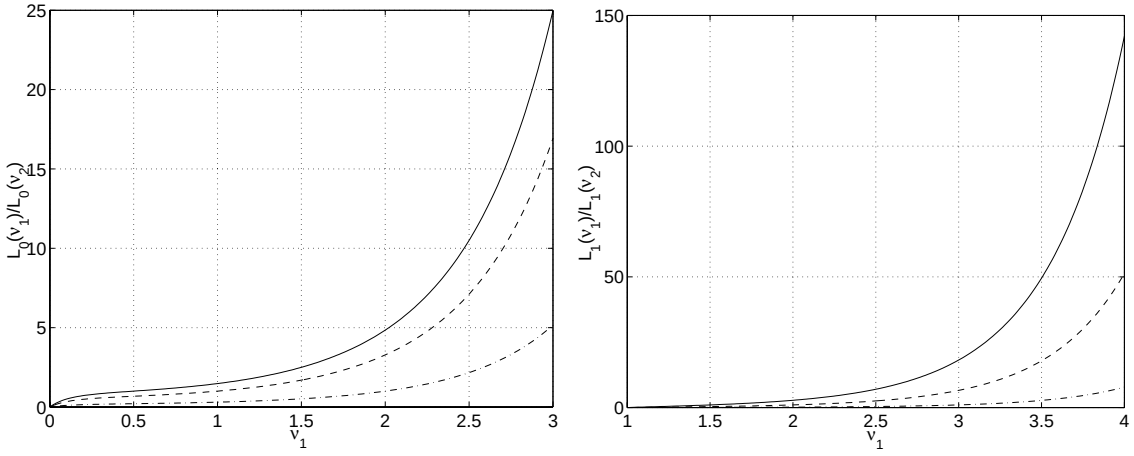
$$\frac{L_{\boldsymbol{k}}(\alpha_1)}{L_{\boldsymbol{k}}(\alpha_2)} = \left(\frac{\alpha_2}{\alpha_1}\right)^d$$

for any given $\boldsymbol{k}$.

For the smoothness parameter ν in the Matérn model we have analogously,

$$\frac{L_{\boldsymbol{k}}(\nu_1)}{L_{\boldsymbol{k}}(\nu_2)} = 2^{\nu_1-\nu_2}\frac{\Gamma(2\nu_1 - |\boldsymbol{k}| + \frac{d}{2})}{\Gamma(2\nu_2 - |\boldsymbol{k}| + \frac{d}{2})}\frac{\Gamma(\nu_2 - |\boldsymbol{k}|)}{\Gamma(\nu_1 - |\boldsymbol{k}|)}\frac{B(\nu_1 + \frac{d}{2}, \nu_1 + \frac{d}{2})}{B(\nu_2 + \frac{d}{2}, \nu_2 + \frac{d}{2})}. \tag{4.18}$$

The ratio (4.18) increases very fast whenever $\nu_2 \ll \nu_1$. The behavior of the fraction (4.18) is reported on Figure 4.2.



(a) $d = 1$, $k = 0$ and $\nu_2 = 0.5$ (full line), $\nu_2 = 1$ (dashed line), $\nu_2 = 2$ (dash-dot line)

(b) $d = 1$, $k = 1$ and $\nu_2 = 1.5$ (full line), $\nu_2 = 2$ (dashed line), $\nu_2 = 3$ (dash-dot line).

Figure 4.2.: Plots of the sensitivity of the limiting imse-update for the Matérn covariance function (equation (4.18))

It can be seen from the comparison of Figures 4.2(a) and 4.2(b) that the precise knowledge about the smoothness parameter ν is more important when using derivatives of higher order.

4.2.6. Generalized version of Theorem 5

When comparing designs, it is clear that there is no need to consider observations of only one type and that the size of the designs to compare can be different. For instance, it is possible to generalize Theorem 5 by assuming that designs consist of observations of diverse derivative fields. The proof can be accomplish in the same way. Let us suppose that a stationary and isotropic random field Z on $\mathbb{R}^d$ is sufficiently differentiable. Let us denote $\{V_m\}_{m=1}^{\infty}$ a sequence of designs in the form

$$V_m(n) = \left\{(\boldsymbol{x}_1^m, \boldsymbol{k}_1), (\boldsymbol{x}_2^m, \boldsymbol{k}_2) \ldots, (\boldsymbol{x}_n^m, \boldsymbol{k}_n)\right\}, \tag{4.19}$$

where $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n$ are d-dimensional multi-indices. Note that by the notation $V_m(n)$, the fact is emphasized that designs to compare using imse-update, can consist of different number of observations. Suppose that the sequence $\{V_m\}_{m=1}^{\infty}$ of designs fulfills the conditions of the proof of Theorem 5, i.e. among others, that correlations between observations vanish as $m \to \infty$. Let us define $\overline{n} = \max\{|\boldsymbol{k}_1|, \ldots, |\boldsymbol{k}_n|\}$ and denote by $n_{|k|}$ the number of multi-indices $\boldsymbol{k}$ in (4.19) such that $k = |\boldsymbol{k}|$. The corresponding formula for (4.14), (4.15) and

(4.16) are

$$\begin{aligned} L(n) &= \sigma^2 \left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \sum_{k=0}^{\overline{n}} n_{|k|} 2^{-k}, \\ L(n) &= \phi\ 2^{\nu-1} \left(\frac{2\sqrt{\pi}}{\alpha}\right)^d B\Big(\nu+\frac{d}{2}, \nu+\frac{d}{2}\Big) \sum_{k=0}^{\overline{n}} n_{|k|} \frac{\Gamma(2\nu-k+\frac{d}{2})}{\Gamma(\nu-k)}, \\ L(n) &= \phi\ 2^{-\nu} \left(\frac{2\sqrt{\pi}}{\alpha}\right)^d \frac{\Gamma(2\nu+1-d)}{\Gamma^2(\nu+1-\frac{d}{2})} \sum_{k=0}^{\overline{n}} n_{|k|} \frac{\Gamma(\nu+k+1)}{\Gamma(2\nu+1+k-\frac{d}{2})}, \end{aligned}$$

respectively. Subsequently, we get the generalized version of formulae (4.11), (4.12) and (4.13)

$$\frac{L(n_1)}{L(n_2)} = \sum_{k=0}^{\overline{n_1}} n_{1|k|} 2^{-k} \Big/ \sum_{k=0}^{\overline{n_2}} n_{2|k|} 2^{-k}, \tag{4.20}$$

$$\frac{L(n_1)}{L(n_2)} = \sum_{k=0}^{\overline{n_1}} n_{1|k|} \frac{\Gamma(2\nu-k+\frac{d}{2})}{\Gamma(\nu-k)} \Big/ \sum_{k=0}^{\overline{n_2}} n_{2|k|} \frac{\Gamma(2\nu-k+\frac{d}{2})}{\Gamma(\nu-k)}, \quad \nu > \overline{n_1}, \overline{n_2}, \tag{4.21}$$

$$\frac{L(n_1)}{L(n_2)} = \sum_{k=0}^{\overline{n_1}} \frac{n_{1|k|}\ \Gamma(\nu+k+1)}{\Gamma(2\nu+1+k-\frac{d}{2})} \Big/ \sum_{k=0}^{\overline{n_2}} \frac{n_{2|k|}\ \Gamma(\nu+k+1)}{\Gamma(2\nu+1+k-\frac{d}{2})}, \quad \nu > \frac{d-1}{2}, \tag{4.22}$$

respectively. The results are no longer as nice and instructive as in Theorem 5, which is the reason why the theorem is stated in a less general setting.

Example 4.2 Consider a stationary, isotropic and differentiable random field Z on $\mathbb{R}^d$. Suppose that an observer can choose between two measurement tools. The first returns observations of Z as output, the other one returns observations of the gradient random field (2.10). Re-call, that the later means that the following d-variate random field

$$\nabla Z = \left(\frac{\partial}{\partial x_1} Z(\boldsymbol{x}),\ \frac{\partial}{\partial x_2} Z(\boldsymbol{x}),\ \ldots,\ \frac{\partial}{\partial x_d} Z(\boldsymbol{x})\right), \qquad \boldsymbol{x} = (x_1, \ldots, x_d)^T,$$

is observed at a particular site. Note that, these observations are mutually uncorrelated, since the field Z is isotropic (see Lemma 2.2). Consider again an n-points designs with vanishing inter-site correlations. Hence, Theorem 5 can be applied and it follows immediately that the efficiency measured by the limiting imse-update ratio is equal

$$\begin{aligned} \frac{L_Z}{L_{\nabla Z}} &= \frac{n\, 2^0}{n\, d\, 2^{-1}} = \frac{2}{d}, \\ \frac{L_Z}{L_{\nabla Z}} &= \frac{2\nu+\frac{d}{2}-1}{d(\nu-1)}, \qquad \nu > 1, \\ \frac{L_Z}{L_{\nabla Z}} &= \frac{2\nu+1-\frac{d}{2}}{d(\nu+1)}, \qquad \nu > \frac{d-1}{2}, \end{aligned}$$

for the Gaussian, Matérn and J-Bessel covariance function, respectively. Particularly, uncorrelated observations of Z and observations of gradients

$$\Big(\frac{\partial}{\partial x_1} Z(\boldsymbol{x}), \frac{\partial}{\partial x_2} Z(\boldsymbol{x})\Big)$$

are equivalent under the Gaussian covariance function in $\mathbb{R}^2$. In this case, the limiting imse-update ratio for a Matérn and J-Bessel class covariance function is

$$\frac{L_Z}{L_{\nabla Z}} = \frac{\nu}{\nu - 1}, \qquad \nu > 1,$$
$$\frac{L_Z}{L_{\nabla Z}} = \frac{\nu}{\nu + 1}, \qquad \nu > \frac{1}{2},$$

respectively.

Example 4.3 The previous example can be extended for compound observations of the field Z and the gradient field ∇Z. Suppose n simultaneous observations of Z and ∇Z which should be compared with n observations of Z. Note that observations of Z and ∇Z are uncorrelated if they are realized at the same site (see Lemma 2.2). From (4.20), (4.21) and (4.22) we get

$$\frac{L_{Z+\nabla Z}}{L_Z} = \left(1 + \frac{d}{2}\right),$$
$$\frac{L_{Z+\nabla Z}}{L_Z} = \left(1 + \frac{d(\nu - 1)}{(2\nu - 1 + \frac{d}{2})}\right), \qquad \nu > 1,$$
$$\frac{L_{Z+\nabla Z}}{L_Z} = \left(1 + \frac{d(\nu + 1)}{(2\nu + 1 - \frac{d}{2})}\right), \qquad \nu > \frac{d-1}{2},$$

for the Gaussian, the Matérn and the J-Bessel covariance function, respectively. Thus, we have a formula for the relative efficiency of compound observations in the case that for given observations of Z gradients are or are not available (compare Example 4.2). It it clear that a number of analogous results can be obtained considering various settings.

4.3. Efficient multiple observation of derivatives

Examples 4.2 and 4.3 introduce a situation in which different random fields can be observed at a particular location. We have profited from the fact that the fields are mutually uncorrelated at the observation site. However, if two observations of different derivative random fields are realized at the same location, they can be correlated. Correlations between observations which are realized at the same location are referred to as the *intra-site correlations*. How does the total imse-update then look like? Should some combinations of derivatives of certain order be preferred? Which one? This section is a survey of this problem.

Let us suppose the design

$$V = \{(\boldsymbol{x}, \boldsymbol{k}_1), \ldots, (\boldsymbol{x}, \boldsymbol{k}_n)\}, \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{4.23}$$

where all multi-indices $\boldsymbol{k}_i, i \in \{1, \ldots, n\}$ are different from each other. Notice from (4.23) that all random fields are observed at a single point simultaneously. The corresponding random vector $\boldsymbol{Z}$ is

$$\boldsymbol{Z} = (Z^{(\boldsymbol{k}_1)}(\boldsymbol{x}), \ldots, Z^{(\boldsymbol{k}_n)}(\boldsymbol{x}))^T. \tag{4.24}$$

Denote

$$\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{Z}, \boldsymbol{Z}^T) = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}, \tag{4.25}$$

the covariance matrix of $\boldsymbol{Z}$. Recall now some well known results and a notation from matrix algebra (see e.g. Lütkepohl (1996)). The inverse $\boldsymbol{\Sigma}^{-1}$ of the matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\det \boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{adj}, \qquad \boldsymbol{\Sigma}^{adj} = \boldsymbol{\Sigma}^{adj\,T} = \begin{bmatrix} \mathrm{cof}(a_{11}) & \dots & \mathrm{cof}(a_{1n}) \\ \vdots & \ddots & \vdots \\ \mathrm{cof}(a_{n1}) & \dots & \mathrm{cof}(a_{nn}) \end{bmatrix},$$

where the matrix $\boldsymbol{\Sigma}^{adj}$ is called *adjoint* of $\boldsymbol{\Sigma}$. *The cofactor of* a_{ij}, $\mathrm{cof}(a_{ij})$, is defined as,

$$\mathrm{cof}(a_{ij}) = (-1)^{i+j} \,\mathrm{minor}(a_{ij}),$$

where *minor of* a_{ij} is the determinant of the matrix $\boldsymbol{\Sigma}$ obtained by deleting the ith row and jth column from $\boldsymbol{\Sigma}$. Note that

$$\det \boldsymbol{\Sigma} = \sum_{i=1}^{n} a_{ij} \,\mathrm{cof}(a_{ij}), \quad \text{for any } j \in \{1, \dots, n\}. \tag{4.26}$$

4.3.1. Decomposition of the imse-update

Lemma 4.1 *Let Z be a stationary, isotropic and sufficiently differentiable random field on $\mathbb{R}^d$ such that the design (4.23) is well defined. Let the first m multi-indices from the set $I = \{\boldsymbol{k}_1, \dots, \boldsymbol{k}_n\}$ be even, $m \leq n$. Let all other multi-indices from the set I be odd (see Definition 2, page 12). Denote*

$$V_e = \{(\boldsymbol{x}, \boldsymbol{k}_1), \dots, (\boldsymbol{x}, \boldsymbol{k}_m)\} \quad \textit{and} \quad V_o = \{(\boldsymbol{x}, \boldsymbol{k}_{m+1}), \dots, (\boldsymbol{x}, \boldsymbol{k}_n)\}.$$

Then

$$U(V_e \cup V_o) = U(V_e) + U(V_o).$$

Proof : Denote $I_e = \{\boldsymbol{k}_1, \dots, \boldsymbol{k}_m\}$ and $I_o = \{\boldsymbol{k}_{m+1}, \dots, \boldsymbol{k}_n\}$. The claim follows immediately from the fact that

$$\mathrm{cov}(Z^{(\boldsymbol{i})}(\boldsymbol{x}), Z^{(\boldsymbol{j})}(\boldsymbol{x})) = 0, \quad \text{for all } \boldsymbol{i} \in I_e \text{ and } \boldsymbol{j} \in I_o$$

(see Lemma 2.2). It means that $\boldsymbol{\Sigma}$ can be written as a block matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_e & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_o \end{bmatrix}, \tag{4.27}$$

where $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_o$ are covariance matrices corresponding to the designs V_e and V_o, respectively. □

Let us define a Boolean function e on the set of all multi-indices, $e : \mathbb{N}_0^d \to \{0,1\}$, $d \in \mathbb{N}$ as follows. For a multi-index $\boldsymbol{k}$

$$e(\boldsymbol{k}) = \begin{cases} 1 & \text{if } \boldsymbol{k} \text{ is even,} \\ 0 & \text{if } \boldsymbol{k} \text{ is odd.} \end{cases}$$

The following claim derives from Lemma 2.2 and it is a supplement to Lemma 4.1.

Remark 4.1 For the covariance matrix (4.25) it holds

$$\operatorname{sgn}(|a_{ij}|) = e(\boldsymbol{k}_i + \boldsymbol{k}_j) \quad \text{for all } i,j \in \{1,\dots,n\}.$$

It means, that there is no zero entry in the matrix $\boldsymbol{\Sigma}_e$ in the proof of Lemma 4.1. Any entry in $\boldsymbol{\Sigma}_e$ can be (up to a sign) interpreted as a variance of a derivative random field. On the other hand, the matrix $\boldsymbol{\Sigma}_o$ can be sparse (see Examples 4.2 and 4.3). It means that $U(V_o)$ can be possibly re-written as a sum of imse-updates corresponding to a particular decomposition of V_o. In particular, if $d = 1$ then there is no zero entry in the matrix $\boldsymbol{\Sigma}_o$.

Denote $\boldsymbol{c}(\boldsymbol{x}_0) = (c_1(\boldsymbol{x}_0), \dots, c_n(\boldsymbol{x}_0))^T$, the vector of covariances between the design point $\boldsymbol{x}$ and a prediction point $\boldsymbol{x}_0 \in \mathbb{R}^d$. It means $c_i(\boldsymbol{x}_0) = \operatorname{cov}(Z^{(\boldsymbol{k}_i)}(\boldsymbol{x}), Z(\boldsymbol{x}_0))$, $\boldsymbol{x}_0 \in \mathbb{R}^d$.

Lemma 4.2 *Suppose that*

$$\int_{\mathbb{R}^d} c_i(\boldsymbol{x}_0) c_j(\boldsymbol{x}_0)\, \mathrm{d}\boldsymbol{x}_0 = a_{ij} g(\boldsymbol{k}_i, \boldsymbol{k}_j), \quad \textit{for all } i,j \in \{1,\dots,n\}, \tag{4.28}$$

where $g : \mathbb{N}_0^d \times \mathbb{N}_0^d \to \mathbb{R}$. Under the notation given in the preceding paragraph it holds

$$U(V) = \sum_{i=1}^n g(\boldsymbol{k}_i, \boldsymbol{k}_i) - \sum_{i=1}^n \sum_{j=1}^{i-1} \frac{a_{ij} \operatorname{cof}(a_{ij})}{\det \boldsymbol{\Sigma}} \Big[g(\boldsymbol{k}_i, \boldsymbol{k}_i) - 2g(\boldsymbol{k}_i, \boldsymbol{k}_j) + g(\boldsymbol{k}_j, \boldsymbol{k}_j) \Big]. \tag{4.29}$$

Proof : Denote $g(\boldsymbol{k}_i) := g(\boldsymbol{k}_i, \boldsymbol{k}_i)$ for short. Using (4.26) and (4.28), we get

$$\begin{aligned} U(V) &= \int_{\mathbb{R}^d} \boldsymbol{c}^T(\boldsymbol{x}_0) \boldsymbol{\Sigma}^{-1} \boldsymbol{c}(\boldsymbol{x}_0)\, \mathrm{d}\boldsymbol{x}_0 = \frac{1}{\det \boldsymbol{\Sigma}} \sum_{i=1}^n \sum_{j=1}^n \operatorname{cof}(a_{ij}) \int_{\mathbb{R}^d} c_i(\boldsymbol{x}_0) c_j(\boldsymbol{x}_0)\, \mathrm{d}\boldsymbol{x}_0 \\ &= \frac{1}{\det \boldsymbol{\Sigma}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \operatorname{cof}(a_{ij}) g(\boldsymbol{k}_i, \boldsymbol{k}_j) \\ &= \sum_{i=1}^n \frac{\sum_{j=1}^n a_{ij} \operatorname{cof}(a_{ij}) g(\boldsymbol{k}_i, \boldsymbol{k}_j) + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \operatorname{cof}(a_{ij}) g(\boldsymbol{k}_i) - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \operatorname{cof}(a_{ij}) g(\boldsymbol{k}_i)}{\sum_{j=1}^n a_{ij} \operatorname{cof}(a_{ij})} \\ &= \sum_{i=1}^n g(\boldsymbol{k}_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij} \operatorname{cof}(a_{ij})}{\det \boldsymbol{\Sigma}} \big[g(\boldsymbol{k}_i, \boldsymbol{k}_j) - g(\boldsymbol{k}_i) \big], \end{aligned}$$

which is equal to (4.29) due to the symmetry of the elements over which is summed. □

Let us note that $\sum_{i=1}^{n} g(\boldsymbol{k}_i, \boldsymbol{k}_i)$ in (4.29) is equal to the limiting imse-update (4.8). Thus, under the given condition the imse-update for the design (4.23) can be splitted into the limiting imse-update and some other part. In the following section, we will discuss simple cases, where Lemmata 4.1 and 4.2 can be used to decide between some designs.

Note that, if $d = 1$, the condition (4.28) is in no way restrictive. In the one-dimensional case this assumption can be always adopted, though if it can be sometimes difficult to find the particular explicite analytical form of the function g.

Examples with 2-points design

Denote by $\Leftrightarrow$ the equivalence of designs with respect to the imse-update (4.5), i.e. for V and V' discrete designs, $V \Leftrightarrow V'$ if and only if $U(V) = U(V')$. Further, define the relation $<$ for discrete designs V and V' as follows: $V < V'$ if and only if $U(V) < U(V')$. The design V' is termed to be *better* than V with respect to the imse-update. Analogously, the relation $\leq$ can be defined.

Consequence 4.2 *Let Z be a stationary, isotropic and sufficiently differentiable random field on $\mathbb{R}^d$ such that the random vector (4.24) is well defined. Let n and m from Lemma 4.1 be $n = 4$ and $m = 2$. (It means that four random fields $Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_3)}$ and $Z^{(\boldsymbol{k}_4)}$ can be observed, where the multi-indices in $I_e = \{\boldsymbol{k}_1, \boldsymbol{k}_2\}$ are even and the multi-indices in $I_o = \{\boldsymbol{k}_3, \boldsymbol{k}_4\}$ are odd.)*

Let (4.28) hold and let

$$g(\boldsymbol{i}, \boldsymbol{i}) - 2g(\boldsymbol{i}, \boldsymbol{j}) + g(\boldsymbol{j}, \boldsymbol{j}) > 0, \quad \text{for all } \boldsymbol{i}, \boldsymbol{j} \in I_e \text{ and for all } \boldsymbol{i}, \boldsymbol{j} \in I_o,\ \boldsymbol{i} \neq \boldsymbol{j}. \tag{4.30}$$

Let the observation area X consists of two points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\boldsymbol{x} \neq \boldsymbol{y}$, such that

$$\operatorname{cov}(Z^{(\boldsymbol{i})}(\boldsymbol{x}), Z^{(\boldsymbol{j})}(\boldsymbol{y})) = 0, \quad \text{for all } \boldsymbol{i}, \boldsymbol{j} \in I_e \cup I_o.$$

If the multi-index $\boldsymbol{k}_3 + \boldsymbol{k}_4$ is even then there are two (structurally different) designs, which are equivalent with respect to the imse-update and they are better than all other designs under the given conditions. Otherwise, there are four (structurally different) designs, which are equivalent with respect to the imse-update and they are better than all other designs.

Proof : Omitting the cases where only $\boldsymbol{x}$ changes with $\boldsymbol{y}$ (this is meant by structurally different designs), the following configurations are possible:

	Observation(s) at $\boldsymbol{x}$	Observation(s) at $\boldsymbol{y}$
Design A:	$Z^{(\boldsymbol{k}_1)}$	$Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_3)}, Z^{(\boldsymbol{k}_4)}$
Design A':	$Z^{(\boldsymbol{k}_2)}$	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_3)}, Z^{(\boldsymbol{k}_4)}$
Design B:	$Z^{(\boldsymbol{k}_3)}$	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_4)}$
Design B':	$Z^{(\boldsymbol{k}_4)}$	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_3)}$
Design D:	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_2)}$	$Z^{(\boldsymbol{k}_3)}, Z^{(\boldsymbol{k}_4)}$
Design C:	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_3)}$	$Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_4)}$
Design C':	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_4)}$	$Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_3)}$
Design D':	$Z^{(\boldsymbol{k}_1)}, Z^{(\boldsymbol{k}_2)}, Z^{(\boldsymbol{k}_3)}\ Z^{(\boldsymbol{k}_4)}$	

Table 4.1.: Structurally different designs for two locations

Take into account the fact that observations of derivatives $Z^{(\boldsymbol{k})}(\boldsymbol{x})$ and $Z^{(\boldsymbol{l})}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$ are uncorrelated, if $\boldsymbol{k}+\boldsymbol{l}$ is odd. It means that the covariance matrix of any design is in the form (4.27), where $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_o$ are 2×2-covariance matrices. The covariance matrices of above designs differ in the values of entries outside the diagonal, a_{12} and a_{34}. Using Lemma 4.1, (4.29) can be re-written as follows

$$\begin{aligned} U(V_e \cup V_o) &= U(V_e) + U(V_o) \\ &= \sum_{i=1}^{4} g(\boldsymbol{k}_i) + \frac{a_{12}^2}{\det \boldsymbol{\Sigma}_e}\big[g(\boldsymbol{k}_1,\boldsymbol{k}_1) - 2g(\boldsymbol{k}_1,\boldsymbol{k}_2) + g(\boldsymbol{k}_2,\boldsymbol{k}_2)\big] \\ &\quad + \frac{a_{34}^2}{\det \boldsymbol{\Sigma}_o}\big[g(\boldsymbol{k}_3,\boldsymbol{k}_3) - 2g(\boldsymbol{k}_3,\boldsymbol{k}_4) + g(\boldsymbol{k}_4,\boldsymbol{k}_4)\big]. \end{aligned} \tag{4.31}$$

It is easy to see that

$$A \Leftrightarrow A', \quad B \Leftrightarrow B', \quad C \Leftrightarrow C' \text{ and } \quad D \Leftrightarrow D'. \tag{4.32}$$

For instance, in the designs A and A' there is $a_{12} = 0$ and a_{34} is the same in the both designs. This can be seen from Table 4.1.

Note that if $\boldsymbol{k}_3 + \boldsymbol{k}_4$ is even then a_{34} can not be zero. Using (4.30), it can be seen from (4.31) that a design maximizes $U(V)$ if both a_{12} and a_{34} are not null. This is true for the designs D and D' only. On the other hand, the designs C and C' are the worst ones, since $a_{12} = 0$ and $a_{34} = 0$. It can be seen from (4.31) that the following relation between designs holds:

$$D > B, A > C.$$

If $\boldsymbol{k}_3 + \boldsymbol{k}_4$ is odd then $a_{34} = 0$ in every case. It means that $D \Leftrightarrow B$ and $A \Leftrightarrow C$. Designs D, D', B, B' are in this case equivalent and better than all other designs (which are also equivalent with each other). □

Consequence 4.2 is an example for the fact that correlated observations can be more efficient than uncorrelated observations. Moreover, the designs where all observations are mutually uncorrelated appears to be the worst one. Note that the correlations between observations at a particular location can be positive or negative. In the both cases some (sub)optimality is reached. Note that, a relation between design B, A can not be established under the given setup. Actually, both situations, $B > A$ and $A < B$ are possible. Note also that if $d = 1$, the situation in that $\boldsymbol{k}_3 + \boldsymbol{k}_4$ is odd can not occur.

Example 4.4 Suppose a stationary stochastic process Z on $\mathbb{R}^1$ with the following covariance function with the compact support $[0, 1]$:

$$\psi_{4,3}(t) = \sigma^2(1-t)_+^7(21t^3 + 19t^2 + 7t + 1), \quad t \geq 0$$

It follows from the theory given in Wendland (1996) that the function $\psi_{4,3}(t)$ is a covariance function and it is 6 times continuously differentiable. (see Theorem 5.2 therein, cf. also section 3.2.4). Note that the range (or support) of the function $\psi_{4,3}$ can be easily changed by a linear transformation of the parameter t.

Since $\psi_{4,3}$ is 6 times differentiable, derivative observations of Z up to the third order can be considered.

Consider the following situation. Our experimental area consists of two points $x \in \mathbb{R}$ and $y \in \mathbb{R}$, such that $|y - x| \geq 1$. We can observe processes $Z^{(0)}$, $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$ only once and we can choose between these points x and y. The resulting design should be optimal with respect to the imse-update.

It is obvious that these experimental conditions correspond to the prerequisites of Consequence 4.2 for $k_1 = 0$, $k_2 = 2$, $k_3 = 1$ and $k_4 = 3$. We only need to check whether (4.30) holds. This can be done by a straightforward but tedious calculation, thus, only the results (resp. their numerical approximations) are presented. The values of the function g for all combinations of indices, which are of interest, are: $g(k_1, k_1) \doteq 0.4161$, $g(k_1, k_2) \doteq 0.2139$, $g(k_2, k_2) \doteq 0.1027$, $g(k_3, k_3) \doteq 0.2129$, $g(k_3, k_4) \doteq 0.1027$, and $g(k_4, k_4) \doteq 0.0298$. Thus, it is obvious that (4.30) holds and therefore D and D' are the best designs under the given conditions.

The covariance matrix of the design D is

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -18 & 0 & 0 \\ -18 & 1008 & 0 & 0 \\ 0 & 0 & 18 & -1008 \\ 0 & 0 & -1008 & 151200 \end{bmatrix}$$

An explicite evaluation of the imse-update in the particular cases yields

$$U(D) \doteq 0.8285, \quad U(B) \doteq 0.8056, \quad U(A) \doteq 0.7854, \quad U(C) \doteq 0.7625.$$

Note that, although Consequence 4.2 is formulated for two design points (x and y) only, the claim can be directly generalized to designs with more design points. Anyway, this generalization is done in the next section in a slightly different context. The simple example-like setup was chosen to avoid useless technicalities.

4.3.2. Multiple observations of derivatives for designs with vanishing inter-site correlations

It clear that the necessary condition, which must be put on the covariance function of Z in Consequence 4.2 is that it must have a compact support. Otherwise, the observations at the points $\boldsymbol{x}$ and $\boldsymbol{y}$ will not be uncorrelated. To obtain similar claims as in Consequence 4.2 for covariance functions without compact support, an analogous setting as in the proof of Theorem 5 has to be adopted.

Conditions for designs with vanishing inter-site correlations

Let a set of multi-indices $I = \{\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4\}$ be given and let us assume that multi-indices in the set $I_e = \{\boldsymbol{k}_1, \boldsymbol{k}_2\}$ are even (see Definition 2, page 12). The multi-indices in the set $I_o = \{\boldsymbol{k}_3, \boldsymbol{k}_4\}$ are supposed to be odd, $I = I_e \cup I_o$.

Let Z be a stationary, isotropic and sufficiently differentiable random field on $\mathbb{R}^d$, such that $Z^{(\boldsymbol{k}_1)}$, $Z^{(\boldsymbol{k}_2)}$, $Z^{(\boldsymbol{k}_3)}$, and $Z^{(\boldsymbol{k}_4)}$ are well defined. Let V^M be the set of all sequences $\{V_u\}_{u=1}^{\infty}$ of regular designs (see (2.18))

$$V^u = \{(\boldsymbol{x}_1^u, \boldsymbol{l}_1), \ldots, (\boldsymbol{x}_M^u, \boldsymbol{l}_M)\}, \quad \boldsymbol{x}_i^u \in \mathbb{R}^d, \boldsymbol{l}_i \in I, \ i \in \{1, \ldots, M\}, \tag{4.33}$$

such that

$$\lim_{u \to \infty} \operatorname{cov}(Z^{(\boldsymbol{l}_i)}(\boldsymbol{x}_i^u), Z^{(\boldsymbol{l}_j)}(\boldsymbol{x}_j^u)) = 0, \tag{4.34}$$

for all $i, j \in \{1 \dots, M\}$, $i \neq j$, such that $\boldsymbol{x}_i^u \neq \boldsymbol{x}_j^u$. Notice from (4.33) that only the locations of the designs depends on n, and not the order of the derivatives. Note that the definition (4.33) allows the design points $\boldsymbol{x}_i^u, i \in \{1, \dots, M\}$ need not to be different from each other. Regularity of the designs assures that a particular pair $(\boldsymbol{x}_i^u, \boldsymbol{l}_i)$ appears in V^M only once. The condition (4.34) means that we avoid (in the limit) any inter-site correlations. On the other hand, nonzero intra-site correlations are possible.

Let $\boldsymbol{\Sigma}^u$ be the covariance matrix corresponding to the design V_u and denote $\boldsymbol{\Sigma} = \lim_{u\to\infty} \boldsymbol{\Sigma}^u$. The elements of $\boldsymbol{\Sigma}$ are denoted as in (4.25). Similarly, we can define $V_{e,u}$ and $V_{o,u}$, i.e. designs with even and odd multi-indices, respectively, such that $V_u = V_{e,u} \cup V_{o,u}$.

Denote $c_i(\boldsymbol{x}_0) = \lim_{u\to\infty} c_i^u(\boldsymbol{x}_0) = \lim_{u\to\infty} \operatorname{cov}(Z^{(\boldsymbol{l}_i)}(\boldsymbol{x}_i^u), Z(\boldsymbol{x}_0))$ for a sequence of design points $\{\boldsymbol{x}_i^u\}_{u=1}^\infty$ and a prediction point $\boldsymbol{x}_0$. Suppose $c_i \in L^2(\mathbb{R}^d)$ for all $i \in \{1, \dots, M\}$. Using the dominated convergence theorem (see e.g. Theorem 8.19 in Kuttler (1998)) we get an analogous claim as in Lemma 4.1 and 4.2.

Subsequently, let us suppose that each random field $Z^{(\boldsymbol{k}_1)}$, $Z^{(\boldsymbol{k}_2)}$, $Z^{(\boldsymbol{k}_3)}$ and $Z^{(\boldsymbol{k}_4)}$ can be observed maximal N times and let $N < M/4$.

Subsequently, as in Consequence 4.2, if $\boldsymbol{k}_3 + \boldsymbol{k}_4$ is even, then two sets $\bar{V}, \bar{\bar{V}} \subset V^M$ of (structurally different) design sequences exist, such that

$$\lim_{u\to\infty} U(D_u) = \lim_{u\to\infty} U(D'_u) > \lim_{u\to\infty} U(B_u)$$

for all $\{D_u\}_{u=1}^\infty \in \bar{V}$, $\{D'_u\}_{u=1}^\infty \in \bar{\bar{V}}$ and $\{B_u\}_{u=1}^\infty \in V^M - \{\bar{V} \cup \bar{\bar{V}}\}$.

Similarly as in Consequence 4.2, if $\boldsymbol{k}_3 + \boldsymbol{k}_4$ is odd, four sets of (structurally different) design sequences can be found.

The designs from the set $\bar{V}$ are constructed so that $Z^{(\boldsymbol{k}_1)}$ and $Z^{(\boldsymbol{k}_2)}$ are observed together in arbitrary N locations. Observations of random fields $Z^{(\boldsymbol{k}_3)}$ and $Z^{(\boldsymbol{k}_4)}$ are also put together, but they are observed at the remaining locations, which are not occupied by $Z^{(\boldsymbol{k}_1)}$ and $Z^{(\boldsymbol{k}_2)}$. It is a situation which corresponds to the design D in Table 4.1. Some more specific description of the locations is not needed, since all designs have in the limit the same (up to a permutation of rows and columns) covariance matrix.

The designs from the set $\bar{\bar{V}}$ are constructed in such a way that all random fields have to be observed at each site up to the number of all possible observations N. This is a situation, which corresponds to the design D' in Table 4.1.

Note that we restrict ourselves to the claim about limiting designs. Thus, for some given $u \in \mathbb{N}$, the above relations between designs need not to be valid. However, the continuity of the imse-update implies that the relations hold for a sufficiently large u too.

Example 4.5 Let a random field Z on $\mathbb{R}^d$ fulfill the above cited conditions and let Z be driven by the Gaussian covariance function (3.15). Let multi-indices $\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3$ and $\boldsymbol{k}_4$ be given as in the preceding paragraph and let $\boldsymbol{k}_3 + \boldsymbol{k}_4$ be even. This is for instance achived if $d = 1$.

For the simplicity, let us suppose that each particular V_u in (4.33) consists of two distinct locations $\boldsymbol{x}^u$ and $\boldsymbol{y}^u$ only, where possibly several random fields can be observed. This is a situation which is completely analogous to that one in Table 4.1.

We need to check whether (4.30) holds. Using (3.21) together with the result (A.54) given in Appendix A.6.1, it can be shown that for any multi-indices $\boldsymbol{k}$ and $\boldsymbol{l}$

$$g(\boldsymbol{k}, \boldsymbol{l}) = \begin{cases} \sigma^2 \left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d 2^{-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}}, & \text{if } e(\boldsymbol{k}+\boldsymbol{l}) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.35}$$

Thus, it can be easily seen that

$$g(\boldsymbol{k},\boldsymbol{k}) - 2g(\boldsymbol{k},\boldsymbol{l}) + g(\boldsymbol{l},\boldsymbol{l}) = \sigma^2\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \left(2^{-|\boldsymbol{k}|} - 2\cdot 2^{-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}} + 2^{-|\boldsymbol{l}|}\right)$$
$$= \sigma^2\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \left(2^{-\frac{|\boldsymbol{k}|}{2}} - 2^{-\frac{|\boldsymbol{l}|}{2}}\right)^2 \geq 0.$$

Therefore, the designs D and D' are not worse than any other design (from the considered class of designs). If moreover $|\boldsymbol{k}| \neq |\boldsymbol{l}|$, these designs are always better (in the limiting sense of the imse-update).

For instance, if $d=1$, $k_1=0$, $k_2=2$, $k_3=1$ and $k_4=3$, we get the following values of the limiting imse-update $L(\cdot)$ for the designs denoted as in Table 4.1.

$$L(D) = b\cdot 2\frac{3}{16}, \quad L(A) = b\cdot 2\frac{1}{16}, \quad L(B) = b\cdot 2, \quad L(C) = b\cdot 1\frac{7}{8},$$

where $b = \sigma^2(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}})$.

We can consider higher derivatives as well. For example, if $k_1 = 1$, $k_2 = 3$, $k_3 = 2$ and $k_4 = 4$, we get

$$L(D) = b\cdot 1\frac{9}{32}, \quad L(B) = b\cdot 1\frac{4}{32}, \quad L(A) = b\cdot 1\frac{3}{32}, \quad L(C) = b\cdot \frac{30}{32}.$$

Note that the relation of designs A and B is converse in the above particular cases.

Example 4.6 Assume the same situation as in Example 4.5, but let us suppose a covariance function from the Matérn class instead of the Gaussian covariance function. Using (3.13) together with the result (A.57) stated in Appendix A.6.2, we get

$$g(\boldsymbol{k},\boldsymbol{l}) = \phi\,\frac{2^{\nu+d-1}}{\pi^{\frac{d}{2}}\alpha^d}\,B\big(\nu+\frac{d}{2},\nu+\frac{d}{2}\big)\,\frac{\Gamma\big(2\nu+\frac{d}{2}-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}\big)}{\Gamma\big(\nu-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}\big)}$$

for all multi-indices $\boldsymbol{k}$ and $\boldsymbol{l}$ such that $e(\boldsymbol{k}+\boldsymbol{l}) = 1$. Without loss of generality, consider $|\boldsymbol{k}| \geq |\boldsymbol{l}|$. Recalling (4.30), we need to check whether

$$\frac{\Gamma\big(2\nu+\frac{d}{2}-|\boldsymbol{k}|\big)}{\Gamma\big(\nu-|\boldsymbol{k}|\big)} - 2\frac{\Gamma\big(2\nu+\frac{d}{2}-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}\big)}{\Gamma\big(\nu-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}\big)} + \frac{\Gamma\big(2\nu+\frac{d}{2}-|\boldsymbol{l}|\big)}{\Gamma\big(\nu-|\boldsymbol{l}|\big)} \tag{4.36}$$

is positive (or nonnegative). Obviously, if $|\boldsymbol{k}| = |\boldsymbol{l}|$ then the above term is equal to zero. Denote $k = |\boldsymbol{k}|$ and $l = |\boldsymbol{l}|$. Since $e(\boldsymbol{k}+\boldsymbol{l}) = 1$, we can write $l = k - 2m$, $m \in \mathbb{N}$ (the trivial case $m = 0$ can be excluded). Define

$$\beta = \frac{\Gamma\big(2\nu+\frac{d}{2}-k\big)}{\Gamma\big(\nu-k\big)}.$$

We can re-write (4.36) as follows

$$\beta\left[1-2\frac{\Gamma(2\nu+\frac{d}{2}-k+m)}{\Gamma(2\nu+\frac{d}{2}-k)}\frac{\Gamma(\nu-k)}{\Gamma(\nu-k+m)}+\frac{\Gamma(2\nu+\frac{d}{2}-k+2m)}{\Gamma(2\nu+\frac{d}{2}-k)}\frac{\Gamma(\nu-k)}{\Gamma(\nu-k+2m)}\right]=$$
$$\beta\left[1-2\frac{(2\nu+\frac{d}{2}-k+m-1)}{(\nu-k+m-1)}\cdots\frac{(2\nu+\frac{d}{2}-k)}{(\nu-k)}+\frac{(2\nu+\frac{d}{2}-k+2m-1)}{(\nu-k+2m-1)}\cdots\frac{(2\nu+\frac{d}{2}-k)}{(\nu-k)}\right]=$$
$$\beta\left[1-2\Big(1+\frac{\nu+\frac{d}{2}}{\nu-k+m-1}\Big)\cdots\Big(1+\frac{\nu+\frac{d}{2}}{\nu-k}\Big)+\Big(1+\frac{\nu+\frac{d}{2}}{\nu-k+2m-1}\Big)\cdots\Big(1+\frac{\nu+\frac{d}{2}}{\nu-k}\Big)\right]=$$
$$\beta\left[1+\Big(1+\frac{\nu+\frac{d}{2}}{\nu-k+m-1}\Big)\cdots\Big(1+\frac{\nu+\frac{d}{2}}{\nu-k}\Big)\left[\Big(\underbrace{1+\frac{\nu+\frac{d}{2}}{\nu-k+2m-1}}_{>2}\Big)\cdots\Big(\underbrace{1+\frac{\nu+\frac{d}{2}}{\nu-k+m}}_{>2}\Big)-2\right]\right].$$

It can be seen that every term in the parentheses is bigger than 2 and therefore (4.36) > 0. Subsequently, the designs D and D' are not worse (in the limiting sense) than any other design from the considered class of designs. If moreover $|\boldsymbol{k}| \neq |\boldsymbol{l}|$, these designs are always better than any other design. This hold for all values of the smoothness parameter ν for which a particular Matérn covariance function is sufficiently differentiable.

Resume

In the first part of this chapter, it is assumed that there is only one observation at a particular site. Otherwise, it is supposed that they are mutually uncorrelated. In the second part, this assumption is weakened. The main focus here is to investigate whether the (limiting) imse-update can be improved by a combination of observations of distinct derivative random fields at one site. In the simplest version of the problem, we can ask: do some pairs (triples etc.) of derivative random fields exist, which should be observed together at one site, if possible? The observations at different sites are supposed to be uncorrelated. Finally, it has been shown that *correlated pairs of derivatives outperform the other possibilities*.

What is the main sense of this survey? Later on, it will be shown on some very simple examples, that the space, where optimization to be proceeded, is very complex and it has many local maxima. Moreover, the values of this maxima are typically close to each other and passing from one maximum to another could be difficult. Therefore, without a good approximation of the solution, the standard optimization numerical methods are usually not successful in search of the optimum.

An application of the presented results consists in a reduction of the complexity of the space of all possible solutions. Seeking for the optimum, we can restrict ourselves to some suitable subspace of the space of all possible solutions. The suggested pairing could be such a suitable restriction in some situations.

However, it will be shown later that this pairing is only of a minor importance if it is possible to choose observation locations arbitrarily. Typically, by choosing the sample locations properly, a bigger improvement of the imse-update can be achived. Anyway, for such designs, the approach of vanishing inter-site correlations is not appropriate, since the correlations between observations are large.

5. An empirical survey of designs with correlated observations

Until now, it was presumed that (derivative) observations at different sites are uncorrelated (in the asymptotic sense, or due to a compact support or due to zero points of the study covariance function). This setting allows to calculate the imse-update analytically in a simple way. If this assumption is weakened, analytical results (if they are possible) can not be understood easily, except some particular very simple cases.

In this chapter, we restrict ourselves mainly to particular analytical results in combination with extensive numerical studies, which support the argumentation. The aim of this section is to give some notion to the following question: Can the imse-update be improved by a suitable choice of design points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$? The answer is as expected positive. Thus, the further focus is to give at least some heuristic rules for a construction of efficient designs.

5.1. Imse-update and convolution

Let us consider a stationary, isotropic and sufficiently differentiable random field on $\mathbb{R}^d$ and let V_n be a design in the form (2.17). The imse-update (4.5) can be written as follows

$$U(V_n) = \int_{\mathbb{R}^d} \boldsymbol{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{c} \, \boldsymbol{dx}_0 = \int_{\mathbb{R}^d} \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{c}\boldsymbol{c}^T) \boldsymbol{dx}_0. \tag{5.1}$$

Since $\boldsymbol{\Sigma}^{-1}$ does not depend on $\boldsymbol{x}_0$ and the trace of a matrix is a linear operator of its argument, we can write

$$U(V_n) = \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \int_{\mathbb{R}^d} \boldsymbol{c}\boldsymbol{c}^T \boldsymbol{dx}_0). \tag{5.2}$$

Now, let us look more detailed at the elements of the $n \times n$-matrix $\boldsymbol{C^*} = \int_{\mathbb{R}^d} \boldsymbol{c}\boldsymbol{c}^T \boldsymbol{dx}_0$. Let $c_i = \mathrm{cov}(Z^{(\boldsymbol{k}_i)}(\boldsymbol{x}_i), Z(\boldsymbol{x}_0)) =: c_{\boldsymbol{k}_i 0}(\boldsymbol{x}_0 - \boldsymbol{x}_i)$ and $c_j = \mathrm{cov}(Z^{(\boldsymbol{k}_j)}(\boldsymbol{x}_j), Z(\boldsymbol{x}_0)) =: c_{\boldsymbol{k}_j 0}(\boldsymbol{x}_0 - \boldsymbol{x}_j)$ be the ith and the jth element of the vector $\boldsymbol{c}$. Since

$$c_{\boldsymbol{k}_j 0}(\boldsymbol{x}_0 - \boldsymbol{x}_j) = \mathrm{cov}(Z^{(\boldsymbol{k}_j)}(\boldsymbol{x}_j), Z(\boldsymbol{x}_0)) = \mathrm{cov}(Z(\boldsymbol{x}_0), Z^{(\boldsymbol{k}_j)}(\boldsymbol{x}_j)) =: c_{0\boldsymbol{k}_j}(\boldsymbol{x}_j - \boldsymbol{x}_0),$$

we can write

$$\begin{aligned} \boldsymbol{C}^*_{ij} = \int_{\mathbb{R}^d} c_i c_j \, \boldsymbol{dx}_0 &= \int_{\mathbb{R}^d} c_{\boldsymbol{k}_i 0}(\boldsymbol{x}_0 - \boldsymbol{x}_i) c_{\boldsymbol{k}_j 0}(\boldsymbol{x}_0 - \boldsymbol{x}_j) \, \boldsymbol{dx}_0 \\ &= \int_{\mathbb{R}^d} c_{\boldsymbol{k}_i 0}(\boldsymbol{x}_0 - \boldsymbol{x}_i) c_{0\boldsymbol{k}_j}(\boldsymbol{x}_j - \boldsymbol{x}_0) \, \boldsymbol{dx}_0. \end{aligned}$$

Denote $\boldsymbol{h} = \boldsymbol{x}_j - \boldsymbol{x}_i$. Because of the stationarity of Z and since we integrate over $\mathbb{R}^d$, we can set $\boldsymbol{x}_i = \boldsymbol{0}$. Thus, we get

$$\boldsymbol{C}^*_{ij} = \int_{\mathbb{R}^d} c_{\boldsymbol{k}_i 0}(\boldsymbol{x}_0) c_{0\boldsymbol{k}_j}(\boldsymbol{h} - \boldsymbol{x}_0) \, \boldsymbol{dx}_0 = c_{0\boldsymbol{k}_j} * c_{\boldsymbol{k}_i 0}(\boldsymbol{h}). \tag{5.3}$$

An exact calculation of the convolution (5.3) for the Gaussian covariance function and for covariance functions from the Matérn and the J-Bessel class can be found in Appendix A (Sections A.6.1 – A.6.3). Note that in general the matrix $\boldsymbol{C}^*$ depends on a particular design.

5.2. Designs with only two observations

Let us suppose that $Z^{(\boldsymbol{k})}$ has been observed at the location $\boldsymbol{x}_1$. The task is to find a point $\boldsymbol{x}_2$, where $Z^{(\boldsymbol{l})}$ should be observed, in order to maximize the imse-update (5.1). Denote $\boldsymbol{h} = \boldsymbol{x}_2 - \boldsymbol{x}_1$ and the imse-update for the two observations as $U_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h})$. Denoting the elements of the covariance matrix of this 2-points design as

$$\boldsymbol{\Sigma} = \begin{bmatrix} a & b(\boldsymbol{h}) \\ b(\boldsymbol{h}) & d \end{bmatrix},$$

it can be seen that

$$U_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h}) = \frac{d\,||c_{\boldsymbol{k}0}||_2^2 - 2b(\boldsymbol{h})\,c_{0\boldsymbol{l}} * c_{\boldsymbol{k}0}(\boldsymbol{h}) + a\,||c_{\boldsymbol{l}0}||_2^2}{ad - b^2(\boldsymbol{h})}. \tag{5.4}$$

Note that only the integral in the middle of the above term depends on $\boldsymbol{h}$.

5.2.1. Gaussian covariance function

If the covariance function of Z is Gaussian (3.15), the function (5.4) has the following form:

$$U_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h}) = \beta\,\frac{Q\,2^{-|\boldsymbol{k}|} - 2\exp(-\frac{3}{2}\,\alpha^2||\boldsymbol{h}||^2)\,2^{-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}}\,\mathcal{H}^{\alpha}_{\boldsymbol{k}\boldsymbol{l}}(\boldsymbol{h}) + Q\,2^{-|\boldsymbol{l}|}}{Q - \exp(-2\alpha^2||\boldsymbol{h}||^2)\,\prod_{i=1}^{d} H^2_{k_i+l_i}(\alpha h_i)}, \tag{5.5}$$

where

$$\beta = \sigma^2\Big(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\Big)^d, \quad Q = \prod_{i=1}^{d}\frac{(2k_i)!}{k_i!}\frac{(2l_i)!}{l_i!}, \quad \mathcal{H}^{\alpha}_{\boldsymbol{k}\boldsymbol{l}}(\boldsymbol{h}) = \prod_{i=1}^{d} H_{k_i+l_i}(\alpha h_i)H_{k_i+l_i}(\frac{\sqrt{2}}{2}\alpha h_i).$$

Here, H_k stands for the Hermite polynomial of the order k (see e.g. Abramowitz and Stegun (1972, Chapter 22)). The equation (5.5) follows from (A.53), (A.55) and (3.21).

Rather than the function (5.5), let us consider the following function

$$U'_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h}) = U_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h}) - \frac{||c_{\boldsymbol{k}0}||_2^2}{a} \tag{5.6}$$

which is only a small modification of the original function. This modified objective function (5.6) corresponds to the influence on the imse-update caused by the additional observation of $Z^{(\boldsymbol{l})}$ only. It is clear that, for a given $\boldsymbol{k}$, a maximization of $U_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h})$ and $U'_{\boldsymbol{k},\boldsymbol{l}}(\boldsymbol{h})$ is equivalent. Note that in the Gaussian case,

$$\frac{||c_{\boldsymbol{k}0}||_2^2}{a} = \beta\,2^{-|\boldsymbol{k}|}$$

(cf. with (4.35)). Formulas, which are necessary to evaluate (5.4) and (5.6) for the Matérn and the J-Bessel class, are provided in Appendix A (Sections A.6.2 and A.6.3) or in Chapter 3. In the following, only the case of the Gaussian covariance function will be discussed, however, an analogous case studies can be done for the Matérn and the J-Bessel class. In particular, Appendix C provides such examples.

Evaluation of the modified imse-update for the Gaussian covariance function in $\mathbb{R}$

For $d = 1$, plots of (5.6) for diverse k and l are given in Figures 5.1 – 5.4. In all graphs some regularities can readily be seen. First, note that if the distance h is big enough, the imse-update (5.6) is practically constant. The value of the constant can be easily derived from Theorem 5. For a large h, it can readily be seen that as l increases the imse-update decreases very fast. This is basically the claim of Theorem 5.

Figures 5.1(b), 5.1(d) and 5.1(f) show that, the imse-update (5.6) is (nearly) the same for $h = 0$ and $h \gg 0$. This is clear since the corresponding observations are (nearly) uncorrelated for $h = 0$ as well as for $h \gg 0$.

Notice from Figures 5.1(c), 5.1(e) 5.2(c) and 5.2(e), that, if $k + l$ is an even number, then the process $Z^{(k)}$ and $Z^{(l)}$ should be observed at the same location. The maximal imse-update is achieved in this case (cf. also the results of Section 4.3).

In Figures 5.1(b), 5.1(d), 5.1(f), 5.2(b), 5.2(d) and 5.2(f), it can be seen that if $k + l$ is an odd number, the imse-update (5.6) has its maximum outside from the origin. It is interesting that the maximum is reached in some distance, say $h^*(k, l)$, and $h^*(k, l) \approx h^*(k, l')$ for a given k and any l and l', such that $k + l$ and $k + l'$ is odd. The whole plots (aside from the different scaling) look similar for l and l', which are not too different.

Note that the situation is analogous for $k > 1$ (see Figures 5.3 and 5.4 among other). The optimal values $h^*(k, l)$ can be numerically estimated for diverse k and l using the standard quasi-Newton method (Fletcher, 1991; Kosmol, 1993). These results are summarized in Table 5.1. In this table, it can be observed for instance that if $k = l$, then $h^*(k, k)$ decreases as k increases etc.

Maybe the most significant feature, which can be seen on the plots of the imse-update (5.6), is that the local maxima appears consecutively rather regularly. See for instance the diagonal of Table 5.1 and 5.3. The values on the diagonal in Table 5.3 are nearly exactly two times bigger than the corresponding values in Table 5.1. This seems to be the reason why some well constructed regular grids perform significantly better than other ones. This issue is discussed in Section 5.4.

Optimal pairs

Table 5.2 shows the value of the modified imse-update (5.6) in the numerical approximation of the global optimum. The table is not symmetric, since the modified imse-update (5.6) is not symmetric in k and l. This tables can be useful to decide which "pairs" of derivatives could be preferred. By a pair it is meant two observations which are mutually optimal located. Let us consider the following very simple example where this idea will be sketched.

Example 5.1 Let us suppose that $Z^{(0)}$ and $Z^{(2)}$ have been observed once and we can observe $Z^{(4)}$ additionally. The distance between observations sites of $Z^{(0)}$ and $Z^{(2)}$ let be such that these observations can be considered as uncorrelated. The additional observations of $Z^{(4)}$ (wherever it is located) will be correlated at most only with one of the previous observations. This is achieved if the distance between $Z^{(0)}$ and $Z^{(2)}$ is sufficiently large. The question is where the observation of $Z^{(4)}$ should be located.

Analogously to Section 4.3, it can be found that the worst case is if observation of $Z^{(4)}$ is uncorrelated with both observations of $Z^{(0)}$ and $Z^{(2)}$. Therefore, it remains the question with which observation should $Z^{(4)}$ be correlated (i.e. what is the optimal pair) and subsequently in which distance.

The answer can be found in the 5th column of Table 5.2. The gain of the imse-update $U'_{2,4}(h^*(2,4))$ observing $Z^{(4)}$ optimally correlated with $Z^{(2)}$ is nearly twice bigger than the gain of imse-update $U'_{0,4}(h^*(0,4))$ observing optimally $Z^{(4)}$ together with $Z^{(0)}$, namely, 0.274162 compared to 0.144425. According to Table 5.1, the optimal distance $h^*(2,4)$ is equal to zero.

k	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$
0	∞	0.560807	0	0.470808	0	0.416437	0
1	0.560807	1.057638	0.450014	0	0.400039	0	0.365590
2	0	0.450014	0.854591	0.388096	0	0.354898	0
3	0.470808	0	0.388096	0.742809	0.346816	0	0.322626
4	0	0.400039	0	0.346816	0.667800	0.316650	0
5	0.416437	0	0.354898	0	0.316650	0.612520	0.293314
6	0	0.365590	0	0.322626	0	0.293314	0.569422

Table 5.1.: Optimal distance $h^*(k,l)$ of two observations in $\mathbb{R}^1$ (numerical approximation) under the Gaussian model (3.15) ($\alpha = 1$).

k	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$
0	1.253314	0.835140	0.469993	0.248563	0.144425	0.072120	0.040812
1	1.461797	0.830408	0.609749	0.391661	0.225445	0.149321	0.081702
2	1.409978	0.923078	0.527314	0.396116	0.274162	0.165231	0.115968
3	1.345213	0.861653	0.552780	0.318295	0.242520	0.176247	0.109185
4	1.319407	0.773770	0.509159	0.320852	0.185932	0.143162	0.107707
5	1.286268	0.736812	0.439394	0.293746	0.182328	0.106172	0.082429
6	1.274544	0.688776	0.409714	0.246266	0.166456	0.102012	0.059623

Table 5.2.: Value of the modified imse-update (5.6) in the first optimum in $\mathbb{R}^1$, i.e. $U'_{k,l}(h^*(k,l))$

Previous example does not mean that the imse-update $U_{2,4}$ is bigger than $U_{0,4}$, but just that $U'_{2,4}(h^*(2,4)) > U'_{0,4}(h^*(0,4))$. These are improvements of the original criterion caused by the additional observation of $Z^{(4)}$.

It can readily be seen from Table 5.2 that, if $Z^{(2)}$ is replaced by $Z^{(3)}$, the result will be the same with the distinction that optimal location will differ, namely $h^*(3,4) \doteq 0.347$. A lot of analogous examples can be easily derived from the both tables.

Multiple maxima of imse-update and sensitivity of designs with derivatives

As k or l increases, multiple local maxima on the plots of the modified imse-update (5.6) can be seen. There is typically only one which is significantly better than the other. This maximum is always the closest one with respect to the origin. The distance corresponding to the second local optimum and the values of (5.6) in this optimum are summarized in Tables 5.3 and 5.4.

As a preview of difficulties, which we meet on the search of the optimal design, let us take Figures 5.4(e) and 5.4(f). First, notice a lot of local maxima which do not differ much from each other, but more important, observe the high sensitivity of the imse-update. If we

over-estimate the distance for the global maximum only a little, then we can arrive at the global minimum. This aspect makes generally the problem of optimal design for derivatives very difficult. Although these effects are milder for lower derivatives, also here a similar tendency can be observed. Note that this high sensitivity is obvious despite the fact that we consider the Gaussian covariance function which is rather regular. In this context, it should be also noted that the imse-update is a very mild criterion, which flattens extrema. It can be conjectured that criteria based on the maximum mean square error (2.23) are much more sensitive and a construction of an optimal design is very difficult.

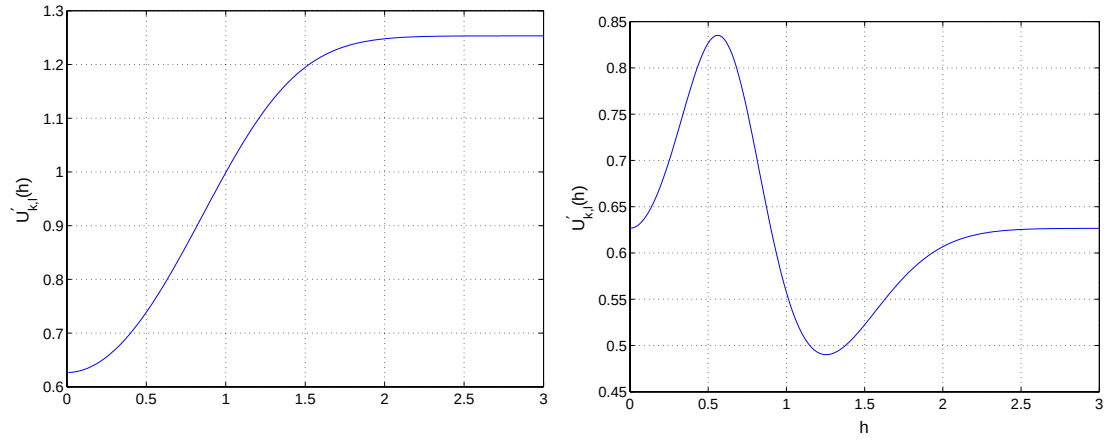
k	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$
0	-	0,∞	1.076588	1.532784	0.899114	1.285173	0.792860
1	0,∞	∞	1.507190	0.865963	1.254993	0.772221	1.118308
2	1.076588	1.507190	1.902469	1.235527	0.750776	1.096047	0.689089
3	1.532784	0.865963	1.235527	1.586559	1.081422	0.673883	0.987978
4	0.899114	1.254993	0.750776	1.081422	1.399993	0.976596	0.617409
5	1.285173	0.772221	1.096047	0.673883	0.976596	1.270463	0.898662
6	0.792860	1.118308	0.689089	0.987978	0.617409	0.898662	1.172966

Table 5.3.: Second optimal distance of two observations in $\mathbb{R}^1$ under the Gaussian model (3.15) ($\alpha = 1$)

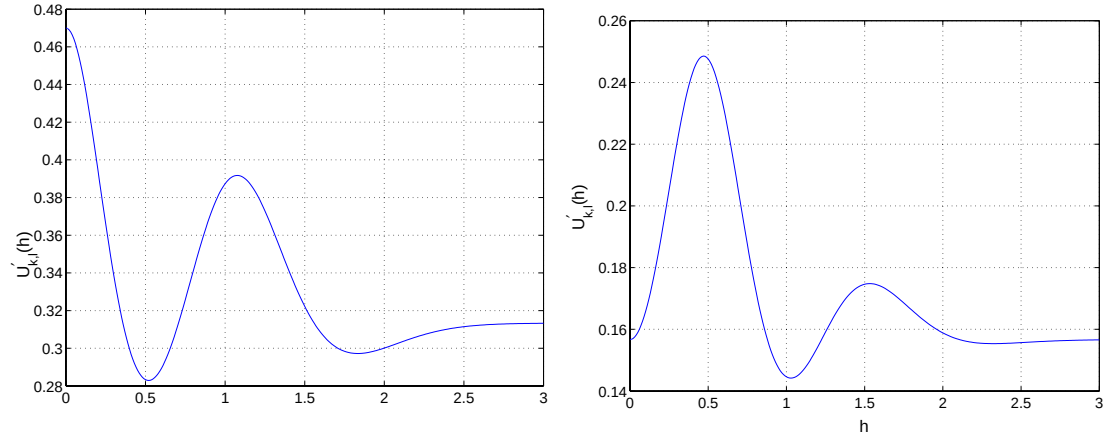
k/l	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$
0	-	0.626657	0.391737	0.174817	0.112790	0.048442	0.031668
1	1.253314	0.626657	0.387584	0.303571	0.130523	0.107172	0.043911
2	1.331723	0.700913	0.341342	0.254392	0.209921	0.095644	0.084184
3	1.271466	0.773564	0.411056	0.199112	0.160010	0.134438	0.065013
4	1.287772	0.678848	0.444918	0.238342	0.116714	0.096769	0.081992
5	1.262590	0.694663	0.369806	0.251937	0.135935	0.067601	0.056820
6	1.265399	0.650985	0.377930	0.202094	0.140741	0.076403	0.038590

Table 5.4.: Value of the modified imse-update (5.6) in the second optimum, $d = 1$)

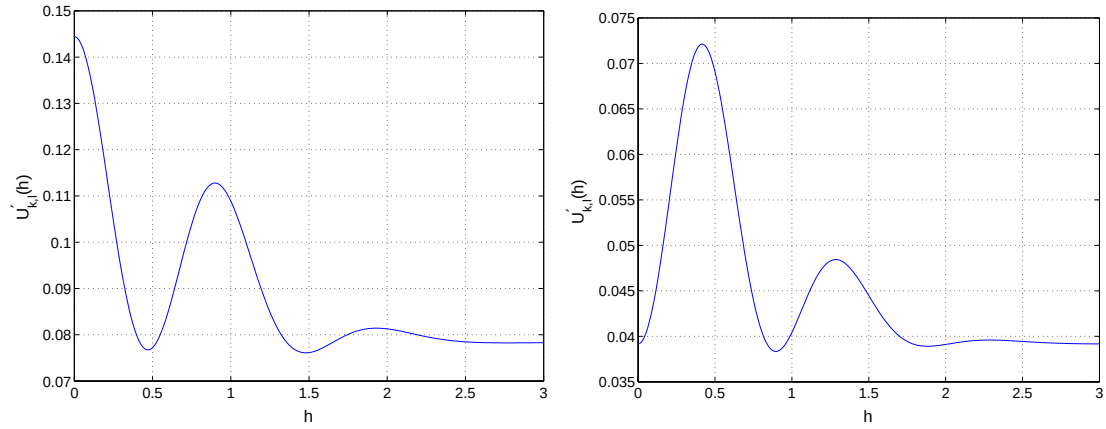
However, although we already know that derivatives of higher order improve the imse-update criterion only very little, a good design can help use this minor information efficiently. For instance, on Figure 5.2(f) the imse-update (5.6) for uncorrelated observations is about 0.02. If we construct the appropriate design optimally, the resulting imse-update (5.6) can be more than four times improved. Despite this improvement, as the order of derivatives increases, the overall performance of the design, characterized by the limiting imse-update, decreases geometrically.



(a) $k = 0$, $l = 0$ (b) $k = 0$, $l = 1$

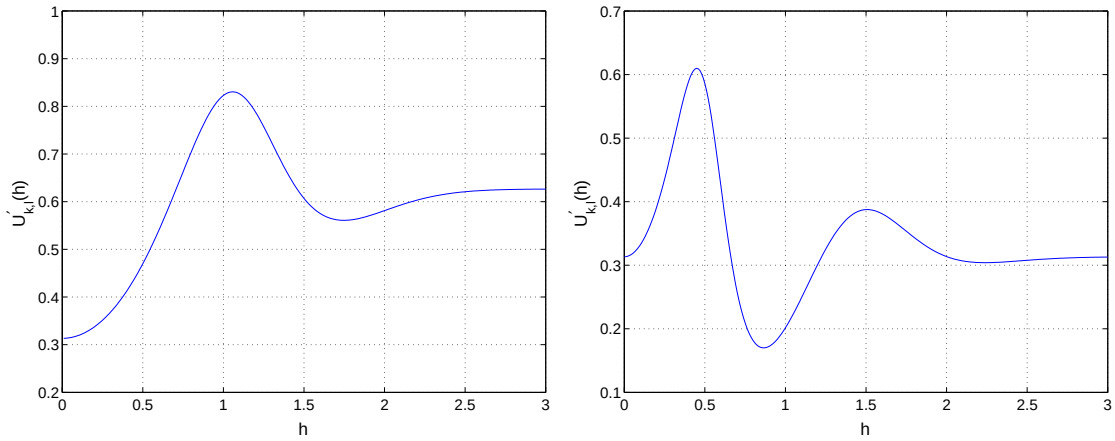


(c) $k = 0$, $l = 2$ (d) $k = 0$, $l = 3$

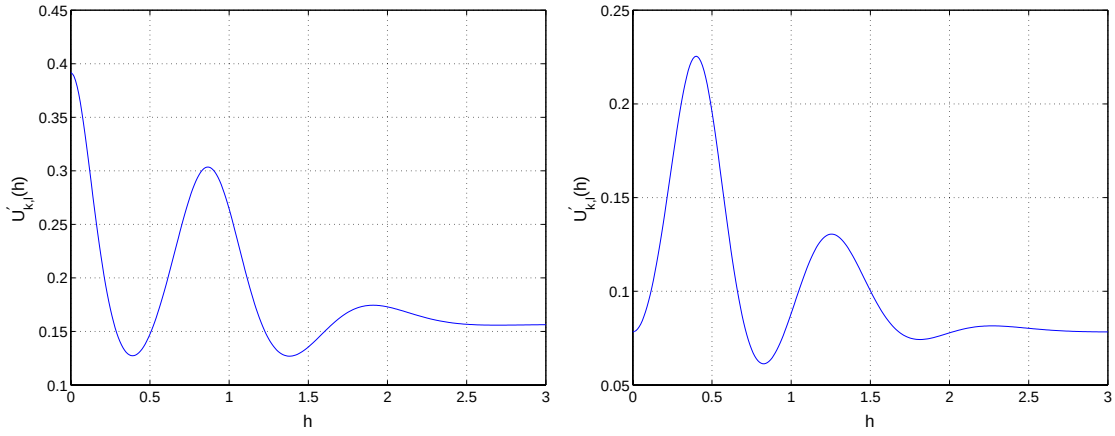


(e) $k = 0$, $l = 4$ (f) $k = 0$, $l = 5$

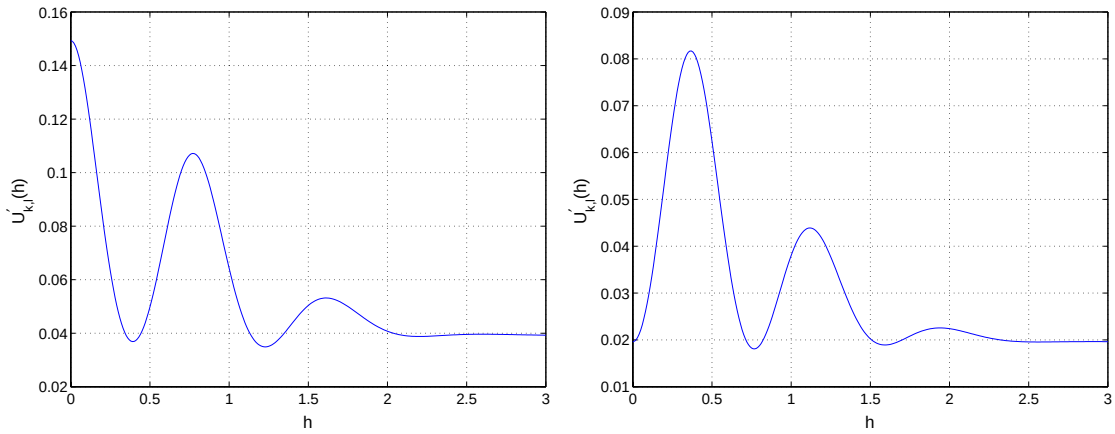
Figure 5.1.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$.



(a) $k = 1$, $l = 1$ (b) $k = 1$, $l = 2$

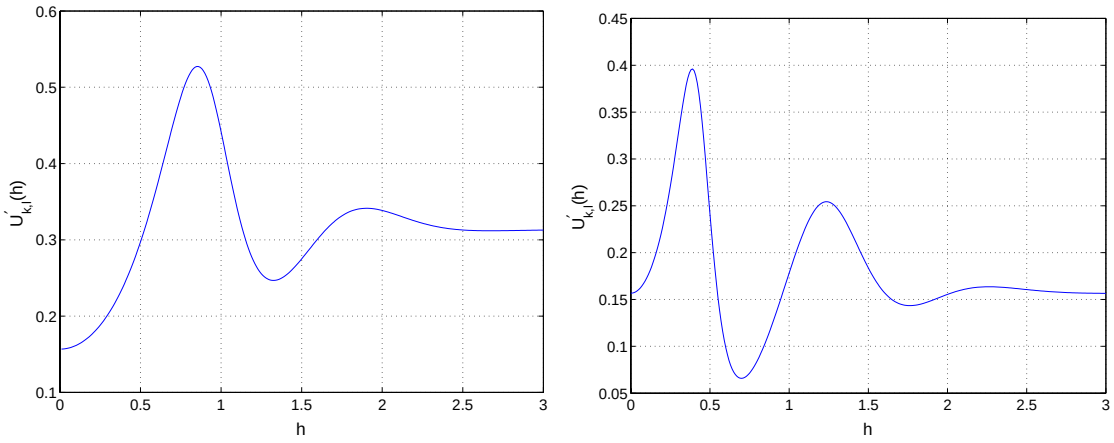


(c) $k = 1$, $l = 3$ (d) $k = 1$, $l = 4$

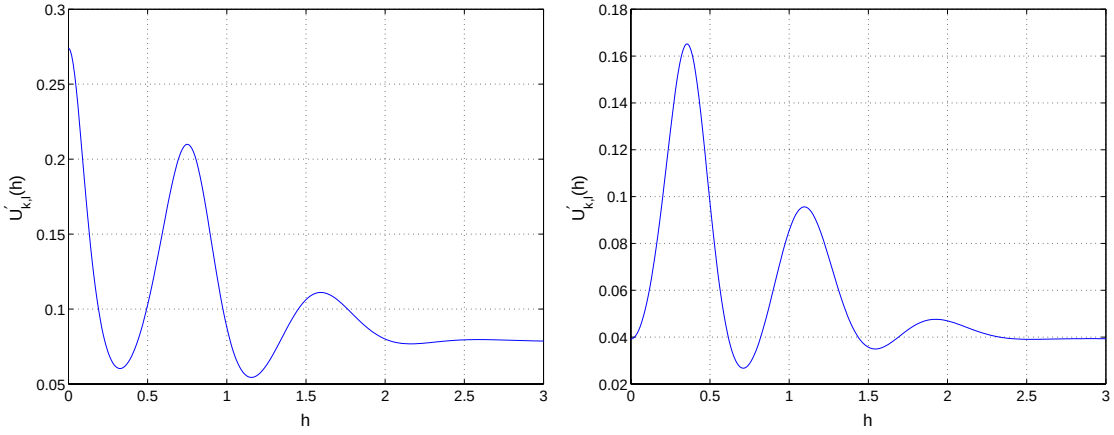


(e) $k = 1$, $l = 5$ (f) $k = 1$, $l = 6$

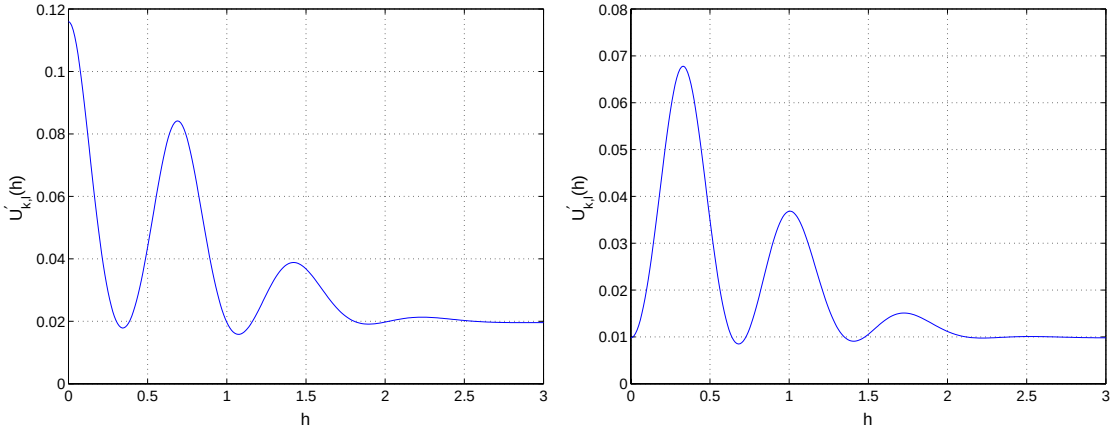
Figure 5.2.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$.



(a) $k = 2$, $l = 2$ (b) $k = 2$, $l = 3$

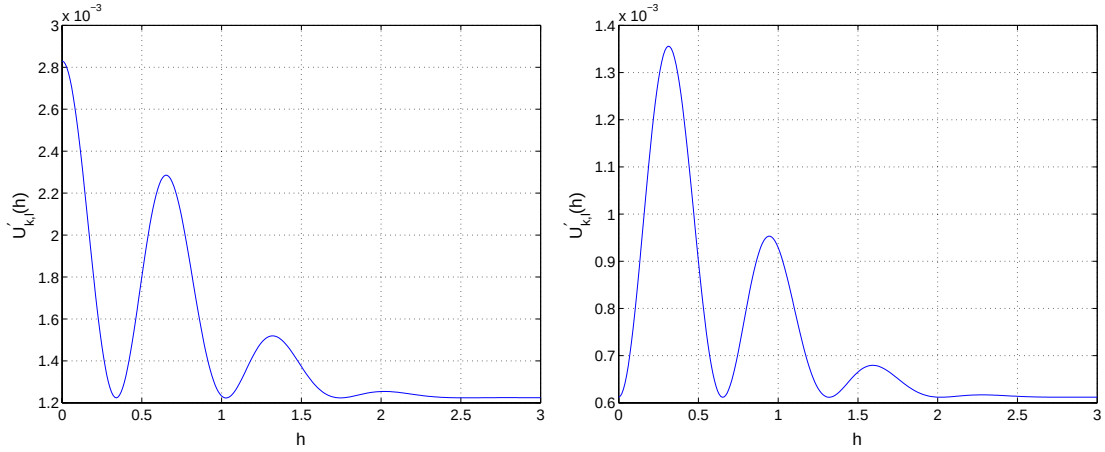


(c) $k = 2$, $l = 4$ (d) $k = 2$, $l = 5$



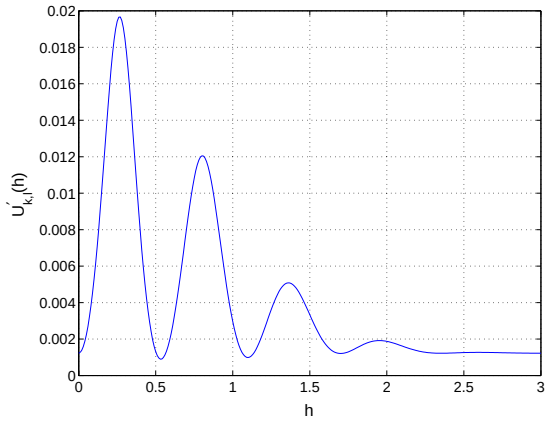
(e) $k = 2$, $l = 6$ (f) $k = 2$, $l = 7$

Figure 5.3.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$.

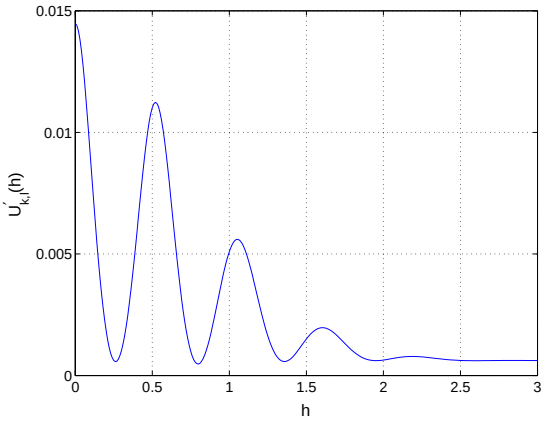


(a) $k = 0$, $l = 10$

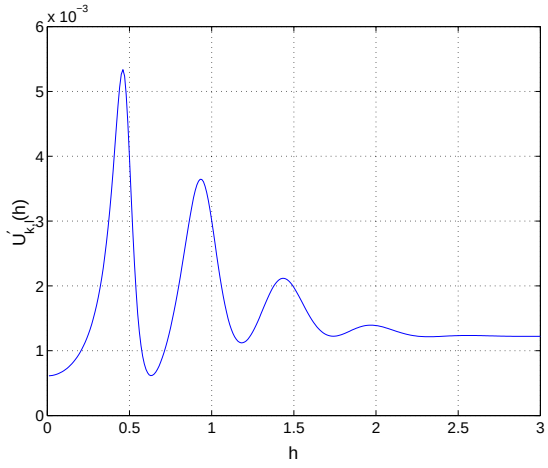
(b) $k = 0$, $l = 11$



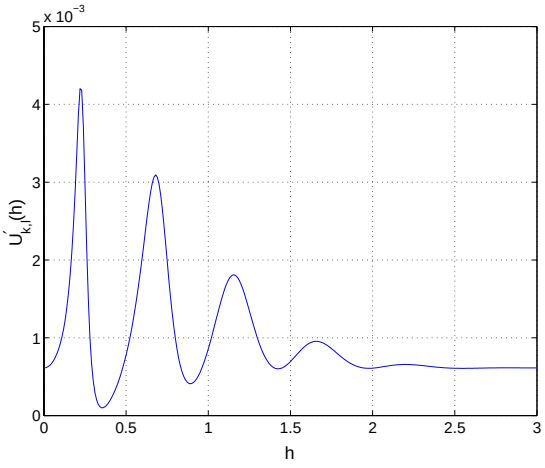
(c) $k = 5$, $l = 10$



(d) $k = 5$, $l = 11$



(e) $k = 10$, $l = 10$



(f) $k = 10$, $l = 11$

Figure 5.4.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$.

Relations between 2-points designs

Now, let us focus on Figures 5.5 – 5.7. Figure 5.5 is nothing else than all sub-figures of the figure 5.1, which are put together. Similarly, Figure 5.6 and 5.7 for Figure 5.2 and 5.3, respectively. It can be seen from Figure 5.5 that in a neighborhood of the origin (where $Z^{(0)}$ is observed) the additional observation of $Z^{(1)}$ performs better than observation of $Z^{(0)}$ observed at the same location. From Figure 5.6 an analogous situation can be seen. Moreover, a small neighborhood of the origin exists, where even an observation of $Z^{(3)}$ performs better than both $Z^{(1)}$ and $Z^{(2)}$. The same can be seen on Figure 5.7. On the basis of the presented figures a lot of relations between diverse 2-points designs can be established.

This could seem interesting in the connection to Theorem 5, but it can not be excepted that this is true in a general case. Putting two observations of $Z^{(0)}$ close to each other means actually to construct a bad design. Thus, it is no longer surprising that a good design of two 'less informative' observations can outperform a bad design of 'more informative' observations. Here, 'less' or 'more informative' is considered by means of the limiting imse-update. It can probably be constructed examples which consist of more than two observations. However, if the observations are spread over a large experimental area so, that correlations between the observations are small, the influence deduced from Theorem 5 will dominate. In this case, derivatives will be mostly much less informative than direct observations of the field to predict.

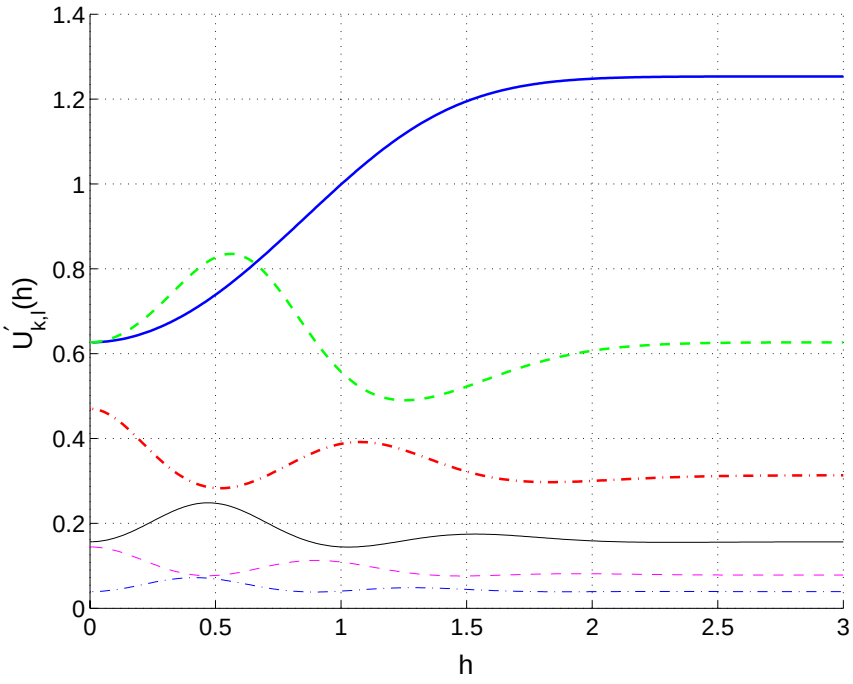


Figure 5.5.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$. Here, $k = 0$, and $l = 0$ (bold solid line), $l = 1$ (bold dashed line), $l = 2$ (bold dash-dot line), $l = 3$ (thin solid line), $l = 4$ (thin dashed line), $l = 5$ (thin dash-dot line).

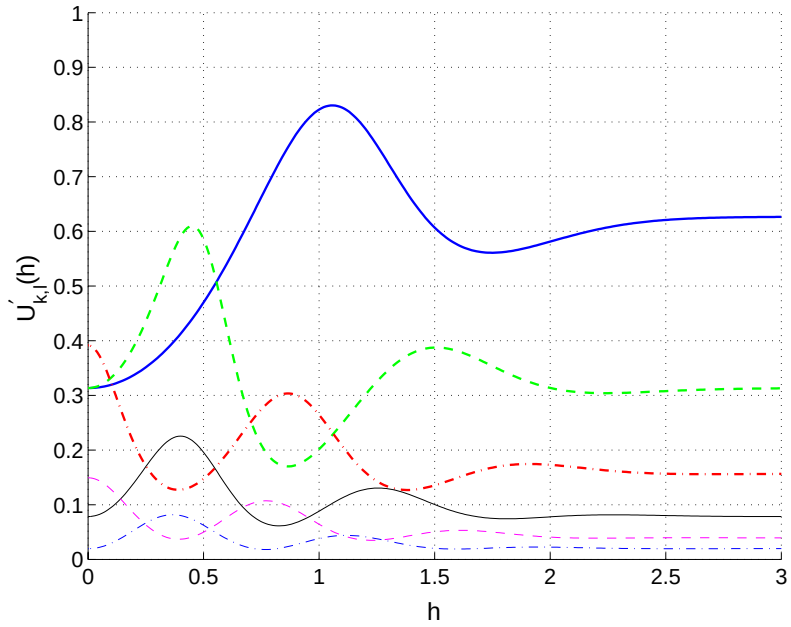


Figure 5.6.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$. Here, $k = 1$, and $l = 1$ (bold full line), $l = 2$ (bold dashed line), $l = 3$ (bold dash-dot line), $l = 4$ (thin solid line), $l = 5$ (thin dashed line), $l = 6$ (thin dash-dot line).

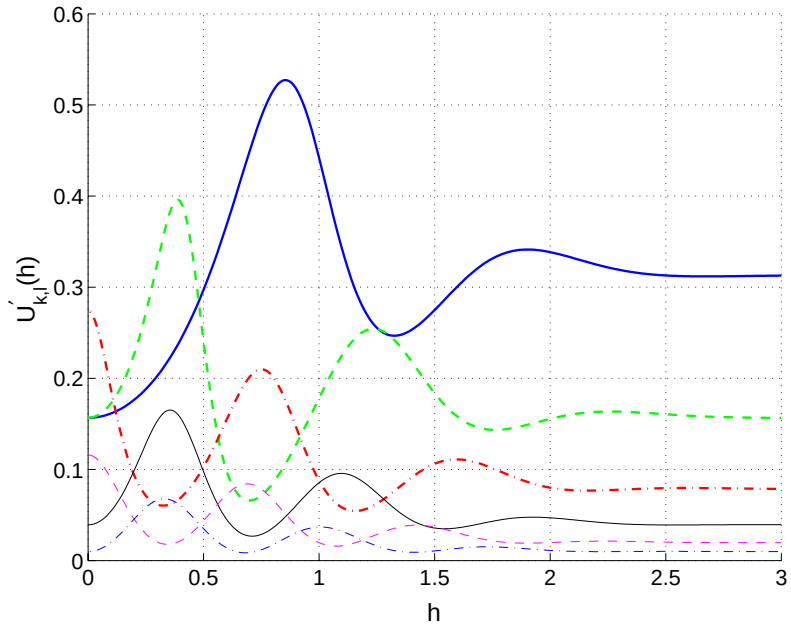


Figure 5.7.: Plots of the imse-update (5.6) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Gaussian model (3.15) with $\alpha = 1$ and $\sigma^2 = 1$ and under the stationarity assumption. Here, $k = 1$, and $l = 1$ (bold solid line), $l = 2$ (bold dashed line), $l = 3$ (bold dash-dot line), $l = 4$ (thin solid line), $l = 5$ (thin dashed line), $l = 6$ (thin dash-dot line).

5.3. Designs with three observations

For designs with three observations, it is not difficult to apply (5.2) and evaluate the imse-update criterion on a very fine grid. Without loss of generality, we can suppose that $Z^{(k)}$ has been observed in the origin and then vary with distances h_1 and h_2 corresponding to observations of $Z^{(l)}$ and $Z^{(m)}$. Since we have two grades of freedom, namely h_1 and h_2, the results are 3-dimensional sceneries. I believe that it can be useful for engineers to visualize such sceneries from different angles and using different visualization techniques. It can be seen many effects that can be interpreted on the basis of experience with only 2-points designs. Nevertheless, in the text, I have no possibility to rotate such figures and therefore I restrict myself only to one example: $k=1$, $l=1$, $m=1$. The experimental area X is the interval $[-3,3]$ and the criterion is the total imse-update (4.5). Figures 5.8 and 5.9 display the surface and the contour plot regarding to this design.

In these figures, it can be seen that certain local optima can be achieved if the observations are related pairwise optimally. This suboptimal solutions build something what can be called "ridges" in the surface. On the basis of symmetry, they can be observed in the horizontal and vertical direction. On the other hand, we can also see configuration, which are bad (some kind of "valleys"), where two observations are pairwise non-optimally located.

However, the reader should be aware just to sum contributions to the imse-update pairwise. It can work good for some simple examples and it could be a good guess sometimes, but it can be probably found examples where it mislead to wrong conclusions.

The main finding from this example is that the optimal (or nearly optimal) solution is a design for which the observations are regularly spaced. The minimal distance between any two observations is only slightly smaller than the optimal distance for 2-points design (see Table 5.1). This is not too surprising because of the symmetry of the problem, however, similar result can be shown also for a 4-points design (when evaluating all possible configurations on a fine grid).

A very similar situation can be seen for designs that consist of derivatives of higher order, like $k=2$, $l=2$, $m=2$ etc. If the design consists of derivatives of diverse order, the situation seems more complicated but in many cases similar tendencies can be observed. For instance, the optimal distances between two observations of different order are close to the corresponding optimal distance in Table 5.1. Some suboptimal regions which correspond the second optimal distance can be seen too.

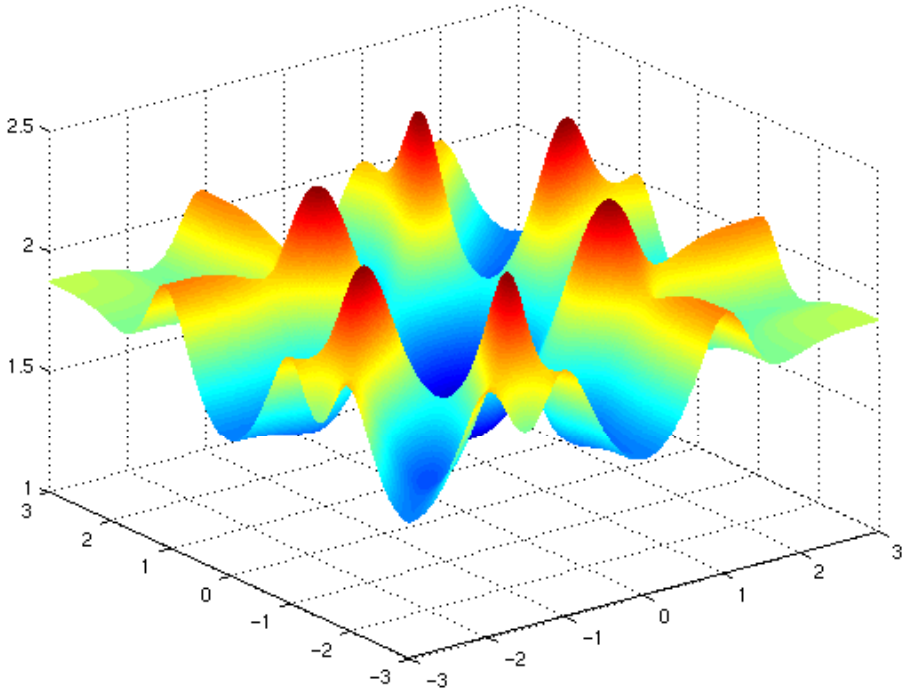


Figure 5.8.: Surface plot: $k = 1$, $l = 1$, $m = 1$

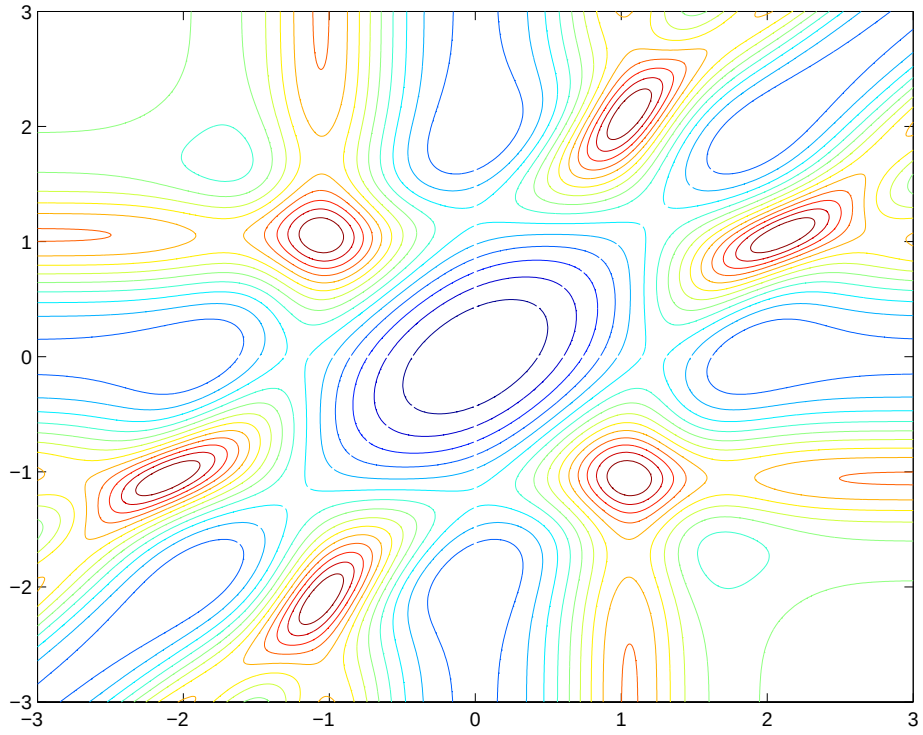


Figure 5.9.: Contour plot: $k = 0$, $l = 1$, $m = 1$

5.4. Regular grids with derivatives

A *regular grid* in $\mathbb{R}^1$

$$V_n^{(k)}(h) = \{(x,k),(x+h,k),(x+2h,k),\ldots,(x+(n-1)h,k)\} \tag{5.7}$$

is clearly characterized by the number $n \in \mathbb{N}$ of derivative observations in the design (grid), the order of the derivatives $k \in \mathbb{N}_0$, the location of the first observation $x \in \mathbb{R}$ and inter-site distance $h \in \mathbb{R}$. Clearly, generalizations to higher dimension, like square, hexagonal and triangular grids in $\mathbb{R}^d$ etc., can be considered, but we concentrate here only to the one-dimensional case. Since stationary processes are supposed and the criterion is the imse-update, x can be chosen arbitrarily.

Some extensive numerical case studies, which were done within the scope of this work, shows that there is some evidence for the conjecture that regularly spaced observations of derivatives of the same order could be nearly optimal designs. Let us refer to the following paragraph just as sketch of realized case studies.

Numerical methods

Typically, results obtained using an adaptation of the fast cooling algorithm of Sacks and Schiller (1988) leads for $k \geq 1$ to configurations that are similar to a regular grid. (In the reality these designs are rather far from the optimal ones.)

The quasi-Newton methods (Fletcher, 1991), when optimizing all locations simultaneously, return nearly regular grids, too. We will refer this method to $qN(\mathbb{R}^{n-1})$ for short, since the minimization proceeds in $(n-1)$-dimensional space. Clearly, for a n-points design, one location can be fixed and it remains $n-1$ sites to optimize. In simple cases, like 5-points designs, these methods are superior to them when we restrict ourselves only to regular grids. However, multiple start and a very good solution approximation are necessary to obtain good results. Due to the discussed high sensitivity, it is a serious problem. For bigger examples some numerical problems can appear (Zabinsky, 1998) and regular grids seems to be more practical solution. The quasi-Newton method restricted to the family of regular grids will be denoted as $qN(\mathrm{Reg})$. This optimization proceeds in $\mathbb{R}$ only, since only the optimal value for h is required.

Example 5.2 For instance, after several starts of the $qN(\mathbb{R}^4)$ method, the following non-regularly spaced solution for 5-points design of the first derivatives has been found

$$U(V_5) = 4.10035939815226.$$

$$V_5^{(1)} = (0,\ 1.04523377697851,\ 2.08507102139555,\ 3.12490825722185,\ 4.17014208577586).$$

The $qN(\mathrm{Reg})$ has found a slightly worse solution:

$$U(V_5^{(1)}(h^*)) = 4.10023769349602 \text{ for the optimal distance } h^* = 1.04243575854999,$$

it means the design

$$V_5^{(1)}(h^*) = (0, 1.04243575854999, 2.08487151709998, 3.12730727564997, 4.16974303419996).$$

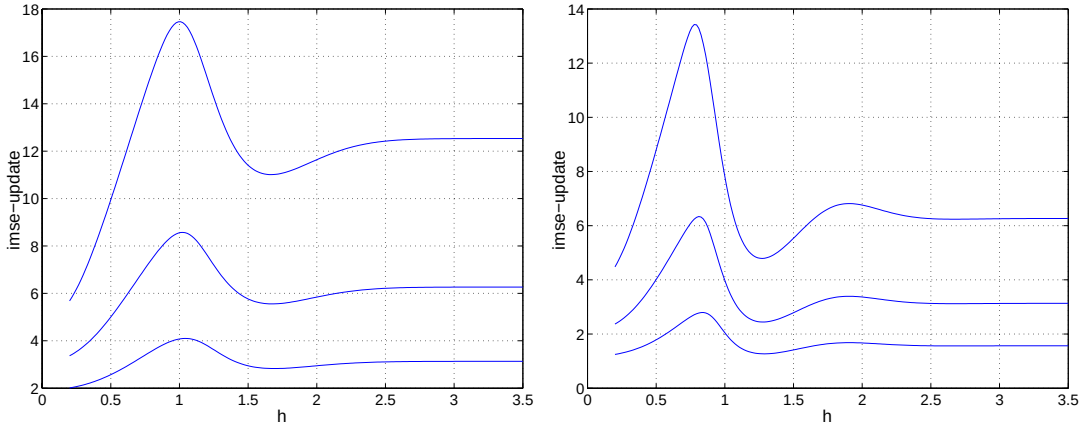
The plots of the imse-update (4.5) for regular grids are presented in Figure 5.10. It can be seen, that for the optimal value of h, the designs significantly outperform other ones. The optimal value of the h depends on n and it shorten as n increases. The estimated optimal distances h^* are reported in Table 5.5. Table 5.6 gives values of the imse-update (4.5) in this optimum. The first line of Table 5.6 is the degenerated case $n = 1$.

n	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
2	1.0576376	0.8545911	0.7428092	0.6678002	0.6125201	0.5694222
3	1.0547248	0.8498840	0.7380173	0.6633214	0.6084289	0.5657049
4	1.0482275	0.8433451	0.7321835	0.6581949	0.6039120	0.5616937
5	1.0424358	0.8372434	0.7265776	0.6531787	0.5994373	0.5576928
6	1.0373001	0.8316147	0.7212908	0.6483774	0.5951125	0.5537896
7	1.0327716	0.8264487	0.7163357	0.6438178	0.5909610	0.5500213
8	1.0287509	0.8216894	0.7116852	0.6394909	0.5869936	0.5463964
9	1.0251556	0.8172876	0.7073114	0.6353810	0.5832013	0.5429147
10	1.0219185	0.8132000	0.7031877	0.6314730	0.5795748	0.5395715
12	1.0163156	0.8058309	0.6956029	0.6242048	0.5727828	0.5332787
15	1.0095584	0.7964125	0.6856249	0.6144963	0.5636256	0.5247362
20	1.0012605	0.7839620	0.6719093	0.6008939	0.5506745	0.5125437
30	0.9911218	0.7671941	0.6520078	0.5805576	0.5310839	0.4949739
40	0.9857393	0.7578321	0.6387049	0.5661452	0.5167842	0.4802243
60	0.9816624	0.7539208	0.6273366	0.5476248	0.4980717	0.4613469

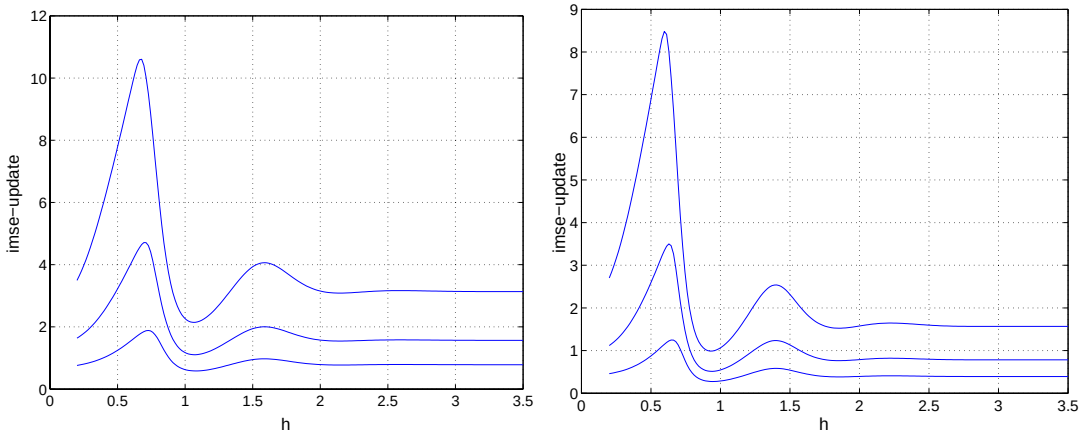
Table 5.5.: The optimal distance h^* of a regular grid in $\mathbb{R}^1$ for the Gaussian covariance function ($\alpha = 1$, $\sigma^2 = 1$).

n	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
1	0.626657	0.313329	0.156664	0.078332	0.039166	0.019583
2	1.457065	0.840643	0.474960	0.264264	0.145339	0.079206
3	2.324733	1.454331	0.891186	0.535993	0.317306	0.185345
4	3.209134	2.112792	1.368200	0.870388	0.544698	0.335905
5	4.100238	2.795761	1.884171	1.250129	0.816641	0.525720
6	4.994273	3.492977	2.425912	1.662919	1.123874	0.749060
7	5.889356	4.198572	2.985009	2.099998	1.458946	1.000560
8	6.784531	4.909026	3.555913	2.555076	1.816003	1.275546
9	7.679279	5.622118	4.134847	3.023564	2.190470	1.570069
10	8.573318	6.336409	4.719177	3.502077	2.578770	1.880845
12	10.358744	7.765085	5.897090	4.479696	3.386250	2.540799
15	13.029721	9.901488	7.671905	5.975249	4.647915	3.598822
20	17.463428	13.433647	10.620439	8.494665	6.818550	5.468202
30	26.279054	20.391302	16.431451	13.518818	11.235556	9.373333
40	35.049742	27.245186	22.133874	18.478064	15.650498	13.349062
60	52.526997	40.831109	33.342335	28.216358	24.376994	21.289246

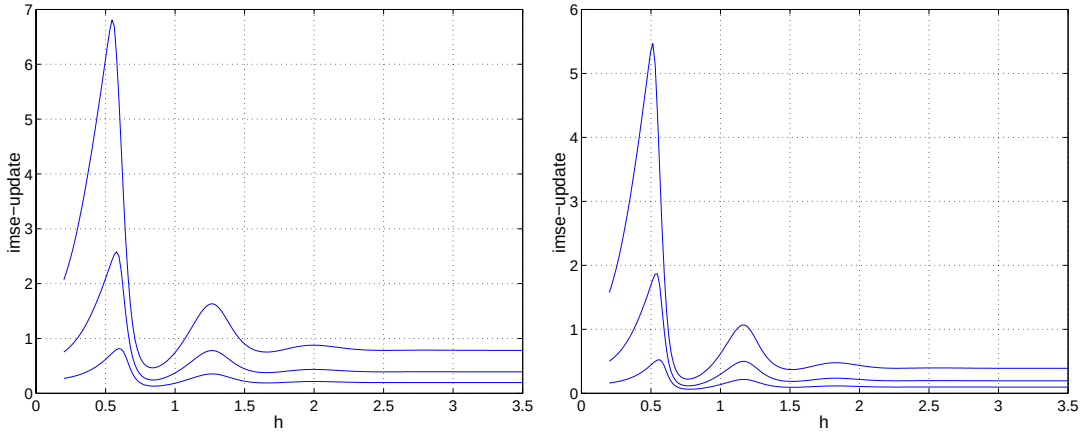
Table 5.6.: The value of the imse-update $U(V_n^{(k)}(h^*))$ for the regular grid $V_n^{(k)}(h^*)$ in $\mathbb{R}^1$ with the corresponding h^* from Table 5.5



(a) $k = 1$ (b) $k = 2$



(c) $k = 3$ (d) $k = 4$



(e) $k = 5$ (f) $k = 6$

Figure 5.10.: Plots of the imse-update (4.5) for a regular grid $V_n^k(h)$ under the Gaussian model (3.15) ($\alpha = 1$, $\sigma^2 = 1$). The number of observations (from the top) is in all cases $n = 20, 10, 5$.

5.4.1. Relative imse-update

If the number of observations is not predefined, it is useful to define *the relative (total) imse-update*

$$U_{\mathrm{rel}}(V_n) = \frac{1}{n}\, U(V_n) \tag{5.8}$$

and compare designs with respect to this criterion. As usual, the aim is to maximize this criterion. The plots of the relative imse-update $U_{\mathrm{rel}}(V_n^k(h^*))$ for regular grids, given the optimal choice h^* (taken from Table 5.5), are presented in Figure 5.11. Note that h^* is a function of k and n. The relative imse-update $U_{\mathrm{rel}}(V_n^k(h^*))$ increases rapidly in all cases for small number of observations till some saturation is reached, where the criterion (5.8) remains nearly constant. It is, however, not clear whether (5.8) is monotonous.

An explanation for these effects could be the fact that first (global) and second (next local) optimum of the pair-wise computed imse-update $U_{k,k}(h)$ and the other optima (if they exist) are more or less regularly spaced (see Figures 5.1 – 5.4, Table 5.1 and Table 5.3). Designs reported in Table 5.5 seem to take advantage of this property. Optimally constructed regular grids have observations located so, that they are efficiently related to the nearest next observation but they are moreover very good related to the second next observation etc. Figures 5.11 and 5.10 demonstrate unique power of these designs. It is reasonable to expect that these designs can be further improved optimizing design points on the border.

For higher derivatives, the saturated state is reached for bigger n than for lower derivatives. This can be explained from the fact that higher derivatives have covariance function with heavier tails. This issue is in particular discussed in Näther and Šimák (2001).

Note that the minimal eigenvalue of $\mathbf{\Sigma}$ (considered as a function of h) has a local minimum close to h^*. This could be an alternative explanation of these effects.

5.4.2. Imse-update ratio for optimal regular grids

Consider the optimal choice of h in a regular grids and suppose that the number of observations in the design is large and the saturated state is reached. Now, we can ask what is the relative performance of such designs for diverse derivatives. It means, the following ratio is to examine:

$$R_{k/l} = \frac{U(V_n^k(h^*(n,k)))}{U(V_n^l(h^*(n,l)))}, \tag{5.9}$$

for a large $n \in \mathbb{N}$. For $k = 0$, the relative imse-update $U(V_n^k(h^*(n,k)))$ can be defined using theoretical value of the limiting imse-update, since $h^*(n,0) \to \infty$. It means that $U_{\mathrm{rel}}(V_n^0(h^*(n,0))) = \frac{1}{\alpha}\sqrt{\frac{\pi}{2}}$

In the above equation, the dependence of h^* on k (resp. l) is emphasized. Note that analogous ratio is studied in Theorem 5 for designs with vanishing inter-site correlations. It is interesting that (5.9) shows rather different behavior from the ratio in Theorem 5. Let us give some examples

$$R_{1/0} \approx 0.6969,\ R_{2/1} \approx 0.7777,\ R_{3/2} \approx 0.8205,\ R_{4/3} \approx 0.8557,\ R_{5/4} \approx 0.8791. \tag{5.10}$$

Thus, although the derivatives of higher order are less informative comparing to the derivatives of lower order (if they are designed optimally on a regular grid), the decrease of the imse-update is much slower than in the case of designs with vanishing inter-site correlations. Recall that for designs with vanishing inter-site correlations, all the above rates are

equal to 1/2. Note also, that two generally non-optimal designs are compared in Theorem 5. In (5.10), such designs are compared which seems rather close to the optimal ones and therefore (5.10) is perhaps of more interest. Notice that the ratio $R_{k+1/k}$ increases as k increases. It means that the difference of the relative performance between derivatives decreases as the order of the derivatives becomes larger. Let us note that the values in (5.10) can vary a little bit for different values of n. They are moreover very sensitive to the estimated value of $h^*(n, \cdot)$.

Here, as the reader can see, we have confined ourselves only to empirical results. From the theoretical point of view, there are more questions than answers. A possible question can be the limiting behavior of regular grids. For example, what are the values of

$$\lim_{n\to\infty} h^*(n,k), \qquad \lim_{n\to\infty} U_{\mathrm{rel}}(V_n^k(h^*(n,k))), \qquad \lim_{k\to\infty} R_{k+1/k}$$

to mention the most obvious ones.

Note to a fast cooling algorithm

In the context of practical methods in the optimal design, a clone of the simulated annealing was proposed by Sacks and Schiller (1988). In Sacks et al. (1989), this algorithm is cited and it is noted that for larger problems the time taken for the annealing process to converge to the optimum is too long. For continuous regions, the authors recommends standard optimization methods (like the quasi-Newton method). Note that in Sacks and Schiller (1988), this algorithm is used for optimization of designs on rather small grids of about 100 grid points.

Using the terminology of Winkler (1995), the algorithm of Sacks and Schiller (1988) can be referred to as a fast cooling with some kind of re-heating mechanism. This algorithm is characterized by several parameters with rather unclear meaning and there is practically no reasonable answer how they should be chosen. However, the proper choice of parameters of a particular algorithm is one of the most important problems in the stochastic relaxation, since they have the critical influence to the convergence of this algorithm. Furthermore, for algorithms like the one of Sacks and Schiller (1988) a proof of convergence is not known.

This algorithm was adopted by Albrecht (1998) for an optimal planing of gradients (2.10) in $[a,b]^2 \subset \mathbb{R}^2$. He used the maximum mean square error (2.23) as a criterion of optimality. However, especially in the connection with the imse-update criterion, I would not recommend the algorithm of Sacks & Schiller for several reasons. One of them is, the absence of the theoretical background supporting this choice. Another point is the structure of the space that need to be explored. It seems that the space of all solutions contains many local minima which are close to each other by means of the criterion value, but the transition from one to another is very unlikely. Typically, the algorithm reaches to one state where all derivative observations build a design that is visually similar to a regular grid, but the distances are not optimal. It means that the rough suboptimal structure of the design has been found. The problem is just to optimize the 'parameters' of this structure, i.e. to find the optimal distances. Unfortunately, a transition to the state where they are reached can be very difficult. In another words, two designs with a small distance in the physical space (see e.g. the designs in Example 5.2) can have a big distance in the space of all solutions by means of an annealing algorithm.

It seems that the problem of optimal design for covariance function models which allows negative correlations between observations is of another nature and more complicated than

optimal design problem with non-negative covariance functions. The former is the case of designs with derivative observations.

Finally, it is possible that some modifications of the annealing which uses a local gradient information could give better results although I would not expect that they will be optimal (see Robert and Casella (1999, Chapter 5) and Zabinsky (1998) for some tips).

5.4.3. Experimental conditions, heuristic rules and recommendations

Let us summarize some of the heuristic rules which can be implied from this text. Let us emphasize that it is not easy to formulate these rules, since any change of experimental conditions can lead to another recommendations. For instance, the assumption that designs possess vanishing inter-site correlations seems to be rather strong. These designs constitute certain class of problems and not all conclusions, which hold for these design, hold for other ones. The influence of the change of the experimental conditions to the construction of the optimal design is demonstrated on the following example.

Example 5.3 Suppose that stationary zero mean stochastic processes $Z^{(0)}$, $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$ can be observed only once and $Z^{(0)}$ has the Gaussian covariance function ($\alpha = 1$ and $\sigma^2 = 1$).

1. At first, let the observations area consists of points, which can be considered as uncorrelated and suppose that at most one process can be observed in each locations. It means, we have the setup of vanishing inter-site correlations (and with no intra-site correlations) as discussed in the first part of Chapter 4. In this case, all design are equivalent and the imse-update is

$$U_1 = \sigma^2 \frac{1}{\alpha} \sqrt{\frac{\pi}{2}} (2^0 + 2^{-1} + 2^{-2} + 2^{-3}) \doteq \underline{2.349964}.$$

2. In the next step, multiple observations in a location are possible (i.e. intra-site correlations are allowed), but the inter-site correlations vanish as in the previous case. This setup is discussed in the second part of Chapter 4. In particular, the results of Example 4.5 yields that the best design is of the type D or D', i.e. that $Z^{(0)}$ and $Z^{(2)}$ have to be observed together in one point and the same must be fulfilled for $Z^{(1)}$ and $Z^{(3)}$. In this case, the imse-update is

$$U_2 = \sigma^2 \frac{1}{\alpha} \sqrt{\frac{\pi}{2}}\, 2\frac{3}{16} \doteq \underline{2.741625}.$$

3. In the following, let us continue developing Example 4.5 and ask what happens if the assumption of vanishing inter-site correlations will be left, but the designs will be like in Table 4.1. It means, we restrict ourselves only to two observation locations, however, unlike Example 4.5, we seek moreover for the optimal distance $h = y - x$. The hope is that the design D remains the best and it can be further improved by the optimal choice of h. The numerical results are summarized in Table 5.7.

Design	Observation(s) at x	Observation(s) at y	$U(h^*)$	$h^* = y - x$
A	$Z^{(0)}$	$Z^{(1)}, Z^{(2)}, Z^{(3)}$	2.859401	0.844992
A'	$Z^{(1)}$	$Z^{(0)}, Z^{(2)}, Z^{(3)}$	2.787016	0.545236
B	$Z^{(2)}$	$Z^{(0)}, Z^{(1)}, Z^{(3)}$	3.050484	0.354003
B'	$Z^{(3)}$	$Z^{(0)}, Z^{(1)}, Z^{(2)}$	2.891803	0.288805
C	$Z^{(0)}, Z^{(1)}$	$Z^{(2)}, Z^{(3)}$	2.741625	0
D	$Z^{(0)}, Z^{(2)}$	$Z^{(1)}, Z^{(3)}$	$\underline{3.087336}$	0.395845
C'	$Z^{(0)}, Z^{(3)}$	$Z^{(1)}, Z^{(2)}$	2.885402	0.291107

Table 5.7.: Optimal designs for two locations

The optimal design in this setting is denoted as D. It can be seen that additional improvement of the design criteria has been achieved.

4. Finally, let us suppose that no constraints are put on the observation area. A numerical solution of this problem has been found the following way. After an exhaustive search over all possible configurations on a fine grid, an approximation of the optimal design has been found. Subsequently, the quasi-Newton method $qN(\mathbb{R}^3)$ yielded the solution (5.11)

$$V_4 = \{(0,0), (0.570657, 1), (1.029996, 2), (1.428088, 3)\}. \tag{5.11}$$

The corresponding imse-update is

$$U(V_4) \doteq \underline{3.356293}.$$

The example demonstrates that there is no golden rule, which can be used for all situation. It can be seen that different conclusions can be drawn from each formulation of experimental conditions. Passing from the most restrictive case to case with no restrictions, the structure of the optimal design changes in every step. In particular, the design V_4 is completely different from the optimal design found in the previous case.

Some rules

However, some recommendations are briefly summarized in this paragraph. If the question is basically what is the most informative part of a multivariate random field, a setup similar to that in the proof of Theorem 5 can be used to establish some relations between the given parts.

If the observations are only of one type (it can also be observations of a multivariate random field) and the imse-update computed pairwise has a maxima not in the infinity, there is a possibility that well constructed regular grid will be rather efficient. However, whether it is true, it could also depend on the properties of the particular covariance function under study and in the consequence on the fact, whether the maxima of the pairwise computed imse-update are regularly spaced. If it is not true, some more complicated chains (or grids) of observations could work better.

If the pairwise imse-update has a maxima in the infinity, the maximin design approach seems to me the right way how to solve such a problem (see page 43 for some references).

If the number of observations is given for every type of observation, some kind of a sequential procedure can be a good solution. This procedure should start with the most

informative observations, find a good solutions for this smaller design problem and then follow with the next most informative component etc. In this case, tables analogous to Tables 5.1 - 5.4 could contain good hints, where to search a solution of every step of such a sequential procedure.

Another approach can be to adopt an "architecture" of a grid (like triangle or square grid) and to assume that there exists an optimal distance between two connected observations of certain type. Thus, the task remains to find the parameters of the grid, i.e. these distances. This approach is similar to the case, where all observations are of one type and we restrict ourselves only to regular grids, i.e. we have to find one parameter of a grid. Here, the situation can be much more complicated since many types of observations can be considered.

5.5. Further research

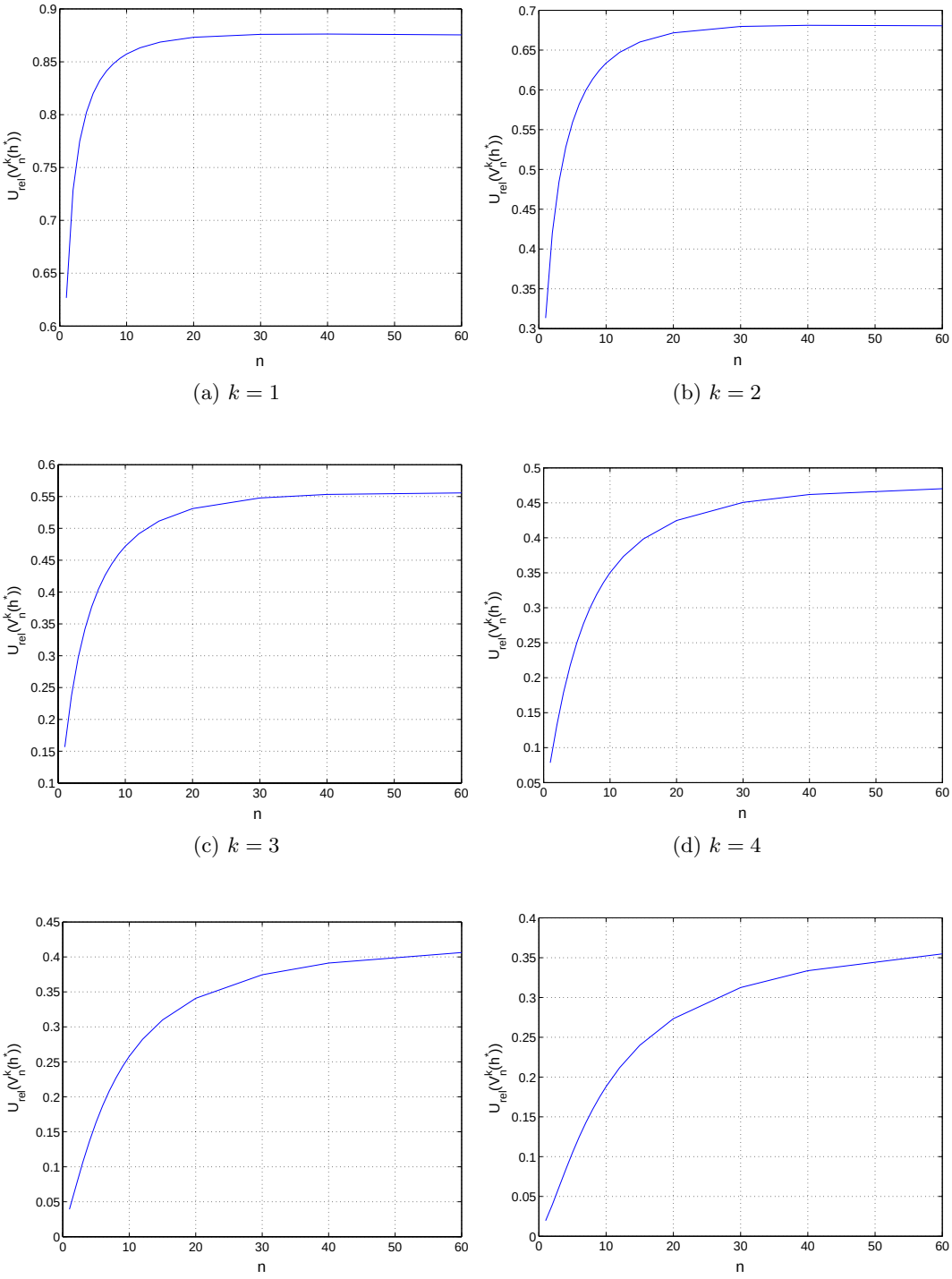
There are surely many possibilities how to continue research connected to the optimal design of derivative observations. One of them is obviously another choice of the objective function. In many application, other criteria can be more appropriate. The influence of a finite prediction area and its shape to the design has not been studied in this work, although it has big impact on the results.

However, in the context of this work, it can be interesting to adopt approach of perturbed covariance matrices (Holtmann, 2001) and adapt this method to the imse-update criterion. Because of the linear nature of the imse-update I think it is possible. Hopefully, it could be a proper way how to throw light on some effects observed in this chapter.

Interesting task could be a comparison of designs proposed in this work with the method suggested by Batsell et al. (1998). The method can be used for solving a problem of optimal design with derivative observations too. Note that the prerequisites of cited method are different from that assumed here. Specifically, we do not allow measurement error (see (2.19)), which is necessary for the approach of Batsell et al. (1998).

Another direction could be a more detailed investigation of properties of different covariance functions by means of Theorem 5 and by means of this chapter. In particular, a comparison of some classical covariance functions with the covariance functions proposed by Wendland (1995); Gaspari and Cohn (1999) and Gneiting (2000) can be interesting.

It is reasonable to suppose that regular grids of derivatives will posses a similar (suboptimal) behavior in $\mathbb{R}^d$, $d > 1$, at least for Gaussian-like or Matérn-like covariance functions. However, there are more possibilities in $\mathbb{R}^d$ how to construct regular grids. We can take square, triangle, hexagonal grid or maybe some other problem-tailored one. A comparison of these possibilities (see Yfantis et al. (1987) in another context), estimation of parameters of the grids under diverse covariance functions etc. could surely be an important reference material for engineers.



(a) $k = 1$ (b) $k = 2$

(c) $k = 3$ (d) $k = 4$

(e) $k = 5$ (f) $k = 6$

Figure 5.11.: Plots of the relative imse-update $U_{\mathrm{rel}}(V_n^k(h^*))$ under the Gaussian model (3.15) ($\alpha = 1$, $\sigma^2 = 1$).

A. Appendix (Proofs)

A.1. Proof of Lemma 2.2

Let us recall the lemma and denote by $\lfloor x \rfloor$ the biggest integer smaller than $x \in \mathbb{R}$.

Lemma A.1 *Let Z be a stationary, isotropic, n-times differentiable random field on $\mathbb{R}^d$ and let $\boldsymbol{k} = (k_1, \dots, k_d)$ and $\boldsymbol{l} = (l_1 \dots, l_d)$ be multi-indices, such that $|\boldsymbol{k}|, |\boldsymbol{l}| \leq n$. Then*

$$\operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x})) = 0, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

if $\boldsymbol{k} + \boldsymbol{l}$ is odd.

Proof : Consider a sufficiently differentiable univariate real function $F : \mathbb{R} \to \mathbb{R}$ and denote

$$F^{(k)}(h^2) = \frac{d^k}{dy^k} F(y)\Big|_{y=h^2}.$$

Thus, it holds (see Lemma A.2, page 83)

$$F^{(n)}(h^2) = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{n!}{(n-2i)!i!} (2h)^{n-2i} F^{(n-i)}(h^2).$$

For multi-indices $\boldsymbol{n} = (n_1, \dots, n_d)$ and $\boldsymbol{i} = (i_1, \dots, i_d)$, define

$$c_{\boldsymbol{i}}^{\boldsymbol{n}} = \frac{n_1!}{(n_1 - 2i_1)!i_1!} \frac{n_2!}{(n_2 - 2i_2)!i_2!} \cdots \frac{n_d!}{(n_d - 2i_d)!i_d!}$$

It can readily be seen that it holds analogously

$$\frac{\partial^{|\boldsymbol{n}|}}{\partial h_1^{n_1} \dots \partial h_d^{n_d}} F(||\boldsymbol{h}||^2) = \sum_{i_1=0}^{\lfloor \frac{n_1}{2} \rfloor} \cdots \sum_{i_d=0}^{\lfloor \frac{n_d}{2} \rfloor} c_{\boldsymbol{i}}^{\boldsymbol{n}} (2h_1)^{n_1 - 2i_1} \dots (2h_d)^{n_d - 2i_d} F^{(|\boldsymbol{n}| - |\boldsymbol{i}|)}(||\boldsymbol{h}||^2) \quad \text{(A.1)}$$

Since Z is stationary and isotropic, its covariance function C can be represented by a function $F : [0, \infty) \to \mathbb{R}$ such that $C(\boldsymbol{x}, \boldsymbol{y}) = F(||\boldsymbol{x} - \boldsymbol{y}||^2)$. Note that the cross-covariance $\operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{y}))$ is (up to a sign) equal to (A.1) for $\boldsymbol{n} = \boldsymbol{k} + \boldsymbol{l}$. Now, let $j \in \mathbb{N}$ be such that i_j is odd. Thus, the right side of (A.1) can be written as

$$2h_j \sum_{i_1=0}^{\lfloor \frac{n_1}{2} \rfloor} \dots \sum_{i_j=0}^{\lfloor \frac{n_j - 1}{2} \rfloor} \dots \sum_{i_d=0}^{\lfloor \frac{n_d}{2} \rfloor} c_{\boldsymbol{i}}^{\boldsymbol{n}} (2h_1)^{n_1 - 2i_1} \dots (2h_j)^{n_j - 1 - 2i_j} \dots (2h_d)^{n_d - 2i_d} F^{(|\boldsymbol{n}| - |\boldsymbol{i}|)}(||\boldsymbol{h}||^2),$$

which is equal to zero in the origin. □

Lemma A.2

$$F^{(n+1)}(x^2) := \frac{d^n}{dx^n} F(x^2) = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{n!}{(n-2i)!i!} (2x)^{n-2i} F^{(n-i)}(x^2), \quad n \in \mathbb{N}. \quad \text{(A.2)}$$

Proof : This lemma can be proved by induction, distinguishing the cases when n is even and odd. For $n = 0$ and $n = 1$, the statement obviously holds.

Let us suppose that n is even, it means $\frac{n}{2} = \lfloor \frac{n}{2} \rfloor = \lfloor \frac{n+1}{2} \rfloor$. Suppose that the claim (A.2) holds. Applying induction step in (A.3), using standard rules for differentiation in (A.4) and shifting the index i in the first sum of (A.5), (A.6) can be obtained. This equation can be re-written as (A.7), which is the claim to prove for $u + 1$ odd.

$$
\begin{aligned}
F^{(n+1)}(x^2) &= \frac{d}{dx} \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{n!}{(n-2i)!i!} (2x)^{n-2i} F^{(n-i)}(x^2) && \text{(A.3)} \\
&= \frac{d}{dx} \sum_{i=0}^{\frac{n}{2}-1} \frac{n!}{(n-2i)!i!} (2x)^{n-2i} F^{(n-i)}(x^2) + \frac{n!}{(n/2)!} \frac{d}{dx} F^{(\frac{n}{2})}(x^2) \\
&= \sum_{i=0}^{\frac{n}{2}-1} \frac{n!}{(n-2i)!i!} 2(n-2i)(2x)^{n-2i-1} F^{(n-i)}(x^2) + \\
&\quad \sum_{i=0}^{\frac{n}{2}-1} \frac{n!}{(n-2i)!i!} (2x)^{n-2i+1} F^{(n+1-i)}(x^2) + \frac{n!}{(n/2)!} 2x\, F^{(\frac{n}{2}+1)}(x^2) && \text{(A.4)} \\
&= \sum_{i=0}^{\frac{n-2}{2}} \frac{2\,n!}{(n-2i-1)!i!} (2x)^{n-2i-1} F^{(n-i)}(x^2) + \\
&\quad \sum_{i=0}^{\frac{n-2}{2}} \frac{n!}{(n-2i)!i!} (2x)^{n+1-2i} F^{(n+1-i)}(x^2) + \frac{n!}{(n/2)!} 2x\, F^{(\frac{n}{2}+1)}(x^2) && \text{(A.5)} \\
&= \sum_{i=1}^{\frac{n}{2}} \frac{2i\,n!}{(n+1-2i)!i!} (2x)^{n+1-2i} F^{(n+1-i)}(x^2) + \\
&\quad \sum_{i=0}^{\frac{n}{2}} \frac{n!}{(n-2i)!i!} (2x)^{n+1-2i} F^{(n+1-i)}(x^2) && \text{(A.6)} \\
&= (2x)^{n+1} F^{(n+1)}(x^2) + \sum_{i=1}^{\frac{n}{2}} \frac{(n+1)!}{(n+1-2i)!i!} (2x)^{n+1-2i} F^{(n+1-i)}(x^2) && \text{(A.7)}
\end{aligned}
$$

Let us suppose that n is odd, it means $\lfloor \frac{n}{2} \rfloor = \lfloor \frac{n-1}{2} \rfloor = \frac{n-1}{2}$. Suppose that (A.2) holds for this n. Application of the induction step in (A.8), using standard rules for differentiation in the next equation and subsequently shifting the index i in the first sum yields (A.9). Finally, (A.9) can be re-written as (A.10), which is the claim to prove for $u + 1$ even.

$$F^{(n+1)}(x^2) = \frac{d}{dx}\sum_{i=0}^{\frac{n-1}{2}} \frac{n!}{(n-2i)!i!}(2x)^{n-2i}F^{(n-i)}(x^2) \tag{A.8}$$

$$= \sum_{i=0}^{\frac{n-1}{2}} \frac{n!}{(n-2i)!i!}\left[2(n-2i)(2x)^{n-2i-1}F^{(n-i)}(x^2) + (2x)^{n-2i+1}F^{(n+1-i)}(x^2)\right]$$

$$= \sum_{i=0}^{\frac{n-1}{2}} \left[\frac{2\,n!}{(n-1-2i)!i!}(2x)^{n-2i-1}F^{(n-i)}(x^2) + \frac{n!}{(n-2i)!i!}(2x)^{n-2i+1}F^{(n+1-i)}(x^2)\right]$$

$$= \sum_{i=1}^{\frac{n-1}{2}} \frac{2i\,n!}{(n+1-2i)!i!}(2x)^{n+1-2i}F^{(n+1-i)}(x^2) + \frac{(n+1)!}{(\frac{n+1}{2})!}F^{(\frac{n+1}{2})}(x^2) +$$

$$(2x)^{n+1}F^{(n+1)}(x^2) + \sum_{i=1}^{\frac{n-1}{2}} \frac{n!}{(n-2i)!i!}(2x)^{n-2i+1}F^{(n+1-i)}(x^2) \tag{A.9}$$

$$= \sum_{i=0}^{\frac{n+1}{2}} \frac{(n+1)!}{(n+1-2i)!i!}(2x)^{n+1-2i}F^{(n+1-i)}(x^2) \tag{A.10}$$

□

A.2. Proof of Lemma 3.1

Lemma A.3 *It holds*

$$\frac{d^u}{dr^u}\left[r^\nu \mathcal{K}_\nu(r)\right] = \sum_{i=0}^{\lfloor \frac{u}{2} \rfloor}(-1)^{u+i}\binom{\lfloor \frac{u}{2} \rfloor}{i}\frac{(2\lfloor \frac{u+1}{2} \rfloor - 1)!!}{(2\lfloor \frac{u+1}{2} \rfloor - 1 - 2i)!!}r^{\nu-i}\mathcal{K}_{\nu-u+i}(r), \quad \textit{for } u \in \mathbb{N}$$

or equivalently

$$\frac{d^u}{dr^u}\left[r^\nu \mathcal{K}_\nu(r)\right] = \sum_{i=0}^{\frac{u}{2}}(-1)^{u+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-1-2i)!!}r^{\nu-i}\mathcal{K}_{\nu-u+i}(r), \quad \textit{for } u \textit{ even} \tag{A.11}$$

$$\frac{d^u}{dr^u}\left[r^\nu \mathcal{K}_\nu(r)\right] = \sum_{i=0}^{\frac{u-1}{2}}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}r^{\nu-i}\mathcal{K}_{\nu-u+i}(r), \quad \textit{for } u \textit{ odd.} \tag{A.12}$$

Proof : The proof can be carried out by the induction over u, distinguishing the cases when u is even and odd. I show the second and the third equation, the first one is just a comprehensive formulation of the claim for a general $u \in \mathbb{N}$.

To shorten the notation, define $f_b^a = r^a\mathcal{K}_b(r)$, $a, b \in \mathbb{R}$, omitting also argument r of the function f_b^a. Using the following well known rule for differentiation of the modified Bessel functions

$$\frac{d}{dz}[z^\nu \mathcal{K}_\nu(z)] = -z^\nu \mathcal{K}_{\nu-1}(z)$$

(see Abramowitz and Stegun (1972), formula 9.6.28), the following simple equation can be shown

$$\begin{aligned}\frac{d}{dr}f_b^a &= \frac{d}{dr}[r^a\mathcal{K}_b(r)] = \frac{d}{dr}[r^{(a-b)}\cdot r^b\mathcal{K}_b(r)]\\ &= (-1)[r^a\mathcal{K}_{b-1}(r) - (a-b)r^{a-1}\mathcal{K}_b(r)]\\ &= (-1)[f_{b-1}^a - (a-b)f_b^{a-1}].\end{aligned} \tag{A.13}$$

Note that, it holds

$$\binom{\frac{u}{2}}{i-1} = \frac{2i}{u-2i+2}\binom{\frac{u}{2}}{i}$$

for an even $u \in \mathbb{N}$ and therefore

$$\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-1-2i)!!} + \binom{\frac{u}{2}}{i-1}\frac{(u-1)!!}{(u+1-2i)!!}(u-2i+2) = \frac{(u+1)!!}{(u+1-2i)!!}\binom{\frac{u}{2}}{i}. \tag{A.14}$$

For $u \in \mathbb{N}$ odd, it holds

$$\binom{\frac{u-1}{2}}{i} = \frac{u+1-2i}{u+1}\binom{\frac{u+1}{2}}{i} \qquad \text{and} \qquad \binom{\frac{u-1}{2}}{i-1} = \frac{2i}{u+1}\binom{\frac{u+1}{2}}{i}. \tag{A.15}$$

Now, the proof can be accomplished. For $u = 0$ and for $u = 1$ the statement holds trivially. Let us suppose that u is even and (A.11) holds. Applying the induction step in (A.16), using (A.13) in (A.17) and by shifting the index i in the second sum of (A.18), (A.19) can be obtained. Finally, by applying (A.14), we get (A.20) which is the claim to prove for $u+1$ odd.

$$\frac{d^{u+1}}{dr^{u+1}}f_\nu^\nu = \frac{d}{dr}\left[\frac{d^u}{dr^u}f_\nu^\nu\right] = \frac{d}{dr}\left[\sum_{i=0}^{\frac{u}{2}}(-1)^{u+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-2i-1)!!}f_{\nu-u+i}^{\nu-i}\right] \tag{A.16}$$

$$= \sum_{i=0}^{\frac{u}{2}}(-1)^{u+1+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-1-2i)!!}\left(f_{\nu-u+i-1}^{\nu-i} - (u-2i)f_{\nu-u+i}^{\nu-i-1}\right) \tag{A.17}$$

$$\begin{aligned}= &(-1)^{u+1}f_{\nu-(u+1)}^\nu + \sum_{i=1}^{\frac{u}{2}}(-1)^{u+1+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-1-2i)!!}f_{\nu-u+i-1}^{\nu-i}\\ &+ \sum_{i=0}^{\frac{u}{2}-1}(-1)^{u+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-1-2i)!!}(u-2i)f_{\nu-u+i}^{\nu-i-1}\end{aligned} \tag{A.18}$$

$$\begin{aligned}= &(-1)^{u+1}f_{\nu-(u+1)}^\nu + \sum_{i=1}^{\frac{u}{2}}(-1)^{u+1+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-1-2i)!!}f_{\nu-u+i-1}^{\nu-i}\\ &+ \sum_{i=1}^{\frac{u}{2}}(-1)^{u+1+i}\binom{\frac{u}{2}}{i-1}\frac{(u-1)!!}{(u+1-2i)!!}(u-2i+2)f_{\nu-(u+1)+i}^{\nu-i}\end{aligned} \tag{A.19}$$

$$= (-1)^{u+1}f_{\nu-(u+1)}^\nu + \sum_{i=1}^{\frac{u}{2}}(-1)^{u+1+i}\binom{\frac{u}{2}}{i}\frac{(u+1)!!}{(u+1-2i)!!}f_{\nu-(u+1)+i}^{\nu-i} \tag{A.20}$$

Let us suppose that u is odd and (A.12) holds. Applying the induction step in (A.21), using (A.13) in (A.22) and by shifting the index i in the second sum of (A.23), (A.24) can

be obtained. Finally, by applying (A.15), we get (A.25) which is the claim to prove for $u+1$ even.

$$\frac{d^{u+1}}{dr^{u+1}}f_\nu^\nu = \frac{d}{dr}\left[\frac{d^u}{dr^u}f_\nu^\nu\right] = \frac{d}{dr}\left[\sum_{i=0}^{\frac{u-1}{2}}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu-u+i}^{\nu-i}\right] \tag{A.21}$$

$$= \sum_{i=0}^{\frac{u-1}{2}}(-1)^{u+1+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}\left(f_{\nu-u+i-1}^{\nu-i} - (u-2i)f_{\nu-u+i}^{\nu-i-1}\right) \tag{A.22}$$

$$= (-1)^{u+1}f_{\nu-(u+1)}^{\nu} + \sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+1+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu-(u+1)+i}^{\nu-i}$$
$$+\sum_{i=0}^{\frac{u-1}{2}-1}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i-2)!!}f_{\nu-u+i}^{\nu-i-1} + (-1)^{u+\frac{u-1}{2}}u!!f_{\nu-\frac{u+1}{2}}^{\nu-\frac{u+1}{2}} \tag{A.23}$$

$$= (-1)^{u+1}f_{\nu-(u+1)}^{\nu} + \sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+1+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu-(u+1)+i}^{\nu-i}$$
$$+\sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+1+i}\binom{\frac{u-1}{2}}{i-1}\frac{u!!}{(u-2i)!!}f_{\nu-(u+1)+i}^{\nu-i} + (-1)^{\frac{u+1}{2}}u!!f_{\nu-\frac{u+1}{2}}^{\nu-\frac{u+1}{2}} \tag{A.24}$$

$$= (-1)^{u+1}f_{\nu-(u+1)}^{\nu} + \sum_{i=1}^{\frac{u+1}{2}-1}(-1)^{u+1+i}\binom{\frac{u+1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu-(u+1)+i}^{\nu-i}$$
$$+(-1)^{\frac{u+1}{2}}u!!f_{\nu-\frac{u+1}{2}}^{\nu-\frac{u+1}{2}} \tag{A.25}$$

□

A.3. Proof of Lemma 3.3

Lemma A.4

$$\frac{d^u}{dr^u}\left[r^{-\nu}J_\nu(r)\right] = \sum_{i=0}^{\lfloor\frac{u}{2}\rfloor}(-1)^{u+i}\binom{\lfloor\frac{u}{2}\rfloor}{i}\frac{(2\lfloor\frac{u+1}{2}\rfloor-1)!!}{(2\lfloor\frac{u+1}{2}\rfloor-1-2i)!!}r^{-(\nu+i)}J_{\nu+u-i}(r), \quad \textit{for } u\in\mathbb{N}$$

or equivalently

$$\frac{d^u}{dr^u}\left[r^{-\nu}J_\nu(r)\right] = \sum_{i=0}^{\frac{u}{2}}(-1)^{u+i}\binom{\frac{u}{2}}{i}\frac{(u-1)!!}{(u-2i-1)!!}r^{-(\nu+i)}J_{\nu+u-i}(r), \quad \textit{for } u \textit{ even} \tag{A.26}$$

$$\frac{d^u}{dr^u}\left[r^{-\nu}J_\nu(r)\right] = \sum_{i=0}^{\frac{u-1}{2}}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}r^{-(\nu+i)}J_{\nu+u-i}(r), \quad \textit{for } u \textit{ odd.} \tag{A.27}$$

Proof : The proof is nearly the same as for Lemma A.3 and it can be carried out by the induction over u, distinguishing the cases when u is even and odd. The argumentation is

the same. I show the second and third equation, the first one is just a more comprehensive formulation of the claim for a general $u \in \mathbb{N}$.

Denote $f_b^a = r^{-a} J_b(r)$, $a, b \in \mathbb{R}$, suppressing also the argument r of the function f_b^a for short. Using the following well known rule for differentiation of the Bessel functions

$$\frac{d}{dz}[z^{-\nu} J_\nu(z)] = -z^{-\nu} J_{\nu+1}(z)$$

(see Abramowitz and Stegun (1972), formula 9.1.30), the following simple equation can be shown

$$\begin{aligned} \frac{d}{dr} f_b^a &= \frac{d}{dr}[r^{-a} J_b(r)] = \frac{d}{dr}[r^{-(a-b)} r^{-b} J_b(r)] \\ &= (-1)[r^{-a} J_{b+1}(r) + (a-b) r^{-(a+1)} J_b(r)] \\ &= (-1)[f_{b+1}^a - (b-a) f_b^{a+1}]. \end{aligned} \tag{A.28}$$

Now, the proof can be accomplished. For $u = 0$ and for $u = 1$ the statement holds trivially. Let us suppose that u is even and (A.26) holds. Applying the induction step in (A.29), using (A.28) in (A.30) and by shifting the index i in the second sum of (A.31), (A.32) can be obtained. Finally, applying (A.14), we get (A.33) which is the claim to prove for $u+1$ odd.

$$\frac{d^{u+1}}{dz^{u+1}} f_\nu^\nu = \frac{d}{dz}\left[\frac{d^u}{dz^u} f_\nu^\nu\right] = \frac{d}{dz}\left[\sum_{i=0}^{\frac{u}{2}} (-1)^{u+i} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-2i-1)!!} f_{\nu+u-i}^{\nu+i}\right] \tag{A.29}$$

$$= \sum_{i=0}^{\frac{u}{2}} (-1)^{u+i} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-2i-1)!!} (-1)\left[f_{\nu+u-i+1}^{\nu+i} - (u-2i) f_{\nu+u-i}^{\nu+i+1}\right] \tag{A.30}$$

$$\begin{aligned} &= (-1)^{u+1} f_{\nu+u+1}^\nu + \sum_{i=1}^{\frac{u}{2}} (-1)^{u+i+1} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-2i-1)!!} f_{\nu+u+1-i}^{\nu+i} \\ &\qquad + \sum_{i=0}^{\frac{u}{2}-1} (-1)^{u+i} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-2i-1)!!} (u-2i) f_{\nu+u-i}^{\nu+i+1} \end{aligned} \tag{A.31}$$

$$\begin{aligned} &= (-1)^{u+1} f_{\nu+u+1}^\nu + \sum_{i=1}^{\frac{u}{2}} (-1)^{u+i+1} \binom{\frac{u}{2}}{i} \frac{(u-1)!!}{(u-2i-1)!!} f_{\nu+u+1-i}^{\nu+i} \\ &\qquad + \sum_{i=1}^{\frac{u}{2}} (-1)^{u+i+1} \binom{\frac{u}{2}}{i-1} \frac{(u-1)!!}{(u-2i+1)!!} (u-2i+2) f_{\nu+u+1-i}^{\nu+i} \end{aligned} \tag{A.32}$$

$$= (-1)^{u+1} f_{\nu+u+1}^\nu + \sum_{i=1}^{\frac{u}{2}} (-1)^{u+i+1} \binom{\frac{u}{2}}{i} \frac{(u+1)!!}{(u+1-2i)!!} f_{\nu+u+1-i}^{\nu+i}. \tag{A.33}$$

Let us suppose that u is odd and (A.27) holds. Applying the induction step in (A.34), using the rule (A.28) in (A.35) and by shifting the index i in the second sum of (A.37), (A.39) can be obtained. Finally, by applying (A.15), we get (A.40) which is the claim to prove for $u+1$ even.

$$\frac{d^{u+1}}{dz^{u+1}}f_\nu^\nu = \frac{d}{dz}\left[\frac{d^u}{dz^u}f_\nu^\nu\right] = \frac{d}{dz}\left[\sum_{i=0}^{\frac{u-1}{2}}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu+u-i}^{\nu+i}\right] \tag{A.34}$$

$$= \sum_{i=0}^{\frac{u-1}{2}}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}(-1)\left[f_{\nu+u-i+1}^{\nu+i} - (u-2i)f_{\nu+u-i}^{\nu+i+1}\right] \tag{A.35}$$

$$= (-1)^{u+1}f_{\nu+u+1}^{\nu} + \sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+i+1}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu+u+1-i}^{\nu+i} \tag{A.36}$$

$$+ \sum_{i=0}^{\frac{u-1}{2}-1}(-1)^{u+i}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i-2)!!}f_{\nu+u-i}^{\nu+i+1} + (-1)^{\frac{u+1}{2}}u!!f_{\nu+\frac{u+1}{2}}^{\nu+\frac{u+1}{2}} \tag{A.37}$$

$$= (-1)^{u+1}f_{\nu+u+1}^{\nu} + \sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+i+1}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu+u+1-i}^{\nu+i} \tag{A.38}$$

$$+ \sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+i+1}\binom{\frac{u-1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu+u+1-i}^{\nu+i} + (-1)^{\frac{u+1}{2}}u!!f_{\nu+\frac{u+1}{2}}^{\nu+\frac{u+1}{2}} \tag{A.39}$$

$$= (-1)^{u+1}f_{\nu+u+1}^{\nu} + \sum_{i=1}^{\frac{u-1}{2}}(-1)^{u+1+i}\binom{\frac{u+1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu+u+1-i}^{\nu+i} + (-1)^{\frac{u+1}{2}+1}u!!f_{\nu+\frac{u+1}{2}}^{\nu+\frac{u+1}{2}}$$

$$= \sum_{i=0}^{\frac{u+1}{2}}(-1)^{(u+1)+i}\binom{\frac{u+1}{2}}{i}\frac{u!!}{(u-2i)!!}f_{\nu+(u+1)-i}^{\nu+i} \tag{A.40}$$

□

A.4. Matérn class: proof of the equation (3.13)

Denote

$$Q = 2^{\nu-1}\phi\Gamma(\nu+\frac{d}{2})\alpha^{2\nu}\pi^{-\frac{d}{2}}$$

in (3.9) and $v_{\boldsymbol{kl}} = \operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x}))$ for short. Denote further $\boldsymbol{d\omega}_n = d\omega_1 \dots d\omega_n$ a volume element of the n-dimensional space $\mathbb{R}^n$. Recall Lemma 2.2, which says that it suffices to consider multi-indices $\boldsymbol{k} = (k_1, \dots, k_d)$, and $\boldsymbol{l} = (l_1, \dots, l_d)$, such that $\boldsymbol{k}+\boldsymbol{l}$ is even. Use the equation (3.6) and obtain (A.41). Use formula 3.251.11 in Gradstein and Ryshik (1981b) and get (A.42). Proceed the same step again $(d-1)$ times and get (A.43), which can be further re-written as (A.44). From properties of beta and gamma function, get finally (A.45).

$$v_{\boldsymbol{kl}} = (-1)^{|\boldsymbol{k}|}(-i)^{|\boldsymbol{k}+\boldsymbol{l}|}Q\int_{\mathbb{R}^d}\omega_1^{k_1+l_1}\dots\omega_d^{k_d+l_d}\big(\alpha^2+\sum_{j=1}^{d}\omega_j^2\big)^{-(\nu+\frac{d}{2})}d\omega_1\dots d\omega_d \tag{A.41}$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}Q\int_{\mathbb{R}^d}\omega_1^{k_1+l_1}\dots\omega_d^{k_d+l_d}\big(\alpha^2+\sum_{j=1}^{d-1}\omega_j^2\big)^{-(\nu+\frac{d}{2})}\Big(1+\frac{\omega_d^2}{\alpha^2+\sum_{j=1}^{d-1}\omega_j^2}\Big)^{-(\nu+\frac{d}{2})}\boldsymbol{d\omega}_d$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}Q\ B\big(\frac{k_d+l_d}{2}+\frac{1}{2},\nu-\frac{k_d+l_d}{s}+\frac{d-1}{2}\big)\times$$

$$\int_{\mathbb{R}^{d-1}}\omega_1^{k_1+l_1}\dots\omega_{d-1}^{k_{d-1}+l_{d-1}}\big(\alpha^2+\sum_{j=1}^{d-1}\omega_j^2\big)^{-(\nu-\frac{k_d+l_d}{2}+\frac{d-1}{2})}\boldsymbol{d\omega}_{d-1} \tag{A.42}$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}\ Q\ B\big(\frac{k_d+l_d}{2}+\frac{1}{2},\nu-\frac{k_d+l_d}{2}+\frac{d-1}{2}\big)\times$$

$$B\big(\frac{k_{d-1}+l_{d-1}}{2}+\frac{1}{2},\nu-\frac{k_d-l_d}{2}-\frac{k_{d-1}+l_{d-1}}{2}+\frac{d-2}{2}\big)\times$$

$$\int_{\mathbb{R}^{d-2}}\omega_1^{k_1+l_1}\dots\omega_{d-2}^{k_{d-2}+l_{d-2}}\big(\alpha^2+\sum_{j=1}^{d-2}\omega_j^2\big)^{-(\nu-\frac{k_d+l_d}{2}-\frac{k_{d-1}+l_{d-1}}{2}+\frac{d-2}{2})}\boldsymbol{d\omega}_{d-2}$$

$$=$$

$$\vdots$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}Q\ \alpha^{-2\nu+|\boldsymbol{k}+\boldsymbol{l}|}\ B\big(\frac{k_d+l_d}{2}+\frac{1}{2},\nu-\frac{k_d+l_d}{2}+\frac{d-1}{2}\big)\times$$

$$B\big(\frac{k_{d-1}+l_{d-1}}{2}+\frac{1}{2},\nu-\frac{k_d+l_d}{2}-\frac{k_{d-1}+l_{d-1}}{2}+\frac{d-2}{2}\big)\dots\times$$

$$\dots\ B\big(\frac{k_j+l_j}{2}+\frac{1}{2},\nu-\sum_{i=j}^{d}\frac{k_i+l_i}{2}+\frac{j-1}{2}\big)\dots B\big(\frac{k_1+l_1}{2}+\frac{1}{2},\nu-\sum_{i=1}^{d}\frac{k_i+l_i}{2}\big) \tag{A.43}$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}\ \frac{2^{\nu-1}\,\phi\,\Gamma(\nu+d/2)}{\pi^{d/2}}\,\alpha^{|\boldsymbol{k}+\boldsymbol{l}|}\ \prod_{j=1}^{d}B\big(\frac{k_j+l_j}{2}+\frac{1}{2},\nu-\sum_{i=j}^{d}\frac{k_i+l_i}{2}+\frac{j-1}{2}\big) \tag{A.44}$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}\ \frac{2^{\nu-1}\ \phi}{\pi^{d/2}}\ \alpha^{|\boldsymbol{k}+\boldsymbol{l}|}\ \Gamma\big(\nu-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}\big)\ \prod_{j=1}^{d}\Gamma\Big(\frac{k_j+l_j}{2}+\frac{1}{2}\Big) \tag{A.45}$$

□

A.5. J-Bessel class: proof of the equation (3.25)

Denote

$$Q = \frac{\phi}{2^{\nu}\pi^{\frac{d}{2}}\alpha^{2\nu}\Gamma(\nu+1-\frac{d}{2})}$$

in (3.24), $v_{\boldsymbol{kl}} = \mathrm{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{x}))$ for short. Denote further $\boldsymbol{d\omega}_n = d\omega_1 \dots d\omega_n$ a volume element of the n-dimensional space $\mathbb{R}^n$. Recall Lemma 2.2, which implies that it suffices to consider multi-indices $\boldsymbol{k}=(k_1,\dots,k_d)$, and $\boldsymbol{l}=(l_1,\dots,l_d)$, such that $\boldsymbol{k}+\boldsymbol{l}$ is even. Use the equation (3.6) and obtain (A.46). Use formula 3.251.1 in Gradstein and Ryshik (1981b) or formula 2.2.4.8 in Prudnikov et al. (1996) and get (A.47). Proceed the same step again $(d-1)$-times and obtain (A.48). From properties of the beta and gamma function, (A.49) can be obtained.

$$v_{\boldsymbol{kl}} = Q\,(-1)^{|\boldsymbol{k}|}(-i)^{|\boldsymbol{k}+\boldsymbol{l}|} \int\limits_{\sum_{j=1}^{d}\omega_j^2\le\alpha^2} \dots \int \omega_1^{k_1+l_1}\dots\omega_d^{k_d+l_d}\big(\alpha^2-\sum_{j=1}^{d}\omega_j^2\big)^{\nu-\frac{d}{2}}\,\boldsymbol{d\omega}_d \tag{A.46}$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} Q\, 2^d \int_0^{\alpha} \omega_1^{k_1+l_1} \int_0^{\alpha^2-\omega_1^2} \omega_2^{k_2+l_2} \dots \int_0^{\alpha^2-\sum_{j=1}^{d-1}\omega_j^2} \omega_d^{k_d+l_d}\big(\alpha^2-\sum_{j=1}^{d}\omega_j^2\big)^{\nu-\frac{d}{2}}\,\boldsymbol{d\omega}_d$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} Q\; 2^{d-1} B\big(\frac{k_d+l_d}{2}+\frac{1}{2}, \nu-\frac{d}{2}+1\big) \times$$

$$\int_0^{\alpha} \omega_1^{k_1+l_1} \dots \int_0^{\alpha^2-\sum_{j=1}^{d-2}\omega_j^2} \omega_{d-1}^{k_{d-1}+l_{d-1}}\big(\alpha^2-\sum_{j=1}^{d-1}\omega_j^2\big)^{\nu+\frac{k_d+l_d}{2}-\frac{d-1}{2}}\,\boldsymbol{d\omega}_{d-1} \tag{A.47}$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} Q\; 2^{d-2} B\big(\frac{k_d+l_d}{2}+\frac{1}{2}, \nu-\frac{d}{2}+1\big) \times$$

$$B\big(\frac{k_{d-1}+l_{d-1}}{2}+\frac{1}{2}, \nu+\frac{k_d+l_d}{2}-\frac{d-1}{2}+1\big) \times$$

$$\int_0^{\alpha} \omega_1^{k_1+l_1} \dots \int_0^{\alpha^2-\sum_{j=1}^{d-3}\omega_j^2} \omega_{d-2}^{k_{d-2}+l_{d-2}}\big(\alpha^2-\sum_{j=1}^{d-2}\omega_j^2\big)^{\nu+k_d+k_{d-1}-\frac{d-2}{2}}\,\boldsymbol{d\omega}_{d-2}$$

$$=$$

$$\vdots$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} Q\; \alpha^{2(\nu+|\boldsymbol{k}|)} B\big(\frac{k_d+l_d}{2}+\frac{1}{2}, \nu-\frac{d}{2}+1\big) \times$$

$$B\big(\frac{k_{d-1}+l_{d-1}}{2}+\frac{1}{2}, \nu+\frac{k_d+l_d}{2}-\frac{d-1}{2}+1\big)\dots \times \tag{A.48}$$

$$\dots B\big(\frac{k_j+l_j}{2}+\frac{1}{2}, \nu+\sum_{i=j+1}^{d}\frac{k_i+l_i}{2}-\frac{j}{2}+1\big)\dots B\big(\frac{k_1+l_1}{2}+\frac{1}{2}, \nu+\sum_{i=2}^{d}\frac{k_i+l_i}{2}-\frac{1}{2}+1\big)$$

$$= (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}} \phi\; \alpha^{|\boldsymbol{k}+\boldsymbol{l}|}\; 2^{-\nu}\; \pi^{-\frac{d}{2}}\; \frac{1}{\Gamma\big(\nu+\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}+1\big)} \prod_{i=1}^{d}\Gamma\big(\frac{k_i+l_i}{2}+\frac{1}{2}\big) \tag{A.49}$$

□

A.6. Matrix of convolutions

Let Z be a stationary, isotropic and sufficiently differentiable random field on $\mathbb{R}^d$ and let $f \in L^1(\mathbb{R}^d)$ be its spectral density. Denote

$$\begin{aligned} c_{\boldsymbol{k}0}(\boldsymbol{x}) &= \operatorname{cov}(Z^{(\boldsymbol{k})}(\boldsymbol{0}), Z^{(0)}(\boldsymbol{x})) \\ c_{0\boldsymbol{l}}(\boldsymbol{h}-\boldsymbol{x}) &= \operatorname{cov}(Z^{(0)}(\boldsymbol{x}), Z^{(\boldsymbol{l})}(\boldsymbol{h})) = \operatorname{cov}(Z^{(\boldsymbol{l})}(\boldsymbol{h}), Z^{(0)}(\boldsymbol{x})) = c_{\boldsymbol{l}0}(\boldsymbol{x}-\boldsymbol{h}) \end{aligned}$$

For the convolution of this covariances, it holds (cf. (3.5))

$$\begin{aligned} c_{0\boldsymbol{l}} * c_{\boldsymbol{k}0}(\boldsymbol{h}) &= \int_{\mathbb{R}^d} c_{0\boldsymbol{l}}(\boldsymbol{h}-\boldsymbol{x}) c_{\boldsymbol{k}0}(\boldsymbol{x})\, \boldsymbol{dx} \\ &= (-1)^{|\boldsymbol{k}|}(2\pi)^d F[(i\omega)^{k_1+l_1}\dots(i\omega)^{k_d+l_d} f^2(\omega)](\boldsymbol{h}) \end{aligned} \tag{A.50}$$

where in the equation (A.50) subsequently equations (2.14) and (3.6) are used. Finally, assuming that $(i\omega)^{k_1+l_1}\dots(i\omega)^{k_d+l_d} f^2(\omega) \in L^1(\mathbb{R}^d)$, the Convolution theorem 4 (page 28) is applied. The above assumption means in consequence that $c_{\boldsymbol{k}0} \in L^2(\mathbb{R}^d)$. Since $\boldsymbol{k}$ can be also equal to zero multi-index, it means that $f \in L^2(\mathbb{R}^d)$. (Note that under f and g in the Theorem 4 the functions $c_{\boldsymbol{k}0}$, and $c_{0\boldsymbol{l}}$ should be understood.)

Since

$$\int_{\mathbb{R}^d} c_{0\boldsymbol{l}}(\boldsymbol{h}-\boldsymbol{x}) c_{\boldsymbol{k}0}(\boldsymbol{x})\, \boldsymbol{dx} = \int_{\mathbb{R}^d} c_{\boldsymbol{k}0}(\boldsymbol{x}) c_{\boldsymbol{l}0}(\boldsymbol{x}-\boldsymbol{h})\, \boldsymbol{dx}$$

we can easily calculate inner product $\langle c_{\boldsymbol{k}0}, c_{\boldsymbol{l}0}\rangle$

$$\langle c_{\boldsymbol{k}0}, c_{\boldsymbol{l}0}\rangle := \int_{\mathbb{R}^d} c_{\boldsymbol{k}0}(\boldsymbol{x}) c_{\boldsymbol{l}0}(\boldsymbol{x})\, \boldsymbol{dx} = c_{\boldsymbol{k}0} * c_{0\boldsymbol{l}}(\boldsymbol{0}), \tag{A.51}$$

and hence

$$||c_{\boldsymbol{k}0}||_2^2 = \langle c_{\boldsymbol{k}0}, c_{\boldsymbol{k}0}\rangle.$$

Remark The product of the spectral densities of two Gaussian covariance functions is again a spectral density of a Gaussian covariance function. The same holds for the Matérn and the J-Bessel class too. In the following, this property is used to establish some formulas, in particular, formulas for (A.51).

A.6.1. Gaussian covariance function

Denote $(i\boldsymbol{\omega})^{\boldsymbol{k}} := (i\omega_1)^{k_1}\dots(i\omega_d)^{k_d}$ for short. Apply (A.50) on the formula (3.16) for the spectral density of the Gaussian covariance function. Further, the formulas (3.6) and (3.21) yield

$$\begin{aligned} c_{0\boldsymbol{l}} * c_{\boldsymbol{k}0}(\boldsymbol{h}) &= (-1)^{|\boldsymbol{k}|}(2\pi)^d \left(\frac{\sigma^2}{(2\sqrt{\pi}\alpha)^d}\right)^2 F\big[(i\boldsymbol{\omega})^{\boldsymbol{k}+\boldsymbol{l}} \exp\Big(-\frac{\omega^2}{2\alpha^2}\Big)\big](\boldsymbol{h}) \\ &= (-1)^{|\boldsymbol{k}|}\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \frac{\sigma^4}{\big(2\sqrt{\pi}\frac{\alpha}{\sqrt{2}}\big)^d} F\big[(i\boldsymbol{\omega})^{\boldsymbol{k}+\boldsymbol{l}} \exp\Big(\frac{-\omega^2}{4(\frac{\alpha}{\sqrt{2}})^2}\Big)\big](\boldsymbol{h}) \\ &= (-1)^{|\boldsymbol{k}|}\sigma^4\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \left[\frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial h^{k_1+l_1}\dots\partial h^{k_d+l_d}} \exp\Big(-\big(\frac{\alpha}{\sqrt{2}}\big)^2||\boldsymbol{h}||^2\Big)\right] \qquad \text{(A.52)} \\ &= (-1)^{|\boldsymbol{l}|}\sigma^4\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \exp\big(-\frac{\alpha^2}{2}||\boldsymbol{h}||^2\big)\, \alpha^{|\boldsymbol{k}+\boldsymbol{l}|}\, 2^{-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}} \prod_{i=1}^d H_{k_i+l_i}\big(\frac{\alpha}{\sqrt{2}}h_i\big) \qquad \text{(A.53)} \end{aligned}$$

Recall (A.51). Using (2.14), (3.6) and (3.21) in (A.52), the following equation can be obtained

$$\langle c_{\boldsymbol{k}0}, c_{\boldsymbol{l}0}\rangle = (-1)^{\frac{|\boldsymbol{l}|-|\boldsymbol{k}|}{2}}\sigma^4\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \alpha^{|\boldsymbol{k}+\boldsymbol{l}|}\ 2^{-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}}\ \prod_{i=1}^{d}\frac{(k_i+l_i)!}{((k_i+l_i)/2)!}, \tag{A.54}$$

and hence

$$||c_{\boldsymbol{k}0}||_2^2 = \sigma^4\left(\frac{1}{\alpha}\sqrt{\frac{\pi}{2}}\right)^d \alpha^{2|\boldsymbol{k}|}\ 2^{-|\boldsymbol{k}|}\ \prod_{i=1}^{d}\frac{(2k_i)!}{k_i!}. \tag{A.55}$$

A.6.2. Matérn class

Using (A.50) and the formula (3.9) for the spectral density of the Matérn class covariance function and formula (3.6), the following equation can be obtained

$$\begin{aligned}
c_{0\boldsymbol{l}} * c_{\boldsymbol{k}0}(\boldsymbol{h}) &= (-1)^{|\boldsymbol{k}|}(2\pi)^d\Big(\frac{\phi 2^{\nu-1}\Gamma(\nu+\frac{d}{2})\alpha^{2\nu}}{\pi^{\frac{d}{2}}}\Big)^2 F\big[(i\boldsymbol{\omega})^{\boldsymbol{k}+\boldsymbol{l}}\frac{1}{(\alpha^2+\omega^2)^{2\nu+d}}\big](\boldsymbol{h}) \\
&= (-1)^{|\boldsymbol{k}|}\frac{\phi}{2}\Big(\frac{\sqrt{2\pi}}{\alpha}\Big)^d B\big(\nu+\frac{d}{2},\nu+\frac{d}{2}\big)\times \\
&\quad \Big(\frac{\phi 2^{2\nu+\frac{d}{2}-1}\Gamma(2\nu+d)\alpha^{2(2\nu+\frac{d}{2})}}{\pi^{\frac{d}{2}}}\Big)F\big[(i\boldsymbol{\omega})^{\boldsymbol{k}+\boldsymbol{l}}\frac{1}{(\alpha^2+\omega^2)^{2\nu+d}}\big](\boldsymbol{h}) \\
&= (-1)^{|\boldsymbol{k}|}\frac{\phi}{2}\Big(\frac{\sqrt{2\pi}}{\alpha}\Big)^d B\big(\nu+\frac{d}{2},\nu+\frac{d}{2}\big)\times \\
&\quad \Big[\frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial h_1^{k_1+l_1}\dots\partial h_d^{k_d+l_d}}\,\phi\,(\alpha||\boldsymbol{h}||)^{2\nu+\frac{d}{2}}\mathcal{K}_{2\nu+\frac{d}{2}}(\alpha||\boldsymbol{h}||)\Big].
\end{aligned} \tag{A.56}$$

If $d=1$, Lemma 3.1 can be applied to (A.53) to get a more explicite formula. However, to obtain the following formulae for the inner product, Lemma 3.1 is not necessary.

Using (2.14), (3.6) and (3.13) in (A.56), the following equation can be obtained

$$\begin{aligned}
\langle c_{\boldsymbol{k}0}, c_{\boldsymbol{l}0}\rangle &= (-1)^{\frac{|\boldsymbol{l}-\boldsymbol{k}|}{2}}\,\phi^2\,\frac{1}{\alpha^d}\,B\big(\nu+\frac{d}{2},\nu+\frac{d}{2}\big)\,2^{2\nu+d-2}\,\alpha^{|\boldsymbol{k}+\boldsymbol{l}|}\times \\
&\quad \Gamma\Big(2\nu+\frac{d}{2}-\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}\Big)\prod_{i=1}^{d}\Gamma\Big(\frac{k_i+l_i}{2}+\frac{1}{2}\Big),
\end{aligned} \tag{A.57}$$

and subsequently

$$||c_{\boldsymbol{k}0}||_2^2 = \phi^2\,\frac{1}{\alpha^d}\,B\big(\nu+\frac{d}{2},\nu+\frac{d}{2}\big)\,2^{2\nu+d-2}\,\alpha^{2|\boldsymbol{k}|}\,\Gamma\Big(2\nu+\frac{d}{2}-|\boldsymbol{k}|\Big)\prod_{i=1}^{d}\Gamma\Big(k_i+\frac{1}{2}\Big). \tag{A.58}$$

A.6.3. J-Bessel class

Using (A.50) on the formula (3.24) for the spectral density of the J-Bessel class covariance function and formula (3.6), the following equation can be obtained

$$\begin{aligned} c_{0\boldsymbol{l}} * c_{\boldsymbol{k}0}(\boldsymbol{h}) &= (-1)^{|\boldsymbol{k}|}(2\pi)^d \left(\frac{\phi}{2^\nu \pi^{\frac{d}{2}} \alpha^{2\nu} \Gamma(\nu+1-\frac{d}{2})}\right)^2 F\big[(i\boldsymbol{\omega})^{\boldsymbol{k}+\boldsymbol{l}}(\alpha^2-\omega^2)^{2\nu-d}\big](\boldsymbol{h}) \\ &= (-1)^{|\boldsymbol{k}|}\phi\Big(\frac{\sqrt{2\pi}}{\alpha}\Big)^d \frac{\Gamma(2\nu+1-d)}{\Gamma^2(\nu+1-\frac{d}{2})} \times \\ &\quad \left(\frac{\phi}{2^{2\nu-\frac{d}{2}}\,\pi^{\frac{d}{2}}\,\alpha^{2(2\nu-\frac{d}{2})}\,\Gamma(2\nu+1-d)}\right) F\big[(i\boldsymbol{\omega})^{\boldsymbol{k}+\boldsymbol{l}}(\alpha^2-\omega^2)^{2\nu-d}\big](\boldsymbol{h}) \\ &= (-1)^{|\boldsymbol{k}|}\phi\Big(\frac{\sqrt{2\pi}}{\alpha}\Big)^d \frac{\Gamma(2\nu+1-d)}{\Gamma^2(\nu+1-\frac{d}{2})} \times \\ &\quad \left[\frac{\partial^{|\boldsymbol{k}+\boldsymbol{l}|}}{\partial h_1^{k_1+l_1}\dots\partial h_d^{k_d+l_d}}\,\phi\,(\alpha||\boldsymbol{h}||)^{-(2\nu-\frac{d}{2})} J_{2\nu-\frac{d}{2}}(\alpha||\boldsymbol{h}||)\right] \end{aligned}$$

In one-dimensional case, Lemma 3.3 can be applied to obtain a more explicite formula.

Using (2.14), (3.6) and (3.25) in (A.59), the following equation can be obtained

$$\langle c_{\boldsymbol{k}0}, c_{\boldsymbol{l}0}\rangle = (-1)^{\frac{|\boldsymbol{l}-\boldsymbol{k}|}{2}}\,\phi^2\,\frac{\Gamma(2\nu+1-d)}{\Gamma^2(\nu+1-\frac{d}{2})}\,\frac{2^{-(2\nu-d)}\alpha^{|\boldsymbol{k}+\boldsymbol{l}|-d}}{\Gamma(2\nu-\frac{d}{2}+\frac{|\boldsymbol{k}+\boldsymbol{l}|}{2}+1)}\,\prod_{i=1}^{d}\Gamma\Big(\frac{k_i+l_i}{2}+\frac{1}{2}\Big),$$

and therefore

$$||c_{\boldsymbol{k}0}||_2^2 = \phi^2\,\frac{\Gamma(2\nu+1-d)}{\Gamma^2(\nu+1-\frac{d}{2})}\,\frac{2^{-(2\nu-d)}\alpha^{2|\boldsymbol{k}|-d}}{\Gamma(2\nu-\frac{d}{2}+|\boldsymbol{k}|+1)}\,\prod_{i=1}^{d}\Gamma\Big(k_i+\frac{1}{2}\Big).$$

B. Appendix (Conditional simulation)

The following section introduces a conditional simulation method and it provides its formal generalization to a simultaneous (conditional) simulation of a random field and its derivative fields. There is only an indirect connection to the problem of optimal design and therefore the section was postponed to the appendix.

Conditional simulation of random fields

It is quite typical in geostatistics that physical measurements are linked with big costs (financial, material or environmental). Therefore, the amount of data available is sometimes much lower than in other disciplines. If real data are not available, it is possible to simulate them. Indeed, validating a method on data where the true values of parameters are known should be an important part of establishing its worth.

Another point is, that in practice complex objectives are required which are difficult to analyze. For instance, geostatisticians in oil prospection could be interested in whether two oil-fields are connected or not.

Another problem could be to estimate the length of stratum across a profile which is observed only at some locations. We can of course use the kriging predictor to interpolate the data, but the result will be much smoother than the real curve. In consequence, we would underestimate its the actual length. Simulations could deliver some insight which scenarios are possible and how their probability distribution functions look like. Thus, if the aim is to obtain some notion about the local structure of the study random field rather than to predict data, simulation can be a useful tool.

Nowadays, simulation of random fields is extensively used in problems like connectivity, fluid flow, transportation and migration of pollutants (see e.g. Chilès and Delfiner (1999)).

A *non-conditional simulation* of a random field Z is simply a realization of Z, randomly selected from the set of all possible realizations. The random field $Z(\boldsymbol{x})$ has infinitely many realizations. Given some observations of the field, we can choose those realizations which have at the sample points the same values as those observed. These realizations can be considered to better represent the reality. In this case, we speak about a *conditional simulation*.

In the following text, we restrict ourselves only to the simplest method for (conditional) simulation using the Cholesky decomposition. It can be shown that in this case there is no difficulty in incorporating observations of derivatives.

B.1. Cholesky decomposition

Recall that the spatial index of the random field under study $Z(\boldsymbol{x})$ varies continuously over some d-dimensional subspace D of $\mathbb{R}^d$. On the other hand, a simulation of the field can be

carried out only at discrete locations, denoted as $G = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. The locations might be a dense regular grid over D or they may be irregularly spaced points in D.

Suppose that the second order random field has the mean $m(\boldsymbol{x})$ and the covariance function $C(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y}))$. Note that neither stationarity nor isotropy is required. Let $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n))^T$ be a vector of pairwise different random variables to be simulated. Denote by $\boldsymbol{m} = (m(\boldsymbol{x}_1), \ldots, m(\boldsymbol{x}_n))^T$ and $\boldsymbol{\Sigma}$ the vector of mean values and the covariance matrix of $\boldsymbol{Z}$, respectively. Recall that the entries of the matrix $\boldsymbol{\Sigma}$ are defined by $\boldsymbol{\Sigma}_{ij} = \mathrm{cov}(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j))$ for $i, j \in \{1, \ldots, n\}$.

Since $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, the Cholesky decomposition allows $\boldsymbol{\Sigma}$ to be decomposed as

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T, \tag{B.1}$$

where $\boldsymbol{L}$ is a lower triangular matrix. Thus, we can simulate $\boldsymbol{Z}$ as

$$\boldsymbol{Z}_{NS} = \boldsymbol{m} + \boldsymbol{L}\boldsymbol{\epsilon}, \tag{B.2}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ is a vector of uncorrelated zero mean random variables with the unit variance. The resulting vector $\boldsymbol{Z}_{NS}$ is a random selection from a distribution which has the mean $\boldsymbol{m}$ and the covariance matrix $\boldsymbol{\Sigma}$ as required. Note that the decomposition (B.1) needs to be done only once and can be reused for simulation of many realizations.

Usually, $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. Gaussian random variables. This implies that $\boldsymbol{Z}$ is multivariate Gaussian. Some other generic models like lognormal, $\exp(Z(\boldsymbol{x}))$, or chi-square random field, $Z^2(\boldsymbol{x})$, can be obtained directly by applying the corresponding transformation to the simulated Gaussian data.

Note that the source of random variation is usually a pseudo-random number generator. The choice of a good one for spatial simulation needs some care, especially for higher dimensional data (see e.g. Robert and Casella (1999)).

B.1.1. Simulation of derivative fields

Since there is no restriction on the vector $\boldsymbol{Z}$ of the simulated random variables (except that Z must be of the second order), simulation of derivative fields can be done with just a formal modification in the algorithm.

Let $\boldsymbol{Z} = (Z^{(\boldsymbol{k}_1)}(\boldsymbol{x}_1), \ldots, Z^{(\boldsymbol{k}_n)}(\boldsymbol{x}_n))^T$ be a vector of the simulated random variables, provided that they are well defined. By means of formulas (2.12) and (2.13), we define $\boldsymbol{m} = (m^{(\boldsymbol{k}_1)}(\boldsymbol{x}_1), \ldots, m^{(\boldsymbol{k}_n)}(\boldsymbol{x}_n))^T$ and $\boldsymbol{\Sigma}_{ij} = \mathrm{cov}(Z^{(\boldsymbol{k}_i)}(\boldsymbol{x}_i), Z^{(\boldsymbol{k}_j)}(\boldsymbol{x}_j))$ for $i, j \in \{1, \ldots, n\}$. The procedure proceeds as above without further modifications.

Now, we can easily get correlated realizations of the underlying process and its derivatives by setting e.g. $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n), Z^{(1)}(\boldsymbol{x}_1), \ldots, Z^{(1)}(\boldsymbol{x}_n))^T$. Examples for a Matérn class covariance function are given in Figures B.1 and B.2.

Note that the main problem of the applicability of non-conditional simulation based on the Cholesky decomposition is, that for large sets G numerical problems turn out and computing time can become excessive. This could particularly be overcome when using covariance functions with a compact support, since nonzero entries in $\boldsymbol{\Sigma}$ will be reduced and algorithms with better performance can be used. Other methods, which are based on a separation of the domain into partially overlapping zones and further improvements of the original algorithm, are reviewed in Chilès and Delfiner (1999, page 465).

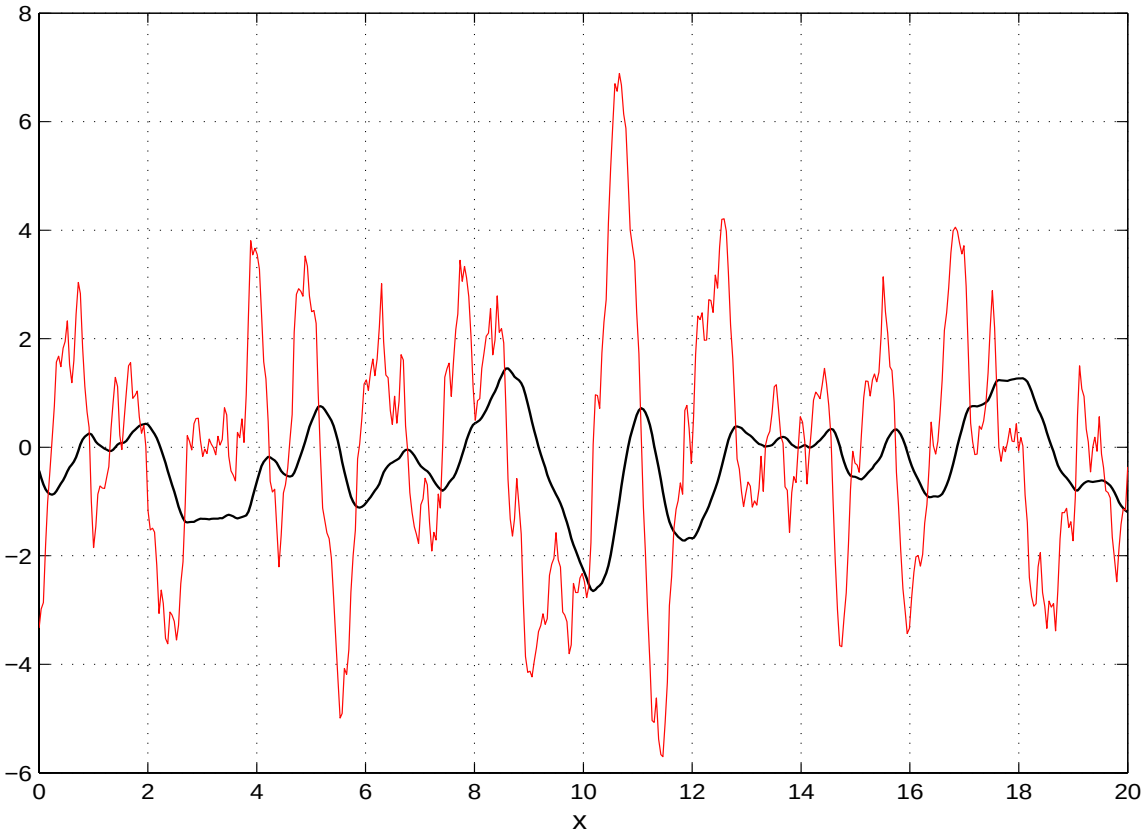


Figure B.1.: Non-conditional simulation of a stationary zero mean Gaussian process Z (thick less varying line) with the Matérn covariance function and its first derivative $Z^{(1)}$ (thin more varying line). The Matérn covariance function (parameterized by (3.10)) has parameters $\sigma^2 = 1$, $\rho = 1$, $\nu = 2$. The grid G consists of 500 points. It means that $\boldsymbol{\Sigma}$ is 1000×1000-matrix. Note that the simulated process $Z^{(0)}$ is the smoothest one (in the Matérn class) which is only once differentiable.

B.2. Conditioning by kriging

In this section, a general method for the transformation of a non-conditional simulation into a conditional one will be recalled. The principle is quite simple.

First, let us assume that we know how to construct a non-conditional simulation. Note that it can be any method for non-conditional simulation and not necessary the one by means of the Cholesky decomposition. Suppose that a random field Z has been observed at the sample points $V = \{\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N\}$. Denote by $\boldsymbol{Z}(V)$ the vector of observations of Z at these sample points, $\boldsymbol{Z}(V) = (Z(\boldsymbol{x}'_1), \ldots, Z(\boldsymbol{x}'_N))^T$. Let us assume that we have a non-conditional simulation $\boldsymbol{Z}_{NS}$ independent of the sample data but with the same covariance function as Z.

Conditioning means passing from Z_{NS} to a conditional simulation Z_{CS} that matches the sample data. Let us begin with the following trivial decomposition which can be found in

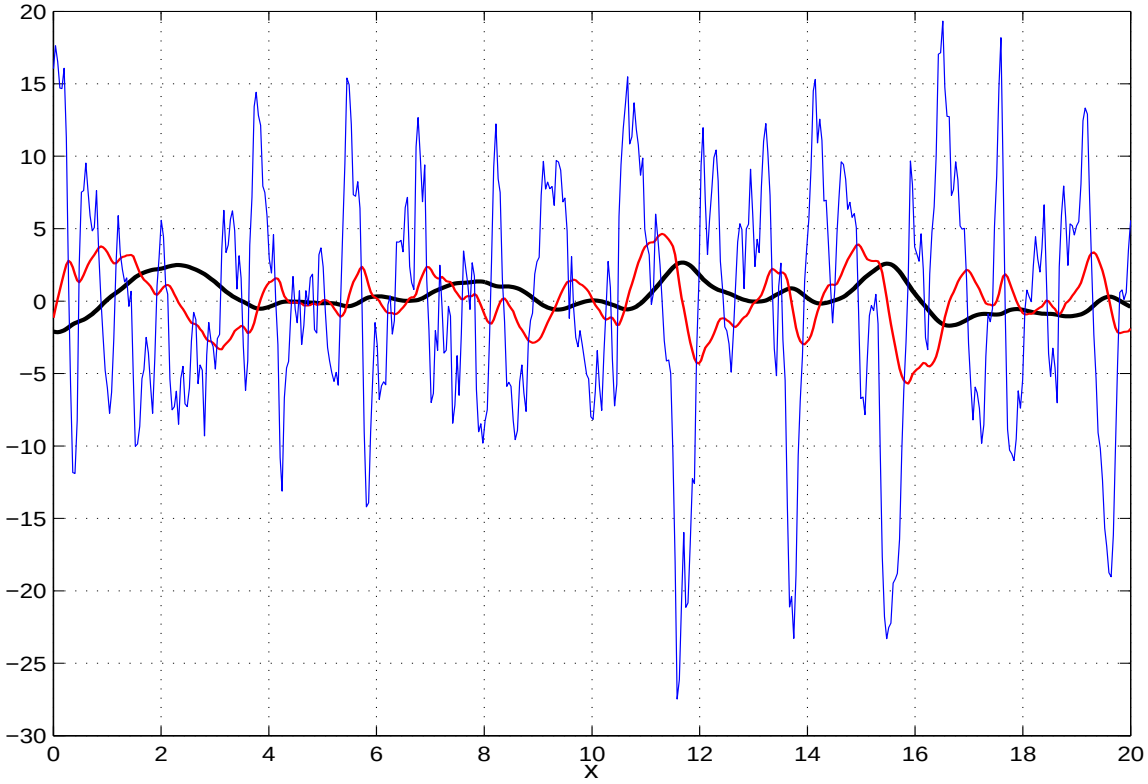


Figure B.2.: Non-conditional simulation of a stationary zero mean Gaussian process Z (thick at least varying line) with the Matérn covariance function and its first two derivatives. The medium line corresponds to $Z^{(1)}$; the thin highly varying line corresponds to the second derivative process $Z^{(2)}$. The Matérn covariance function (3.10) has the parameters $\sigma^2 = 1$, $\rho = 1$, $\nu = 3$. The grid G consists of 500 points. It means that $\boldsymbol{\Sigma}$ is 1500×1500-matrix. Note that the simulated process $Z^{(0)}$ is the smoothest one (in the Matérn class) which is only twice differentiable.

Chilès and Delfiner (1999) and Cressie (1991)

$$\begin{array}{rcccl} Z(\boldsymbol{x}) & = & p_L^*(Z(\boldsymbol{x}); \boldsymbol{Z}(V)) & + & \big[Z(\boldsymbol{x}) \ - \ p_L^*(Z(\boldsymbol{x}); \boldsymbol{Z}(V))\big]. \\ \text{true value} & = & \text{simple kriging predictor} & + & \text{prediction error} \end{array} \tag{B.3}$$

Since $Z(\boldsymbol{x})$ is not known for $\boldsymbol{x} \notin V$, the prediction error is not known. Note that

$$\operatorname{cov}\left(p_L^*(Z(\boldsymbol{x}); \boldsymbol{Z}(V)), \big[Z(\boldsymbol{x}) \ - \ p_L^*(Z(\boldsymbol{x}); \boldsymbol{Z}(V))\big]\right) = 0.$$

Similarly, we can also decompose Z_{NS}

$$Z_{NS}(\boldsymbol{x}) = p_L^*(Z_{NS}(\boldsymbol{x}); \boldsymbol{Z}_{NS}(V)) + \big[Z_{NS}(\boldsymbol{x}) \ - \ p_L^*(Z_{NS}(\boldsymbol{x}); \boldsymbol{Z}_{NS}(V))\big]. \tag{B.4}$$

The predictor and the prediction error are orthogonal as in the previous case. Here, on the other hand, the prediction error is known. Finally, define the conditional simulation by

replacing the prediction error in (B.3) with the prediction error from (B.4):

$$\begin{aligned} Z_{CS}(\boldsymbol{x}) &= p_L^*(Z(\boldsymbol{x}); \boldsymbol{Z}(V)) + \big[Z_{NS}(\boldsymbol{x}) - p_L^*(Z_{NS}(\boldsymbol{x}); \boldsymbol{Z}_{NS}(V))\big]. \\ \text{true value} &= \text{simple kriging predictor} + \text{simulation of predition error} \end{aligned} \tag{B.5}$$

It can easily be seen that Z_{CS} interpolates data at the sample points, since the simple kriging is an exact interpolator. Moreover, it holds

$$\operatorname{cov}(Z_{CS}(\boldsymbol{x}), Z_{CS}(\boldsymbol{y})) = \operatorname{cov}(Z_{NS}(\boldsymbol{x}), Z_{NS}(\boldsymbol{y})) = \operatorname{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) \tag{B.6}$$

as required.

B.2.1. Conditioning with derivative data

Since we already know how to simulate a random field and its derivatives simultaneously, the conditional simulation, in case when derivative data are observed, can be used without big modifications in the algorithm. Let us assume that data can be in the form $\boldsymbol{Z}(V) = (Z^{(\boldsymbol{k}_1)}(\boldsymbol{x}'_1), \ldots, Z^{(\boldsymbol{k}_N)}(\boldsymbol{x}'_N))^T$ whenever $\boldsymbol{Z}(V)$ is well defined, i.e. if the random field Z is differentiable sufficiently. To proceed conditional simulation correctly, we need to simulate new data $\boldsymbol{Z}_{NS}(V)$, together with the random field Z at the grid points G. The vector $\boldsymbol{Z}$ which is used for non-conditional simulation by means of the Cholesky decomposition (B.2) is as follows

$$\boldsymbol{Z} = (\underbrace{Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n)}_{(1)}, \underbrace{Z^{(\boldsymbol{k}_1)}(\boldsymbol{x}'_1), \ldots, Z^{(\boldsymbol{k}_N)}(\boldsymbol{x}'_N)}_{(2)}). \tag{B.7}$$

Simulating new data corresponding to (B.7), from the first part of these data values $Z_{NS}(\boldsymbol{x})$ in (B.5) are taken. On the other hand, the second part is used in predictor as $\boldsymbol{Z}_{NS}(V)$ only. If $V \subset G$ then all random variables which are common to the second and the first part of (B.7) need to be eliminated from the second part to assure regularity of the matrix $\boldsymbol{\Sigma} = \operatorname{cov}(\boldsymbol{Z}, \boldsymbol{Z}^T)$. However, the simulated values from (1) which correspond to the eliminated random variables in (2) need to be used for prediction.

Conditional simulations of a Gaussian process with the Matérn covariance function are presented in Figure B.3. The characteristics of the process are the same as in Figure B.1. On the original process (thin darker line) some observations have been proceeded, precisely at the points $1, 5, 8, 9, 10, 15, 17$. The circle $\odot$ means an observation of the process $Z^{(0)}$; the triangle $\triangle$ means an observation of the first derivative process $Z^{(1)}$, which was generated simultaneously with $Z^{(0)}$ as described in the previous text. Finally, the asterisk $*$ means that both the process and the first derivative process has been observed. The thick line represents a conditional simulation with respect to the given observations.

Resume

Following Chilès and Delfiner (1999), conditional simulations are useful qualitatively, to obtain realistic pictures of spatial variability. Quantitatively, they are a tool of choice to evaluate the impact of spatial uncertainty on the results of complex procedures, such as numerical modeling of dynamical systems or economic optimization of the exploitation of natural resources.

In this section, it was demonstrated that non-conditional simulation by means of the Cholesky decomposition can be straightforward generalized for simultaneous simulation of derivative fields. The same holds for the conditioning by the kriging, too. A question remains, whether non-conditional simulation of derivative fields can be adapted also for other simulation methods.

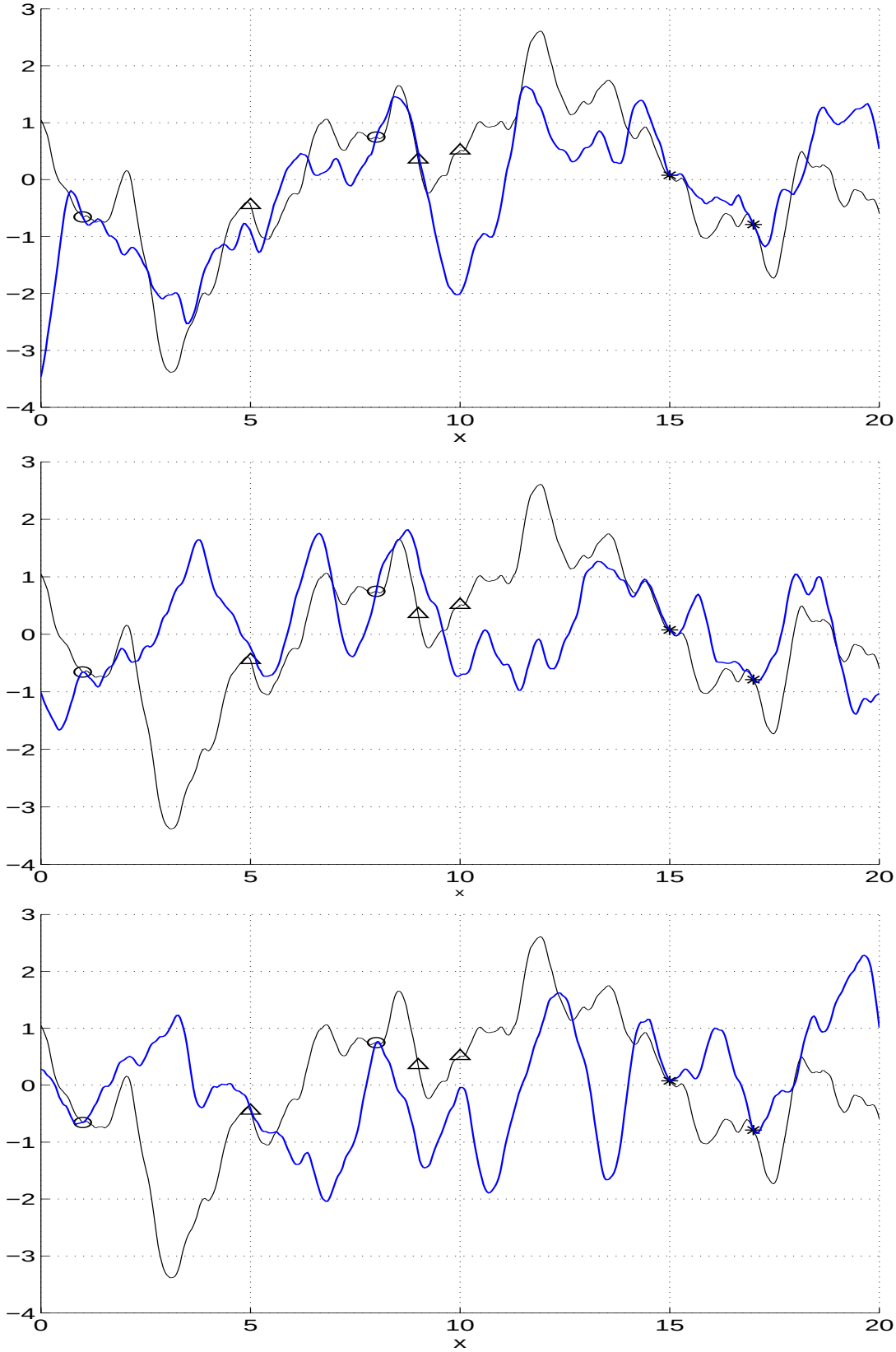


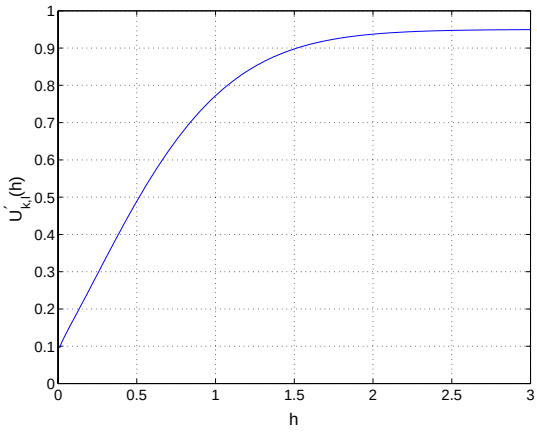
Figure B.3.: Three conditional simulation of a process (for a description see page 99)

C. Appendix (Figures)

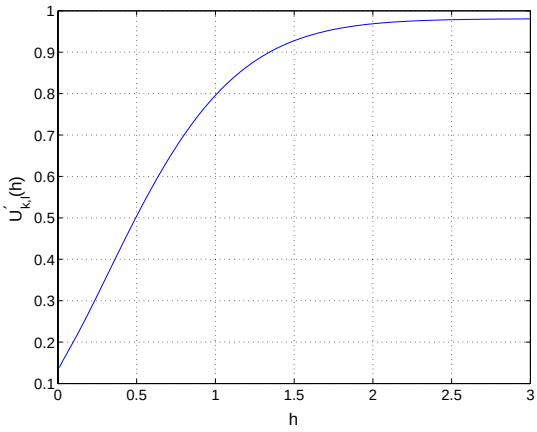
This chapter supplements the empirical results given in Chapter 5. It can be seen that the plots of (5.4) for the Gaussian covariance function are analogous to (5.4) for the Matérn class. Therefore, it can be conjectured that some optimally constructed regular grids perform very good for the Matérn class too.

On the other hand, the case of the J-Bessel class is more complicated. For larger ν (see Figure 3.2), an analogy to the Gaussian covariance function can be seen (see Figures C.8 and C.9). For small values of ν, (5.4) strongly varies. Moreover, in some figures, some effects can be observed, which have no analogy under the Gaussian model. In particular, see Figures C.7(b) and C.7(f), and observe that the global maximum is not reached in the origin.

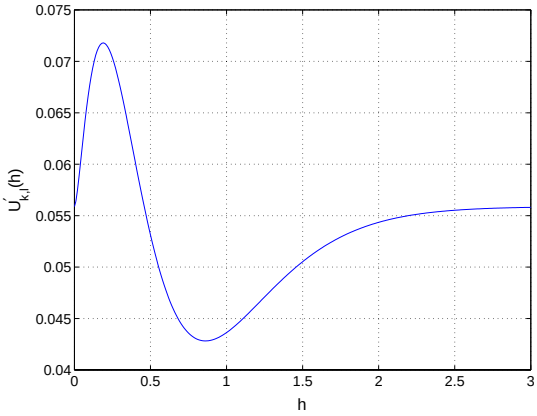
The following plots are based on the results given in Appendix A.6.2 and A.6.3 together with Lemmata 3.1 and 3.3.



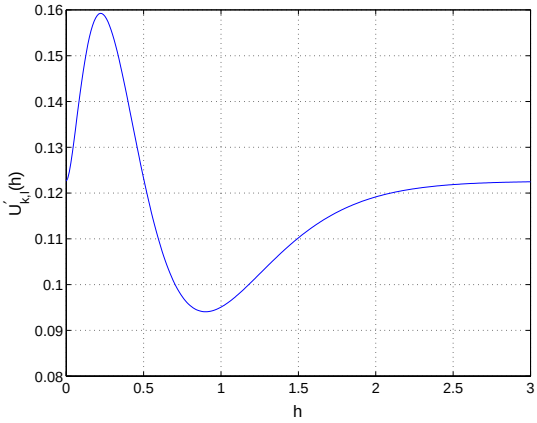
(a) $\nu = 1.1,\ k = 0,\ l = 0$



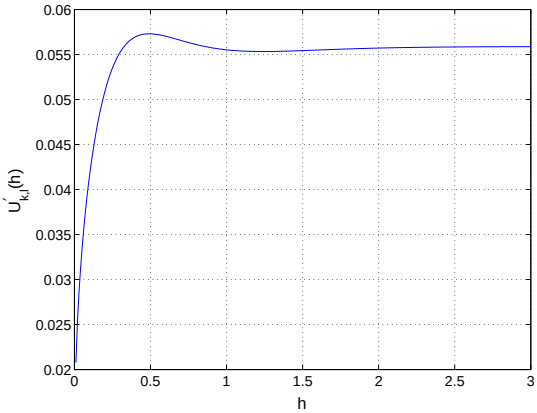
(b) $\nu = 1.25,\ k = 0,\ l = 0$



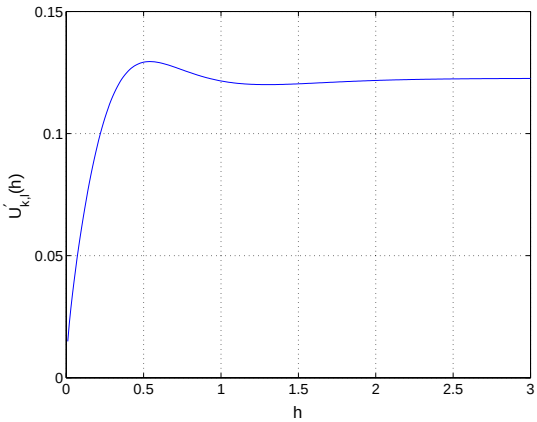
(c) $\nu = 1.1,\ k = 0,\ l = 1$



(d) $\nu = 1.25,\ k = 0,\ l = 1$

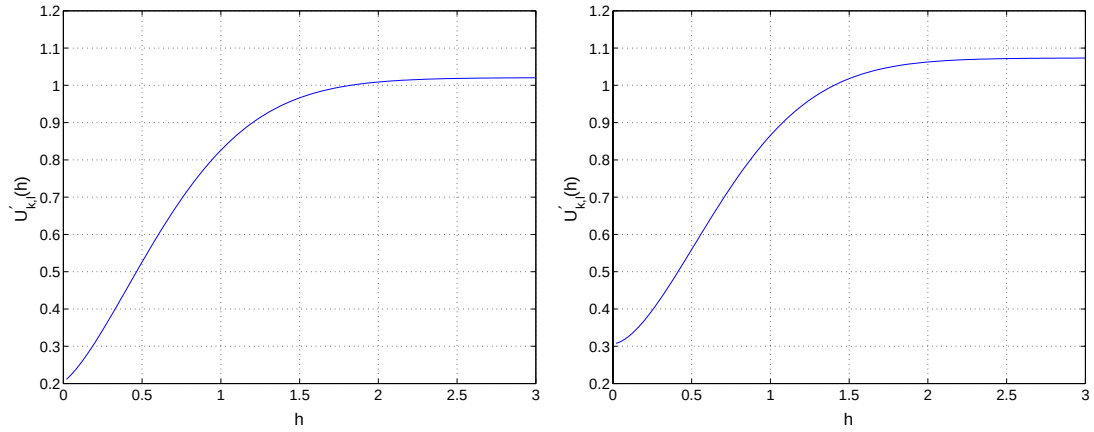


(e) $\nu = 1.1,\ k = 1,\ l = 1$



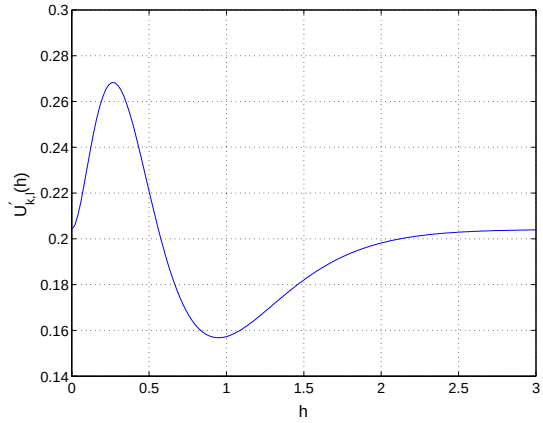
(f) $\nu = 1.25,\ k = 1,\ l = 1$

Figure C.1.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Matérn model (3.10) with $\rho = 1$ and $\sigma^2 = 1$.

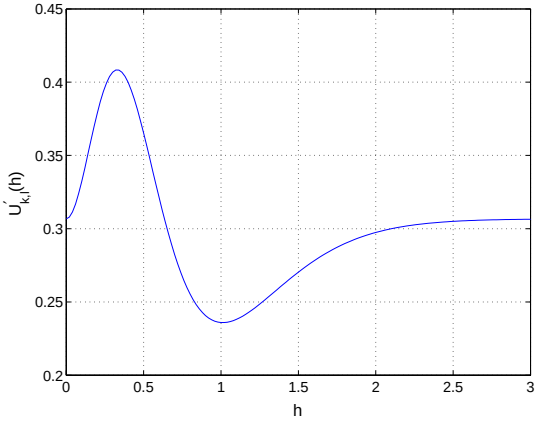


(a) $\nu = 1.5$, $k = 0$, $l = 0$

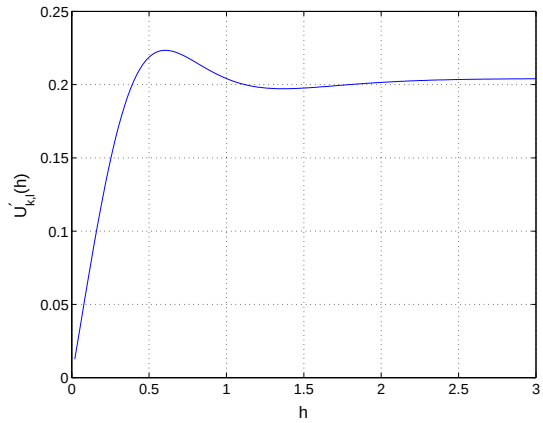
(b) $\nu = 2$, $k = 0$, $l = 0$



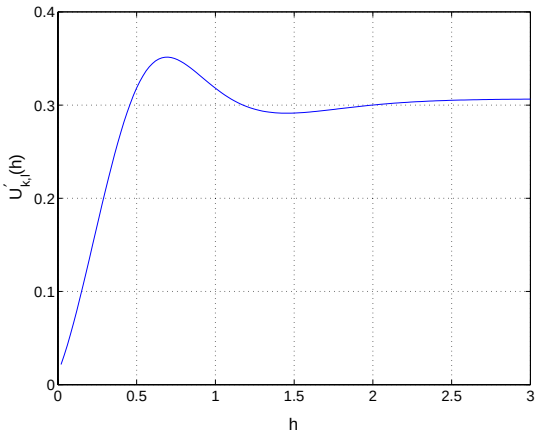
(c) $\nu = 1.5$, $k = 0$, $l = 1$



(d) $\nu = 2$, $k = 0$, $l = 1$

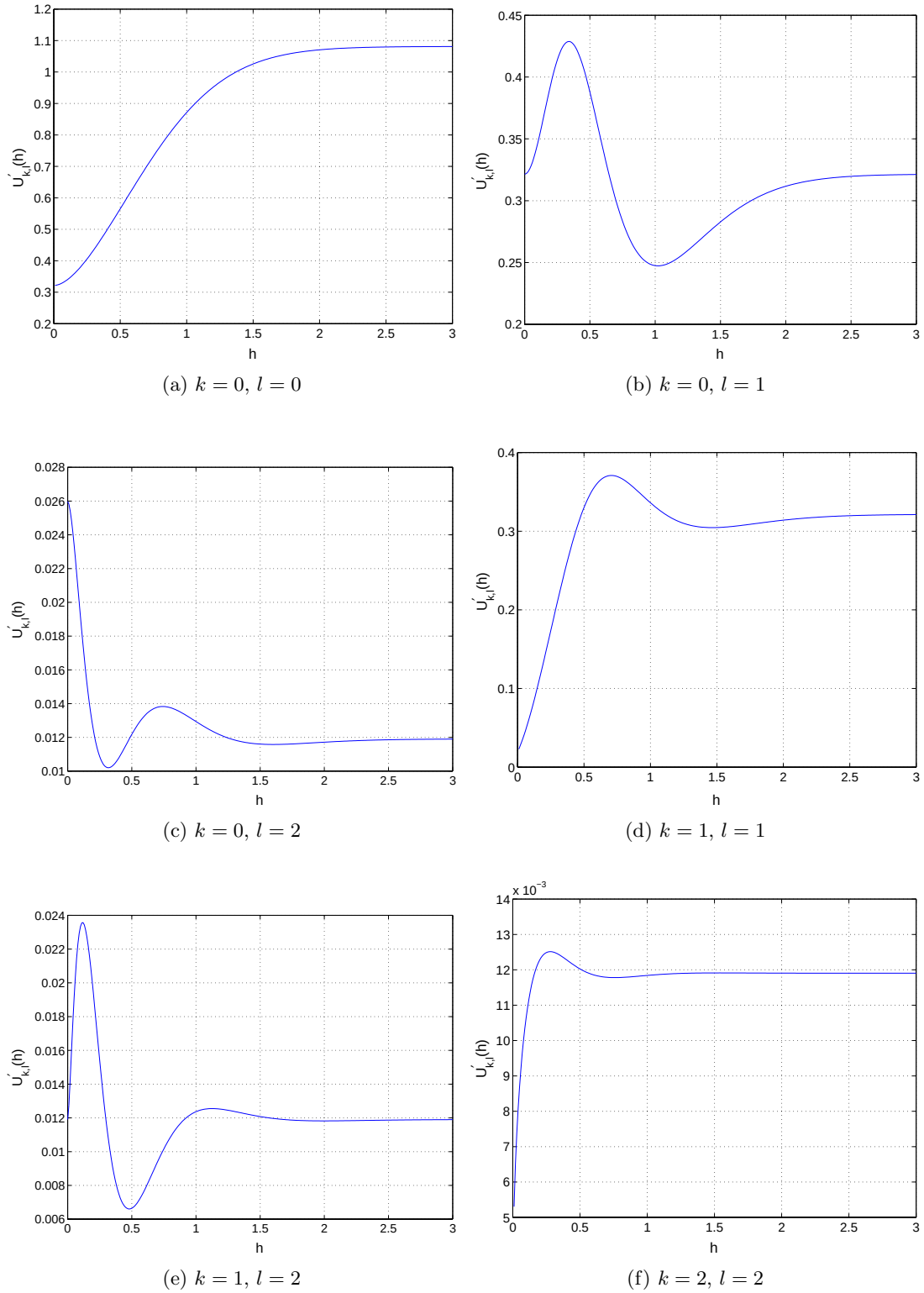


(e) $\nu = 1.5$, $k = 1$, $l = 1$



(f) $\nu = 2$, $k = 1$, $l = 1$

Figure C.2.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Matérn model (3.10) with $\rho = 1$ and $\sigma^2 = 1$.



(a) $k = 0$, $l = 0$

(b) $k = 0$, $l = 1$

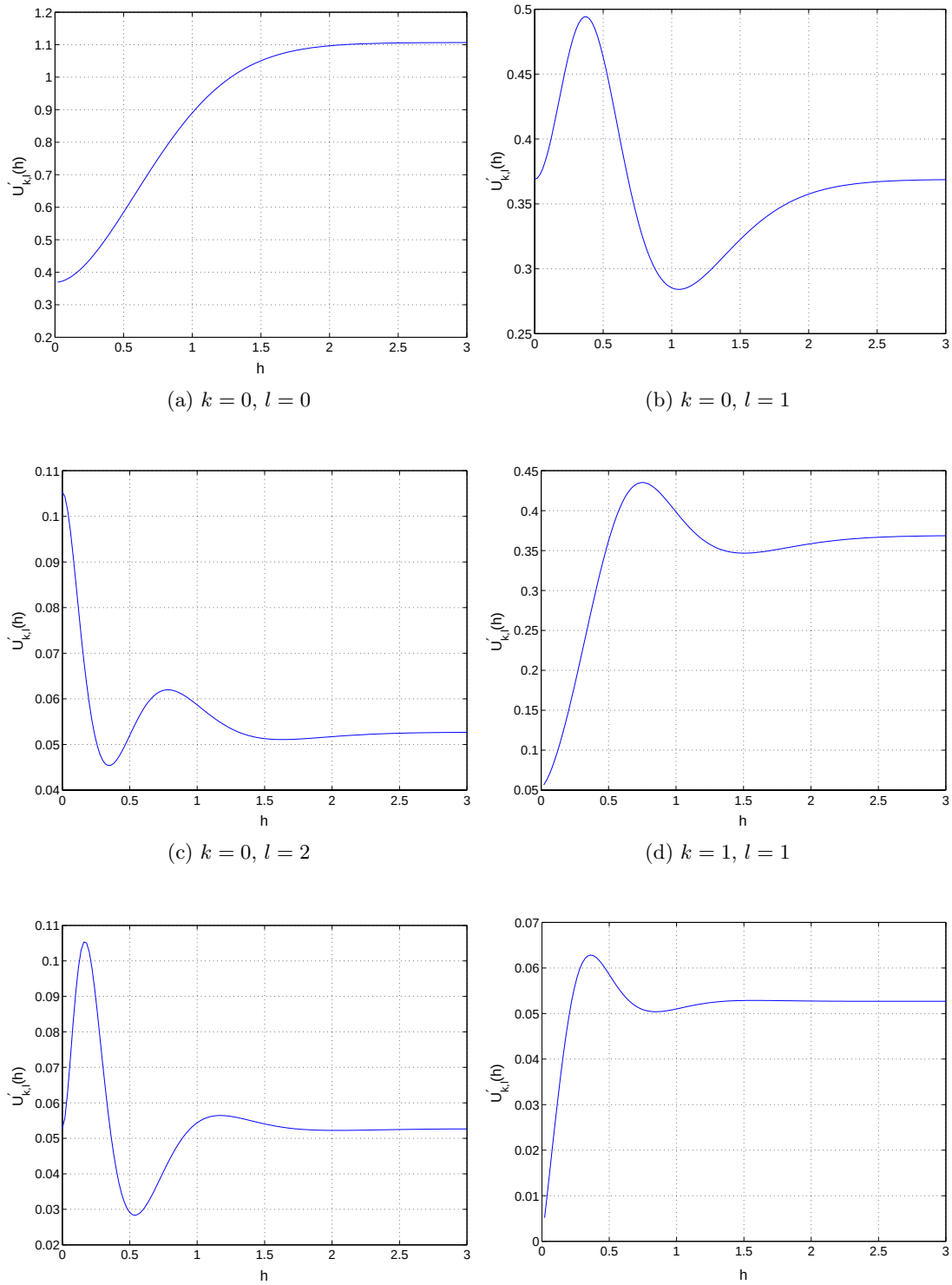
(c) $k = 0$, $l = 2$

(d) $k = 1$, $l = 1$

(e) $k = 1$, $l = 2$

(f) $k = 2$, $l = 2$

Figure C.3.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Matérn model (3.10) with $\nu = 2.1$, $\rho = 1$ and $\sigma^2 = 1$.



(a) $k = 0$, $l = 0$

(b) $k = 0$, $l = 1$

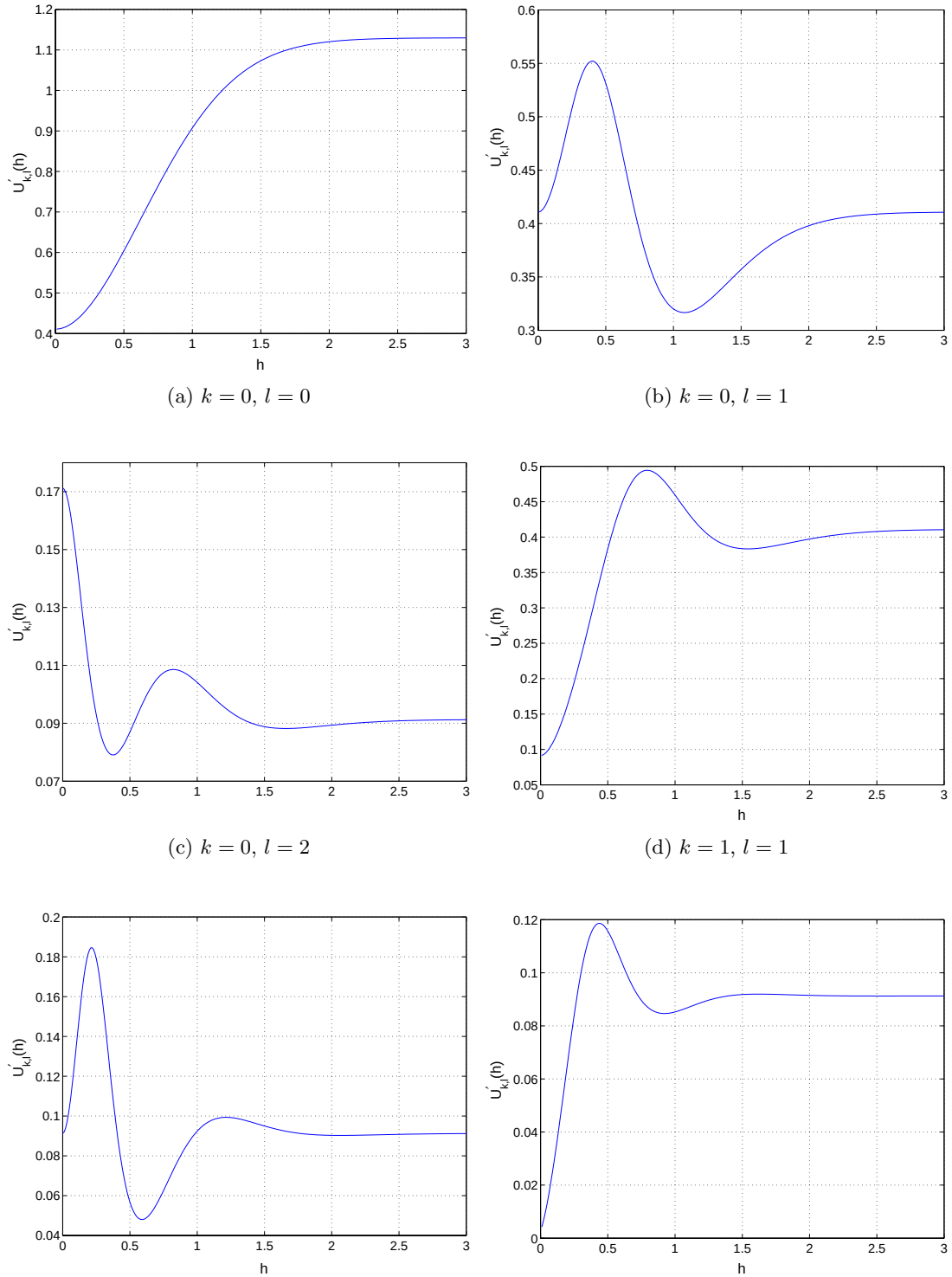
(c) $k = 0$, $l = 2$

(d) $k = 1$, $l = 1$

(e) $k = 1$, $l = 2$

(f) $k = 2$, $l = 2$

Figure C.4.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Matérn model (3.10) with $\nu = 2.5$, $\rho = 1$ and $\sigma^2 = 1$.



(a) $k = 0$, $l = 0$

(b) $k = 0$, $l = 1$

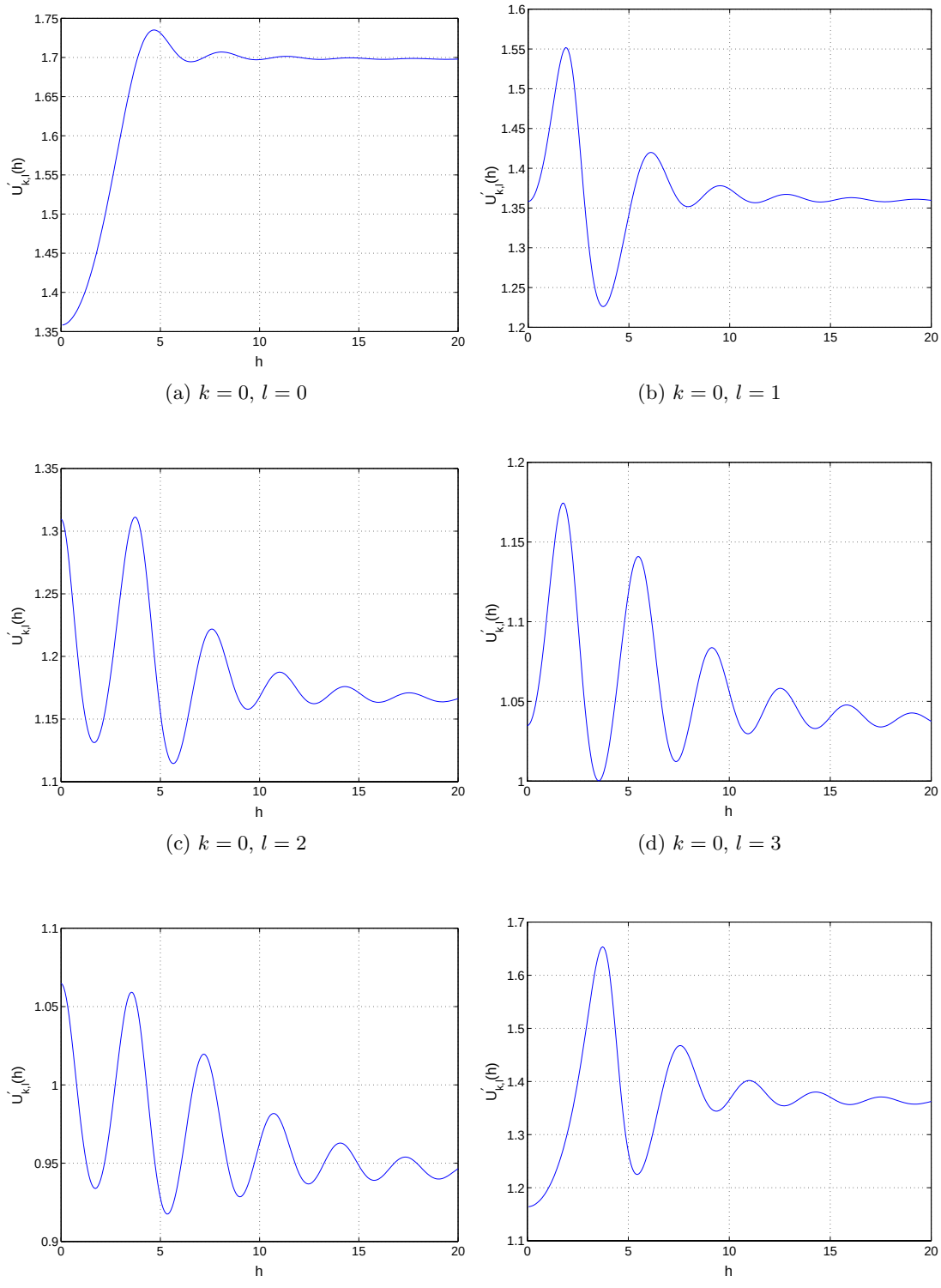
(c) $k = 0$, $l = 2$

(d) $k = 1$, $l = 1$

(e) $k = 1$, $l = 2$

(f) $k = 2$, $l = 2$

Figure C.5.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the Matérn model (3.10) with $\nu = 3$, $\rho = 1$ and $\sigma^2 = 1$.



(a) $k = 0$, $l = 0$

(b) $k = 0$, $l = 1$

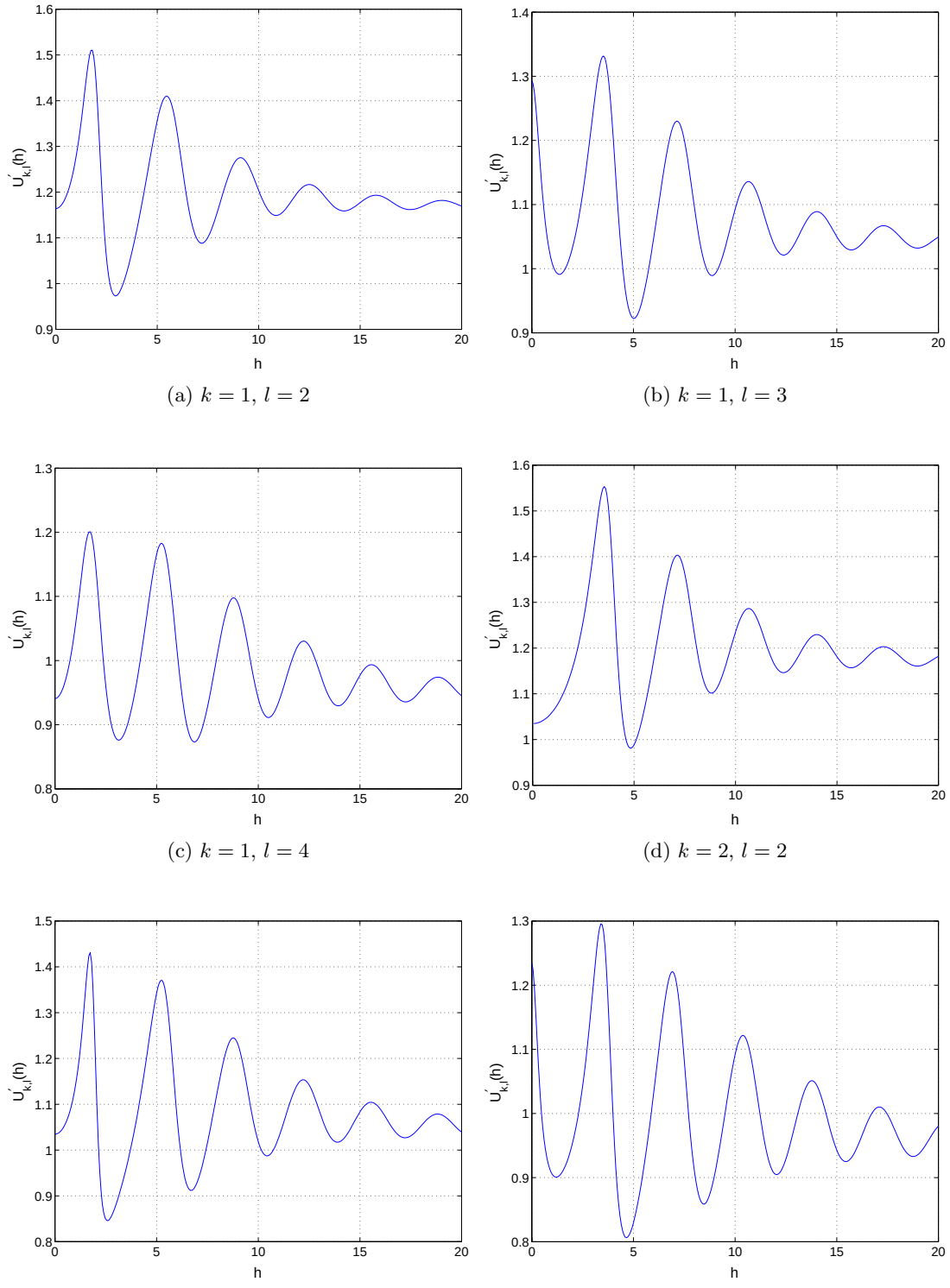
(c) $k = 0$, $l = 2$

(d) $k = 0$, $l = 3$

(e) $k = 0$, $l = 4$

(f) $k = 1$, $l = 1$

Figure C.6.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the J-Bessel model with $\nu = 1$, $\phi = 1$ and $\alpha = 1$.



(a) $k = 1$, $l = 2$

(b) $k = 1$, $l = 3$

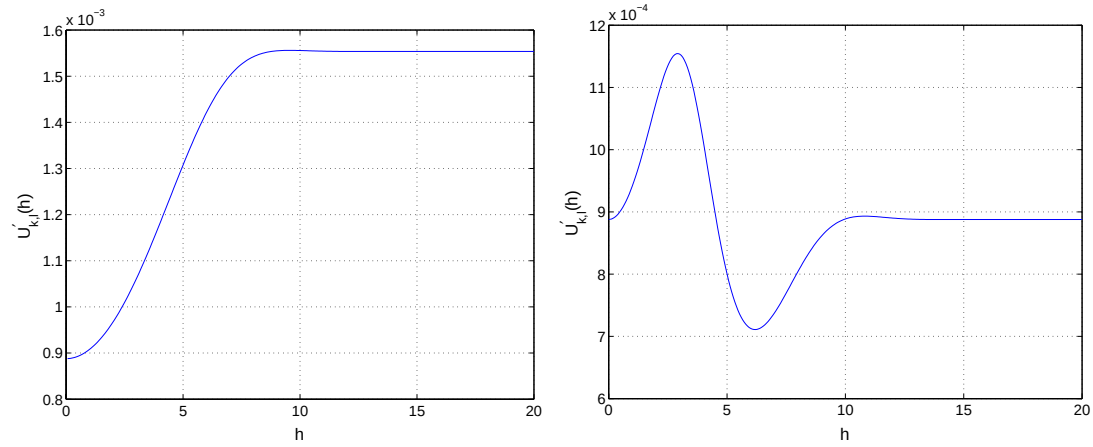
(c) $k = 1$, $l = 4$

(d) $k = 2$, $l = 2$

(e) $k = 2$, $l = 3$

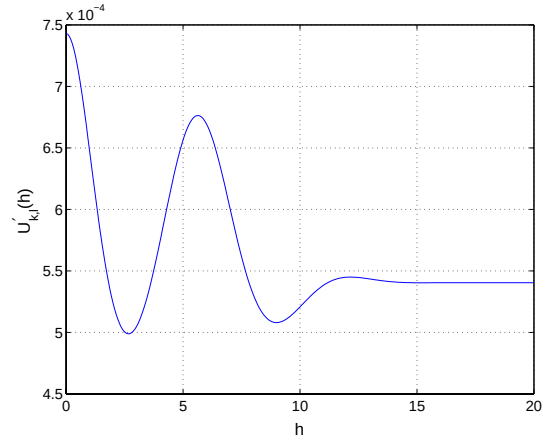
(f) $k = 2$, $l = 4$

Figure C.7.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the J-Bessel model (3.23) with $\nu = 1$, $\phi = 1$ and $\alpha = 1$.

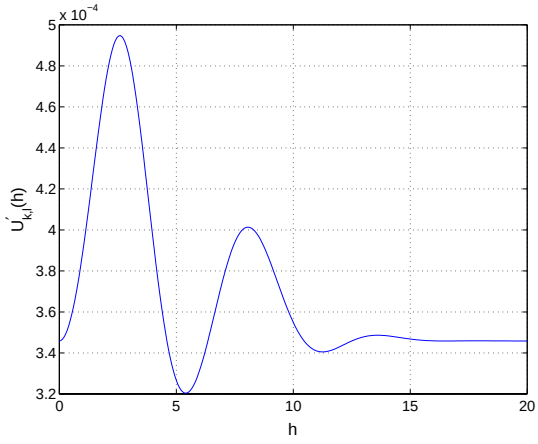


(a) $k = 0$, $l = 0$

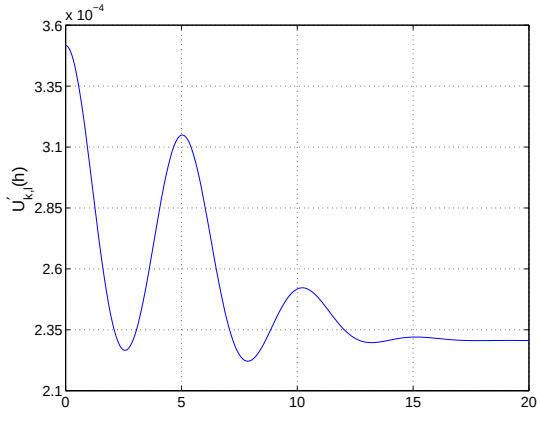
(b) $k = 0$, $l = 1$



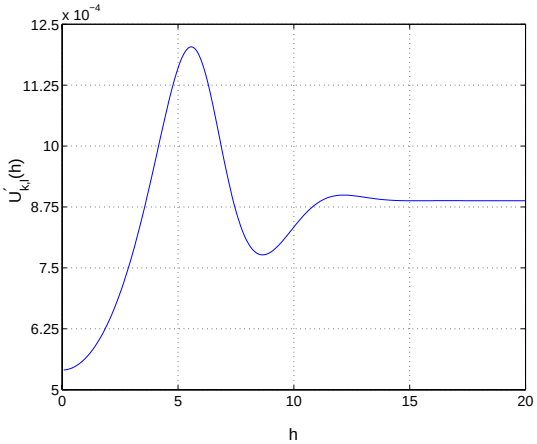
(c) $k = 0$, $l = 2$



(d) $k = 0$, $l = 3$

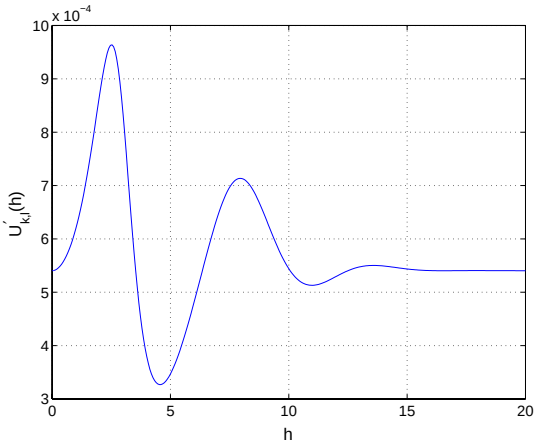


(e) $k = 0$, $l = 4$

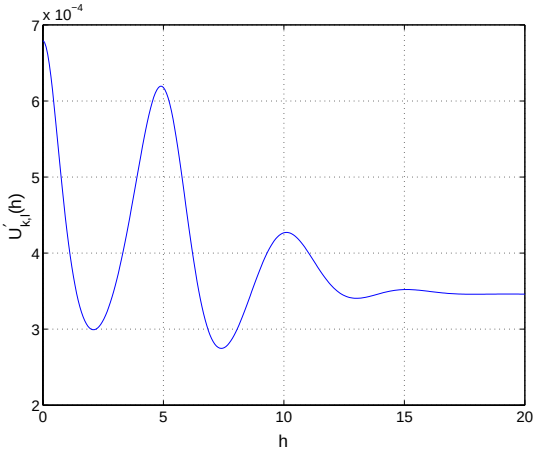


(f) $k = 1$, $l = 1$

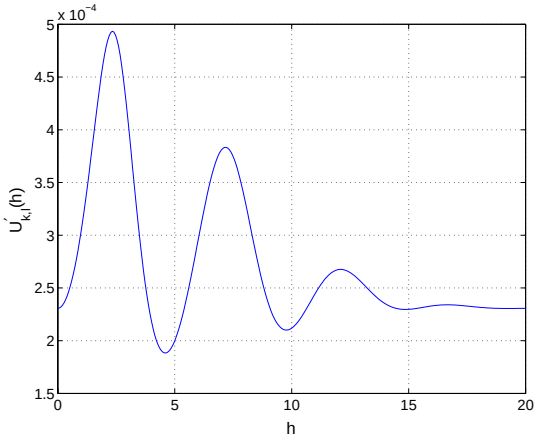
Figure C.8.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the J-Bessel model (3.23) with $\nu = 5$, $\phi = 1$ and $\alpha = 1$.



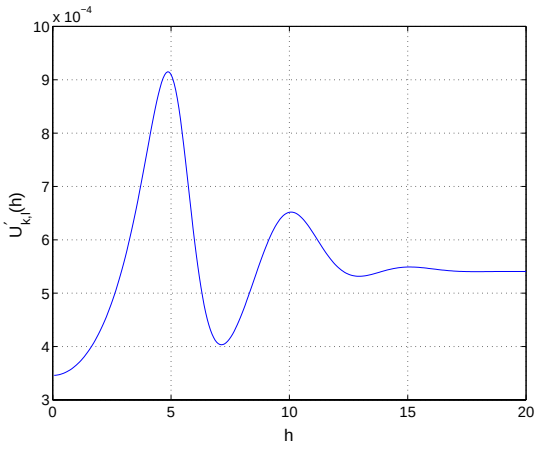
(a) $k = 1$, $l = 2$



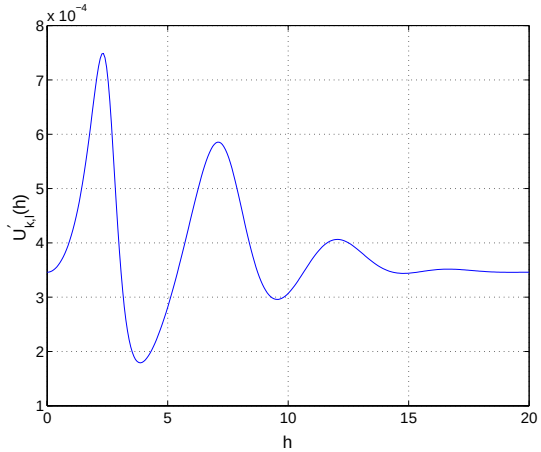
(b) $k = 1$, $l = 3$



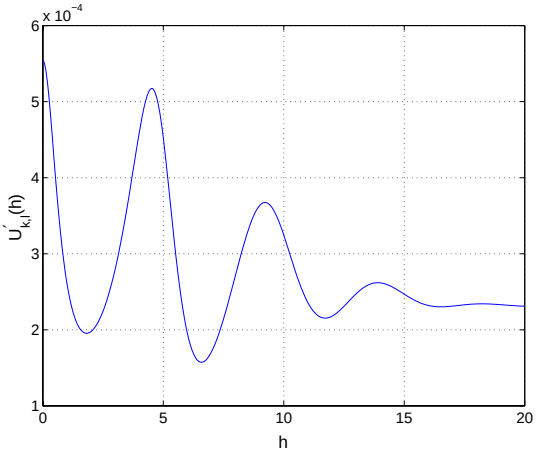
(c) $k = 1$, $l = 4$



(d) $k = 2$, $l = 2$



(e) $k = 2$, $l = 3$



(f) $k = 2$, $l = 4$

Figure C.9.: Plots of the imse-update (5.4) for two observations $Z^{(k)}(0)$ and $Z^{(l)}(h)$ under the J-Bessel model (3.23) with $\nu = 5$, $\phi = 1$ and $\alpha = 1$.

Bibliography

Abramowitz, M. and Stegun, I., editors (1972). *Handbook of mathematical functions*. Dover Publications, Inc., New York.

Abt, M. (1992). Some exact optimal designs for linear covariance functions in one dimension. *Communications in statistics / Theory and methods*, 21(7):2059–2069.

Abt, M. (1999). Estimating the prediction mean squared error in gaussian stochastic processes. *Scand. J. Stat.*, 26(4):563–578.

Albrecht, M. (1998). Geostatistische Meßnetzplanung für das Gradientenkriging. Forschungsteilbericht zum Thema "Markscheiderisch-geologische Komplexauswertung unter geostatistischen Modellannahmen" DFG III 03 - Me 1111/2-1, TU Bergakademie Freiberg.

Aldworth, J. and Cressie, N. (1999). Sampling designs and prediction methods for Gaussian spatial processes. In Ghosh, S., editor, *Multivariate analysis, design of experiments, and survey sampling*, pages 1–54. Marcel Dekker.

Aspie, D. and Barnes, R. (1990). Infill-sampling design and the cost of classification errors. *Mathematical Geology*, 22(8):915–932.

Bandemer, H. a. (1977). *Theorie und Anwendung der optimalen Versuchsplanung I. Handbuch zur Theorie.* Akademie-Verlag, Berlin.

Barnes, R. J. (1989). Sample design for geologic site characterisation. In Armstrong, M., editor, *Geostatistics*, volume 2, pages 809–822. Kluwer Academic Publisher.

Batsell, S., Fedorov, V., and Flanagan, D. (1998). Multivariate prediction: Selection of the most informative components to measure. In Atkinson, A., Pronzato, L., and Wynn, H., editors, *MODA 5 - Advances in Model-Oriented Data Analysis end Experimental Design*, pages 215–222.

Benhenni, K. and Cambanis, S. (1992). Sampling designs for estimating integrals of stochastic processes using quadratic mean derivatives. *Lect. Notes Pure Appl. Math.*, 138:93–123.

Biedermann, S. and Dette, H. (2001). Minimal optimal designs for nonparametric regression - a further optimality property of the uniform distribution. In Atkinson, A., Hackl, P., and Müller, W., editors, *MODA 6 - Advances in Model-Oriented Data Analysis and Experimental Design*, pages 13–20.

Chilès, J. and Delfiner, P. (1999). *Geostatistics*. Academic Press, New York.

Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian prediction of deterministic function with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963.

Daley, R. (1996). *Atmospheric data analysis*. Cambridge University Press, Cambridge.

Deutsch, C. and Journel, A. (1992). *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Fedorov, V. (1996). Design of spatial experiments: Model fitting and prediction. In Ghosh, S. and Rao, C., editors, *Handbook of Statistics: Design and Analysis of Experiment*, volume 13, pages 515–553. Elsevier Science B.V.

Fletcher, R. (1991). *Practical Methods of Optimization*. Wiley, Chichester, 2 edition.

Freeden, W. (1999). *Multiscale modelling of spaceborne geodata*. Teubner, Stuttgart.

Gao, H., Wang, J., and Zhao, P. (1996). The updated kriging variance and optimal sample design. *Mathematical Geology*, 28(3):295–313.

Gaspari, G. and Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, 125:723–757.

Gikhman, I. and Skorokhod, A. (1996). *Introduction to the Theory of Random Processes*. W. B. Saunders Company, Philadelphia.

Gneiting, T. (1999a). Correlation functions for atmospheric data analysis. *Q. J. R. Meteorol. Soc.*, 125:2449–2464.

Gneiting, T. (1999b). On the derivatives of radial positive definite functions. *Journal of Mathematical Analysis and Applications*, 236:86–93.

Gneiting, T. (2000). Compactly supported correlation functions. Technical Report TRS No. 045, NRCSE.

Gradstein, I. and Ryshik, I. (1981a). *Summen-, Produkt- und Integral- Tafeln*, volume 2. Verlag Harri Deutsch, Thun.

Gradstein, I. and Ryshik, I. (1981b). *Summen-, Produkt- und Integral- Tafeln*, volume 1. Verlag Harri Deutsch, Thun.

Handcock, M. and Wallis, J. (1994). An approach to statistical spatial temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association*, 89:368–390.

Harville, D. (1985). Decomposition of prediction error. *Journal of the American Statistical Association*, 80(389):132–138.

Harville, D. and Jeske, D. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87(419):724–731.

Holtmann, M. (2001). *Effektive Beobachtung von zufälligen Funktionen unter besonderer Berücksichtigung von Ableitungen*. Dissertation, TU Bergakademie Freiberg.

Jäkel, M. (1997). Gradientengewinnung in den Geowissenschaften. Forschungsteilbericht zum Thema "Markscheiderisch-geologische Komplexauswertung unter geostatistischen Modellannahmen" DFG III 03 - Me 1111/2-1, TU Bergakademie Freiberg.

Johnson, M., Moore, L., and Ylvisaker, D. (1990). Minimax and maximin distance design. *Journal of Statistical Planning and Interference*, 26:131–148.

Jordan, W., Eggert, O., and Kneißl, M. (1967). *Handbuch der Vermessungskunde*. Metzler, Stuttgart.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics*. Academic Press, London, 7 edition.

Khuri, A. and Cornell, J. (1996). *Response surfaces: design and analysis*. Marcel Dekker, New York.

Koehler, J. and Owen, A. (1996). Computer experiments. In Ghosh, S. and Rao, C., editors, *Handbook of Statistics: Design and Analysis of Experiments*, volume 13, pages 261–308. Elsevier Science B.V.

Kosmol, P. (1993). *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben*. Teubner, Stuttgart, 2 edition.

Kuttler, K. (1998). *Modern analysis*. CRC Press, Boca Raton.

Lütkepohl, H. (1996). *Handbook of matrices*. Wiley.

Mitchell, T., Morris, M., and Ylvisaker, D. (1994). Asymptotically optimum experimental designs for prediction of deterministic functions given derivative information. *Journal of Statistical Planning and Interference*, 41:377–389.

Morris, M., Mitchell, T., and Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255.

Müller, W. (2001). *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*. Physica-Verlag, Heidelberg, 2 edition.

Näther, W. (1985). *Effective Observation of Random Fields*. Teubner, Leipzig.

Näther, W. and Šimák, J. (2001). Effective observation of random processes using derivatives. *submitted to Metrika*, pages 1–18.

Pázman, A. (1986). *Foundations of optimum experimental design*. Reidel, Dordrecht.

Prudnikov, A., Brychkov, Y., and Marichev, O. (1990). *Integrals and series: Special functions*, volume 2. Gordon and Breach science Publishers, Thun, 3 edition.

Prudnikov, A., Brychkov, Y., and Marichev, O. (1996). *Integrals and series: Elementary functions*, volume 1. Gordon and Breach science Publishers, Thun, 3 edition.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Rao, R. (1973). *Linear statistical inference and its applications*. Wiley, New York.

Rao, R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives*. Springer, New York.

Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Röttig, A. (1996). *Gradientenkriging - eine integrierende geostatische Methode zur einheitlichen Auswertung von absoluten und relativen Meßdaten*. Dissertation, TU Bergakademie Freiberg.

Sacks, J. and Schiller, S. (1988). Spatial designs. *Statistical decision theory and related topics IV, Pap. 4th Purdue Symp.; West Lafayette/Indiana 1986*, 2:385–399.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments (with discussion). *Statistical Science*, 4(4):409–435.

Sacks, J. and Ylvisaker, D. (1966). Design for regression problems with correlated errors. *The Annals of Mathematical Statistics*, 37:66–89.

Sacks, J. and Ylvisaker, D. (1968). Design for regression problems with correlated errors; many parameters. *The Annals of Mathematical Statistics*, 39(1):49–69.

Sacks, J. and Ylvisaker, D. (1970). Design for regression problems with correlated errors III. *The Annals of Mathematical Statistics*, 41(6):2057–2074.

Schott, J. (1997). *Matrix Analysis for Statistics*. Wiley, New York.

Schwabe, R. (1996). *Optimum Designs for Multi-Factor Models*. Springer-Verlag, New York.

Skorokhod, A. (1996). *Lectures on the Theory of Stochastic Processes*. VSP, Utrecht, The Netherlands.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.

Todorovic, P. (1992). *An introduction to stochastic processes and their applications*. Springer-Verlag, New York.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.*, 4:389–396.

Wendland, H. (1996). *Konstruktion und Untersuchung radialer Basisfunktionen mit kompaktem Träger*. Dissertation, University of Göttingen.

Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer-Verlag, Berlin.

Yaglom, A. (1987a). *Correlation Theory of Stationary and Related Random Functions I*. Springer-Verlag, New York.

Yaglom, A. (1987b). *Correlation Theory of Stationary and Related Random Functions II*. Springer-Verlag, New York.

Yfantis, E., Flatman, G., and Behar, J. (1987). Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology*, 19(3):183–205.

Zabinsky, Z. (1998). Stochastic methods for practical global optimization. *Journal of Global Optimization*, 13(4):433–444.

Zimmerman, D. and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, 44(1):27–43.