SODANKYLÄ GEOPHYSICAL OBSERVATORY PUBLICATIONS

OULUN YLIOPISTO
UNIVERSITY of OULU

No. 109

DISCRETISATION-INVARIANT AND COMPUTATIONALLY EFFICIENT CORRELATION PRIORS FOR BAYESIAN INVERSION

LASSI ROININEN

Sodankylä 2015

SODANKYLÄ GEOPHYSICAL OBSERVATORY
PUBLICATIONS

OULUN YLIOPISTO
UNIVERSITY of OULU

No. 109

DISCRETISATION-INVARIANT AND COMPUTATIONALLY EFFICIENT CORRELATION PRIORS FOR BAYESIAN INVERSION

LASSI ROININEN

Academic dissertation
University of Oulu Graduate School
Faculty of Science

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu, in Polaria lecture hall of the Sodankylä Geophysical Observatory on 16 June 2015 at 12 o'clock.

Sodankylä 2015

SODANKYLÄ GEOPHYSICAL OBSERVATORY PUBLICATIONS
Editor: Dr Thomas Ulich
Sodankylä Geophysical Observatory
FI-99600 SODANKYLÄ, Finland

This publication is the continuation of the former series
"Veröffentlichungen des geophysikalischen Observatoriums
der Finnischen Akademie der Wissenschaften"

Sodankylä Geophysical Observatory Publications

ISBN 978-952-62-0753-7 (paperback)
ISBN 978-952-62-0754-4 (pdf)
ISSN 1456-3673

OULU UNIVERSITY PRESS
Sodankylä 2015

Abstract

We are interested in studying Gaussian Markov random fields as correlation priors for Bayesian inversion. We construct the correlation priors to be discretisation-invariant, which means, loosely speaking, that the discrete priors converge to continuous priors at the discretisation limit. We construct the priors with stochastic partial differential equations, which guarantees computational efficiency via sparse matrix approximations. The stationary correlation priors have a clear statistical interpretation through the autocorrelation function.

We also consider how to make structural model of an unknown object with anisotropic and inhomogeneous Gaussian Markov random fields. Finally we consider these fields on unstructured meshes, which are needed on complex domains.

The publications in this thesis contain fundamental mathematical and computational results of correlation priors. We have considered one application in this thesis, the electrical impedance tomography. These fundamental results and application provide a platform for engineers and researchers to use correlation priors in other inverse problem applications.

Keywords: Bayesian statistical inverse problems, Gaussian Markov random fields, convergence, discretisation, stochastic partial differential equations

Acknowledgments

This study has been carried out in the Sodankylä Geophysical Observatory of the University of Oulu, Finland. I thank Professor Markku Lehtinen for providing scientific problems, finance and supervision of the thesis. I would also like to thank Professor Valeri Serov for being the second supervisor of the thesis. For evaluating the thesis, I thank Professor Håvard Rue for pre-examination and for being the first opponent, Professor Jouko Lampinen for being the second opponent, and, Professor Ville Kolehmainen for pre-examination.

This thesis is based on four articles. I thank the co-authors of these articles: Dr Sari Lasanen, Dr Petteri Piiroinen, Dr Mikko Orispää, Dr Janne Huttunen, Dr Simo Särkkä and Mr Markku Markkanen. I thank Docent Jyrki Manninen and Dr Markus Harju for the PhD thesis follow-up group work. I have had the pleasure to work with all the Observatory staff, Finnish Inverse Problems Society, Centre of Excellence in Inverse Problems Research, EISCAT Scientific Association, Bahir Dar University and all the other colleagues. Hence, thanks to all the collaboration partners. I thank especially Rector Baylie Damtie, Docent Thomas Ulich, Dr Juha Vierinen, Dr Antti Kero, Dr Ilkka Virtanen, Mr Johannes Norberg and Mr Derek McKay-Bukowski.

This thesis has been funded by Academy of Finland through Finnish Centre of Excellence in Inverse Problems Research and Vilho, Yrjö and Kalle Väisälä foundation.

Contents

Original publications

This thesis consists of an introductory part and the following original papers:

I L. Roininen, M. Lehtinen, S. Lasanen, M. Orispää and M. Markkanen, *Correlation priors*, Inverse Problems and Imaging, **5** (2011), 167–184.

II L. Roininen, P. Piiroinen and M. Lehtinen, *Constructing continuous stationary covariances as limits of the second-order stochastic difference equations*, Inverse Problems and Imaging, **7** (2013), 611–647.

III L. Roininen, J. M. Huttunen and S. Lasanen, *Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography*, Inverse Problems and Imaging, **8** (2014), 561–586.

IV L. Roininen, S. Lasanen, M. Orispää and S. Särkkä, *Sparse approximations of fractional Matérn fields*, Scandinavian Journal of Statistics, submitted March 2015.

In the text, the original papers will be referred to by their Roman numerals.

The contributions of the author to the original publications are as follows:

Paper I: Correlation priors

This is the fundamental paper of this thesis and provides all the notations of one-dimensional correlation priors. The construction is based on stochastic difference equations and on the discussion of how the discrete autocovariance function converges to the continuous autocovariance function at the discretisation limit. The author presented the fundamental idea of the paper and has written significant parts of the paper. All numerical simulations were carried out by the author.

Paper II: Constructing continuous stationary covariances as limits of the second-order stochastic difference equations

The strongest convergence results of the correlation priors are presented in this paper. We discuss a number of items: Discretisation schemes of stochastic processes, the strong-weak convergence of probability measures, discretisation-invariance in Bayesian inversion and computation of a number of autocovariance functions of the correlation priors. The starting point for this paper are the stochastic partial differential equations, in contrast to stochastic difference equations in Paper I. The author was behind the original idea of the paper. Dr Petteri Piiroinen developed the convergence theorems and discussion on discretisation-invariance. The author contributed especially to the formulation of continuous and discrete real and complex models and to the calculation of autocovariance functions.

Paper III: Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography

We showed how to apply correlation prior formalism to electrical impedance tomography. Here we use finite element discretisation schemes instead of lattice approximations. The author was behind the original idea of implementing the correlation priors on finite element meshes and has written significant parts of the text. Dr Janne Huttunen is responsible for all the finite element simulations.

Paper IV: Sparse approximations of fractional Matérn fields

In this paper, we considered how to efficiently approximate Matérn fields with fractional spectrum. The study is based on Taylor approximations and studying band-limited spectrum. For the discretisation, we use the same methods as in Paper I. Convergence results are done by Dr Sari Lasanen. The author formulated the original idea of the paper and has written significant parts of the paper.

Chapter 1

Introduction

Inverse problems are the mathematical theory and practical interpretation of noise-perturbed indirect observations. The specific field of Bayesian statistical inverse problems is the effort to formulate real-world inverse problems as Bayesian statistical estimation problems [3, 11, 33]. Applications include for example atmospheric remote sensing, near-space studies, medical imaging and ground prospecting.

In Bayesian inversion, a priori probability distribution is, in practice, the only tuneable parameter in the estimation algorithm. Prior distribution is subjective information of the unknown before any actual measurements are done. The better prior information we have, the better estimates we will get.

In this thesis, we consider sparse matrix approximations of continuous Gaussian Markov random field priors, the *correlation priors*. These priors have three benefits:

1. They have a *clear statistical interpretation* as stationary Gaussian random fields.
2. The priors can be represented as systems of stochastic partial differential equations and approximated with sparse difference matrices, hence providing *computational efficiency.*
3. The priors can be constructed to be *discretisation-invariant*, which means, loosely speaking, that the discrete covariances converge to continuous covariances at the discretisation limit.

We define Gaussian Markov random fields as zero-mean stationary random fields with a covariance function

$$\mathcal{C}(x,x') = \mathcal{C}(x-x') = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \frac{1}{\sum_{k=0}^K c_k|\xi|^{2k}} \exp(-i\xi\cdot(x-x'))d\xi,$$

where $x, x' \in \mathbb{R}^d$, $d = 1, 2, \ldots$ is the dimensionality, $c_0 > 0$ and $c_k \in \mathbb{R}$ and $K > 0$ is some integer. The polynomial $P(\xi) := \sum_{k=0}^K c_k|\xi|^{2k} > 0$ is our object of interest, because we can relate the polynomial to the study of stochastic partial differential equations. Consider $P(\xi) = (1+\xi^2)^2$, where $\xi \in \mathbb{R}$. We note that this corresponds to a stochastic partial differential equation of the form

$$\sqrt{P(\xi)}\widehat{\mathcal{X}} = (1+\xi^2)\widehat{\mathcal{X}} = \widehat{\mathcal{W}} \quad \Leftrightarrow \quad (1-\Delta)\mathcal{X} = \mathcal{W}, \tag{1.1}$$

where $\mathcal{X}$ is the unknown of interest, $\mathcal{W}$ is white noise, Δ is the Laplacian, and 'hat'-notation denotes Fourier-transformed objects.

We can also calculate the autocorrelation function of the random field defined through Equation (1.1) in closed-form. Then the autocorrelation function is

$$\mathcal{C}(x-x') = \frac{1}{4}\left(1+|x-x'|\right)\exp\left(-|x-x'|\right).$$

Let us denote by the vector X the discrete approximation of the continuous unknown $\mathcal{X}$. By using finite-differences, we can make a discrete approximation of $\mathcal{X}$ in Equation (1.1) as

$$X_j - \frac{X_{j-1} - 2X_j + X_{j+1}}{h^2} = W_j \sim \mathcal{N}\left(0, \frac{1}{h}\right), \tag{1.2}$$

where we have used the discretisation $x \approx jh$, where $j \in \mathbb{Z}$ and discretisation step $h > 0$. Equation (1.2) can be written as a sparse matrix approximation $LX = W$.

Later, we will show that the covariance of the discrete approximation (1.2) converges to the continuous covariance at the discretisation limit $h \to 0$. Lasanen 2012 [13, 14] showed that if the prior distributions converge, then in most cases the posterior distributions also converge. Hence, from a computational point of view this means that the posterior estimates are essentially independent of the discretisation, i.e. estimators in different lattices look essentially the same.

These concepts comprise the scope of this thesis.

Outline of the thesis

This thesis is organised as follows:

Chapter 2 contains the main results of the thesis. We start by giving an introduction to Bayesian statistical inverse problems and make preliminary notes on Gaussian stationary fields. Then, we define Gaussian Markov random fields in the continuous domain. After that, we discuss discretisation of white noise and Gaussian Markov random fields. We form the autocovariances defined through stochastic partial difference equations. We then consider how the discrete autocovariances converge to the continuous ones. Then we consider band-limited approximations.

In Chapter 3, we consider how to form anisotropic and inhomogeneous priors. We put an emphasis on how to model structural properties of an unknown field with correlation lengths. Numerical examples are given. Finally we study the finite element approach to correlation priors, i.e. how to construct the priors on unstructured meshes.

In Chapter 4, we conclude the study and make some suggestion for future research.

Chapter 2

Continuous and discrete correlation priors

We are interested in modelling prior distributions as function-valued stochastic processes and fields. For this purpose, we consider Gaussian Markov random fields, as we can construct them through sparse matrix approximations [19, 24, 26]. Our specific interest is in the interplay between the sparse matrix approximations and the continuous random fields. We will start by defining basic concepts of Bayesian statistical inverse problems and a number of concepts of stochastics, and then discuss the continuous and discrete Gaussian Markov random fields.

2.1 Bayesian statistical inverse problems

In a typical Bayesian statistical inverse problem, the objective is to estimate the posterior distribution of an unknown object $\mathcal{X}$ from its noise-perturbed indirect observations. Formally the observations are described as

$$m = \mathcal{A}(\mathcal{X}) + e, \tag{2.1}$$

where m is a known finite-dimensional vector of measurements, $\mathcal{A}$ is a known linear or a non-linear mapping between function spaces, e.g. between separable Hilbert spaces and a projection to some vector-space. The noise term e is a Gaussian random vector with known statistical properties, i.e. we know its mean and covariance.

The solution of a Bayesian statistical inverse problem is an a posteriori probability distribution. For a definition of posterior distributions and the required conditions for Bayes' formula in infinite-dimensional spaces, see Lasanen [13]. We give it as

$$D(\mathrm{d}\mathcal{X}|m) = \frac{D(m|\mathcal{X})}{D(m)} D(\mathrm{d}\mathcal{X}).$$

The notation $\mathrm{d}\mathcal{X}$ refers to integration with respect to a measure. $D(m|\mathcal{X})$ is the likelihood density describing the observations and the statistical properties of noise. $D(m)$ is the probability density of the observations which, for fixed m, can be treated as a normalisation constant. The prior distribution $D(\mathrm{d}\mathcal{X})$ reflects our subjective information on the unknown $\mathcal{X}$ before any actual measurements are taken.

We are interested in sparse approximations of Gaussian Markov random field priors. These priors can be best modelled with continuous models, as we can give autocorrelation functions in closed form. Later in this chapter, we consider discretised Gaussian Markov random field priors and the convergence of the discrete GMRF autocorrelation function to the continuous autocorrelation function. Lasanen 2012 [13, 14] showed that provided the convergence of discrete Gaussian prior covariances, then also the posterior solutions converge, i.e. the solutions of Bayesian statistical inverse problems converge also. From the computational point of view, this means in practice that solutions to inverse problems on different (dense enough) computational meshes are practically the same. We refer to this property as the discretisation-invariant Bayesian statistical inversion.

As noted in Paper II, the concept of discretisation-invariance, or discretisation-independence, is still under debate. The literature on discretisation-invariant inversion is rather vast, see for example [9, 12, 15, 13, 14, 18, 23]. Edge-preserving discretisation-invariance studies include e.g. [16, 17]. From the point of view of this thesis, discretisation-invariance falls down to the category of studying the interplay between discrete and continuous equations [4, 5, 6, 28, 29, 30].

For a practical solution of an inverse problem, we have to apply some numerical discretisation scheme to the observations (2.1) and to the prior. Let us denote the discrete approximation of $\mathcal{X}$ by X and the discrete version of the continuous observation model in Equation (2.1) as

$$m = A(X) + e, \quad e \sim \mathcal{N}(0, \Sigma).$$

Then, the finite-dimensional posterior distribution is

$$D(X|m) = \frac{D(m|X)}{D(m)} D(X) \propto D(m|X)D(X), \tag{2.2}$$

where $D(X)$ is the prior density. Then given discretised observations of (2.1) and the discrete GMRF prior, we can write the posterior distribution as (2.2) as an unnormalised probability density

$$D(X|m) \propto \exp\left(-\frac{1}{2}(m - A(X))^T \Sigma^{-1}(m - A(X)) - \frac{1}{2}(X - \mu)^T C^{-1}(X - \mu)\right),$$

where the prior is distributed as $\mathcal{N}(\mu, C)$. We omit $D(m)$, as it is merely a normalisation constant.

We note that visualisation of the posterior density is difficult if the number of unknowns is higher than two. We follow the standard practice and compute the estimates of the unknown from the posterior distribution. A natural estimate from the posterior distribution is the conditional mean

$$X_{\mathrm{CM}} = \frac{\int X D(m|X)D(X)\mathrm{d}X}{\int D(m|X)D(X)\mathrm{d}X}. \tag{2.3}$$

Another commonly used – and usually computationally simpler estimate – is the maximum a posteriori (MAP) estimate. It is, in a sense, the maximum point of the posterior distribution. The MAP estimate can be computed as a solution of the minimisation problem

$$X_{\mathrm{MAP}} = \arg\min_X \left\{ (m - A(X))^T \Sigma^{-1} (m - A(X)) + (X - \mu)^T C^{-1} (X - \mu) \right\}, \quad (2.4)$$

which can be solved using iterative methods such as Gauss-Newton or sequential quadratic programming methods. A numerically stable and efficient form for the minimisation problem (2.4) can be found by expressing the inverse covariance matrices using matrix square roots (e.g. Cholesky factors)

$$\Sigma^{-1} = H^T H, \quad C^{-1} = L^T L.$$

The minimisation problem can then be cast in a form of a non-linear least-squares problem

$$\min_X \|Z - B(X)\|^2, \quad \text{where } Z = \begin{pmatrix} Hm \\ L\mu \end{pmatrix}, \quad B(X) = \begin{pmatrix} HA(X) \\ LX \end{pmatrix}.$$

This optimisation problem can be efficiently solved with any computational package with for example as a sequence of linearised least-squares problems by applying QR-decomposition or recursive methods such as the GMRES method [8].

We note that in the computation of the estimators (2.3) and (2.4), we typically need to compute the inverse covariance matrices Σ^{-1} and C^{-1}, or their Cholesky factorisations. This is also the case in the simulation of the random field, because we are interested in performing matrix-vector operations of the form $L^{-1}v$, where v is some given vector. In many applications, the error covariance matrix Σ is low-dimensional, possibly even diagonal. Hence, computation of Σ^{-1} is rather easy. However, the number of elements of X can often be big, especially in higher dimensions. Hence, computation of the inverse prior covariance matrix can be computationally expensive. Therefore we aim to construct directly the inverse covariance matrix C^{-1} or its Cholesky factor L with Gaussian Markov random field approximations. When L is a sparse matrix, this can be efficiently evaluated without explicitly computing the (full) matrix inverse L^{-1}. Although the matrix L can be computed via factoring C^{-1}, for maximal numerical accuracy it is beneficial to compute it directly without computing C^{-1}. This is because the number of bits required for a given floating point precision for constructing C^{-1} is twice the required bits for L.

Our main motivation, in this thesis, for studying Gaussian Markov random fields is in applying them as prior distributions in Bayesian statistical inverse problems. In Papers I, II and IV, we have considered Gaussian Markov random fields within the framework of Bayesian statistical inverse problems and applied the methodology to an electrical impedance tomography problem in Paper III. Studies of very high-dimensional prior distributions arising from spatially sampled values of random fields in Bayesian inversion are reported by Lasanen 2012 [13] and Stuart 2010 [31]. Särkkä et al. 2013 [25] and Solin et al. 2013 [27] applied Matérn and other types of spatio-temporal Gaussian random fields to functional MRI brain imaging and prediction of local precipitation, and in Hiltunen et al. 2011 [10] it was applied to diffuse optical tomography. Other applications of Matérn fields include for example spatial interpolation carried out by Lindgren et al. [19]. They considered Matérn fields and the weak convergence to the full stochastic partial differential equation solutions.

2.2 Preliminaries from stochastics

In the theory of Bayesian statistical inverse problems, the measurements m, noise e and unknown $\mathcal{X}$ are modelled as statistical objects. Hence, it is necessary to define some basic concepts of stochastics. Let $(\Omega, \mathscr{B}, P)$ be complete probability space. $\mathscr{B}(\mathbb{R}^d)$ is the Borel σ-algebra of $\mathbb{R}^d$. We denote a set of random fields

$$\left\{ \mathcal{X}(x) : x \in \mathbb{R}^d \right\}.$$

The expectation of the random field is defined as

$$\mathbf{E}\left(\mathcal{X}(x)\right) = \int_\Omega \mathcal{X}(x) dP.$$

The covariance of the set of random field is

$$\mathcal{C}(x, x') := \mathbf{E}\left(\mathcal{X}(x) - \mathbf{E}\left(\mathcal{X}(x)\right)\right)\left(\mathcal{X}(x') - \mathbf{E}\left(\mathcal{X}(x')\right)\right)$$

Following ([7] p. 200, Definition 5), we define stationarity of the random fields.

Definition 1 (Wide-sense stationarity). *A continuous real-valued Gaussian field $\mathcal{X}(x)$ is called wide-sense stationary, if its expectation and autocorrelation function $\mathrm{ACF}_{\mathcal{X}}(s)$ can be given as*

$$\begin{aligned} \mathbf{E}(\mathcal{X}(x)) &= \mu = \text{Constant}, \\ \mathbf{E}(\mathcal{X}(x) - \mu)(\mathcal{X}(x') - \mu) &= \mathrm{ACF}_{\mathcal{X}}(s), \end{aligned}$$

where $s := x - x'$ and $\mu \in \mathbb{R}$ is some constant. We also assume that $\mathrm{ACF}_{\mathcal{X}}(s)$ is an absolutely integrable function.

In the following, because of notational simplicity, we choose $\mu = 0$.

Wide-sense stationary processes and fields can be analysed by studying spectral presentations of the random functions. In order to define spectral density, we first need to define the concept of Fourier transforms.

Definition 2 (Fourier transform pair). *Let f be some absolutely integrable continuous function. Then the Fourier transform of object f is given by an integral transform*

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x) \exp(i\xi \cdot x) dx$$

and the inverse Fourier transform is defined by

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) \exp(-i\xi \cdot x) d\xi$$

where $x, \xi \in \mathbb{R}^d$.

Let us assume that the autocorrelation function $\mathrm{ACF}_{\mathcal{X}}(x)$ is a continuous function in $\mathbb{R}^d$. According to Bochner's theorem [7], then there exists a probability measure P on $\mathbb{R}^d$ which satisfies

$$\mathrm{ACF}_{\mathcal{X}}(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-i\xi \cdot x) dP(\xi).$$

Let us define power spectrum $S(\xi) := P'(\xi)$, where by prime we denote differentiation. Then we can obtain a simplified form of Bochner's theorem, which is called Wiener-Khinchin theorem. Following Gikhman and Skorokhod [7], we summarise the theorem as a remark.

Remark 1 (Wiener-Khinchin Theorem)**.** *Power spectrum and autocorrelatiom function of a wide-sense stationary process $\mathcal{X}$ are a Fourier transform pair*

$$\begin{aligned} S(\xi) &= \int_{\mathbb{R}^d} \mathrm{ACF}_{\mathcal{X}}(x) \exp(i\xi \cdot x) dx \\ \mathrm{ACF}_{\mathcal{X}}(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S(\xi) \exp(-i\xi \cdot x) d\xi. \end{aligned}$$

Discrete objects are defined similarly with discrete Fourier transforms. This will be dealt within subsequent sections.

2.3 Gaussian Markov random fields

We are motivated by using Gaussian Markov random fields as correlation priors for Bayesian statistical inverse problems. The goal is to construct sparse matrix approximations of these priors. The approximations, as discussed later, are related to polynomials in the continuous and to trigonometric polynomials in the discrete approximations.

Definition 3 (Gaussian Markov random field)**.** *Let*

$$P(t) := \sum_{k=0}^{K} c_k t^{2k} > 0, \quad \text{for all } t \in \mathbb{R}$$

where $c_0 > 0$ and $c_k \in \mathbb{R}$ and $K \in \mathbb{Z}^+$. We call a random function $\mathcal{X}(x)$ a Gaussian Markov random field, if it is wide-sense stationary with autocorrelation function

$$\begin{aligned} \mathrm{ACF}_{\mathcal{X}}(x) &= \frac{\sigma^2}{(2\pi)^d} \int \frac{1}{P(|\xi|)} \exp\left(-ix \cdot \xi\right) d\xi \\ &= \frac{\sigma^2}{(2\pi)^d} \int \frac{1}{\sum_{k=0}^{K} c_k |\xi|^{2k}} \exp\left(-ix \cdot \xi\right) d\xi \end{aligned} \tag{2.5}$$

and $K < \infty$.

The case $K = \infty$ was considered in Paper IV, we will consider that in Section 2.4.

Let us turn our attention to the so-called Whittle-Matérn priors, known simply as Matérn priors, named after the seminal work by Whittle [35] and Matérn [20]. These priors were considered in Papers II, III and IV.

Example 1 (Matérn fields)**.** *A Gaussian Markov random field is called a Matérn field, if its autocorrelation function is*

$$\mathrm{ACF}_{\mathcal{X}}(x) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|x|}{\ell}\right)^{\nu} K_\nu\left(\frac{|x|}{\ell}\right), \quad x \in \mathbb{R}^d \tag{2.6}$$

where $\nu > 0$ *is the smoothness parameter,* ℓ *is the correlation length,* Γ *is the gamma-function,* K_ν *the modified Bessel function of the second kind of order* ν*, and* $|x|$ *is the Euclidean distance. The power spectrum of the Matérn field is*

$$S(\xi) = \frac{2^d \pi^{d/2} \Gamma(\nu + d/2)}{\Gamma(\nu)\ell^{2\nu}} \left(\frac{1}{\ell^2} + |\xi|^2\right)^{-(\nu+d/2)}.$$

Let us consider the Matérn fields as solutions of stochastic partial differential equations. The following results are presented in more detail in Paper III.

Lemma 1. *A Matérn prior can be given as a solution of a stochastic pseudodifferential equation.*

Proof. Let us denote by $\mathcal{W}$ as continuous white noise. If

$$\widehat{\mathcal{X}} = \sigma\sqrt{S(\xi)}\,\widehat{\mathcal{W}} \tag{2.7}$$

in the sense of distributions, then $\mathcal{X}$ is a Gaussian random field with an autocorrelation function (2.6). Then, by first dividing (2.7) by $\sqrt{S(\xi)}$ and carrying out an inverse Fourier transform will give us

$$\left(1 - \ell^2\Delta\right)^{(\nu+d/2)/2} \mathcal{X} = \sqrt{\alpha\ell^d}\,\mathcal{W}, \tag{2.8}$$

where the constant α is

$$\alpha := \sigma^2 \frac{2^d \pi^{d/2} \Gamma(\nu + d/2)}{\Gamma(\nu)}.$$

The operator $\left(1 - \ell^2\Delta\right)^{(\nu+d/2)/2}$ is a pseudodifferential operator defined by its Fourier transform. ■

Discretisation schemes of pseudodifferential equations often lead to full matrix approximations. Therefore we prefer to work with elliptic operators.

Corollary 1. *A special case of the Matérn prior can be given as a stochastic partial differential equation*

$$\left(I - \ell^2 \Delta\right) \mathcal{X} = \sqrt{\alpha \ell^d}\, \mathcal{W}. \tag{2.9}$$

Proof. We will fix the smoothness parameter by setting $\nu = 2 - d/2$ and choose $d = 1, 2, 3$. Then the operators are elliptic operators instead of pseudodifferential operators. ■

2.4 Systems of stochastic difference equations

Instead of the Matérn field, let us consider a rather more general construction. These were studied in detail in Paper I. Instead of a single matrix equation, we consider system of infinite-dimensional matrix difference equations. Given the Gaussian Markov random field as in Equation (2.5), we give it equivalently as a system of equations

$$\begin{pmatrix} L_0 \\ L_1 \\ \vdots \\ L_K \end{pmatrix} X = \begin{pmatrix} W^{(0)} \\ W^{(1)} \\ \vdots \\ W^{(K)} \end{pmatrix},$$

where L_0 is an infinite-dimensional diagonal matrix and L_k-matrices are infinite-dimensional $k^{(\mathrm{th})}$-order difference matrices. For example

$$L_1 X = \begin{pmatrix} \ddots & \ddots & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & \ddots & \ddots \end{pmatrix} \begin{pmatrix} \vdots \\ X_{j-1} \\ X_j \\ X_{j+1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ W^{(1)}_{j-1} \\ W^{(1)}_{j} \\ W^{(1)}_{j+1} \\ \vdots \end{pmatrix}.$$

Higher order difference matrices are constructed similarly. The infinite-dimensional covariance matrix of X, in the sense of least-squares, is given as

$$C = \left(\sum_{k=0}^{K} L_k^T \Sigma_k^{-1} L_k \right)^{-1} = \left(\sum_{k=0}^{K} \sigma_k^2 L_k^T L_k \right)^{-1}, \tag{2.10}$$

where $W^{(k)} \sim \mathcal{N}(0, \sigma_k^2 I)$ and σ_k^2 are some scaling factors.

In Paper I, we defined these priors on the whole lattice $h\mathbb{Z}$. This definition was given as:

Definition 4 (One-dimensional discrete correlation priors)**.** *Discrete correlation priors with a power α and a correlation length ℓ are certain zero-mean Gaussian processes on $j \in \mathbb{Z}$. Then the correlation priors are defined by combining white-noise measurements*

$$X_j \sim \mathcal{N}(0, c_0^{-1} \alpha \ell / h)$$

with some of the following virtual convolution measurements (which are statistically independent)

$$\Delta X_j \sim \mathcal{N}(0, c_1^{-1}\alpha h/\ell)$$
$$\vdots$$
$$\Delta^K X_j \sim \mathcal{N}(0, c_K^{-1}\alpha (h/\ell)^{2K-1})$$

where Δ^k *is a* $k^{(\mathrm{th})}$*-order difference operator and* $\Delta X_j = X_j - X_{j-1}$*. The correlation prior with the highest difference of* $K^{(\mathrm{th})}$ *order is called a* $K^{(\mathrm{th})}$*-order correlation prior.*

It is apparent, that priors defined through these equations lead to similar formulas as in Equation (2.10). We note that in Definition 4, we have taken the discretisation parameter h into account, because we require the discrete correlation priors to converge to continuous ones at the continuous limit. We note that for a fixed correlation function form, the choices of c_k, ℓ and α are not unique. Therefore, we find it practical, to first fix the c_k terms. Then the correlation length ℓ scales the discretisation parameter h and the power α scales the variance.

Combining white-noise measurements with virtual difference measurements might seem obscure, if one thinks separately of the effect of white noise and difference measurements to the posterior distribution. The catch is in how the combination is carrued iyt and what is the corresponding prior covariance. The covariance might be counterintuitive to the prior beliefs related to virtual measurements of different orders.

Autocorrelation function

Now our objective is to calculate the continuous limits of the discrete processes via their autocorrelation functions. First, we let $X := \{X_j\}_{j=-\infty}^{\infty}$ have values in the space of doubly-infinite sequences of real numbers $\mathbb{R}^{\mathbb{Z}}$. We defined stationarity for continuous random processes in Definition 1. For discrete processes definition is analogous. Discrete stationary random processes have a constant mean and a covariance satisfying

$$\mathbf{E}\left(X_j X_{j'}\right) = \mathbf{E}\left(X_{j-j'} X_0\right) =: \mathrm{ACF}_X(j-j'),$$

where $j, j' \in \mathbb{Z}$.

Let $w_k = \{w_{k,j}\}_{j=-\infty}^{\infty} \in \mathbb{R}^{\mathbb{Z}}$ have only finitely many non-zero elements for $k = 1, ..., K$. Consider virtual independent measurements

$$(w_k * X)_j = \sum_{m=-\infty}^{\infty} w_{k,(j-m)} X_m \sim N(0, \sigma_k^2),\ k = 0, .., K,\ j \in \mathbb{Z}$$

on X.

Definition 5 (Discrete Fourier transform). *The discrete Fourier transform of an absolutely summable sequence $w \in \mathbb{R}^{\mathbb{Z}}$ is*

$$\widehat{w}(\xi) := \sum_{j=-\infty}^{\infty} \exp(i\xi j) w_j, \quad \xi \in [-\pi, \pi)$$

and the inverse discrete Fourier transform is

$$w_j := \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-i\xi j) \widehat{w}(\xi) d\xi.$$

Recall, that X is a random function on $\mathbb{Z}$ if $(X_{j_{n_1}}, ..., X_{j_{n_d}})$ is a random vector for any finite collection of indices $j_{n_k} \in \mathbb{Z}$. The random function is called Gaussian if the above random vectors have a multivariate Gaussian distribution.

Let X be a Gaussian random function on $\mathbb{Z}$ having zero-mean and a stationary covariance $\mathrm{ACF}_X(j)$ with a property $\sum_{k=-\infty}^{\infty} \mathrm{ACF}_X(j) < \infty$. The spectrum S_X of the stationary process X is the Fourier transform of the autocorrelation function

$$\mathrm{ACF}_X(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-i\xi j) S_X(\xi) d\xi.$$

Definition 6. *If $P_X(\xi) S_X(\xi) = 1$ almost everywhere, then $P_X(\xi)$ is called Fourier domain Fisher information of X.*

In the following lemma we use the term *additional measurements* for any measurement combined with measurement X.

Lemma 2. *The Fourier domain Fisher information $P_X(\xi) > 0$ of X is additive with respect to virtual convolution measurements in the sense that $P_X(\xi)$ increases by $\sigma^{-2}|\widehat{w}|^2$ when an additional convolution measurement $w * X \sim \mathcal{N}(0, \sigma^2)$ is given.*

For the proof of Lemma 2, see Paper I.

Theorem 1. *If $|\widehat{w}_k|^2 > 0$ for some $k \in \{0, ..., K\}$ the prior covariance operator formed from virtual measurements (2.4) of X is*

$$\mathrm{ACF}_X(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\exp(-i\xi j)}{\sum_{k=0}^{K} \sigma_k^{-2} |\widehat{w}_k(\xi)|^2} d\xi. \tag{2.11}$$

Proof. In the above lemma, choose as the original Fourier-domain Fisher information $|\widehat{w}_k|^{-2}$. Use additivity of the Fourier domain Fisher informations for other convolution measurements to obtain the result. ∎

Continuous limit of the autocorrelation function

The prior covariances corresponding to virtual measurements are calculated by formula (2.11).

Theorem 2. *The discrete autocovariance corresponding to a system of stochastic difference equations*

$$X_j \sim \mathcal{N}(0, c_0^{-1}\alpha\ell/h)$$
$$X_j - X_{j-1} \sim \mathcal{N}(0, c_1^{-1}\alpha h/\ell)$$

is given as

$$\mathrm{ACF}_X(j) = \frac{\alpha}{2\pi}\int_{-\pi}^{\pi}\frac{\exp(-i\xi j)}{h/\ell + \ell/h\,(2-2\cos(\xi))}d\xi.$$

Proof. We first define $w_0 = \delta_j$ and $w_1 = \delta_j - \delta_{j-1}$. Then by simple algebra

$$\widehat{w}_0 = 1, \quad \text{and} \quad \sigma_0^{-2}|\widehat{w}_0|^2 = h/(\alpha\ell)$$
$$\widehat{w}_1 = 1 - \exp(-i\xi), \quad \text{and} \quad \sigma_1^{-2}|\widehat{w}_1|^2 = \ell/(\alpha h)\,(2-2\cos(\xi)).$$

Claim follows from using the additivity of the Fourier domain Fisher information (Lemma 2) and by using Equation (2.11). ■

Now we want to have convergence of the discrete autocorrelation function at the discretisation limit $h \to 0$.

Theorem 3. *The discrete autocovariance given in Theorem 2, converges to a continuous stationary process with autocorrelation function*

$$\mathrm{ACF}_{\mathcal{X}}(x) = \frac{\alpha}{2}\exp\left(-\frac{|x|}{\ell}\right).$$

Proof. We defined the autocorrelation function as

$$\begin{aligned}\mathrm{ACF}_X(x) &= \frac{\alpha}{2\pi}\int_{-\pi}^{\pi}\frac{1}{h/l + l/h\,(2-2\cos(\xi)}\exp(-i\xi x)d\xi \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\alpha}{h/l + l/h\,(1+\xi^2 B(\xi))}\exp(-i\xi x)d\xi,\end{aligned}$$

where $B(\xi) = ((2-2\cos(\xi))/\xi^2 - 1)/\xi^2$. This function has asymptotic behaviour $\mathcal{O}(1)$ as ξ approaches zero. Taking $\xi' = (l/h)\xi$ as the integration variable we obtain

$$\mathrm{ACF}_X(x) = \frac{\alpha}{2\pi}\int_{-\pi l/h}^{\pi l/h}\frac{1}{1+\xi'^2(1+(\xi' h/l)^2 B(\xi' h/l))}\exp(-i\xi' x)d\xi'.$$

For any h, the integrand is dominated by $(1+\xi'^2 4/\pi^2)^{-1}$. By denoting $\simeq$ the nearest integer, we can use Lebesgue's dominated convergence theorem to obtain the continuous limit

$$\mathrm{ACF}_{\mathcal{X}}(x) = \lim_{h\to 0}\mathrm{ACF}_X(jh)|_{j\simeq x/h} = \frac{\alpha}{2\pi}\int_{-\infty}^{\infty}\frac{\exp(-i\xi' x/l)}{1+\xi'^2}d\xi'.$$

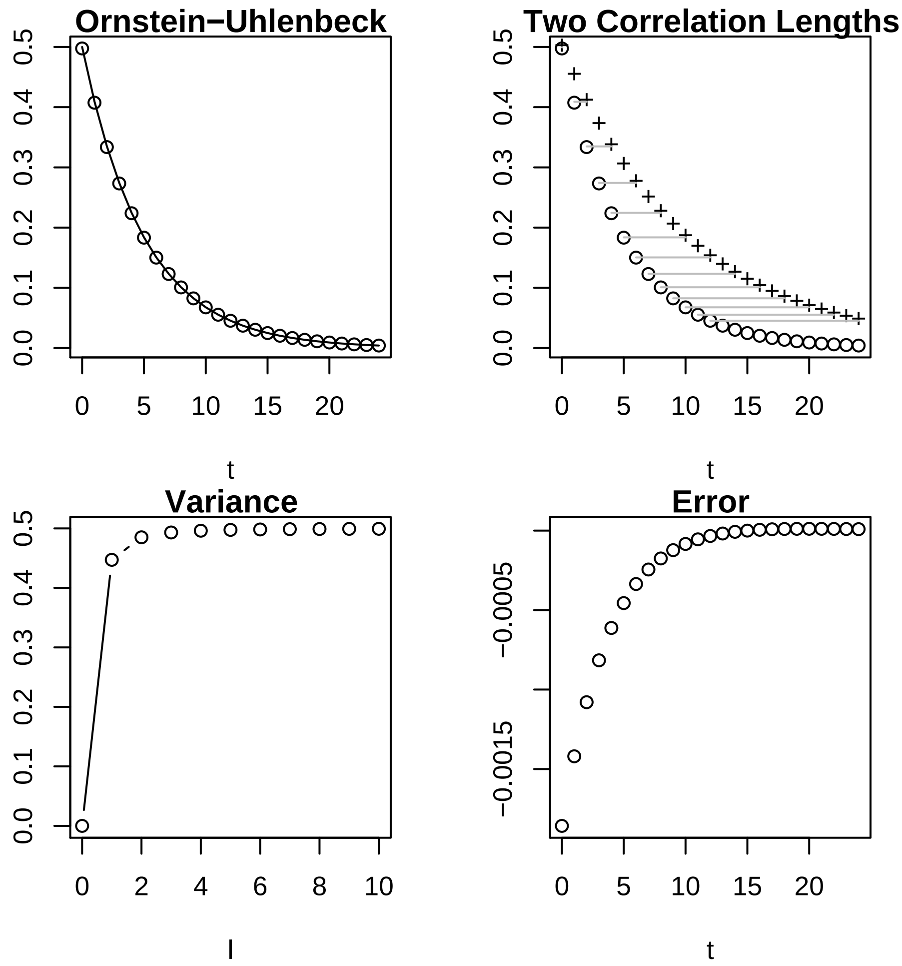


Figure 2.1: The correlation profile of Ornstein-Uhlenbeck prior obtained by correlation priors. On the top left panel, the continuous limit $\mathrm{ACF}_{\mathcal{X}}(x)$ is denoted by the solid line and the discrete correlation prior $\mathrm{ACF}_X(jh)|_{j\simeq x/h}$ pointwise by the circles. The top right panel shows the correlation priors $\mathrm{ACF}_X(jh)$ with correlation lengths 5 plotted by circles and 10 by crosses. If scaling works as we want, every second point of the longer correlation length plot should correspond to every point of the shorter correlation length plot. The bottom left panel shows the behaviour of the variance $\mathrm{ACF}_X(0,\ell)$ as a function of correlation length ℓ when the discretisation is $h = 1$. The bottom right panel shows the absolute difference of the plot with correlation length 5 and every second point of the plot with correlation length 10 $\mathrm{ACF}_X(j,\ell = 5) - \mathrm{ACF}_X(2j-1,\ell = 10)$.

Hence, the claim follows by evaluating the integral with for example calculus of residues. ■

We note that this is the covariance of the Ornstein-Uhlenbeck process. We have studied the limit behaviour numerically in Figure 2.1 (originally in Paper I). Similar techniques apply to more complicated correlation priors, as demonstrated in Papers I and IV.

Band-limited fields

In Papers I, II and III, we were mostly interested in Gaussian Markov random fields characterised by polynomials $P(t) = \sum_{k=0}^{K} c_k |t|^{2k}$, where $c_0 > 0$ and $c_k \geq 0$ when $k = 1, \ldots, K$. In paper IV, we considered also the case when $c_k \in \mathbb{R}$. The reason for this is that we wish to consider

$$P(t) := \left(\kappa^2 + t^2\right)^{\alpha},$$

where $t \in \mathbb{R}$ and $\alpha > 0$ is fractional. The function P has the well-known Taylor series

$$\left(\kappa^2 + t^2\right)^{\alpha} = \sum_{k=0}^{\infty} a_k \kappa^{2(\alpha-k)} t^{2k}, \tag{2.12}$$

where

$$\begin{aligned} a_0 &= 1, \\ a_k &= \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!} \quad \text{for} \quad k \geq 1. \end{aligned}$$

We note that the series (2.12) converges for $|t| \leq \kappa$ and diverges for $|t| > \kappa$. This leads us to the study of band-limited Matérn fields

$$\mathcal{C}(s) = \frac{\sigma^2}{(2\pi)^d} \int_{|\xi| \leq \kappa} \frac{\exp\left(-i\xi \cdot s\right)}{\sum_{k=0}^{K} c_k |\xi|^{2k}} d\xi.$$

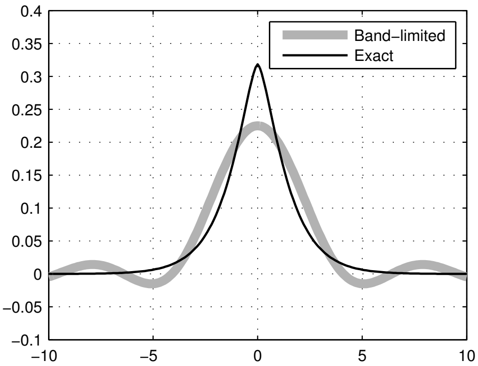
The discrete presentation is

$$C(jh) = \frac{\sigma^2}{(2\pi)^d} \int_{(-\pi,\pi)^d} \frac{\exp\left(-i\xi \cdot jh\right)}{\sum_{k=0}^{K} c_k h^{d-2k} \left(\sum_{p=1}^{d} (2 - 2\cos(\xi_p))^k\right)} d\xi.$$

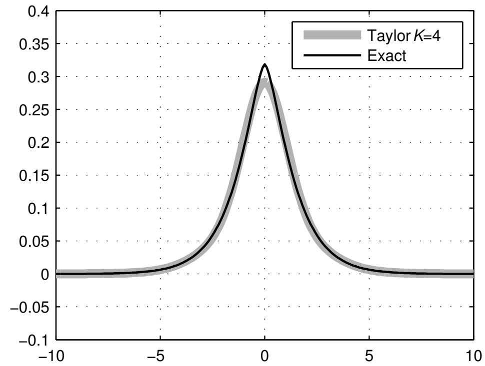
Following the Definition 4 of correlation priors, we can write the $k^{(\mathrm{th})}$-order difference matrix L_k in a similar way to Equation (2.10). The corresponding covariance matrices Σ_k are obtained from the constants $c_k h^{d-2k}$ and they are $\Sigma_k = \frac{h^{2k-1}}{c_k} I$. Using the additivity of the precision matrix (Lemma 2), we may then write the discrete covariance with a matrix equation as

$$C = \left(L^T L\right)^{-1} = \left(\sum_{k=0}^{K} L_k^T \Sigma_k^{-1} L_k\right)^{-1} = \left(\sum_{k=0}^{K} a_k \kappa^{2(\alpha-k)} h^{1-2k} L_k^T L_k\right)^{-1}. \tag{2.13}$$

Given the matrices L_k and Σ_K in Equation (2.13), our aim is to construct an upper triangular sparse matrix L. The full covariance matrix C in Equation (2.13), has both $c_k > 0$ and $c_k < 0$ terms, which we relate to constructing L with Cholesky decomposition. We choose this construction, as we aim to construct the upper triangular matrix L term-by-term, that is, we recursively apply the Cholesky decomposition in order to get the required presentation. Cholesky decomposition algorithms are covered in standard literature [8].



(a) Plain band-limited



(b) Taylor expansion

Figure 2.2: (a) Covariance function of a band-limited approximation to one-dimensional Matérn spectral density. (b) Covariance function of a $4^{(\mathrm{th})}$ order Taylor series expansion. Although the plain band-limited approximation is quite inaccurate, the truncated Taylor series on the whole $\mathbb{R}$ is quite accurate.

As an example, we choose $d = 1$, $\sigma^2 = 1$, $\alpha = 3/2$ and $\kappa = 1$ with truncation parameter $K = 4$, and set

$$P(t) = 1 + \frac{3}{2}t^2 + \frac{3}{8}t^4 - \frac{1}{16}t^6 + \frac{3}{128}t^8. \tag{2.14}$$

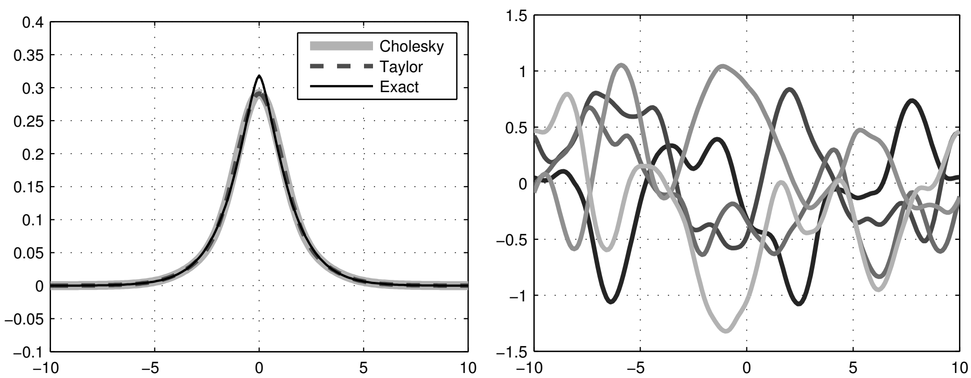
The polynomial is clearly everywhere positive and hence the spectral density is valid in the whole $\mathbb{R}$. Thus we can extend the integration area to the whole space. Figure 2.2 (from Paper IV) illustrates the resulting approximation.

Let us consider partitioning the precision matrix C^{-1} as

$$P_+ - P_- := \sum_{a_k \geq 0} a_k \kappa^{2(\alpha-k)} h^{1-2k} L_k^T L_k - \sum_{a_k < 0} |a_k \kappa^{2(\alpha-k)}| h^{1-2k} L_k^T L_k,$$

where the partitioned precision matrices P_+ and P_- correspond to the parts to be sequentially updated with positive and negative signs, respectively. When making the Cholesky decomposition, we first loop over the positive c_k coefficients and carry out Cholesky updates with $\sqrt{a_k \kappa^{2(\alpha-k)} h^{1-2k}}\, L_k$. Then we do the same for the negative coefficients with the so-called Cholesky downdates. We note that it is advisable not to mix the updates with positive and negative signs, because this might break the positive-definiteness property of the covariance matrix. This might break the algorithm and hence, we propose to carry out updates with positive signs first and downdates with negative signs in the final part of the algorithm.

Figure 2.3 (from Paper IV) shows an example of a covariance function approximation formed with the above procedure as well as example realisations of the process.



(a) Approximation (b) Realisations

Figure 2.3: (a) Exact covariance function of the one-dimensional example of Equation (2.14) ($\sigma^2 = 1$, $\alpha = 3/2$, $\kappa = 1$, $K = 4$), the truncated Taylor series approximation and its finite-difference approximation with discretisation step $h = 0.1$. In the finite-difference computations, we have used periodic boundary conditions in an extended domain and cropped the image. (b) Realisations from the process simulated via the discretised approximation.

2.5 Correlation priors on torus

Previously, we have discussed correlation priors on the whole lattice $\mathbb{Z}$. For practical inverse problems, we would prefer to define correlation priors on some finite interval. However, then we would have two boundary points and these would cause boundary effects. We overcome this problem by periodic boundary conditions and hence mitigate the boundary effects. Thus, instead of the whole space $\mathbb{Z}$, we consider the problem on a circle $\mathbb{Z}/2N$.

Following Paper II, we start by defining a discretisation sequence

$$\mathscr{L}(N,h) := \{jh \mid j \in \mathbb{Z} \cap [-N,N)\},$$

where the discretisation step $h > 0$ and the number of elements $\#\mathscr{L}(N,h) = 2N$. The idea is to define correlation priors on circle and let $N \to \infty$ and $h \to 0$. In higher dimensions, we study problem on torus $(\mathbb{Z}/2N)^d$.

When we change from whole space to circle, or torus, we end up studying circulant matrices. The importance of circulant matrices is that they are diagonalised by the

Discrete Fourier Transform and hence, all their properties are defined by the first row. The other option would be to use symmetric Toeplitz matrices. However, because of the convolution theorem, circulant matrices are closed under matrix multiplication. The Toeplitz matrices are not.

Let us consider discretisation of the Matérn field in Equation (2.9). We choose discretisation sequence $\mathscr{L}(N,h)$. For simplicity, let us start with $d=1$. Then the continuous equation is

$$\mathcal{L}\mathcal{X} := \left(I - \ell^2 \frac{\partial^2}{\partial x^2}\right)\mathcal{X} = \mathcal{W}. \tag{2.15}$$

We have two objects to discretise, the linear operator in the left-hand-side and the white noise in the right-hand-side. Here, we start from the results and then go through discretisation in steps. As we assumed periodic boundary conditions, we set $X_N = X_{-N}$. With a certain finite-difference scheme, we can give an approximation of (2.15) with $d=1$ as in (1.2) as

$$\begin{pmatrix} 1+2\frac{\ell^2}{h^2} & -\frac{\ell^2}{h^2} & & & -\frac{\ell^2}{h^2} \\ -\frac{\ell^2}{h^2} & 1+2\frac{\ell^2}{h^2} & -\frac{\ell^2}{h^2} & & \\ & \ddots & & \ddots & \\ & & -\frac{\ell^2}{h^2} & 1+2\frac{\ell^2}{h^2} & -\frac{\ell^2}{h^2} \\ -\frac{\ell^2}{h^2} & & & -\frac{\ell^2}{h^2} & 1+2\frac{\ell^2}{h^2} \end{pmatrix} \begin{pmatrix} X_{-N} \\ X_{-N+1} \\ \vdots \\ X_{N-1} \end{pmatrix} = \begin{pmatrix} W_{-N} \\ W_{-N+1} \\ \vdots \\ W_{N-1} \end{pmatrix},$$

where $W_j \sim \mathcal{N}(0, \alpha\ell/h)$. This equation is of the form we are searching for, i.e. $LX = W$ with a sparse matrix L. This makes the construction suitable for efficient computer solvers [22], i.e. the matrix-vector-product $L^{-1}W$ is fast to compute.

Following Paper III, we can make similar constructions in higher dimensions. Let us take $d=2$ and denote

$$A := \begin{pmatrix} 2\frac{\ell^2}{h^2} & -\frac{\ell^2}{h^2} & & & -\frac{\ell^2}{h^2} \\ -\frac{\ell^2}{h^2} & 2\frac{\ell^2}{h^2} & -\frac{\ell^2}{h^2} & & \\ & \ddots & & \ddots & \\ & & -\frac{\ell^2}{h^2} & 2\frac{\ell^2}{h^2} & -\frac{\ell^2}{h^2} \\ -\frac{\ell^2}{h^2} & & & -\frac{\ell^2}{h^2} & 2\frac{\ell^2}{h^2} \end{pmatrix}.$$

Let us denote by $\otimes$ the Kronecker product and let J be an identity matrix. Then we can write

$$LX = (I + A \otimes J + J \otimes A)X = W$$

where $W_{j_1,j_2} \sim \mathcal{N}(0, \alpha\ell^2/h^2)$ for all $j_1, j_2 \in \mathbb{Z}$. Hence, we obtained the sparse matrix approximation of the two-dimensional Matérn field we are looking for.

Now, let us study the discretisation of an operator equation

$$\mathcal{I}\mathcal{X} = \mathcal{X} = \mathcal{W},$$

where $\mathcal{I}$ is an identity operator and $\mathcal{W}$ is continuous white noise. The discretisation of this formulation was considered in Paper II. The discretised white noise W is a Gaussian multivariate random variable with identity matrix I as its covariance. Hence, we can consider it also as wide-sense stationary process. This implies that X is also a finite stationary random process.

We want to obtain a convergence to a continuous object when making discretisation denser and denser, i.e. we want to have $X \to \mathcal{X}$ in the discretisation limit. Hence, the limit is basically a continuum. This was considered in detail in Paper II. Here we consider the most important building blocks. First of all, we need to take the topology of the continuum into account and embed the discretisation in a corresponding continuum.

First, we define the intervals as

$$\ll a..b \gg := [a, b) \cap \mathscr{L}(N, h)$$

and a length measure

$$\mathrm{Len}(\ll a..b \gg) := h\# \ll a..b \gg .$$

The notation $[\cdot]$ is the Iverson bracket

$$[A] := \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise.} \end{cases}$$

The reason for the notation is that we can use it as an indicator function. In Lemma 4.4 in Paper II, we had white noise on lattice $\mathscr{L}(N, h)$ embedded in $\mathbb{R}$.

Lemma 3. *If W is white noise with the parameter set $\mathscr{L}(N, h)$, then W' defined by*

$$W'(\ll a..b \gg) := \sqrt{h} \sum_{j \in \mathscr{L}(N,h)} W(jh)[jh \in \ll a..b \gg]$$

is white noise on $\mathscr{L}(N, h)$ embedded in $\mathbb{R}$.

In order to obtain convergence, as we work on lattice $\mathscr{L}(N, h)$, we will link the lattice parameters $h = \sqrt{\pi/N}$ or as $N = \lfloor \pi/h^2 \rfloor$. We know that as $h \to 0$ and $N \to \infty$, we will cover the whole real line $\mathbb{R}$. Now we can argue that there has to be some kind of convergence as $N \to \infty$. Construction is based on strong-weak convergence of probability measures. This was carried out in detail in Paper II. Here it is enough to note that the limit exists.

Let us denote by $\mathscr{L}_d(N, h)$ a d-dimensional discretisation sequence. The d-dimensional generalisation of Lemma 3, was given in Lemma 4.5 (Paper II). It says:

Lemma 4. *If W is white noise with parameter set $\mathscr{L}_d(N, h)$, then W' defined by*

$$W'(\ll \mathbf{a}..\mathbf{b} \gg) := h^{d/2} \sum_{\mathbf{j} \in \mathcal{L}_d(N,h)} W(\mathbf{j})[\mathbf{j} \in \ll \mathbf{a}..\mathbf{b} \gg]$$

is white noise on $\mathscr{L}_d(N, h)$ embedded in $\mathbb{R}^d$.

The discretisation of the linear operator in Equation (2.15) was considered in detail in Section 4.3 in Paper II, where the differential operator was parameterised there as $(-\lambda_0 I + \lambda_1 \Delta)\mathcal{X} = \mathcal{W}$, where $\lambda_0\lambda_1 > 0$. Let us give it here with the same parameterisation and discretisation on lattice $\mathscr{L}(N,h)$ as

$$-\lambda_0 h X_j - \lambda_1 h^{-1}(X_{j-1} - 2X_j + X_{j+1}) = W'(\ll jh \gg).$$

If we embed the discretisation step h in the parameters λ_0 and λ_1, and if we remember that we have one $h^{1/2}$ inside the noise term, we can write the discrete equation in the interior points as

$$\lambda_1 X_j - (\lambda_0 + 2\lambda_1)(X_{j-1} - \lambda_1 X_j + X_{j+1}) = W_j' \sim \mathcal{N}(0,1), \tag{2.16}$$

where $\lambda_1 = \mu_1 h^{-3/2}$ and $\lambda_0 = \mu_0 h^{1/2}$.

We want to have again convergence of the autocorrelation function in the discretisation limit. This was covered in Theorem 5.1 in Paper II.

Theorem 4. *Suppose $\lambda_0\lambda_1 > 0$. Then the discrete stationary processes obtained corresponding to the Equation* (2.16) *converge in the strong-weak topology to a stationary Gaussian process with the autocorrelation function*

$$\mathrm{ACF}_{\mathcal{X}}(\lambda_0,\lambda_1)(x) = \frac{1}{4\alpha\lambda_0^2}(1 + |x|/\alpha)\exp(-|x|/\alpha),$$

where $\alpha = \sqrt{\lambda_1/\lambda_0}$. If the discretisation for the Laplacian is given by three lattice points, we have

$$\mathrm{ACF}_N(\lambda_0,\lambda_1)(\widetilde{x}) = \mathrm{ACF}(\lambda_0,\lambda_1)(\widetilde{x}) + \mathscr{O}N^{-\alpha}$$

with $\alpha = 3/8$. The rate of convergence is $\alpha = 3/5$ with the five-point stencil. The optimal α is obtained with stencil length $n = 7$ and $\alpha = 3/4$.

The proof was given in Section 5 in Paper II in four lemmas. The autocorrelation function of the discrete stationary was calculated in Lemma 5.3. The asymptotic estimate with convergence rates were covered in Lemmas 5.4 and 5.5 with different stencils. The closed-form limit of the autocorrelation was given in Lemma 5.6.

If we denote $\ell := \alpha = \sqrt{\lambda_1/\lambda_0}$ and choose $\lambda_0 = \ell^{-1/2}$ we get $\lambda_1 = \ell^{3/2}$ and

$$\mathrm{ACF}_{\mathrm{mod}}(l)(x) := \mathrm{ACF}(\ell^{-1/2},\ell^{3/2})(x) = \frac{1}{4}(1 + |x|/\ell)\exp(-|x|/\ell).$$

We note that this is up to a scaling factor the same result as in [32].

The two-dimensional Matérn field was considered in Theorem 6.1 in Paper II.

Another example in Paper II was considered when we study the linear opetor $(-i\lambda_0 + \lambda_1\Delta)$ instead $(-\lambda_0 + \lambda_1\Delta)$. Then we study complex-valued stochastic field

$$(-i\lambda_0 + \lambda_1\Delta)\mathcal{X} = \mathcal{W}.$$

As shown in Paper II, this corresponds to the case when study a pair of equations

$$\begin{cases} \lambda_1 \Delta X = W^{(1)} \\ \lambda_0 X = W^{(2)}. \end{cases}$$

where $W^{(1)} \sim W^{(2)}$ are independent real Gaussian white noises. This is of course a special case of the continuous limit of discrete correlation priors of Definition 4. In Paper II, one-dimensional case was considered in Theorem 7.1 and two-dimensional case in Theorem 7.2.

2.6 Table of analytical formulas for correlation priors

Let us conclude the results of this Chapter with a number of analytical formulas for correlation priors. We give them with respect to to the Fourier domain Fisher information $P(|\xi|)$, because it immediately shows the relationship of the prior to its system of partial differential equation representation, and hence to the fast approximation with difference matrices. The counterpart is the autocorrelation function, i.e. the Fourier transform of the reciprocal of $P(|\xi|)$.

We first give a number of one-dimensional priors, studied especially in Papers I and II as

$$\begin{aligned} P(\xi) = 1 + |\xi|^2 \;&\Leftrightarrow\; \mathrm{ACF}_{\mathcal{X}}(x) = \frac{1}{2}\exp\left(-|x|\right) \\ P(\xi) = 1 + |\xi|^4 \;&\Leftrightarrow\; \mathrm{ACF}_{\mathcal{X}}(x) = \frac{1}{2}\exp\left(-\frac{|x|}{\sqrt{2}}\right)\sin\left(\frac{|x|}{\sqrt{2}} + \frac{\pi}{4}\right) \\ P(\xi) = (1 + |\xi|^2)^2 \;&\Leftrightarrow\; \mathrm{ACF}_{\mathcal{X}}(x) = \frac{1}{4}(1+|x|)\exp\left(-|x|\right). \end{aligned}$$

The two-dimensional correlation priors were covered in Paper II. Hence, we give a special case of the Matérn field as

$$P(\xi) = 1 + |\xi|^4 \;\Leftrightarrow\; \mathrm{ACF}_{\mathcal{X}}(x) = \frac{1}{4\pi}|x|K_1(|x|),$$

where K_1 is the modified Bessel function of the second kind. We give the second one as

$$P(\xi) = (1 + |\xi|^2)^2 \;\Leftrightarrow\; \mathrm{ACF}_{\mathcal{X}}(x) = -\frac{1}{2\pi}\mathrm{kei}(|x|),$$

where kei is the Thomson function $x \mapsto \mathrm{Imag}\,(K_0(x\exp(i\pi/4)))$.

We could also consider the similar Fourier domain presentation in a two-dimensional case as for the Ornstein-Uhlenbeck. Formally, in the sense of distributions, we can write

$$P(\xi) = 1 + |\xi|^2 \;\Leftrightarrow\; \mathrm{ACF}_{\mathcal{X}}(x) = 2\pi K_0(|x|).$$

The autocorrelation function has a logarithmic singularity at $x = 0$. This was studied in detail in [2].

A d-dimensional Matérn field c¡n be presented

$$P(\xi) = \frac{\Gamma(\nu)\ell^{2\nu}}{2^d \pi^{d/2} \Gamma(\nu + d/2)} \left(\frac{1}{\ell^2} + |\xi|^2 \right)^{\nu + d/2} \Leftrightarrow \mathrm{ACF}_{\mathcal{X}}(x) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|x|}{\ell} \right)^{\nu} K_\nu \left(\frac{|x|}{\ell} \right).$$

Finally, we give the squared exponential function with a Taylor expansion at $\xi = 0$:

$$P(\xi) = \exp\left(\frac{1}{2}\xi^2\right) \approx 1 + \frac{1}{2}\xi^2 + \frac{1}{8}\xi^4 + \frac{1}{48}\xi^6 + \dots \Leftrightarrow \mathrm{ACF}_{\mathcal{X}}(x) = \exp\left(-\frac{1}{2}x^2\right).$$

We can make a Gaussian Markov random field approximation by truncating the Taylor expansion. Otherwise the sparse matrix approximation would become a full matrix.

Chapter 3

Non-stationary correlation priors and unstructured meshes

In this chapter, we start from the isotropic correlation priors and go to anisotropic and inhomogeneous priors. Finally, we consider finite element formulation, as with them we can make inversion on complex domains on unstructured meshes. This chapter is almost completely based on Papers II and III. With anisotropic and inhomogeneous correlation priors, we can make advanced correlation structures of the unknown based on, for example, physical models or empirical observations. These were considered in Paper III for electrical impedance tomography. Similarly, in [21, 34] we applied the methodology for ionospheric tomography.

Many of the results in this chapter do not yet have rigorous mathematical proofs, i.e. we have not yet studied rigorously the convergence of discrete inhomogeneous correlation priors to continuous priors. The same applies for the correlation priors on unstructured meshes. Therefore, here, the approach is more computationally oriented than in the previous chapter. However, we conjecture that the inhomogeneous correlation priors are discretisation-invariant. This has been studied numerically (see Paper III), and the numerical results agree with the conjecture. Rigorous convergence studies are outside the scope of this thesis and hence we leave them for future studies.

3.1 Anisotropic priors

In Chapter 2, all the discussions concentrated on stationary random fields. We call such priors isotropic, because the lengths of the covariance ellipsoid axes are the same for all coordinate directions. As an example, consider a two-dimensional Matérn field with fixed correlation length ℓ in both coordinate directions. Then our only structural model for the unknown is to change the isotropic correlation length ℓ. In Figure 3.1 (from Paper III), we have plotted four realisations of the Matérn fields with varying correlation lengths. Hence, we can model the unknown by changing the isotropic correlation length ℓ.

The first step in constructing more flexible priors is to change from isotropic to

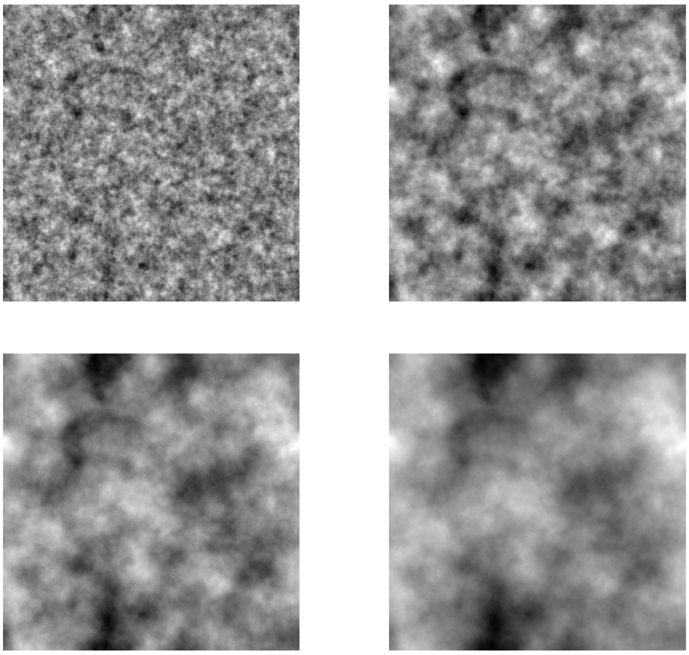


Figure 3.1: Samples of an isotropic correlation prior on 512×512 lattice with varying correlation lengths $\ell = 5, 15, 40, 100$ and discretisation step $h = 1$.

anisotropic priors. An example of such a construction is given in Lemma 6.2 in Paper II, where we considered the anisotropic two-dimensional Matérn field. The idea is that we replace the linear differential operators by weighted operators $(\ell_d \partial_n)^k$, where ℓ_d is the correlation length in the $d^{(\mathrm{th})}$ coordinate direction. The example in Paper II then had autocorrelation function

$$\mathrm{ACF}_{\mathcal{X}}(\ell_1, \ell_2)(x) = \frac{\mathrm{ACF}_{\mathcal{X}}(x_1/\ell_1, x_2/\ell_2)}{\ell_1 \ell_2}.$$

Hence, it is a scaled autocorrelation function and as such other properties remain similar to the previous chapter, i.e. we can make sparse approximation of such fields. The autocorrelation function then has covariance ellipses of the form

$$\frac{x_1^2}{\ell_1^2} + \frac{x_2^2}{\ell_2^2} = c^2.$$

Examples of two such priors are in Figure 3.2, where in the top panel, we have two realisations of an anisotropic Matérn prior.

The next step is to rotate the covariance ellipse, i.e. we tilt the covariance ellipse by a rotation on a plane with angle θ. The contours of this function are of the form

$$\frac{x_{1\theta}^2}{\ell_{1\theta}^2} + \frac{x_2^2}{\ell_2^2} = c^2$$

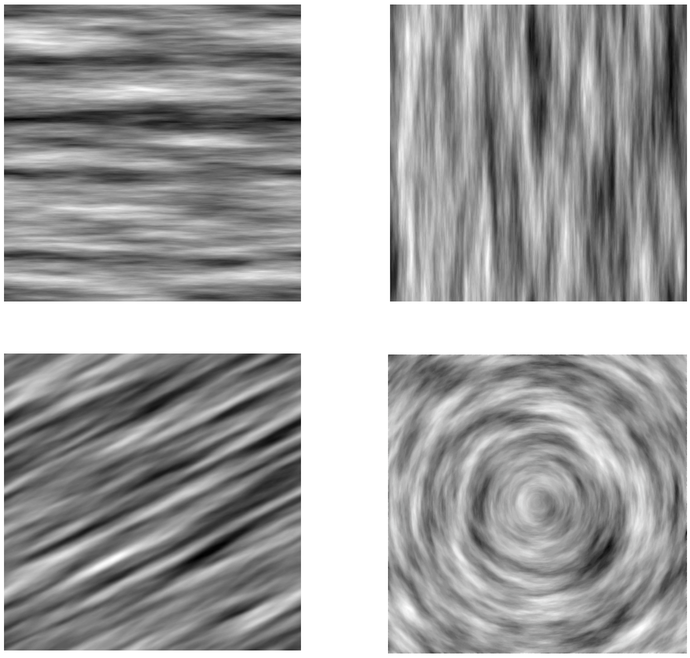


Figure 3.2: Three samples of an anisotropic correlation prior and, bottom right, one sample from an inhomogeneous correlation prior with spatially varying θ. Correlation lengths in all samples $\ell_1 = 10$, $\ell_2 = 100$ and discretisation step $h = 1$. All samples are on 512×512 lattice.

which in the original coordinate system is an ellipse with tilt θ.

As an example, we consider a two-dimensional case in which we would like to have correlation length ℓ_1 to the angle θ with respect to the x-axis and to the direction orthogonal to the previous one. The Whittle-Matérn field for such case can be modelled using rotations (see Paper II). Since

$$
\begin{aligned}
R^T R &= \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \ell_1^2 & 0 \\ 0 & \ell_2^2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \\
&= \begin{pmatrix} a_\theta^2 + b_\theta^2 & a_\theta c_\theta - b_\theta d_\theta \\ a_\theta c_\theta - b_\theta d_\theta & c_\theta^2 + d_\theta^2 \end{pmatrix},
\end{aligned}
$$

we write the weighted formulation as a stochastic partial differential equation

$$
\left(I - (a_\theta^2 + b_\theta^2) \frac{\partial^2}{\partial x_1^2} + (c_\theta^2 + d_\theta^2) \frac{\partial^2}{\partial x_2^2} + 2(a_\theta c_\theta - b_\theta d_\theta) \frac{\partial^2}{\partial x_1 \partial x_2} \right) \mathcal{X} = \sqrt{\alpha \ell_1 \ell_2}\, \mathcal{W}. \quad (3.1)
$$

The only new object to discretise is the operator $\frac{\partial^2}{\partial x_1 \partial x_2}$. Any regular finite-difference approximation can be used as an approximation for this operator. We have plotted samples of the anisotropic prior with varying θ in the bottom right panel in Figure 3.2.

To summarise, anisotropic priors can be obtained through change of variables and rotation of the covariance ellipse. Hence, results discussed in Chapter 2 are technically rather easy to generalise to anisotropic cases, as discussed in Paper II.

3.2 Inhomogeneous priors

Let us now consider inhomogeneous priors in $d = 2$. We first define parameter fields $\ell_1 = \ell_1(x)$, $\ell_2 = \ell_2(x)$ and $\theta = \theta(x)$. We then construct the prior by discretising Equation (3.1) on some lattice with variable correlation lengths and tilt. As an example, let us choose that ℓ_1 and ℓ_2 constants with $\ell_1 > \ell_2$. In order to demonstrate the flexibility of the inhomogeneous correlation priors, let us choose the tilt angle to change in such a way that we get onion-shaped features. The bottom right panel in Figure 3.2, shows the realisation of such a prior.

As another example of an inhomogeneous prior, let us consider a two-dimensional case with a squared exponential autocovariance $\exp(-|x|^2/\ell^2)$. The Fourier transform of a squared exponential is also a squared exponential, hence we have the power spectrum

$$S(\xi) = \frac{\alpha\pi}{\ell_1\ell_2} \exp\left(-\frac{\ell_1^2\xi_1^2}{4} - \frac{\ell_2^2\xi_2^2}{4}\right). \tag{3.2}$$

We approximate the squared exponential covariance via the power spectrum by a twice differentiable approximation. Instead of (3.2), we use unscaled squared exponential function (which we can obtain by a change of variables). By choosing $c_k = (1/2)^k/k!$, we construct the approximation by a truncated series, i.e. as a polynomial

$$\exp\left(-\frac{1}{2}|\xi|^2\right) = \frac{1}{\sum_{k=0}^{\infty} c_k|\xi|^{2k}} \approx \frac{1}{\sum_{k=0}^{2} c_k|\xi|^{2k}}. \tag{3.3}$$

Similarly to Definition 4, we note that the $(2k)^{(\mathrm{th})}$-order polynomials in Equation (3.3) correspond to the $k^{(\mathrm{th})}$ order linear differential operator $\mathcal{L}_k$, i.e. convolution operators. In order to obtain a sparse matrix approximation for the two-dimensional prior, this additive object corresponds to a system of stochastic partial differential equations of the form

$$\begin{cases} \sqrt{c_0}\mathcal{X}(x) = \sqrt{\alpha\ell_1\ell_2}\mathcal{W}^{(0)}, \\ \sqrt{c_1}\ell_1\frac{\partial}{\partial x_1}\mathcal{X}(x) = \sqrt{\alpha\ell_1\ell_2}\mathcal{W}^{(1,2)}, \\ \sqrt{c_1}\ell_2\frac{\partial}{\partial x_2}\mathcal{X}(x) = \sqrt{\alpha\ell_1\ell_2}\mathcal{W}^{(1,2)}, \\ \sqrt{c_2}\left(\ell_1^2\frac{\partial^2}{\partial x_1^2} + \ell_2^2\frac{\partial^2}{\partial x_2^2}\right)\mathcal{X}(x) = \sqrt{\alpha\ell_1\ell_2}\mathcal{W}^{(2)}, \end{cases} \tag{3.4}$$

where $\mathcal{W}^{(\cdot)}$ are formal continuous white noises that are independent of each other. We omit here π, which was in the power spectrum in Equation (3.2), as it is merely a constant scaling term.

Then, the discrete approximation of the continuous operator equations (3.4) can be given as a system of stochastic partial difference equations,

$$\begin{cases} X_{j_1,j_2} = \sqrt{\frac{\alpha \ell_1 \ell_2}{c_0 h_1 h_2}} W^{(0)}_{j_1,j_2}, \\ X_{j_1,j_2} - X_{j_1-1,j_2} = \sqrt{\frac{\alpha h_1 \ell_2}{c_1 \ell_1 h_2}} W^{(1,1)}_{j_1,j_2}, \\ X_{j_1,j_2} - X_{j_1,j_2-1} = \sqrt{\frac{\alpha \ell_1 h_2}{c_1 h_1 \ell_2}} W^{(1,2)}_{j_1,j_2}, \\ \frac{\ell_1^2}{h_1^2}\left(X_{j_1+1,j_2} - 2X_{j_1,j_2} + X_{j_1-1,j_2}\right) + \frac{\ell_2^2}{h_2^2}\left(X_{j_1,j_2+1} - 2X_{j_1,j_2} + X_{j_1,j_2-1}\right) \\ \quad = \sqrt{\frac{\alpha \ell_1 \ell_2}{c_2 h_1 h_2}} W^{(2)}_{j_1,j_2}. \end{cases}$$

For an inhomogeneous prior, we write

$$\begin{cases} X_{j_1,j_2} = \sqrt{\frac{\alpha_{j_1,j_2} \ell_{1,(j_1,j_2)} \ell_{2,(j_1,j_2)}}{c_0 h_1 h_2}} W^{(0)}_{j_1,j_2} \\ X_{j_1,j_2} - X_{j_1-1,j_2} = \sqrt{\frac{\alpha_{j_1,j_2} h_1 \ell_{2,(j_1,j_2)}}{c_1 \ell_{1,(j_1,j_2)} h_2}} W^{(1,1)}_{j_1,j_2} \\ X_{j_1,j_2} - X_{j_1,j_2-1} = \sqrt{\frac{\alpha_{j_1,j_2} \ell_{1,(j_1,j_2)} h_2}{c_1 h_1 \ell_{2,(j_1,j_2)}}} W^{(1,2)}_{j_1,j_2} \\ \frac{\ell_{1,(j_1,j_2)}^2}{h_1^2}\left(X_{j_1+1,j_2} - 2X_{j_1,j_2} + X_{j_1-1,j_2}\right) + \frac{\ell_{2,(j_1,j_2)}^2}{h_2^2}\left(X_{j_1,j_2+1} - 2X_{j_1,j_2} + X_{j_1,j_2-1}\right) \\ \quad = \sqrt{\frac{\alpha_{j_1,j_2} \ell_{1,(j_1,j_2)} \ell_{2,(j_1,j_2)}}{c_1 h_1 h_2}} W^{(2)}_{j_1,j_2}. \end{cases}$$

By using a stacked matrix, and then calculating the covariance, we end up with a similar formulation as in Equation (2.13), but for the inhomogeneous covariance.

3.3 Correlation priors on unstructured meshes

Previous discussion concentrated on using finite-difference methods. Such methods are computationally efficien,t especially for linear inverse problems on rather simple domains, such as rectangles or similar. However, if we wanted to study non-linear inverse problems on non-trivial domains, such as electrical impedance tomography in clinical imaging, we prefer to work on unstructured meshes and finite element methods. These were covered in detail in Paper III. Let us review the construction of the Matérn priors in such cases. Other good references on Matérn fields on unstructured meshes include for example [19].

For a continuous random field $\mathcal{X}$, we define

$$\langle \mathcal{X}, \phi \rangle := \int \mathcal{X}(x)\phi(x)\mathrm{d}x,$$

where ϕ belongs to the Schwartz space $\mathcal{S}(\mathbb{R}^d)$. Generalised random variable white noise is defined as

$$\mathbf{E}\left\langle \mathcal{W}, \phi \right\rangle = 0 \quad \text{and} \quad \mathbf{E}\left(\left\langle \mathcal{W}, \phi \right\rangle \left\langle \mathcal{W}, \varphi \right\rangle\right) = \int \phi\varphi \, \mathrm{d}x. \tag{3.5}$$

for all $\phi, \varphi \in \mathcal{S}(\mathbb{R}^d)$.

Constructing the Matérn field according to Equation (2.8) corresponds to finding an $\mathcal{X}$ such that

$$\left\langle (I - \ell^2\Delta)\mathcal{X}, \phi \right\rangle = \left\langle \sqrt{\alpha\,\ell^d}\,\mathcal{W}, \phi \right\rangle = \left\langle \mathcal{W}, \sqrt{\alpha\,\ell^d}\,\phi \right\rangle \tag{3.6}$$

for all $\phi \in \mathcal{S}(\mathbb{R}^d)$. For practical purposes, we limit ourselves to a bounded domain. Hence, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. The problem now is to find an $\mathcal{X}$ on Ω such that it satisfies (3.6). However, this problem is not uniquely solvable, since if $\mathcal{X}$ is a solution of (3.6), also $\mathcal{X} + g$ is also a solution when $g(x) = e^{a\cdot x}$ with $a \in \mathbb{R}^d$ such that $|a| = \ell^{-1}$. Therefore, we have to specify additional conditions to make the solution unique. Following Paper III, we specify some of these common boundary conditions:

$$\begin{aligned} \mathcal{X}|_{\partial\Omega} &= 0 && \text{(Dirichlet condition)} \\ \left.\frac{\partial \mathcal{X}}{\partial \mathbf{n}}\right|_{\partial\Omega} &= 0 && \text{(Neumann condition)} \\ \left.\left(\mathcal{X} + \lambda\frac{\partial \mathcal{X}}{\partial \mathbf{n}}\right)\right|_{\partial\Omega} &= 0 && \text{(Robin condition)} \end{aligned} \tag{3.7}$$

where $\mathbf{n}$ is the unit outward normal vector on the boundary and λ is a constant. We note that specifying boundary conditions leads to a change of the correlation properties. These issues were addressed in detail in Paper III.

Now we derive a weak bilinear approximation of the problem. Let us consider the Neumann boundary condition (3.7). For the derivation of the weak bilinear form, we assume for a moment that $\mathcal{X} \in H^2(\Omega)$. Then Green's first identity applied to (3.6) gives

$$\begin{aligned} \left\langle (I - \ell^2\Delta)\mathcal{X}, \phi \right\rangle &= \int_\Omega (I - \ell^2\Delta)\mathcal{X}\phi \, \mathrm{d}x \\ &= \int_\Omega \mathcal{X}\phi \, \mathrm{d}x + \ell^2 \left(\int_\Omega \nabla\mathcal{X} \cdot \nabla\phi \, \mathrm{d}x - \int_{\partial\Omega} \frac{\partial \mathcal{X}}{\partial \mathbf{n}} \phi \, \mathrm{d}\sigma \right) \\ &= \int_\Omega \mathcal{X}\phi \, \mathrm{d}x + \ell^2 \int_\Omega \nabla\mathcal{X} \cdot \nabla\phi \, \mathrm{d}x. \end{aligned} \tag{3.8}$$

The problem can be formulated as the following weak problem:

$$\text{find } \mathcal{X}\text{:} \quad a(\mathcal{X}, \phi) = \left\langle \mathcal{W}, \sqrt{\alpha\ell^d}\,\phi \right\rangle \ \text{ for all } \phi \in H^1(\Omega)$$

where a is a bilinear function defined as

$$a(\varphi, \phi) = \int_\Omega \varphi\phi \mathrm{d}x + \ell^2 \int_\Omega \nabla\varphi \cdot \nabla\phi \, \mathrm{d}x, \quad \varphi, \phi \in H^1(\Omega).$$

However, the realisations of $\mathcal{X}$ may not belong to $H^1(\Omega)$ and therefore above derivation is not valid. In order to overcome this technical problem, we first approximate $\mathcal{X}$ on a finite-element basis and apply the derivation above of the bilinear form to the approximation. A finite-dimensional approximation of the unknown object is achieved by approximating $\mathcal{X}$ by

$$\mathcal{X} \approx \sum_{j=1}^{N} X_j \psi_j$$

where X_j are random variables and ψ_j are basis functions $H^1(\Omega)$, for example piece-wise linear or polynomial functions. We substitute this approximation into (3.6) and apply the Green's theorem as in (3.8). Furthermore, we make the usual Galerkin's choice $\phi = \psi_i$, which gives the following approximation for the problem:

$$\text{find } \mathcal{X} \approx \sum_j X_j \psi_j\text{:} \quad a(\mathcal{X}, \psi_i) = \left\langle \mathcal{W}, \sqrt{\alpha \ell^d}\, \psi_i \right\rangle \text{ for all } i = 1, \ldots, N.$$

The above problem can be formulated as a matrix equation

$$LX = \left(M + \ell^2 S\right) X = W,$$

where $X = (X_j)$ and the components of the matrices M and S and the vector W are

$$M_{i,j} = \int_\Omega \psi_j \psi_i \, dx, \quad S_{i,j} = \int_\Omega \nabla\psi_j \cdot \nabla\psi_i \, dx, \quad W_i = \left\langle W, \sqrt{\alpha \ell^d}\, \psi_i \right\rangle.$$

By (3.5), $W \sim \mathcal{N}(0, \Gamma)$, where

$$\Gamma_{i,j} = \int_\Omega \alpha \ell^d \, \psi_j \psi_i \, dx.$$

Therefore $X \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = L^{-1} \Gamma L^{-1} \quad \text{and} \quad \Sigma^{-1} = L \Gamma^{-1} L = (RL)^T RL$$

and where Γ, usually a sparse matrix with a proper computational mesh, is easy to invert and the Cholesky R of Γ^{-1} is easy to compute.

Chapter 4

Discussion and conclusion

Gaussian smoothness priors are widely used in Bayesian statistical inverse problems. Similar constructions are used in deterministic Tikhonov regularisation via L^2-norms. In spite of their popularity, these methods typically do not provide proper integrable smoothness priors. In this thesis, we have considered Gaussian Markov random field priors on structured lattices with finite difference methods and on unstructured lattice with finite element methods. In addition to being proper integrable priors, they are also discretisation-invariant and computationally efficient.

We have discussed discretisation of the Gaussian Markov random fields through systems of stochastic partial difference equations. Via studying convergence of the discrete random fields to the continuous ones, we have shown the discretisation-invariance of the random fields. In conjunction with Lasanen's 2012 studies [13, 14], we know that then also the posterior distributions converge. In addition, we have discussed modelling anisotropic and inhomogeneous priors on unstructured meshes, with which we can model complex-shaped unknown.

A fundamental open problem is to study the convergence of the discrete inhomogeneous priors to continuous inhomogeneous priors. Another open problem is the boundary condition terms, which would guarantee stationarity of the random field.

An interesting extension of the topic is to study Lévy alpha-stable distributions instead of limiting oneself to Gaussian cases. Such priors provide edge-preserving inversion in the case when $\alpha = 1$, i.e. a Cauchy prior.

Using correlation priors is not very common in Bayesian inversion. Besides the Papers of this thesis, we have used GMRF priors in ionospheric tomography. Other examples of GMRF priors in Bayesian inversion includes for example. Bardsley [1]. GMRF priors have been rather popular for example in spatial statistics and machine learning. Hence, different inverse problems applications with GMRF priors may become fruitful future research areas.

Bibliography

[1] J. M. Bardsley. Gaussian markov random field priors for inverse problems. *Inverse Problems and Imaging*, 7:397–416, 2013.

[2] J. Besag and D. Mondal. First-order intrinsic autoregressions and the de wijs process. *Biometrica*, 92:909–920, 2005.

[3] D. Calvetti and E. Somersalo. *Introduction to Bayesian Scientific Computing – Ten Lectures on Subjective Computing*. Springer, 2007.

[4] M. D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 1951.

[5] J. L. Doob. *Stochastic processes*. John Wiley & Sons, New York, 1953.

[6] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet forms and symmetric Markov processes*. Walter de Gruyter & Co., Berlin, 1994.

[7] I. I. Gikhman and A. V. Skorokhod. *The Theory of Stochastic Processes I*. Springer, 1974.

[8] G. H. Golub and C. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition edition, 1996.

[9] T. Helin. On infinite-dimensional hierarchical probability models in statistical inverse problems. *Inverse Problems and Imaging*, 4:567–597, 2009.

[10] P. Hiltunen, S. Särkkä, I. Nissilä, A. Lajunen, and J. Lampinen. State space regularization in the nonstationary inverse problem for diffuse optical tomography. *Inverse Problems*, 27:2:025009, 2011.

[11] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[12] S. Lasanen. *Discretizations of Generalized Random Variables with Applications to Inverse Problems*. PhD thesis, University of Oulu, 2002.

[13] S. Lasanen. Non-gaussian statistical inverse problems. part i: Posterior distributions. *Inverse Problems and Imaging*, 6:215–266, 2012.

[14] S. Lasanen. Non-gaussian statistical inverse problems. part ii: Posterior convergence for approximated unknowns. *Inverse Problems and Imaging*, 6(267-287), 2012.

[15] S. Lasanen and L. Roininen. Statistical inversion with green's priors. In *5th Int. Conf. Inv. Prob. Eng. Proc.*, 2005.

[16] M. Lassas, E. Saksman, and S. Siltanen. Discretization invariant bayesian inversion and besov space priors. *Inverse Problems and Imaging*, 3:87–122, 2009.

[17] M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving bayesian inversion? *Inverse Problems*, 20:1537–1563, 2004.

[18] M. S. Lehtinen, L. Päivärinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5:599–612, 1989.

[19] F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B*, 73:423–498, 2011.

[20] B. Matérn. *Spatial Variation. Vol. 36 of Lecture Notes in Statistics.* Springer-Verlag, New York, 2nd edition edition, 1986.

[21] J. Norberg, L. Roininen, J. Vierinen, O. Amm, D. McKay-Bukowski, and M. Lehtinen. Ionospheric tomography in bayesian framework with gaussian markov random field priors. *Radio Science*, 2015.

[22] M. Orispää and M. Lehtinen. Fortran linear inverse problem solver. *Inverse Problems and Imaging*, 4:482–503, 2010.

[23] P. Piiroinen. *Statistical Measurements, Experiments and Applications.* PhD thesis, University of Helsinki, 2005.

[24] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications.* Chapman&Hall/CRC, 2005.

[25] S. Särkkä, A. Solin, and J. Hartikainen. Spatio-temporal learning via infinite-dimensional bayesian filtering and smoothing. *IEEE Signal Processing Magazine*, 30(4):41–61, 2013.

[26] D. Simpson. *Krylov subspace methods for approximating functions of symmetric positive definite matrices with applications to applied statistics and models of anomalous diffusion.* PhD thesis, Queensland University of Technology, Brisbane, Queensland, Australia, 2009.

[27] A. Solin and S. Särkkä. Infinite-dimensional bayesian filtering for detection of quasi-periodic phenomena in spatio-temporal data. *Physical Review E*, 88:5:052909, 2013.

[28] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients. i. *Comm. Pure Appl. Math.*, 22:345–400, 1969.

[29] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients. ii. *Comm. Pure Appl. Math.*, 22(479-530), 1969.

[30] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971.

[31] A. M. Stuart. Inverse problems: a bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[32] C. H. Su and D. Lucor. Covariance kernel representations of multidimensional second-order stochastic processes. *J. Comp. Phys.*, 217:82–99, 2006.

[33] A. Tarantola. *Inverse problem theory. Methods for data fitting and model parameter estimation.* Elsevier, 1987.

[34] J. Vierinen, J. Norberg, M. S. Lehtinen, O. Amm, L. Roininen, A. Väänänen, P. J. Erickson, and D. McKay-Bukowski. Beacon satellite receiver for ionospheric tomography. *Radio Science*, 49:1141–1152, 2014.

[35] P. Whittle. Stochastic processes in several dimensions. *Bull. Inst. Int. Statist.*, 40:974–994, 1963.

ISBN 978-952-62-0753-7 (paperback)
ISBN 978-952-62-0754-4 (pdf)
ISSN 1456-3673