NTNU

Norwegian University of
Science and Technology

Spatial Modelling and Inference with SPDE-based GMRFs

Geir-Arne Fuglstad

Master of Science in Physics and Mathematics
Submission date: June 2011
Supervisor: Håvard Rue, MATH
Co-supervisor: Finn Lindgren, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

The purpose of this thesis is to investigate modelling and inference issues for SPDE-based GMRF-models for use in spatial statistics.

Assignment given: 24. January 2011
Supervisor: Håvard Rue, MATH
Co-supervisor: Finn Lindgren, MATH

Preface

This master's thesis was written during the last semester of my master of science degree at the Norwegian University of Science and Technology (NTNU). It marks the end of the five year study programme "Applied Physics and Mathematics", with specialization in "Industrial Mathematics".

The work has been interesting and rewarding and has provided challenges both in theoretical and practical work. I wrote the code used for the simulations and the generation of the results in C++. I used the Eigen library (Guennebaud et al. 2010) for simple sparse matrix operations, the Cholmod library (Chen et al., 2008) for the calculation of the Cholesky decomposition of sparse matrices and the NLopt package (Johnson, 2011) for optimization, specifically the MMA algorithm (Svanberg, 2002).

I would like to thank my supervisors Håvard Rue and Finn Lindgren for their help and guidance.

Geir-Arne Fuglstad
Trondheim, June 2011

Abstract

In recent years, stochastic partial differential equations (SPDEs) have been shown to provide a useful way of specifying some classes of Gaussian random fields. The use of an SPDE allows for the construction of a Gaussian Markov random field (GMRF) approximation, which has very good computational properties, of the solution. In this thesis this kind of construction is considered for a specific spatial SPDE with non-constant coefficients, a form of diffusion equation driven by Gaussian white noise.

The GMRF approximation is derived from the SPDE by a finite volume method. The diffusion matrix in the SPDE provides a way of controlling the covariance structure of the resulting GMRF. By using different diffusion matrices, it is possible to construct simple homogeneous isotropic and anisotropic fields and more interesting inhomogeneous fields.

Moreover, it is possible to introduce random parameters in the coefficients of the SPDE and consider the parameters to be part of a hierarchical model. In this way one can devise a Bayesian inference scheme for the estimation of the parameters. In this thesis two different parametrizations of the diffusion matrix and corresponding parameter estimations are considered.

The results show that the use of an SPDE with non-constant coefficients provides a useful way of creating inhomogeneous spatial GMRFs.

Contents

Chapter 1

Introduction

The construction of spatial models based on Gaussian Markov random fields directly from the conditional distributions can be an effective tool, but is not always easy to achieve. When constructing spatial models, one often wants to introduce parameters and hyperparameters which control the local or global behaviour in some way. This can be a difficult task since the conditional distributions must give valid joint distributions for all legal parameter choices. A second problem is that the model should be consistent in the sense that when the distances between the positions decrease, the model should approach some continuous model.

Lindgren and Rue (2008) demonstrated that by specifying second-order random walk as the solution of a stochastic differential equation and approximating the solution with a Gaussian Markov random field, one gets a spatially consistent model for second-order random walk. This type of construction was later extended to the very useful class of Matérn fields in Lindgren et al. (2011). This showed that the combination of stochastic differential equations and Gaussian Markov random fields could be a useful tool for creating consistent models.

In this thesis this kind of construction is done for a stochastic differential equation with non-constant coefficients. The goal is to introduce parameters in the coefficients of the stochastic differential equation and to study how the Gaussian Markov random field approximation can be used to do inference on these parameters.

The thesis starts with basic theory on Gaussian Markov random fields and stochastic differential equations in Chapter 2. The chapter serves as a short introduction to each of the concepts and contains definitions and results that are needed in the later chapters.

Chapter 3 covers the actual construction of the Gaussian Markov random field approximation. This includes the discretization by a finite volume method and

the resulting precision matrix. The chapter ends with some examples of homogeneous and inhomogeneous random fields constructed with the approximation.

In Chapter 4 the Gaussian Markov random field from Chapter 3 is used to construct an inference scheme for parameters in the coefficients of the stochastic differential equation. In addition, different parametrizations of the coefficients of the stochastic differential equation are discussed and the corresponding results are shown.

The thesis ends with discussion of the results and the relevance for spatial modelling in Chapter 5 and conclusions and some remarks on possible further work in Chapter 6.

Chapter 2

Theory

This chapter introduces the two fundamental theoretical parts on which the rest of the chapters depend, the stochastic differential equations used to model Gaussian fields and the Gaussian Markov random fields used to do computations with the models.

2.1 Gaussian Markov random fields

2.1.1 Definition

A Gaussian Markov random field can be seen as a way of emphasizing the conditional independence structure of a Gaussian random field on a finite set of points. Gaussian random fields are usually defined through their finite dimensional distributions.

Definition 2.1.1 (Gaussian random field (GRF))**.** Let $\mathcal{D} \subset \mathbb{R}^n$, then the random field $\{x(\boldsymbol{t}) : \boldsymbol{t} \in \mathcal{D}\}$ is said to be a Gaussian random field if for all $m \in \mathbb{N}$ for all choices of points $t_1, \ldots, t_m \in \mathcal{D}$, $(x(t_1), \ldots, x(t_m))$ has a multivariate Gaussian distribution.

From this definition one can see that the defining property of a GRF is that all finite dimensional distributions of the field are multivariate Gaussian. For a Gaussian field on a finite set, this means that the joint distribution of all the variables is a multivariate Gaussian distribution. This distribution can be characterized by its mean $\boldsymbol{\mu}$ and its covariance matrix $\boldsymbol{\Sigma}$. However, in the setting of Gaussian Markov random fields, the *precision matrix* $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ is a more useful matrix than the covariance matrix. The reason for this is that the off-diagonal elements of the precision matrix characterizes the conditional dependence properties of the distribution.

Theorem 2.1.1 (Conditional independence and the precision matrix)**.** *Let the set* $\mathcal{V} = \{1, 2, \ldots, n\}$ *and let* x *be a Gaussian field on* $\mathcal{V}$ *with mean* $\boldsymbol{\mu}$ *and precision matrix* $\mathbf{Q} > 0$*. Then for* $i \neq j$*,* $x(i)|\{x(k) : k \in \mathcal{V} - \{i, j\}\}$ *is independent of* $x(j)|\{x(k) : k \in \mathcal{V} - \{i, j\}\}$ *if and only if* $Q_{i,j} = 0$*.*

Proof. Can be found in Rue and Held (2005, Theorem 2.2). □

Loosely speaking, a Gaussian Markov random field is a multivariate Gaussian distribution that satisfy a certain conditional independence structure. To make this more rigorous the concept of a graph is needed. For the purpose of this thesis all graphs are finite and unordered unless otherwise specified, and a graph is defined to have these properties. Since the graph is finite, its vertices can without loss of generality be numbered from 1 to n, where n is the number of vertices.

Definition 2.1.2 (Graph)**.** A graph is an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ comprising a finite set $\mathcal{V}$ of vertices and a set $\mathcal{E}$ of edges, where $\mathcal{E}$ is a set of unordered pairs of elements of $\mathcal{V}$. If $\mathcal{V} = \{1, 2, \ldots, n\}$, the graph is called a labelled graph.

Visually, one can think of this kind of graph as a set of n points with a line between two points, i and j, if and only if $\{i, j\} \in \mathcal{E}$. Let $\boldsymbol{x}$ have a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$. Then one can define a conditional independence structure on $\boldsymbol{x}$ from a graph by demanding x_i and x_j conditionally independent if and only if there is no edge between vertex i and vertex j. Theorem 2.1.1 shows that this condition can be specified through the non-zero elements of $\mathbf{Q}$.

Definition 2.1.3 (Gaussian Markov random field (GMRF))**.** A random vector $\boldsymbol{x} \in \mathbb{R}^n$ is called a Gaussian Markov random field with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q} > 0$, if and only if its probability density is of the form

$$\pi(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{Q} (\boldsymbol{x} - \boldsymbol{\mu})\right), \tag{2.1}$$

and for all $i \neq j$, $Q_{i,j} \neq 0$ if and only if $\{i, j\} \in \mathcal{E}$.

Since it is cumbersome to always talk about the graph and the graph can be directly inferred from the non-zero elements of the precision matrix, the graph is suppressed in the notation unless it is explicitly needed.

Note that Definition 2.1.3 only allows $\mathbf{Q} > 0$, which is the requirement for a proper distribution. It is useful to also allow positive semi-definite matrices, which are not positive definite. This gives improper distributions for which the probability density function in Equation (2.1) is not applicable. The problem is

that there are certain linear combinations of the variables that have unbounded variance. This means that it is not meaningful to sample from the distribution unless one conditions on that the linear combinations with unbounded variance have some fixed values. This is not a problem if the distribution is part of some model in which the final distributions become proper when conditioned on the observations.

From the definition of GMRFs one can show that a GMRF satisfies certain Markov properties. The first one is used explicitly in the definition of a GMRF, namely the pairwise Markov property. There are two other types of Markov properties that are also useful. The following theorem from Rue and Held (2005, Theorem 2.4) shows that they are equivalent for a GMRF.

Theorem 2.1.2 (Markov properties of a GMRF)**.** *Let $\boldsymbol{x}$ be a GMRF with respect to the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let* $\mathrm{ne}(i)$ *denote the set of neighbours of i. In addition, use the notation $\boldsymbol{x}_A = \{x_i : i \in A\}$ and $\boldsymbol{x}_{-A} = \{x_i : i \in \mathcal{V} - A\}$ for $A \subset \mathcal{V}$. Then the following are equivalent.*

(i) The pairwise Markov property:
$x_i | \boldsymbol{x}_{-\{i,j\}}$ is independent of $x_j | \boldsymbol{x}_{-\{i,j\}}$ if $\{i, j\} \in \mathcal{E}$ and $i \neq j$.

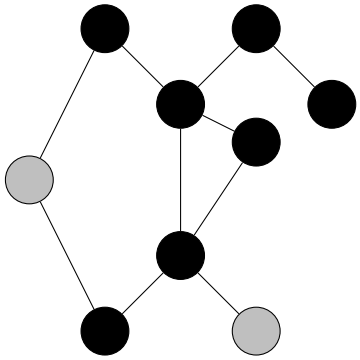
(ii) The local Markov property:
$x_i | \boldsymbol{x}_{\mathrm{ne}(i)}$ is independent of $\boldsymbol{x}_{-\{i,\mathrm{ne}(i)\}} | \boldsymbol{x}_{\mathrm{ne}(i)}$ for all $i \in \mathcal{V}$.

(iii) The global Markov property:
$\boldsymbol{x}_A | \boldsymbol{x}_C$ is independent of $\boldsymbol{x}_B | \boldsymbol{x}_C$ for all pairwise disjoint sets $A, B, C \subset \mathcal{V}$, where C separates A and B, and A and B are non-empty.

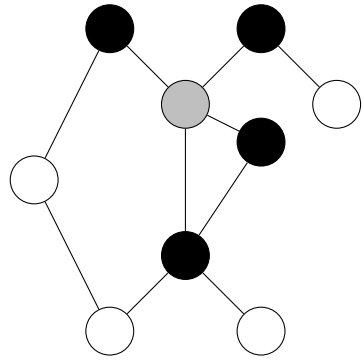
These Markov properties show that the graph gives a useful way of visualizing the conditional dependencies in the GMRF. These properties do not require the knowledge of the actual values of the elements of the precision matrix, only the knowledge of which elements are non-zero, i.e. the graph of the GMRF. Figure 2.1 illustrates each of the Markov properties in Theorem 2.1.2.

2.1.2 Numerical benefits

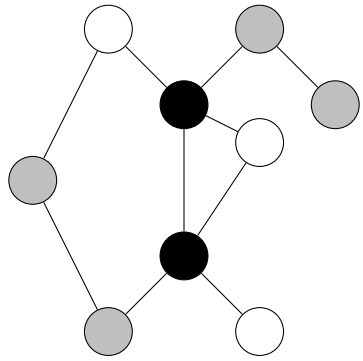
The main reason that GMRFs are used in this thesis is that the problems considered result in spatial GMRFs with each variable only conditionally dependent on the variables closest to itself. This in turn means that the precision matrix is very sparse, and in comparison the covariance matrix is (typically) dense. The sparseness of the precision matrix can be exploited to do both simulations from the GMRF and calculations of the probability density much faster than based on the dense covariance matrix. Additionally, it also requires less memory to store a sparse matrix than a dense matrix.



(a) Pairwise Markov property.



(b) Local Markov property.



(c) Global Markov property.

Figure 2.1: Illustration of different Markov properties. (a) The two grey nodes have no direct edge and are independent given the black nodes. (b) The grey node and the white nodes are independent given the neighbours of the grey node (the black nodes). (c) The two collections of grey nodes are independent given the black nodes which separate them.

One general way to do a simulation from a GMRF is to first calculate the Cholesky decomposition of the precision matrix, and then use the Cholesky factor to transform a standard multivariate Gaussian variable into the correct distribution. Theorem 2.1.3 shows that a simulation can be found by solving a lower triangular linear system of equations. The proof of the theorem is omitted, but only requires that one verifies that the mean and the covariance are correct.

Theorem 2.1.3 (Simulation)**.** *Let $\boldsymbol{x}$ be a GMRF with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q} > 0$, and let $\mathbf{L}$ be the Cholesky factor of $\mathbf{Q}$. Then*

$$\mathbf{L}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

This means that the complexity of finding the Cholesky factor and the complexity of solving the linear system determine the total complexity of doing a simulation. The Cholesky decomposition is the most expensive of these two, and the more efficiently one can do this step, the more efficiently one can do simulations. If the precision matrix is dense, the complexity of the Cholesky decomposition is $\mathcal{O}(n^3)$, where n is the dimension of the distribution. This is also the general complexity when using a dense covariance matrix. This becomes very expensive for the type of dimensions that is used in this thesis, therefore one should rather take advantage of the sparseness of the precision matrix. A second point is that the space complexity for dealing with a dense matrix is $\mathcal{O}(n^2)$, which may be larger than the available memory on the computer for the size of n that is needed.

A very nice property of the Cholesky decomposition is that the Cholesky factor usually is sparse if the original matrix is sparse. In fact, if a matrix is banded, then the Cholesky factor is banded with the same (lower) bandwidth. Further, one can also make statements in the case that the matrix is sparse, but not banded. See Rue and Held (2005, Section 2.4) for details. The important point is that one can implement algorithms that are faster than for dense matrices. For the spatial GMRFs the computational complexity is typically $\mathcal{O}(n^{3/2})$, which is a significant gain from the dense matrices.

One point that should be noted is that when working with precision matrices, one no longer has the marginal properties of the distribution directly available as for the covariance matrices. Thus computations are required to find, for example, marginal variances. It turns out that one can take advantage of the sparseness of the precision matrix also in this case. See Gelfand et al. (2010, Chapter 12) for further details.

Note that the implementation of Cholesky factorization of sparse matrices is a well known problem in numerics and that there exist libraries that do this. The same is true for sparse matrix operations such as addition and multiplication. The calculation of the marginal variances as presented in Gelfand et al. (2010, Chapter 12), however, has to be implemented.

2.2 Stochastic differential equations

2.2.1 Preliminaries

To talk about stochastic differential equations, it is necessary to introduce the concepts of stochastic processes and the calculus associated with these processes. In addition, the concept of a generalized stochastic process is briefly mentioned due to the very important generalized stochastic process called white noise. This is not an extensive introduction and most of the details are skipped. Since only Gaussian processes are needed in the thesis, this discussion is limited to these. Additionally, only real-valued stochastic processes are considered.

A *Gaussian stochastic process* is simply another name for a Gaussian random field. However, for the calculus setting that is considered here, the random field has to be indexed by a set on which it is meaningful to do calculus. Without dwelling on what this actually means, it can be sets like $\mathbb{R}^n$ or $[0,1]^n$, for some $n > 0$. In short, things that it is possible to do calculus on for deterministic functions.

If one ignores the sample path properties of a Gaussian stochastic process, it is uniquely defined by the mean value of the Gaussian variable at each position and the covariance between any pair of Gaussian variables.

Definition 2.2.1 (Covariance and mean value functions)**.** Let $\{u(\boldsymbol{t}) : \boldsymbol{t} \in E\}$ be a Gaussian stochastic process, then the function $c : E \times E \to \mathbb{R}$ defined by

$$c(\boldsymbol{t}, \boldsymbol{s}) = \mathrm{Cov}[u(\boldsymbol{t}), u(\boldsymbol{s})]$$

is called the covariance function of the process and the function $m : E \to \mathbb{R}$ defined by

$$m(\boldsymbol{t}) = \mathrm{E}[u(\boldsymbol{t})]$$

is called the mean value function of the process.

Since these functions define all mean values and covariances, they uniquely define all finite dimensional distributions of the Gaussian stochastic process. If the domain of the stochastic process is obvious from the context, the stochastic process $\{u(\boldsymbol{t}) : \boldsymbol{t} \in E\}$ is simply called the stochastic process u.

It is possible to introduce different types of calculi, for example, from sample path properties or from quadratic mean properties. As mentioned above, the sample path properties of a Gaussian stochastic process is not determined by the finite dimensional distributions. In particular, one can construct a Gaussian stochastic process with continuous sample paths and one with discontinuous sample paths, but both with the same covariance function and mean value function. Loosely speaking, one can call these processes different versions of each other and

always "select" the one with continuous sample paths if it exists. For example, when a Wiener process is used, it is always the continuous version.

However, as the quadratic mean version of integration and differentiation is used in what follows, this is not a big concern. Differentiation can be defined as for deterministic functions, only using a different sense of convergence.

Definition 2.2.2 (Quadratic mean derivative)**.** Let $\{u(\boldsymbol{t}) : \boldsymbol{t} = (t_1, \cdots, t_n) \in \mathbb{R}^n\}$ be a Gaussian stochastic process. Then the Gaussian stochastic process $\{y_i(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^n\}$ is a quadratic mean derivative of u with respect to t_i if

$$y_i(\boldsymbol{t}) = \lim_{h \to 0} \frac{u(\boldsymbol{t} + h\boldsymbol{e}_i) - u(\boldsymbol{t})}{h}, \qquad \forall t_i \in \mathbb{R},$$

where $\boldsymbol{e}_i$ is the vector with 1 at position i and 0 at all other positions and the limit is taken in quadratic mean.

The covariance functions and the mean value functions of the resulting Gaussian fields can easily be calculated. The mean value functions of the derivatives are simply the corresponding derivatives of the mean value function of the stochastic process. For the covariance functions and cross-covariance functions see Abrahamsen (1997, p. 22) for references. The results in Abrahamsen (1997, p. 22) can be summarized in a theorem.

Theorem 2.2.1 (Covariance and cross-covariance of derivatives)**.** *Let $\{u(\boldsymbol{t}) : \boldsymbol{t} = (t_1, \cdots, t_n) \in \mathbb{R}^n\}$ be a Gaussian stochastic process with covariance function c. Then the following hold*

(i)

$$\operatorname{Cov}\left[u(\boldsymbol{t}), \frac{\partial u}{\partial s_i}(\boldsymbol{s})\right] = \frac{\partial c}{\partial s_i}(\boldsymbol{t}, \boldsymbol{s}), \qquad i = 1, \ldots, n.$$

(ii)

$$\operatorname{Cov}\left[\frac{\partial u}{\partial t_i}(\boldsymbol{t}), \frac{\partial u}{\partial s_j}(\boldsymbol{s})\right] = \frac{\partial^2 c}{\partial t_i \partial s_j}(\boldsymbol{t}, \boldsymbol{s}), \qquad i, j = 1, \ldots, n.$$

This theorem indicates that there is going to be a problem if the covariance function is not sufficiently differentiable, because the covariance function as defined in the theorem fails to exist. It can in fact be shown that a process, with mean 0, is quadratic mean differentiable if and only if the covariance function is sufficiently differentiable. See Abrahamsen (1997, Theorem 2.4). It is not stated as a theorem here, as it is also meaningful to consider derivatives that are not regular stochastic processes.

The next part of quadratic mean calculus that is needed is integration. There exist more advanced types of constructions of integrals and there are some remarks about this in Section 2.2.5, but the type of integrals that are needed can be

constructed in the same way as for deterministic integrals. Assume for simplicity that the mean value functions of the processes are 0.

Let u be a stochastic process on $\mathbb{R}^n$ and let $A \subset \mathbb{R}^n$, then integrals such as

$$\int_A g(\boldsymbol{t})u(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t},$$

where g and the covariance function of u satisfy certain conditions, can be evaluated as usual Riemann integrals.

Theorem 2.2.2 (Quadratic mean integration)**.** *Let $\{u(\boldsymbol{t}) : \boldsymbol{t} \in E \subset \mathbb{R}^n\}$, where $E = [0, A_1]\times\cdots\times[0, A_n]$ for positive constants $A_1, \ldots, A_n$, be a Gaussian stochastic process with covariance function c and mean 0. Then if c is continuous and $g : E \to \mathbb{R}$ is such that*

$$Q_1 = \int_E \int_E g(\boldsymbol{t})g(\boldsymbol{s})c(\boldsymbol{t}, \boldsymbol{s})\,\mathrm{d}\boldsymbol{t}\,\mathrm{d}\boldsymbol{s} < \infty, \tag{2.2}$$

then the integral

$$J_1 = \int_E g(\boldsymbol{t})x(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}$$

exists as a quadratic mean limit of the usual construction of the Riemann integral for deterministic functions and

$$J_1 \sim \mathcal{N}(0, Q_1).$$

Proof. A partition of E can be constructed from a partition of each interval $[0, A_i]$ into $\mathcal{P}_i$ by taking

$$\mathcal{P} = \{I_1 \times \cdots \times I_n : I_1 \in \mathcal{P}_1, \ldots, I_n \in \mathcal{P}_n\}.$$

Based on this partition one gets a sum

$$J = \sum_{\alpha\in\mathcal{P}} g(\boldsymbol{t}_\alpha)x(\boldsymbol{t}_\alpha)M(\alpha),$$

where $\boldsymbol{t}_\alpha$ is any point in α and $M(\alpha)$ is the volume of α. To show convergence of J_1 in quadratic mean, one must show that for any sequence of partitions $\{\mathcal{P}^{(i)}\}_{i=1}^{\infty}$ where the diameter of the largest element tends to zero and for any choice of points $\boldsymbol{t}_\alpha^i$ in the elements of the partitions, the quadratic mean limits of the resulting sums tend to the same random variable.

Take any such sequence $\{\mathcal{P}^{(i)}\}_{i=1}^{\infty}$ and any choice of points in the elements of the partitions and write the corresponding sums as

$$J^{(i)} = \sum_{\alpha\in\mathcal{P}^{(i)}} g(\boldsymbol{t}_\alpha^i)x(\boldsymbol{t}_\alpha^i)M(\alpha), \qquad i = 1, 2, \ldots.$$

From the Loève criterion this sequence converges to some random variable in quadratic mean if and only if $\mathrm{E}[J^{(m)}J^{(n)}]$ converges to some finite limit. Writing out the expression for the expected value gives

$$\begin{aligned}\mathrm{E}[J^{(m)}J^{(n)}] &= \mathrm{E}\left[\sum_{\alpha\in\mathcal{P}^{(m)}}\sum_{\beta\in\mathcal{P}^{(n)}} g(\boldsymbol{t}_\alpha^m)g(\boldsymbol{t}_\beta^n)u(\boldsymbol{t}_\alpha^m)u(\boldsymbol{t}_\beta^n)M(\alpha)M(\beta)\right] \\ &= \sum_{\alpha\in\mathcal{P}^{(m)}}\sum_{\beta\in\mathcal{P}^{(n)}} g(\boldsymbol{t}_\alpha^m)g(\boldsymbol{t}_\beta^n)c(\boldsymbol{t}_\alpha^m,\boldsymbol{t}_\beta^n)M(\alpha)M(\beta).\end{aligned} \tag{2.3}$$

But this sequence is approximating the integral in Equation (2.2) and since the partition size tends to 0 when n and m tends to infinity, this must converge to Q_1. Thus $J^{(n)}$ converges to some random variable for any sequence of partitions when n tends to infinity.

Put $m = n$ in Equation (2.3) and again this sum converges to Q_1, but it must also converge to the variance of J_1. This means that

$$J_1 \sim \mathcal{N}(0, Q_1).$$

□

This means that these type of integrals can be evaluated by the deterministic integral in Equation (2.2). This theorem used a box set for the integration, but can be extended to more general sets.

2.2.2 Gaussian white noise

The theory in the previous section defines differentiation for Gaussian stochastic processes which have covariance functions that are sufficiently differentiable. But it is desirable to extend what is meant by differentiation to also allow for differentiation of processes that do not have derivatives in the usual sense.

The perhaps most important example is the standard Wiener process, W, on $\mathbb{R}$. It has almost surely unbounded variation on any interval and is almost surely nowhere differentiable. But it is still desirable to give meaning to something like

$$\int_a^b f(t)\frac{\mathrm{d}W}{\mathrm{d}t}(t)\,\mathrm{d}t,$$

where $a < b$ and f is a deterministic function. It is useful to give it an interpretation as a weak derivative. In the case of the standard Wiener process, the derivative can be interpreted through a Riemann-Stieltjes integral

$$\int_a^b f(t)\frac{\mathrm{d}W}{\mathrm{d}t}(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}W(t), \tag{2.4}$$

which is meaningful in quadratic mean sense even though W is (almost surely) not of bounded variation. Thus one can think of the derivative of the Wiener process as a distribution valued stochastic processes even though the process is not differentiable in the usual sense. These type of stochastic processes are called generalized stochastic processes.

The most important generalized stochastic process is the white noise process. It can not be defined through its finite dimensional distributions, as the pointwise evaluations needed to define these are not meaningful. It is necessary to describe its effect on sets, specifically sets which have a meaningful measure of "size". Adler and Taylor (2007, Theorem 1.4.3) shows that the following defines a valid stochastic process.

Definition 2.2.3 (Gaussian white noise)**.** Let $(E, \mathcal{E}, \nu)$ be a σ-finite measure space and let $\mathcal{T}_\nu$ denote the family of sets of $\mathcal{E}$ of finite ν-measure. Then a Gaussian white noise based on ν is a random field $\mathcal{W} : \mathcal{T}_\nu \to \mathbb{R}$ such that for all $A, B \in \mathcal{T}_\nu$,

(i) $\mathcal{W}(A) \sim \mathcal{N}(0, \nu(A))$,

(ii) $A \cap B = \emptyset$ implies that $\mathcal{W}(A \cup B) = \mathcal{W}(A) + \mathcal{W}(B)$,

(iii) $A \cap B = \emptyset$ implies that $\mathcal{W}(A)$ and $\mathcal{W}(B)$ are independent.

A simple example of this construction is to let $E = \mathbb{R}^n$, for some $n > 0$, $\mathcal{E}$ be the Borel σ-algebra on $\mathbb{R}^n$ and ν be the Lebesgue measure on $\mathbb{R}^n$. This gives what is called a *standard Gaussian white noise process* on $\mathbb{R}^n$. Note that the set E does not have to be Euclidean and can be something like a sphere or a torus and ν an appropriate measure of the area of a set on the sphere or the torus.

Let $\mathcal{W}$ be a standard Gaussian white noise process on $\mathbb{R}^n$ for some $n > 0$. Then $\mathcal{W}$ can be considered as a form of random measure and one can give meaning to expressions such as

$$\int_A f(\boldsymbol{s})\mathcal{W}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s},$$

where f satisfies certain conditions. In this expression $\mathcal{W}$ is written as a function of position in the same manner as a Dirac delta function is written as a function of position even though it is a generalized function. One can show that the following theorem is true, but the proof is not given here.

Theorem 2.2.3. *Let $\mathcal{W}$ be a standard Gaussian white noise process on $\mathbb{R}^n$, for some $n > 0$, and let $L^2(\mathbb{R}^n)$ be the set of Lebesgue square-integrable functions from $\mathbb{R}^n$ to $\mathbb{R}$. Then the following holds for all $f, g \in L^2(\mathbb{R}^n)$.*

(i) $\int_{\mathbb{R}^n} f(\boldsymbol{s})\mathcal{W}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$ *has a Gaussian distribution.*

(ii) $\mathrm{E}\left[\int_{\mathbb{R}^n} f(\boldsymbol{s})\mathcal{W}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right] = 0.$

(iii) $\mathrm{E}\left[\left(\int_{\mathbb{R}^n} f(\boldsymbol{s})\mathcal{W}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right)\cdot\left(\int_{\mathbb{R}^n} g(\boldsymbol{s})\mathcal{W}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right)\right] = \int_{\mathbb{R}^n} f(\boldsymbol{s})g(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}.$

Both this theorem and Definition 2.2.3 give the properties needed for the finite volume method approximation used in the next chapter. Namely, that for a Lebesgue measurable subset A of $\mathbb{R}^n$, for some $n > 0$,

$$\int_A \mathcal{W}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} \sim \mathcal{N}(0, |A|)$$

and that for two disjoint Lebesgue measurable subsets A and B of $\mathbb{R}^n$ the integral over A and the integral over B are independent. In the remaining chapters the statement that the sets are Lebesgue measurable is dropped, because there is no need to consider sets which fail to fulfil this condition.

One can show that the use of the derivative of the standard Wiener process as in Equation (2.4) has the same properties as a standard Gaussian white noise process on $\mathbb{R}$. Therefore, one often says that the derivative of a standard Wiener process is a standard Gaussian white noise process. Walsh (1986) has more information on the connection between Brownian sheets, which are the generalization of the Wiener process to higher dimensions, and Gaussian white noise.

2.2.3 Definition

A *stochastic differential equation*, abbreviated SDE, is a differential equation which involves one or more stochastic processes. The stochastic processes may also be generalized stochastic processes. Further, the expression *stochastic partial differential equation*, abbreviated SPDE, can be used to emphasize that there is more than one free variable. For the purpose of this thesis only linear SPDEs are considered. Let u be the stochastic process that the SPDE describes, then it is a *linear SPDE* if the left hand side is linear in u and its derivatives and the right hand side is a known stochastic process.

Example 2.2.1. A simple example of a linear SPDE for u over $\mathbb{R}^2$ is

$$\sum_{k=0}^{n}\sum_{l=0}^{m} A_{k,l}(x,y)\frac{\partial^{k+l}}{\partial x^k \partial y^l}u(x,y) = v(x,y), \qquad (x,y) \in \mathbb{R}^2, \tag{2.5}$$

where n and m are non-negative integers, $A_{k,l}$ is a deterministic function for $k = 0, \ldots, n$ and $l = 0, \ldots, m$ and v is a known Gaussian stochastic process.

It is clear that a solution u of the SPDE in Equation (2.5) should be a stochastic process which makes the left hand side and the right hand side equal, but it is not clear in what sense they should be equal. Since the derivatives previously were defined in the quadratic mean sense, the same is done for the equality. If v is a Gaussian stochastic process, then a *solution* is a stochastic process u such that the left hand side and the right hand side have the same mean value function and covariance function. Note that some of the derivatives may be generalized stochastic processes.

If v is a generalized stochastic process, the covariance function does not exist in the usual sense and a different approach must be taken. Since the only generalized stochastic process that is used on the right hand side in this thesis is a standard Gaussian white noise process, this is the only one that is considered. Simply put, in this case a *solution* is a stochastic process which gives the left hand side the same properties as a standard Gaussian white noise process. That is the left hand side should satisfy the properties in Theorem 2.2.3.

2.2.4 Random coefficients

In the previous section the linear SPDEs have deterministic coefficients. This does not have to be the case. Also SPDEs with random coefficients are encountered in later chapters. But the randomness can in the specific cases encountered in a sense be ignored when solving the SPDEs.

Consider the SPDE

$$f(t, A_1, \ldots, A_k)u'(t) = \mathcal{W}(t), \qquad u(0) = 0,$$

where $\mathcal{W}$ is a Gaussian white noise process, f is a deterministic function and $A_1, A_2, \ldots, A_k$ are random variables independent of the Gaussian white noise process. Here the fact that the coefficient in front of the u' term is random does not impose any difficulties. Because it can be interpreted in the sense that for each value of the variables $A_1, A_2, \ldots, A_k$ there is a solution which can be found in the usual sense. And one can say that the resulting joint distribution is the "solution" of the SPDE. This is the useful way of considering the SPDE for the inference in Chapter 4.

2.2.5 Further notes

The theory presented in the previous sections did not use any "new" form of calculus. The only thing that was changed in the integration and the differentiation was the sense in which the limits should hold. For more general integrands this

is not possible. A simple example is

$$\int_a^b W(t)\,\mathrm{d}W(t),$$

where $a < b$ and W is a standard Wiener process. This is a Wiener process integrated with respect to itself and it is not meaningful to define it as a usual Riemann-Stieltjes integral. This is because it is not a well defined construction, the resulting random variable depends on how the sums are constructed. The problem is that if one creates a partition $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$ and writes the sum as

$$\sum_{k=1}^{n} W(t_k^*)(W(t_k) - W(t_{k-1})),$$

the two choices $t_k^* = t_{k-1}$ and $t_k^* = (t_{k-1} + t_k)/2$ give different results as the size of the partition tends to zero. These choices are connected with Itō calculus and Stratonovich integrals, respectively, and more details can be found in Øksendal (2003).

For SDEs it becomes a problem for equations such as

$$u(t) - \frac{\partial u}{\partial t}(t) = \sigma(u(t), t)\mathcal{W}(t),$$

where $\mathcal{W}(t)$ is a standard Gaussian white noise process, in which one of the coefficients depends on the solution itself. But no equations such as this is encountered in later chapters. Therefore, there is no further discussion of the problem.

Chapter 3

Modelling inhomogeneity through diffusion

The first goal of this chapter is to derive a GMRF approximation for the solution of the SPDE

$$\kappa^2(\boldsymbol{s})u(\boldsymbol{s}) - \nabla \cdot \mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in [0, A] \times [0, B], \tag{3.1}$$

where A and B are strictly positive constants, κ^2 is a scalar function, $\mathbf{H}$ is a 2×2 matrix valued function, $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ and $\mathcal{W}$ is a standard Gaussian white noise process. In addition, κ^2 is assumed to be a continuous, strictly positive function and $\mathbf{H}$ is assumed to be a continuously differentiable function which gives a positive definite matrix $\mathbf{H}(\boldsymbol{s})$ for all $\boldsymbol{s} \in [0, A] \times [0, B]$.

Further, periodic boundary conditions are used. This means that opposite sides of the rectangle $[0, A] \times [0, B]$ are identified. The variable loops around the rectangle in the sense that $\boldsymbol{s} + (A, 0)$ and $\boldsymbol{s} + (0, B)$ are the same point as $\boldsymbol{s}$. This means that the space can be viewed as a torus parametrized as the Cartesian product of two circles, with the distance defined as the square root of the sum of the squared angular distances. This gives additional requirements for κ^2 and $\mathbf{H}$. The values of κ^2 must agree on opposite edges and the values of $\mathbf{H}$ and its first order derivatives must agree on opposite edges.

The second goal of this chapter is to study some of the GMRFs that can be built by using different choices of functions for $\mathbf{H}$.

3.1 GMRF approximation

3.1.1 Finite volume methods

Finite volume methods are useful for creating discretizations of conservative laws of the form

$$\frac{\partial q}{\partial t}(\boldsymbol{x}, t) + \nabla \cdot \boldsymbol{F}(\boldsymbol{x}, t) = f(\boldsymbol{x}, t),$$

where $\nabla\cdot$ is the spatial divergence operator. This equation relates the change of q in time to the spatial divergence of the flux $\boldsymbol{F}$ and the sink-/source-term f. In this thesis there is no time dependence, but the method is still useful by considering the differential equation to describe a system that has reached a steady state for a right hand side that does not depend on time.

The main tool in these methods is the use of the divergence theorem

$$\int_E \nabla \cdot \boldsymbol{F} \, \mathrm{d}V = \oint_{\partial E} \boldsymbol{F} \cdot \boldsymbol{n} \, \mathrm{d}\sigma, \tag{3.2}$$

where $\boldsymbol{n}$ is the outer normal vector of the surface ∂E relative to E.

The main idea is to divide the domain of the SPDE in Equation (3.1) into smaller parts and consider the resulting "flow" between the different parts. A lengthy treatment of finite volume methods is not given, but a comprehensive treatment of the method for non-stochastic differential equations can be found in Eymard et al. (2000).

3.1.2 Discretization scheme

The discretization of the SPDE in Equation (3.1) is done by a finite volume method. To keep the calculations simple the domain is divided into a regular grid of rectangular cells. Let there be N cells along the x-coordinate and M cells along the y-coordinate, then for each cell the sides parallel to the x-axis have length $h_x = A/N$ and the sides parallel to the y-axis have length $h_y = B/M$. Number the cells by (i, j), where i is the column of the cell (along the x-axis) and j is the row of the cell (along the y-axis). Call the lowest row 0 and the leftmost column 0, then cell (i, j) is

$$E_{i,j} = [ih_x, (i+1)h_x] \times [jh_y, (j+1)h_y].$$

Using this notation the set of cells, $\mathcal{I}$, is given by

$$\mathcal{I} = \{E_{i,j} : i = 0, 1, \ldots, N-1, j = 0, 1, \ldots, M-1\}.$$

Figure 3.1 shows an illustration of the discretization of $[0, A] \times [0, B]$ into the cells $\mathcal{I}$.

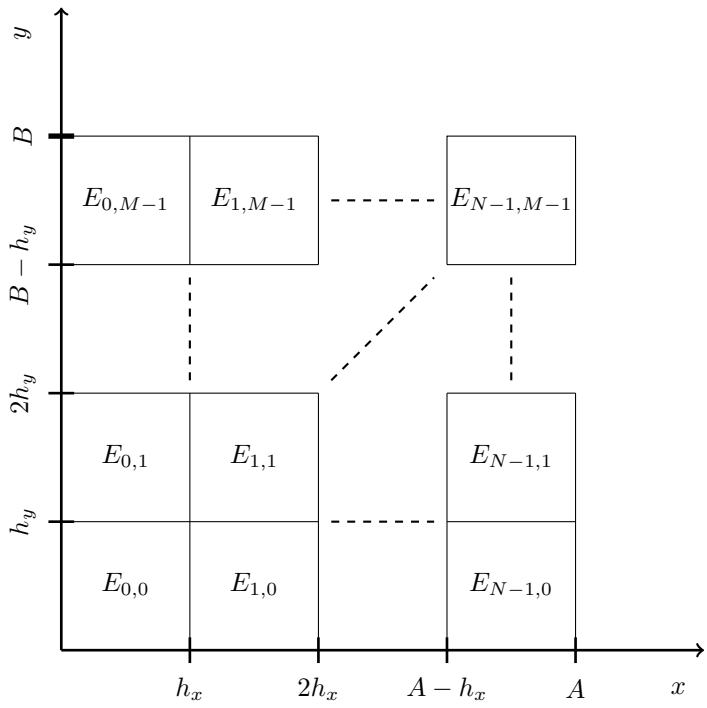


Figure 3.1: Illustration of the division of $[0, A] \times [0, B]$ into a regular $N \times M$ grid of rectangular cells.

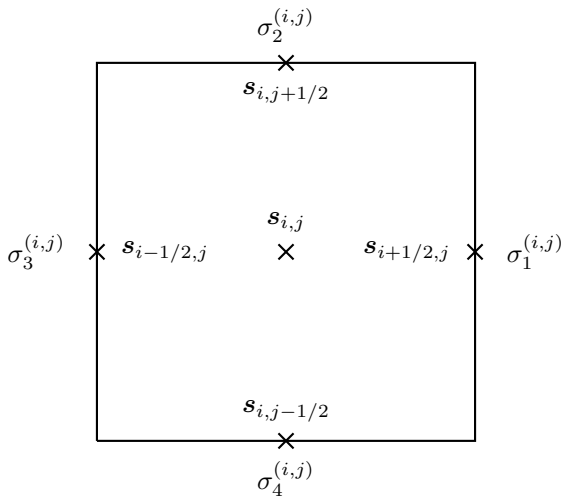


Figure 3.2: One cell, $E_{i,j}$, of the discretization with faces $\sigma_1^{(i,j)}$, $\sigma_2^{(i,j)}$, $\sigma_3^{(i,j)}$ and $\sigma_4^{(i,j)}$, centroid $\boldsymbol{s}_{i,j}$ and centres of the faces $\boldsymbol{s}_{i-1/2,j}$, $\boldsymbol{s}_{i,j-1/2}$, $\boldsymbol{s}_{i+1/2,j}$ and $\boldsymbol{s}_{i,j+1/2}$.

Let $\boldsymbol{s}_{i,j}$ denote the centroid of cell $E_{i,j}$, then

$$\boldsymbol{s}_{i,j} = ((i+1/2)h_x, (j+1/2)h_y), \quad (i,j) \in \{0,\dots,N-1\} \times \{0,\dots,M-1\}.$$

Further, let the expression for $\boldsymbol{s}_{i,j}$ also hold for $i = -1/2, 1/2, 3/2, \dots, N-1/2$ and $j = -1/2, 1/2, 3/2, \dots, M-1/2$. That is the centres of the faces of the cells. For cell $E_{i,j}$ the centres of the faces are $\boldsymbol{s}_{i-1/2,j}$, $\boldsymbol{s}_{i,j-1/2}$, $\boldsymbol{s}_{i+1/2,j}$ and $\boldsymbol{s}_{i,j+1/2}$. Because of the periodic boundary conditions the i-index and the j-index in $\boldsymbol{s}_{i,j}$ are modulo N and modulo M respectively.

Each cell has four faces, two parallel to the x-axis (top and bottom) and two parallel to the y-axis (left and right). Let the right face, top face, left face and bottom face of cell $E_{i,j}$ be denoted $\sigma_1^{(i,j)}$, $\sigma_2^{(i,j)}$, $\sigma_3^{(i,j)}$ and $\sigma_4^{(i,j)}$ respectively. Further, let $\sigma(E_{i,j})$ be the set of faces of cell $E_{i,j}$. Figure 3.2 shows one cell $E_{i,j}$ with the centroid and the faces marked on the figure.

In addition, let the value of the function u at the centroid of cell $E_{i,j}$ be denoted by $u_{i,j} = u(\boldsymbol{s}_{i,j})$ for all the centroids, and denote the area of $E_{i,j}$ by $V_{i,j}$. Since the grid is regular, $V_{i,j} \equiv V = h_x h_y$.

To derive the finite volume scheme, begin by integrating Equation (3.1) over a cell, $E_{i,j}$. This gives

$$\int_{E_{i,j}} \kappa^2(\boldsymbol{s}) u(\boldsymbol{s}) \,\mathrm{d}A - \int_{E_{i,j}} \nabla \cdot \mathbf{H}(\boldsymbol{s}) \nabla u(\boldsymbol{s}) \,\mathrm{d}A = \int_{E_{i,j}} \mathcal{W}(\boldsymbol{s}) \,\mathrm{d}A, \tag{3.3}$$

where $\mathrm{d}A$ is an area element.

As discussed in Section 2.2.2, the integral on the right hand side is distributed as a Gaussian variable with mean 0 and variance V for each (i, j). Further, the integrals are independent, because all $E_{i,j}$ are disjoint. Thus Equation (3.3) can be written as

$$\int_{E_{i,j}} \kappa^2(\boldsymbol{s}) u(\boldsymbol{s}) \,\mathrm{d}A - \int_{E_{i,j}} \nabla \cdot \mathbf{H}(\boldsymbol{s}) \nabla u(\boldsymbol{s}) \,\mathrm{d}A = \sqrt{V} z_{i,j},$$

where $z_{i,j}$ is a standard Gaussian variable for all (i, j) and the Gaussian variables are independent.

By the divergence theorem in Equation (3.2), the second integral on the left hand side can be turned into an integral over the boundary of the cell. This results in

$$\int_{E_{i,j}} \kappa^2(\boldsymbol{s}) u(\boldsymbol{s}) \,\mathrm{d}A - \oint_{\partial E_{i,j}} (\mathbf{H}(\boldsymbol{s}) \nabla u(\boldsymbol{s}))^{\mathrm{T}} \boldsymbol{n}(\boldsymbol{s}) \,\mathrm{d}\sigma = \sqrt{V} z_{i,j}, \tag{3.4}$$

where $\boldsymbol{n}$ is the exterior normal vector of $\partial E_{i,j}$ with respect to $E_{i,j}$ and $\mathrm{d}\sigma$ is a line element. It is useful to divide the integral over the boundary in Equation (3.4) into integrals over each face

$$\int_{E_{i,j}} \kappa^2(\boldsymbol{s}) u(\boldsymbol{s}) \,\mathrm{d}A - \sum_{l=1}^{4} W_l^{(i,j)} = \sqrt{V} z_{i,j}, \tag{3.5}$$

where $W_l^{(i,j)} = \int_{\sigma_l^{(i,j)}} (\mathbf{H}(\boldsymbol{s}) \nabla u(\boldsymbol{s}))^{\mathrm{T}} \boldsymbol{n}(\boldsymbol{s}) \,\mathrm{d}\sigma$.

The first integral on the left hand side of Equation (3.5) is approximated by

$$\int_{E_{i,j}} \kappa^2(\boldsymbol{s}) u(\boldsymbol{s}) \,\mathrm{d}A = V \kappa_{i,j}^2 u(\boldsymbol{s}_{i,j}), \tag{3.6}$$

where $\kappa_{i,j}^2 = \frac{1}{V} \int_{E_{i,j}} \kappa^2(\boldsymbol{s}) \,\mathrm{d}A$. For the purpose of this thesis κ^2 is assumed to be continuous and $\kappa_{i,j}^2$ is approximated by $\kappa^2(\boldsymbol{s}_{i,j})$.

The second part of Equation (3.5) requires the approximation of the surface integral over each face of a given cell. The values of $\mathbf{H}$ are in general not diagonal, so it is necessary to estimate both components of the gradient on each face of the cell. For simplicity, it is assumed that the gradient is constant on each face and that it is identically equal to the value at the centre of the face. On a face parallel to the y-axis the estimate of the partial derivative with respect to x is simple since the centroid of each of the cells which share the face have the same y-coordinate. The problem is the estimate of the partial derivative with respect to y. The reverse is true for the top and bottom face of the cell.

It is important to use a scheme which gives the same estimate of the gradient for a given face no matter which of the two neighbouring cells are chosen. For the right face of $E_{i,j}$, that is $\sigma_1^{(i,j)}$, the approximation used is

$$\frac{\partial}{\partial y}u(\boldsymbol{s}_{i+1/2,j}) \approx \frac{1}{h_y}(u(\boldsymbol{s}_{i+1/2,j+1/2}) - u(\boldsymbol{s}_{i+1/2,j-1/2})),$$

where the value of u at $\boldsymbol{s}_{i+1/2,j+1/2}$ and $\boldsymbol{s}_{i+1/2,j-1/2}$ are linearly interpolated from the values at the four closest cells. More precisely, because of the regularity of the grid the mean of the four closest cells are used. This gives

$$\frac{\partial}{\partial y}u(\boldsymbol{s}_{i+1/2,j}) \approx \frac{1}{4h_y}(u_{i+1,j+1} + u_{i,j+1} - u_{i,j-1} - u_{i+1,j-1}). \tag{3.7}$$

Note that this formula can be used for the partial derivative with respect to y on any face parallel to the y-axis, by suitably changing the i and j indices. The partial derivative with respect to x on a face parallel to the y-axis can be approximated directly by

$$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i+1/2,j}) \approx \frac{1}{h_x}(u_{i+1,j} - u_{i,j}). \tag{3.8}$$

In more or less exactly the same way the two components of the gradient on the top face of cell $E_{i,j}$ can be approximated by

$$\frac{\partial}{\partial x}u(\boldsymbol{s}_{i,j+1/2}) \approx \frac{1}{4h_x}(u_{i+1,j+1} + u_{i+1,j} - u_{i-1,j} - u_{i-1,j+1}) \tag{3.9}$$

and

$$\frac{\partial}{\partial y}u(\boldsymbol{s}_{i,j+1/2}) \approx \frac{1}{h_y}(u_{i,j+1} - u_{i,j}). \tag{3.10}$$

These approximations can be used on any side parallel to the x-axis by changing the indices appropriately.

The approximations for the partial derivatives on each face are collected in Table 3.1. Using this table one can find the approximations needed for the second part of Equation (3.5). It is helpful to write

$$W_l^{(i,j)} = \int_{\sigma_l^{(i,j)}} (\mathbf{H}(\boldsymbol{s})\nabla u(\boldsymbol{s}))^{\mathrm{T}}\boldsymbol{n}(\boldsymbol{s})\,\mathrm{d}\sigma = \int_{\sigma_l^{(i,j)}} (\nabla u(\boldsymbol{s}))^{\mathrm{T}}(\mathbf{H}(\boldsymbol{s})\boldsymbol{n}(\boldsymbol{s}))\,\mathrm{d}\sigma,$$

where the symmetry of $\mathbf{H}$ is used to avoid transposing the matrix. Assuming that the gradient is identically equal to the value at the centre of the face, one finds

$$W_l^{(i,j)} \approx (\nabla u(\boldsymbol{c}_l^{(i,j)}))^{\mathrm{T}} \int_{\sigma_l^{(i,j)}} \mathbf{H}(\boldsymbol{s})\boldsymbol{n}(\boldsymbol{s})\,\mathrm{d}\sigma,$$

Table 3.1: Finite difference schemes for the partial derivative with respect to x and y at the different faces of cell $E_{i,j}$.

Face	$\frac{\partial}{\partial x}u(s)$	$\frac{\partial}{\partial y}u(s)$
$\sigma_1^{(i,j)}$	$\frac{u_{i+1,j}-u_{i,j}}{h_x}$	$\frac{u_{i,j+1}+u_{i+1,j+1}-u_{i,j-1}-u_{i+1,j-1}}{4h_y}$
$\sigma_2^{(i,j)}$	$\frac{u_{i+1,j}+u_{i+1,j+1}-u_{i-1,j}-u_{i-1,j+1}}{4h_x}$	$\frac{u_{i,j+1}-u_{i,j}}{h_y}$
$\sigma_3^{(i,j)}$	$\frac{u_{i,j}-u_{i-1,j}}{h_x}$	$\frac{u_{i-1,j+1}+u_{i,j+1}-u_{i-1,j-1}-u_{i,j-1}}{4h_y}$
$\sigma_4^{(i,j)}$	$\frac{u_{i+1,j}+u_{i+1,j-1}-u_{i-1,j-1}-u_{i-1,j}}{4h_x}$	$\frac{u_{i,j}-u_{i,j-1}}{h_y}$

where $\boldsymbol{c}_l^{(i,j)}$ is the centre of face $\sigma_l^{(i,j)}$.

Since the cells form a regular grid, $\boldsymbol{n}$ is constant on each face. Let $\mathbf{H}$ be approximated by its value at the centre of the face, then

$$W_l^{(i,j)} \approx m(\sigma_l^{(i,j)})(\nabla u(\boldsymbol{c}_l^{(i,j)}))^{\mathrm{T}}(\mathbf{H}(\boldsymbol{c}_l^{(i,j)})\boldsymbol{n}(\boldsymbol{c}_l^{(i,j)})), \tag{3.11}$$

where $m(\sigma_l^{(i,j)})$ is the length of the face and $\boldsymbol{c}_l^{(i,j)}$ is the centre of the face. One can observe that the length of the face is either h_x or h_y and that the normal vector is parallel to the x-axis or the y-axis.

Let

$$\mathbf{H}(\boldsymbol{s}) = \begin{bmatrix} H^{11}(\boldsymbol{s}) & H^{12}(\boldsymbol{s}) \\ H^{21}(\boldsymbol{s}) & H^{22}(\boldsymbol{s}) \end{bmatrix},$$

then using Table 3.1 one finds the approximations

$$\begin{aligned} \hat{W}_1^{(i,j)} = & \\ & h_y \left[H^{11}(\boldsymbol{s}_{i+1/2,j}) \frac{u_{i+1,j} - u_{i,j}}{h_x} \right] + \\ & h_y \left[H^{21}(\boldsymbol{s}_{i+1/2,j}) \frac{u_{i,j+1} + u_{i+1,j+1} - u_{i,j-1} - u_{i+1,j-1}}{4h_y} \right], \end{aligned}$$

$$\begin{aligned} \hat{W}_2^{(i,j)} = & \\ & h_x \left[H^{12}(\boldsymbol{s}_{i,j+1/2}) \frac{u_{i+1,j+1} + u_{i+1,j} - u_{i-1,j+1} - u_{i-1,j}}{4h_x} \right] + \\ & h_x \left[H^{22}(\boldsymbol{s}_{i,j+1/2}) \frac{u_{i,j+1} - u_{i,j}}{h_y} \right], \end{aligned}$$

$$\hat{W}_3^{(i,j)} = h_y \left[H^{11}(\boldsymbol{s}_{i-1/2,j}) \frac{u_{i-1,j} - u_{i,j}}{h_x} \right] + h_y \left[H^{21}(\boldsymbol{s}_{i-1/2,j}) \frac{u_{i,j-1} + u_{i-1,j-1} - u_{i-1,j+1} - u_{i,j+1}}{4h_y} \right]$$

and

$$\hat{W}_4^{(i,j)} = h_x \left[H^{12}(\boldsymbol{s}_{i,j-1/2}) \frac{u_{i-1,j} + u_{i-1,j-1} - u_{i+1,j} - u_{i+1,j-1}}{4h_x} \right] + h_x \left[H^{22}(\boldsymbol{s}_{i,j-1/2}) \frac{u_{i,j-1} - u_{i,j}}{h_y} \right].$$

These approximations can be combined with the approximation in Equation (3.6) and inserted into Equation (3.5) to give

$$V\kappa_{i,j}^2 u_{i,j} - \sum_{l=1}^{4} \hat{W}_l^{(i,j)} = \sqrt{V} z_{i,j}.$$

Stacking the variables $u_{i,j}$ row-wise in a vector $\boldsymbol{u}$, that is first row 0, then row 1 and so on, gives the linear system of equations,

$$\mathbf{D}_V \mathbf{D}_{\kappa^2} \boldsymbol{u} - \mathbf{A}_{\mathbf{H}} \boldsymbol{u} = \mathbf{D}_V^{1/2} \boldsymbol{z}, \tag{3.12}$$

where $\mathbf{D}_V = V\mathbf{I}_{MN}$, $\mathbf{D}_{\kappa^2} = \text{diag}(\kappa_{0,0}^2, \ldots, \kappa_{N-1,0}^2, \kappa_{0,1}^2, \ldots, \kappa_{N-1,M-1}^2)$, $\boldsymbol{z}$ is a standard multivariate Gaussian variable of dimension MN and $\mathbf{A}_{\mathbf{H}}$ is a matrix considered more closely in what follows.

The construction of the matrix $\mathbf{A}_{\mathbf{H}}$, which depends on the function $\mathbf{H}$, requires only that one writes out the sum

$$\sum_{l=1}^{4} \hat{W}_l^{(i,j)}$$

and collects the coefficients of the different $u_{a,b}$ terms. This is not difficult, but requires many lines of equations. Therefore, only the resulting coefficients are given. Fix (i, j) and consider the equation for cell $E_{i,j}$. For convenience, let i_p and i_n be the column left and right of the current column respectively and let j_n and j_p be the row above and below the current row respectively. These rows and columns are 0-indexed and due to the periodic boundary conditions one has, for example, that column 0 is to the right of column $N - 1$. Further, number the rows and columns of the matrix $\mathbf{A}_{\mathbf{H}}$ from 0 to $MN - 1$.

For row $jN + i$ the coefficient of $u_{i,j}$ itself is given by

$$\begin{aligned}(\mathbf{A_H})_{jN+i,jN+i} = \\ &- \frac{h_y}{h_x}\left[H^{11}(\boldsymbol{s}_{i+1/2,j}) + H^{11}(\boldsymbol{s}_{i-1/2,j})\right] \\ &- \frac{h_x}{h_y}\left[H^{22}(\boldsymbol{s}_{i,j+1/2}) + H^{22}(\boldsymbol{s}_{i,j-1/2})\right].\end{aligned}$$

The four closest neighbours have coefficients

$$\begin{aligned}(\mathbf{A_H})_{jN+i,jN+i_p} &= \frac{h_y}{h_x}H^{11}(s_{i-1/2,j}) - \frac{1}{4}\left[H^{12}(\boldsymbol{s}_{i,j+1/2}) - H^{12}(\boldsymbol{s}_{i,j-1/2})\right], \\ (\mathbf{A_H})_{jN+i,jN+i_n} &= \frac{h_y}{h_x}H^{11}(\boldsymbol{s}_{i+1/2,j}) + \frac{1}{4}\left[H^{12}(\boldsymbol{s}_{i,j+1/2}) - H^{12}(\boldsymbol{s}_{i,j-1/2})\right], \\ (\mathbf{A_H})_{jN+i,j_nN+i} &= \frac{h_x}{h_y}H^{22}(\boldsymbol{s}_{i,j+1/2}) + \frac{1}{4}\left[H^{21}(\boldsymbol{s}_{i+1/2,j}) - H^{21}(\boldsymbol{s}_{i-1/2,j})\right], \\ (\mathbf{A_H})_{jN+i,j_pN+i} &= \frac{h_x}{h_y}H^{22}(\boldsymbol{s}_{i,j-1/2}) - \frac{1}{4}\left[H^{21}(\boldsymbol{s}_{i+1/2,j}) - H^{21}(\boldsymbol{s}_{i-1/2,j})\right].\end{aligned}$$

Lastly, the four diagonally closest neighbours have coefficients

$$\begin{aligned}(\mathbf{A_H})_{jN+i,j_pN+i_p} &= +\frac{1}{4}\left[H^{12}(\boldsymbol{s}_{i,j-1/2}) + H^{21}(\boldsymbol{s}_{i-1/2,j})\right], \\ (\mathbf{A_H})_{jN+i,j_pN+i_n} &= -\frac{1}{4}\left[H^{12}(\boldsymbol{s}_{i,j-1/2}) + H^{21}(\boldsymbol{s}_{i+1/2,j})\right], \\ (\mathbf{A_H})_{jN+i,j_nN+i_p} &= -\frac{1}{4}\left[H^{12}(\boldsymbol{s}_{i,j+1/2}) + H^{21}(\boldsymbol{s}_{i-1/2,j})\right], \\ (\mathbf{A_H})_{jN+i,j_nN+i_n} &= +\frac{1}{4}\left[H^{12}(\boldsymbol{s}_{i,j+1/2}) + H^{21}(\boldsymbol{s}_{i+1/2,j})\right].\end{aligned}$$

The rest of the elements of row $jN + i$ are 0.

Based on Equation (3.12) one can write

$$\boldsymbol{z} = \mathbf{D}_V^{-1/2}\mathbf{A}\boldsymbol{u}, \tag{3.13}$$

where $\mathbf{A} = \mathbf{D}_V\mathbf{D}_{\kappa^2} - \mathbf{A_H}$. This gives the joint distribution of $\boldsymbol{u}$,

$$\begin{aligned}\pi(\boldsymbol{u}) &\propto \pi(\boldsymbol{z}) \propto \exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\right) \\ \pi(\boldsymbol{u}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{D}_V^{-1}\mathbf{A}\boldsymbol{u}\right) \\ \pi(\boldsymbol{u}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}\boldsymbol{u}\right),\end{aligned} \tag{3.14}$$

where $\mathbf{Q} = \mathbf{A}^{\mathrm{T}}\mathbf{D}_V^{-1}\mathbf{A}$.

3.1.3 Appropriate domain size

The SPDE in Equation (3.1) has domain $\mathcal{D} = [0, A] \times [0, B]$ with periodic boundary conditions. As mentioned in the start of the chapter, the periodic boundary conditions mean that opposite sides of the rectangular domain are identified with each other. It is clear that the sizes of the parameters A and B affect the properties of the solution of the SPDE. To study these effects closer, consider the SPDE

$$u(\boldsymbol{s}) - \Delta u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in [0, A] \times [0, B], \tag{3.15}$$

where Δ is the Laplace operator and $\mathcal{W}$ is a standard Gaussian white noise process, with periodic boundary conditions. Note that the solution of this SPDE must be homogeneous, which in particular means the same marginal variance everywhere.

First the marginal variances' dependency on A and B is considered for the SPDE in Equation (3.15). By integrating Equation (3.15) over the domain $\mathcal{D}$, one finds

$$\int_{\mathcal{D}} u \, \mathrm{d}A - \int_{\mathcal{D}} \Delta u \, \mathrm{d}A \stackrel{d}{=} \mathcal{N}(0, V_1),$$

where V_1 is the area of $\mathcal{D}$. It is clear that the integral of the Laplacian of u over the full domain must be zero since there can be no net flux into an area with no boundary. It follows that

$$\int_{\mathcal{D}} u \, \mathrm{d}A \stackrel{d}{=} \mathcal{N}(0, V_1),$$

or in other words

$$\frac{1}{V_1} \int_{\mathcal{D}} u \, \mathrm{d}A \stackrel{d}{=} \mathcal{N}(0, 1/V_1).$$

Combining this statement with the statement that the marginal variances are the same at each point, one sees that if the area of the domain becomes small, the point-wise variances must become large. In fact, they can be made arbitrarily large by decreasing the area. This shows that the marginal variances do not only depend on the chosen κ^2 and $\mathbf{H}$ in the SPDE, but also on A and B.

However, if A and B are so large that the correlation becomes more or less 0 for distances greater than $\min(A, B)/2$, the marginal variances should be approximately the same as for the homogeneous solution of the corresponding SPDE over $\mathbb{R}^2$. For the SPDE in Equation (3.15) this corresponds to the Matérn case in Lindgren et al. (2011). To see what happens when A and B become small compared to the correlation range, it is useful to consider the spectra of the solution and the white noise. Since the SPDE in Equation (3.15) has homogeneous solutions, the spectral representation of homogeneous stochastic processes can be used. See Adler and Taylor (2007) for details about spectral representations for homogeneous processes.

Assume, for simplicity, that A and B are equal, say $A = B = T$. Since the solution and the white noise are Gaussian fields on the same box domain with periodic boundaries, they have discrete spectra with spectral mass at the same frequencies. The covariance function of the white noise is $\delta(\boldsymbol{s})$, which can be represented by its Fourier series as the double sum

$$\delta(\boldsymbol{s}) = \sum_{n,m=-\infty}^{\infty} \frac{1}{T^2} \exp(2\pi i(nx + my)/T).$$

From this expression one sees that the white noise on the torus has a spectral mass function

$$f_W(2\pi n/T, 2\pi m/T) = \begin{cases} \frac{1}{T^2}, & n, m \in \mathbb{Z}, \\ 0, & \text{otherwise.} \end{cases}$$

In addition, the SPDE in Equation (3.15) has a (spectral) transfer function

$$g(w_x, w_y) = \frac{1}{1 + w_x^2 + w_y^2}. \tag{3.16}$$

By combining the spectral mass function of the white noise and the above transfer function, one finds the spectral mass function of the solution,

$$f_s(2\pi n/T, 2\pi m/T) = \begin{cases} \left(\frac{T^2}{T^2+4\pi^2n^2+4\pi^2m^2}\right)^2 \cdot \frac{1}{T^2}, & n, m \in \mathbb{Z}, \\ 0, & \text{otherwise,} \end{cases}$$

or

$$f_s(2\pi n/T, 2\pi m/T) = \begin{cases} \frac{T^2}{(T^2+4\pi^2(n^2+m^2))^2}, & n, m \in \mathbb{Z}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.17}$$

In theory, the marginal variance could be calculated by taking the sum of the spectral mass function in Equation (3.17) over all $(n, m) \in \mathbb{Z}^2$, but in practice this is not easy to compute. But one can see that when T tends to zero $f_s(0, 0)$ behaves asymptotically as $1/T^2$ and that f_s goes to zero as some constant times T^2 for any other fixed (n, m). It is known that the sum of f_s over all integer pairs converges, thus the sum must behave asymptotically as $1/T^2$ when T tends to zero. Or in other words, the marginal variance must be asymptotically $1/T^2$ for small T. Thus one should not make T too small or the marginal variance will be completely dominated by the $(0, 0)$ frequency of the white noise. These calculations used $A = B$, but are possible to do also without this assumption. However, choosing first a good value for T and then taking A and B greater than this value should be sufficient.

Since it is not practical to use f_s to calculate the marginal variance, it is reasonable to try to approximate the marginal variance with the marginal variance

of the homogeneous solution of the corresponding SPDE on $\mathbb{R}^2$. Intuitively, one expects this to be a good approximation if A and B are large. However, what "large" means will most likely depend on κ^2 and $\mathbf{H}$. When κ^2 and $\mathbf{H}$ are constant, the corresponding marginal variance on $\mathbb{R}^2$ can be found from a simple formula.

Proposition 3.1.1. *Let u be a homogeneous solution of the SPDE*

$$\kappa^2 u(x,y) - \nabla \cdot \mathbf{H} \nabla u(x,y) = \mathcal{W}(x,y), \qquad (x,y) \in \mathbb{R}^2, \tag{3.18}$$

where $\mathcal{W}$ is a standard Gaussian white noise process, $\kappa^2 > 0$ is a constant, $\mathbf{H}$ is a positive definite 2×2 matrix and $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$.

Then u has marginal variance

$$\sigma_m^2 = \frac{1}{4\pi\kappa^2\sqrt{\det(\mathbf{H})}}.$$

Proof. Firstly, it is clear that rotating the coordinate system cannot change the marginal variances of the process. Secondly, from Section 2.2.2 it is clear that $\mathcal{W}$ is not changed by rotation since the measures of sets are not changed by rotations. Thus the coordinate axes can be chosen to be parallel to the eigenvectors of H. In addition, since the solution is homogeneous, Gaussian white noise is homogeneous and the SPDE has constant coefficients, the SPDE is acting as a linear filter. Thus one can use spectral theory to find the marginal variance.

In the rotated coordinate system in which the axes are parallel to the eigenvectors of $\mathbf{H}$, the SPDE becomes

$$\kappa^2 u(x',y') - \lambda_1 \frac{\partial^2 u}{\partial x'^2}(x',y') - \lambda_2 \frac{\partial^2 u}{\partial y'^2}(x',y') = \mathcal{W}(x',y'), \qquad (x',y') \in \mathbb{R}^2,$$

where λ_1 and λ_2 are the eigenvalues of $\mathbf{H}$. Section 3.1.4 gives details as to why this is true, but in the rotated coordinate system the effect of the operator $\nabla \cdot \mathbf{H} \nabla$ is simply to scale each of the components of the gradient and take the divergence of the scaled gradient.

The transfer function of the rotated SPDE is

$$g(w_1,w_2) = \frac{1}{\kappa^2 + \lambda_1 w_1^2 + \lambda_2 w_2^2}.$$

Further, the spectral density of a standard Gaussian white noise process on $\mathbb{R}^2$ is identically equal to $1/(2\pi)^2$. Thus the spectral density of the solution is

$$f_S(w_1,w_2) = \left(\frac{1}{2\pi}\right)^2 \frac{1}{(\kappa^2 + \lambda_1 w_1^2 + \lambda_2 w_2^2)^2}.$$

Table 3.2: The marginal variance, σ_a^2, of the approximate solution on a 100×100 grid of $[0, T]^2$ as a function of T. The corresponding marginal variance on $\mathbb{R}^2$ is 0.0796.

T	σ_a^2
0.25	16.0
0.50	4.00
1.00	1.00
2.00	0.263
4.00	0.0804
16.0	0.0809
100	0.0903
1000	0.00926

From the spectral density it is only a matter of integrating the density over $\mathbb{R}^2$,

$$\sigma_m^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_S(w_1, w_2)\, \mathrm{d}w_1\, \mathrm{d}w_2 = \frac{1}{4\pi\kappa^2\sqrt{\lambda_1\lambda_2}}.$$

Thus

$$\sigma_m^2 = \frac{1}{4\pi\kappa^2\sqrt{\det(\mathbf{H})}}.$$

□

From the proposition above one can calculate the marginal variance of the homogeneous solution of the SPDE in Equation (3.15) solved over $\mathbb{R}^2$. This gives the value 0.0796. To see how good this approximation is, it is compared with the values found by the GMRF approximation for different domain sizes. Let $A = B = T$ and consider a 100×100 grid for the GMRF approximation. Table 3.2 shows the marginal variances found for the GMRF approximation for different values of T. The table shows that for small T the marginal variance is higher than for the homogeneous solution of the SPDE on $\mathbb{R}^2$ and that the asymptotic behaviour $1/T^2$ for small T fits well with the values of $T \leq 1$ in the table. This increase in marginal variance is caused by increased boundary effects as the size of the domain decreases. For $T = 4$ to $T = 16$ the estimate is very good, for $T = 100$ the grid most likely becomes to coarse for the solution to be well approximated by the GMRF and for $T = 1000$ the grid is not good enough. By increasing the grid size to 200×200 and using $T = 100$, one finds a marginal variance of 0.0861 for the GMRF, which is closer, and one should expect to get even closer if the grid size is increased further.

The numerical example above uses a specific choice of values for κ^2 and $\mathbf{H}$. If the value of κ^2 or the value $\mathbf{H}$ is changed, the value of T required to avoid

boundary effects on the marginal variances may change. There may also be cases where one wants there to be significant correlations also for distances as large as A or B. In this case one can not expect the approximation from solving the SPDE over $\mathbb{R}^2$ to be any good. Further, when κ^2 and $\mathbf{H}$ are not constant, there is no approximation from solving the SPDE over $\mathbb{R}^2$, but one should take care in selecting the correlation ranges.

3.1.4 Interpretation of H

The function $\mathbf{H}$ in Equation (3.1) affects the diffusive operator $\nabla \cdot \mathbf{H}\nabla$. To see how the behaviour of that operator differs from the usual ∇^2 operator, it is useful to split it into the flux part $\mathsf{D}_j = \mathbf{H}\nabla$ and the divergence part $\nabla \cdot \mathsf{D}_j$. Let u be a function on some two-dimensional domain into $\mathbb{R}$ which is twice continuously differentiable and $\mathbf{H}$ a function which is continuously differentiable, then it is possible to calculate $\nabla \cdot \mathbf{H}\nabla u$.

Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a function on the Cartesian coordinate system (x, y), then the gradient of u at $\boldsymbol{s} \in \mathbb{R}^2$ is

$$\nabla u(\boldsymbol{s}) = \begin{bmatrix} \frac{\partial u}{\partial x}(\boldsymbol{s}) \\ \frac{\partial u}{\partial y}(\boldsymbol{s}) \end{bmatrix}. \tag{3.19}$$

This representation hides some of the true underlying meaning of the gradient. Namely, that the gradient at $\boldsymbol{s}$ is the linear transformation that takes a direction $\boldsymbol{v} \in \mathbb{R}^2$ and maps it to the directional derivative $\boldsymbol{v} \cdot \nabla u(\boldsymbol{s})$. The 2×1 matrix in Equation (3.19) is a representation of this linear transformation in the basis consisting of $\frac{\partial}{\partial x}$, change in the x-coordinate, and $\frac{\partial}{\partial y}$, change in y-coordinate. For the same function u in a rotated and scaled version of the coordinate system, the matrix which represents the linear transformation may be different, but it is the same linear transformation represented in a different basis. This is the key to understanding what applying $\mathbf{H}(\boldsymbol{s})$ to $\nabla u(\boldsymbol{s})$ means.

Since $\mathbf{H}$ is symmetric positive definite, $\mathbf{H}$ has a real spectral decomposition given by

$$\mathbf{H}(\boldsymbol{s}) = \mathbf{U}(\boldsymbol{s})\mathbf{\Lambda}(\boldsymbol{s})\mathbf{U}(\boldsymbol{s})^{\mathrm{T}},$$

where $\mathbf{\Lambda}(\boldsymbol{s}) = \operatorname{diag}(\lambda_1(\boldsymbol{s}), \lambda_2(\boldsymbol{s}))$ and

$$\mathbf{U}(\boldsymbol{s}) = \begin{bmatrix} \boldsymbol{e}_1(\boldsymbol{s}) & \boldsymbol{e}_2(\boldsymbol{s}) \end{bmatrix},$$

where $\boldsymbol{e}_1(\boldsymbol{s})$ and $\boldsymbol{e}_2(s)$ are the orthogonal normalized eigenvectors of $\mathbf{H}(\boldsymbol{s})$ and $\lambda_1(\boldsymbol{s})$ and $\lambda_2(\boldsymbol{s})$ are the corresponding eigenvalues. In what follows it is assumed that all these functions are evaluated at a fixed $\boldsymbol{s}$, so the simpler notation $\mathbf{H}$, $\mathbf{U}$, $\mathbf{\Lambda}$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, λ_1 and λ_2, where $\boldsymbol{s}$ is omitted, is used. Further, expressions like ∇u are also assumed to be evaluated at the fixed point.

One way to interpret $\mathsf{D}_j u$ is through

$$\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}\nabla u.$$

The rightmost part of this expression, $\mathbf{U}^{\mathrm{T}}\nabla u$, can be interpreted as the gradient of u in a different (Cartesian) coordinate system. ∇u is the gradient in the basis consisting of $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ and $\mathbf{U}^{\mathrm{T}}$ transforms the gradient into the the basis $\frac{\partial}{\partial \boldsymbol{e}_1}$ and $\frac{\partial}{\partial \boldsymbol{e}_2}$. This follows from the fact that $\mathbf{U}^{\mathrm{T}} : \mathbb{R}^2 \to \mathbb{R}^2$ maps a vector in (x, y)-coordinates into $(\boldsymbol{e}_1, \boldsymbol{e}_2)$-coordinates, which also means that $\mathbf{U}^{\mathrm{T}}$ transform ∇u from $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ coordinates into $(\frac{\partial}{\partial \boldsymbol{e}_1}, \frac{\partial}{\partial \boldsymbol{e}_2})$ coordinates.

The next part of the $\mathsf{D_j}$ operator is to scale the component of $\mathbf{U}^{\mathrm{T}}\nabla u$ in direction $\frac{\partial}{\partial \boldsymbol{e}_1}$ by λ_1 and the component in direction $\frac{\partial}{\partial \boldsymbol{e}_2}$ by λ_2. Then the $\mathbf{U}$ factor transforms the scaled gradient represented by derivatives in the directions $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ to the same gradient represented by derivatives in the directions x and y. In other words a function represented in the basis $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$.

Thus the interpretation of $\mathsf{D}_j u$ for a fixed spatial point is similar to the interpretation of the usual gradient of a function at a fixed point. It no longer takes a direction $\boldsymbol{v} \in \mathbb{R}^2$ to the directional derivative, but to a transformed version of the directional derivative. The $\mathbf{H}$ matrix is defined relative to some coordinate system, and if one changes the coordinate system to the orthogonal normalized eigenvectors of $\mathbf{H}$, $\mathbf{H}$ becomes the diagonal matrix $\mathrm{diag}(\lambda_1, \lambda_2)$. Thus it only scales the components of the gradient separately when expressed in this coordinate system.

This fact can be used to understand why one requires $\mathbf{H}(\boldsymbol{s})$ to be positive definite for all $\boldsymbol{s}$ in the domain. A negative eigenvalue would give something quite different from what is called diffusion. It would amplify differences in the first order derivative. To see this effect it is necessary to include a time derivative. Consider the partial differential equation

$$\frac{\partial}{\partial t}u = -\frac{\partial^2}{\partial x^2}u,$$

which corresponds to a one dimensional spatial domain and a 1×1 matrix $\mathbf{H}$ with a negative eigenvalue, and

$$\frac{\partial}{\partial t}u = \frac{\partial^2}{\partial x^2}u,$$

which corresponds to a positive eigenvalue. The first differential equation tends to amplify differences in the spatial first order derivative, whereas the latter tends to equalize differences in the spatial first order derivative. Only the latter can be considered a diffusion. This generalizes to higher dimensions.

The divergence of a vector field at a given point can be interpreted as the volume density of the net outward flow at that point. If (x, y) is a Cartesian

coordinate system and $\boldsymbol{F} = (F_1, F_2) : \mathbb{R}^2 \to \mathbb{R}^2$ is a differentiable function expressed in that coordinate system, the divergence of $\boldsymbol{F}$ at $\boldsymbol{s} \in \mathbb{R}^2$ is simply

$$\nabla \cdot \boldsymbol{F}(\boldsymbol{s}) = \frac{\partial}{\partial x} F_1(\boldsymbol{s}) + \frac{\partial}{\partial y} F_2(\boldsymbol{s}).$$

Loosely speaking, one can say that the divergence of $\boldsymbol{F}$ at a fixed point is the net outward flow from that point according to the vector field $\boldsymbol{F}$.

From this point on, the spatial point is no longer considered fixed. The last part of the $\nabla \cdot \mathsf{D}_j u$ expression is to take the divergence expressed in the xy-coordinate system of the flux $\mathsf{D}_j u$ also expressed in the xy-coordinate system. This involves the local net outward flux associated with $\mathsf{D}_j u$ at the different points, that is a measure of how much the transformed gradient is changing at a given point.

From the spectral representation of $\mathbf{H}$ it is possible to construct operators which scale the the derivative in a chosen direction with a higher factor than the other directions. This penalizes changes in the derivative more in one direction than another. Choose a normalized vector $\boldsymbol{v}$ and construct

$$\mathbf{H} = \gamma \mathbf{I}_2 + \beta \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}},$$

where $\gamma > 0$ and $\beta \geq 0$ are constants. This matrix results in a scaling by $\gamma + \beta$ of the derivative in the directions parallel to $\boldsymbol{v}$ and a scaling of γ in the directions orthogonal to $\boldsymbol{v}$. In the context of the SPDE in Equation (3.1) this means that the solution is more regular in direction $\boldsymbol{v}$ than in the direction orthogonal to $\boldsymbol{v}$. This means that it can be used to give an anisotropic covariance function.

For a general $\mathbf{H}$, which is allowed to vary with $\boldsymbol{s}$, it is not enough to only consider the local properties as above, but the same type of effects should be present. But there are effects both from the local spectral decomposition and from the changing scaling factors and directions in the spectral decomposition as a function of spatial position.

3.1.5 Interpretation of κ^2

By assumption, κ^2 is a strictly positive function, so Equation (3.1) can be written as

$$u(\boldsymbol{s}) - \frac{1}{\kappa^2(\boldsymbol{s})} \nabla \cdot H(\boldsymbol{s}) \nabla u(\boldsymbol{s}) = \frac{1}{\kappa^2(\boldsymbol{s})} \mathcal{W}(\boldsymbol{s}). \tag{3.20}$$

From this one can see that the effect of increasing κ^2 is both to reduce the coefficient of the diffusive term relative to the u term and to decrease the variance of the driving Gaussian white noise. Only reducing the coefficient of the diffusive term relative to the u term, would give more irregular solutions and higher

marginal variances. Only reducing the coefficient of the Gaussian white noise process relative to the other coefficients, would give some sort of scaling of the solution. Multiplying the right hand side with a constant C simply implies that the solutions are scaled by C as well. For a non-constant factor things become more involved.

However, the focus of the thesis is on the function $\mathbf{H}$ and only constant functions are used for κ^2.

3.2 Numerical examples

This section has examples of some of the types of behaviour which can be achieved by different $\mathbf{H}$. The function κ^2 is constant in each of the examples. The section consists of three parts. The parts are homogeneous isotropic GMRFs, homogeneous anisotropic GMRFs and inhomogeneous GMRFs.

3.2.1 Homogeneous isotropic GMRFs

As discussed in Section 3.1.3, there is an increase in marginal variances for small A and B. The domain is chosen large enough to avoid this. For the purpose of this section, A and B in Equation (3.1) are both chosen to be equal to 20.

The construction of a homogeneous isotropic GMRF requires both that κ^2 and $\mathbf{H}$ are constant in the domain and that $\mathbf{H}$ is equal to a constant times the identity matrix. This means that one only needs to control the parameters C in $\kappa^2(\boldsymbol{s}) \equiv C$ and γ in $\mathbf{H}(\boldsymbol{s}) \equiv \gamma \mathbf{I}_2$. This type of GMRF does not have very exciting behaviour, but could be useful as a null hypothesis when checking whether there should be inhomogeneity or anisotropy. They also serve well as first examples of the type of behaviour that can be achieved from the SPDE.

Note that if one is only interested in the solution up to a scalar factor, one only needs one parameter. Reparametrize to $\sigma = 1/C$ and $\hat{\gamma} = \gamma/C$. Then σ is only a scale factor for the solution and all homogeneous isotropic solutions can be found by solving

$$\hat{u}(\boldsymbol{s}) - \hat{\gamma}\left(\frac{\partial^2}{\partial x^2}\hat{u}(\boldsymbol{s}) + \frac{\partial^2}{\partial y^2}\hat{u}(\boldsymbol{s})\right) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in [0, A] \times [0, B], \tag{3.21}$$

where $\mathcal{W}$ is a standard Gaussian white noise process, and setting $u = \sigma\hat{u}$. Therefore, this is the SPDE that is considered in the following example.

Example 3.2.1 (Homogeneous isotropic GMRF)**.** The example uses the SPDE in Equation (3.21) with domain $[0, 20]^2$ and periodic boundary conditions. A 200×200 regular grid is used with the precision matrix found in Section 3.1.2.

Figure 3.3 shows one simulation with $\hat{\gamma} = 1$ and one simulation with $\hat{\gamma} = 5$. For $\hat{\gamma} = 1$ the marginal variance of the GMRF is 0.0802 and for $\hat{\gamma} = 5$ the marginal variance of the GMRF is 0.0160. The formula in Proposition 3.1.1 gives the values 0.0796 and 0.0159 for the same values of $\hat{\gamma}$, so the formula gives very good estimates in these cases.

Comparing Figure 3.3(a) and Figure 3.3(b) one sees that the behaviour for $\hat{\gamma} = 5$ is more regular than the behaviour for $\hat{\gamma} = 1$. Comparing with the corresponding PDE without the white noise, this is what one should expect since large values of $\hat{\gamma}$ penalize large values of the second order derivatives. One should expect that the correlations increase when $\hat{\gamma}$ is increased.

This is in fact what happens. Figure 3.4 shows the covariances for the variable at $(9.95, 9.95)$ with every other point in the grid. Since the marginal variances are the same everywhere, the covariances can be scaled with the marginal variance so that they are correlations. This means in practice that the same colours in Figure 3.4(a) and Figure 3.4(b) are given the same values. Under that assumption one can see that $\hat{\gamma} = 5$ gives a much larger area with high correlation than $\hat{\gamma} = 1$.

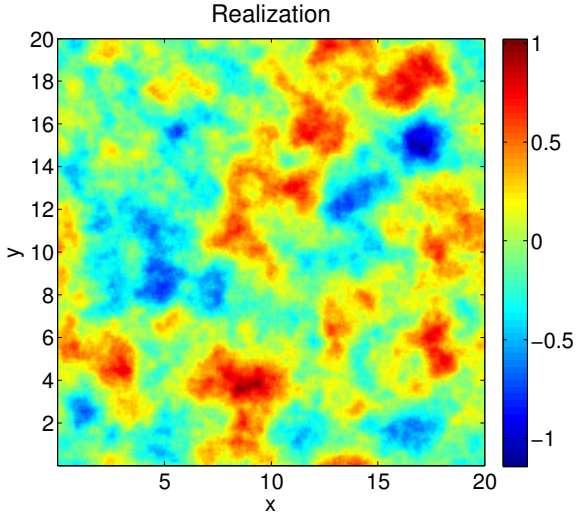
In addition, the covariances appear to be quite isotropic for both $\hat{\gamma} = 1$ and $\hat{\gamma} = 5$. The use of the point $(9.95, 9.95)$ for the calculation of the covariances is not of importance, because of the homogeneity all points have the same covariance function. Note that the isotropy gives level curves which are circles around the point.

This example demonstrates that changing $\hat{\gamma}$ changes the correlation function of the field. In addition one can use the σ parameter to get the desired marginal variances. In this example the correlations were close to 0 for distances greater than or equal to 10, and the formula for the marginal variances in Proposition 3.1.1 gave good estimates for the marginal variances.

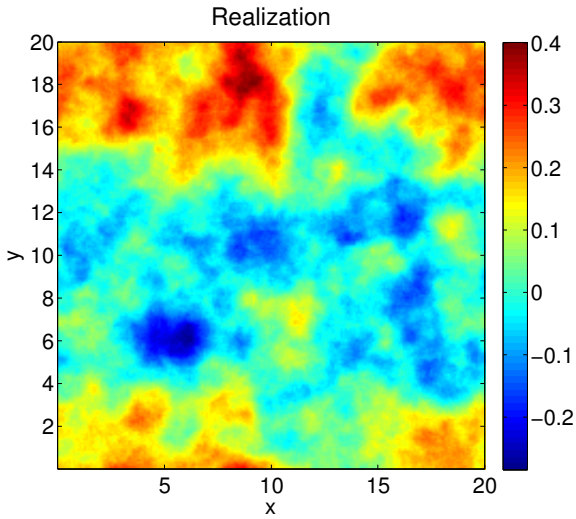
3.2.2 Homogeneous anisotropic GMRFs

To give the SPDE in Equation (3.1) homogeneous solutions, one should use a constant κ^2 and a constant $\mathbf{H}$. With constant κ^2 and $\mathbf{H}$ one has anisotropy if and only if the eigenvalues of $\mathbf{H}$ are different. If the SPDE were solved over $\mathbb{R}^2$, a non-diagonal $\mathbf{H}$ would correspond to a rotation of the domain. Therefore, all homogeneous cases could be reduced to a constant κ^2 and a constant diagonal $\mathbf{H}$.

For the case of a rectangular domain with periodic boundary conditions the same is not true. However, if the correlation is small at distances of 10 and greater, the effect of the bounded domain is small and one only needs to consider a diagonal $\mathbf{H}$ and rotate the resulting solution. In that case one can use the parameter C in $\kappa^2 \equiv C$ and parameters H^{11} and H^{22} in $\mathbf{H} \equiv \text{diag}(H^{11}, H^{22})$ to describe the homogeneous anisotropic GMRFs up to rotation.

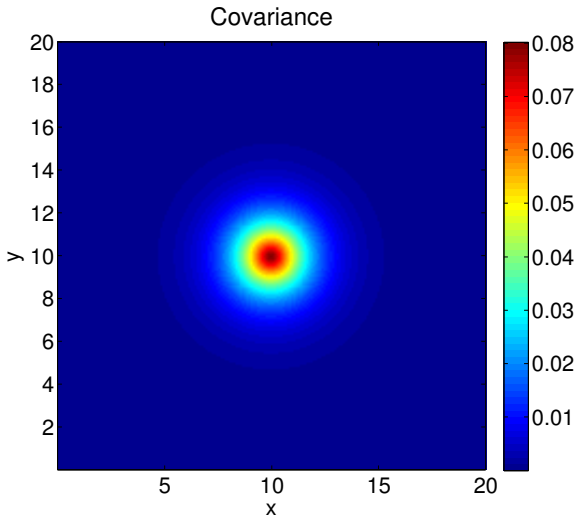


(a) $\hat{\gamma} = 1$.

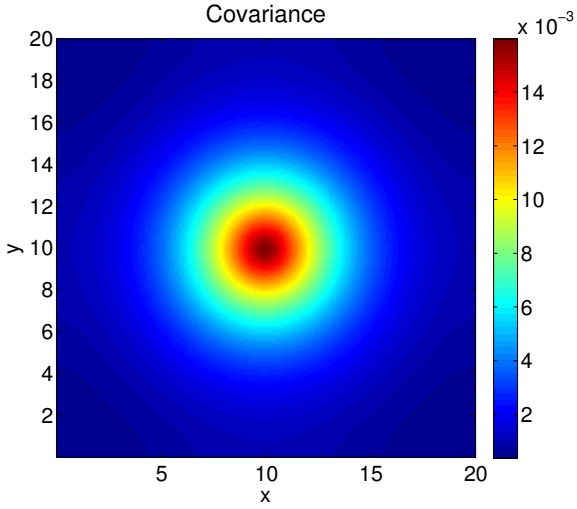


(b) $\hat{\gamma} = 5$.

Figure 3.3: Realizations from the SPDE in Equation (3.21) with a 200×200 grid of $[0, 20]^2$ and periodic boundary conditions for two different values of $\hat{\gamma}$.



(a) $\hat{\gamma} = 1$.



(b) $\hat{\gamma} = 5$.

Figure 3.4: Covariances of $(9.95, 9.95)$ with every other point for the solution of the SPDE in Equation (3.21) with a 200×200 grid of $[0, 20]^2$ and periodic boundary conditions for two different values of $\hat{\gamma}$.

In the same manner as in the previous section, reparametrize as $\gamma_1 = H^{11}/C$, $\gamma_2 = H^{22}/C$ and $\sigma = 1/C$ to get the SPDE

$$u(\boldsymbol{s}) - \gamma_1 \frac{\partial^2}{\partial x^2} u(\boldsymbol{s}) - \gamma_2 \frac{\partial^2}{\partial y^2} u(\boldsymbol{s}) = \sigma \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in [0, A] \times [0, B], \tag{3.22}$$

where $\mathcal{W}$ is a standard Gaussian white noise process, with periodic boundary conditions. Then σ is just a scale parameter and one can solve the SPDE with $\sigma = 1$ to get $\hat{u}$, and set $u = \hat{\sigma}\hat{u}$ to find a solution for $\sigma = \hat{\sigma}$.

If one does not assume that $\mathbf{H}$ is diagonal, one has an additional parameter γ_{12} corresponding to the off-diagonal entry in $\mathbf{H}$ divided by C. This gives the SPDE

$$u(\boldsymbol{s}) - \left[\gamma_1 \frac{\partial^2}{\partial x^2} + 2\gamma_{12} \frac{\partial^2}{\partial x \partial y} + \gamma_2 \frac{\partial^2}{\partial y^2}\right] u(\boldsymbol{s}) = \sigma \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in [0, A] \times [0, B], \tag{3.23}$$

where $\mathcal{W}$ is a standard Gaussian white noise process, with periodic boundary conditions. Comparing Equation (3.22) with Equation (3.23) one sees that the only difference is the cross derivative with the $-2\gamma_{12}$ coefficient.

The SPDE in Equation (3.22) and the SPDE in Equation (3.21) are similar, and with $\gamma_1 = \gamma_2 = \hat{\gamma}$ (and $\sigma = 1$) they are equal. This means that the only difference is the introduction of the possibility of different coefficients for the second order partial derivatives. This gives the ability to introduce different regularity in different directions. In the following example both a diagonal $\mathbf{H}$ and a non-diagonal $\mathbf{H}$ is considered.

Example 3.2.2 (Homogeneous anisotropic GMRF)**.** The purpose of this example is to consider the effect of using a $\mathbf{H}$ with different eigenvalues. Both with the SPDE in Equation (3.22) and the SPDE in Equation (3.23). The domain $[0, 20]^2$ is used with periodic boundary conditions and is discretized by a regular 200×200 grid. The case with $\mathbf{H}$ having equal eigenvalues is done in Example 3.2.1 and to get comparable results with that example one of the eigenvalues of $\mathbf{H}$ is chosen equal to 1.

Figure 3.5(a) shows one realization of the solution of the SPDE in Equation (3.22) with $\sigma = 1$, $\gamma_1 = 1$ and $\gamma_2 = 3$. This means that the eigenvalues of $\mathbf{H}$ are 1 and 3. Figure 3.5(b) shows one realization of the solution of the SPDE in Equation (3.23) with $\sigma = 1$, $\gamma_1 = 5$, $\gamma_{12} = 4$ and $\gamma_2 = 5$. This corresponds to

$$\mathbf{H} = \mathbf{I}_2 + 8\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}},$$

where $\boldsymbol{v} = (1, 1)/\sqrt{2}$, and gives the eigenvalue 9 for direction $\boldsymbol{v}$ and the eigenvalue 1 for the direction $(1, -1)/\sqrt{2}$.

Figure 3.5 shows that both cases give a realization which looks anisotropic. For the first case the realization looks more regular in the y-direction than the x-direction and for the latter case the realization looks most regular in the direction

given by $\boldsymbol{v}$ and least regular in the direction orthogonal to $\boldsymbol{v}$. The difference in regularity seems much greater for the case with the non-diagonal matrix than for the case with the diagonal matrix. This should be expected since the difference between the eigenvalues is larger for this case.

To demonstrate that the solution of the SPDE is anisotropic for each of the cases, the covariances between a fixed point and all others are calculated for each case. Note that this is sufficient to describe all the covariances since the solutions are homogeneous. Figure 3.6 shows the covariances between the variable at $(9.95, 9.95)$ and all other points in the grid for both cases. One can immediately note that the non-diagonal $\mathbf{H}$ gives what appears to be elliptic constant level curves with semi-axes along $\boldsymbol{v}$ and the direction orthogonal to $\boldsymbol{v}$, whereas the diagonal $\mathbf{H}$ gives constant level curves that appears to be ellipses with semi-axes along the x-axis and the y-axis. It is clear from the figure that the solutions are not isotropic.

Since the solutions are homogeneous, one can consider Figure 3.6(a) and Figure 3.6(b) to be rescaled versions of the correlation functions. From this one can see that the correlation decreases most slowly and most quickly along the directions used to specify $\mathbf{H}$, with slowest decrease along the direction with the highest parameter value. For the diagonal $\mathbf{H}$ the correlation decreases most slowly in the y-direction and decreases most quickly in the x-direction. Similarly, for the non-diagonal $\mathbf{H}$ the correlation decreases slowest in the direction given by $\boldsymbol{v}$ and most rapidly in the direction orthogonal to $\boldsymbol{v}$.

As previously mentioned, the constant level curves seem to be ellipses for both cases. It is interesting to note that if one compares the two correlation functions found in this example with the one in Figure 3.4(a), one can observe that they all seem to have the same length for the minor-axis of the constant level curves. Further, increasing the largest eigenvalue of $\mathbf{H}$ gives larger semi-axis along the chosen direction, and compared with Figure 3.4(b) increasing both eigenvalues increases both semi-axes.

This example indicates that the constant level curves of the correlation function are approximately ellipses. It can be shown that this is the case for a homogeneous anisotropic solution of the corresponding SPDE on $\mathbb{R}^2$ by doing a change of variables from the standard equation with $\gamma_1 = 1$ and $\gamma_2 = 1$. No further investigation of this is done.

From this section one can see that compared to the previous section, which only uses one parameter $\hat{\gamma}$ to control the correlation, the introduction of two parameters γ_1 and γ_2 allows for the creation of GMRFs which are more regular in one direction than another. One can use the parameters γ_1, γ_{12} and γ_2 to decide the correlation function and σ to get the desired marginal variance.

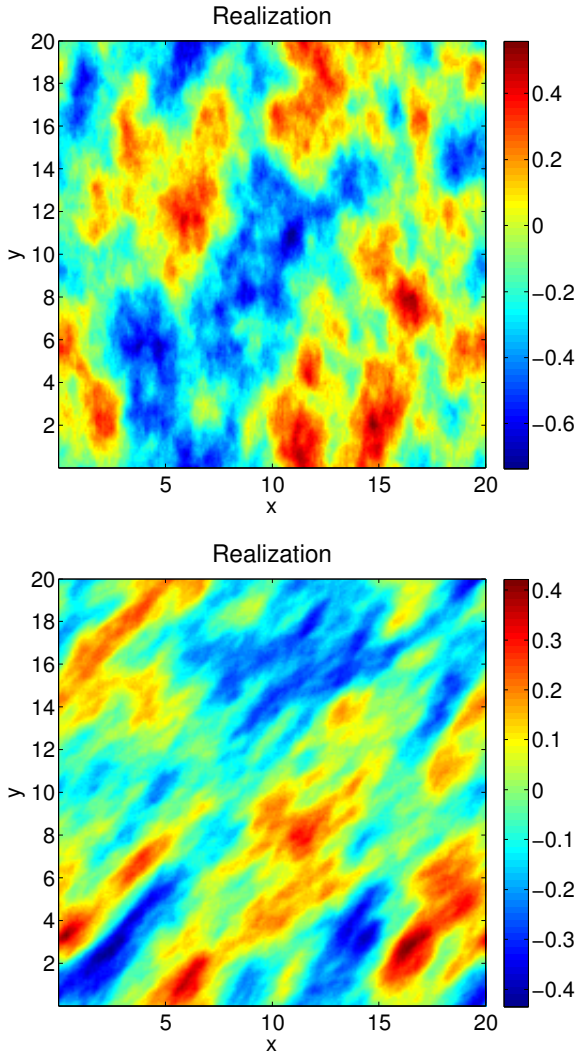


Figure 3.5: (a) Realization from the SPDE in Equation (3.22) on $[0, 20]^2$ with a 200×200 grid and periodic boundary conditions with $\sigma = 1$, $\gamma_1 = 1$ and $\gamma_2 = 3$. (b) Realization from the SPDE in Equation (3.23) on $[0, 20]^2$ with a 200×200 grid and periodic boundary conditions with $\sigma = 1$, $\gamma_1 = 5$, $\gamma_{12} = 4$ and $\gamma_2 = 5$.

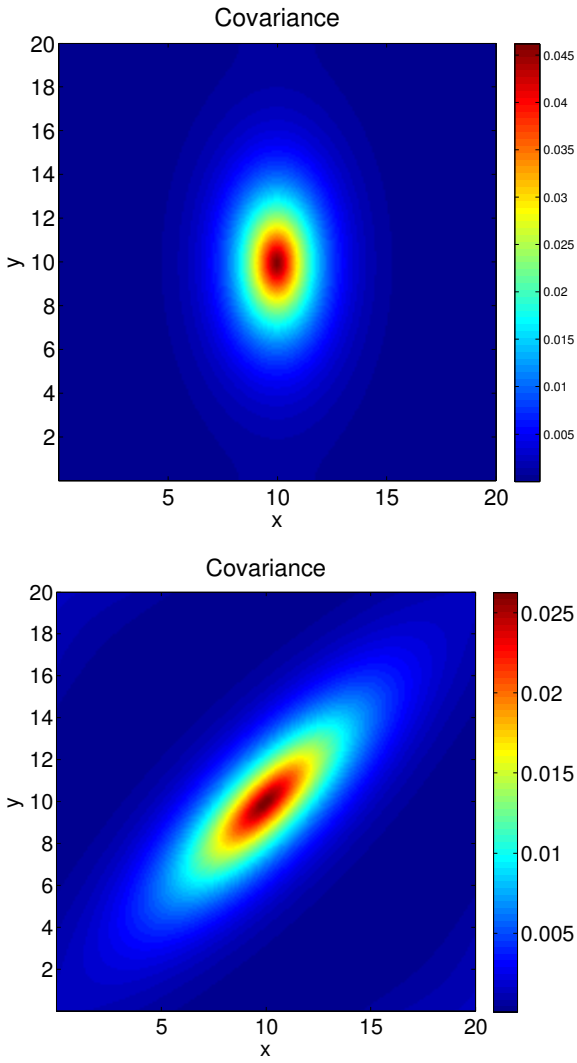


Figure 3.6: (a) Covariances of the centre with all other points for the solution of the SPDE in Equation (3.22) on $[0, 20]^2$ with a 200×200 grid and periodic boundary conditions with $\sigma = 1$, $\gamma_1 = 1$ and $\gamma_2 = 3$. (b) Covariances of the centre with all other points for the SPDE in Equation (3.23) on $[0, 20]^2$ with a 200×200 grid and periodic boundary conditions with $\sigma = 1$, $\gamma_1 = 5$, $\gamma_{12} = 4$ and $\gamma_2 = 5$.

3.2.3 Inhomogeneous GMRFs

The two previous sections only have homogeneous GMRFs. To make the solutions of the SPDE in Equation (3.1) inhomogeneous, either κ^2 or $\mathbf{H}$ has to be a non-constant function. One way to achieve inhomogeneity is by letting $\mathbf{H}$ be non-constant through

$$\mathbf{H}(\boldsymbol{s}) = \gamma \mathbf{I}_2 + \beta \boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}},$$

where $\boldsymbol{v}$ is a vector field on $[0, A] \times [0, B]$ which satisfy the periodic boundary conditions and $\gamma > 0$ and $\beta > 0$ are constants.

Example 3.2.3 (Inhomogeneous GMRF)**.** Use the domain $[0, 20]^2$ with a 200×200 grid and periodic boundary conditions for the SPDE in Equation (3.1). Let κ^2 be identically equal to 1 and let $\mathbf{H}$ be given as

$$\mathbf{H}(\boldsymbol{s}) = \gamma \mathbf{I}_2 + \beta \boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}},$$

where $\boldsymbol{v}$ is a 2-dimensional vector field on $[0, 20]^2$ which satisfies the periodic boundary conditions and $\gamma > 0$ and $\beta > 0$ are constants.

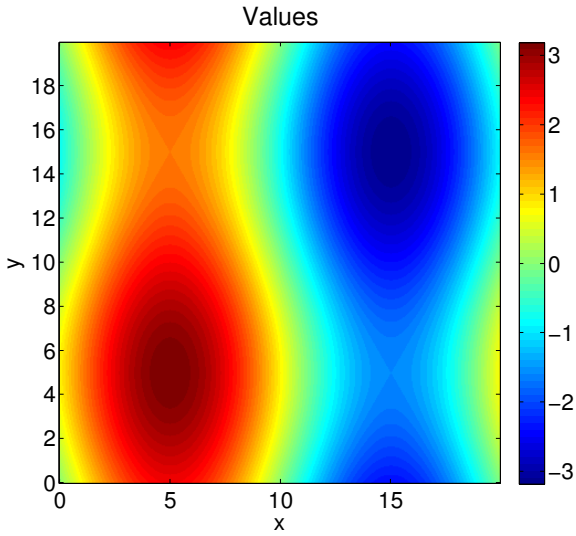
To create an interesting vector field, start with the function $f : [0, 20]^2 \to \mathbb{R}$ defined by

$$f(x, y) = \left(\frac{10}{\pi}\right)\left(\frac{3}{4}\sin(2\pi x/20) + \frac{1}{4}\sin(2\pi y/20)\right).$$

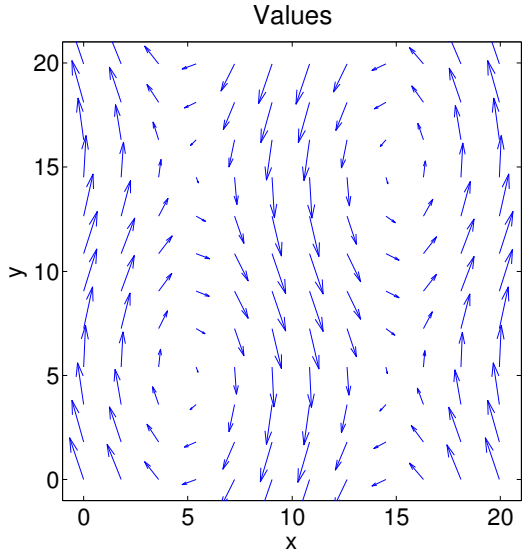
Then calculate the gradient ∇f and let $\boldsymbol{v} : [0, 20]^2 \to \mathbb{R}^2$ be the gradient rotated $90\,^\circ$ counter-clockwise at each point. Figure 3.7(a) shows the values of the function f and Figure 3.7(b) shows the resulting vector field $\boldsymbol{v}$. The vector field is calculated on a 400×400 regular grid, because the values between neighbouring cells of the discretization is needed.

Figure 3.8(a) shows one realization from the resulting GMRF with $\gamma = 0.1$ and $\beta = 25$. A much higher value for β than γ is chosen to illustrate the connection between the vector field and the resulting covariance structure. From the realization it is clear that there is stronger dependence along the directions of the vector field shown in Figure 3.7(b) at each point than in the other directions. In addition, from Figure 3.8(b) it seems that positions with large values for the norm of the vector field has smaller marginal variance than positions with small values and vice versa.

From Figure 3.9 one can see that the covariances depend on the direction and norm of the vector field, and that there is clearly inhomogeneity. Figure 3.9(a) shows the covariances of the variable at position $(9.95, 9.95)$ with all other points in the grid and Figure 3.9(b) shows the covariance of the variable at $(4.95, 2.05)$ with all other points in the grid. At both of these points there is much higher covariance along the vector field than in the direction orthogonal to the vector

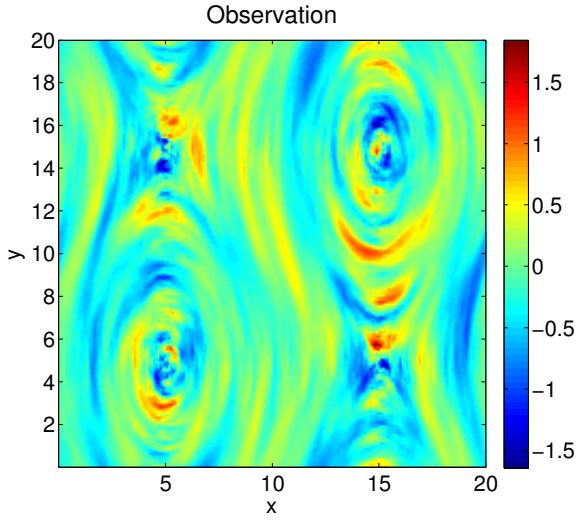


(a) The function used to create the vector field.

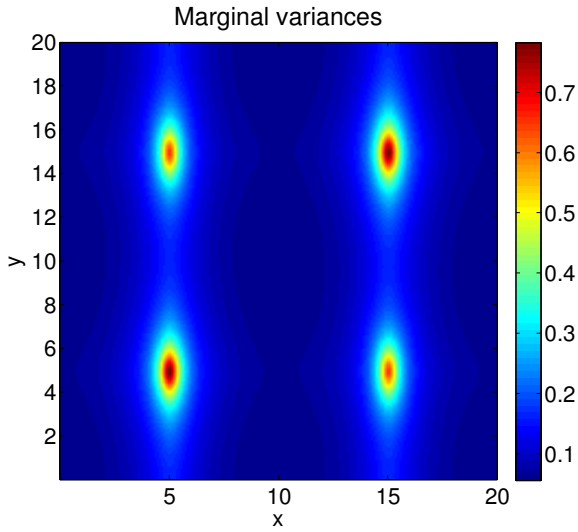


(b) The resulting vector field.

Figure 3.7: The gradient of the function illustrated in (a) is calculated and rotated 90 ° counter-clockwise at each point to give the vector field illustrated in (b).



(a) One realization.

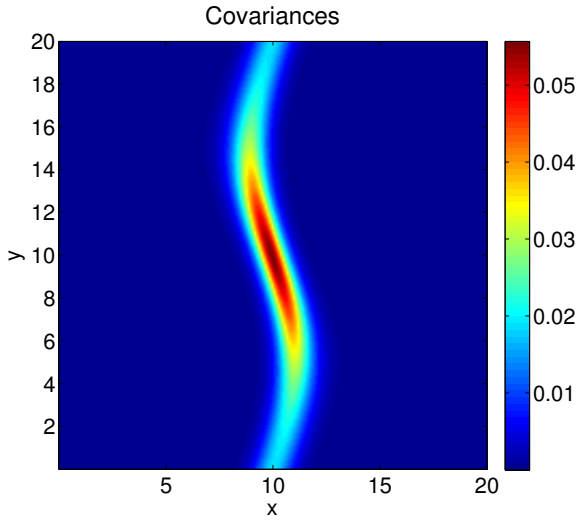


(b) Marginal variances.

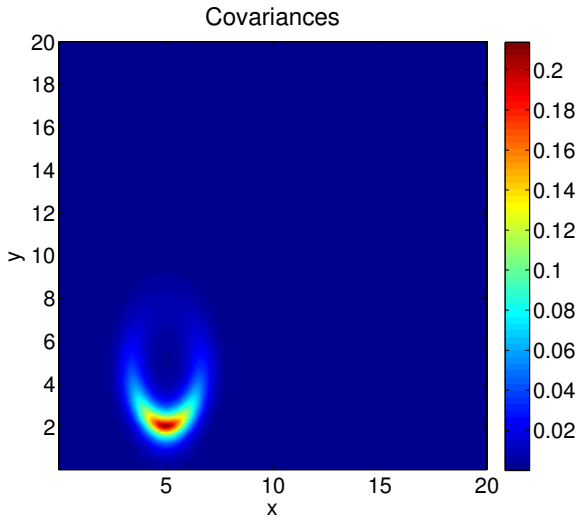
Figure 3.8: One observation and the marginal variances of the solution of the SPDE in Equation (3.1) on a 200×200 regular grid of $[0, 20]^2$ with periodic boundary conditions, $\kappa^2 \equiv 1$ and $\mathbf{H} = 0.1\mathbf{I}_2 + 25\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$, where $\boldsymbol{v}$ is the vector field described in Example 3.2.3.

field. In Figure 3.9(a) one can see that the covariance follows the curving of the vector field and in Figure 3.9(b) one can see that the high covariance values extend around the point $(4.95, 4.95)$.

From this example one can see that allowing $\mathbf{H}$ to be non-constant means that one can vary the dependence structure in more interesting ways than the homogeneous anisotropic fields. Secondly, using a vector field to control how $\mathbf{H}$ varies means that the resulting correlation structure can be partially visualized from the vector field. Thirdly, when $\gamma > 0$ this construction guarantees that $\mathbf{H}$ is everywhere positive definite.



(a) Covariances with position $(9.95, 9.95)$.



(b) Covariances with position $(4.95, 2.05)$.

Figure 3.9: Covariances of two different points with all other points in the grid for the solution of the SPDE in Equation (3.1) on a 200×200 regular grid of $[0, 20]^2$ with periodic boundary conditions, $\kappa^2 \equiv 1$ and $\mathbf{H} = 0.1\mathbf{I}_2 + 25\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$, where $\boldsymbol{v}$ is the vector field in described in Example 3.2.3.

Chapter 4

Inference with the GMRF approximation

In the previous chapter a GMRF approximation is derived for the solution of the SPDE in Equation (3.1). The goal of this chapter is to do inference on the coefficients in the SPDE in a Bayesian setting by using the derived GMRF.

4.1 Inference scheme

To do the inference on the coefficients in the SPDE in Equation (3.1), parameters are introduced in both κ^2 and $\mathbf{H}$. Let κ^2 depend on parameters $\boldsymbol{\theta}_1$ and $\mathbf{H}$ depend on parameters $\boldsymbol{\theta}_2$, and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}})$. In addition, give $\boldsymbol{\theta}$ a prior distribution $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

Then for a given value of $\boldsymbol{\theta}$, use the discretization in Section 3.1.2 to find the GMRF $\boldsymbol{u}|\boldsymbol{\theta} \sim \mathcal{N}(0, \mathbf{Q}(\boldsymbol{\theta})^{-1})$. Combine this with the prior of $\boldsymbol{\theta}$ to find the joint distribution of the parameters and $\boldsymbol{u}$.

Lastly, introduce a third step $\boldsymbol{y}|\boldsymbol{u} \sim \pi(\boldsymbol{y}|\boldsymbol{u})$ which specifies how $\boldsymbol{y}$ is observed from the underlying GMRF. The model for how $\boldsymbol{y}$ is related to $\boldsymbol{u}$ is chosen to be particularly simple, namely that linear combinations of $\boldsymbol{u}$ are observed with Gaussian noise,

$$\boldsymbol{y}|\boldsymbol{u} \sim \mathcal{N}(\mathbf{A}\boldsymbol{u}, \mathbf{Q}_N^{-1}).$$

This is a three level hierarchical model, where the goal is to do inference on $\boldsymbol{\theta}$ based on an observation of $\boldsymbol{y}$. To complete the scheme the posterior distribution of $\boldsymbol{\theta}$ given an observation of $\boldsymbol{y}$ has to be calculated.

4.1.1 Posterior distribution

Assume that N rectangles are used in the discretization in Section 3.1.2. Then $\boldsymbol{u}|\boldsymbol{\theta}$ is an N-dimensional multivariate Gaussian distribution with probability density

$$\pi(\boldsymbol{u}|\boldsymbol{\theta}) = \left(\frac{1}{2\pi}\right)^{N/2} |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}(\boldsymbol{\theta})\boldsymbol{u}\right),$$

where $\mathbf{Q}(\boldsymbol{\theta})$ is the precision matrix at the end of Section 3.1.2 for parameter value $\boldsymbol{\theta}$. Further, $\boldsymbol{y}$ is a k-dimensional random variable defined through

$$\boldsymbol{y}|\boldsymbol{u} \sim \mathcal{N}(\mathbf{A}\boldsymbol{u}, \mathbf{Q}_N^{-1}),$$

where $\mathbf{A}$ is a $k \times N$ matrix and $\mathbf{Q}_N$ is a $k \times k$ positive definite matrix. The probability density of $\boldsymbol{y}|\boldsymbol{u}$ is

$$\pi(\boldsymbol{y}|\boldsymbol{u}) = \left(\frac{1}{2\pi}\right)^{k/2} |\mathbf{Q}_N|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\mathbf{A}\boldsymbol{u})^{\mathrm{T}}\mathbf{Q}_N(\boldsymbol{y}-\mathbf{A}\boldsymbol{u})\right).$$

Note that since the density of $\boldsymbol{y}|\boldsymbol{u}$ is independent of $\boldsymbol{\theta}$, the probability density of $\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\theta}$ is equal to the probability density of $\boldsymbol{y}|\boldsymbol{u}$.

It is useful to first calculate the probability density of $\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}$. It can be found from the probability densities above,

$$\begin{aligned}
\pi(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}) &\propto \pi(\boldsymbol{u},\boldsymbol{y}|\boldsymbol{\theta}) \\
&= \pi(\boldsymbol{u}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\theta}) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}(\boldsymbol{\theta})\boldsymbol{u} + (\boldsymbol{y}-\mathbf{A}\boldsymbol{u})^{\mathrm{T}}\mathbf{Q}_N(\boldsymbol{y}-\mathbf{A}\boldsymbol{u})\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}(\mathbf{Q}(\boldsymbol{\theta}) + \mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\mathbf{A})\boldsymbol{u} - 2\boldsymbol{u}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\boldsymbol{y}\right]\right).
\end{aligned}$$

Let $\mathbf{Q}_C(\boldsymbol{\theta}) = \mathbf{Q}(\boldsymbol{\theta}) + \mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\mathbf{A}$ and $\boldsymbol{\mu}_C(\boldsymbol{\theta}) = \mathbf{Q}_C(\boldsymbol{\theta})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\boldsymbol{y}$, then

$$\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{C}}(\boldsymbol{\theta}), \mathbf{Q}_C(\boldsymbol{\theta})^{-1}).$$

This is an N-dimensional multivariate Gaussian distribution.

The probability density of $\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}$ can be used to integrate out $\boldsymbol{u}$ from the joint density of $\boldsymbol{u}$, $\boldsymbol{y}$ and $\boldsymbol{\theta}$ by calculating

$$\begin{aligned}
\pi(\boldsymbol{\theta},\boldsymbol{y}) &= \frac{\pi(\boldsymbol{\theta},\boldsymbol{u},\boldsymbol{y})}{\pi(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta})} \\
&= \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{u}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\theta})}{\pi(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta})}.
\end{aligned}$$

This expression gives

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \\ \pi(\boldsymbol{\theta})\frac{|\mathbf{Q}(\boldsymbol{\theta})|^{1/2}|\mathbf{Q}_N|^{1/2}}{|\mathbf{Q}_C(\boldsymbol{\theta})|^{1/2}}\exp\left(-\frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}(\boldsymbol{\theta})\boldsymbol{u}\right]\right)\times \\ \times\exp\left(-\frac{1}{2}\left[(\boldsymbol{y}-\mathbf{A}\boldsymbol{u})^{\mathrm{T}}\mathbf{Q}_N(\boldsymbol{y}-\mathbf{A}\boldsymbol{u})\right]\right)\times \\ \times\exp\left(+\frac{1}{2}\left[(\boldsymbol{u}-\boldsymbol{\mu}_C(\boldsymbol{\theta}))^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{\theta})(\boldsymbol{u}-\boldsymbol{\mu}_C(\boldsymbol{\theta}))\right]\right).\end{aligned}$$

Observe that the quadratic terms in the exponential functions can be greatly simplified. Write out the two first quadratic terms to find

$$-\frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}(\boldsymbol{\theta})\boldsymbol{u}+\boldsymbol{y}^{\mathrm{T}}\mathbf{Q}_N\boldsymbol{y}-2\boldsymbol{u}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\boldsymbol{y}+\boldsymbol{u}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\mathbf{A}\boldsymbol{u}\right].$$

But $\mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\boldsymbol{y}=\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{\mu}_C(\boldsymbol{\theta})$ and $\mathbf{Q}(\boldsymbol{\theta})+\mathbf{A}^{\mathrm{T}}\mathbf{Q}_N\mathbf{A}=\mathbf{Q}_C(\boldsymbol{\theta})$, so the terms the terms can be written as

$$-\frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{u}+\boldsymbol{y}^{\mathrm{T}}\mathbf{Q}_N\boldsymbol{y}-2\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{\mu}_C(\boldsymbol{\theta})\right].$$

Further, the quadratic term in the third exponential function can be written as

$$-\frac{1}{2}\left[-\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{u}+2\boldsymbol{u}^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{\mu}_C(\boldsymbol{\theta})-\boldsymbol{\mu}_C(\boldsymbol{\theta})\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{\mu}_C(\boldsymbol{\theta})\right].$$

Adding these two sums and removing everything that does not depend on $\boldsymbol{\theta}$ gives

$$\begin{aligned}\log(\pi(\boldsymbol{\theta}|\boldsymbol{y})) = \\ \text{Const}+\log(\pi(\boldsymbol{\theta}))+\frac{1}{2}\log(|\mathbf{Q}(\boldsymbol{\theta})|) \\ -\frac{1}{2}\log(|\mathbf{Q}_C(\boldsymbol{\theta})|)+\frac{1}{2}\boldsymbol{\mu}_C(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{Q}_C(\boldsymbol{\theta})\boldsymbol{\mu}_C(\boldsymbol{\theta}).\end{aligned}\tag{4.1}$$

From the above expression one can see that the posterior distribution of $\boldsymbol{\theta}$ contains terms which are hard to handle analytically. It is hard to say anything about both the determinants and the quadratic term as functions of $\boldsymbol{\theta}$. Therefore, the inference is done numerically.

An additional distribution that can also be of interest is $\pi(\boldsymbol{u}|\boldsymbol{y})$. This is not as easily calculated as the distribution of $\boldsymbol{\theta}|\boldsymbol{y}$ and usually has to be calculated numerically. It can be found from the expression

$$\pi(\boldsymbol{u}|\boldsymbol{y})=\int_{\mathbb{R}^n}\pi(\boldsymbol{u}|\boldsymbol{\theta},\boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\theta},$$

where n is the dimension of $\boldsymbol{\theta}$.

4.1.2 Approximate inference

As mentioned in the previous section, the posterior distribution $\boldsymbol{\theta}|\boldsymbol{y}$ is hard to work with analytically. Therefore, approximate numerical inference is used. If the posterior distribution were a multivariate Gaussian distribution, it could be characterized by the mode of its probability distribution and the Hessian of the logarithm of its probability distribution at the mode. The mode would correspond to the mean and the Hessian is connected to the covariance matrix. If $\mathbf{K}$ is the Hessian of the logarithm of a multivariate Gaussian probability density at the mode and $\boldsymbol{\Sigma}$ is the covariance matrix, then

$$\boldsymbol{\Sigma} = (-\mathbf{K})^{-1}.$$

$\boldsymbol{\theta}|\boldsymbol{y}$ typically does not have a multivariate Gaussian distribution, but these two statistics are used to give an estimate of the parameter and the associated uncertainty of the estimate.

For the actual implementation, the mode is first found by numerical optimization of the logarithm of the posterior distribution, and then the Hessian is found by a numerical differentiation at the estimated mode. There are numerical issues to each of these operations. Firstly, there may actually be more than one local maximum. Secondly, the estimation of the Hessian involves estimating second order derivatives, which may be hard if the posterior is not well behaved. These are problems connected with optimization and numerical differentiation in general.

Further, there may be restrictions on the parameters, $\boldsymbol{\theta}$. For example, one parametrization that is considered later is

$$\mathbf{H}(\boldsymbol{s}) = \gamma \mathbf{I}_2 + \sum_{k=1}^{n} \beta_k \boldsymbol{v}_k(\boldsymbol{s}) \boldsymbol{v}_k(\boldsymbol{s})^{\mathrm{T}}.$$

Since $\mathbf{H}$ must be positive definite at all positions, not all parameter choices are legal. As discussed later, one reasonable choice here is to take $\gamma > 0$ and $\beta_k \geq 0$ for $k = 1, \ldots, n$. This means that the prior distribution is zero for all other parameter values. One should use an optimization method which can handle simple boundaries on the domain such as these.

4.1.3 Choice of prior

The selection of the prior for the parameters depends on the actual parametrization used. For the examples in this chapter there is no real prior knowledge about the parameters, except that some of the parameters must be positive and so on. The examples only use arbitrarily chosen parameter values and there is no "physical" interpretation of the parameters. Therefore, an improper prior which

specifies which values that are allowed is used. This gives maximum likelihood estimates.

4.2 First parametrization of H

The SPDE in Equation (3.1) uses periodic boundary conditions. This means that **H** has to be periodic. One possible parametrization is

$$\mathbf{H}(\boldsymbol{s}) = \gamma \mathbf{I}_2 + \sum_{k=1}^{n} \beta_k \boldsymbol{v}_k(\boldsymbol{s}) \boldsymbol{v}_k(\boldsymbol{s})^{\mathrm{T}},$$

where $\boldsymbol{v}_k$ is a vector field on the rectangular domain that satisfy the periodic boundary conditions for $k = 1, \ldots, n$. In this parametrization one can consider γ to specify a baseline effect. If $\beta_k = 0$ for all $k = 1, \ldots, n$, then **H** specifies an isotropic field and γ controls the regularity of the isotropic field. The β_k parameter can be considered to specify an additional regularity in the direction of vector field $\boldsymbol{v}_k$. With this interpretation one should have $\gamma > 0$ and $\beta_k \geq 0$ for $k = 1, \ldots, n$.

4.2.1 Examples

In a realistic situation one can not expect to have observed the field at all points (in the grid) and one can not expect to have observed the values exactly. One could imagine that general linear combinations of $\boldsymbol{u}$ were observed, but the focus is on the situation that the field is observed on a subset of the grid cells. This means that **A** is a $k \times N$ matrix, where k is the number of grid cells observed and N is the total number of grid cells. Further, **A** has exactly one 1 on each row, where a 1 in column j on row i means that element i of $\boldsymbol{y}$ is an observation of grid cell number j. In this section $\mathbf{Q}_N = \mathbf{I}_k/\sigma_\epsilon^2$ is used, that means that the values are observed with Gaussian white noise with mean 0 and standard deviation σ_ϵ. The improper prior used is

$$\pi(\gamma, \beta_1, \cdots, \beta_n) = \begin{cases} 1, & (\gamma_1, \beta_1, \cdots, \beta_n) \in [0, \infty)^{n+1}, \\ 0, & \text{otherwise}, \end{cases}$$

which corresponds to maximum likelihood estimates.

Consider first the situation in which one has observed the whole GMRF without noise. The grid size is decreased to 100×100 compared to 200×200 in the examples in Chapter 3 to save computation time.

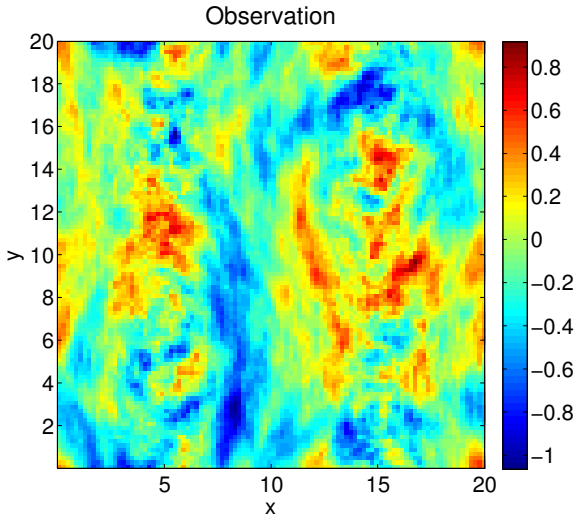


Figure 4.1: An observation of the SPDE in Equation (3.1) on a 100×100 regular grid of $[0, 20]^2$ with periodic boundary conditions, $\kappa^2 = 1$ and $\mathbf{H}(\boldsymbol{s}) = 0.5\mathbf{I}_2 + 5\boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}}$, where $\boldsymbol{v}$ is the vector field in Example 3.2.3.

Example 4.2.1 (γ and one vector field)**.** Use a 100×100 grid of $[0, 20]^2$ and periodic boundary conditions for the SPDE in Equation (3.1). Let κ^2 be identically equal to 1 and let $\mathbf{H}$ be parametrized as

$$\mathbf{H}(\boldsymbol{s}) = \gamma\mathbf{I}_2 + \beta\boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}},$$

where $\boldsymbol{v}$ is the vector field from Example 3.2.3.

Figure 4.1 shows one observation of the solution with $\gamma = 0.5$ and $\beta = 5$. In this case one expects that it is possible to make accurate estimates about γ and β as the solution is observed at all grid points.

The inference is done with the posterior distribution $\boldsymbol{\theta}|\boldsymbol{y}$ from Section 4.1.1, with $\mathbf{Q}_N = \mathbf{I}_{10000}/\sigma^2$, where $\sigma^2 = 10^{-6}$, and $\mathbf{A} = I_{10000}$. This means all values are assumed to be observed nearly exactly. This gives the estimates in Table 4.1. From the table one can see that the estimates for both γ and β are quite accurate, which is reflected both in the actual value of the estimates and the approximated standard deviations. The estimates for both γ and β are accurate to 2 digits.

The example shows that when using only the $\gamma\mathbf{I}_2$ term and one vector field for $\mathbf{H}$, the estimates for the parameters are quite accurate. The accuracy of the estimates for β and γ will of course depend on the vector field used.

Table 4.1: Posterior inference on parameters in Example 4.2.1.

Parameter	**True value**	**Estimate**	**Std.dev.**
γ	0.5	0.5012	0.0081
β	5	5.014	0.084

Table 4.2: Posterior inference on parameters in Example 4.2.2.

Parameter	**True value**	**Estimate**	**Std.dev.**
γ	0.5	0.492	0.021
β_1	5	4.984	0.086
β_2	0	0.000	0.020
β_3	0	0.027	0.025

To make it harder to estimate γ, one can add more vector fields which are known not to be present. In this case one can imagine that one does not know that these vector fields are not present, and see how this changes the posterior results. Care must be taken such that the choice of vector fields actually gives matrices, $\boldsymbol{v}_k(\boldsymbol{s})\boldsymbol{v}_k(\boldsymbol{s})^{\mathrm{T}}$, which are linearly independent. If not, there is infinitely many modes and the estimate of the covariance is not useful.

Example 4.2.2 (γ and three vector fields)**.** Continue with the same situation as in Example 4.2.1, but change the parametrization to

$$\mathbf{H}(\boldsymbol{s}) = \gamma \mathbf{I}_2 + \beta_1 \boldsymbol{v}_1(\boldsymbol{s})\boldsymbol{v}_1(\boldsymbol{s})^{\mathrm{T}} + \beta_2 \boldsymbol{v}_2(\boldsymbol{s})\boldsymbol{v}_2(\boldsymbol{s})^{\mathrm{T}} + \beta_3 \boldsymbol{v}_3(\boldsymbol{s})\boldsymbol{v}_3(s)^{\mathrm{T}},$$

where $\boldsymbol{v}_1$ is the vector field from Example 4.2.1, $\boldsymbol{v}_2(x, y) = (1+0.3\cos(2\pi x/20), 0)$ and $\boldsymbol{v}_3(x, y) = (0, 1 + 0.3\cos(2\pi x/20))$. In addition, use the same observation as in Example 4.2.1. This corresponds to $\gamma = 0.5$, $\beta_1 = 5$, $\beta_2 = 0$ and $\beta_3 = 0$ with the parametrization for $\mathbf{H}$ used in this example.

The inference is done as in Example 4.2.1. This gives the estimates in Table 4.2. From the table one can see that the estimate for γ is less exact than in Example 4.2.1 and that γ has a higher standard deviation. The estimate of γ is exact to 1 digit, compared to the estimate in Example 4.2.1 which is exact to 2 digits. For β_1 the estimate has the same accuracy.

Further the estimates for β_2 and β_3 are both nearly within one standard deviation of 0. This means that the inference gives reason to suspect that they are not significant. Since it is known that they are in fact 0, this is the result that one wants.

As can be seen from the example, introducing additional vector fields can change the degree to which one can estimate the parameters. It is harder to

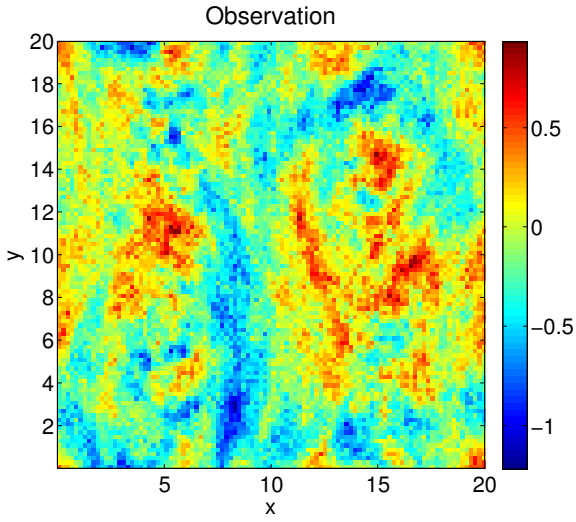


Figure 4.2: The observation in Figure 4.1 with Gaussian white noise with variance 0.01 added to each grid cell.

estimate more parameters as there is a broader range of parameters which is likely to give the observed behaviour.

As noted in the beginning of the section, there may be noise in the observation. Use the same vector fields and observation as in the previous example, but add Gaussian white noise to the observation and repeat the inference.

Example 4.2.3 (All points observed with $\sigma_\epsilon = 0.1$.)**.** The construction of the grid and the parametrization of $\mathbf{H}$ is the same as in Example 4.2.2, but now the observation from that example has Gaussian white noise with standard deviation 0.1 added to each grid cell. The true parameter values are $\gamma = 0.5$, $\beta_1 = 5$, $\beta_2 = 0$ and $\beta_3 = 0$. The original observation is shown in Figure 4.1 and the observation with added noise is shown in Figure 4.2.

The inference is done with the posterior distribution $\boldsymbol{\theta}|\boldsymbol{y}$ from Section 4.1.1 with $\mathbf{Q}_N = \mathbf{I}_{10000}/\sigma^2$, where $\sigma^2 = 0.01$, and $\mathbf{A} = \mathbf{I}_{10000}$. This means that all values are assumed to be observed with iid Gaussian noise with variance 0.01. This gives the estimates in Table 4.3.

As expected, compared to Example 4.2.2 the approximated standard deviations for the parameters are larger than when the observation is exact. But all estimates are approximately within 2 standard deviations from the true values. Since these are approximate standard deviations and the exact distribution is not multivariate Gaussian, it is hard to say exactly what 2 standard deviations

Table 4.3: Posterior inference on parameters in Example 4.2.3.

Parameter	**True value**	**Estimate**	**Std.dev.**
γ	0.5	0.436	0.043
β_1	5	4.9	0.29
β_2	0	0.039	0.0407
β_3	0	0.108	0.052

Table 4.4: Posterior inference on parameters in Example 4.2.4.

Parameter	**True value**	**Estimate**	**Std.dev.**
γ	0.5	0.44	0.26
β_1	5	5.1	4.3
β_2	0	0.02	0.19
β_3	0	0.6	1.2

means. But the standard deviations give indications on the order of error to be expected. The inference still suggests that the parameters β_2 and β_3 are not significant.

The example demonstrates that adding Gaussian white noise to the cells increases the standard deviations of the estimates. Consider now the situation that the field is only observed at some of the grid cells. This may be the case when data is missing.

Example 4.2.4. Use the same SPDE and parametrization of **H** as in Example 4.2.2. As in that example, a 100×100 grid is used, but 1000 points were chosen randomly from the observation in Example 4.2.2. This means only 10% of the grid cells are observed.

From Table 4.4 one can see that there is a lot of uncertainty for all of the parameters. None of the parameters are more than 2 standard deviations from 0. These standard deviations are only approximations, but indicate that there is not enough information to estimate the 4 parameters.

For all the above examples the inference on the parameters in the SPDE works as desired. The hierarchical model works well both with and without noise.

4.2.2 Shortcomings

In the previous section one can see that the examples all have good estimates of the parameters. But consider the simple case that an observation is made without noise and that the true **H** is identically equal to $(\cos(\alpha), \sin(\alpha))(\cos(\alpha), \sin(\alpha))^{\mathrm{T}}$

for some $\alpha \in [0, 2\pi]$. In this case it is impossible to choose a finite set $\{\boldsymbol{v}_k\}_{k=1}^n$ such that for any α there exists a set $\{\beta_k\}_{k=1}^n$ of coefficient such that

$$\begin{bmatrix}\cos(\alpha)\\ \sin(\alpha)\end{bmatrix}\begin{bmatrix}\cos(\alpha) & \sin(\alpha)\end{bmatrix} = \sum_{k=1}^{n} \beta_k \boldsymbol{v}_k(\boldsymbol{s})\boldsymbol{v}_k(\boldsymbol{s})^{\mathrm{T}}.$$

This indicates that the previous parametrization of $\mathbf{H}$ is most appropriate when specifying a model in which the effect of different vector fields are "added". The reason for the problems is that each of the vector fields $\boldsymbol{v}_k$ has a fixed direction at each position. In fact, if $\boldsymbol{v}_k$ is not parallel to $(\cos(\alpha), \sin(\alpha))$ everywhere, β_k must be 0. Since there are uncountably many angles, it follows that no finite set of vector fields can be sufficient.

4.3 Second parametrization of H

Based on the comments in Section 4.2.2, it is useful to look for a different parametrization for inference applications in which the parametric form of $\mathbf{H}$ is not known. One choice is

$$\mathbf{H}(\boldsymbol{s}) = \gamma(\boldsymbol{s})\mathbf{I}_2 + \boldsymbol{v}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})^{\mathrm{T}}, \tag{4.2}$$

where γ is a strictly positive function and $\boldsymbol{v}$ is some vector field. In this situation $\gamma(\boldsymbol{s})$ is equal to the minimum of the eigenvalues of $\mathbf{H}(\boldsymbol{s})$, $\boldsymbol{v}(\boldsymbol{s})$ is in the direction of the eigenvector corresponding to the largest eigenvalue and the length of $\boldsymbol{v}(\boldsymbol{s})$ is equal to the difference between the largest and smallest eigenvalue. In this setting, all constant $\mathbf{H}$ only requires three parameters. Use γ_1, v_1 and v_2 as parameters and write

$$\mathbf{H}(\boldsymbol{s}) \equiv \gamma_1\mathbf{I}_2 + \begin{bmatrix}v_1\\ v_2\end{bmatrix}\begin{bmatrix}v_1 & v_2\end{bmatrix}.$$

All constant $\mathbf{H}$ can be written in this form.

To use Equation (4.2), one has to parametrize the function γ and the vector field $\boldsymbol{v}$ in some manner. γ has to be strictly positive in order for $\mathbf{H}$ to be positive definite, so a parametrization which ensures this is necessary. For $\boldsymbol{v}$ any vector field is possible, so a basis which can generate any vector field is appropriate. Additionally, the function and the vector field must satisfy the periodic boundary conditions.

4.3.1 Parametrization of the vector field

Let the domain be $[0, A] \times [0, B]$ and assume that $\boldsymbol{v}$ is a differentiable, periodic vector field on the domain. Then each component of the vector field can be

written as

$$\sum_{(k,l)\in\mathbb{Z}^2} C_{k,l} \exp\left[2\pi i\left(\frac{k}{A}x + \frac{l}{B}y\right)\right],$$

where i is the imaginary unit. But since the components are real-valued, one can write each of them as a real 2-dimensional Fourier series of the form

$$A_{0,0} + \sum_{(k,l)\in E}\left[A_{k,l}\cos\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right] + B_{k,l}\sin\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right]\right],$$

where the set $E \subset \mathbb{Z}^2$ is given by

$$E = (\mathbb{N}\times\mathbb{Z}) \cup (\{0\}\times\mathbb{N}).$$

Use this Fourier series for each component of $\boldsymbol{v}$ to find the representation

$$\begin{aligned}\boldsymbol{v}(\boldsymbol{s}) = &\\ &\begin{bmatrix}A^{(1)}_{0,0}\\ A^{(2)}_{0,0}\end{bmatrix} + \sum_{(k,l)\in E}\begin{bmatrix}A^{(1)}_{k,l}\\ A^{(2)}_{k,l}\end{bmatrix}\cos\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right] + \\ &\sum_{(k,l)\in E}\begin{bmatrix}B^{(1)}_{k,l}\\ B^{(2)}_{k,l}\end{bmatrix}\sin\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right],\end{aligned} \tag{4.3}$$

where $A^{(1)}_{k,l}$ and $B^{(1)}_{k,l}$ are the coefficients for the first component of $\boldsymbol{v}$ and $A^{(2)}_{k,l}$ and $B^{(2)}_{k,l}$ are the coefficients of the second component.

Introduce new coefficients $\hat{A}^{(1)}_{k,l}$ and $\hat{A}^{(2)}_{k,l}$ through

$$A^{(1)}_{k,l} = \frac{k}{A}\hat{A}^{(1)}_{k,l} - \frac{l}{B}\hat{A}^{(2)}_{k,l}, \tag{4.4}$$

$$A^{(2)}_{k,l} = \frac{l}{B}\hat{A}^{(1)}_{k,l} + \frac{k}{A}\hat{A}^{(2)}_{k,l}, \tag{4.5}$$

for $(k,l) \neq (0,0)$. Then the cosine term in Equation (4.3) can for term (k,l) be written as

$$\left(\hat{A}^{(1)}_{k,l}\begin{bmatrix}\frac{k}{A}\\ \frac{l}{B}\end{bmatrix} + \hat{A}^{(2)}_{k,l}\begin{bmatrix}-\frac{l}{B}\\ \frac{k}{A}\end{bmatrix}\right)\cos\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right].$$

It is easy to verify that the first part gives something with zero curl and that the second part gives something with zero divergence. Exactly the same can be done for the sine term. The result is

$$\left(\hat{B}^{(1)}_{k,l}\begin{bmatrix}\frac{k}{A}\\ \frac{l}{B}\end{bmatrix} + \hat{B}^{(2)}_{k,l}\begin{bmatrix}-\frac{l}{B}\\ \frac{k}{A}\end{bmatrix}\right)\sin\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right],$$

where the first part has zero curl and the second part has zero divergence.

Lemma 4.3.1. *Let* $\mathbf{M} : \mathbb{R}^2 \to \mathbb{R}^2$ *be defined by*

$$\mathbf{M}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{k}{A} & -\frac{l}{B} \\ \frac{l}{B} & \frac{k}{A} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix},$$

where A *and* B *are positive constants. Then for* $(k, l) \in E$*, where*

$$E = (\mathbb{N} \times \mathbb{Z}) \cup (\{0\} \times \mathbb{N}),$$

$\mathbf{M}$ *is an invertible linear transformation.*

Proof. The expression for $\mathbf{M}$ gives

$$\det(\mathbf{M}) = \frac{k^2}{A^2} + \frac{l^2}{B^2}.$$

This is greater than 0 for all $(k, l) \neq (0, 0)$. So the linear transformation is invertible. □

The lemma shows that the new and the old coefficients uniquely determine each other. Therefore, these new coefficients together with $A^{(1)}_{0,0}$ and $A^{(2)}_{0,0}$ provide an alternative way of representing $\boldsymbol{v}$. The details can be summarized in a proposition.

Proposition 4.3.2. *Let* A *and* B *be positive constants and let* $\boldsymbol{v}$ *be a differentiable vector field on* $[0, A] \times [0, B]$*, such that the vector field and its first order derivatives agree on opposite edges. Then there exist vector fields* $\hat{\boldsymbol{v}}$ *and* $\boldsymbol{w}$ *that satisfies the same conditions as* $\boldsymbol{v}$ *such that*

$$\boldsymbol{v}(\boldsymbol{s}) = \bar{\boldsymbol{v}} + \hat{\boldsymbol{v}}(\boldsymbol{s}) + \boldsymbol{w}(\boldsymbol{s}), \qquad \forall \boldsymbol{s} \in [0, A] \times [0, B],$$

where $\bar{\boldsymbol{v}}$ *is the mean of* $\boldsymbol{v}$*,* $\hat{\boldsymbol{v}}$ *has mean zero and* $\operatorname{curl}(\hat{\boldsymbol{v}}) = 0$ *and* $\boldsymbol{w}$ *has mean zero and* $\operatorname{div}(\boldsymbol{w}) = 0$*.*

Let

$$E = (\mathbb{N} \times \mathbb{Z}) \cup (\{0\} \times \mathbb{N}),$$

then $\hat{\boldsymbol{v}}$ *can be written as*

$$\begin{aligned} \hat{\boldsymbol{v}}(\boldsymbol{s}) = & \\ & \sum_{(k,l)\in E} \hat{A}^{(1)}_{k,l} \begin{bmatrix} \frac{k}{A} \\ \frac{l}{B} \end{bmatrix} \cos\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right] + \\ & \sum_{(k,l)\in E} \hat{B}^{(1)}_{k,l} \begin{bmatrix} \frac{k}{A} \\ \frac{l}{B} \end{bmatrix} \sin\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right], \end{aligned}$$

where $\hat{A}^{(1)}_{k,l}$ *and* $\hat{B}^{(1)}_{k,l}$ *are unique real coefficients, and* $\boldsymbol{w}$ *can be written as*

$$\begin{aligned}\boldsymbol{w}(\boldsymbol{s}) = &\\ &\sum_{(k,l)\in E} \hat{A}^{(2)}_{k,l} \begin{bmatrix} -\frac{l}{B} \\ \frac{k}{A} \end{bmatrix} \cos\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right] + \\ &\sum_{(k,l)\in E} \hat{B}^{(2)}_{k,l} \begin{bmatrix} -\frac{l}{B} \\ \frac{k}{A} \end{bmatrix} \sin\left[2\pi\left(\frac{k}{A}x + \frac{l}{B}y\right)\right].\end{aligned}$$

where $\hat{A}^{(2)}_{k,l}$ *and* $\hat{B}^{(2)}_{k,l}$ *are unique real coefficients.*

Proof. A Fourier series provides a unique decomposition for $\boldsymbol{v}$, and Lemma 4.3.1 shows that there is a one-to-one correspondence between the coefficients in the decomposition in this proposition and the coefficients in the Fourier series. Since the basis is still the same cosine and sine functions, the decomposition in this proposition exists for all differentiable, periodic vector fields and the coefficients are unique. □

The fact that a differentiable, periodic vector field with mean $\mathbf{0}$ can be decomposed uniquely in a curl free component and a divergence free component can also be seen from the fact that the only solutions of the Laplace equation on a rectangle with periodic boundaries are the constant functions.

The decomposition in Proposition 4.3.2 can be useful for, for example, determining if the vector field in $\mathbf{H}$ is curl free or divergence free. And if one knows that the vector field is curl free one can reduce the number of parameters needed.

4.3.2 Parametrization of γ

The function γ must be strictly positive. One idea for achieving this is to instead use γ^2, but this does not help as keeping γ^2 from becoming 0 is the same as keeping γ positive or negative everywhere. A possible solution is to take

$$\gamma(\boldsymbol{s}) = \exp\left(\sum_{k=1}^{\infty} \gamma_k(\boldsymbol{s})\right),$$

where $\{\gamma_k\}_{k=1}^{\infty}$ is, for example, the Fourier basis. This function is strictly positive.

4.3.3 Examples

This section gives examples for the homogeneous anisotropic case with $\mathbf{H}$ constant. This can not be estimated well with the previous scheme, but only requires three parameters with this parametrization. Three parameters is what one should

Table 4.5: Parameter estimates for Example 4.3.1.

Parameter	**True value**	**Estimate**	**Std.dev.**
γ	3	2.971	0.071
v_1	0.707	0.730	0.049
v_2	1.225	1.237	0.039

expect since it requires three parameters to specify all positive definite 2×2 matrices. The following example only uses the minimum number of parameters.

Example 4.3.1. Use the SPDE

$$u(\boldsymbol{s}) - \nabla \cdot \mathbf{H} \nabla u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in [0, 20] \times [0, 20], \tag{4.6}$$

where $\mathcal{W}$ is a standard Gaussian white noise process and $\mathbf{H}$ is a 2×2 matrix, with periodic boundary conditions. Let

$$\mathbf{H} = 3\mathbf{I}_2 + 2\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}},$$

with $\boldsymbol{v} = (1, \sqrt{3})/2$. This means that $\mathbf{H}$ has eigenvector v with eigenvalue 5 and an eigenvector orthogonal to v with eigenvalue 3. Use a 100×100 grid for the construction of the GMRF.

One observation of the solution is shown in Figure 4.3. Assume that the fact that $\mathbf{H}$ is constant is known, but that its value is not. Then using the decomposition from the previous sections one writes

$$\mathbf{H} = \gamma \mathbf{I}_2 + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix},$$

where γ, v_1 and v_2 are the parameters. Use an improper prior that disallows $\gamma < 0$. This gives the estimates shown in Table 4.5. From the table one can see that all the estimates are accurate to one digit, and within one standard deviation of the true value.

In the example the estimates are close to the correct values, and the computation time required is small. However, if one increases the number of parameters used in the decomposition in Proposition 4.3.2, the computational task is much higher and it is harder to find the global maximum of the posterior distribution for the parameters. This is demonstrated in the following example.

Example 4.3.2. This is a continuation of Example 4.3.1 and uses the same SPDE and grid. In addition, the same observation is used. The previous example uses only three parameters. Assume that it is not known that $\mathbf{H}$ is constant and use in addition the terms from Proposition 4.3.2 for $(k, l) = (1, 0)$, $(k, l) = (0, 1)$ and

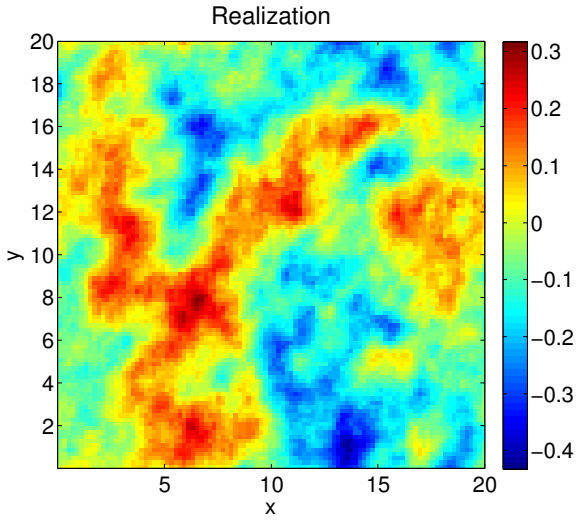


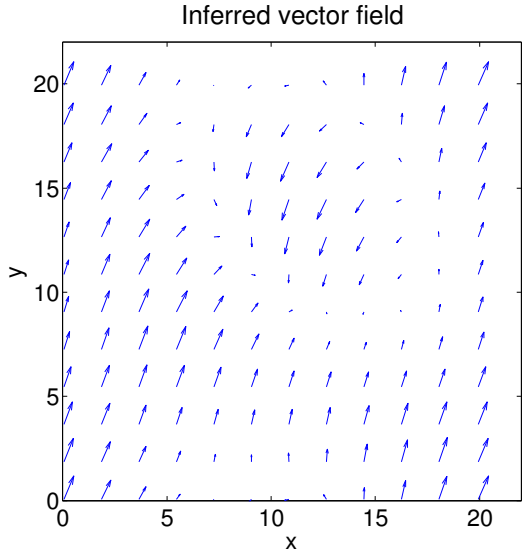
Figure 4.3: One realization of the solution of the SPDE in Equation (4.6).

$(k, l) = (1, 1)$. This give 12 extra parameters, 4 additional parameters for each frequency.

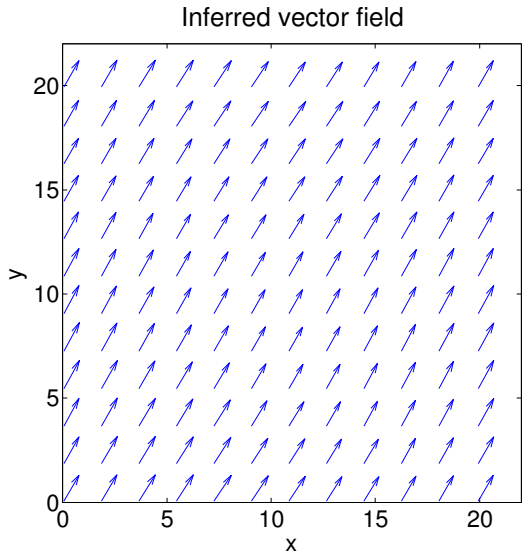
First two arbitrary starting positions are chosen for the optimization. The first is $\gamma = 3.0$ and all other parameters at 0.1. And the second is $\gamma = 3.0$, $A_{0,0} = 0.1$, $\hat{A}_{0,0} = 0.1$ and all other parameters equal to 0. For both of these starting points the optimization converges to more or less the same non-global maximum. Parameter estimates and approximate standard deviations are calculated, but requires a large table. So instead only the resulting vector field $\boldsymbol{v}$ is shown in Figure 4.4(a). Note that the vectors (a, b) and $-(a, b)$ have the same effect on $\mathbf{H}$, so only the direction up to a factor of -1 is important.

A third optimization is done with starting values close to the correct parameter values. This gives the vector field shown in Figure 4.4(b). This gives estimates for γ, $A^{(1)}_{0,0}$ and $A^{(2)}_{0,0}$ that agree with the ones in Example 4.3.1 to two digits. The other frequencies all had coefficients close to zero, with the largest having an absolute value of 0.058.

Comparing the two vector fields in Figure 4.4 it is clear that there are major differences. The vector field in Figure 4.4(b) has low variations from the mean value, whereas the one in Figure 4.4(a) has a lot of variation. In this example the problem is that there are at least two maxima for the posterior (actually at least 4, due to symmetry). It is hard to avoid these alternative maxima, but part of the problem is likely to be that there are 15 parameters, but only one



(a) Wrong maximum.



(b) Correct maximum.

Figure 4.4: Two different local maxima found for the vector field. The vector field in (a) gives a lower value for the posterior distribution of the parameters than the vector field in (b).

observation.

The example shows that introducing many parameters makes the optimization much harder, but that when starting close to the actual parameter values one gets an estimate close to the actual vector field. In addition, there is a large increase in computation time when increasing the parameter space to be 15-dimensional, compared to a 3-dimensional parameter space. The computation time required is increased by a factor of 10.

4.4 Model verification

In this chapter there is an underlying assumption that it is possible to specify $\mathbf{H}$ in the chosen way. But one does in general not expect to be able to specify a parametric form that catches the real value of $\mathbf{H}$, when the real value of $\mathbf{H}$ is not know. The inference scheme presented in the previous sections requires a correct specification, and if the specification is incorrect, the estimated values may give no indication of it. A simple example is to simply specify an isotropic model with $\mathbf{H} = \gamma \mathbf{I}_2$. The posterior estimate of the standard deviation of γ might in this case not give any indication of how appropriate this specification is. Therefore, the inference done in this chapter should be combined with some kind of model verification. However, this is not something which have been studied extensively for these type of SPDE-based GMRFs yet.

Chapter 5

Discussion

The goal of this thesis is to study SPDE-based GMRFs for use in spatial statistics. To this end, a specific SPDE is chosen in Chapter 3 with rectangular domain and periodic boundary conditions. These choices are based on a desire to not introduce unnecessary complications. For a more general domain, for example, a convex domain with a boundary consisting of a finite number of straight line segments, it is in general not possible to divide the domain into rectangles. But this problem can be solved with a triangulation of the domain. Triangulation is a known problem in numerics, and the finite volume method can be applied to the triangles in the triangulation in a similar manner as for the regular discretization in the thesis. Thus the SPDE-based GMRFs are not limited to rectangular domains. In fact, in Lindgren et al. (2011) it is demonstrated that even a 2-sphere can be used. For the SPDE in Equation (3.1) the diffusive operator has to be modified slightly to make it work for a non-Cartesian coordinate system, but it is possible.

On the other hand, the boundary conditions must be changed for a more general domain. The periodic boundary conditions provide a simple way to avoid worrying about boundary conditions for the rectangle, but there is no clear way to extend it to other domains. In Lindgren et al. (2011) Neumann boundary conditions are used to give a solution which behaves close to the desired Matérn field. For the SPDE considered in this thesis something similar can be done. The normal derivative does not provide the desired behaviour. The diffusion matrix can give an anisotropic diffusion operator, so setting the normal derivative equal to zero does not stop "things" from leaving or entering the domain. A reasonable solution is to use the directions provided by the diffusion matrix at the boundary. Thus there are different boundary conditions that can be applied for more general boundaries, but this is something to be considered in further work.

An additional point worth mentioning is that the work in the thesis is restricted to a constant function κ^2 in the SPDE. One thing one can note from Example 3.2.3 is that the diffusion matrix $\mathbf{H}$ controls the covariance structure according to the vector field used to construct it, but there is also an effect on the marginal variances. It would be desirable to also be able to keep the marginal variances nearly constant while still varying the diffusion matrix. To which degree κ^2 can be used to do exactly that is a question not answered in this thesis, but it is an interesting point for further studies.

From the examples in Section 3.2 it is clear that diverse spatial GMRFs can be constructed from the SPDE. It is interesting to note that the isotropic case in Section 3.2.1 is controlled by only two parameters. One parameter controls the correlation range and the other is a scale factor for the solution. Thus one can control correlation range and marginal variance. For the homogeneous anisotropic case in Section 3.2.2 two new parameters are introduced. The example indicates that one can specify the elliptic level curves of the correlation. One can use two parameters to specify the shape and rotation of the ellipse, one parameter to specify the rate at which the correlation decreases and one parameter to control the marginal variance.

For the inhomogeneous GMRFs in Section 3.2.3 it is not possible to reduce to a finite number of parameters. Any sufficiently differentiable diffusion matrix can be used. As demonstrated in Section 3.2.3, the diffusion matrix can be constructed from a vector field that everywhere specifies directions with additional dependence. The example shows that this gives covariances that curve along the vector field. In the specific example the effect is large because the direction of the vector field is given much higher dependence than the orthogonal direction. Overall, the use of a vector field seems to have clear advantages when visualizing the effect of the diffusion matrix. One possible problem is that changing the diffusion matrix also changes the marginal variances.

For the SPDE-based GMRFs to be useful, it must also be possible to introduce useful parameters which control the GMRF. In Chapter 4 observations from the GMRF are used as way of estimating the diffusion matrix. Section 4.2 and Section 4.3 introduce two different parametrizations of the diffusion matrix. For each of the parametrizations one is guaranteed to get a well defined precision matrix for the GMRF. As long as the parametrization gives a everywhere positive definite, differentiable diffusion matrix, the resulting precision matrix is valid. Thus for the SPDE in this thesis it is easy to introduce parameters in a valid way.

The parametrizations in Section 4.2 and Section 4.3 are used to estimate parameters based on observations of the GMRF with and without noise. The examples for the first parametrization in Section 4.2.1 all show the good results for the parameter estimation. But from Section 4.2.2 it is clear that this parametrization

is not appropriate unless one knows what kind of vector fields that control the diffusion matrix. For instance, there is no general choice of vector fields that can estimate a constant diffusion matrix. This specific problem is better solved with the parametrization in Section 4.3. This parametrization uses only one vector field and instead parametrizes that one vector field. The effect of this is that the vector field can take any constant direction with only two parameters, and one can estimate constant diffusion matrices with only three parameters. In Example 4.3.1 the use of three parameters gives good results. But when 15 parameters are used in Example 4.3.2, the inference becomes much harder. For a very good starting point it still converges to the correct solution, but it is increasingly hard to do the numerical inference. An additional point is that there is a limit to how many parameters one should try to estimate based on a single observation. But this should be given closer consideration in further work.

A question only briefly touched upon in the thesis is the choice of prior. The posterior distributions are found, but the prior is only used to limit the valid parameter choices for the maximum likelihood estimations. The main reason for this is that in all the examples the parameters are chosen arbitrarily. In this setting a choice of prior would be somewhat artificial unless one gives some meaning to the parameters. For futher study of the choice of priors, it seems useful to consider real-life problems that one wants to solve.

From the above points there are many arguments for using a SPDE to specify a GMRF. Firstly, the SPDE in this thesis can be used to specify many interesting covariance structures from different diffusion matrices and the resulting GMRFs possess a sparseness which enables fast simulation and numerical inference. Secondly, it is simple to introduce parameters that guarantee valid precision matrices and the simple examples show that the hierarchical inference scheme gives reasonable results. Thirdly, the GMRFs are approximating some continuous solution and one can increase resolution without resulting to ad hoc modifications of the precisions.

Necessarily, the approach is not perfect. The SPDE used in this thesis can not give all possible covariance structures. Further, there are issues with how to parametrize the coefficients and numerical issues for the actual implementation. But, most importantly, the approach should be tested on a real-life example where the parameters can be given a meaning and there are other established methods to compare it with.

Chapter 6

Conclusion

An SPDE with non-constant coefficients provides a useful way of specifying an inhomogeneous GMRF. Specifically, the spatial diffusion equation studied in this thesis admits the construction of a GMRF through a finite volume method applied to the SPDE, and it is simple to create interesting inhomogeneous GMRFs by varying the diffusion matrix.

Further, the introduction of parameters in the coefficients of the SPDE provides a way of estimating the coefficients. The sparseness of the GMRF approximation implies that the approximate inference can be done considerably faster than for a dense covariance matrix. But there are some issues with choosing an appropriate parametrization of the coefficients and with numerical optimization.

In summary, the SPDE-based GMRFs show good potential for spatial models.

6.1 Future work

The work in this thesis focuses on the diffusion matrix in the SPDE in Equation (3.1), but it would be interesting to also extend the work to a non-constant κ^2. In addition, there is more work to be done on the choice of the parametrization for the diffusion matrix and on the prior distributions for the parameters. This is especially interesting for real-world applications where there is prior knowledge available.

Further, the SPDE can be extended to also include a time derivative $\frac{\partial u}{\partial t}$ and a transport term $\nabla(bu)$, where b is a scalar function. This raises interesting questions as to what the difference between the diffusive term and the transport term is for the resulting spatio-temporal model.

Bibliography

Abrahamsen, P. 1997. *A Review of Gaussian Random Fields and Correlation Functions: Second Edition*, Technical Report 917, Norwegian Computing Center.

Adler, R.J. and J.E. Taylor. 2007. *Random Fields and Geometry*, Springer Verlag.

Chen, Y., T.A. Davis, W.W. Hager, and S. Rajamanickam. 2008. *Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate*, ACM Trans. Math. Softw. **35**, 22:1–22:14.

Eymard, R., T. Gallouët, and R. Herbin. 2000. *Finite Volume Methods*, Solution of Equations in $\mathbb{R}^n$ (part 3), Techniques of Scientific Computing (Part 3), pp. 713 –1018.

Gelfand, A.E., M. Fuentes, P. Diggle, and P. Guttorp. 2010. *Handbook of Spatial Statistics*, CRC.

Guennebaud, G., B. Jacob, et al. 2010. *Eigen v3*, `http://eigen.tuxfamily.org`.

Johnson, S.G. 2011. *The NLopt nonlinear-optimization package*, `http://ab-initio.mit.edu/nlopt`.

Lindgren, F. and H. Rue. 2008. *On the Second-Order Random Walk Model for Irregular Locations*, Scandinavian journal of statistics **35**, no. 4, 691–700.

Lindgren, F., H. Rue, and J. Lindström. 2011. *An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion)*, Journal of the Royal Statistical Society, Series B. (in press).

Øksendal, B.K. 2003. *Stochastic Differential Equations: An Introduction with Applications*, Sixth Edition, Springer Verlag.

Rue, H. and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall.

Svanberg, K. 2002. *A class of globally convergent optimization methods based on conservative convex separable approximations*, SIAM journal on optimization **12**, no. 2, 555–573.

Walsh, J. 1986. *An introduction to stochastic partial differential equations*, École d'Été de Probabilités de Saint Flour XIV-1984, 265–439.