FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Lars Lau Rakêt

2D Functional Data Analysis

with applications to image analysis

Thesis for the Master degree in Statistics
Department of Mathematical Sciences

Supervisors: Bo Markussen & Helle Sørensen

Hand in date: July 27, 2010

Note

This document is a slightly revised version of the original thesis. A number of minor typos have been corrected, and references have been updated. Note in particular that the accompanying R package is no longer available. The main scientific contributions of this thesis can also be found in improved form in the paper Rakêt & Markussen (2014).

Lars Lau Rakêt
May 5, 2013

Abstract

This thesis considers functional data analysis on two dimensional rectangular domains. The central problem of the work is that of regularization or smoothing, where a roughness measure is introduced as a penalizing term in the likelihood function. In certain cases a corresponding Bayesian model producing equivalent estimates can be specified, and this is the principal model considered in this thesis. The estimates in this model can be computed explicitly, but requires calculation of the Green's function corresponding to the roughness measure. The main results of the thesis are in Chapter 5 where a number of approximations to the problem are introduced. The idea behind the approximations is to move the calculations (time intensive matrix operations) to the functional domain. These approximations make it possible to do smoothing of large datasets (e.g. several million observations.) The approximation results are new, and comes as a generalization of equivalent one-dimensional approximations by Markussen (2013). Finally the model and results are applied on 2D electrophoresis images. In connection with the thesis an R package (R Core Team 2013) named `laplacesmooth` has been developed, which implements most methods presented in this thesis.

Preface

This text and the R software package `laplacesmooth` constitute my thesis for the MSc (Cand.Scient) degree in Statistics. The thesis has been written in the period February-July under supervision of Bo Markussen and Helle Sørensen, and handed in July 27, 2010 at Department of Mathematical Sciences, University of Copenhagen.
The thesis is mainly based on Markussen (2013), generalizing a number of results to two dimensions.

Structure

Chapter 1 introduces and motivates the penalized likelihood smoothing problem in two dimensions. Then a Bayesian model with a Gaussian prior is introduces, and it is shown that this model produces equivalent estimates to those of the penalized likelihood approach.

In Chapter 2 rigorous definitions of Gaussian fields and Green's functions are given, and some results linking covariance functions to continuity and smoothness of Gaussian fields are given. Finally the standard reproducing kernel Hilbert space approach to smoothing using generalized splines is covered.

Chapter 3 deals with the actual calculations in the model introduced in Chapter 1. The Laplace operator is covered first and an interpretation of the meaning of using the Laplace operator for smoothing is given. Next a number of Green's functions related to the Laplacian are found. The most important is the Green's function for the iterated Laplacian $(-\Delta)^\ell$, which is the main smoothing operator throughout the thesis. At the end of the chapter the proposed model is reviewed in relation to other models in the literature.

In Chapter 4 three different simulated data examples are considered, and methods from the previous chapter are used.

Chapter 5 holds the main results of this thesis. Here the computationally expensive matrix computations are approximated by continuous counterparts, that are both pleasing from the perspective that the data is only a discrete sample of a continuous function, and because most of the operations can be implemented to run in constant or linear time as opposed e.g. cubic. The approximations are inspired by the one-dimensional counterparts coined of Markussen (2013).

Chapter 6 is as the title says, "Chapter 5 in practice". Closed formulae (i.e. convergent infinite double series) for the approximations in the previous chapter are calculated, the quality of the approximations is studied and one of the data examples from Chapter 4 is revisited in the approximated setting.

In Chapter 7 smoothing is performed on 2D electrophoresis images using the approximations, different kinds of model control is done, and a likelihood ratio test for testing wether a spot on the image is a real protein spot or noise is introduced.

The final chapter reviews possibilities and future directions.

R package

The thesis has a companion R package called `laplacesmooth`. The package is documented in the standard R help system, where examples on use of all functions can be found as well. A number of listings containing examples of use of the package is also supplied in this text.

Thanks

First and foremost I would like to thank Bo Markussen for coming up with the subject of the thesis, for his help and guidance throughout the writing process, and for patiently answering all my questions. I would also like to thank Helle Sørensen for being my internal supervisor and for good advice on thesis writing. Finally I would like to thank my dearest Line Skotte for thorough proofreading and for bringing me along to Berkeley during the writing of this thesis.

CONTENTS

CHAPTER 1

INTRODUCTION

This thesis will present methods for statistical inference for image data viewed as two-dimensional functional data. The methods are based on the one-dimensional approach presented in Markussen (2013). Since these methods are not standard, this chapter will work as an initial motivation, that introduces the problem, corresponding statistical models, and the resulting estimators.

1.1 Motivation

Consider the function $\theta : [0,1] \times [0,1] \to \mathbb{R}$, which we will refer to as an *image*, and the differential operator $\mathscr{K}$ of order k that can be used to measure the roughness of θ by the integral of squares

$$\int_0^1 \int_0^1 |\mathscr{K}\theta(s,t)|^2 \,\mathrm{d}s\,\mathrm{d}t.$$

The methods that will be presented will mainly focus on the squared operator $\mathscr{L} = \mathscr{K}^\dagger \mathscr{K}$ which we can take to be defined on $\mathcal{C}^{2k}([0,1] \times [0,1], \mathbb{R})$. The roughness measure can be expressed in terms of $\mathscr{L}$:

$$\int_0^1 \int_0^1 |\mathscr{K}\theta(s,t)|^2 \,\mathrm{d}s\,\mathrm{d}t = \langle \mathscr{K}\theta, \mathscr{K}\theta \rangle = \langle \theta, \mathscr{K}^\dagger \mathscr{K}\theta \rangle = \langle \theta, \mathscr{L}\theta \rangle.$$

$\mathscr{L}$ is obviously positive semidefinite, but not neccesaraly positive definite on the space on which it is defined. We will need that $\mathscr{L}$ is positive definite, and therefore we will restrict the choices of θ to some subspace $\mathscr{H}$ such that this is the case. Let us consider an example.

Example 1.1. Let $\mathscr{K} = \alpha\partial_s + \mathbb{I}$ where $\alpha \in \mathbb{R}$. ∂_s denotes the partial differential operator with respect to the first coordinate and $\mathbb{I}$ is the identity operator. Expanding the inner integral of the roughness measure, one gets that

$$\int_0^1 |(\alpha\partial_s + \mathbb{I})\,\theta(s,t)|^2 \,\mathrm{d}s = \int_0^1 (\alpha\partial_s\theta(s,t))^2 \,\mathrm{d}s + \int_0^1 \theta(s,t)^2 \,\mathrm{d}s + 2\int_0^1 \theta(s,t)\alpha\partial_s\theta(s,t)\,\mathrm{d}s.$$

Applying integration by parts on the first integral it is obtained that

$$\int_0^1 (\alpha\partial_s\theta(s,t))^2 \,\mathrm{d}s = \alpha^2 \Big[\theta(s,t)\partial_s\theta(s,t)\Big]_0^1 - \int_0^1 \theta(s,t)\alpha^2\partial_s^2\theta(s,t)\,\mathrm{d}s,$$

and so, if the boundary condition

$$\theta(0,t) = \theta(1,t) = 0$$

is imposed for all $t \in [0,1]$, the roughness measure is given by

$$\int_0^1 \int_0^1 |\mathscr{K}\theta(s,t)|^2 \,\mathrm{d}s\,\mathrm{d}t = \int_0^1 \int_0^1 \theta(s,t)(-\alpha^2\partial_s^2 + \alpha\partial_s + \mathbb{I})\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

I.e. for $\theta \in \mathscr{H}$ where $\mathscr{H}$ is given by

$$\mathscr{H} = \left\{ \theta \in \mathcal{C}^{2k}([0,1] \times [0,1], \mathbb{R}) \,\middle|\, \theta(0,t) = \theta(1,t) = 0 \right\},$$

the squared operator $\mathscr{K}^\dagger\mathscr{K}$ is given by $\mathscr{L} = -\alpha^2\partial_s^2 + \alpha\partial_s + \mathbb{I}$. It is easily verified that $\mathscr{L}$ is positive definite on $\mathscr{H}$. ∘

We have only introduced a single differential operator for measuring roughness, but if we have n differential operators $\mathscr{K}_1, \ldots, \mathscr{K}_n$, each measuring a different aspect of roughness, we can put

$$\mathscr{L} = \sum_{i=1}^{n} \mathscr{K}_i^\dagger \mathscr{K}_i,$$

which, as long as $\mathscr{H}$ is chosen such that $\mathscr{L}$ is positive definite, will work just as seen until now.

Now consider the simple statistical model of having noisy observations y_{ij} of an image in the grid $(s_i, t_j)_{ij}$ with $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, and where the noise terms ε_{ij} are assumed to be independent and $\mathcal{N}(0, \sigma^2)$-distributed:

$$y_{ij} = \theta(s_i, t_j) + \varepsilon_{ij}, \qquad \theta \in \mathscr{H}. \tag{1.1}$$

It is problematic to infer θ in this model as $\mathscr{H}$ will typically be a very versatile class of functions, and thus overfitting problems arise with standard likelihood methods. One way to handle this problem is to introduce a penalty in the form of the roughness measure to the negative log-likelihood. This gives the negative penalized log-likelihood $\ell_y^{\mathscr{L}}$:

$$\ell_y^{\mathscr{L}}(\theta, \sigma^2, \lambda) = \frac{n_1 n_2}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (y_{ij} - \theta(s_i, t_j))^2 + \lambda \int_0^1 \int_0^1 \theta(s,t) \mathscr{L} \theta(s,t) \, \mathrm{d}s \, \mathrm{d}t,$$

where λ is a smoothing parameter.

Before we move on to the estimation of θ, the next example will address how $\mathscr{L}$ can be chosen.

Example 1.2. Suppose that $\theta(s,t)$ is the temperature of a one-dimensional metal string at the point s to time t, with initial condition that the entire string has zero temperature at the beginning of the measurements, and the condition that the endpoints are kept at zero temperature at all times:

$$\theta(s, 0) = \theta(0, t) = \theta(1, t) = 0.$$

If we observe θ in the grid $(s_i, t_j)_{i,j}$ with independent standard normal distributed errors, we are in a setup corresponding to the model (1.1). We are interested in estimating the evolution of the distribution of heat over time, and perhaps also in estimating the partial derivatives. The question is then how $\mathscr{L}$ should be chosen. It is known from physics that the distribution of heat in metal is usually described well by the solution to the heat equation

$$\partial_t \theta(s,t) - \alpha^2 \partial_s^2 \theta(s,t) = 0,$$

where α^2 is a constant known as the thermal diffusivity. This way $\mathscr{K} = \partial_t - \alpha^2 \partial_s^2$ corresponds to giving a penalty to all functions that do not satisfy the heat equation.

The above approach can be used to easily choose the differential operator $\mathscr{L}$ when we are observing structures, that are (approximately) governed by a differential equation. If such information is not available, $\mathscr{L}$ will typically be chosen to be a sum of squared partial derivatives up to some order k, which will produce smoother estimates up to the order k. ○

Let us now consider the problem of estimating θ using the penalized likelihood function. It should be noted that this problem is only interesting as long as there is no function $\theta \in \mathscr{H}$ with $\mathscr{L}\theta = 0$ which also produces a perfect fit. Denote by

$$\frac{\delta}{\delta\theta} \ell_y^{\mathscr{L}}(\theta, \sigma^2, \lambda)$$

the functional derivative of the negative penalized log-likelihood. As $\ell_y^{\mathscr{L}}(\theta, \sigma^2, \lambda)$ is a strictly convex functional of $\theta \in \mathscr{H}$, the unique minimum is the root of $\frac{\delta}{\delta\theta} \ell_y^{\mathscr{L}}(\theta, \sigma^2, \lambda)$. Remembering that $\mathscr{L}$ is linear

and self-adjoint, it is easily derived that

$$\frac{\delta}{\delta\theta(s,t)}\ell_y^{\mathscr{L}}(\theta(s,t),\sigma^2,\lambda) = \frac{1}{\sigma^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\theta(s_i,t_j)-y_{ij})\delta_{(s_i,t_j)}(s,t) + 2\lambda\mathscr{L}\theta(s,t), \tag{1.2}$$

where $\delta_{(s_i,t_j)}$ is the 2-dimensional Dirac delta function. Now assume that $\mathscr{L}$ has Green's function $\mathcal{G}$, i.e.

$$\mathscr{L}^{-1}f = \int_0^1\int_0^1 \mathcal{G}(\,\cdot\,,(v,w))f(v,w)\,\mathrm{d}v\,\mathrm{d}w.$$

Considering

$$\mathscr{L}^{-1}\delta_{(s,t)} = \int_0^1\int_0^1 \mathcal{G}(\,\cdot\,,(v,w))\,\delta_{(s,t)}(v,w)\,\mathrm{d}v\,\mathrm{d}w = \mathcal{G}(\,\cdot\,,(s,t))$$

it is obtained that

$$\mathscr{L}\mathcal{G}(\,\cdot\,,(s,t)) = \delta_{(s,t)},$$

and with this in mind it is easily verified that

$$\left\{\frac{1}{2\lambda}\mathcal{G}(\,\cdot\,,(s_k,t_l))\right\}_{(k,l)}\left\{\frac{1}{2\lambda}\mathcal{G}((s_i,t_j)(s_k,t_l)) + \sigma^2\mathbb{1}_{(i,j)=(k,l)}\right\}^{-1}_{(i,j),(k,l)} y, \tag{1.3}$$

where $y = \{y_{ij}\}_{(i,j)}$, is a root of (1.2), and therefore the penalized likelihood estimate of θ.

In the next section the problem of estimating the image will be presented in an alternative setup, where it turns out that the minimum mean square error estimator is identical to the penalized maximum likelihood estimator (1.3).

1.2 Model

If we consider the same setup as in the previous section, i.e. noisy observations $y = \{y_{ij}\}_{(i,j)}$ of an image and some differential operator $\mathscr{L}$, invertible on a space $\mathscr{H}$, which can be used for measuring the roughness of the image, one could propose the model

$$y_{ij} = x(s_i,t_j) + \varepsilon_{ij} \tag{1.4}$$

where the ε_{ij}'s are independent and identically $\mathcal{N}(0,\sigma^2)$-distributed random variables and x is a Gaussian process (or Gaussian random field), independent of the ε_{ij}'s, with zero mean, a property that we will refer to as *centeredness*, and covariance function $\tau^2\mathcal{G}$ where $\mathcal{G}$ still denotes the Green's function corresponding to $\mathscr{L}$.[1]

In this model x plays the role of the image θ from the model (1.1), but the two models seem very different. In (1.1) θ is a smooth surface that is continuously differentiale up to an order $2k$, whereas in (1.4) the image-component x need not even be differentiable. Also it seems unlikely that a Gaussian process with drift 0 should exhibit all the traits of a particular image θ. Undeterred by these facts, we will continue the analysis of the model (1.4), where x is thought of as a prior for θ in a Bayesian point of view.

With a slightly generalized notation, we have that

$$\begin{pmatrix} y \\ x \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \{\tau^2\mathcal{G}((s_i,t_j)(s_k,t_l)) + \sigma^2\mathbb{1}_{(i,j)=(k,l)}\}_{(i,j),(k,l)} & \{\tau^2\mathcal{G}((s_i,t_j),\,\cdot\,)\}_{(i,j)} \\ \{\tau^2\mathcal{G}(\,\cdot\,,(s_k,t_l))\}_{(k,l)} & \tau^2\mathcal{G} \end{pmatrix}\right).$$

[1] Note that since $\mathscr{L}$ is positive definite, $\mathcal{G}$ is also positive definite and thus a valid covariance function as long as $\mathcal{G}$ does not contain any singularities.

It is well known that the unique maximum a posteriori estimate of x is given by $\mathrm{E}[\,x\,|\,y\,]$, which is a quantity that is easily identified from standard results of multivariate analysis.

$$\mathrm{E}[\,x\,|\,y\,] = \left\{\tau^2\mathcal{G}(\,\cdot\,,(s_k,t_l))\right\}_{(k,l)} \left\{\tau^2\mathcal{G}((s_i,t_j)(s_k,t_l)) + \sigma^2\mathbb{1}_{(i,j)=(k,l)}\right\}^{-1}_{(i,j),(k,l)} y. \tag{1.5}$$

Furthermore we find that

$$\mathrm{Var}[\,x\,|\,y\,] = \tau^2\mathcal{G} - \left\{\tau^2\mathcal{G}(\,\cdot\,,(s_k,t_l))\right\}_{(k,l)} \left\{\tau^2\mathcal{G}((s_i,t_j)(s_k,t_l)) + \sigma^2\mathbb{1}_{(i,j)=(k,l)}\right\}^{-1}_{(i,j),(k,l)} \cdot \left\{\tau^2\mathcal{G}((s_i,t_j),\,\cdot\,)\right\}_{(i,j)}.$$

Interestingly the Bayes estimator $\mathrm{E}[\,x\,|\,y\,]$ corresponds exactly to (1.3) with $\tau^2 = 1/2\lambda$, which means that the Bayesian a priori model (1.4) will produce estimates identical to those of the penalized likelihood approach if the smoothing parameter λ and the variance component τ^2 are related by the previously mentioned equation.

It is possible to write up the negative log-likelihood:

$$\begin{aligned}\ell_y(\sigma^2,\tau^2) &= \frac{1}{2}\log\det\Sigma + \frac{1}{2}\left\{y_{ij}\right\}^{\top}_{(i,j)}\Sigma^{-1}\left\{y_{ij}\right\}_{(i,j)} \\ &= \frac{1}{2}\log\det\Sigma + \frac{1}{2\sigma^2}\left\{y_{ij}\right\}^{\top}_{(i,j)}\left\{y_{ij} - \mathrm{E}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)},\end{aligned} \tag{1.6}$$

where

$$\Sigma = \left\{\tau^2\mathcal{G}((s_i,t_j)(s_k,t_l)) + \sigma^2\mathbb{1}_{(i,j)=(k,l)}\right\}_{(i,j),(k,l)}.$$

In principle this allows for maximum likelihood estimation of σ^2 and τ^2, but the matrix inversion and determinant computation will restrict the cases where this is possible, either because of the involved numerics or because of the computation time, to sparsely observed images.

The use of maximum likelihood estimation can also be problematic, because the model (1.4) may be in clear violation with some structure of the image, as mentioned earlier. In that case, even though the penalized likelihood and Bayes approach yields structurally identical estimates, the estimation of τ^2 may give a very different result compared to the estimate of $1/2\lambda$ by e.g. generalized cross-validation (Craven & Wahba 1978) or a similar technique, and thus give different estimates of the image.

An easy way to deal with the potential problems in applying this model is to apply it only to the residual image, where one has already modelled the expectation structure. We will propose to take the approach of Markussen (2013), where the expectation structure is modeled by low-dimensional crude representations of the image. This approach can be aided by standard algorithms developed in image analysis. This problem will be revisited in Chapter 7, and for now the model will be explored in its purity.

CHAPTER 2

GAUSSIAN FIELDS, GREEN'S FUNCTIONS AND SPLINES

In this chapter some relevant theory about Gaussian random fields on $[0,1] \times [0,1]$ will be developed. After this some details on Green's functions will be mentioned. In the following chapters these two subjects will be connected by specifying properties of a centered Gaussian random field x with covariance given by the Green's function $\mathcal{G}$ corresponding to some differential operator $\mathscr{L}$. Finally the model (1.1) will be revisited once more with a specific roughness penalty, and the usual thin plate spline solution, which is also connected to Green's functions, will be derived.

2.1 Gaussian fields

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. A measurable mapping $x : \Omega \to \mathbb{R}^{[0,1]\times[0,1]}$ is called a Gaussian random field or simply a Gaussian field, if its finite-dimensional distributions given by the vectors of the form $(x(s_1,t_1), \ldots, x(s_{n_1}, t_{n_2}))$ are multivariate Gaussian for all $n_1, n_2 \in \mathbb{N}$ and all $s_1, \ldots, s_{n_1}, t_1, \ldots, t_{n_2} \in [0,1]$. We will assume that all random fields we encounter are centered, i.e.

$$\mathrm{E}[x(s,t)] = 0$$

and have covariance function C:

$$C((s,t),(v,w)) = \mathrm{E}\left[x(s,t)x(v,w)\right].$$

With this definition, Kolmogorov's consistency theorem will guarantee that there exists a Gaussian field on $[0,1] \times [0,1]$ for any positive definite covariance function C.

Consider the model (1.4). As previously mentioned, there might be problems in using x as a prior for θ because x might have properties that do not match those of θ at all. Much of this section will be devoted to finding sufficient conditions for Gaussian fields to be bounded, continuous and differentiable, so that we are able to control the "look" of x.

Let us start with boundedness and continuity. Since we are working on a compact space, continuity will imply boundedness. We will therefore have to find conditions such that

$$P\left(\lim_{(v,w)\to(s,t)} |x(s,t) - x(v,w)| = 0,\ \forall s,t \in [0,1]\right) = 1,$$

which we will take as our definition of continuity of a Gaussian process x. The following theorem gives a sufficient condition for continuous sample paths.

Theorem 2.1. A centered Gaussian field x is continuous if for some positive c and $\alpha, \eta > 0$

$$C((s,t),(s,t)) + C((v,w),(v,w)) - 2C((s,t),(v,w)) \leq \frac{c}{|\log \|(s,t)-(v,w)\||^{1+\alpha}} \tag{2.1}$$

for all s,t with $\|(s,t)-(v,w)\| < \eta$.

Proof. This is a special case of Theorem 1.4.1 in Adler & Taylor (2007), where it has been used that

$$\mathrm{E}[|x(s,t) - x(v,w)|^2] = C((s,t),(s,t)) + C((v,w),(v,w)) - 2C((s,t),(v,w)).$$

□

What Theorem 2.1 tells us is that practically any Gaussian field with continuous covariance function is itself continuous. Consider the function

$$f(t) = \frac{1}{(-\log t)^{1+\alpha}}, \qquad t \in (0,1).$$

We see that

$$\lim_{t\to 0} f(t) = 0 \quad \text{and} \quad \lim_{t\to 1} f(t) = \infty,$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} f(t) = \frac{1+\alpha}{t(-\log(t))^{2+\alpha}} \to \infty \quad \text{as } t \to 0.$$

For $s_n \to s$ and $t_n \to t$ it seems that one should make an effort to find a continuous covariance function C where $C((s,t),(s_n,t_n))$ will converge to $C((s,t),(s,t))$ so slowly that (2.1) will not be met.

Now let us move to the question of differentiability. Define the Gaussian fields x_h^1 and x_h^2 by

$$x_h^1(s,t) = \frac{x(s+h,t) - x(s,t)}{h} \quad \text{or} \quad x_h^2(s,t) = \frac{x(s,t+h) - x(s,t)}{h}.$$

We say that x has L^2 partial derivative at (s,t) in respectively the s or t-directon if

$$\frac{\partial_{L_2}}{\partial s} x(s,t) = \lim_{h\to 0} x_h^1(s,t) \quad \text{and} \quad \frac{\partial_{L_2}}{\partial t} x(s,t) = \lim_{h\to 0} x_h^2(s,t)$$

exist in L^2. We call x L^2-differentiable if both partial derivatives exist in L^2 for all $(s,t) \in [0,1]\times[0,1]$. Since we are working with centered Gaussian fields, L^2 results can easily be expressed in terms of the covariance function C. The next theorem considers the question of continuously differentiable sample paths.

Theorem 2.2. A centered Gaussian field x with covariance function C for which

$$\frac{\partial^2}{\partial s \partial v} C((s,t),(v,w)) \quad \text{and} \quad \frac{\partial^2}{\partial t \partial w} C((s,t),(v,w))$$

exists for all $((s,t),(v,w))$, is continuously differentiable with probability one if there exists a positive c and $\rho, \delta, h_0 > 0$ such that for $\alpha_1, \alpha_2, \eta > 0$

$$\frac{\partial^2}{\partial s \partial v} \left(C((s,t),(s,t)) + C((v,w),(v,w)) - 2C((s,t),(v,w))\right) \le \frac{c}{|\log \|(s,t)-(v,w)\||^{1+\alpha_1}}$$

and

$$\frac{\partial^2}{\partial t \partial w} \left(C((s,t),(s,t)) + C((v,w),(v,w)) - 2C((s,t),(v,w))\right) \le \frac{c}{|\log \|(s,t)-(v,w)\||^{1+\alpha_2}}$$

for all $(s,t),(v,w)$ with $\|(s,t)-(v,w)\| < \eta$.

Proof. From the first assumption on C we get that x is L^2-differentiable. Since the characteristic function of a multivariate Gaussian with zero-mean and covariance C_n is

$$\phi(\theta) = \exp\left(-\frac{\theta^\top C_n \theta}{2}\right),$$

it follows from Lévy's continuity theorem that if $(X_n)_{n\ge 1}$ is an L^2 convergent sequence of multivariate, zero-mean Gaussians with covariance C_n, then the limit X must also be Gaussian with covariance $C^* = \lim_n C_n$. From this fact it is easily seen that $\frac{\partial}{\partial s} x(s,t)$ must be centered Gaussian with covariance $\frac{\partial^2}{\partial s \partial v} C((s,t),(v,w))$ and $\frac{\partial}{\partial t} x(s,t)$ is centered Gaussian with covariance $\frac{\partial^2}{\partial t \partial w} C((s,t),(v,w))$. The result now follows by using Theorem 2.1 on these two processes. □

2.2 Green's functions

Green's functions have already been introduced in Chapter 1 as integral kernels $\mathcal{G}$ that determine the inverse of the differential operator $\mathscr{L}$

$$\mathscr{L}^{-1}f(s,t) = \int_0^1 \int_0^1 \mathcal{G}((s,t),(v,w))f(v,w)\,\mathrm{d}v\,\mathrm{d}w. \tag{2.2}$$

It turns out that the proper definition of $\mathcal{G}$ is as the distribution on the space consisting of the infinitely differentiable functions in $\mathscr{H}$ (the space on which $\mathscr{L}$ is defined) for which the identity

$$\mathscr{L}\mathcal{G}((s,t),(v,w)) = \delta_{(s,t)}(v,w), \tag{2.3}$$

holds. This definition may require some fiddling where functions in $\mathscr{H}$ are thought of as though they were continuously embedded in a larger space, e.g. $\mathbb{R}^2$ or $[-\varepsilon, 1+\varepsilon]\times[-\varepsilon,1+\varepsilon]$ for $\varepsilon > 0$, in such a manner that they have compact support and disappear on the boundary of the larger space, but we will leave that as unimportant technicalities. It has already been established that the definition from Chapter 1 implies (2.3). If on the other hand $\mathcal{G}$ is a distribution for which (2.3) holds and f is an infinitely differentiable function

$$\mathscr{L}\langle \mathcal{G}((s,t),\cdot), f\rangle = \langle \mathscr{L}\mathcal{G}((s,t),\cdot), f\rangle = \langle \delta_{(s,t)}, f\rangle = f(s,t),$$

so the new definition makes sense. Whenever possible we will pretend that $\mathcal{G}$ is a usual function.

2.3 Generalized splines

Consider the problem of minimizing

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\theta(s_i,t_j)-y_{ij})^2 + \lambda\int_0^1\int_0^1\sum_{\ell=1}^{k}\binom{k}{\ell}|\partial_s^\ell\partial_t^{k-\ell}\theta(s,t)|^2\,\mathrm{d}s\,\mathrm{d}t \tag{2.4}$$

over θ. It turns out that the natural space of functions θ for this problem is the generalized *Beppo-Levi space*

$$\mathscr{H} = \{\theta \in \mathscr{D}'([0,1]\times[0,1]) \,|\, \partial_s^\ell\partial_t^{k-\ell}\theta \in L_2 \text{ for } \ell = 0,\ldots,k\}$$

where $\mathscr{D}'([0,1]\times[0,1])$ denotes the space of distributions (extended as above) on $[0,1]\times[0,1]$ and where the derivatives are understood in the distributional sense.

Now let $\{\phi_\nu\}_{1\le\nu\le M}$ denote the $M = \binom{k+1}{2}$ polynomials of total degree less than or equal to $k-1$, e.g. if $k=2$ the polynomials are

$$\phi_1(s,t) = 1, \qquad \phi_2(s,t) = s, \qquad \text{and} \qquad \phi_3(s,t) = t.$$

These polynomials are all infinitely smooth in terms of the roughness penalty (the last term in (2.4).) We can decompose $\mathscr{H}$ as follows

$$\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_1$$

where $\mathscr{H}_0$ is the M-dimensional function space with $\{\phi_\nu\}_{1\le\nu\le M}$ as basis and $\mathscr{H}_1$ consists of what is left (i.e. $\mathscr{H}_1 = \mathscr{H}/\mathscr{H}_0$, or alternatively, the orthogonal complement to $\mathscr{H}_0$ with respect to the inner product defined below.) Denote by P_1 the orthogonal projection from $\mathscr{H}$ onto $\mathscr{H}_1$. Remember that $(\mathbb{I}-P_1)\theta = 0$ if $\theta\in\mathscr{H}_1$. We now equip $\mathscr{H}$ with the inner product given by

$$\langle\theta,\vartheta\rangle = \sum_{i=1}^{M}(\mathbb{I}-P_1)\theta(v_i,w_i)(\mathbb{I}-P_1)\vartheta(v_i,w_i) + \sum_{\ell=1}^{k}\binom{k}{\ell}\int_0^1\int_0^1 \partial_s^\ell\partial_t^{k-\ell}\theta(s,t)\,\partial_s^\ell\partial_t^{k-\ell}\vartheta(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

The corresponding norm becomes

$$\|\theta\|^2 = \sum_{i=1}^{M}((\mathbb{I}-P_1)\theta)^2(v_i,w_i) + \sum_{\ell=1}^{k}\binom{k}{\ell}\int_0^1\int_0^1 \left|\partial_s^\ell \partial_t^{k-\ell}\theta(s,t)\right|^2 \,\mathrm{d}s\,\mathrm{d}t.$$

With this inner product $\mathscr{H}$ is a Hilbert space and $\mathscr{H}_0 \perp \mathscr{H}_1$.

In the following we will need that point evaluation of θ is a continuous linear functional. It turns out that this is the case when $k > 1$ (this follows from Theorem 1 of Meinguet (1979)) so from now on we will assume that $k > 1$. Because point evaluation is continuous, we know by the Riesz representation theorem that there exists representers $\eta_{st} \in \mathscr{H}$ such that

$$\theta(s,t) = \langle \eta_{st}, \theta\rangle.$$

We will denote the *reproducing kernel* of $\mathscr{H}$ by R:

$$R((s,t),(s',t')) = \langle \eta_{st}, \eta_{s't'}\rangle.$$

Note that

$$R((s,t),(s',t')) = R_0((s,t),(s',t')) + R_1((s,t),(s',t'))$$

where $R_0 = (\mathbb{I}-P_1)R$ and $R_1 = P_1 R$ are reproducing kernels on $\mathscr{H}_0$ and $\mathscr{H}_1$ respectively with the obvious choices of inner products.

Using the notation $\eta_{ij} = \eta_{s_i t_j}$ the minimization problem can be rewritten as

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(\langle \eta_{ij},\theta\rangle - y_{ij})^2 + \lambda\|P_1\theta\|^2. \tag{2.5}$$

And since $\mathscr{H}$ is a Hilbert space we can write

$$\theta = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}\xi_{ij} + \sum_{\nu=1}^{M} d_\nu \phi_\nu + \rho$$

where $\xi_{ij} = P_1\eta_{ij}$ and $\rho \in \mathscr{H}_1$ is perpendicular to all ξ_{ij}'s and ϕ_ν's. Substituting this representation into (2.5), and using that ρ is perpendicular to η_{ij} as well as the fact $P_1\phi_\nu = 0$ for all ν gives the minimiziation problem

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(\left\langle \eta_{ij}, \sum_{i'=1}^{n_1}\sum_{j'=1}^{n_2} c_{i'j'}\xi_{i'j'}\right\rangle + \left\langle \eta_{ij}, \sum_{\nu=1}^{M} d_\nu\phi_\nu\right\rangle - y_{ij}\right)^2 + \lambda\left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}\xi_{ij} + \rho\right\|^2. \tag{2.6}$$

It is seen that only the roughness penalty part depends on ρ. Since ρ is perpendicular to all ξ_{ij}'s it follows that ρ must be 0 in the minimization problem.

Clearly $\langle(\mathbb{I}-P_1)\eta_{ij},\xi_{i'j'}\rangle = 0$, so $\langle\eta_{ij},\xi_{i'j'}\rangle = \langle\xi_{ij},\xi_{i'j'}\rangle$, so we can rewrite the minimization even further as

$$\frac{1}{2\sigma^2}\left\|y - (H\boldsymbol{c} + T\boldsymbol{d})\right\|_{n_1n_2}^2 + \lambda\left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}\xi_{ij}\right\|^2$$

where

$$H = \left\{\langle\xi_{ij},\xi_{i'j'}\rangle\right\}_{(i,j),(i',j')} \quad \text{and} \quad T = \left\{\phi_\nu(s_i,t_j)\right\}_{(i,j),\nu}$$

and $\boldsymbol{c} = \{c_1,\ldots,c_{n_1n_2}\}$, $\boldsymbol{d} = \{d_1,\ldots,d_M\}$. Now since $\xi_{ij}(s,t) = R_1((s_i,t_j),(s,t))$ we have that

$$\lambda\left\|\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}\xi_{ij}\right\|^2 = \lambda\boldsymbol{c}^\top H\boldsymbol{c},$$

and the problem is reduced to

$$\frac{1}{2\sigma^2}\left\|y-(H\boldsymbol{c}+T\boldsymbol{d})\right\|_{n_1n_2}^2+\lambda\boldsymbol{c}^\top H\boldsymbol{c}.$$

If we assume that all observations do not lie on a straight line, the above problem is solved by setting

$$\boldsymbol{c}=(H+2\sigma^2\lambda I)^{-1}(I-T(T^\top(H+2\sigma^2\lambda I)^{-1}T)^{-1}T^\top(H+2\sigma^2\lambda I)^{-1})y$$

and

$$\boldsymbol{d}=(T^\top(H+2\sigma^2\lambda I)^{-1}T)^{-1}T^\top(H+2\sigma^2\lambda I)^{-1}y.$$

This can be written more elegantly as

$$(H+2\sigma^2\lambda I))^{-1}\boldsymbol{c}+T\boldsymbol{d}=y,\qquad\text{and}\qquad T^\top\boldsymbol{c}=0.$$

We have hereby shown that the minimizer of (2.4) is on the form

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}c_{ij}\xi_{ij}+\sum_{\nu=1}^{M}d_\nu\phi_\nu \tag{2.7}$$

where the coeffecients $\boldsymbol{c}$, $\boldsymbol{d}$ and polynomials ϕ are known. This function is often refered to as a generalized spline. All there is left is to determine the representers ξ_{ij}. I.e. we need to find a distribution ξ_{ij} such that for $\theta\in\mathscr{H}_1$

$$\theta(v,w)=\langle\xi_{vw},\theta\rangle=\sum_{\ell=1}^{k}\binom{k}{\ell}\int_0^1\int_0^1\partial_s^\ell\partial_t^{k-\ell}\xi_{vw}(s,t)\,\partial_s^\ell\partial_t^{k-\ell}\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

If we impose boundary conditions[1] on $\mathscr{H}$ such that the boundary terms arising from integration by parts vanishes, the above integral is equal to

$$\int_0^1\int_0^1\theta(s,t)(\partial_s^2+\partial_t^2)^k\xi_{vw}(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

We get the identity

$$(\partial_s^2+\partial_t^2)^k\xi_{vw}(s,t)=\delta_{(v,w)}(s,t).$$

Comparing this with (2.3) we see that ξ_{ij} must be the Green's function corresponding to $(\partial_s^2+\partial_t^2)^k$.

[1] Another alternative which is used in Wahba & Wendelberger (1980) is to extend the area on which we integrate to $\mathbb{R}^2$.

CHAPTER 3

CALCULATIONS AND EXAMPLES

A Green's function naturally appear in (1.3), but apart from the mathematical tricks that lead us to a similar expression for the estimate in the model (1.4), one could rightfully ask if Green's functions give rise to natural covariance structures for Gaussian processes. The next example is encouraging in the sense that it shows that the most basic Gaussian fields corresponds to the simplest choices of penalty differential operator $\mathscr{K}$.

Example 3.1. Markussen (2013) shows that in one dimension the simplest possible penalty (corresponding to setting $\mathscr{K}$ equal to the first derivative) produces the covariance of Brownian motion or the Brownian bridge depending on the boundary conditions. A similar result holds in two dimensions.

The table below holds some of the most basic bivariate Gaussian fields and their corresponding covariance functions as well as some boundary conditions the covariance functions meet.

Process	$C((s,t),(v,w))$	Boundary conditions
Brownian sheet	$(s \wedge v)(t \wedge w)$	$C((0,t),(v,w)) = 0$ $C((s,0),(v,w)) = 0$ $\partial_s C((1,t),(v,w)) = 0$ $\partial_t C((s,1),(v,w)) = 0$
Brownian bridge	$(s \wedge v)(t \wedge w) - svtw$	$C((0,t),(v,w)) = 0$ $C((s,0),(v,w)) = 0$ $C((1,1),(v,w)) = 0$ $\partial_s \partial_t^2 C((1,t),(v,w)) = 0$ $\partial_s^2 \partial_t C((s,1),(v,w)) = 0$
Tied down Brownian bridge	$(s \wedge v - sv)(t \wedge w - tw)$	$C((0,t),(v,w)) = 0$ $C((1,t),(v,w)) = 0$ $C((s,0),(v,w)) = 0$ $C((s,1),(v,w)) = 0$

Observe that

$$\partial_s(s \wedge v) = 1_{(0,v)}(s) \quad \text{and} \quad \partial_s 1_{(0,v)}(s) = \delta_v(s) = \delta_s(v).$$

If we let $\mathscr{K} = \partial_s \partial_t$, then it holds for any of the boundary conditions in the above table that $\mathscr{L} = \mathscr{K}^\dagger \mathscr{K} = \partial_s^2 \partial_t^2$. It is easily obtained that

$$\mathscr{L}(s \wedge v)(t \wedge w) = \mathscr{L}\left((s \wedge v)(t \wedge w) - svtw\right) = \mathscr{L}(s \wedge v - sv)(t \wedge w - tw) = \delta_{(s,t)}(v,w),$$

and so each of the covariance functions corresponds to the Green's function of $\mathscr{L}$ on the space of continuously differentiable functions subject to the boundary conditions given in the above table.

For these fields continuity and differentiability properties are well-known. They are all continuous and neither are differentiable. Using a higher order differential operator $\mathscr{K}$ should hopefully produce smoother fields. The relation between the order of $\mathscr{K}$ and smoothness of the corresponding Gaussian field will be explored further in Example 3.9. ∘

The rest of this chapter will focus on examples of the model (1.4). First we will derive some Green's functions and later they will be considered in a concrete setup. We will mainly focus on penalties involving the the negative Laplace operator $-\Delta$, where

$$\Delta u(s,t) = \frac{\partial^2}{\partial s^2} u(s,t) + \frac{\partial^2}{\partial t^2} u(s,t).$$

The simplest case is $\mathscr{L} = -\Delta$, but it turns out that this model has some undesirable properties and more general Laplacian penalties will be considered. First an interpretation of a Laplacian penalty will be given. This is followed by a number of computations of Green's functions. In connection with these computations, different boundary conditions will be imposed on $\mathscr{H}$, and techniques for finding Green's functions will be explored.

3.1 The Laplace operator

Proceeding with the notation from the previous chapters, ∂_s and ∂_t will denote the partial differential operators with respect to the first and second coordinate.

The following example gives an interpretation of the roughness measure when we use the negative Laplacian as our smoothing operator $\mathscr{L}$.

Example 3.2. Assume that $\mathscr{H}$ is the subspace of $\mathcal{C}^2([0,1] \times [0,1], \mathbb{R})$ which is restricted to those functions θ that meet the boundary conditions

$$\theta(0,t) = \theta(1,t) = \theta(s,0) = \theta(s,1) = 0. \tag{3.1}$$

These types of boundary conditions are called *Dirichlet boundary conditions.* For $\theta \in \mathscr{H}$ it holds that

$$\begin{aligned} -\int_0^1 \int_0^1 \theta(s,t)\Delta\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t &= -\int_0^1 \int_0^1 \theta(s,t)(\partial_s^2 + \partial_t^2)\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t \\ &= -\int_0^1 \int_0^1 \theta(s,t)\partial_s^2\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t - \int_0^1 \int_0^1 \theta(s,t)\partial_t^2\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_0^1 \int_0^1 \left|\partial_s\theta(s,t)\right|^2 \mathrm{d}s\,\mathrm{d}t + \int_0^1 \int_0^1 \left|\partial_t\theta(s,t)\right|^2 \mathrm{d}s\,\mathrm{d}t, \end{aligned} \tag{3.2}$$

where integration by parts was used for the last equality. This means that one can think of $\mathscr{L} = -\Delta$ as the operator that measures roughness in the s-direction and t-direction independently. I.e. $\mathscr{L} = \mathscr{K}_1^\dagger \mathscr{K}_1 + \mathscr{K}_2^\dagger \mathscr{K}_2$, where $\mathscr{K}_1 = \partial_s$ and $\mathscr{K}_2 = \partial_t$.

Another interpretation is possible. Consider for $\theta \in \mathscr{H}$ the operation

$$\max_{\varphi \in [0,2\pi]} \left|\partial_\varphi \theta(s,t)\right|^2 = \max_{\varphi \in [0,2\pi]} \left|\cos(\varphi)\partial_s\theta(s,t) + \sin(\varphi)\partial_t\theta(s,t)\right|^2.$$

It is easily verified that the maximizer of this expression is $\varphi = \arctan\left(\frac{\partial_t\theta(s,t)}{\partial_s\theta(s,t)}\right)$, and thus it can be

written as

$$\begin{aligned}
&\left|\cos\left(\arctan\left(\frac{\partial_t\theta(s,t)}{\partial_s\theta(s,t)}\right)\right)\partial_s\theta(s,t)+\sin\left(\arctan\left(\frac{\partial_t\theta(s,t)}{\partial_s\theta(s,t)}\right)\right)\partial_t\theta(s,t)\right|^2\\
&=\left|\frac{1}{\sqrt{1+\left(\frac{\partial_t\theta(s,t)}{\partial_s\theta(s,t)}\right)^2}}\partial_s\theta(s,t)+\frac{\frac{\partial_t\theta(s,t)}{\partial_s\theta(s,t)}}{\sqrt{1+\left(\frac{\partial_t\theta(s,t)}{\partial_s\theta(s,t)}\right)^2}}\partial_t\theta(s,t)\right|^2\\
&=\left|\frac{|\partial_s\theta(s,t)|^2+|\partial_t\theta(s,t)|^2}{\sqrt{|\partial_s\theta(s,t)|^2+|\partial_t\theta(s,t)|^2}}\right|^2\\
&=|\partial_s\theta(s,t)|^2+|\partial_t\theta(s,t)|^2.
\end{aligned}$$

Comparting this with (3.2), one realise that

$$\max_{\varphi\in[0,2\pi]}\partial_\varphi\theta(s,t),$$

is a weak square root of Δ, which means that the penalty given by Δ is the first order roughness of θ in the direction in which it is maximal. By (3.2) it is easily verified that $-\Delta$ is positive definite on $\mathscr{H}$, so it is known that $\mathscr{K}=(-\Delta)^{1/2}$ exists, but the above operator is not linear, so it does not correspond to $\mathscr{K}$.

∘

Note that the previous example also works if we imposed *Neumann boundary conditions*

$$\partial_s\theta(0,t)=\partial_s\theta(1,t)=\partial_t\theta(s,0)=\partial_t\theta(s,1)=0$$

even though these only make $-\Delta$ positive semidefinite.

3.2 Green's functions in practice

A pleasing interpretation of using $\mathscr{L}=-\Delta$ for measuring the roughness of an image has been found in the cases of Dirichlet and Neumann boundary conditions. In this section and the next chapter, examples related to the model (1.4) will be explored. It can be a great challenge to find suitable covariance functions and verify that they are in fact positive definite. The way things are presented here, all there is to do is to show that $\mathscr{L}$ is positive definite, which automatically ensures that $\mathcal{G}$ is also positive definite, and to actually calculate $\mathcal{G}$, which is in general very hard to do. For the operators we will consider, it turns out that $\mathcal{G}$ can always be constructed from eigenfunction expansions. The first example considers the simplest possible case.

Example 3.3. The Green's function $\mathcal{G}$ corresponding to $-\Delta$ will be computed. Assume that $\mathscr{H}$ is the space of twice continuously differentiable functions u subject to the Dirichlet boundary conditions

$$u(0,t)=u(1,t)=u(s,0)=u(s,1)=0.$$

The Green's function can be derived by finding a solution to Poisson's equation

$$-\Delta u(s,t)=f(s,t),\qquad u\in\mathscr{H}$$

with the structure (cf. the simple definition (2.2))

$$u(s,t)=\int_0^1\int_0^1\mathcal{G}((s,t),(v,w))f(v,w)\,\mathrm{d}(v,w).$$

One easily gets the idea of considering trigonometric functions for the solution. The function

$$\sin(n\pi s)\sin(m\pi t), \qquad n, m \in \mathbb{N}$$

satisfies the boundary conditions. Furthermore it holds that

$$-\Delta \sin(n\pi s)\sin(m\pi t) = \pi^2(n^2+m^2)\sin(n\pi s)\sin(m\pi t),$$

so this function is an *eigenfunction* of $-\Delta$. Thus if one puts

$$u(s,t) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} c_{nm}\sin(n\pi s)\sin(m\pi t),$$

and assumes that the double series and $-\Delta$ can be interchanged, it holds that

$$-\Delta u(s,t) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} c_{nm}\pi^2(n^2+m^2)\sin(n\pi s)\sin(m\pi t).$$

To solve Poisson's equation we will need that $-\Delta u(s,t) = f(s,t)$. One can multiply both sides of this equation by $\sin(k\pi s)\sin(l\pi t)$ and integrate to get that

$$\int_0^1\int_0^1 \sin(k\pi s)\sin(l\pi t) f(s,t)\,\mathrm{d}s\,\mathrm{d}t = c_{kl}\pi^2(k^2+l^2)\,\frac{1}{4}.$$

From this the constants c_{nm} can be determined to be

$$c_{nm} = \frac{4}{\pi^2(n^2+m^2)}\int_0^1\int_0^1 \sin(n\pi v)\sin(m\pi w) f(v,w)\,\mathrm{d}v\,\mathrm{d}w.$$

The question is now if this choice of c_{nm} will make $-\Delta u$ convergent and if the limit is in fact f. But since f is bounded it clearly holds that $f \in L^2([0,1]\times[0,1],\mathbb{R})$, and because the double sine terms make up a complete orthonormal sequence in this space, it is known that $-\Delta u = f$. Therefore the solution to Poisson's equation subject to the boundary conditions (3.1) is given by

$$u(s,t) = \int_0^1\int_0^1 \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{4}{\pi^2(n^2+m^2)}\sin(n\pi v)\sin(m\pi w)\sin(n\pi s)\sin(m\pi t) f(v,w)\,\mathrm{d}v\,\mathrm{d}w,$$

and thus the Green's function becomes

$$\mathcal{G}((s,t),(v,w)) = \frac{4}{\pi^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{n^2+m^2}\sin(n\pi v)\sin(m\pi w)\sin(n\pi s)\sin(m\pi t). \tag{3.3}$$

Unfortunately this function has an undesirable property, namely that

$$\mathcal{G}((s,t),(s,t)) = +\infty,$$

for all s and t. For the details see Example 3.6. Using this Green's function as a covariance kernel seems problematic, even though the process might be well defined, and similar covariance kernels are considered for *generalized* Gaussian fields. An illustration of the shape of $\mathcal{G}$ can be found in Figure 3.1.

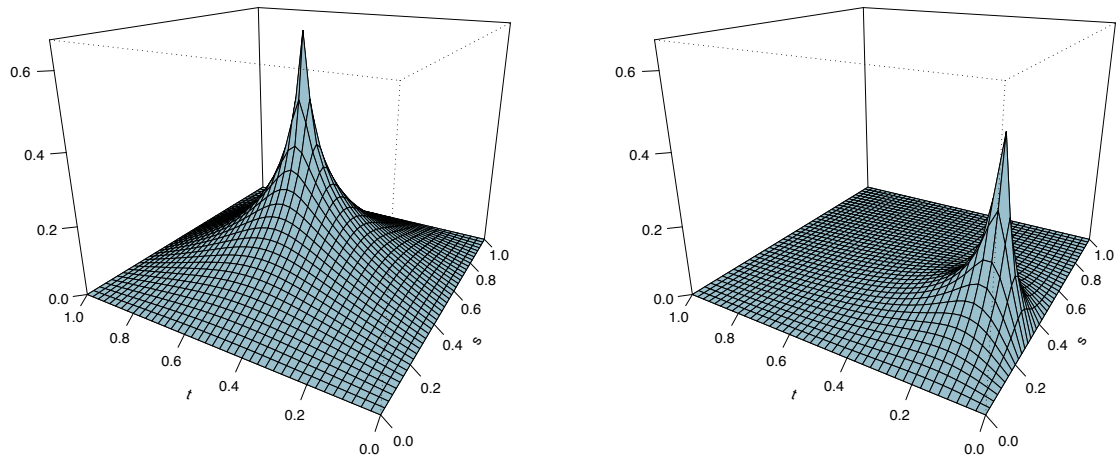

Figure 3.1: Plots of $\mathcal{G}(\,\cdot\,,(v,w))$ corresponding to $-\Delta$ with Dirichlet boundary condtions. (v,w) has been set to (0.5, 0.5) in the left plot and (0.3, 0.1) in the right. The sums in (3.3) were both cut off at their 40th term.

∘

Example 3.4. Consider again the problem of finding the Green's function corresponding to $-\Delta$, but this time subject to the Neumann boundary condtions given by

$$\partial_s u(0,t) = \partial_s u(1,t) = \partial_t u(s,0) = \partial_t u(s,1) = 0,$$

for all $u \in \mathscr{H}$. Let us try to use the definition (2.3) to identify $\mathcal{G}$. Taking a hint from the previous example, one could make a guess that $\mathcal{G}$ had the form

$$\mathcal{G}((s,t),(v,w)) = \frac{4}{\pi^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{n^2+m^2}\cos(n\pi s)\cos(m\pi t)\cos(n\pi v)\cos(m\pi w), \tag{3.4}$$

which clearly satisfies the boundary conditions, and it is easily verified that

$$-\Delta\mathcal{G}((s,t),(v,w)) = 4\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\cos(n\pi s)\cos(m\pi t)\cos(n\pi v)\cos(m\pi w) = \delta_{(v,w)}(s,t),$$

where the derivative is understood in the weak sense. It is clear that this Green's function also suffers from the problem of being singular when $s = v$ and $t = w$, as it basically has same properties as the corresponding Green's function for the Dirichlet problem. There is also another problem however, $-\Delta$ is not neccesaraly positive definite on $\mathscr{H}$, since any constant function c is in $\mathscr{H}$ and $-\Delta c = 0$. Because of this, $\mathcal{G}$ is not a true inverse, but only a right inverse. This means that $\mathcal{G}$ might in fact be positive definite, but it is not easily verified, neither does it seem very likely. The rest of this example will consider how to deal with the problem caused by the lack of positive definiteness.

One solution to the problem could be to impose the extra condition

$$u(0,0) = 0$$

on $\mathscr{H}$ (the choice of $(s,t) = (0,0)$ is arbitrary.) This would ensure that $-\Delta$ became positive definite, but the corresponding Green's function is very hard to compute explicitly.

Another soulution could be to add a positive definite differential operator to $-\Delta$. For example one

could consider $-\Delta + \varepsilon\mathbb{I}$ with $\varepsilon > 0$, which is clearly positive definite. To find the corresponding Green's function consider the equation

$$(-\Delta + \varepsilon\mathbb{I})u(s,t) = f(s,t) \tag{3.5}$$

which is a *Helmholtz* type equation.[1] It is clear that this equation has the same eigenfunctions as the corresponding Poisson equation, only now

$$(-\Delta + \varepsilon\mathbb{I})\cos(n\pi s)\cos(m\pi t) = (\pi^2 n^2 + \pi^2 m^2 + \varepsilon)\cos(n\pi s)\cos(m\pi t),$$

and with an operation identical to that of Example 3.3, it is obtained that the Green's function is given by

$$\mathcal{G}_\varepsilon((s,t),(v,w)) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{4}{\pi^2 n^2 + \pi^2 m^2 + \varepsilon}\cos(n\pi s)\cos(m\pi t)\cos(n\pi v)\cos(m\pi w) \tag{3.6}$$

which is positive definite. Choosing ε close to 0 will give make $\mathcal{G}_\varepsilon \approx \mathcal{G}$ and for all practial purposes, one can consider $\mathcal{G}$ to be positive definite. The shape of $\mathcal{G}$, which is identical to the shape of $\mathcal{G}_\varepsilon$ for ε small, can be found in Figure 3.2.

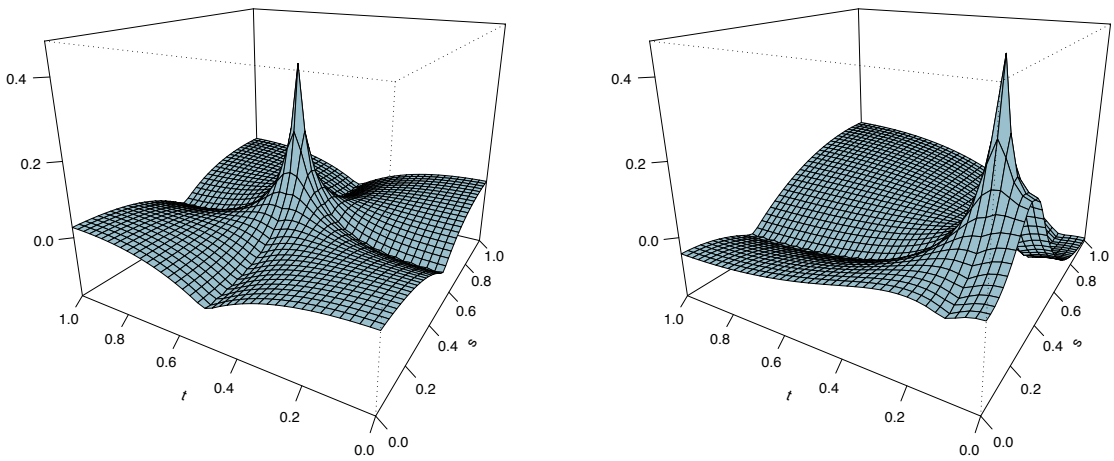

Figure 3.2: Plots of $\mathcal{G}(\,\cdot\,,(v,w))$ corresponding to $-\Delta$ with Neumann boundary condtions. (v,w) has been set to (0.5, 0.5) in the left plot and (0.3, 0.1) in the right. The sums in (3.4) were both cut off at their 40th term.

○

Before we move on to other examples, we will go a little in depth with the Green's function (3.3) exploring both the representation and singularity a little further. Even though this Green's function is not of great practical use, the method of the next example is also very useful for other examples, but much easier to understand in this simple setting.

[1] The equation (3.5) is perhaps more correctly referred to as the *screened Poisson equation* or the *time-independent Klein-Gordon equation*, nevertheless we refer to it as the Helmholtz equation.

Example 3.5. Consider again the function

$$\mathcal{G}((s,t),(v,w)) = \frac{4}{\pi^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{n^2+m^2}\sin(n\pi v)\sin(m\pi w)\sin(n\pi s)\sin(m\pi t).$$

The double sum representation converges very slowly to $\mathcal{G}$, which unfortunately makes the representation computationally intensive – this is a great problem because one typically has to evaluate these kinds of functions millions of times. Another representation with better convergence properties is available, however. If we start by only focussing on the two first boundary conditions

$$u(0,t) = u(1,t) = 0,$$

one could propose to write the Green's function on the form

$$\mathcal{G}((s,t),(v,w)) = \sum_{n=1}^{\infty} G_n(t,w)\sin(n\pi s)\sin(n\pi v),$$

which we already know works. It holds that

$$-\Delta\mathcal{G}((s,t),(v,w)) = \delta_{(s,t)}(v,w),$$

and multiplying both sides with $4\sin(k\pi s)\sin(k\pi v)$ and integrating over s and v, it is obtained that

$$k^2\pi^2 G_k(t,w) - \frac{\partial^2}{\partial t^2}G_k(t,w) = 2\delta_t(w),$$

which means that the expansion coefficients G_n are themselves Green's functions for the above ordinary differential equation, with the boundary conditions for solutions g that

$$g(0) = g(1) = 0.$$

With some calculations it can be obtained that

$$G_n(t,w) = \frac{2\sinh(n\pi(1-(t\vee w)))\sinh(n\pi(t\wedge w))}{n\pi\sinh(n\pi)},$$

which gives the alternative expression for the Green's function (3.3)

$$\mathcal{G}((s,t),(v,w)) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sinh(n\pi(1-(t\vee w)))\sinh(n\pi(t\wedge w))}{n\sinh(n\pi)}\sin(n\pi s)\sin(n\pi v), \tag{3.7}$$

which has somewhat better convergence properties than the previous found expression. ∘

We have derived an alternative expression of the Green's function for Poisson's equation with Dirichlet boundary conditions. This expression has better computational properties, but the nature of the singularity is still not quite clear from the form of the function. This next example goes into the exact nature of the singularity, and in addition we arrive at an expression for $\mathcal{G}$ with convergence properties that are very impressive. It is quite an unaesthetic example, and the reader should not feel forced to go through it, but is encouraged to look at the new expression for the Green's function (3.9).

Example 3.6. We will use a method of Melnikov (1999)[2] to get a second alternative expression for the Green's function (3.3), we start out with the single-series representation (3.7). We can rewrite the G_n

[2] Originally introduced for the Laplacian in Melnikov, Yu. A., *Green's Functions in Applied Mechanics*, WIT Press 1995.

terms as

$$\begin{aligned}\pi n G_n(t,w) &= \frac{\left(e^{n\pi(1-(t\vee w))} - e^{-n\pi(1-(t\vee w))}\right)\left(e^{n\pi(t\wedge w)} - e^{-n\pi(t\wedge w)}\right)}{e^{n\pi} - e^{-n\pi}} \\ &= \frac{e^{n\pi(1-(t\vee w)+(t\wedge w)))} - e^{n\pi(1-(t\vee w)-(t\wedge w))} - e^{n\pi((t\vee w)-1+(t\wedge w))} + e^{n\pi((t\vee w)-1-(t\wedge w))}}{e^{n\pi} - e^{-n\pi}} \\ &= \frac{e^{n\pi}\left(e^{-n\pi|t-w|} - e^{-n\pi(t+w)}\right) - e^{-n\pi}\left(e^{n\pi(t+w)} - e^{n\pi|t-w|}\right)}{e^{n\pi} - e^{-n\pi}},\end{aligned}$$

which we can split up as

$$\begin{aligned}\pi n \mathcal{G}_n(t,w) &= e^{-n\pi|t-w|} - e^{-n\pi(t+w)} - \frac{e^{n\pi(t+w)} - e^{n\pi|t-w|} - e^{-n\pi|t-w|} + e^{-n\pi(t+w)}}{2e^{n\pi}\sinh(n\pi)} \\ &= e^{-n\pi|t-w|} - e^{-n\pi(t+w)} - \frac{2\sinh(n\pi t)\sinh(n\pi w)}{2e^{n\pi}\sinh(n\pi)}.\end{aligned}$$

From this expansion, we can write $\mathcal{G}$ as

$$\begin{aligned}\mathcal{G}((s,t),(v,w)) = \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\exp(-n\pi|t-w|) - \exp(-n\pi(t+w))}{n}\sin(n\pi s)\sin(n\pi v) \\ -\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sinh(n\pi t)\sinh(n\pi w)}{ne^{n\pi}\sinh(n\pi)}\sin(n\pi s)\sin(n\pi v).\end{aligned}$$

From the sine product-to-sum identity the first term can be rewritten as

$$\frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{\exp(-n\pi|t-w|) - \exp(-n\pi(t+w))}{n}\left(\cos(n\pi(s-v)) - \cos(n\pi(s+v))\right).$$

Remembering that for $|q| < 1$ it holds that

$$\sum_{n=1}^{\infty}\frac{q^n}{n}\cos(n\gamma) = -\log\sqrt{1-2q\cos(\gamma)+q^2} = -\log\sqrt{(1-qe^{i\gamma})(1-qe^{-i\gamma})}, \tag{3.8}$$

we get that the first term can be rewritten as

$$\begin{aligned}\frac{1}{2\pi}\log\Bigg(\sqrt{\frac{(1-e^{-\pi(t+w)}e^{\pi i(s-v)})(1-e^{-\pi(t+w)}e^{-\pi i(s-v)})}{(1-e^{-\pi|t-w|}e^{\pi i(s-v)})(1-e^{-\pi|t-w|}e^{-\pi i(s-v)})}} \\ \times\sqrt{\frac{(1-e^{-\pi|t-w|}e^{\pi i(s+v)})(1-e^{-\pi|t-w|}e^{-\pi i(s+v)})}{(1-e^{-\pi(t+w)}e^{\pi i(s+v)})(1-e^{-\pi(t+w)}e^{-\pi i(s+v)})}}\Bigg),\end{aligned}$$

which if L denotes the function

$$L(x+iy) = |e^{\pi(x+iy)} - 1| = \sqrt{(e^{\pi(x+iy)}-1)(e^{\pi(x-iy)}-1)},$$

where $x, y \in \mathbb{R}$, can be reduced to

$$\frac{1}{2\pi}\log\left(\frac{L(z+\bar{\varsigma})L(z-\bar{\varsigma})}{L(z-\varsigma)L(z+\varsigma)}\right)$$

with $z = t + is$ and $\varsigma = w + iv$. We have obtained a new expression for the Green's function

$$\mathcal{G}((s,t),(v,w)) = \frac{1}{2\pi}\log\left(\frac{L(z+\bar{\varsigma})L(z-\bar{\varsigma})}{L(z-\varsigma)L(z+\varsigma)}\right) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sinh(n\pi t)\sinh(n\pi w)}{ne^{n\pi}\sinh(n\pi)}\sin(n\pi s)\sin(n\pi v)$$

which clearly shows the logarithmic singularity when $s = v$ and $t = w$ in the first term. This representation has much better convergence properties than (3.7), but there is still one slight problem in terms of convergence of the series. It can be realized by rewriting the coeffecient

$$\begin{aligned}\frac{\sinh(n\pi t)\sinh(n\pi w)}{ne^{n\pi}\sinh(n\pi)} &= \frac{\cosh(n\pi(t+w)) - \cosh(n\pi(t-w))}{2ne^{n\pi}\sinh(n\pi)} \\ &= \frac{\cosh(n\pi(t+w))}{n(e^{2n\pi}-1)} - \frac{\cosh(n\pi(t-w))}{n(e^{2n\pi}-1)}.\end{aligned}$$

When $t + w \approx 2$ the last term is negligible and

$$\cosh(n\pi(t+w)) = \frac{e^{n\pi(t+w)} - e^{-n\pi(t+w)}}{2} \approx \frac{e^{2n\pi}}{2}$$

and so

$$\frac{\cosh(n\pi(t+w))}{n(e^{2n\pi}-1)} \approx \frac{1}{n},$$

which makes the convergence of the series slow when $t + w \approx 2$, luckily we can do something about it. Write

$$\begin{aligned}\frac{\sinh(n\pi t)\sinh(n\pi w)}{ne^{n\pi}\sinh(n\pi)} &= \frac{\sinh(n\pi t)\sinh(n\pi w)}{ne^{n\pi}\sinh(n\pi)} - \frac{\cosh(n\pi(t+w))}{ne^{2n\pi}} + \frac{\cosh(n\pi(t+w))}{ne^{2n\pi}} \\ &= S_n(s,v) + \frac{\cosh(n\pi(t+w)))}{ne^{2n\pi}},\end{aligned}$$

where S_n can be rewritten as

$$\begin{aligned}S_n(s.v) &= \frac{\sinh(n\pi t)\sinh(n\pi w)}{ne^{n\pi}\sinh(n\pi)} - \frac{\cosh(n\pi(t+w))}{ne^{2n\pi}} \\ &= \frac{e^{n\pi}\sinh(n\pi t)\sinh(n\pi w) - \cosh(n\pi(t+w))\sinh(n\pi)}{ne^{2n\pi}\sinh(n\pi)} \\ &= \frac{e^{n\pi t}\sinh(n\pi(w-1)) - e^{-n\pi t}\sinh(n\pi(w+1))}{2ne^{2n\pi}\sinh(n\pi)}.\end{aligned}$$

It is obvious that the series

$$-\frac{2}{\pi}\sum_{n=1}^{\infty} S_n(t,w)\sin(n\pi s)\sin(n\pi v)$$

converges uniformly, so now we only need to compute the last part of the series

$$\begin{aligned}-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\cosh(n\pi(t+w))}{ne^{2n\pi}}\sin(n\pi s)\sin(n\pi v) &= -\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{e^{n\pi(t+w)} + e^{-n\pi(t+w)}}{ne^{2n\pi}}\sin(n\pi s)\sin(n\pi v) \\ &= -\frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{e^{n\pi(t+w-2)}}{n}(\cos(n\pi(s-v)) - \cos(n\pi(s+v))) \\ &\quad -\frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{e^{-n\pi(t+w+2)}}{n}(\cos(n\pi(s-v)) - \cos(n\pi(s+v))).\end{aligned}$$

using (3.8), we get that this equals

$$
\begin{aligned}
-\frac{1}{2\pi}\Bigg(&-\log\sqrt{(1-e^{\pi(t+w-2)}e^{in\pi(s-v)})(1-e^{\pi(t+w-2)}e^{-in\pi(s-v)})} \\
&+\log\sqrt{(1-e^{\pi(t+w-2)}e^{in\pi(s+v)})(1-e^{\pi(t+w-2)}e^{-in\pi(s+v)})} \\
&-\log\sqrt{(1-e^{-\pi(t+w+2)}e^{in\pi(s-v)})(1-e^{-\pi(t+w+2)}e^{-in\pi(s-v)})} \\
&+\log\sqrt{(1-e^{-\pi(t+w+2)}e^{in\pi(s+v)})(1-e^{-\pi(t+w+2)}e^{-in\pi(s+v)})}\Bigg)
\end{aligned}
$$

which we can write in terms of the L-function as

$$
\frac{1}{2\pi}\log\left(\frac{L(z_1+\bar{\varsigma}_1)L(z_2+\bar{\varsigma}_2)}{L(z_1+\varsigma_1)L(z_2+\varsigma_2)}\right)
$$

where $z_1 = t+1+is$, $z_2 = t-1+is$, $\varsigma_1 = w+1+iv$ and $\varsigma_2 = w-1+iv$. This leads to the final form of the Green's function:

$$
\begin{aligned}
\mathcal{G}((s,t),(v,w)) = \frac{1}{2\pi}\log&\left(\frac{L(z+\bar{\varsigma})L(z-\bar{\varsigma})L(z_1+\bar{\varsigma}_1)L(z_2+\bar{\varsigma}_2)}{L(z-\varsigma)L(z+\varsigma)L(z_1+\varsigma_1)L(z_2+\varsigma_2)}\right) \\
&-\frac{2}{\pi}\sum_{n=1}^{\infty} S_n(t,w)\sin(n\pi s)\sin(n\pi v),
\end{aligned}
\tag{3.9}
$$

which is far superior to any of the previous expressions in terms of convergence properties. ○

Because of the singularity in the Green's functions for $-\Delta$, another differential operator for smoothing is needed. For this one could consider the differential operator $\Delta\Delta$, which gives rise to the biharmonic equation

$$
\Delta\Delta u(s,t) = f(s,t). \tag{3.10}
$$

Since

$$
\Delta\Delta = \partial_s^4 + \partial_t^4 + 2\partial_s^2\partial_t^2,
$$

for suitable boundary conditions one would expect that the corresponding Green's function would at least result in continuous samples, since the last term corresponds to the penalty operator for the Brownian fields, and the derivatives in both directions are even further penalized. Again this will be considered in detail in Example 3.9, but before that we will consider the *ℓ-harmonic* or *iterated Laplacian* equation

$$
(-\Delta)^{\ell} u(s,t) = f(s,t), \tag{3.11}
$$

where a higher value of ℓ corresponds to a smoother model. In the following ℓ will be referred to as the *order of smoothing* and it will be assumed that ℓ is an integer greater than 1.

Example 3.7. Neither Dirichlet nor Neumann boundary conditions will make $(-\Delta)^{\ell}$ positive definite when $\ell > 2$. As before $\varepsilon\mathbb{I}$ is added to the operator which results in the equation

$$
((-\Delta)^{\ell} + \varepsilon\mathbb{I})u(s,t) = f(s,t) \tag{3.12}
$$

where $\ell \geq 2$. For Dirichlet boundary conditions, i.e. $u \in \mathscr{H}$ where

$$
\mathscr{H} = \left\{ u \in \mathcal{C}^{2\ell}([0,1]\times[0,1],\mathbb{R}) \,\middle|\, u(0,t) = u(1,t) = u(s,0) = u(s,1) = 0 \right\},
$$

the ℓ harmonic differential operator $(-1)^\ell \Delta^\ell + \varepsilon \mathbb{I}$ has the same eigenfunctions as in example 3.3, and the Green's function becomes

$$\mathcal{G}_\varepsilon((s,t),(v,w)) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{4}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon} \sin(n\pi s)\sin(m\pi t)\sin(n\pi v)\sin(m\pi w) \tag{3.13}$$

and for Neumann boundary conditions the Green's function is

$$\mathcal{G}_\varepsilon((s,t),(v,w)) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{4}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon} \cos(n\pi s)\cos(m\pi t)\cos(n\pi v)\cos(m\pi w). \tag{3.14}$$

Both Green's functions are clearly bounded when $\ell \geq 2$. ∘

One could easily allow the ℓ in the Green's functions to be real, as long as the Green's function contains no singularities – wether these expressions will be Green's functions for $(-\Delta)^\ell$ is not obvious, but it would allow for greater flexibility in the degree of smoothness. These thoughts will not be further explored here.
The case that will receive the most intensive treatment is the biharmonic operator $\Delta\Delta$ corresponding to $\ell = 2$ and the triharmonic operator $-\Delta^3$. The Green's function corresponding to the biharmonic operator with $\varepsilon = 1$, subject to the Dirichlet boundary conditions can be found in Figure 3.3, and for the corresponding Neumann problem in Figure 3.4. The correlation structure along the boundary introduced by the Green's function for the Neumann problem seem somewhat unusual. This is most likely caused by the discontinuity of the boundary conditions at the corners, that makes this type of boundary conditions somewhat unnatural.

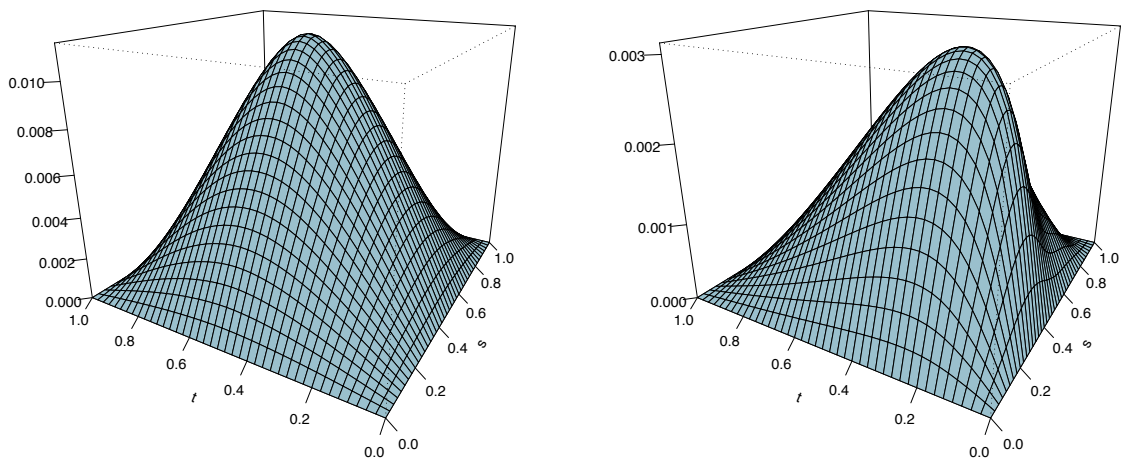

Figure 3.3: Plots of $\mathcal{G}_1(\,\cdot\,,(v,w))$ corresponding to $\Delta\Delta$ with Dirichlet boundary condtions. (v,w) has been set to (0.5, 0.5) in the left plot and (0.3, 0.1) in the right. The sums in (3.13) were both cut off at their 40th term.

Example 3.8. An alternative expression for the Green's function (3.13) when $\ell = 2$ is given in Nicholson & Bergman (1985) to be

$$G_\varepsilon((s,t),(v,w)) = \frac{1}{\sqrt{\varepsilon}} \sum_{n=1}^{\infty} \Phi_n(t\wedge w, t\vee w)\sin(n\pi s)\sin(n\pi v) \tag{3.15}$$

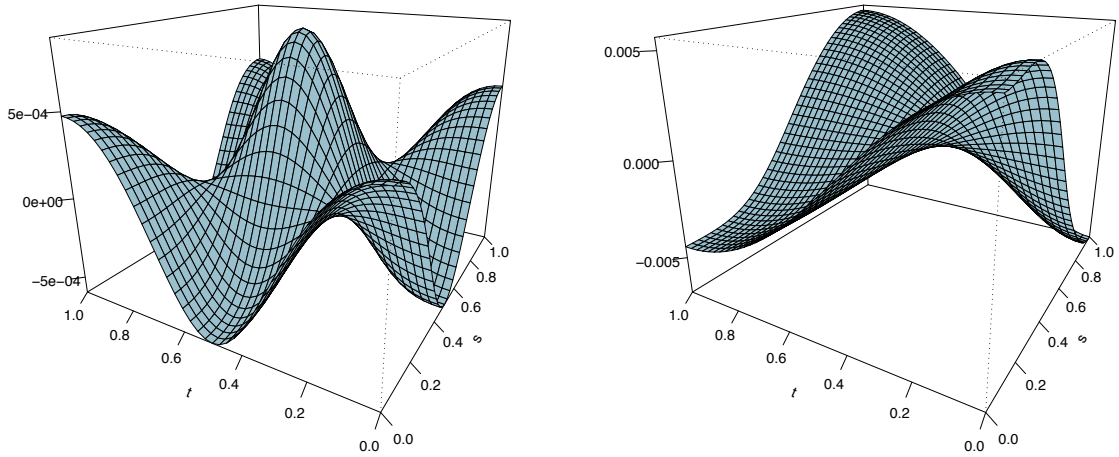

Figure 3.4: Plots of $\mathcal{G}_1(\,\cdot\,,(v,w))$ corresponding to $\Delta\Delta$ with Neumann boundary condtions. (v,w) has been set to (0.5, 0.5) in the left plot and (0.3, 0.1) in the right. The sums in (3.14) were both cut off at their 40th term.

where

$$\Phi_n(t,w) = -\frac{\Psi_{1,n}(1-w)\Psi_{1,n}(t)}{\Psi_{1,n}(1)} + \frac{\Psi_{2,n}(1-w)\Psi_{2,n}(t)}{\Psi_{2,n}(1)}$$

with

$$\Psi_{1,n}(t) = \frac{\sinh(t\sqrt{n^2\pi^2+\sqrt{\varepsilon}})}{\sqrt{n^2\pi^2+\sqrt{\varepsilon}}}$$

and

$$\Psi_{2,n}(t) = \begin{cases} \dfrac{\sinh(t\sqrt{|n^2\pi^2-\sqrt{\varepsilon}|})}{\sqrt{|n^2\pi^2-\sqrt{\varepsilon}|}} & \text{if } \sqrt{\varepsilon} \neq n\pi \\ t & \text{if } \sqrt{\varepsilon} = n\pi. \end{cases}$$

It is clear that if (3.15) needs N function evaluations to achive some relative error E_r, the representation (3.13) requires $\Omega(N^2)$ function evaluations to achieve the same relative error, which means that (3.15) is considerably faster to use in terms of computation time.

A similar expression can be calculated for Neumann boundary conditions with some tedious calculations. This will not be pursued here. ○

Finally, let us review some geometrical properties of Gaussian fields with the different choices of Green's functions as covariances.

Example 3.9. Let x be a centered Gaussian field with covariance function $\mathcal{G}_\varepsilon$ given by (3.13). In the following we will make use of the results that

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{n+m}{(n^2+m^2)^\ell} < \infty \tag{3.16}$$

for all $\ell \geq 2$, and also that

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{n^{2k}(n+m)}{(n^2+m^2)^\ell} < \infty \tag{3.17}$$

when $\ell \geq k+2$ and it equals $+\infty$ when $\ell = k+1$. Neither of these results are trivial, but can be established from properties of the Hurwitz zeta function.

Let us first consider continuity properties of x for different values of ℓ. Since we have that $\mathcal{G}_0 \geq \mathcal{G}_\varepsilon$ for all $\varepsilon > 0$ we will work with $\mathcal{G}_0$. Denote by $\mathcal{D}$ the function

$$\mathcal{D}((s,t),(v,w)) = \mathcal{G}_0((s,t),(s,t)) + \mathcal{G}_0((v,w),(v,w)) - 2\mathcal{G}_0((s,t),(v,w)).$$

By Theorem 2.1 x is continuous if we can show that

$$\mathcal{D}((s,t),(v,w)) \leq \frac{1}{\log(|s-v|+|t-w|)^2} \tag{3.18}$$

for $|s-v|+|t-w| < \eta$.

For any $(s,t) \in [0,1]\times[0,1]$ and any unit vector (α_1,α_2) we have that

$$\begin{aligned}\frac{\partial}{\partial\eta}\mathcal{D}((s,t),(s+\alpha_1\eta,t+\alpha_2\eta)) = 8\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{n+m}{\pi^{2\ell}(n^2+m^2)^\ell}\Big(&\sin(n\pi s)\sin(n\pi t)) \\ -\sin(n\pi(s+\alpha_1\eta)\sin(m\pi(t+\alpha_2\eta))\Big)\Big(&-\alpha_1 n\pi\cos(n\pi(s+\alpha_1\eta)\sin(m\pi(t+\alpha_2\eta)) \\ &-\alpha_2 m\pi\sin(n\pi(s+\alpha_1\eta))\cos(m\pi(t+\alpha_2\eta))\Big)\end{aligned}$$

which is convergent by (3.16). Since the above funtion is bounded (as opposed to the derivative of the inverse logarithm in (3.18)) and by the fact that $\mathcal{D}$ is zero on the diagonal, there exists a rectangle of (v,w)'s close to (s,t) such that the equation (3.18) holds.

Now let us move to the question of differentiability. We have that

$$\frac{\partial^{2k}}{\partial s^k\partial v^k}\mathcal{D}((s,t),(v,w)) = -8\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{n^{2k}}{\pi^{2(\ell-k)}(n^2+m^2)^\ell}g_k(n\pi s)\sin(m\pi t)g_k(n\pi v)\sin(m\pi w)$$

where

$$g_k(s) = \begin{cases}\sin(s) & \text{for } k \text{ even}\\ \cos(s) & \text{for } k \text{ odd.}\end{cases}$$

A simple calculation shows that

$$\frac{\partial}{\partial\eta}\left(\frac{\partial^{2k}}{\partial s^k\partial v^k}\mathcal{D}((s,t),(v,w))\Big|_{\substack{v=s+\alpha_1\eta\\ w=t+\alpha_2\eta}}\right)$$

is finite if (3.17) holds. From this we obtain the following theorem.

Theorem 3.1. A Gaussian process $x \sim \mathcal{N}(0,\mathcal{G})$, where $\mathcal{G}$ is the Green's function for $(-\Delta)^\ell$ is continuous for $\ell \geq 2$. For $\ell \geq 3$ it furthermore holds that x is (at least) $\ell-2$ times continuously differentiable.

Note that it has not been shown that the Gaussian process in the above theorem is exactly $\ell-2$ times continously differentiable, but the author hypothesize that this is in fact the case. ∘

As the final example of this chapter we will consider an alternative roughness penalty.

Example 3.10. Instead of confining the differential operators to $[0,1]\times[0,1]$, one could let them be defined on the entire plane $\mathbb{R}^2$ with the boundary conditions translated to plus and minus infinity in both coordinates, and argue that we are just watching a window of an image. $[0,1]\times[0,1]$ is not an

ideal window since it is not symmetric around $(0,0)$, and it is better to think of the images as being defined on $[-1,1]\times[-1,1]$. If we use $\mathscr{L}=(-\Delta)^\ell$ one would get

$$\int_{-1}^{1}\int_{-1}^{1}\theta(s,t)(-\Delta)^\ell\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t$$

as our roughness penalty (corresponding to putting θ equal to zero outside $[-1,1]\times[-1,1]$, where we have no observations) but the corresponding Green's function will have a much simpler form than what has been seen so far. Meinguet (1979) shows that the corresponding Green's function is

$$\mathcal{G}((s,t),(v,w))=\frac{1}{2^{2\ell-1}\pi((\ell-1)!)^2}\|(s,t)-(v,w)\|^{2(\ell-1)}\log\|(s,t)-(v,w)\|, \tag{3.19}$$

where $\|\cdot\|$ denotes the Euclidian norm. This Green's function has some very appealing properties, firstly that it does not involve any infinite series and secondly that it is stationary. In terms of the model (1.4) however, it does not make much sense as a covariance function. ∘

3.3 Relation to other work

The one-dimensional version of the model (1.1) with a roughness penalty is the standard model in functional data analysis (see e.g. Chapter 5 in the main reference on functional data analysis Ramsay & Silverman (2005).) It is of course also the natural model to start out from in higher dimensions, and there are numerous texts that take this approach. The usual approach to this problem would be to use some sort of spline smoothing (e.g. using (2.7)), and to infer the smoothing parameter λ and perhaps also the order ℓ by generalized cross validation. Algorithms to do this has been developed, and a number of optimizations of the necessary matrix operations has been suggested, some of which are discussed in Wahba (1990). The one-dimensional version of the model (1.4) is also introduced in Kimeldorf & Wahba (1970) and the corresponding Bayesian estimate is shown to be equal to the spline solving the smoothing problem. To the author's knowledge is has not been pursued in higher dimensions. This is probably due to the fact that the Green's function used for smoothing, likely due to ease of calculation, is usually given by (3.19), which is not very convincing as a covariance. Alternative one-dimensional models are presented in Kimeldorf & Wahba (1971) where both prediction and filtering are related to the (spline) solution of the smoothing problem. The filtering problem is also used as an alternative model for the multi-dimensional version of (1.1) in Wahba (1978).

Dyn et al. (1979) suggest the second-order smoothness criterion

$$\iint\Delta\Delta\theta(s,t)\,\mathrm{d}s\,\mathrm{d}t$$

as opposed to a criterion without any mixed derivatives, and some results on this choice is shown. Wahba & Wendelberger (1980) introduce the penalty $J_\ell(\theta)$, given by

$$J_\ell(\theta)=\iint\sum_{k=0}^{\ell}\binom{\ell}{k}\left|\partial_s^k\partial_t^{\ell-k}\theta(s,t)\right|^2\,\mathrm{d}s\,\mathrm{d}t$$

which corresponds to the choice $\mathscr{L}=(-\Delta)^k$ with suitable boundary conditions. They consider $k=1$ but note that the solution to the problem involves the Green's function for the Laplacian which is unbounded, and instead proceed with the case $k=2$. In Wahba & Wendelberger (1980) the idea of smoothing (meteorological) images with the Laplacian is credited to Sasaki[3] who apparently used the Laplacian, even though he noted that the corresponding Green's function was unbounded. His approach

[3] Sasaki, U., *An objective analysis for determining initial conditions for the primary equations,* Technical Report 208, Department of Oceanography and Meteorology, A&M College of Texas (1960)

involved specifying an elliptic partial differential equation with some boundary conditions, which sounds similar to our approach.

In Wahba & Wendelberger (1980) (p. 1128) it is mentioned that derivatives of the generalized smoothing spline function can be obtained by analytical differentiation up to the order $\ell - 2$, which seems to fit very well with Theorem 3.1. The result is stated without any proof or reference.

CHAPTER 4

DATA EXAMPLES

Example 4.1. A noisy image y is observed, where

$$y_{ij} = \theta(s_i, t_j) + \varepsilon_{ij}, \qquad i, j = 1, \ldots, 21$$

with

$$\theta(s, t) = 70s^2(s-1)t(t-1)$$

and the ε_{ij}'s are independent and identically normal distributed random variables with standard deviation 0.5. A plot of θ and the (very) noisy observation y can be found in Figure 4.1.

θ satisfies Dirichlet boundary conditions, so smoothing will be done with $(-\Delta)^\ell$ on the space of 2ℓ times differentiable functions that are zero along the border. It should be noted that using observations on the border provides additional information and results in significantly better estimates than what is obtained without using the boundary information.
The smoothing of y (`pic`) with $(-\Delta)^\ell + \varepsilon\mathbb{I}$ subject to Dirichlet boundary conditions with the boundary information used, can be done with the `laplacesmooth` R package using the code in Listing 4.1.

Listing 4.1: Smoothing an image and plotting the smooth

```
size <- dim(pic)
vcov <- covmatrix(n1 = size[1], n2 = size[2], order = 3, epsilon = epsilon,
                  type = "dirichlet")
est <- estimate(pic, vcov)$par
persp(x = steps, y = steps, cematrix(pic, est[1], est[2], vcov))
```

Figure 4.2 holds a plot of the smoothed images for $\ell = 1, \ldots, 4$ where both sums in (3.13) have been cut off at their 20th term and $\varepsilon = 0$.

The corresponding (maximum likelihood) parameter estimates in the Bayesian model (1.4) are given in the table below where τ is the scaling parameter of the Gaussian field x.[1]

ℓ	$\hat{\sigma}$	$\hat{\tau}$
1	0.4146035	0.8477593
2	0.4870721	8.064982
3	0.4936229	70.20042
4	0.4969383	533.5105

The estimates of σ become better (increase) as ℓ increases, and the τ estimates increase with ℓ, which is caused by the decreasing scales of the corresponding Green's functions. 100 samples have been simulated for each of the four orders of smoothing, and the parameters have been estimated. A boxplot of the σ estimates can be found in Figure 4.3. The median of the estimates becomes closer to the real value of 0.5 as ℓ increase, but from $\ell = 3$ to $\ell = 4$ the width of the interval of estimated values also increase. This is not a general feature however. The result is interesting, since the original function is in the null space of e.g. $-\Delta^3$, but it still gives better results than $\ell = 1$ and 2. It is not clear that a similar plot of

[1] Note that maximum likelihood estimation in the case $\ell = 1$ is only possible due to the censoring of the double series, as it would otherwise have resulted in infinite variance.

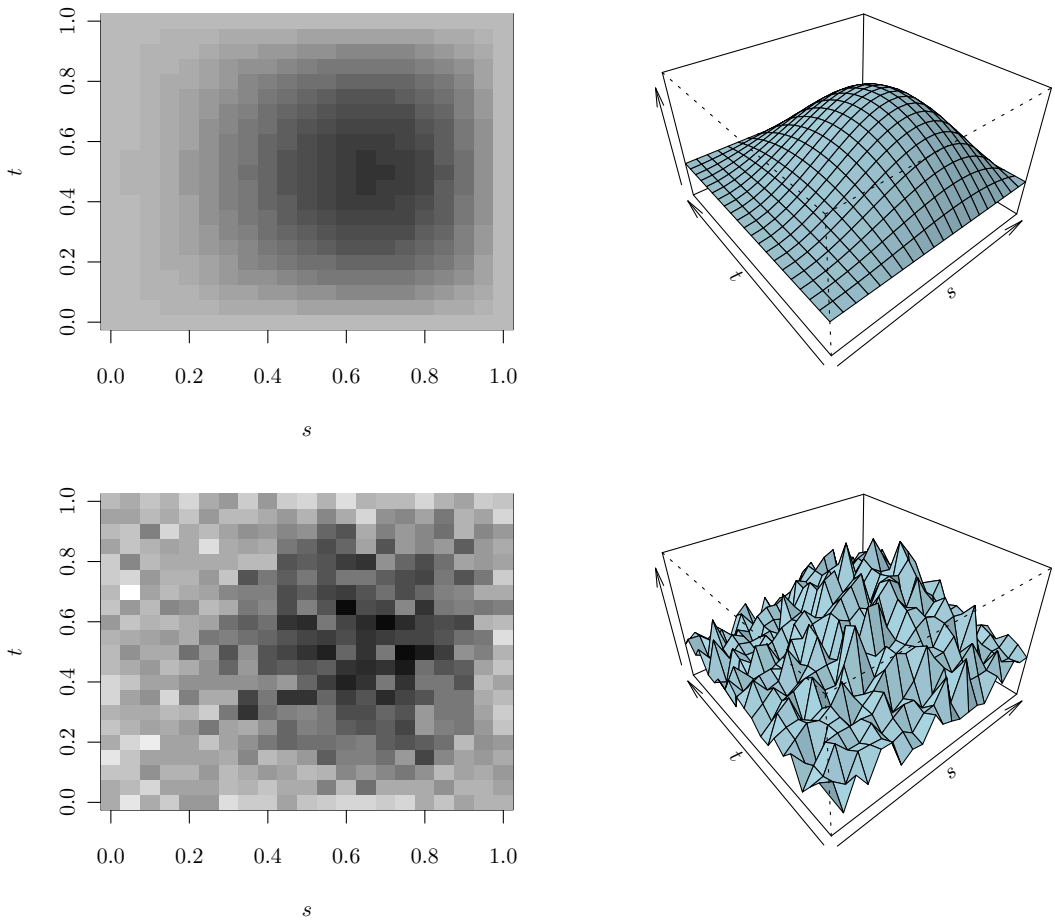

Figure 4.1: The function θ in a grayscale representation and in three dimensions (top plots) and the observed image (bottom plots.)

the τ estimates would provide any valuable information, but it should be noted that the estimates of τ become less reliable for higher values of ℓ (> 5), which will also produce less reliable σ estimates. This is most likely due to a numerical problem caused by the corresponding Green's functions taking values very close to zero.

Now let us consider the problem of estimating the partial derivatives of θ. From (1.5) it is obtained that the estimate of the partial derivative $\partial_s \mathrm{E}[\, x \,|\, y \,]$ in every observation point is given by

$$\left\{\tau^2 \partial_s \mathcal{G}((s,t),(s_k,t_l))\right\}_{(k,l)} \left\{\tau^2 \mathcal{G}((s_i,t_j)(s_k,t_l)) + \sigma^2 \mathbb{1}_{(i,j)=(k,l)}\right\}^{-1}_{(i,j),(k,l)} y$$

and likewise for other derivatives. Figure 4.4 holds plots of $\partial_s \theta(s,t)$ and $\partial_t \theta(s,t)$. For $\ell > 1$ we know that the smoothed images $\mathrm{E}[\, x \,|\, y \,]$ is always 2ℓ times differentiable, but Theorem 3.1 only tells us that the Gaussian process x is $\ell - 2$ times continuously differentiable, this would suggest that if the model were in fact true we would only have information about derivatives up to order $\ell - 2$, so to be on the safe side one should refrain from estimating derivatives of higher order.
Looking at Figure 4.2 it is clear that $\ell = 1$ produces a smooth that is quite rough, and thus we will not bother to estimate partial derivatives, as their behaviour will be too extreme. This way the partial

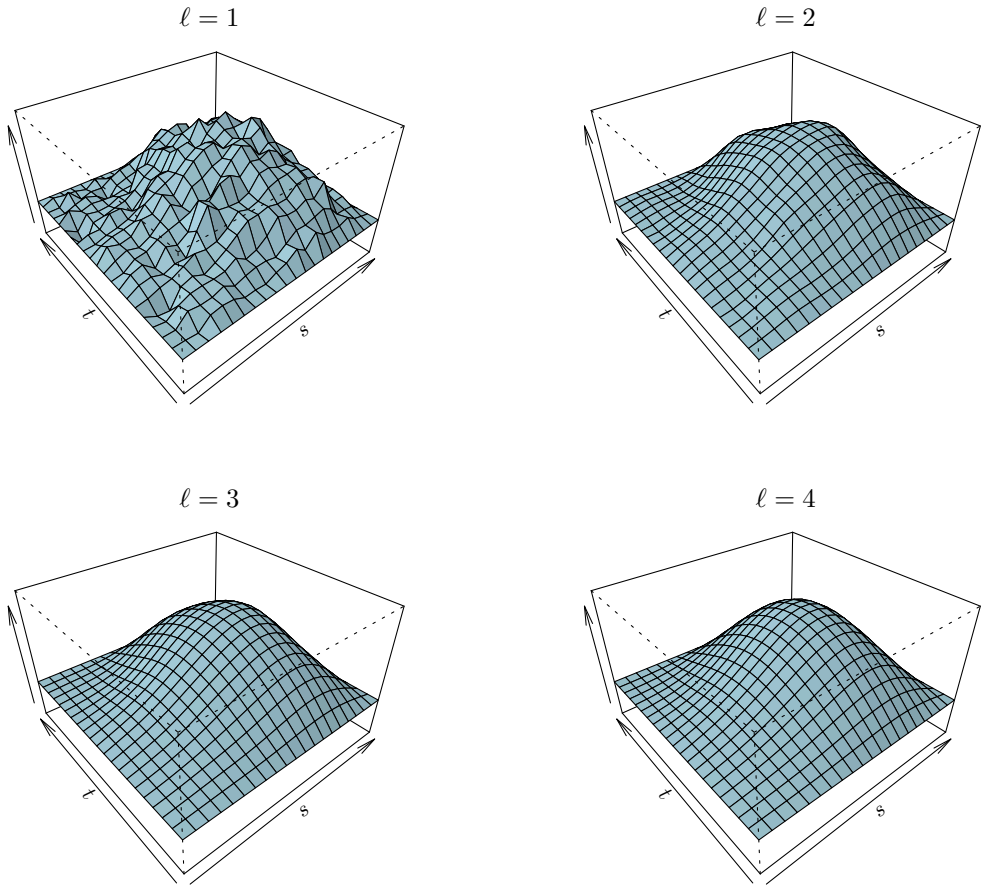

Figure 4.2: The smooths of the observed image for different values of ℓ. $\ell = 3$ and $\ell = 4$ produce smooths that are indistinguishable from θ on this scale.

derivatives have only been estimated for $\ell \geq 2$ and can be found in Figure 4.5. All estimates capture the shape of the partial derivatives reasonably well, but it is only in the cases $\ell = 3$ and $\ell = 4$ that the estimates could be considered good, with the latter producing a quite nice estimate. The biggest problem in the estimates is in the estimate of $\partial_s \theta$ along the edge $s = 1$. It is no surprise that the biggest problem is along here as this is the edge with the greatest deviation from 0.
Measuring the absolute difference from the estimated pictures and the true picture by

$$\sum_{i=1}^{21} \sum_{j=1}^{21} |\theta(s_i, t_j) - \mathrm{E}\,[\,x(s_i, t_j)\,|\,y\,]|$$

one gets the values summarized in the Table 4.1.
Interestingly the estimate of $\partial_s \theta$ seem to be better for $\ell = 3$ than $\ell = 4$. In other replications this phenomenon is rarely observed for $\ell = 4$, but is typically seen when going from $\ell = 4$ to $\ell = 5$ or from 5 to 6. This makes sense in relation to the overfitting allowed by these higher-order roughness penalties.

ℓ	∂	Absolute difference
1	–	72.30071
2	–	34.43044
3	–	29.3492
4	–	28.97725
2	∂_s	407.5034
3	∂_s	266.8843
4	∂_s	268.7481
2	∂_t	372.576
3	∂_t	226.5541
4	∂_t	205.6528

Table 4.1: Absolute difference between true (Figure 4.4) and estimated derivatives (Figure 4.5).

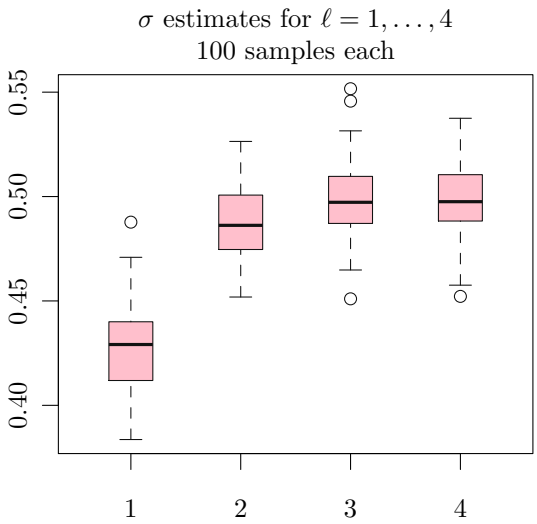

Figure 4.3: Boxplot of σ estimates for different values of ℓ. Each box is based on 100 independent replications.

∘

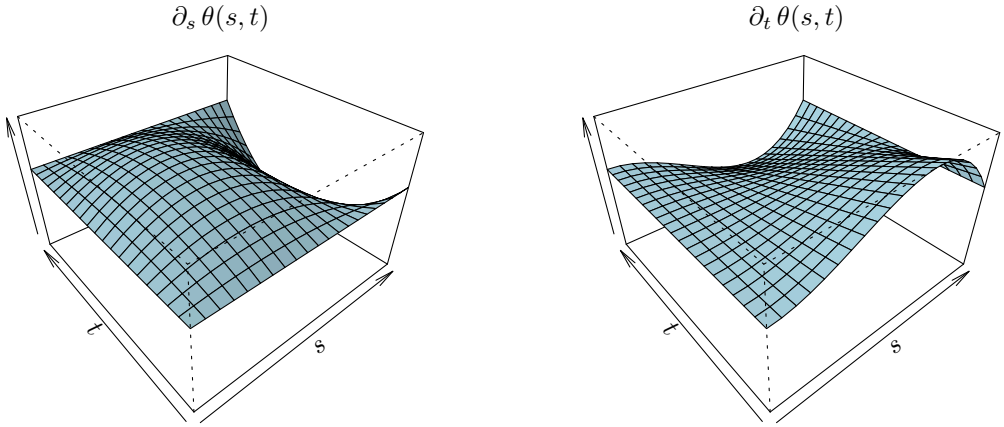

Figure 4.4: Partial derivatives of θ.

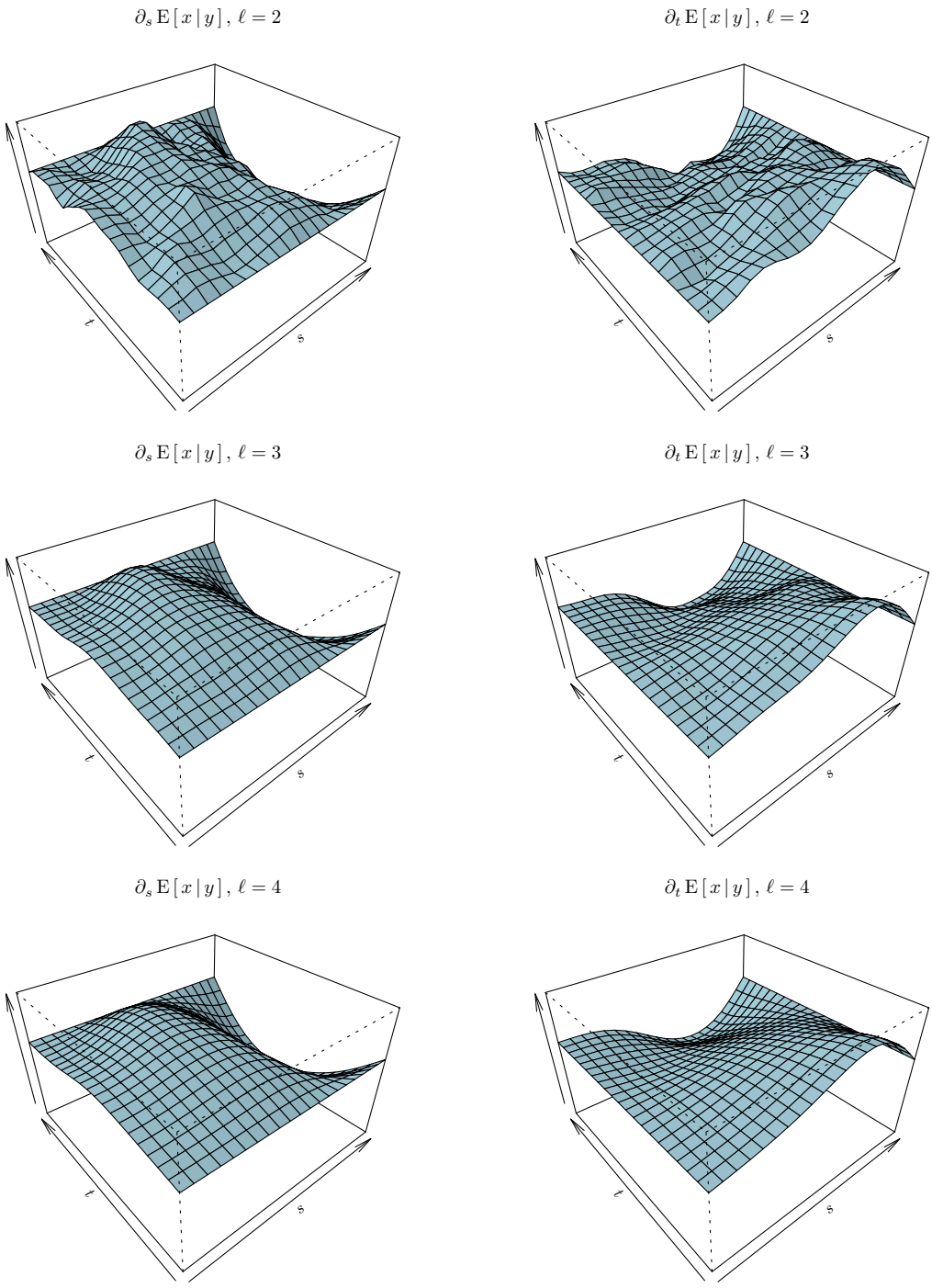

Figure 4.5: Estimated partial derivatives of θ.

Example 4.2. Now consider the observation of an image y where

$$y_{ij} = x(s_i, t_j) + \varepsilon_{ij}, \qquad i, j = 1, \ldots, 21 \tag{4.1}$$

with x centered Gaussian with covariance $\tau^2\mathcal{G}$, where $\mathcal{G}$ is the Green's function corresponding to $(-\Delta)^\ell$ and the ε_{ij}'s are independent $\mathcal{N}(0,\sigma^2)$ distributed variables. An image of such an x can be simulated with the R code from Listing 4.2 using the `laplacesmooth` package. The simulation is done by using the eigenvalues and eigenvectors from the relevant covariance matrix to produce a square root of the matrix, which is then multiplied to a vector of standard normal distributed variables, then converting the resulting vector to an $n_1 \times n_2$ matrix.

Listing 4.2: Simulating and plotting a centered Gaussian process with covariance $\mathcal{G}$

```
size <- 21
steps <- seq(0, 1, length = size)
vcov <- covmatrix(n1 = size, n2 = size, order = 3)
pic <- sim(vcov)
persp(x = steps, y = steps, pic)
```

Simulations of centered Gaussian fields x with covariance $\mathcal{G}$ for different choices of ℓ and boundary conditions are found in Figure 4.6. The fields have been simulated from the same seed (i.e. the same observation of an image of independent standard normal distributed random variables) and the different choices of Green's functions change the behaviour of the fields in exactly the ways one would expect. For $\ell = 2$ the difference between Dirichlet and Neumann boundary conditions is that the image is tied down for Dirichlet conditions, and for Neumann conditions the "inner motion" of the image is carried on to the edges. For $\ell = 3$ the fields have similar geometrical features, but are smoother (and are scaled down.) The problem with the continuity of the Neumann boundary conditions is somewhat noticeable by the shape of the Neumann processes. The figure seems to suggest that the difference of scale is approximately a factor 10. Therefore we will put $\tau = 1$ in (4.1) when $\ell = 2$ and $\tau = 10$ when $\ell = 3$. σ will be fixed at 0.01. The parameters were decently recovered as maximum likelihood estimates and can be found in the table below.

ℓ	Boundary conditions	$\hat{\sigma}$	$\hat{\tau}$
2	Dirichlet	0.01089435	0.96460265
2	Neumann	0.01036528	0.98541738
3	Dirichlet	0.01067704	9.54557709
3	Neumann	0.01058705	9.41531745

To obtain any conclusions a more careful normalization procedure has to be developed. E.g. one could consider

$$\int_0^1 \mathrm{Var}_\ell[x(t,t)]\,\mathrm{d}t$$

as a measure of the scale of the corresponding process. For a general ℓ and Dirichlet boundary conditions the above integral becomes

$$\frac{4}{\pi^{2\ell}}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{(n^2+m^2)^\ell}\int_0^1 \sin^2(n\pi t)\sin^2(m\pi t)\,\mathrm{d}t = \frac{1}{\pi^{2\ell}}\sum_{n=1}^{\infty}\sum_{\substack{m=1\\ m\neq n}}^{\infty}\frac{1}{(n^2+m^2)^\ell} + \frac{3}{2^{2\ell+1}\pi^{2\ell}}\sum_{n=1}^{\infty}\frac{1}{n^{2\ell}}.$$

It is seen that the same result holds for Neumann boundary conditions, so only the parameter ℓ determines the scale. The table on the next page holds numerically computed values for different values of ℓ and the associated choices of τ that makes the processes τx comparable across the different ℓ's.

ℓ	$\int_0^1\int_0^1 \mathrm{Var}_\ell[x(s,t)]\,\mathrm{d}s\,\mathrm{d}t$	τ
2	$1.84\cdot 10^{-3}$	1
3	$2.41\cdot 10^{-5}$	8.73
4	$4.10\cdot 10^{-7}$	67.0
5	$7.62\cdot 10^{-9}$	491

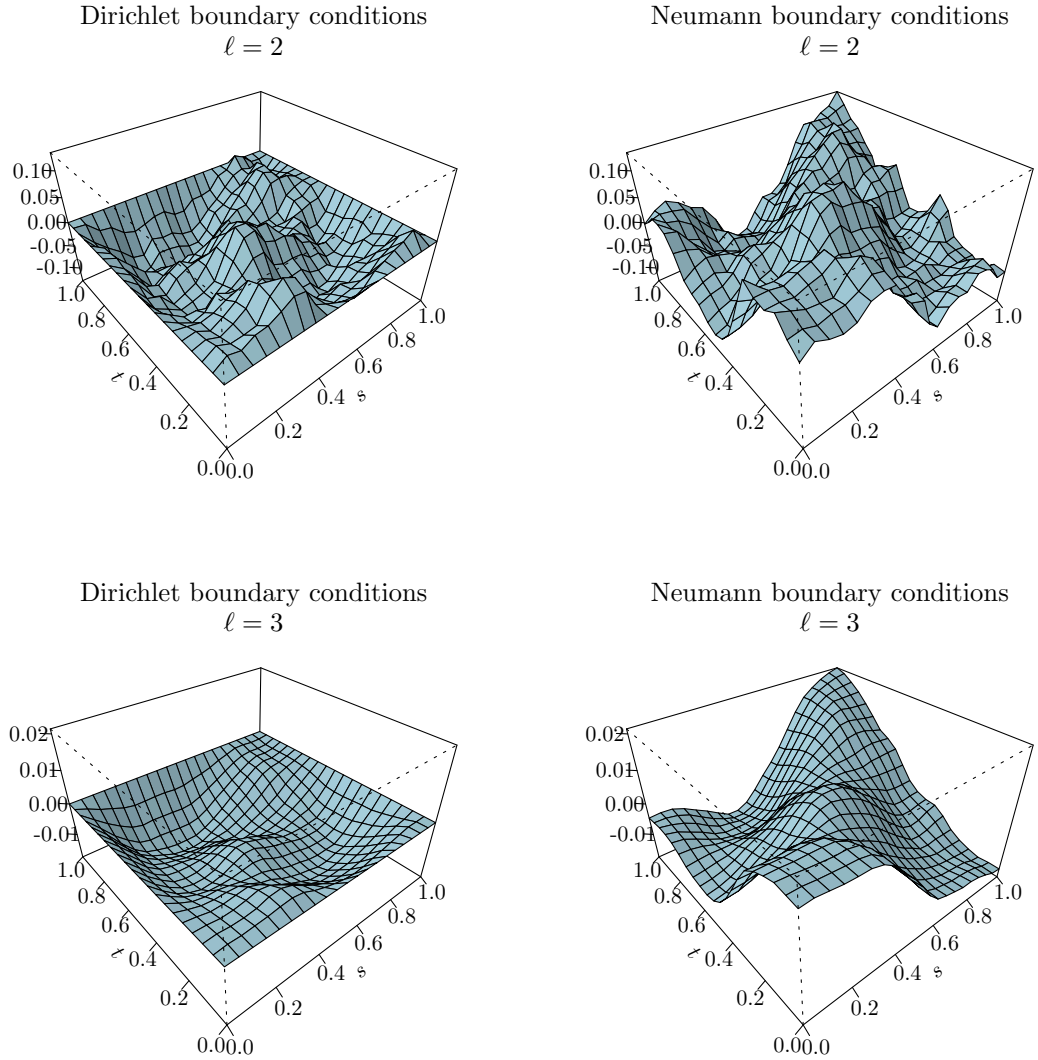

Figure 4.6: Simulated centered Gaussian processes with covariances given by the Green's functions $\mathcal{G}$ corresponding to $(-\Delta)^\ell$ with different choices of boundary conditions. The processes have been generated using the same seed.

τ is computed as

$$\tau = \sqrt{\frac{\int_0^1 \int_0^1 \mathrm{Var}_2[x(s,t)]\,\mathrm{d}s\,\mathrm{d}t}{\int_0^1 \int_0^1 \mathrm{Var}_\ell[x(s,t)]\,\mathrm{d}s\,\mathrm{d}t}}.$$

Two boxplots of the estimated values of σ and τ in model (4.1) with τ given in the above table and $\sigma = 0.01$ is found in Figure 4.7. The estimated values of σ are quite reasonable. For the τ estimates it seems that the variance of the estimator increase drastically with ℓ. Wether this is a natural condition or caused by numerical instability is not clear.

∘

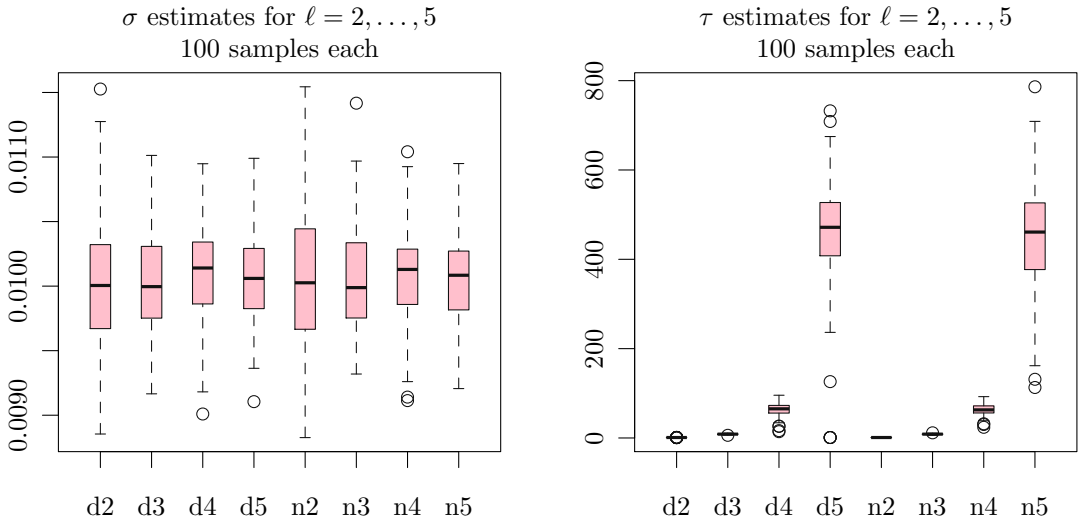

Figure 4.7: Boxplots of the estimated σ and τ values. Each box corresponds to 100 independent simulated Gaussian processes with normal distributed noise with standard deviation $\sigma = 0.01$. The box names indicate the choice of boundary conditions and value of ℓ, i.e. 'd2' means Dirichlet boundary conditions with $\ell = 2$.

Example 4.3. As our final example in this section we will consider a complex fixed image. We have the model

$$y_{ij} = \theta(s_i, t_j) + \varepsilon_{ij}, \qquad i, j = 1, \ldots, 32$$

with

$$\theta(s,t) = \cos\left(\frac{3\pi}{4}((2s-1)^2 + (2t-1)^2)\right) - \cos(\pi(2s-1))\sin(\pi(2t-1))$$

and the ε_{ij}'s are independent and identically normal distributed random variables with zero mean and standard deviation 0.09. This image and the choice of standard deviation is identical to an example found in O'Sullivan (1991), where another method is used for the estimation/smoothing. A plot of θ and the noisy observation y can be found in Figure 4.8.

The image θ neither satisfies Dirichlet or Neumann boundary conditions. To deal with that we will assume that the observations are all within the square $[0.25, 0.75] \times [0.25, 0.75]$. If `pic` is a matrix object containing the observations y, the analysis can be done with the `laplacesmooth` package using the code from Listing 4.3.

Listing 4.3: Specifying an alternative boundary

```
cov3d <- covmatrix(n1 = 32, n2 = 32, slow = 0.25, shigh = 0.75, tlow = 0.25,
                   thigh = 0.75, order = 3, type = "dirichlet")
est3d <- estimate(pic, cov3d)$par
smooth3d <- cematrix(pic, est3d[1], est3d[2], cov3d)
persp(x = steps, y = steps, smooth3d)
```

In Figure 4.9 the smooths for various combinations of boundary conditions and degrees of smoothing can be found. The specific maximum likelihood estimates are of less interest than the shape of the

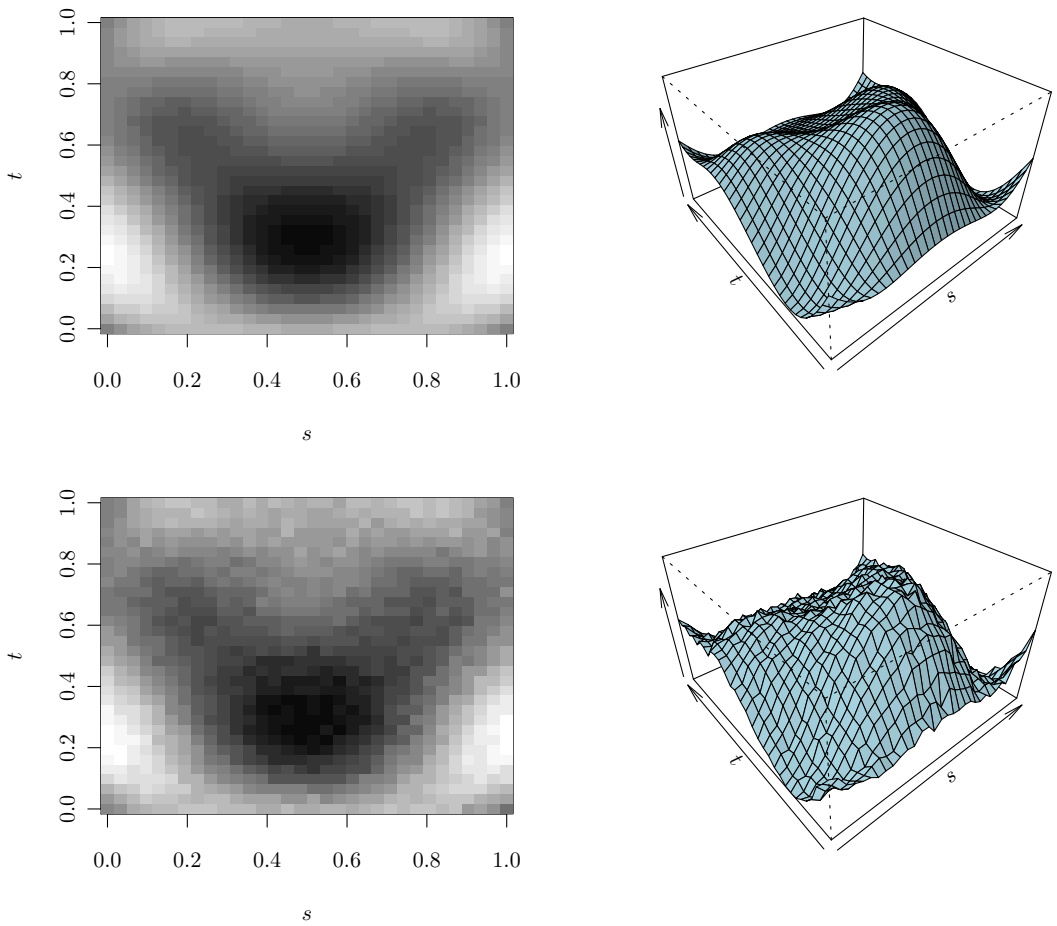

Figure 4.8: The function θ in a grayscale representation and in three dimensions (top plots) and the observed image (bottom plots.)

images. The four smooths are indistinguishable to the naked eye, but if one measures the absolute difference from the estimated pictures and the true picture by

$$\sum_{i=1}^{32}\sum_{j=1}^{32} |\theta(s_i, t_j) - \mathrm{E}\,[\,x(s_i, t_j)\,|\,y\,]|$$

one gets the values summerized in the table below

ℓ	Boundary conditions	Absolute difference
3	Dirichlet	22.77934
3	Neumann	23.2517
4	Dirichlet	19.03713
4	Neumann	20.03809

It seems that Dirichlet boundary conditions give slightly better fits judged by this measure, and also that $\ell = 4$ is better than $\ell = 3$. It also seems that the slight increase in the number of observations has affected the quality of the smooths compared to e.g. Example 4.1.

∘

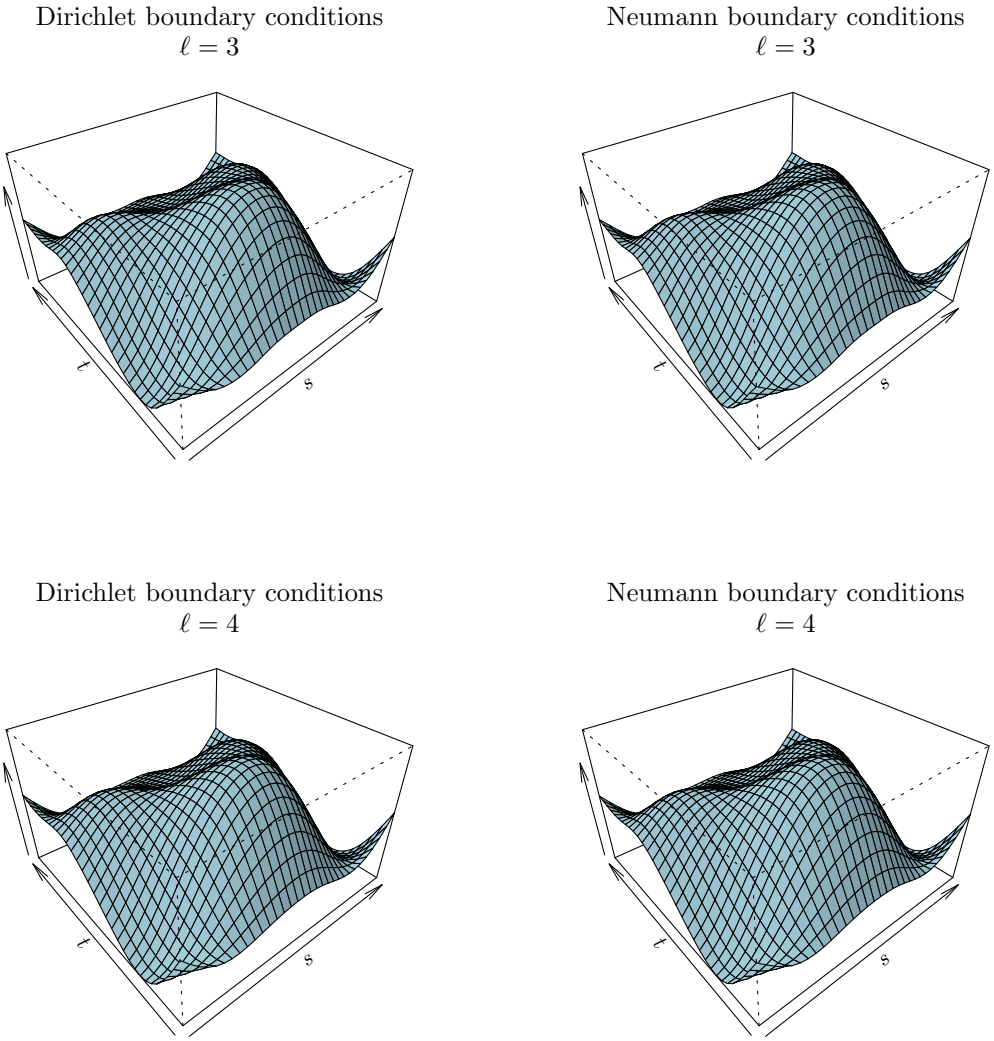

Figure 4.9: The estimates of θ for various combinations of boundary conditions and degrees of smoothing.

CHAPTER 5

APPROXIMATIONS

The formulae from Chapter 1 will only allow for maximum likelihood estimation for sparsely observed images, as already mentioned. In this chapter the computationally intensive matrix operations will be approximated by continuous operations that can be implemented to run in constant or linear time. These are all similar to the one-dimensional approximations found in Markussen (2013).

5.1 Approximations of matrix computations

We want to approximate the operation of the matrix $\{\mathcal{G}((s_i,t_j),(s_k,t_l))\}_{(i,j),(k,l)}$ in terms of the integral operator $\mathscr{G}$ given by

$$\mathscr{G}f(s,t) = \int_0^1 \int_0^1 \mathcal{G}((s,t),(v,w))f(v,w)\,\mathrm{d}v\,\mathrm{d}w.$$

For this we will need to specify the exact setup in which we are working. Let $(s_k,t_l)_{kl}$ be a grid of points in $[0,1]\times[0,1]$ where $0 \le k \le n_1+1$ and $0 \le l \le n_2+1$, with $s_0 = t_0 = 0$ and $s_{n_1+1} = t_{n_2+1} = 1$, and assume furthermore that it holds that

$$s_{k+1} - s_k = \frac{1}{n_1+1}, \qquad t_{l+1} - t_l = \frac{1}{n_2+1}$$

for all $0 \le k \le n_1$ and $0 \le l \le n_2$. I.e. the points in the grid $(s_k,t_l)_{kl}$ are equidistantly spaced. For $z \in \mathbb{R}^{(n_1+2)\times(n_2+2)}$ denote by $\mathscr{E}_z$ the piecewise linear embedding of z into $\mathcal{C}([0,1]\times[0,1],\mathbb{R})$ with anchors given by the grid $(s_k,t_l)_{kl}$. I.e. $\mathscr{E}_z$ is the piecewise linear function (linear interpolation) for which

$$\mathscr{E}_z(s_i,t_j) = z_{ij}$$

for $i = 0,\dots,n_1+1$ and $j = 0,\dots,n_2+1$.

The next proposition establishes the connection between the matrix $\{\mathcal{G}((s_i,t_j),(s_k,t_l))\}$ and the integral operator $\mathscr{G}$.

Proposition 5.1. Let $\mathcal{G}((s,t),(\,\cdot\,,\,\cdot\,)) : [0,1]\times[0,1] \to \mathbb{R}$ be a function that is twice differentiable on the squares $[s_k,s_{k+1}]\times[t_l,t_{l+1}]$ with continuous second derivatives. If

$$z_{0,l} = z_{n_1+1,l} = z_{k,0} = z_{k,n_2+1} = 0,$$

for all l,k there exist $\xi_{1k} \in [s_{k-1},s_k]$ and $\xi_{2l} \in [t_{l-1},t_l]$ such that

$$\sum_{l=1}^{n_2}\sum_{k=1}^{n_1} \mathcal{G}((s,t),(s_k,t_l))z_{kl} - (n_1+1)(n_2+1)\int_0^1\int_0^1 \mathcal{G}((s,t),(v,w))\mathscr{E}_z(v,w)\,\mathrm{d}v\,\mathrm{d}w,$$

equals

$$\begin{aligned}
&\frac{1}{12(n_2+1)^2}\sum_{l=1}^{n_2+1}\sum_{k=1}^{n_1}\Big(\mathcal{G}((s,t)(s_k,\,\cdot\,))\mathscr{E}_z(s_k,\,\cdot\,)\Big)^{(2)}(\xi_{2l}) \\
&\qquad + \frac{1}{12(n_1+1)^2}\sum_{l=1}^{n_2}\sum_{k=1}^{n_1+1}\Big(\mathcal{G}((s,t)(\,\cdot\,,t_l))\mathscr{E}_z(\,\cdot\,,t_l)\Big)^{(2)}(\xi_{1k}) \\
&\qquad - \frac{1}{144(n_1+1)^2(n_2+1)^2}\sum_{l=1}^{n_2+1}\sum_{k=1}^{n_1+1}\Big(\mathcal{G}((s,t)(\,\cdot\,,\,\cdot\,))\mathscr{E}_z\Big)^{(2,2)}(\xi_{1k},\xi_{2l}),
\end{aligned} \tag{5.1}$$

where the superscript numbers in parentheses indicate (partial) derivatives.

Proof. The Trapezoidal rule of integration give points $\xi_{1k} \in [s_{k-1}, s_k]$ such that the scaled integral $(n_1+1)(n_2+1)\int_0^1 \mathcal{G}((s,t),(v,w))\mathscr{E}_z(v,w)\,\mathrm{d}v$ equals

$$(n_2+1)\sum_{k=1}^{n_1}\mathcal{G}((s,t)(s_k,w))\mathscr{E}_z(s_k,w) - \frac{(n_2+1)}{12(n_1+1)^2}\sum_{k=1}^{n_1+1}\Big(\mathcal{G}((s,t)(\,\cdot\,,w))\mathscr{E}_z(\,\cdot\,,w)\Big)^{(2)}(\xi_{1k}).$$

Using the Trapezoidal rule again, points $\xi_{2l} \in [t_{l-1}, t_l]$ are obtained such that the integral of the above function with regard to w equals

$$\begin{aligned}\sum_{l=1}^{n_2}\sum_{k=1}^{n_1}\mathcal{G}((s,t),(s_k,t_l))\mathscr{E}_z(s_k,t_l) - \frac{1}{12(n_2+1)^2}\sum_{l=1}^{n_2+1}\sum_{k=1}^{n_1}\Big(\mathcal{G}((s,t)(s_k,\,\cdot\,))\mathscr{E}_z(s_k,\,\cdot\,)\Big)^{(2)}(\xi_{2l})\\ -\frac{1}{12(n_1+1)^2}\sum_{l=1}^{n_2}\sum_{k=1}^{n_1+1}\Big(\mathcal{G}((s,t)(\,\cdot\,,t_l))\mathscr{E}_z(\,\cdot\,,t_l)\Big)^{(2)}(\xi_{1k})\\ +\frac{1}{144(n_1+1)^2(n_2+1)^2}\sum_{l=1}^{n_2+1}\sum_{k=1}^{n_1+1}\Big(\mathcal{G}((s,t)(\,\cdot\,,\,\cdot\,))\mathscr{E}_z\Big)^{(2,2)}(\xi_{1k},\xi_{2l})\end{aligned}$$

where it has been used that

$$z_{0,l} = z_{N_1+1,l} = z_{k,0} = z_{k,N_2+1} = 0.$$

□

Note that the error (i.e. the difference of the two expressions) in Proposition 5.1 can be explicitly written out in terms of the second derivatives of $\mathcal{G}$ and the ξ_{kl}'s. What is gained in enlightenment from this expression can be discussed however, and it is left to the interested reader.

Denote by G the $n_1n_2 \times n_1n_2$ matrix

$$\Big\{\mathcal{G}((s_i,t_j)(s_k,t_l))\Big\}_{(i,j),(k,l)}$$

and by $\overline{G}$ the $(n_1+2)(n_2+2)\times(n_1+2)(n_2+2)$ where the border entries are included. If z denotes the $n_1n_2 \times 1$ matrix of observations

$$z = \Big\{z_{kl}\Big\}_{(k,l)}$$

and $\overline{z}$ denotes the $(n_1+2)(n_2+2)\times 1$ matrix with zeros at the entries corresponding to border observations, then proposition 5.1 suggests the approximation

$$Gz \approx \Big\{\mathscr{G}\mathscr{M}\mathscr{E}_{\overline{z}}(s_i,t_j)\Big\}_{(i,j)} \tag{5.2}$$

where $\mathscr{M}$ is the multiplication operator

$$\mathscr{M}f = (n_1+1)(n_2+1)f.$$

We can identify z with the function $\mathscr{M}\mathscr{E}_{\overline{z}} \in \mathcal{C}([0,1]\times[0,1],\mathbb{R})$. With this identification the roles of G, $\overline{G}$ and $\mathcal{G}$ will be similar on their respective spaces. A schematic representation of the approximation setup can be found in Figure 5.1.

Since inversion is a continuous operation, the conditional expectation can be approximated by

$$\widehat{\mathrm{E}}\,[\,x\,|\,y\,] = \tau^2\mathscr{G}\mathscr{M}\left(\tau^2\mathscr{G}\mathscr{M}+\sigma^2\mathbb{I}\right)^{-1}\mathscr{E}_{\overline{y}} = \left(\mathbb{I}+\frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)^{-1}\mathscr{E}_{\overline{y}}.$$

$$\begin{tikzcd}
\mathbb{R}^{n_1 n_2} \arrow[r, leftrightarrow] \arrow[d, "G"'] & \mathbb{R}^{(n_1+2)(n_2+2)} \arrow[r, "\mathscr{M}\mathscr{E}"] \arrow[d, "\overline{G}"'] & \mathcal{C}([0,1]\times[0,1],\mathbb{R}) \arrow[d, "\mathscr{G}"'] \\
\mathbb{R}^{n_1 n_2} \arrow[r, leftrightarrow] & \mathbb{R}^{(n_1+2)(n_2+2)} & \mathcal{C}([0,1]\times[0,1],\mathbb{R}) \arrow[l, "\text{point evaluation}"']
\end{tikzcd}$$

Figure 5.1: The structure of the approximation (5.2). Notice that the diagram does not commute.

Unfortunately we have no knowledge of the quality of the approximation of the inverse term, other than that the error will vanish as the mesh of the observation grid disappear. Interestingly we have that

$$\begin{aligned}
\frac{1}{\tau^2}\mathscr{L}\widehat{\mathrm{E}}\left[\,x\,|\,y\,\right] &= \mathscr{M}\left(\tau^2\mathscr{G}\mathscr{M}+\sigma^2\mathbb{I}\right)^{-1}\mathscr{E}_{\overline{y}} \\
&= \frac{1}{\sigma^2}\mathscr{M}\left(\mathbb{I}-\tau^2\mathscr{G}\mathscr{M}\left(\tau^2\mathscr{G}\mathscr{M}+\sigma^2\mathbb{I}\right)^{-1}\right)\mathscr{E}_{\overline{y}} \\
&= \frac{1}{\sigma^2}\mathscr{M}\left(\mathscr{E}_{\overline{y}}-\widehat{\mathrm{E}}\left[\,x\,|\,y\,\right]\right), \qquad (5.3)
\end{aligned}$$

which shows us that even though our estimate is differentiable up to the order of $\mathscr{L}$, the derivative $\mathscr{L}\widehat{\mathrm{E}}\left[\,x\,|\,y\,\right]$ is only a scaling of the residual, and thus does not contain any information that is not already in $\widehat{\mathrm{E}}\left[\,x\,|\,y\,\right]$. The corresponding approximation of the conditional variance is

$$\begin{aligned}
\widehat{\mathrm{V}}\left[\,x\,|\,y\,\right] &= \tau^2\mathscr{G}-\tau^2\mathscr{G}\mathscr{M}\left(\tau^2\mathscr{G}\mathscr{M}+\sigma^2\mathbb{I}\right)^{-1}\tau^2\mathscr{G} \\
&= \tau^2\mathscr{G}-\left(\mathbb{I}+\frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)^{-1}\tau^2\mathscr{G} \\
&= \tau^2\mathscr{G}-\left(\mathbb{I}-\left(\mathbb{I}+\frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)^{-1}\frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)\tau^2\mathscr{G} \\
&= \tau^2\left(\frac{\tau^2}{\sigma^2}\mathscr{M}+\mathscr{L}\right)^{-1}. \qquad (5.4)
\end{aligned}$$

Recall that the likelihood can be written as

$$\ell_y(\sigma^2,\tau^2) = \frac{1}{2}\log\det\Sigma + \frac{1}{2\sigma^2}\left\{y_{ij}\right\}_{(i,j)}^{\top}\left\{y_{ij}-\mathrm{E}[\,x(s_i,t_j)\;|\,y\,]\right\}_{(i,j)},$$

with

$$\Sigma = \left\{\tau^2\mathcal{G}((s_i,t_j)(s_k,t_l))+\sigma^2\mathbb{1}_{(i,j)=(k,l)}\right\}_{(i,j),(k,l)}.$$

The quadratic term is approximated by

$$\frac{1}{2\sigma^2}\left\{y_{ij}\right\}_{(i,j)}^{\top}\left\{y_{ij}-\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)} \qquad (5.5)$$

which can be numerically unstable, but writing it as

$$\frac{1}{2\sigma^2}\left\{y_{ij}-\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}^{\top}_{(i,j)}\left\{y_{ij}-\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)} \\ +\frac{1}{2\sigma^2}\left\{\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}^{\top}_{(i,j)}\left\{y_{ij}-\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)}$$

and using (5.3) on the last matrix, we get that the term equals

$$\frac{1}{2\sigma^2}\left\{y_{ij}-\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}^{\top}_{(i,j)}\left\{y_{ij}-\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)} \\ +\frac{1}{2\tau^2}\left\{\widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}^{\top}_{(i,j)}\left\{\mathscr{M}^{-1}\mathscr{L}\widehat{\mathrm{E}}\left[\,x(s_i,t_j\,|\,y\,\right]\right\}_{(i,j)}.$$

Remembering that $\mathscr{L}=\mathscr{K}^{\dagger}\mathscr{K}$ one realise that the last matrix product is a square, and then we only have positive terms, which should solve the problem.

Note: In the `laplacesmooth` package the quadratic term approximation is implemented using (5.5).

For the determinant things are a bit more tricky, and one has to come up with a clever trick to make a reasonable approximation. First of all we have that

$$\begin{aligned}\log\det\Sigma &= \log\det\left\{\tau^2\mathcal{G}((s_i,t_j)(s_k,t_l))+\sigma^2\mathbb{1}_{(i,j)=(k,l)}\right\}_{(i,j),(k,l)}\\ &= \log\det\left\{\tau^2/\sigma^2\mathcal{G}((s_i,t_j)(s_k,t_l))+\mathbb{1}_{(i,j)=(k,l)}\right\}_{(i,j),(k,l)}+n_1n_2\log\sigma^2. \end{aligned}\tag{5.6}$$

Denote by Ξ the $n_1n_2\times n_1n_2$ matrix

$$\Xi=\left\{\tau^2/\sigma^2\mathcal{G}((s_i,t_j)(s_k,t_l))+\mathbb{1}_{(i,j)=(k,l)}\right\}.$$

By $\overline{\Xi}$ we will denote the $(n_1+2)(n_2+2)\times(n_1+2)(n_2+2)$ which also has elements corresponding to the boundary. Now introduce the matrix function

$$\Xi(\alpha)=\left\{\alpha\tau^2/\sigma^2\mathcal{G}((s_i,t_j)(s_k,t_l))+\mathbb{1}_{(i,j)=(k,l)}\right\}.$$

Differentiating $\log\det\Xi(\alpha)$ with respect to α yeilds

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\det\Xi(\alpha)=\Big(\det\Xi(\alpha)\Big)^{-1}\det\Xi(\alpha)\,\mathrm{tr}\left(\Xi(\alpha)^{-1}\frac{\mathrm{d}}{\mathrm{d}\alpha}\Xi(\alpha)\right),$$

where Jacobi's formula has been used. Now introduce the basis vectors $e_\ell=\{\mathbb{1}_{\ell=n}\}_n\in\mathbb{R}^{n_1n_2}$. Further denote by E_{ij} the $(n_1+2)\times(n_2+2)$ matrix with zero's everywhere except at the entry i,j (with the coordinate nummeration starting with 0) where the value is 1. It is easily obtained that

$$\log\det\Xi=\int_0^1\sum_{\ell=1}^{n_1n_2}e_\ell^{\top}\Xi(\alpha)^{-1}\left\{\tau^2/\sigma^2\mathcal{G}((s_i,t_j),(s_k,t_l)\right\}_{(i,j),(k,l)}e_\ell\,\mathrm{d}\alpha,$$

which can be approximated by

$$\int_0^1\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(\alpha\tau^2/\sigma^2\mathscr{G}\mathscr{M}+\mathbb{I}\right)^{-1}\tau^2/\sigma^2\mathscr{G}\mathscr{M}\mathscr{E}_{E_{ij}}(s_i,t_j)\,\mathrm{d}\alpha,$$

where the second basis vector acts as a matrix operation that is approximated by $\mathscr{M}\mathscr{E}_{E_{ij}}$ and the first as the point evaluation. One can further approximate this expression by substituting the sums with

integrals. For this an extra normalization operator $\mathscr{M}$ will need to be used on the $\mathscr{E}_{E_{ij}}$ term. As the mesh of the grid vanishes, the function $\mathscr{M}\mathscr{E}_{E_{ij}}$ converges to a Dirac delta function. The approximation becomes

$$\int_0^1 \int_{\frac{1}{n_2+1}}^{\frac{n_2}{n_2+1}} \int_{\frac{1}{n_1+1}}^{\frac{n_1}{n_1+1}} \left(\alpha\mathbb{I} + \frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)^{-1} \delta_{(s,t)}(s,t)\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha.$$

The obvious next step is to extend the range of integration so the expression becomes

$$\int_0^1 \int_0^1 \int_0^1 \left(\alpha\mathbb{I} + \frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)^{-1} \delta_{(s,t)}(s,t)\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha.$$

When the number of observations is large, this will make little difference, but for smaller datasets the difference (that E_{ij} has zero's along the boundary and $\delta_{(s,t)}$ does not share this characteristic) will be noticeable. Looking closer at what has happened, one realize that the above approximation is in fact of $\log\det\overline{\Xi}$.

The differential operator $\alpha\mathbb{I} + \frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}$ is invertible, so if we denote its Greens function by $\overline{\mathcal{G}}_\alpha$ we have that

$$\begin{aligned}\left(\alpha\mathbb{I} + \frac{\sigma^2}{\tau^2}\mathscr{M}^{-1}\mathscr{L}\right)^{-1} \delta_{(s,t)}(s,t) &= \int_0^1 \int_0^1 \overline{\mathcal{G}}_\alpha((s,t),(v,w))\delta_{(s,t)}(s,t)\,\mathrm{d}v\,\mathrm{d}w \\ &= \overline{\mathcal{G}}_\alpha((s,t),(s,t))\end{aligned}$$

so we have the following approximation

$$\log\det\overline{\Xi} \approx \int_0^1 \int_0^1 \int_0^1 \overline{\mathcal{G}}_\alpha((s,t),(s,t))\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha.$$

If one supposes that $\mathcal{G}$ satisfies Dirichlet boundary conditions, $\overline{\Xi}$ will have block structure with zero boundaries on the blocks (plus some ones on the diagonal.) Calculating the determinant of $\overline{\Xi}$ iteratively using the Laplace expansion and rules for determinants of block diagonal matrices, one easily realize that

$$\det\Xi = \det\overline{\Xi},$$

so in the case of Dirichlet boundary conditions the above approximation works well even for low observation numbers. If $\mathcal{G}$ does not satisfy Dirichlet boundary conditions, one can think of Ξ as being $\overline{\Xi}$ for a problem with $(n_1-2)(n_2-2)$ observations. Going through the approximation procedure we get that

$$\log\det\Xi \approx \int_0^1 \int_0^1 \int_0^1 \mathcal{G}^*_\alpha((s,t),(s,t))\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha.$$

where $\mathcal{G}^*_\alpha$ denotes the Greens function corresponding to the differential operator

$$\frac{\sigma^2}{\tau^2}\mathscr{M}_*^{-1}\mathscr{L} + \alpha\mathbb{I}$$

where $\mathscr{M}_* f = (n_1-1)(n_2-1)f$.

To conclude, in the general case we approximate the negative log likelihood $\hat{\ell}_y(\sigma^2,\tau^2)$ by

$$\begin{aligned}\hat{\ell}_y(\sigma^2,\tau^2) =& \frac{1}{2}\int_0^1 \int_0^1 \int_0^1 \mathcal{G}^*_\alpha((s,t),(s,t))\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha + \frac{n_1 n_2}{2}\log\sigma^2 \\ &+ \frac{1}{2\sigma^2}\left\{y_{ij}\right\}^\top_{(i,j)} \left\{y_{ij} - \widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)}\end{aligned} \tag{5.7}$$

and in the case of Dirichlet boundary conditions we use the approximation

$$\begin{aligned}\hat{\ell}_y(\sigma^2,\tau^2) =& \frac{1}{2}\int_0^1\int_0^1\int_0^1 \overline{\mathcal{G}}_\alpha((s,t),(s,t))\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha + \frac{n_1 n_2}{2}\log\sigma^2 \\ &+ \frac{1}{2\sigma^2}\left\{y_{ij}\right\}_{(i,j)}^{\top}\left\{y_{ij} - \widehat{\mathrm{E}}\left[\,x(s_i,t_j)\,|\,y\,\right]\right\}_{(i,j)}. \end{aligned} \tag{5.8}$$

CHAPTER 6

CHAPTER 5 IN PRACTICE

Example 6.1. To approximate the determinant as done in the previous section when $\mathscr{L} = (-\Delta)^\ell + \varepsilon\mathbb{I}$, we have to find the Green's function corresponding to

$$\frac{\sigma^2}{\tau^2}\mathscr{M}_*^{-1}\mathscr{L} + \alpha\mathbb{I} = \frac{\sigma^2}{\tau^2}\mathscr{M}_*^{-1}\left(\mathscr{L} + \frac{\alpha\tau^2}{\sigma^2}\mathscr{M}_*\right),$$

which is $\frac{\tau^2}{\sigma^2}\mathscr{M}_*$ times the Green's function for

$$(-\Delta)^\ell + \left(\varepsilon + \frac{\alpha\tau^2}{\sigma^2}\mathscr{M}_*\right)\mathbb{I},$$

but the corresponding Green's function $\mathcal{G}$ has already been found in Example 3.7 for both Dirichlet and Neumann boundary conditions, and so the Dirichlet Green's function becomes

$$\mathcal{G}_\alpha^*((s,t),(v,w)) = \frac{\tau^2}{\sigma^2}\mathscr{M}_*\sum_{n=1}^\infty\sum_{m=1}^\infty \frac{4}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon + \frac{\alpha\tau^2(n_1-1)(n_2-1)}{\sigma^2}} \times \sin(n\pi s)\sin(m\pi t)\sin(n\pi v)\sin(m\pi w)$$

and for the Neumann problem it becomes

$$\mathcal{G}_\alpha^*((s,t),(v,w)) = \frac{\tau^2}{\sigma^2}\mathscr{M}_*\sum_{n=1}^\infty\sum_{m=1}^\infty \frac{4}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon + \frac{\alpha\tau^2(n_1-1)(n_2-1)}{\sigma^2}} \times \cos(n\pi s)\cos(m\pi t)\cos(n\pi v)\cos(m\pi w).$$

We need to calculate the triple integral

$$\int_0^1\int_0^1\int_0^1 \mathcal{G}_\alpha^*((s,t),(s,t))\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha.$$

If $\ell \geq 2$ the Green's function becomes integrable on the diagonal, and for both the Dirichlet problem and the Neumann problem one obtains that

$$\int_0^1\int_0^1 \mathcal{G}_\alpha^*((s,t)(s,t))\,\mathrm{d}s\,\mathrm{d}t = \frac{\tau^2}{\sigma^2}\mathscr{M}_*\sum_{n=1}^\infty\sum_{m=1}^\infty \frac{1}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon + \frac{\alpha\tau^2(n_1-1)(n_2-1)}{\sigma^2}}.$$

Performing the final integration it is obtained that

$$\begin{aligned}\int_0^1\int_0^1\int_0^1 \mathcal{G}_\alpha^*((s,t)(s,t))\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}\alpha &= \frac{\tau^2}{\sigma^2}\mathscr{M}_*\sum_{n=1}^\infty\sum_{m=1}^\infty\int_0^1 \frac{1}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon + \frac{\alpha\tau^2(n_1-1)(n_2-1)}{\sigma^2}}\,\mathrm{d}\alpha \\ &= \sum_{n=1}^\infty\sum_{m=1}^\infty \log\left(1 + \frac{\tau^2}{\sigma^2}\frac{(n_1-1)(n_2-1)}{(\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon)}\right). \qquad (6.1)\end{aligned}$$

The above expression is clearly convergent for all $\varepsilon \geq 0$, as the fundamental inequality of the logarithm gives that

$$\log\left(1 + \frac{\tau^2}{\sigma^2}\frac{(n_1-1)(n_2-1)}{(\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon)}\right) \leq \frac{\tau^2}{\sigma^2}\frac{(n_1-1)(n_2-1)}{(\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon)}.$$

The role of ε in (6.1) is negligible compared to the other parameters, if it is chosen small, and in actual computations one can safely use $\varepsilon = 0$.

The specific approximation for Dirichlet boundary conditions becomes

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \log\left(1 + \frac{\tau^2}{\sigma^2}\frac{(n_1+1)(n_2+1)}{(\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon)}\right). \tag{6.2}$$

∘

Example 6.2. Now let us move to the conditional expectation approximation $\widehat{\mathrm{E}}\,[\,x\,|\,y\,]$. With the choice $\mathscr{L} = (-\Delta)^\ell + \varepsilon\mathbb{I}$ we have that

$$\widehat{\mathrm{E}}\,[\,x\,|\,y\,]\,(s,t) = \frac{\tau^2}{\sigma^2}\mathscr{M}\left((-\Delta)^\ell + \left(\varepsilon + \frac{\tau^2}{\sigma^2}\mathscr{M}\right)\mathbb{I}\right)^{-1}\mathscr{E}_y(s,t).$$

For Dirichlet boundary conditions we get that $\left((-\Delta)^\ell + \left(\varepsilon + \frac{\tau^2}{\sigma^2}\mathscr{M}\right)\mathbb{I}\right)^{-1}\mathscr{E}_y(s,t)$ equals

$$\sum_{n=1}^{\infty}\sum_{n=1}^{\infty}\frac{4}{\pi^{2\ell}(n^2+m^2)^\ell + \varepsilon + \frac{\tau^2}{\sigma^2}(n_1+1)(n_2+1)}$$
$$\times\int_0^1\int_0^1 \sin(n\pi s)\sin(m\pi t)\sin(n\pi v)\sin(m\pi w)\mathscr{E}_y(v,w)\,\mathrm{d}v\,\mathrm{d}w.$$

Since $\mathscr{E}_y(\,\cdot\,,w)$ is linear on the segments $[s_i, s_{i+1}]$ for $0 \le i \le n_1$ there exist a and b (functions of i and j) such that

$$\begin{aligned}\int_0^1 \sin(n\pi v)\mathscr{E}_y(v,w)\,\mathrm{d}v &= \sum_{i=0}^{n_1}\int_{s_i}^{s_{i+1}} \sin(n\pi v)\mathscr{E}_y(v,w)\,\mathrm{d}v\\ &= \sum_{i=0}^{n_1}\int_{s_i}^{s_{i+1}} \sin(n\pi v)(a(v-s_i)+b)\,\mathrm{d}v\\ &= \sum_{i=0}^{n_1}\Bigg(\frac{a}{n^2\pi^2}(\sin(n\pi s_{i+1}) - \sin(n\pi s_i)) - \frac{a}{n\pi}(s_{i+1}-s_i)\cos(n\pi s_{i+1})\\ &\qquad - \frac{b}{n\pi}(\cos(n\pi s_{i+1}) - \cos(n\pi s_i))\Bigg).\end{aligned}$$

The line between (s_i, t_j) and (s_i, t_{j+1}) is given by

$$a_1(w-t_j) + b_1 = \frac{y_{i,j+1} - y_{ij}}{t_{j+1}-t_j}(w-t_j) + y_{ij}$$

and between (s_{i+1}, t_j) and (s_{i+1}, t_{j+1}) it is given by

$$a_2(w-t_j) + b_2 = \frac{y_{i+1,j+1} - y_{i+1,j}}{t_{j+1}-t_j}(w-t_j) + y_{i+1,j}.$$

From this we have that if $w \in [t_j, t_{j+1}]$ then

$$a = \frac{a_2(w-t_j) + b_2 - a_1(w-t_j) - b_1}{s_{i+1}-s_i} = \frac{(a_2-a_1)(w-t_j) + b_2 - b_1}{s_{i+1}-s_i}$$

and

$$b = a_1(w-t_j) + b_1.$$

We have that

$$\int_0^1 \int_0^1 \sin(n\pi v)\sin(m\pi w)\mathscr{E}_y(v,w)\,\mathrm{d}v\,\mathrm{d}w$$
$$= \sum_{j=0}^{n_2}\sum_{i=0}^{n_1}\int_{t_j}^{t_{j+1}}\left(\frac{a}{n^2\pi^2}(\sin(n\pi s_{i+1})-\sin(n\pi s_i)) - \frac{a}{n\pi}(s_{i+1}-s_i)\cos(n\pi s_{i+1})\right.$$
$$\left. - \frac{b}{n\pi}(\cos(n\pi s_{i+1})-\cos(n\pi s_i))\right)\sin(m\pi w)\,\mathrm{d}w,$$

where the integral of the a terms $\int_{t_j}^{t_{j+1}} a\sin(m\pi w)\,\mathrm{d}w$ can be computed as

$$\frac{1}{s_{i+1}-s_i}\int_{t_j}^{t_{j+1}}((a_2-a_1)(w-t_j)+b_2-b_1)\sin(m\pi w)\,\mathrm{d}w$$
$$= \frac{1}{s_{i+1}-s_i}\left(\frac{a_2-a_1}{m^2\pi^2}(\sin(m\pi t_{j+1})-\sin(m\pi t_j)) - \frac{(a_2-a_1)(t_{j+1}-t_j)}{m\pi}\cos(m\pi t_{j+1})\right.$$
$$\left. - \frac{b_2-b_1}{m\pi}(\cos(m\pi t_{j+1})-\cos(m\pi t_j))\right).$$

For the b term we have that $\int_{t_j}^{t_{j+1}} b\sin(m\pi w)\,\mathrm{d}w$ equals

$$\int_{t_j}^{t_{j+1}}(a_1(w-t_j)+b_1)\sin(m\pi w)\,\mathrm{d}w$$
$$= \left(\frac{a_1}{m^2\pi^2}(\sin(m\pi t_{j+1})-\sin(m\pi t_j)) - \frac{a_1(t_{j+1}-t_j)}{m\pi}\cos(m\pi t_{j+1})\right.$$
$$\left. - \frac{b_1}{m\pi}(\cos(m\pi t_{j+1})-\cos(m\pi t_j))\right).$$

So (prepare for the horror) we have that

$$\widehat{\mathrm{E}}\,[\,x\,|\,y\,]\,(s,t) = \frac{\tau^2}{\sigma^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{4(n_1+1)(n_2+1)}{\pi^{2\ell}(n^2+m^2)^\ell+\varepsilon+\frac{\tau^2}{\sigma^2}(n_1+1)(n_2+1)}\sin(n\pi s)\sin(m\pi t)$$
$$\times\sum_{j=0}^{n_2}\sum_{i=0}^{n_1}\left(\frac{1}{s_{i+1}-s_i}\left(\frac{a_2-a_1}{m^2\pi^2}(\sin(m\pi t_{j+1})-\sin(m\pi t_j))\right.\right.$$
$$\left. - \frac{(a_2-a_1)(t_{j+1}-t_j)}{m\pi}\cos(m\pi t_{j+1}) - \frac{b_2-b_1}{m\pi}(\cos(m\pi t_{j+1})-\cos(m\pi t_j))\right)$$
$$\times\left(\frac{1}{n^2\pi^2}(\sin(n\pi s_{i+1})-\sin(n\pi s_i)) - \frac{1}{n\pi}(s_{i+1}-s_i)\cos(n\pi s_{i+1})\right)$$
$$-\left(\frac{a_1}{m^2\pi^2}(\sin(m\pi t_{j+1})-\sin(m\pi t_j)) - \frac{a_1(t_{j+1}-t_j)}{m\pi}\cos(m\pi t_{j+1})\right.$$
$$\left.\left. - \frac{b_1}{m\pi}(\cos(m\pi t_{j+1})-\cos(m\pi t_j))\right)\frac{1}{n\pi}(\cos(n\pi s_{i+1})-\cos(n\pi s_i))\right).$$

With the notation

$$\mathfrak{c}_i(n) = \cos(n\pi s_{i+1}) - \cos(m\pi s_i) = -2\sin\left(\frac{n\pi}{2(n_1+1)}\right)\sin\left(\frac{n\pi(2i+1)}{2(n_1+1)}\right),$$

$$\mathfrak{c}_j(m) = \cos(m\pi t_{j+1}) - \cos(m\pi t_j) = -2\sin\left(\frac{m\pi}{2(n_2+1)}\right)\sin\left(\frac{m\pi(2j+1)}{2(n_2+1)}\right),$$

$$\mathfrak{sc}_i(n) = \frac{n_1+1}{n\pi}\big(\sin(n\pi s_{i+1}) - \sin(n\pi s_i)\big) - \cos(n\pi s_{i+1}),$$

$$\mathfrak{sc}_j(m) = \frac{n_2+1}{m\pi}\big(\sin(m\pi t_{j+1}) - \sin(m\pi t_j)\big) - \cos(m\pi t_{j+1}),$$

it is obtained (with some calculations) that

$$\begin{aligned}\widehat{\mathrm{E}}\,[\,x\,|\,y\,]\,(s,t) &= \frac{\tau^2}{\sigma^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{4(n_1+1)(n_2+1)}{\pi^{2\ell}(n^2+m^2)^\ell+\varepsilon+\frac{\tau^2}{\sigma^2}(n_1+1)(n_2+1)}\sin(n\pi s)\sin(m\pi t)\\ &\times\frac{1}{nm\pi^2}\sum_{j=0}^{n_2}\sum_{i=0}^{n_1}\Bigg((y_{i+1,j+1}-y_{i+1,j})\mathfrak{sc}_i(n)\mathfrak{sc}_j(m)-(y_{i,j+1}-y_{ij})(\mathfrak{sc}_i(n)+\mathfrak{c}_i(n))\mathfrak{sc}_j(m)\\ &\qquad -(y_{i+1,j}-y_{ij})\mathfrak{c}_j(m)\mathfrak{sc}_i(n)+y_{ij}\mathfrak{c}_i(n)\mathfrak{c}_j(m)\Bigg).\end{aligned} \tag{6.3}$$

For Neumann boundary conditions the expression is

$$\begin{aligned}\widehat{\mathrm{E}}\,[\,x\,|\,y\,]\,(s,t) &= \frac{\tau^2}{\sigma^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{4(n_1+1)(n_2+1)}{\pi^{2\ell}(n^2+m^2)^\ell+\varepsilon+\frac{\tau^2}{\sigma^2}(n_1+1)(n_2+1)}\sin(n\pi s)\sin(m\pi t)\\ &\times\frac{1}{nm\pi^2}\sum_{j=0}^{n_2}\sum_{i=0}^{n_1}\Bigg((y_{i+1,j+1}-y_{i+1,j})\mathfrak{cs}_i(n)\mathfrak{cs}_j(m)-(y_{i,j+1}-y_{ij})(\mathfrak{cs}(n)-\mathfrak{s}_i(n))\mathfrak{cs}_j(m)\\ &\qquad +(y_{i+1,j}-y_{ij})\mathfrak{s}_j(m)\mathfrak{cs}_i(n)+y_{ij}\mathfrak{s}_i(n)\mathfrak{s}_j(m)\Bigg)\end{aligned} \tag{6.4}$$

where

$$\mathfrak{s}_i(n) = \sin(n\pi s_{i+1}) - \sin(m\pi s_i) = 2\sin\left(\frac{n\pi}{2(n_1+1)}\right)\cos\left(\frac{n\pi(2i+1)}{2(n_1+1)}\right),$$

$$\mathfrak{s}_j(m) = \sin(m\pi t_{j+1}) - \sin(m\pi t_j) = 2\sin\left(\frac{m\pi}{2(n_2+1)}\right)\cos\left(\frac{m\pi(2j+1)}{2(n_2+1)}\right),$$

$$\mathfrak{cs}_i(n) = \frac{n_1+1}{n\pi}\big(\cos(n\pi s_{i+1}) - \cos(n\pi s_i)\big) + \sin(n\pi s_{i+1}),$$

$$\mathfrak{cs}_j(m) = \frac{n_2+1}{m\pi}\big(\cos(m\pi t_{j+1}) - \cos(m\pi t_j)\big) + \sin(m\pi t_{j+1}).$$

Note: $\widehat{\mathrm{E}}\,[\,x\,|\,y\,]$ is only implemented with the Dirichlet boundary conditions in the `laplacesmooth` package. ◦

Example 6.3. The conditional residuals in the model are given by

$$y_{ij} - \widehat{\mathrm{E}}\,[\,x(s_i,t_j)\,|\,y\,] = \widehat{\mathrm{E}}\,[\,y_{ij} - x(s_i,t_j)\,|\,y\,]\,.$$

We have that

$$\begin{aligned}\operatorname{Cov}\left[\widehat{\mathrm{E}}\left[\,y_{ij}-x(s_i,t_j)\,|\,y\,\right],\widehat{\mathrm{E}}\left[\,y_{kl}-x(s_k,t_l)\,|\,y\,\right]\right] &= \operatorname{Cov}\left[y_{ij}-x(s_i,t_j),y_{kl}-x(s_k,t_l)\right]\\ &\quad-\mathrm{E}\left[\operatorname{Cov}\left[y_{ij}-x(s_i,t_j),y_{kl}-x(s_k,t_l)\middle|\,y\right]\right]\\ &= \sigma^2\mathbb{1}_{(i,j)=(k,l)}-\operatorname{Cov}\left[x(s_i,t_j),x(s_k,t_l)\middle|\,y\right],\end{aligned}$$

where the law of total covariance has been used, as well as the fact that conditional covariances given the observation y does not depend on y in Gaussian models. This covariance can be approximated by

$$\sigma^2\mathbb{1}_{(i,j)=(k,l)}-\widehat{\mathrm{V}}\left[\,x\,|\,y\,\right]\delta_{(s_k,t_l)}(s_i,t_j).$$

For $\mathscr{L}=(-\Delta)^\ell+\varepsilon\mathbb{I}$ we have that

$$\widehat{\mathrm{V}}\left[\,x\,|\,y\,\right]=\tau^2\left((-\Delta)^\ell+\left(\varepsilon+\frac{\tau^2}{\sigma^2}\mathscr{M}\right)\mathbb{I}\right)^{-1},$$

so for Dirichlet boundary conditions we have that

$$\begin{aligned}\widehat{\mathrm{V}}\left[\,x\,|\,y\,\right]\delta_{(s_k,t_l)}(s_i,t_j)=\tau^2\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}&\frac{4}{\pi^{2\ell}(n^2+m^2)^\ell+\varepsilon+\frac{\tau^2(n_1+1)(n_2+1)}{\sigma^2}}\\ &\times\sin(n\pi s_i)\sin(m\pi t_j)\sin(n\pi s_k)\sin(m\pi t_l).\end{aligned}\tag{6.5}$$

The approximation for Neumann boundary conditions is similar.

We can now introduce the approximated two-dimensional autocovariance function given by

$$\widehat{\gamma}(n,m)=\sigma^2\mathbb{1}_{(n,m)=(0,0)}-\frac{1}{(n_1-n)(n_2-m)}\sum_{k=1}^{n_1-n}\sum_{l=1}^{n_2-m}\widehat{\mathrm{V}}\left[\,x\,|\,y\,\right]\delta_{(s_k,t_l)}(s_{k+n},t_{l+m})\tag{6.6}$$

and the associated two-dimensional autocorrelation function

$$\widehat{\rho}(n,m)=\frac{\widehat{\gamma}(n,m)}{\widehat{\gamma}(0,0)}.\tag{6.7}$$

These can be compared to their empirical counterparts to do model validation. The empirical auto covariance can be defined by

$$\begin{aligned}\widehat{\widehat{\gamma}}(n,m)=&\frac{1}{(n_1-n)(n_2-m)}\\ &\times\sum_{k=1}^{n_1-n}\sum_{l=1}^{n_2-m}\left(y_{kl}-\widehat{\mathrm{E}}\left[\,x(s_k,t_l)\,|\,y\,\right]-\overline{r}\right)\left(y_{k+n,l+m}-\widehat{\mathrm{E}}\left[\,x(s_{k+n},t_{l+m})\,|\,y\,\right]-\overline{r}\right),\end{aligned}\tag{6.8}$$

where $\overline{r}$ is an estimate of the overall expectation of $y-\widehat{\mathrm{E}}\left[\,x\,|\,y\,\right]$, it could for example be the average value of $y-\widehat{\mathrm{E}}\left[\,x\,|\,y\,\right]$ in all observation points. ∘

We now have all the ingredients for performing functional data analysis using the approximations, in the next example we will consider the quality of the approximations.

Example 6.4. We will now consider the benefits and quality of the approximations introduced in the previous chapter. We will assume that $\mathscr{L}=(-\Delta)^\ell$ with suitable boundary conditions. As previously mentioned

$$\log\det\Sigma=\log\det\left\{\tau^2/\sigma^2\mathcal{G}((s_i,t_j)(s_k,t_l))+\mathbb{1}_{(i,j)=(k,l)}\right\}_{(i,j),(k,l)}+n_1n_2\log\sigma^2.$$

We will compare $\log\det\Sigma - n_1 n_2 \log\sigma^2$ to the approximations (6.1) and (6.2), which are all functions of τ^2/σ^2. In Figure 6.1 $\log\det\Sigma - n_1 n_2 \log\sigma^2$ is plotted as a function of τ^2/σ^2 for both Dirichlet and Neumann boundary conditions and the approximations are also plotted. The approximation error can be found in Figure 6.2. It is seen that the two approximations are almost indistinguishable from the exact expression already when the number of observations is 100. The convergence seems a bit faster for higher values of ℓ, and it is seen that for $\ell = 3$ the approximation for Dirichlet boundary conditions is much better than the corresponding approximation for the Neumann boundary condition.

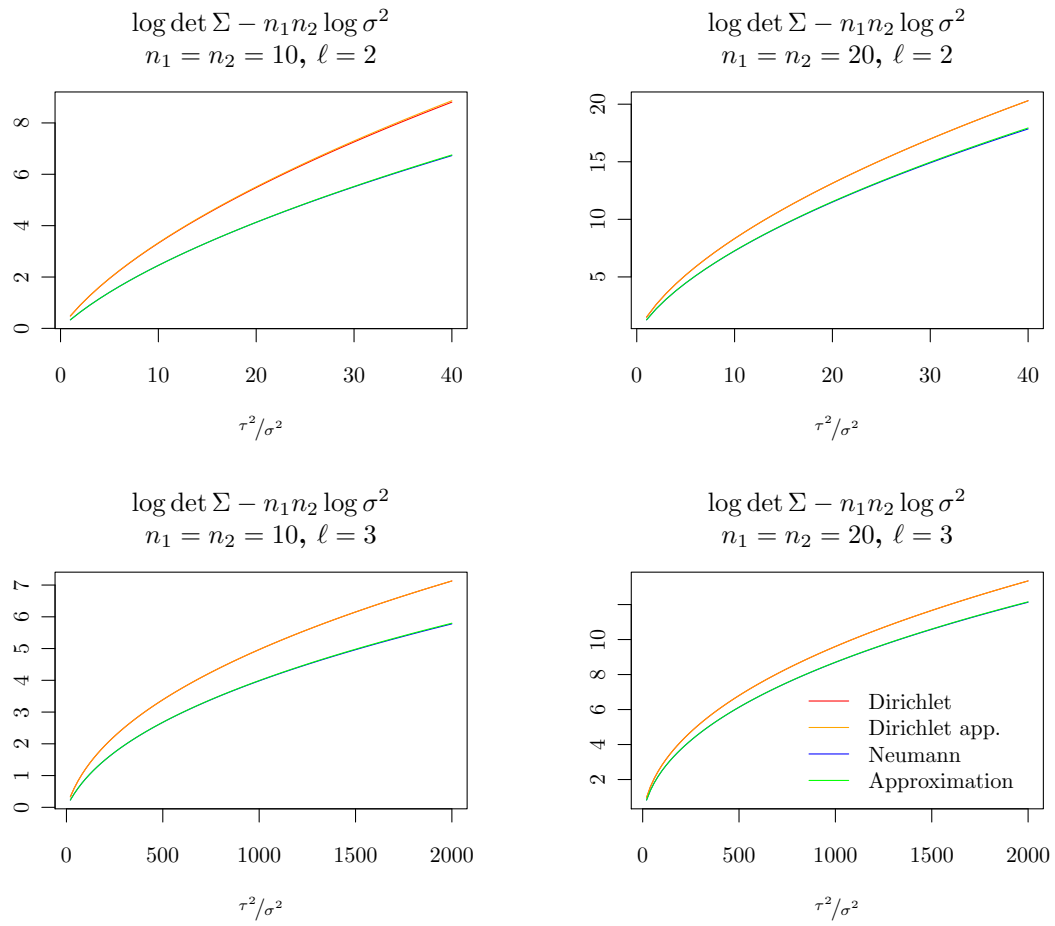

Figure 6.1: Comparisons of the exact and approximated log-determinant as functions of τ^2/σ^2.

Determinant computations are typically implemented with time complexity $O(n^3)$ where n is the height or length of the matrix (i.e. $n_1 n_2$), but the approximations requires only a constant time computation. In Figure 6.3 the computation times are compared, and it is clear that the approximation (green) is an amazing improvement.

Now let us move to the conditional expectation $\mathrm{E}[\,x\,|\,y\,]$ and the corresponding approximation $\widehat{\mathrm{E}}[\,x\,|\,y\,]$. The computation of $\mathrm{E}[\,x\,|\,y\,]$ in every observation point requires a matrix multiplication of two $n_1 n_2 \times n_1 n_2$ matrices, a lower order matrix computation and one inversion of a $n_1 n_2 \times n_1 n_2$ matrix. Although asymptotically faster algorithms exist the two dominating operations are typically both implemented to have running time $O(n_1^3 n_2^3)$, which is then the overall running time of the computation. $\widehat{\mathrm{E}}[\,x\,|\,y\,]$ on the other hand can be implemented to run in linear time $O(n_1 n_2)$. In Figure 6.4 a comparison of the running time of the two expression can be found. In all fairness the matrix computations done in the

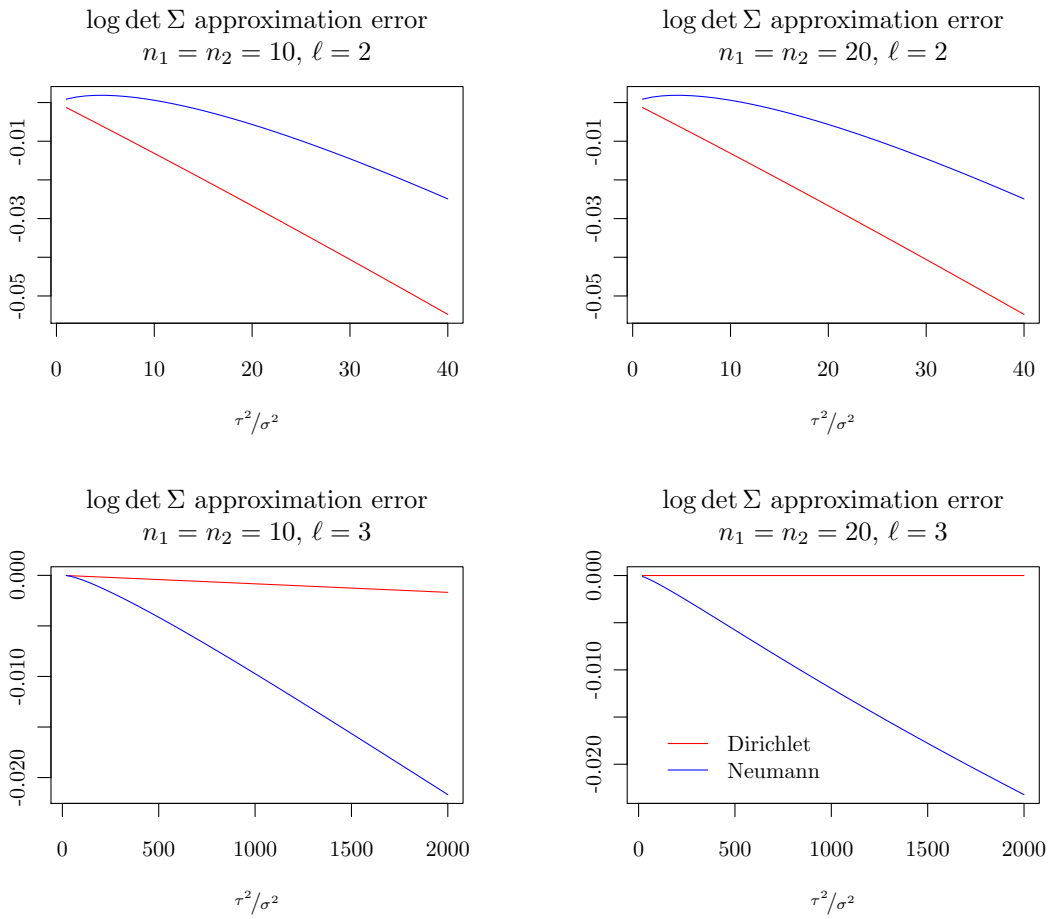

Figure 6.2: Approximation error from Figure 6.1.

calculation of $\mathrm{E}[\,x\,|\,y\,]$ is implemented to take use of multithreading, and for the larger matrices the actual experienced running time will be divided by the number of cores in the CPU. The quality of the corresponding approximation will be reviewed in the next example.

One aspect has been left out so far. For the two previous types of exact computations the construction of a covariance matrix is needed. Of course this is a one time investment, but actually it can be a rather costly one as can be seen in Figure 6.5. It is not only costly to construct a covariance matrix, but it also has storage demand $O(n_1^2 n_2^2)$, which can be problematic for very big images. This just speaks in favor of using the approximations.

Of course the implementation found in the `laplacesmooth` package is not optimal, but it somewhat optimized. A number of methods have been implemented in C which has decreased running times for some functions up to a factor 150. The speed of the various matrix computations could be decreased further by using e.g. QR decomposition. A further speed increase in the range of a factor 10-100 could probably be achieved by doing the computations on graphics processing units rather than on CPUs, since these are specially designed for matrix algebra. Even with these optimizations it is still infeasible to do exact smoothing on pictures in the megapixel range, and the presented approximations should be a long-awaited step forward in the realm of doing statistics on images.

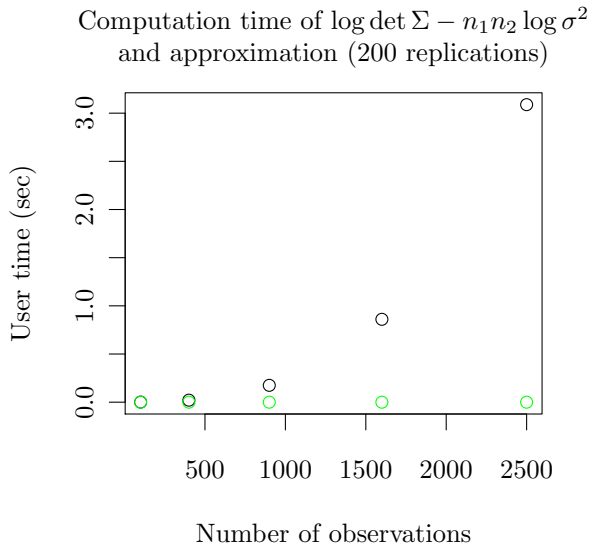

Figure 6.3: Computation time of $\log\det\Sigma - n_1 n_2 \log\sigma^2$ (black) compared to one of the approximations (6.1) (green). Values are averages of 200 replications.

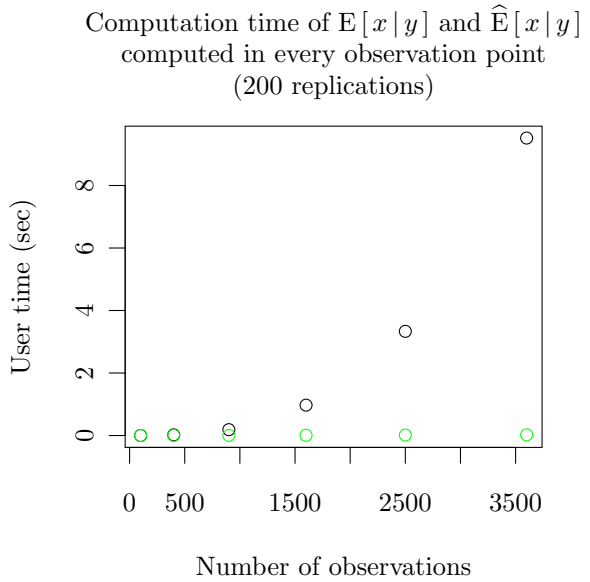

Figure 6.4: Computation time of $\mathrm{E}[x \mid y]$ (black) compared to the approximation $\widehat{\mathrm{E}}[x \mid y]$ (green). Values are averages of 200 replications.

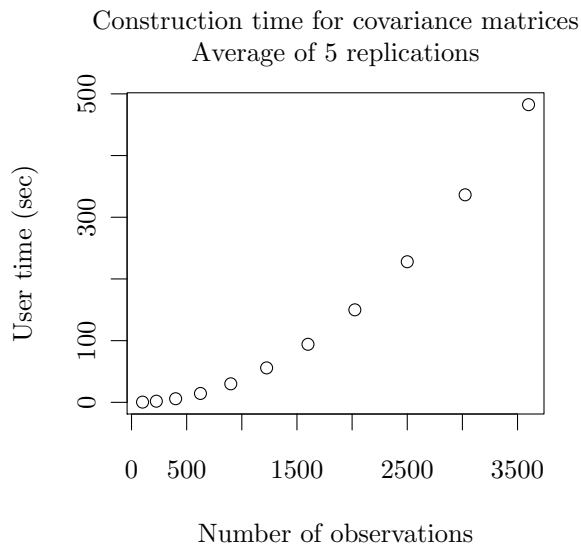

Figure 6.5: Construction time of a covariance matrix.

○

Example 6.5. We consider again the noisy image from Example 4.1

$$y_{ij} = \theta(s_i, t_j) + \varepsilon_{ij}, \qquad i, j = 1, \ldots, 21$$

where

$$\theta(s, t) = 70s^2(s-1)t(t-1)$$

and the ε_{ij}'s are independent and identically normal distributed random variables with standard deviation 0.5. Plots of θ and the noisy observation can be found in Figure 4.1 in Example 4.1. A plot of θ can also be found in Figure 6.6 on the next page. We will focus on the performance of the approximations introduced in the previous chapter. The infinite double series occurring in all approximations has been cut off at each sums 20th term in the following. Using the approximated likelihood the parameters are estimated for the observation from Example 4.1, and can be found in the table below.

ℓ	$\hat{\sigma}$	$\hat{\tau}$
2	0.5037203	7.003465
3	0.5101308	58.34387
4	0.5113562	449.5905

Figure 6.6 holds a plot of the approximated smooths for $\ell = 2, 3, 4$.

The parameter estimates and smooths are produced with the `laplacesmooth` package using the code in Listing 6.1 where `pic` refers to the matrix of observations

Listing 6.1: Approximated smoothing an image and plotting the smooth

```
size <- dim(pic)
est <- appestimate(pic, order = 2)
persp(x = steps, y = steps, z = ceamatrix(pic, est$par[1], est$par[2],  order = 2))
```

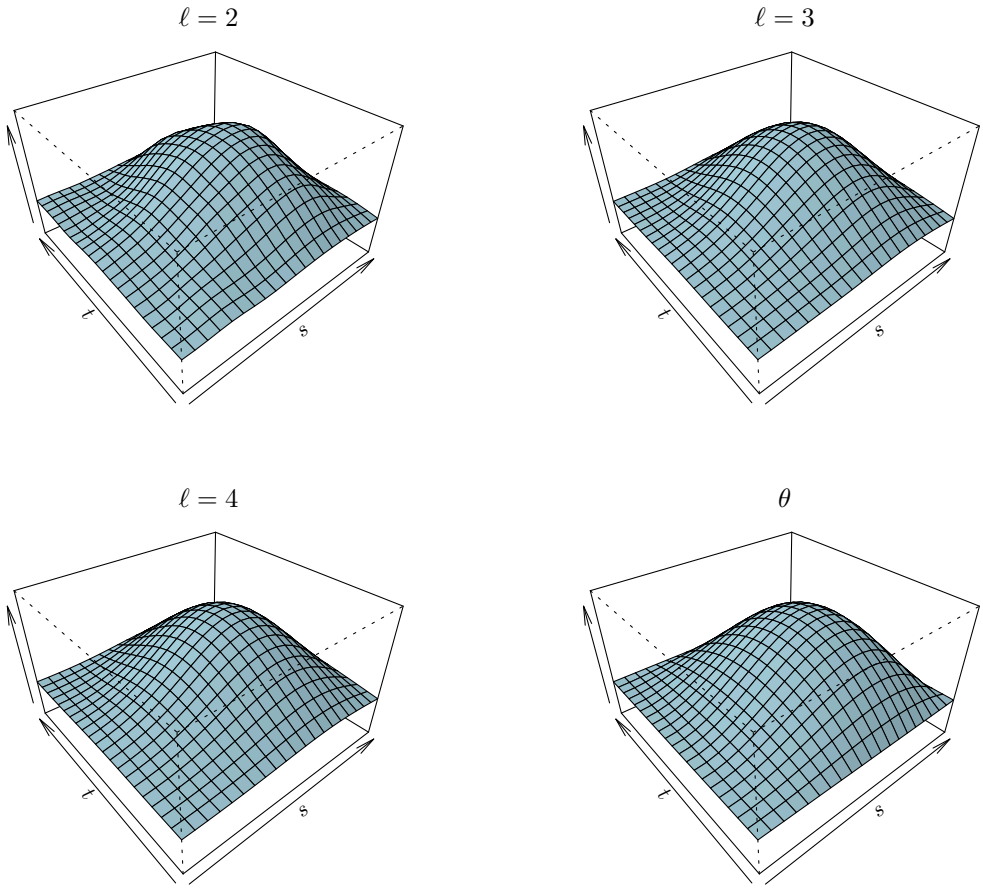

Figure 6.6: The approximated smooths of the observed image for different values of ℓ and the true image.

The partial derivatives of $\widehat{\mathrm{E}}\,[\,x\,|\,y\,]$ are easily calculated from the formula (6.3) and the corresponding estimates can be found in Figure 6.7 and 6.8. It is seen that even though the number of observations is very low compared to standard data sizes in image analysis, the smooths and partial derivatives are almost as good as the ones in Example 4.1, with the only difference noticeable to the naked eye being a slightly worse estimate of $\partial_t\,\theta$ when $\ell = 4$.

Considering the absolute difference between the true and the estimated picture, we get the values in the table below

ℓ	∂	Absolute difference	Absolute difference per observation
2	–	46.12313	0.1045876
3	–	44.90465	0.1018246
4	–	44.64823	0.1012432
2	∂_s	432.1687	0.9799744
3	∂_s	362.8343	0.8227535
4	∂_s	348.2625	0.7897109
2	∂_t	423.5122	0.9603451
3	∂_t	375.8897	0.8523576
4	∂_t	365.572	0.8289615

These numbers are somewhat worse than the corresponding numbers in Example 4.1, but this is not

alarming. The number of observations is quite small in this example, and results are better when the number of observations is higher. Also, the profit in terms of reduced computation time is really worth to notice. Looking an identical example with 100×100 observations instead of the 21×21, one get the absolute differences in the below table

ℓ	∂	Absolute difference	Absolute difference per observation
2	–	314.4681	0.03144681
3	–	277.3654	0.02773654
4	–	250.4005	0.02504005
2	∂_s	5652.123	0.5652123
3	∂_s	3898.208	0.3898208
4	∂_s	2978.498	0.2978498
2	∂_t	4430.046	0.4430046
3	∂_t	2652.161	0.2652161
4	∂_t	1799.494	0.1799494

These numbers are considerably better, although the improvement measured on the partial derivatives is not as impressive as on the image itself.

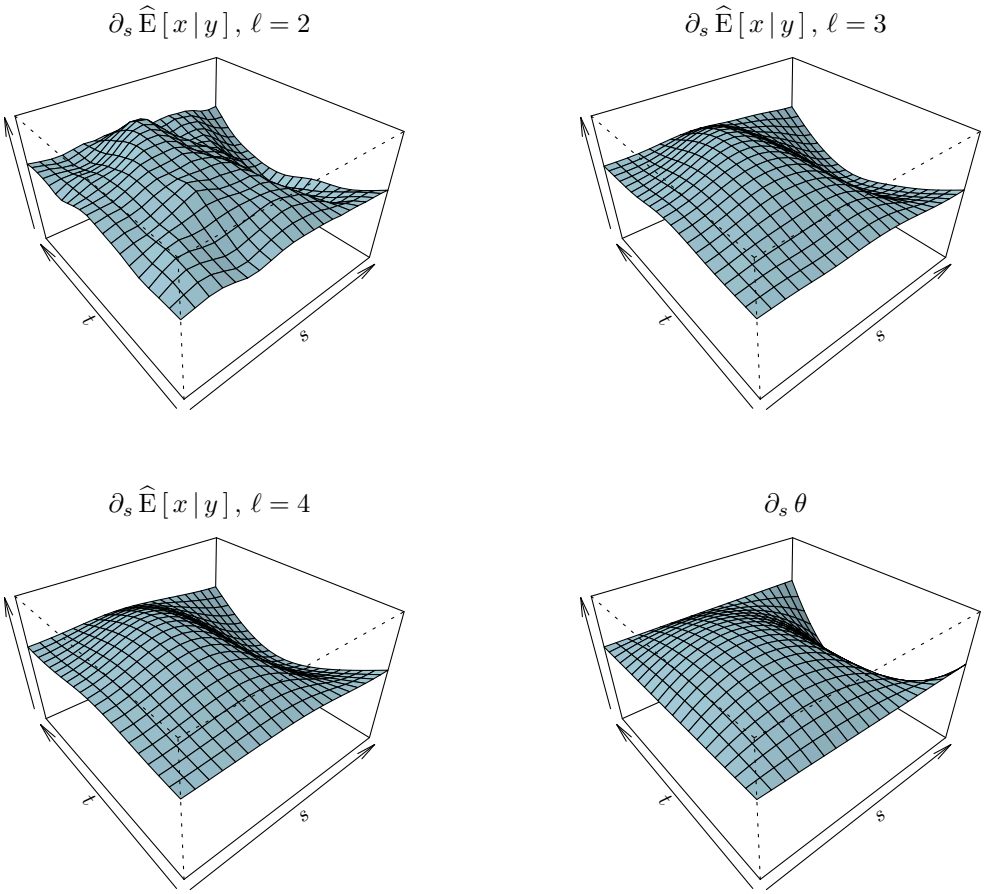

Figure 6.7: The partial derivatives in the s direction of the approximated smooths for different values of ℓ and the true partial derivative of θ.

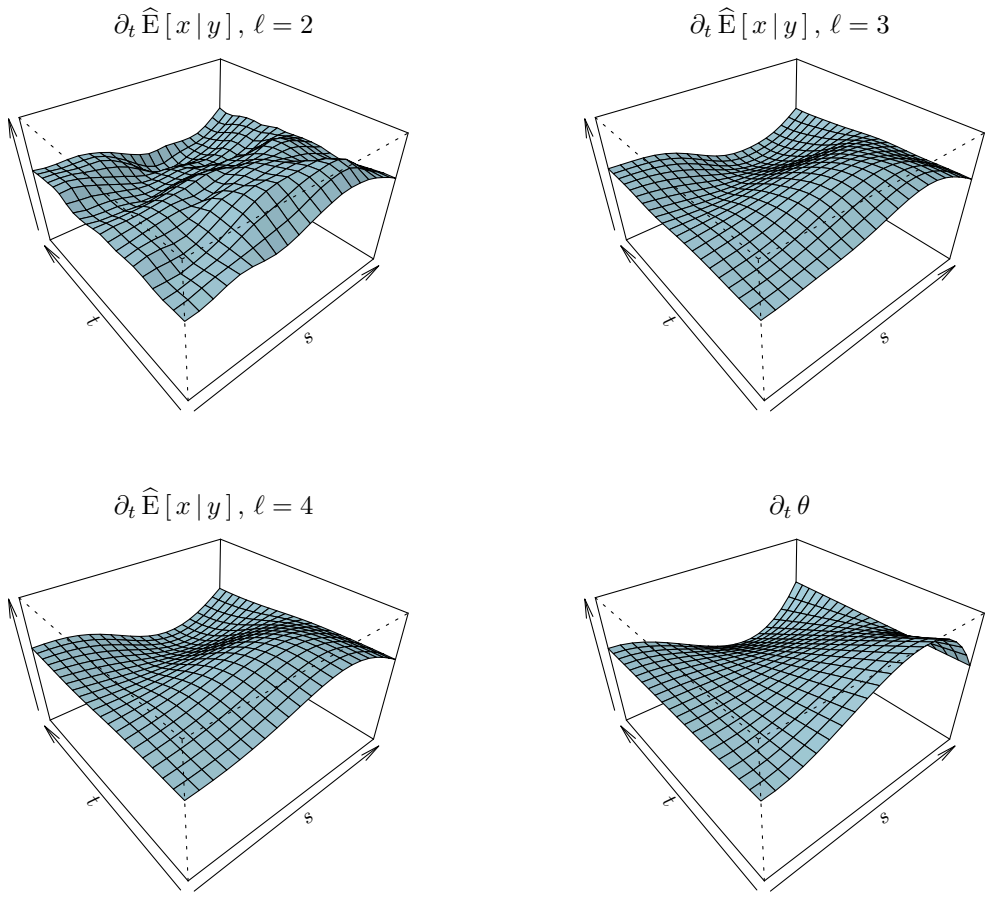

Figure 6.8: The partial derivatives in the t direction of the approximated smooths for different values of ℓ and the true partial derivative of θ.

○

Example 6.6. For the final example of this chapter, we will look at the autocovariance functions. We simulate a 40×40 Gaussian image x with covariance function $\tau^2\mathcal{G}$, where $\mathcal{G}$ is the Green's function for the biharmonic operator subject to Dirichlet boundary conditions, and add standard normal distributed noise to the image. A simulated image where $\tau = 100$ is found in Figure 6.9.

The smoothing is done in the approximated framework, and the parameter estimates $\hat{\sigma}$ and $\hat{\tau}$ are 1.197052 and 77.875697 respectively. The smooth and residuals can be found in Figure 6.10

We can now take a look at the theoretical and the empirical autocovariance images given by (6.6) and (6.8) respectively, which can be found in Figure 6.11. One notices that the empirical autocovariance becomes quite rough for n and m large. This is natural since the estimates of the covariances are based on fewer and fewer observations as n and m increase, and one should only compare the theoretical and empirical functions for n and m reasonably small. Finally it should be noted that the estimation of the autocovariance function becomes much better for larger numbers of observations.

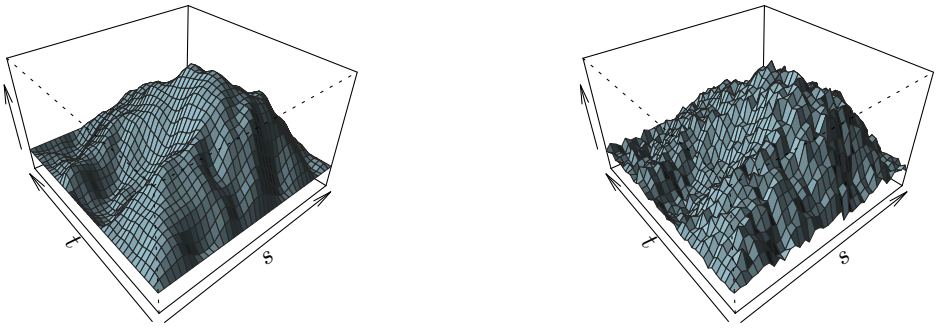

Figure 6.9: The Gaussian process x (left) and the process plus standard normal distributed noise.

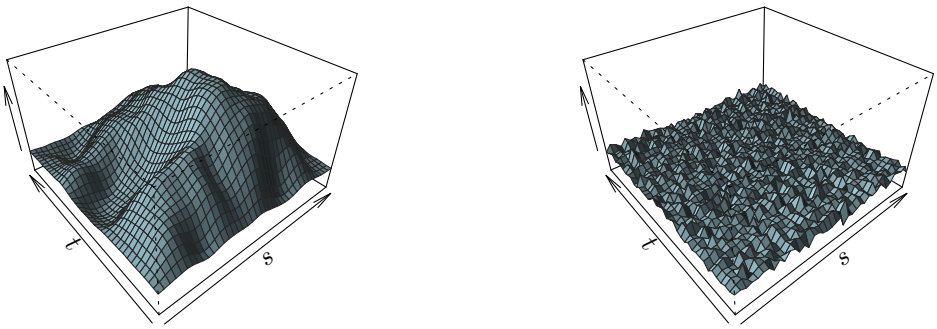

Figure 6.10: The smooth of the image from Figure 6.9 (left) and the corresponding residuals.

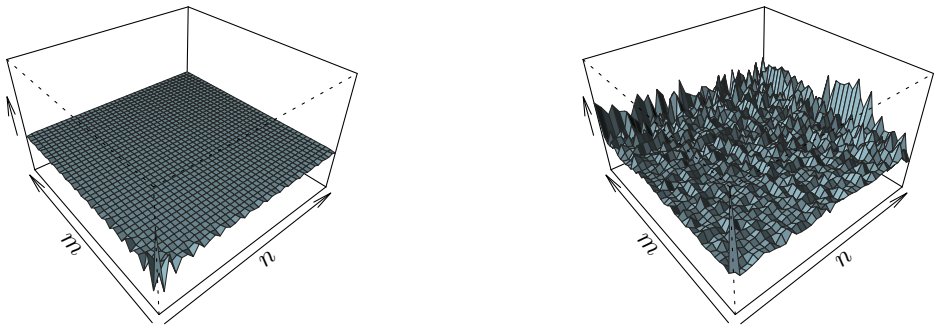

Figure 6.11: Theoretical autocovariance (left) and empirical for $n = 1, \ldots, 39$ and $m = 1, \ldots, 39$.

∘

CHAPTER 7

APPLICATIONS TO 2D ELECTROPHORESIS IMAGES

Two-dimensional (gel) electrophoresis is a technique used for separating proteins, DNA or RNA by two properties. The two properties could be isoelectric point and protein mass, but other choices are possible. Two segments from different 2D electrophoresis images of pig tissue (Lametsch et al. 2006) can be found in Figure 7.1. The chosen regions are nice in the sense that there are no extreme errors (e.g. black lines or thick white lines) as often seen in this kind of images. The images are quite noisy and there is for example an obvious error in the lower right corner of the left picture. A problem in the analysis of such images is determining wether something is in fact a protein spot or if it is just noise. That problem will be addressed in this chapter.

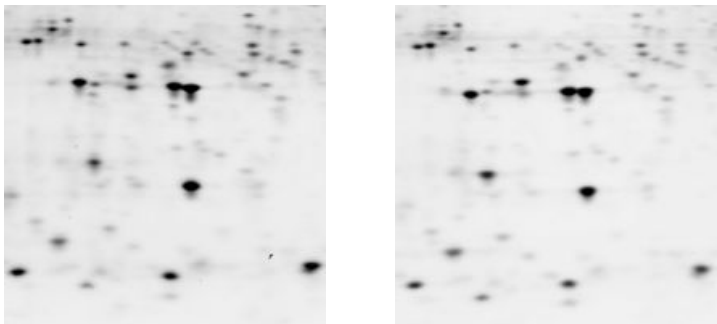

Figure 7.1: 400×400 pixel subregions of two 2D electrophoresis images. The protein is from pork.

The images in Figure 7.1 are plotted as surfaces in Figure 7.2. It is clear that the surfaces are not Gaussian, so something else should probably be done. We propose the model of the image y with N spots to be

$$y_{ij} = \sum_{k=1}^{N} \theta_k(s_i, t_j) + x(s_i, t_j) + \varepsilon_{ij} \tag{7.1}$$

where θ_k models spot k, x is Gaussian with covariance $\tau^2\mathcal{G}$ and the ε_{ij}'s are independent normal distributed variables with zero mean and variance σ^2. The obvious choice of θ_k that comes to mind is a Gaussian function

$$\theta_k(s,t) = \lambda_k \exp\left(-\frac{(s-\mu_k^1)^2}{2(\sigma_k^1)^2} - \frac{(t-\mu_k^2)^2}{2(\sigma_k^2)^2}\right)$$

where $(\mu_k^1, \mu_k^2) \in \mathbb{R}^2$ describes the center of the spot, $(\sigma_k^1, \sigma_k^2) \in (0,\infty)^2$ describes the spread, and $\lambda_k \in [0,\infty)$ is height of the peak.

The idea in the model is to use classical blob detection methods in image analysis to determine all possible spots, and their respective five parameters, and then model the residual image $y_{ij} - \sum_{k=1}^{N} \theta_k(s_i, t_j)$ with the Gaussian model used in the previous chapters. This will inevitably lead to overfitting, where noise is modelled as spots, but one strength of this model is that it allows for testing of the hypothesis $H_k : \lambda_k = 0$, for which one can do a likelihood ratio test. Testing the hypotheses of the form H_k will inherently lead to multiple testing problems, and some sort of adjustment of the p-values will be needed.

Looking at the images in Figure 7.1, one realizes that there are some problems in model (7.1). First of all the spots might be better described by something else than Gaussian functions, but more importantly there are some problems with using the Gaussian process x to model the residual. The correlated

Figure 7.2: The images from Figure 7.1 plotted as surfaces.

noise seems to be particularly strong around the protein spots and on vertical or horizontal lines with big/many spots, so the correlated noise is somewhat dependent on the protein spots.

We will leave the modelling of protein spots to people who are interested in that kind of thing, and just consider the left image in Figures 7.1 in its purity, i.e. model (7.1) with all $\lambda_k = 0$. If we use the Green's function for the biharmonic operator $\Delta\Delta$ as the covariance $\mathcal{G}$, we can estimate σ and τ using the approximated likelihood with Dirichlet boundary conditions. This is of course somewhat silly as the images are clearly not Gaussian. Also there is a small problem with using Dirichlet boundary conditions, as these are only partly satisfied, but this will be ignored for now. Unless otherwise mentioned, the cutoff of the double series in the approximations has been set to 40. A plot of the negative log likelihood function can be found In Figure 7.3.

Approximated likelihood function
$\tau = 0, \ldots, 20,$
$\sigma = 0.016, \ldots, 0.300$

Approximated likelihood function
$\tau = 10, \ldots, 14,$
$\sigma = 0.028, \ldots, 0.122$

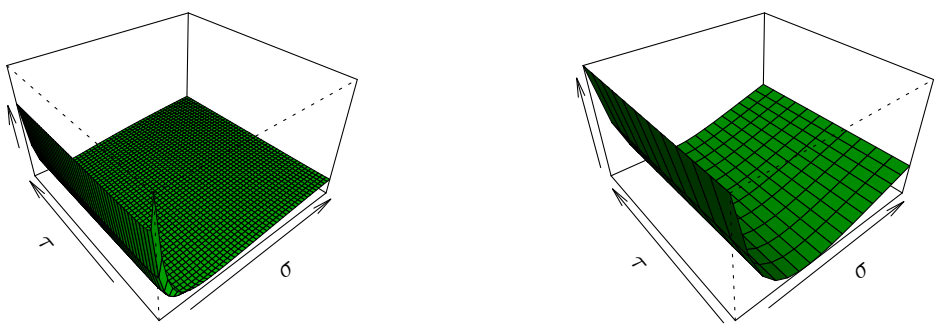

Figure 7.3: The likelihood function for the left image in Figure 7.1.

The maximum likelihood estimate of σ is 0.0465678 and for τ it is 11.9800299. The corresponding smooth can be found in Figure 7.4. As one can see the smooth describes the original image well. The two most noticeable differences between the real image and the smooth are that the spikes corresponding to the spots does not have the original flat tops (due to censoring in the dynamic range of the original image) in the smooth, and that the noise between the spots has been filtered away. The residual image can be found in Figure 7.5.

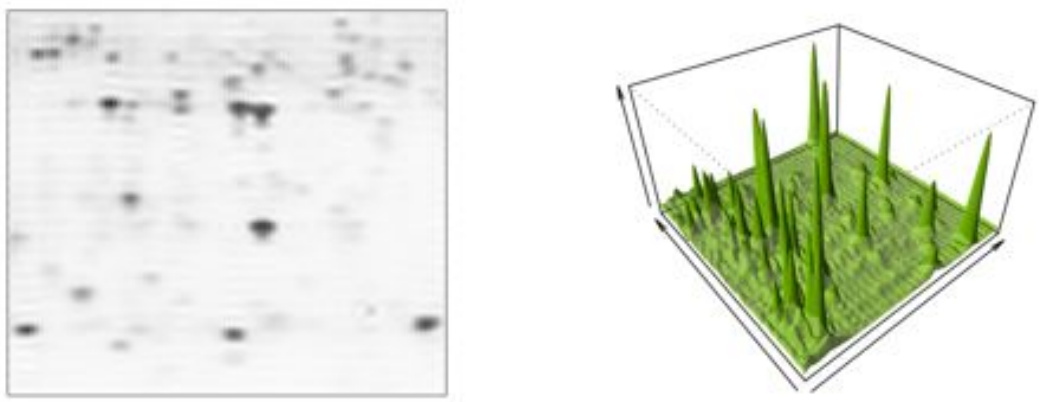

Figure 7.4: The smooths of the left image from Figure 7.1.

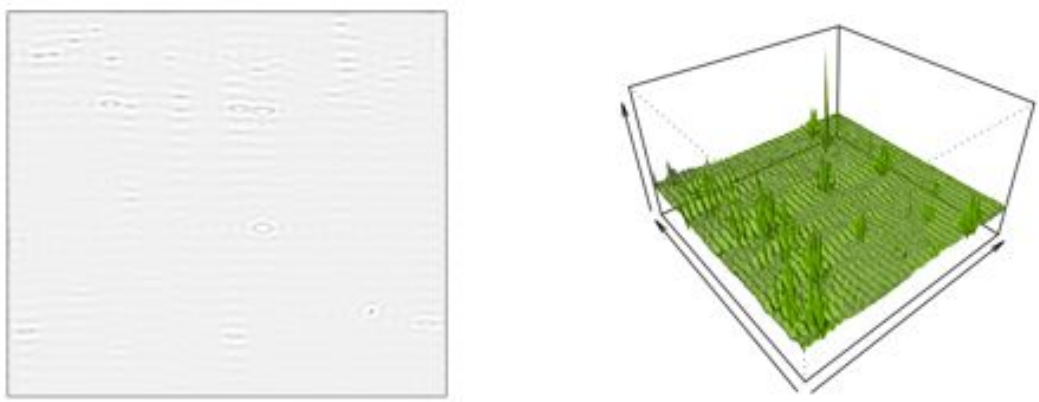

Figure 7.5: The residual $y - \widehat{\mathrm{E}}\,[\,y\,|\,x\,]$, i.e. the left images in Figures 7.1 and 7.2 with the corresponding images from Figure 7.4 subtracted.

One very interesting feature of the smooth is the structure of the surface between the spikes, which is predominantly ridges parallel to the axes – a feature well known from the Brownian sheet. This structure is also noticeable in the original image, but as one can see in the residual image this structure might

be a bit exaggerated in the smooth – especially anlong one of the axes. In fact these axis parallelities are a feature of the electrophoresis image, as the separation mechanisms in the electrophoresis process are done one at the time, and the protein spots will then leave a traces parallel to the axes. This also explains why the structure parallel to the axes seem to be more dominant along one of the axes. It is clear (and not surprising) that there are some problems with the residuals. A normal QQ plot of the residuals can be found in Figure 7.6, which clearly shows that the residuals are not normal distributed. The residual plot tell us a number of interesting things. First of all the green points corresponding to observations within 5 pixels from the border lie reasonably well, although there are some problems with both positive and negative residuals due to small protein spots partly lying on the border. With fits around 0 and 0.2 we also observe some quite extreme residuals. These are due to one of the bigger spots which is quite close to the border, and should disappear with a proper modelling of the corresponding spot. Finally there is a very noticable line structure for fitted values between 0.6 and 1.0. What this line tells us is that the smooth produces spikes that are more narrow and sometimes slightly higher than the observed. This is probably caused by the previously mentioned censoring in the dynamic range of the image, and this is certainly a problem that should disappear with proper modelling of the protein spots. If these three problems were dealt with, one could imagine that the residuals became somewhat reasonable.

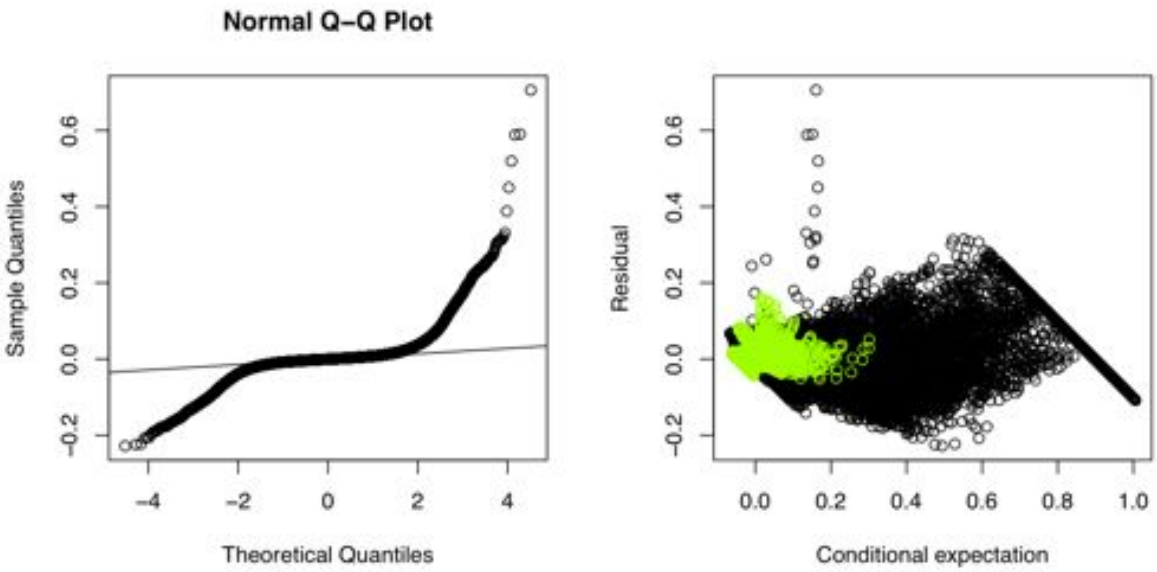

Figure 7.6: Normal QQ plot of the residuals from Figure 7.5 (left) and residuals plotted against the fitted values. Green points corresponds to observations within 5 pixels from the border.

Let us finally look at the autocovariance function. In Figure 7.7 a plot of the autocovariance function under the model and its empirical counterpart can be found (computed using (6.6) with the maximum likelihood estimates as parameters and (6.8) respectively.) The two plots are very different. First of all they are on different scales, where the theoretical autocovariance takes values in the interval $(-5 \cdot 10^{-6}, 10^{-6})$ for $n, m > 0$, the empirical takes values in the interval $(-2.5 \cdot 10^{-4}, 3 \cdot 10^{-4})$ for $1 \leq n, m \leq 200$. The value for $n = m = 0$ is 0.001957549 for the theoretical image, and in the empirical image it is 0.0004246596, again quite a difference. These differences are caused by the residual image clearly not being centered Gaussian with the correct covariance. One should notice that the ripple in the two images are not of the same nature. In the theoretical image the ripple is mainly due to the approximation error in the double sine series, that is heavily amplified. In the empirical autocovariance image the ripple along the n-axis of the empirical autocovariance image is caused by the corresponding

ripple in the residual image. It is hard to say how much better the residuals would be with a proper modelling, but it is sure that the modelling of tops would help on some of the observed problems.

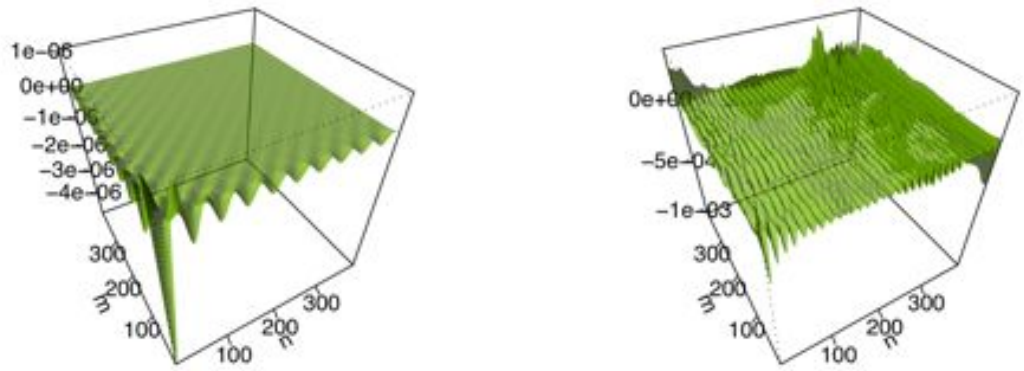

Figure 7.7: Autocovariance image under the model (left) and observed autocovariance image, both for $n, m = 1, \ldots, 399$. The waves in the left autocovariance image are mainly due to an error in the approximation process (the series in (6.6) were cut of at their 20th terms) that has been amplified. As previously mentioned it is expected that the empirical autocovariance image has wild behaviour for large values of n and m.

A transformation of the data might produce an even better correspondence between the data and the proposed model. We will try a log-transformation. This model requires that we add a constant to the new data, so it corresponds better with the Dirichlet boundary conditions. I.e. if z denotes the image data, we will work on $y = \log(z) + c$ with $c \in \mathbb{R}$, where c is chosen such that the average deviation from zero along the border is minimized. The log-transformed images can be seen in Figure 7.8 and the corresponding (negative log) likelihood function can be found in Figure 7.9.

The maximum likelihood estimates of σ and τ are 0.1381845 and 40.3267746, and the corresponding smooth can be found in Figure 7.10. The residuals which can be found in Figure 7.11 seem quite a bit nicer than the ones in Figure 7.5. The normal QQ plot of the residuals and the plot of residuals against fitted values in Figure 7.12 also seem a bit prettier than for the non-log-transformed images, but they still suffer from the same problems. The same is the case for the autocorrelation functions found in Figure 7.13. One can still hope that proper modelling of the tops will solve the worst of these problems.

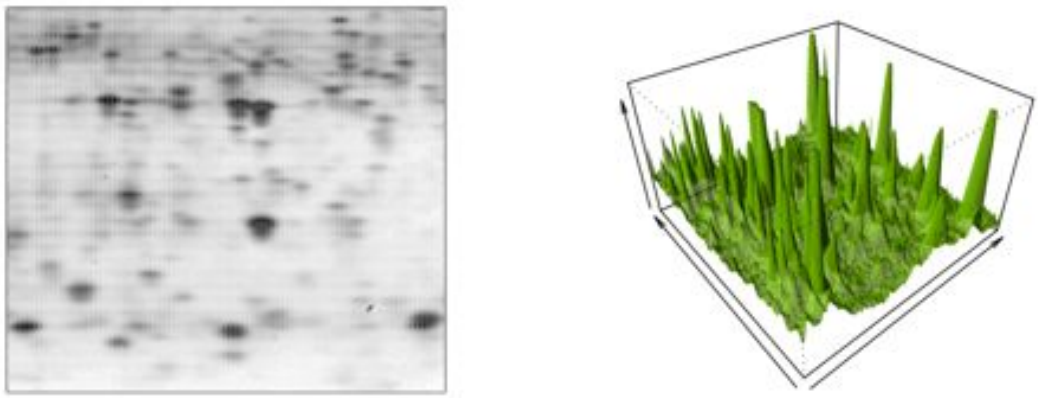

Figure 7.8: The log-transformed version of the left electrophoresis image from Figures 7.1 and 7.2.

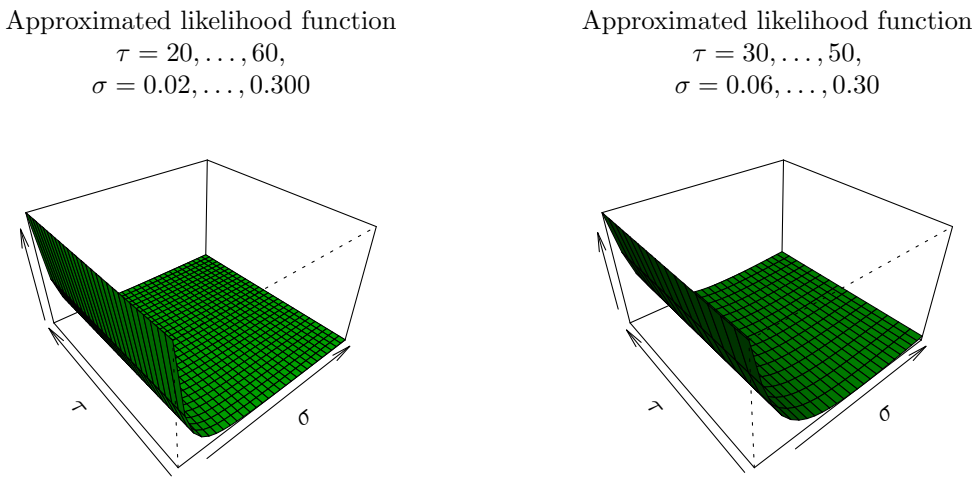

Figure 7.9: The likelihood function for the image in Figure 7.8.

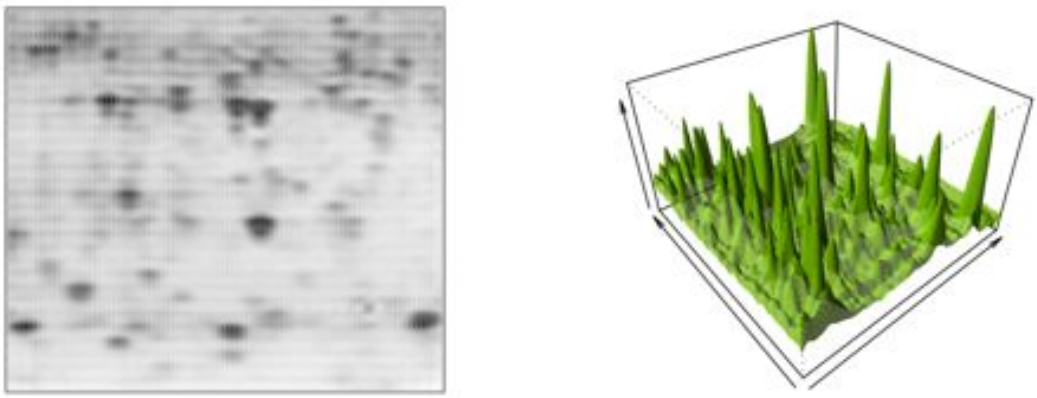

Figure 7.10: The smooths of the image from Figure 7.8.

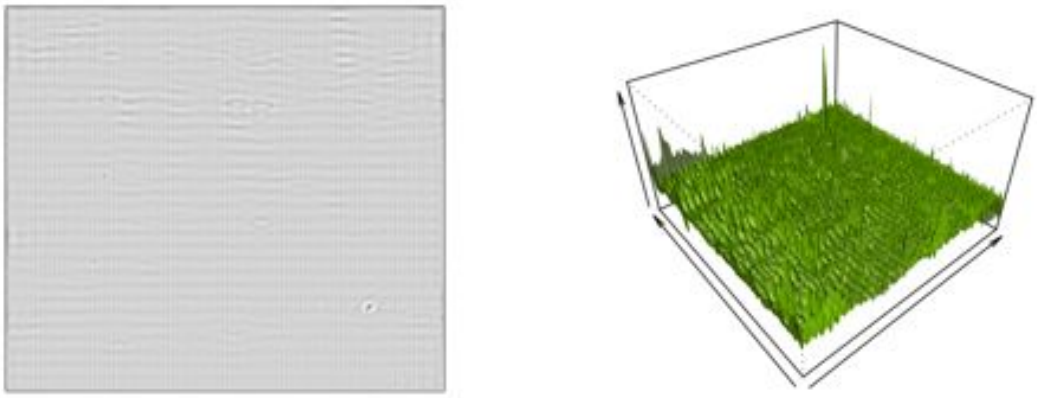

Figure 7.11: The residual $y - \widehat{\mathrm{E}}[\,y\,|\,x\,]$, where y is the image in Figure 7.8 and $\widehat{\mathrm{E}}[\,x\,|\,y\,]$ is the image from Figure 7.10.

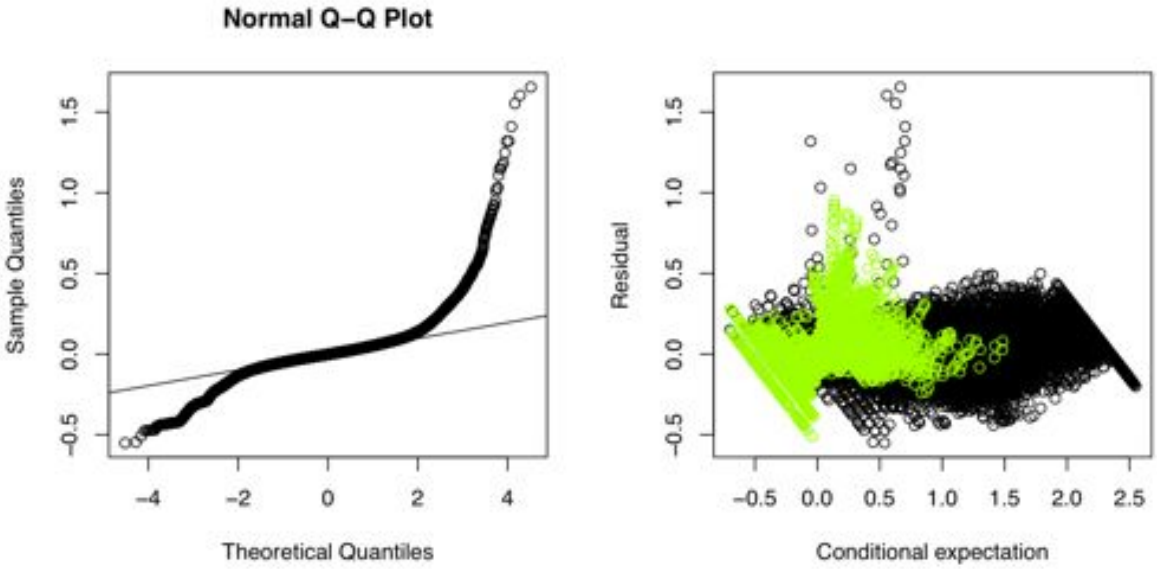

Figure 7.12: Normal QQ plot of the residuals from Figure 7.9 (left) and residuals plotted against the fitted values. Green points corresponds to observations within 5 pixels from the border.

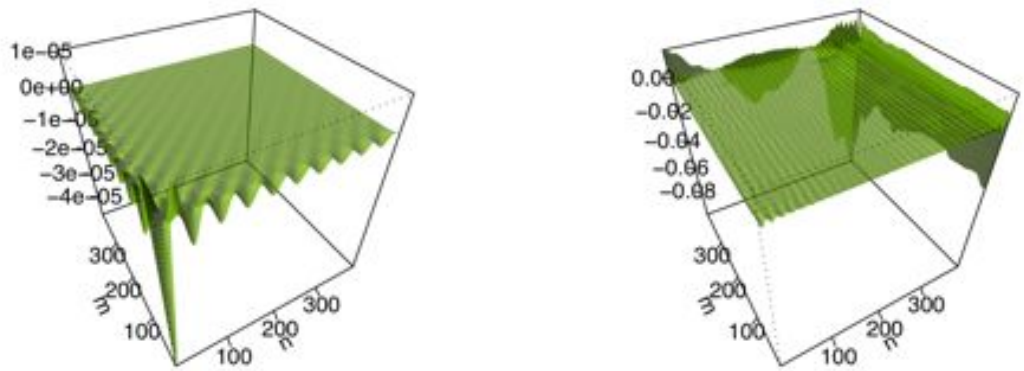

Figure 7.13: Autocovariance image under the (log-transformed) model (left) and observed for $n, m = 1, \ldots, 399$. The waves in the left autocovariance image are mainly due to an error in the approximation process (the series in (6.6) were cut of at their 20th terms) that has been amplified.

CHAPTER 8

FUTURE DIRECTIONS

The presented methods should be an interesting addition to the field of two-dimensional functional data analysis. The results throughout this thesis are essentially dimensionless, and are easily generalized to higher dimensions. This perhaps makes the results even more interesting, since the field of multidimensional functional analysis is somewhat underdeveloped and needs computationally efficient and interpretable solutions (see e.g. Ramsay & Silverman (2005) Section 22.2.3.)

Another interesting addition would be to generalize the theory to handle multivariate images, as this would make it possible to do functional data analysis on e.g. color images.

Below a number of specific things and choices in this thesis will be reviewed.

The idea to specify the image model in a Bayesian framework may be problematic. Especially if the prior is improper in some sense. It is possible to specify alternative models and produce similar smooths using these. Markussen (2013) specifies the model in a linear mixed effects model framework and Wahba (1978) specifies an equivalent model as a stochastic filtering problem. An interesting question is what the effect would be if one used alternative prior distributions in the estimation. It would especially be interesting if one could use priors that had different continuity properties than the Gaussians. This would result in different estimates, but the approximation methods, where calculations is moved to the functional domain, should still be applicable.

The use of maximum likelihood estimation for the parameters can be criticized. Instead of using maximum likelihood estimation on the prior distribution, one could perhaps develop an approximated generalized cross validation method to estimate the parameters. This would allow for use of the approximation in a setup that would please people with a stronger frequentist mindset. Another possibility is to use e.g. leave-one-out cross validation. This would require that one looked at approximations for non-equidistant observations, but only leaving one out would still result in closed form approximations. In fact it should still be possible to do the approximations if the observations are not equidistantly spaced, but the prettiness of the resulting formulae will suffer.

Another important choice made in this thesis is the choice of using iterated Laplacians for the roughness measure. Alternative roughness measures such as $(\partial_s \partial_t)^\ell$ could result in Green's functions with much better computational properties, and perhaps also more pleasing continuity properties.

The Green's functions considered have been computed for function spaces subject to certain boundary conditions. The two different types we have considered – Dirichlet and Neumann – are of course not the only choices, nor are they always optimal. Dirichlet boundary conditions are easily interpretable and result in "nice" covariance functions. For Neumann boundary conditions, the interpretation is that the "inner flow" of the surface should continue at the boundary, but the boundary conditions give problems in the corners where a discontinuity of the boundary conditions occurs. A solution to this could be to find a Green's function for "radial" Neumann conditions, i.e. where the conditions are not defined in terms of the normal derivatives on the boundary, but rather the derivative in the the direction from the center of the image. Such a Green's function may be very hard to find, and an alternative is to translate the boundary conditions to the outer circle of the square. Such a Green's function should be directly computable with some effort. An alternative type of boundary conditions that we have not considered is mixed boundary conditions, i.e. Dirichlet boundary conditions on some borders and Neumann conditions on others. As long as two Neumann conditioned borders does not meet in a corner this will fix the previously mentioned problem.

These and more things will be considered in future work by the author in collaboration with Bo Markussen.

BIBLIOGRAPHY

Adler, R. J. & Taylor, J. E. (2007), *Random fields and geometry*, Springer.

Craven, P. & Wahba, G. (1978), 'Smoothing noisy data with spline functions', *Numerische Mathematik* **31**(4), 377–403.

Dyn, N., Wahba, G. & Wong, W. (1979), 'Smooth pycnophylactic interpolation for geographical regions: Comment', *Journal of the American Statistical Association* **74**(367), 530–545.

Kimeldorf, G. S. & Wahba, G. (1970), 'A correspondence between Bayesian estimation on stochastic processes and smoothing by splines', *The Annals of Mathematical Statistics* **41**(2), 495–502.

Kimeldorf, G. & Wahba, G. (1971), 'Some results on Tchebycheffian spline functions', *Journal of Mathematical Analysis and Applications* **33**(1), 82–95.

Lametsch, R., Kristensen, L., Larsen, M. R., Therkildsen, M., Oksbjerg, N. & Ertbjerg, P. (2006), 'Changes in the muscle proteome after compensatory growth in pigs', *Journal of animal science* **84**(4), 918–924.

Markussen, B. (2013), 'Functional data analysis in an operator-based mixed-model framework', *Bernoulli* **19**(1), 1–17.

Meinguet, J. (1979), 'Multivariate interpolation at arbitrary points made simple', *Zeitschrift für angewandte Mathematik und Physik* **30**(2), 292–304.

Melnikov, Y. A. (1999), *Influence functions and matrices*, CRC Press.

Nicholson, J. & Bergman, L. (1985), 'On the efficacy of the modal series representation for the green functions of vibrating continuous structures', *Journal of Sound and Vibration* **98**, 299–304.

O'Sullivan, F. (1991), 'Discretized Laplacian smoothing by Fourier methods', *Journal of the American Statistical Association* **86**(415), 634–642.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Rakêt, L. L. & Markussen, B. (2014), 'Approximate inference for spatial functional data on massively parallel processors', *Computational Statistics & Data Analysis* **72**, 227 – 240.

Ramsay, J. O. & Silverman, B. W. (2005), *Functional data analysis*, second edn, Springer.

Wahba, G. (1978), 'Improper priors, spline smoothing and the problem of guarding against model errors in regression', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 364–372.

Wahba, G. (1990), *Spline models for observational data*, Society for industrial and applied mathematics.

Wahba, G. & Wendelberger, J. (1980), 'Some new mathematical methods for variational objective analysis using splines and cross validation', *Monthly Weather Review* **108**, 1122.