AALTO UNIVERSITY
SCHOOL OF SCIENCE

Arno Solin

HILBERT SPACE METHODS IN INFINITE-DIMENSIONAL KALMAN FILTERING

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology in the Degree Programme in Engineering Physics and Mathematics.

Espoo 2012

Supervisor:
Prof. Jouko Lampinen

Instructor:
D.Sc. (Tech.) Simo Särkkä

Aalto University
School of Science

AALTO UNIVERSITY SCHOOL OF SCIENCE P.O. Box 1100, FI-00076 AALTO http://www.aalto.fi	ABSTRACT OF THE MASTER'S THESIS

Author:	Arno Solin		
Title:	Hilbert Space Methods in Infinite-Dimensional Kalman Filtering		
Degree programme:	Engineering Physics and Mathematics		
Major subject:	Systems and Operations Research	Minor subject:	Computational Science and Engineering
Chair (code):	S-114		
Supervisor:	Prof. Jouko Lampinen		
Instructor:	D.Sc. (Tech.) Simo Särkkä		

Many physical and biological processes include both spatial and temporal features. Spatio-temporal modeling under the machine learning paradigm of Gaussian process (GP) regression has demonstrated prominent results. However, the appealing Bayesian treatment by GP regression is often difficult in practical problems due to computational complexity.

In this thesis, methods for writing spatio-temporal Gaussian process regression as infinite-dimensional Kalman filtering and Rauch–Tung–Striebel smoothing problems are presented. These scale linearly with respect to the number of time steps as opposed to the cubic scaling of the direct GP solution. Spatio-temporal covariance functions are formulated as infinite-dimensional stochastic differential equations. Furthermore, it is presented how infinite-dimensional models can be combined with a finite number of observations to an approximative solution. For this, a truncated eigenfunction expansion of the Laplace operator is formed in various domains, of which the n-dimensional hypercube and hypersphere are explicitly written out.

The approach in this thesis is primarily application-driven, and therefore three real-world case studies are presented as proof of concept. The feasibility of infinite-dimensional Kalman filtering is demonstrated by forming a spatio-temporal resonator model which is applied to temperature data in two spatial dimensions, and a novel way of modeling the space–time structure of physiological noise in functional brain imaging data is considered in both two and three spatial dimensions.

Date:	April 24, 2012	Language:	English	Number of pages:	64

Keywords: Infinite-dimensional Kalman filter, Distributed parameter system, Gaussian process regression, Spatio-temporal model, Eigenfunction expansion

Aalto-yliopisto
Perustieteiden korkeakoulu

AALTO-YLIOPISTO PERUSTIETEIDEN KORKEAKOULU PL 1100, FI-00076 AALTO http://www.aalto.fi	DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Arno Solin		
Työn nimi:	Hilbert-avaruusmenetelmät ääretönulotteisessa Kalman-suodatuksessa		
Tutkinto-ohjelma:	Teknillinen fysiikka ja matematiikka		
Pääaine:	Systeemi- ja operaatiotutkimus	Sivuaine:	Laskennallinen tiede ja tekniikka
Opetusyksikön koodi:	S-114		
Valvoja:	Prof. Jouko Lampinen		
Ohjaaja:	TkT Simo Särkkä		

Monet fysikaaliset ja biologiset mallit ovat sidottuja sekä paikkaan että aikaan. Koneoppimislähtöinen spatiotemporaalinen mallinnus gaussisten prosessien (GP) avulla on osoittautunut hyväksi lähestymistavaksi. Laskennallisen raskauden vuoksi gaussisten prosessien tarjoaman bayesilaisen malliperheen käyttö ei kuitenkaan usein käytännössä onnistu.

Tässä työssä tarkastellaan menetelmiä, joissa spatiotemporaalinen GP-regressio kirjoitetaan ääretönulotteisen Kalman-suodatuksen ja Rauch–Tung–Striebel-silotuksen avulla. Näiden menetelmien laskenta-aika skaalautuu lineaarisesti aikapisteiden määrän suhteen, kun taas suorassa GP-ratkaisussa laskenta skaalautuu kuutiollisesti. Työssä käytetyssä lähestymistavassa spatiotemporaaliset kovarianssifunktiot esitetään ääretönulotteisina stokastisina differentiaaliyhtälöinä. Lisäksi tutkittiin, miten ääretönulotteiset mallit voidaan yhdistää mittausarvoihin ja saada aikaan äärellisulotteinen approksimaatio. Tähän käytettiin katkaistua ominaisfunktiohajotelmaa, joka esitetään eksplisiittisesti Laplace-operaattorille n-kuutiossa ja n-pallossa.

Työn sovelluslähtöisyyden vuoksi esitellään kolme sovellusta, joissa ääretönulotteista Kalman-suodatusta voidaan käyttää. Tätä varten muodostetaan spatiotemporaalinen resonaattorimalli, jolla mallinnetaan lämpötilaa maapallon pinnalla kahdessa spatiaaliulottuvuudessa. Mallia sovelletaan myös fysiologisen kohinan mallintamiseen aivoissa kahdessa ja kolmessa spatiaaliulottuvuudessa.

Päivämäärä:	24.4.2012	Kieli:	Englanti	Sivumäärä:	64
Avainsanat:	Ääretönulotteinen Kalman-suodin, Jakautuneet järjestelmät, Gaussiset prosessit, Spatiotemporaalinen malli, Ominaisfunktiohajotelma				

Preface

This work was carried out in the Bayesian Statistical Methods group in the Department of Biomedical Engineering and Computational Science at Aalto University, Finland.

I wish to express my gratitude to my instructor Simo Särkkä and supervisor Jouko Lampinen for their support and guidance during the project. I also thank Jouni Hartikainen for valuable prior work, and Aki Vehtari for all over expertise. Furthermore, Aapo Nummenmaa, Toni Auranen, Simo Vanni and Fa-Hsuan Lin have all been part of the DRIFTER project and helped out with things beyond my knowledge.

Finally, I would like to thank my fiancée Kaisa and my family for their support and patience during the occasionally quite intensive writing process.

Otaniemi, 2012

Arno Solin

Contents

Symbols and Abbreviations

Matrices are capitalized and vectors are in bold type. Operators are slanted. We do not generally distinguish between probabilities and probability densities.

Operators and miscellaneous notation

$1:k$	$1, 2, \ldots, k$
$p(\mathbf{f} \mid \mathbf{y})$	Conditional probability density of $\mathbf{f}$ given $\mathbf{y}$
$\mathbf{f}_{k\mid k-1}$	Conditional value of $\mathbf{f}_k$ given observations up to step $k-1$
$\mathbb{R}$, $\mathbb{C}$ and $\mathbb{N}$	The real, complex and natural numbers
$\mathcal{N}(\mathbf{m}, \mathbf{C})$	Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{C}$
$\mathbf{I}$	Identity matrix
$\mathbf{A}^\mathsf{T}$	Matrix transpose of $\mathbf{A}$
$\mathscr{F}_t[f(t)](\omega)$	Fourier transform of $f(t)$
$\langle \cdot, \cdot \rangle$	Inner product
$\mathscr{H}$	Hilbert space
$\mathcal{L}$	Linear operator
$\psi_n(\mathbf{x})$	Eigenfunction
λ_n	Eigenvalue
Ω and $\partial\Omega$	Domain and domain boundary

General notation

$\mathbf{f} \in \mathbb{R}^s$	System state
$\mathbf{x} \in \mathbb{R}^n$	Spatial coordinate
$\mathbf{y} \in \mathbb{R}^d$	Observation
k	Time step index
T	Final time step

Abbreviations

GP	Gaussian process
SDE	Stochastic differential equation
SPDE	Stochastic partial differential equation
RTS	Rauch–Tung–Striebel (smoother)
fMRI	Functional magnetic resonance imaging
TR	Repetition time, interval between subsequent scans (T_R)
RMSE	Root-mean-square error

1 Introduction

Models derived for physical processes often involve variability in both space and time. Such spatio-temporal models can, for example, be applied to neural activity in the brain or the weather — just to name the applications that are addressed further on. For analysis, the Bayesian treatment provided by *Gaussian process* (GP) regression (O'Hagan, 1978) is often eligible. GP regression is a supervised machine learning (Alpaydin, 2010) paradigm where learning amounts to computing the posterior process from a given set of measurements. However, large datasets and the modeling of space, time and spatio-temporal interactions raise difficulties that are often impossible to deal with. One issue with GP methods is that their computational complexity is $\mathcal{O}(n^3)$, due to the inversion of an $n \times n$ matrix. This renders the basic approach prohibitive as the amount of data grows.

The celebrated Kalman filter (Kalman, 1960; Jazwinski, 1970; Grewal and Andrews, 2001) can be used for computing the Bayesian solutions to a general class of temporal Gaussian processes observed through a Gaussian linear model. While GP models are given in terms of a mean and covariance function, Kalman filter models are constructed as solutions to linear stochastic differential equations. The Kalman filter actually only provides the forward-time posterior, and the full posterior is given by some smoothing method such as the Rauch–Tung–Striebel smoother (Rauch *et al.*, 1965; Grewal and Andrews, 2001). The solutions can be written in closed-form, and the computational complexity scales linearly with respect to the number of measurements in the temporal dimension.

The appealing properties of the Kalman filter can be exploited in postulating an *infinite-dimensional* Kalman filter (see, e.g., Curtain, 1975; Cressie and Wikle, 2002; Wikle and Cressie, 1999) or a *distributed parameter* Kalman filter (as referred to in Tzafestas, 1978; Omatu and Seinfeld, 1989), where the state is actually an element in an infinite-dimensional Hilbert space. With 'Hilbert space methods' in the title, we refer to tools from functional analysis that can be used to form finite-dimensional approximations and combine them with observations. This can give actual solutions to real-world estimation problems. In this context Hilbert spaces methods are used as synonyms for basis function approximations.

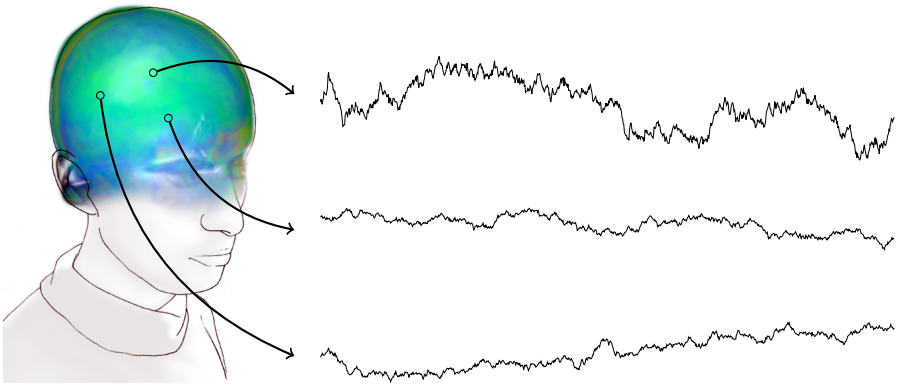


Figure 1: An illustration of spatio-temporal data, where functional brain data is visualized as a three-dimensional density. The three time series correspond to three different spatial locations.

As mentioned, application areas can be found in different *spatio-temporal* models. The term 'spatio-temporal' or 'space–time' model refers any mathematical model that combines space (as in place) and time. Instead of directly considering four-dimensional GP regression, we review evolution-type models. These are interpreted with space being three-dimensional — also one- and two-dimensional spatial domains are considered here — and time playing the role of a fourth dimension that is different from the spatial dimensions.

The main interest in this work is put on the following time evolution prototype model which is presented here to ease the reading further on. Also the demonstrations fall under this formulation which we presented in the form

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = \mathcal{A}\, f(\mathbf{x},t) + L\, w(\mathbf{x},t)$$
$$\mathbf{y}_k = \mathcal{H}_k\, f(\mathbf{x},t) + \mathbf{r}_k$$

where $f(\mathbf{x},t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ is the state, $\mathcal{A}$ some spatial (linear) operator and $w(\mathbf{x},t)$ a space–time white noise process. The observation model is set up by a functional $\mathcal{H}_k$. Values of $\mathbf{y}_k$ are noisy measurements of the phenomenon at discrete time steps t_k, where the noise term $\mathbf{r}_k$ is assumed Gaussian. Details and generalizations of the model can be found in Section 2.4.

This work is meant to be primarily application-driven, which means that we put interest in practically feasible solutions. Therefore we also present some illustrative real-world examples where the methods have been applied to actual data. In the two different data sets, the state function $f(\mathbf{x}, t)$ corresponds to the value of the phenomenon (i.e. outdoor temperature or brain activity) and $\mathbf{y}_k$ to some observations of it (i.e. thermometer readings or magnetic resonance imaging data of the brain). Figure 1 shows an illustrative sketch of a spatial three-dimensional density field. Three temporal time series are extracted from different spatial locations. If we consider all possible time series in all spatial locations, we have the spatio-temporal data.

1.1 Literature Review

We present a brief literature review covering the themes discussed in this thesis. The context, in which each of the sources is used, is easier seen by following the references throughout the text.

The mathematical definitions and properties concerning Hilbert spaces in the scope of functional analysis are covered in, e.g., Kreyszig (1978), and a more detailed presentation of spectral theory and differential operators is given in Davies (1995). Hilbert space methods in partial differential equations are presented by Showalter (1977).

Stochastic differential equations (SDE) are discussed extensively in Øksendal (2003) and stochastic partial differential equations (SPDE) further in Holden *et al.* (1996) and Chow (2007). An extensions to this is the pseudodifferential operator presentation in Shubin (1987). The infinite-dimensional dynamic perspective offered in Robinson (2001) and Da Prato and Zabczyk (1992) bring up the evolution type models.

Tools for combining SPDE equations with indirect observations are provided by Bayesian inference, a concept which is readily presented by Gelman *et al.* (2004). The Bayesian outlook sets the backdrop of the study, which leads to the use of Gaussian process regression. Gaussian process models (O'Hagan, 1978) are introduced from a machine learning perspective by Rasmussen and Williams (2006). They are formally equivalent to random field based *kriging* in geostatistics (Cressie, 1993; Christakos, 2005). The link

was shown by Hartikainen and Särkkä (2010) and Särkkä and Hartikainen (2012), the work on which this thesis is primarily based.

Kalman filtering (Kalman, 1960; Jazwinski, 1970; Grewal and Andrews, 2001) and Rauch–Tung–Striebel smoothing (Rauch *et al.*, 1965) for linear state space systems act as a backdrop to the infinite–dimensional filtering theory. The concept of infinite-dimensional Kalman filtering is not a new one (for an early survey, see, Curtain, 1975), as it can be seen as a very intuitive extension to the standard linear Kalman filter. Infinite-dimensional filtering has been covered by Falb (1967), and later by Wikle and Cressie (1999) and Cressie and Wikle (2002). Recently the non-stationary inverse problem viewpoint (Kaipio and Somersalo, 2004) in this setting has been addressed by Pikkarainen (2006).

Fields where applications of infinite-dimensional Kalman filtering have been considered include earth sciences and geostatistics (see Christakos, 2005), and also biomedical applications, such as in electrical impedance tomography (e.g. Pikkarainen, 2005), exist.

1.2 Objectives and Scope

In the first section of this thesis we will go through the concepts of stochastic equations in infinite dimensions and Hilbert space valued stochastic processes. The link between Gaussian process (GP) regression and the state space form of stochastic partial differential equations (SPDE) is gone through in some detail to provide an insight to the spatio-temporal state space formulation.

The next step is to introduce optimal estimation — i.e. Kalman filtering and Rauch–Tung–Striebel smoothing — in finite dimensions. We show how a continuous-time stochastic differential equation can be discretized to fall under the Kalman filtering formulation. We present the infinite-dimensional Kalman filter as a generalization of the standard discrete-time linear filtering solution.

For practical implementations, the infinite-dimensional Kalman filter has to be used in combination with some approximative methods in order to do the computations feasible. This aspect is dealt with in the third section of this thesis, where we present eigenfunction expansions of the Laplace operator in various domains — n-dimensional hypercubes and spheres —

subject to Dirichlet boundary conditions. Here we restrict our interest to problems which can be solved in these domains and formulated in terms of the Laplace operator.

We also go through the practical numerical use of infinite-dimensional Kalman filtering. To do this, we extend the resonator model that was introduced in Särkkä *et al.* (2012a) to a space–time form. We demonstrate the use of this spatio-temporal resonator model by presenting three case studies: a hourly temperature model on the surface of a sphere, and two examples for fMRI data analysis in both two-dimensional polar and three-dimensional spherical coordinates. In the latter examples, we construct a space–time filtering model for modeling physiological noise in fMRI data, which acts as a spatial extension to the recently published DRIFTER method (Särkkä *et al.*, 2012a). The results and approaches are discussed in some detail and future extensions are suggested.

The main contributions of this study are to (i) verify previous work, (ii) unify different notation and conceptual approaches from distinctive fields of space–time modeling, and (iii) provide a backdrop in spatio-temporal models that can be subject to further extensions.

2 Infinite-Dimensional Methods in Kalman Filtering

2.1 Hilbert Spaces

Hilbert spaces (see, e.g., Kreyszig, 1978, for a good introduction) are a generalization of the concept of two- or three-dimensional Euclidean spaces. Hilbert spaces extend the vector algebra and calculus of finite-dimensional spaces to any finite or infinite number of dimensions. A Hilbert space is an abstract vector space characterized by an inner product and a norm which define concepts as 'length' and 'angle' in the space.

To be more precise, we assign a *vector space* an inner product. This forms an *inner product space*, where the inner product is here defined by

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, \mathrm{d}x,$$

where $f, g : [a, b] \to \mathbb{R}$. The inner product induces a norm to the space, so an inner product space is can be made a normed vector space. The norm defines the metrics of the space, for which we can write $\|f\|^2 = \langle f, f \rangle$. All the quadratically integrable or *square-integrable* functions, for which

$$\int_a^b \|f(x)\|^2 \, \mathrm{d}x < \infty,$$

form a *complete metric space* (a space with all points defined), which is hence a *Banach space*. A complete space with an inner product and a norm $\|f(x)\|^2 = \langle f, f \rangle$ is called a *Hilbert space*. This Hilbert space (the functional space of square-integrable functions) is conventionally denoted $L^2[a, b]$ in the bounded interval a to b.

We present some concepts related to linear operator theory that will be referred to throughout this work. Hereafter we denote operators by slanted calligraphic symbols, e.g. $\mathcal{L}$. In general, a *compact operator* is a linear operator $\mathcal{L}$ from a Banach space to another Banach space, such that the image under $\mathcal{L}$ of any bounded subset of X is a relatively compact subset of Y. Such an operator is necessarily a bounded operator, and so continuous.

Equivalently, a definition of a compact operator $\mathcal{T}$ on a Hilbert space $\mathscr{H}$ can be given such that $\mathcal{T} : \mathscr{H} \to \mathscr{H}$ is said to be compact if (and only

if) it can be written in the form

$$\mathcal{T} = \sum_{n=1}^{N} \lambda_n \langle f_n, \cdot \rangle \, g_n,$$

where $f_1, f_2, \ldots, f_N \in \mathscr{H}$ and $g_1, g_2, \ldots, g_N \in \mathscr{H}$ for $1 \leq N < \infty$ are (not necessarily complete) orthonormal sets. The coefficients $\lambda_1, \lambda_2, \ldots, \lambda_N$ are a sequence of positive numbers which are the singular values of the operator. As we will put our main interest in the unbounded *Laplace operator* $\Delta = \nabla^2$, this setup is strictly speaking unsatisfactory. However, the methods can be rigorously generalized to unbounded operators and the results would be analogous.

A linear operator $\mathcal{L}$ on a Hilbert space $\mathscr{H}$ is called *symmetric* if $\langle \mathcal{L}x, y \rangle = \langle x, \mathcal{L}y \rangle$ for all elements x and y in the domain of $\mathcal{L}$. A symmetric operator that is defined everywhere is also *self-adjoint*, which means that the operator is equal to its own adjoint $\mathcal{L}^*$. In the scope of this study, it is noteworthy that if (and only if) the Hilbert space is finite-dimensional and a self-adjoint operator $\mathcal{L}$ has been written in terms of an orthonormal basis, the matrix $\mathbf{L}$ describing $\mathcal{L}$ is Hermitian (it equals its own conjugate transpose $\mathbf{L} = \mathbf{L}^*$ or transpose $\mathbf{L} = \mathbf{L}^\mathsf{T}$ in the real case).

The main interest in constructing real-world solutions to infinite-dimensional models in this study is put on expanding the infinite-dimensional operator equations to truncated series approximations. The theory behind this is based on the *Hilbert–Schmidt theorem* which is also known more casually as the *eigenfunction expansion theorem*.

To present the eigenfunction expansion of a bounded compact self-adjoint operator $\mathcal{L} : \mathscr{H} \to \mathscr{H}$ let $\lambda_n, n = 1, 2, \ldots, N$, be a sequence of non-zero real eigenvalues such that $|\lambda_n|$ is monotonically non-increasing. If $N = \infty$ then $\lim_{n\to\infty} \lambda_n = 0$. We furthermore assume that each eigenvalue is repeated in the sequence according to its multiplicity. Now we can say that there exists a set $\psi_n, n = 1, 2, \ldots, N$, of corresponding eigenfunctions such that

$$\mathcal{L}\psi_n = \lambda_n \psi_n, \quad \text{for } n = 1, 2, \ldots, N.$$

This enables us to consider the case where $\mathcal{L}$ operates on some function $u(\cdot)$

by writing

$$\mathcal{L}u = \sum_{n=1}^{N} \lambda_n \langle \psi_n, u \rangle \psi_n, \quad \text{for all } u \in \mathscr{H},$$

where ψ_n forms an orthonormal basis for the range of $\mathcal{L}$. This relates to the so called *spectral decomposition* of an operator which we write in the form of the following mapping

$$f(x) \mapsto \int_a^b k(x,y) f(y) \ \mathrm{d}y,$$

where $k(\cdot,\cdot)$ is a continuous function symmetric in x and y. The resulting eigenfunction expansion expresses the *kernel function* $k(\cdot,\cdot)$ of $\mathcal{L}$ as a series of the form

$$k(x,y) = \sum_n \lambda_n \psi_n(x) \psi_n(y),$$

where the functions ψ_n are orthonormal in the sense that $\langle \psi_n, \psi_m \rangle = 0$ for all $n \neq m$. More generally, kernels of unbounded operators comprise delta functions and their derivatives.

2.2 Stochastic Equations in Infinite Dimensions

The rest of this section is dedicated to stochastic processes in finite and infinite-dimensions. We go through the idea behind Gaussian process models, time evolution models and the connection between them.

2.2.1 Spatio-Temporal Gaussian Processes

A linear finite-dimensional regression problem can be written as a vector $\mathbf{f} \in \mathbb{R}^s$ being a draw from a normal prior $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$. The observed value $\mathbf{y} \in \mathbb{R}^d$ of $\mathbf{f}$ can be affected by some zero mean Gaussian measurement noise, $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, with covariance $\mathbf{R}$. This linear regression problem can be given in the form

$$\begin{aligned} \mathbf{f} &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0) \\ \mathbf{y} &= \mathbf{H}\,\mathbf{f} + \mathbf{r}, \end{aligned} \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{d \times s}$ is the linear observation model matrix.

We extend the linear regression model in Equation (1) to account for an infinite-dimensional process. Gaussian process (GP) regression (O'Hagan, 1978; Rasmussen and Williams, 2006) is a machine learning paradigm in which a process is seen as realizations of a Gaussian random process prior. A Gaussian process model is characterized by its prior mean $m_0(\mathbf{x})$ and prior covariance function $C_0(\mathbf{x}, \mathbf{x}')$. Let us consider that $f(\mathbf{x}) \in \mathscr{H}(\mathbb{R}^n)$ is an element in an infinite-dimensional Hilbert space, and thus values $f(\mathbf{x})$ correspond to outputs of the process with different inputs:

$$\begin{aligned} f(\mathbf{x}) &\sim \mathcal{GP}\left(m_0(\mathbf{x}), C_0(\mathbf{x}, \mathbf{x}')\right) \\ \mathbf{y} &= \mathcal{H} f(\mathbf{x}) + \mathbf{r}, \end{aligned} \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^d$ is the observation and $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is the measurement noise component as earlier. The measurement model is defined by $\mathcal{H}$, a functional which defines the discrete observations. The GP model is actually equivalent to the kriging model of Cressie (1993) as presented for example by Särkkä and Hartikainen (2012).

The GP model in Equations (2) can be seen as a spatial regression model. If we include a separate dependent variable t to the model to account for the temporal structure, we get a spatio-temporal GP model, which can be written rather straight-forwardly as

$$\begin{aligned} f(\mathbf{x}, t) &\sim \mathcal{GP}\left(m_0(\mathbf{x}, t), C_0(\mathbf{x}, t; \mathbf{x}', t')\right) \\ \mathbf{y}_k &= \mathcal{H}_k f(\mathbf{x}, t_k) + \mathbf{r}_k, \end{aligned} \tag{3}$$

where the functions are dependent on t as well (such that $f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$), and we assume that the values are observed at discrete time points $t_k, k = 1, 2, \ldots, T$. The observation functional $\mathcal{H}_k$ and the dimension of the observation $\mathbf{y}_k \in \mathbb{R}^{d_k}$ as well as the time-white measurement noise $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, can all depend on the step index k.

2.2.2 Covariance Functions

In the previous section we saw that a Gaussian process is characterized by its mean $m(\mathbf{x}) = \mathrm{E}[f(\mathbf{x})]$ and its covariance function $C(\mathbf{x}, \mathbf{x}') = \mathrm{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\right]$. The covariance function encodes the similarity of data between separate locations in space (or time).

A covariance function is a function of input pairs $\mathbf{x}$ and $\mathbf{x}'$ where the function gives information about the relation of the two inputs. A *stationary* covariance function is a function of $\mathbf{x} - \mathbf{x}'$ (the difference between the point locations), and if the covariance is only function of $\|\mathbf{x} - \mathbf{x}'\|$ (the distance between points), it is called *isotropic* (further details can be found, e.g., in Rasmussen and Williams, 2006).

In general, a function of two arguments mapping the relation between $\mathbf{x} \in \Omega$ and $\mathbf{x}' \in \Omega$ is called a *kernel* $k(\mathbf{x}, \mathbf{x}')$. This relates directly to covariance functions, and is also familiar from earlier as the same notation arises in theory of integral operators. We can consider an operator $\mathcal{T}$ with a kernel $k(\cdot, \cdot)$ such that

$$\mathcal{T} f(\mathbf{x}) = \int_\Omega k(\mathbf{x}, \mathbf{x}') f(\mathbf{x}') \, \mathrm{d}\mathbf{x}'.$$

Rasmussen and Williams (2006) offer a more thorough introduction to the subject. A kernel is said to be *symmetric* if $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$.

The connection between covariance functions and covariance matrices is that if we are given a set of input points $\{\mathbf{x}_i \mid i = 1, 2, \ldots, n\}$ we can compute the so called *Gram matrix* $\mathbf{K} \in \mathbb{R}^{n \times n}$, which has elements $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. If $k(\cdot, \cdot)$ is a covariance function, then the Gram matrix $\mathbf{K}$ is the corresponding covariance matrix. Because the covariance function is symmetric, the corresponding covariance matrix is also symmetric.

A kernel is said to be positive semidefinite, if

$$\int_\Omega \int_\Omega k(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) f(\mathbf{x}') \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{x}' \geq 0$$

for all square-integrable functions, $f \in L^2(\Omega)$. Similarly a positive semidefinite matrix $\mathbf{K}$ satisfies $\mathbf{x}^\mathsf{T} \mathbf{K} \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Furthermore a symmetric matrix is positive semidefinite if and only if all its eigenvalues are non-negative. We can also state that any Gram matrix (valid covariance matrix) is positive semidefinite.

Bochner's theorem (see, e.g., Da Prato and Zabczyk, 1992) states that a complex-valued function $k : \mathbb{R}^n \to \mathbb{C}$ is the covariance function $k(\cdot, \cdot)$ of a weakly stationary, $\mathbf{r} = \mathbf{x} - \mathbf{x}'$, mean-square continuous complex-valued

random process on $\mathbb{R}^n$, if and only if it can be represented as

$$k(\mathbf{r}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\boldsymbol{\omega}\cdot\mathbf{r}} \,\mathrm{d}\mu(\boldsymbol{\omega}),$$

where μ is a positive finite measure. In a rigorous sense, this is problematic because the white noise measure is not finite, but the theory can be made exact through generalizations. However, if the measure $\mu(\boldsymbol{\omega})$ has a density, this is the *spectral density* $S(\boldsymbol{\omega})$ corresponding to the kernel (i.e. the covariance function). This gives rise to the Fourier duality of covariance and spectral density, which is known as the *Wiener–Khintchine theorem* (see, e.g., Rasmussen and Williams, 2006) and defines

$$k(\mathbf{r}) = \frac{1}{(2\pi)^n} \int S(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}\cdot\mathbf{r}} \,\mathrm{d}\boldsymbol{\omega} \quad \text{and} \quad S(\boldsymbol{\omega}) = \int k(\boldsymbol{\omega}) e^{-i\boldsymbol{\omega}\cdot\mathbf{r}} \,\mathrm{d}\mathbf{r}.$$

In practice the Fourier transforms are not that often needed to be calculated explicitly. If the kernel $k(\cdot,\cdot)$ is a covariance function, we denote it by $C(\cdot,\cdot)$ and the corresponding covariance matrix by $\mathbf{C}$.

In this study, we are concerned with isotropic stationary covariance functions. As an illustrative example we consider a covariance functions of the Matérn class (Matérn, 1960). This class of stationary isotropic (if the norm is the Euclidean distance) covariance functions is widely used in many applications as it is rather simple and the parameters have somewhat understandable interpretations. A Matérn covariance function can be written as

$$C(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\,\frac{r}{l}\right)^{\nu} K_\nu \left(\sqrt{2\nu}\,\frac{r}{l}\right), \tag{4}$$

where $r = \|\mathbf{x} - \mathbf{x}'\|$, $\Gamma(\cdot)$ is the Gamma function and $K_\nu(\cdot)$ is the modified Bessel function. The covariance function is characterized by three parameters: a smoothness parameter ν, distance scale parameter l and strength (magnitude) parameter σ, all of which are positive. The Matérn class is especially interesting as it features two commonly used covariance functions as special cases: if $\nu \to \infty$, we get the *squared exponential* covariance function $C(r) = \sigma^2 \exp\left(-r^2/(2l^2)\right)$, and if $\nu = \frac{1}{2}$, we get the *exponential* covariance function $C(r) = \sigma^2 \exp\left(-r/l\right)$.

Figure 2 shows three covariance functions of the Matérn class with different parameter values for ν, one-dimensional draws from Gaussian distri-

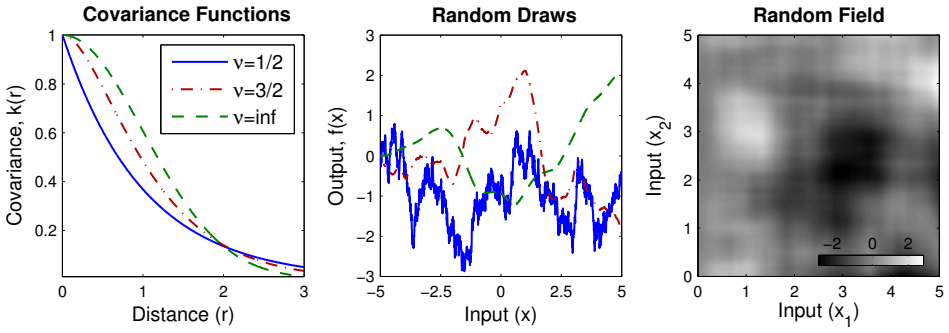


Figure 2: Matérn covariance functions and random functions draws from Gaussian processes with the corresponding covariance functions, for different values of ν. On far right a Gaussian random field with Matérn covariance ($\nu = 5/2$). In all the figures $l = 1$.

butions with the corresponding covariance matrices, and a two-dimensional draw that is a Gaussian random field. The one-dimensional draws are done by discretizing the x-axis into 2000 equally-spaced points. The draw in the rightmost figure is based on a 256×256 equally-spaced grid.

2.2.3 Converting Covariance Functions to Stochastic Equations

As was discussed earlier, the covariance function of the Gaussian process encodes the overall structure of the solution by including the dependencies between the values in the model. The space–time covariances characterize a random field in a subspace. Intuitively the same effects should be possible to be described by a suitable differential equation — or more precisely, a stochastic differential equation.

In this section we consider one-dimensional — the intuitive interpretation being the temporal dimension — stationary isotropic covariance functions $C(t, t')$ which can be written in terms of being only a function of the norm of the difference of the points, that is $C(t, t') = C(\tau)$, where $\tau = |t - t'|$. Following the procedure described by Hartikainen and Särkkä (2010), we consider a stationary scalar covariance function $C(\tau)$ for a process $f(t), t \in \mathbb{R}$. We try to find the corresponding stochastic differential equation (SDE) with approximately the same covariance function. By Fourier transform we can compute the spectral density $S(\omega)$ of the model, where $\omega \in \mathbb{R}$. We try to find a function $G(i\omega)$, so that $S(\omega) \approx G(i\omega)\, G(-i\omega)$.

Following the derivation in Hartikainen and Särkkä (2010) we consider a linear time-invariant (LTI) stochastic differential equation (SDE). The differential equation can be written as

$$\frac{\mathrm{d}^m f(t)}{\mathrm{d}t^m} + a_{m-1}\frac{\mathrm{d}^{m-1} f(t)}{\mathrm{d}t^{m-1}} + \cdots + a_1\frac{\mathrm{d}f(t)}{\mathrm{d}t} + a_0 f(t) = w(t),$$

where $a_0, \ldots, a_{m-1}$ are known constants and $w(t)$ is a white noise process with spectral density $S_w(\omega) = q$. The model is an mth order scalar LTI SDE, where the process is characterized by its derivatives up to order m and the stochastic variation comes from the white noise term. The linear SDE can be written as a matrix equation

$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\,\mathbf{f}(t) + \mathbf{L}\,w(t), \tag{5}$$

where the state $\mathbf{f}(t) = \left(f(t), \frac{\mathrm{d}}{\mathrm{d}t}f(t), \ldots, \frac{\mathrm{d}^{m-1}}{\mathrm{d}t^{m-1}}f(t)\right)$ contains the derivatives of $f(t)$ up to order $m-1$. The dynamic model matrix $\mathbf{F} \in \mathbb{R}^{m\times m}$ and the process noise propagation matrix $\mathbf{L} \in \mathbb{R}^{m\times 1}$ can be given as

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & \cdots & -a_{m-2} & -a_{m-1} \end{bmatrix} \quad \text{and} \quad \mathbf{L} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

This is called the 'companion form' (see Grewal and Andrews, 2001) for higher-order differential equations expressed in terms of first-order differential equations, and it is especially useful in the Kalman filtering context as it is in the form of a linear matrix equation. The SDE in Equation (5) can also be given using the Itô differential notation, where it would be $\mathrm{d}\mathbf{x}(t) = \mathbf{F}\,\mathbf{f}(t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}W(t)$, where $W(t)$ is a Wiener process or Brownian motion (see, e.g., Øksendal, 2003).

Even though we have written the stochastic differential equation in vector form, we still are primarily interested in the value of the process itself. The value of $f(t)$ can be extracted by the linear observation model $f(t) = \mathbf{H}\,\mathbf{f}(t)$, where $\mathbf{H} = [1\ 0\ \ldots\ 0]^{\mathsf{T}}$. Using this identity and formally

Fourier transforming both sides in Equation (5) yield

$$-i\omega \mathscr{F}_t[\mathbf{f}(t)](\omega) = \mathbf{F}\, \mathscr{F}_t[\mathbf{f}(t)](\omega) + \mathbf{L}\, \mathscr{F}_t[w(t)](\omega).$$

The next step is to substitute the Fourier transform by the spectral density of the noise process $|\mathscr{F}_t[w(t)](\omega)|^2 = S_w(\omega) = q$ (from earlier), and to denote the Fourier transform as the spectral density of the process. Rearranging the terms gives us

$$S(\omega) = \mathbf{H}\,(\mathbf{F} + i\omega\mathbf{I})^{-1}\,\mathbf{L}\,q\,\mathbf{L}^{\mathsf{T}}\left[(\mathbf{F} - i\omega\mathbf{I})^{-1}\right]^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}.$$

When the process has reached a stationary state (i.e. run an infinite period of time) the covariance function of $f(t)$ is given by the inverse Fourier transform of the spectral density:

$$C(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega\tau}\,\mathrm{d}\omega.$$

According to Hartikainen and Särkkä (2010), this can be calculated as

$$C(\tau) = \begin{cases} \mathbf{H}\,\mathbf{C}_\infty\,\mathbf{U}^{\mathsf{T}}(\tau)\,\mathbf{H}^{\mathsf{T}}, & \text{if } \tau \geq 0 \\ \mathbf{H}\,\mathbf{U}(\tau)\,\mathbf{C}_\infty\,\mathbf{H}^{\mathsf{T}}, & \text{if } \tau < 0 \end{cases}$$

where $\mathbf{U}(\tau) = \exp(\mathbf{F}\tau)$ and $\mathbf{C}_\infty$ is the stationary covariance of $\mathbf{f}(t)$. The matrix Riccati equation (see Grewal and Andrews, 2001)

$$\frac{\mathrm{d}\mathbf{C}}{\mathrm{d}t} = \mathbf{F}\,\mathbf{C} + \mathbf{C}\,\mathbf{F}^{\mathsf{T}} + \mathbf{L}\,q\,\mathbf{L}^{\mathsf{T}} = 0$$

can be used to solve the stationary covariance $\mathbf{C}_\infty$.

2.2.4 Space–Time Covariance Functions as Evolution Models

In this section we go through a rather general method for converting space–time covariances into stochastic differential equations. The following procedure is presented by Särkkä and Hartikainen (2012). We assume the covariance functions to be stationary, which enables us to write the covariance function $C(\mathbf{x}, \mathbf{x}'; t, t')$ as $C(\mathbf{x} - \mathbf{x}', t - t')$ and further $C(\mathbf{x}, t)$. Once again, we consider a stationary scalar covariance function $C(\mathbf{x} - \mathbf{x}', t - t')$ for a

spatio-temporal process $f(\mathbf{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$. Intuitively, there should be a stochastic (partial differential) equation which behaves similarly.

Fourier transforming the covariance function gives us the corresponding spectral density $S(\boldsymbol{\omega}_x, \omega_t)$, where $\boldsymbol{\omega}_x \in \mathbb{R}^n$ and $\omega_t \in \mathbb{R}$. Next we need to find a function $G(i\boldsymbol{\omega}_x, i\omega_t)$ which is stable in forward time and rational for variables $i\omega_t$ such that

$$G(i\boldsymbol{\omega}_x, i\omega_t) = \frac{b_0(i\boldsymbol{\omega}_x)}{(i\omega_t)^N + a_{N-1}(i\boldsymbol{\omega}_x)(i\omega_t)^{N-1} + \cdots + a_0(i\boldsymbol{\omega}_x)}. \tag{6}$$

The absolute value of this function should approximate the spectral density well, $S(\boldsymbol{\omega}_x, \omega_t) \approx G(i\boldsymbol{\omega}_x, i\omega_t)G(-i\boldsymbol{\omega}_x, -i\omega_t)$. This can, for example, be done by forming a Taylor expansion of the inverse spectral density function in terms of $(i\omega_t)^2$. This results in a polynomial approximation of order $2N$, which is of form

$$\frac{1}{S(\boldsymbol{\omega}_x, \omega_t)} \approx c_0(i\boldsymbol{\omega}_x) + c_2(i\boldsymbol{\omega}_x)(i\omega_t)^2 + c_4(i\boldsymbol{\omega}_x)(i\omega_t)^4 + \cdots$$

We can now use the rational approximation in (6) to form the Fourier transform of $f(\mathbf{x}, t)$, which is

$$F(i\boldsymbol{\omega}_x, i\omega_t) = G(i\boldsymbol{\omega}_x, i\omega_t)N(i\boldsymbol{\omega}_x, i\omega_t),$$

where $N(i\boldsymbol{\omega}_x, i\omega_t)$ is the formal Fourier transform of a space–time white noise with unit spectral density. Consequently the spectral density of $F(i\boldsymbol{\omega}_x, i\omega_t)$ is $|F(i\boldsymbol{\omega}_x, i\omega_t)|^2 = G(i\boldsymbol{\omega}_x, i\omega_t)G(-i\boldsymbol{\omega}_x, -i\omega_t) \approx S(\boldsymbol{\omega}_x, \omega_t)$. The inverse Fourier transform with respect to time is denoted by $\tilde{f}(\boldsymbol{\omega}_x, t) = \mathscr{F}_t^{-1}[F(i\boldsymbol{\omega}_x, i\omega_t)]$. This gives

$$\frac{\mathrm{d}\tilde{\mathbf{f}}(\boldsymbol{\omega}_x, t)}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0(i\boldsymbol{\omega}_x) & -a_1(i\boldsymbol{\omega}_x) & \cdots & -a_{N-1}(i\boldsymbol{\omega}_x) \end{bmatrix} \tilde{\mathbf{f}}(\boldsymbol{\omega}_x, t) + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tilde{w}(\boldsymbol{\omega}_x, t),$$

where the process of interest is the first component $\tilde{f} = \tilde{\mathbf{f}}_1$ and $\tilde{w}(\boldsymbol{\omega}_x, t)$ is a scalar white noise process with constant spectral density $|b_0(i\boldsymbol{\omega}_x)|^2$.

Taking a second inverse Fourier transform $\mathscr{F}_x^{-1}[\cdot]$, now with respect to

the spatial variable $\mathbf{x}$, yields the following stochastic evolution equation,

$$\frac{\mathrm{d}\mathbf{f}(\mathbf{x},t)}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -\mathcal{A}_0 & -\mathcal{A}_1 & \cdots & -\mathcal{A}_{N-1} \end{bmatrix} \mathbf{f}(\mathbf{x},t) + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} w(\mathbf{x},t),$$

where $w(\mathbf{x},t)$ is a Hilbert space valued white noise process (see, e.g., Da Prato and Zabczyk, 1992) with a stationary spectral density operator $Q_c(\mathbf{x},\mathbf{x}') \triangleq Q_c(\mathbf{x}) = \mathscr{F}^{-1}\left[|b_0(i\boldsymbol{\omega}_x)|^2\right]$. The linear operators $\mathcal{A}_j$ are defined in terms of their Fourier transforms, such that $\mathcal{A}_j = \mathscr{F}_x^{-1}\left[a_j(i\boldsymbol{\omega}_x)\right]$, for $j = 0, 1, \ldots, N-1$.

Särkkä and Hartikainen (2012) point out that if the terms $a_j(i\boldsymbol{\omega}_x)$ are rational functions, the operators are so called integro-differential operators — and further, if they are polynomials, the equation becomes a stochastic partial differential equation of evolution type. Even if the functions are neither polynomials nor rational functions, the operators are so called pseudo-differential operators and the equation becomes a stochastic pseudo-differential equation or a fractional stochastic equation.

2.3 Optimal Estimation in Finite Dimensions

The term *optimal estimation* refers to the methods that are used to estimate the underlying state of a time-varying system of which there exist only indirectly observed noisy measurements. In many cases Kalman filter and Rauch–Tung–Striebel smoother (see, e.g., Grewal and Andrews, 2001; Särkkä, 2006; Solin, 2010) algorithms are the ones referred to with optimal estimation. These two algorithms can be used for computing the exact Bayesian posterior filtering distributions of the state in discrete-time linear Gaussian state space models.

2.3.1 Time Discretization

Before using the discrete-time methods, we start by considering *continuous-time* linear stochastic differential equations (SDEs) (see, e.g., Øksendal, 2003). This is because the time discretization plays and important role in the handling of continuous-time linear operator equations further on.

We consider the stochastic process defined by the following continuous-discrete state space model, where the dynamics are defined by a differential equation and the measurements are discrete in time,

$$\begin{aligned} \frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} &= \mathbf{F}(t)\,\mathbf{f}(t) + \mathbf{L}(t)\mathbf{w}(t) \\ \mathbf{y}(t_k) &= \mathbf{H}_k\,\mathbf{f}(t_k) + \mathbf{r}_k, \end{aligned} \tag{7}$$

where $\mathbf{y}(t_k) \in \mathbb{R}^d$ is the observation of the process at time $t_k, k = 0, 1, 2, \ldots$ and matrix $\mathbf{H}_k \in \mathbb{R}^{d\times s}$ defines the measurement model. $\mathbf{w}(t) \in \mathbb{R}^q$ is a q-dimensional white noise process with spectral density $\mathbf{Q}_c$. Because we consider linear time-invariant (LTI) models, we drop off the dependence of t in the dynamic model $\mathbf{F}(t)$. We will come back to time-dependency later on.

Linear continuous-time models can be handled by optimal estimation techniques by first discretizing the dynamics (see, e.g., Särkkä, 2006) of the model. If we assume that the model is time-invariant, the sampling period is Δt, and we define $t_k = k\,\Delta t$, then the *weak solution* (Øksendal, 2003) to this continuous-time stochastic differential equation can be expressed as

$$\mathbf{f}(t_{k+1}) = \exp(\Delta t\,\mathbf{F})\,\mathbf{f}(t_k) + \int_{t_k}^{t_{k+1}} \exp((t_{k+1} - s)\,\mathbf{F})\,\mathbf{L}\,\mathbf{w}(s)\ \mathrm{d}s. \tag{8}$$

The second integral above is just a Gaussian random variable with covariance

$$\mathbf{Q}_k = \int_0^{\Delta t} \exp((\Delta t - \tau)\,\mathbf{F})\,\mathbf{L}\,\mathbf{Q}_c\,\mathbf{L}^\mathsf{T} \exp((\Delta t - \tau)\,\mathbf{F})^\mathsf{T}\ \mathrm{d}\tau. \tag{9}$$

Thus, if we define $\mathbf{A}_k = \exp(\Delta t\,\mathbf{F})$, the model becomes a discrete-time state space model. This leads to the reformulation of Equation (7), which gives us the discrete-time state space model

$$\begin{aligned} \mathbf{f}(t_{k+1}) &= \mathbf{A}_k\,\mathbf{f}(t_k) + \mathbf{q}_k \\ \mathbf{y}(t_k) &= \mathbf{H}_k\,\mathbf{f}(t_k) + \mathbf{r}_k, \end{aligned} \tag{10}$$

where $\mathbf{f}(t_k) \in \mathbb{R}^s$ is the state at time t_k, where $k = 0, 1, 2, \ldots$, $\mathbf{y}(t_k) \in \mathbb{R}^d$ is the measurement at time t_k, $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ is the Gaussian process noise, and $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is the Gaussian measurement noise. Matrix $\mathbf{A}_k$ is the state transition matrix and $\mathbf{H}_k$ is the measurement model matrix.

2.3.2 Linear Estimation of Discrete-Time Models

As the model and state now are discrete in time, we drop the function representation of the state in Equation (10) for a more conveniently indexed presentation. Thus we may start by considering a linear stochastic state space model

$$\begin{aligned} \mathbf{f}_k &= \mathbf{A}_{k-1}\mathbf{f}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{f}_k + \mathbf{r}_k, \end{aligned} \tag{11}$$

where $\mathbf{f}_k \in \mathbb{R}^s$ is the state, $\mathbf{y}_k \in \mathbb{R}^d$ is the measurement of the state at time step k, $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0},\, \mathbf{Q}_k)$ is the Gaussian process noise of the dynamic model, $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0},\, \mathbf{R}_k)$ is the Gaussian noise process of the measurement model, $\mathbf{A}_k$ is the dynamic model translation matrix and $\mathbf{H}_k$ is the measurement model matrix. The time steps k run from 0 to T, and at time step $k = 0$ only the prior distribution is given, $\mathbf{f}_0 \sim \mathcal{N}(\mathbf{m}_0,\, \mathbf{C}_0)$.

The dynamic model defines the system dynamics and its uncertainties as a *Gauss–Markov sequence*. The discrete-time state space model presented in Equations (11) can be written equivalently in terms of probability distributions as a recursively defined probabilistic model of the form

$$\begin{aligned} p(\mathbf{f}_k \mid \mathbf{f}_{k-1}) &= \mathcal{N}(\mathbf{f}_k \mid \mathbf{A}_{k-1}\mathbf{f}_{k-1},\, \mathbf{Q}_{k-1}) \\ p(\mathbf{y}_k \mid \mathbf{f}_k) &= \mathcal{N}(\mathbf{y}_k \mid \mathbf{H}_k\mathbf{f}_k,\, \mathbf{R}_k). \end{aligned} \tag{12}$$

The model is assumed to be Markovian in the sense that it incorporates the *Markov property*, which means that the current state is conditionally independent from the past given the previous state. Additionally all the measurements of the separate states are assumed to be conditionally independent of each other given the state.

In this approach we bluntly divide the concept of Gaussian optimal estimation into three marginal distributions of interest (see, *e.g.*, Särkkä, 2006):

Filtering distributions $p(\mathbf{f}_k \mid \mathbf{y}_{1:k})$ that are the marginal distributions of current state $\mathbf{f}_k$ given all previous measurements $\mathbf{y}_{1:k} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k)$.

Prediction distributions $p(\mathbf{f}_k \mid \mathbf{y}_{1:k-1})$ that are the marginal distributions of forthcoming states.

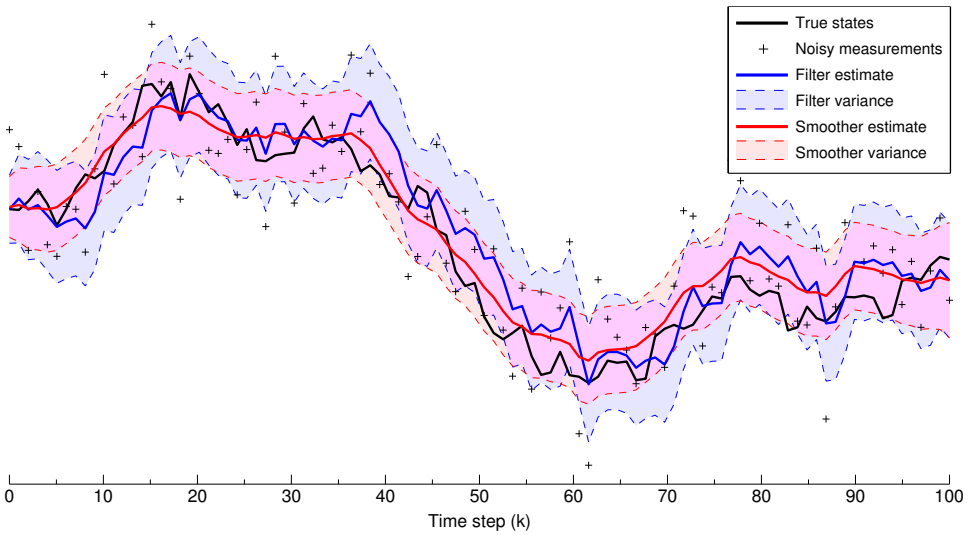


Figure 3: An illustrative example of the filtering and smoothing results for a linear Gaussian random-walk model. The variances are presented with the help of the 95 % confidence intervals.

Smoothing distributions $p(\mathbf{f}_k \mid \mathbf{y}_{1:T})$ that are the marginal distributions of the states $\mathbf{f}_k$ given measurements $\mathbf{y}_{1:T}$ such that $T > k$.

At time step k the prediction distribution utilizes less than k measurements, whereas the filtering solution uses exactly k measurements and the smoothing distribution more than k measurements.

An illustrative example of the differences between filtering and smoothing is shown in Figure 3. The black solid line in the figure demonstrates a realization of a Gaussian random walk process. The blue line together with the bluish patch following the line show the filtered solution obtained by using the noisy measurements in the figure. Similarly the red line and the reddish patch depict the smoothed solution. As the smoother has access to more measurements, it follows the original states more strictly and has a smaller variance than the filtering solution.

2.3.3 Kalman Filter Equations

The *Kalman filter* is a closed-form solution to the linear filtering problem in Equation (11) — or equivalently in (12). As the Kalman filter is conditional to all measurements up to time step k, the recursive filtering algorithm can

be seen as a two-step process that first includes calculating the marginal distribution of the next step using the known system dynamics (see, e.g., Bar-Shalom *et al.*, 2001). This is called the **Prediction step:**

$$\begin{aligned} \mathbf{m}_{k|k-1} &= \mathbf{A}_{k-1}\mathbf{m}_{k-1|k-1} \\ \mathbf{C}_{k|k-1} &= \mathbf{A}_{k-1}\mathbf{C}_{k-1|k-1}\mathbf{A}_{k-1}^{\mathsf{T}} + \mathbf{Q}_{k-1}. \end{aligned} \tag{13}$$

The algorithm then uses the observation to update the distribution to match the new information obtained by the measurement at step k. This is called the **Update step**:

$$\begin{aligned} \mathbf{S}_k &= \mathbf{H}_k\mathbf{C}_{k|k-1}\mathbf{H}_k^{\mathsf{T}} + \mathbf{R}_k \\ \mathbf{K}_k &= \mathbf{C}_{k|k-1}\mathbf{H}_k^{\mathsf{T}}\mathbf{S}_k^{-1} \\ \mathbf{m}_{k|k} &= \mathbf{m}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{m}_{k|k-1}) \\ \mathbf{C}_{k|k} &= \mathbf{C}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^{\mathsf{T}}. \end{aligned} \tag{14}$$

As a result, the filtered distribution at step k is given by $p(\mathbf{f}_k \mid \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{f}_k \mid \mathbf{m}_{k|k}, \mathbf{C}_{k|k})$. The difference $\mathbf{y}_k - \mathbf{H}_k\mathbf{m}_{k|k-1}$ in Equation (14) is called the innovation or the residual. It basically reflects the deflection between the actual measurement and the predicted measurement. The innovation is weighted by the *Kalman gain*. This term minimizes the *a posteriori* error covariance by weighting the residual with respect to the prediction step covariance $\mathbf{C}_{k|k-1}$ (see Maybeck, 1979, 1982; Welch and Bishop, 1995).

The linear Kalman filter solution coincides with the optimal least squares solution which is exactly the posterior mean $\mathbf{m}_{k|k}$. For derivation and further discussion on the matter see, for example, Kalman (1960), Maybeck (1979) and Särkkä (2006).

2.3.4 Rauch–Tung–Striebel Smoother Equations

We take a brief look at fixed-interval optimal smoothing. The purpose of optimal smoothing is to obtain the marginal posterior distribution of the state $\mathbf{f}_k$ at time step k, which is conditional on all the measurements $\mathbf{y}_{1:T}$, where $k \in [1, \ldots, T]$ is a fixed interval.

Similarly as the discrete-time linear Kalman filter gives a closed-form filtering solution, the discrete-time *Rauch–Tung–Striebel (RTS) Smoother* (see, e.g., Rauch *et al.*, 1965; Särkkä, 2006) gives a closed-form solution to

the linear smoothing problem. That is, the smoothed state is given as

$$p(\mathbf{f}_k \mid \mathbf{y}_{1:T}) = \mathcal{N}(\mathbf{f}_k \mid \mathbf{m}_{k|T},\, \mathbf{C}_{k|T}).$$

The RTS equations are written so that they utilize the Kalman filtering results $\mathbf{m}_{k|k}$ and $\mathbf{C}_{k|k}$ as a forward sweep, and then perform a backward sweep to update the estimates to use the forthcoming observations (see, e.g., Särkkä, 2006). The forward sweep is already presented in Equations (13) and (14). The smoother's backward sweep may be written as

$$\begin{aligned}
\mathbf{m}_{k+1|k} &= \mathbf{A}_k \mathbf{m}_{k|k} \\
\mathbf{C}_{k+1|k} &= \mathbf{A}_k \mathbf{C}_{k|k} \mathbf{A}_k^\mathsf{T} + \mathbf{Q}_k \\
\mathbf{G}_k &= \mathbf{C}_{k|k} \mathbf{A}_k^\mathsf{T} \mathbf{C}_{k+1|k}^{-1} \\
\mathbf{m}_{k|T} &= \mathbf{m}_{k|k} + \mathbf{G}_k (\mathbf{m}_{k+1|T} - \mathbf{m}_{k+1|k}) \\
\mathbf{C}_{k|T} &= \mathbf{C}_{k|k} + \mathbf{G}_k (\mathbf{C}_{k+1|T} - \mathbf{C}_{k+1|k}) \mathbf{G}_k^\mathsf{T},
\end{aligned} \tag{15}$$

where $\mathbf{m}_{k|T}$ is the smoothed mean and $\mathbf{C}_{k|T}$ the smoothed covariance at time step k. The RTS smoother can be seen as a discrete-time forward–backward filter, as the backward sweep utilizes information from the forward filtering sweep. When performing the backward recursion, the time steps run from T to 0.

2.4 Optimal Estimation in Infinite Dimensions

Next we consider the infinite-dimensional counterpart of the continuous-time state space model in Equation (7). We denote the space–time state by $\mathbf{f}(\mathbf{x}, t)$, where $\mathbf{x} \in \Omega$ (for some domain $\Omega \subseteq \mathbb{R}^n$) denotes the spatial variable and $t \in \mathbb{R}_+$ stands for time. We consider the case where the linear matrix evolution equation from before is replaced by a linear differential operator equation. We can then define the following stochastic equation (see, e.g., Särkkä and Hartikainen, 2012) in an infinite-dimensional state space form:

$$\begin{aligned}
\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial t} &= \boldsymbol{\mathcal{F}}\, \mathbf{f}(\mathbf{x}, t) + \mathbf{L}\mathbf{w}(\mathbf{x}, t) \\
\mathbf{y}_k &= \boldsymbol{\mathcal{H}}_k\, \mathbf{f}(\mathbf{x}, t) + \mathbf{r}_k,
\end{aligned} \tag{16}$$

where $\mathbf{x} \mapsto f_j(\mathbf{x}, t) \in \mathscr{H}(\mathbb{R}^n)$ for $j = 1, 2, \ldots, s$, and $\boldsymbol{\mathcal{F}}$ is an $s \times s$ matrix of linear operators operating on $\mathbf{x}$ with elements $\boldsymbol{\mathcal{F}}_{i,j} : \mathscr{H}(\mathbb{R}^n) \to \mathscr{H}(\mathbb{R}^n)$. The stochastic part is given by the matrix $\mathbf{L} \in \mathbb{R}^{s \times q}$ and $\mathbf{w}(\mathbf{x}, t)$ is a q-dimensional vector of Hilbert space $\mathscr{H}(\mathbb{R}^n)$ valued white noise processes with the joint spectral density operator $\mathbf{Q}_c(\mathbf{x}, \mathbf{x}')$.

The observation model in (16) is defined by the $d_k \times s$ -dimensional matrix $\boldsymbol{\mathcal{H}}_k$ of functionals operating on $\mathbf{x}$ with elements $\mathcal{H}_{i,j} : \mathscr{H}(\mathbb{R}^n) \to \mathbb{R}$. The observations are given as a vector $\mathbf{y}_k \in \mathbb{R}^{d_k}$, which corresponds to d_k observations at distinctive locations $\mathbf{x}_{i,k}^{\text{obs}} \in \Omega, i = 1, 2, \ldots, d_k$ at time step t_k. The measurement noise $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is a zero-mean Gaussian random variable.

The dynamic model above is an infinite-dimensional linear stochastic differential equation (Da Prato and Zabczyk, 1992). If $\boldsymbol{\mathcal{A}}$ is a differential operator, the Equation (16) is an *evolution type* stochastic partial differential equation (SPDE, see Chow, 2007; Pikkarainen, 2006). However, the same formulation also apply to a wider class of equations, where the operators are pseudo-differential operators (Shubin, 1987; Särkkä and Hartikainen, 2012).

2.4.1 Time Discretization

In Equation (16) we have written the spatio-temporal model as an evolution type SPDE, where we treat the temporal variable separately. The reason for this is to enable us to use infinite-dimensional optimal estimation methods. These methods are however meant for discrete time estimation, and thus we need to discretize the evolution equation with respect to time.

The discrete-time version of Equation (16) can be calculated similarly as in the finite-dimensional case in Section 2.3. We first form the evolution operator

$$\boldsymbol{\mathcal{U}}(\Delta t) = \exp\left(\Delta t \, \boldsymbol{\mathcal{F}}\right),$$

where $\exp(\cdot)$ is the operator exponential function. A solution to the stochastic equation can now be given as (Särkkä and Hartikainen, 2012)

$$\mathbf{f}(\mathbf{x}, t_{k+1}) = \boldsymbol{\mathcal{U}}(t_{k+1} - t_k)\, \mathbf{f}(\mathbf{x}, t_k) + \int_{t_k}^{t_{k+1}} \boldsymbol{\mathcal{U}}(t_{k+1} - \tau)\, \mathbf{L}\, \mathbf{w}(\mathbf{x}, \tau)\, \mathrm{d}\tau, \quad (17)$$

where t_{k+1} and $t_k < t_{k+1}$ are arbitrary. The second term is Gaussian process with covariance function $\mathbf{Q}(\mathbf{x}, \mathbf{x}'; t, t') = \int_{t'}^{t} \boldsymbol{\mathcal{U}}(t - \tau)\, \mathbf{L}\, \mathbf{Q}_c\, \mathbf{L}^{\mathsf{T}}\, \boldsymbol{\mathcal{U}}^*(t - \tau)\, \mathrm{d}\tau$.

This leads to the following discrete-time model

$$\begin{aligned} \mathbf{f}(\mathbf{x}, t_k) &= \boldsymbol{\mathcal{U}}(\Delta t_k)\, \mathbf{f}(\mathbf{x}, t_{k-1}) + \mathbf{q}_k(\mathbf{x}) \\ \mathbf{y}_k &= \boldsymbol{\mathcal{H}}_k\, \mathbf{f}(\mathbf{x}, t) + \mathbf{r}_k, \end{aligned} \tag{18}$$

where $\Delta t_k = t_k - t_{k-1}$ and $\mathbf{q}_k(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{Q}(\mathbf{x}, \mathbf{x}'; t_k, t_{k-1}))$.

This discretization is not an approximation, but the so called *mild solution* to the infinite-dimensional differential equation. The mild solution is a weaker solution concept than the weak solution of a stochastic process. However, it is worth noting, that in many circumstances the mild and weak solutions coincide (see Da Prato and Zabczyk, 1992, for proofs and discussion).

2.4.2 Infinite-Dimensional Kalman Filter

The *infinite-dimensional Kalman filter* (see Tzafestas, 1978; Omatu and Seinfeld, 1989; Cressie and Wikle, 2002) is a closed-form solution to the infinite-dimensional linear filtering problem. As in the finite-dimensional case, we present a two-step scheme that first includes calculating the marginal distribution of the next step using the known system dynamics. The following formulation uses a similar notation as Särkkä and Hartikainen (2012) and can be compared to the finite-dimensional filter in Equations (13) and (14).

The infinite-dimensional **Prediction step** can be written as follows:

$$\begin{aligned} \mathbf{m}_{k|k-1}(\mathbf{x}) &= \boldsymbol{\mathcal{U}}(\Delta t_k)\, \mathbf{m}_{k-1|k-1}(\mathbf{x}) \\ \mathbf{C}_{k|k-1}(\mathbf{x}, \mathbf{x}') &= \boldsymbol{\mathcal{U}}(\Delta t_k)\, \mathbf{C}_{k-1|k-1}(\mathbf{x}, \mathbf{x}')\, \boldsymbol{\mathcal{U}}^*(\Delta t_k) + \mathbf{Q}(\mathbf{x}, \mathbf{x}'; t_k, t_{k-1}), \end{aligned} \tag{19}$$

where $(\cdot)^{-1}$ denotes the matrix or operator inverse and $(\cdot)^*$ denotes an adjoint which in practice swaps the roles of inputs $\mathbf{x}$ and $\mathbf{x}'$ and operates from the right. The operator adjoint can be compared to the matrix transpose. The recursive iteration is initialized by presenting the prior in the form $\mathbf{f}(\mathbf{x}, t_0) \sim \mathcal{GP}\left(\mathbf{m}_0(\mathbf{x}), \mathbf{C}_0(\mathbf{x}, \mathbf{x}')\right)$.

The algorithm then uses the observation to update the distribution to match the new information obtained by the measurement at step k. This is

the infinite-dimensional **Update step**:

$$\begin{aligned}
\mathbf{S}_k &= \boldsymbol{\mathcal{H}}_k \, \mathbf{C}_{k|k-1}(\mathbf{x}, \mathbf{x}') \, \boldsymbol{\mathcal{H}}_k^* + \mathbf{R}_k \\
\mathbf{K}_k(\mathbf{x}) &= \mathbf{C}_{k|k-1}(\mathbf{x}, \mathbf{x}') \, \boldsymbol{\mathcal{H}}_k^* \, \mathbf{S}_k^{-1} \\
\mathbf{m}_{k|k}(\mathbf{x}) &= \mathbf{m}_{k|k-1}(\mathbf{x}) + \mathbf{K}_k(\mathbf{x}) \, \left(\mathbf{y}_k - \boldsymbol{\mathcal{H}}_k \mathbf{m}_{k|k-1}(\mathbf{x}) \right) \\
\mathbf{C}_{k|k}(\mathbf{x}, \mathbf{x}') &= \mathbf{C}_{k|k-1}(\mathbf{x}, \mathbf{x}') - \mathbf{K}_k(\mathbf{x}) \, \mathbf{S}_k \, \mathbf{K}_k^*(\mathbf{x}).
\end{aligned} \tag{20}$$

As a result the filtered forward-time posterior process at step k (time t_k) is given by $\mathbf{f}_{k|k}(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{m}_{k|k}(\mathbf{x}),\, \mathbf{C}_{k|k}(\mathbf{x}, \mathbf{x}')\right)$.

2.4.3 Infinite-Dimensional Rauch–Tung–Striebel Smoother

The infinite-dimensional Rauch–Tung–Striebel smoother equations are written so that they utilize the Kalman filtering results $\mathbf{m}_{k|k}(\mathbf{x})$ and $\mathbf{C}_{k|k}(\mathbf{x}, \mathbf{x}')$ as a forward sweep, and then perform a backward sweep to update the estimates to match the forthcoming observations. The smoother's backward sweep may be written with the following infinite-dimensional **RTS smoothing equations** (Särkkä and Hartikainen, 2012):

$$\begin{aligned}
\mathbf{m}_{k+1|k}(\mathbf{x}) &= \boldsymbol{\mathcal{U}}(\Delta t_k) \, \mathbf{m}_{k|k}(\mathbf{x}) \\
\mathbf{C}_{k+1|k}(\mathbf{x}, \mathbf{x}') &= \boldsymbol{\mathcal{U}}(\Delta t_k) \, \mathbf{C}_{k|k}(\mathbf{x}, \mathbf{x}') \, \boldsymbol{\mathcal{U}}^*(\Delta t_k) + \mathbf{Q}_k(\mathbf{x}, \mathbf{x}'; t_k, t_{k-1}) \\
\mathbf{G}_k(\mathbf{x}) &= \mathbf{C}_{k|k}(\mathbf{x}, \mathbf{x}') \, \boldsymbol{\mathcal{U}}^*(\Delta t_k) \, \left[\mathbf{C}_{k+1|k}(\mathbf{x}, \mathbf{x}') \right]^{-1} \\
\mathbf{m}_{k|T}(\mathbf{x}) &= \mathbf{m}_{k|k}(\mathbf{x}) + \mathbf{G}_k(\mathbf{x}, \mathbf{x}') \, \left[\mathbf{m}_{k+1|T}(\mathbf{x}) - \mathbf{m}_{k+1|k}(\mathbf{x}) \right] \\
\mathbf{C}_{k|T}(\mathbf{x}, \mathbf{x}') &= \mathbf{C}_{k|k}(\mathbf{x}, \mathbf{x}') + \mathbf{G}_k(\mathbf{x}) \left(\mathbf{C}_{k+1|T} \left(\mathbf{x}, \mathbf{x}' \right) - \mathbf{C}_{k+1|k}(\mathbf{x}, \mathbf{x}') \right) \mathbf{G}_k^*(\mathbf{x}).
\end{aligned} \tag{21}$$

The discrete-time backward sweep utilizes information from the forward filtering steps, and thus the time steps run from T to 0.

Now that we have run both the Kalman filtering and Rauch–Tung–Striebel sweeps on the model given the observed data, we have the marginal posterior that can be given as the Gaussian process

$$\mathbf{f}(\mathbf{x}, t_k \mid \mathbf{y}_{1:T}) \sim \mathcal{GP}\left(\mathbf{m}_{k|T}(\mathbf{x}), \mathbf{C}_{k|T}(\mathbf{x}, \mathbf{x}')\right),$$

where the observed values $\mathbf{y}_k \in \mathbb{R}^{d_k}$ are given on discrete time points $t_k, k = 1, 2, \ldots, T$, and measured at known locations $\mathbf{x}_{i,k}^{\text{obs}} \in \Omega, i = 1, \ldots, d_k$.

The observant reader might have noticed that during the estimation the infinite-dimensional Kalman filtering approach evaluates values only for inference with respect to the observations, that is for known locations $\mathbf{x}_{i,k}^{\text{obs}}$. However the resulting process functions can be evaluated at any test points $\mathbf{x}_* \in \Omega$ by simply considering an appropriate measurement functional $\mathcal{H}$. The marginal posterior of the value of $\mathbf{f}(\mathbf{x}_*, t_k)$ in $\mathbf{x}_*$ at time instant t_k is thus

$$p\left(\mathbf{f}(\mathbf{x}_*, t_k) \mid \mathbf{y}_{1:T}\right) = \mathcal{N}\left(\mathbf{f}(\mathbf{x}_*, t_k) \mid \mathbf{m}_{k|T}(\mathbf{x}_*), \mathbf{C}_{k|T}(\mathbf{x}_*, \mathbf{x}_*)\right). \tag{22}$$

Predicting values at more time steps could also be included. A test time point t_* should be taken into account when doing the time discretization and the state of the system $\mathbf{f}(\mathbf{x}, t_*)$ would be predicted on this step, but as there is no data, no update step would be needed.

As a noteworthy detail we point out the connection between the standard (in this case) spatial GP model and the evolution type state space SPDE. The model in Equation (16) coincide with the GP formulation in Equation (3). If we leave out the temporal evolution model, that is $\boldsymbol{\mathcal{F}} = 0$ and $\mathbf{Q}_c(\mathbf{x}, \mathbf{x}') = 0$, the estimation task for this model could be solved by considering only one measurement step and using the same equations.

3 Approximative Numerical Solutions

Practical implementations of infinite-dimensional Kalman filtering require some sort of approximations to be used. In this study we take a basis function (Hilbert space) approach, which shows beneficial in the examples further on.

Hereafter we concentrate our interest on the Laplace operator in domains that are subject to certain symmetries and can be easily dealt with in numerical implementations. We start by showing how the eigenfunction expansion of the Laplacian operator subject to Dirichlet boundary conditions can be given using orthonormal basis functions subject to the L^2 inner product.

Furthermore, we show how the linear operator equation models can be approximated using the eigenfunction expansion and thereby applied in the infinite-dimensional Kalman filtering context. This is used at the end of this section where we form finite-dimensional approximations to the infinite-dimensional models.

3.1 Eigenfunction Expansions of the Laplacian Subject to Dirichlet Boundary Conditions

We consider an arbitrary domain Ω that has a boundary $\partial\Omega$. Inside the domain some real-valued process can be given in terms of a function $f(\mathbf{x}, t)$, for $(\mathbf{x}, t) \in \Omega \times \mathbb{R}_+$, where t stands for time. Even though the function f could be a function of several other variables as well, we only consider one spatial variable $\mathbf{x} \in \Omega$ and one temporal variable.

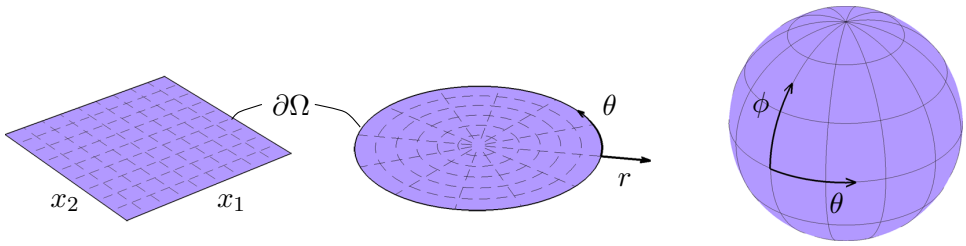


Figure 4: We consider the following two-dimensional surface domains Ω; on far left a rectangular domain given in Cartesian coordinates, in the middle a disk given in polar coordinates, and on far right the hull of an S^2 sphere given in angular coordinates.

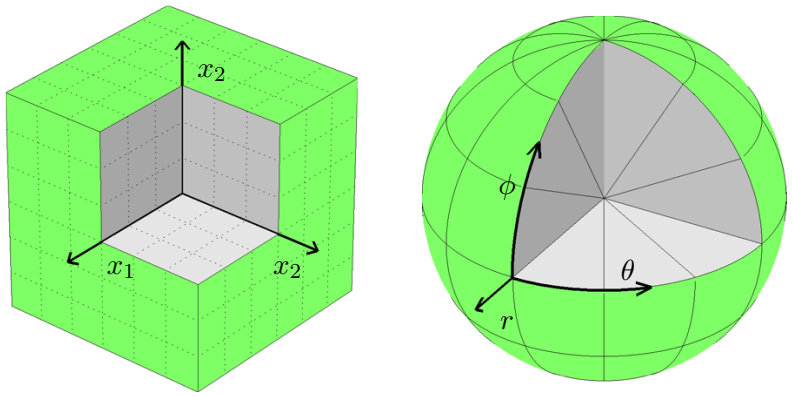


Figure 5: We consider the following three-dimensional domains Ω; on the left side a box given in Cartesian coordinates, and on the right a sphere given spherical coordinates.

Furthermore, we suppose the function to be zero everywhere on the boundary $\mathbf{x} \in \partial\Omega$. This property is commonly referred to as the *Dirichlet boundary condition* and is in practice the simplest boundary condition. A function $f(\mathbf{x}, t)$ is harmonic if operating on the function with the Laplace operator $\nabla^2 = \nabla \cdot \nabla = \Delta$ yields zero for all $t \in \mathbb{R}_+$. We write the problem as

$$\begin{aligned} \nabla^2 f(\mathbf{x}, t) &= 0, && (\mathbf{x}, t) \in \Omega \times \mathbb{R}_+ \\ f(\mathbf{x}, t) &= 0, && (\mathbf{x}, t) \in \partial\Omega \times \mathbb{R}_+. \end{aligned}$$

We start by considering a one-dimensional domain $\Omega \subset \mathbb{R}$, where $\Omega = \{x \mid -L < x < L\}$. The value on the boundary is $f(L) = f(-L) = 0$. The eigenvalue problem can be given as

$$\nabla^2 \psi_n(x) = \lambda_n \psi_n(x),$$

where $\psi_n(x)$ is the nth eigenfunction and λ_n the corresponding eigenvalue. Solving the problem yields the solution

$$\lambda_n = \frac{n\pi}{2L} \quad \text{and} \quad \psi_n(x) = \sqrt{\frac{1}{L}} \sin\left(\frac{n\pi(x+L)}{2L}\right). \tag{23}$$

For different values of n, these eigenfunctions are all possible eigenfunctions and they form a complete orthonormal basis that can be used for evaluating

any function, with sufficient continuity and smoothness properties, over the domain. As discussed at the very beginning of this study, this means that the one-dimensional Laplacian can be associated with the *formal kernel* (even though the sum does not converge)

$$k(x, x') = \sum_n \lambda_n \psi_n(x) \psi_n(x'),$$

such that

$$\nabla^2 f(x,t) = \int k(x, x') f(x', t) \, \mathrm{d}x'.$$

We continue by considering several higher-dimensional domains in the next sections. Figure 4 shows two-dimensional domains, a rectangular area in Cartesian coordinates, a disk in polar coordinates and the surface of a sphere in angular coordinates. Further on, we will also present two three-dimensional domains that are visualized in Figure 5: a three-dimensional cube in Cartesian coordinates and a sphere in spherical coordinates.

3.1.1 In an n-Dimensional Hypercube

Both the two dimensional rectangle in Figure 4 and the three-dimensional cube in Figure 5 fall under the same formulation of n-dimensional hypercubes. Hereafter we denote the dimensionality by d and reserve n_i for indexing the eigenvalues. The Laplace operator in d-dimensional Cartesian coordinates can be given as $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$. Assuming separable solutions, the one-dimensional results can be extended rather straightforward to higher dimensions from Equation (23).

We first consider a two-dimensional rectangle $\Omega = \{(x_1, x_2) \mid -L_1 \leq x_1 \leq L_1, -L_2 \leq x_2 \leq L_2\} \subset \mathbb{R}^2$. By assuming separable solutions, the eigenfunctions and eigenvalues from the one-dimensional solution in (23) can be generalized to two dimensions (see, e.g., Pivato, 2010),

$$\begin{aligned} \lambda_{n_1,n_2} &= \frac{n_1 \pi}{2L_1} + \frac{n_2 \pi}{2L_2} \quad \text{and} \\ \psi_{n_1,n_2}(\mathbf{x}) &= \sqrt{\frac{1}{L_1 L_2}} \sin\left(\frac{n_1 \pi (x_1 + L_1)}{2L_1}\right) \sin\left(\frac{n_2 \pi (x_2 + L_2)}{2L_2}\right), \end{aligned} \tag{24}$$

for index pairs $(n_1, n_2) \in \mathbb{N}^2$. Now, the eigenvalues and eigenfunctions in a

d-dimensional hypercube are given by

$$\begin{aligned} \lambda_{n_1,n_2,\ldots} &= \sum_{i=1}^{d} \frac{n_i \pi}{2L_i} \quad \text{and} \\ \psi_{n_1,n_2,\ldots}(\mathbf{x}) &= \prod_{i=1}^{d} \sqrt{\frac{1}{L_i}} \sin\left(\frac{n_i \pi (x_i + L_i)}{2L_i}\right), \end{aligned} \tag{25}$$

where $\mathbf{x} \in \Omega \subset \mathbb{R}^d$, for index combinations $(n_1, n_2, \ldots, n_d) \in \mathbb{N}^d$.

3.1.2 In an n-Dimensional Hypersphere

In the following sections we consider domains that are given in different spherical coordinates: polar, angular and spherical polar. These solutions are in general subject to more complicated structures and given in terms of different orthogonal functions. However, they provide an effective basis for real-world problems, as will be demonstrated further on. Details can be found, for example, in Pivato (2010).

In Polar Coordinates

From Figure 4 we consider a circular disk in two dimensions given by $\Omega = \{(x, y) \mid x^2 + y^2 \leq L^2\} \subset \mathbb{R}^2$, where we choose the disk to have radius L. We study the problem as earlier, subject to Dirichlet boundary conditions $f(\mathbf{x}, t) = 0$ for $\mathbf{x} \in \partial\Omega = \{(x, y) \mid x^2 + y^2 = L^2\}$.

Polar coordinates give each point in a two-dimensional plane as a distance from a fixed point, typically origin, and an angle from a fixed direction. The coordinates are given as a pair (r, θ), where $r \in \mathbb{R}_+$ is the radial coordinate and $\theta \in [0, 2\pi)$ is the angular coordinate or azimuth. We use the same notation in higher dimensions as well, as will be explained further on.

The relationship between Cartesian (x, y) and polar coordinates (r, θ) is given by the trigonometric identities: $x = r\cos\theta$, $y = r\sin\theta$ and $r = \sqrt{x^2 + y^2}, \theta = \arctan\left(\frac{y}{x}\right)$. Changing to polar coordinates yields the domain $\Omega = \{r \mid r \leq L\}$ and the Laplace operator can be given as (Arfken and Weber, 2001)

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}. \tag{26}$$

Solving the problem by assuming separable solutions $\psi(r, \theta) = R(r)\Theta(\theta)$, yields the following parts: an angular part $\Theta(\theta)$ and a radial part $R(r)$. The

angular part has the boundary condition $\Theta(0) = \Theta(2\pi)$. If it satisfies the Laplace's equation $\nabla^2\psi = 0$ using Equation (26), it yields the angular part $\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\Theta(\theta) = -m^2\Theta(\theta)$, which has solutions $\exp(mi\theta)$, where $m \in \mathbb{Z}$ is an integer.

Bessel functions (Abramowitz and Stegun, 1964) arise as the solutions $R(r) \triangleq y(x)$ of the Bessel differential equation, $x^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}y(x) + x\frac{\mathrm{d}}{\mathrm{d}x}y(x) + (x^2 - \alpha^2)y(x) = 0$, and for different values of $\alpha \in \mathbb{C}$ (the order of the Bessel function). For integer orders $\alpha = m \in \mathbb{Z}$ these solutions are commonly denoted by $J_m(x)$ and the Bessel functions called Bessel functions of first kind.

A Bessel function of first kind, $J_m(x)$, can be defined by its Taylor expansion around $x = 0$,

$$J_m(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,\Gamma(k+m+1)}\left(\tfrac{1}{2}x\right)^{2k+m}, \tag{27}$$

where $\Gamma(\cdot)$ is the gamma function and can be replaced by the factorial $(k+m+1)!$ for functions of integer order. Another definition of the Bessel functions of first kind is the integral definition (see, e.g., Abramowitz and Stegun, 1964)

$$J_m(x) = \frac{1}{\pi}\int_0^{\pi} \cos(m\theta - x\sin\theta)\,\mathrm{d}\theta.$$

The Bessel functions are not periodic, which also implies that their roots are not periodic. However one can show that for the kth root $\alpha_{k,m}$ of the Bessel function $J_m(x)$ it holds that $\alpha_{k,m} \approx (k + \frac{1}{2}m - \frac{1}{4})\pi$ as $k \to \infty$ (see, e.g., Olver, 2012).

If we consider the radial part of the eigenvalue problem — which is occasionally referred to by using the Helmholtz equation, $\nabla^2\psi + k^2\psi = 0$ — the eigenfunctions are the Bessel functions of first kind $J_m(\alpha_{n,m}r)$ and the eigenvalues are the square of the positive zeros of the Bessel functions; $\lambda_{n,m} = \alpha_{n,m}^2, n = 1, 2, \ldots$ and $m = 0, 1, \ldots$. Taking the angular part into account yields the polar eigenfunctions in a disk,

$$\psi_{n,m}(r,\theta) = \begin{cases} J_m(\alpha_{n,m}r/L)\cos|m|\theta, & \text{when} \quad m = 0, 1, \ldots \\ J_m(\alpha_{n,m}r/L)\sin|m|\theta, & \text{when} \quad m = -1, -2, \ldots \end{cases} \tag{28}$$

for which the corresponding eigenvalues are $\lambda_{n,m} = \alpha_{n,m}^2$. A few of the first

Table 1: Table of the six first positive roots $\alpha_{m,k}, k = 1, 2, \ldots, 6$, of the Bessel functions of first kind $J_m(x)$ for $m = 0, 1, \ldots, 5$. The roots have been numerically solved, and more extensive tables can easily be found in literature.

$\alpha_{n,m}$	Bessel function order m					
nth root	0	1	2	3	4	5
1	2.4048	3.8317	5.1356	6.3802	7.5883	8.7715
2	5.5201	7.0156	8.4172	9.7610	11.0647	12.3386
3	8.6537	10.1735	11.6198	13.0152	14.3725	15.7002
4	11.7915	13.3237	14.7960	16.2235	17.6160	18.9801
5	14.9309	16.4706	17.9598	19.4094	20.8269	22.2178
6	18.0711	19.6159	21.1170	22.5827	24.0190	25.4303

roots $\alpha_{n,m}$ are given in Table 1. The truncated expansion is in this work given in terms of indices $n = 1, 2, \ldots, N$ and $m = -M, \ldots, -1, 0, 1, \ldots, M$.

In Spherical Coordinates

Spherical coordinates define a coordinate system, which can be seen as a generalization of the polar coordinates to three dimensions. Each point in $\mathbb{R}^3$ can be written with the help of two angles and a radial coordinate, the Euclidean distance from origin (further interpretation can be found in, e.g., Arfken and Weber, 2001).

In notation, some care has to be taken, because different conventions in denoting the angles is used in different fields of science. In this study, we will use the notation $\mathbf{x} = (r, \theta, \phi)$, where θ is the azimuthal (longitudinal) coordinate with $\theta \in [0, 2\pi)$, and ϕ the polar (colatitudal) coordinate, $\phi \in [0, \pi]$, that ranges from the polar axis. This notation is often used in mathematics, whereas physicists prefer the alternative notation, where θ and ϕ are reversed, and θ is latitudal, ranging from the equator. For graphical interpretation, refer to Figure 5.

The radial coordinate defines the distance from origin and is within $r \in [0, L]$, where L is the radius of the sphere. The domain $\Omega \to \mathbb{R}^3$ as $r \to \infty$. We may write the transformations between Cartesian coordinates

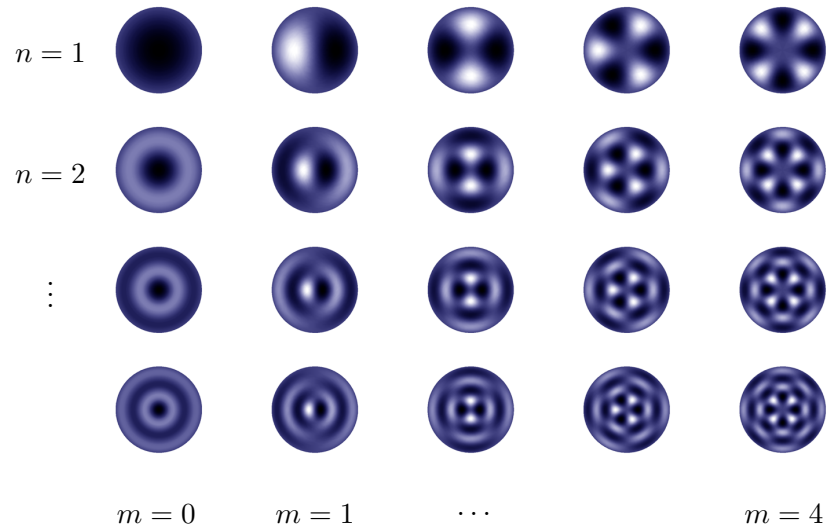


Figure 6: Visualisations of a few first eigenfunctions of the Laplace operator in a unit disc. The visualizations corresponding to negative values of m look similar but are rotated by 90 degrees.

and spherical polar coordinates as:

$$\begin{aligned} x &= r\sin\phi\,\cos\theta & r &= \sqrt{x^2+y^2+z^2} \\ y &= r\sin\phi\,\sin\theta & \theta &= \arctan\left(\frac{y}{x}\right) \\ z &= r\cos\phi & \phi &= \arccos\left(\frac{z}{r}\right) \end{aligned}$$

where one has to take the quadrant into account when taking the inverse tangent. The volume element in spherical coordinates can be written as $\mathrm{d}\Omega = r^2\sin\phi\,\mathrm{d}\phi\,\mathrm{d}\theta\,\mathrm{d}r$.

We study a sphere of radius $r = L$, which we denote by Ω. We try define the Diriclet problem in the sphere similarly as in the disk earlier. In this case the Diriclet boundary condition defines that the function value $f(r,\theta,\phi)$ on the surface of the sphere $\partial\Omega$ (i.e. when $r = L$) to zero.

The Laplacian in spherical coordinates can be given as (see, e.g., Arfken and Weber, 2001)

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) \tag{29}$$

The Helmholtz equation $\nabla^2\psi + k^2\psi = 0$ is separable in spherical coordinates.

Table 2: Table of the six first positive roots $\sigma_{n,m}, n = 1, 2, \ldots, 6$, of the Spherical Bessel functions of first kind $S_m(x)$ for $m = 0, 1, \ldots, 5$. The roots have been numerically solved, and more extensive tables can easily be found in literature.

$\sigma_{n,m}$	Spherical Bessel function order m					
nth root	0	1	2	3	4	5
1	3.1416	4.4934	5.7635	6.9879	8.1826	9.3558
2	6.2832	7.7253	9.0950	10.4171	11.7049	12.9665
3	9.4248	10.9041	12.3229	13.6980	15.0397	16.3547
4	12.5664	14.0662	15.5146	16.9236	18.3013	19.6532
5	15.7080	17.2208	18.6890	20.1218	21.5254	22.9046
6	18.8496	20.3713	21.8539	23.3042	24.7276	26.1278

That means that one may write the solution $\psi(r, \theta, \phi)$ as a product of three functions, $\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi)$, where $R(r)$ is a radial function, $\Theta(\theta)$ is a longitudal function, and $\Phi(\phi)$ a function only depending on the colatitudal polar coordinate.

The solutions for the Helmholtz equation corresponds to the solving the eigenfunction of the negative Laplace operator in a three-dimensional sphere. The radial part $R(r)$ is determined by Spherical Bessel functions, such that $R(r) = S_m(\sigma r/L)$, where σ is a zero of the function such that the boundary condition $R(L) = 0$ is satisfied.

The *Spherical Bessel function* S_m of order $m \geq 0$ is defined by the formula (for definitions and more detailed discussion see, e.g., Abramowitz and Stegun, 1964; Olver, 2012; Arfken and Weber, 2001)

$$S_m(x) = \sqrt{\frac{\pi}{2x}} J_{m+1/2}(x), \tag{30}$$

where J_m is the Bessel function of first kind as in Equation (27). Unlike the Bessel functions of integer order, the spherical Bessel functions are elementary functions. For example the spherical Bessel function of order $m = 0$ is

$$S_0(x) = \frac{\sin x}{x},$$

which means that the roots are evenly spaced at $\pi, 2\pi, \ldots$. The higher order spherical Bessel functions are given by the recurrence relation $S_{m+1}(x) = -\frac{\mathrm{d}}{\mathrm{d}x} S_m(x) + \frac{m}{x} S_m(x)$ (Abramowitz and Stegun, 1964). Therefore the next

three spherical Bessel functions are given by

$$S_1(x) = -\frac{\cos x}{x} + \frac{\sin x}{x^2},$$
$$S_2(x) = -\frac{\sin x}{x} - \frac{3\cos x}{x^2} + \frac{3\sin x}{x^3}, \quad \text{and}$$
$$S_3(x) = \frac{\cos x}{x} - \frac{6\sin x}{x^2} + \frac{15\cos x}{x^3} + \frac{15\sin x}{x^4}.$$

Taking into account the homogeneous Dirichlet boundary conditions requires the solution to be zero on the boundary. We use the zeros of the Spherical Bessel functions to rescale the functions to meet this requirement. The positive roots of the spherical Bessel function $S_m(x)$ are denoted by $\sigma_{n,m}$, where n denotes the nth positive root. These values satisfy $S_m(\sigma_{n,m}) = 0$, for all $n = 1, 2, \ldots$ and $m = 0, 1, \ldots$.

The angular part $\Theta(\theta)\Phi(\phi)$ is given by the Laplace spherical harmonics that are described in detail in the next section. We denote this part as $\Theta(\theta)\Phi(\phi) = Y_m^k(\theta, \phi)$ for indices $m = 0, 1, \ldots, M$ and $k = -m, 1 - m, \ldots, m - 1, m$.

Now we may construct the separable eigenfunctions of the Helmholtz equation by putting together the spherical harmonics and spherical Bessel functions

$$\psi_{n,m,k}(r, \theta, \phi) = S_m(\sigma_{n,m} r/L)\, Y_m^k(\theta, \phi) \tag{31}$$

and corresponding eigenvalues

$$\lambda_{n,m} = \sigma_{n,m}^2, \tag{32}$$

for $n = 1, 2, \ldots, N$, $m = 0, 1, 2, \ldots, M$ and $k = -m, \ldots, -1, 0, 1, \ldots, m$. For values of $\sigma_{n,m}$ refer to Table 2, where some of the first are explicitly shown. In this study the truncated expansion is given in terms of specifying upper bounds for indices n and m by N and M, respectively. The number of eigenfunctions is thus $N(M + 1)^2$.

Spherical Harmonics

Spherical harmonics $Y_m^k(\theta, \phi)$ are the angular part of the solution to Laplace's equation in spherical coordinates; solutions to Laplace's equation $\nabla^2 f = 0$ are called 'harmonic' functions. The Laplace's spherical harmonics form an orthogonal basis, and are therefore an important tool

in many fields of science. In this study spherical harmonics are used in combination with other orthogonal functions to form the eigensolutions to the Laplacian in n-dimensional spheres. They are however presented separately, because solving systems on the surface of a 2-sphere S^2, where $S^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\| = L\}$, is useful as well. In Laplace's equation (or Helmholz's for that matter) the angular dependencies come entirely from the Laplacian operator, and by using the definition of the Laplacian in spherical coordinates — see Equation (31) — the solution can be given as (Arfken and Weber, 2001)

$$\frac{\Theta(\theta)}{\sin\phi}\frac{\mathrm{d}}{\mathrm{d}\phi}\left(\sin\phi\frac{\mathrm{d}\Phi(\phi)}{\mathrm{d}\phi}\right) + \frac{\Phi(\phi)}{\sin^2\phi}\frac{\mathrm{d}^2\Theta(\theta)}{\mathrm{d}\theta^2} + m(m+1)\Phi(\phi)\Theta(\theta) = 0, \quad (33)$$

where m is an integer.

The reader is reminded that here the notation defines $\theta \in [0, 2\pi]$ as the azimuthal coordinate and $\phi \in [0, \pi)$ as the colatitudal polar coordinate, which differs from, for example, the notation in Arfken and Weber (2001). Visual interpretation can be found in Figure 4.

Separation of variables yields for the azimuthal part

$$\frac{1}{\Theta(\theta)}\frac{\mathrm{d}^2\Theta(\theta)}{\mathrm{d}\theta^2} = -m^2,$$

with solutions $\Theta(\theta) = \exp(im\theta)$, where m is an integer. This defines the solutions to be complex-valued and also features the complex conjugates for each solution. We can also define real solutions that are $\Theta(\theta) = \sin m\theta$ and $\Theta(\theta) = \cos m\theta$, where m is an integer as earlier. To ease the notation we use the following indexing of the real solutions

$$\Theta_m(\theta) = \begin{cases} \cos|m|\theta, & \text{for non-negative } m \\ \sin|m|\theta, & \text{for negative } m \end{cases}$$

where $m \in \mathbb{Z}$.

The remaining polar angle (ϕ) dependence in Equation (33) leads to the general Legendre differential equation, $(1-x^2)\frac{\mathrm{d}^2}{\mathrm{d}x^2}y(x) - 2x\frac{\mathrm{d}}{\mathrm{d}x}y(x) + \left(m(m+1) - \frac{k^2}{1-x^2}\right)y(x) = 0$. Solutions to this equation are the associated Legendre polynomials $P_m^k(x) : [-1, 1] \to \mathbb{R}$ (Abramowitz and Stegun, 1964),

which means that the polar solution can be given as

$$\Phi_{m,k}(\phi) = P_m^k(\cos\phi),$$

where m and k are integers such that $-m \leq k \leq m$.

Strictly speaking, the solutions of the Legendre differential equation are polynomials only for those with $k = 0$. The first few associated Legendre functions are

$$\begin{aligned}
P_0^0(x) &= 1 & P_2^{-2}(x) &= \frac{1}{24}P_2^2(x) \\
& & P_2^{-1}(x) &= -\frac{1}{6}P_2^1(x) \\
P_1^{-1}(x) &= -\frac{1}{2}P_1^1(x) & P_2^0(x) &= \frac{1}{2}(3x^2-1) \\
P_1^0(x) &= x & P_2^1(x) &= -3x(1-x^2)^{1/2} \\
P_1^1(x) &= -(1-x^2)^{1/2} & P_2^2(x) &= 3(1-x^2)
\end{aligned}$$

Now the functions $\Theta_m(\theta)$ (the sines and cosines) are orthonormal with respect to azimuthal angle θ, and the function $\Phi_{m,k}(\phi)$ (the associated Legendre functions) are orthonormal with respect to the polar angle ϕ, if the normalization of the functions is chosen accordingly. The spherical harmonics are defined by combining the two angular parts $\Theta(\theta)$ and $\Phi(\phi)$ and applying the normalization. The *complex-valued spherical harmonics* can be given as

$$Y_m^k(\theta,\phi) = \sqrt{\frac{2m+1}{4\pi}\frac{(m-k)!}{(m+k)!}}\,P_m^k(\cos\phi)e^{ik\theta} \tag{34}$$

or similarly the *real spherical harmonics*

$$Y_m^k(\theta,\phi) = \sqrt{\frac{2m+1}{2\pi}\frac{(m-k)!}{(m+k)!}}\begin{cases} P_m^{|k|}(\cos\phi)\cos|k|\theta, & \text{for } k \geq 0 \\ P_m^{|k|}(\cos\phi)\sin|k|\theta, & \text{for } k < 0 \end{cases} \tag{35}$$

where the real form only requires associated Legendre functions $P_m^k(x)$ of non-negative $|k|$. The corresponding eigenvalues for the negative Laplace operator on the surface of a sphere are thus $\lambda_{m,k} = m(m+1)$.

The normalization is chosen so that functions of two angles, $\theta \in [0, 2\pi)$ and $\phi \in [0, \pi]$, and two indices $m, k \in \mathbb{N}_0$ and $-m \leq k \leq m$, are orthonormal

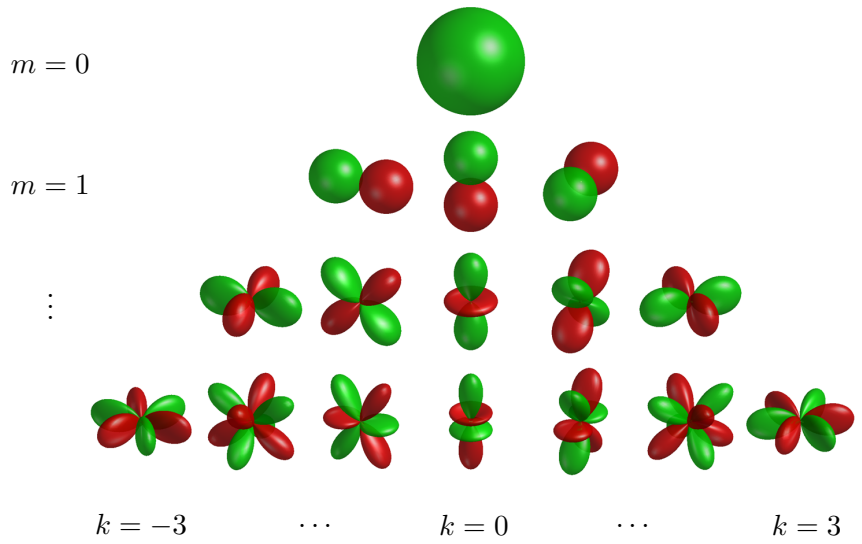


Figure 7: An illustrative visualization of the first few spherical harmonics Y_m^k. The figure visualizes $|Y_m^k(\theta, \phi)|$ in the case of the real spherical harmonics. The green regions are positive values and the red ones correspond to negative function values. The view angle is the same in each, but the scaling is not.

over the surface of a sphere. The orthogonality condition is satisfied by

$$\int_0^{2\pi} \int_0^{\pi} Y_m^k(\theta, \phi) Y_{m'}^{k'}(\theta, \phi) \sin \phi \, \mathrm{d}\phi \, \mathrm{d}\theta = \delta_{m,m'} \delta_{k,k'}, \tag{36}$$

where $\delta_{i,j}$ denotes the Kronecker delta function that is one if the indices i and j are equal and zero otherwise. The above equation holds for the real spherical harmonics, whereas in the case of the complex spherical harmonics, one of the spherical harmonics has to be the complex conjugate $\overline{Y}_m^k$.

Sometimes an extra $(-1)^k$ is included in the definition of the spherical harmonics. This does not actually affect their properties, but can be useful in certain applications in quantum mechanics in physics. The coefficient $(-1)^k$ is called the Condon–Shortley phase factor (Arfken and Weber, 2001). We ignore the Condon–Shortley phase in this study.

A visual demonstration of the first few spherical harmonic functions is shown in Figure 7. The plots have different interpretations depending on whether we consider the real or complex spherical harmonics. For complex-valued $Y_m^k(\theta, \phi)$ from Equation (34) the negative ks show the real part of the spherical harmonics $r = \mathrm{Re}\left[Y_m^k\right]$, and for non-negative k the plots are

of the imaginary part $r = \mathrm{Im}\left[Y_m^k\right]$.

An alternative interpretation of Figure 7 is available for the real spherical harmonics. Now the spherical figure is given by $r = |Y_m^k(\theta, \phi)|$. The green and red coloring denotes the sign of the value, red being negative and green positive. As can be seen, the spherical harmonics for positive and negative k are alike, but rotated by 90° with respect to the azimuthal coordinate θ.

We introduced spherical harmonics originally as a part of the solution of the Helmholtz equation in a sphere. However, the spherical harmonics can be used on their own. The Sturm–Liouville form of the Laplace's equation provides the completeness property, which leads to the fact that any function with sufficient continuity properties evaluated over the surface of a sphere can be expanded in a uniformly convergent series of spherical harmonics (Arfken and Weber, 2001). Such a Laplace's series can be given by

$$f(\theta, \phi) = \sum_m \sum_{k=-m}^{m} a_{m,k} Y_m^k(\theta, \phi), \tag{37}$$

where f is the function and $a_{m,k}$ are the representation in the new basis. if $f(\theta, \phi)$ is known, the coefficients can be evaluated by

$$a_{m,k} = \int_0^{2\pi} \int_0^{\pi} Y_m^k(\theta, \phi) f(\theta, \phi) \, \mathrm{d}\phi \, \mathrm{d}\theta, \tag{38}$$

for each (m, k).

Higher-Dimensional Polar Coordinates

The polar coordinate system in two-dimensional and the spherical coordinates in three-dimensional Euclidean space can be extended to a coordinate system in an n-dimensional Euclidean space, in which $\Omega \subset \mathbb{R}^n$. The coordinates consist of a radial coordinate, $r \in [0, R] \subset \mathbb{R}$, and $n - 1$ angular coordinates $\phi_1, \phi_2, \ldots, \phi_{n-1}$ where ϕ_{n-1} ranges over $[0, 2\pi)$ radians and the other angles range over $[0, \pi]$ radians. If x_i are the Cartesian coordinates,

we may compute $x_1, \ldots, x_n$ from $r, \phi_1, \ldots, \phi_{n-1}$ with:

$$\begin{aligned}
x_1 &= r\cos(\phi_1) \\
x_2 &= r\sin(\phi_1)\cos(\phi_2) \\
x_3 &= r\sin(\phi_1)\sin(\phi_2)\cos(\phi_3) \\
&\vdots \\
x_{n-1} &= r\sin(\phi_1)\cdots\sin(\phi_{n-2})\cos(\phi_{n-1}) \\
x_n &= r\sin(\phi_1)\cdots\sin(\phi_{n-2})\sin(\phi_{n-1}).
\end{aligned}$$

The polar and spherical coordinates in Sections 3.1.2 and 3.1.2 can be seen as special cases of the above for $n = 2$ and $n = 3$. The inverse transformation can be similarly given as:

$$\begin{aligned}
r &= \sqrt{{x_n}^2 + {x_{n-1}}^2 + \cdots + {x_2}^2 + {x_1}^2} \\
\phi_1 &= \operatorname{arccot}\frac{x_1}{\sqrt{{x_n}^2 + {x_{n-1}}^2 + \cdots + {x_2}^2}} \\
\phi_2 &= \operatorname{arccot}\frac{x_2}{\sqrt{{x_n}^2 + {x_{n-1}}^2 + \cdots + {x_3}^2}} \\
&\vdots \\
\phi_{n-2} &= \operatorname{arccot}\frac{x_{n-2}}{\sqrt{{x_n}^2 + {x_{n-1}}^2}} \\
\phi_{n-1} &= 2\operatorname{arccot}\frac{\sqrt{{x_n}^2 + {x_{n-1}}^2} + x_{n-1}}{x_n}
\end{aligned}$$

where if $x_k \neq 0$ for some k but all of $x_{k+1}, \ldots, x_n$ are zero then $\phi_k = 0$ when $x_k > 0$, and $\phi_k = \pi$ radians when $x_k < 0$. There are some special cases where the inverse transform is not unique; ϕ_k for any k will be ambiguous whenever all of $x_k, x_{k+1}, \ldots, x_n$ are zero; in this case ϕ_k may be chosen to be zero.

The Laplacian in n-dimensional spherical coordinates can be given as (see, e.g., Chavel, 1984)

$$\Delta = \frac{1}{r^{n-1}}\frac{\partial}{\partial r}\left(r^{n-1}\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\Delta_{S^{n-1}}, \tag{39}$$

where the angular part is called the Laplace–Beltrami operator that is given

recursively as

$$\Delta_{S^{n-1}} = \sin^{2-n}\phi\,\frac{\partial}{\partial\phi}\left(\sin^{n-2}\phi\,\frac{\partial}{\partial\phi}\right) + \sin^{-2}\phi\,\Delta_{S^{n-2}}, \tag{40}$$

where ϕ is an angular coordinate. Again, the polar and spherical Laplace operators can be seen as special cases of the general n-dimensional case. Eigenfunctions of the general case could be formulated by assuming separable solutions, but these are omitted here, as they are not of primary interest in this work.

3.2 Numerical Evaluation of Infinite-Dimensional Filtering

In Section 2.4 we saw how the infinite-dimensional Kalman filter could be seen as the optimal filtering solution of linear evolution equations written in terms of linear operators. This theoretical backdrop can now be combined with a more practical approach in this section. We consider the eigenfunction expansion of the linear operator and combine it with the infinite-dimensional framework. By truncating the expansion, we get a finite-dimensional approximative solution that can be evaluated.

Each of the domains considered in Section 3.1 — i.e. the n-cube and n-sphere — was combined with the Laplace operator subject to Diriclet boundary conditions. This resulted in each case in an eigenfunction equation of form $\nabla^2\psi_n(\mathbf{x}) = \lambda_n\psi_n(\mathbf{x})$, where $\psi_n(\mathbf{x})$ is an eigenfunction and λ_n the corresponding eigenvalue for each n and spatial coordinate $\mathbf{x} \in \Omega$. The eigenfunctions form an orthonormal basis in each of the cases, and we may use the Hilbert space methods considered at the very beginning of this study in Section 2.1 to make the computations feasible. Our solution $\mathbf{f}(\mathbf{x}, t)$ will be transformed to a new basis that is given by the eigendecomposition of the linear operator in Ω. This new basis decodes the spatial structure so that we are only left with $\tilde{\mathbf{f}}(t)$, a finite-dimensional approximation of $\mathbf{f}(\mathbf{x}, t)$.

3.2.1 Finite-Dimensional Approximation of Dynamics

Let us now consider an $s \times s$ -matrix of linear operators $\boldsymbol{\mathcal{F}}$ such that each linear operator can be given in terms of the Laplace operator, meaning the eigenfunctions $\psi_n(\mathbf{x})$ of ∇^2 are also eigenfunctions of each operator in $\boldsymbol{\mathcal{F}}$. Say, one of the operators is some operator $\mathcal{A}$. Then the corresponding

eigenvalues λ_n of $\mathcal{A}$ are solved such that $\mathcal{A}\,\psi_n(\mathbf{x}) = \lambda_n \psi_n(\mathbf{x})$ holds for all n. Following the derivation in the supplementary material of Särkkä and Hartikainen (2012), we go through the procedure here. A change of basis yields that the evolution is not subject to an infinite-dimensional Hilbert space but to a (truncated) finite-dimensional space defined by the orthonormal basis $\{\psi_1(\mathbf{x}), \psi_2(\mathbf{x}), \ldots, \psi_N(\mathbf{x})\}$. This transforms the stochastic evolution equation in Equation (16) to

$$\begin{aligned} \frac{\mathrm{d}\tilde{\mathbf{f}}(t)}{\mathrm{d}t} &= \mathbf{F}\,\tilde{\mathbf{f}}(t)\,\mathrm{d}t + \mathbf{L}\,\tilde{\mathbf{w}}(t) \\ \mathbf{y}_k &= \tilde{\mathbf{H}}_k\,\tilde{\mathbf{f}}(t) + \mathbf{r}_k, \end{aligned}$$

where $\tilde{\mathbf{f}}$ is an sN-dimensional state vector — s being the dimension of the operator matrix and N the number of eigenvalues in the eigenfunction expansions of all the linear operators in $\boldsymbol{\mathcal{F}}$. This means that we have expanded the process values $f_j(\mathbf{x}, t)$, each component $j = 1, 2, \ldots, s$ in $\mathbf{f}(\mathbf{x}, t)$, such that it can be given as

$$f_j(\mathbf{x}, t) = \sum_n \tilde{f}_{j,n}(t)\psi_n(\mathbf{x}).$$

The evolution matrix $\mathbf{F}$ can be formed simply by using the eigenfunction expansion of the $s \times s$ matrix of linear operators combined by the change of basis, which yields a block-diagonal matrix $\mathbf{F}$ of size $sN \times sN$, where the operators are replaced by their eigenvalue counterparts.

A formal series expansion of the noise term $w(x, t)$ can now be given as

$$w(\mathbf{x}, t) = \sum_n \tilde{w}_n(t)\psi_n(\mathbf{x}) \quad \text{and} \quad \tilde{w}_n(t) = \int_\Omega w(\mathbf{x}, t)\psi_n(\mathbf{x})\,\mathrm{d}\mathbf{x}. \tag{41}$$

The joint spectral density $\tilde{\mathbf{Q}}$ for the process noise can be derived as follows by taking the expectation

$$\begin{aligned} \mathrm{E}[\tilde{w}_n(t)\tilde{w}_m(s)] &= \mathrm{E}\left[\int_\Omega \int_\Omega w(\mathbf{x}, t)\psi_n(\mathbf{x})w(\mathbf{x}', s)\psi_m(\mathbf{x}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}'\right] \\ &= \int_\Omega \int_\Omega \psi_n(\mathbf{x})\mathrm{E}\left[w(\mathbf{x}, t)w(\mathbf{x}', s)\right]\psi_m(\mathbf{x}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}' \\ &= \int_\Omega \int_\Omega \psi_n(\mathbf{x})Q_c(\mathbf{x} - \mathbf{x}')\psi_m(\mathbf{x}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}'\,\delta(t - s), \end{aligned} \tag{42}$$

where $\delta(\cdot)$ denotes the Dirac delta function. This means that the spectral density can be given as

$$\tilde{\mathbf{Q}}_{nm} = \int_{\Omega}\int_{\Omega} \psi_n(\mathbf{x})\,\mathbf{L}\,Q_c(\mathbf{x}-\mathbf{x}')\,\mathbf{L}^{\mathsf{T}}\psi_m(\mathbf{x}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}' \tag{43}$$

The observation model matrix $\tilde{\mathbf{H}}_k \in \mathbb{R}^{d_k \times sN}$ evaluates the observations at each $\mathbf{x}_i^{\text{obs}}, i = 1, 2, \ldots, d_k$. This means $\tilde{\mathbf{H}}_k$ is a matrix of $(\psi_1(\mathbf{x}_i), \psi_2(\mathbf{x}_i), \ldots, \psi_N(\mathbf{x}_i)) \otimes \mathbf{h}$, where $\otimes$ denotes the Kronecker product and the vector $\mathbf{h}$ is a vector of ones and zeros defining which component to observe.

3.2.2 Time Discretization

Now we need to reconsider the time discretization step we presented for the infinite-dimensional evolution equation in Section 2.4.1. The operator exponential that defines the discrete-time dynamic evolution operator matrix was given as

$$\mathcal{U}(t) = \exp(\Delta t \boldsymbol{\mathcal{F}}).$$

Now that we have made the change of basis and replaced the operator matrix $\boldsymbol{\mathcal{F}}$ with the block-diagonal eigenfunction expansion counterpart $\mathbf{F}$, we can write the operator exponential function as a matrix exponential

$$\mathbf{A}(\Delta t) = \exp(\Delta \mathbf{F}) = \begin{bmatrix} \exp(\Delta t \mathbf{F}_1) & & & \\ & \exp(\Delta t \mathbf{F}_2) & & \\ & & \ddots & \\ & & & \exp(\Delta t \mathbf{F}_N) \end{bmatrix},$$

where the evaluation of the matrix exponential has been broken down to being evaluated separately for each N blocks in the block-diagonal matrix $\mathbf{F}$. Consequently $\mathbf{A}(\Delta t)$ is a square matrix of size $sN \times sN$.

The discrete process noise covariance matrix $\tilde{\mathbf{Q}}(t, s)$ is given by the integral $\tilde{\mathbf{Q}}(t,s) = \int_s^t \mathbf{A}(t-\tau)\,\mathbf{L}\,\mathbf{Q}_c\mathbf{L}^{\mathsf{T}}\,\mathbf{A}^{\mathsf{T}}(t-\tau)\,\mathrm{d}\tau$. In numerical implementation the *matrix fraction decomposition* (Grewal and Andrews, 2001) can be used for efficiency.

In other words, we have formed the discrete-time mild solution (in terms of $\mathbf{A}(t_{k+1}-t_k)$, $\tilde{\mathbf{Q}}(t_{k+1}, t_k)$ and $\tilde{\mathbf{H}}_k$) to the stochastic differential equation

which is a projection of the original SPDE of interest. Now the estimation can be accomplished by the Kalman filtering equations in Section 2.3.3 — Equations (13–15). This is now comparable to doing the infinite-dimensional estimation which is set up by the Equations (19–21).

3.2.3 Connection to GP Models

The formulation of infinite-dimensional Kalman filtering started by consideration of GP models in Section 2.2. This connection can now be restated in terms of the approximative solution. The covariance function of the model in terms of the new basis can be written as

$$\mathbf{C}_\mathbf{f}(t) = \mathrm{E}\left[\mathbf{f}(t)\mathbf{f}^\mathsf{T}(t+\tau)\right] = \begin{cases} \mathbf{C}_\infty \mathbf{A}^\mathsf{T}(\tau), & \text{for } \tau \geq 0 \\ \mathbf{A}(\tau)\,\mathbf{C}_\infty, & \text{for } \tau < 0, \end{cases} \tag{44}$$

where $\mathbf{C}_\infty$ is the solution to the Riccati equation as in Section 2.2.3. Let us also consider the change of basis that gives us the values of the actual state $f(\mathbf{x},t) = \sum_n \tilde{f}_n(t)\psi_n(\mathbf{x}) = \boldsymbol{\psi}^\mathsf{T}(\mathbf{x})\,\mathbf{H}\,\tilde{\mathbf{f}}(t)$, where $\boldsymbol{\psi}(\mathbf{x}) = (\psi_1(\mathbf{x}), \psi_2(\mathbf{x}), \ldots, \psi_N(\mathbf{x}))$ and $\mathbf{H}$ is a matrix with elements picking out the desired components.

The covariance with respect to both spatial and temporal variables is thus

$$\begin{aligned} \mathrm{E}\left[f(\mathbf{x},t)f(\mathbf{x}+\mathbf{x}',t+\tau)\right] &= \mathrm{E}\left[\boldsymbol{\psi}^\mathsf{T}(\mathbf{x})\,\mathbf{H}\,\tilde{\mathbf{f}}(t)\,\boldsymbol{\psi}^\mathsf{T}(\mathbf{x}+\mathbf{x}')\,\mathbf{H}\,\tilde{\mathbf{f}}(t+\tau)\right] \\ &= \boldsymbol{\psi}^\mathsf{T}(\mathbf{x})\,\mathbf{H}\,\mathrm{E}\left[\tilde{\mathbf{f}}(t)\,\boldsymbol{\psi}^\mathsf{T}(\mathbf{x}+\mathbf{x}')\right]\mathbf{H}\,\tilde{\mathbf{f}}(t+\tau) \\ &= \boldsymbol{\psi}^\mathsf{T}(\mathbf{x})\,\mathbf{H}\,\mathbf{C}_\mathbf{f}(\tau)\,\mathbf{H}^\mathsf{T}\boldsymbol{\psi}(\mathbf{x}+\mathbf{x}'). \end{aligned} \tag{45}$$

Given this, we can approximate the covariance function $C_f(\mathbf{x},t) \approx \boldsymbol{\psi}^\mathsf{T}(\mathbf{0})\,\mathbf{H}\,\mathbf{C}_\mathbf{f}(t)\,\mathbf{H}^\mathsf{T}\boldsymbol{\psi}(\mathbf{x})$ (see Särkkä and Hartikainen, 2012, for details).

4 Case Studies

In this section we present a spatio-temporal resonator model that can be formulated in terms of a continuous-time linear infinite-dimensional state space system that fits under the infinite-dimensional Kalman filtering framework. We then apply the model to three real-world datasets in order to demonstrate the methods from the previous sections.

In the first demonstration we use hourly temperature measurements around the globe and model the short-term variation of temperature by a spatio-temporal resonator model with constant frequencies on the two-dimensional surface of a sphere. In the second and third demonstration we model cardiac- and respiration-induced periodic noise in brain data. We use the same spatio-temporal resonator model, but in these two cases with time-varying frequencies. The model is constructed in a two-dimensional polar slice and in spherical polar coordinates for the whole head.

4.1 Spatio-Temporal Resonator Model

The recent article 'Dynamic Retrospective Filtering of Physiological Noise in BOLD fMRI: DRIFTER' by Särkkä *et al.* presented a method for eliminating periodic noise induced by respiration and cardiac activity. The formulation of DRIFTER is based on presenting the dynamic phenomena as a superposition of several resonators with known angular velocities ω_j (i.e. frequencies), but unknown phases and amplitudes. These were modelled as spatially independent realisations of stochastic processes. The sum $\sum_{j=1}^{N} f_j(\mathbf{x}, t)$ of the oscillatory components $f_j(\mathbf{x}, t)$ can be defined through separate state space models. In Särkkä *et al.* (2012a) this was presented as a partial differential equation

$$\frac{\partial^2 f_j(\mathbf{x}, t)}{\partial t^2} + \gamma_j \frac{\partial f_j(\mathbf{x}, t)}{\partial t} + \omega_j^2 f_j(\mathbf{x}, t) = \xi_j(\mathbf{x}, t), \tag{46}$$

where $\xi_j(\mathbf{x}, t)$ is spatially and temporally white noise. The above formulation also features a damping factor γ_j that was assumed zero in the article. A straightforward way to extend this formulation to also account for spatial structure, is to assume that the local derivative depends not only on time, but also on surrounding data through some spatial linear operator. By

including such arbitrary linear operators affecting both the oscillation and damping, we get

$$\frac{\partial^2 f_j(\mathbf{x},t)}{\partial t^2} + \mathcal{A}_j \frac{\partial f_j(\mathbf{x},t)}{\partial t} + \mathcal{B}_j f_j(\mathbf{x},t) = \xi_j(\mathbf{x},t). \tag{47}$$

This model features three types of spatial dependency. Choosing operators $\mathcal{A}_j$ and $\mathcal{B}_j$ suitably defines spatial coupling through the first and second temporal derivative. We can also assume some spatial and temporal structure in the process noise term $\xi_j(\mathbf{x},t)$ through a correlation structure

$$C_j(\mathbf{x},\mathbf{x}') = \mathrm{E}[\xi_j(\mathbf{x},t)\xi_j(\mathbf{x}',t')] = C_{\xi,j}(\mathbf{x},\mathbf{x}')\,\delta(t-t'). \tag{48}$$

4.1.1 Choosing Spatial Operators

If we assume the operators $\mathcal{A}_j$ and $\mathcal{B}_j$ translation and time invariant, we can calculate the corresponding transfer functions $A_j(i\boldsymbol{\nu}_x)$ and $B_j(i\boldsymbol{\nu}_x)$. Taking both spatial and temporal Fourier transforms of Equation (47) leaves us with

$$(i\nu_t)^2 F_j(i\nu_t, i\boldsymbol{\nu}_x) + (i\nu_t) A_j(i\boldsymbol{\nu}_x) F_j(i\nu_t, i\boldsymbol{\nu}_x) + B_j(i\boldsymbol{\nu}_x) F_j(i\nu_t, i\boldsymbol{\nu}_x) = \Xi_j(i\nu_t, i\boldsymbol{\nu}_x).$$

By solving F_j from above we get

$$F_j(i\nu_t, i\boldsymbol{\nu}_x) = \frac{\Xi_j(i\nu_t, i\boldsymbol{\nu}_x)}{(i\nu_t)^2 + (i\nu_t)A_j(i\boldsymbol{\nu}_x) + B_j(i\boldsymbol{\nu}_x)},$$

which corresponds to the spectral density

$$S_j(i\nu_t, i\boldsymbol{\nu}_x) = \frac{Q_j(\boldsymbol{\nu}_x)}{[(i\nu_t)^2 + (i\nu_t)A_j(i\boldsymbol{\nu}_x) + B_j(i\boldsymbol{\nu}_x)][(i\nu_t)^2 + (i\nu_t)A_j(i\boldsymbol{\nu}_x) + B_j(i\boldsymbol{\nu}_x)]^*},$$

where $Q_j(\boldsymbol{\nu}_x) = |\Xi_j(i\nu_t, i\boldsymbol{\nu}_x)|^2$ is the spectral density of ξ_j. If we assume that the operators $\mathcal{A}_j$ and $\mathcal{B}_j$ are formally hermitian (bounded and symmetric), the identities $A_j(i\boldsymbol{\nu}_x) = A_j(-i\boldsymbol{\nu}_x)$ and $B_j(i\boldsymbol{\nu}_x) = B_j(-i\boldsymbol{\nu}_x)$ hold, which simplifies the spectral density to

$$S_j(i\nu_t, i\boldsymbol{\nu}_x) = \frac{Q_j(\boldsymbol{\nu}_x)}{\left[\nu_t^2 - B_j(i\boldsymbol{\nu}_x)\right]^2 + \nu_t^2 A_j^2(i\boldsymbol{\nu}_x)}.$$

The divisor derivative zeros of the system suggest that the system has a

temporal resonance of $\nu_t^2 = B_j(i\boldsymbol{\nu}_x) - A_j^2(i\boldsymbol{\nu}_x)/2$. We include the temporal oscillation with angular velocity of ω_j by setting $B_j(i\boldsymbol{\nu}_x) = A_j^2(i\boldsymbol{\nu}_x)/2 + \omega_j^2$. This gives us a spectral density of form

$$S_j(i\nu_t, i\boldsymbol{\nu}_x) = \frac{Q_j(\boldsymbol{\nu}_x)}{(\nu_t^2 - A_j^2(i\boldsymbol{\nu}_x)/2 - \omega_j^2)^2 + \nu_t^2 A_j^2(i\boldsymbol{\nu}_x)}.$$

According to Bochner's theorem (see Section 2.2.2) every positive definite function is the Fourier transform of a positive finite Borel measure. This requires the spectral density to be positive everywhere (in order to be a valid Fourier transform of a covariance function). This condition is fulfilled if $Q_j(\boldsymbol{\nu}_x)$ is a positive function (i.e. a valid spectral density). To ensure the causality and stability of the system we have to choose $A_j(i\boldsymbol{\nu}_r)$ such that it is a positive function, which corresponds to the operator $\mathcal{A}_j$ being positive (semi)definite. We as well choose the operator $\mathcal{B}_j$ to be positive, which gives us the condition $A_j^2(i\boldsymbol{\nu}_x)/2 + \omega_j^2 \geq 0$. This holds, if A_j is real and positive. Zero values in the spectrum corresponds to infinite peaks. However, this does not seem a problem, because if both the operators are zero the model falls back to a spatially independent model, where the only spatial structure comes from the process noise term $\xi(\mathbf{x}, t)$.

To actually make the model useful, some choices have to be made. The operator $\mathcal{A}_j$ has to be positive semidefinite. Examples of such operators are the identity operator $\mathcal{I}$ and the negative Laplacian $-\Delta = -\nabla^2$. We therefore consider the following operator structure

$$\begin{aligned} \mathcal{A}_j &= \gamma_j \mathcal{I} - \chi_j \nabla^2 \\ \mathcal{B}_j &= \frac{\gamma_j^2}{2} - \gamma_j \chi_j \nabla^2 + \frac{\chi_j^2}{2}\nabla^4 + \omega_j^2 \\ &= \frac{1}{2}(\gamma_j - \chi_j \nabla^2)^2 + \omega_j^2, \end{aligned} \tag{49}$$

where $\gamma_j, \chi_j \geq 0$ are some non-negative constants and ∇^4 is the so-called biharmonic operator. These choices define the following resonator model

$$\begin{aligned} \frac{\partial^2 f_j(\mathbf{x},t)}{\partial t^2} + \gamma_j \frac{\partial f_j(\mathbf{x},t)}{\partial t} - \chi_j \nabla^2 \frac{\partial f_j(\mathbf{x},t)}{\partial t} + \frac{\gamma_j^2}{2} f_j(\mathbf{x},t) \\ - \gamma_j \chi_j \nabla^2 f_j(\mathbf{x},t) + \frac{\chi_j^2}{2}\nabla^4 f_j(\mathbf{x},t) + \omega_j^2 f_j(\mathbf{x},t) = \xi_j(\mathbf{x},t). \end{aligned} \tag{50}$$

We have to decide a covariance function for $\xi_j(\mathbf{x}, t)$ which can be virtually any spatial stationary covariance function $C_{\xi,j}(r)$ for which

$$\mathrm{E}[\xi_j(\mathbf{x}, t)\xi_j(\mathbf{x}', t')] = C_{\xi,j}\,\delta(t - t').$$

The covariance functions of the Matérn class, that were presented earlier in Section 2.2.2, are useful in this context and will be used in the demonstrations.

4.1.2 Modeling Spatio-Temporal Data

Combining all the components in the model gives us the solution as a superposition of all the oscillator components $f(\mathbf{x}, t) = \sum_{j=1}^{N} f_j(\mathbf{x}, t)$. The oscillator component $f_j(\mathbf{x}, t)$ is defined by a stochastic partial differential equations with Dirichlet boundary conditions

$$\begin{aligned} \frac{\partial^2 f_j(\mathbf{x}, t)}{\partial t^2} + \mathcal{A}_j \frac{\partial f_j(\mathbf{x}, t)}{\partial t} + \mathcal{B}_j f_j(\mathbf{x}, t) &= \xi_j(\mathbf{x}, t) && (\mathbf{x}, t) \in \mathbb{R}_+ \times \Omega \\ f_j(\mathbf{x}, t) &= 0 && (\mathbf{x}, t) \in \mathbb{R}_+ \times \partial\Omega \end{aligned}$$

for all $j = 1, 2, \ldots, N$. We define the state of the system as a combination of the periodic oscillating fields and their first temporal derivatives

$$\mathbf{f}(\mathbf{x}, t) = \left[f_1(\mathbf{x}, t) \quad \tfrac{\partial}{\partial t} f_1(\mathbf{x}, t) \quad \ldots \quad f_N(\mathbf{x}, t) \quad \tfrac{\partial}{\partial t} f_N(\mathbf{x}, t) \right]^\mathsf{T}.$$

This leads us to the linear state space model that can be written in the following form

$$\begin{aligned} \frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial t} &= \boldsymbol{\mathcal{F}}\,\mathbf{f}(\mathbf{x}, t) + \mathbf{L}\,\boldsymbol{\xi}(\mathbf{x}, t) \\ \mathbf{y}_k &= \boldsymbol{\mathcal{H}}_k \mathbf{f}(\mathbf{x}, t_k) + \mathbf{r}_k, \end{aligned} \tag{51}$$

where $\boldsymbol{\mathcal{F}}$ is a block-diagonal matrix of linear operators such that each i blocks consist of a 2×2 matrix of linear operators and $\mathbf{L}$ is a block-column matrix such that

$$\boldsymbol{\mathcal{F}}_j = \begin{bmatrix} 0 & \mathcal{I} \\ -\mathcal{B}_j & -\mathcal{A}_j \end{bmatrix} \quad \text{and} \quad \mathbf{L}_j = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

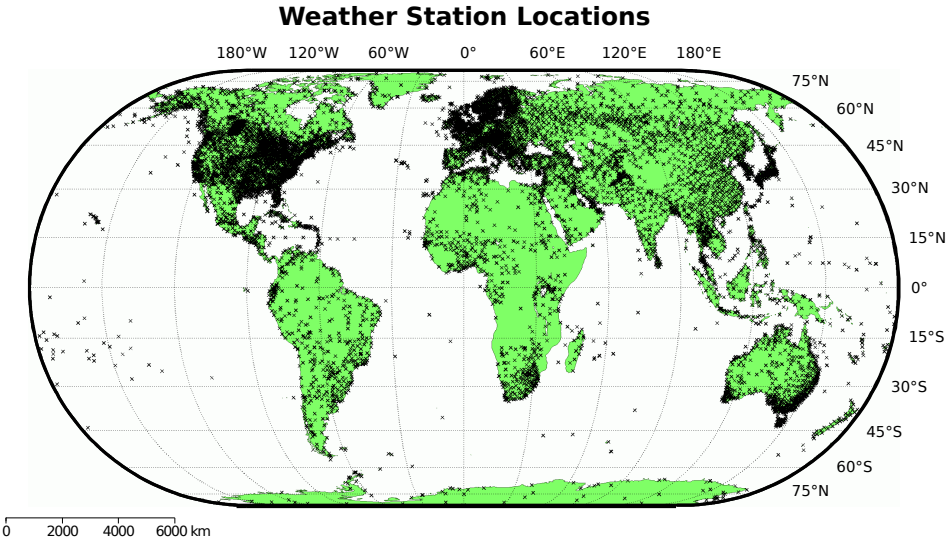


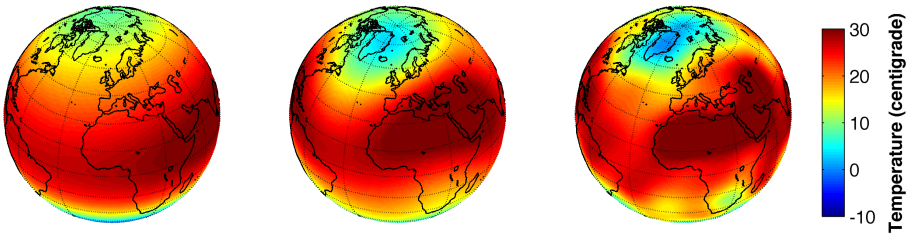
Figure 8: Spatial locations of the 11 344 weather observation stations all over the globe.

We define the measurement model by defining a linear operator $\boldsymbol{\mathcal{H}}_k$ through which the model is observed at discrete time steps t_k and known locations $\mathbf{x}_i^{\text{obs}} \in \Omega, i = 1, 2, \ldots, d_k$. The measurement noise term $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ in Equation (51) is a Gaussian random variable of dimension d_k. On step k the observed values are $\mathbf{y}_k \in \mathbb{R}^{d_k}$. For notational convenience, we have omitted the possibility of $\boldsymbol{\mathcal{F}}$ depending on time. However, this will be the case in the two latter demonstrations, where the oscillation frequencies $\omega_j(t)$ change over time.

The presented spatio-temporal resonator model can be solved by applying the numerical evaluation techniques presented in Section 3.2. We apply the eigenfunction expansion theorem to the operators $\mathcal{A}_j$ and $\mathcal{B}_j$, do the time discretization step, and apply the finite-dimensional Kalman filter to the resulting state space model. The time variant case is done by considering the dynamics piece-wise constant, and doing the eigenfunction decomposition from scratch on each time step.

4.2 Spatio-Temporal Oscillation of Temperatures

We present an illustrative example of a two-dimensional application of the oscillator model by applying it to temperature data. Different types of Kalman filtering approaches exist for weather prediction, and here this idea



$M = 3$ $M = 6$ $M = 12$

Figure 9: This illustration shows how the number of basis functions affects the smoothness of the temperature surface. The temperatures correspond to the day mean of July 8, 2011.

is presented only with interest in demonstrating the oscillator model on the surface of a sphere. We use hourly observations of temperature readings in centigrades that were collected worldwide by *National Environmental Satellite, Data, and Information Services* (NESDIS). The dataset is available for download through U.S. National Climatic Data Center[1].

We consider a subset of the data that consists of hourly temperature measurements for one month (July, 2011) resulting in a time series of 745 temporal points. The temperatures were recorded at 11 344 different spatial locations, of which the longitudinal and latitudinal coordinates are known and assumed exact. The locations of the stations are shown in Figure 8. However, not all stations provide hourly measurements and there are altogether only 5 637 501 measurements, meaning a proportion of 33.3 % of missing values. About half of the time series are complete with 24 measurements a day, and the second most common observation rate is three times per day.

We use the oscillator model that was introduced in the previous section to model the daily variation of the temperature. We use a three-component setup, where the first component, a bias term, accounts for the slow drifting of the mean temperature, and the other two are oscillatory components for the daily variation of temperature. The first oscillator oscillates at the constant base frequency of $f_2 = 1/\text{day}$, and the other is the first harmonic ($f_3 = 2f_2$). The bias term is constructed as an oscillator with zero frequency, which can be seen as a spatio-temporal Wiener velocity model.

[1] `http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd` (accessed December 12, 2011)

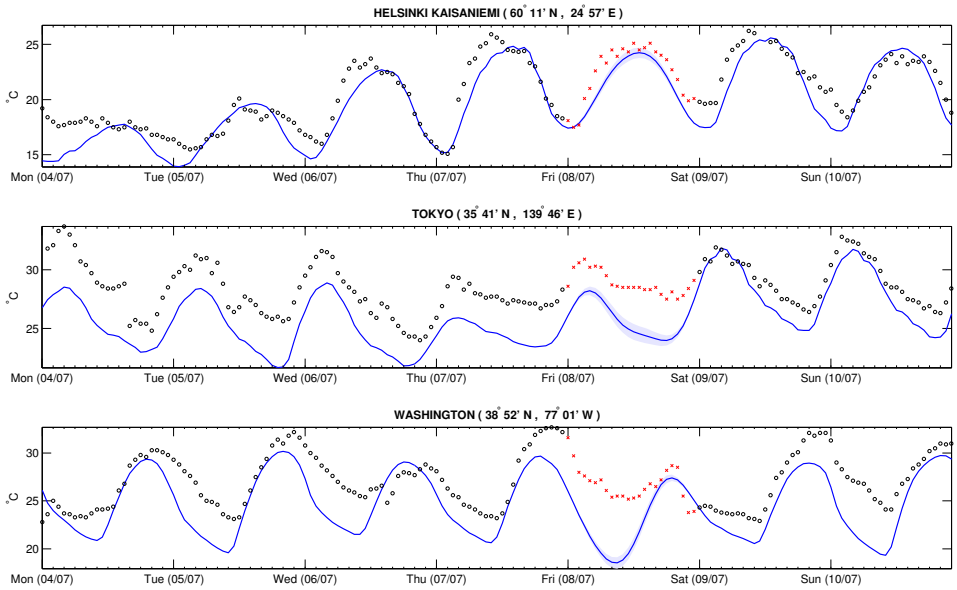


Figure 10: Three one-week time series of temperatures in Helsinki, Tokyo and Washington DC. The estimate mean is shown in blue, and the shaded patch shows the 95 % confidence interval. The measurements during the 8th of July were excluded from the estimation and shown here by crosses.

This test setup is subject to many simplifications and assumptions that affect the results; the surface of the earth is actually not a symmetric sphere and we disregard the evident fact that the fluctuation covariance structure is not stationary. Furthermore, in this demonstration we are not concerned with finding the best possible parameters for the model. Therefore we just choose appropriate values for the parameters. For each of the components in the model we used the same number of eigenfunctions in the eigenfunction expansion, $M = 24$. Figure 9 shows how the number of eigenfunctions in the expansion affects the spatial resolution for $M = 3, 6$, and 12. The damping constants in the model were fixed to $\gamma_1 = [M(M+1)]^{-1} = .0017$ (from the eigenvalues) and $\chi_1 = 1$ for the bias and $\gamma_{2,3} = .0017$ and $\chi_{2,3} = 1$ for the oscillators.

We chose to use the squared exponential covariance function to model the spatial correlation of the dynamic noise term. The distances were calculated as great circle distances on the surface of the sphere in degrees. Exploratory analysis on the data suggested that the scale parameter would be $l = 1°$ (approx. 111 km). The magnitude was fixed to $\sigma = 0.3$ for the bias and

$\sigma = 0.003$ for the oscillators.

We present the results for a test setup where we exclude all measurements for the 8th of July. The goal is to estimate these measurements using the remaining data. The results show that the root-mean-square error for all temperature observations during July 8th was 2.81 °C (mean error 2.00 °C). The results for three observations stations in Helsinki, Tokyo and Washington DC are shown in Figure 10. The uncertainty is visualized by blue shading, but the envelope widths are dependent on the choice of parameters.

As can be seen in Figure 10 the results are not exactly following the observed values. The smoothness of the estimated surface makes it difficult to account for local jitter. Making the model actually useful for weather prediction and analysis would require the parameters to be tuned. Additionally, altitude, terrain and other properties affecting the local climate could be included. These could perhaps be encoded partly through a non-stationary covariance function. However, even with this simplified weather model, the results indicate that a spatio-temporal resonator model can provide an easy-to-understand and very efficient way to model the hourly weather around the globe.

4.3 Spatio-Temporal Modeling of One Slice of fMRI Data

The next two demonstrations of spatio-temporal modeling using infinite-dimensional Kalman filtering are both concerned with separation of oscillatory noise from functional brain imaging data. The rapidly evolving methodology of functional magnetic resonance imaging (fMRI, Ogawa *et al.*, 1990; Belliveau *et al.*, 1991; Kwong *et al.*, 1992) techniques calls for means to restrain noise levels and improve sampling rates. In Särkkä *et al.* (2012a) it was shown that eliminating oscillating physiological noise components could be done by using a resonator model setup combined with Kalman filtering. This method was named DRIFTER.

DRIFTER relies on modeling each voxel signal as an independent time series, which leads to an effective implementation and little risk of the model based approach interfering with the data too much. The problem is that the sampling rate of each voxel needs to be high in order to not hit the Nyquist frequency of the signal. In most fMRI studies the main interest is in the spatial coverage of the data, which leads to slow temporal sampling rates.

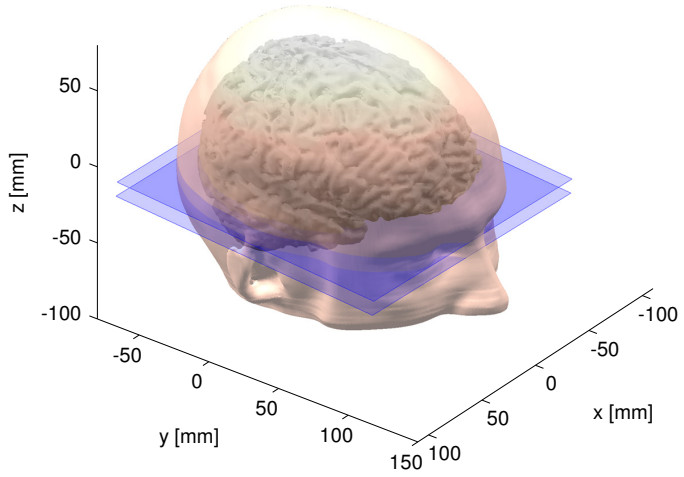


Figure 11: The head profile and brain image segmented from the anatomical MRI images and two overlays showing the orientation of the fast-sampled two fMRI slices.

The idea in DRIFTER was to use fast-sampled reference signals to find the time-dependent frequencies of heart beats and respiration cycles. This information provides a remedy to the separation of physiological noises in fMRI data. However, as was presented by Särkkä *et al.* (2012b), the resulting cardiac- and respiration induced noise estimates also feature clear spatial structure. This suggests that the estimation and separation of physiological noises could be included by extending the DRIFTER method to spatio-temporal modeling.

We consider two runs of empirical *functional magnetic resonance imaging* (fMRI) data. The first of these is fast-sampled data of one slice which we use to demonstrate the spatio-temporal resonator model in two-dimensional polar-coordinates. The second run of data comprises 29 slices (the whole brain) and is used in demonstrating the method in spherical three-dimensional coordinates. The runs are the same as the ones labelled '11' and '12' in Särkkä *et al.* (2012a).

This fMRI data, together with anatomical images, for one volunteer was obtained with a 3.0 T scanner (Signa HDxt; General Electric) located at the *Advanced Magnetic Imaging Centre* (AMI) of Aalto University School of Science using an 8-channel (MRI Devices Corporation) receive-only head coil. For the functional imaging, the major parameters were two different

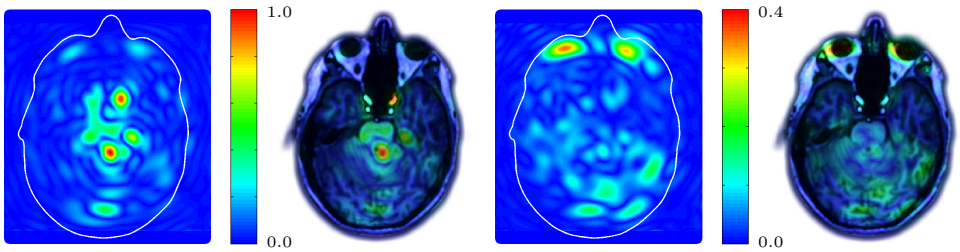
repetition times (TR) 100 ms and 1800 ms; echo time (TE) 20 ms; flip angle (FA) 60°; field-of-view (FOV) 20 cm; and matrix size 64×64. The visual stimuli were presented with a 3-micromirror data projector (Christie X3; Christie Digital Systems) using the Presentation software (Neurobehavioral Systems).

The stimuli consisted of 50 achromatic photographs of familiar objects presented in the center of the visual field of the volunteer at a distance of 37 cm from the eyes. The stimulus condition was contrasted with fixation alone. The runs, both roughly 240 seconds in length, comprised of similar blocks (∼15 s of stimulus-on and ∼7 s of stimulus-off). The heart and respiratory signals were recorded time-locked to the fMRI data during the runs. The measurements were done as part of AMI Centre's local technical methods development research and conformed to the guidelines of the Declaration of Helsinki. The research was approved by the ethical committee in the Hospital District of Helsinki and Uusimaa.

Figure 11 shows the orientation of two slices of fast-sampled fMRI data overlaid atop of head and brain profiles that were segmented from the anatomical MRI data. The orientation of the slices was chosen so that they cut through the high-order object sensitive visual cortex, where the stimulus response is expected evident. In this study we put little interest in the stimulus, the main interest is in demonstrating that the physiological oscillations can be modelled using spatio-temporal state space models.

The oscillation frequency of the physiological noise components are not exactly periodic, but quasi-periodic, which means that the frequencies change over time. We use external reference signals with the *interacting multiple model* (IMM) approach presented in Särkkä *et al.* (2012a) to estimate the frequency time series of the heart beats and respiration cycles. The cardiac frequency alternates between 65–80 bpm and the respiratory frequency between 12–24 cycles per minute.

We use only one slice of fast-sampled fMRI data. The sampling interval (TR) is 0.1 s and we observe the whole 64×64 matrix at each time step. We consider a two-dimensional disk from Section 3.1.2 with radius $L \approx 155$ mm. The observed values are assumed to be measured in a square that is centered in the disk. The disk boundary was extended to 110 % of the observed point with largest radius. This ensures that the boundary effects will not practically affect the estimation result.



(a) Cardiac signal amplitude **(b)** Respiratory signal amplitude

Figure 12: Mean amplitude maps for both the cardiac- and respiration-induced noise components. In both cases the left side figure shows the mean amplitude field with an anatomical contour and the right side figure the same field overlaid atop of the corresponding anatomical slice.

We use a spatio-temporal resonator model with three components. The first is the slowly moving cleaned brain (blood-oxygen-level-dependent, BOLD) signal — that also includes scanner drift and other slow phenomena — which is modeled using a spatio-temporal Wiener velocity model. This is done, as in the temperature example in the previous section, by including an oscillator model with zero frequency. The two remaining components are space–time resonators oscillating at the time-dependent cardiac and respiratory frequencies, respectively. Here we only include resonators for the base frequencies, and more complex signals could be accounted for by including harmonics. For all three components, we use an eigenfunction decomposition with $N = 12$ and $M = 12$ (refer to Section 3.1.2 for details).

By studying estimation results from DRIFTER we come up with parameters for the spatial process noise covariance functions. We use the squared-exponential covariance function. In the slow moving cleaned BOLD signal the parameters were $\sigma_1 = 10^{-7}$ and $l_1 = 2\,\text{mm}$, and for both the oscillators $\sigma_{2,3} = 10^{-8}$ and $l_{2,3} = 2\,\text{mm}$. The units for magnitude parameter σ_j are arbitrary, as the scaling of the fMRI signal affects the value. However, the length scale parameter l_j has a clear interpretation of correlation length scale (given in millimeters here). We assume no coupling between the derivatives (damping) by putting $\gamma_j = \chi_j = 0$ for all three j.

The estimation results in four components — the cleaned BOLD, cardiac- and respiration-induced noises, and a measurement noise estimate — all of which are four-dimensional. To assess the estimation outcome we therefore calculate and show the mean amplitudes with respect to time of the spatio-

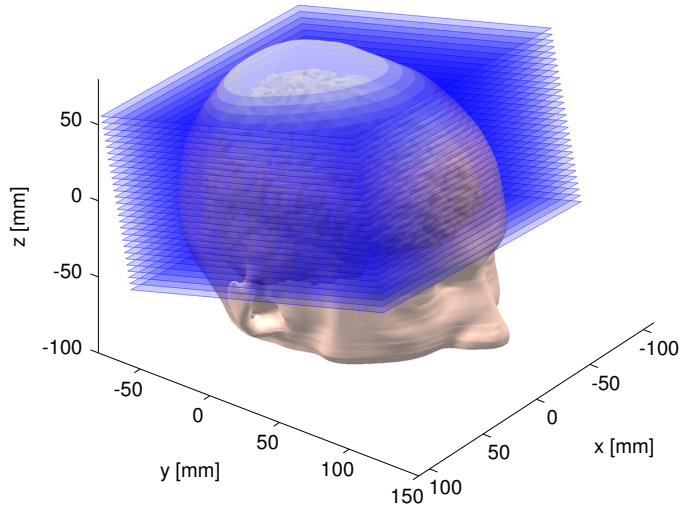


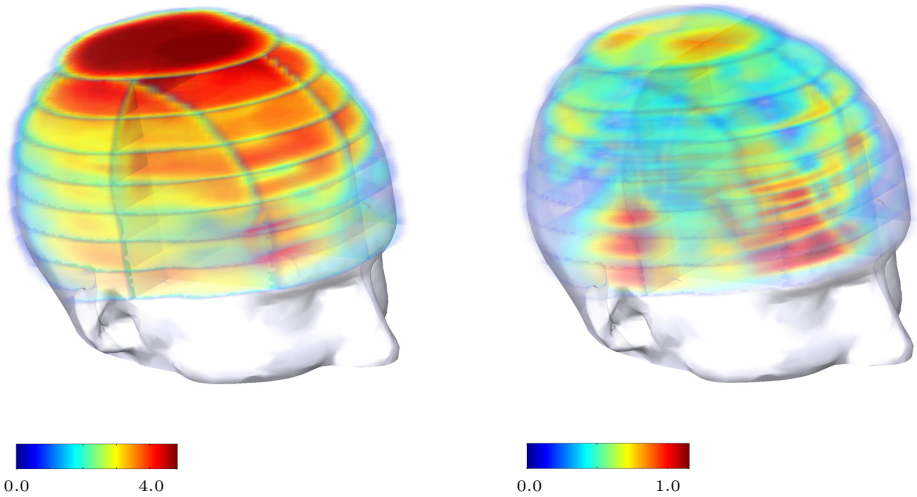
Figure 13: The head profile and brain image segmented from the anatomical MRI images and 29 overlays showing the orientation of the whole-brain fMRI slices.

temporal oscillator fields. Figure 12 shows mean amplitude maps for both the heart beat and respiration related signals in the fMRI data. Both the maps are presented as such and overlaid on an anatomical slice image. The scale has been normalized to one.

The amplitude maps shows virtually no contribution of the physiological noises outside the head, which is clearly desired in this case. Furthermore, the results in Figure 12 are very similar to the ones presented in Särkkä *et al.* (2012b), where the oscillators were treated as spatially independent and only the final results were spatially smoothed. This suggest that the method is able to capture the space–time structure of the oscillations.

4.4 Spatio-Temporal Modeling of the Whole Head

In the previous section we demonstrated that the spatio-temporal model shows prominent ways in modelling periodic noise structures in the brain. However, the perusal was limited to one slice only, and therefore the spatial effects that could be accounted for were restricted to two dimensions. In this section we use the second run of fMRI data that was presented in the previous section. The data consists of 29 slices with matrix size 64×64 that were acquired with TR $= 1800$ ms. The slice orientations are visualized in Figure 13.



(a) Cardiac signal amplitude **(b)** Respiratory signal amplitude

Figure 14: Three-dimensional amplitude density maps for both respiration- and cardiac-induced noise. The viewport is the same as in the previous figures and the amplitudes are shown on a logarithmic scale.

Accurate modeling of physiological oscillations in fMRI with slow TR is difficult. As discussed by Särkkä *et al.* (2012a), the sampling theorem by Nyquist and Shannon (see, e.g., Oppenheim *et al.*, 1999) states that reconstructing a signal is only possible if the sampling frequency is at least twice the fundamental frequency of the resonator. For a typical 72 bpm cardiac signal this would be roughly 400 ms. The theorem is actually more of a limitation only in the frequency estimation stage, but slow TRs cause aliasing that makes it difficult to distinct between several oscillatory components in the fMRI data.

A remedy to the problems related to the slow sampling rate can be found in the spatial correlation between voxels. Even though the whole brain is observed only every 1.8 s, the consecutive slices are observed with $\Delta t = 1.800\,\mathrm{s}/29 \approx 0.062\,\mathrm{s}$. Thus, we can apply the infinite-dimensional filtering approach to identify the spatio-temporal oscillators from the fMRI data. The slice observations are interleaved so that odd slices are observed first and thereafter the even-numbered.

Similarly as in the previous section, we use a spatio-temporal resonator model with three components. The cardiac and respiratory frequency trajectories are again estimated using the IMM approach (heart beat rate 64–

79 bpm, respiration cycles 10–21 cpm). The data has now three spatial dimensions, so we use spherical polar coordinates, with the boundary radius $L \approx 165\,\text{mm}$. The boundary is extended by adding 10 % the farthest data point from origin. For all three components, we use an eigenfunction decomposition with $N = 8$ and $M = 8$ (refer to Section 3.1.2 for details), which results in a total of $N(M+1)^2 = 648$ eigenfunctions in each expansion. For the process noise we use the squared exponential covariance function. In the slow moving cleaned BOLD signal the parameters were $\sigma_1 = 10^{-7}$ and $l_1 = 4\,\text{mm}$, and for both the oscillators $\sigma_{2,3} = 10^{-8}$ and $l_{2,3} = 4\,\text{mm}$. Again, we assume no coupling between the derivatives by putting $\gamma_j = \chi_j = 0$ for all three j.

In the two-dimensional example we studied the mean amplitudes of the physiological noise signals by visualizing them in the slice. Figure 14 shows similar visualization, but now the amplitudes constitute a three-dimensional amplitude field. The field is actually visualized by slice planes cutting through the density, which causes some bulkiness. In general, the results seem to match the one-slice example in the previous section, but further research is still needed to confirm the accuracy of the method.

5 Discussion and Conclusions

In this thesis infinite-dimensional Kalman filtering was presented through Gaussian process regression and Hilbert space valued stochastic processes. This resulted in infinite-dimensional state space models. These linear operator equations were applied to the infinite-dimensional Kalman filtering theory in the same manner as linear matrix equations are applied in the traditional Kalman filtering scheme. This resulted in an elegant infinite-dimensional formulation of the Kalman filter, even though the derivation was not entirely rigorous.

However, the main interest in this thesis was to use the infinite-dimensional Kalman filtering formulation in practice. We used the truncated eigenfunction expansion of the Laplace operator to form a finite-dimensional basis over the spatial domain, which enabled us to revert to the traditional Kalman filtering scheme. The eigenfunction expansions of the Laplace operator in n-dimensional spherical and Cartesian coordinates were presented.

The final step was to bring the methods to an application level, which was done by deriving a spatio-temporal resonator model and applying it to real-world datasets. The three case studies showed how the space–time structure of temperature variation on the earth's surface and physiological noise in brain data could be modeled. The case studies acted only as proof of concept examples, and many interesting questions remain unanswered in each of the applications.

For example, the results in the temperature example suggested that the lack of short-scale variability was a problem. The basis function approach tend to make the model spatially smooth, a problem which has been tackled in many ways under the GP regression scheme before (see, e.g., Rasmussen and Williams, 2006; Vanhatalo and Vehtari, 2008). Several methodological extensions could be considered as well, such as the possibility of including non-stationary covariance functions to the process noise model. Efficient parameter estimation under the infinite-dimensional Kalman filtering perspective also remains an open question, and is of great interest even tough it was outside the scope of this study.

The spatio-temporal resonator model has prospects. The two short demonstrations showed promising results of the identification of physiolog-

ical noise components in fMRI data. In brain imaging applications, the method is perhaps overly complex for physiological noise elimination alone, but it might be beneficial when combined with inversion based fMRI methods (Lin *et al.*, 2006, 2008), or further extensions of the DRIFTER method as discussed in Särkkä *et al.* (2012a).

We have showed that the rather complex infinite-dimensional Kalman filtering framework provides an efficient tool for future applications, and that the methodology calls for further improvements and extensions to meet up with the need for spatio-temporal modeling tools.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth Dover printing, tenth GPO edition.

Alpaydin, E. (2010). *Introduction to Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, second edition.

Arfken, G. B. and Weber, H. J. (2001). *Mathematical Methods for Physicists*. Academic press New York, fifth edition.

Bar-Shalom, Y., Li, X.-R., and Kirubarajan, T. (2001). *Estimation with Applications to Tracking and Navigation*. Wiley Interscience.

Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., and Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716.

Chavel, I. (1984). *Eigenvalues in Riemannian Geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press.

Chow, P. (2007). *Stochastic Partial Differential Equations*, volume 11 of *Chapman & Hall/CRC applied mathematics and nonlinear science series*. Chapman & Hall/CRC Press.

Christakos, G. (2005). *Random Field Models in Earth Sciences*. Dover Books on Earth Sciences. Dover Publications, corrected Dover edition. Originally published by Academic Press in 1992.

Cressie, N. and Wikle, C. K. (2002). Space-time Kalman filter. In El-Shaarawi, A. H. and Piegorsch, W. W., editors, *Encyclopedia of Environmetrics*, volume 4, pages 2045–2049. John Wiley & Sons, Ltd, Chichester.

Cressie, N. A. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, corrected edition.

Curtain, R. (1975). A survey of infinite-dimensional filtering. *SIAM Review*, 17(3):395–411.

Da Prato, G. and Zabczyk, J. (1992). *Stochastic Equations in Infinite Dimensions*, volume 45 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press.

Davies, E. (1995). *Spectral Theory and Differential Operators*, volume 42 of *Cambridge studies in advanced mathematics*. Cambridge University Press.

Falb, P. (1967). Infinite-dimensional filtering: The Kalman–Bucy filter in Hilbert space. *Information and Control*, 11(1/2):102–137.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Texts in Statistical Science. Chapman & Hall / CRC press, second edition.

Grewal, M. S. and Andrews, A. P. (2001). *Kalman Filtering: Theory and Practice Using MATLAB*. Wiley-Intersciece, second edition.

Hartikainen, J. and Särkkä, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 379–384.

Holden, H., Øksendal, B., Ubøe, J., and Zhang, T. (1996). *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*. Probability and Its Applications. Birkhäuser, Boston.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Kaipio, J. and Somersalo, E. (2004). *Statistical and Computational Inverse Problems*. Number 160 in Applied Mathematical Sciences. Springer.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. John Wiley & Sons, New York.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain

activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12):5675–5679.

Lin, F.-H., Wald, L. L., Ahlfors, S. P., Hämäläinen, M. S., Kwong, K. K., and Belliveau, J. W. (2006). Dynamic magnetic resonance inverse imaging of human brain function. *Magnetic Resonance in Medicine*, 56:787–802.

Lin, F.-H., Witzel, T., Mandeville, J. B., Polimeni, J. R., Zeffiro, T. A., Greve, D. N., Wiggins, G., Wald, L. L., and Belliveau, J. W. (2008). Event-related single-shot volumetric functional magnetic resonance inverse imaging of visual processing. *NeuroImage*, 42:230–247.

Matérn, B. (1960). Spatial variation: stochastic models and their applications to some problems in forest surveys and other sampling investigations. *Meddelanden från statens skogsforskningsinstitut*, 49:1–144.

Maybeck, P. S. (1979). *Stochastic Models, Estimation and Control*, volume 1. Academic Press.

Maybeck, P. S. (1982). *Stochastic Models, Estimation and Control*, volume 2. Academic Press.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(1):1–42.

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag, sixth edition.

Olver, P. J. (2012). Introduction to Partial Differential Equations. Book draft available online: `http://www.math.umn.edu/~olver/pdn.html` (accessed January 1, 2012).

Omatu, S. and Seinfeld, J. H. (1989). *Distributed Parameter Systems: Theory and Applications*. Clarendon Press / Ohmsha.

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (1999). *Discrete-Time Signal Processing*. Prentice Hall, second edition.

Pikkarainen, H. K. (2005). *A Mathematical Model for Electrical Impedance Process Tomography*. Doctoral dissertion, Helsinki University of Technology.

Pikkarainen, H. K. (2006). State estimation approach to nonstationary inverse problems: discretization error and filtering problem. *Inverse problems*, 22:365–379.

Pivato, M. (2010). *Linear Partial Differential Equations and Fourier Theory*. Cambridge University Press.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

Rauch, H., Tung, F., and Striebel, C. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450.

Robinson, J. C. (2001). *Infinite-Dimensional Dynamical Systems: An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors*, volume 28 of *Cambridge Texts in Applied Mathematics*. Cambridge University Press.

Showalter, R. (1977). *Hilbert Space Methods for Partial Differential Equations*, volume 1. Pitman Publishing London.

Shubin, M. A. (1987). *Pseudodifferential Operators and Spectral Theory*. Springer Series in Soviet Mathematics. Springer-Verlag.

Solin, A. (2010). *Cubature Integration Methods in Non-Linear Kalman Filtering and Smoothing*. Bachelor's thesis, Faculty of Information and Natural Sciences, Aalto University, Finland.

Särkkä, S. (2006). *Recursive Bayesian inference on stochastic differential equations*. Doctoral dissertion, Helsinki University of Technology.

Särkkä, S. and Hartikainen, J. (2012). Infinite-dimensional Kalman filtering approach to spatio-temporal Gaussian process regression. In *JMLR Workshop and Conference Proceedings Volume 22: AISTATS 2012*, pages 993–1001.

Särkkä, S., Solin, A., Nummenmaa, A., Vehtari, A., Auranen, T., Vanni, S., and Lin, F.-H. (2012a). Dynamical retrospective filtering of physiological noise in BOLD fMRI: DRIFTER. *NeuroImage*, 60:1517–1527.

Särkkä, S., Solin, A., Nummenmaa, A., Vehtari, A., Auranen, T., Vanni, S., and Lin, F.-H. (2012b). Identification of spatio-temporal oscillatory signal structure in cerebral hemodynamics using DRIFTER. In *Proceedings of ISMRM 2012*. To appear as abstract and e-poster in ISMRM 2012.

Tzafestas, S. G. (1978). Distributed parameter state estimation. In Ray, W. H. and Lainiotis, D. G., editors, *Distributed Parameter Systems: Identification, Estimation and Control*. Marcel Dekker, Inc., New York.

Vanhatalo, J. and Vehtari, A. (2008). Modelling local and global phenomena with sparse Gaussian processes. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI 2008)*, pages 571–578.

Welch, G. and Bishop, G. (1995). An introduction to the Kalman filter. Technical report, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Wikle, C. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829.