Björn Sprungk

Numerical Methods for Bayesian Inference in Hilbert Spaces

Björn Sprungk

Numerical Methods for Bayesian Inference in Hilbert Spaces

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Universitätsverlag Chemnitz
2017

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über http://dnb.d-nb.de abrufbar.

Titelgrafik: Björn Sprungk
Satz/Layout: Björn Sprungk

Technische Universität Chemnitz/Universitätsbibliothek
Universitätsverlag Chemnitz
09107 Chemnitz
http://www.tu-chemnitz.de/ub/univerlag

readbox unipress
in der readbox publishing GmbH
Am Hawerkamp 31
48155 Münster
http://unipress.readbox.net

ISBN 978-3-96100-028-9

http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-226748

TECHNISCHE UNIVERSITÄT
CHEMNITZ
Fakultät für Mathematik

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

Numerical Methods for Bayesian Inference in Hilbert Spaces

vorgelegt von Dipl.-Math. Björn Sprungk,
geboren am 31. Mai 1985 in Freital

Tag der Einreichung: 01. Februar 2017

Betreuer: Prof. Dr. Oliver Ernst, Technische Universität Chemnitz

1. Gutachter: Prof. Dr. Oliver Ernst, Technische Universität Chemnitz
2. Gutachter: Prof. Dr. Andrew M. Stuart, California Institute of Technology
3. Gutachter: Prof. Dr. Fabio Nobile, École Polytechnique Fédérale de Lausanne

Tag der öffentlichen Prüfung: 09. Juni 2017

Contents

Chapter 1
Introduction

> I beseech you, in the bowels of Christ, think it possible that you may be mistaken.
>
> *Oliver Cromwell*

Uncertainty and probabilistic predictions are common in science, engineering and everday life: from tomorrow's weather forecast including the chance of precipitation to failure probabilities of networks and structures to Heisenberg's uncertainty principle in quantum mechanics. In practice, the uncertainty in predictions can have various sources:

- the (mathematical) model employed for the prediction is a simplified version of the real underlying process missing small or unresolvable features and effects,
- parameters or coefficients within the model are not known exactly such as material properties, or they are considered to be random by nature such as wind speed and direction or manufacturing imperfections,
- the model is solved by numerical methods and, thus, the resulting prediciton is affected by discretization errors.

Analyzing and estimating the latter is a classical task in numerical analysis. In recent years, particularly since Ghanem and Spanos [70], also the second item has gained more and more interest within the scientific computing community and several numerical methods for quantifying the uncertainty arising from incomplete knowledge about model parameters or random model coefficients have been developed and examined.

Although, as Smith already remarks in the preface of [160], "uncertainty quantification [...] is as old as the disciplines of probability and statistics", the innovation

in the last decades lies in the interaction and resulting synergy between statistics, probability theory and numerical analysis. In particular, an increase in computational power as well as improved algorithms enable us to estimate the resulting uncertain behaviour of complex systems modeled, e.g., by (systems of) partial differential equations (PDEs), given uncertain input data. This thesis contributes to the analysis and development of such algorithms.

In the following, we present the key ideas in uncertainty quantification (UQ) and outline how Bayesian inference is an essential part of UQ. For the sake of clarity we omit technical details at this point.

Uncertainty Quantification. We will consider an elliptic boundary value problem on a bounded domain $D \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, as the model problem describing, e.g., a stationary groundwater flow:

$$-\nabla \cdot (a(x)\, \nabla p(x)) = 0 \text{ on } D, \qquad p(x) = g(x) \text{ on } \partial D. \tag{1.1}$$

Here, a denotes the spatially varying (hydraulic) conductivity, g the (Dirichlet) boundary data and p the resulting (groundwater) pressure head. In particular, in subsurface physics the knowledge about a (as well as g) is limited and based only on finitely many measurement data and geological information. Scientific simulations of and predictions for the solution p of (1.1) or the associated flow $\boldsymbol{u} = -a\nabla p$ have to account for the incomplete knowledge about a and g. Although there exist different mathematical concepts to describe uncertainty such as fuzzy sets and interval arithmetics, see, e.g., Bandemer [8] or Oberguggenberger [127] for an overview, we will focus on probabilistic methods in this thesis. Thus, the limited knowledge about a and g is mathematically described by stochastic models such as random functions or probability measures on function spaces which leads, in turn, to a random solution p of (1.1).

Problems and resulting numerical methods considered within the field of uncertainty quantification can roughly be classified into two groups:

1. *Forward problems*: Given a stochastic model for the uncertainty in the input data of a PDE such as a and g in (1.1), quantify the resulting uncertainty in the solution p by computing the resulting probability distribution of the latter.

2. *Inverse problems*: Given noisy measurements of (functionals of) the solution of a PDE such as p in (1.1), adjust the current stochastic model for the uncertain input, e.g., a and g in (1.1), w.r.t. the additional information provided by these measurements.

In pratice, both classes of problems can appear in combined form, e.g., often measurements of both, the conductivity coefficient a as well as the pressure head p in (1.1), are available and, of course, all observational data is taken into account for constructing a stochastic model describing the available knowledge about a (inverse problem). On the other hand, the remaining uncertainty in a may then be propagated through a PDE such as (1.1) to compute the resulting probability laws of quantities of interest of p or $\boldsymbol{u} = -a\nabla p$, e.g., breakthrough times of pollutants, for predicitons or decision making (forward problem). In this thesis we will mainly study numerical algorithms for the inverse problem in UQ. Besides that, we will illustrate the whole UQ procedure for a real-word problem in Chapter 7.

The Inverse Problem in UQ and Bayesian Inference. Let us assume that we already have a stochastic model describing our current state of knowledge (or our current uncertainty) about the coefficient a in (1.1). Further, suppose that we are given new noisy observational data of the solution p of (1.1), i.e.,

$$y_i = p(x_i) + \varepsilon_i, \qquad i = 1, \dots, k, \tag{1.2}$$

where the ε_i denote random measurement errors following a known distribution. We then would like to incorporate this new information into our stochastic model for a in a consistent way and obtain an updated stochastic model. This is exactly the basic principle of *Bayesian inference*: to update prior beliefs or probabilities, respectively, as more information becomes available, see, e.g., Hoff [89] or Jaynes [94] for an introduction and further references. Let μ denote the prior probability distribution of a associated to our current stochastic model for a. Here, we assume that a belongs to a function space such as $L^\infty(D)$ or $C(\bar{D})$ which can be continuously embedded in a Hilbert space $\mathcal{H}$, e.g., $\mathcal{H} = L^2(D)$, and that μ is defined on the Borel σ-algebra of $\mathcal{H}$. Then, merging the prior knowledge described by μ with the new information provided by the measurement $y := (y_1, \dots, y_k) \in \mathbb{R}^k$ is mathematically done by conditioning the distribution μ on the event of observing y. The resulting *posterior* probability distribution μ^y of a is then explicitly given by *Bayes' rule* which under mild assumptions on the random errors ε_i takes the form

$$\mu^y(\mathrm{d}a) = \frac{1}{Z}\,\mathrm{e}^{-\Phi(a)}\;\mu(\mathrm{d}a), \tag{1.3}$$

where $\Phi\colon \mathcal{H} \to [0, \infty)$ denotes a measurable mapping and Z denotes the typically unknown normalizing constant. We then may be interested in statistics of the measure μ^y such as the mean and the covariance or we may want to generate samples

according to μ^y. The latter is particularly interesting when employing μ^y as the updated stochastic model for the uncertain input a in a succeeding forward problem.

Numerical Methods for Bayesian Inference and Contributions of the Thesis. Since the posterior measure μ^y can have a very complicated form, direct sampling methods are usually not available or feasible. However, a well-established method in Bayesian statistics to generate samples, which are (approximately) distributed according to μ^y, is the *Markov chain Monte Carlo* (MCMC) method, see, e.g., Tierney [170]. The basic idea behind MCMC methods is to construct a Markov chain, i.e., a sequence of random variables with Markov property, such that the distribution of the nth state of the chain converges to the posterior μ^y as $n \to \infty$. In the setting outlined above, i.e., Bayesian inference for a coefficient function a appearing in a PDE (1.1), the Markov chain has to run in an infinite dimensional state space, i.e., the function space $\mathcal{H}$. Thus, we require MCMC methods which are well-defined in general Banach or Hilbert spaces. Such MCMC algorithms have recently been developed, e.g., the *preconditioned Crank-Nicolson (pCN) Metropolis* algorithm, see Cotter et al. [35]. In fact, many other common MCMC methods are only defined in finite dimensional state spaces and show a deteriorating efficiency as the dimension increases, see, e.g. Roberts and Rosenthal [141].

One goal of this thesis is to combine MCMC algorithms for function spaces with another recent idea to improve the efficiency of MCMC: to allow the algorithm to exploit available "geometric" information about the posterior distribution such as (approximations of) the posterior covariance, see, e.g., Girolami and Calderhead [73]. In particular, the posterior covariance provides information about the "spread" of the posterior distribution in the different directions or dimensions of $\mathcal{H}$, respectively, and can guide the MCMC algorithm, e.g., to take steps of appropriate size in the corresponding directions. In Chapters 5 and 6 we will propose and analyze a generalization of the pCN Metropolis algorithm which is able to incorporate approximations of the posterior covariance.

Moreover, numerical experiments conducted with the new algorithm show its robust efficiency w.r.t. the variance of the observational noise, i.e., the variance of ε_i in (1.2). Such a robustness is surprising, since a smaller noise variance implies a higher concentration of the posterior distribution μ^y which in turn usually affects the performance of common MCMC methods. The observation made in Chapter 5 motivates the study of the efficiency of MCMC methods for decreasing noise variance. To the author's knowledge, there exists, so far, just one other publication considering this topic: the work of Beskos et al. [16]. In Chapter 6 we develop a different approach than Beskos et al. and prove a first result on the variance inde-

pendent efficiency of MCMC methods.

Moreover, we examine in Chapter 4 two other algorithms proposed for Bayesian inference and UQ: the *ensemble Kalman filter (EnKF)*, see, e.g., Evensen [62], and the *polynomial chaos Kalman filter (PCKF)*, see, e.g., Rosić et al. [143]. Both methods are extensions of the classical Kalman filter [97]. Although it is known that the EnKF does not provide samples which are distributed according to the posterior, a characterization of its outcome within the framework of Bayesian inference is still missing in the literature. However, recently, starting with Le Gland et al. [75], several convergence results for the EnKF were established under various assumptions. So far, no convergence results are known in case of the PCKF. Another contribution of the thesis is to fill the mentioned theoretical gaps for the EnKF and PCKF when both are applied to Bayesian inference problems.

1.1. Outline of the Thesis

The thesis is structured as follows:

Chapter 2: This chapter provides an introduction to important mathematical concepts for random functions including random fields and function space-valued random variables. We discuss the relation between both and consider elliptic PDEs with random fields as coefficients. Furthermore, recent numerical approximation methods for solutions of random PDEs are outlined.

Chapter 3: We present the Bayesian approach to statistical inference and inverse problems in separable Hilbert spaces. Besides the definition as a conditional measure and the stability of the posterior measure w.r.t. perturbations in the observed data or observational functionals, we also cover the concept of Bayes estimators and make some comments about the relation to the regularizational approach to inverse problems.

Chapter 4: This chapter analyzes two recent numerical Kalman filtering methods for Bayesian inference in Hilbert spaces, i.e., the ensemble Kalman filter and the polynomial chaos Kalman filter. For both methods we establish convergence results for the large ensemble and large polynomial basis limit, respectively, and, moreover, provide a characterization of their outcome within the Bayesian methodology.

Chapter 5: The Markov chain Monte Carlo method for approximate sampling of (posterior) measures on Hilbert spaces is outlined and a new Metropolis-

Hastings algorithm is proposed. The latter is a generalization of the well-known pCN Metropolis algorithm and allows to incorporate information about the posterior covariance. A geometric convergence result for the proposed algorithm is proven where the proof involves spectral gap theory and a comparison arguement for Metropolis-Hastings algorithms. Furthermore, in numerical experiments we show a superior performance of the new algorithm compared to the pCN Metropolis algorithm.

Chapter 6: This chapter is devoted to a deeper analysis of the phenomenon observed in the numerical simulations in Chapter 5: a robust performance of the gpCN Metropolis algorithm w.r.t. the (observational) noise or likelihood variance, respectively, in the Bayesian inference problem. A positive result is shown for the simple but already nontrivial case of linear observations with Gaussian prior and noise. Numerical experiments illustrate the theoretical findings.

Chapter 7: We perform Bayesian inference for a real-world groundwater flow model including real data.To this end, the methods developed and analyzed in Chapters 4 and 5 are applied to estimate the posterior distribution of breakthrough times of pollutants.

The content of the Chapters 3 to 5 is based on already published articles of the author. The way of presentation was, however, modified and extended for this thesis. The corresponding publications are:

- [56] O. Ernst, B. Sprungk, and H.-J. Starkloff. Bayesian inverse problems and Kalman filters. In S. Dahlke et al., editor, *Extraction of Quantifiable Information from Complex Systems*, volume 102 of *Lecture Notes in Computational Science and Engineering*, pages 133–159. Springer, 2014.

- [57] O. Ernst, B. Sprungk, and H.-J Starkloff. Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. *SIAM/ASA J. Uncertainty Quantification*, 3(1):823–851, 2015.

- [146] D. Rudolf and B. Sprungk. On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm. *Found. Comput. Math.*, 2016. doi:10.1007/s10208-016-9340-x.

1.2. Notation

Subsequently, we introduce some basic notations used in this thesis. By $\Rightarrow$ and $\Leftrightarrow$ we denote logical implication and equivalence, respectively. As usual, the sets of all natural, real and complex numbers will be denoted by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$, respectively, and we set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Furthermore, we will use $\mathbb{R}^{\mathbb{N}} := \bigtimes_{m=1}^{\infty} \mathbb{R}$ as symbol for the countably infinite Cartesian product of $\mathbb{R}$. The Euclidean norm in $\mathbb{R}^n$, $n \in \mathbb{N}$, is denoted by $|\cdot|$ and for a symmetric and positive definite matrix $W \in \mathbb{R}^{n\times n}$ we define

$$|x|_W := \left(x^\top W x\right)^{1/2}, \qquad x \in \mathbb{R}^n.$$

On the other hand, if M is a set $|M|$ denotes the cardinality of M.

Throughout the thesis $\mathcal{X}$ denotes an arbitrary real Banach space with norm $\|\cdot\|_{\mathcal{X}}$ and $\mathcal{H}$ an arbitrary separable real Hilbert space with norm $\|\cdot\|_{\mathcal{H}}$ and inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. The (topological) dual of a Banach space $\mathcal{X}$ is denoted by $\mathcal{X}^*$ and the corresponding duality pairing by $\langle\cdot,\cdot\rangle_{\mathcal{X}^*}\colon \mathcal{X}^* \times \mathcal{X} \to \mathbb{R}$.

Given a linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$ we denote its range by $\mathrm{rg}(A)$, its adjoint operator by $A^*\colon \mathcal{Y}^* \to \mathcal{X}^*$, i.e.,

$$\langle f, Ax\rangle_{\mathcal{Y}^*} = \langle A^* f, x\rangle_{\mathcal{X}^*}, \qquad \forall x \in \mathcal{X},\ \forall f \in \mathcal{Y}^*,$$

and its norm by

$$\|A\| := \sup_{\|x\|_{\mathcal{X}}=1} \|Ax\|_{\mathcal{Y}}$$

without any further subscripts. The space of all bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $\mathcal{L}(\mathcal{X},\mathcal{Y})$ where we set $\mathcal{L}(\mathcal{X}) := \mathcal{L}(\mathcal{X},\mathcal{X})$. For a bounded linear operator $A \in \mathcal{L}(\mathcal{X})$ we denote the spectrum of A (in $\mathcal{X}$) by $\mathrm{spec}(A \mid \mathcal{X})$. Given two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ let $\mathcal{L}^1(\mathcal{H}_1,\mathcal{H}_2)$ and $\mathcal{L}^2(\mathcal{H}_1,\mathcal{H}_2)$ be the space of all nuclear and Hilbert-Schmidt operators, respectively, from $\mathcal{H}_1$ to $\mathcal{H}_2$. We refer to Appendix A for the corresponding definitions. Moreover, the tensor product of $\mathcal{H}_1$ and $\mathcal{H}_2$ is denoted by $\mathcal{H}_1 \otimes \mathcal{H}_2$. More details on tensor products are given in Appendix B.

We often assume an underlying probability space which we denote by $(\Omega, \mathcal{A}, \mathbb{P})$. For the expectation and covariance w.r.t. $\mathbb{P}$ we use the notation $\mathbb{E}[\cdot]$ and $\mathrm{Cov}(\cdot,\cdot)$, respectively, and for the correlation between two random variables X and Y we write $\mathrm{Corr}(X,Y)$. For a metric space $\mathcal{E}$ we denote by $\mathcal{B}(\mathcal{E})$ the Borel σ-algebra induced by the underlying metric and by $\mathcal{P}(\mathcal{E})$ the set of all probability measures on $(\mathcal{E},\mathcal{B}(\mathcal{E}))$. We will usually denote measures on metric spaces by greek letters such as μ, ν or η. Moreover, for $x \in \mathcal{E}$ the Dirac measure at x is denoted by δ_x (and

should not be confused with the Kronecker delta δ_{ij} for $i, j \in \mathbb{N}_0$). The normal (or Gaussian) distribution on $\mathbb{R}$ with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$ is denoted by $N(m, \sigma^2)$ and an analogous notation is used for multivariate normal distributions. By $\mathrm{Uni}(a, b)$ we denote the uniform distribution on a finite interval $[a, b] \subset \mathbb{R}$. Further notation concerning (probability) measures is specified in Section 3.1. Given a random variable $X\colon (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{E}, \mathcal{B}(\mathcal{E}))$ and a $\mu \in \mathcal{P}(\mathcal{E})$ the notation $X \sim \mu$ means that the resulting probability distribution of X on $\mathcal{E}$ is μ. Random variables will mainly be denoted by capital letters.

Acknowledgements

First of all, I would like to express my deep gratitude to my supervisor, Oliver Ernst, for his great support over the last six years, in particular for his enduring encouragement and patience, his constant availability, and for the many opportunities he provided to present and discuss my research with others.

I am also very grateful to Hans-Jörg Starkloff for the enlightening discussions we had and his numerous hints to helpful literature. Another person who truly deserves great thanks is my dear colleague, Daniel Rudolf, who introduced me to the beauty of spectral gaps and with whom I share not only a pleasant collaboration, but also a precious friendship. Furthermore, I would like to thank Andrew Stuart for his many helpful comments and the opportunity of an inspiring four-week stay at the University of Warwick. I also express my gratitude to Albrecht Böttcher for providing help with functional calculus and operator theory. Moreover, I would like to thank my colleagues Elisabeth Ullmann and Ingolf Busch for allowing me to build upon their MATLAB code packages.

Finally, I am greatly thankful to my beloved family and friends for their continuous support. In particular, a very special thanks goes to my dear friend Hanna Rudolph who boosted confidence and motivation whenever needed.

Chapter 2

Random Fields and Random Elliptic PDEs

This chapter provides a brief introduction into random fields, their interpretation as Banach- or Hilbert-space valued random variables and elliptic partial differential equations with random fields as coefficients. We assume the reader is familiar with basic probability theory and real- or vector-valued random variables as well as with standard function spaces such as the space $C(D)$ of all continuous functions on a bounded domain $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, the Lebesgue space $L^2(D)$ of all square integrable functions on D and the Sobolev space $H^1(D)$ of all square integrable functions f on D with square integrable weak derivatives.

Our model problem throughout the chapter is the following elliptic boundary value problem (BVP) stated on a bounded domain $D \subseteq \mathbb{R}^d$ with boundary ∂D:

$$-\nabla \cdot (a(x)\,\nabla u(x)) = f(x) \text{ on } D, \qquad u(x) = g(x) \text{ on } \partial D. \tag{2.1}$$

Here the function $a\colon D \to \mathbb{R}_+$ describes a physical conductivity, $f\colon D \to \mathbb{R}$ a forcing term and $g\colon \partial D \to \mathbb{R}$ the (Dirichlet) boundary data for the solution $u\colon D \to \mathbb{R}$. Such an equation describes, e.g., a stationary groundwater flow or a stationary temperature field in a material. In practice the coefficients a, f and g are often not known exactly, e.g., the conductivity may vary with x in an incompletely known way. This chapter provides the mathematical foundations to treat such PDEs with uncertain coefficients.

Maybe the most natural and convenient approach to model probabilistically the uncertainty about an incompletely known function are stochastic processes. These are by definition families $\{X_t : t \in I\}$ of random variables X_t indexed by a parameter t varying within an index set I. For example, in financial mathematics stochastic processes indexed over a time intervall $I = [t_1, t_2]$ are used to model the (time-continuous) evolution of stock prices. If the index set I is a subset of $\mathbb{R}^d$ with $d > 1$

one often calls the stochastic process a *random field*. Random fields are applied, for instance, in geophysics and geostatistics to model uncertain physical properties of subsurface layers.

Elliptic BVPs such as (2.1) are usually treated in their weak formulation which are variational equations in function spaces, e.g., for (2.1)

$$\langle a\nabla u, \nabla v\rangle_{L^2(D)} = \langle f, v\rangle_{L^2(D)} \qquad \forall v \in H_0^1(D), \tag{2.2}$$

where $\langle\cdot,\cdot\rangle_{L^2(D)}$ denotes the inner product in $L^2(D)$ and we require $a \in L^\infty(D)$ and $f \in L^2(D)$. The functional analytic approach to solve and analyze the variational equation (2.2) suggests to view uncertain coefficients a and f as $L^\infty(D)$- and $L^2(D)$-valued random variables, respectively. Banach or Hilbert space-valued random variables and random fields are related but slightly different concepts to describe random functions and we will explain their relation in detail in Section 2.2. In particular, Hilbert space-valued random variables, such as $L^2(D)$-valued random variables, can be represented by a series expansion with random coefficients w.r.t. a complete orthonormal system (CONS) of the underlying Hilbert space. This expansion allows for a convenient approximation via truncation and for a parametrization of the random variable in terms of the random coefficients. This parametric perspective yields parametric reformulations of PDEs with random coefficients which we will outline in Section 2.3. Moreover, it forms the basis for many modern approximation methods for random PDEs. We provide a short introduction to the latter in Section 2.3.2.

In summary, this chapter focuses on the *forward problem* of uncertainty quantification, i.e., how does the uncertainty about coefficient functions propagate to the solution of the corresponding PDE and how can we numerically quantify it by approximation methods. For further reading on this topic, we refer to the textbooks by Lord et al. [114], Le Maitre and Knio [108] and Xiu [179].

2.1. Random Fields

Random fields are stochastic processes on domains D in $\mathbb{R}^d$, $d \geq 1$, and can be used as a probabilistic model for an uncertain function on D. Particularly, they are used in geophysics and geostatistics to describe limited knowledge about material properties such as, e.g., hydraulic or electrical conductivity of subsurface layers. We provide only the very basic definitions and concepts about random fields in the following and refer for more details to, e.g., Adler [2]. We recall, that $(\Omega, \mathcal{A}, \mathbb{P})$ denotes an underlying probability space.

Definition 2.1 (Random field, modification)**.** A *(real-valued) random field* a on a domain $D \subseteq \mathbb{R}^d$ is a mapping $a\colon D \times \Omega \to \mathbb{R}$ such that $a(x,\cdot)\colon \Omega \to \mathbb{R}$ is a random variable for each $x \in D$. Two random fields a_1 and a_2 on D are *modifications* of each other if for each $x \in D$ there holds $a_1(x) = a_2(x)$ $\mathbb{P}$-almost surely.

Thus, a random field can be seen as a collection of real-valued random variables indexed by the spatial coordinate $x \in D$. On the other hand, for each $\omega \in \Omega$ we can consider $a(\cdot,\omega)\colon D \to \mathbb{R}$ as a function on D depending on ω, i.e., a *random function*. By virtue of a basic result on stochastic processes, see Kallenberg [96, Lemma 3.1], we obtain

Proposition 2.2. Let $D \subseteq \mathbb{R}^d$ and a be a random field on D. Then the mapping $\omega \mapsto a(\cdot,\omega)$ from $(\Omega,\mathcal{A})$ to $(\mathbb{R}^D,\mathcal{S})$ is measurable where

$$\mathcal{S} := \sigma(\pi_x : x \in D), \qquad \pi_x(f) := f(x), \quad f \in \mathbb{R}^D, \tag{2.3}$$

denotes the σ-algebra generated by all point evaluation functionals $\pi_x\colon \mathbb{R}^D \to \mathbb{R}$, $x \in D$, i.e., the smallest σ-algebra on $\mathbb{R}^D$ such that for each $x \in D$ the mapping π_x is measurable.

We will later adopt this point of view by considering random fields as random elements in function spaces such as $C(D)$. Proposition 2.2 justifies the following

Definition 2.3 (Realization, path)**.** Let a be a random field on $D \subseteq \mathbb{R}^d$. For each $\omega \in \Omega$ the function $a(\cdot,\omega)\colon D \to \mathbb{R}$ is called *realization* or *path* or *sample* of the random field a.

We further define several important properties of a random field.

Definition 2.4 (Stationarity)**.** A random field a on $D \subseteq \mathbb{R}^d$ is called *stationary* if for any number $k \in \mathbb{N}$, any points $x_1,\dots,x_k \in D$ and any $h \in \mathbb{R}$ such that $x_1 + h,\dots,x_k + h \in D$ the random vectors $(a(x_1),\dots,a(x_k))^\top$ and $(a(x_1+h),\dots,a(x_k+h))^\top$ are identically distributed.

Definition 2.5 (Second order, mean and covariance function)**.** A random field a on $D \subseteq \mathbb{R}^d$ is a *second-order random field* if for each $x \in D$ there holds $a(x) \in L^2(\Omega;\mathbb{R})$. In this case the *mean field* or *mean function* of a is the mapping $m\colon D \to \mathbb{R}$ given by $m(x) := \mathbb{E}[a(x)]$ and the *covariance function* $c\colon D \times D \to \mathbb{R}$ of a is defined by $c(x,y) := \operatorname{Cov}(a(x),a(y))$.

Hence, a stationary second-order random field has a constant mean function and its covariance function is also constant along the diagonal $\{(x,x) : x \in D\}$. An even

stronger property than stationarity is that of *isotropy* which extends the translation invariance of the finite-dimensional distributions $(a(x_1), \dots, a(x_n))$ to invariance w.r.t. rotations or reflections.

Definition 2.6 (Isotropy)**.** A stationary random field a on $D \subseteq \mathbb{R}^d$ is called *isotropic* if for any number $k \in \mathbb{N}$, any points $x_1, \dots, x_k \in D$ and any orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ such that $Qx_1, \dots, Qx_k \in D$ the random vectors $(a(x_1), \dots, a(x_k))^\top$ and $(a(Qx_1), \dots, a(Qx_k))^\top$ are identically distributed.

For an isotropic random field the covariance function depends only on the distance $|x - y|$, i.e., there exists a function $\tilde{c} \colon [0, \infty) \to \mathbb{R}$ such that

$$c(x, y) = \mathrm{Cov}(a(x), a(y)) = \tilde{c}(|x - y|), \qquad \forall x, y \in D.$$

In the following, we will identify the covariance function c of an isotropic random field a with the mapping $\tilde{c}$ and simply write $c(|x - y|)$.

The smoothness of the mean and covariance functions m and c will determine the smoothness of the realizations of a second-order random field a. The deep theoretical result behind this issue is

Theorem 2.7 (Kolmogorov-Chentsov theorem, [96, Theorem 3.23])**.** Let a denote a second-order random field on $D \subseteq \mathbb{R}^d$ and let there exist positive constants s and K such that

$$\mathbb{E}\left[|a(x) - a(y)|^2\right] \leq K|x - y|^{d+s}, \qquad x, y \in D.$$

Then there exists a modification $\tilde{a}$ of a which is $\mathbb{P}$-a.s. Hölder continuous for any exponent less than $\frac{s}{2}$.

In general, we will identify a with its pathwise continuous modification if the latter exists. Now, due to

$$\begin{aligned}\mathbb{E}\left[|a(x) - a(y)|^2\right] &= \mathrm{Var}(a(x) - a(y)) + \mathbb{E}\left[a(x) - a(y)\right]^2 \\ &= \mathrm{Var}(a(x)) + \mathrm{Var}(a(y)) - 2\,\mathrm{Cov}(a(x), a(y)) \\ &\quad + \left(\mathbb{E}\left[a(x)\right] - \mathbb{E}\left[a(y)\right]\right)^2\end{aligned}$$

we see that an isotropic second-order random field a on $D \subseteq \mathbb{R}^d$ satisfies the assumptions of Theorem 2.1, if there exist positive constants s and K such that for its mean and covariance function m and c there holds

$$c(0) - c(|x - y|) \leq \frac{K}{2}|x - y|^{d+s}, \qquad |m(x) - m(y)|^2 \leq K|x - y|^{d+s}, \qquad x, y \in D.$$

Example 2.8 (Matérn covariance)**.** In geostatistics a common class of covariance functions for isotropic random fields are *Matérn covariance functions*. These functions are parametrized by three parameters: $\sigma^2 \in (0, \infty)$ determines the pointwise variance, $\rho \in (0, \infty)$ the so-called *correlation length* and $\nu \in (0, \infty)$ the smoothness of the random field. They are given by, see, e.g., Stein [164, Section 2.10, p. 50],

$$c_{\sigma^2,\rho,\nu}(r) := \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}|r|}{\rho} \right) K_\nu \left(\frac{2\sqrt{\nu}|r|}{\rho} \right), \qquad r \in \mathbb{R}, \tag{2.4}$$

where K_ν denotes the modified Bessel function of second kind and Γ the gamma function. The covariance function $c_{\sigma^2,\rho,\nu}$ is $2k$ times differentiable if $\nu > k$. This implies that a mean-zero random field a with $c_{\sigma^2,\rho,\nu}$ as its covariance function is k times *mean square differentiable* if $\nu > k$, see Stein [164, Section 2.6] for details and definitions. Moreover, by a recent result of Potthoff [134, Theorem 3.2], mean square differentiability implies under mild additional assumptions also pathwise differentiability. In particular, [134, Theroem 3.2] yields that for a mean-zero random field a with covariance function $c_{\sigma^2,\rho,\nu}$ there exists a modification of a with $k-1$ times differentiable realizations if $\nu > k$. If a is in addition also a *Gaussian random field*, see below, then there exists a modification of a with $k-1$ times differentiable realizations if $\nu > k$, see [134, Corollary 4.4.] in combination with Stein [164, Chapter 2, Equation (16)].

In the remainder of the thesis we will be particularly interested in *Gaussian random fields*.

Definition 2.9 (Gaussian and lognormal random field)**.** A random field a on a domain $D \subseteq \mathbb{R}^d$ is a *Gaussian random field* (GRF) if for any $k \in \mathbb{N}$ and any $x_1, \ldots, x_k \in D$ the joint distribution of $(a(x_1), \ldots, a(x_k))^\top$ is multivariate Gaussian. It is a *lognormal random field* if $\log a$ is a GRF.

Gaussian random fields are uniquely determined by their mean and covariance function as is the case for finite dimensional Gaussian distributions. The deep mathematical result behind this is the *Kolmogorov extension theorem* which states conditions for the existence of a random field a given all finite dimensional distributions $(a(x_1), \ldots, a(x_k))^\top$, see Adler [2, Section 1.5].

Example 2.10 (Brownian motion and Brownian bridge)**.** Maybe the most well-known Gaussian random field or Gaussian process is the (standard) *Brownian motion* B on $[0, \infty)$. This Gaussian process is given by the constant mean function $m(x) \equiv 0$ and the covariance function $c(x, y) = \min(x, y)$. Thus, we have $B(x) \sim N(0, x)$ as well

as $B(x) - B(y) \sim N(0, |x-y|)$, since

$$\begin{aligned}\mathrm{Var}(B(x) - B(y)) &= \mathrm{Var}(B(x)) - 2\,\mathrm{Cov}(B(x), B(y)) + \mathrm{Var}(B(y)) \\ &= x - 2\min(x,y) + y = |x-y|.\end{aligned}$$

By an analogous reasoning, we can also conclude that the increments $B(x) - B(y)$ and $B(u) - B(v)$ for $u < v < x < y$ are uncorrelated and, thus, independent. A related and probably equally well-known random field is the (standard) *Brownian bridge* BB on $[0,1]$ which is given by $BB(x) = B(x) - xB(1)$ or, equivalently, by a mean $m(x) \equiv 0$ and a covariance $c(x,y) = \min(x,y) - xy$. For both random fields B and BB it can be verified by Therorem 2.1 that the paths are Hölder continuous with an exponent less than $\frac{1}{2}$. Moreover, the Brownian motion or Brownian bridge can be easily extended to $[0,\infty)^d$ or $[0,1]^d$, respectively, by considering products of corresponding univariate random fields.

Furthermore, the statement of the Kolmogorov-Chentsov theorem can be improved if we consider Gaussian random fields. In particular, a mean-zero Gaussian random field a on a bounded domain $D \subset \mathbb{R}^d$ is $\mathbb{P}$-a.s. Hölder continuous with a Hölder exponent less than $s/2$ given that

$$\mathbb{E}\left[|a(x) - a(y)|^2\right] \leq K|x-y|^s, \qquad x, y \in D,$$

for some $0 < K < \infty$, see [114, Theorem 7.68]. Thus, a mean-zero isotropic Gaussian random field has $\mathbb{P}$-a.s. continuous realizations whenever there exist a $K < \infty$ and an $s > 0$ such that the covariance function c of the Gaussian random field a satisfies

$$c(0) - c(|x-y|) \leq \frac{K}{2}|x-y|^s \qquad \forall x, y \in D. \tag{2.5}$$

We remark that the Matérn covariance functions $c_{\sigma^2,\rho,\nu}$ satisfy (2.5) for $\nu \geq 1/2$.

Remark 2.11 (Sampling Gaussian random fields)**.** There are several ways to generate samples of a GRF a at finitely many points $x_i \in D$, $i = 1, \ldots, n$. Assume for simplicity that $\mathbb{E}[a(x)] \equiv 0$ and $c(x,y) = \mathrm{Cov}(a(x), a(y))$. Then the random vector $\boldsymbol{a} := (a(x_i))_{i=1,\ldots,n}$ is multivariate normally distributed $\boldsymbol{a} \sim N(0, C)$ with covariance matrix $C = (c(x_i, x_j))_{i,j=1}^n \in \mathbb{R}^{n\times n}$ and can be sampled via $\boldsymbol{a} = L\boldsymbol{\xi}$ where $\boldsymbol{\xi} \sim N(0, I)$ and $LL^\top = C$, cf. Proposition 2.20. Here, L can be computed via, e.g., a *Cholesky decomposition* of C. In case of an isotropic GRF a and if the x_i form a regular rectangular grid over D, the resulting Toeplitz structure of C can be exploited by *circulant embedding* and the fast Fourier transform to generate samples of $\boldsymbol{a}$ much faster. We refer to Dietrich and Newsam [45] for details. Another, rather classi-

cal, method for sampling random fields is the *turning bands method*, see Stein [164] for details. Finally, the Karhunen-Loève expansion of random fields a, see Section 2.2.2, provides an easy way to generate samples of a on D. To this end, we only have to sample the real-valued random coefficients appearing in the expansion. Of course, in numerical simulations we have to truncate the expansion after finitely many terms, i.e., we only obtain approximate samples of a. However, the error can be made arbitrarily small by suitably choosing the truncation index.

2.2. Banach and Hilbert Space Valued Random Variables

We recall the notation introduced in Section 1.2, e.g., $\mathcal{X}$ denoting a real Banach space with norm $\|\cdot\|_{\mathcal{X}}$ and $\mathcal{H}$ a separable Hilbert space with norm $\|\cdot\|_{\mathcal{H}}$ and inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$.

Definition 2.12 (Random variable, strongly measurable)**.** A mapping $X\colon (\Omega,\mathcal{A}) \to (\mathcal{X},\mathcal{B}(\mathcal{X}))$ is a *($\mathcal{X}$-valued) random variable* if it is measurable, i.e., if for each $B \in \mathcal{B}(\mathcal{X})$ there holds $X^{-1}(B) \in \mathcal{A}$. The mapping X is a *simple* ($\mathcal{X}$-valued) random variable if it takes the form

$$X(\omega) = \sum_{k=1}^{n} x_k \, \mathbf{1}_{A_k}(\omega), \qquad \omega \in \Omega,$$

where $x_k \in \mathcal{X}$ and $A_k \in \mathcal{A}$ for $k = 1,\dots,n$. An $\mathcal{X}$-valued random variable X is *strongly measurable* if there exists a sequence $(X_n)_{n\in\mathbb{N}}$ of simple $\mathcal{X}$-valued random variables such that $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ holds $\mathbb{P}$-almost surely.

For details about Banach space-valued mappings and Bochner integrals we refer, e.g., to Yosida [180, Chapter V] and state only the relevant facts. There is also the notion of *weakly measurable* $\mathcal{X}$-valued random variables which means that for each $f \in \mathcal{X}^*$ the mapping $\langle f, X\rangle_{\mathcal{X}^*}$ is an $\mathbb{R}$-valued random variable. It is clear that measurability implies weak measurability since $\langle f, X\rangle_{\mathcal{X}^*}$ is then a composition of the continuous mapping $\langle f,\cdot\rangle_{\mathcal{X}^*}\colon \mathcal{X} \to \mathbb{R}$ and the measurable mapping $X\colon \Omega \to \mathcal{X}$, thus, measurable. Moreover, from classical measure theory we know that the pointwise limit of measurable functions is again measurable, see, e.g., Kallenberg [96, Lemma 1.10], thus, strong measurability implies measurability. A useful relation between strong and weak measurability is given by *Pettis' theorem*, see Yosida [180, Section V.4], which states that $X\colon (\Omega,\mathcal{A}) \to (\mathcal{X},\mathcal{B}(\mathcal{X}))$ is strongly measurable iff it is weakly measurable and $\mathbb{P}$-almost surely *separably valued*. The latter means that

there exists a $\mathbb{P}$-null set $A_0 \in \mathcal{A}$ such that $X(\Omega \setminus A_0) \subseteq \mathcal{X}$ is separable. An immediate consequence of Pettis' theorem is

Proposition 2.13. Let $\mathcal{X}$ be a separable Banach space. Then each $\mathcal{X}$-valued random variable is strongly measurable.

We will mainly work with separable spaces later on. However, for elliptic PDEs with random diffusion coefficients we will also consider $L^\infty(D)$-valued random variables where D denotes a bounded domain in $\mathbb{R}^d$. In that case, the random variables under consideration will be given by expansions which ensures strong measurability, see also Proposition 2.22.

Analogously to the classical Lebesgue integral for real-valued functions we can define the Bochner integral of $\mathcal{X}$-valued strongly measurable mappings. This leads to *Lebesgue–Bochner* spaces.

Definition 2.14 (Lebesgue–Bochner space)**.** By $L^p(\Omega, \mathcal{A}, \mathbb{P}; \mathcal{X})$, or shorter $L^p(\Omega; \mathcal{X})$, $p \in [1, \infty)$, we denote the *Lebesgue–Bochner space* of all $\mathcal{X}$-valued strongly measurable random variables X with

$$\|X\|^p_{L^p(\Omega;\mathcal{X})} := \int_\Omega \|X(\omega)\|^p_{\mathcal{X}} \, \mathbb{P}(\mathrm{d}\omega) < \infty. \tag{2.6}$$

We will often use the shorter notation $\|\cdot\|_{L^p}$ instead of $\|\cdot\|_{L^p(\Omega;\mathcal{X})}$. Moreover, for $p = 2$ we define an inner product by

$$\langle X, Y \rangle_{L^2} := \int_\Omega \langle X(\omega), Y(\omega) \rangle_{\mathcal{H}} \, \mathbb{P}(\mathrm{d}\omega), \qquad X, Y \in L^2(\Omega; \mathcal{H}). \tag{2.7}$$

Condition (2.6) with $p \geq 1$ ensures that the strongly measurable random variable X is *Bochner integrable*, i.e., the integral $\int_\Omega X(\omega) \, \mathbb{P}(\mathrm{d}\omega) \in \mathcal{X}$ is defined. The Lebesgue–Bochner spaces $L^p(\Omega; \mathcal{X})$ are again Banach spaces and the space $L^2(\Omega; \mathcal{H})$ is again a Hilbert space w.r.t. its inner product.

Remark 2.15 (Separability of Lebesgue–Bochner spaces)**.** If the Banach space $\mathcal{X}$ itself is separable and the σ-algebra $\mathcal{A}$ is *countably generated*, i.e., there exist $A_n \subset \Omega$, $n \in \mathbb{N}$, such that $\mathcal{A} = \sigma(A_n : n \in \mathbb{N})$, then the Lebesgue–Bochner space $L^p(\Omega; \mathcal{X})$ is also separable. The first condition is obviously necessary and the second condition ensures that the usual Lebesgue spaces $L^p(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R})$ are separable, see Schilling [153, Lemma 23.19]. Moreover, $\mathcal{A}$ being countably generated is a rather mild assumption, e.g., the Borel σ-algebra of any separable metric space is countably generated.

Definition 2.16 (Mean)**.** Let $X \in L^1(\Omega; \mathcal{X})$, then the *mean* or *expectation* $\mathbb{E}[X]$ of X is given by the Lebesgue–Bochner integral

$$\mathbb{E}[X] := \int_\Omega X(\omega)\, \mathbb{P}(\mathrm{d}\omega) \in \mathcal{X}. \tag{2.8}$$

Definition 2.17 (Covariance)**.** For $X \in L^2(\Omega; \mathcal{X})$ and $Y \in L^2(\Omega; \mathcal{Y})$, where $\mathcal{Y}$ denotes another separable Banach space, we define the *covariance* $\mathrm{Cov}(X, Y)$ of X and Y as the bounded linear operator $C\colon \mathcal{Y}^* \to \mathcal{X}$ uniquely determined by

$$\langle f, Cg\rangle_{\mathcal{X}^*} = \mathrm{Cov}\left(\langle f, X\rangle_{\mathcal{X}^*}, \langle g, Y\rangle_{\mathcal{Y}^*}\right) \qquad \forall f \in \mathcal{X}^*, g \in \mathcal{Y}^*. \tag{2.9}$$

We set $\mathrm{Cov}(X) := \mathrm{Cov}(X, X) \in \mathcal{L}(\mathcal{X}^*, \mathcal{X})$.

In the following we will mainly work with covariance operators for Hilbert space-valued random variables. Due to the Riesz representation theorem we will identify the covariance $\mathrm{Cov}(X, Y)$ of $X \in L^2(\Omega; \mathcal{H}_1)$ and $Y \in L^2(\Omega; \mathcal{H}_2)$, where $\mathcal{H}_1, \mathcal{H}_2$ are two separable Hilbert spaces, with the bounded linear operator $C \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ given by

$$\langle x, Cy\rangle_{\mathcal{H}} = \mathrm{Cov}\left(\langle x, X\rangle_{\mathcal{H}_1}, \langle Y, y\rangle_{\mathcal{H}_2}\right) \qquad \forall x \in \mathcal{H}_1, y \in \mathcal{H}_2. \tag{2.10}$$

The covariance operator $\mathrm{Cov}(X)$ inherits the properties of covariance matrices, i.e., symmetry and positive definiteness. Moreover, $\mathrm{Cov}(X)$ is a *trace class operator*, see Appendix A for the corresponding definition.

Proposition 2.18 (cf. [40, Proposition 1.8])**.** For each $X \in L^2(\Omega; \mathcal{H})$ the covariance operator $\mathrm{Cov}(X)$ is self-adjoint, positive and trace class. Moreover, for $X \in L^2(\Omega; \mathcal{H}_1)$ and $Y \in L^2(\Omega; \mathcal{H}_2)$ the covariance operator $\mathrm{Cov}(X, Y)$ is Hilbert-Schmidt.

Proof. The statements about $\mathrm{Cov}(X)$ are proven in Da Prato and Zabczyk [40, Proposition 1.8], thus, we only show that $C := \mathrm{Cov}(X, Y)$ is Hilbert-Schmidt. For simplicity, we assume that X and Y have mean zero. Let $\{e_k : k \in \mathbb{N}\}$ be a CONS for $\mathcal{H}_1$ and $\{f_n : n \in \mathbb{N}\}$ denote a CONS for $\mathcal{H}_2$. Then there holds

$$\begin{aligned}
\sum_{n=1}^{\infty} \|Cf_n\|_{\mathcal{H}_1}^2 &= \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \langle e_k, Cf_n\rangle_{\mathcal{H}_1}^2 = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mathrm{Cov}\left(\langle e_k, X\rangle_{\mathcal{H}_1}, \langle f_n, Y\rangle_{\mathcal{H}_2}\right)^2 \\
&\le \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mathbb{E}\left[\langle e_k, X\rangle_{\mathcal{H}_1}^2\right] \mathbb{E}\left[\langle f_n, Y\rangle_{\mathcal{H}_2}^2\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{\infty} \langle e_k, X\rangle_{\mathcal{H}_1}^2\right] \mathbb{E}\left[\sum_{n=1}^{\infty} \langle f_n, Y\rangle_{\mathcal{H}_2}^2\right] = \|X\|_{L^2}^2 \|Y\|_{L^2}^2 < \infty,
\end{aligned}$$

where we have applied the Cauchy-Schwarz inequality and dominated convergence in the second line. Hence, C is Hilbert-Schmidt. □

It could be that the covariance $\mathrm{Cov}(X, Y)$ is even a nuclear operator, however, we were not able to prove it nor did we find a corresponding result in the literature. Anyway, the statement of Proposition 2.18 is sufficient for our purposes. In particular, Proposition 2.18 allows us to exploit the isomorphy $\mathcal{L}^2(\mathcal{H}_2, \mathcal{H}_1) \simeq \mathcal{H}_1 \otimes \mathcal{H}_2$, see Proposition B.3 in Appendix B, in order to express the covariance of Hilbert space-valued random variables as tensor products:

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X]) \otimes (Y - \mathbb{E}[Y])\right], \tag{2.11}$$

where $(X - \mathbb{E}[X]) \otimes (Y - \mathbb{E}[Y])$ is now a (mean-zero) $\mathcal{H}_1 \otimes \mathcal{H}_2$-valued random variable.

As for random fields, we are particularly interested in Gaussian random variables taking values in Banach or Hilbert spaces. To this end, we recall that one characterization of multivariate normally distributed random vectors $X \sim N(\boldsymbol{m}, \Sigma)$ with mean $\boldsymbol{m} \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ is that

$$\boldsymbol{a}^\top X \sim N(\boldsymbol{a}^\top \boldsymbol{m}, \boldsymbol{a}^\top \Sigma \boldsymbol{a}) \qquad \forall \boldsymbol{a} \in \mathbb{R}^n.$$

This extends naturally to Banach spaces:

Definition 2.19 (Gaussian random variable)**.** An $\mathcal{X}$-valued random variable $X \in L^2(\Omega; \mathcal{X})$ with mean $\mathbb{E}[X] = m \in \mathcal{X}$ and covariance $\mathrm{Cov}(X) = C \in \mathcal{L}(\mathcal{X}^*; \mathcal{X})$ is called *Gaussian* if

$$\langle f, X \rangle_{\mathcal{X}^*} \sim N\left(\langle f, m \rangle_{\mathcal{X}^*}, \langle f, \mathrm{Cov}(X) f \rangle_{\mathcal{X}^*}\right) \qquad \forall f \in \mathcal{X}^*. \tag{2.12}$$

We then denote $X \sim N(m, C)$.

In the Hilbert space case (2.12) reads as

$$\langle x, X \rangle_{\mathcal{H}} \sim N\left(\langle x, \mathbb{E}[X] \rangle_{\mathcal{H}}, \langle x, \mathrm{Cov}(X) x \rangle_{\mathcal{H}}\right) \qquad \forall x \in \mathcal{H}.$$

We will consider *Gaussian measures* on Hilbert spaces in more detail in Section 3.1 and Appendix C.

As for multivariate normally distributed random vectors the class of Gaussian $\mathcal{X}$-valued random variables is invariant w.r.t. linear transformations:

Proposition 2.20 ([39, Proposition 1.2.3])**.** Let $X \sim N(m, C)$ be an $\mathcal{X}$-valued Gaussian random variable. Then for each $b \in \mathcal{Y}$ and $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, where $\mathcal{Y}$ denotes another separable Banach space, there holds

$$b + AX \sim N(b + Am, ACA^*) \tag{2.13}$$

with $A^* \in \mathcal{L}(Y^*, X^*)$ denoting the adjoint of A.

Although, Da Prato and Zabczyk [39, Proposition 1.2.3] stated the result only for Hilbert spaces, it can also easily be verified for Banach spaces.

2.2.1. Expansions of Hilbert Space-Valued Random Variables

Since the Hilbert space $\mathcal{H}$ is assumed to be separable, it has a complete orthonormal system (CONS), say, $\{\phi_m : m \in \mathbb{N}\}$. Then, for each $X \in L^2(\Omega; \mathcal{H})$, we can define the real-valued random variables

$$\xi_m(\omega) := \langle \phi_m, X(\omega) \rangle_{\mathcal{H}}, \qquad m \in \mathbb{N}.$$

Thus, $\mathbb{P}$-almost surely there holds

$$X(\omega) = \sum_{m=1}^{\infty} \xi_m(\omega)\, \phi_m. \tag{2.14}$$

By construction we also have $\xi_m \in L^2(\Omega; \mathbb{R})$, since $|\xi_m(\omega)|^2 \leq \|X(\omega)\|_{\mathcal{H}}^2$ holds $\mathbb{P}$-almost surely, and, in particular, the series above converges also in $L^2(\Omega; \mathcal{H})$ to X: it is known, see Kallenberg [96, Proposition 4.12], that a sequence of random variables which converges $\mathbb{P}$-a.s. also converges in the L^p-sense if their L^p-norms are uniformly bounded and for the above series we obtain for $p = 2$

$$\begin{aligned}
\left\| \sum_{m=1}^{M} \xi_m(\omega)\, \phi_m \right\|_{L^2}^2 &= \mathbb{E}\left[\langle \sum_{m=1}^{M} \xi_m(\omega)\, \phi_m, \sum_{n=1}^{M} \xi_n(\omega)\, \phi_n \rangle_{\mathcal{H}} \right] \\
&= \mathbb{E}\left[\sum_{m=1}^{M} \xi_m^2(\omega) \right] = \sum_{m=1}^{M} \mathbb{E}\left[\xi_m^2(\omega) \right] \\
&\leq \sum_{m=1}^{\infty} \mathbb{E}\left[\xi_m^2(\omega) \right] = \mathbb{E}\left[\langle X(\omega), X(\omega) \rangle_{\mathcal{H}} \right] \\
&= \|X\|_{L^2}^2 < \infty.
\end{aligned}$$

Concerning the distribution of the random variables ξ_m, $m \in \mathbb{N}$, we easily see that

$$\mathbb{E}[\xi_m] = \langle \mathbb{E}[X], \phi_m \rangle_{\mathcal{H}}, \qquad \mathrm{Cov}(\xi_m, \xi_n) = \langle \phi_m, \mathrm{Cov}(X)\phi_n \rangle_{\mathcal{H}}, \qquad m, n \in \mathbb{N}.$$

Recalling that covariance operators are selfadjoint and compact, we can apply the spectral theorem for compact operators, see, e.g., Dunford and Schwartz [50, Chapter VII], and choose the eigenfunctions of $\mathrm{Cov}(X)$ as a CONS of $\mathcal{H}$. This yields

Theorem 2.21 ([156, Theorem C.29])**.** Let $X \in L^2(\Omega; \mathcal{H})$ and let $(\lambda_m, \phi_m)_{m\in\mathbb{N}}$ denote the eigenpairs of $\mathrm{Cov}(X) \in \mathcal{L}(\mathcal{H})$. Then there exist uncorrelated mean-zero random variables $\xi_m \in L^2(\Omega; \mathbb{R})$, $m \in \mathbb{N}$, with unit variance such that

$$X = \mathbb{E}[X] + \sum_{m=1}^{\infty} \sqrt{\lambda_m}\phi_m\xi_m \tag{2.15}$$

holds in $L^2(\Omega; \mathcal{H})$ and also $\mathbb{P}$-almost surely. Moreover, if X is Gaussian, then $\xi_m \sim N(0, 1)$ i.i.d. .

The representation (2.15) is an abstract version of the well-known *Karhunen-Loève expansion* for stochastic processes or random fields which we will encounter later. In particular, (2.15) guides us to a construction of finite dimensional approximations of a Hilbert space-valued random variable X by truncating the series in (2.15)

$$X_M := \mathbb{E}[X] + \sum_{m=1}^{M} \sqrt{\lambda_m}\phi_m\xi_m, \qquad M \in \mathbb{N}, \tag{2.16}$$

where the resulting error in $L^2(\Omega; \mathcal{H})$ is then given by

$$\|X - X_M\|_{L^2}^2 = \sum_{m>M}^{\infty} \lambda_m. \tag{2.17}$$

As it turns out the approximation (2.16) is the best M-term approximation of X, where by the former we mean approximations given by $\sum_{m=1}^{M} \tilde{\phi}_m \tilde{\xi}_m$ for $\tilde{\phi}_m \in \mathcal{H}$ and $\tilde{\xi}_m \in L^2(\Omega; \mathbb{R})$ for $m \geq 1$. We will not give a proof of this optimality statement but refer to Ghanem and Spanos [70, Section 2.3] for a discussion in the case of $\mathcal{H} = L^2(D)$. Another consequence of Theorem 2.21 is that for $X \in L^2(\Omega; \mathcal{H})$ we can write

$$\mathrm{Cov}(X) = \sum_{m=1}^{\infty} \lambda_m \phi_m \otimes \phi_m, \tag{2.18}$$

where (λ_m, ϕ_m) are again the eigenpairs of $\mathrm{Cov}(X)$, see, e.g., Schwab and Gittelson [156, Corollary C.28] for a proof. We close this subsection with two kind of converse

statements to Theorem 2.21, i.e., constructing Banach and Hilbert space-valued random variables via expansions.

Proposition 2.22. Let $\xi_m \in L^2(\Omega;\mathbb{R})$, $m \in \mathbb{N}$, be stochastically independent with $\mathbb{E}[\xi_m] = 0$ and $\mathrm{Var}(\xi_m) = 1$ for $m \in \mathbb{N}$. Given $\{\phi_m\}_{m\in\mathbb{N}} \subseteq \mathcal{X}$ where $\mathcal{X}$ denotes a Banach space and

$$\sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{X}} < \infty,$$

then the limit $X := \sum_{m=1}^{\infty} \phi_m \xi_m$ exists $\mathbb{P}$-a.s. and defines an $\mathcal{X}$-valued strongly measurable random variable with $X \in L^2(\Omega;\mathcal{X})$, $\mathbb{E}[X] = 0$ and $\mathrm{Cov}(X)\colon \mathcal{X}^* \to \mathcal{X}$ given by

$$\mathrm{Cov}(X) f = \sum_{m=1}^{\infty} \langle f, \phi_m \rangle_{\mathcal{X}^*} \phi_m, \qquad \forall f \in \mathcal{X}^*.$$

Moreover, if all ξ_m, $m \in \mathbb{N}$, are normally distributed, then X defines a Gaussian random variable.

Proof. We start with

$$\left\| \sum_{m=1}^{\infty} \phi_m \xi_m(\omega) \right\|_{\mathcal{X}} \leq \sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{X}} \, |\xi_m(\omega)| = \sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{X}} \, \mathbb{E}[|\xi_m|] + \sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{X}} \, \zeta_m,$$

where we set $\zeta_m := |\xi_m| - \mathbb{E}[|\xi_m|]$. Since $\mathbb{E}[|\xi_m|] \leq \mathbb{E}\left[|\xi_m|^2\right]^{1/2} = \mathrm{Var}(\xi_m)^{1/2} = 1$, we obtain due to the assumption

$$\sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{X}} \, \mathbb{E}[|\xi_m|] \leq \sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{X}} < \infty.$$

Moreover, the random variables ζ_m are stochastically independent, mean-zero and $\mathrm{Var}(\zeta_m) = \mathbb{E}\left[|\xi_m|^2\right] - \mathbb{E}[|\xi_m|]^2 \leq \mathbb{E}\left[|\xi_m|^2\right] \leq 1$. Thus, due to the assumption there holds

$$\sum_{m\geq 1} \mathrm{Var}(\|\phi_m\|_{\mathcal{X}} \, \zeta_m) = \sum_{m\geq 1} \|\phi_m\|_{\mathcal{X}}^2 \, \mathrm{Var}(\zeta_m) \leq \sum_{m\geq 1} \|\phi_m\|_{\mathcal{X}}^2 \leq \sum_{m\geq 1} \|\phi_m\|_{\mathcal{X}} < \infty.$$

This, in turn, implies that $\sum_{m\geq 1} \|\phi_m\|_{\mathcal{X}} \, \zeta_m$ converges $\mathbb{P}$-almost surely, see Kallenberg [96, Lemma 4.16]. Hence, we have

$$\left\| \sum_{m=1}^{\infty} \phi_m \xi_m(\omega) \right\|_{\mathcal{X}} < \infty \qquad \mathbb{P}\text{-a.s.}$$

and, therefore, X is well-defined and measurable. The strong measurability follows from the fact that X is weakly measurable and clearly separably valued.

Next, we show that $X \in L^2(\Omega;\mathcal{X})$: due to the independence of the ξ_m there holds for arbitrary $M \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{m=1}^{M}\phi_m\xi_m\right\|_{\mathcal{X}}^2\right] &\leq \mathbb{E}\left[\left(\sum_{m=1}^{M}\|\phi_m\|_{\mathcal{X}}\,|\xi_m|\right)^2\right] = \sum_{m,n=1}^{M}\|\phi_m\|_{\mathcal{X}}\,\|\phi_n\|_{\mathcal{X}}\,\mathbb{E}\left[|\xi_m||\xi_n|\right] \\
&= \sum_{m=1}^{M}\sum_{n\neq m,n=1}^{M}\|\phi_m\|_{\mathcal{X}}\,\|\phi_n\|_{\mathcal{X}}\,\mathbb{E}\left[|\xi_m|\right]\,\mathbb{E}\left[|\xi_n|\right] \\
&\quad + \sum_{m=1}^{M}\|\phi_m\|_{\mathcal{X}}^2\,\mathbb{E}\left[|\xi_m|^2\right] \\
&\leq \left(\sum_{m=1}^{M}\|\phi_m\|_{\mathcal{X}}\,\mathbb{E}\left[|\xi_m|\right]\right)\left(\sum_{n=1}^{M}\|\phi_n\|_{\mathcal{X}}\,\mathbb{E}\left[|\xi_n|\right]\right) + \sum_{m=1}^{M}\|\phi_m\|_{\mathcal{X}}^2 \\
&\leq \left(\sum_{m=1}^{\infty}\|\phi_m\|_{\mathcal{X}}\right)^2 + \sum_{m=1}^{\infty}\|\phi_m\|_{\mathcal{X}} < \infty
\end{aligned}$$

which implies that $\mathbb{E}\left[\|X\|_{\mathcal{X}}^2\right] < \infty$. Exploiting again the independence of the ξ_m we obtain for $f, g \in \mathcal{X}^*$

$$\begin{aligned}
\operatorname{Cov}\left(\langle f, X\rangle_{\mathcal{X}^*}, \langle g, X\rangle_{\mathcal{X}^*}\right) &= \sum_{m,n=1}^{\infty}\langle f,\phi_m\rangle_{\mathcal{X}^*}, \langle g,\phi_n\rangle_{\mathcal{X}^*}\operatorname{Cov}(\xi_m,\xi_n) \\
&= \left\langle g, \sum_{m=1}^{\infty}\langle f,\phi_m\rangle_{\mathcal{X}^*}\phi_m\right\rangle_{\mathcal{X}^*}
\end{aligned}$$

where all series converge under the given assumption that $\sum_{m=1}^{\infty}\|\phi_m\|_{\mathcal{X}}^2 < \infty$. If the ξ_m are Gaussian, then with $X_M := \sum_{m=1}^{M}\phi_m\xi_m$ we obtain for each $f \in \mathcal{X}^*$

$$\langle f, X_m\rangle_{\mathcal{X}^*} = \sum_{m=1}^{M}\langle f,\phi_m\rangle_{\mathcal{X}^*}\xi_m \sim N\left(0, \sum_{m=1}^{M}\langle f,\phi_m\rangle_{\mathcal{X}^*}^2\operatorname{Var}(\xi_m)\right).$$

Now, due to $\|X - X_M\|_{L^2(\Omega;\mathcal{X})} \to 0$, it follows that $\langle f, X_m\rangle_{\mathcal{X}^*} \to \langle f, X\rangle_{\mathcal{X}^*}$ in $L^2(\Omega;\mathbb{R})$ which implies, in particular, convergence in distribution. Thus, the assertion follows by noticing that $\lim_{M\to\infty}\langle f, X_m\rangle_{\mathcal{X}^*}$ is normally distributed with mean 0 and variance $\sum_{m=1}^{\infty}\langle f,\phi_m\rangle_{\mathcal{X}^*}^2$. Again, the latter converges due to the summability of $\|\phi_m\|_{\mathcal{X}}^2$, $m \in \mathbb{N}$. □

We remark, that if the assumption on the stochastic independence of the ξ_m in Proposition 2.22 would be omitted, then $X := \sum_{m=1}^{\infty}\phi_m\xi_m$ would still define an $\mathcal{X}$-valued random variable but in the sense of an $L^2(\Omega;\mathcal{X})$-limit. Moreover, in the case of series expansions for Hilbert space-valued random variables, we can obtain

a well-defined $L^2(\Omega;\mathcal{H})$-limit under a milder assumption on the summability of the norms $\|\phi_m\|_{\mathcal{H}}$ given orthogonality.

Proposition 2.23. Assume that for $m \in \mathbb{N}$ we have $\xi_m \in L^2(\Omega;\mathbb{R})$ with $\mathbb{E}[\xi_m] = 0$ and $\mathrm{Var}(\xi_m) = 1$. Then, for an orthogonal system $\{\phi_m\}_{m\in\mathbb{N}} \subseteq \mathcal{H}$ with

$$\sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{H}}^2 < \infty,$$

the limit $X := \sum_{m=1}^{\infty} \phi_m \xi_m$ in $L^2(\Omega;\mathcal{H})$ exists with $\mathbb{E}[X] = 0$ and $\mathrm{Cov}(X)\colon \mathcal{H} \to \mathcal{H}$ given by

$$\mathrm{Cov}(X) = \sum_{m,n=1}^{\infty} \mathrm{Cov}(\xi_m, \xi_n)\, \phi_m \otimes \phi_n.$$

Moreover, if all ξ_m, $m \in \mathbb{N}$, are normally distributed, then X defines a Gaussian random variable.

Proof. We notice that for arbirary $M \in \mathbb{N}$

$$\mathbb{E}\left[\left\|\sum_{m=1}^{M} \phi_m \xi_m\right\|_{\mathcal{H}}^2\right] = \sum_{m=1}^{M} \|\phi_m\|_{\mathcal{H}}^2\, \mathbb{E}\left[\xi_m^2\right] = \sum_{m=1}^{M} \|\phi_m\|_{\mathcal{H}}^2 = \sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{H}}^2 < \infty.$$

This implies, that the series $X = \sum_{m=1}^{\infty} \phi_m \xi_m$ converges in $L^2(\Omega;\mathcal{H})$. Moreover, for $f, g \in \mathcal{H}$ we obtain

$$\begin{aligned}\mathrm{Cov}\left(\langle f, X\rangle_{\mathcal{H}}, \langle g, X\rangle_{\mathcal{H}}\right) &= \sum_{m,n=1}^{\infty} \langle f, \phi_m\rangle_{\mathcal{H}} \langle g, \phi_n\rangle_{\mathcal{H}} \mathrm{Cov}(\xi_m, \xi_n) \\ &= \sum_{m,n=1}^{\infty} (\phi_m \otimes \phi_n)(f, g)\, \mathrm{Cov}(\xi_m, \xi_n)\end{aligned}$$

where all series converge under the given assumption that $\sum_{m=1}^{\infty} \|\phi_m\|_{\mathcal{H}}^2 < \infty$. The proof for the Gaussianity of X follows analogously to Proposition 2.22. □

2.2.2. Random Fields as Banach and Hilbert Space-Valued Random Variables

As mentioned earlier, we can view random fields on $D \subseteq \mathbb{R}^d$ as random functions from D to $\mathbb{R}$ where the σ-algebra on the linear space $\mathbb{R}^D$ of all mappings $f\colon D \to \mathbb{R}$ is specified in Proposition 2.2. Thus, it is tempting to assume that random fields with realizations which belong $\mathbb{P}$-a.s. to a certain function space such as $\mathcal{X} = C(D)$ are $\mathcal{X}$-valued random variables. This requires the measurability of the mapping

$\Omega \to \mathcal{X}$. In the following we will outline when random fields are $C(D)$- and $L^2(D)$-valued random variables. A first and classical result is

Proposition 2.24 ([96, Lemma 16.1], [136, Theorem 1])**.** Let a be a random field on a compact domain $\bar{D} \subset \mathbb{R}^d$ with $\mathbb{P}$-almost surely continuous paths. Then $a\colon \Omega \to C(\bar{D})$ is a $C(\bar{D})$-valued random variable. Conversely, each $C(\bar{D})$-valued random variable defines a random field on $\bar{D}$ with $\mathbb{P}$-almost surely continuous paths. Moreover, each Gaussian random field on $\bar{D}$ with $\mathbb{P}$-a.s. continous paths defines a Gaussian $C(\bar{D})$-valued random variable and vice versa.

The first part of the proposition follows by virtue of a result in Kallenberg [96, Lemma 16.1] which tells us that $\mathcal{B}(C(\bar{D})) = \mathcal{S}$ with $\mathcal{S} = \sigma(\pi_x : x \in D)$ as given in Proposition 2.2. The second part about Gaussian random fields follows by a result of Rajput and Cambanis [136, Theorem 1] which states that a GRF on $\bar{D}$ with $\mathbb{P}$-a.s. continuous paths yields a Gaussian measure on $C(\bar{D})$ and vice versa. A similar result to the first part of Proposition 2.24 holds also for random fields on $[0,\infty)$ with $\mathbb{P}$-a.s. rcll (right continuous with lefthand limits) realizations, i.e., they are then $\mathcal{D}(0,\infty)$-valued random variables with $\mathcal{D}(0,\infty)$ denoting the space of all rlcc function on $[0,\infty)$, see Kallenberg [96, Theorem A2.2] for more details.

However, we would like to work with Hilbert space-valued random variables, since this will be more convenient. If a has $\mathbb{P}$-almost surely continuous paths a natural candidate for a Hilbert space would be $\mathcal{H} = L^2(D)$, since then the realizations of a are also elements of $L^2(D)$. As before we have to ensure the measurability of the mapping $a\colon (\Omega, \mathcal{A}) \to (L^2(D), \mathcal{B}(L^2(D)))$, but this can be easily concluded from the continuous embedding of $C(\bar{D})$ in $L^2(D)$ for a bounded domain $D \subset \mathbb{R}^d$:

Proposition 2.25. Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces such that $\mathcal{X}$ is continuously embedded in $\mathcal{Y}$ and let $X\colon (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B}(\mathcal{X}))$ be measurable. Then also the mapping $X\colon (\Omega, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is measurable. Moreover, if X is a Gaussian $\mathcal{X}$-valued random variable, then X is also a Gaussian $\mathcal{Y}$-valued random variable.

Proof. Since the embedding is continuous the inclusion map $\iota\colon \mathcal{X} \to \mathcal{Y}$ is continuous and, thus, measurable. Therefore, the composition $\iota \circ X\colon \Omega \to \mathcal{Y}$ is again measurable. The second assertion follows by the linearity of ι and Proposition 2.20. □

By the continuous embedding of $C(\bar{D})$ into $L^p(D)$, $p \in [1,\infty]$, for a bounded domain $D \subset \mathbb{R}^d$ and applying Proposition 2.25, we obtain

Proposition 2.26. There holds:

- If a is a random field on a compact domain $\bar{D} \subset \mathbb{R}^d$ with $\mathbb{P}$-a.s. continuous realizations, then it defines an $L^p(D)$-valued random variable for $p \in [1,\infty]$.

- Let a be a second-order random field such that $a(\cdot,\omega) \in L^2(D)$ holds $\mathbb{P}$-a.s. and $a\colon \Omega \to L^2(D)$ is measurable. Then $a \in L^2(\Omega; L^2(D))$ if the mean field m and covariance function c of the random field a satisfy $m \in L^2(D)$ and $\int_D c(x,x)\,\mathrm{d}x < \infty$.

Moreover, if a is a Gaussian random field in any of the above cases, then it defines a Gaussian $L^p(D)$-random variable with p corresponding to the above cases.

We highlight that this time a converse statement as in Proposition 2.24 does, in general, not hold, i.e., an $L^p(D)$-valued random variable with $p \in [1,\infty]$ does not, in general, yield a random field, because pointwise evaluations are not defined. Even if a mapping $a\colon D \times \Omega \to \mathbb{R}$ is given and we have $a \in L^2(\Omega; L^p(D))$, $p \in [1,\infty]$, then $a(x,\cdot)\colon \Omega \to \mathbb{R}$ does not necessarily need to be measurable for each $x \in D$. However, if a is a Sobolev space-valued random variable, for instance, an $H_0^{d/2+\varepsilon}(D)$-valued random variable with $\varepsilon > 0$, then by Sobolev embedding theorems, see Gilbarg and Trudinger [72, Section 7.7], and Propositon 2.24 and 2.25 we obtain that a defines again a random field with $\mathbb{P}$-a.s. paths.

We now derive a representation for the covariance operator $\mathrm{Cov}(a)\colon L^2(D) \to L^2(D)$ of a second-order random field a which satisfies the assumptions of Proposition 2.26. For simplicity we assume a to be mean-zero and that the covariance function c of a satisfies $c \in L^2(D \times D)$. Then, by applying Fubini's theorem, we obtain for each $f, g \in L^2(D)$

$$\begin{aligned}
\mathrm{Cov}\left(\langle a, f\rangle_{L^2(D)}, \langle a, g\rangle_{L^2(D)}\right) &= \mathbb{E}\left[\int_D a(x,\omega)\, f(x)\,\mathrm{d}x \int_D a(y,\omega)\, g(y)\,\mathrm{d}y\right] \\
&= \int_D \int_D f(x)\mathbb{E}\left[a(x,\omega)\, a(y,\omega)\right] g(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_D \int_D f(x)\ \mathrm{Cov}(a(x), a(y))\, g(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_D f(x) \int_D \mathrm{Cov}(a(x), a(y))\, g(y)\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned}$$

Proposition 2.27. Let a be a second-order random field on $D \subset \mathbb{R}$ with mean m and covariance function c. If a is an $L^2(D)$-valued random variable and $m \in L^2(D)$ as well as $c \in L^2(D \times D)$, then the covariance operator of $a\colon \Omega \to L^2(D)$ is given by the integral operator $C\colon L^2(D) \to L^2(D)$ defined as

$$Cf(x) := \int_D c(x,y)\, f(y)\,\mathrm{d}y, \qquad f \in L^2(D). \tag{2.19}$$

Note, that $c \in L^2(D \times D)$ ensures that $Cf \in L^2(D)$ for $f \in L^2(D)$ which can be easily seen by applying the Cauchy-Schwarz inequality.

As an immediate consequence of Theorem 2.21 we get

Theorem 2.28 (Karhunen-Loève expansion (KLE))**.** Let a be a second-order random field on $D \subset \mathbb{R}$ satisfying the assumptions of Proposition 2.27. If (λ_m, ϕ_m), $m \in \mathbb{N}$, denote the eigenpairs of the integral operator given in (2.19) then there exist mutually uncorrelated mean zero random variables ξ_m, $m \in \mathbb{N}$, with unit variance such that

$$a(x,\omega) = \mathbb{E}[a(x)] + \sum_{m=1}^{\infty} \sqrt{\lambda_m}\phi_m(x)\xi_m(\omega) \tag{2.20}$$

holds in $L^2(\Omega; L^2(D))$ and $\mathbb{P}$-almost surely w.r.t. $\|\cdot\|_{L^2(D)}$. Moreover, if a is a Gaussian random field, then $\xi_m \sim N(0,1)$ i.i.d. .

A convenient property of the expansion (2.20) is that the spatial and random variations are separated.

Theorem 2.29 (Mercer's theorem, [114, Theorem 1.80 & Theorem 7.53])**.** Let a be a second-order random field on $\bar{D} \subset \mathbb{R}$ with $\mathbb{P}$-almost surely continuous paths. If its mean m and covariance function c are continuous, then the series given in (2.20) converges in $L^2(\Omega; C(\bar{D}))$ and $\mathbb{P}$-a.s. in $C(\bar{D})$, and there holds uniformly in x and y that

$$c(x,y) = \sum_{m=1}^{\infty} \lambda_m \phi_m(x)\phi_m(y), \qquad x, y \in \bar{D}.$$

Thus, under mild assumptions random fields can be represented via a Karhunen-Loève expansion. The other way round, i.e., defining a random field via such an expansion is also possible due to Proposition 2.22:

Proposition 2.30. Let $D \subset \mathbb{R}^d$, $\phi_m \in C(\bar{D})$ for $m \geq 0$ and $\xi_m \in L^2(\Omega; \mathbb{R})$ for $m \geq 1$ be stochastically independent and have zero mean and unit variance. If $(\|\phi_m\|_{C(\bar{D})})_{m\in\mathbb{N}} \in \ell^1(\mathbb{N})$, then by

$$a(x,\omega) = \phi_0(x) + \sum_{m=1}^{\infty} \phi_m(x)\,\xi_m(\omega) \qquad \forall x \in D,\ \mathbb{P}\text{-a.s.},$$

a second-order random field with $\mathbb{P}$-a.s. continuous paths, continuous mean and covariance function is given. Moreover, if the ξ_m are Gaussian, then a is a Gaussian random field.

Proof. That $a(\cdot,\omega) \in C(\bar{D})$ is well-defined and a $C(\bar{D})$-valued random variable follows by Proposition 2.22 which, moreover, also implies $a \in L^2(\Omega; C(\bar{D}))$. Thus, a is a second-order random field. The continuity of the mean is obvious and concerning

the covariance function we get, due to independence,

$$\begin{aligned}\operatorname{Cov}(a(x),a(y)) &= \operatorname{Cov}\left(\phi_0(x)+\sum_{m=1}^{\infty}\phi_m(x)\,\xi_m,\ \phi_0(y)+\sum_{m=1}^{\infty}\phi_m(y)\,\xi_m\right)\\ &= \sum_{m=1}^{\infty}\phi_m(x)\phi_m(y).\end{aligned}$$

The series on the right hand side defines a function in $C(D\times D)$, since for the functions $f_m(x,y) := \phi_m(x)\phi_m(y)$ there holds $f_m \in C(\bar{D}\times\bar{D})$ and

$$\sum_{m=1}^{\infty}\|f_m\|_{C(\bar{D}\times\bar{D})} = \sum_{m=1}^{\infty}\|\phi_m\|^2_{C(\bar{D})} < \infty.$$

The last assertion is obvious, since, e.g.,

$$\phi_0(x)+\sum_{m=1}^{M}\phi_m(x)\,\xi_m(\omega) \sim N\left(\phi_0(x),\sum_{m=1}^{M}\phi_m^2(x)\right).$$

□

Example 2.31 (Karhunen-Loève expansions of Brownian motion and bridge)**.** We recall the two Gaussian processes Brownian motion B and Brownian bridge BB introduced in Example 2.10. Since both processes are mean-zero with continuous covariance function and possess $\mathbb{P}$-a.s. continuous paths on finite intervals, they can be treated as $C([0,1])$- or $L^2([0,1])$-valued random variables. Their covariance operators on $L^2(0,1)$ are given by

$$C_B f(x) = \int_0^1 \min(x,y)\,f(y)\,\mathrm{d}y, \qquad C_{BB} f(x) = \int_0^1 (\min(x,y)-xy)\,f(y)\,\mathrm{d}y,$$

where $f \in L^2(D)$, respectively, which yields the following Karhunen-Loève expansions

$$\begin{aligned}B(x,\omega) &= \sum_{m=1}^{\infty}\frac{1}{(m+1/2)\pi}\,\sqrt{2}\,\sin((m+1/2)\pi x)\,\xi_m(\omega), \quad \xi_m \sim N(0,1)\text{ i.i.d.},\\ BB(x,\omega) &= \sum_{m=1}^{\infty}\frac{1}{m\pi}\,\sqrt{2}\,\sin(m\pi x)\,\xi_m(\omega), \qquad\qquad \xi_m \sim N(0,1)\text{ i.i.d.},\end{aligned}$$

see Lord et al. [114, Chapter 5]. Note, that due to $BB(x) = B(x) - xB(1)$ or $B(x) = BB(x) + xB(1)$, respectively, and

$$\operatorname{Cov}(B(1),BB(x)) = \operatorname{Cov}(B(1),B(x)) - x\operatorname{Cov}(B(1),B(1)) = 0,$$

we get another expansion for the Brownian motion on $[0,1]$ by

$$B(x,\omega) = x\xi_0(\omega) + \sum_{m=1}^{\infty} \frac{1}{m\pi}\,\sqrt{2}\,\sin(m\pi x)\,\xi_m(\omega), \qquad \xi_m \sim N(0,1) \text{ i.i.d.}$$

Remark 2.32 (Decay of eigenvalues for Matérn covariance functions)**.** As we have motivated in the discussion following Theorem 2.21, the series of the remaining eigenvalues $\sum_{m>M} \lambda_m$ quantifies the $L^2(\Omega; L^2(D))$-error of a truncated KLE of M terms. Thus, the faster the eigenvalues decay, the fewer terms we need for a "good" approximation of a second-order random field. In case of a random field with Matérn covariance function $c_{\sigma^2,\rho,\nu}$, as given in Example 2.8, the asymptotic decay rate is known to be

$$\lambda_m \leq C_{\sigma^2,\nu}\, m^{-\frac{d+2\nu}{d}},$$

see Widom [174] or Lord et al. [114, Example 7.59], where ν denotes the smoothness parameter of the Matérn covariance function and d the dimension of the domain $D \subset \mathbb{R}^d$. We mention that for $\nu \geq \frac{1}{2}$ a random field with continous mean and Matérn covariance function $c_{\sigma^2,\rho,\nu}$ is a $C(\bar{D})$- and, thus, $L^2(D)$-valued random variable by means of Proposition 2.26 and the discussion at the end of Section 2.1.

2.3. Elliptic Partial Differential Equations with Random Coefficients

Since we know how to treat random fields and function space-valued random variables, we will consider PDEs with random fields or function space-valued random variables as coefficients. The motivation behind these random PDEs is the in practice often incomplete knowledge of, e.g., material properties, boundary conditions or forcing terms which is modelled by constructing random fields describing the limited knowledge about the uncertain coefficients. The resulting solutions of such random PDEs are again function space-valued random variables which can be used to quantify the uncertainty about the state of the physical systems described by the corresponding PDE. In this thesis we focus on elliptic problems of second-order, namely, for a given bounded domain $D \subset \mathbb{R}^d$, we consider the BVP

$$-\nabla \cdot (a(x,\omega)\,\nabla u(x,\omega)) = f(x,\omega) \qquad \text{in } D, \quad \mathbb{P}\text{-a.s.}, \tag{2.21a}$$

$$p(x,\omega) = g(x,\omega) \qquad \text{on } \partial D, \quad \mathbb{P}\text{-a.s.}, \tag{2.21b}$$

with a, f and g random fields on D and ∂D, respectively. The differential operator ∇ acts w.r.t. the spatial variable x and we require the equations to hold $\mathbb{P}$-almost surely. For details on the underlying concepts and spaces concerning PDEs we refer to, e.g., Gilbarg and Trudinger [72]. For simplicity and without loss of generality, we will assume $g \equiv 0$ $\mathbb{P}$-a.s. in the following.

For numerical simulations of (2.21), particularly by the Galerkin method, see Remark 2.33 below, one often considers the *pathwise weak formulation* of (2.21) given by: find $u : \Omega \to H_0^1(D)$ such that

$$\langle a(\omega)\nabla u(\omega), \nabla v\rangle_{L^2(D)} = \langle f(\omega), v\rangle_{L^2(D)} \qquad \forall v \in H_0^1(D), \quad \mathbb{P}\text{-a.s.}, \tag{2.22}$$

where a is understood as an $L^\infty(D)$- and f as an $L^2(D)$-valued random variable.

Remark 2.33 (Galerkin method and finite elements)**.** Solving (2.22) numerically for a fixed realization ω is usually done by the *Galerkin method*: we construct a finite dimensional subspace of $V_h \subset H_0^1(D)$ and compute $u_h \in V_h$ which satisfies

$$\langle a(\omega)\nabla u_h(\omega), \nabla v\rangle_{L^2(D)} = \langle f(\omega), v\rangle_{L^2(D)} \qquad \forall v \in V_h.$$

This yields a system of linear equations for the coefficients of $u_h(\omega)$ w.r.t. a basis of V_h. If we increase the dimension of V_h then the error $\|u(\omega) - u_h(\omega)\|_{H_0^1(D)}$ decreases. In the *finite element Galerkin method* the finite dimensional subspace $V_h \subset H_0^1(D)$ is constructed via *finite elements*: we decompose D by, e.g., a triangular mesh of meshsize h into finitely many elements and define V_h as the span of (continuous) elementwise polynomials. We refer to Ern and Guermond [53] for more details.

The random variational problem (2.22) is well studied in the literature on PDEs with random data, see, e.g., [5, 6, 27]. Less often considered, but maybe more relevant in pratice, is the *mixed form* of (2.21)

$$a^{-1}(x,\omega)\,\boldsymbol{u}(x,\omega) = \nabla p(x,\omega) \qquad \text{in } D, \quad \mathbb{P}\text{-a.s.}, \tag{2.23a}$$

$$\nabla\cdot\,\boldsymbol{u}(x,\omega) = -f(x,\omega) \qquad \text{in } D, \quad \mathbb{P}\text{-a.s.}, \tag{2.23b}$$

$$p(x,\omega) = 0 \qquad \text{on } \partial D, \quad \mathbb{P}\text{-a.s.}, \tag{2.23c}$$

respectively, its weak form: find $(\boldsymbol{u}, p)\colon \Omega \to H(\mathrm{div}; D) \times L^2(D)$ such that

$$\langle a^{-1}(\omega)\boldsymbol{u}(\omega), \boldsymbol{v}\rangle_{L^2(D)} - \langle p(\omega), \nabla\cdot\,\boldsymbol{v}\rangle_{L^2(D)} = 0 \quad \forall \boldsymbol{v} \in H(\mathrm{div}; D),\ \mathbb{P}\text{-a.s.}, \tag{2.24a}$$

$$-\langle v, \nabla\cdot\,\boldsymbol{u}(\omega)\rangle_{L^2(D)} = \langle f(\omega), v\rangle_{L^2(D)} \quad \forall v \in L^2(D),\ \mathbb{P}\text{-a.s.}, \tag{2.24b}$$

where we require $\mathbb{P}$-a.s. that $a(x,\omega) > 0$ for each $\boldsymbol{x} \in D$ and introduce

$$H(\mathrm{div}; D) := \left\{ \boldsymbol{v} \in L^2(D;\mathbb{R}^d) : \nabla \cdot \boldsymbol{v} \in L^2(D;\mathbb{R}) \right\}. \tag{2.25}$$

The reason why (2.23) or (2.24), respectively, are favorable for pratical purposes is that the flux $\boldsymbol{u}$ is often the relevant quantity, e.g., when we consider transport of pollutants. By simulating and discretizing (2.24), by, e.g., Galerkin methods, we obtain immediately an approximation of $\boldsymbol{u}$. In the remainder of the chapter we will focus on (2.22), but revisit the mixed problem (2.24) in Chapter 7.

Concerning the existence of solutions to (2.22) or (2.24), we can apply the corresponing theory for deterministic variational problems pathwise. In case of (2.22) the key theoretical tool is the *Lax-Milgram lemma*

Lemma 2.34 (Lax-Milgram lemma [72, Theorem 5.8])**.** Let $B\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a continuous bilinear form on a Hilbert space $\mathcal{H}$ which is also *coercive*, i.e., there exists an $\alpha > 0$ such that

$$B(u,u) \geq \alpha \|u\|_{\mathcal{H}}^2 \qquad \forall u \in \mathcal{H}.$$

Then for each bounded linear form $F\colon \mathcal{H} \to \mathbb{R}$ there exists a unique $u^* \in \mathcal{H}$ such that

$$B(u^*, v) = F(v) \qquad \forall v \in \mathcal{H}$$

and there holds $\|u^*\|_{\mathcal{H}} \leq \|F\|/\alpha$.

Existence results of a pathwise solution to (2.24) can already be found in Babuška et al. [5] in case of a lower bound $a(x,\omega) \geq a_{\min} > 0$ for all $x \in D$ which holds $\mathbb{P}$-almost surely, and later in Charrier [27] without this assumption. By verifying the assumptions of Lemma 2.34 for ($\mathbb{P}$-almost) each ω, we can define the pathwise solution $u(\omega) \in H_0^1(D)$ of (2.22). Its continuous dependence on a and f yields measurability and by Hölder's inequality we easily obtain

Theorem 2.35 (cf. [27, Proposition 2.4])**.** Let $f \in L^p(\Omega; L^2(D))$ for a $p > 2$ and let a be an $L^\infty(D)$-valued random variable such that for

$$a_{\min}(\omega) := \operatorname*{ess\,inf}_{x \in D} a(x,\omega) \tag{2.26}$$

there holds $\mathbb{P}(a_{\min} > 0) = 1$ and $a_{\min}^{-1} \in L^q(\Omega;\mathbb{R})$ with $q > \frac{2p}{p-2}$, then (2.22) possesses a unique solution $u \in L^2(\Omega; H_0^1(D))$.

The same approach can be applied for the mixed problem (2.24). We refer, e.g., to Ernst and Sprungk [55] and the recent paper by Graham et al. [77]. For applications where the random field a models, e.g., an uncertain material conductivity,

a common modelling assumption is that a is a lognormal random field, see Definition 2.9. This ensures the $\mathbb{P}$-a.s. positiveness of $a(x)$, $x \in D$, and is often justified by measurement data, see, e.g., Freeze [64, p. 728] and the references therein. For these special random fields there holds

Proposition 2.36 ([27, Proposition 2.2])**.** Let a be a lognormal random field on $D \subset \mathbb{R}^d$ with $\mathbb{P}$-a.s. continuous paths. Then for $a_{\min}$ as in (2.26) there holds $a_{\min}^{-1} \in L^q(\Omega;\mathbb{R})$ for each $q \geq 1$. Thus, if $f \in L^p(\Omega;L^2(D))$ with $p > 1$, then (2.22) possesses a unique solution $u \in L^q(\Omega;H_0^1(D))$ for $q < p$.

Remark 2.37 (Monte Carlo FEM)**.** In Remark 2.33 the Galerkin and finite element method (FEM) for the pathwise numerical solution of (2.22) was outlined. If we are interested in certain moments such as the mean or the variance of the solution $u \in L^q(\Omega;H_0^1(D))$ of (2.22) or of functionals $Q\colon H_0^1(D) \to \mathbb{R}$ of u, we can apply the *Monte Carlo FEM* as described in Babuška et al. [5]. To this end, we generate samples $a(\omega_n)$ and $f(\omega_n)$, $n = 1,\dots,N$, of the random fields (or random variables) a and f, respectively, cf. Remark 2.11, and compute the corresponding Monte Carlo estimate of, e.g., $\mathbb{E}[Q(u)]$

$$E_Q(h,N) := \frac{1}{N}\sum_{n=1}^{N} Q\left(u_h(\omega_n)\right).$$

For the resulting mean squared error there holds typically

$$\mathbb{E}\left[\left|E_Q(h,N) - \mathbb{E}\left[Q(u)\right]\right|^2\right] \leq C_{Q,u}\left(N^{-1} + h^{-r}\right),$$

where the decay rate r of the FEM error depends on the choice of the underlying finite elements. We refer to Babuška et al. [5] for more details. In recent years, several improvements of the basic Monte Carlo FEM have been developed. We mention the *multilevel Monte Carlo FEM* (MLMC FEM), see, e.g., Cliffe et al. [32] and Teckentrup et al. [169], which combines independent MC FEM estimates obtained for several mesh sizes h_ℓ, $\ell = 0,\dots,L$, in order to reduce the variance of the resulting MLMC FEM estimate, and the *Quasi Monte Carlo FEM* (QMC FEM), see, e.g., Graham et al. [76], which employs Quasi Monte Carlo sampling instead of plain Monte Carlo sampling and in this way obtains higher convergence rates w.r.t. N.

Besides the Monte Carlo approaches described in the previous remark to compute expectations of functionals of the solution u of (2.22), there now exist also numerous numerical methods to approximate the random variable u in $L^2(\Omega;H_0^1(D))$. These methods rely particularly on the fact that we can represent random fields and

Hilbert space-valued random variables by (abstract) Karhunen-Loève expansions which yield a more convenient parametric reformulation of the random PDE (2.22).

2.3.1. Parametric Reformulation

The assumptions of Theorem 2.35 and Proposition 2.36 allow us to apply Theorem 2.28 and 2.29 to a and f, i.e., to represent these $\mathbb{P}$-a.s. by their Karhunen-Loève expansions. Of course, in general, the random variables appearing in the KLE of a and f are different ones, but to ease notation and also w.l.o.g. we make the following

Assumption 2.38. There exist stochastically independent $\xi_m \in L^2(\Omega;\mathbb{R})$ for $m \in \mathbb{N}$ as well as $\phi_m \in L^\infty(D)$ and $\psi_m \in L^2(D)$ for $m \in \mathbb{N}_0$ such that $\mathbb{P}$-a.s. there holds in $L^\infty(D)$ and $L^2(D)$, respectively,

$$\log a(x,\omega) = \phi_0(x) + \sum_{m=1}^{\infty} \phi_m(x)\xi_m(\omega), \quad f(x,\omega) = \psi_0(x) + \sum_{m=1}^{\infty} \psi_m(x)\xi_m(\omega). \tag{2.27}$$

Let μ_m denote the distribution of the random variables $\xi_m \sim \mu_m$, $m \in \mathbb{N}$, appearing in Assumption 2.38. Then for the random sequence $\boldsymbol{\xi}\colon \Omega \to \mathbb{R}^{\mathbb{N}}$ given by

$$\boldsymbol{\xi}(\omega) := (\xi_m(\omega))_{m\in\mathbb{N}}, \qquad \omega \in \Omega,$$

there holds $\boldsymbol{\xi} \sim \mu$ with μ given as the following product measure on the product measurable space $(\mathbb{R}^{\mathbb{N}}, \bigotimes_{m\geq 1} \mathcal{B}(\mathbb{R}))$:

$$\mu(\mathrm{d}\boldsymbol{\xi}) := \bigotimes_{m=1}^{\infty} \mu_m(\mathrm{d}\xi_m). \tag{2.28}$$

For more details about the notation of measures used in this thesis, we refer the reader to Section 3.1. Further, we note that $\boldsymbol{\xi}\colon (\Omega,\mathcal{A}) \to (\mathbb{R}^{\mathbb{N}}, \bigotimes_{m\geq 1} \mathcal{B}(\mathbb{R}))$ as given above is measurable by construction.

Thus, concerning the random fields (or random variables, respectively) a and f, we can reformulate them – with a slight abuse of notation – as mappings of $\boldsymbol{\xi}$ rather than ω:

$$\log a(x,\boldsymbol{\xi}) = \phi_0 + \sum_{m=1}^{\infty} \phi_m(x)\xi_m, \quad f(x,\boldsymbol{\xi}) = \psi_0 + \sum_{m=1}^{\infty} \psi_m(x)\xi_m, \quad \mu\text{-a.e.} \tag{2.29}$$

In other words, by the change of variables $\omega \mapsto \boldsymbol{\xi}(\omega)$ we switch from the abstract

probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the more convenient product probability space

$$\left(\mathbb{R}^{\mathbb{N}}, \bigotimes_{m\geq 1} \mathcal{B}(\mathbb{R}), \mu\right) = \bigotimes_{m=1}^{\infty} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_m). \tag{2.30}$$

Concerning Lebesgue–Bochner spaces for this new probability space we will use the notation

$$L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{X}) := \Big\{ f\colon \mathbb{R}^{\mathbb{N}} \to \mathcal{X} \,\Big|\, f \text{ is strongly measurable and } \|f\|^2_{L^2_\mu} := \int_{\mathbb{R}^{\mathbb{N}}} \|f(\boldsymbol{\xi})\|^2_{\mathcal{X}}\, \mu(\mathrm{d}\boldsymbol{\xi}) < \infty \Big\}$$

where $\mathcal{X}$ denotes again an arbitrary Banach space. It follows that if a and f given by (2.27) belong to the Lebesgue–Bochner spaces $L^2(\Omega; L^\infty(D))$ and $L^2(\Omega; L^2(D))$, respectively, then a and f given in (2.29) belong to $L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^\infty(D))$ and $L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))$, respectively, and vice versa.

Assume that a and f given in Assumption 2.38 satisfy the assumptions of Theorem 2.35, then the pathwise solution $u\colon \Omega \to H^1_0(D)$ of (2.22) exists. Let $S\colon L^\infty(D) \times L^2(D) \to H^1_0(D)$ denote the solution operator associated to the (deterministic) variational problem (2.22), i.e., $u(\omega) = S(a(\omega), f(\omega))$ $\mathbb{P}$-almost surely. By virtue of the Lax-Milgram lemma, the mapping S is continuous and, therefore, measurable. Hence, also the pathwise solution u can be represented by a measurable mapping of $\boldsymbol{\xi}$, i.e., there exists a measurable mapping $\hat{u}\colon \mathbb{R}^{\mathbb{N}} \to H^1_0(D)$ such that we have $u(\omega) = \hat{u}(\boldsymbol{\xi}(\omega))$ $\mathbb{P}$-almost surely. This mapping $\hat{u}$ coincides with the solution of the *parametric variational problem*: find $u\colon \mathbb{R}^{\mathbb{N}} \to H^1_0(D)$ such that

$$\langle a(\boldsymbol{\xi})\nabla u(\boldsymbol{\xi}), \nabla v\rangle_{L^2(D)} = \langle f(\boldsymbol{\xi}), v\rangle_{L^2(D)} \qquad \forall v \in H^1_0(D), \quad \mu\text{-a.e.} \tag{2.31}$$

We consider (2.31) as the *parametric reformulation* of (2.22) given Assumption 2.38. We state a corresponding existence result for (2.31) adapted from Bachmayr et al. [6] where the authors considered only deterministic forcing terms f in (2.31).

Theorem 2.39 (cf. [6, Theorem 2.1 & Corollary 2.1])**.** Let Assumption 2.38 be satisfied with $\xi_m \sim N(0,1)$ if $\phi_m \not\equiv 0$. Further assume that there exists a sequence $(\rho_m)_{m\in\mathbb{N}}$ of positive numbers $\rho_m > 0$ and a $p > 0$ such that

$$\sum_{m\geq 1} \exp(-\rho_m^2) < \infty, \quad \sum_{m\geq 1} \rho_m \|\phi_m\|_{L^\infty(D)} < \infty, \quad \sum_{m\geq 1} \|\psi_m\|^p_{L^2(D)} \operatorname{Var}(\xi_m)^p < \infty.$$

Then, there exists a unique solution u of (2.31) and $u \in L^q_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ for $q < p$.

The assumptions of Theorem 2.39 imply, in particular, that a is a lognormal random field with $a \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ and that $f \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))$, but f need not necessarily be a Gaussian random field.

Remark 2.40 (On truncation errors)**.** For numerical simulations we have to truncate the expansions given in (2.27) after, say, $M \in \mathbb{N}$ terms. Let us denote the resulting random fields by a_M and f_M, respectively. Then the solution $u_M \colon \mathbb{R}^{\mathbb{N}} \to H^1_0(D)$ of

$$\langle a_M(\boldsymbol{\xi}) \nabla u_M(\boldsymbol{\xi}), \nabla v \rangle_{L^2(D)} = \langle f_M(\boldsymbol{\xi}), v \rangle_{L^2(D)} \qquad \forall v \in H^1_0(D), \quad \mu\text{-a.e.} \tag{2.32}$$

differs from the solution u of (2.31). In Charrier [27] and Charrier and Debussche [28] the error $u - u_M$ has been investigated in case of a deterministic f and lognormal a. Under suitable assumptions on the ϕ_m in (2.27) they were able to estimate the error $u - u_M$ in $L^p_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ by certain functionals of the remainder term $a - a_M$, i.e., for $M \to \infty$ they showed $u_M \to u$ in $L^p_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ for any $p \geq 1$. We refer to their works for details.

The parametric problem (2.31) allows now for approximation methods, since we deal with functions depending on a parameter $\boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}}$ rather than with random fields depending on an abstract ω. In the next section we provide a brief overview on existing approximation methods for the solution of parametric elliptic PDEs in the form (2.31).

2.3.2. Approximation Methods

In this section we outline some basic methods to approximate objects in $L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ with $\mathcal{H}$ a separable Hilbert space and μ as in (2.28), for example, the solution $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ of the (random) parametric variational problem (2.31). We will focus on approximating the dependence on $\boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}}$ rather than spatial approximations in $\mathcal{H}$ or $H^1_0(D)$, respectively, because the latter are well-known, see Remark 2.33.

Since $L^2_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ is again a separable Hilbert space, there exists a CONS $\{\varphi_n : n \in \mathbb{N}\}$ of $L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathbb{R})$ and we can represent each $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ by

$$u(\boldsymbol{\xi}) = \sum_{n=1}^{\infty} u_n \varphi_n(\boldsymbol{\xi}), \tag{2.33}$$

where the equality holds in $L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ and

$$u_n := \int_{\mathbb{R}^{\mathbb{N}}} u(\boldsymbol{\xi})\, \varphi_n(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}) \in \mathcal{H}. \tag{2.34}$$

In order to contruct a CONS of the space $L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathbb{R})$, we can exploit once more the underlying product structure (2.30) of the probability space $(\mathbb{R}^{\mathbb{N}}, \bigotimes_{m\geq 1}\mathcal{B}(\mathbb{R}), \mu)$ and obtain

Theorem 2.41 ([156, Theorem 2.12])**.** For each $m \in \mathbb{N}$ let $\{\varphi_n^{(m)} : n \in \mathbb{N}_0\}$ denote a CONS of $L^2_{\mu_m}(\mathbb{R};\mathbb{R})$ with $\varphi_0^{(m)} \equiv 1$. For each multiindex $\boldsymbol{\alpha} = (\alpha_m)_{m\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ define

$$|\boldsymbol{\alpha}|_0 := |\{j \in \mathbb{N} \colon \alpha_j > 0\}| \tag{2.35}$$

and set

$$\mathcal{F} := \{\boldsymbol{\alpha} \in \mathbb{N}_0^{\mathbb{N}} \colon |\boldsymbol{\alpha}|_0 < \infty\}. \tag{2.36}$$

Then $\{\boldsymbol{\varphi}_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}\}$ is a CONS of $L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathbb{R})$ with μ as in (2.28) and

$$\boldsymbol{\varphi}_{\boldsymbol{\alpha}} := \prod_{m\geq 1} \varphi_{\alpha_m}, \qquad \boldsymbol{\alpha} \in \mathcal{F}.$$

We note, that for $\boldsymbol{\alpha} \in \mathcal{F}$ the function $\boldsymbol{\varphi}_{\boldsymbol{\alpha}}$ is actually given by a finite product, since $|\boldsymbol{\alpha}|_0 < \infty$ and $\varphi_0^{(m)} \equiv 1$. A common choice of orthonormal systems in the spaces $L^2_{\mu_m}(\mathbb{R};\mathbb{R})$ are the orthogonal polynomials w.r.t. the measure μ_m. However, they need not be complete in $L^2_{\mu_m}(\mathbb{R};\mathbb{R})$, see Ernst et al. [54] for a discussion. For many common cases, such as $\mu_m = N(c,\sigma^2)$ or $\mu_m = U[a,b]$, the corresponding orthogonal polynomials, i.e., *Hermite* or *Legendre* polynomials, respectively, form a CONS of $L^2_{\mu_m}(\mathbb{R};\mathbb{R})$. In the following we will focus on the standard Gaussian case, i.e., $\mu_m = N(0,1)$ for $m \in \mathbb{N}$, and work with Hermite polynomials. For details on the latter we refer to Szegő [168] and Gautschi [66].

Definition 2.42 ((Wiener-Hermite) Polynomial chaos expansion)**.** Let $\mu_m = N(0,1)$ for $m \in \mathbb{N}$ and $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})$ with μ as in (2.28). Moreover, let $H_n \colon \mathbb{R} \to \mathbb{R}$ denote the L^2-normalized Hermite polynomial of degree $n \in \mathbb{N}_0$ w.r.t. the measure $N(0,1)$ on $\mathbb{R}$ and define for $\boldsymbol{\alpha} \in \mathcal{F}$ with $\mathcal{F}$ as in (2.36) the multivariate Hermite polynomial $\boldsymbol{H}_{\boldsymbol{\alpha}} \colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ by $\boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) := \prod_{m\geq 1} H_{\alpha_m}(\xi_m)$. Then the expansion

$$u(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}\in\mathcal{F}} u_{\boldsymbol{\alpha}} \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \qquad u_{\boldsymbol{\alpha}} := \int_{\mathbb{R}^{\mathbb{N}}} u(\boldsymbol{\xi})\, \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}), \tag{2.37}$$

which converges in $L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})$ is called *(Wiener-Hermite) polynomial chaos expansion* (PCE) of u and the $u_{\boldsymbol{\alpha}} \in \mathcal{H}$ are called *(polynomial) chaos coefficients* of u.

The term "Wiener-Hermite" relates, besides to Charles Hermite, to Norbert Wiener, who introduced such polynomial expansions in his work [175]. More recently, Xiu and Karniadakis [178] introduced the notion *generalized polynomial chaos expan-*

sion for expansions such as (2.37) but based on orthogonal polynomials w.r.t. non-Gaussian probability measures μ_m.

We will focus in the following on numerical methods which yield approximations $\hat{u}_F$ of $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ of the form

$$\hat{u}_F(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha} \in F} \hat{u}_{\boldsymbol{\alpha}} \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \qquad F \subset \mathcal{F},\ |F| < \infty, \tag{2.38}$$

where $\hat{u}_{\boldsymbol{\alpha}} \in \mathcal{H}$ are approximations of the chaos coefficients $u_{\boldsymbol{\alpha}}$ of u since the latter are typically unknown or not exactly computable. The error $u - \hat{u}_F$ in $L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ is then given by

$$\|u - \hat{u}_F\|^2_{L^2_\mu} = \sum_{\boldsymbol{\alpha} \in F} \|u_{\boldsymbol{\alpha}} - \hat{u}_{\boldsymbol{\alpha}}\|^2_{\mathcal{H}} + \sum_{\boldsymbol{\alpha} \in \mathcal{F} \setminus F} \|u_{\boldsymbol{\alpha}}\|^2_{\mathcal{H}}. \tag{2.39}$$

Thus, for proving convergence rates of such approximations the decay of the chaos coefficients $\|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}}$ plays an important role. As in classical Fourier analysis, there is a relation between the smoothness of u w.r.t. $\boldsymbol{\xi}$ and the rate of decay of $\|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}}$ which we will discuss later on. In the following we outline several common numerical methods for computing such approximations of $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$.

Best N-term approximations. Let $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$, then following DeVore [44, Section 2] an *N-term (polynomial chaos) approximation* of u is a function $\hat{u}_F$ as given in (2.38) with $|F| = N$. Let us denote the set of all such functions by $S_N \subset L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$, i.e.,

$$S_N := \left\{ \sum_{\boldsymbol{\alpha} \in F} v_{\boldsymbol{\alpha}} \boldsymbol{H}_{\boldsymbol{\alpha}} : v_{\boldsymbol{\alpha}} \in \mathcal{H},\ |F| \le N \right\}.$$

Then the *best N-term (polynomial chaos) approximation* u^*_N of u is defined as

$$u^*_N := \operatorname*{argmin}_{v \in S_N} \|u - v\|_{L^2_\mu}.$$

Obviously, see (2.39), u^*_N is given by

$$u^*_N(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha} \in F^*_N} u_{\boldsymbol{\alpha}} \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$$

where F^*_N denotes the set of the multiindices $\boldsymbol{\alpha} \in \mathcal{F}$ corresponding to the N largest $\|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}}$. The resulting error is

$$\|u - u^*_N\|_{L^2_\mu} = \left(\sum_{\boldsymbol{\alpha} \in \mathcal{F} \setminus F^*_N} \|u_{\boldsymbol{\alpha}}\|^2_{\mathcal{H}} \right)^{1/2}$$

and convergence rates can be established via *Stechkin's lemma*, see Cohen and DeVore [33, Lemma 3.6]: let $(\|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}})_{\boldsymbol{\alpha}\in\mathcal{F}} \in \ell^p(\mathcal{F})$ with $p < 2$, then

$$\left(\sum_{\boldsymbol{\alpha}\in\mathcal{F}\setminus F_N^*} \|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}}^2\right)^{1/2} \leq C\, N^{-\frac{1}{p}+\frac{1}{2}}.$$

For further details on best N-term approximations on solutions u of parametric PDEs and variational problems such as (2.31) we refer to Cohen et al. [34], Hoang and Schwab [88] and Bachmayr et al. [7, 6]. Of course the Hermite coefficients $u_{\boldsymbol{\alpha}}$ are usually unknown, thus, best N-term approximations are usually not computable in practice. However, their convergence rates provide an upper bound for the convergence rate of other numerical methods.

Stochastic Galerkin methods. One way to approximate the chaos coefficients of the solution $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ of the parametric variational problem (2.31) are Galerkin methods. To this end, we test the parametric variational equation

$$\langle a(\boldsymbol{\xi})\nabla u(\boldsymbol{\xi}), \nabla v\rangle_{L^2(D)} = \langle f(\boldsymbol{\xi}), v\rangle_{L^2(D)} \qquad \forall v \in H^1_0(D) \quad \mu\text{-a.e.}$$

with the real-valued multivariate Hermite polynomials $\boldsymbol{H}_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in \mathcal{F}$, and, hence, obtain the following stochastic variational formulation:

$$\int_{\mathbb{R}^{\mathbb{N}}} \langle a(\boldsymbol{\xi})\nabla u(\boldsymbol{\xi}), \nabla v\rangle_{L^2(D)}\, \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}) = \int_{\mathbb{R}^{\mathbb{N}}} \langle f(\boldsymbol{\xi}), v\rangle_{L^2(D)}\, \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}) \tag{2.40}$$

for all $v \in H^1_0(D)$ and all $\boldsymbol{\alpha} \in \mathcal{F}$. We notice that on both sides of (2.40) the inner product in $L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))$ appears. In order to ensure that the integrals or inner products, respectively, in (2.40) are finite, we require that $f \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))$ and $a\nabla u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))$. The latter is, for instance, satisfied if $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D))$ and if there exists a finite constant K such that μ-a.e. $\|a(\boldsymbol{\xi})\|_{L^\infty(D)} \leq K$. Since the Hermite polynomials $\{\boldsymbol{H}_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}\}$ form a CONS of $L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathbb{R})$, the formulation (2.40) motivates the following *stochastic variational problem*: given the energy space

$$\mathcal{V}_a := \left\{ v \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; H^1_0(D)) \colon a\nabla v \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D)) \right\}$$

equipped with the inner product $\langle u, v\rangle_a := \langle a\nabla u, \nabla v\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))}$, find $u \in \mathcal{V}_a$ such that

$$\langle a\nabla u, \nabla v\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))} = \langle f, v\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^2(D))} \qquad \forall v \in \mathcal{V}_a. \tag{2.41}$$

If there exists a lower bound $a(x,\boldsymbol{\xi}) \geq a_{\min} > 0$, then $\mathcal{V}_a$ is again a Hilbert space which is continuously embedded in $L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathbb{R})$ due to

$$\|v\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}};H^1_0(D))} \leq \frac{1}{a_{\min}}\|v\|_a \qquad \forall v \in \mathcal{V}_a,$$

where $\|v\|_a := \langle v, v\rangle_a$. Moreover, in this case the Lax-Milgram lemma can be applied to show the existence of a unique solution $u \in \mathcal{V}_a \subset L^2_\mu(\mathbb{R}^{\mathbb{N}};H^1_0(D))$ of the stochastic variational problem (2.41) which will coincide with the pathwise solution of (2.31) whenever the latter belongs to $\mathcal{V}_a$.

However, in case of a lognormal random field a, i.e., when a satisfies Assumption 2.38, then there exists no deterministic lower bound $a(x,\boldsymbol{\xi}) \geq a_{\min} > 0$ which leads to several technical difficulties. For instance, the bilinear form $\langle a\nabla\cdot, \nabla\cdot\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}};L^2(D))}$ appearing on the left hand side of (2.41) is then not coercive, i.e., there exists no constant $\alpha > 0$ such that

$$\langle a\,\nabla v, \nabla v\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}};L^2(D))} \geq \alpha\,\|v\|_a^2 \qquad \forall v \in \mathcal{V}_a,$$

and, thus, the Lax-Milgram lemma can not be applied. However, by a suitable change of measure $\mu \mapsto \nu$, a theory for stochastic variational problems can be established. We refer to Galvis and Sarkis [65], Gittelson [74] and Mugler and Starkloff [121] for more details.

Given there exists a solution $u \in \mathcal{V}_a$ of (2.41), we can again apply the Galerkin method to approximate it. To this end, let us assume that $v\boldsymbol{H}_{\boldsymbol{\alpha}} \in \mathcal{V}_a$ for each $v \in H^1_0(D)$ and $\boldsymbol{\alpha} \in \mathcal{F}$. Then, for each finite-dimensional subspace $\mathcal{V}_F \subset L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathbb{R})$, e.g.,

$$\mathcal{V}_F := \operatorname{span}(\boldsymbol{H}_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in F), \qquad F \subset \mathcal{F},\ |F| < \infty,$$

and each a finite subspace of $V_h \subset H^1_0(D)$, there holds

$$V_h \otimes \mathcal{V}_F = \left\{ \sum_{\boldsymbol{\alpha}\in F} u_{\boldsymbol{\alpha},h}\boldsymbol{H}_{\boldsymbol{\alpha}} : u_{\boldsymbol{\alpha},h} \in V_h \text{ for all } \boldsymbol{\alpha} \in F \right\} \subset \mathcal{V}_a$$

and we can compute the solution $\hat{u}_{F,h} \in V_h \otimes \mathcal{V}_F$ of the discretized stochastic variational problem

$$\langle a\nabla\hat{u}_{F,h}, \nabla v\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}};L^2(D))} = \langle f, v\rangle_{L^2_\mu(\mathbb{R}^{\mathbb{N}};L^2(D))} \qquad \forall v \in V_h \otimes \mathcal{V}_F. \tag{2.42}$$

This leads again to a finite dimensional linear system for the finitely many chaos coefficients of $\hat{u}_{F,h}$ represented in a basis of V_h which can then be solved numerically.

A comprehensive introduction to the stochastic Galerkin method is given, e.g., by Ghanem and Spanos [70] and by Le Maitre and Knio [108].

Collocation methods. Another way to approximate a given mapping $u\colon \mathbb{R}^{\mathbb{N}} \to \mathcal{H}$, $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$, is by *sparse grid collocation methods* which are based on Lagrange interpolation. To this end, let us introduce $\mathcal{I}_n$ as the univariate Lagrange interpolation operator based on $n+1$ distinct nodes $\Xi_n := \{\xi_{0,n}, \xi_{1,n}, \dots, \xi_{n,n}\}$, $\Xi_n \subset \mathbb{R}$, i.e.,

$$(\mathcal{I}_n u)(\xi) = \sum_{i=0}^{n} u(\xi_{i,n})\, L_{i,n}(\xi), \qquad u\colon \mathbb{R} \to \mathbb{R},$$

where $L_{i,n}$ denotes the canonical ith Lagrange polynomial of degree n associated to the nodes Ξ_n. For a multiindex $\boldsymbol{\alpha} \in \mathcal{F}$ we can then define the tensorized operator $\mathcal{I}_{\boldsymbol{\alpha}} := \bigotimes_{m\in\mathbb{N}} \mathcal{I}_{\alpha_m}$, i.e.,

$$(\mathcal{I}_{\boldsymbol{\alpha}} u)(\boldsymbol{\xi}) = \sum_{\boldsymbol{i}:\, i_m \le \alpha_m} u(\boldsymbol{\xi}_{\boldsymbol{i},\boldsymbol{\alpha}}) L_{\boldsymbol{i},\boldsymbol{\alpha}}(\boldsymbol{\xi}), \qquad u : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}, \tag{2.43}$$

where $L_{\boldsymbol{i},\boldsymbol{\alpha}}(\boldsymbol{\xi}) := \prod_{m\ge 1} L_{i_m,\alpha_m}(\xi_m)$ and $\boldsymbol{\xi}_{\boldsymbol{i},\boldsymbol{\alpha}} = (\xi_{i_m,\alpha_m})_{m\in\mathbb{N}}$. Note, that $L_{\boldsymbol{i},\boldsymbol{\alpha}}$ is well-defined due to $\boldsymbol{\alpha} \in \mathcal{F}$ and $L_{0,0} \equiv 1$. Moreover, the set of all *grid points* $\boldsymbol{\xi}_{\boldsymbol{i},\boldsymbol{\alpha}}$ is then given by

$$\Xi_{\boldsymbol{\alpha}} := \bigtimes_{m\in\mathbb{N}} \Xi_{\alpha_m}. \tag{2.44}$$

Again, due to $\boldsymbol{\alpha} \in \mathcal{F}$ and $|\Xi_0| = 1$ the set $\Xi_{\boldsymbol{\alpha}}$ is finite with $|\Xi_{\boldsymbol{\alpha}}| = \prod_{m\in\mathbb{N}} |1+\alpha_m|$. The resulting interpolating approximation $\mathcal{I}_{\boldsymbol{\alpha}} u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ for an $\mathcal{H}$-valued mapping $u\colon \mathbb{R}^{\mathbb{N}} \mapsto \mathcal{H}$ is called *full grid collocation approximation* of u and corresponds again to an approximation of the form (2.38). Full grid collocation methods for approximating solutions of parametric PDEs such as (2.31) were introduced by Xiu and Hesthaven [177] and later analyzed by, e.g., Babuška et al. [4]. The setting in both cases was a finite dimensional one, i.e., they considered finite expansions for the random or parametric coefficients in the PDE and, thus, the corresponding solution u depended only on finitely many $\xi_1, \dots, \xi_N$. In general, full grid collocation is not suited for high-dimensional (or infinite-dimensional) approximation due to the fast growth of $|\Xi_{\boldsymbol{\alpha}}|$ for increasing $\boldsymbol{\alpha}$.

Based on sparse grids and hierarchical approximations, see Bungartz and Griebel [23], Nobile et al. [125, 124] introduced *sparse grid collocation approximations*, again in finite dimensions. These methods were later extended to the infinite dimensional setting by Chkifa et al. [30, 31] and they also established convergence results for sparse grid collocation approximations of functions $u\colon [-1,1]^{\mathbb{N}} \to \mathcal{H}$. Sim-

ilar results were recently obtained by Ernst et al. [58] for the unbounded case, i.e., for sparse grid collocation applied to approximate functions $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ such as the solution u to the parametric (lognormal) elliptic BVP (2.31). We briefly outline the idea of sparse grid collocation and refer to the publications above for more details and motivation. First, we introduce the univariate *detail operators* $\Delta_n := \mathcal{I}_n - \mathcal{I}_{n-1}$ for $n > 0$ and set $\Delta_0 := \mathcal{I}_0$. Analogously to $\mathcal{I}_{\boldsymbol{\alpha}}$ the tensorized operator $\Delta_{\boldsymbol{\alpha}} := \bigotimes_{m\in\mathbb{N}} \Delta_{\alpha_m}$ can be defined for $\boldsymbol{\alpha} \in \mathcal{F}$. Then, given a finite subset $F \subset \mathcal{F}$ we set

$$\mathcal{S}_F u := \sum_{\boldsymbol{\alpha}\in F} \Delta_{\boldsymbol{\alpha}} u, \qquad u : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}, \tag{2.45}$$

and call $\mathcal{S}_F u$ a *sparse grid collocation approximation* of u. For computational as well as analytical reasons, it is often assumed that the finite set F is *monotone*, i.e., for each $\boldsymbol{\alpha} \in F$ also $\boldsymbol{\alpha} - \boldsymbol{e}_m \in F$, $m \in \mathbb{N}$, where $\boldsymbol{e}_m$ denotes the mth unit vector in $\mathbb{R}^{\mathbb{N}}$. Given this assumption on F the approximation $\mathcal{S}_F u$ can be written in the form (2.38) and the corresponding set F in (2.38) is the same as in (2.45), see Ernst et al. [58, Section 2.1]. The advantage of the sparse grid collocation is that we are more flexible w.r.t. refining the approximation: for the full grid collocation approximation $I_{\boldsymbol{\alpha}} u$ an increase in one of the components of $\boldsymbol{\alpha}$ yields a possibly dramatic increase in the number of grid points $|\Xi_{\boldsymbol{\alpha}}| = \prod_{m\in\mathbb{N}} |1 + \alpha_m|$, whereas we refine $\mathcal{S}_F u$ by adding a new multiindex $\boldsymbol{\alpha} \in \mathcal{F}$ to F which under the condition that $F \cup \{\boldsymbol{\alpha}\}$ is again monotone yields only a moderate growth of the associated *sparse grid*

$$\Xi_F := \bigcup_{\boldsymbol{\alpha}\in F} \Xi_{\boldsymbol{\alpha}}, \qquad |\Xi_F| \le \sum_{\boldsymbol{\alpha}\in F} |\Xi_{\boldsymbol{\alpha}}|. \tag{2.46}$$

For example, for *activating* a dimension, say, n, i.e., allowing the collocation approximation to depend (non-constantly) on ξ_n , we have to increase in case of a full grid collocation approximation $\mathcal{I}_{\boldsymbol{\alpha}} u$ the component α_n from 0 to 1 which immediately doubles the number of grid points in the associated full grid. On the other hand, for a sparse grid collocation approximation $\mathcal{S}_F u$ we just have to add the multiindex $\boldsymbol{\alpha} = \boldsymbol{e}_n$ to the set F which increases the number grid points in the associated sparse grid at most by $|\Xi_{e_n}| = 2$. Moreover, given some regularity of the function of interest $u\colon \mathbb{R}^{\mathbb{N}} \to \mathcal{H}$, convergence of $\mathcal{S}_F u$ to u in $L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ can be shown. In particular, Ernst et al. [58, Theorem 19] established under some regularity assumptions on $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ – which were motivated by the results of Bachmayr et al. [6] – that for sparse grid collocation based on Gauss-Hermite nodes, there exists a sequence

of nested, monotone subsets $F_n \subset \mathcal{F}$, $n \in \mathbb{N}$, with $|F_n| = n$ such that

$$\|u - \mathcal{S}_{F_n} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})} \leq C_u \, |\Xi_{F_n}|^{-s},$$

where the constant C_u depends on u and the rate $s > 0$ on the regularity of u, cf. the next paragraph.

Regularity w.r.t. ξ and decay of chaos coefficients. In classical harmonic analysis, smoothness or regularity of a function $f\colon [-\pi, \pi) \to \mathbb{R}$ relates to the decay rate of its Fourier coefficients, i.e., the order of weak differentiability corresponds to the algebraic decay of the coefficients. Similar results hold for the decay of Hermite coefficients of functions $f\colon \mathbb{R} \to \mathbb{R}$. Perhaps the first one to obtained results on the decay of Hermite coefficients was Hille [85, 86]. He considered functions which are anlytic in a strip $\{z \in \mathbb{C}\colon |\Im(z)| < \tau\}$ of the complex plane and established necessary and sufficient conditions for the pointwise convergence of Hermite expansions within this strip. If these assumptions are fulfilled then the Hermite coefficients $f_n \in \mathbb{C}$, $n \in \mathbb{N}_0$, of $f\colon \mathbb{C} \to \mathbb{C}$ decay like $\exp(-(\tau + \varepsilon)\sqrt{2n+1})$ where $\varepsilon > 0$ is arbitrary. Hille's result was later extended and refined by Boyd [20, 21] who showed that also the order of decay of f on the real line influences the decay of its Hermite coefficients. However, to the author's knowledge there exist so far no extensions of Hille's or Boyd's work to analytic functions of several variables.

Concerning finite differentiability, we get similar results as for classical Fourier coefficients which in this case can be extended to the case of countably many variables as shown by Bachmayr et al. [6]: Let $f\colon \mathbb{R} \to \mathbb{R}$ have a weak derivative $f^{(k)} \in L^2_{\mu_1}(\mathbb{R})$ with $\mu_1 = N(0,1)$ and $k \geq 0$. Then by the *Rodrigues' formula* for Hermite polynomials, see Abramowitz and Stegun [1, Section 22.11],

$$H_n(\xi) = \frac{(-1)^n}{\sqrt{n!}} \, \mathrm{e}^{\xi^2/2} \frac{\mathrm{d}^n}{\mathrm{d}\xi^n} \, \mathrm{e}^{-\xi^2/2}, \qquad n \in \mathbb{N}_0,$$

and integration by parts, we obtain for the Hermite coefficient f_n, $n > k$, of f

$$\begin{aligned} f_n &= \int_{\mathbb{R}} f(\xi) H_n(\xi) \, \mu_1(\mathrm{d}\xi) = \int_{\mathbb{R}} f(\xi) H_n(\xi) \, \mathrm{e}^{-\xi^2/2} \frac{\mathrm{d}\xi}{\sqrt{2\pi}} \\ &= \sqrt{\frac{(n-k)!}{n!}} \int_{\mathbb{R}} f^{(k)}(\xi) H_{n-k}(\xi) \, \mathrm{e}^{-\xi^2/2} \frac{\mathrm{d}\xi}{\sqrt{2\pi}} \\ &= \sqrt{\frac{(n-k)!}{n!}} f^{(k)}_{n-k} \end{aligned}$$

where $f^{(k)}_{n-k}$ denotes the $(n-k)$th Hermite coefficient of $f^{(k)}$. In other words, there

holds $f_n^{(k)} = \sqrt{(n+k)!/n!}\, f_{n+k}$ for $n \in \mathbb{N}$ and, thus, by Parseval's identity we obtain

$$\|f^{(k)}\|_{L^2_{\mu_1}}^2 = \sum_{n\geq 0} \frac{(n+k)!}{n!} |f_{n+k}|^2.$$

We note that $(n+k)!/n! = n \cdot (n-1) \cdots (n-k+1) \in \mathcal{O}(n^k)$. Thus, if $f \in L^2_{\mu_1}(\mathbb{R})$ has a weak derivative $f^{(k)} \in L^2_{\mu_1}(\mathbb{R})$ — i.e., if f has finite Sobolev regularity — then there exists a constant $C_f < \infty$ such that

$$|f_n| \leq C_f\, n^{-(k+1)/2} \qquad \forall n \in \mathbb{N}_0.$$

This reasoning can easily be extended to several dimensions and Bachmayr et al. [6] showed that a weighted ℓ^2-summability (with increasing weights) of the Hermite coefficients $\|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}}$ of $u \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ is equivalent to a weighted ℓ^2-summability (with other increasing weights) of the L^2_μ-norms of the partial derivatives $\|\partial^{\boldsymbol{j}} u\|_{L^2_\mu}$, $\boldsymbol{j} \in \mathcal{F}$ and $\max_{m\in\mathbb{N}} j_m \leq k$. We refer to their work for details. Again, the weighted ℓ^2-summability of $(\|u_{\boldsymbol{\alpha}}\|_{\mathcal{H}})_{\boldsymbol{\alpha}\in\mathcal{F}}$ implies a certain decay analogously to above which can be used to establish convergence rates of approximation methods via Stechkin's lemma.

Chapter 3
Bayesian Inference

This chapter is based on the publications [56, 57], but the presentation has been modified and many details and remarks as well as some new theoretical results have been added.

In this and the subsequent chapters we consider the *inverse problem* in uncertainty quantification. This time we do not propagate uncertainty through a forward map such as the solution operator of an elliptic PDE as in the previous chapter. We rather modify our uncertainty about coefficient functions by statistical inference given noisy observations of a random variable defined by such a forward map. From the UQ perspective the inverse problem is of tremendous interest, since incorporating any available information into the probability law for an uncertain quantity may reduce the uncertainty and lead to improved stochastic models and predictions.

Assume that we have made a finite-dimensional noisy observation modeled as

$$y = G(u) + \varepsilon, \tag{3.1}$$

where $u \in \mathcal{H}$ represents the unknown in a separable Hilbert space $\mathcal{H}$, $G\colon \mathcal{D}_G \to \mathbb{R}^d$ denotes the deterministic unknown-to-observation map with domain $\mathcal{D}_G \subseteq \mathcal{H}$ and ε the observational noise. Of course, we need to assume that $u \in \mathcal{D}_G$ in (3.1). Moreover, we suppose that our current state of knowledge about u, which may be based on physical reasoning, expert knowledge or previously collected data, is described by a *prior* probability measure μ_0 on $\mathcal{H}$. Besides that, the observational noise ε is assumed to be random with a known distribution based, again, on physical or statistical reasoning. Thus, a more appropriate model for the observation is

$$Y = G(U) + \varepsilon, \tag{3.2}$$

where now $U \sim \mu_0$ denotes a random variable in $\mathcal{H}$ distributed according to the prior measure μ_0, ε denoting the random variable for the observational noise and,

hence, Y the $\mathbb{R}^d$-valued random variable describing our prior probabilistic model for the observation (before we have made it). Here, we need to assume $\mu_0(\mathcal{D}_G) = 1$ to give Y in (3.2) meaning which in turn requires $\mathcal{D}_G$ to be a measurable set.

The observation $Y = y$ then provides new information to us which we would like to take into account and merge with our prior belief or prior uncertainty about u. This can be realized mathematically by conditioning the prior probability measure μ_0 on the event of observing $Y = y$. Such conditional measures are rooted in Kolmogorov's fundamental concept of *conditional expectation*, see, e.g., Rao [137], which we will explain later in detail. Furthermore, the famous *Bayes' rule* provides an analytic expression for the resulting conditional probability measure or *posterior measure* under certain mild assumptions. It is this posterior measure which represents our remaining uncertainty about u after having made the observation $Y = y$. In fact, computing the resulting posterior measure via Bayes' rule is nothing else than performing *Bayesian inference* for u given the prior μ_0 and the noisy data y. Bayesian inference is a type of statistical inference – the other common one is called *Frequentist inference* – where the unknown parameters which we want to infer are considered as random parameters following a postulated prior probability distribution. Again, this prior measure is supposed to represent our current knowledge about the parameters before we infer them given the observational data. Both types of statistical inference are based on a corresponding interpretation of probability which led to many, still ongoing debates between Bayesians and Frequentists. In light of uncertainty quantification, perhaps the Bayesian point of view, that probability is a mathematical representation of one's *subjective* beliefs, seems to be more appropriate than the Frequentist interpretation as an (objective) limit of relative frequencies for infinitely many trials. However, we will not deepen this philosophical discussion, since, as Williams [176, p. 220] remarks, "Enough paper has already been devoted to the topic", and refer the interested reader to Jaynes [94] and Lindley [112].

In this chapter, we will outline the Bayesian approach to inferring knowledge about an uncertain function given indirect observations of it and we will relate the Bayesian methodology to the concept of conditioning. Moreover, we derive the basic results such as Bayes' rule and the continuous dependence of the resulting posterior measure on the data in a quite general setting. Besides this we introduce Bayes estimates and Bayes estimators which are also an important part of Bayesian statistics. At the end of the chapter, we make a few remarks about the relation between the Bayesian approach and regularization theory for (deterministic) inverse problems and provide an overview of numerical methods for Bayesian inference.

Before we start recalling the necessary basics on probability measures, we illus-

trate the abstract observational model (3.1) and (3.2), respectively, with our running example.

Example 3.1 (Elliptic PDE)**.** We consider an elliptic BVP stated on a bounded domain $D \subset \mathbb{R}^3$

$$-\nabla \cdot (\mathrm{e}^u \nabla p) = f \quad \text{on } D, \qquad p \equiv 0 \quad \text{on } \partial D, \tag{3.3}$$

or its weak formulation, respectively, where p may describe a stationary groundwater pressure head and $u \in L^\infty(D)$ the logarithm of the uncertain hydraulic conductivity of a porous medium. Given observations of the pressure head p at d locations $x_1, \dots, x_d \in D$, e.g., by measurements at boreholes, we would like to infer the log conductivity u. Thus, here we have $u \in L^\infty(D) \subset L^2(D) =: \mathcal{H}$ as the unknown and the observation map $G\colon L^\infty(D) \to \mathbb{R}^d$ is given by the mapping $u \mapsto (\ell_1(p), \dots, \ell_d(p))$, where the linear functionals $\ell_i(p) := \int_{|y-x_i|\le\epsilon} p(y)\,\mathrm{d}y$ model local averages of p around the borehole centers $x_i \in D$.

In pratice one often has already some limited prior knowledge about the log conductivity u, e.g., by geological information, which may be encoded in a pathwise continuous Gaussian random field model for u, i.e., $u(\cdot, \omega) \in C(\bar{D})$ $\mathbb{P}$-almost surely. As we know from Chapter 2, this yields an $L^2(D)$-valued random variable U whose distribution μ_0 on $\mathcal{H} = L^2(D)$ serves as prior measure and satisfies $\mu_0(C(\bar{D})) = \mu_0(L^\infty(D)) = 1$. Furthermore, measurement errors ε are usually modeled as Gaussian. Given the accuracy of the measurement instrument in terms of a standard deviation σ, we may assume $\varepsilon \sim N(0, \sigma^2 I_d)$ with I_d denoting the d-dimensional identity matrix.

We can also apply the Karhunen-Loève expansion of the Gaussian random field u, see Theorem 2.28,

$$u(x, \omega) = \phi_0(x) + \sum_{m=1}^{\infty} \phi_m(x) \xi_m(\omega),$$

where $\|\phi_m\|_{L^2(D)} = 1$ and the $\xi_m \sim N(0, \lambda_m)$ are stochastically independent with $(\lambda_m)_{m\in\mathbb{N}} \in \ell^2(\mathbb{N})$, and equivalently infer the random coefficients $\boldsymbol{\xi} = (\xi_m)_{m\in\mathbb{N}}$ given the observational data. The underlying Hilbert space is then $\ell^2(\mathbb{N})$, since for the ξ_m as above there holds $\mathbb{P}$-almost surely $\boldsymbol{\xi}(\omega) \in \ell^2(\mathbb{N})$, see, e.g., Schwab and Gittelson [156, Proposition C.12].

Remark 3.2. Given Example 3.1 where the prior measure μ_0 is supported on a separable Banach space $C(\bar{D})$ one might ask why we consider the larger Hilbert space $\mathcal{H} = L^2(D)$ as the basic setting for the Bayesian inference. The reason here is that it is easier to work with measures on Hilbert spaces rather than with measures on

Banach spaces, although all of the results in this chapter are extendable to separable Banach spaces. Moreover, when considering Kalman filter methods in Chapter 4, we actually need an inner product space.

3.1. Preliminaries on Probability Measures

We first recall some basic notation from probability theory. For two measures μ_1 and μ_2 on a separable Hilbert space $\mathcal{H}$ we define the notation

$$\mu_1(\mathrm{d}u) \propto \mu_2(\mathrm{d}u) \quad \Longleftrightarrow \quad \exists c \in \mathbb{R}\ \forall A \in \mathcal{B}(\mathcal{H}) :\ \mu_1(A) = c\mu_2(A),$$

and, given a measurable function $\rho\colon \mathcal{H} \to [0,\infty)$,

$$\mu_1(\mathrm{d}u) = \rho(u)\mu_2(\mathrm{d}u) \quad \Longleftrightarrow \quad \mu_1(A) = \int_A \rho(u)\mu_2(\mathrm{d}u) \quad \forall A \in \mathcal{B}(\mathcal{H}),$$

i.e., $\mathrm{d}u$ serves as a placeholder for measurable sets. We say ν is a *dominating measure* of μ if μ is absolutely continuous w.r.t. ν, i.e., there exists a measurable $\rho\colon \mathcal{H} \to [0,\infty)$ such that $\mu(\mathrm{d}u) = \rho(u)\,\nu(\mathrm{d}u)$. In the following, let $\mathcal{Y}$ denote a second separable Hilbert space and recall that $\mathcal{P}(\mathcal{H})$ and $\mathcal{P}(\mathcal{Y})$ denote the sets of all probability measures on $\mathcal{H}$ and $\mathcal{Y}$, respectively. Given two probability measures $\mu_1 \in \mathcal{P}(\mathcal{H})$ and $\mu_2 \in \mathcal{P}(\mathcal{Y})$ we denote by $\mu_1 \otimes \mu_2$ the product measure on the measurable space $(\mathcal{H} \times \mathcal{Y}, \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{Y}))$, i.e., for any $A \in \mathcal{B}(\mathcal{H})$ and $B \in \mathcal{B}(\mathcal{Y})$ we have

$$\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B).$$

Here, $\mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{Y})$ denotes the tensor-product σ-algebra on $\mathcal{H} \times \mathcal{Y}$ generated by all sets $A \times B$, $A \in \mathcal{B}(\mathcal{H})$ and $B \in \mathcal{B}(\mathcal{Y})$. In particular, if two random variables $X \sim \mu_1$ and $Y \sim \mu_2$ are independent, we have $(X, Y) \sim \mu_1 \otimes \mu_2$ and vice versa.

For a measurable mapping $f\colon \mathcal{H} \to \mathcal{Y}$ we denote by $f_*\mu$ the *pushforward measure* of μ under f on $\mathcal{Y}$, i.e., $f_*\mu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{B}(\mathcal{Y})$. Besides this we will sometimes use the notation

$$\begin{aligned} \mathbb{E}_\mu[f] &:= \int_{\mathcal{H}} f(u)\,\mu(\mathrm{d}u), \\ \mathrm{Cov}_\mu(f) &:= \int_{\mathcal{H}} \big(f(u) - \mathbb{E}_\mu[f]\big) \otimes \big(f(u) - \mathbb{E}_\mu[f]\big)\ \mu(\mathrm{d}u), \end{aligned}$$

for the expectation and covariance of f w.r.t. μ, respectively. For a real-valued $f\colon \mathcal{H} \to \mathbb{R}$ we denote by $\mathrm{Var}_\mu(f)$ the resulting variance of f w.r.t μ. Furthermore,

analogously to Definition 2.14 we will use the notation

$$L^p_\mu(\mathcal{H};\mathcal{Y}) := \left\{ f\colon \mathcal{H} \to \mathcal{Y} \,|\, f \text{ measurable and } \|f\|^p_{L^p_\mu} := \int_{\mathcal{H}} \|f(u)\|^p_{\mathcal{Y}}\, \mu(\mathrm{d}u) < \infty \right\},$$

where $p \in [1,\infty)$, and denote the inner product in $L^2_\mu(\mathcal{H};\mathcal{Y})$ by

$$\langle f, g\rangle_{L^2_\mu} := \int_{\mathcal{H}} \langle f(u), g(u)\rangle_{\mathcal{Y}}\, \mu(\mathrm{d}u), \qquad f, g \in L^2_\mu(\mathcal{H};\mathcal{Y}).$$

We will also use the short notation $L^p_\mu(\mathcal{H}) := L^p_\mu(\mathcal{H};\mathbb{R})$.

Definition 3.3 (Mean and covariance of probability measures)**.** For $q \in \mathbb{N}$ the set of all probability measures μ on $\mathcal{H}$ with qth absolute moment is denoted by

$$\mathcal{P}^q(\mathcal{H}) := \left\{ \mu \in \mathcal{P}(\mathcal{H}) :\ \int_{\mathcal{H}} \|u\|^q_{\mathcal{H}}\, \mu(\mathrm{d}u) < \infty \right\}.$$

For a measure $\mu \in \mathcal{P}^1(\mathcal{H})$ its *mean* $m \in \mathcal{H}$ is given by the Bochner integral

$$m = \int_{\mathcal{H}} u\, \mu(\mathrm{d}u)$$

and for $\mu \in \mathcal{P}^2(\mathcal{H})$ its *covariance (operator)* is defined as the unique bounded linear operator $C \in \mathcal{L}(\mathcal{H})$ satisfying

$$\langle u_1, Cu_2\rangle_{\mathcal{H}} = \int_{\mathcal{H}} \langle u_1, v - m\rangle_{\mathcal{H}}\, \langle u_2, v - m\rangle_{\mathcal{H}}\, \mu(\mathrm{d}v) \qquad \forall u_1, u_2 \in \mathcal{H}.$$

There holds a similar statement to Proposition 2.18, i.e., for $\mu \in \mathcal{P}^2(\mathcal{H})$ its covariance operator C is self-adjoint, positive and trace class. In the following, we will denote the set of all linear self-adjoint, positive and trace class operators $A\colon \mathcal{H} \to \mathcal{H}$ by $\mathcal{L}^1_+(\mathcal{H})$ and refer to Appendix A for more details on subspaces of linear operators.

We also require the notion of distance between probability measures. In this thesis we will mainly work with the *total variation distance* d_{TV} and *Hellinger distance* d_{H} as well as with the weak convergence of measures. For a survey on metrics for probability measures and their relations we refer to Gibbs and Su [71].

Definition 3.4 (Weak convergence)**.** Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{H})$ be a sequence of probability measures on $\mathcal{H}$. We say the measures μ_n *converge weakly* to a measure $\mu \in \mathcal{P}(\mathcal{H})$, written as $\mu_n \xrightarrow{w} \mu$, if

$$\lim_{n\to\infty} \int_{\mathcal{H}} f(u)\, \mu_n(\mathrm{d}u) = \int_{\mathcal{H}} f(u)\, \mu(\mathrm{d}u) \qquad \forall f \in C_b(\mathcal{H};\mathbb{R}), \tag{3.4}$$

where $C_b(\mathcal{H};\mathbb{R})$ denotes the set of all bounded, continuous functions $f\colon \mathcal{H} \to \mathbb{R}$.

Weak convergence of measures on Polish spaces is induced by the *Lévy-Prokhorov metric*, we refer to Klenke [99, Remark 13.14] for details. Moreover, there are several equivalent characterizations of weak convergence usually summarized under the name *Portmanteau theorem*. For example, we can replace $C_b(\mathcal{H};\mathbb{R})$ in (3.4) by the set of all bounded, Lipschitz continuous functions $f \in \mathrm{Lip}_b(\mathcal{H};\mathbb{R})$, see again Klenke [99, Theorem 13.16], and obtain the same topology on $\mathcal{P}(\mathcal{H})$. This particular statement of the Portmanteau theorem will be used in Section 4.2.2.

Definition 3.5 (Total variation distance)**.** The *total variation distance* between two probability measures $\mu_1, \mu_2 \in \mathcal{P}(\mathcal{H})$ is given by

$$d_{\mathrm{TV}}(\mu_1,\mu_2) := \sup_{A\in\mathcal{B}(\mathcal{H})} |\mu_1(A) - \mu_2(A)|\,. \tag{3.5}$$

The definition of the total variation distance is quite natural, since it measures the difference of probabilities for the same event. Moreover, there holds for $\mu, \mu_n \in \mathcal{P}(\mathcal{H})$, $n \in \mathbb{N}$, that

$$d_{\mathrm{TV}}(\mu_n,\mu) \to 0 \quad \Longrightarrow \quad \mu_n \xrightarrow{w} \mu, \tag{3.6}$$

see Gibbs and Su [71], i.e., convergence in total variation implies weak convergence.

Definition 3.6 (Hellinger distance)**.** For two probability measures $\mu_1, \mu_2 \in \mathcal{P}(\mathcal{H})$ their *Hellinger distance* is defined as

$$d_{\mathrm{H}}(\mu_1,\mu_2) := \left[\int_{\mathcal{H}} \left(\sqrt{\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu}(u)} - \sqrt{\frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}(u)}\right)^2 \nu(\mathrm{d}u)\right]^{1/2},$$

where ν is a dominating measure for μ_1 and μ_2, e.g., $\nu = (\mu_1+\mu_2)/2$.

It can be easily checked that the definition of the Hellinger distance is independent of the dominating measure. Furthermore, there holds, see again Gibbs and Su [71],

$$\frac{d_{\mathrm{H}}^2(\mu_1,\mu_2)}{2} \le d_{\mathrm{TV}}(\mu_1,\mu_2) \le d_{\mathrm{H}}(\mu_1,\mu_2), \qquad \mu_1,\mu_2 \in \mathcal{P}(\mathcal{H}), \tag{3.7}$$

i.e., they induce the same topology on $\mathcal{P}(\mathcal{H})$. The Hellinger distance possesses the convenient property that continuity w.r.t. d_{H} implies the continuity of moments as stated in Dashti and Stuart [43]:

Lemma 3.7 ([43, Lemma 7.14])**.** Let $\mu_1, \mu_2 \in \mathcal{P}(\mathcal{H})$ and $f \in L^2_{\mu_1}(\mathcal{H};\mathcal{Y}) \cap L^2_{\mu_2}(\mathcal{H};\mathcal{Y})$.

Then there holds

$$\left\|\mathbb{E}_{\mu_1}[f] - \mathbb{E}_{\mu_2}[f]\right\|_{\mathcal{Y}} \leq \left(2\|f\|^2_{L^2_{\mu_1}} + 2\|f\|^2_{L^2_{\mu_2}}\right)^{1/2} d_{\mathrm{H}}(\mu_1, \mu_2). \tag{3.8}$$

Gaussian measures on Hilbert spaces. As for random fields and Hilbert space-valued random variables we will often work with Gaussian measures on Hilbert spaces and exploit their properties. Gaussian measures are very convenient to work with. We refer to Bogachev [18] for a comprehensive presentation as well as to Da Prato and Zabczyk [39, Chapter 1] and Hairer [80, Section 3] for lucid introductions.

Definition 3.8 (Gaussian measure)**.** A measure $\mu \in \mathcal{P}(\mathcal{H})$ is called a *Gaussian measure* with mean $m \in \mathcal{H}$ and covariance operator $C \in \mathcal{L}^1_+(\mathcal{H})$, denoted by $\mu = N(m, C)$, if for each $u \in \mathcal{H}$ there holds with $f_u \colon \mathcal{H} \to \mathbb{R}$ given as $f_u(v) := \langle u, v\rangle_{\mathcal{H}}$ that

$$(f_u)_* \mu \sim N\left(\langle m, u\rangle_{\mathcal{H}}, \langle u, Cu\rangle_{\mathcal{H}}\right),$$

where $(f_u)_*\mu \in \mathcal{P}(\mathbb{R})$ denotes the push-forward measure of μ under the mapping f_u.

Needless to say, that the distribution on $\mathcal{H}$ of a Gaussian $\mathcal{H}$-valued random variable X, i.e., the pushforward measure $\mu = X_*\mathbb{P} \in \mathcal{P}(\mathcal{H})$, is Gaussian with mean $\mathbb{E}[X]$ and covariance $\mathrm{Cov}(X)$.

Remark 3.9. An equivalent characterization of Gaussian measures can be obtained via *characteristic functions*, i.e., $\mu = N(m, C)$ iff

$$\int_{\mathcal{H}} \mathrm{e}^{\mathrm{i}\langle u, v\rangle_{\mathcal{H}}}\, \mu(\mathrm{d}v) = \mathrm{e}^{\mathrm{i}\langle m, u\rangle_{\mathcal{H}} - \frac{1}{2}\langle Cu, u\rangle_{\mathcal{H}}}, \qquad \forall u \in \mathcal{H}.$$

Moreover, analogously to Definition 2.19, we can define Gaussian meaures also in Banach spaces. However, in this thesis we focus on the Hilbert space case.

Gaussian measures are uniquely determined by their mean and covariance, i.e., for any $m \in \mathcal{H}$ and any $C \in \mathcal{L}^1_+(\mathcal{H})$ there exists a unique Gaussian measure $\mu = N(m, C)$ on $\mathcal{H}$, see, e.g., Da Prato and Zabczyk [38, Proposition 2.18]. Concerning equivalence of Gaussian measures, there holds that $\mu_1 = N(m_1, C_1)$ and $\mu_2 = N(m_2, C_2)$ are either singular or equivalent, see Da Prato and Zabczyk [38, Theorem 2.23]. More details on the latter case are provided in Appendix C.

In the following, we will use upper case latin letters such as X, Y, U to denote random variables on Hilbert spaces and lower case latin letters such as x, y, u for

elements in these Hilbert spaces or realizations of the associated random variables. The greek letter ε will be used to denote random observational noise as well as its realization, and μ and ν (with various subscripts) will denote measures on the Hilbert space $\mathcal{H}$ and on $\mathbb{R}^d$, respectively.

3.2. Conditional Measures

Bayesian inference consists in updating prior knowledge on an unknown quantity modeled as a random variable U, reflecting a gain in knowledge due to new observations. The distribution of U, characterized by the probabilities $\mathbb{P}(U \in B)$ for $B \in \mathcal{B}(\mathcal{H})$, quantifies, in stochastic terms, our knowledge about the uncertainty associated with U. When new information becomes available, such as knowing that the event $Y = y$ has occurred, this is reflected in our quantitative description as the "conditional distribution of U given $\{Y = \mathrm{y}\}$", denoted $\mathbb{P}(U \in B \mid Y = \mathrm{y})$.

Before we present the general approach to define conditional distributions via conditional expectation, we illustrate the procedure of conditioning for a very simple case: assume we know that an event A with positive probability $\mathbb{P}(A) > 0$ has occured. Then, the conditional probability $\mathbb{P}(U \in B \mid A)$ of $U \in B$, $B \in \mathcal{B}(\mathcal{H})$, given the event A is defined as

$$\mathbb{P}(U \in B \mid A) := \frac{\mathbb{P}(\{\omega : U(\omega) \in B\} \cap A)}{\mathbb{P}(A)},$$

and Bayes' rule in its simplest form then reads

$$\mathbb{P}(U \in B \mid A) = \frac{\mathbb{P}(A \mid U \in B)}{\mathbb{P}(A)} \mathbb{P}(U \in B),$$

which can be verified by a simple calculation. Here, the term $\mathbb{P}(A \mid U \in B)$ represents the probability (or later *likelihood*) for A to occur given that $U \in B$. Thus, the change from the prior probability $\mathbb{P}(U \in B)$ to the conditional or *posterior* probability $\mathbb{P}(U \in B \mid A)$ is a simple reweighting by the *normalized likelihood* $\mathbb{P}(A \mid U \in B)/\mathbb{P}(A)$. This structure will remain the same also in a more general setting.

If the occured event A is not assigned a positive probability, e.g., we observed $Y = y$ where the distribution of Y is absolutly continuous w.r.t. Lebesgue measure, the approach above is clearly not well-defined. In finite dimensions, i.e., $\mathcal{H} = \mathbb{R}^n$, we can then define *conditional densities* w.r.t. the Lebesgue measure to make sense of conditional probabilities. However, we omit the corresponding statements here and continue with the general approach of *conditional distributions* which also works in infinite dimensional spaces. The key concept here is that of *conditional expectation*.

Definition 3.10 (Conditional expectation, conditional probability)**.** Let $X \in L^1(\Omega; \mathcal{H})$ and $Y\colon \Omega \to S$ be random variables, where $(S, \mathcal{S})$ denotes an arbitrary measurable space. We define the *conditional expectation* $\mathbb{E}[X|Y]$ of X given Y as any $\sigma(Y)$-measurable mapping $\mathbb{E}[X|Y]\colon \Omega \to \mathcal{H}$ which satisfies

$$\int_A \mathbb{E}[X|Y]\, \mathbb{P}(\mathrm{d}\omega) = \int_A X\, \mathbb{P}(\mathrm{d}\omega) \qquad \forall A \in \sigma(Y).$$

For any $B \in \mathcal{B}(\mathcal{H})$ the *conditional probability* $\mathbb{P}(X \in B|Y)$ of $X \in B$ given Y is defined as any $\sigma(Y)$-measurable mapping $\mathbb{P}(X \in B|Y)\colon \Omega \to [0,1]$ satisfying

$$\mathbb{P}(X \in B|Y) = \mathbb{E}\left[\mathbf{1}_{\{X \in B\}}|Y\right] \qquad \mathbb{P}\text{-almost surely.}$$

Thus, conditional expectation and conditional probability are unqiuely defined only up to $\mathbb{P}$-null sets. By the Doob-Dynkin Lemma, see Kallenberg [96, Lemma 1.13], there exists a measurable function $\phi\colon S \to \mathcal{H}$ such that $\mathbb{E}[X|Y] = \phi(Y)$ holds $\mathbb{P}$-almost surely. Again, we note that this does not determine a unique function ϕ but rather an equivalence class of measurable functions where $\phi_1 \sim \phi_2$ iff $\mathbb{P}(\phi_1(Y) \neq \phi_2(Y)) = 0$. For a specific realization $y \in S$ of Y (and a specific ϕ), we set

$$\mathbb{E}[X|Y = y] := \phi(y), \qquad y \in S,$$

and analogously

$$\mathbb{P}(X \in B|Y = y) := \mathbb{E}[\mathbf{1}_{\{U \in B\}}|Y = y], \qquad y \in S,$$

and, thus, obtain mappings $\mathbb{E}[X|Y = \cdot]\colon S \to \mathcal{H}$ and $\mathbb{P}(X \in B|Y = \cdot)\colon S \to [0,1]$ representing conditional expectation and probability. Concerning the latter, one could now ask for a family of probability measures $\mathbb{P}(X \in \cdot|Y = y)\colon \mathcal{B}(\mathcal{H}) \to [0,1]$ parametrized by the realization y of Y, providing the conditional distribution of X resulting from the observation $Y = y$. Before stating the corresponding definition we introduce the concept of a *stochastic kernel* which plays an important role in many fields of probability theory, particularly, in the theory of Markov chains which we will consider in detail in Chapter 5.

Definition 3.11 (Stochastic kernel)**.** A mapping $K\colon \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0,1]$ is called a *stochastic kernel* or *Markov kernel on* $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, or simply *on* $\mathcal{H}$, if

- $K(u, \cdot)\colon \mathcal{B}(\mathcal{H}) \to [0,1]$ is a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ for each $u \in \mathcal{H}$,
- $K(\cdot, A)\colon \mathcal{H} \to [0,1]$ is a measurable function for each $A \in \mathcal{B}(\mathcal{H})$.

Thus, a stochastic kernel K can be viewed as a parametrized probability measure, i.e., as a mapping from $\mathcal{H}$ into $\mathcal{P}(\mathcal{H})$. The relation to the idea of conditional distributions as above is obvious and we state

Definition 3.12 (Conditional distribution)**.** For two random variables $X\colon (\Omega, \mathcal{A}) \to (\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and $Y\colon (\Omega, \mathcal{A}) \to (S, \mathcal{S})$, where $(S, \mathcal{S})$ denotes an arbibtrary measurable space, the *regular conditional distribution of X given Y* is a stochastic kernel

$$\mu_{X|Y}\colon S \times \mathcal{B}(\mathcal{H}) \to [0,1]$$

such that for each $B \in \mathcal{B}(\mathcal{H})$ there holds

$$\mu_{X|Y}(Y(\omega), B) = \mathbb{P}(X \in B|Y)(\omega) \qquad \mathbb{P}\text{-almost surely.}$$

Thus, given a regular conditional distribution $\mu_{X|Y}$ we can employ the probability measure $\mu_{X|Y}(y, \cdot)$ as the distribution of X conditioned on the event $Y = y$.

Remark 3.13 (On the existence of regular conditional distributions)**.** The existence of regular conditional distributions can be ensured under very general assumptions. In particular, if $X\colon \Omega \to T$ and $Y\colon \Omega \to S$ are random variables with S as a general measurable space and T a Polish space, i.e., a complete and separable metric space, there exists a regular conditional distribution of X given Y, see Kallenberg [96, Theorem 6.3]. For cases where regular conditional distributions do not exist we refer to Rao [137].

3.3. Bayes' Rule and the Posterior Measure

We will now provide an expression for the conditional distribution of U given Y if Y is defined as in (3.2) by Bayes' rule and, moreover, discuss stability results for the resulting posterior measure. We mention that similar statements can be found in Stuart [167] for Gaussian random variables U and ε in (3.2) and in Dashti and Stuart [43] for a more general setting. To this end, we make the following assumptions for the model equation (3.2).

Assumption 3.14.

1. Let $U \sim \mu_0$, $\varepsilon \sim \nu_\varepsilon$ and $(U, \varepsilon) \sim \mu_0 \otimes \nu_\varepsilon$, i.e., U and ε are independent.

2. The distribution ν_ε of ε is absolutely continuous w.r.t. the Lebesgue measure in $\mathbb{R}^d$ with density $\rho(\varepsilon) \propto \mathrm{e}^{-\ell(\varepsilon)}$ where $\ell\colon \mathbb{R}^d \to [0, \infty)$.

3. For $G\colon \mathcal{D}_G \to \mathbb{R}^d$ there holds that $\mathcal{D}_G \in \mathcal{B}(\mathcal{H})$ with $\mu_0(\mathcal{D}_G) = 1$ and that G is measurable w.r.t. the Borel σ-algebras $\mathcal{B}(\mathcal{H})$ and $\mathcal{B}(\mathbb{R}^d)$.

In the following, when we consider integrals over $\mathcal{H}$ w.r.t. μ_0 where the integrand involves G, we understand them in the sense that they are taken over $\mathcal{D}_G$ or, equivalently, the G appearing in the integral is an extension of G to $\mathcal{H}$ with $G(u) = 0$ for $u \notin \mathcal{D}_G$.

Since the random variable Y in (3.2) is, by Assumption 3.14, the sum of two independent random variables, $G(U)$ and ε, its distribution ν_Y is given as

$$\nu_Y(B) = \int_{\mathcal{H}} \nu_\varepsilon(B - G(u))\, \mu_0(\mathrm{d}u) = \int_B \int_{\mathcal{H}} \rho(y - G(u))\, \mu_0(\mathrm{d}u)\, \mathrm{d}y, \qquad B \in \mathcal{B}(\mathbb{R}^d),$$

see Kallenberg [96, Lemma 1.28 & Corollary 3.12], hence,

$$\nu_Y(\mathrm{d}y) = C\,\gamma(y)\mathrm{d}y, \quad \gamma(y) := \int_{\mathcal{H}} \mathrm{e}^{-\ell(y-G(u))}\ \mu_0(\mathrm{d}u),\ C^{-1} := \int_{\mathbb{R}^d} \gamma(y)\,\mathrm{d}y. \quad (3.9)$$

We remark that $\gamma(y)$ is well-defined and strictly positive, since $0 < |\,\mathrm{e}^{-\ell(y-G(u))}\,| \leq 1$, and that $\gamma \in L^1(\mathbb{R}^d)$ due to Fubini's theorem, see Kallenberg [96, Theorem 1.27]. In particular, the distribution ν_Y of $\mathcal{Y}$ is also absolutely continuous w.r.t. the Lebesgue measure. Moreover, it follows by construction that the conditional probability of $Y \in B$, $B \in \mathcal{B}(\mathbb{R}^d)$, given $U = u \in \mathcal{D}_G$ is

$$\mathbb{P}(Y \in B | U = u) \propto \int_B \mathrm{e}^{-\ell(y-G(u))}\,\mathrm{d}y, \quad (3.10)$$

and, in particular, that the joint distribution μ of (U, Y) on $\mathcal{H} \times \mathbb{R}^d$ is given by

$$\mu(\mathrm{d}u, \mathrm{d}y) = C\,\mathrm{e}^{-\ell(y-G(u))}\ \mu_0(\mathrm{d}u) \otimes \mathrm{d}y. \quad (3.11)$$

Definition 3.15 (Potential)**.** Given Assumption 3.14 we define the *potential* Φ as the mapping $\Phi\colon \mathcal{H} \times \mathbb{R}^d \to [0, \infty)$ given by

$$\Phi(u, y) := \ell(y - G(u)) \quad (3.12)$$

where $G(u) := 0$ for $u \notin \mathcal{D}_G$.

Note, that Φ represents the negative log likelihood of observing $Y = y$ given $U = u$ and that Φ is a measurable function. We now show that under Assumption 3.14 Bayes' rule yields a regular conditional distribution $\mu_{U|Y}$ of U given Y where Y is defined by (3.2).

Theorem 3.16 (Bayes' rule)**.** Consider the model (3.2) and let Assumption 3.14 be satisfied. For each $y \in \mathbb{R}^d$ we define a probability measure μ^y on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ by

$$\mu^y(\mathrm{d}u) := \frac{1}{\gamma(y)} \exp(-\Phi(u,y))\ \mu_0(\mathrm{d}u) \tag{3.13}$$

where $\gamma(y)$ is as in (3.9). Then the mapping $\mu_{U|Y} \colon \mathbb{R}^d \times \mathcal{B}(\mathcal{H}) \to [0,1]$ given by

$$\mu_{U|Y}(y,B) := \mu^y(B) \qquad \forall B \in \mathcal{B}(\mathcal{H})$$

is a regular conditional distribution of U given Y.

Proof. First, we verify that $\mu_{U|Y}$ with $\mu_{U|Y}(y,B) := \mu^y(B)$ for $y \in \mathbb{R}^d$ and $B \in \mathcal{B}(H)$ is a stochastic kernel. That $\mu_{U|Y}(y,\cdot)$ is a probability measure for each $y \in \mathbb{R}^d$ follows by construction. Moreover, we argue that for a fixed $B \in \mathcal{B}(H)$ the mapping $y \mapsto \mu^y(B)$ is measurable. To this end, we recall that for any measurable function $f \colon \mathcal{H} \times \mathbb{R}^d \to \mathbb{R}$ and any probability measure η on $\mathcal{H}$ the mapping

$$y \mapsto \int_{\mathcal{H}} f(u,y)\, \eta(\mathrm{d}u)$$

is again measurable, see, e.g., Kallenberg [96, Lemma 1.41]. Thus, since the potential $\Phi \colon \mathcal{H} \times \mathbb{R}^d \to [0,\infty)$ is measurable by assumption, we obtain that

$$y \mapsto \int_B \exp(-\Phi(u,y))\, \mu_0(\mathrm{d}u) = \int_{\mathcal{H}} \mathbf{1}_B(u) \exp(-\Phi(u,y))\, \mu_0(\mathrm{d}u)$$

is measurable for any $B \in \mathcal{B}(\mathcal{H})$ which implies that also

$$y \mapsto \frac{\int_B \exp(-\Phi(u,y))\, \mu_0(\mathrm{d}u)}{\int_{\mathcal{H}} \exp(-\Phi(u,y))\, \mu_0(\mathrm{d}u)} = \mu^y(B).$$

is measurable. Hence, $\mu_{U|Y}$ is a stochastic kernel.

Next, we prove that for any $B \in \mathcal{B}(\mathcal{H})$ there holds $\mathbb{P}$-a.s. $\mu_{U|Y}(Y,B) = \mathbb{P}(U \in B|Y)$. To this end, we verify that

$$\int_A \mu_{U|Y}(Y(\omega),B)\, \mathbb{P}(\mathrm{d}\omega) = \int_A \mathbf{1}_B(U(\omega))\, \mathbb{P}(\mathrm{d}\omega) \qquad \forall A \in \sigma(Y).$$

Recall the joint distribution μ of (U,Y) given in (3.11). There holds

$$\mu(\mathrm{d}u,\mathrm{d}y) = C\,\mathrm{e}^{-\Phi(u,y)}\ \mu_0(\mathrm{d}u) \otimes \mathrm{d}y = C\gamma(y)\, \mu_{U|Y}(y,\mathrm{d}u)\, \mathrm{d}y = \mu_{U|Y}(y,\mathrm{d}u)\ \nu_Y(\mathrm{d}y)$$

where ν_Y denotes the distribution of Y given in (3.9). Since $\sigma(Y) = \{Y^{-1}(A) : A \in$

$\mathcal{B}(\mathbb{R}^d)\}$, we choose an arbitrary $A \in \mathcal{B}(\mathbb{R}^d)$ and obtain

$$\begin{aligned}\int_{Y^{-1}(A)} \mathbf{1}_B(U(\omega))\,\mathbb{P}(\mathrm{d}\omega) &= \int_{B\times A} \mu(\mathrm{d}u,\mathrm{d}y) = \int_A \mu_{U|Y}(y,B)\;\nu_Y(\mathrm{d}y)\\ &= \int_{Y^{-1}(A)} \mu_{U|Y}(Y(\omega),B)\,\mathbb{P}(\mathrm{d}\omega)\end{aligned}$$

which concludes the proof. □

Definition 3.17 (Prior measure, posterior measure)**.** Let Assumption 3.14 be satisfied for the model (3.2). Then the measure μ_0 is called the *prior measure* of U and the measure μ^y given in (3.13) the *posterior measure* of U given $Y = y$.

It is helpful to maintain a clear distinction between *conditional* and *posterior* quantities in the following: the former contain the – as yet unrealized – observation as a parameter, while in the latter the observation has been made. Specifically, $\mu_{X|Y}$ is the conditional measure of X conditioned on Y, whereas $\mu_{X|Y}(y,\cdot)$ denotes the posterior measure of X for the observation $Y = y$. Moreover, *prior* quantities do not depend on the observation random variable Y or its realizations at all.

Stability of the posterior measure. We will now study the continuity of the posterior measure w.r.t. the observed data and its stability w.r.t. approximations of the forward map G. Again, these issues were already adressed by Stuart [167] for Gaussian priors and Gaussain noise and were recently extended by Dashti and Stuart [43] to a more general setting including ours. Nevertheless, we will not just cite the corresponding results in [43], but try to explain the underlying mathematics. The key observation for the stability analysis of the posterior is

Proposition 3.18 (cf. Stuart [167, Theorem 4.6])**.** Let $\mu_0 \in \mathcal{P}(\mathcal{H})$ and let $\Phi, \tilde{\Phi}\colon \mathcal{H} \to [0,\infty)$ be measurable functions and such that $\Phi, \tilde{\Phi} \in L^2_{\mu_0}(\mathcal{H})$. Then there holds for the two probability measures $\mu, \tilde{\mu} \in \mathcal{P}(\mathcal{H})$ given by

$$\begin{aligned}\mu(\mathrm{d}u) &:= \frac{1}{Z}\,\mathrm{e}^{-\Phi(u)}\;\mu_0(\mathrm{d}u), & Z &:= \int_{\mathcal{H}} \mathrm{e}^{-\Phi(u)}\;\mu_0(\mathrm{d}u),\\ \tilde{\mu}(\mathrm{d}u) &:= \frac{1}{\tilde{Z}}\,\mathrm{e}^{-\tilde{\Phi}(u)}\;\mu_0(\mathrm{d}u), & \tilde{Z} &:= \int_{\mathcal{H}} \mathrm{e}^{-\tilde{\Phi}(u)}\;\mu_0(\mathrm{d}u),\end{aligned}$$

that

$$d_{\mathrm{H}}(\mu,\tilde{\mu}) \leq \frac{1}{\min(Z,\tilde{Z})}\,\|\Phi - \tilde{\Phi}\|_{L^2_{\mu_0}}, \tag{3.14}$$

and also $|Z - \tilde{Z}| \leq \|\Phi - \tilde{\Phi}\|_{L^2_{\mu_0}}$.

Proof. Analogous to the proof of [167, Theorem 4.6] we start with

$$\begin{aligned} d_{\mathrm{H}}(\mu,\tilde{\mu})^2 &= \int_{\mathcal{H}} \left(\frac{\mathrm{e}^{-\Phi(u)/2}}{\sqrt{Z}} - \frac{\mathrm{e}^{-\tilde{\Phi}(u)/2}}{\sqrt{\tilde{Z}}} \right)^2 \mu_0(\mathrm{d}u) \\ &\leq 2 \int_{\mathcal{H}} \left(\frac{\mathrm{e}^{-\Phi(u)/2}}{\sqrt{Z}} - \frac{\mathrm{e}^{-\tilde{\Phi}(u)/2}}{\sqrt{Z}} \right)^2 + \left(\frac{\mathrm{e}^{-\tilde{\Phi}(u)/2}}{\sqrt{Z}} - \frac{\mathrm{e}^{-\tilde{\Phi}(u)/2}}{\sqrt{\tilde{Z}}} \right)^2 \mu_0(\mathrm{d}u) \\ &= I_1 + I_2 \end{aligned}$$

where

$$I_1 := \frac{2}{Z} \int_{\mathcal{H}} \left(\mathrm{e}^{-\Phi(u)/2} - \mathrm{e}^{-\tilde{\Phi}(u)/2} \right)^2 \mu_0(\mathrm{d}u), \qquad I_2 := \frac{2}{Z} \left(\sqrt{\tilde{Z}} - \sqrt{Z} \right)^2.$$

Since $|\,\mathrm{e}^{-x} - \mathrm{e}^{-y}\,| = \mathrm{e}^{-\min(x,y)}\,|1 - \mathrm{e}^{-|x-y|}\,| \leq 1 \cdot |x-y|$ for any $x, y \geq 0$, we get

$$I_1 \leq \frac{2}{Z} \int_{\mathcal{H}} \frac{\left|\Phi(u) - \tilde{\Phi}(u)\right|^2}{4} \mu_0(\mathrm{d}u) = \frac{1}{2Z} \|\Phi - \tilde{\Phi}\|_{L^2_{\mu_0}}^2$$

and since $|x^{1/2} - y^{1/2}| \leq \frac{1}{2} \min(x,y)^{-1/2}\,|x-y|$ we obtain

$$I_2 \leq \frac{1}{2Z\,\min(Z,\tilde{Z})}\,|Z - \tilde{Z}|^2.$$

Now, as for I_1 we get

$$|Z - \tilde{Z}|^2 \leq \int_{\mathcal{H}} \left| \mathrm{e}^{-\Phi(u)} - \mathrm{e}^{-\tilde{\Phi}(u)} \right|^2 \mu_0(\mathrm{d}u) \leq \|\Phi - \tilde{\Phi}\|_{L^2_{\mu_0}}^2$$

and due to $Z, \tilde{Z} \leq 1$ we have

$$\frac{1}{2Z} + \frac{1}{2Z\min(Z,\tilde{Z})} \leq \frac{1}{2\min(Z,\tilde{Z})^2} + \frac{1}{2\min(Z,\tilde{Z})^2} = \frac{1}{\min(Z,\tilde{Z})^2}$$

which concludes the proof. □

Proposition 3.18 guides us to establish assumptions under which we can show continuity in the Hellinger distance of μ^y w.r.t. y or w.r.t. G, since for μ^y we have $\Phi(u) = \Phi(u,y) = \ell(y - G(u))$. In particular, the regularity of ℓ will determine the corresponding modulus of continuity. Moreover, due to the term $\min(Z,\tilde{Z})^{-3/2}$ on the righthand side of (3.14) we will usually obtain only a local modulus of continuity, e.g., only a local Hölder continuity of μ^y w.r.t. y. We make the following assumption about the continuity of ℓ.

Assumption 3.19. The negative log likelihood $\ell\colon \mathbb{R}^d \to [0,\infty)$ is locally Hölder continuous with exponent $\alpha > 0$. In particular, there exists a nondecreasing function $L\colon [0,\infty) \to [0,\infty)$ such that

$$|\ell(y) - \ell(\tilde{y})| \leq L(r)\,|y-\tilde{y}|^\alpha \qquad \forall\, |y|, |\tilde{y}| \leq r.$$

This assumption is satisfied by many standard probability distribution functions. For example, if the negative log likelihood ℓ is continuously differentiable Assumption 3.19 holds with $\alpha = 1$ and $L(r) := \max_{|y|\leq r} |\nabla \ell(y)|$. In order to establish continuity of μ^y w.r.t. y, we then have to ensure a specific integrability of the function L in Assumption 3.19.

Theorem 3.20 (cf. Stuart [167, Theorem 4.2], Dashti and Stuart [43, Theorem 4.5])**.** Let Assumption 3.14 and 3.19 be satisfied. Assume further that for the mapping L in Assumption 3.19 there holds

$$\int_{\mathcal{H}} L^2(r + |G(u)|)\,\mu_0(\mathrm{d}u) \leq C_r < \infty \qquad \forall r \geq 0 \tag{3.15}$$

where C_r denotes a finite constant depending on r. Moreover, we assume that there exists a measurable function $f\colon \mathcal{H} \to \mathbb{R}_+$ such that

$$\ell(y - G(u)) \leq c_r + f(u) \qquad \forall |y| \leq r \tag{3.16}$$

where c_r denotes another finite constant depending on r. Then the measure μ^y in (3.13) is locally Hölder continuous in the Hellinger distance w.r.t. y with exponent α, i.e., for any $y, \tilde{y} \in \mathbb{R}^d$ with $|y|, |\tilde{y}| \leq r$ there holds

$$d_{\mathrm{H}}(\mu^y, \mu^{\tilde{y}}) \leq K_r\,|y - \tilde{y}|^\alpha,$$

with a constant K_r independent of y and $\tilde{y}$.

Proof. We define

$$\Phi(u) := \ell(y - G(u)), \qquad \tilde{\Phi}(u) := \ell(\tilde{y} - G(u))$$

and let Z and $\tilde{Z}$ be as in Proposition 3.18. Then, we get due to Assumption 3.19 and (3.15) as well as $|y - G(u)|, |\tilde{y} - G(u)| \leq r + |G(u)|$ that

$$\begin{aligned}\|\Phi - \tilde{\Phi}\|^2_{L^2_{\mu_0}} &= \int_{\mathcal{H}} |\ell(y - G(u)) - \ell(\tilde{y} - G(u))|^2\ \mu_0(\mathrm{d}u) \\ &\leq |y - \tilde{y}|^{2\alpha} \int_{\mathcal{H}} L(r + |G(u)|)^2\,\mu_0(\mathrm{d}u) \leq C_r |y - \tilde{y}|^{2\alpha}.\end{aligned}$$

Moreover, by assumption (3.16) we obtain

$$\min(Z, \tilde{Z}) \geq \int_{\mathcal{H}} \exp\left(-c_r - f(u)\right) \mu_0(\mathrm{d}u) =: \gamma_r > 0.$$

Thus, by Proposition 3.18 the assertion follows with $K_r = \sqrt{C_r}/\gamma_r$. □

Theorem 3.21 (cf. Stuart [167, Corollary 4.9], Dashti and Stuart [43, Theorem 4.8])**.** Let Assumption 3.14 and 3.19 be satisfied and let $G_h \colon \mathcal{D}_G \to \mathbb{R}^d$, $h > 0$, be an approximation of the forward map $G \colon \mathcal{D}_G \to \mathbb{R}^d$ such that

$$|G(u) - G_h(u)| \leq K(u)\, \psi(h) \qquad \forall u \in \mathcal{D}_G,$$

where $K \colon \mathcal{H} \to [0, \infty)$ and $\psi \colon [0, \infty) \to [0, \infty)$ denote measurable mappings. If there exists a finite constant C_r depending on $r \geq 0$ such that

$$\int_{\mathcal{H}} K^{2\alpha}(u)\; L^2(r + |G(u)| + |G_h(u)|)\, \mu_0(\mathrm{d}u) \leq C_r < \infty \qquad \forall r \geq 0 \tag{3.17}$$

and if there exists a measurable function $f \colon \mathcal{H} \to \mathbb{R}_+$ and another finite constant c_r such that there holds (3.16) and

$$\ell(y - G_h(u)) \leq c_r + f(u) \qquad \forall |y| \leq r, \tag{3.18}$$

then for the posterior measures μ^y as in (3.13) and $\mu_h^y(\mathrm{d}u) \propto \mathrm{e}^{-\ell(y - G_h(u))}\, \mu_0(\mathrm{d}u)$ there holds

$$d_{\mathrm{H}}(\mu^y, \mu_h^y) \leq 2K_r\, \psi(h)^\alpha, \qquad |y| \leq r,$$

with K_r depending only on r.

Proof. We denote $\Phi(u) := \ell(y - G(u))$ and $\tilde{\Phi}(u) := \ell(y - G_h(u))$ and let Z, $\tilde{Z}$, μ and $\tilde{\mu}$ be again as in Proposition 3.18. Due to assumption (3.16) and (3.18) we get analogously

$$\min(Z, \tilde{Z}) \geq \int_{\mathcal{H}} \exp\left(-c_r - f(u)\right) \mu_0(\mathrm{d}u) =: \gamma_r > 0$$

and by Assumption 3.19 and (3.17) we have due to $|y| \leq r$

$$\begin{aligned} \|\Phi - \tilde{\Phi}\|_{L^2_{\tilde{\mu}_0}}^2 &= \int_{\mathcal{H}} |\ell(y - G(u)) - \ell(y - G_h(u))|^2\, \mu_0(\mathrm{d}u) \\ &\leq \int_{\mathcal{H}} L^2(r + |G(u)| + |G_h(u)|)\, |G(u) - G_h(u)|^{2\alpha}\, \mu_0(\mathrm{d}u) \leq C_r \psi(h)^{2\alpha}. \end{aligned}$$

Thus, by Proposition 3.18 the assertion follows with $K_r = \sqrt{C_r}/\gamma_r$. □

In Theorem 3.20 and Theorem 3.21 we applied the analogous convention as before that $G_h(u) := 0$ for $u \notin \mathcal{D}_{\tilde{G}}$. Next, we investigate how Assumption 3.19 looks like for the Gaussian case and try to recover the corresponding assumptions made in Stuart [167] for the continuity of the posterior w.r.t. y.

Example 3.22 (Gaussian prior and noise)**.** Let $\mu_0 = N(0, C)$ and $\varepsilon \sim N(0, I_d)$, i.e., we have $\ell(y) = \frac{1}{2}|y|^2$ for $y \in \mathbb{R}^d$, hence,

$$|\ell(y) - \ell(\tilde{y})| \leq \max\{|y|, |\tilde{y}|\}\ |y - \tilde{y}|,$$

i.e., concerning Assumption 3.19 the Hölder exponent α is one and the function L is $L(r) = r$. Then, if for any $\beta > 0$ there exists a constant K_β such that

$$|G(u)| \leq \exp\left(\beta \|u\|_{\mathcal{H}}^2 + K_\beta\right), \qquad \forall u \in \mathcal{D}_G, \tag{3.19}$$

which is exactly [167, Assumption 2.7 (i)], we can ensure (3.15) by Fernique's theorem, see, e.g., Da Prato and Zabczyk [38, Theorem 2.6]. Analogously, (3.19) ensures (3.16), since

$$\ell(y - G(u)) = \frac{1}{2}|y - G(u)|^2 \leq |y|^2 + |G(u)|^2 \leq r^2 + \exp\left(2\beta \|u\|_{\mathcal{H}}^2 + 2K_\beta\right).$$

Moreover, if the bound (3.19) also holds for an approximation G_h of G and if

$$|G(u) - G_h(u)| \leq \exp\left(\beta \|u\|_{\mathcal{H}}^2 + K_\beta\right)\ \psi(h), \qquad u \in \mathcal{D}_G,$$

which is the assumption (4.11) in [167, Corollary 4.9], then also (3.17) and (3.18) are ensured by the same arguments.

As mentioned before continuity w.r.t. the Hellinger distance implies continuity of moments.

Corollary 3.23 (Continuity of posterior moments)**.** Assume for the prior measure that $\mu_0 \in \mathcal{P}^2(\mathcal{H})$. Then there hold the following statements:

- Let the assumptions of Theorem 3.20 be satisfied and let m^y and C^y denote the mean and covariance of μ^y and $m^{\tilde{y}}$ and $C^{\tilde{y}}$ the mean and covariance of $\mu^{\tilde{y}}$. Then there exists a constant K_r depending on $r = \max(|y|, |\tilde{y}|)$ such that

$$\|m^y - m^{\tilde{y}}\|_{\mathcal{H}} \leq K_r\, |y - \tilde{y}|^\alpha, \quad \|C^y - C^{\tilde{y}}\| \leq K_r\, |y - \tilde{y}|^\alpha.$$

- Let the assumptions of Theorem 3.21 be satisfied and let m_h^y and $C_h^{\tilde{y}}$ denote the mean and covariance of μ_h^y. Then there exists a constant K_r depending on

$r = |y|$ such that

$$\|m^y - m_h^y\|_{\mathcal{H}} \leq K_r \psi(h)^\alpha, \quad \|C^y - C_h^y\| \leq K_r \psi(h)^\alpha.$$

Proof. First of all, due to

$$\int_{\mathcal{H}} \|u\|_{\mathcal{H}}^2 \, \mu^y(\mathrm{d}u) = \frac{1}{\gamma(y)} \int_{\mathcal{H}} \|u\|_{\mathcal{H}}^2 \, \mathrm{e}^{-\Phi(u,y)} \, \mu_0(\mathrm{d}u) \leq \frac{1}{\gamma(y)} \int_{\mathcal{H}} \|u\|_{\mathcal{H}}^2 \, \mu_0(\mathrm{d}u) < \infty$$

we have $\mu^y \in \mathcal{P}^2(\mathcal{H})$ and the mean and covariance of μ^y exist for each $y \in \mathbb{R}^d$. Hence, the function $f(u) := u$ belongs to $L^2_{\mu^y}(\mathcal{H}; \mathcal{H}) \cap L^2_{\mu^{\tilde{y}}}(\mathcal{H}; \mathcal{H})$ and Lemma 3.7 can be applied which yields by

$$\|f\|^2_{L^2_{\mu^y}} + \|f\|^2_{L^2_{\mu^{\tilde{y}}}} \leq \frac{2}{\gamma_r} \|f\|^2_{L^2_{\mu_0}}$$

where γ_r is as in the proof of Theorem 3.20, the assertion $\|m^y - m^{\tilde{y}}\|_{\mathcal{H}} \leq K_r \, |y - \tilde{y}|^\alpha$. For the covariances of μ^y and $\mu^{\tilde{y}}$ there holds

$$\begin{aligned} \left\| C^y - \tilde{C}^y \right\| &= \left\| \mathbb{E}_{\mu^y}[u \otimes u] - \mathbb{E}_{\mu^y}[u] \otimes \mathbb{E}_{\mu^y}[u] - \left(\mathbb{E}_{\mu^{\tilde{y}}}[u \otimes u] - \mathbb{E}_{\mu^{\tilde{y}}}[u] \otimes \mathbb{E}_{\mu^{\tilde{y}}}[u] \right) \right\| \\ &\leq \left\| \mathbb{E}_{\mu^y}[u \otimes u] - \mathbb{E}_{\mu^{\tilde{y}}}[u \otimes u] \right\| + \left\| \mathbb{E}_{\mu^y}[u] \otimes \mathbb{E}_{\mu^y}[u] - \mathbb{E}_{\mu^{\tilde{y}}}[u] \otimes \mathbb{E}_{\mu^{\tilde{y}}}[u] \right\|. \end{aligned}$$

Since $g(u) := u \otimes u$ belongs to $L^2_{\mu^y}(\mathcal{H}; \mathcal{H} \otimes \mathcal{H}) \cap L^2_{\mu^{\tilde{y}}}(\mathcal{H}; \mathcal{H} \otimes \mathcal{H})$ – due to $\mu^y \in \mathcal{P}^2(\mathcal{H})$ for each $y \in \mathbb{R}^d$ – we can again apply Lemma 3.7 and get analogously to above

$$\left\| \mathbb{E}_{\mu^y}[u \otimes u] - \mathbb{E}_{\mu^{\tilde{y}}}[u \otimes u] \right\| \leq \frac{2}{\sqrt{\gamma_r}} \|g\|_{L^2_{\mu_0}} |y - \tilde{y}|^\alpha.$$

Moreover, due to $a \otimes a - b \otimes b = (a + b) \otimes (a - b)$, we get

$$\begin{aligned} \|\mathbb{E}_{\mu^y}[u] \otimes \mathbb{E}_{\mu^y}[u] - \mathbb{E}_{\mu^{\tilde{y}}}[u] \otimes \mathbb{E}_{\mu^{\tilde{y}}}[u]\| &= \|\mathbb{E}_{\mu^y}[u] + \mathbb{E}_{\mu^{\tilde{y}}}[u]\|_{\mathcal{H}} \, \|\mathbb{E}_{\mu^y}[u] - \mathbb{E}_{\mu^{\tilde{y}}}[u]\|_{\mathcal{H}} \\ &\leq \frac{2}{\gamma_r} \|\mathbb{E}_{\mu_0}[u]\|_{\mathcal{H}} \, K_r |y - \tilde{y}|^\alpha \end{aligned}$$

which yields the corresponding statement on $\|C^y - \tilde{C}^y\|$. The second part of the corollary follows analogously. □

3.4. Bayes Estimators

Although the posterior measure μ^y is the solution object to the Bayesian inference problem of inferring U from noisy observations of Y, it is usually not easy to com-

pute in practice and needs to be approximated. Moreover, when the dimension of $\mathcal{H}$ is large or infinite, visualizing, exploring or using μ^y for post-processing are demanding tasks.

More accessible quantities from Bayesian statistics than the posterior measure itself are *point estimates* for the unknown u, see, e.g., Bernardo [10]. In the Bayesian setting a point estimate is a "best guess" $\hat{u}$ of u based on the posterior knowledge, i.e., the posterior measure μ^y. Here "best" is determined by a cost function $\mathrm{c}\colon \mathcal{H} \to [0,\infty)$ which describes the loss or costs $\mathrm{c}(u-\hat{u})$ incurred when $\hat{u}$ is substituted for (the true) u for post-processing or decision making.

Definition 3.24. A mapping $\mathrm{c}\colon \mathcal{H} \to [0,\infty)$ is called a *cost function* if it satisfyies $\mathrm{c}(0) = 0$ and $\mathrm{c}(u) \leq \mathrm{c}(\lambda u)$ for any $u \in \mathcal{H}$ and $\lambda \geq 1$. For $y \in \mathbb{R}^d$ and μ^y given by (3.13) we define the *(posterior) Bayes cost* of an estimate $\hat{u} \in \mathcal{H}$ w.r.t. c by

$$\mathrm{B}_{\mathrm{c}}(\hat{u};y) := \int_{\mathcal{H}} \mathrm{c}(u-\hat{u})\,\mu^y(\mathrm{d}u).$$

We remark that also more general forms of a cost function than stated in Definition 3.24 are possible, see Berger [9] and Bernardo [10].

Definition 3.25. Under the assumptions of Definition 3.24 we define the *Bayes estimate* $\hat{u}_{\mathrm{c}}(y)$ as the minimizer

$$\hat{u}_{\mathrm{c}}(y) := \operatorname*{argmin}_{\hat{u}\in\mathcal{H}} \mathrm{B}_{\mathrm{c}}(\hat{u};y),$$

assuming a unique minimizer exists. Moreover, the *Bayes estimator* is defined as the mapping $\hat{u}_{\mathrm{c}}\colon \mathbb{R}^d \to \mathcal{H}$ which assigns to an observation $y \in \mathbb{R}^d$ the associated Bayes estimate $\hat{u}_{\mathrm{c}}(y)$.

So far, it does not seem easier to determine Bayes estimates or Bayes estimators than sampling w.r.t. the posterior, since for the former we have to solve a minimization problem involving integrals w.r.t. the posterior. However, because $\hat{u}_{\mathrm{c}}$ obtains for each $y \in \mathbb{R}^d$ the minimum of the posterior Bayes cost, it will also minimize an averaged Bayes cost.

Definition 3.26. Under the assumptions of Definition 3.24 and Assumption 3.14 the *prior Bayes cost* of a measurable mapping $\phi\colon \mathbb{R}^d \to \mathcal{H}$ is given by

$$\mathrm{B}_{\mathrm{c}}(\phi) := \mathbb{E}\left[\mathrm{B}_{\mathrm{c}}(\phi(Y);Y)\right] = \int_{\mathbb{R}^d}\int_{\mathcal{H}} \mathrm{c}(u-\phi(y))\,\mu^y(\mathrm{d}u)\,\nu_Y(\mathrm{d}y) = \mathbb{E}\left[\mathrm{c}(U-\phi(Y))\right].$$

where U and Y are as in (3.2).

Thus, assuming measurability of the Bayes estimator $\hat{u}_c$ in the following, we can characterize $\hat{u}_c$ by

$$\hat{u}_c := \operatorname*{argmin}_{\phi\colon \mathbb{R}^d \to \mathcal{H} \text{ measurable}} \mathrm{B}_c(\phi), \tag{3.20}$$

or, equivalently,

$$\mathbb{E}\left[c(U - \hat{u}(Y))\right] \leq \mathbb{E}\left[c(U - \phi(Y))\right] \qquad \forall \text{ measurable } \phi\colon \mathbb{R}^d \to \mathcal{H}. \tag{3.21}$$

Since the expectations in (3.21) are w.r.t. prior measures it is possible to determine the estimator $\hat{u}_c$ without actually computing the posterior measure μ^y. Therefore, Bayes estimators are typically easier to compute or approximate than μ^y. Next, we will recall two very common Bayes estimators.

Posterior Mean Estimator For the cost function $c(u) = \|u\|_{\mathcal{H}}^2$ the posterior Bayes cost

$$\mathrm{B}_c(\hat{u}; y) = \int_{\mathcal{H}} \|u - \hat{u}\|_{\mathcal{H}}^2 \, \mu^y(\mathrm{d}u)$$

is minimized by the posterior mean $\mathbb{E}[U|Y = y] = \int_{\mathcal{H}} u \, \mu^y(\mathrm{d}u)$. This is due to

Proposition 3.27. Let $X \in L^2(\Omega; \mathcal{H})$ and define a functional $J_X\colon \mathcal{H} \to [0, \infty)$ by

$$J_X(x) := \mathbb{E}[\|X - x\|_{\mathcal{H}}^2].$$

Then we have

$$\mathbb{E}[X] = \operatorname*{argmin}_{x \in \mathcal{H}} J_X(x).$$

Proof. Let $0 \neq x \in \mathcal{H}$ be arbitrary. Then we obtain by linearity

$$\begin{aligned} J_X(\mathbb{E}[X] + x) &= \mathbb{E}[\|X - \mathbb{E}[X] - x\|_{\mathcal{H}}^2] = \mathbb{E}[\langle X - \mathbb{E}[X] - x, X - \mathbb{E}[X] - x\rangle_{\mathcal{H}}] \\ &= \mathbb{E}[\langle X - \mathbb{E}[X], X - \mathbb{E}[X]\rangle_{\mathcal{H}}] - 2\mathbb{E}[\langle X - \mathbb{E}[X], x\rangle_{\mathcal{H}}] + \mathbb{E}[\langle x, x\rangle_{\mathcal{H}}] \\ &= J_X(\mathbb{E}[X]) - 2\langle \underbrace{\mathbb{E}[X - \mathbb{E}[X]]}_{\equiv 0}, x\rangle_{\mathcal{H}} + \langle x, x\rangle_{\mathcal{H}} > J_X(\mathbb{E}[X]). \end{aligned}$$

Thus, $\mathbb{E}[X]$ is the unique minimizer of J_X. □

Remark 3.28. If we consider Banach spaces $\mathcal{X}$ instead of Hilbert spaces $\mathcal{H}$ then Proposition 3.27 does not hold anymore. For example, let $\mathcal{X} = \mathbb{R}^2$, $\|v\|_{\mathcal{X}} = |v_1| + |v_2|$ and $X = (X_1, X_2)$ with independent random variables X_1, X_2 such that

$$\mathbb{P}(X_i = -1) = p_i, \quad \mathbb{P}(X_i = 1) = 1 - p_i, \qquad i = 1, 2.$$

Here $\mathbb{E}[X]$ minimizes $\mathbb{E}[\|X - v\|_{\mathcal{H}}^2]$ iff $p_1 = p_2 = 0.5$. In fact, one can show $\mathbb{E}[X] = \operatorname{argmin}_{v \in \mathcal{X}} \mathbb{E}[\|X - v\|_{\mathcal{X}}^2]$ if X is distributed symmetrically w.r.t. its mean, i.e., if there holds $\mathbb{P}(X - \mathbb{E}[X] \in A) = \mathbb{P}(\mathbb{E}[X] - X \in A)$ for all $A \in \mathcal{B}(\mathcal{X})$.

Definition 3.29. The *conditional mean estimator* or *posterior mean estimator* is defined by

$$u_{\mathrm{CM}}(y) := \int_{\mathcal{H}} u \, \mu^y(\mathrm{d}u), \qquad y \in \mathbb{R}^d, \tag{3.22}$$

provided that $\mu^y \in \mathcal{P}^1(\mathcal{H})$ for each $y \in \mathbb{R}^d$.

By Corollary 3.23 we immediately obtain

Proposition 3.30. Let the assumptions of Corollary 3.23 be satisfied, then the mapping $u_{\mathrm{CM}} \colon \mathbb{R}^d \to \mathcal{H}$ given by (3.22) is continuous.

Thus, the posterior mean estimator is in particular measurable. Let us recall that $\mathbb{E}[U|Y]$ is the best approximation of U in $L^2(\Omega, \sigma(Y), \mathbb{P}; \mathcal{H})$ w.r.t. the norm in $L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathcal{H})$. Hence, we view the posterior mean estimator $u_{\mathrm{CM}}(y)$ as the best L^2-approximation to U based on the information $\sigma(Y)$ provided by observations of Y.

Remark 3.31. If $\mathcal{H} = \mathbb{R}^n$ and μ^y is unimodal, then the posterior mean is the Bayes estimator for any symmetric, convex cost function c, see Lewis et al. [110, Remark 16.2.2] or Speyer [162].

Maximum A Posteriori Estimator Another common estimator in Bayesian statistics is the *maximum a posteriori estimator* (MAPE) u_{MAP}. For finite-dimensional spaces $\mathcal{H} = \mathbb{R}^n$ and priors μ_0 which are absolutely continuous w.r.t. the Lebesgue measure, i.e., $\mu_0(\mathrm{d}u) = \pi_0(u) \, \mathrm{d}u$, the MAPE is defined as

$$u_{\mathrm{MAP}}(y) := \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\Phi(u, y) - \log \pi_0(u) \right), \tag{3.23}$$

provided the minimum exists for all $y \in \mathbb{R}^d$. Thus, the estimate $u_{\mathrm{MAP}}(y)$ is the point in $\mathbb{R}^n$ where the posterior density obtains its maximum. We, therefore, can view $u_{\mathrm{MAP}}(y)$ as the Bayesian counterpart of the *maximum likelihood estimate* u_{ML} from Frequentist statistics

$$u_{\mathrm{ML}}(y) := \operatorname*{argmin}_{u \in \mathbb{R}^n} \Phi(u, y).$$

For the corresponding cost function for which u_{MAP} provides the Bayes estimator we refer to Burger and Lucka [24]. In the older literature the MAPE is usually

introduced as the minimizer of the limit of a sequence of cost functions, see e.g., Lewis et al. [110, Section 16.2]. Burger and Lucka [24] were the first to show that it can be defined as a proper Bayes estimator in case of linear forward maps G.

If $\mathcal{H}$ is infinite dimensional there exists no posterior density w.r.t. the Lebesgue measure, since the latter does also not exist. However, in case of Gaussian priors $\mu_0 = N(m, C)$ and under certain assumptions on Φ it is shown by Dashti et al. [41], that the MAPE can be defined analogously to above by

$$u_{\mathrm{MAP}}(y) = \operatorname*{argmin}_{u \in \mathrm{rg}(C^{1/2})} \Phi(u, y) - \|C^{-1/2}u\|_{\mathcal{H}}. \tag{3.24}$$

3.5. Relation to Regularizational Approaches to Inverse Problems

In the deterministic setting the usual approach to solve the inverse problem of reconstructing u given y as in (3.1) is by regularized least squares problems. The reason for the least squares approach is that in general $y \notin \mathrm{rg}(G)$ due to the noise ε. Hence, determining $u = G^{-1}(y)$ is replaced by

$$u = \operatorname*{argmin}_{v \in \mathcal{D}_G} |\mathrm{y} - G(v)|^2,$$

but the solution to the least squares problem does, in general, not depend continuously on the data y. Therefore, the problem is regularized by incorporating additional prior information about the desired u by introducing a regularizing functional $\mathrm{R}\colon \mathcal{H} \to [0, \infty]$, see, e.g., Engl et al. [52], and to solve for

$$u_\alpha = \operatorname*{argmin}_{v \in \mathcal{D}_G} |\mathrm{y} - G(v)|^2 + \alpha\, \mathrm{R}(v), \tag{3.25}$$

where $\alpha \in [0, \infty)$ serves as a regularization parameter. For example, we may regularize by restricting u to a subset or subspace $\tilde{\mathcal{H}} \subset \mathcal{H}$ which can be realized by using a stronger norm than $\|\cdot\|_{\mathcal{H}}$ as the regularizing functional R.

We immediately notice the common structure between the regularized solution u_α in (3.25) and the MAP estimate given in (3.23). In particular, if $\mathcal{H} = \mathbb{R}^n$ and the regularizing functional $\mathrm{R}\colon \mathbb{R}^n \to [0, \infty)$ satisfies

$$\int_{\mathbb{R}^n} \exp\left(-\frac{\alpha}{\sigma^2}\,\mathrm{R}(u)\right) \mathrm{d}u < +\infty,$$

then the regularized solution u_α coincides with the MAP estimate $u_{\mathrm{MAP}}(y)$ for a

prior

$$\mu_0(\mathrm{d}u) \propto \exp\left(-\frac{\alpha}{\sigma^2}\,\mathrm{R}(u)\right)\,\mathrm{d}u$$

and $\varepsilon \sim N(0, \sigma^2 I)$. The same holds in infinite dimensions if the regularizing functional takes a form

$$\mathrm{R}(u) = \begin{cases} \|A(u - u_{\mathrm{ref}})\|_{\mathcal{H}}^2, & u - u_{\mathrm{ref}} \in \mathcal{D}_A, \\ \infty, & \text{otherwise}, \end{cases}$$

where $\mathcal{D}_A$ denotes the domain of the operator $A\colon \mathcal{D}_A \to \mathcal{H}$. If A is the inverse of a bounded, positive, nonsingular and self-adjoint Hilbert-Schmidt operator, i.e., $A^{-2} \in \mathcal{L}_+^1(\mathcal{H})$, then the resulting regularized solution u_α coincides again with the MAP estimate given a Gaussian prior $\mu_0 = N(u_{\mathrm{ref}}, \frac{\alpha}{\sigma^2} A^{-2})$ and $\varepsilon \sim N(0, \sigma^2 I)$ as before. We refer to Kaipio and Somersalo [95], Stuart [167] and Dashti and Stuart [43] for a more detailed discussion about the relation of both approaches.

Broadly speaking, the Bayesian approach may be viewed as a probabilistic regularization where prior information on u is modeled by a prior measure μ_0 on the Hilbert space $\mathcal{H}$. Thus, a quantitative preference of some solutions u over others is given by assigning higher and lower probabilities. However, the Bayesian approach is not only dedicated to the task of *identifying* one specific $u \in \mathcal{H}$ which explains the observed data best. The goal in Bayesian inference is to learn from the observed data in a statistical or probabilistic sense by adjusting our prior belief μ_0 about u in accordance with the available data. Of couse, as we have seen, the task of identification may also be achieved within the Bayesian framework by Bayes estimates and Bayes estimators. But besides this, the Bayesian approach provides also a quantification of our remaining uncertainty on u in terms of the posterior measure.

3.6. Computational Methods for Bayesian Inference

In special cases, the posterior measure is given in a closed form. For instance, if the forward map G is linear and the prior μ_0 as well as the noise distribution ν_ε are Gaussian measures, then the resulting posterior is also Gaussian and explicit formulas for its mean and covariance are available, see Theorem 4.3 in the next chapter. Furthermore, in finite dimensions $\mathcal{H} = \mathbb{R}^N$ given the noise distribution we may sometimes choose a *conjugate prior* which means that the resulting posterior density function (w.r.t. the Lebesgue measure) belongs to the same family of probability density functions as the prior. In these cases, the corresponding parameters of the posterior density, e.g., mean, variance, skewness or shape, are often given in an

analytic form. We refer to Hoff [89] for more details. Aside from these special situations the posterior μ^y and its statistics can only be approximated. In the following we provide a brief overview of existing numerical methods for Bayesian inference – with no claim of exhaustiveness.

Methods for approximate sampling from the posterior. Perhaps the simplest and most natural idea to provide approximations to a measure is by empirical measures, i.e., by sampling. Typically, the posterior measure is of a complicated form and its density (w.r.t. a reference or prior measure) only available upon pointwise evaluation. Thus, direct sampling is unfeasible or even impossible and approximate sampling methods have to be applied.

Maybe the most established method for this purpose is the *Markov Chain Monte Carlo* method which we will discuss in detail in Chapter 5. The basic idea of MCMC is to construct a sequence of random variables which converge in distribution to the posterior measure.

Besides MCMC another common method in Bayesian inference are *particle filters* or *sequential Monte Carlo methods*, see, e.g., Kaipio and Somersalo [95, Section 4.3] or Doucet et al. [48]. These methods are well suited for *sequential data assimilation*, e.g., when the observational data arrives sequentially in time and, thus, a recursive comptutation or approximation of the posterior is desirable. The basic principle behind particle filters is as follows: we generate samples according to the prior and assign to each sample a weight according to posterior density evaluated for this sample. If new data arrives, we can employ the previous posterior weighted samples as initial ensemble and update their weights by multiplying them with the likelihood function associated to the new observations. For more details, we refer to the references above.

One extension of particle filters are *Gaussian mixture filters*, see, e.g., Stordal et al. [166]. The idea there is to approximate the posterior density by a weighted mean of Gaussian kernels located at each sample and in addition to the weights also the location of the samples are updated according to the posterior.

A further technique for sampling from the posterior is presented by El Moselhy and Marzouk [51]: a mapping $F\colon \mathcal{H} \to \mathcal{X}$ is constructed in such a way that $F(U) \sim \mu^y$ for a random variable $U \sim \mu_0$. Given F, which is obtained by solving an optimal transport problem, samples according to μ^y can then easily be generated by evaluating F for samples from the prior. The mapping F can usually be only approximated yielding, thus, samples which are again only approximately distributed according to μ^y.

Numerical integration w.r.t. the posterior. Since many posterior statistics such as mean, covariance or even probabilites, are given by expectations w.r.t. the posterior, we can apply numerical integration methods to compute those. Concerning probabilites the integration would involve an indicator function, thus, numerical integration which requires some smoothness of the integrand would not be appropriate. Nonetheless, for posterior moments of smooth functionals of interest efficient numerical quadrature methods have been developed in recent years. Often the idea is to perform quadrature w.r.t. the known prior measure and to include the unnormalized posterior density $\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp(\Phi(u,y))$ in the integrand. Of course, then the normalizing constant has to be computed as well by quadrature. We mention Schwab and Stuart [157] and Schillings and Schwab [154] who investigated sparse quadrature formulas for Bayesian inference. Due to the assumed smoothness of the likelihood $\mathrm{e}^{-\Phi(u,y)}$ w.r.t. u, these methods can yield faster convergence rates than Monte Carlo or Markov chain Monte Carlo integration and are also suited to infinite dimensions. Beside these works a recent paper by Scheichl et al. [151] presents *quasi* and *multilevel Monte Carlo* methods for estimating posterior expectations.

Methods for Bayesian state estimation. In dynamical systems where the underlying state is unknown but noisy observations of related quantites are available, the main interest is often to estimate the current state of the system. In the Bayesian framework this corresponds to computing Bayes estimates for state. A well-known and widely applied computational method is the *Kalman filter* and its extensions which provide approximations for the posterior mean estimate and which we will consider in detail in Chapter 4. Similar to particle filters these methods are adapted to sequentially arriving data and allow for a recursive computation of the estimate.

Alternatively, one can use the MAP estimate for the unknown state of a dynamical system. By definition we can apply numerical optimization methods to compute the MAP estimate, see, e.g., Vogel [173]. Two popular algorithms in, e.g., weather prediction, are *3DVar* and *4DVar*. Both methods compute the MAP estimate, but the difference between them is that 3DVar treats the sequentially arriving data recursively, while 4DVar performs the optimization w.r.t. the entire data set at once, see also Lewis et al. [110] or Law and Stuart [105].

Chapter 4

Kalman Filter Methods for Bayesian Inference

The content of this chapter is based on the publications [56, 57], but the way of presentation has been modified and many details and remarks added.

This chapter is devoted to Kalman filter methods and their application to Bayesian inference for an unknown u in a separable Hilbert space $\mathcal{H}$ given noisy observations

$$y = G(u) + \varepsilon \in \mathbb{R}^d \tag{4.1}$$

where $G\colon \mathcal{D}_G \to \mathbb{R}^d$, $\mathcal{D}_G \subseteq \mathcal{H}$, denotes the forward map and ε the noise. We will follow the same notations and assumptions as in the previous chapter.

Kalman filter techniques are often applied in practice, since they are easy to implement and require less computational work than, e.g., MCMC methods, in terms of fewer evaluations of the forward map G. In particular, if G involves solving a PDE, then evaluations of G constitute the main computational work for Bayesian inference. Moreover, filtering methods are well adapted to inference in dynamical systems where observational data arrives sequentially in time, see, e.g., Stuart [167, Section 5.4] for a more detailed discussion of this issue.

The *Kalman filter* (KF) was introduced by Kalman [97] and developed for state estimation in linear dynamics. Already in Kalman's seminal paper the principle for deriving the filter equations was based on minimizing the average cost of an estimate w.r.t. quadratic loss functions, i.e., Kalman employed the terminology of Bayes estimators. Later, the KF was extended to nonlinear dynamics resulting in the *extended Kalman filter* (EKF), see, e.g., Kaipio and Somersalo [95, Section 4.2.2]. The EKF is based on local linearizations of the underlying nonlinear dynamics or forward maps. A few years later, Evensen [59] proposed another extension of the KF to nonlinearities, called the *ensemble Kalman filter* (EnKF). This method employs

Monte Carlo sampling according to a prior measure for the unknown state, propagates the ensemble through the dynamics and updates each ensemble member given observational data by a procedure analogous to the KF. The mean of the ensemble then serves as an estimate for the unknown state. The EnKF is widely used by practicioners and seems to perform well for state and parameter estimation, see, e.g., Evensen [62].

In recent years, the EnKF has also drawn the attention of the growing UQ community and has been investigated in the context of inverse problems and Bayesian inference by, e.g., Iglesias et al. [93, 92] and Law and Stuart [105]. Since the EnKf yields not only a single estimate but an *analysis ensemble* of states it is tempting to use the latter for quantifying the uncertainty about the unknown state. Although it is known, see, e.g., Apte et al.[3], Le Gland et al. [75] or Evensen and Van Leeuwen [63], that the analysis ensemble generated by the EnKF is in general not distributed according to the posterior measure defined by Bayes' rule, the relation between the empirical measure associated with the EnKF analysis ensemble and the posterior measure is yet missing in the literature.

As a further improvement of the KF, the authors of [149, 148, 17, 129, 144, 143] have combined the idea of the EnKF with the computationally attractive representation of random variables by a polynomial chaos expansion.Their method represents the uncertain state not by an ensemble drawn according to a prior measure but by a truncated PCE of a prior distributed random variable. The PC coefficients are then updated — again in a similar way as the original KF update — given the observational data y which results in a vector of *analysis PC coefficients*. We will refer to this approach in the following as the *Polynomial Chaos Kalman filter* (PCKF). It was mainly the study of the PCKF which led to the content of this chapter. Although the authors of the above mentioned papers provide a motivation for deriving their algorithm, they do not clearly characterized how to understand the random variable determined by the analysis PC coefficients, i.e., if its distribution may provide a reasonable approximation to the posterior measure.

The results presented in this chapter try to fill this gap. In particular, we clarify the stochastic model underlying the EnKF and PCKF in the special case of a single update given data y as in (4.1). We show that both methods, the EnKF and the PCKF, provide different types of approximations to the same random variable, which we term the *analysis variable*. Specifically, the empirical distribution of the EnKF ensemble provides an approximation to the distribution of the analysis variable whereas the PCKF constructs an approximation to the PCE of the analysis variable. This will be made precise by two convergence results in Section 4.2 where we show that in the large ensemble limit and the large polynomial basis limit the outcome of the

EnKF and the PCKF will converge to the distribution and the PCE of the analysis variable, respectively. To the authors' knowledge, a convergence analysis for PCKF is missing in the literature so far. In case of the EnKF, convergence results were already given by Le Gland et al. [75], Mandel et al. [116] and Law et al. [106]. In each of these works the general setting was sequential data assimilation for a discrete-time dynamical system with linear or (locally) Lipschitz continuous drift and linear observation operators, but the convergence analysis was always done w.r.t. a different type of convergence (details will be given later). Our result comes closest to that of Le Gland et al. [75], but we consider an abstract Hilbert space setting and a nonlinear parameter-to-observation map G which need not satisfy a Lipschitz condition.

Furthermore, we will provide an explicit characterization of the analysis variable in the context of Bayesian inference in Section 4.3. Therefore, we will determine how the analysis ensemble of the EnKF and the analysis PCE of the PCKF relate to the posterior measure and Bayes estimators. In fact, both fail to approximate the posterior measure and rather approximate a certain Bayes estimator and its prior error. This insight explains the observations made in numerical experiments by Law and Stuart [105] and yield implications for the usage of Kalman filter methods in uncertainty quantification

4.1. The Kalman Filter and its Generalizations

In this section we introduce the classical Kalman filter as well as the Ensemble and Polynomial Chaos Kalman filter. Although these methods are desigend for data assimilation in discrete-time dynamical systems, we will apply them to the nonlinear Bayesian inference problem of inferring $U \sim \mu_0$ given the event $Y = y$ where $y \in \mathbb{R}^d$ and

$$Y := G(U) + \varepsilon \tag{4.2}$$

with $G\colon \mathcal{D}_G \to \mathbb{R}^d$ and $\varepsilon \sim \nu_\varepsilon$ as in Assumption 3.14 in Chapter 3. We assume an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ in the following and require

Assumption 4.1. Assumption 3.14 holds with $\mu_0 \in \mathcal{P}^2(\mathcal{H})$ and $\nu_\varepsilon \in \mathcal{P}^2(\mathbb{R}^d)$ as well as

$$\int_{\mathcal{H}} |G(u)|^2 \, \mu_0(\mathrm{d}u) < \infty.$$

This assumption implies the existence of the first and second moments of U, ε and Y appearing in the model (4.2).

4.1.1. The Kalman Filter

The Kalman filter [97] is a well-known method for sequential state estimation for incompletely observable, linear discrete-time dynamics of the form

$$\begin{aligned} U_n &= F_n U_{n-1} + \eta_n, \qquad n = 1, 2, \dots, \\ Y_n &= G_n U_n + \varepsilon_n, \qquad n = 1, 2, \dots, \end{aligned} \tag{4.3}$$

where $(U_n)_{n\in\mathbb{N}}$ denotes the unobservable state and $(Y_n)_{n\in\mathbb{N}}$ the observable process. The linear operators $F_n \in \mathcal{L}(\mathcal{H})$ and $G_n \in \mathcal{L}(\mathcal{H}, \mathbb{R}^d)$ are mappings in state space and from state to observation space, respectively, and the noise processes $(\eta_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ on $\mathcal{H}$ and $\mathbb{R}^d$, respectively, are usually assumed to have zero mean and known covariances. In addition, the mean and covariance of U_0 need to be known and the random variables U_0, η_n, ε_n, $n \in \mathbb{N}$, are supposed to be mutually independent. The filtering problem consists then of computing an estimate of the current unknown state U_n given the sequentially arrived observations $Y_1 = y_1, \dots, Y_n = y_n$. Particularly, we ask for the minimum variance estimate $u_n^a = u_n^a(y_1, \dots, y_n)$ defined by

$$\mathbb{E}\left[\|U_n - u_n^a(Y_1, \dots, Y_n)\|_{\mathcal{H}}^2\right] \leq \mathbb{E}\left[\|U_n - \phi(Y_1, \dots, Y_n)\|_{\mathcal{H}}^2\right]$$

for all measurable $\phi\colon \mathbb{R}^{d\times n} \to \mathcal{H}$. The superscript a in the notation of the estimate u_n^a refers to the term *analysis*, since in the filtering literature, particularly in the literature on the EnKF, the incorporation of observational data is called *analysis step*. As we know from the previous chapter, the minimal variance estimate u_n^a as characterized above corresponds to the posterior mean of U_n given $Y_1 = y_1, \dots, Y_n = y_n$. Under the assumption that U_0, η_n, ε_n, $n \in \mathbb{N}$ are Gaussian, Kalman [97] then derived a coupled system of recursive equations for the estimates u_n^a and their error covariances

$$C_n^a := \mathrm{Cov}(U_n - u_n^a(Y_1, \dots, Y_n)), \qquad n \in \mathbb{N}.$$

The recursive structure of the Kalman filter represents a main advantage, because for computing the estimate u_n^a we only have to use the former estimate u_{n-1}^a, its error covariance C_{n-1}^a and the new observation y_n — besides F_n, G_n and the covariances of η_n and ε_n — rather than taking into account all previous observations $y_1, \dots, y_n$. Again we refer to Stuart [167, Section 5.4] for details about that issue. However, we will not state the Kalman filter equations for dynamical system here and only provide the reasoning for deriving the equation for $n = 1$ in the following paragraph. For a comprehensive introduction and discussion of the Kalman filter we refer to Catlin [26] and Simon [159].

The Kalman filter for a time-independent Bayesian inference problem. As mentioned above we will focus on the application of the Kalman filter and its generalizations to time-independent Bayesian inference problems of the form (4.2). For applying the classical Kalman filter, we restrict the forward $G\colon \mathcal{H} \to \mathbb{R}^d$ to be linear:

$$Y = GU + \varepsilon, \qquad (U, \varepsilon) \sim \mu_0 \otimes \nu_\varepsilon \tag{4.4}$$

where we assume $\mu_0 = N(0, C)$ and $\nu_\varepsilon = N(0, \Sigma)$ are zero-mean Gaussian measures on $\mathcal{H}$ and $\mathbb{R}^d$, respectively, with invertible covariances $C \in \mathcal{L}(\mathcal{H})$ and $\Sigma \in \mathbb{R}^{d\times d}$.

Remark 4.2. We note that (4.2) or (4.4) can be seen as one step of the dynamical system (4.3) for $F_n \equiv I$, $\eta_n \equiv 0$ and $G_n = G$. Conversely, the estimation problem for the initial U_0 in (4.3) given $Y_1 = y_1, \dots, Y_n = y_n$ can be reformulated as (4.4) with $Y := (Y_1, \dots, Y_n)^\top$ and $G := (G_1, \dots, G_n)$ where $G_j = F_j \circ \cdots \circ F_1$. However, the task of inferring the current state U_n given $Y_1 = y_1, \dots, Y_n = y_n$ does, in general, not fit the form of (4.4), since there does not necessarily exist a mapping relating U_n and Y_j for $j \leq n$. Only if the dynamics F_n are invertible for all $n \in \mathbb{N}$, then we have $Y_j = G_j F_j^{-1} \cdots F_n^{-1} U_n + \tilde{\varepsilon}_j$ where $\tilde{\varepsilon}_j$ is then the sum of ε_j and $G_j F_j^{-1} \cdots F_k^{-1} \eta_k$, $j \leq k \leq n$.

Given the model (4.4) for the observable random variable Y it follows by Proposition 2.20 that (U, Y) is jointly Gaussian

$$\begin{pmatrix} U \\ Y \end{pmatrix} = \begin{pmatrix} I & 0 \\ G & I \end{pmatrix} \begin{pmatrix} U \\ \varepsilon \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C & CG^* \\ GC & GCG^* + \Sigma \end{pmatrix} \right).$$

Moreover, it is well known that the conditional distribution of jointly Gaussian random variables is again Gaussian, see, e.g., Mandelbaum [117]:

Theorem 4.3 ([117, Corollary 2])**.** Let X and Y be two Gaussian random variables on separable Hilbert spaces $\mathcal{H}$ and $\mathcal{Y}$ with mean m_X and m_Y and covariance operators C_X and C_Y, respectively, where C_Y is invertible. Further let $C_{XY} := \operatorname{Cov}(X, Y)$ and $C_{YX} := \operatorname{Cov}(Y, X)$. Then a regular conditional distribution $\mu_{X|Y}\colon \mathcal{Y} \to \mathcal{P}(\mathcal{H})$ of X given Y is given by

$$\mu_{X|Y}(y) = N\left(m_X + C_{XY} C_Y^{-1}(y - m_Y), C_X - C_{XY} C_Y^{-1} C_{YX} \right).$$

Mandelbaum [117] actually considers two Gaussian random variables on the same Hilbert space $\mathcal{H}$, but the result can be extended to the setting in Theorem 4.3. Recall now, that the Kalman filter was designed to compute the minimial vari-

ance estimate of the unknown U given the observation $Y = y$. Since this estimate is the posterior mean of U given $Y = y$ we end up with

Definition 4.4 (Kalman Filter)**.** Given the model (4.4), the *Kalman filter* for estimating U given $Y = y$ is as follows

1. **Initialization:** Set as inital estimate $u_0 := \mathbb{E}[U]$ and as initial error covariance $C_0 := \mathrm{Cov}(U)$.

2. **Forecast:** Set as forecast $y_0 := Gu_0$ and as forecast covariances $C_Y := GC_0G^* + \Sigma$ and $C_{UY} := C_0G^*$.

3. **Analysis:** Compute $\mathrm{K} := C_{UY}C_Y^{-1}$ and the *analysis estimate* and *error covariance*

$$\begin{aligned} u^a &:= u_0 + \mathrm{K}(y - y_0), \\ C^a &:= C_U - \mathrm{K}C_{UY}^*, \end{aligned} \tag{4.5}$$

where the operator $\mathrm{K}\colon \mathbb{R}^d \to \mathcal{H}$ is called *Kalman gain*.

Note, that C_Y^{-1} exists, since Σ is positive definite and GC_0G^* positive semi-definite. Moreover, by linearity we easily see that there holds

$$y_0 = \mathbb{E}[Y], \qquad C_Y = \mathrm{Cov}(Y), \qquad C_{UY} = \mathrm{Cov}(U, Y),$$

as well as

$$u^a = u^a(y) = \mathbb{E}[U \mid Y = y], \qquad C^a = \mathrm{Cov}(U \mid Y = y),$$

i.e., the estimate $u^a = u^a(y)$ generated by the Kalman filter is the posterior mean and the error covariance C^a coincides with posterior covariance of U given $Y = y$ as stated in Theorem 4.3. We observe also that the covariance C^a does not depend on the data y by construction — a fact which will reappear in our Bayesian interpretation of the EnKF and PCKF.

Remark 4.5 (On error covariances)**.** The reason why all the appearing covariances in the Kalman filter algorithm are called error covariances is that they can be understood as covariances of the errors of the associated estimates, i.e.,

$$\begin{aligned} C_0 &= \mathrm{Cov}(U) = \mathrm{Cov}(U - u_0), & C_Y &= \mathrm{Cov}(Y) = \mathrm{Cov}(Y - y_0), \\ C_{UY} &= \mathrm{Cov}(U, Y) = \mathrm{Cov}(U - u_0, Y - y0), & C^a &= \mathrm{Cov}(U - u^a(Y)), \end{aligned}$$

where the last equality follows by a straight forward calculation noting that $u^a(Y) = u_0 + \mathrm{K}(Y - y_0)$:

$$\begin{aligned}\mathrm{Cov}(U - u^a(Y)) &= \mathrm{Cov}(U,U) - \mathrm{Cov}(U,u^a(Y)) - \mathrm{Cov}(u^a(Y),U) \\ &\quad + \mathrm{Cov}(u^a(Y),u^a(Y)) \\ &= C_0 - \mathrm{Cov}(U,Y)\mathrm{K}^* - \mathrm{K}\,\mathrm{Cov}(Y,K) + \mathrm{K}\,\mathrm{Cov}(Y)\mathrm{K}^* \\ &= C_0 - C_{UY}C_Y^{-1}C_{UY}^* = C^a.\end{aligned}$$

We see that by incorporating the observations y into the estimate for the unknown we obtain an improvement in the error covariance, since $\mathrm{K}C_{UY}^*$ is a positive operator. Hence, the trace of C^a — which coincides with the mean squared error $\mathbb{E}\left[\|U - u^a(Y)\|_{\mathcal{H}}^2\right]$ — is smaller than the trace of C_0 — which coincides with the mean squared error $\mathbb{E}\left[\|U - u_0\|_{\mathcal{H}}^2\right]$.

Since the Kalman filter yields the posterior mean and the posterior covariance of U given $Y = y$ in case of model (4.4) and since the posterior is then Gaussian and determined by its first two moments, one can say, that the Kalman filter also provides the posterior distribution of U given $Y = y$. However, without the assumption that μ_0 and ν_ε are Gaussian the Kalman filter will, in general, not yield the first two posterior moments, nor is the posterior measure necessarily Gaussian. In that case the outcome of the Kalman filter represents the linear minimal variance estimate for U given $Y = y$ where the term linear refers to the dependence on y. This was already mentioned in the original work by Kalman [97]. Again we will recall this fact later in Section 4.3.

Remark 4.6 (On the extension to dynamical systems)**.** In case of a dynamical system (4.3) the estimate u_n^a of U_n given $Y_1 = y_1, \ldots, Y_n = y_n$ is propagated through the dynamics to serve as initial estimate for the next time step, i.e., $F_n u_n^a$ would be the initial estimate for U_{n+1} before taking into account the new observation y_{n+1} of $Y_{n+1} = G_{n+1}U_{n+1} + \varepsilon_{n+1}$ in a similar procedure as explained above. The inital error covariance of the estimate $F_n u_n^a$ is then given by $F_n C_n^a F_n^* + \mathrm{Cov}(\eta_n)$.

4.1.2. The Ensemble Kalman Filter

The Ensemble Kalman filter was first introduced by Evensen [59] and its final form was published by Burgers et al. [25]. The EnKF extends the Kalman filter procedure to nonlinear dynamical systems (and nonlinear observation operators) employing an ensemble methodology. Since its introduction the EnKF has been investigated and evaluated in many publications, see, e.g., [60, 62, 61, 122]. However, the focus is

usually on its application to state or parameter estimation, see Evensen [62], rather than for Bayesian inference and uncertainty quantification. Again, we focus on the application of the EnKF to the Bayesian inference problem (4.2).

Definition 4.7 (Ensemble Kalman Filter)**.** Given the model (4.2), the *Ensemble Kalman filter* for estimating U given $Y = y$ is as follows:

1. **Initial ensemble:** Choose a sample size $M \in \mathbb{N}$ and draw samples $u_1, \dots, u_M$ of $U \sim \mu_0$. Set $\boldsymbol{u} := (u_1, \dots, u_M)$.

2. **Forecast:** Draw M samples $\varepsilon_1, \dots, \varepsilon_M$ of $\varepsilon \sim \nu_\varepsilon$, compute

$$y_j := G(u_j) + \varepsilon_j, \qquad j = 1, \dots, M,$$

   and set $\boldsymbol{y} := (y_1, \dots, y_M)$. Compute the empirical covariances $\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{y})$ and $\mathrm{Cov}(\boldsymbol{y})$, i.e.,

$$\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{y}) = \frac{1}{M-1} \sum_{j=1}^{M} (u_j - \bar{u}_M) \otimes (y_j - \bar{y}_M),$$

   where $\bar{u}_M = \frac{1}{M}(u_1 + \cdots + u_M)$ and $\bar{y}_M = \frac{1}{M}(y_1 + \cdots + y_M)$ and analogously for $\mathrm{Cov}(\boldsymbol{y}) = \mathrm{Cov}(\boldsymbol{y}, \boldsymbol{y})$.

3. **Analysis:** Compute $\tilde{\mathrm{K}}_M := \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{y}) \, \mathrm{Cov}(\boldsymbol{y})^{-1}$ and

$$u_j^a := u_j + \tilde{\mathrm{K}}_M (y - y_j), \qquad j = 1, \dots, M, \tag{4.6}$$

   and set $\boldsymbol{u}^a := (u_1^a, \dots, u_M^a)$ as well as $\tilde{C}_M^a := \mathrm{Cov}(\boldsymbol{u}^a)$.

The ensemble $\boldsymbol{u}^a$ is called *analysis ensemble*.

In Definition 4.7 we simply assume the invertibility of the empirical covariance $\mathrm{Cov}(\boldsymbol{y})$ which requires, e.g., $M > d$. Furthermore, we highlight that $\mathbb{P}$-almost surely there holds $u_j \in \mathcal{D}_G$ for all $j = 1, \dots, M$. Thus, the ensemble $\boldsymbol{y}$ generated in the forecast step of the EnKF consists of samples of Y by construction. Furthermore, the empirical mean

$$\bar{u}_M^a = \frac{1}{M}(u_1^a + \cdots + u_M^a)$$

of the analysis ensemble $\boldsymbol{u}^a$ serves as an estimate for the unknown U and the empirical covariance

$$\mathrm{Cov}(\boldsymbol{u}^a) = \frac{1}{M-1} \sum_{j=1}^{M} (u_j^a - \bar{u}_M^a) \otimes (u_j^a - \bar{u}_M^a)$$

of $\boldsymbol{u}^a$ is usually viewed as the error covariance of the estimate. The outcome of the EnKF, i.e., the analysis ensemble, is random due to the sampling involved in the algorithm. Therefore, in the subsequent convergence analysis of the EnKF we will consider $\mathcal{H}$-valued random variables U_j^a, $j = 1, \dots, M$, such that the analysis ensemble generated by the EnKF algorithm is a realization of $\boldsymbol{U}^a := (U_1^a, \dots, U_M^a)$. Moreover, we introduce the empirical measure associated with the (random) analysis ensemble.

Definition 4.8 (Empirical analysis measure)**.** Let U_j^a for $j = 1, \dots, M$ denote $\mathcal{H}$-valued random variables such that samples of $\boldsymbol{U}^a := (U_1^a, \dots, U_M^a)$ are generated by the EnKF algorithm as described in Definition 4.7. Then the *empirical analysis measure* is defined as

$$\tilde{\mu}_M^a(\mathrm{d}u) := \frac{1}{M} \sum_{j=1}^{M} \delta_{U_j^a}(\mathrm{d}u), \tag{4.7}$$

where $\delta_{u_j^a}$ denotes the Dirac measure at $u_j^a \in \mathcal{H}$.

Empirical measures such as $\tilde{\mu}_M^a$ in (4.7) are by definition *random probability measures*, i.e., $\tilde{\mu}_M^a(A)$ is a real-valued random variable for each $A \in \mathcal{B}(\mathcal{H})$. We will also investigate to which (deterministic) probability measure the empirical analysis measures $\tilde{\mu}_M^a$ converge in the large ensemble limit $M \to \infty$ in Section 4.2.2.

Remark 4.9 (On extension to dynamical systems)**.** For dynamical systems such as (4.3), each member u_j^a of the analysis ensemble $\boldsymbol{u}^a$ would be propagated through the system dynamics and the resulting ensemble would then be used as the initial ensemble for incorporating the new observational data.

4.1.3. The Polynomial Chaos Kalman Filter

Saad et al. [149, 148] as well as later Blanchard et al. [17] and Rosić et al. [143, 144] proposed a sampling-free Kalman filtering scheme for nonlinear systems. Rather than updating samples of the unknown U, the updating is carried out for the coefficient vector of a polynomial chaos expansion of the unknown U — we refer to Section 2.3.2 and Definition 2.42 for the introduction of PCEs. This approach necessitates the construction of a PCE for $U \sim \mu_0$ based on certain real-valued basis random variables $\xi_m \colon \Omega \to \mathbb{R}$, $m \in \mathbb{N}$. To this end, we can apply Theorem 2.21: since due to Assumption 4.1 we have $U \in L^2(\Omega; \mathcal{H})$ and $\varepsilon \in L^2(\Omega; \mathbb{R}^d)$, there exist mutually uncorrelated mean-zero random variables $\xi_m^{(u)} \in L^2(\Omega; \mathbb{R})$ for $m \in \mathbb{N}$ and $\xi_k^{(\varepsilon)} \in L^2(\Omega; \mathbb{R})$ for $k = 1, \dots, d$, as well as $\phi_m \in \mathcal{H}$, $m \in \mathbb{N}$, and $e_k \in \mathbb{R}^d$,

$k = 1, \dots, d$, such that there holds

$$U = \mathbb{E}[U] + \sum_{m=1}^{\infty} \phi_m \xi_m^{(u)}, \qquad \varepsilon = \mathbb{E}[\varepsilon] + \sum_{k=1}^{d} e_m \xi_k^{(\varepsilon)},$$

$\mathbb{P}$-a.s. as well as in $L^2(\Omega; \mathcal{H})$ and $L^2(\Omega; \mathbb{R}^d)$, respectively. These abstract Karhunen-Loève expansions are already PCEs which consist only of constant and linear polynomials in $\xi_m^{(u)}$ and $\xi_m^{(\varepsilon)}$, respectively. However, it is not always easy to compute the corresponding $\phi_m \in \mathcal{H}$, for example, if U itself is the solution of a PDE with random fields as coefficients. Therefore, and in order to ease the notation we make the following

Assumption 4.10. There exist countably many independent real-valued random variables $\boldsymbol{\xi} := (\xi_m)_{m \in \mathbb{N}}$ with $\xi_m \sim \nu_m \in \mathcal{P}(\mathbb{R})$ as well as mappings $f \in L^2_\nu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$ and $g \in L^2_\nu(\mathbb{R}^{\mathbb{N}}; \mathbb{R}^d)$, where $\nu := \bigotimes_{m \geq 1} \nu_m$, such that for the random variables U and ε in (4.2) there holds

$$U = f(\boldsymbol{\xi}), \quad \varepsilon = g(\boldsymbol{\xi}) \qquad \mathbb{P}\text{-almost surely.}$$

Furthermore, the normalized orthogonal polynomials $\{P_\alpha^{(m)}\}_{\alpha \in \mathbb{N}_0}$ in $L^2_{\nu_m}(\mathbb{R}; \mathbb{R})$ form a complete systems of $L^2_{\nu_m}(\mathbb{R}; \mathbb{R})$.

We recall that in the case of U and ε being Gaussian, e.g., U representing a Gaussian random field, Assumption 4.10 is satisfied. Moreover, we refer to Ernst et al. [54] for a discussion in which cases the second part of Assumption 4.10 is satisfied. Given 4.10 we can represent U and ε by PCEs

$$U = \sum_{\boldsymbol{\alpha} \in \mathcal{F}} u_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \qquad \varepsilon = \sum_{\boldsymbol{\alpha} \in \mathcal{F}} \varepsilon_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \tag{4.8}$$

where $u_{\boldsymbol{\alpha}} \in \mathcal{H}$ and $\varepsilon_{\boldsymbol{\alpha}} \in \mathbb{R}^d$ denote the chaos coefficients of U and ε, respectively and

$$\mathcal{F} = \{\boldsymbol{\alpha} \in \mathbb{N}_0^{\mathbb{N}} : \alpha_j \neq 0 \text{ for only finitely many } j\}, \quad P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) := \prod_{m \geq 1} P_{\alpha_m}^{(m)}(\xi_m) \ \forall \alpha \in \mathcal{F}.$$

Remark 4.11. In Assumption 4.10 we require U and ε already to be given, but an alternative assumption would be the existence of chaos coefficients $u_{\boldsymbol{\alpha}}$ and $\varepsilon_{\boldsymbol{\alpha}}$, $\alpha \in \mathcal{F}$, such that

$$\left(\sum_{\boldsymbol{\alpha} \in \mathcal{F}} u_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \sum_{\boldsymbol{\alpha} \in \mathcal{F}} \varepsilon_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \right) \sim \mu_0 \otimes \nu_\varepsilon.$$

Then, equation (4.8) would define the random variables U and ε appearing in (4.2).

Moreover, an expansion in polynomials $P_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is not crucial for constructing and applying the PCKF. In principle, any countable CONS $(\Psi_\alpha)_{\alpha\in\mathbb{N}}$ of the space $L^2_\nu(\mathbb{R}^{\mathbb{N}};\mathbb{R})$ such that $\left(\sum_\alpha u_\alpha\Psi_\alpha(\boldsymbol{\xi}), \sum_\alpha \varepsilon_\alpha\Psi_\alpha(\boldsymbol{\xi})\right) \sim \mu_0 \otimes \nu_\varepsilon$ would be suitable.

For numerical simulations and, particularly, for the PCKF we will have to truncate the PCE and, therefore, introduce

Definition 4.12. Let $X \in L^2(\Omega;\mathcal{H})$ be given as $X = f(\boldsymbol{\xi})$ with $\boldsymbol{\xi}$ as in Assumption 4.10 and $f\colon \mathbb{R}^{\mathbb{N}} \to \mathcal{H}$ measurable. Then, denoting the chaos coefficients of X by $x_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in \mathcal{F}$, we define for a subset $J \subseteq \mathcal{F}$

$$\mathrm{P}_J X := \sum_{\boldsymbol{\alpha}\in J} x_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}).$$

In the construction of the PCKF we will also have to employ the PCE of deterministic objects, such as the observed data $y \in \mathbb{R}^d$. It is clear, that only the chaos coefficient associated with the constant polynomial can be non-zero for such deterministic quantities - otherwise they would be random. In the following we will use the Kronecker symbol for multi-indices

$$\delta_{\boldsymbol{\alpha},\boldsymbol{\beta}} := \prod_{m=1}^{\infty} \delta_{\alpha_j,\beta_j} = \begin{cases} 1, & \boldsymbol{\alpha} = \boldsymbol{\beta} \\ 0, & \text{otherwise,} \end{cases} \tag{4.9}$$

to describe the vector of chaos coefficients $z_{\boldsymbol{\alpha}}$ of deterministic quantities $z \in \mathbb{R}^d$:

$$(z_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathcal{F}} = (\delta_{\boldsymbol{\alpha}\mathbf{0}} z)_{\boldsymbol{\alpha}\in\mathcal{F}} = (z, 0, 0, \ldots).$$

Definition 4.13 (Polynomial Chaos Kalman Filter)**.** Given the model (4.2) and Assumption 4.10, the *Polynomial Chaos Kalman filter* for estimating U given $Y = y$ is as follows:

1. **Initialization:** Choose a finite subset $J \subset \mathcal{F}$ and compute the chaos coefficients $u_{\boldsymbol{\alpha}}$ of $U \sim \mu_0$ for $\boldsymbol{\alpha} \in J$. Set $U_J := \mathrm{P}_J U$.

2. **Forecast:** Compute the chaos coefficients $g_{J,\boldsymbol{\alpha}}$ of $G(U_J)$ for $\boldsymbol{\alpha} \in J$ and set

$$y_{J,\boldsymbol{\alpha}} := g_{J,\boldsymbol{\alpha}} + \varepsilon_{\boldsymbol{\alpha}} \qquad \forall \boldsymbol{\alpha} \in J,$$

   where $\varepsilon_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in \mathcal{J}$, denote the chaos coefficients of ε. Set $Y_J := \mathrm{P}_J(G(U_J) + \varepsilon)$ and compute the covariances $\mathrm{Cov}(U_J, Y_J)\colon \mathbb{R}^d \to \mathcal{H}$ and $\mathrm{Cov}(Y_J)\colon \mathbb{R}^d \to \mathbb{R}^d$

where, e.g., $\mathrm{Cov}(U_J, Y_J)$ is determined by

$$\mathrm{Cov}(U_J, Y_J)x = \sum_{\boldsymbol{\alpha}\in J\setminus\{\mathbf{0}\}} \left(y_{J,\boldsymbol{\alpha}}^\top x\right)\, u_{\boldsymbol{\alpha}}, \qquad x \in \mathbb{R}^d.$$

3. **Analysis:** Compute $\mathrm{K}_J := \mathrm{Cov}(U_J, Y_J)\,\mathrm{Cov}(Y_J)^{-1} \in \mathcal{L}(\mathbb{R}^d, \mathcal{H})$ and

$$u_{J,\boldsymbol{\alpha}}^a := u_{\boldsymbol{\alpha}} + \mathrm{K}_J\,(\delta_{\boldsymbol{\alpha},0} y - y_{J,\boldsymbol{\alpha}}) \qquad \forall \boldsymbol{\alpha} \in J. \tag{4.10}$$

The coefficients $u_{\boldsymbol{\alpha}}^a$, $\boldsymbol{\alpha} \in J$, are called *analysis PC coefficients* and we define

$$U_J^a := \sum_{\boldsymbol{\alpha}\in J} u_{J,\boldsymbol{\alpha}}^a P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{4.11}$$

In Definition 4.13 we assume again the invertibility of the covariance $\mathrm{Cov}(Y_J)$ which is, for example, ensured if $\mathrm{Cov}(\varepsilon)$ is invertible and J sufficiently large such that $\mathrm{P}_J\,\varepsilon = \varepsilon$. Furthermore, we note that the random variable $G(U_J)$ is well-defined, since $U_J \in \mathcal{D}_G$ $\mathbb{P}$-almost surely for a finite set J:

$$U_J = \sum_{\boldsymbol{\alpha}\in J} u_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \qquad u_{\boldsymbol{\alpha}} = \mathbb{E}\left[P_{\boldsymbol{\alpha}}(\boldsymbol{\xi})U\right] \in \mathcal{D}_G,$$

where the latter holds due to $U \in \mathcal{D}_G$ $\mathbb{P}$-almost surely. However, we need to assume that $G(U_J) \in L^2(\Omega;\mathbb{R}^d)$ in the definition of the PCKF, in order to ensure the existence of the PC coefficients $g_{J,\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in \mathcal{F}$. Moreover, by linearity the coefficients $y_{J,\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in J$, are indeed the chaos coefficients of $Y_J = \mathrm{P}_J(G(U_J) + \varepsilon)$, but due to the nonlinearity of G there holds in general $\mathrm{P}_J\,G(U) \neq G(\mathrm{P}_J\,U) \neq \mathrm{P}_J\,G(U_J)$, and, hence, $Y_J \neq \mathrm{P}_J\,Y$. For numerical methods to compute the chaos coefficients $g_{J,\boldsymbol{\alpha}}$ of $G(U_J)$ we refer to Section 2.3.2. In the next section we will analyze to which random variable U_J^a converges if the finite set J tends to $\mathcal{F}$.

Remark 4.14 (On extension to dynamical systems)**.** Again, the PCKF can also be applied to dynamical systems. Then, given nonlinear dynamics $F_n\colon \mathcal{H} \to \mathcal{H}$, i.e., $U_{n+1} = F_n(U_n) + \eta_n$ we have to compute the chaos coefficients of $F_n(U_{J,n}^a) + \eta_n$ in order to obtain the initial PC vector for incorporating the observations y_{n+1} of $G(U_{n+1}) + \varepsilon_{n+1}$.

Remark 4.15 (On computational complexity)**.** Although a detailed complexity analysis of the EnKF and PCKF is beyond the scope of this work, we mention that the EnKF requires M evaluations of the forward map with M denoting the ensemble size, whereas the PCKF requires the computation of the chaos coefficients of $G(U)$

by, e.g., the stochastic Galerkin method, cf. Section 2.3.2. Thus the former yields, in general, many small systems to solve, whereas the latter typically requires the solution of a large coupled system. Moreover, we emphasize the computational savings by applying Kalman filters compared to a "full Bayesian update", i.e., sampling from the posterior measure by MCMC methods. In particular, each MCMC simulation may require hundreds of thousands evaluations of the forward map G, i.e., (at least) one for each iteration.

4.2. Convergence of Generalized Kalman Filters

The algorithmic descriptions of the EnKF and PCKF look quite similiar. Indeed, both algorithms perform discretized versions of an update for random variables, namely,

$$U^a := U + \mathrm{K}(y - Y), \qquad \mathrm{K} := \mathrm{Cov}(U, Y)\,\mathrm{Cov}(Y)^{-1}, \tag{4.12}$$

where Y is as in (4.2) and $(U, \varepsilon) \sim \mu_0 \otimes \nu_\varepsilon$. The difference between the EnKF and the PCKF is that the former provides samples and the latter chaos coefficients of U^a.

Definition 4.16 (Analysis variable)**.** The random variable U^a given by (4.12) is called *analysis variable* and the linear operator $\mathrm{K} \in \mathcal{L}(\mathbb{R}^d; \mathcal{H})$ in (4.12) is called *Kalman gain*.

The output of the EnKF and the PCKF is corrupted by the approximation of the Kalman gain operator K by the empirical covariances $\tilde{\mathrm{K}}_M$ and the operator K_J, respectively. In the following we will prove that both methods do indeed converge to U^a in some sense for increasing sample size $M \in \mathbb{N}$ or an increasing subset $J \subset \mathcal{F}$ of chaos coefficients.

4.2.1. Convergence of the PCKF

We start with the convergence of the PCKF for increasing subsets $J_n \subset \mathcal{F}$. In particular, we assume in the following a nested and exhaustive sequence of finite subsets $(J_n)_{n\in\mathbb{N}}$, i.e, $J_m \subset J_n$ for $m \leq n$ and $J_n \uparrow \mathcal{F}$. An example of such a sequence is

$$J_n := \left\{ \boldsymbol{\alpha} \in \mathcal{F} : \alpha_j = 0 \ \forall j > n, \sum_{j=1}^{\infty} |\alpha_j| \leq n \right\}.$$

Given such a sequence $(J_n)_{n\in\mathbb{N}}$ the error $\|U - U_{J_n}\|_{L^2(\Omega;\mathcal{H})}$, where $U_{J_n} = \mathrm{P}_{J_n} U$,

will tend to zero:

$$\|U - U_{J_n}\|^2_{L^2(\Omega;\mathcal{H})} = \sum_{\boldsymbol{\alpha}\in\mathcal{F}\setminus J_n} \|u_{\boldsymbol{\alpha}}\|^2_{\mathcal{H}} \xrightarrow{J_n\uparrow\mathcal{F}} 0.$$

The same will, in general, not hold for the error $\|G(U) - \mathrm{P}_{J_n} G(U)\|_{L^2(\Omega;\mathbb{R}^d)}$ — even if G is smooth — since the L^2-convergence is not preserved under continuous mappings (unlike convergence in the almost sure sense, in probability and in distribution). This implies that also $\mathrm{K}_J \to \mathrm{K}$ can not be ensured in general. However, under an additional assumption, we can prove the following result.

Theorem 4.17. Consider the model (4.2) and let Assumption 4.10 be satisfied. Let $(J_n)_{n\in\mathbb{N}}$ be a nested and exhaustive sequence of finite subsets of $\mathcal{F}$ with $\mathbf{0} \in J_1$ and let

- $G\colon \mathcal{D}_G \to \mathbb{R}^d$ be continuous,
- there exists constans $\delta > 0$ and $C < +\infty$ such that

$$\mathbb{E}\left[|G(U_{J_n})|^{2+\delta}\right] \leq C \qquad \forall n \in \mathbb{N}. \tag{4.13}$$

Then for $Y_{J_n} := \mathrm{P}_{J_n}(G(U_{J_n}) + \varepsilon)$ and

$$U^a_{J_n} = \sum_{\boldsymbol{\alpha}\in J_n} u^a_{J_n,\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}),$$

denoting the random variable generated by the PCKF in the analysis step for the subset $J = J_n$, there holds

$$\lim_{n\to\infty} \|Y - Y_{J_n}\|_{L^2(\Omega;\mathbb{R}^d)} = 0$$

and

$$\|U^a - U^a_{J_n}\|_{L^2(\Omega;\mathcal{H})} \in \mathcal{O}\left(\|U - U_{J_n}\|_{L^2(\Omega;\mathcal{H})} + \|Y - Y_{J_n}\|_{L^2(\Omega;\mathbb{R}^d)}\right), \tag{4.14}$$

which means, in particular, that $U^a_{J_n} \to U^a$ in $L^2(\Omega;\mathcal{H})$ as $n \to \infty$.

Proof. In the following we use $\|\cdot\|_{L^2}$ as shorthand for $\|\cdot\|_{L^2(\Omega;\mathcal{H})}$ and $\|\cdot\|_{L^2(\Omega;\mathbb{R}^d)}$, respectively. Since the sequence $(J_n)_{n\in\mathbb{N}}$ is exhaustive, we obtain $U_{J_n} \to U$ in $L^2(\Omega;\mathcal{H})$, and hence $U_{J_n} \xrightarrow{\mathbb{P}} U$, where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability. Since G is continuous, it follows by the continuous mapping theorem, see Kallenberg [96, Lemma 4.3] that also $G(U_{J_n}) \xrightarrow{\mathbb{P}} G(U)$. Now the boundedness assumption

(4.13) implies the uniform integrability of the random variables $|G(U_{J_n})|^2$, $n \in \mathbb{N}$, see again Kallenberg [96, p. 67], and by [96, Proposition 4.12] we then obtain $G(U_{J_n}) \to G(U)$ in $L^2(\Omega; \mathcal{H})$. Thus,

$$\|Y - Y_{J_n}\|_{L^2} \leq \underbrace{\|Y - \mathrm{P}_{J_n} Y\|_{L^2}}_{\to 0} + \underbrace{\|\mathrm{P}_{J_n}(Y - G(U_{J_n}) - \varepsilon)\|_{L^2}}_{\leq \|G(U) - G(U_{J_n})\|_{L^2} \to 0} \to 0.$$

Next, consider J_n as fixed. Since $U^a = U + \mathrm{K}(y - Y)$ and $U^a_{J_n} = U_{J_n} + \mathrm{K}_{J_n}(y - Y_{J_n})$, we have

$$\|U^a - U^a_{J_n}\|_{L^2} \leq \|U - U_{J_n}\|_{L^2} + \|\mathrm{K} - \mathrm{K}_{J_n}\| \, \|y - Y_{J_n}\|_{L^2} + \|\mathrm{K}\| \, \|Y - Y_{J_n}\|_{L^2},$$

where the norm for K and $\mathrm{K} - \mathrm{K}_{J_n}$ is the usual operator norm for linear mappings from $\mathbb{R}^d \to \mathcal{H}$. Due to assumption (4.13) there exists a finite constant which we will also denote by C such that

$$\|Y_{J_n}\|_{L^2} = \|\mathrm{P}_{J_n}(G(U_{J_n}) + \varepsilon)\|_{L^2} \leq \|G(U_{J_n}) + \varepsilon\|_{L^2} \leq \|G(U_{J_n})\|_{L^2} + \|\varepsilon\|_{L^2} \leq C.$$

Hence, we can estimate $\|y - Y_{J_n}\|_{L^2} \leq |y| + C < \infty$. Considering $\|\mathrm{K} - \mathrm{K}_{J_n}\|$, we can further split this error into

$$\begin{aligned} \|\mathrm{K} - \mathrm{K}_{J_n}\| \leq{}& \|\operatorname{Cov}(U, Y) - \operatorname{Cov}(U_{J_n}, Y_{J_n})\| \, \|\operatorname{Cov}^{-1}(Y)\| \\ &+ \|\operatorname{Cov}(U_{J_n}, Y_{J_n})\| \, \|\operatorname{Cov}^{-1}(Y) - \operatorname{Cov}^{-1}(Y_{J_n})\|. \end{aligned}$$

Next, we recall that the covariance $\operatorname{Cov}(X, Z)$ of two random variables depends continuously on X and Z. In particular, for zero-mean Hilbert space-valued random variable $X_1, X_2 \in L^2(\Omega; \mathcal{H})$ and $Z_1, Z_2 \in L^2(\Omega; \mathbb{R}^d)$ we obtain

$$\begin{aligned} \|\operatorname{Cov}(X_1, Z_1) - \operatorname{Cov}(X_2, Z_2)\| &= \|\mathbb{E}\left[X_1 \otimes Z_1\right] - \mathbb{E}[X_2 \otimes Z_2]\,\| \\ &\leq \mathbb{E}\left[\|(X_1 - X_2) \otimes Z_1\| + \|X_2 \otimes (Z_1 - Z_2)\|\right] \\ &= \mathbb{E}\left[\|X_1 - X_2\|_{\mathcal{H}} \, |Z_1|\right] + \mathbb{E}[\|X_2\|_{\mathcal{H}} \, |Z_1 - Z_2|] \\ &\leq (\|Z_1\|_{L^2} + \|X_2\|_{L^2}) \, (\|X_1 - X_2\|_{L^2} + \|Z_1 - Z_2\|_{L^2}), \end{aligned}$$

where we have used Jensen's and the triangle inequality in the second line and the Cauchy-Schwarz inequality in the last line. Since $\operatorname{Cov}(X, Z) = \operatorname{Cov}(X - \mathbb{E}[X], Z - \mathbb{E}[Z])$ and $\|X - \mathbb{E}[X]\|_{L^2} \leq \|X\|_{L^2}$ the above estimate holds also for non-zero-mean random variables. Thus, we get

$$\|\operatorname{Cov}(U, Y) - \operatorname{Cov}(U_{J_n}, Y_{J_n})\| \leq (\|U\|_{L^2} + \|Y\|_{L^2}) \, (\|U - U_{J_n}\|_{L^2} + \|Y - Y_{J_n}\|_{L^2}),$$

since $\|U_{J_n}\|_{L^2} \leq \|U\|_{L^2}$, and

$$\|\operatorname{Cov}(Y) - \operatorname{Cov}(Y_{J_n})\| \leq 4C \, \|Y - Y_{J_n}\|_{L^2},$$

due to $\|Y_{J_n}\|_{L^2} \leq C$. Now, we exploit that the sequence $(J_n)_{n\in\mathbb{N}}$ is nested and exhaustive and recall that, by taking a sufficiently large n, the error $\|U - U_{J_n}\|_{L^2}$ and $\|Y - Y_{J_n}\|_{L^2}$ can be made arbitrarily small. Thus, also $\|\operatorname{Cov}(Y) - \operatorname{Cov}(Y_{J_n})\|$ will tend to zero as $n \to \infty$. We then apply the continuity of the matrix inverse to estimate $\|\operatorname{Cov}^{-1}(Y) - \operatorname{Cov}^{-1}(Y_{J_n})\|$. Recall that $\operatorname{Cov}(Y), \operatorname{Cov}(Y_{J_n}) \in \mathbb{R}^{d\times d}$. Let n be sufficiently large such that

$$\|\operatorname{Cov}(Y) - \operatorname{Cov}(Y_{J_n})\| < \frac{1}{2\|\operatorname{Cov}^{-1}(Y)\|},$$

then there holds, see Horn and Johnson [90, Section 5.8],

$$\|\operatorname{Cov}^{-1}(Y) - \operatorname{Cov}^{-1}(Y_{J_n})\| \leq 2\|\operatorname{Cov}^{-1}(Y)\|^2 \|\operatorname{Cov}(Y) - \operatorname{Cov}(Y_{J_n})\|.$$

Summing up all previous estimates, we obtain

$$\|\mathrm{K} - \mathrm{K}_{J_n}\| \;\leq\; C_1(\|U - U_{J_n}\|_{L^2} + \|Y - Y_{J_n}\|_{L^2}) + C_2\|Y - Y_{J_n}\|_{L^2},$$

with $C_1 = \|\operatorname{Cov}^{-1}(Y)\|(\|U\|_{L^2} + \|Y\|_{L^2})$ and $C_2 = 8C^2\|U\|_{L^2} \, \|\operatorname{Cov}^{-1}(Y)\|^2$ where we have used

$$\|\operatorname{Cov}(U_{J_n}, Y_{J_n})\| \leq \|U_{J_n}\|_{L^2} \, \|Y_{J_n}\|_{L^2} \leq C\|U\|_{L^2}$$

to obtain C_2. Finally, we arrive at

$$\begin{aligned} \|U^a - U^a_{J_n}\|_{L^2} &\leq \|U - U_{J_n}\|_{L^2} + (|y| + C)\|\mathrm{K} - \mathrm{K}_{J_n}\| + \|\mathrm{K}\| \, \|Y - Y_{J_n}\|_{L^2} \\ &\leq C_3(\|U - U_{J_n}\|_{L^2} + \|Y - Y_{J_n}\|_{L^2}), \end{aligned}$$

with $C_3 = 1 + \|\mathrm{K}\| + |y| + C + C_1 + C_2$, and the assertion follows. □

Remark 4.18. Since for many applications evaluating the forward map G corresponds to solving a differential or integral equation, an additional error arises due to numerical approximations G_h of G. This error affects the filters by sampling or computing chaos coefficients of $Y_h = G_h(U) + \varepsilon$ instead of Y. We neglect this error in our analysis since it is beyond the scope of this work. However, if G is the solution operator for differential equations, we expect that (4.13) could be verified in many cases, such as for elliptic boundary value problems with U a random diffusion coefficient or source term, cf. Section 2.3.

4.2.2. Convergence of the EnKF

A first large ensemble convergence analysis for the EnKF was carried out by Le Gland et al. [75]. The authors considered filtering for finite-dimensional locally Lipschitz dynamical systems. They proved that the empirical mean of $f(u_j^a)$, $j = 1, \dots, M$, where $f\colon \mathbb{R}^n \to \mathbb{R}$ denotes a locally Lipschitz continuous function and u_j^a the members of the analysis ensemble, converges $\mathbb{P}$-almost surely to the expectation of f w.r.t to a certain limit distribution. The randomness which is meant here by the term "$\mathbb{P}$-a.s." refers to the randomness in the EnKF algorithm, i.e., the sampling involved in generating the analysis ensemble $\boldsymbol{u}^a$. In our case this limit distribution coincides with the distribution of the analysis variable U^a. Furthermore, the authors of [75] proved an L^p-type convergence for the empirical analysis measures associated with the analysis ensembles. Mandel et al. [116] proved the L^p-convergence of the ensemble members to random variables distributed according to the posterior measure in case of linear dynamics. More recently, Law et al. [106] extended the L^p-convergence result of Le Gland et al. [75] to more general forms of dynamical systems and stated them in terms of a suitably chosen metric for random measures. However, large ensemble limits are rather of academic interest, since in practice the number of ensemble members is usually at most a few hundred. We mention the work of Schillings and Stuart [155] where the behaviour of the EnKF with a finite number of samples is analyzed in case of linear dynamical systems.

In the following, we derive a result similar to Le Gland et al. [75] for the considered Bayesian inference problem setting, i.e., we extend their result in case of one update to Hilbert spaces and data obtained by general nonlinear continuous observation operators. Although the proof employs similar ideas as the one of [75, Theorem 5.1] we present it for completion.

Theorem 4.19. We consider the model (4.2) and assume that $G\colon \mathcal{D}_G \to \mathbb{R}^d$ is continuous and Assumption 4.1 is satisfied. For a fixed $M \in \mathbb{N}$ let $U_{j,M}^a$, $j = 1, \dots, M$, denote the M random variables whose realizations form the analysis ensemble generated by the EnKF algorithm in Definition 4.7. Furthermore, let μ^a denote the distribution on $\mathcal{H}$ of the analysis variable U^a given in Definition 4.16. Then for each $j \in \mathbb{N}$ the random variable $U_{j,M}^a$ converges in distribution to U^a for $M \to \infty$. Moreover, we have

$$\lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} f(U_{j,M}^a) = \int_{\mathcal{H}} f(u)\, \mu^a(\mathrm{d}u) \qquad \mathbb{P}\text{-a.s.}$$

for any function $f\colon \mathcal{H} \to \mathcal{Y}$, where $\mathcal{Y}$ denotes an arbitrary separable Hilbert space,

which satisfies

$$\|f(u) - f(v)\|_{\mathcal{Y}} \leq C(1 + \|u\|_{\mathcal{H}} + \|v\|_{\mathcal{H}}) \, \|u - v\|_{\mathcal{H}} \qquad \forall u, v \in \mathcal{H},$$

This implies, in particular, for $\bar{U}_M^a := \frac{1}{M} \sum_{j=1}^M U_{j,M}^a$ and $\boldsymbol{U}_M^a := (U_{1,M}^a, \dots, U_{M,M}^a)$

$$\lim_{M\to\infty} \bar{U}_M^a = \mathbb{E}[U^a] \quad \text{and} \quad \lim_{M\to\infty} \mathrm{Cov}(\boldsymbol{U}_M^a) = \mathrm{Cov}(U^a) \qquad \mathbb{P}\text{-almost surely.}$$

Proof. Let us denote by U_j and ε_j, $j \in \mathbb{N}$, i.i.d. random variables such that $(U_j, \varepsilon_j) \sim \mu_0 \otimes \nu_\varepsilon$. We set $Y_j := G(U_j) + \varepsilon_j$ and assume w.l.o.g. that for $j \leq M$

$$U_{j,M}^a = U_j + \tilde{\mathrm{K}}_M(y - Y_j), \qquad \tilde{\mathrm{K}}_M = \mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M) \, \mathrm{Cov}^{-1}(\boldsymbol{Y}_M),$$

where $\mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M)$ and $\mathrm{Cov}(\boldsymbol{Y}_M)$ are empirical covariances, e.g.,

$$\mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M) = \frac{1}{M-1} \sum_{j=1}^M (U_j - \bar{U}_M) \otimes (Y_j - \bar{\boldsymbol{Y}}_M)$$

with $\bar{U}_M = \frac{1}{M}(U_1 + \cdots + U_M)$ and $\bar{\boldsymbol{Y}}_M = \frac{1}{M}(Y_1 + \cdots + Y_M)$. Further, we define

$$U_j^a := U_j + \mathrm{K}(y - Y_j), \qquad \mathrm{K} = \mathrm{Cov}(U, Y) \, \mathrm{Cov}^{-1}(Y), \qquad j \in \mathbb{N},$$

i.e., the random variables U_j^a are i.i.d. copies of the analysis variable U^a. We estimate for each $j = 1, \dots, M$

$$\|U_{j,M}^a - U_j^a\|_{\mathcal{H}} \leq \|\mathrm{K} - \tilde{\mathrm{K}}_M\| \, |y - Y_j| \qquad \mathbb{P}\text{-a.s.},$$

where we can further split

$$\begin{aligned} \mathrm{K} - \tilde{\mathrm{K}}_M = {} & \big(\mathrm{Cov}(U, Y) - \mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M)\big) \, \mathrm{Cov}^{-1}(Y) \\ & + \mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M) \, \Big(\mathrm{Cov}^{-1}(Y) - \mathrm{Cov}^{-1}(\boldsymbol{Y}_M)\Big). \end{aligned}$$

Then, we recall that the empirical covariance converges $\mathbb{P}$-almost surely to the true covariance which follows easily by writing

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M) = {} & \frac{1}{M-1} \sum_{j=1}^M (U_j - \mathbb{E}[U]) \otimes (Y_j - \mathbb{E}[Y]) \\ & - \frac{M}{M-1} (\bar{U}_M - \mathbb{E}[U]) \otimes (\bar{\boldsymbol{Y}}_M - \mathbb{E}[Y]) \end{aligned}$$

and applying the strong law of large numbers (SLLN) for i.i.d. Hilbert space-valued

random variables, see Padgett and Taylor [128], which yields $\mathbb{P}$-almost surely

$$\frac{1}{M-1}\sum_{j=1}^{M}(U_j - \mathbb{E}[U]) \otimes (Y_j - \mathbb{E}[Y]) \xrightarrow{M\to\infty} \mathbb{E}[(U - \mathbb{E}[U]) \otimes (Y - \mathbb{E}[Y])$$

as well as

$$\frac{M}{M-1}(\bar{U}_M - \mathbb{E}[U]) \otimes (\bar{\boldsymbol{Y}}_M - \mathbb{E}[Y]) \xrightarrow{M\to\infty} 0.$$

Thus, we have $\mathbb{P}$-almost surely

$$\mathrm{Cov}(U, Y) - \mathrm{Cov}(\boldsymbol{U}_M, \boldsymbol{Y}_M) \xrightarrow{M\to\infty} 0, \qquad \mathrm{Cov}(Y) - \mathrm{Cov}(\boldsymbol{Y}_M) \xrightarrow{M\to\infty} 0.$$

Now, since the matrix inverse is a continuous mapping, there follows also $\mathbb{P}$-a.s.

$$\mathrm{Cov}^{-1}(Y) - \mathrm{Cov}^{-1}(\boldsymbol{Y}_M) \xrightarrow{M\to\infty} 0$$

and, hence, $\mathbb{P}$-almost surely $\mathrm{K}_M \xrightarrow{M\to\infty} \mathrm{K}$. Thus, for any $j \in \mathbb{N}$ we get that $\mathbb{P}$-a.s.

$$\lim_{M\to\infty} \|U_{j,M}^a - U_j^a\|_{\mathcal{H}} \le |y - Y_j| \lim_{M\to\infty} \|\mathrm{K} - \tilde{\mathrm{K}}_M\| = 0,$$

i.e., as $M \to \infty$ we have $\mathbb{P}$-a.s. $U_{j,M}^a \to U_j^a$, hence, $U_{j,M}^a \to U_j^a$ also in distribution.

Next, for any $f\colon \mathcal{H} \to \mathcal{Y}$ satisfying the assumptions stated in the theorem, we have

$$\frac{1}{M}\sum_{j=1}^{M} f(U_{j,M}^a) \;=\; \frac{1}{M}\sum_{j=1}^{M}\Big(f(U_{j,M}^a) - f(U_j^a)\Big) + \frac{1}{M}\sum_{j=1}^{M} f(U_j^a).$$

Due to the SLLN there holds

$$\lim_{M\to\infty} \frac{1}{M}\sum_{j=1}^{M} f(U_j^a) = \mathbb{E}\left[f(U_1^a)\right] = \mathbb{E}\left[f(U^a)\right] \qquad \mathbb{P}\text{-a.s.}$$

Hence, we need only ensure that

$$\begin{aligned}\left\|\frac{1}{M}\sum_{j=1}^{M}\Big(f(U_{j,M}^a) - f(U_j^a)\Big)\right\|_{\mathcal{Y}} &\le \frac{1}{M}\sum_{j=1}^{M} C(1 + \|U_j^a\|_{\mathcal{H}} + \|U_{j,M}^a\|_{\mathcal{H}})\, \|U_{j,M}^a - U_j^a\|_{\mathcal{H}} \\ &\le \left(\frac{C}{M}\sum_{j=1}^{M}(1 + \|U_j^a\|_{\mathcal{H}} + \|U_{j,M}^a\|_{\mathcal{H}})^2\right)^{1/2} \cdot \\ &\qquad \left(\frac{C}{M}\sum_{j=1}^{M}\|U_{j,M}^a - U_j^a\|_{\mathcal{H}}^2\right)^{1/2}\end{aligned}$$

converges $\mathbb{P}$-a.s. to 0 as $M \to \infty$ to prove the second statement. Since by the SLLN we have

$$\lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} |y - Y_j|^p = \mathbb{E}[|y - Y|^p] \qquad \mathbb{P}\text{-a.s.},$$

we get by the above reasoning that $\mathbb{P}$-a.s.

$$\lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \|U_{j,M}^a - U_j^a\|_{\mathcal{H}}^p \leq \lim_{M\to\infty} \|\mathrm{K} - \tilde{\mathrm{K}}_M\|^p \, \frac{1}{M} \sum_{j=1}^{M} |y - Y_j|^p = 0 \cdot \mathbb{E}[|y - Y|^p] = 0.$$

Moreover, there holds

$$\begin{aligned}(1 + \|U_j^a\|_{\mathcal{H}} + \|U_{j,M}^a\|_{\mathcal{H}})^2 &\leq (1 + 2\|U_j^a\|_{\mathcal{H}} + \|U_{j,M}^a - U_j^a\|_{\mathcal{H}})^2 \\ &\leq 2(1 + 2\|U_j^a\|_{\mathcal{H}})^2 + 2\|U_{j,M}^a - U_j^a\|_{\mathcal{H}}^2\end{aligned}$$

and, hence,

$$\begin{aligned}\frac{1}{M} \sum_{j=1}^{M} (1 + \|U_j^a\|_{\mathcal{H}} + \|U_{j,M}^a\|_{\mathcal{H}})^2 &\leq \frac{2}{M} \sum_{j=1}^{M} (1 + 2\|U_j^a\|_{\mathcal{H}})^2 + \frac{2}{M} \sum_{j=1}^{M} \|U_{j,M}^a - U_j^a\|_{\mathcal{H}}^2 \\ &\xrightarrow{M\to\infty} 2\mathbb{E}[(1 + 2\|U^a\|_{\mathcal{H}})^2]\end{aligned}$$

$\mathbb{P}$-almost surely. Thus, we finally obtain

$$\left\| \frac{1}{M} \sum_{j=1}^{M} \Big(f(U_{j,M}^a) - f(U_j^a) \Big) \right\|_{\mathcal{Y}} \xrightarrow{M\to\infty} 0 \qquad \mathbb{P}\text{-a.s.},$$

proving the second statement of the theorem. The remaining statements follow immediately. □

It is possible to extend the statement of Theorem 4.19, similiar to Le Gland et al. [75, Theorem 5.1], to functions $f\colon \mathcal{H} \to \mathcal{Y}$ with polynomially growing local Lipschitz constant

$$\|f(u) - f(v)\|_{\mathcal{Y}} \leq C(1 + \|u\|_{\mathcal{H}}^p + \|v\|_{\mathcal{H}}^p) \, \|u - v\|_{\mathcal{H}}, \qquad u, v \in \mathcal{H},$$

where $p \geq 1$, if for the prior holds $\mu_0 \in \mathcal{P}^p(\mathcal{H})$ and $\int_{\mathcal{H}} |G(u)|^p \, \mu_0(\mathrm{d}u) < \infty$.

Moreover, the almost sure convergence of the random members $U_{j,M}^a$ of the analysis ensemble to the i.i.d. copies U_j^a of the analysis variable might suggest that $\frac{1}{M} \sum_{j=1}^{M} f(U_{j,M}^a)$ converges $\mathbb{P}$-almost surely to $\mathbb{E}\left[f(U^a)\right]$ for any continuous function f. However, as we have seen in the proof of Theorem 4.19, we need to show the $\mathbb{P}$-almost sure convergence of $\|\frac{1}{M} \sum_{j=1}^{M} f(U_{j,M}^a) - f(U_j^a)\|_{\mathcal{H}}$ to zero. The result

which we will establish in the next paragraph implies the $\mathbb{P}$-a.s. convergence of $\frac{1}{M}\sum_{j=1}^{M} f(U_{j,M}^a)$ to $\mathbb{E}\left[f(U^a)\right]$ for each bounded, continuous function f.

Almost sure convergence of empirical analysis measures. We recall the empirical analysis measure $\tilde{\mu}_M^a$ associated with the (random) analysis ensemble $\boldsymbol{U}_M^a := (U_{1,M}^a, \dots, U_{M,M}^a)$ as stated in Definition 4.8. Then Theorem 4.19 implies that for any Lipschitz continuous function $f\colon \mathcal{H} \to \mathcal{Y}$ we have

$$\lim_{M\to\infty} \int_{\mathcal{H}} f(u)\, \tilde{\mu}_M^a(\mathrm{d}u) = \int_{\mathcal{H}} f(u)\, \mu^a(\mathrm{d}u) \qquad \mathbb{P}\text{-a.s.},$$

where μ^a denotes the distribution of the analysis variable U^a on $\mathcal{H}$. Moreover, we know due to Klenke [99, Theorem 13.16] that weak convergence of probability measures $\mu_n \xrightarrow{w} \mu$ is equivalent to

$$\mu_n(f) \to \mu(f) \qquad \forall f \in \mathrm{Lip}_b(\mathcal{H};\mathbb{R}),$$

where for each $\nu \in \mathcal{P}(\mathcal{H})$ we set $\nu(f) := \int_{\mathcal{H}} f(u)\, \nu(\mathrm{d}u)$ and where $\mathrm{Lip}_b(\mathcal{H};\mathbb{R})$ denotes the space of all bounded, Lipschitz continuous functions $f\colon \mathcal{H} \to \mathbb{R}$. Thus, one might assume that Theorem 4.19 immediately implies

$$\tilde{\mu}_M^a \xrightarrow{w} \mu^a \qquad \mathbb{P}\text{-almost surely.} \tag{4.15}$$

However, this is a nontrivial implication, because the $\mathbb{P}$-a.s. weak convergence of $\tilde{\mu}_M^a$ to μ^a means that

$$\mathbb{P}\big(\tilde{\mu}_M^a(f) \to \mu^a(f) \quad \forall f \in f \in \mathrm{Lip}_b(\mathcal{H};\mathbb{R})\big) = 1 \tag{4.16}$$

whereas Theorem 4.19 only states that

$$\mathbb{P}\big(\tilde{\mu}_M^a(f) \to \mu^a(f)\big) = 1 \qquad \forall f \in f \in \mathrm{Lip}_b(\mathcal{H};\mathbb{R}). \tag{4.17}$$

Since the space $\mathrm{Lip}_b(\mathcal{H};\mathbb{R}) \subset C(\mathcal{H};\mathbb{R})$ is not separable w.r.t. the supremum norm $\|\cdot\|_{C(\mathcal{H};\mathbb{R})}$ if $\mathcal{H}$ is not compact, the condition (4.17) does not per se imply (4.16). However, Berti et al. [11] were able to prove the following result.

Theorem 4.20 ([11, Theorem 2.2])**.** Let $\mathcal{E}$ be a Polish space with Borel σ-algebra $\mathcal{B}(\mathcal{E})$ and let $(\Omega, \mathcal{A}, \mathbb{P})$ denote a probability space. Further, let $\mu\colon \Omega \to \mathcal{P}(\mathcal{E})$ and $\mu_n\colon \Omega \to \mathcal{P}(\mathcal{E})$, $n \in \mathbb{N}$, be random probability measures on $\mathcal{E}$, i.e., such that $\omega \mapsto \mu(\omega)(A)$ and $\omega \mapsto \mu_n(\omega)(A)$, $n \in \mathbb{N}$, are measurable for each $A \in \mathcal{B}(\mathcal{E})$. Then the following conditions are equivalent:

1. $\mu_n \xrightarrow{w} \mu$ converges $\mathbb{P}$-almost surely,

2. $\mu_n(f) \to \mu(f)$ converges $\mathbb{P}$-almost surely for each $f \in C_b(\mathcal{E}; \mathbb{R})$.

By virtue of this theorem we obtain the following corollary.

Corollary 4.21 (Almost sure weak convergence of the empirical analysis measures)**.** Under the assumptions and with the notations of Theorem 4.21 the statement (4.15) holds, i.e., the empirical analysis measures $\tilde{\mu}_M^a$ associated with the analysis ensemble of size M generated by the EnKF algorithm converge $\mathbb{P}$-a.s. weakly to the distribtution μ^a of the analysis variable U^a.

Proof. Let $(\Omega, \mathcal{A}, \mathbb{P})$ denote the underlying probability space. By Theorem 4.21 we know that there exists a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ such that for each $\omega \in A$ there holds

$$\tilde{\mu}_M^a(\omega)(f) \to \mu^a(f) \quad \forall f \in f \in \mathrm{Lip}_b(\mathcal{H}; \mathbb{R}),$$

where $\tilde{\mu}_M^a(\omega)$ denotes now the realization of $\tilde{\mu}_M^a$ for ω, i.e., a (deterministic) probability measure on $\mathcal{H}$. By Klenke [99, Theorem 13.16] the latter is equivalent to $\tilde{\mu}_M^a(\omega) \xrightarrow{w} \mu^a$ or

$$\tilde{\mu}_M^a(\omega)(f) \to \mu^a(f) \quad \forall f \in f \in C_b(\mathcal{H}; \mathbb{R}).$$

Thus, we have that $\tilde{\mu}_M^a(f) \to \mu^a(f)$ converges $\mathbb{P}$-almost surely for each $f \in C_b(\mathcal{H}; \mathbb{R})$ which with Theorem 4.20 yields the assertion. □

Remark 4.22 (On rates of convergence)**.** The results in Theorem 4.19 and Corollary 4.21 do not state any rate of convergence which is usually hard to obtain for $\mathbb{P}$-almost sure convergence. On the other hand, when considering L^p-convergence as done by Le Gland et al. [75, Theorem 5.2], Mandel et al. [116, Corollary 1] and Law et al. [106, Theorem 5.2] rates of convergence can be derived. Of course, the rate depends on the norm involved, but in case of the above results, the authors obtained the usual Monte Carlo rate of $M^{-1/2}$. At this point we would like to highlight that the PCKF can yield faster rates of convergence, since it employs spectral approximations, cf. Theorem 4.17. However, since our goal is to understand and analyze the common principle behind both generalized Kalman filters rather than comparing their convergence rates (w.r.t. different types of convergence) or computational cost we do not deepen this discussion and leave it for futher research.

4.3. Bayesian Interpretation of Generalized Kalman Filters

In the previous section we have characterized the limit of the EnKF and PCKF approximations for increasing sample size or polynomial degree, respectively. We now investigate how this limit, the analysis variable U^a, may be understood in the context of Bayesian inference. By analyzing the properties of this random variable we are able to characterize the approximations provided by the two Kalman filtering methods. In particular, we show that the EnKF and the PCKF do, in general, not provide sensible approximations to the posterior distribution. They are rather related to a linear approximation of the conditional mean estimator u_{CM}, see Definition 3.29, and the associated estimation error.

4.3.1. The Linear Conditional Mean

We recall from Section 3.4 that the conditional mean estimator $u_{\mathrm{CM}}\colon \mathbb{R}^d \to \mathcal{H}$ for a $\mathcal{H}$-valued random variable U given realizations of an random vector Y is characterized by

$$u_{\mathrm{CM}} := \operatorname*{argmin}_{\phi\colon \mathbb{R}^d \to \mathcal{H} \text{ measurable}} \mathbb{E}\left[\|U - \phi(Y)\|_{\mathcal{H}}^2\right].$$

In general, the computation of the corresponding Bayes estimate $u_{\mathrm{CM}}(y)$ can be costly. By restricting to linear maps $\phi\colon \mathbb{R}^d \to \mathcal{H}$ one obtains a new estimator which is explicitly computable in the Hilbert space setting.

Definition 4.23 (Linear conditional mean estimator)**.** The *linear conditional mean estimator* or *linear posterior mean estimator* for a random variable U on $\mathcal{H}$ given a random variable Y on another separable Hilbert space $\mathcal{Y}$ is defined as

$$u_{\mathrm{LCM}} := \operatorname*{argmin}_{\phi \in \mathcal{P}_1(\mathcal{Y};\mathcal{H})} \mathbb{E}\left[\|U - \phi(Y)\|_{\mathcal{H}}^2\right], \tag{4.18}$$

where $\mathcal{P}_1(\mathcal{Y};\mathcal{H}) = \{\phi : \phi(z) = b + Az \text{ with } b \in \mathcal{H}, A \in \mathcal{L}(\mathcal{Y},\mathcal{H})\}$ denotes the set of all linear mappings from $\mathcal{Y}$ to $\mathcal{H}$. The random variable $u_{\mathrm{LCM}}(Y)$ is called the *linear conditional mean*.

Again, we assume a unique minimizer of (4.18). The linear posterior mean estimator is the Bayesian equivalent of the best linear unbiased estimator (BLUE) known in Frequentist statistics. Furthermore, we recall that the conditional mean $u_{\mathrm{CM}}(Y) = \mathbb{E}[U \mid Y]$ is the best approximation of U in $L^2(\Omega, \sigma(Y), \mathbb{P}; \mathcal{H})$ w.r.t. the $L^2(\Omega;\mathcal{H})$-norm. Thus, the linear conditional mean $u_{\mathrm{LCM}}(Y)$ can be seen as the

best approximation of U in the subspace $\mathcal{P}_1(Y;\mathcal{H}) \subset L^2(\Omega,\sigma(Y),\mathbb{P};\mathcal{H})$, where $\mathcal{P}_1(Y;\mathcal{H}) := \{\phi(Y)\colon \phi \in \mathcal{P}_1(\mathcal{Y};\mathcal{H})\}$.

Lemma 4.24. The linear conditional mean as defined in (4.18) is given by

$$u_{\mathrm{LCM}}(y) = \mathbb{E}[U] + \mathrm{Cov}(U,Y)\,\mathrm{Cov}(Y)^{-1}(y - \mathbb{E}[Y]), \qquad y \in \mathcal{Y}.$$

Proof. The assertion follows by verifying that

$$u_{\mathrm{LCM}}(Y) = \mathbb{E}[U] + \mathrm{K}(Y - \mathbb{E}[Y]), \qquad \mathrm{K} = \mathrm{Cov}(U,Y)\,\mathrm{Cov}(Y)^{-1},$$

coincides with the orthogonal projection of U to $\mathcal{P}_1(Y;\mathcal{H})$. To do so, we will show that $U - u_{\mathrm{LCM}}(Y)$ is orthogonal to $\mathcal{P}_1(Y;\mathcal{H})$ w.r.t. the inner product in $L^2(\Omega;\mathcal{H})$. Let $b \in \mathcal{H}$ and $A \in \mathcal{L}(\mathcal{Y},\mathcal{H})$ be arbitrary. Then there holds

$$\begin{aligned}
\mathbb{E}[\langle U - u_{\mathrm{LCM}}(Y), b + AY\rangle_{\mathcal{H}}] &= \underbrace{\mathbb{E}[\langle U - \mathbb{E}[U], b\rangle_{\mathcal{H}}]}_{=0} - \underbrace{\mathbb{E}[\langle \mathrm{K}(Y - \mathbb{E}[Y]), b\rangle_{\mathcal{H}}]}_{=0} \\
&\quad + \mathbb{E}[\langle U - \mathbb{E}[U], AY\rangle_{\mathcal{H}}] - \mathbb{E}[\langle \mathrm{K}(Y - \mathbb{E}[Y]), AY\rangle_{\mathcal{H}}] \\
&= \mathbb{E}[\langle U - \mathbb{E}[U], A(Y - \mathbb{E}[Y])\rangle_{\mathcal{H}}] \\
&\quad - \mathbb{E}[\langle \mathrm{K}(Y - \mathbb{E}[Y]), A(Y - \mathbb{E}[Y]\rangle_{\mathcal{H}}] \\
&= \mathrm{Cov}(U,Y)A^* - \mathrm{K}\,\mathrm{Cov}(Y)A^* = 0,
\end{aligned}$$

since

$$\mathbb{E}[\langle U - \mathbb{E}[U], A\mathbb{E}[Y]\rangle_{\mathcal{H}}] = \mathbb{E}[\langle \mathrm{K}(Y - \mathbb{E}[Y]), A\mathbb{E}[Y]\rangle_{\mathcal{H}}] = 0$$

and $\mathrm{Cov}(AX,BY) = A\,\mathrm{Cov}(X,Y)B^*$ for Hilbert space valued random variable X, Y and bounded, linear operators A, B. □

This result is not entirely new. For example in finite dimensions similiar results were already stated by Luenberger [115, Section 4.5]. We mention though that Lemma 4.24 fails to hold in Banach spaces $\mathcal{X}$, since then the expectation $\mathbb{E}[U]$ and covariance $\mathrm{Cov}(U,Y)$ no longer minimize $\mathbb{E}[\|U-b\|_{\mathcal{X}}^2]$, $b \in \mathcal{X}$, and $\mathbb{E}[\|U-AY\|_{\mathcal{X}}^2]$, $A \in \mathcal{L}(\mathcal{Y},\mathcal{X})$, respectively; see also Remark 3.28.

4.3.2. Bayesian Interpretation of the Analysis Variable

Lemma 4.24 immediately yields a characterization of the analysis variable U^a defined in (4.12).

Theorem 4.25. Let Assumption 4.1 be satisfied for the model (4.2). Then for any $y \in \mathbb{R}^d$ the analysis variable $U^a = U + \mathrm{K}(y - Y)$, $\mathrm{K} = \mathrm{Cov}(U,Y)\,\mathrm{Cov}(Y)^{-1}$, coincides

with

$$U^a = u_{\mathrm{LCM}}(y) + (U - u_{\mathrm{LCM}}(Y)).$$

In particular, there holds

$$\mathbb{E}\,[U^a] = u_{\mathrm{LCM}}(y) \quad \text{and} \quad \mathrm{Cov}(U^a) = \mathrm{Cov}(U) - \mathrm{K}\,\mathrm{Cov}(Y, U).$$

We summarize the consequences of Theorem 4.25 as follows:

1. The analysis variable U^a, to which the EnKF and the PCKF provide approximations, is the sum of a Bayes estimate $u_{\mathrm{LCM}}(y)$ and the (prior) error $U - u_{\mathrm{LCM}}(Y)$ of the corresponding Bayes estimator u_{LCM}.

2. The mean of the EnKF analysis ensemble or PCKF analysis vector provide approximations to the linear posterior mean estimate. How far the latter deviates from the true posterior mean depends on the model and observation y.

3. The covariance approximated by the empirical covariance of the EnKF analysis ensemble, as well as that of the PCKF analysis vector, is independent of the actual observational data $y \in \mathbb{R}^d$. Therefore, it constitutes a prior rather than a posterior measure of uncertainty.

4. In particular, the randomness in U^a is entirely determined by the prior measures μ_0 of U and ν_ε of ε. Only the location, i.e., the mean, of U^a is influenced by the observational data y; the randomness of U^a is independent of the data y and determined only by the prior projection error $U - u_{\mathrm{LCM}}(Y)$.

5. In view of the last two items, the analysis variable U^a, and therefore the EnKF analysis ensemble or the result of the PCKF, are in general not distributed according to the posterior measure μ^y. Moreover, the difference between μ^y and the distribution of U^a depends on the data y and can become quite large for nonlinear problems, see Example 4.27.

Remark 4.26. We mention that the second and third item above explain the observations made for the EnKF by Law and Stuart [105], i.e., that "[...] (i) with appropriate parameter choices, approximate filters can perform well in reproducing the mean of the desired probability distribution, (ii) they do not perform as well in reproducing the covariance [...] ".

In the following section we will demonstrate the second and fourth item from above in numcerical examples. In order to illustrate the conceptual difference between the distribution of the analysis variable U^a and the posterior measure μ^y we

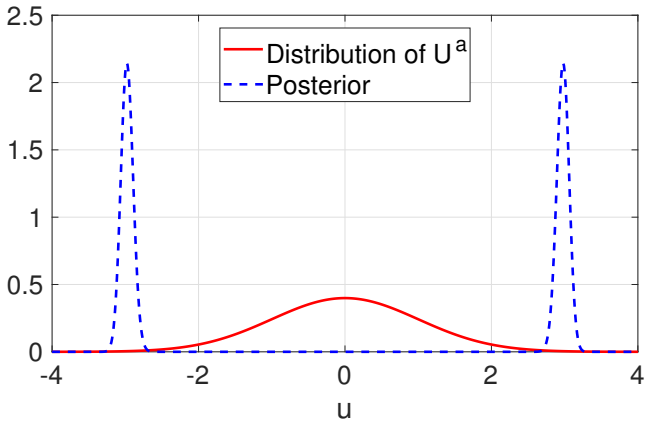

Figure 4.1.: Density of the posterior μ^y (*dashed, blue line*) and the probability density of the analysis variable U^a (*solid, red line*) for $y = 9$ and $\sigma = 0.5$.

provide a simple yet striking example, cf. also Apte et al. [3, Section 7], Le Gland et al. [75, Example 2.2] or Evensen and Van Leeuwen [63, Section 5]:

Example 4.27. We consider $U \sim N(0,1)$, $\varepsilon \sim N(0,\sigma^2)$ and $G(u) = u^2$. Given data $y \in \mathbb{R}$, the posterior measure defined by Bayes' rule is

$$\mu^y(\mathrm{d}u) = C \exp\left(-\frac{\sigma^2 u^2 + (y-u^2)^2}{2\sigma^2}\right) \mathrm{d}u.$$

Due to the symmetry of μ^y we have $u_{\mathrm{CM}}(y) = \int_{\mathcal{H}} u\, \mu^y(\mathrm{d}u) = 0$ for any $y \in \mathbb{R}^d$. Thus, $\mathbb{E}[U \mid Y] \equiv 0$ and $u_{\mathrm{LCM}} \equiv u_{\mathrm{CM}}$. In particular, we have $\mathrm{K} = 0$ due to

$$\mathrm{Cov}(U,Y) = \mathrm{Cov}(U,U^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u(u^2-1)\, \mathrm{e}^{-u^2/2}\, \mathrm{d}u = 0,$$

which in turn yields $U^a = U \sim N(0,1)$. Thus, the analysis variable is distributed according to the prior measure μ_0. This is not surprising as, by definition, its mean is the best linear approximation to the posterior mean and its fluctuation is simply the prior estimation error $U - u_{\mathrm{LCM}}(Y) = U - 0 = U$. This illustrates that U^a is suited for approximating the posterior mean, but not appropriate as a method for uncertainty quantification for the nonlinear inverse problem. As displayed in Figure 4.1, the distribution of U^a can be markedly different from the true posterior distribution.

4.4. Numerical Examples

To illustrate the application of the EnKF and PCKF to simple Bayesian inference problems, we consider in the following a one-dimensional elliptic boundary value

problem and a time-dependent RLC circuit model as dynamical system. Although in other papers much more complicated problems such as elasto-plastic deformations or Lorenz systems are considered to demonstration, we have chosen these rather simple model problems in order to illustrate the basic limitations of Kalman filter methods.

4.4.1. 1D Elliptic Boundary Value Problem

Let $D = [0,1]$ and

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(u_1)\,\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = f(x), \qquad p(0) = p_0, \quad p(1) = u_2, \tag{4.19}$$

be given where $u = (u_1, u_2)$ are unknown scalar parameters. The solution of (4.19) is

$$p(x) = p_0 + (u_2 - p_0)x + \exp(-u_1)\,(S_x(F) - S_1(F)\,x)\,, \tag{4.20}$$

where $S_x(g) := \int_0^x g(y)\,\mathrm{d}y$ and $F(x) = S_x(f) = \int_0^x f(y)\,\mathrm{d}y$. For simplicity we choose $f \equiv 1$, $p_0 = 0$ in the following and assume that noisy measurements of p have been made at $x_1 = 0.25$ and $x_2 = 0.75$ with values $y = (27.5, 79.7)$. We seek to infer u based on this data and on a priori information modelled by $(u_1, u_2) \sim N(0,1) \otimes \mathrm{Uni}(90, 110)$, where $\mathrm{Uni}(a,b)$ denotes the uniform distribution on the interval $[a,b]$. Thus the forward map here is $G(u) = (p(x_1), p(x_2))$, where p is given in (4.20) with $f \equiv 1$ and $p_0 = 0$. For the measurement noise we assume $\varepsilon \sim N(0, 0.01\,I_2)$.

Applying the EnKF. In Figure 4.2 we show the level curves of the prior and posterior densities as well as $1,000$ ensemble members of the initial and analysis ensemble obtained by the EnKF. A total ensemble size of $M = 10^5$ was chosen in order to reduce the sampling error to a negligible level. It can be seen, however, that the analysis EnKF-ensemble does not follow the posterior distribution, although its mean $(-2.92, 105.14)$ is quite close to the true posterior mean $(-2.65, 104.5)$ (computed by quadrature). To illustrate the difference between the distribution of the analysis ensemble/variable and the true posterior distribution, we present the marginal posterior distributions of u_1 and u_2 in Figure 4.3. The posterior marginals were evaluated by quadrature, whereas for the analysis ensemble we show a relative frequency plot.

We remark that slightly changing the observational data to $\tilde{y} = (23.8, 71.3)$ moves the analysis ensemble as well as the distribution of the analysis random variable much closer to the true posterior, as shown in Figure 4.4. Moreover, for these mea-

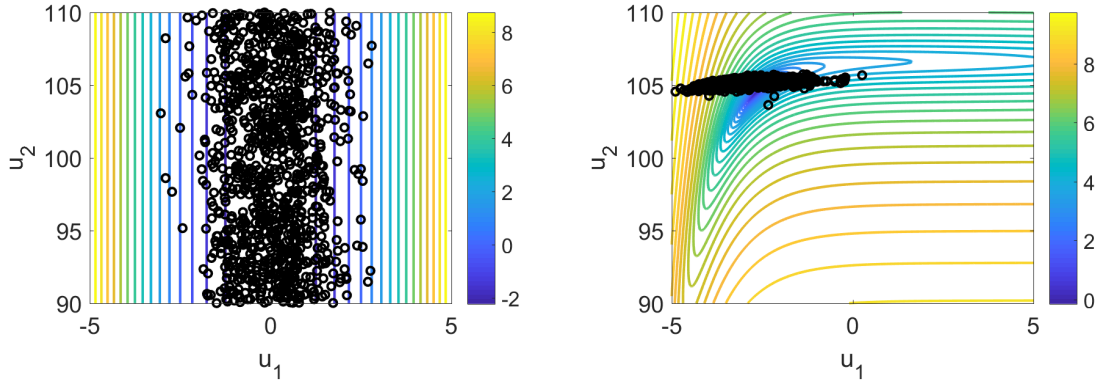

Figure 4.2.: Left: Contour plot of the negative logarithm of the prior density and the locations of 1000 ensemble members of the initial EnKF-ensemble.
Right: Contour plot of the logarithm of the negative logarithm of the posterior density and the locations of the updated 1,000 ensemble members in the analysis EnKF-ensemble.

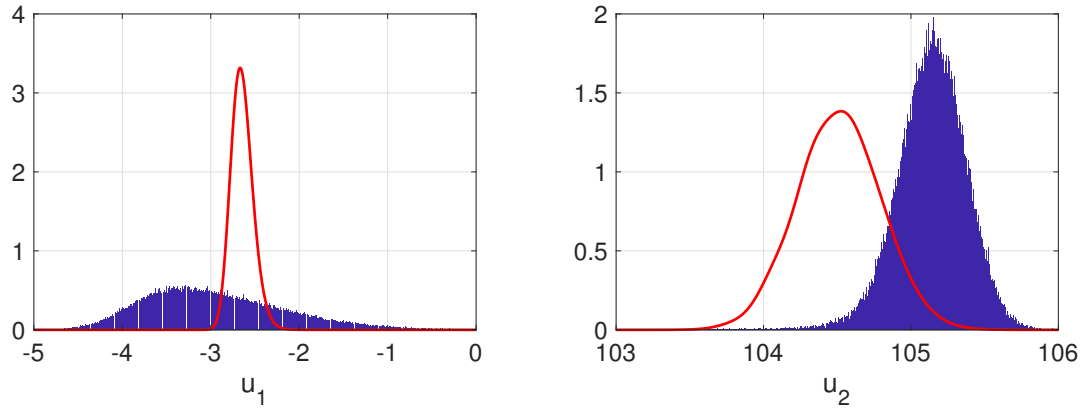

Figure 4.3.: Posterior marginals and correspoding relative frequencies in the analysis ensemble for u_1 (left) and u_2 (right).

surement values the mean of the analysis ensemble $(0.33, 94.94)$ provides a better fit to the true posterior mean $(0.33, 94.94)$.

To reaffirm the fact that only the mean of the analysis variable U^a depends on the actual data, we show density estimates for the marginals of u_1 and u_2 of U^a in Figure 4.5 obtained from the observational data $y = (27.5, 79.7)$ (blue, solid lines) and $\tilde{y} = (23.8, 71.3)$ (red, dashed lines), respectively. The density estimates were obtained by normal kernel density estimation (KDE, in this case MATLAB's `ksdensity` routine) based on the resulting analysis ensembles $(\boldsymbol{u}_1^a, \boldsymbol{u}_2^a)$ and $(\tilde{\boldsymbol{u}}_1^a, \tilde{\boldsymbol{u}}_2^a)$ for the data sets y and $\tilde{y}$, respectively. We observe that the marginal distributions of the centered ensembles coincide, in agreement with Theorem 4.25.

In addition, whenever the prior and, thus, also the posterior support for u_2 is bounded — as in this example — the EnKF may generate members in the anal-

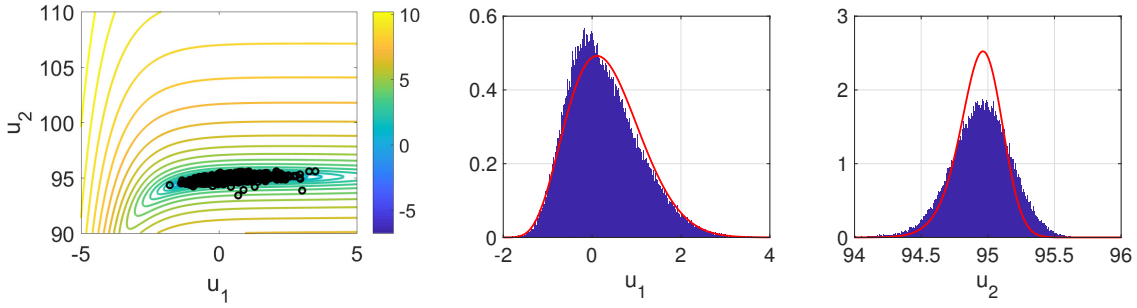

Figure 4.4.: Left: Contours of the logarithm of the negative log posterior density and locations of 1,000 members of the analysis EnKF-ensemble.
Middle, Right: Posterior marginals and corresponding relative frequencies in the analysis ensemble for u_1 (middle) and u_2 (right).

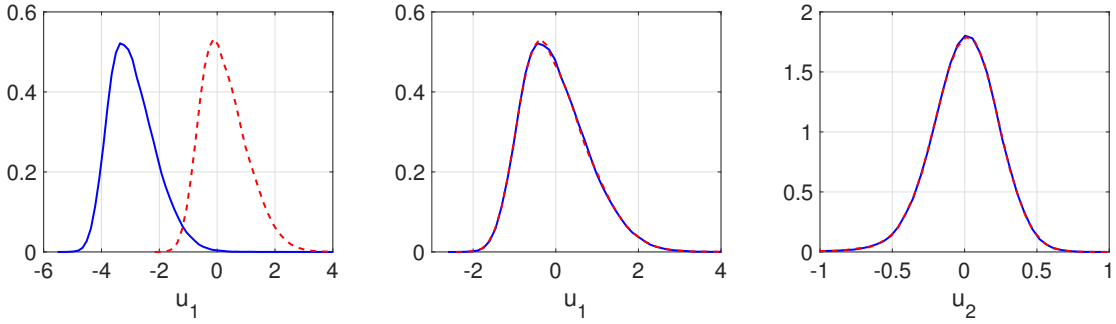

Figure 4.5.: Left: Kernel density estimates for $\boldsymbol{u}_1^a$ (*blue, solid line*) and $\tilde{\boldsymbol{u}}_1^a$ (*red, dashed line*).
Middle, Right: Kernel density estimates for $\boldsymbol{u}_i^a - \mathbb{E}[\boldsymbol{u}_i^a]$ (*blue, solid*) and $\tilde{\boldsymbol{u}}_i^a - \mathbb{E}[\tilde{\boldsymbol{u}}_i^a]$ (*red, dashed*), $i = 1, 2$.

ysis ensemble which lie outside of this support. This is a further consequence of Theorem 4.25: Since the analysis ensemble of the EnKF follows the distribution of the analysis variable rather than the true posterior distribution, ensemble members lying outside the posterior support can always occur whenever the support of the analysis variable is not a subset of the support of the posterior.

Finally, we emphasize that whether or not the distribution of the analysis variable is a good fit to the true posterior distribution depends entirely on the observed data — which can neither be controlled nor be known a priori.

Applying the PCKF. The calculations for applying the PCKF to this simple example problem can be carried out analytically: we require four independent random variables $\xi_1 \sim N(0,1)$, $\xi_2 \sim \mathrm{Uni}(0,1)$, $\xi_3 \sim N(0,1)$ and $\xi_4 \sim N(0,1)$ to define PCEs which yield random variables distributed according to the prior and error distributions:

$$U := (\xi_1,\ 90 + 20\xi_2)^\top \sim \mu_0, \qquad \varepsilon := (0.1\xi_3,\ 0.1\xi_4)^\top \sim \nu_\varepsilon.$$

Moreover, due to (4.20), $G(U)$ is also available in closed form as

$$G(U) = \begin{pmatrix} c_{11}(90 + 20\xi_2) + c_{12} \sum_{n=0}^{\infty} (-1)^n \frac{\sqrt{e}}{\sqrt{n!}} H_n(\xi_1) \\ c_{21}(90 + 20\xi_2) + c_{22} \sum_{n=0}^{\infty} (-1)^n \frac{\sqrt{e}}{\sqrt{n!}} H_n(\xi_1) \end{pmatrix},$$

where H_n denotes the nth normalized Hermite polynomial and $c_{11}, c_{12}, c_{21}, c_{22}$ can be deduced from inserting $x = 0.25$ and $x = 0.75$ into (4.20). Here we have used the Hermite expansion of $\exp(-\xi)$, see also Ullmann [172, Example 2.2.7]. Thus, the chaos coefficient vectors of U and $G(U) + \varepsilon$ w.r.t. the polynomials

$$P_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = H_{\alpha_1}(\xi_1)\, L_{\alpha_2}(\xi_2)\, H_{\alpha_3}(\xi_3)\, H_{\alpha_4}(\xi_4), \qquad \boldsymbol{\alpha} \in \mathbb{N}_0^4,$$

can be obtained explicitly where H_α and L_α denote the normalized Hermite and Legendre polynomials of degree α, respectively. In particular, the nonvanishing chaos coefficients involve only the basis polynomials

$$P_0(\boldsymbol{\xi}) \equiv 1, \quad P_1(\boldsymbol{\xi}) = L_1(\xi_2), \quad P_2(\boldsymbol{\xi}) = H_1(\xi_3), \quad P_3(\boldsymbol{\xi}) = H_1(\xi_4)$$

and $P_\alpha(\boldsymbol{\xi}) = H_{\alpha-3}(\xi_1)$ for $\alpha \geq 4$. Arranging the resulting chaos coefficients $u_\alpha \in \mathbb{R}^2$ and $g_\alpha \in \mathbb{R}^2$, $\alpha \neq 0$, of U and $G(U)$, respectively, as column vectors in matrices $\boldsymbol{U}, \boldsymbol{G} \in \mathbb{R}^{2\times\mathbb{N}}$ we obtain

$$\mathrm{K} = \boldsymbol{U}\boldsymbol{G}^\top \left(\boldsymbol{G}\boldsymbol{G}^\top + 0.01 I_2 \right)^{-1}.$$

Thus, the only numerical error incurred in applying the PCKF in this example is the truncation of the PCE. We have carried out a simulation using a truncated PCE of length $J = 4 + 50$ according to the reduced basis above. In particular, we evaluated the approximation K_J to K by using only the first 53 columns of $\boldsymbol{G}$ in the formula above and then performed the update of the chaos coefficients according to (4.10). Subsequently $M = 10^5$ samples of the resulting random variable U_J^a were drawn, but, since the empirical distributions were essentially indistinguishable from those obtained by the EnKF described previously, they are omitted here.

4.4.2. Dynamical System: RLC circuit

This time we consider sequential data assimilation in a simple dynamical system: a damped LC-circuit or RLC-circuit. Denoting the initial voltage by U_0, the resistance by R, the inductance by L and the capacitance by C, and assuming $R < 2\sqrt{LC}$, the

voltage and current in the circuit can be modelled as

$$U(t) = U_0 \ \mathrm{e}^{\delta t} \ \big(\cos(w_e t) + \frac{\delta}{w_e}\sin(w_e t)\big), \tag{4.21a}$$

$$I(t) = -\frac{U_0}{w_e L} \ \mathrm{e}^{\delta t} \ \sin(w_e t), \tag{4.21b}$$

where $\delta = R/(2L)$, $w_e = \sqrt{w_0^2 - \delta^2}$ and $w_0 = 1/\sqrt{LC}$. The data assimilation setting is now as follows. We observe the state of the system (4.21) at four time points $t_n = 5n$, $n = 1, \dots, 4$, where all observations $z \in \mathbb{R}^8$ are corrupted by measurement noise $\varepsilon \sim N(0, \mathrm{diag}(\sigma_1^2, \dots, \sigma_8^2))$. Here, we set σ_n^2 to be 10% of the true, undisturbed observations. We want to infer U_0 and L based on these observations, i.e, the unknown is $u = (U_0, L)$, and we take as the prior $(U_0, L) \sim N(0.5, 0.25) \otimes \mathrm{Uni}(1, 5)$.

We will only apply the EnKF in this example for simplicity. Given the observations $y \in \mathbb{R}^8$ we compare two assimilation strategies using the EnKF:

- *Simultaneous*: We apply the EnKF to the inverse problem

  $$y = G(u) + \varepsilon,$$

  where G maps (U_0, L) to the states $(U(t_1), I(t_1), \dots, U(t_4), I(t_4)) \in \mathbb{R}^8$. Thus, we perform one EnKF update using all the available data at once, resulting in one EnKF analysis ensemble.

- *Sequential*: We apply the EnKF to the inverse problem

  $$y_n = G_n(u) + \varepsilon_n, \qquad n = 1, \dots, 4,$$

  where G_n maps (U_0, L) to the state $(U(t_n), I(t_n)) \in \mathbb{R}^2$. In particular, we will perform four EnKF updates using at each update only the corrupted data $y_n = (U(t_n) + \varepsilon_{2n-1}, I(t_n) + \varepsilon_{2n})$. This yields, for each update, one EnKF analysis ensemble which serves as the initial ensemble for the next update.

Again we use two different data sets $y, \tilde{y}$[1], obtained by two realizations of ε given the solution of (4.21) for $U_0 = 0.75, R = 0.5, L = 1.5, C = 0.5$. The resulting posteriors and EnKF analysis ensembles for the simultaneous and sequential update are presented in Fig. 4.6. We make again the observation, that for different data sets the analysis ensemble follows a distribution which is in one case quite close and in the other case quite far away from the true posterior distribution. This can be

[1] $y = (0.505, 0.237, 0.014, 0.096, 0.036, 0.011, -0.002, -0.003)$ and $\tilde{y} = (0.265,\ 0.066,\ 0.058,\ 0.002,\ 0.021, 0.012, 0.007, -0.01)$

also seen in Table 4.1, where we compare the empirical means of the EnKF analysis ensembles with the true posterior mean for both data sets y and $\tilde{y}$.

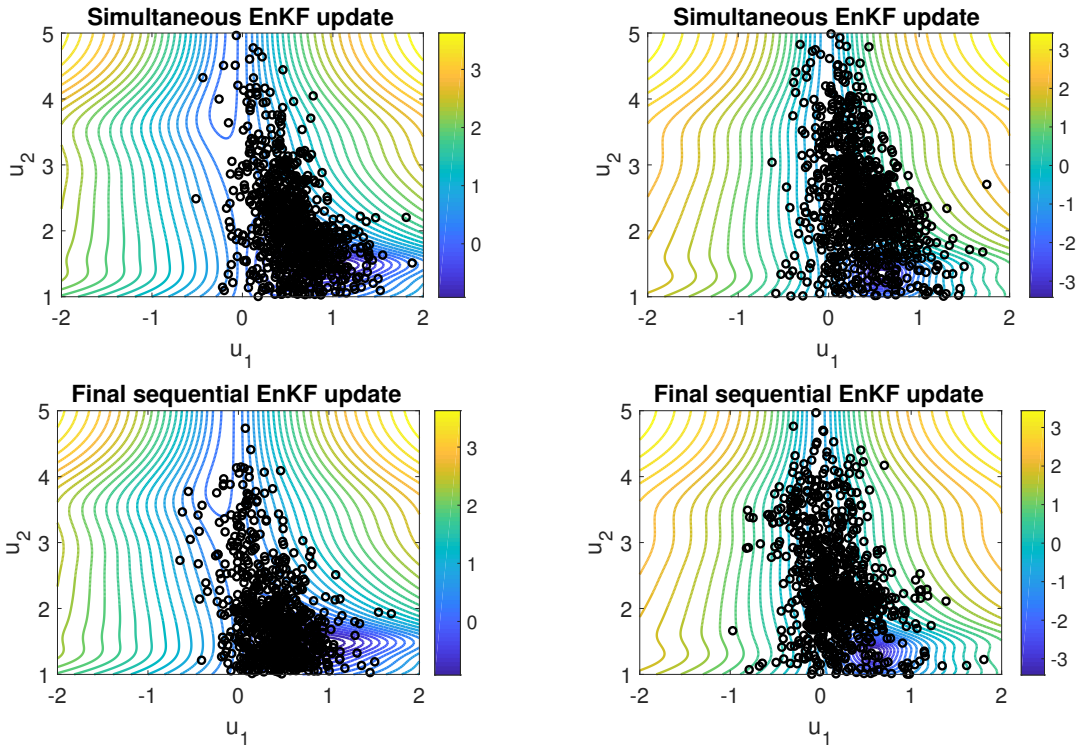

Figure 4.6.: Contours of the logarithm of the negative log posterior density and locations of 1,000 members of the analysis ensembles resulting from simultaneous and sequential EnKF updates given the observational data y (left column) and $\tilde{y}$, respectively (right column).

update	EnKF mean for data y	posterior mean for data y	EnKF mean for data $\tilde{y}$	posterior mean for data $\tilde{y}$
1	(0.42, 1.56)	(0.42, 2.42)	(0.27, 2.25)	(0.35, 2.61)
2	(0.44, 1.53)	(0.39, 2.36)	(0.20, 2.20)	(0.32, 2.56)
3	(0.43, 1.59)	(0.38, 2.34)	(0.19, 2.26)	(0.31, 2.52)
4	(0.43, 1.59)	(0.38, 2.32)	(0.19, 2.24)	(0.30, 2.50)
Simu.	(0.58, 1.84)	(0.38, 2.32)	(0.38, 2.40)	(0.30, 2.50)

Table 4.1.: Means of the EnKF analysis ensembles and corresponding true posterior means.

Finally, we are again interested in the marginals of the posterior and the associated histograms of the EnKF analysis ensembles which give us a rough impression of the difference between the distribution of the analysis variable and the true posterior. In Fig. 4.7 and Fig. 4.8 we compare for both data sets the marginals of u_2 and the corresponding relative frequencies in the analysis ensemble resulting from sequential and simultaneous updating. The distribution of the simultaneous EnKF analysis ensemble does not depend on the data (as predicted by our theory)

whereas the distribution of the final EnKF analysis ensemble for the sequential updating clearly does in this example. The latter is probably caused by the nonlinearity of the forward map G: in the sequential updating the former analysis variable U_n^a serves as initial one for the current update step $n+1$, therefore, the difference in the mean of the former analysis variables $U_n^a, \tilde{U}_n^a$ for different data sets $y, \tilde{y}$ may yield different forecast random variables $G(U_n^a), G(\tilde{U}_n^a)$ due to the nonlinearity of G which in turn yields different next analysis variables $U_{n+1}^a, \tilde{U}_{n+1}^a$.

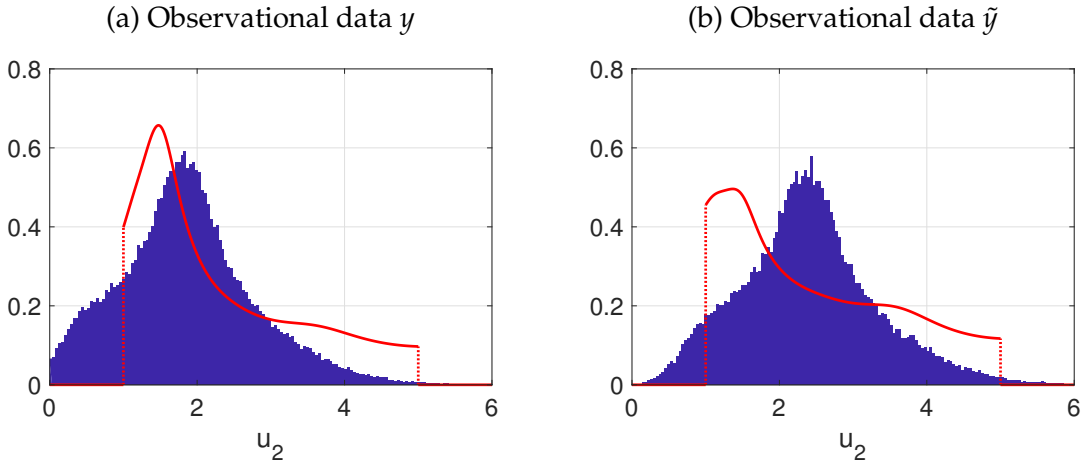

Figure 4.7.: Posterior marginal distribution of u_2 (red line) and corresponding relative frequencies in the analysis ensemble resulting from the simultaneous EnKF update.

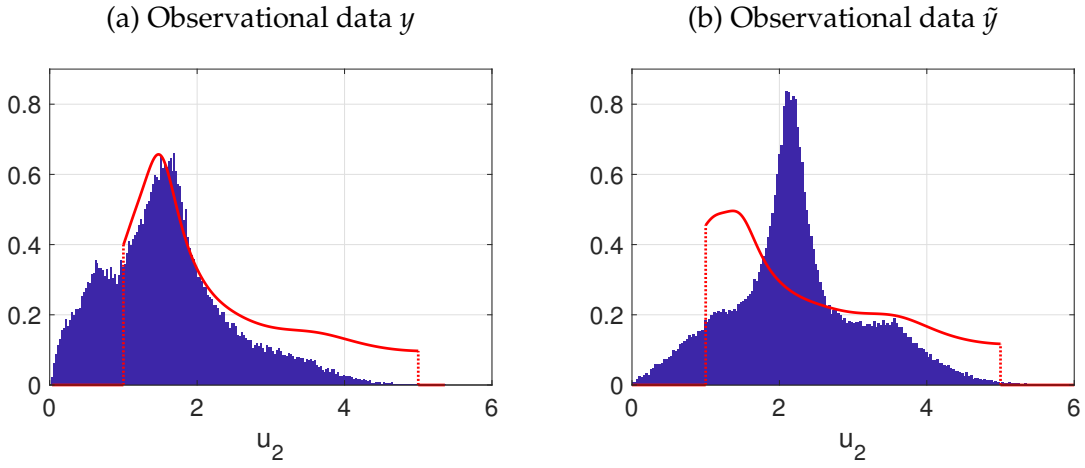

Figure 4.8.: Posterior marginal distribution of u_2 (red line) and corresponding relative frequencies in the final analysis ensemble resulting from sequential EnKF updates.

Chapter 5

Markov Chain Monte Carlo Methods

This chapter is based on [146]. Although the presented mathematical content is in general the same, the presentation and explanation has been adapted and extended and also some new numerical results added.

We will consider Markov chain Monte Carlo methods for (approximate) sampling of a *target probability measure* μ defined on a separable Hilbert space $\mathcal{H}$ by

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u)), \qquad u \in \mathcal{H}, \tag{5.1}$$

where μ_0 denotes a Gaussian *reference measure*, e.g., $\mu_0 = N(0, C)$, on $\mathcal{H}$ and $\Phi\colon \mathcal{H} \to \mathbb{R}_+$ a measurable mapping. Such probability measures μ arise as posterior distributions in Bayesian inference with μ_0 as a Gaussian prior as described in Chapter 3. Although we will sometimes recall the Bayesian inference setting, we rather use the terminology of target and reference measure for μ and μ_0, respectively, as it is common in the literature on MCMC methods.

Since the measure μ is only known up to a normalizing constant and Φ is in general only available in the form of function evaluations, a direct sampling of μ is often not feasible. Another idea to generate samples which are (approximately) distributed according to μ, which goes back at least to Metropolis et al. [119], is to construct a *Markov chain*, i.e., a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ on $\mathcal{H}$ satisfying the Markov property, with limit distribution μ. Then X_n for n sufficiently large follows a distribution close to μ and given *ergodicity* — details will be given in the subsequent section — averages taken along a path of the Markov chain will converge to expectations w.r.t. μ. However, in comparison to basic Monte Carlo simulations the samples generated by a Markov chain Monte Carlo simulation are typically correlated. In this case, the MCMC sampling is statistically less efficient than basic MC sampling and it is the *autocorrelation* of the Markov chain which serves for example for comparing different MCMC algorithms.

We will focus on a particular class of MCMC methods: *Metropolis-Hastings (MH) algorithms*. These algorithms go back to the seminal paper of Metropolis et al. [119] and were later generalized by W. K. Hastings [84]. They are usually easy to implement and quite popular. As a historical note, at the beginning MH algorithms were mainly employed for computations in statistical mechanics and physics. Just since Gelfand and Smith [67] and Tierney [170] they have also became one of the standard computational methods in Bayesian inference. Their principal idea to construct Markov chains with μ as their limit distribution is as follows: Given the current state of the Markov chain $X_n = u$ a new state v is drawn according to a measure $P(u, \mathrm{d}v)$ depending on u. Thus, P itself is a stochastic or *Markov kernel* as stated in Defintion 3.11 which in this context is typically called *proposal kernel*. Then the new state y is only accepted, i.e., $X_{n+1} = v$, with a certain probability $\alpha(u, v)$, otherwise the Markov chain remains where it is, i.e., $X_{n+1} = u$. This *acceptance probability* $\alpha \colon \mathcal{H} \times \mathcal{H} \to [0, 1]$ is chosen in such a way that the resulting Markov chain has μ as its *invariant* measure. The latter means that if $X_n \sim \mu$, then also $X_{n+1} \sim \mu$ — a necessary property for μ to be the limit distribution of the Markov chain. Given a proposal kernel P the choice of the acceptance probability α follows the guidelines given by Tierney [171] and, thus, the construction of the proposal P remains as the only parameter of design for MH algorithms.

A simple proposal kernel in finite dimensions is the *(Gaussian) random walk proposal* $P(u) = N(u, s^2 C_P)$ where $s > 0$ denotes a (proposal) stepsize parameter and $C_P \in \mathbb{R}^{N \times N}$ a covariance matrix, e.g., the covariance $C_P = C$ of the Gaussian reference measure μ_0. The efficiency of the MH algorithm based on this proposal as well as of other common MH algorithms suffers in high dimensional state spaces: as the state-space dimension N increases we have to decrease the stepsize parameter s of the corresponding proposals in order to maintain the same *average acceptance rate*, i.e., the mean of $\alpha(u, v)$ w.r.t. the measure $P(u, \mathrm{d}v)\mu(\mathrm{d}u)$, see Roberts and Rosenthal [141]. Thus, either by decreasing s or by the decreased average acceptance rate the Markov chain will show a higher autocorrelation and the statistical efficiency of the MCMC scheme will deteriorate. This is a clear drawback for many applications such as Bayesian inference for functions, since then the unknown to infer is an element of an infinite dimensional Banach or Hilbert space, respectively, or at least from a high dimensional space stemming from numerical discretizations.

These kind of applications motivated the recent research on MH algorithms for sampling from target measures in infinite dimensional Hilbert spaces. For example, as shown by Cotter et al. [35] the simple random walk proposal mentioned above does not yield a well-defined MH algorithm in infinite dimensions, since then a corresponding acceptance probability according to Tierney [171] does not

exist. Beskos et al. [15] suggested a modified Gaussian random walk proposal $P(u) = N(\sqrt{1-s^2}u, s^2C)$ which is μ_0-reversible, i.e., it satisfies the *detailed balance equation* $P(u, \mathrm{d}v)\mu_0(\mathrm{d}u) = P(v, \mathrm{d}u)\mu_0(\mathrm{d}v)$ in the sense of measures on $\mathcal{H} \times \mathcal{H}$. The μ_0-reversibility leads to a well-defined acceptance probability following Tierney [171] and, thus a well-defined MH algorithm in arbitrary separable Hilbert spaces. This particular proposal was referred to by Cotter et al. [35] as *preconditioned Crank-Nicolson (pCN)* proposal. Furthermore, it was shown by Hairer et al. [81] that the Markov chain of the resulting *pCN Metropolis* algorithm has *dimension-independent efficiency* stated in terms of a dimension-independent *spectral gap* of the associated Markov operators – details about spectral gaps will be given below and in Section 5.3.

The main purpose of this chapter is to extend the pCN proposal in order to allow for other proposal covariances than the covariance C of the reference measure μ_0. Thus, we will combine the idea of dimension-independent MH algorithms in Hilbert spaces with another recent development in Markov chain Monte Carlo methods: exploiting *geometric information* about the target measure μ. Such information can, for instance, be the varying concentration of the target marginal distribution in different directions provided by the (anisotropy of the) target covariance operator, or the local curvature of the log density Φ which can be employed as a metric tensor measuring distances on the "manifold generated by μ". By using such geometric information for proposing new states, one might obtain a larger average step size for the resulting Markov chain and, thus, a faster state space exploration. However, this idea is not entirely new. It is already mentioned by Tierney [170] who suggested to choose a proposal covariance matrix which is similar to the target covariance matrix. Later in [73] Girolami and Calderhead explain how to propose new states in finite dimensions using general local metric tensors which relate to local curvatures of the Lebesgue density of μ. Moreover, in [118] Martin et al. employ the Hessian of the negative log density Φ of μ as such a local curvature information to design a stochastic Newton MH method in finite dimensions, and Cui et al. [37] and Law [104] outline a Gauss-Newton variant for capturing global curvature in an infinite dimensional setting.

Our approach for adapting the pCN proposal to the target measure μ has a similiar motivation as the proposals considered by Cui et al. [37] and Law [104]. Based on approximating the nonlinear forward map in a Bayesian inference problem by local linearization which leads to a Gaussian approximation of the posterior measure with a particular covariance, we consider Gaussian proposals with covariances of the form $(C + \Gamma)^{-1}$ where Γ denotes an arbitrary self-adjoint and positive bounded linear operator. By enforcing μ_0-reversibility we derive our class

of *generalized pCN (gpCN) proposal kernels* P_Γ. Besides proving well-definedness of the resulting Metropolis-Hastings algorithm in the infinite dimensional setting in Section 5.2 we also present a geometric ergodicity result for the *gpCN Metropolis* algorithm in Section 5.3. The latter roughly means that the distribution of the nth step of the Markov chain converges exponentially fast to its invariant measure. The proof is based on an L^2-spectral gaps approach which appears to be a common strategy in literature: each μ-reversible Markov chain is associated with a self-adjoint bounded linear operator on $L^2_\mu(\mathcal{H})$ – its *Markov operator* – and it is well known, see Roberts and Rosenthal [139], that the L^2_μ-geometric ergodicity of a μ-reversible Markov chain is equivalent to a positive (L^2_μ-)spectral gap of its associated Markov operator. For details we refer to Section 5.3. In particular, we derive and employ a new comparison theorem for spectral gaps of Markov operators. By verifying the assumptions of this comparison result for Markov chains generated by the pCN and gpCN Metropolis algorithms, we arrive at our main theoretical result Theorem 5.45. This states that whenever the pCN Metropolis algorithm yields a nonzero L^2_μ-spectral gap, see Hairer et al. [81] for conditions, then a restriction of the gpCN Metropolis algorithm which targets a restriciton μ_R of μ to an arbitrary R-ball yields a nonzero $L^2_{\mu_R}$-spectral gap. Although the mentioned restricted version has a slightly different invariant measure, μ_R, the difference between both in total variation distance can be made arbitrarly small by choosing R sufficiently large. Thus, we will show the exponentially fast convergence of the gpCN Metropolis algorithm to an arbitrarily close approximation of the target.

Moreover, our analysis enables us to extend the gpCN Metropolis to allow also state-dependent proposal covariances which will be outlined in Section 5.5. Such MH algorithms with state-dependent proposal covariances have also become an active field of research in recent years, we refer, e.g., to the works [73, 118, 113, 102, 12]. In particular, Beskos et al. [12] follow similar ideas and constructions as ours.

Finally, we present some numerical illustrations for the developed algorithm in Section 5.6 and show that it outperforms other classical MH algorithms including the pCN Metropolis. In particular, we observe a performance of the new algorithm which seems to be independent of dimension and robust to the noise or likelihood variance. The latter refers to the variance of the observational noise in a Bayesian inference setting and typically a decreased noise variance leads to more concentrated posterior measures which are then usually harder to sample by MH algorithms. However, our numerical results indicate that mimicking the behavior of the posterior covariance in the proposal covariances as described above seems to yield a larger robustness w.r.t. decreasing noise variance. This fact is also adressed in Chapter 6 in a mathematically rigorous way.

5.1. Preliminaries and Metropolis-Hastings Algorithms

We refer again to the basic notation of this thesis as introduced in Section 1.2, i.e., $\mathcal{H}$ denotes a separable Hilbert space with inner-product and norm $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\| \cdot \|_{\mathcal{H}}$, respectively, and $(\Omega, \mathcal{A}, \mathbb{P})$ an underlying probability space. Further, let $\mu_0 = N(0, C)$ denote a Gaussian reference measure on $\mathcal{H}$ with $C \in \mathcal{L}^1_+(\mathcal{H})$ being nonsingular, i.e., $\ker C = \{0\}$. Moreover, in the remainder of the chapter we assume that the target measure $\mu \in \mathcal{P}(\mathcal{H})$ is given by (5.1).

5.1.1. Markov Chains and Markov Chain Monte Carlo

We provide only a short introduction to Markov chains and Markov chain Monte Carlo (MCMC) methods on general state spaces. For more details, we refer to, e.g., Meyn and Tweedie [120].

Definition 5.1. A *Markov chain* in $\mathcal{H}$ is a sequence of $\mathcal{H}$-valued random variables $(X_n)_{n \in \mathbb{N}}$ satisfying *the Markov property*, i.e.,

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in A \mid X_n) \qquad \mathbb{P}\text{-a.s.}$$

for each $n \in \mathbb{N}$ and $A \in \mathcal{B}(\mathcal{H})$. A stochastic or Markov kernel $K \colon \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0, 1]$ is the *transition kernel* of a Markov chain $(X_n)_{n \in \mathbb{N}}$ if for all $n \in \mathbb{N}$ and $A \in \mathcal{B}(\mathcal{H})$ there holds

$$K(X_n, A) = \mathbb{P}(X_{n+1} \in A \mid X_n) \qquad \mathbb{P}\text{-a.s.}$$

Most properties of a Markov chain can be expressed as properties of its transition kernel. We therefore introduce the following notions.

Definition 5.2. Let K be a Markov kernel on $\mathcal{H}$ and $\nu \in \mathcal{P}(\mathcal{H})$. Then by νK we denote the probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ given by

$$(\nu K)(A) := \int_{\mathcal{H}} K(v, A)\, \nu(\mathrm{d}v) \qquad \forall A \in \mathcal{B}(\mathcal{H}). \tag{5.2}$$

Moreover, for $n \in \mathbb{N}$ we define in a recursive manner the Markov kernel K^n on $\mathcal{H}$ by

$$K^n(u, A) := \int_{\mathcal{H}} K^{n-1}(v, A)\, K(u, \mathrm{d}v) \qquad \forall u \in \mathcal{H},\ \forall A \in \mathcal{B}(\mathcal{H}). \tag{5.3}$$

By this notation, we can express the distribution of the nth state X_n of a Markov chain with transition kernel K and initial distribution $X_1 \sim \nu$ simply by $X_n \sim \nu K^{n-1}$.

Remark 5.3 (On the notation νK)**.** The definition of νK is of course a slight abuse of notation, since K denotes a Markov kernel but in νK the symbol K rather plays the role of a mapping from $\mathcal{P}(\mathcal{H})$ to $\mathcal{P}(\mathcal{H})$. Given this point of view, it seems odd to place ν on the left hand side of K instead of writing $K\boldsymbol{u}$. However, the notation νK is quite common in the Markov chain literature. It likely stems from Markov chains in discrete state spaces where the transition kernel K is simply a stochastic matrix $K \in \mathbb{R}^{N\times N}$ — typically a row stochastic matrix — and, thus, for a column vector $\boldsymbol{\nu} \in \mathbb{R}^N$ of initial probabilities, the vector given by $\boldsymbol{\nu}^\top K$ describes the distribution of the next state of the Markov chain.

Definition 5.4. Let $\mu \in \mathcal{P}(\mathcal{H})$ and K be the transition kernel of a Markov chain $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$. The measure μ is called an *invariant measure* of the Markov chain $(X_n)_{n\in\mathbb{N}}$ or *invariant w.r.t.* K if $\mu = \mu K$. Furthermore, the kernel K is called *μ-reversible* if it satisfies the *detailed balance condition*

$$K(u, \mathrm{d}v)\,\mu(\mathrm{d}u) = K(v, \mathrm{d}u)\,\mu(\mathrm{d}v) \tag{5.4}$$

where equality holds in the sense of measures on $\mathcal{H} \times \mathcal{H}$.

It is easily verified that (5.4) implies the invariance of μ w.r.t. K, i.e., $\mu = \mu K$: for any $A \in \mathcal{B}(\mathcal{H})$ there holds

$$\begin{aligned}(\mu K)(A) &= \int_{\mathcal{H}} K(u, A)\,\mu(\mathrm{d}u) = \int_{\mathcal{H}}\int_A K(u, \mathrm{d}v)\,\mu(\mathrm{d}u) = \int_{\mathcal{H}}\int_A K(v, \mathrm{d}u)\,\mu(\mathrm{d}v)\\ &= \int_A\int_{\mathcal{H}} K(v, \mathrm{d}u)\,\mu(\mathrm{d}v) = \int_A K(v, \mathcal{H})\,\mu(\mathrm{d}v) = \int_A \mu(\mathrm{d}v) = \mu(A).\end{aligned}$$

We will now introduce a notion of geometric convergence of Markov chains to their stationary distribution.

Definition 5.5. A Markov chain $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ with transition kernel K is $L^2_\mu(\mathcal{H})$-*geometrically ergodic* if there exists a number $r \in [0, 1)$ such that for any probability measure ν which has a density $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2_\mu(\mathcal{H})$ w.r.t. μ there holds

$$d_{\mathrm{TV}}(\nu K^n, \mu) \le C_\nu\, r^n \qquad \forall n \in \mathbb{N}.$$

Remark 5.6 (On orders of convergence)**.** Consider again Markov chain in a finite state space with a row-stochastic matrix $K \in \mathbb{R}^{N\times N}$ representing its transition kernel. Then the convergence of the distribution $\nu^\top K^n \in \mathbb{R}^N$ of the nth state of the

Markov chain given an initial distribution $\nu \in \mathbb{R}^N$ is typically a geometric convergence as in Definition 5.5 where r depends on the eigenvalues of K. However, already for Markov chains in continuous state space such as $\mathbb{R}^N$ we can distinguish between *uniform ergodicity*, meaning that there exists a constant $C < \infty$ and an $r \in [0, 1)$ such that

$$d_{\mathrm{TV}}(\delta_x K^n, \mu) \leq C\, r^n \qquad \forall n \in \mathbb{N}, \quad \forall x \in \mathbb{R}^d,$$

and *geometric ergodicity*, meaning that there exists a measurable function $C\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$d_{\mathrm{TV}}(\delta_x K^n, \mu) \leq C(x)\, r^n \qquad \forall n \in \mathbb{N}, \quad \forall x \in \mathbb{R}^d.$$

Besides that, there exist also results on other orders of convergence such as subgeometric convergence for Markov chains in general state spaces, see, e.g., Douc et al. [47] or Kovchegov and Michalowksi [100].

If the distribution of X_n converges to μ, then the Markov chain $(X_n)_{n\in\mathbb{N}}$ can be used for approximate sampling from μ. This leads to *Markov chain Monte Carlo* methods for the computation of expectations. In particular, the expectation $\mathbb{E}_\mu(f)$ of a function $f\colon \mathcal{H} \to \mathbb{R}$ w.r.t. μ can then be approximated by the path average

$$S_{n,n_0}(f) := \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0}), \tag{5.5}$$

where n is the sample size and n_0 a burn-in parameter to decrease the influence of the initial distribution. In fact, a strong law of large numbers and also a central limit theorem holds for the path average S_{n,n_0} under appropriate assumptions.

Theorem 5.7 (Central limit theorem for reversible Markov chains [139, Corollary 2.1], [98])**.** Let $(X_n)_{n\in\mathbb{N}}$ be a μ-reversible and $L^2_\mu(\mathcal{H})$-geometrically ergodic Markov chain and $f \in L^2_\mu(\mathbb{R})$. Then there holds

$$\sqrt{n}\left(S_{n,n_0}(f) - \mathbb{E}_\mu[f]\right) \xrightarrow{\mathcal{D}} N(0, \sigma_f^2)$$

where σ_f^2 denotes the *asymptotic variance* $\sigma_f^2 := \lim_{n\to\infty} n \operatorname{Var}\left(S_{n,n_0}(f)\right)$ which, in this case, satisfies

$$\sigma_f^2 = \operatorname{Var}(f(X_1)) + 2\sum_{k=1}^{\infty} \operatorname{Cov}(f(X_1), f(X_{1+k})) < \infty. \tag{5.6}$$

We want to motivate the specific form (5.6) of the asymptotic variance and, first,

consider the variance of $S_{n,n_0}(f)$:

$$\begin{aligned}\operatorname{Var}\left(S_{n,n_0}(f)\right) &= \operatorname{Var}\left(\frac{1}{n}\sum_{j=1}^{n} f(X_{j+n_0})\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Cov}\left(f(X_{i+n_0}), f(X_{j+n_0})\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\operatorname{Var}(f(X_{j+n_0})) + \frac{1}{n^2}\sum_{\substack{i,j=1 \\ i\neq j}}^{n}\operatorname{Cov}\left(f(X_{i+n_0}), f(X_{j+n_0})\right).\end{aligned}$$

If we now assume that $(X_n)_{n\in\mathbb{N}}$ is μ-reversible and $X_1 \sim \mu$, then $X_n \sim \mu$ for any $n \in \mathbb{N}$ which further implies that (X_n, X_{n+k}) for $n \in \mathbb{N}$ follows the same distribution as (X_1, X_{1+k}). Hence, there holds $\operatorname{Var}(f(X_n)) = \operatorname{Var}(f(X_1))$ as well as $\operatorname{Cov}(f(X_{i+n_0}), f(X_{j+n_0})) = \operatorname{Cov}(f(X_1), f(X_{1+|i-j|}))$ and we get

$$\operatorname{Var}\left(S_{n,n_0}(f)\right) = \frac{1}{n}\operatorname{Var}(f(X_1)) + \frac{2}{n}\sum_{k=1}^{n}\operatorname{Cov}\left(f(X_1), f(X_{1+k})\right).$$

Of course, the assumption $X_1 \sim \mu$ is rather academic and, in general, not given in practice. However, since the Markov chain in Theorem 5.7 is assumed to be $L^2_\mu(\mathcal{H})$-geometrically ergodic, the distribution of its nth state X_n converges exponentially fast to μ as $n \to \infty$.

By Theorem 5.7 and the reasoning above the asymptotic variance σ_f^2 in (5.6) provides a measure for the statistical efficiency of Markov chain Monte Carlo methods and can therefore be used to compare them. We introduce two common terms related to σ_f^2 in the following definition which we will also employ in the numerical experiments in Section 5.6.

Definition 5.8 (Integrated autocorrelation time, effective sample size)**.** Under the assumptions and with the notations of Theorem 5.7 the *integrated autocorrelation time* τ_f of the stochastic process $(f(X_n))_{n\in\mathbb{N}}$ is given by

$$\tau_f := \frac{\sigma_f^2}{\operatorname{Var}_\mu(f)} = 1 + 2\sum_{k=1}^{\infty}\operatorname{Corr}(f(X_1), f(X_{1+k})),$$

and for the path average $S_{n,n_0}(f)$ as given in (5.5) we define the associated *effective sample size* by

$$\mathrm{ESS} = \mathrm{ESS}(n, f, (X_k)_{k\in\mathbb{N}}) := \frac{n}{\tau_f}.$$

Remark 5.9. The value of $\mathrm{ESS}(n, f, (X_k)_{k\in\mathbb{N}})$ corresponds to the number of independent samples $\tilde{X}_k \sim \mu$ which yield the same mean squared error as the MCMC estimator $S_{n,n_0}(f)$ for computing $\mathbb{E}_\mu(f)$. This can be justified under the assumption

that $X_{n_0} \sim \mu$, since then a result by Rudolf [145, Proposition 3.26] yields

$$\lim_{n\to\infty} n \cdot \mathbb{E}\left|S_{n,n_0}(f) - \mathbb{E}_\mu(f)\right|^2 = \sigma_f^2 = \tau_f \cdot \mathrm{Var}_\mu(f)$$

and

$$n \cdot \mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n} \tilde{X}_k - \mathbb{E}_\mu(f)\right|^2\right] = \mathrm{Var}_\mu(f).$$

5.1.2. Metropolis-Hastings Algorithms and the pCN Metropolis Algorithm

We will focus on Markov chains generated by Metropolis-Hastings algorithms.

Definition 5.10 (Metropolis-Hastings algorithm)**.** Let P denote a Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and $\alpha\colon \mathcal{H}\times\mathcal{H} \to [0,1]$ be a measurable function. Then a *Metropolis-Hastings algorithm* with *proposal kernel* P and *acceptance probability* α generates recursively realizations of a Markov chain $(X_n)_{n\in\mathbb{N}}$ in the following way:

1. Given the current state $X_n = u$, draw independently a sample v of a random variable $V \sim P(u,\cdot)$ and a sample a of a random variable $A \sim \mathrm{Uni}(0,1)$.
2. If $a < \alpha(u,v)$, then set $X_{n+1} = v$, otherwise set $X_{n+1} = u$.

A Markov chain which is generated by a Metropolis-Hastings algorithm possesses a transition kernel of a particular form.

Definition 5.11 (Metropolis kernel)**.** A *Metropolis kernel* M on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ is a Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ which can be written as

$$M(u, \mathrm{d}v) = \alpha(u,v)P(u,\mathrm{d}v) + \delta_u(\mathrm{d}v)\int_{\mathcal{H}} (1-\alpha(u,w))\,P(u,\mathrm{d}w) \tag{5.7}$$

where P denotes another Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and $\alpha\colon \mathcal{H}\times\mathcal{H} \to [0,1]$ a measurable function.

Proposition 5.12. If a Metropolis-Hastings algorithm uses a proposal kernel P and an acceptance probability α the transition kernel of the resulting Markov chain is given by (5.7).

Remark 5.13 (On notation)**.** In this chapter we use the following notational convention: K denotes a general Markov or transition kernel, P denotes a Markov kernel employed as proposal kernel in a Metropolis-Hastings algorithm and M denotes the transition kernel of a Markov chain generated by a Metropolis-Hastings algorithm.

We hope that this notation makes it easier for the reader to follow the presentation and statements of results.

It is well known, see Tierney [171], that a Metropolis kernel M is reversible w.r.t. μ if the associated acceptance probability $\alpha(\cdot,\cdot)$ is chosen in a specific way. In order to state the latter we first define two measures on $(\mathcal{H}\times\mathcal{H}, \mathcal{B}(\mathcal{H})\otimes\mathcal{B}(\mathcal{H}))$

$$\eta(\mathrm{d}u,\mathrm{d}v) := P(u,\mathrm{d}v)\,\mu(\mathrm{d}u), \qquad \eta^\top(\mathrm{d}u,\mathrm{d}v) := \eta(\mathrm{d}v,\mathrm{d}u),$$

given a proposal kernel P. Assume now the two measures η and $\eta^\top$ are equivalent, i.e., the Radon-Nikodym derivative $\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}$ exists and is positive on $\mathcal{H}$. Then by choosing the acceptance probability as

$$\alpha(u,v) = \min\left\{1, \frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}(u,v)\right\}, \qquad u,v\in\mathcal{H}, \tag{5.8}$$

the resulting Metropolis kernel M employing P and α is μ-reversible: let $A,B\in\mathcal{B}(\mathcal{H})$ with $A\cap B=\emptyset$ for simplicity, then

$$\begin{aligned}
\int_A\int_B M(u,\mathrm{d}v)\,\mu(\mathrm{d}u) &= \int_A\int_B \alpha(u,v)P(u,\mathrm{d}v)\,\mu(\mathrm{d}u)\\
&= \int_A\int_B \min\left\{1, \frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}(u,v)\right\}\,\eta(\mathrm{d}u,\mathrm{d}v)\\
&= \int_A\int_B \min\left\{\frac{\mathrm{d}\eta}{\mathrm{d}\eta^\top}(u,v), 1\right\}\,\eta^\top(\mathrm{d}u,\mathrm{d}v)\\
&= \int_B\int_A \min\left\{\frac{\mathrm{d}\eta}{\mathrm{d}\eta^\top}(v,u), 1\right\}\,\eta(\mathrm{d}u,\mathrm{d}v)\\
&= \int_B\int_A \min\left\{\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}(u,v), 1\right\}\,\eta(\mathrm{d}u,\mathrm{d}v)\\
&= \int_B\int_A M(u,\mathrm{d}v)\,\mu(\mathrm{d}u),
\end{aligned}$$

where the last line follows from $\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}(u,v) = \frac{\mathrm{d}\eta}{\mathrm{d}\eta^\top}(v,u)$ for $u,v\in\mathcal{H}$.

Remark 5.14. The choice of α given in (5.8) is not the only one which yields μ-reversibility, see, e.g., Peskun [131] and the references therein for further examples. However, (5.8) is the optimal admissible choice in the sense that it leads to the smallest asymptotic variance σ_f^2, $f\in L^2_\mu(\mathcal{H})$, for the resulting Markov chain, see Tierney [171].

In finite dimensional state spaces the equivalence of η and $\eta^\top$ can often be verified by considering the densities of η and $\eta^\top$ w.r.t. Lebesgue measure. However, in

infinite dimensional separable Hilbert spaces there exists no Lebesgue measure and equivalence of measures becomes a more delicate issue. As pointed out by Beskos et al. [15] a possible way to ensure the existence of $\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}$ is to choose a proposal kernel P which is μ_0-reversible, i.e.,

$$\eta_0(\mathrm{d}u, \mathrm{d}v) := P(u, \mathrm{d}v)\, \mu_0(\mathrm{d}u) \;=\; P(v, \mathrm{d}u)\, \mu_0(\mathrm{d}v) =: \eta_0^\top(\mathrm{d}u, \mathrm{d}v). \tag{5.9}$$

Then, due to the fact that $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}$ and $\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu}$ exist, see (5.1), it follows that

$$\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}(u,v) = \frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(v) \underbrace{\frac{\mathrm{d}\eta_0^\top}{\mathrm{d}\eta_0}(u,v)}_{\equiv 1} \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu}(u) = \exp(\Phi(u) - \Phi(v))$$

and, hence,

$$\alpha(u,v) = \min\left\{1, \exp(\Phi(u) - \Phi(v))\right\}. \tag{5.10}$$

Remark 5.15. In fact, μ_0-reversibility of the proposal kernel P is not necessary for $\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}$ to exist. It is sufficient that P be reversible w.r.t. a measure ν which is equivalent to μ, i.e., $\mu(\mathrm{d}u) = \rho(u)\, \nu(\mathrm{d}u)$ with a positive density $\rho\colon \mathcal{H} \to (0, \infty)$. Then

$$P(u, \mathrm{d}v)\, \nu(\mathrm{d}u) = P(v, \mathrm{d}u)\, \nu(\mathrm{d}v)$$

analogously implies

$$\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}(u,v) = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(v)\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(u) = \frac{\rho(v)}{\rho(u)}$$

and choosing $\alpha(u,v) = \min\left\{1, \frac{\rho(v)}{\rho(u)}\right\}$ will yield again a μ-reversible Metropolis kernel.

In the following paragraph we will introduce a common μ_0-reversible proposal, the *preconditioned Crank-Nicolson* (pCN) proposal, but beforehand we make a short remark on history and notation:

Remark 5.16 (On Metropolis and Metropolis-Hastings algorithms)**.** In the original Metropolis algorithm as stated by Metropolis et al. [119] the proposal kernel was chosen in a symmetric way, e.g., if $\mathcal{H} = \mathbb{R}^N$ and $P(u)$ has a Lebesgue density $p(u, \cdot)$, then $p(u,v) = p(v,u)$ for each $u, v \in \mathbb{R}^N$. This yields that the acceptance probability α as in (5.8) only involves ratios of the density of μ. Hastings [84] generalized the Metropolis algorithm to allow also for nonsymmetric proposals. Therefore, the term $\frac{\mathrm{d}\eta^\top}{\mathrm{d}\eta}$ in (5.8) is sometimes called *Hastings ratio*. Moreover, MH algorithms with

an acceptance probability $\alpha(\cdot,\cdot)$ which only involves ratios of densities of μ (e.g., w.r.t. μ_0) are simply called *Metropolis algorithms*.

The preconditioned Crank-Nicolson Metropolis algorithm First, we describe a general approach to construct proposal kernels based on discretization schemes for stochastic differential equations (SDE) which leads, in particular, to the construction of the preconditioned Crank-Nicolson proposal.

Assume a given SDE of *Langevin type*

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{5.11}$$

where W_t denotes a Q-Brownian motion in $\mathcal{H}$, i.e., a Brownian motion with increments $W_t - W_s \sim N(0, |t-s|Q)$ and $Q \in \mathcal{L}^1_+(\mathcal{H})$, and $f\colon \mathcal{H} \to \mathcal{H}$ a measurable mapping. If we apply appropriate time-stepping schemes, e.g., linear one-step schemes, to (5.11), then the resulting time-discrete solution is again a Markov chain. The transition kernel of this Markov chain can then be employed within an MH algorithm, e.g., as the proposal kernel. The idea behind this SDE-based approach is, that by suitable choices of f and σ in (5.11) one can specify the limit distribution of the solution process X_t which, in turn, might be inherited by the discretized solution. For example, by setting $f(x) = -(u + C\nabla\Phi(u))$, $\sigma = \sqrt{2}$ and $Q = C$ the limit distribution of X_t is μ as given in (5.1) provided some assumptions on Φ and $\nabla\Phi$ hold, see Hairer et al. [82, Theorem 3.6]. By applying the forward Euler scheme to the resulting SDE we obtain a Markov chain and may hope that this Markov chain also has μ as its invariant measure. This would allow to omit the accept/reject step in the MH algorithm and, thus, yield a less correlated chain. However, as it turns this is not the case, i.e., the resulting Markov chain does not possess μ as an invariant measure and still requires a *Metropolization*, i.e., the accept/reject procedure as described earlier, see, e.g., Roberts and Tweedie [142] for details. The resulting algorithm is also known as *Metropolis adjusted Langevin algorithm (MALA)*.

Now, we recall that our previous considerations for well-defined Metropolis-Hastings algorithms in infinite dimensions called for a μ_0-reversible proposal kernel. Hence, if we can construct an SDE with μ_0 as the limit distribution of its solution, then perhaps we can obtain μ_0-reversible proposal kernels by applying suitable time-discretization scheme to the corresponding SDE. Such an SDE is given by an *Ornstein-Uhlenbeck process* X_t with equilibirum state 0:

$$\mathrm{d}X_t = (0 - X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t$$

where W_t is again a C-Brownian motion, C denoting the covariance operator of $\mu_0 = N(0, C)$, see Hairer et al. [82] for more details. Moreover, as it turns out, see Cotter et al. [35, Theorem 6.2], the only linear one-step scheme which leads to a μ_0-reversible Markov chain is given by applying the Crank-Nicolson scheme to the drift term combined with an Euler-Maruyama step for the diffusion:

$$X_{n+1} = X_n - h\frac{X_n + X_{n+1}}{2} + \sqrt{2h}\xi_n \tag{5.12}$$

where $h > 0$ is the time stepsize and the $\xi_n \sim N(0, C)$ are i.i.d. $\mathcal{H}$-valued random variables. The transition kernel of the resulting Markov chain $(X_n)_{n\in\mathbb{N}}$ is given by

$$P_0(u, \cdot) = N(\sqrt{1-s^2}u, s^2C), \tag{5.13}$$

where $s \in [0, 1]$ relates to h in (5.12) by $s = \frac{\sqrt{8h}}{2+h}$. It is straightforward to verify that P_0 is μ_0-reversible: we have to show that for $\eta_0(\mathrm{d}u, \mathrm{d}v) = P_0(u, \mathrm{d}v)\,\mu_0(\mathrm{d}u)$ there holds $\eta_0 = \eta_0^\top$. We prove this by constructing a random vector $(U, V)^\top \sim \eta_0$ and then verifying that $(V, U)^\top \sim \eta_0$. Let $U \sim \mu_0$ and $W \sim \mu_0$ independently, then

$$\begin{pmatrix} U \\ V \end{pmatrix} := \begin{pmatrix} U \\ \sqrt{1-s^2}U + sW \end{pmatrix} = \begin{pmatrix} I & 0 \\ \sqrt{1-s^2}I & sI \end{pmatrix} \begin{pmatrix} U \\ W \end{pmatrix} \sim \eta_0,$$

and by applying Proposition 2.20 we obtain particularly

$$\eta_0(\mathrm{d}u, \mathrm{d}v) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} C & \sqrt{1-s^2}C \\ \sqrt{1-s^2}C & C \end{bmatrix}\right)$$

as well as

$$\begin{pmatrix} V \\ U \end{pmatrix} = \begin{pmatrix} \sqrt{1-s^2}I & sI \\ I & 0 \end{pmatrix} \begin{pmatrix} U \\ W \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} C & \sqrt{1-s^2}C \\ \sqrt{1-s^2}C & C \end{bmatrix}\right),$$

i.e., $\eta_0 = \eta_0^\top$.

Definition 5.17 (pCN proposal, pCN Metropolis)**.** Let $\mu_0 = N(0, C)$ and μ be given as in (5.1). The Markov kernel P_0 in (5.13) is called *preconditioned Crank Nicolson (pCN) proposal kernel*. The Metropolis algorithm based on the proposal P_0 and the acceptance probability (5.10) is called *pCN Metropolis algorithm* or simply *pCN Metropolis* and its Metropolis kernel will be denoted by M_0.

Remark 5.18. As mentioned by Cotter et al. [35] the pCN proposal kernel (5.13) was already suggested and employed by Neal [123, Equation (15)] for MCMC sam-

pling in finite dimensions with Gaussian priors. Although Neal noticed the prior reversibility of the proposal (5.13), he did not comment on the well-definedness of the resulting Metropolis algorithm in infinite dimensions nor did he provide a motivation or a derivation of the proposal. However, he remarked that the proposal (5.13) "is a bit faster and seems to work somewhat better" than the basic random walk proposal $P(u) = N(u, s^2C)$.

5.2. A Metropolis Algorithm with Generalized pCN Proposal

In recent years many authors have proposed and pursued the idea to construct proposals which try to exploit geometrical features of the target measure, see for example [73, 118, 104, 37]. In the following we will propose a generalized pCN (gpCN) proposal which allows to incorporate approximations to the covariance operator of the target measure μ. First, we prove a brief motivation before establishing the well-definedness of the resulting gpCN Metropolis algorithm in the Hilbert space $\mathcal{H}$.

5.2.1. Motivation from Bayesian Inference

As mentioned earlier there are numerous hints in the literature dating back at least to Tierney [170] which suggest that in case of a (Gaussian) random walk proposal $P(u) = N(u, Q)$ it is beneficial to choose the covariance operator Q as a scaled version of the covariance operator C_μ of the target measure μ. Although a rigourous proof of $Q \propto C_\mu$ being always optimal, e.g., in terms of the resulting asymptotic variance, is missing, numerical experiments, e.g., [118, 73, 104], and theoretical results from scaling theory, see Roberts and Rosenthal [139], strengthen this conjecture.

We provide a simple, graphical motivation why the choice $Q \propto C_\mu$ can be beneficial. Assume that the target measure μ is a bivariate Gaussian measure and that we employ two Gaussian random walk proposals for MH algorithms: one with covariance $Q_1 = s_1^2 I$ and the other with covariance $Q_2 = s_2^2 C_\mu$, where the parameters s_1, s_2 are chosen such that $\mathrm{tr}\,(Q_1) = \mathrm{tr}\,(Q_2)$[1], i.e., the total variance of both proposals is the same. Then, as indicated in Figure 5.1, the latter proposal will yield a higher acceptance rate, since the intersection of its level sets with regions of certain or high acceptance are larger than for the choice Q_1. This, in turn, leads to a less correlated

[1] By $\mathrm{tr}\,(Q)$ we denote the trace of an operator Q, see Appendix A for more details.

Markov chain and a faster exploration of the state space.

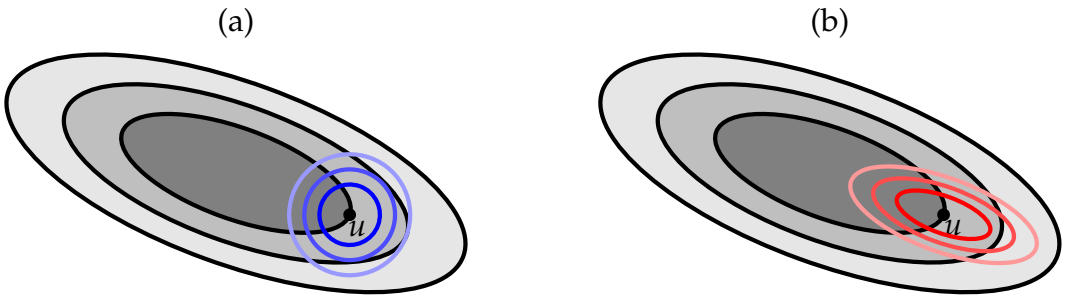

Figure 5.1.: For a Gaussian target measure $\mu = N(m_\mu, C_\mu)$ and current state u the region of acceptance $\{v : \alpha(u,v) = 1\}$ (dark grey region) as well as two regions of possible rejection $\{v : \underline{p} \leq \alpha(u,v) < \overline{p} \leq 1\}$ (lighter grey regions) are displayed. Moreover, we present the contour lines (blue and red, resp.) of Gaussian proposals $N(u, s^2Q)$ with covariance $Q = I$ in part (a) and target covariance $Q = C_\mu$ in part (b).

Based on this guideline, many authors have developed different strategies to obtain estimates on C_μ and to employ these within MH algorithms. For example the *adaptive MH algorithm* by Haario et al. [78] computes empirical estimates of C_μ given the history of the Markov chain and employs this for proposing new states.

We follow rather the approach of Martin et al. [118] to approximate C_μ in the case where μ results from Bayesian inference by linearization of the corresponding forward map. We briefly recall the setting of Bayesian inference which was descibed in more detail in Chapter 3: let U be a random variable in $\mathcal{H}$ with distribution $\mu_0 = N(0, C)$ and let Y be a random variable on $\mathbb{R}^d$ given by

$$Y = G(U) + \varepsilon \tag{5.14}$$

with a continuous map $G\colon \mathcal{H} \to \mathbb{R}^d$ and $\varepsilon \sim N(0, \Sigma)$, independent of U, with $\Sigma \in \mathbb{R}^{d\times d}$. Then, given some observation $y \in \mathbb{R}^d$ of Y we want to infer U, i.e., we are interested in the conditional distribution of U given the event $Y = y$ which admits a representation of the form (5.1) with

$$\Phi(u) = \frac{1}{2}|y - G(u)|^2_{\Sigma^{-1}}, \qquad u \in \mathcal{H}. \tag{5.15}$$

A special situation results if $G(u) = Lu + b$ with $L \in \mathcal{L}(\mathcal{H}; \mathbb{R}^d)$ and $b \in \mathbb{R}^d$. Then, it is known, see Mandelbaum [117] or Theorem 4.3, that $\mu = N(m_\mu, C_\mu)$ with

$$m_\mu = CL^*(LCL^* + \Sigma)^{-1}(y - b), \qquad C_\mu = (C^{-1} + L^*\Sigma^{-1}L)^{-1}, \tag{5.16}$$

where L^* denotes the adjoint operator of L.

The affine case indicates how we can construct good Gaussian proposal kernels if the map G is nonlinear but smooth. For a fixed $u_0 \in \mathcal{H}$ local linearization leads to

$$G(u) = G(u_0) + \nabla G(u_0)\,(u - u_0) + r(u)$$

with a remainder term $r(u) \in \mathbb{R}^d$. For a sufficiently smooth G the remainder r is small (in a neighborhood of u_0), so that

$$\widetilde{G}(u) = G(u_0) + \nabla G(u_0)\,(u - u_0)$$

is close to $G(u)$ (in a neighborhood of u_0). The substitution of G by $\widetilde{G}$ in the model (5.14) leads to a Gaussian target measure $\widetilde{\mu} = N(\widetilde{m}, \widetilde{C})$ with covariance

$$\widetilde{C} = (C^{-1} + L^*\Sigma^{-1}L)^{-1}, \qquad L = \nabla G(u_0).$$

By the fact that G and $\widetilde{G}$ are close, we also have that the measures μ and $\widetilde{\mu}$ are close as well, cf. Theorem 3.21. Then, it is reasonable to use the covariance operator $\widetilde{C}$ for proposing new states in a Metropolis algorithm. Of course, there might be other choices besides a simple linearization of G at one point. For example, averaging linearizations at several points $u_1, \dots, u_n \in \mathcal{H}$ leads to

$$\widetilde{C} = \Big(C^{-1} + \frac{1}{N}\sum_{n=1}^{N} L_n^*\,\Sigma^{-1}L_n\Big)^{-1}, \qquad L_n = \nabla G(u_n).$$

Natural candidates for the points $u_1, \dots, u_N$ are samples drawn according to the prior or samples taken from a short run of a preliminary Markov chain with the posterior as stationary measure, cf. the adaptive method in Cui et al. [37, Section 3.4]. In view of the suggested approximations $\widetilde{C}$ of C_μ above we will consider in the following proposals which use covariances of the form $C_\Gamma = (C^{-1} + \Gamma)^{-1}$ for suitably chosen operators Γ.

5.2.2. Well-Defined gpCN Proposals

In this section we introduce the gpCN proposal kernel and prove that the Metropolis algorithm with this proposal is well-defined in the sense that it leads to a μ-reversible transition kernel. Let $\mathcal{L}_+(\mathcal{H})$ denote the set of all bounded, self-adjoint and positive linear operators $\Gamma\colon \mathcal{H} \to \mathcal{H}$. Then, we define the operators

$$C_\Gamma := (C^{-1} + \Gamma)^{-1}, \qquad \Gamma \in \mathcal{L}_+(\mathcal{H}), \tag{5.17}$$

motivated in Section 5.2.1, where C denotes the covariance operator of the prior measure $\mu_0 = N(0, C)$, for which we also use the equivalent representation

$$C_\Gamma = C^{1/2}\,(I + H_\Gamma)^{-1}\,C^{1/2}, \qquad H_\Gamma := C^{1/2}\Gamma C^{1/2}. \tag{5.18}$$

Next, we show that the operators C_Γ are again covariance operators and can, therefore, be used for constructing Gaussian proposal kernels.

Proposition 5.19. Let C be a nonsingular covariance operator on $\mathcal{H}$, $\Gamma \in \mathcal{L}_+(\mathcal{H})$, and C_Γ and H_Γ given as in (5.18). Then $H_\Gamma \in \mathcal{L}_+(\mathcal{H})$ is trace class and C_Γ is also a nonsingular covariance operator on $\mathcal{H}$.

Proof. We have $H_\Gamma \in \mathcal{L}_+(\mathcal{H})$ by construction. Moreover, we note that $C^{1/2}$ is a Hilbert-Schmidt operator, because C itself is trace class, see Theorem A.7. Hence, H_Γ is a composition of two Hilbert-Schmidt operators ($C^{1/2}$) and one bounded operator (Γ) and, therefore, by virtue of Theorem A.7 it is trace class.

We now show the second assertion, i.e., that C_Γ is nonsingular, selfadjoint, positive and trace class: since H_Γ is selfadjoint and compact, we have from Fredholm operator theory that the operator $I + H_\Gamma$ is invertible if and only if $\ker H_\Gamma = \{0\}$. The latter is the case since H_Γ is positive which implies $\langle (I + H_\Gamma)u, u\rangle_{\mathcal{H}} \geq \langle u, u\rangle_{\mathcal{H}}$. Hence, the inverse $(I + H_\Gamma)^{-1}$ exists and, moreover, $(I + H_\Gamma)^{-1} \in \mathcal{L}_+(\mathcal{H})$ with $\|(I + H_\Gamma)^{-1}\| \leq 1$. The self-adjointness and positivity of C_Γ follows immediately and since C_Γ is a composition of two nonsingular Hilbert-Schmidt operators and a nonsingular bounded operator, $C^{1/2}$ and $(I + H_\Gamma)^{-1}$, respectively, it is trace class and nonsingular as well. □

Proposition 5.19 allows us to define and consider proposal kernels of the following form:

$$P(u, \cdot) = N(Au, s^2 C_\Gamma), \qquad s \in [0, 1),\ \Gamma \in \mathcal{L}_+(\mathcal{H}), \tag{5.19}$$

with $A \in \mathcal{L}(\mathcal{H})$. We would like to choose A such that P is μ_0-reversible, which implies that a Metropolis kernel with proposal P is μ-reversible, see Section 5.1.2. By applying Proposition 2.20 to

$$\begin{pmatrix} U \\ V \end{pmatrix} := \begin{pmatrix} U \\ AU + sW \end{pmatrix} = \begin{pmatrix} I & 0 \\ A & sI \end{pmatrix} \begin{pmatrix} U \\ W \end{pmatrix}, \qquad (U, W) \sim \mu_0 \otimes N(0, C_\Gamma),$$

and $(V, U)^\top$, we obtain in this setting

$$P(u, \mathrm{d}v)\,\mu_0(\mathrm{d}u) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} C & CA^* \\ AC & ACA^* + s^2 C_\Gamma \end{bmatrix}\right)$$

and

$$P(v, \mathrm{d}u)\, \mu_0(\mathrm{d}v) = N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}ACA^* + s^2 C_\Gamma & AC\\ CA^* & C\end{bmatrix}\right).$$

Thus, for satisfying (5.9) we need to choose A so that

$$AC = CA^*, \qquad ACA^* + s^2 C_\Gamma = C. \tag{5.20}$$

By straightforward calculation we obtain as the formal solution to (5.20)

$$A = A_\Gamma = C^{1/2}\ \sqrt{I - s^2\,(I + H_\Gamma)^{-1}}\, C^{-1/2}. \tag{5.21}$$

The following lemma ensures that this choice of A yields a well-defined bounded linear operator on $\mathcal{H}$.

Lemma 5.20. Let the assumptions of Proposition 5.19 be satisfied and let $s \in [0, 1)$. Then (5.21) defines a bounded linear operator $A_\Gamma \colon \operatorname{rg}(C^{1/2}) \to \mathcal{H}$.

Proof. From the proof of Proposition 5.19 we know that $(I + H_\Gamma)^{-1} \colon \mathcal{H} \to \mathcal{H}$ is self-adjoint and that $\|(I + H_\Gamma)^{-1}\| \leq 1$. Thus, $I - s^2(I + H_\Gamma)^{-1}$ is also a self-adjoint, bounded and positive operator on $\mathcal{H}$ and its square root operator appearing in (5.21) exists. This yields the well-definedness of $A_\Gamma \colon \operatorname{rg}(C^{1/2}) \to \mathcal{H}$. We now prove that A_Γ is a bounded operator on $\operatorname{rg}(C^{1/2})$. For $s = 0$ we get $A_\Gamma = I$ and the assertion follows, so that we assume $s \in (0, 1)$. Let us now define $f \colon \mathbb{C} \setminus \{-1\} \to \mathbb{C}$ by

$$f(z) = \sqrt{1 - s^2(1 + z)^{-1}}.$$

The function f is analytic in the complex half plane $\{z \in \mathbb{C} : \Re(z) > s^2 - 1\}$, since $\Re(1 + z) > s^2$ implies

$$\Re\left((1 + z)^{-1}\right) = \frac{\Re(1 + z)}{|1 + z|^2} \leq \frac{1}{\Re(1 + z)} < \frac{1}{s^2}.$$

Denoting $\gamma := \|H_\Gamma\| \in \mathbb{R}$ the spectrum of $H_\Gamma = C^{1/2}\Gamma C^{1/2}$ is contained in $[0, \gamma]$. Then, since $s < 1$ we have that f is analytic in a neighborhood, say, $\mathcal{N}[0, \gamma]$ of $[0, \gamma]$. Hence, by the holomorphic functional calculus, see Dunford and Schwartz [50, Section VII.3], we obtain

$$\sqrt{I - s^2\,(I + H_\Gamma)^{-1}} = f(H_\Gamma) = \frac{1}{2\pi \mathrm{i}} \int_{\partial \mathcal{N}[0,\gamma]} f(\zeta)\,(\zeta I - H_\Gamma)^{-1}\, \mathrm{d}\zeta.$$

Due to analyticity we can approximate f by a sequence of polynomials p_n with degree n which converge uniformly on $\mathcal{N}[0, \gamma]$ to f for $n \to \infty$. Then, by [50,

Lemma VII.3.13] holds

$$\|p_n(H_\Gamma) - f(H_\Gamma)\| \to 0,$$

for $n \to \infty$. Since the polynomials p_n can be represented as $p_n(z) = \sum_{k=0}^{n} a_k^{(n)} z^k$, we obtain further

$$C^{1/2}\, p_n(H_\Gamma) = C^{1/2} \sum_{k=0}^{n} a_k^{(n)}\, (C^{1/2}\Gamma C^{1/2})^k = p_n(C\Gamma)\, C^{1/2}.$$

By a result of Hladnik and Omladič [87, Proposition 1] we have

$$\operatorname{spec}(C\Gamma \mid \mathcal{H}) = \operatorname{spec}(C^{1/2}\Gamma C^{1/2} \mid \mathcal{H}) \subseteq [0, \gamma]$$

where $\operatorname{spec}(\cdot \mid \mathcal{H})$ denotes the spectrum on $\mathcal{H}$, and, thus, we can conclude $\|p_n(C\Gamma) - f(C\Gamma)\| \to 0$ as $n \to \infty$ again by [50, Lemma VII.3.13]. Hence,

$$C^{1/2} f(H_\Gamma) = \lim_{n\to\infty} C^{1/2}\, p_n(H_\Gamma) = \lim_{n\to\infty} p_n(C\Gamma)\, C^{1/2} = f(C\Gamma) C^{1/2}$$

and

$$A_\Gamma = C^{1/2} f(H_\Gamma) C^{-1/2} = f(C\Gamma) C^{1/2} C^{-1/2} = f(C\Gamma)$$

where $f(C\Gamma)$ is by construction a bounded operator on $\mathcal{H}$. □

Remark 5.21. Although the well-definedness of $A_\Gamma \colon \operatorname{rg}(C^{1/2}) \to \mathcal{H}$ follows rather easily, its boundedness is not trivial. Namely, in general, there exist operators $B \in \mathcal{L}(\mathcal{H})$ such that $C^{1/2} B C^{-1/2}$ is unbounded on $\operatorname{rg}(C^{1/2})$. For instance, let (λ_n, e_n), $n \in \mathbb{N}$, denote the eigenpairs of C and assume $\lambda_n \propto n^{-p}$ for a $p > 1$, then the operator $B \colon \mathcal{H} \to \mathcal{H}$ given by

$$Be_n = \begin{cases} e_{\sqrt{n}}, & \text{if } \sqrt{n} \in \mathbb{N}, \\ 0, & \text{otherwise,} \end{cases}$$

is bounded with norm $\|B\| = 1$, but $\|C^{1/2} B C^{-1/2} e_{n^2}\|_{\mathcal{H}} = n^{p/2} \to \infty$ as $n \to \infty$.

Lemma 5.20 allows us now to extend A_Γ to $\mathcal{H}$ by continuation, because $\operatorname{rg}(C^{1/2})$ is a dense subspace of $\mathcal{H}$ if C is a nonsingular trace class operator. For simplicity we denote this continuous extension again by $A_\Gamma \colon \mathcal{H} \to \mathcal{H}$.

Definition 5.22 (gpCN proposal)**.** For $s \in [0, 1)$ and $\Gamma \in \mathcal{L}_+(\mathcal{H})$ the *generalized pCN proposal kernel* is given by

$$P_\Gamma(u, \cdot) := N(A_\Gamma u, s^2 C_\Gamma). \tag{5.22}$$

For the zero operator $\Gamma = 0$ we recover the pCN proposal. By Lemma 5.20 and the arguments given in Section 5.1.2 we obtain the following important result.

Corollary 5.23. Let $\mu_0 = N(0, C)$ and μ be given by (5.1). Let the assumptions of Lemma 5.20 be satisfied. Then, a gpCN proposal kernel P_Γ given by (5.22) and an acceptance probability as in (5.10) induce a μ-reversible Metropolis kernel denoted by M_Γ.

For simplicity we will sometimes call the Metropolis algorithm with transition kernel M_Γ just gpCN Metropolis. There are relations of the gpCN Metropolis to other recently developed Metropolis algorithms for general Hilbert spaces which also use more sophisticated choices for the proposal than the pCN proposal. The following two remarks address these relations.

Remark 5.24. The gpCN proposals form a subclass of the *operator weighted proposals* introduced by Law [104] and Cui et al. [37]. The particular form of the gpCN proposal allows us to derive properties such as boundedness of the "proposal mean operator" A_Γ and the convergence of the resulting Markov chain, see Section 5.4. These issues were left open in [37, 104].

Remark 5.25. Pinski et al. [133] compute a Gaussian measure $\mu_* = N(m_*, C_*)$ which comes closest to μ w.r.t. the Kullback-Leibler distance. The admissible class of Gaussian measures considered there is closely related to our parametrized proposal covariances C_Γ, although their class of Gaussian measures is slightly larger. The measure μ_* is then used to construct a proposal kernel $P_*(u, \cdot) = N(m_* + \sqrt{1 - s^2}(u - m_*), s^2 C_*)$ for Metropolis algorithms. Note that P_* is not μ_0-reversible but μ_*-reversible, since it is a pCN proposal given the prior μ_*. In order to obtain a μ-reversible Metropolis kernel the authors need to adapt the acceptance probability by including terms of $\frac{\mathrm{d}\mu_*}{\mathrm{d}\mu_0}$, cf. Remark 5.15. Thus, Pinski et al. [133] also use a different covariance operator than the prior covariance in a pCN proposal in order to increase the efficiency of the Metropolis algorithm. The difference to our approach is the way they ensure the μ-reversibility of the algorithm. They keep the mean of the original pCN proposal and modify the acceptance probability whereas we modify also the mean of the proposal to maintain its μ_0-reversibility and, therefore, can leave the acceptance probability unchanged.

5.3. Spectral Gaps and Geometric Ergodicity

In this section we provide a brief summary of the spectral gap approach for proving L^2_μ-geometric ergodicity of Markov chains and the concept of conductance. Based

on the latter we then develop a general comparison result for spectral gaps of Metropolis algorithms with equivalent proposals. Although this comparison result is of interest in its own right, our main motivation for it is to apply it for proving the L^2_μ-geometric ergodicity of the gpCN Metropolis algorithm which will be done in the next section. In particular, our strategy there is to relate the existence of a spectral gap for the gpCN to the existence of a spectral gap of the pCN Metropolis. Here it is worth mentioning that Hairer et al. [81] established sufficient conditions for the latter under additional regularity assumptions on the function Φ in (5.1). However, with our comparative approach we do not need to rely on those conditions and will benefit from any improvement of the results stated in [81].

5.3.1. Spectral Gaps of Markov Operators

We will explain the relation between the L^2_μ-geometric ergodicity of a Markov chain and spectral properties of the associated Markov operator.

Definition 5.26. Let K be a μ-reversible Markov kernel where μ denotes a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. Then the associated *Markov operator* $\mathrm{K}\colon L^2_\mu(\mathcal{H}) \to L^2_\mu(\mathcal{H})$ is given by

$$(\mathrm{K}f)(u) = \int_{\mathcal{H}} f(v)\, K(u, \mathrm{d}v) \qquad \forall f \in L^2_\mu(\mathcal{H})$$

where $L^2_\mu(\mathcal{H}) = L^2_\mu(\mathcal{H}; \mathbb{R})$ is the Hilbert space of measurable functions $f\colon \mathcal{H} \to \mathbb{R}$ which are square integrable w.r.t. μ equipped with the inner product

$$\langle f, g \rangle_{L^2_\mu} = \int_{\mathcal{H}} f(u) g(u)\, \mu(\mathrm{d}u).$$

By the μ-reversibility of K we have that $\mathrm{K}\colon L^2_\mu(\mathcal{H}) \to L^2_\mu(\mathcal{H})$ is a self-adjoint bounded linear operator with norm 1: there holds

$$\begin{aligned}\langle \mathrm{K}f, g \rangle_{L^2_\mu} &= \int_{\mathcal{H}} (\mathrm{K}f)(u)\, g(u)\, \mu(\mathrm{d}u) = \int_{\mathcal{H}} \int_{\mathcal{H}} f(v)\, g(u)\, K(u, \mathrm{d}v)\, \mu(\mathrm{d}u) \\ &= \int_{\mathcal{H}} \int_{\mathcal{H}} f(v)\, g(u)\, K(v, \mathrm{d}u)\, \mu(\mathrm{d}v) = \int_{\mathcal{H}} f(v)\, (\mathrm{K}g)(v)\, \mu(\mathrm{d}v) = \langle f, \mathrm{K}g \rangle_{L^2_\mu},\end{aligned}$$

and by Jensen's inequality and the invariance of μ w.r.t. K, i.e., $\mu = \mu K$,

$$\begin{aligned}\|\mathrm{K}f\|^2_{L^2_\mu} &= \int_{\mathcal{H}} \left(\int_{\mathcal{H}} f(v)\, K(u, \mathrm{d}v) \right)^2 \mu(\mathrm{d}u) \leq \int_{\mathcal{H}} \int_{\mathcal{H}} f^2(v)\, K(u, \mathrm{d}v)\, \mu(\mathrm{d}u) \\ &= \int_{\mathcal{H}} f^2(v)\, \mu(\mathrm{d}v) = \|f\|^2_{L^2_\mu}.\end{aligned}$$

We will now define the (L^2_μ-)spectral gap of a Markov operator. The term "gap"

relates, roughly said, to the distance between the largest eigenvalue of K — which is $\lambda = 1$, see below — and its second largest eigenvalue in absolute terms. In the following, we will use the notation as in Rudolf [145] and define the (L^2_μ-)spectral gap in terms of an operator norm. To this end, we introduce a linear subspace of $L^2_\mu(\mathcal{H})$ which excludes the nonzero constant functions, since they are the trivial eigenfunctions to the eigenvalue $\lambda = 1$: let $f \in L^2_\mu(\mathcal{H})$ with $f \equiv c \in \mathbb{R}$ then

$$\mathrm{K}f(u) = \int_{\mathcal{H}} f(v)\, K(u, \mathrm{d}v) = c \int_{\mathcal{H}} K(u, \mathrm{d}v) = c = f(u).$$

Definition 5.27. We set

$$L^2_{\mu,0}(\mathcal{H}) := \left\{ f \in L^2_\mu(\mathcal{H}) \;\middle|\; \int_{\mathcal{H}} f(u)\, \mu(\mathrm{d}u) = 0 \right\}$$

and define for any linear bounded operator $A\colon L^2_\mu(\mathcal{H}) \to L^2_\mu(\mathcal{H})$

$$\|A\|_\mu := \sup_{f \in L^2_{\mu,0}(\mathcal{H}),\, f \neq 0} \frac{\|Af\|_{L^2_\mu}}{\|f\|_{L^2_\mu}}.$$

The term $\|A\|_\mu$ is nothing else than the usual operator norm $\|A|_{L^2_{\mu,0}(\mathcal{H})}\|$ of the restriction of A to $L^2_{\mu,0}(\mathcal{H})$, but we find the notation $\|A\|_\mu$ more convenient.

We note that for a Markov operator K and $f \in L^2_{\mu,0}(\mathcal{H})$ holds $\mathrm{K}f \in L^2_{\mu,0}(\mathcal{H})$ since

$$\int_{\mathcal{H}} \mathrm{K}f(u)\, \mu(\mathrm{d}u) = \int_{\mathcal{H}} \int_{\mathcal{H}} f(v)\, K(u, \mathrm{d}v)\, \mu(\mathrm{d}u) = \int_{\mathcal{H}} f(v)\, \mu(\mathrm{d}u).$$

Thus, we can state

Definition 5.28 (L^2_μ-spectral gap)**.** For a Markov operator K given as in Definition 5.26 we define its *(L^2_μ-)spectral gap* by

$$\mathrm{gap}_\mu(\mathrm{K}) := 1 - \|\mathrm{K}\|_\mu . \tag{5.23}$$

The relation between the spectral gap of a Markov operator K and the convergence of the distribution of the states of the related Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K is given in the next result.

Theorem 5.29 ([139, Theorem 2.1])**.** Let K be a μ-reversible Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ with associated Markov operator $\mathrm{K}\colon L^2_\mu(\mathcal{H}) \to L^2_\mu(\mathcal{H})$. Then a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K is L^2_μ-geometrically ergodic iff

$$\mathrm{gap}_\mu(\mathrm{K}) > 0.$$

Moreover, we can express autocovariances $\mathrm{Cov}(f(X_1), f(X_{1+k}))$ for $f \in L^2_\mu(\mathcal{H};\mathbb{R})$ of Markov chains with μ-reversible transition kernel K starting at stationarity $X_1 \sim \mu$ by the associated Markov operator. The joint distribution of consecutive states (X_k, X_{k+1}) is then given by the measure $\eta(\mathrm{d}u, \mathrm{d}v) := K(u, \mathrm{d}v)\mu(\mathrm{d}u)$ and we obtain

$$\begin{aligned}\mathrm{Cov}(f(X_1), f(X_2)) &= \int_{\mathcal{H}}\int_{\mathcal{H}} (f(u) - \mathbb{E}_\mu[f])\,(f(v) - \mathbb{E}_\mu[f])\,K(u,\mathrm{d}v)\,\mu(\mathrm{d}u) \\ &= \int_{\mathcal{H}} (f(u) - \mathbb{E}_\mu[f]) \left(\int_{\mathcal{H}} (f(v) - \mathbb{E}_\mu[f])\,K(u,\mathrm{d}v) \right) \mu(\mathrm{d}u) \\ &= \langle \mathrm{K}(f - \mathbb{E}_\mu[f]), (f - \mathbb{E}_\mu[f]) \rangle_\mu.\end{aligned}$$

By recursion, we get

$$\mathrm{Cov}(f(X_1), f(X_{1+k})) = \langle \mathrm{K}^k(f - \mathbb{E}_\mu[f]), (f - \mathbb{E}_\mu[f]) \rangle_\mu, \qquad k \in \mathbb{N}, \tag{5.24}$$

and, thus,

$$\sigma_f = \mathrm{Var}_\mu(f) + 2\sum_{k=1}^{\infty} \langle \mathrm{K}^k(f - \mathbb{E}_\mu[f]), f - \mathbb{E}_\mu[f] \rangle_\mu. \tag{5.25}$$

Assuming now that $\mathrm{gap}_\mu(\mathrm{K}) > 0$ we can express and bound the asymptotic variance σ_f^2 of $f \in L^2_\mu(\mathcal{H};\mathbb{R})$ appearing in the Markov chain CLT in Theorem 5.7 by

$$\sigma_f^2 = \langle (I + \mathrm{K})(I - \mathrm{K})^{-1}(f - \mathbb{E}_\mu[f]), (f - \mathbb{E}_\mu[f]) \rangle_\mu \leq \frac{2\,\|f\|^2_{L^2_\mu}}{\mathrm{gap}_\mu(\mathrm{K})}, \tag{5.26}$$

cf. Kipnis and Varadhan [98] and Rudolf [145]. The latter establishes also a non-asymptotic bound for the mean square error $\mathbb{E}\left[\left|S_{n,n_0}(f) - \mathbb{E}_\mu[f]\right|^2\right]$,

$$\sup_{\|f\|_{L^4_\mu} \leq 1} \mathbb{E}\left[\left|S_{n,n_0}(f) - \mathbb{E}_\mu[f]\right|^2\right] \leq \frac{2}{n \cdot \mathrm{gap}_\mu(\mathrm{K})} + \frac{C_\nu\,\|\mathrm{K}\|_\mu^{n_0}}{n^2 \cdot \mathrm{gap}_\mu(\mathrm{K})^2}$$

where the constant $C_\nu \geq 0$ depends on the initial distribution ν. This emphasizes once more the importance of $\mathrm{gap}_\mu(\mathrm{K})$ in the study of Markov chains and the numerical analysis of MCMC methods.

5.3.2. Conductance and Spectral Gaps

A useful concept in analyzing spectral gaps is the conductance of a Markov kernel and Cheeger's inequality.

Definition 5.30. Let K be a μ-reversible Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. Then its

conductance (w.r.t. μ) is given by

$$\varphi(K) := \inf_{\mu(A)\in(0,1/2]} \frac{\int_A K(u, A^c)\mu(\mathrm{d}u)}{\mu(A)}.$$

The definition of $\varphi(K)$ maybe seems a bit cryptic at first glance, but $\varphi(K)$ basically provides a lower bound for the (conditional) probability that a Markov chain with transition kernel K switches between A and its complement within one step, since then with $X_n \sim \mu$ we have

$$\frac{\int_A K(u, A^c)\mu(\mathrm{d}u)}{\mu(A)} = \mathbb{P}(X_{n+1} \in A^c \mid X_n \in A), \qquad \forall A : \mu(A) \notin \{0, 1\}.$$

This interpretation might explain why $\varphi(K)$ is called conductance in the literature. Moreover, by μ-reversibility of K the infimum in Definition 5.30 remains the same when taken over all sets A with $\mu(A) \in (0, 1)$:

$$\begin{aligned}\int_A K(u, A^c)\mu(\mathrm{d}u) &= \int_A \int_{A^c} K(u, \mathrm{d}v)\mu(\mathrm{d}u) = \int_{A^c} \int_A K(u, \mathrm{d}v)\mu(\mathrm{d}u) \\ &= \int_{A^c} K(u, A)\mu(\mathrm{d}u).\end{aligned}$$

The conductance of Markov kernel can be used to obtain bounds on the second largest eigenvalue of the associated Markov operator:

Theorem 5.31 (Cheeger's inequality [107])**.** Let K be μ-reversible Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. Then there holds

$$\frac{\varphi(K)^2}{2} \leq 1 - \Lambda(\mathrm{K}) \leq 2\varphi(K) \tag{5.27}$$

where

$$\Lambda(\mathrm{K}) := \sup\{\lambda \colon \lambda \in \mathrm{spec}(\mathrm{K} \mid L^2_{\mu,0}(\mathcal{H}))\}.$$

denotes the spectral abscissa of K in $L^2_{\mu,0}(\mathcal{H})$.

Hence, provided the Markov operator is positive, Theorem 5.31 enables us to bound its spectral gap by the conductance of its Markov kernel.

5.3.3. Comparison of Conductance and Spectral Gaps

In the following we will derive results on comparing the conductance and later also the spectral gaps of two μ-reversible Metropolis kernels M_1 and M_2 which

differ only in the proposal — as it is the case for the pCN and the gpCN Metropolis algorithm.

The main idea here is that provided the proposals P_1 and P_2 of the two Metropolis algorithms admit a Radon-Nikodym derivative, we can use it in the definition of the conductance in order to relate $\varphi(M_1)$ and $\varphi_2(M_2)$. For example, assume that the Radon-Nikodym derivative $\frac{\mathrm{d}P_1(u)}{\mathrm{d}P_2(u)}(v)$ exists for any $u \in \mathcal{H}$ and is uniformly (w.r.t. u and v) bounded by $0 < c < \infty$, then due to

$$\begin{aligned}\int_A M_1(u, A^c)\mu(\mathrm{d}u) &= \int_A \int_{A^c} \alpha(u,v)\, P_1(u,\mathrm{d}v)\mu(\mathrm{d}u) \\ &= \int_A \int_{A^c} \alpha(u,v)\, \frac{\mathrm{d}P_1(u)}{\mathrm{d}P_2(u)}(v)\, P_2(u,\mathrm{d}v)\mu(\mathrm{d}u) \\ &\leq c \int_A \int_{A^c} \alpha(u,v)\, P_2(u,\mathrm{d}v)\mu(\mathrm{d}u) = c \int_A M_2(u, A^c)\mu(\mathrm{d}u),\end{aligned}$$

for any $A \in \mathcal{B}(\mathcal{H})$ we get $\varphi(M_1) \leq c\varphi(M_2)$. The following result extends this approach to the case where $\frac{\mathrm{d}P_1(u)}{\mathrm{d}P_2(u)}$ is not necessarily bounded but satisfies certain integrability conditions.

Lemma 5.32. Let μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and for $i = 1, 2$ let

$$M_i(u, \mathrm{d}v) = \alpha(u,v)P_i(u,\mathrm{d}v) + \delta_u(\mathrm{d}v) \int_{\mathcal{H}} (1 - \alpha(u,w))\, P_i(u,\mathrm{d}w)$$

be μ-reversible Metropolis kernels. Assume that for any $u \in \mathcal{H}$ the Radon-Nikodym derivative of $P_1(u,\mathrm{d}v)$ w.r.t. $P_2(u,\mathrm{d}v)$ exists, i.e., the proposal kernels admit a density

$$\rho(u,v) = \frac{\mathrm{d}P_1(u)}{\mathrm{d}P_2(u)}(v), \qquad u, v \in \mathcal{H}.$$

If for a number $p > 1$ we have

$$\kappa_p := \sup_{\mu(A)\in(0,1/2]} \frac{\int_A \int_{A^c} \rho(u,v)^p P_2(u,\mathrm{d}v)\, \mu(\mathrm{d}u)}{\mu(A)} < \infty, \tag{5.28}$$

then

$$\varphi(M_1) \leq \kappa_p^{1/p}\, \varphi(M_2)^{(p-1)/p}.$$

Proof. Let $A \in \mathcal{B}(\mathcal{H})$ with $\mu(A) \in (0, 1/2]$. Further, let $q = p/(p-1)$ such that $1/q + 1/p = 1$. Then

$$\begin{aligned}\int_A M_1(u, A^c)\, \mu(\mathrm{d}u) &= \int_{\mathcal{H}} \int_{\mathcal{H}} \mathbf{1}_{A^c}(v)\, \mathbf{1}_A(u)\, \alpha(u,v)\, P_1(u,\mathrm{d}v)\, \mu(\mathrm{d}u) \\ &= \int_{\mathcal{H}} \int_{\mathcal{H}} \mathbf{1}_{A^c}(v)\, \mathbf{1}_A(u)\, \alpha(u,v)\, \rho(u;v)\, P_2(u,\mathrm{d}v)\, \mu(\mathrm{d}u).\end{aligned}$$

Note that $P_2(u, \mathrm{d}v)\mu(\mathrm{d}u)$ is a probability measure on $(\mathcal{H} \times \mathcal{H}, \mathcal{B}(\mathcal{H} \times \mathcal{H}))$ and we can apply Hölder's inequality according to this measure with parameters p and q. Thus, by using $\alpha(u, v) = \alpha(u, v)^{1/q}\alpha(u, v)^{1/p}$ we obtain

$$\begin{aligned}
&\int_A M_1(u, A^c)\, \mu(\mathrm{d}u) \\
&\leq \left(\int_A M_2(u, A^c)\, \mu(\mathrm{d}u) \right)^{1/q} \left(\int_A \int_{A^c} \rho(u, v)^p \alpha(u, v)\, P_2(u, \mathrm{d}v)\, \mu(\mathrm{d}u) \right)^{1/p} \\
&\leq \left(\int_A M_2(u, A^c)\, \mu(\mathrm{d}u) \right)^{1/q} \left(\int_A \int_{A^c} \rho(u, v)^p P_2(u, \mathrm{d}v)\, \mu(\mathrm{d}u) \right)^{1/p}
\end{aligned}$$

Dividing by $\mu(A)$, applying $\mu(A)^{-1} = \mu(A)^{-1/q}\, \mu(A)^{-1/p}$ and taking the infimum yields

$$\varphi(M_1) \leq \varphi(M_2)^{1/q} \kappa_p^{1/p}.$$

□

We would like to mention that employing comparison inequalities in terms of the conductance is not an entirely new idea, see for example Lee and Łatuszyński [109, Proof of Theorem 4]. There the authors obtained a conductance inequality for transition kernels with mutually bounded Radon-Nikodym derivatives. An immediate consequence of Lemma 5.32 and (5.27) is the following theorem.

Theorem 5.33 (Spectral gap comparison)**.** Let the assumptions of Lemma 5.32 be satisfied and let the Markov operators associated with M_1 and M_2 be positive on $L^2_\mu(\mathcal{H})$. Then

$$\left(\frac{\mathrm{gap}_\mu(\mathrm{M}_1)}{2} \right)^p \leq \kappa_p \left(2\, \mathrm{gap}_\mu(\mathrm{M}_2) \right)^{(p-1)/2}.$$

Proof. The assertion follows by Theorem 5.31 and Lemma 5.32, since then

$$\frac{\mathrm{gap}_\mu(\mathrm{M}_1)}{2} \leq \varphi(M_1) \leq \varphi(M_2)^{(p-1)/p} \kappa_p^{1/p} \leq \left(2\mathrm{gap}_\mu(\mathrm{M}_2) \right)^{(p-1)/(2p)} \kappa_p^{1/p}.$$

□

Theorem 5.33 serves as a guideline to prove our convergence result for the gpCN Metropolis. In particular, in the next section we will investigate if the conditions of Theorem 5.33 are fulfilled for the pCN and the gpCN Metropolis kernel.

5.4. Geometric Ergodicity of the (Restricted) gpCN Metropolis Algorithm

We summarize the assumptions and statement of Theorem 5.33 with P_1 as the pCN Metropolis and P_2 as the gpCN Metropolis kernel: if

1. the Markov operators M_0 and M_Γ, $\Gamma \in \mathcal{L}_+(\mathcal{H})$, are positive,
2. the Radon-Nikodym derivative $\rho_\Gamma(u,v) := \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v)$ exists for each $u \in \mathcal{H}$,
3. and this derivative ρ_Γ satisfies for a $p > 1$

$$\sup_{\mu(A)\in(0,1/2]} \frac{\int_A \int_{A^c} \rho_\Gamma(u,v)^p \, P_\Gamma(u,\mathrm{d}v)\, \mu(\mathrm{d}u)}{\mu(A)} < \infty,$$

then $\mathrm{gap}_\mu(M_0) > 0$ implies $\mathrm{gap}_\mu(M_\Gamma) > 0$.

In the following two subsections we show that the first two assumptions are satisfied and also provide a bound for the integral $\int_{\mathcal{H}} \rho_\Gamma(u,v)^p \, P_\Gamma(u,\mathrm{d}v)$. Unfortunately, the latter will not enable us to verify the third assumption. We therefore introduce and study restrictions of measures and Metropolis kernels to balls in $\mathcal{H}$ to which we can then apply Theorem 5.33 and prove our convergence result for the restricted gpCN Metropolis algorithm.

5.4.1. Positivity of the gpCN Metropolis Kernel

In the following we will equivalently use the phrase that a Markov kernel is positive meaning that its associated Markov operator is positive. We will state positivity results for Metropolis algorithms with general Gaussian proposals at the beginning and verify afterwards the positivity of the gpCN Metropolis.

Lemma 5.34 (Positivity of proposal kernels)**.** Let $\mu_0 = N(0,C)$ be a Gaussian measure on a separable Hilbert space $\mathcal{H}$ and let $P(u,\cdot) = N(Au,Q)$ be a μ_0-reversible proposal kernel with $A \in \mathcal{L}(\mathcal{H})$. If there exists an operator $B \in \mathcal{L}(\mathcal{H})$ such that

$$B^2 = A, \qquad BC = CB^*,$$

and $D := C - BCB^*$ is positive and trace class, then, the Markov operator associated with the proposal P is positive on $L^2_{\mu_0}(\mathcal{H})$.

Proof. Because of the assumptions on B and D we obtain that the proposal kernel $P_1(u,\cdot) = N(Bu, D)$ is well-defined. Further, since $BCB^* + D = C$ we derive

$$P_1(u, \mathrm{d}v)\mu_0(\mathrm{d}u) = N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}C & CB^*\\ BC & C\end{bmatrix}\right),$$

which leads by $BC = CB^*$ to the μ_0-reversibility of P_1 and, thus, to the self-adjointness of its associated Markov operator in $L^2_{\mu_0}(\mathcal{H})$. It remains to prove that $\mathrm{P}_1^2 = \mathrm{P}$ holds for the associated Markov operators which then immediately yields the assertion. The equality of the Markov operators is equivalent to the equality of the measures $P_1^2(u,\cdot)$ and $P(u,\cdot)$ for all $u \in \mathcal{H}$, since by Fubini's theorem

$$\begin{aligned}\mathrm{P}_1^2 f(u) &= \int_{\mathcal{H}} \mathrm{P}_1 f(v)\, P_1(u, \mathrm{d}v) = \int_{\mathcal{H}}\int_{\mathcal{H}} f(w)\, P_1(v, \mathrm{d}w)\, P_1(u, \mathrm{d}v)\\ &= \int_{\mathcal{H}}\int_{\mathcal{H}} f(w)\, P_1(u, \mathrm{d}v)\, P_1(v, \mathrm{d}w) = \int_{\mathcal{H}} f(w) \int_{\mathcal{H}} P_1(u, \mathrm{d}v)\, P_1(v, \mathrm{d}w)\\ &= \int_{\mathcal{H}} f(w) P_1^2(u, \mathrm{d}w).\end{aligned}$$

In order to show that $P_1^2(u,\cdot) = P(u,\cdot)$ for all $u \in \mathcal{H}$, we take $(\xi_n)_{n\in\mathbb{N}}$ to be an i.i.d. sequence with $\xi_1 \sim N(0, D)$ and construct an auxiliary Markov chain by

$$X_{n+1} = BX_n + \xi_n, \quad n \geq 1,$$

where $X_1 = u$ for an arbitrary $u \in \mathcal{H}$. The transition kernel of the Markov chain $(X_n)_{n\in\mathbb{N}}$ is the kernel P_1. In particular, for $G \in \mathcal{B}(\mathcal{H})$ holds $\mathbb{P}(X_3 \in G) = P_1^2(u, G)$. By

$$X_3 = BX_2 + \xi_2 = B^2u + B\xi_1 + \xi_2$$

and $B\xi_1 + \xi_2 \sim N(0, BDB^* + D)$ we obtain $X_3 \sim N(B^2u, BDB^* + D)$. Due to the assumptions we have $B^2 = A$ and

$$BDB^* + D = B(C - BCB^*)B^* + C - BCB^* = C - ACA^*.$$

The last step $C - ACA^* = Q$ follows by the assumed μ_0-reversibility of P, because we know from Section 5.2.2 that P being μ_0-reversible is equivalent to A and Q satisfying $AC = CA^*$ and $ACA^* + Q = C$. We thus arrive at $X_3 \sim N(Au, Q)$ which proves $P_1^2(u,\cdot) = P(u,\cdot)$. □

The next lemma extends the previous result to Markov operators associated with Metropolis algorithms. The proof follows along the same line of arguments as developed by Rudolf and Ullrich [147, Section 3.4] and is therefore omitted.

Lemma 5.35 (Positivity of Metropolis kernels, cf. [147, Section 3.4])**.** Let μ be a measure on $\mathcal{H}$ given by (5.1) and let P be a μ_0-reversible proposal kernel whose associated Markov operator is positive on $L^2_{\mu_0}(\mathcal{H})$. Then the Markov operator associated with a μ-reversible Metropolis kernel

$$M(u, \mathrm{d}v) = \alpha(u, v)P(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \int_{\mathcal{H}} (1 - \alpha(u, w))P(u, \mathrm{d}w)$$

with $\alpha(u, v) = \min\{1, \frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(v)\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu}(u)\}$ is positive on $L^2_\mu(\mathcal{H})$.

The previous two lemmas lead to the following result about the gpCN Metropolis kernel.

Theorem 5.36 (Positivity of gpCN Metropolis kernel)**.** Let $\mu_0 = N(0, C)$ and μ as in (5.1) and let M_Γ denote the gpCN Metropolis kernel as in Corollary 5.23. Then the associated Markov operator M_Γ is self-adjoint and positive on $L^2_\mu(\mathcal{H})$.

Proof. It is enough to verify the assumptions of Lemma 5.34 for the gpCN proposal. Recall that $P_\Gamma(u, \cdot) = N(A_\Gamma u, s^2 C_\Gamma)$ which is μ_0-reversible by construction with bounded $A_\Gamma = C^{1/2}\sqrt{I - s^2(I + H_\Gamma)^{-1}}C^{-1/2}$. By choosing

$$B := C^{1/2}\sqrt[4]{I - s^2(I + H_\Gamma)^{-1}}C^{-1/2},$$

we obtain $B^2 = A_\Gamma$ and $BC = CB^*$. Moreover,

$$D = C - BCB^* = C^{1/2}(I - \sqrt{I - s^2(I + H_\Gamma)^{-1}})C^{1/2}.$$

The eigenvalues of $I - \sqrt{I - s^2(I + H_\Gamma)^{-1}}$ take the form $1 - \sqrt{1 - \frac{s^2}{1+\lambda}} \geq 0$ with $\lambda \geq 0$ being an eigenvalue of H_Γ. Thus, $I - \sqrt{I - s^2(I + H_\Gamma)^{-1}}$ is positive and bounded which yields D being positive and trace class since D is then a product of two Hilbert-Schmidt and one bounded operator. Thus, the conditions of Lemma 5.34 are satisfied and the assertion follows. □

5.4.2. The Density between the pCN and gpCN Proposal

We now show that for any state $u \in \mathcal{H}$ the gpCN proposal is equivalent to the pCN proposal in the sense of measures. Moreover, we will also derive an integrability result for the corresponding density. For proving the equivalence we need the following technical result the proof of which is similar to the proof of Lemma 5.20.

Lemma 5.37. Let the assumptions of Corollary 5.23 be satisfied and define the bounded, linear operator $\Delta_\Gamma \colon \mathcal{H} \to \mathcal{H}$ by

$$\Delta_\Gamma := A_0 - A_\Gamma = \sqrt{1-s^2} I - C^{1/2} \sqrt{I - s^2 (I + H_\Gamma)^{-1}} C^{-1/2}. \tag{5.29}$$

Then $\operatorname{rg}(\Delta_\Gamma) \subseteq \operatorname{rg}(C^{1/2})$, i.e., $C^{-1/2}\Delta_\Gamma$ is a bounded operator on $\mathcal{H}$.

Proof. By Douglas [49, Theorem 1] the relation $\operatorname{rg}(\Delta_\Gamma) \subseteq \operatorname{rg}(C^{1/2})$ holds iff there exists a bounded operator $B \colon \mathcal{H} \to \mathcal{H}$ such that $\Delta_\Gamma = C^{1/2} B$. Thus, $\operatorname{rg}(\Delta_\Gamma) \subseteq \operatorname{rg}(C^{1/2})$ is equivalent to $C^{-1/2}\Delta_\Gamma$ being bounded on $\mathcal{H}$. In order to construct and analyze the operator B, we define $f \colon \mathbb{C} \setminus \{-1\} \to \mathbb{C}$ by

$$f(z) := \sqrt{1 - s^2 (1+z)^{-1}} - \sqrt{1-s^2},$$

which is analytic in $\{z \in \mathbb{C} : \Re(z) > s^2 - 1\}$, cf. the proof of Lemma 5.20, and particularly in

$$V = \{z \in \mathbb{C} : \operatorname{dist}(z, [0, \gamma]) \leq \varepsilon\}, \qquad 0 < \varepsilon < 1 - s^2,$$

where $\gamma := \|H_\Gamma\|$ and $\operatorname{dist}(z, A) := \inf_{a \in A} |z - a|$ for any $A \subset \mathbb{C}$. We have the following representation

$$\begin{aligned} -\Delta_\Gamma = A_\Gamma - \sqrt{1-s^2} I &= C^{1/2} \left(\sqrt{I - s^2 (I + H_\Gamma)^{-1}} - \sqrt{1-s^2} I \right) C^{-1/2} \\ &= C^{1/2} f(H_\Gamma) C^{-1/2} \end{aligned}$$

with

$$f(H_\Gamma) = \frac{1}{2\pi \mathrm{i}} \int_{\partial V} f(\zeta) (\zeta I - H_\Gamma)^{-1} \, \mathrm{d}\zeta$$

see Dunford and Schwartz [50, Chapter VII.3]. Hence, if we can prove that $B = -f(H_\Gamma) C^{-1/2}$ is a bounded operator on $\mathcal{H}$, we have shown the assertion.

To this end let $p_n(z) = \sum_{k=0}^n a_k^{(n)} z^k$ be polynomials of degree n, with $n \in \mathbb{N}$, which converge uniformly on V to f. Such polynomials exist due to the analyticity of f and by the fact that $f(0) = 0$ we can assume w.l.o.g. that $a_0^{(n)} = 0$ for all $n \in \mathbb{N}$. This leads to

$$\begin{aligned} p_n(H_\Gamma) &= C^{1/2} \Gamma^{1/2} \left(\sum_{k=1}^n a_k^{(n)} (\Gamma^{1/2} C \Gamma^{1/2})^{k-1} \right) \Gamma^{1/2} C^{1/2} \\ &= C^{1/2} \Gamma^{1/2} \, q_{n-1}(\Gamma^{1/2} C \Gamma^{1/2}) \, \Gamma^{1/2} C^{1/2} \end{aligned}$$

with $q_{n-1}(z) := \sum_{k=1}^{n} a_k^{(n)} z^{k-1} = p_n(z)/z$. Now, a result by Hladnik and Omladič [87, Proposition 1] implies that the operators $C^{1/2}\Gamma C^{1/2}$ and $\Gamma^{1/2}C\Gamma^{1/2}$ share the same spectrum, since C and Γ are positive. Thus, $\operatorname{spec}(\Gamma^{1/2}C\Gamma^{1/2} \mid \mathcal{H}) \subset [0, \gamma]$ and we have

$$q_n(\Gamma^{1/2}C\Gamma^{1/2}) = \frac{1}{2\pi i} \int_{\partial V} q_n(\zeta)\, (\zeta I - \Gamma^{1/2}C\Gamma^{1/2})^{-1}\, \mathrm{d}\zeta, \qquad n \in \mathbb{N}.$$

Moreover, the polynomials q_n are a Cauchy sequence in $C(\partial V)$, since

$$\begin{aligned} \sup_{\zeta\in\partial V} |q_n(\zeta) - q_m(\zeta)| &\leq \sup_{\zeta\in\partial V} \frac{|\zeta|}{\min_{\eta\in\partial V} |\eta|} |q_n(\zeta) - q_m(\zeta)| \\ &= \frac{1}{\min_{\eta\in\partial V} |\eta|} \sup_{\zeta\in\partial V} |\zeta q_n(\zeta) - \zeta q_m(\zeta)| \\ &= \frac{1}{\min_{\eta\in\partial V} |\eta|} \sup_{\zeta\in\partial V} |p_{n+1}(\zeta) - p_{m+1}(\zeta))| \end{aligned}$$

where $\min_{\eta\in\partial V} |\eta| = \varepsilon > 0$ due to our choice of V. Thus, the polynomials q_n converge uniformly on ∂V to a function g. This implies that the operators $q_n(\Gamma^{1/2}C\Gamma^{1/2})$ converge in the operator norm to a bounded operator

$$g(\Gamma^{1/2}C\Gamma^{1/2}) := \frac{1}{2\pi i} \int_{\partial V} g(\zeta)\, (\zeta I - \Gamma^{1/2}C\Gamma^{1/2})^{-1}\, \mathrm{d}\zeta$$

cf. Dunford and Schwartz [50, Lemma VII.3.13]. We arrive at

$$\begin{aligned} f(H_\Gamma) &= \lim_{n\to\infty} p_n(C^{1/2}\Gamma C^{1/2}) = \lim_{n\to\infty} C^{1/2}\Gamma^{1/2}\, q_{n-1}(\Gamma^{1/2}C\Gamma^{1/2})\, \Gamma^{1/2}C^{1/2} \\ &= C^{1/2}\Gamma^{1/2}\, g(\Gamma^{1/2}C\Gamma^{1/2})\, \Gamma^{1/2}C^{1/2}, \end{aligned}$$

which yields

$$B = -f(H_\Gamma)C^{-1/2} = -C^{1/2}\Gamma^{1/2}\, g(\Gamma^{1/2}C\Gamma^{1/2})\Gamma^{1/2}$$

being bounded on $\mathcal{H}$. □

Lemma 5.37 ensures that we can apply the Cameron-Martin theorem, Theorem C.5 in Appendix C, in the proof of the following result. The other main tool for deriving the next theorem is a variant of the Feldman-Hajek theorem as stated in Theorem C.8 in Appendix C.

Theorem 5.38 (Density of pCN w.r.t. gpCN)**.** With the notation and assumptions of Corollary 5.23 there holds:

1. the measures $\mu_0 = N(0, C)$ and $\mu_\Gamma = N(0, C_\Gamma)$ are equivalent with

$$\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_\Gamma}(v) = \frac{\exp\left(\frac{1}{2}\langle \Gamma v, v\rangle_{\mathcal{H}}\right)}{\sqrt{\det(I + H_\Gamma)}} =: \pi_\Gamma(v), \tag{5.30}$$

2. For $u \in \mathcal{H}$ the measures $P_0(u, \cdot)$ and $P_\Gamma(u, \cdot)$ are equivalent with density

$$\frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v) = \pi_{\mathrm{CM}}\Big(\Delta_\Gamma u, \frac{1}{s}(v - A_\Gamma u)\Big)\, \pi_\Gamma\Big(\frac{1}{s}(v - A_\Gamma u)\Big) \tag{5.31}$$

where Δ_Γ as in (5.29) and

$$\pi_{\mathrm{CM}}(h, v) := \exp\left(-\frac{1}{2}\|C^{-1/2}h\|_{\mathcal{H}}^2 + \langle C^{-1}h, v\rangle_{\mathcal{H}}\right). \tag{5.32}$$

(The subscript in π_{CM} refers to the Cameron-Martin formula.)

Proof. We prove (5.30) by verifying the assumptions of Theorem C.8. We observe

$$I - C^{-1/2} C_\Gamma C^{-1/2} = I - (I + H_\Gamma)^{-1}$$

and set $T_\Gamma := I - (I + H_\Gamma)^{-1}$. The eigenvalues $(t_n)_{n\in\mathbb{N}}$ of the self-adjoint operator T_Γ are given by

$$t_n = 1 - \frac{1}{1 + \lambda_n} = \frac{\lambda_n}{1 + \lambda_n} < 1$$

where $(\lambda_n)_{n\in\mathbb{N}}$ are the eigenvalues of the positive trace class operator H_Γ. Thus, T_Γ is also trace class and satisfies $\langle T_\Gamma u, u\rangle_{\mathcal{H}} < \|u\|_{\mathcal{H}}^2$ for any $u \in \mathcal{H}$. Then, the assertion follows by Theorem C.8 and

$$T_\Gamma (I - T_\Gamma)^{-1} = \left(I - (I + H_\Gamma)^{-1}\right)(I + H_\Gamma) = H_\Gamma$$

as well as

$$\langle H_\Gamma\, C^{-1/2} v, C^{-1/2} v\rangle_{\mathcal{H}} = \langle \Gamma v, v\rangle_{\mathcal{H}} \qquad \forall v \in \mathcal{H}.$$

To show the equivalence of $P_0(u, \cdot)$ and $P_\Gamma(u, \cdot)$ for any $u \in \mathcal{H}$ we introduce the auxiliary kernel $K_\Gamma(u, \cdot) = N(A_\Gamma u, s^2 C)$. The first assertion and a simple change of variables, see Lemma C.9 in the appendix, lead to

$$\frac{\mathrm{d}K_\Gamma(u)}{\mathrm{d}P_\Gamma(u)}(v) = \pi_\Gamma\left(\frac{1}{s}\,(v - A_\Gamma u)\right), \qquad u, v \in \mathcal{H}.$$

Thus, it remains to prove the equivalence of $K_\Gamma(u, \cdot)$ and $P_0(u, \cdot)$ for any $u \in \mathcal{H}$. By

the Cameron-Martin formula, see Theorem C.5, this holds iff

$$\mathrm{rg}(A_\Gamma - \sqrt{1-s^2}I) \subseteq \mathrm{rg}(C^{1/2})$$

which was shown in Lemma 5.37. Now Theorem C.5 combined with a change of variables, see Lemma C.9, yields

$$\frac{\mathrm{d}P_0(u)}{\mathrm{d}K_\Gamma(u)}(v) = \pi_{\mathrm{CM}}\left((\sqrt{1-s^2}I - A_\Gamma)u, \frac{1}{s}(v - A_\Gamma u)\right)$$

and the assertion follows by

$$\frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v) = \frac{\mathrm{d}P_0(u)}{\mathrm{d}K_\Gamma(u)}(v)\frac{\mathrm{d}K_\Gamma(u)}{\mathrm{d}P_\Gamma(u)}(v). \qquad \square$$

Note, that Theorem 5.38 implies for any $\Gamma_1, \Gamma_2 \in \mathcal{L}_+(\mathcal{H})$ the existence of $\frac{\mathrm{d}P_{\Gamma_1}(u)}{\mathrm{d}P_{\Gamma_2}(u)}$.

Theorem 5.39 (Integrability of gpCN density)**.** Let the assumptions of Lemma 5.37 be satisfied and set

$$\rho_\Gamma(u,v) := \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v), \qquad u, v \in \mathcal{H}.$$

Then, for any $0 < p < 1 + \frac{1}{2\|H_\Gamma\|}$ there exist constants $c = c(p, H_\Gamma) < \infty$ and $b = b(p, \|C^{-1/2}\Delta_\Gamma\|) < \infty$ such that

$$\int_{\mathcal{H}} \rho_\Gamma^p(u,v)\, P_\Gamma(u, \mathrm{d}v) \le c\, \exp\left(\frac{b}{2}\|u\|_{\mathcal{H}}^2\right)$$

where $b \le 0$ for $0 < p \le \frac{1}{2}$ and $b > 0$ for $p > \frac{1}{2}$.

Proof. We employ the same notation as in Theorem 5.38, i.e., let $\mu_0 = N(0, C)$ and $\mu_\Gamma = N(0, C_\Gamma)$ as well as π_Γ and π_{CM} be as in (5.30) and (5.32), respectively. By Theorem 5.38 we know

$$\rho_\Gamma(u,v) = \pi_{\mathrm{CM}}\Big(\Delta_\Gamma u, \frac{1}{s}(v - A_\Gamma u)\Big)\ \pi_\Gamma\Big(\frac{1}{s}(v - A_\Gamma u)\Big).$$

By first applying a change of variables, see Lemma C.9, and then the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}\int_{\mathcal{H}} \rho_\Gamma^p(u,v)\, P_\Gamma(u, \mathrm{d}v) &= \int_{\mathcal{H}} \pi_{\mathrm{CM}}^p(\Delta_\Gamma u, v)\ \pi_\Gamma^{p-1}(v)\, \mu_0(\mathrm{d}v) \\ &\le \left(\int_{\mathcal{H}} \pi_{\mathrm{CM}}^{2p}(\Delta_\Gamma u, v)\mu_0(\mathrm{d}v)\right)^{1/2} \left(\int_{\mathcal{H}} \pi_\Gamma^{2p-2}(v)\mu_0(\mathrm{d}v)\right)^{1/2}.\end{aligned}$$

Furthermore, we have by applying Proposition C.4 from Appendix C

$$\begin{aligned}\int_{\mathcal{H}} \pi_{\mathrm{CM}}^{2p}\Big(\Delta_\Gamma u, v\Big)\,\mu_0(\mathrm{d}v) &= \int_{\mathcal{H}} \mathrm{e}^{-\frac{2p}{2}\|C^{-1/2}\Delta_\Gamma u\|_{\mathcal{H}}^2}\,\mathrm{e}^{2p\ \langle C^{-1}\Delta_\Gamma u, v\rangle_{\mathcal{H}}}\ \mu_0(\mathrm{d}v)\\ &= \exp\Big((2p^2-p)\|C^{-1/2}\Delta_\Gamma u\|_{\mathcal{H}}^2\Big).\end{aligned}$$

We apply $\|C^{-1/2}\Delta_\Gamma u\|_{\mathcal{H}} \le \|C^{-1/2}\Delta_\Gamma\|\,\|u\|_{\mathcal{H}}$ and set

$$b := (2p^2-p)\,\|C^{-1/2}\Delta_\Gamma\|^2.$$

Note, that $b \le 0$ for $p \le \frac{1}{2}$. Due to the assumptions on p we have

$$\langle (2p-2)H_\Gamma v, v\rangle_{\mathcal{H}} < \frac{\langle H_\Gamma v, v\rangle_{\mathcal{H}}}{\|H_\Gamma\|} \le \|v\|_{\mathcal{H}}^2, \qquad v \in \mathcal{H}.$$

Thus, we can apply Proposition C.7 and get

$$\begin{aligned}\int_{\mathcal{H}} \pi_\Gamma^{2p-2}(v)\mu_0(\mathrm{d}v) &= \int_{\mathcal{H}} \frac{\exp\Big(\frac{1}{2}\langle (2p-2)H_\Gamma\, C^{-1/2}v, C^{-1/2}v\rangle\Big)}{\det(I+H_\Gamma)^{(2p-2)/2}}\,\mu_0(\mathrm{d}v)\\ &= \Big(\det(I-(2p-2)H_\Gamma)\ \det(I+H_\Gamma)^{2p-2}\Big)^{-1/2}\\ &=: c^2.\end{aligned}$$

Since H_Γ is positive and trace class, $\det(I+H_\Gamma)$ is well-defined (see Definition A.5 in Appendix A) and $\det(I+H_\Gamma) \in [1,\infty)$. Furthermore, due to $\langle (2p-2)H_\Gamma v, v\rangle_{\mathcal{H}} < \|v\|_{\mathcal{H}}^2$, the eigenvalues of $(2p-2)H_\Gamma$ lie within $(-\infty, 1)$ which ensures that $\det(I-(2p-2)H_\Gamma) > 0$ and, hence $0 < c^2 < \infty$. This proves the assertion. □

Thus, the above theorem allows us to estimate the integral in (5.28). We obtain for $1 < p < 1 + 1/(2\|H_\Gamma\|)$ that

$$\int_A \int_{A^c} \rho_\Gamma(u;v)^p P_\Gamma(u,\mathrm{d}v)\,\mu(\mathrm{d}u) \le c\int_A \exp\left(\frac{b}{2}\,\|u\|_{\mathcal{H}}^2\right)\mu(\mathrm{d}u), \qquad b>0,\ c<\infty.$$

Unfortunately, if we divide the right-hand side by $\mu(A)$ and take the supremum over all $\{A : 0 < \mu(A) \le 0.5\}$ this will be unbounded. This can be seen by choosing the sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{B}(\mathcal{H})$ with $A_n := \{u \in \mathcal{H}\colon \|u\|_{\mathcal{H}} > 2n\}$: since $\mu(A_n) \propto \mu_0(A_n) \to 0$ as $n \to \infty$ there exists an n_0 such that $\mu(A_n) \le 0.5$ for $n \ge n_0$ but on the other hand we also have

$$\int_{A_n} \exp\left(\frac{b}{2}\,\|u\|_{\mathcal{H}}^2\right)\mu(\mathrm{d}u) \ge \mu(A_n)\exp(bn).$$

Thus, we can not yet verify (5.28). In the next section we introduce restrictions of the target measure and the pCN and gpCN Metropolis kernel for which we can circumvent this problem of unboundedness.

5.4.3. Restrictions of the Target Measure and Restricted Markov Kernels

In order to prove the boundedness of κ_p in (5.28) for the gpCN proposal we consider restrictions of the target measure to bounded sets. Let us mention here that restricted measures appear, for example, also in Bou-Rabee and Hairer [19, Equation (3.5)] and in the recent work by Hu et al. [91], in order to analyze the convergence of Metropolis-Hastings algorithms.

Definition 5.40 (Restricted measure)**.** Let $R \in (0, \infty]$ and μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. We set

$$\mathcal{H}_R := \{u \in \mathcal{H} \colon \|u\|_{\mathcal{H}} < R\}.$$

and define the *restriction of μ to $\mathcal{H}_R$* as the probability measure μ_R on $\mathcal{H}$ given by

$$\mu_R(\mathrm{d}u) := \frac{1}{\mu(\mathcal{H}_R)} \mathbf{1}_{\mathcal{H}_R}(u)\mu(\mathrm{d}u) \tag{5.33}$$

For sufficiently large R the measure μ_R is close to μ, because

$$d_{\mathrm{TV}}(\mu_R, \mu) = \int_{\mathcal{H}} \left| \frac{\mathrm{d}\mu_R}{\mathrm{d}\mu}(u) - 1 \right| \mu(\mathrm{d}u) = \mu(\mathcal{H}_R^c) + 1 - \mu(\mathcal{H}_R) = 2\mu(\mathcal{H}_R^c)$$

and since μ is a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ there exists for any $\varepsilon > 0$ a number $R > 0$ such that $2\mu(\mathcal{H}_R^c) < \varepsilon$.

We ask now whether good convergence properties of a μ-reversible Markov kernel K are inherited on a suitably modified μ_R-reversible Markov kernel K_R.

Definition 5.41 (Restricted Markov kernel)**.** Let K be a Markov kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and $R \in (0, \infty]$. Then the *restricted Markov kernel* $K_R \colon \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0, 1]$ is defined by

$$K_R(u, \mathrm{d}v) := \mathbf{1}_{\mathcal{H}_R}(v)\, K(u, \mathrm{d}v) + K(u, \mathcal{H}_R^c)\, \delta_u(\mathrm{d}v). \tag{5.34}$$

The next result shows that restricting Markov kernels preserves the Metropolis form (5.7) and also reversibility (w.r.t. an appropriate measure).

Proposition 5.42. Let μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and K be a μ-reversible Markov kernel. Then for any $R > 0$ the Markov kernel K_R given in (5.34) is μ_R-reversible with μ_R as in (5.33). Moreover, for a Metropolis kernel M of the form

(5.7) the corresponding restricted kernel M_R is again a Metropolis kernel

$$M_R(u, \mathrm{d}v) = \alpha_R(u, v)P(u, \mathrm{d}v) + \delta_u(\mathrm{d}v)\left(1 - \int_{\mathcal{H}} \alpha_R(u, w)P(u, \mathrm{d}w)\right)$$

with $\alpha_R(u, v) := \mathbf{1}_{\mathcal{H}_R}(v)\alpha(u, v)$.

Proof. Recall that K is μ-reversible iff

$$\int_A K(u, B)\,\mu(\mathrm{d}u) = \int_B K(u, A)\,\mu(\mathrm{d}u), \qquad \forall A, B \in \mathcal{B}(\mathcal{H}).$$

Let $A, B \in \mathcal{B}(\mathcal{H})$. We have

$$\begin{aligned}\int_A K_R(u, B)\,\mu_R(\mathrm{d}u) &= \int_A K(u, B \cap \mathcal{H}_R)\,\mu_R(\mathrm{d}u) + \int_{A\cap B} K(u, \mathcal{H}_R^c)\,\mu_R(\mathrm{d}u) \\ &= \frac{1}{\mu(\mathcal{H}_R)}\int_{A\cap\mathcal{H}_R} K(u, B \cap \mathcal{H}_R)\,\mu(\mathrm{d}u) + \int_{A\cap B} K(u, \mathcal{H}_R^c)\,\mu_R(\mathrm{d}u).\end{aligned}$$

Because of the μ-reversibility of K we can interchange A and B which leads to the first assertion. The second statement follows by

$$\begin{aligned}M_R(u, \mathrm{d}v) &= \mathbf{1}_{\mathcal{H}_R}(v)M(u, \mathrm{d}v) + \delta_u(\mathrm{d}v)M(u, \mathcal{H}_R^c) \\ &= \mathbf{1}_{\mathcal{H}_R}(v)\alpha(u, v)P(u, \mathrm{d}v) \\ &\qquad + \delta_u(\mathrm{d}v)\left(1 - \int_{\mathcal{H}} \alpha(u, w)P(u, \mathrm{d}w) + \int_{\mathcal{H}_R^c} \alpha(u, w)P(u, \mathrm{d}w)\right) \\ &= \mathbf{1}_{\mathcal{H}_R}(v)\alpha(u, v)P(u, \mathrm{d}v) + \delta_u(\mathrm{d}v)\left(1 - \int_{\mathcal{H}_R} \alpha(u, w)P(u, \mathrm{d}w)\right).\end{aligned}$$

□

Now we want to investigate if a spectral gap of K on $L^2_\mu(\mathcal{H})$ implies a spectral gap of the Markov operator associated with the restricted kernel K_R on $L^2_{\mu_R}(\mathcal{H})$. We first observe that

$$\mathrm{K}_R f(u) = \int_{\mathcal{H}} f(v)\,K_R(u, \mathrm{d}v) = \int_{\mathcal{H}_R} f(v)\,K(u, \mathrm{d}v) + f(u)\,K(u, \mathcal{H}_R^c).$$

This yields the following relation between $\|K_R\|_{\mu_R}$ and $\|K\|_\mu$, and $\mathrm{gap}_{\mu_R}(\mathrm{K}_R)$ and $\mathrm{gap}_\mu(\mathrm{K})$, respectively.

Lemma 5.43. Under the assumptions of Proposition 5.42 holds

$$\mathrm{gap}_{\mu_R}(\mathrm{K}_R) \geq \mathrm{gap}_\mu(\mathrm{K}) - \sup_{u\in\mathcal{H}_R} K(u, \mathcal{H}_R^c). \tag{5.35}$$

Furthermore, if the Markov operator K is positive on $L^2_\mu(\mathcal{H})$, then K_R is also positive on $L^2_{\mu_R}(\mathcal{H})$.

Proof. For $f \in L^2_{\mu_R}(\mathcal{H})$ let

$$(Ef)(u) := \mathbf{1}_{\mathcal{H}_R}(u) f(u) \in L^2_\mu(\mathcal{H}).$$

Note that $\|f\|_{L^2_{\mu_R}} = \frac{1}{\sqrt{\mu(\mathcal{H}_R)}} \|Ef\|_{L^2_\mu}$ and for $\int_{\mathcal{H}_R} f \,\mathrm{d}\mu_R = 0$ follows $\int_{\mathcal{H}} Ef \,\mathrm{d}\mu = 0$. Further, for any $f \in L^2_{\mu_R}(\mathcal{H})$ we have

$$\begin{aligned}
\|\mathrm{K}_R f\|^2_{L^2_{\mu_R}} &= \int_{\mathcal{H}_R} \left| \int_{\mathcal{H}_R} f(v)\, K(u, \mathrm{d}v) + f(u)\, K(u, \mathcal{H}_R^c) \right|^2 \mu_R(\mathrm{d}u) \\
&= \int_{\mathcal{H}_R} \left| \int_{\mathcal{H}} Ef(v)\, K(u, \mathrm{d}v) + Ef(u)\, K(u, \mathcal{H}_R^c) \right|^2 \mu_R(\mathrm{d}u) \\
&= \|\mathrm{K}(Ef) + g\, Ef\|^2_{L^2_{\mu_R}}
\end{aligned}$$

with $g(u) := \mathbf{1}_{\mathcal{H}_R}(u)\, K(u, \mathcal{H}_R^c)$. Then

$$\begin{aligned}
\frac{\|\mathrm{K}_R f\|_{L^2_{\mu_R}}}{\|f\|_{L^2_{\mu_R}}} &= \frac{\|\mathrm{K}(Ef) + g\, Ef\|_{L^2_{\mu_R}}}{\|Ef\|_{L^2_{\mu_R}}} = \frac{\|E(\mathrm{K}(Ef)) + g\, Ef\|_{L^2_\mu}}{\|Ef\|_{L^2_\mu}} \\
&\leq \frac{\|\mathrm{K}(Ef)\|_{L^2_\mu} + \|g\, Ef\|_{L^2_\mu}}{\|Ef\|_{L^2_\mu}} \leq \frac{\|\mathrm{K}(Ef)\|_{L^2_\mu}}{\|Ef\|_{L^2_\mu}} + \sup_{u \in \mathcal{H}_R} K(u, \mathcal{H}_R^c),
\end{aligned}$$

where we applied $\|Ef\|_{L^2_\mu} \leq \|f\|_{L^2_\mu}$ in the first inequality. By taking the supremum over all $f \in L^2_{\mu_R,0}(\mathcal{H})$ and because of $E(L^2_{\mu_R,0}(\mathcal{H})) \subseteq L^2_{\mu,0}(\mathcal{H})$ we obtain

$$\|\mathrm{K}_R\|_{\mu_R} \leq \|\mathrm{K}\|_\mu + \sup_{u \in \mathcal{H}_R} K(u, \mathcal{H}_R^c)$$

and the first assertion follows. Moreover, we have for $f \in L^2_{\mu_R}(\mathcal{H})$ that

$$\begin{aligned}
\langle \mathrm{K}_R f, f \rangle_{\mu_R} &= \int_{\mathcal{H}} \mathrm{K}_R f(u)\, f(u)\, \mu_R(\mathrm{d}u) \\
&= \int_{\mathcal{H}} \left(\int_{\mathcal{H}_R} f(v)\, K(u, \mathrm{d}v) + f(u)\, K(u, \mathcal{H}_R^c) \right) f(u)\, \mu_R(\mathrm{d}u) \\
&= \int_{\mathcal{H}} \int_{\mathcal{H}} (Ef)(v)\, K(u, \mathrm{d}v)\; (Ef)(u)\, \frac{\mu(\mathrm{d}u)}{\mu(\mathcal{H}_R)} + \int_{\mathcal{H}} f^2(u)\, K(u, \mathcal{H}_R^c)\, \mu_R(\mathrm{d}u).
\end{aligned}$$

The second term is always positive since $f^2(u)\, K(u, \mathcal{H}_R^c) \geq 0$ for all $u \in \mathcal{H}$ and the first term coincides with $\langle \mathrm{K}(Ef), Ef \rangle_\mu / \mu(\mathcal{H}_R)$. Thus, the second statement is proven. □

Lemma 5.43 tells us that the Markov operator K_R associated with the restricted Markov kernel K_R possesses an $L^2_{\mu_R}$-spectral gap if K has an L^2_μ-spectral gap and if $\sup_{u\in\mathcal{H}_R} K(u,\mathcal{H}_R^c)$ is sufficiently small. We can now apply this result to the pCN Metropolis algorithm as done in the next subsection.

5.4.4. The Spectral Gap of the Restricted gpCN Metropolis Kernel

Now we can combine all our previous results to establish our main convergence result for the restricted gpCN-Metropolis algorithm. But let us first consider the restricted pCN Metropolis which we need later for applying the comparison theorem for spectral gaps, Theorem 5.33.

Theorem 5.44 (Spectral gap of restricted pCN Metropolis)**.** Let μ be as in (5.1) and let M_0 denote the μ-reversible pCN Metropolis kernel. If there holds $\mathrm{gap}_\mu(\mathrm{M}_0) > 0$, then for any $\varepsilon > 0$ there exists a number $R \in (0,\infty)$ such that

$$\mathrm{gap}_{\mu_R}(\mathrm{M}_{0,R}) \geq \mathrm{gap}_\mu(\mathrm{M}_0) - \varepsilon,$$

where μ_R is as in (5.33) and $M_{0,R}$ according to Definition 5.41.

Proof. Given the results of Proposition 5.42 and Lemma 5.43 it suffices to prove that for any $\varepsilon > 0$ there exists an $R > 0$ such that $\sup_{u\in\mathcal{H}_R} M_0(u,\mathcal{H}_R^c) \leq \varepsilon$. We recall that the proposal kernel of M_0 is $P_0(u,\cdot) = N(\sqrt{1-s^2}u, s^2C)$ and obtain with $\mu_0^s := N(0,s^2C)$ that

$$\begin{aligned}
\sup_{u\in\mathcal{H}_R} M_0(u,\mathcal{H}_R^c) &\leq \sup_{u\in\mathcal{H}_R} P_0(u,\mathcal{H}_R^c) = \sup_{u\in\mathcal{H}_R} \int_{\|\sqrt{1-s^2}u+v\|_{\mathcal{H}}\geq R} \mathrm{d}\mu_0^s(v) \\
&\leq \sup_{u\in\mathcal{H}_R} \int_{\|\sqrt{1-s^2}u\|_{\mathcal{H}}+\|v\|_{\mathcal{H}}\geq R} \mathrm{d}\mu_0^s(v) \\
&= \sup_{u\in\mathcal{H}_R} \int_{\|v\|_{\mathcal{H}}\geq R-\sqrt{1-s^2}\|u\|_{\mathcal{H}}} \mathrm{d}\mu_0^s(v) \\
&\leq \int_{\|v\|_{\mathcal{H}}\geq(1-\sqrt{1-s^2})R} \mathrm{d}\mu_0^s(v) = \mu_0(\mathcal{H}_{R_s}^c)
\end{aligned}$$

where $R_s = \frac{1-\sqrt{1-s^2}}{s}R$ and $\mu_0 = N(0,C)$. Again, since μ_0 is a probability measure on $\mathcal{H}$ we know that there exists a number R, such that $\mu_0(\mathcal{H}_{R_s}^c) \leq \varepsilon$. □

We apply now Theorem 5.33 to the restricted pCN and restricted gpCN Metropolis and obtain:

Theorem 5.45 (Convergence of restricted gpCN Metropolis)**.** Let μ be as in (5.1) and assume that the pCN Metropolis kernel possesses a spectral gap in $L^2_\mu(\mathcal{H})$, i.e., $\mathrm{gap}_\mu(\mathrm{M}_0) > 0$. Then, for any $\Gamma \in \mathcal{L}_+(\mathcal{H})$ and any $\varepsilon \in (0, \mathrm{gap}_\mu(M_0))$ there exists a number $R_0 = R_0(\varepsilon) \in (0, \infty)$ such that for any $R \geq R_0$ holds

$$d_{\mathrm{TV}}(\mu_R, \mu) < \varepsilon \quad \text{and} \quad \mathrm{gap}_{\mu_R}(\mathrm{M}_{\Gamma,R}) > 0$$

where μ_R is as in (5.33) and $M_{\Gamma,R}$ according to Definition 5.41.

Proof. By Theorem 5.44 we have that for any $\varepsilon \in (0, \mathrm{gap}_\mu(\mathrm{M}_0))$ there exists a number $R_0 \in (0, \infty)$ such that for any $R \geq R_0$ holds

$$d_{\mathrm{TV}}(\mu_R, \mu) \leq \varepsilon \quad \text{and} \quad \mathrm{gap}_\mu(\mathrm{M}_{0,R}) > 0.$$

Next, we will verify the assumptions of Theorem 5.33 for $M_1 := M_{0,R}$ and $M_2 := M_{\Gamma,R}$ which then yields the assertion. By Proposition 5.42 we know that $M_{\Gamma,R}$ is again a Metropolis kernel with proposal P_Γ and acceptance probability α_R for any $\Gamma \in \mathcal{L}_+(\mathcal{H})$. Thus, M_1 and M_2 employ the same acceptance probability and proposal kernels $P_1 = P_0$ and $P_2 = P_\Gamma$, respectively. Moreover, by Theorem 5.38 we know that

$$\frac{\mathrm{d}P_1(u)}{\mathrm{d}P_2(u)}(v) = \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v) = \rho_\Gamma(u, v)$$

exists for each $u \in \mathcal{H}$. Since M_1 and M_2 are μ_R-reversible due to Proposition 5.42, we are left to verify that for a $p > 1$ there holds

$$\kappa_{p,R} := \sup_{\mu_R(A) \in (0,1/2]} \frac{\int_A \int_{A^c} \rho_\Gamma(u, v)^p \, P_\Gamma(u, \mathrm{d}v) \, \mu_R(\mathrm{d}u)}{\mu_R(A)} < \infty$$

and that the associated Markov operators $\mathrm{M}_1 = \mathrm{M}_{0,R}$ and $\mathrm{M}_2 = \mathrm{M}_{\Gamma,R}$ are positive on $L^2_{\mu_R}(\mathcal{H})$. The latter follows immediately by Lemma 5.43 in combination with Theorem 5.36. And by Theorem 5.39 we have for any $p < 1 + \frac{1}{2\|H_\Gamma\|}$ that

$$\kappa_{p,R} \leq \sup_{\mu_R(A) \in (0,1/2]} \frac{\int_A c \exp\left(\frac{b}{2} \|u\|^2_{\mathcal{H}}\right) \mu_R(\mathrm{d}u)}{\mu_R(A)} \leq c \exp\left(\frac{b}{2} R^2\right) < \infty.$$

Thus, Theorem 5.33 can be applied to M_1 and M_2 and yields

$$\mathrm{gap}_{\mu_R}(\mathrm{M}_{\Gamma,R})^{(p-1)/2} \geq \frac{1}{2^{(3p-1)/2}} \frac{\mathrm{gap}_{\mu_R}(\mathrm{M}_{0,R})^p}{\kappa_{p,R}} > 0$$

for any $p < 1 + \frac{1}{2\|H_\Gamma\|}$ which concludes the proof. □

Theorem 5.45 tells us that the corresponding restricted gpCN Metropolis converges exponentially fast to any, arbitrarily close, restriction μ_R of μ whenever the pCN Metropolis has a spectral gap, e.g., under the conditions of Hairer et al. [81, Theorem 2.14]. In particular, Theorem 5.45 is a statement about the inheritance of geometric convergence from the pCN to the restricted gpCN Metropolis. We emphasize that a quantitative comparison of their spectral gaps is not proven. We provide a lower bound for the spectral gap of $\mathrm{M}_{\Gamma,R}$ in nonlinear terms of the spectral gap of the pCN Metropolis. Additionally, the stated estimate behaves rather poorly w.r.t. R, more precisely, it decays exponentially as $R \to \infty$.

Although we argued in the above theorem with restrictions of μ in order to bound κ_p from Theorem 5.33, let us mention that, in simulations when R is sufficiently large one cannot distinguish between μ and μ_R as well as between Markov chains with transition kernels M_Γ and $M_{\Gamma,R}$.

Moreover, we conjecture that the gpCN Metropolis targeting μ has a strictly positive spectral gap whenever the pCN Metropolis has one. In particular, regarding the results of the numerical simulations in Section 5.6 we even conjecture that the spectral gap of the gpCN Metropolis with suitably chosen $\Gamma \in \mathcal{L}_+(\mathcal{H})$ is much larger than the one of the pCN Metropolis.

5.5. A gpCN Metropolis Algorithm with State-Dependent Proposal Covariance

So far, the gpCN proposal employs one fixed covariance operator which is supposed to approximate the covariance of the target measure. We extend now the gpCN proposal in order to allow for state-dependent proposal covariances. The advantage of such a state-dependent approach is that the resulting Metropolis algorithm might be even better adapted to the target measure by allowing locally different proposal covariances. For an illustrative motivation for state-dependent proposal covariances we refer to Girolami and Calderhead [73] and Martin et al. [118]. First theoretical results on the geometric ergodicity of random walk proposals with state-dependent covariances were recently obtained by Livingstone [113] in the case of finite dimensional state spaces. Moreover, we mention the work by Beskos et al. [12] where the authors construct MALA and Hamiltonian Monte Carlo (HMC) algorithms in Hilbert spaces which employ local metric tensors for proposing new states. Their algorithms are derived by inserting state-dependent metric tensors into the stochastic and Hamiltonian differential equations underlying the standard MALA and HMC algorithm, respectively, and then applying the same

semi-implicit time stepping scheme as done for the pCN Metropolis, cf. Section 5.1.2. We will take a slightly different approach to define our *local gpCN* and *pCN Metropolis algorithm*. In particular, we will exploit the existence of the density between the gpCN and pCN proposal as shown in Theorem 5.38.

Definition 5.46. Given a measureable mapping $\mathcal{H} \ni u \mapsto \Gamma(u) \in \mathcal{L}_+(\mathcal{H})$ and $\mu_0 = N(0, C)$ on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ we define the *local gpCN proposal* by

$$P_{\text{loc}}(u, \cdot) = N(A_{\Gamma(u)}u, s^2 C_{\Gamma(u)}) \tag{5.36}$$

where $C_{\Gamma(u)} = C_\Gamma = (C^{-1} + \Gamma(u))^{-1}$ and

$$A_{\Gamma(u)} = C^{1/2} \sqrt{I - s^2 \left(I + H_{\Gamma(u)}\right)^{-1}} C^{-1/2}, \qquad H_{\Gamma(u)} = C^{1/2}\Gamma(u)C^{1/2}.$$

Remark 5.47. Following the heuristic presented in Section 5.2.1 for Bayesian inference problems where Φ in (5.1) is of the form (5.15), we could choose for instance

$$\Gamma(u) = \nabla G(u)^* \, \Sigma^{-1} \, \nabla G(u). \tag{5.37}$$

A similar idea appears in the paper by Beskos et al. [12]: motivated by the stochastic Newton algorithm by Martin et al. [118], they consider $\Gamma(u)$ as (a suitable positive modification of) the Hessian of $\Phi(u) := |y - G(u)|^2_{\Sigma^{-1}}$.

In order to derive the appropriate acceptance probability for a Metropolis algorithm with the local gpCN proposal, we define the measure

$$\eta_{\text{loc}}(\mathrm{d}u, \mathrm{d}v) := P_{\text{loc}}(u, \mathrm{d}v)\mu_0(\mathrm{d}u).$$

We notice that η_{loc} is no longer a Gaussian measure due to the dependence of Γ on u. However, to construct a μ-reversible Metropolis kernel with a proposal P_{loc} as above, we can apply the same trick as Beskos et al. [15, Theorem 4.1]. Namely, with $\rho_\Gamma(u, v) = \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v)$ as given in Theorem 5.38 we obtain

$$\begin{aligned} P_{\text{loc}}(u, \mathrm{d}v)\mu_0(\mathrm{d}u) &= \frac{1}{\rho_{\Gamma(u)}(u, v)} P_0(u, \mathrm{d}v)\mu_0(\mathrm{d}u) = \frac{1}{\rho_{\Gamma(u)}(u, v)} P_0(v, \mathrm{d}u)\mu_0(\mathrm{d}v) \\ &= \frac{\rho_{\Gamma(v)}(v, u)}{\rho_{\Gamma(u)}(u, v)} P_{\text{loc}}(v, \mathrm{d}u)\mu_0(\mathrm{d}v), \end{aligned}$$

where we used the μ_0-reversibility of the pCN proposal P_0. Hence, according to the general Metropolis kernel construction outlined in Section 5.1.2 we obtain the following result.

Corollary 5.48. Let $\mu_0 = N(0, C)$, μ be given by (5.1) and let P_{loc} be a local gpCN proposal as in Definition 5.46. Then

$$M_{\text{loc}}(u, \mathrm{d}v) := \alpha_{\text{loc}}(u, v) P_{\text{loc}}(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \int_{\mathcal{H}} (1 - \alpha_{\text{loc}}(u, w))\, P_{\text{loc}}(u, \mathrm{d}w)$$

with

$$\alpha_{\text{loc}}(u, v) = \min \left\{ 1, \exp(\Phi(u) - \Phi(v)) \, \frac{\rho_{\Gamma(u)}(u, v)}{\rho_{\Gamma(v)}(v, u)} \right\}, \tag{5.38}$$

where $\rho_\Gamma(u, v) = \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v)$ as given in Theorem 5.38, defines a μ-reversible Metropolis kernel which we call *local gpCN Metropolis kernel*.

Note, that the same construction can analogously be applied to local variants of the pCN proposals.

Definition 5.49. Under the same assumptions as in Definition 5.46 we define the *local pCN proposal* by

$$P'_{\text{loc}}(u, \cdot) := N(\sqrt{1 - s^2} u, s^2 C_{\Gamma(u)}). \tag{5.39}$$

Corollary 5.50. Let $\mu_0 = N(0, C)$, μ be as in (5.1) and P'_{loc} denote a local pCN proposal according to Definition 5.49. Then the *local pCN Metropolis kernel* given by

$$M'_{\text{loc}}(u, \mathrm{d}v) := \alpha'_{\text{loc}}(u, v) P'_{\text{loc}}(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \int_{\mathcal{H}} (1 - \alpha'_{\text{loc}}(u, w))\, P'_{\text{loc}}(u, \mathrm{d}w),$$

where

$$\alpha'_{\text{loc}}(u, v) = \min \left\{ 1, \exp(\Phi(u) - \Phi(v)) \, \frac{\pi_{\Gamma(u)}(\frac{1}{s}[v - A_0 u])}{\pi_{\Gamma(v)}(\frac{1}{s}[u - A_0 v])} \right\} \tag{5.40}$$

with π_Γ as stated in Theorem 5.38, is μ-reversible.

Thus, we defined in the above corollaries two Metropolis-Hastings algorithms employing state-dependent proposal covariances which are well-posed in infinite dimensions. Unfortunately, the tools and results developed and presented in Section 5.3 are not sufficient to prove also spectral gaps of these MH algorithms. The main reason for this is the missing reversibility of the proposals P_{loc} and P'_{loc} w.r.t. the prior measure μ_0. This reversibility condition played a key role in proving Theorem 5.33 and, therefore, the analysis of Section 5.4 which was driven by this theorem is not applicable to M_{loc} and M'_{loc}. This could be a topic for future research.

Of course, also the question arises if the additional computational cost of evaluating $\Gamma(u)$ and, maybe even more costly, evaluating $\rho_{\Gamma(u)}$ or $\pi_{\Gamma(u)}$ in each step pays

off in a significantly higher statistical efficiency — see also the next section for numerical experiments with the local pCN and the local gpCN Metropolis algorithm.

Remark 5.51. Related to the concern of computational work, one could think of substituting $\nabla G(u)$ in (5.37) by a cheaper approximation in order to reduce the computational work. This might help to make MH algorithms with local proposal covariances feasible and is again left for future research.

5.6. Numerical Experiments

We illustrate the gpCN Metropolis algorithm for approximate sampling of a posterior distribution in Bayesian inference. In particular, we compare different Metropolis algorithms and examine which of those perform independently of the state space dimension and of the variance of the involved observational noise.

Remark 5.52 (Error due to numerical discretizations)**.** In every numerical simulation we have to apply discretizations and approximations, e.g., of functions or operators. In the setting of Bayesian inference where we condition on observations $y = G(U) + \varepsilon$ with $G\colon \mathcal{H} \to \mathbb{R}^d$ and U being a $\mathcal{H}$-valued random variable, we typically have to employ finite dimensional subspaces of $\mathcal{H}$ as well as approximations of the forward map G. Both will contribute to the error in the resulting approximation of the true posterior. Concerning an error analysis for projections to finite dimensional subspaces, we refer to Dashti and Stuart [42, Section 2.3] whereas the error due to approximating G is estimated in Theorem 3.21.

5.6.1. Problem Setting

We consider the same setting and inference problem as Pinski et al. [133, Section 6.1]: given noisy observations $y_j = p(0.2j) + \varepsilon_j$ with $j = 1, \dots, 4$, of the solution p of

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{u(x)}\,\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 0 \ \text{ on } D = [0,1], \qquad p(0) = 0,\ p(1) = 2, \tag{5.41}$$

we want to infer u. Here, the ε_j are independent realizations of the normal distribution $N(0, \sigma_\varepsilon^2)$. We place a Gaussian prior $N(0, -\Delta^{-1})$ with $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ on the completion $\mathcal{H}_c$ of $H_0^1(D) \cap H^2(D)$ in $L^2(D)$. Recall the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $U\colon \Omega \to \mathcal{H}_c \subset L^2(D)$ be a random function with distribution $N(0, \Delta^{-1})$. This allows us to represent the random function U as

$$U(\omega)(x) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty} \xi_k(\omega)\sin(k\pi x), \qquad \xi_k \sim N(0, k^{-2}), \tag{5.42}$$

$\mathbb{P}$-a.s. where all random variables ξ_k are independent. In other words U describes a Brownian bridge, see Example 2.31. Thus, inferring U is equivalent to inferring $\boldsymbol{\xi} = (\xi_k)_{k\in\mathbb{N}}$. This leads to the prior $\mu_0 = N(0, C)$ with $C = \operatorname{diag}(k^{-2} : k \in \mathbb{N})$ on $\mathcal{H} := \ell^2(\mathbb{N})$, since $\mathbb{P}$-a.s. $\boldsymbol{\xi}(\omega) \in \ell^2(\mathbb{N})$, see Schwab and Gittelson [156, Proposition C.12]. Further, we denote by μ the resulting conditional distribution of $\boldsymbol{\xi}$ given the observed data $y = (y_1, \dots, y_4)$. The measure μ is given by a density of the form (5.1) with Φ as in (5.15) where $\Sigma = \sigma_\varepsilon^2 I$ and $G(\boldsymbol{\xi})$ is the mapping

$$\boldsymbol{\xi} \mapsto u(\cdot, \boldsymbol{\xi}) \mapsto p(\cdot, \boldsymbol{\xi}) \mapsto (p(0.2j, \boldsymbol{\xi}))_{j=1}^4.$$

The solution $p(\boldsymbol{\xi})$ of (5.41) for a diffusion coefficient $u(\boldsymbol{\xi})$ is given by

$$p(x, \boldsymbol{\xi}) = \frac{2}{S_1(\mathrm{e}^{-u(\xi)})} S_x(\mathrm{e}^{-u(\xi)}) \tag{5.43}$$

with $S_x(f) = \int_0^x f(y)\mathrm{d}y$. For numerical simulations we use a uniform discretization of $[0, 1]$ with $\Delta x = 2^{-10}$ and apply the trapezoidal rule for evaluating $S_x(f)$ and integrals w.r.t. $\mathrm{d}x$, respectively. Furthermore, we truncate the expansion (5.42) after N terms where we vary N in order to test the Metropolis algorithms for dimension independent performance. The data y is generated by the specific choice $u(x) = 2\sin(2\pi x)$. We also consider different noise levels σ_ε to examine the effect of smaller variances σ_ε^2, leading to more concentrated posterior distributions μ, on the performance of the Metropolis algorithms.

We choose four quantities of interest

$$f_1(\boldsymbol{\xi}) := \int_0^1 \mathrm{e}^{u(x,\boldsymbol{\xi})}\,\mathrm{d}x, \quad f_2(\boldsymbol{\xi}) := \max_{0\le x\le 1} \mathrm{e}^{u(x,\boldsymbol{\xi})}, \quad f_3(\boldsymbol{\xi}) := p(0.5, \boldsymbol{\xi}), \quad f_4(\boldsymbol{\xi}) := \xi_1,$$

for our experiments. However, the obtained results for the comparison of MCMC methods will be essentially the same for any of the mentioned quantities and we will sometimes only present the results for the first quantity f_1.

5.6.2. Comparison of Different Metropolis Algorithms

We apply four Metropolis algorithms denoted by RW, pCN, GN-RW and gpCN resulting from the following four proposal kernels:

- RW: Gaussian random walk proposal $P_1(\boldsymbol{\xi}, \cdot) = N(\boldsymbol{\xi}, s^2 C)$,
- pCN: pCN proposal $P_2(\boldsymbol{\xi}, \cdot) = N(\sqrt{1 - s^2}\boldsymbol{\xi}, s^2 C)$,
- GN-RW: Gauss-Newton random walk proposal $P_3(\boldsymbol{\xi}, \cdot) = N(\boldsymbol{\xi}, s^2 C_\Gamma)$,

- gpCN: gpCN proposal $P_4(\boldsymbol{\xi}, \cdot) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma)$.

We choose $\Gamma = \sigma_\varepsilon^{-2} L L^\top$ with $L = \nabla G(\boldsymbol{\xi}_{\mathrm{MAP}})$ and

$$\boldsymbol{\xi}_{\mathrm{MAP}} = \operatorname*{argmin}_{\xi \in \mathrm{rg}(C^{1/2})} \left(\sigma_\varepsilon^{-2} |\boldsymbol{y} - G(\boldsymbol{\xi})|^2 + \|C^{-1/2}\boldsymbol{\xi}\|_{\mathcal{H}}^2 \right). \tag{5.44}$$

Since we observe linear functionals of p, the gradient $\nabla G(\boldsymbol{\xi})$ can be obtained by differentiating the explicit formula (5.43) for p w.r.t. $\boldsymbol{\xi}$. In particular, we obtain by the linearity of $S_x(\cdot)$

$$\begin{aligned}
\frac{\partial}{\partial \xi_k} p(x, \boldsymbol{\xi}) &= \left(\frac{\partial}{\partial \xi_k} \frac{2}{S_1(\mathrm{e}^{-u(\xi)})} \right) S_x\left(\mathrm{e}^{-u(\xi)}\right) + \frac{2}{S_1(\mathrm{e}^{-u(\xi)})} S_x\left(\frac{\partial}{\partial \xi_k} \mathrm{e}^{-u(\xi)} \right) \\
&= 2 \frac{S_1(\phi_k\, \mathrm{e}^{-u(\xi)})}{\left[S_1(\mathrm{e}^{-u(\xi)})\right]^2} S_x\left(\mathrm{e}^{-u(\xi)}\right) - \frac{2}{S_1(\mathrm{e}^{-u(\xi)})} S_x\left(\phi_k\, \mathrm{e}^{-u(\xi)}\right) \\
&= \frac{S_1(\phi_k\, \mathrm{e}^{-u(\xi)})}{S_1(\mathrm{e}^{-u(\xi)})} p(x, \boldsymbol{\xi}) - \frac{2}{S_1(\mathrm{e}^{-u(\xi)})} S_x\left(\phi_k\, \mathrm{e}^{-u(\xi)}\right),
\end{aligned}$$

where $\phi_k(x) := \frac{2}{\pi} \sin(k\pi x)$. Again, we evaluate the appearing integrals numerically by the trapezoidal rule mentioned above. We apply the Levenberg-Marquardt algorithm to solve the above optimization problem for the MAP estimator $\boldsymbol{\xi}_{\mathrm{MAP}}$ (5.44). Specifically, we used MATLAB's `lsqnonlin` function to do so.

Remark 5.53. In general, elliptic PDEs can be solved in a weak sense by variational methods, see Section 2.3. Then, adjoint methods known from PDE constrained optimization and parameter identification can be employed to compute $\nabla G(\boldsymbol{\xi})$, see Vogel [173, Chapter 6] for more details and Section 7.3 for an example.

As a metric for comparing the performance and efficiency of MCMC algorithms we consider and estimate the *effective sample size*

$$\mathrm{ESS} = \mathrm{ESS}(n, f, (\boldsymbol{\xi}_k)_{k \in \mathbb{N}}) := n \left[1 + 2 \sum_{k \geq 0} \gamma_f(k) \right]^{-1}$$

where n is the number of samples taken from a Markov chain $(\boldsymbol{\xi}_k)_{k \in \mathbb{N}}$ generated by an Metropolis algorithms and γ_f denotes the autocorrelation function for a quantity of interest $f \in L^2_\mu(\mathcal{H})$:

$$\gamma_f(k) = \mathrm{Corr}(f(\boldsymbol{\xi}_{n_0}), f(\boldsymbol{\xi}_{n_0+k})).$$

To estimate the ESS we compute the empirical autocorrelation function

$$\hat{\gamma}_f(k) := \frac{1}{n-k} \sum_{i=1}^{n-k} \Big(f(\boldsymbol{\xi}_{n_0+i}) - S_{n_0,n}(f) \Big) \Big(f(\boldsymbol{\xi}_{n_0+i+k}) - S_{n_0,n}(f) \Big), \qquad k < n,$$

with $S_{n_0,n}(f)$ denoting the path average as in (5.5), and use the initial monotone sequence estimators (IMSE) proposed by Geyer [68, Section 3.3]. For robustness reasons we also employ batch means to estimate the ESS [68, Section 3.2]. Since this led to similar results we only present the estimates obtained by the former method unless stated otherwise.

Remark 5.54 (On IMSE)**.** The IMSE is based on the fact that for a μ-reversible Markov chain starting at its stationary measure, the mapping $\Gamma_f(k) := \gamma_f(2k) + \gamma_f(2k-1)$, $k \in \mathbb{N}$, is strictly positive, strictly decreasing and strictly convex, see Geyer [68, Theorem 3.1]. In particular, in order to estimate the ESS the IMSE takes into account only the first K terms of the empirical values $\hat{\Gamma}_f(k) := \hat{\gamma}_f(2k) + \hat{\gamma}_f(2k-1)$ such that $(\hat{\Gamma}_f(k))_{k=1}^{K}$ is strictly positive and monotone but $(\hat{\Gamma}_f(k))_{k=1}^{K+1}$ is not.

Remark 5.55 (On batch means)**.** The idea behind batch means is that the n realizations along a path of the Markov chain, e.g., $(x_1, \dots, x_n)$, which are used for MCMC integration are divided into m batches of same size k, i.,e., $(x_1, \dots, x_k)$, $(x_{k+1}, \dots, x_{2k})$ and so on. Then the empirical mean of each batch converges in distribution to i.i.d. normal random variables with $\mathbb{E}_\mu[f]$ as mean and σ_f^2 from (5.6) as their variance. Hence, σ_f^2 can be esimated by the empirical variance of the batch means.

For all Metropolis algorithms we tune the step size parameter s such that the average acceptance rate is about 0.25, since the empirical performance of each algorithm was best for this particular tuning. In all cases we take $n_0 = 10^5$ as burn-in length and $n = 10^6$ as sample size.

Remark 5.56 (On tuning the average acceptance rate)**.** For the random walk MH algorithm the usual rule of thumb is that the stepsize s should be chosen such that the average acceptance rate is approximately 0.234, see Roberts and Rosenthal [141]. However, their result is an asymptotic result for Metropolis-Hastings algorithms which are not well-defined in infinite dimensions and the stepsize s of which must, therefore, deteriorate with increasing dimension — otherwise the Markov chains will reject the proposed new state more and more often. Thus, for the pCN and gpCN Metropolis algorithm this rule of thumb does not really apply. However, our numerical experiments suggest that an average acceptance rate of approximately

25% leads to the best performance of pCN and gpCN in terms of the ESS or integrated autocorrelation time, see Table 5.1.

	$N = 50, \sigma_\varepsilon = 0.1$			$N = 400, \sigma_\varepsilon = 0.01$		
MH proposal	$\overline{\alpha} \approx 0.25$	$\overline{\alpha} \approx 0.5$	$\overline{\alpha} \approx 0.75$	$\overline{\alpha} \approx 0.25$	$\overline{\alpha} \approx 0.5$	$\overline{\alpha} \approx 0.75$
RW	115.2	169.3	598.6	1332.8	2013.0	4157.9
pCN	60.4	80.4	265.9	382.0	650.3	2324.3
GN-RW	127.9	195.1	591.8	1098.0	1451.5	3868.7
gpCN	24.4	34.4	109.1	17.2	25.5	82.1

Table 5.1.: Estimated integrated autocorrelation times for quantity f_1 based on Metropolis algorithms based on the four proposals and tuned to certain average acceptance rates $\overline{\alpha}$ for two different settings of number of dimensions N and noise variance σ_ε.

The final results of the simulations are illustrated in Figure 5.2, Figure 5.3 and Figure 5.4. The first one displays the empirical autocorrelation functions $\hat{\gamma}_{f_1}$ resulting from the four Metropolis algorithms for various combinations of state space dimension N and noise variances σ_ε^2. We can already observe some interesting behavior in this figure:

- The two random walk Metropolis algorithms (RW, GN-RW) seem to yield more strongly correlated chains with increasing dimension — comparing (a) with (c) or (b) with (d) — whereas the autocorrelation of the Markov chains generated by the pCN and gpCN Metropolis seem not to be negatively affected by increasing N.

- If we compare (a) with (b) or (c) with (d), we detect a slower decaying autocorrelation for the Markov chains generated by the RW and the pCN Metropolis algorithm. Again the gpCN Metropolis and also the GN-RW seem to be less affected by a decreased noise variance.

We investigate these two observations a bit further and display in Figure 5.3 the estimated ESS for varying state space dimensions $N = 50, 100, 200, 400, 800$, for a fixed noise standard deviation $\sigma_\varepsilon = 0.1$. This time we present the resulting ESS for all four quantities of interest ((a) to (d)) and include the estimates for ESS obtained by the batch means method for comparison (as dashed lines). The observation made in 5.2 is confirmed in Figure 5.3: the pCN and the gpCN Metropolis show an efficiency independent of the dimension N, whereas the random walk Metropolis algorithms show the well-known, see Roberts and Rosenthal [141], deterioating efficiency which seems to decay roughly like N^{-1}. This holds for each of the four quantities of interest.

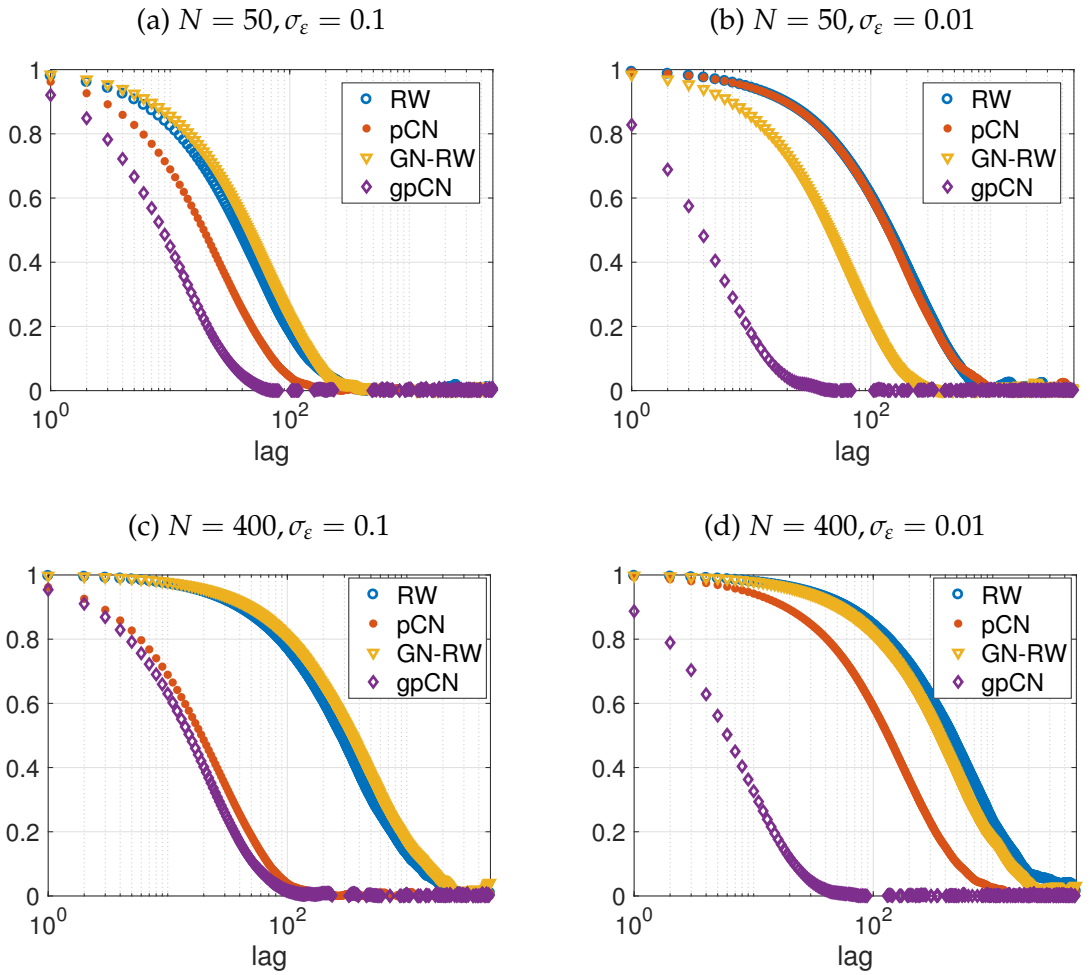

Figure 5.2.: Empirical autocorrelation of f_1 given samples generated by the four Metropolis algorithms denoted by RW, pCN, GN-RW and gpCN for various choices of state dimension N and noise standard deviation σ_ε.

The second observation from above is confirmed by Figure 5.4 which displays again the estimated ESS for $f_1, \ldots, f_4$, but now for varying noise standard deviation $\sigma_\varepsilon = 0.5, 0.25, 0.1, 0.05, 0.025, 0.01, 0.005, 0.0025, 0.001$, and a fixed state space dimension $N = 100$. We see that the gpCN and the GN-RW Metropolis algorithm perform more robust w.r.t. σ_ε. The latter shows for quite a range of σ_ε an almost constant ESS whereas the gpCN even improves its ESS when σ_ε drops from 0.1 to 0.01. However, at the end, when σ_ε becomes very small all four algorithms show a decaying efficiency.

Summarizing, the gpCN Metropolis seems to combine both desirable properties of dimension independent performance and robustness w.r.t. the noise variance whereas the other algorithms suffer from one or the other. Moreover, the gpCN performs best among the four algorithms also in absolute terms of the ESS.

Remark 5.57. The effect of a decreasing noise or likelihood variance on the perfor-

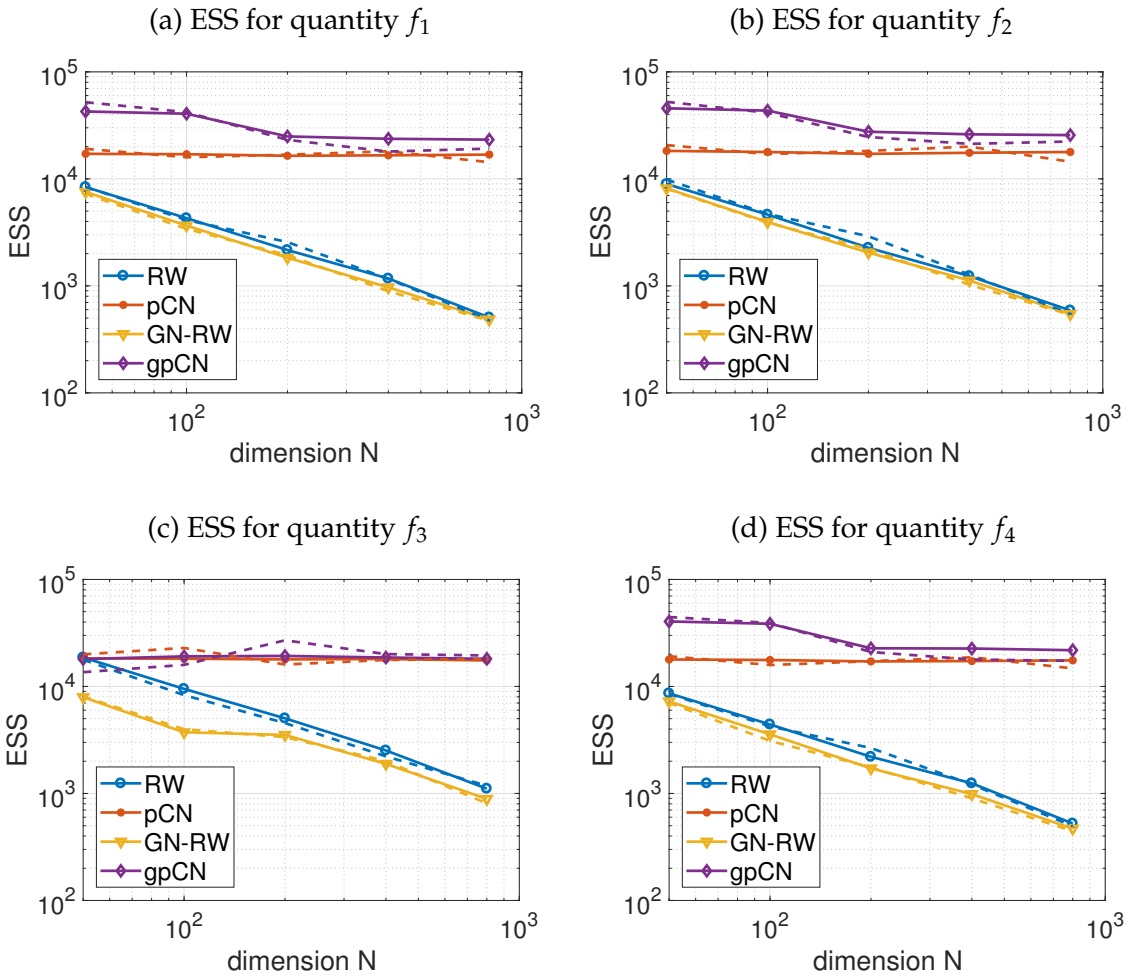

Figure 5.3.: Dependence of empirical ESS — estimated by IMSE (markers, solid lines) and batch means (dashed lines) — for each Metropolis algorithm RW, pCN, GN-RW and gpCN w.r.t. state dimension N with fixed noise variance $\sigma_\varepsilon^2 = 0.01$.

mance of MH algorithms is, surprisingly, a rather less studied issue in the MCMC community. However, it seems to gain more attention recently, see Beskos et al. [16]. We will revisit this issue in Chapter 6 where we will prove a specific notion of variance independent performance for the GN-RW Metropolis in case of a linear forward map G.

5.6.3. Performance of Metropolis Algorithms with State-Dependent Proposal Covariances

We also apply the local gpCN and the local pCN Metropolis algorithm for approximate sampling of the posterior and, in particular, compare their performance to the corresponding "nonlocal" counterparts, i.e., the gpCN and pCN Metropolis algo-

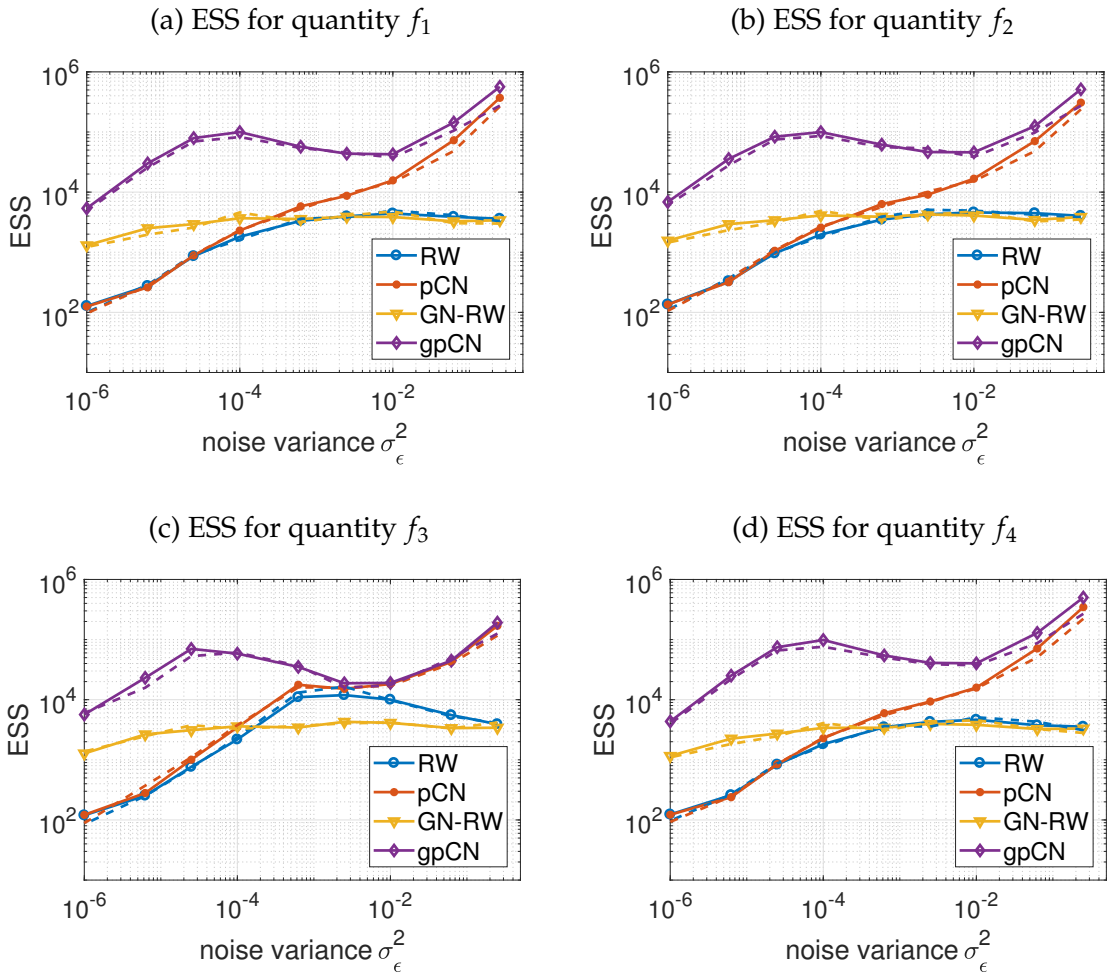

Figure 5.4.: Dependence of empirical ESS — estimated by IMSE (solid lines) and batch means (dashed lines) — for each Metropolis algorithm RW, pCN, GN-RW and gpCN w.r.t. noise variance σ_ε^2 with fixed state dimension $N = 100$.

rithm. We choose the mapping

$$\Gamma(\boldsymbol{\xi}) := \sigma_\varepsilon^{-2} \nabla G(\boldsymbol{\xi}) \nabla G(\boldsymbol{\xi})^\top$$

for both local Metropolis algorithms as motivated in Remark 5.47. Again, the gradient of G is evaluated as explained in Subsection 5.6.2. Moreover, we compute in each iteration of the local pCN and gpCN Metropolis algorithm the singular value decomposition (SVD) of $H_{\Gamma(\boldsymbol{\eta})} = C^{1/2}\Gamma(\boldsymbol{\eta})C^{1/2}$, where $\boldsymbol{\eta}$ denotes the proposed new state, by MATLAB's `svd` routine. This might be kind of a computational overkill, however, it makes the computation of $A_{\Gamma(\boldsymbol{\eta})}$ and $C_{\Gamma(\boldsymbol{\eta})}$ as well as the evaluation of the densities

$$\pi_{\Gamma(\boldsymbol{\eta})}\left(\frac{1}{s}[\boldsymbol{\xi}_n - A_0\boldsymbol{\eta}]\right) = \frac{\exp\left(\frac{1}{2s^2}[\boldsymbol{\xi}_n - A_0\boldsymbol{\eta}]^\top \Gamma(\boldsymbol{\eta})[\boldsymbol{\xi}_n - A_0\boldsymbol{\eta}]\right)}{\sqrt{\det(I + H_{\Gamma(\boldsymbol{\eta})})}}$$

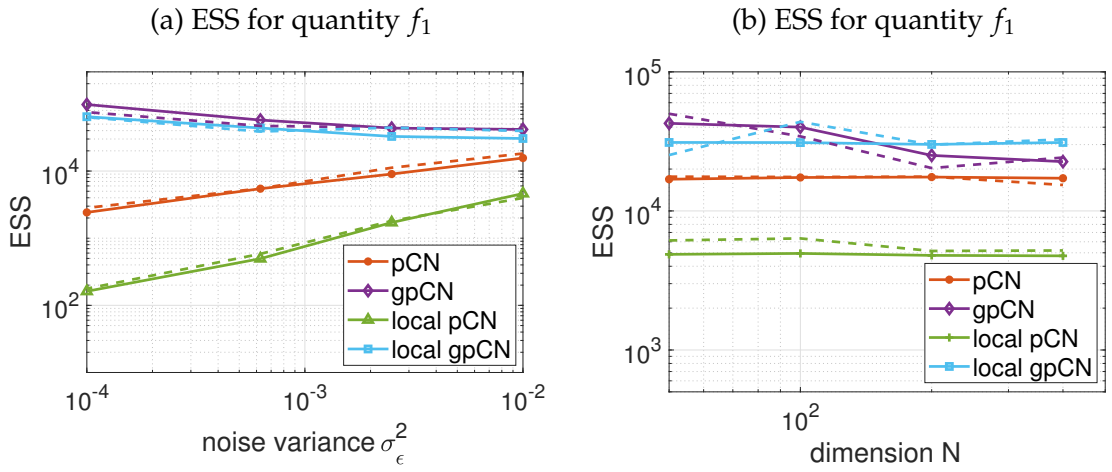

Figure 5.5.: Dependence of empirical ESS — estimated by IMSE (solid lines) and batch means (dashed lines) — for the pCN and gpCN Metropolis algorithm and their local variants w.r.t. noise variance (left) and w.r.t. state space dimension (right).

and $\rho_{\Gamma(\boldsymbol{\eta})}(\boldsymbol{\xi}_n)$ quite comfortable where we recall that $\pi_{\Gamma(\boldsymbol{\eta})}$ and $\rho_{\Gamma(\boldsymbol{\eta})}$ appear in the acceptance probability of the local pCN and the local gpCN Metropolis, respectively. For example, given a $\boldsymbol{\eta} \in \mathbb{R}^N$ and $H_{\Gamma(\boldsymbol{\eta})} = C^{1/2}\Gamma(\boldsymbol{\eta})C^{1/2} = V\Lambda V^\top$ with $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ and orthogonal $V \in \mathbb{R}^{N\times N}$, we have

$$\begin{aligned} A_{\Gamma(\boldsymbol{\eta})} &= C^{1/2}\sqrt{I - s^2\left(I + H_{\Gamma(\boldsymbol{\eta})}\right)^{-1}}C^{-1/2} \\ &= C^{1/2}\,V\sqrt{I - s^2\operatorname{diag}\left((1+\lambda_1)^{-1}, \dots (1+\lambda_n)^{-1}\right)}V^\top C^{-1/2} \end{aligned}$$

and

$$\det(I + H_{\Gamma(\boldsymbol{\eta})}) = \exp\left(\sum_{i=1}^{n}(1+\lambda_i)\right).$$

Remark 5.58 (On additional computational work.)**.** The computational work required for the local pCN and local gpCN Metropolis algorithms is quite large compared to their nonlocal counterparts. The most dramatic additional computations in case of the local pCN Metropolis are the evaluation of $\nabla G(\boldsymbol{\eta})$ — usually at the cost of one forward solve of the adjoint of G — and the computation of the determinant $\det(I + H_{\Gamma(\boldsymbol{\eta})})$ in each iteration. We display an algorithmic description of one transition step of the local pCN Metropolis in Algorithm 5.1. There, Lines 3 to 8 represent the additional work compared to the pCN Metropolis. For the local gpCN Metropolis the computatinal effort is even larger, since also $A_{\Gamma(\cdot)}$ and $\rho_{\Gamma(\cdot)}$ have to be calculated.

In Figure 5.5 we display the estimated ESS for the quantity f_1 resulting simulations with the pCN, gpCN, local pCN and local gpCN Metropolis algorithm. In the

Input : current state $\boldsymbol{\xi}_n$, value $\Phi(\boldsymbol{\xi}_n)$, matrix $\Gamma(\boldsymbol{\xi}_n)$ and determinant $d_{\boldsymbol{\xi}_n} = \det(I + H_{\Gamma(\boldsymbol{\xi}_n)})$

Draw $\boldsymbol{\zeta} \sim N(0, I)$;

Compute proposed state $\boldsymbol{\eta} := \sqrt{1 - s^2}\boldsymbol{\xi}_n + s \cdot \sqrt{C_{\Gamma(\boldsymbol{\xi}_n)}} \cdot \boldsymbol{\zeta}$;

Compute $\nabla G(\boldsymbol{\eta})$;

Compute $\Gamma(\boldsymbol{\eta}) := \sigma_\varepsilon^{-2} \nabla G(\boldsymbol{\eta}) \nabla G(\boldsymbol{\eta})^\top$ and $H_{\Gamma(\boldsymbol{\eta})} := \sqrt{C} \cdot \Gamma(\boldsymbol{\eta}) \cdot \sqrt{C}$;

Compute $d_{\boldsymbol{\eta}} := \det(I + H_{\Gamma(\boldsymbol{\eta})})$;

Set $\Delta_{\boldsymbol{\xi}\boldsymbol{\eta}} := \frac{1}{s}[\boldsymbol{\eta} - A_0\boldsymbol{\xi}_n]$ and $\Delta_{\boldsymbol{\eta}\boldsymbol{\xi}} := [\boldsymbol{\xi}_n - A_0\boldsymbol{\eta}]$;

Compute $q_{\boldsymbol{\xi}\boldsymbol{\eta}} := \pi_{\Gamma(\boldsymbol{\xi}_n)}(\frac{1}{s}[\boldsymbol{\eta} - A_0\boldsymbol{\xi}_n]) = \frac{1}{d_{\boldsymbol{\xi}_n}} \exp\left(\frac{1}{2}\Delta_{\boldsymbol{\xi}\boldsymbol{\eta}}^\top \Gamma(\boldsymbol{\xi}) \Delta_{\boldsymbol{\xi}\boldsymbol{\eta}}\right)$;

Compute $q_{\boldsymbol{\eta}\boldsymbol{\xi}} := \pi_{\Gamma(\boldsymbol{\eta})}(\frac{1}{s}[\boldsymbol{\xi}_n - A_0\boldsymbol{\eta}]) = \frac{1}{d_{\boldsymbol{\eta}}} \exp\left(\frac{1}{2}\Delta_{\boldsymbol{\eta}\boldsymbol{\xi}}^\top \Gamma(\boldsymbol{\eta}) \Delta_{\boldsymbol{\eta}\boldsymbol{\xi}}\right)$;

Evaluate $G(\boldsymbol{\eta})$ and $\Phi(\boldsymbol{\eta}) := \frac{1}{2\sigma_\varepsilon^2}|y - G(\boldsymbol{\eta})|^2$;

Calculate the acceptance probability $\alpha := \min\left\{1, \exp(\Phi(\boldsymbol{\xi}) - \Phi(\boldsymbol{\eta})) \cdot \frac{q_{\boldsymbol{\xi}\boldsymbol{\eta}}}{q_{\boldsymbol{\eta}\boldsymbol{\xi}}}\right\}$.;

Draw $a \sim \mathrm{Uni}(0, 1)$;

if $a < \alpha$ **then**

  $\boldsymbol{\xi}_{n+1} \leftarrow \boldsymbol{\eta}$, $\Phi(\boldsymbol{\xi}_{n+1}) \leftarrow \Phi(\boldsymbol{\eta})$, $\Gamma_{\boldsymbol{\xi}_{n+1}} \leftarrow \Gamma_{\boldsymbol{\eta}}$, $d_{\boldsymbol{\xi}_{n+1}} \leftarrow d_{\boldsymbol{\eta}}$, ;

else

  $\boldsymbol{\xi}_{n+1} \leftarrow \boldsymbol{\xi}_n$, $\Phi(\boldsymbol{\xi}_{n+1}) \leftarrow \Phi(\boldsymbol{\xi}_n)$, $\Gamma_{\boldsymbol{\xi}_{n+1}} \leftarrow \Gamma_{\boldsymbol{\xi}_n}$, $d_{\boldsymbol{\xi}_{n+1}} \leftarrow d_{\boldsymbol{\xi}_n}$, ;

end

Output: next state $\boldsymbol{\xi}_{n+1}$, value $\Phi(\boldsymbol{\xi}_{n+1})$, matrix $\Gamma(\boldsymbol{\xi}_{n+1})$ and determinant $d_{\boldsymbol{\xi}_{n+1}}$

Algorithm 5.1: One transition step of the local pCN Metropolis algorithm.

left panel we fix the state space dimension to $N = 100$ and vary the noise standard deviation $\sigma_\varepsilon = 0.1, 0.05, 0.025, 0.01$, whereas in the right panel we fix $\sigma_\varepsilon = 0.1$ and let $N = 50, 100, 200, 400$. The smaller range of values for varying σ_ε and N is due to the increased computational cost for the local pCN and local gpCN Metropolis. Again, we can observe a dimension independent performance of all algorithms as we expected due to their construction. Moreover, the local gpCN seems to perform as robust as the nonlocal gpCN w.r.t. noise variance whereas the local pCN (as the nonlocal pCN) is strongly affected by decreasing σ_ε. However, in this example we observe that the local versions do not pay off. Actually, they seem to perform worse than their nonlocal counterparts. In particular, the local pCN Metropolis performs quite poorly compared to the other methods. A possible reason why state-dependent proposal covariances do not improve the performance, might be that in this example the resulting posterior measure is close to a Gaussian measure. However, in other situations the application of Metropolis algorithms with state-dependent proposal covariances may be beneficial — if the computational costs of the local pCN and local gpCN Metropolis can be reduced.

Chapter 6

Variance Independence of Metropolis-Hastings Algorithms

In this chapter we investigate and, at least partly, explain the observation made in Section 5.6, i.e., that the performance of the random walk and gpCN Metropolis algorithm — measured in terms of the effective sample size — did not change or only slightly changed for decreasing measurement noise variance when they employ (approximations of) the target covariance for proposing new states. To this end, we consider target measures

$$\mu_\sigma(\mathrm{d}u) \propto \exp\left(-\frac{1}{2\sigma^2}\Phi(u)\right)\,\mu_0(\mathrm{d}u), \qquad \sigma > 0, \tag{6.1}$$

where $\Phi\colon \mathcal{H} \to [0,\infty)$, on the finite-dimensional Hilbert space $\mathcal{H} = \mathbb{R}^n$ and examine the efficiency of MH algorithms targeting μ_σ when σ drops to 0, i.e., when the measure μ_σ becomes more concentrated. The reason for the finite dimensional setting is that the random walk MH algorithm, which we also want to investigate, is not well-defined in infinite dimensions. However, some auxillary results and definitions will be stated for general Hilbert spaces $\mathcal{H}$ and we will also comment on the generalization of the obtained results for the gpCN MH algorithm to infinite dimensions.

Intuitively, a variance independent performance of MH algorithms is rather surprising: for a small σ the target measure μ_σ will concentrate around the manifold $\mathcal{M} := \{u \in \mathbb{R}^n : \Phi(u) = 0\}$ — given that $\min_{u\in\mathbb{R}^n} \Phi(u) = 0$ — and proposed new states are likely to be rejected unless they are close to this manifold (given the current state is also close to $\mathcal{M}$). The usual approach to increase then the acceptance probability is to decrease the proposal variance which, however, leads to smaller moves of the Markov chain and, thus, a higher autocorrelation. Therefore, except when the proposal kernel is informed about the manifold $\mathcal{M}$, we would expect that the efficiency of a MH algorithm to be worse for smaller values of σ.

On the other hand, variance independence or variance robustness (i.e., only small changes in efficiency for decreasing σ) is a desirable property of MH algorithms. Recall the setting of Bayesian inference for an unknown $U \sim \mu_0$ given realizations $y \in \mathbb{R}^d$ of an observable random variable

$$Y = G(U) + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2\Sigma), \tag{6.2}$$

where this time the Gaussian noise ε is scaled by a variance parameter $\sigma > 0$. In this setting the posterior measure of U given $Y = y$ takes a form as in (6.1) with $\Phi(u) = \frac{1}{2}|y - G(u)|^2_{\Sigma^{-1}}$. If the variance parameter σ is small, then the measurements of Y are quite accurate measurements of $G(U)$ and, thus, the realization y is informative about $G(U)$ and U, respectively. In the other case, if σ is relatively large compared to the magnitude of $G(U)$, then the observational data is corrupted by a large noise and we mainly observe realizations of ε, i.e., the observed data y carries less information about the unknown U. Hence, we prefer MH algorithms which also perform well for highly informative data, i.e., small values of σ in (6.2). This motivates the study of the behaviour of MH algorithms applied for approximate sampling of targets μ_σ as in (6.1) for σ decaying to 0.

Besides the work presented here, the only related research on MH algorithms for approximate sampling of increasingly concentrated target measures known to the author is the work by Beskos et al. [16]. In their article Beskos et al. assume that Φ varies only on a linear subspace of $\mathbb{R}^n$, i.e., $\mathbb{R}^n = \mathcal{U}_1 \oplus \mathcal{U}_2$ such that for $u = u_1 + u_2$, $u_i \in \mathcal{U}_i$, there holds (with slight abuse of notation)

$$\Phi(u) = \Phi(u_1, u_2) = \Phi(u_2). \tag{6.3}$$

Thus, the target measure μ_σ will concentrate on the manifold $\mathcal{M} = \{u \in \mathbb{R}^n : \Phi(u_2) = 0\}$ as $\sigma \to 0$. Beskos et al. then investigate optimal scalings for Gaussian random walk proposals $P(u) = N(u, s^2(\sigma)I)$ where $s(\sigma)$ denotes a proposal stepsize which is allowed to depend on σ. By considering diffusion limits of Markov chains they show under some assumptions that the proposal stepsize has to satisfy $s(\sigma) \in \mathcal{O}(\sigma)$ in order to ensure a non-deterioating acceptance rate in the resulting MH algorithm. This yields that the resulting Markov chains will make smaller and smaller steps as $\sigma \to 0$ and, thus, yields an increasing autocorrelation. Hence, such a rescaling does not seem to yield a variance independent or robust performance, e.g., in terms of the effective sample size.

An obvious drawback of the scaled proposals considered by Beskos et al. [16] is that the proposal stepsize decreases also in dimensions which are not affected

by the likelihood Φ, i.e., the subspace $\mathcal{U}_1$. By keeping the proposal stepsize fixed in the dimensions corresponding to $\mathcal{U}_1$ and only rescaling the proposal stepsize in the subspace affected by the likelihood Φ should also lead to non-deterioating Metropolis algorithms. As it turns out, such a modified rescaling is implicitly done by considering random walk proposals $P(u) = N(u, s^2 C_\sigma)$ where C_σ denotes the covariance matrix of the target measure μ_σ.

In this chapter, we will first discuss several concepts for variance independent performance of MH algorithms and then focus on analyzing variance independence of the *expected squared jump distance* of Markov chains generated by MH algorithms. This quantity is often examined in the MCMC literature, see, e.g., [13, 14, 126, 130, 140], since it is easier to estimate than, for instance, spectral gaps or effective samples sizes. In particular, we will prove that for a Gaussian target $\mu_\sigma = N(m_\sigma, C_\sigma)$ the Metropolis algorithms based on the random walk proposal $P(u) = N(u, s^2 C_\sigma)$, $s > 0$, or the gpCN proposal $P(u) = N(A_{\Gamma_\sigma} u, s^2 C_{\Gamma_\sigma})$, $s \in (0, 1)$, with $C_{\Gamma_\sigma} = C_\sigma$, yield a variance independent expected squared jump distance. Of course, a Gaussian target is a rather academic example, but analyzing variance independence in this setting is already non-trivial. Furthermore, we present numerical simulations illustrating our analysis and suggesting that the proven theoretical result also holds for non-Gaussian target measures in certain cases.

6.1. Variance Independent Performance of Metropolis-Hastings Algorithms

To ease the distinction between indexed vectors and their components in the remainder of the chapter, we will denote vectors in $\mathbb{R}^n$ by bold symbols such as $\boldsymbol{u}, \boldsymbol{v}$ and their ith or jth component by u_i or v_j, respectively. However, we denote random vectors as well as real-valued random variables by capital letters such as X and Y. Moreover, we apply the same notation as in Chapter 5, i.e., by K we denote a general Markov kernel as given in Definition 3.11, by M the transition kernel of a Markov chain generated by a MH algorithm as in Definition 5.11, and by P and α the proposal kernel and acceptance probability employed in a MH algorithm, see Definition 5.10.

In the following the target measure μ_σ on $\mathbb{R}^n$ is assumed to be given as in (6.1) for $\sigma > 0$ where $\mu_0 \in \mathcal{P}^2(\mathbb{R}^n)$ denotes an arbitrary reference measure and $\Phi\colon \mathbb{R}^n \to [0, \infty)$ a measurable mapping. We will investigate how MH algorithms perform when σ tends to zero, i.e., how a specific measure of performance for a μ_σ-reversible Metropolis kernel M_σ behaves for $\sigma \to 0$.

6.1.1. Notions of Variance Independent Performance

As already outlined in Chapter 5, there are several measures of efficiency or performance of MH algorithms. Maybe the strongest one is the spectral gap of the associated Markov operator, see Definition 5.28, because this quantity controls the rate of convergence to the limit distribution and also provides an upper bound on the resulting error $\mathbb{E}_\mu[f] - S_{n,n_0}(f)$ of the MCMC integration, see (5.26). Thus, we could ask for MH algorithms which generate μ_σ-reversible Metropolis kernels M_σ such that

$$\lim_{\sigma\to 0} \operatorname{gap}_{\mu_\sigma}(\mathrm{M}_\sigma) \geq \beta > 0, \tag{6.4}$$

i.e., for $\sigma \to 0$ the spectral gap associated with M_σ is bounded from below by a positive constant β.

Another measure of performance is the effective sample size or, equivalently, the integrated autocorrelation time τ_f given a function of interest $f\colon \mathbb{R}^n \to \mathbb{R}$, see Definition 5.8. This quantity was examined in the numerical experiments in Section 5.6. Let $\tau_f(K_\sigma)$ denote the integrated autocorrelation time for $f \in L^2_{\mu_\sigma}(\mathbb{R}^n;\mathbb{R})$ associated with a μ_σ-reversible Markov chain with transition kernel K_σ. Recall, that $\tau_f(K_\sigma)$ can be represented as

$$\tau_f(K_\sigma) = 1 + 2\sum_{k=1}^{\infty} \frac{\langle \mathrm{K}_\sigma^k(f - \mathbb{E}_{\mu_\sigma}(f)), f - \mathbb{E}_{\mu_\sigma}(f)\rangle_{\mu_\sigma}}{\operatorname{Var}_{\mu_\sigma}(f)},$$

see Chapter 5. Then, another notion of variance independent performance of a MH algorithm which generates μ_σ-reversible Metropolis kernels M_σ, could be

$$\lim_{\sigma\to 0} \tau_f(M_\sigma) \leq \beta_f < \infty \qquad \forall f \in \bigcap_{\sigma>0} L^2_{\mu_\sigma}(\mathbb{R}^n;\mathbb{R}), \tag{6.5}$$

where the finite constant β_f is allowed to depend on f. Note, that $\bigcap_{\sigma>0} L^2_{\mu_\sigma}(\mathbb{R}^n;\mathbb{R})$ is not empty, in particular, there holds $L^2_{\mu_0}(\mathbb{R}^n;\mathbb{R}) \subseteq \bigcap_{\sigma>0} L^2_{\mu_\sigma}(\mathbb{R}^n;\mathbb{R})$. Moreover, we mention that (6.5) is weaker than (6.4), since the latter imples a (w.r.t. f) uniform upper bound for $\lim_{\sigma\to 0}\tau_f(M_\sigma)$ due to (5.26).

However, both conditions (6.4) and (6.5) are hard to prove, since spectral gaps and integrated autocorrelation times are not easy to estimate. Therefore, we will settle for something simpler. Namely, we will consider the *expected squared jump distance* (ESJD) of Metropolis kernels and prove lower bounds for these as $\sigma \to 0$. The ESJD is considered by several authors, see, e.g., [13, 14, 126, 130, 140], since it is more convenient to analyze and still a relevant measure of efficiency. In particular, it relates to the lag one autocovariance for the functions $f_i(\boldsymbol{v}) = v_i$, $i = 1,\dots,n$.

Definition 6.1 (Expected squared jump distance)**.** Let $\mu \in \mathcal{P}(\mathbb{R}^n)$ and let K denote a μ-reversible Markov kernel. The *expected squared jump distance* of K is given by

$$\mathrm{ESJD}(K) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |\boldsymbol{u} - \boldsymbol{v}|^2 \, K(\boldsymbol{u}, \mathrm{d}\boldsymbol{v}) \, \mu(\mathrm{d}\boldsymbol{u}). \tag{6.6}$$

Furthermore, we define the *expected squared jump distance in the ith dimension* by

$$\mathrm{ESJD}_i(K) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (u_i - v_i)^2 \, K(\boldsymbol{u}, \mathrm{d}\boldsymbol{v}) \, \mu(\mathrm{d}\boldsymbol{u}), \qquad i = 1, \dots, n. \tag{6.7}$$

We see that $\mathrm{ESJD}(K) = \mathrm{ESJD}_1(K) + \cdots + \mathrm{ESJD}_n(K)$, i.e., a lower bound for $\mathrm{ESJD}_i(K)$ for at least one $i \in \{1, \dots, n\}$ implies a lower bound for $\mathrm{ESJD}(K)$, but if $\mathrm{ESJD}(K)$ is bounded from below, there can still exist a dimension i with $\mathrm{ESJD}_i(K) = 0$. Thus, we will be interested in lower bounds for each $\mathrm{ESJD}_i(K_\sigma)$ of a μ_σ-reversible Markov kernel K_σ as $\sigma \to 0$. The next paragraph provides already some insights on what we can hope for.

The ESJD and the lag one autocorrelation. Let us consider the functions $f_i(\boldsymbol{v}) := v_i$, $i = 1, \dots, n$, and let $(X_k^{(\sigma)})_{k \in \mathbb{N}}$ denote a Markov chain with μ_σ-reversible transition kernel K_σ starting at stationarity $X_1^{(\sigma)} \sim \mu_\sigma$. We recall that for real-valued random variables X and Y, there holds $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y)$. Thus, due to stationarity of the Markov chain there holds

$$\mathrm{Cov}\left(f_i(X_k^{(\sigma)}), f_i(X_{k+1}^{(\sigma)})\right) = \mathrm{Var}_{\mu_\sigma}(f_i) - \frac{1}{2}\,\mathrm{Var}\left(f_i(X_k^{(\sigma)}) - f_i(X_{k+1}^{(\sigma)})\right)$$

and $\mathbb{E}\left[f_i(X_k^{(\sigma)})\right] = \mathbb{E}\left[f_i(X_{k+1}^{(\sigma)})\right]$, i.e., the associated lag one autocorrelation reads as

$$\mathrm{Corr}\left(f_i(X_k^{(\sigma)}), f_i(X_{k+1}^{(\sigma)})\right) = 1 - \frac{\mathrm{ESJD}_i(K_\sigma)}{2\,\mathrm{Var}_{\mu_\sigma}(f_i)}. \tag{6.8}$$

This yields, in particular, that

$$\mathrm{ESJD}_i(K_\sigma) \leq 4\,\mathrm{Var}_{\mu_\sigma}(f_i),$$

which in turn implies that there exists no positive lower bound for $\mathrm{ESJD}_i(K_\sigma)$ if $\mathrm{Var}_{\mu_\sigma}(f_i) \to 0$ for $\sigma \to 0$, i.e., if the marginal of μ_σ in the ith dimension converges to a Dirac distribution. Thus, the best we can hope for is that $\mathrm{ESJD}_i(K_\sigma)$ decays not faster than $\mathrm{Var}_{\mu_\sigma}(f_i)$ as $\sigma \to 0$, i.e., that there exist a $\beta > 0$ such that

$$\lim_{\sigma \to 0} \frac{\mathrm{ESJD}_i(K_\sigma)}{\mathrm{Var}_{\mu_\sigma}(f_i)} \geq \beta \qquad \forall i = 1, \dots, n. \tag{6.9}$$

This seems to be a reasonable and, moreover, feasible concept of a variance independent ESJD which we will pursue in the remainder of the chapter. Obviously (6.9) is equivalent to

$$\lim_{\sigma\to 0} \operatorname{Corr}\left(f_i(X_k^{(\sigma)}), f_i(X_{k+1}^{(\sigma)})\right) \leq 1 - \frac{\beta}{2} \qquad \forall i = 1, \ldots, n, \tag{6.10}$$

i.e., that the limit for $\sigma \to 0$ of the corresponding lag one autocorrelations for the functions f_i is bounded away from 1. This, in turn, has some interesting implications: if the lag one autocorrelation is smaller than one, then this also holds for autocorrelations of larger lags under some additional assumptions:

Proposition 6.2. Let K be a μ-reversible Markov kernel on a separable Hilbert space $\mathcal{H}$ with positive associated Markov operator K. Then for any function $f \in L^2_\mu(\mathcal{H})$ there holds for $k, l \in \mathbb{N}$ with $k \leq l$ that

$$\langle \mathrm{K}^l f, f\rangle_\mu \leq \langle \mathrm{K}^k f, f\rangle_\mu.$$

Hence, for a Markov chain $(X_k)_{k\in\mathbb{N}}$ with transition kernel K starting at $X_1 \sim \mu$ the autocorrelation function $k \mapsto \operatorname{Corr}(f(X_1), f(X_{1+k}))$ is nonincreasing.

Proof. Since K is self-adjoint there exists a spectral measure $E\colon \mathcal{B}(\mathcal{H}) \to \mathcal{L}(L^2_\mu(\mathcal{H}))$ such that

$$\langle \mathrm{K}^k f, f\rangle_\mu = \int_{\operatorname{spec}(\mathrm{K}\,|\,L^2_\mu(\mathcal{H}))} \lambda^k \, E_f(\mathrm{d}\lambda),$$

where $E_f\colon \mathcal{B}(\mathcal{H}) \to [0,\infty)$ denotes the resulting pushforward measure $E_f(\cdot) := \langle E(\cdot) f, f\rangle_\mu$, see, e.g., Halmos [83]. Since K is positive, we have

$$\operatorname{spec}(\mathrm{K}\,|\,L^2_\mu(\mathcal{H})) \subseteq [0,1]$$

and the first statement follows by $\lambda^l \leq \lambda^k$ for $\lambda \in [0,1]$ and $k \leq l$. The second statement is a consequence of the representation (5.24) for the autocovariances and applying the first assertion to $f - \mathbb{E}_\mu[f]$. □

Thus, under the assumptions of Proposition 6.2 we get the implication

$$\frac{\mathrm{ESJD}_i(K_\sigma)}{\operatorname{Var}_{\mu_\sigma}(f_i)} \geq \beta > 0 \quad \Rightarrow \quad \operatorname{Corr}\left(f_i(X_1^{(\sigma)}), f_i(X_{1+k}^{(\sigma)})\right) \leq 1 - \frac{\beta}{2} \quad \forall k \geq 1.$$

Unfortunately, the right-hand side does not yield any upper bound for the associated integrated autocorrelation time τ_{f_i} nor does (6.9) or (6.10) imply a lower bound for the spectral gap as in (6.4). However, we state the following two relations between spectral gaps and lag one autocorrelations.

Proposition 6.3. For $\sigma \to 0$ let μ_σ be given by (6.1), let $\{K_\sigma\}_{\sigma>0}$ denote a family of μ_σ-reversible Markov kernels with positive associated Markov operators K_σ and let $(X_k^{(\sigma)})_{k\in\mathbb{N}}$ denote a Markov chain with transition kernel K_σ starting at $X_1^{(\sigma)} \sim \mu_\sigma$. Then if $\lim_{\sigma\to 0} \operatorname{gap}_{\mu_\sigma}(\mathrm{K}_\sigma) > 0$, we get for any $f \in L^2_{\mu_0}(\mathbb{R}^n)$ with $\operatorname{Var}_{\mu_\sigma}(f) > 0$ for all $\sigma > 0$ that

$$\lim_{\sigma\to 0} \operatorname{Corr}\left(f(X_k^{(\sigma)}), f(X_{k+1}^{(\sigma)})\right) < 1.$$

Equivalently, if there exists a function $f \in L^2_{\mu_0}(\mathbb{R}^n)$ such that $\operatorname{Var}_{\mu_\sigma}(f) > 0$ for all $\sigma > 0$ and

$$\lim_{\sigma\to 0} \operatorname{Corr}\left(f(X_k^{(\sigma)}), f(X_{k+1}^{(\sigma)})\right) = 1,$$

then $\lim_{\sigma\to 0} \operatorname{gap}_{\mu_\sigma}(\mathrm{K}_\sigma) = 0$.

Proof. We prove the first assertion. Let $f \in L^2_{\mu_0}(\mathbb{R}^n)$ and set

$$\bar{f}_\sigma := \frac{f - \mathbb{E}_{\mu_\sigma}[f]}{\operatorname{Var}_{\mu_\sigma}(f)^{1/2}}.$$

Thus, we have $\mathbb{E}_{\mu_\sigma}\left[\bar{f}_\sigma\right] = 0$ and $\|\bar{f}_\sigma\|_{L^2_{\mu_\sigma}} = \operatorname{Var}_{\mu_\sigma}(\bar{f}_\sigma) = 1$ and, in particular, $\bar{f}_\sigma \in L^2_{\mu_\sigma,0}(\mathbb{R}^n)$. Then, we get with $\mathrm{K}_\sigma^{1/2}$ denoting the self-adjoint root operator of K_σ

$$\begin{aligned}
\operatorname{Corr}\left(f(X_k^{(\sigma)}), f(X_{k+1}^{(\sigma)})\right) &= \frac{\langle \mathrm{K}_\sigma(f - \mathbb{E}_{\mu_\sigma}[f]), f - \mathbb{E}_{\mu_\sigma}[f]\rangle_{\mu_\sigma}}{\langle \mathrm{f} - \mathbb{E}_{\mu_\sigma}[f], f - \mathbb{E}_{\mu_\sigma}[f]\rangle_{\mu_\sigma}} \\
&= \langle \mathrm{K}_\sigma \bar{f}_\sigma, \bar{f}_\sigma\rangle_{\mu_\sigma} = \|\mathrm{K}_\sigma^{1/2}\bar{f}_\sigma\|^2_{L^2_{\mu_\sigma}} \\
&\leq \|\mathrm{K}_\sigma^{1/2}\|^2_{\mu_\sigma} = \|\mathrm{K}_\sigma\|_{\mu_\sigma} = 1 - \operatorname{gap}_{\mu_\sigma}(\mathrm{K}_\sigma)
\end{aligned}$$

which yields the first assertion. □

Variance independence of the ESJD. In the subsequent definition of variance independence of the ESJD we will allow for a change of basis or coordinate systems, respectively. This is mainly motivated by the increased flexibility for verifying the resulting condition in practice. We will provide an interpretation of the definition afterwards, but first we introduce the concept of the pushforward Markov kernel as an analogue to the pushforward measure.

Definition 6.4 (Pushforward Markov kernel)**.** Let $K\colon \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0,1]$ be a Markov kernel on a separable Hilbert space $\mathcal{H}$ and let $T\colon (\mathcal{H}, \mathcal{B}(\mathcal{H})) \to (S, \mathcal{S})$ denote a bijective and measurable mapping to another measurable space. Then the *pushforward Markov kernel* $T_*K\colon S \times \mathcal{S} \to [0,1]$ of K under T is defined by

$$T_*K(x, A) := K(T^{-1}(x), T^{-1}(A)), \qquad x \in S,\ A \in \mathcal{S}.$$

We note that if K is μ-reversible, then T_*K is $T_*\mu$-reversible, since for $A, B \in \mathcal{S}$ there holds

$$\begin{aligned}\int_A \int_B T_*K(x, \mathrm{d}y)\, T_*\mu(\mathrm{d}x) &= \int_{T^{-1}(A)} K(u, T^{-1}(B))\, \mu(\mathrm{d}u) \\ &= \int_{T^{-1}(B)} K(u, T^{-1}(A))\, \mu(\mathrm{d}u) \\ &= \int_B \int_A T_*K(x, \mathrm{d}y)\, T_*\mu(\mathrm{d}x).\end{aligned}$$

In particular, we have for $\mathcal{H} = \mathbb{R}^n$ and a regular $T \in \mathbb{R}^{n\times n}$ that

$$\begin{aligned}\mathrm{ESJD}(T_*K) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |\boldsymbol{u} - \boldsymbol{v}|^2\, T_*K(\boldsymbol{u}, \mathrm{d}\boldsymbol{v})\, T_*\mu(\mathrm{d}\boldsymbol{u}) \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |T\boldsymbol{u} - T\boldsymbol{v}|^2\, K(\boldsymbol{u}, \mathrm{d}\boldsymbol{v})\, \mu(\mathrm{d}\boldsymbol{u}).\end{aligned}$$

We are now ready to state

Definition 6.5 (Variance independent ESJD)**.** Let $\mu_\sigma \in \mathcal{P}(\mathbb{R}^n)$ be given as in (6.1). A MH algorithm which generates μ_σ-reversible Metropolis kernels M_σ yields a *variance independent ESJD* if there exists a regular matrix $T \in \mathbb{R}^{n\times n}$ and a constant $\beta > 0$, both independent of σ, such that for $f_i(\boldsymbol{v}) := v_i$, $i = 1, \ldots, n$ there holds

$$\lim_{\sigma\to 0} \frac{\mathrm{ESJD}_i(T_*M_\sigma)}{\mathrm{Var}_{T_*\mu_\sigma}(f_i)} \geq \beta \qquad \forall i = 1, \ldots, n. \tag{6.11}$$

As motivated above, the matrix T in Definition 6.5 represents a change of basis, i.e., the condition (6.11) means that the ESJD in the ith coordinate of the resulting Markov chain $(TX_k^{(\sigma)})_{k\in\mathbb{N}}$ does not decay faster than the associated posterior marginal variance. Of course, Definition 6.5 depends on the choice of T, but by (6.8) and Proposition 6.3 we have for each regular $T \in \mathbb{R}^{n\times n}$ that

$$\lim_{\sigma\to 0} \mathrm{gap}_{\mu_\sigma}(\mathrm{M}_\sigma) > 0 \quad \Rightarrow \quad \max_{i=1,\ldots,n} \lim_{\sigma\to 0} \frac{\mathrm{ESJD}_i(T_*M_\sigma)}{\mathrm{Var}_{T_*\mu_\sigma}(f_i)} > 0.$$

Moreover, it seems likely that, if (6.11) holds for a specific regular $T \in \mathbb{R}^{n\times n}$, then it may hold also for (many) other changes of basis. However, we were not able to prove such statements, see the following remark for some comments.

Remark 6.6 (On a stronger notion of variance independent ESJD via Rayleigh quotients)**.** In order to get rid of the dependence on the choice of T in Definition 6.5, one can require (6.11) to hold for any regular matrix $T \in \mathbb{R}^{n\times n}$. This, in turn, is

equivalent to

$$\frac{\boldsymbol{v}^\top \operatorname{Cov}(X_k^{(\sigma)}, X_{k+1}^{(\sigma)})\boldsymbol{v}}{\boldsymbol{v}^\top \operatorname{Cov}(X_k^{(\sigma)})\boldsymbol{v}} \leq 1-\beta \qquad \forall \boldsymbol{v} \in \mathbb{R}^n, \tag{6.12}$$

where $X_k^{(\sigma)}$ denotes again the kth state of a Markov chain with transition kernel M_σ starting at stationarity $X_1^{(\sigma)} \sim \mu_\sigma$. The quotient on the left-hand side in (6.12) is a Rayleigh quotient $R(\boldsymbol{v})$ of the, in general, nonsymmetric matrix $\operatorname{Cov}(X_k^{(\sigma)}, X_{k+1}^{(\sigma)}) \in \mathbb{R}^{n\times n}$ w.r.t. the inner product induced by $\operatorname{Cov}(X_k^{(\sigma)})$. Thus, (6.12) is equivalent to bounding the largest eigenvalue $\lambda_{\max} = \lambda_{\max}(\sigma)$ of the generalized eigenproblem

$$\frac{1}{2}\left(\operatorname{Cov}(X_k^{(\sigma)}, X_{k+1}^{(\sigma)}) + \operatorname{Cov}(X_k^{(\sigma)}, X_{k+1}^{(\sigma)})^\top\right)\boldsymbol{v} = \lambda \operatorname{Cov}(X_k^{(\sigma)})\boldsymbol{v}$$

by $1-\beta$. However, we were not able to obtain theoretical results for this stronger notion of variance independent ESJD and, therefore, leave it for future research.

The expected acceptance probability. We consider another common measure for the performance of MH algorithms:

Definition 6.7 (Expected acceptance probability)**.** Let $\mu \in \mathcal{P}(\mathbb{R}^n)$ and let M denote a μ-reversible Metropolis kernel with proposal kernel P and acceptance probability α. The *expected acceptance probability* (EAP) of M is defined as

$$\operatorname{EAP}(M) := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \alpha(\boldsymbol{u},\boldsymbol{v})\, P(\boldsymbol{u}, \mathrm{d}\boldsymbol{v})\, \mu(\mathrm{d}\boldsymbol{u}).$$

Hence, a corresponding notion of variance independent performance is given by

Definition 6.8 (Variance independent EAP)**.** Let $\mu_\sigma \in \mathcal{P}(\mathbb{R}^n)$ be given as in (6.1). A MH algorithm generating μ_σ-reversible Metropolis kernels M_σ yields a *variance independent EAP* if

$$\lim_{\sigma\to 0} \operatorname{EAP}(M_\sigma) = \beta > 0. \tag{6.13}$$

Let P_σ and α_σ denote the proposal kernel and the acceptance probability, respectively, of a μ_σ-reversible Metropolis kernel M_σ on $\mathbb{R}^n$, then

$$\operatorname{ESJD}_i(M_\sigma) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (u_i - v_i)^2\, \alpha_\sigma(\boldsymbol{u},\boldsymbol{v})\, P_\sigma(\boldsymbol{u}, \mathrm{d}\boldsymbol{v})\, \mu_\sigma(\mathrm{d}\boldsymbol{u}). \tag{6.14}$$

Hence, if $\operatorname{EAP}(M_\sigma)$ decays to zero as $\sigma \to 0$, then so will $\operatorname{ESJD}(M_\sigma)$ in usual situations: for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ from any bounded subset of $\mathbb{R}^n \times \mathbb{R}^n$ the squared jumpsize $|\boldsymbol{u}-\boldsymbol{v}|^2$ is also bounded and the only possibility that then $\operatorname{ESJD}(M_\sigma) \not\to 0$ is that larger and larger jumps become more probable as $\sigma \to 0$. The latter seems rather

unlikely, since this would at least require an increasing proposal variance as well as that high probability regions of μ_σ are drifting further apart from each other as $\sigma \to 0$. Moreover, by the same reasoning a lower bound for $\mathrm{ESJD}_i(M_\sigma)$ as $\sigma \to 0$ should usually imply a variance independent EAP.

Thus, we conjecture, that under suitable assumptions a variance independent ESJD as defined in (6.11) implies a variance independent EAP as given in (6.13), but so far we could not to prove this conjecture rigorously. On the other hand, as we will see in the next section, condition (6.13) does, in general, not ensure that condition (6.11) holds. Hence, we may say, that (6.13) is a weaker notion of variance independent performance than (6.11). Moreover, by applying Cheeger's inequality, we can state

Proposition 6.9. For $\sigma > 0$ let μ_σ be given by (6.1) and let M_σ denote a μ_σ-reversible Metropolis kernel. If (6.13) does not hold, then

$$\lim_{\sigma \to 0} \mathrm{gap}_{\mu_\sigma}(\mathrm{M}_\sigma) = 0.$$

Proof. We recall the definition of the conductance $\varphi(M_\sigma)$ of M_σ and obtain

$$\varphi(M_\sigma) = \inf_{\mu_\sigma(A) \in (0,1/2]} \frac{\int_A \int_{A^c} \alpha_\sigma(\boldsymbol{u}, \boldsymbol{v})\, P_\sigma(\boldsymbol{u}, \mathrm{d}\boldsymbol{v}) \mu_\sigma(\mathrm{d}\boldsymbol{u})}{\mu_\sigma(A)} \leq 2\, \mathrm{EAP}(M_\sigma)$$

which yields by Theorem 5.31 that

$$\mathrm{gap}_{\mu_\sigma}(\mathrm{M}_\sigma) \leq 4\, \mathrm{EAP}(M_\sigma).$$

The assertion follows. □

Remark 6.10. Summarizing, we have discussed four notions of variance independent performance of MH algorithms in this subsection, which were based on spectral gaps (6.4), integrated autocorrelation times (6.5), ESJD or lag one autocorrelation (6.11) and EAP (6.13). Given $\mu_0 \in \mathcal{P}^2(\mathbb{R}^n)$ they are related as follows:

$$(6.4) \;\Rightarrow\; (6.5) \;\Rightarrow\; (6.11), \qquad (6.4) \;\Rightarrow\; (6.13).$$

6.1.2. Main Result on Variance Independent ESJD for Gaussian Target Measure

For the following result we require that the reference measure μ_0 and the target measure μ_σ are Gaussian:

Assumption 6.11. Let μ_0 and Φ in (6.1) be given by $\mu_0 = N(0, C)$, with a regular covariance matrix $C \in \mathbb{R}^{n\times n}$, and

$$\Phi(\boldsymbol{u}) = \frac{1}{2}|\boldsymbol{y} - G\boldsymbol{u}|^2_{\Sigma^{-1}}, \qquad \boldsymbol{u} \in \mathbb{R}^n,$$

for a $\boldsymbol{y} \in \mathbb{R}^d$, $d \leq n$, a matrix $G \in \mathbb{R}^{d\times n}$, and a symmetric, positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$, respectively. Moreover, we define $r := \operatorname{rank}(G) \leq d$.

Assumption 6.11 implies by Theorem 4.3 that the target μ_σ given by (6.1) is also Gaussian, i.e., $\mu_\sigma = N(\boldsymbol{m}_\sigma, C_\sigma)$ with

$$\boldsymbol{m}_\sigma := CG^\top(GCG^\top + \sigma^2\Sigma)^{-1}\boldsymbol{y} = \sigma^{-2}\left(C^{-1} + \sigma^{-2}G^\top\Sigma^{-1}G\right)^{-1}G^\top\Sigma^{-1}\boldsymbol{y} \tag{6.15}$$

and

$$C_\sigma := (C^{-1} + \sigma^{-2}G^\top\Sigma^{-1}G)^{-1}. \tag{6.16}$$

Theorem 6.12. Let Assumption 6.11 be satisfied and for $\sigma > 0$ let μ_σ be given by (6.1). Then there holds:

- The random walk Metropolis algorithm that generates μ-reversible Metropolis kernels M_σ with proposal kernel

  $$P_\sigma(\boldsymbol{u}) := N(\boldsymbol{u}, s^2C_\sigma),$$

  with C_σ as in (6.16) and $s > 0$, yields a variance independent ESJD as in (6.11) and a variance independent EAP as in (6.13). In both cases the resulting limit β depends only on s and the state space dimension n.

- The gpCN Metropolis algorithm that generates μ-reversible Metropolis kernels M_{Γ_σ} with gpCN proposal kernel

  $$P_{\Gamma_\sigma}(\boldsymbol{u}) := N(A_{\Gamma_\sigma}\boldsymbol{u}, s^2C_{\Gamma_\sigma}),$$

  where $\Gamma_\sigma = \sigma^{-2}G^\top\Sigma^{-1}G$, i.e., $C_{\Gamma_\sigma} = C_\sigma$, and $s \in (0, 1)$, yields also a variance independent ESJD as in (6.11) and a variance independent EAP as in (6.13). This time, the resulting limit β depends in both cases only on s and the rank r of G.

Moreover, if $r < n$, then we even have

$$\lim_{\sigma\to 0}\operatorname{ESJD}(M_\sigma) > 0, \qquad \lim_{\sigma\to 0}\operatorname{ESJD}(M_{\Gamma_\sigma}) > 0.$$

We highlight, that for the gpCN Metropolis the lower bound β in (6.11) and (6.13) is independent of the state space dimension n — which is not the case for the random walk Metropolis M_σ defined in Theorem 6.12. This is a consequence of the reversibility of the gpCN proposal w.r.t. the reference (or prior) measure μ_0.

We will prove Theorem 6.12 in Section 6.3. In the next section we will construct the basis $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ of $\mathbb{R}^n$, respectively the regular matrix $T \in \mathbb{R}^{n\times n}$, w.r.t. which we will verify condition (6.11).

Remark 6.13 (On negative results for scaled Gaussian proposals)**.** We show that for the scaled Gaussian random walk proposals $\widetilde{P}_\sigma(\boldsymbol{u}) = N(\boldsymbol{u}, s^2(\sigma)I)$ with $s(\sigma) = \sigma s_0$, $s_0 > 0$, as examined by Beskos et al. [16] there holds no variance independence of the ESJD under Assumption 6.11 with $r < n$. Let $\widetilde{M}_\sigma$ denote the μ_σ-reversible Metropolis kernel resulting from the proposal $\widetilde{P}_\sigma$. Then for any regular matrix $T \in \mathbb{R}^{n\times n}$ we get

$$\begin{aligned}
\mathrm{ESJD}(T_*\widetilde{M}_\sigma) &\le \|T\|\,\mathrm{ESJD}(\widetilde{M}_\sigma) \\
&= \|T\| \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |\boldsymbol{u}-\boldsymbol{v}|^2\, \widetilde{\alpha}_\sigma(\boldsymbol{u},\boldsymbol{v})\, \widetilde{P}_\sigma(\boldsymbol{u}, \mathrm{d}\boldsymbol{v})\, \mu_\sigma(\mathrm{d}\boldsymbol{u}) \\
&\le \|T\| \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |\boldsymbol{u}-\boldsymbol{v}|^2\, \widetilde{P}_\sigma(\boldsymbol{u}, \mathrm{d}\boldsymbol{v})\, \mu_\sigma(\mathrm{d}\boldsymbol{u}) \\
&= \|T\| \int_{\mathbb{R}^n} n\, s_0^2\, \sigma^2 \mu_\sigma(\mathrm{d}\boldsymbol{u}) = n\,\|T\|\, s_0^2\, \sigma^2,
\end{aligned}$$

i.e., $\mathrm{ESJD}_i(T_*\widetilde{M}_\sigma) \to 0$ as $\sigma \to 0$ for each $i = 1, \dots, n$. On the other hand, if Assumption 6.11 is satisfied with $r < n$, then $\mathrm{rank}(G^\top \Sigma^{-1} G) \le r < n$. As we will see in the next section, particularly Proposition 6.14, this yields that the trace of the resulting posterior covariance C_σ will not become 0 as $\sigma \to 0$, i.e., there exists a $\beta > 0$ such that $\mathrm{tr}\,(C_\sigma) \to \beta$ as $\sigma \to 0$. Then, for any regular $T \in \mathbb{R}^{n\times n}$ we obtain for the trace of the covariance matrix $TC_\sigma T^\top$ of the resulting pushfoward measure $T_*\mu_\sigma$ that

$$\begin{aligned}
\mathrm{tr}\,(TC_\sigma T^\top) &= \sum_{i=1}^n \mathrm{Var}([TX^{(\sigma)}]_i) = \mathbb{E}\left[\left|T(X^{(\sigma)} - \mathbb{E}\left[X^{(\sigma)}\right])\right|^2\right] \\
&\ge \frac{1}{\|T^{-1}\|}\,\mathbb{E}\left[\left|X^{(\sigma)} - \mathbb{E}\left[X^{(\sigma)}\right]\right|^2\right] = \frac{\mathrm{tr}\,(C_\sigma)}{\|T^{-1}\|}
\end{aligned}$$

where $X^{(\sigma)} \sim \mu_\sigma$. Moreover, since $\lim_{\sigma\to 0} \mathrm{tr}\,(C_\sigma) = \beta > 0$, this implies that there exists at least one $i \in \{1, \dots, n\}$ such that $\mathrm{Var}([TX^{(\sigma)}]_i) \ge \beta/\|T^{-1}\| > 0$. Thus, for this particular i we have

$$\lim_{\sigma\to 0} \frac{\mathrm{ESJD}_i(T_*\widetilde{M}_\sigma)}{\mathrm{Var}_{T_*\mu_\sigma}(f_i)} = 0,$$

i.e., condition (6.11) is not satisfied by the Metropolis kernels $\widetilde{M}_\sigma$ for any $T \in \mathbb{R}^{n\times n}$.

6.2. Intrinsic Structure of a Gaussian Target Measure and Its Implications

The fact that the forward map G is linear and the target measure μ_σ is again Gaussian allows us to exploit the structure of its covariance matrix C_σ for constructing an appropriate basis of $\mathbb{R}^n$ and, thus, an appropriate regular matrix $T \in \mathbb{R}^{n\times n}$ to prove our main result.

We follow an approach due to Cui et al. [37] and Spantini et al. [161] and consider, motivated by the structure of the inverse of C_σ,

$$C_\sigma^{-1} = C^{-1} + \sigma^{-2}H, \qquad H := G^\top \Sigma^{-1} G, \tag{6.17}$$

the C^{-1}-orthonormal system $\{\boldsymbol{v}_j : j \in \mathbb{N}\}$ consisting of the eigenvectors of the generalized eigenproblem

$$H\boldsymbol{v} = \lambda C^{-1}\boldsymbol{v}. \tag{6.18}$$

The vectors $\boldsymbol{v}_j$ are given by $\boldsymbol{v}_j = C^{1/2}\tilde{\boldsymbol{v}}_j$ where $(\lambda_j, \tilde{\boldsymbol{v}}_j)$ are the eigenpairs of the operator $C^{1/2}HC^{1/2}$, i.e.,

$$C^{1/2}HC^{1/2}\tilde{\boldsymbol{v}}_j = \lambda_j \tilde{\boldsymbol{v}}_j, \qquad j = 1, \dots, n.$$

In the following, we assume an ordering $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$ and $\lambda_j = 0$ for $j > r$, i.e.,

$$\ker H = \ker G = \operatorname{span}(\boldsymbol{v}_j : j > r).$$

By construction we get

$$C_\sigma^{-1}\boldsymbol{v}_j = (C^{-1} + \sigma^{-2}H)\boldsymbol{v}_j = (1 + \sigma^{-2}\lambda_j)C^{-1}\boldsymbol{v}_j.$$

This implies two things: (a) unless C is a scaled identity matrix the vectors $\boldsymbol{v}_j$, $j = 1, \dots, n$, are not the eigenvectors of the target covariance C_σ, but (b) they are a C_σ^{-1}-orthogonal basis of $\mathbb{R}^n$, i.e.,

$$\boldsymbol{v}_j^\top C_\sigma^{-1}\boldsymbol{v}_k = (1 + \sigma^{-2}\lambda_k)\boldsymbol{v}_j^\top C^{-1}\boldsymbol{v}_k = (1 + \sigma^{-2}\lambda_k)\delta_{jk}, \tag{6.19}$$

where δ_{jk} denotes the Kronecker delta. A consequence of the C_σ^{-1}-orthogonality of the vectors $\boldsymbol{v}_j$ is that the Gaussian target measure μ_σ factorizes w.r.t. the projections onto $\operatorname{span}(\boldsymbol{v}_j)$. To make this statement precise, let us define the matrix $T \in \mathbb{R}^{n\times n}$ for the corresponding change of basis. Since the eigenvectors $\{\boldsymbol{v}_j : j = 1, \dots, n\}$ of

(6.18) are a C^{-1}-orthonormal system in $\mathbb{R}^n$, there holds

$$T := [\boldsymbol{v}_1 \dots \boldsymbol{v}_n]^{-1} = \begin{bmatrix} \boldsymbol{v}_1^\top \\ \vdots \\ \boldsymbol{v}_n^\top \end{bmatrix} C^{-1}, \tag{6.20}$$

where $[\boldsymbol{v}_1 \dots \boldsymbol{v}_n]$ denotes an $n \times n$-matrix with column vectors $\boldsymbol{v}_j$, $j = 1, \dots, n$. We note that

$$\boldsymbol{v}_j^\top C^{-1} = \frac{1}{\boldsymbol{v}_j^\top C_\sigma^{-1} \boldsymbol{v}_j} \boldsymbol{v}_j^\top C_\sigma^{-1}, \qquad j = 1, \dots, n,$$

which will prove useful in the following. Furthermore, we introduce the following notation for convenience:

$$\mathbf{v} = (\mathrm{v}_1, \dots, \mathrm{v}_n) := T\boldsymbol{v}, \qquad \boldsymbol{v} \in \mathbb{R}^n,$$

i.e., for a vector $\boldsymbol{v} \in \mathbb{R}^n$ we denote by the corresponding upright bold letter $\mathbf{v}$ the vector of its coordinates w.r.t. the basis $\{\boldsymbol{v}_j : j = 1, \dots, n\}$ consisting of the eigenvectors $\boldsymbol{v}_j$ of (6.18). Since the vectors $\boldsymbol{v}_j$ are, in general, not orthogonal w.r.t. the Euclidean inner product, we have $|\mathbf{v}| = |T\boldsymbol{v}| \neq |\boldsymbol{v}|$. However, there holds

$$\frac{1}{\|T^{-1}\|} |\boldsymbol{v}| \leq |\mathbf{v}| \leq \|T\| \, |\boldsymbol{v}|, \qquad \boldsymbol{v} \in \mathbb{R}^n. \tag{6.21}$$

We now analyze the pushforward measure of the target μ_σ given the change of basis matrix $T \in \mathbb{R}^{n \times n}$ in (6.20).

Proposition 6.14. Let Assumption 6.11 be satisfied and μ_σ be given as in (6.1). Then the pushforward measure $T_*\mu_\sigma$ with T as in (6.20) is given by

$$T_*\mu_\sigma = \bigotimes_{j=1}^{n} N\left(\mathrm{m}_{j,\sigma}, \gamma_{j,\sigma}^2\right), \qquad \mathrm{m}_{j,\sigma} := \frac{\boldsymbol{v}_j^\top G^\top \Sigma^{-1} \boldsymbol{y}}{(\sigma^2 + \lambda_j)}, \quad \gamma_{j,\sigma}^2 := \frac{1}{1 + \sigma^{-2}\lambda_j}, \tag{6.22}$$

where $j = 1, \dots, n$.

Proof. Let $U \sim \mu_\sigma$, then the distribution of TU is $T_*\mu_\sigma$. By Proposition 2.20 $T_*\mu_\sigma$ is again Gaussian with mean $T\boldsymbol{m}_\sigma$ and covariance matrix $TC_\sigma T^\top$ where $\boldsymbol{m}_\sigma$ and C_σ are as in (6.15) and (6.16), respectively. We obtain for the mean $\mathrm{m}_\sigma = (\mathrm{m}_{1,\sigma}, \dots, \mathrm{m}_{n,\sigma}) = T\boldsymbol{m}_\sigma$

$$\mathrm{m}_{j,\sigma} = \frac{\boldsymbol{v}_j^\top C_\sigma^{-1} \boldsymbol{m}_\sigma}{\boldsymbol{v}_j^\top C_\sigma^{-1} \boldsymbol{v}_j} = \frac{\sigma^{-2} \boldsymbol{v}_j^\top C_\sigma^{-1} C_\sigma G^\top \Sigma^{-1} \boldsymbol{y}}{1 + \sigma^{-2}\lambda_j} = \frac{\boldsymbol{v}_j^\top G^\top \Sigma^{-1} \boldsymbol{y}}{\sigma^2 + \lambda_j}.$$

Furthermore, the (j,k)-entry of the matrix $TC_\sigma T^\top$ is given by

$$\frac{\boldsymbol{v}_j^\top C_\sigma^{-1} C C_\sigma^{-1} \boldsymbol{v}_k}{(\boldsymbol{v}_j^\top C_\sigma^{-1} \boldsymbol{v}_j)(\boldsymbol{v}_k^\top C_\sigma^{-1} \boldsymbol{v}_k)} = \frac{(1+\sigma^{-2}\lambda_j)\delta_{jk}}{(1+\sigma^{-2}\lambda_j)(1+\sigma^{-2}\lambda_k)} = \frac{1}{1+\sigma^{-2}\lambda_k}\delta_{jk}.$$

Thus, $TC_\sigma T^\top$ is diagonal, and, hence, the components $[TU]_j$ of TU are independent with $[TU]_j \sim N(\mathrm{m}_{j,\sigma}, \gamma_{j,\sigma}^2)$. □

Recalling that $\lambda_j > 0$ for $j \leq r$ equation (6.22) implies that for $j = 1, \dots, r$, the variance $\gamma_{j,\sigma}$ of the target marginal in u_j will converge to 0 as $\sigma \to 0$ and, thus, the marginal itself will converge to a Dirac measure at $\boldsymbol{v}_j^\top G^\top \Sigma^{-1} \boldsymbol{y}/\lambda_j$. Moreover, we see that for $j > r$ the marginal variance of u_j is equal to 1 due to $\lambda_j = 0$ for $j > r$. In the next paragraph we show that also $\mathrm{m}_{j,\sigma}$ is independent of σ for $j > r$.

Representation of the data y and the mapping Φ. We will state a representation of the data $\boldsymbol{y} \in \mathbb{R}^d$, which appears in Assumption 6.11, w.r.t. a specific basis of $\mathbb{R}^d$. Namely, let $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_d\}$ denote a Σ^{-1}-orthonormal basis of $\mathbb{R}^d$ such that $\mathrm{rg}(G) = \mathrm{span}(\boldsymbol{w}_1, \dots, \boldsymbol{w}_r)$. This yields the following decomposition

$$\boldsymbol{y} = \boldsymbol{y}_G + \boldsymbol{y}_\perp, \qquad \boldsymbol{y}_G \in \mathrm{rg}(G),\ \boldsymbol{y}_\perp \in \mathrm{span}(\boldsymbol{w}_{r+1}, \dots, \boldsymbol{w}_d). \tag{6.23}$$

Moreover, by construction there exists a unique vector $\boldsymbol{u}^\dagger \in \mathbb{R}^n$ satisfying

$$G\boldsymbol{u}^\dagger = \boldsymbol{y}_G, \qquad \boldsymbol{u}^\dagger \in \mathrm{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_r). \tag{6.24}$$

Remark 6.15 (SVD of G)**.** Let us equip $\mathbb{R}^n$ with the inner product w.r.t. C^{-1} and $\mathbb{R}^d$ with the inner product w.r.t. Σ^{-1}. Then the eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are the corresponding right-singular vectors of G, $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}$ the associated singular values and the vectors $\boldsymbol{w}_1, \dots, \boldsymbol{w}_d$ are the corresponding left-singular vectors.

Proposition 6.16. Let Assumption 6.11 be satisfied, T be as in (6.20), $\boldsymbol{m}_\sigma$ as in (6.15) and $\boldsymbol{u}^\dagger$ as in (6.24). Then there holds with $\mathbf{m}_\sigma = T\boldsymbol{m}_\sigma$ and $\mathbf{u}^\dagger = T\boldsymbol{u}^\dagger$

$$\mathrm{m}_{j,\sigma} = \frac{\lambda_j \mathrm{u}_j^\dagger}{\sigma^2 + \lambda_j}, \qquad \forall j = 1, \dots, n,$$

thus, in particular, $\mathrm{m}_{j,\sigma} = 0$ for $j > r$. Moreover, with $\mathbf{u} = T\boldsymbol{u}$, $\boldsymbol{u} \in \mathbb{R}^n$, we have

$$\Phi(\boldsymbol{u}) = \frac{1}{2}|\boldsymbol{y}_\perp|_{\Sigma^{-1}}^2 + \frac{1}{2}\sum_{j=1}^r \lambda_j (\mathrm{u}_j^\dagger - \mathrm{u}_j)^2.$$

Proof. The first assertion can be shown by a straightforward calculation:

$$\begin{aligned}\mathrm{m}_{j,\sigma} &= \frac{\boldsymbol{v}_j^\top G^\top \Sigma^{-1}\boldsymbol{y}}{\sigma^2+\lambda_j} = \frac{\boldsymbol{v}_j^\top G^\top \Sigma^{-1}(\boldsymbol{y}_G+\boldsymbol{y}_\perp)}{\sigma^2+\lambda_j} = \frac{\boldsymbol{v}_j^\top G^\top \Sigma^{-1}\boldsymbol{y}_G}{\sigma^2+\lambda_j} = \frac{\boldsymbol{v}_j^\top G^\top \Sigma^{-1} G\boldsymbol{u}^\dagger}{\sigma^2+\lambda_j} \\ &= \frac{\boldsymbol{v}_j^\top H\boldsymbol{u}^\dagger}{\sigma^2+\lambda_j} = \frac{\lambda_j \boldsymbol{v}_j^\top C^{-1}\boldsymbol{u}^\dagger}{\sigma^2+\lambda_j} = \frac{\lambda_j \mathrm{u}_j^\dagger}{\sigma^2+\lambda_j}\,.\end{aligned}$$

Moreover, since $\boldsymbol{u}^\dagger \in \operatorname{span}(\boldsymbol{v}_1,\dots,\boldsymbol{v}_r)$, there holds for $j > r$ that $\boldsymbol{v}_j C^{-1}\boldsymbol{u}^\dagger = 0$ and, hence, $0 = \mathrm{u}_j^\dagger = \mathrm{m}_{j,\sigma}$. Furthermore, for $\boldsymbol{u} = \sum_{j=1}^n \mathrm{u}_j \boldsymbol{v}_j \in \mathbb{R}^n$ we obtain

$$\begin{aligned}2\Phi(\boldsymbol{u}) &= (\boldsymbol{y}-G\boldsymbol{u})^\top \Sigma^{-1}(\boldsymbol{y}-G\boldsymbol{u}) = \boldsymbol{y}_\perp^\top \Sigma^{-1}\boldsymbol{y}_\perp + (\boldsymbol{u}^\dagger-\boldsymbol{u})^\top H(\boldsymbol{u}^\dagger-\boldsymbol{u}) \\ &= |\boldsymbol{y}_\perp|^2_{\Sigma^{-1}} + \sum_{i,j=1}^n (\mathrm{u}_i^\dagger-\mathrm{u}_i)(\mathrm{u}_j^\dagger-\mathrm{u}_j)\boldsymbol{v}_i^\top H\boldsymbol{v}_j = |\boldsymbol{y}_\perp|^2_{\Sigma^{-1}} + \sum_{j=1}^r \lambda_j(\mathrm{u}_j^\dagger-\mathrm{u}_j)^2,\end{aligned}$$

since $\boldsymbol{v}_i^\top H\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_i^\top C^{-1}\boldsymbol{v}_j = \lambda_j\delta_{ij}$ and $\lambda_j = 0$ for $j > r$. □

An immediate consequence of Proposition 6.16 (and Proposition 6.14) and $\lambda_j = 0$ for $j > r$ is that

$$T_*\mu_\sigma = \bigotimes_{j=1}^r N\left(\frac{\lambda_j \mathrm{u}_j^\dagger}{\sigma^2+\lambda_j}, \frac{\sigma^2}{\sigma^2+\lambda_j}\right) \otimes \bigotimes_{j=r+1}^n N(0,1)\,. \tag{6.25}$$

This special structure of the pushforward target measure is a special case of the assumption made by Beskos et al. [16] where Φ decomposes as in (6.3): here, the *affected* or *informed subspace* is $\mathcal{U}_2 = \operatorname{span}(\boldsymbol{v}_1,\dots,\boldsymbol{v}_r)$. Its C^{-1}-orthogonal complement $\mathcal{U}_1 = \operatorname{span}(\boldsymbol{v}_{r+1},\dots,\boldsymbol{v}_n)$ represents the *unaffected* or *uninformed subspace*, i.e., the subspace where the target marginal actually coincides with the reference marginal. Such a decomposition (6.25) is also exploited in the mentioned works by Cui et al. [37] and Spantini et al. [161]. Furthermore, Proposition 6.16 implies

$$\lim_{\sigma\to 0} \boldsymbol{m}_\sigma = \sum_{j=1}^r \left(\boldsymbol{v}_j^\top \boldsymbol{u}^\dagger\right)\boldsymbol{v}_j = \boldsymbol{u}^\dagger. \tag{6.26}$$

Pushforward proposal kernels given the transformation T. In order to apply the change of variables $\boldsymbol{u} \mapsto T\boldsymbol{u} = \mathbf{u}$ in the proof of Theorem 6.12 we also need a representation of the corresponding pushforward proposal kernels T_*P_σ and $T_*P_{\Gamma_\sigma}$.

Proposition 6.17. Let Assumption 6.11 be satisfied and T defined as in (6.20). Then there holds for the proposal kernel P_σ and the gpCN proposal P_{Γ_σ} as given in The-

orem 6.12 that

$$T_*P_\sigma(\mathbf{u}) = \bigotimes_{j=1}^{n} N\left(\mathrm{u}_j, s^2\gamma_{j,\sigma}^2\right), \qquad T_*P_{\Gamma_\sigma}(\mathbf{u}) = \bigotimes_{j=1}^{n} N\left(a_{j,\sigma}\mathrm{u}_j, s^2\gamma_{j,\sigma}^2\right),$$

where we set $a_{j,\sigma} := \sqrt{1 - s^2\gamma_{j,\sigma}^2}$.

Proof. Recall the notation $\mathbf{u} = T\boldsymbol{u}$ for $\boldsymbol{u} \in \mathbb{R}^n$. Then, since $T_*P(\mathbf{u}) = P(T^{-1}(\mathbf{u})) \circ T^{-1}$ and $P_\sigma(\boldsymbol{u}) = N(\boldsymbol{u}, s^2C_\sigma)$, we can apply the same reasoning as in the proof of Proposition 6.14 and obtain

$$T_*P_\sigma(\mathbf{u}) = P_\sigma(\boldsymbol{u}) \circ T^{-1} = N(T\boldsymbol{u}, s^2TC_\sigma T^\top),$$

where $s^2TC_\sigma T^\top = s^2\,\mathrm{diag}(\gamma_{1,\sigma}^2, \ldots, \gamma_{n,\sigma}^2)$, see the proof of Proposition 6.14 for details. Hence, the first statement is shown. Moreover, for $P_{\Gamma_\sigma}(\boldsymbol{u}) = N(A_{\Gamma_\sigma}\boldsymbol{u}, s^2C_\sigma)$ there holds

$$T_*P_{\Gamma_\sigma}(\mathbf{u}) = P_{\Gamma_\sigma}(\boldsymbol{u}) \circ T^{-1} = N(TA_{\Gamma_\sigma}\boldsymbol{u}, s^2TC_\sigma T^\top).$$

Thus, it remains to show that $TA_{\Gamma_\sigma}\boldsymbol{u} = (a_{1,\sigma}\mathrm{u}_1, \ldots, a_{n,\sigma}\mathrm{u}_n)^\top$. For each $k = 1, \ldots, n$ we get for the eigenvector $\boldsymbol{v}_k$ of (6.18)

$$\begin{aligned}
A_{\Gamma_\sigma}\boldsymbol{v}_k &= C^{1/2}\sqrt{I - s^2(I + \sigma^{-2}C^{1/2}HC^{1/2})^{-1}}C^{-1/2}\boldsymbol{v}_k \\
&= C^{1/2}\sqrt{I - s^2(I + \sigma^{-2}C^{1/2}HC^{1/2})^{-1}}\tilde{\boldsymbol{v}}_k \\
&= C^{1/2}\left(\sqrt{I - s^2(1 + \sigma^{-2}\lambda_k)^{-1}}\tilde{\boldsymbol{v}}_k\right) \\
&= \sqrt{I - s^2(1 + \sigma^{-2}\lambda_k)^{-1}}\boldsymbol{v}_k \\
&= \sqrt{1 - s^2\gamma_{\sigma,k}^2}\boldsymbol{v}_k,
\end{aligned}$$

which yields for $j = 1, \ldots, n$,

$$\boldsymbol{v}_j^\top C^{-1}A_{\Gamma_\sigma}\boldsymbol{u} = \boldsymbol{v}_j^\top C^{-1}A_{\Gamma_\sigma}\left(\sum_{k=1}^{n}\mathrm{u}_k\,\boldsymbol{v}_k\right) = \boldsymbol{v}_j^\top C^{-1}\left(\sum_{k=1}^{n}\mathrm{u}_k\,a_{k,\sigma}\,\boldsymbol{v}_k\right) = a_{j,\sigma}\,\mathrm{u}_j,$$

since $\boldsymbol{v}_j^\top C^{-1}\boldsymbol{v}_k = \delta_{jk}$. Hence, $TA_{\Gamma_\sigma}\boldsymbol{u} = (a_{1,\sigma}\mathrm{u}_1, \ldots, a_{n,\sigma}\mathrm{u}_n)^\top$ for each $\boldsymbol{u} \in \mathbb{R}^n$. □

Thus, Proposition 6.17 shows that the proposal kernels P_σ and P_{Γ_σ} decrease their proposal variance only in the subspace $\mathrm{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$ as $\sigma \to 0$. This is exactly the subspace on which the target measure will concentrate for decreasing σ.

6.3. Proof of the Main Result

The proof is splitted into two parts corresponding to the two proposals P_σ and P_{Γ_σ}. In both parts we will follow a similar approach to Norton and Fox [126, Theorem 3.4 and 4.2] who examine the limit of the EAP and the ESJD of Gaussian proposals for an increasing state space dimension $n \to \infty$. The adapted approach for the first part consists of constructing two n-dimensional random vectors $\mathrm{U}_\sigma = (\mathrm{U}_{1,\sigma}, \dots, \mathrm{U}_{n,\sigma})$ and $\mathrm{V}_\sigma = (\mathrm{V}_{1,\sigma}, \dots, \mathrm{V}_{n,\sigma})$ such that $(\mathrm{U}_\sigma, \mathrm{V}_\sigma)$ follows the distribution $\eta_\sigma^T(\mathrm{d}\mathbf{u}, \mathrm{d}\mathbf{v}) := T_* P_\sigma(\mathbf{u}, \mathrm{d}\mathbf{v})\, T_*\mu_\sigma(\mathrm{d}\mathbf{u})$. In other words, the combined random vector $(\mathrm{U}_\sigma, \mathrm{V}_\sigma)$ models two consecutive states of the Markov chain in $\mathbb{R}^n$ generated by $T_* M_\sigma$ starting at stationarity (here $T_*\mu_\sigma$). Then, we prove the assertion for M_σ by studying

$$\mathrm{ESJD}_i(T_* M_\sigma) = \mathbb{E}\left[|\mathrm{U}_{i,\sigma} - \mathrm{V}_{i,\sigma}|^2\, \alpha_\sigma^T(\mathrm{U}_\sigma, \mathrm{V}_\sigma)\right], \qquad i = 1, \dots, n,$$

where α_σ^T denotes the reformulation of the acceptance probability α given the coordinate transform T. We proceed analogously in the second part.

6.3.1. Proof for the Random Walk Proposal P_σ

Let $p_\sigma(\boldsymbol{u}; \cdot)\colon \mathbb{R}^n \to [0, \infty)$ denote the probability density funtion of the proposal kernel $P_\sigma(\boldsymbol{u}) = N(\boldsymbol{u}, s^2 C_\sigma)$, i.e., a multivariate normal density function. Then, we easily see, that $p_\sigma(\boldsymbol{u}; \boldsymbol{v}) = p_\sigma(\boldsymbol{v}; \boldsymbol{u})$ holds for each $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. This implies, that the resulting acceptance probability α_σ according to (5.8) of the μ_σ-reversible Metropolis kernel M_σ takes the form

$$\alpha_\sigma(\boldsymbol{u}, \boldsymbol{v}) = 1 \wedge \frac{\pi_\sigma(\boldsymbol{v})}{\pi_\sigma(\boldsymbol{u})}, \qquad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n,$$

where $a \wedge b$ denotes $\min(a, b)$ and $\pi_\sigma\colon \mathbb{R}^n \to [0, \infty)$ denotes the probability density function of $\mu_\sigma = N(\boldsymbol{m}_\sigma, C_\sigma)$. Hence,

$$\alpha_\sigma(\boldsymbol{u}, \boldsymbol{v}) = 1 \wedge \exp\left(\frac{1}{2}\left[|\boldsymbol{u} - \boldsymbol{m}_\sigma|^2_{C_\sigma^{-1}} - |\boldsymbol{v} - \boldsymbol{m}_\sigma|^2_{C_\sigma^{-1}}\right]\right), \qquad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n.$$

Furthermore, with $\mathbf{u} = T\boldsymbol{u}$, $\mathbf{v} = T\boldsymbol{v}$ and $\mathbf{m}_\sigma = T\boldsymbol{m}_\sigma$ we get

$$\alpha_\sigma(\boldsymbol{u}, \boldsymbol{v}) = 1 \wedge \exp\left(\frac{1}{2}\sum_{j=1}^n \gamma_{j,\sigma}^{-2}\left[(\mathrm{m}_{j,\sigma} - \mathrm{u}_j)^2 - (\mathrm{m}_{j,\sigma} - \mathrm{v}_j)^2\right]\right) =: \alpha_\sigma^T(\mathbf{u}, \mathbf{v}).$$

Now, let $\mathrm{W} = (\mathrm{W}_1, \dots, \mathrm{W}_n) \sim N(0, I_n)$ and $\mathrm{Z} = (\mathrm{Z}_1, \dots, \mathrm{Z}_n) \sim N(0, I_n)$ be independent. Then, we define two random vectors $\mathrm{U}_\sigma = (\mathrm{U}_{1,\sigma}, \dots, \mathrm{U}_{n,\sigma})$ and

$\mathrm{V}_\sigma = (\mathrm{V}_{1,\sigma}, \dots, \mathrm{V}_{n,\sigma})$ by

$$\mathrm{U}_{j,\sigma} = \mathrm{m}_{j,\sigma} + \gamma_{j,\sigma}\mathrm{W}_j, \quad \mathrm{V}_{j,\sigma} = \mathrm{m}_{j,\sigma} + \gamma_{j,\sigma}\mathrm{W}_j + s\gamma_{j,\sigma}\mathrm{Z}_j, \qquad j = 1, \dots, n,$$

and notice that by construction there holds with $\eta_\sigma^T(\mathrm{d}\mathbf{u}, \mathrm{d}\mathbf{v}) = T_* P_\sigma(\mathbf{u}, \mathrm{d}\mathbf{v})\, T_*\mu_\sigma(\mathrm{d}\mathbf{u})$ that $\mathrm{U}_\sigma \sim T_*\mu_\sigma$ and $(\mathrm{U}_\sigma, \mathrm{V}_\sigma) \sim \eta_\sigma^T$. Thus, we get

$$\begin{aligned}\mathrm{ESJD}_i(T_*M_\sigma) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (\mathrm{u}_i - \mathrm{v}_i)^2 \alpha_\sigma^T(\mathbf{u}, \mathbf{v})\, \eta_\sigma^T(\mathrm{d}\mathbf{u}, \mathrm{d}\mathbf{v}) \\ &= \mathbb{E}\left[|\mathrm{U}_{i,\sigma} - \mathrm{V}_{i,\sigma}|^2\, \alpha_\sigma^T(\mathrm{U}_\sigma, \mathrm{V}_\sigma)\right].\end{aligned}$$

It is easy to see that $|\mathrm{U}_{i,\sigma} - \mathrm{V}_{i,\sigma}|^2 = s^2\gamma_{i,\sigma}^2\mathrm{Z}_i^2$ and

$$\begin{aligned}\alpha_\sigma^T(\mathrm{U}_\sigma, \mathrm{V}_\sigma) &= 1 \wedge \exp\left(\frac{1}{2}\sum_{j=1}^n \gamma_{j,\sigma}^{-2}\left[(\mathrm{m}_{j,\sigma} - \mathrm{U}_{j,\sigma})^2 - (\mathrm{m}_{j,\sigma} - \mathrm{V}_{j,\sigma})^2\right]\right) \\ &= 1 \wedge \exp\left(\frac{1}{2}\sum_{j=1}^n \left[\mathrm{W}_j^2 - (\mathrm{W}_j^2 + s\mathrm{Z}_j)^2\right]\right).\end{aligned}$$

Hence, recalling that $U_\sigma \sim T_*\mu_\sigma$ we get $\mathrm{Var}_{T_*\mu_\sigma}(\mathrm{u}_i) = \mathrm{Var}(U_{i,\sigma}) = \gamma_{i,\sigma}^2$ and

$$\begin{aligned}\frac{\mathrm{ESJD}_i(T_*M_\sigma)}{\mathrm{Var}(U_{i,\sigma})} &= \frac{\gamma_{i,\sigma}^2\mathbb{E}\left[s^2\mathrm{Z}_i^2\,\left(1 \wedge \mathrm{e}^{\frac{1}{2}|\mathrm{W}|^2 - \frac{1}{2}|\mathrm{W}+s\mathrm{Z}|^2}\right)\right]}{\gamma_{i,\sigma}^2} \\ &= s^2\mathbb{E}\left[\mathrm{Z}_i^2\,\left(1 \wedge \mathrm{e}^{\frac{1}{2}|\mathrm{W}|^2 - \frac{1}{2}|\mathrm{W}+s\mathrm{Z}|^2}\right)\right]\end{aligned}$$

where the right-hand side is independent of σ and strictly positive, i.e., we have proven a variance independent ESJD according to Definition 6.5 for the Metropolis kernels M_σ. The calculation above implies, in particular, that

$$\mathrm{ESJD}(T_*M_\sigma) = \sum_{i=1}^n \mathrm{ESJD}_i(T_*M_\sigma) = s^2\mathbb{E}\left[|\mathrm{Z}|^2\,\left(1 \wedge \mathrm{e}^{\frac{1}{2}|\mathrm{W}|^2 - \frac{1}{2}|\mathrm{W}+s\mathrm{Z}|^2}\right)\right] > 0$$

and by (6.21) we get

$$\mathrm{ESJD}(M_\sigma) \geq \frac{\mathrm{ESJD}(T_*M_\sigma)}{\|T\|^2} > 0.$$

Moreover, we easily see that

$$\mathrm{EAP}(M_\sigma) = \mathbb{E}\left[1 \wedge \mathrm{e}^{\frac{1}{2}|\mathrm{W}|^2 - \frac{1}{2}|\mathrm{W}+s\mathrm{Z}|^2}\right] > 0.$$

Thus, we have proven the assertions for M_σ.

6.3.2. Proof for the gpCN Poposal P_{Γ_σ}

For the gpCN proposal $P_{\Gamma_\sigma}(\boldsymbol{u}) = N(A_{\Gamma_\sigma}\boldsymbol{u}, s^2 C_{\Gamma_\sigma})$ we get according to (5.8) the acceptance probability $\alpha_\sigma(\boldsymbol{u},\boldsymbol{v}) = 1 \wedge \exp\left(-\frac{1}{2\sigma^2}(\Phi(\boldsymbol{u}) - \Phi(\boldsymbol{v}))\right)$, $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^n$. With $\mathbf{u} = T\boldsymbol{u}$, $\mathbf{v} = T\boldsymbol{v}$ and $\mathbf{u}^\dagger = T\boldsymbol{u}^\dagger$ this yields due to Proposition 6.16

$$\alpha_\sigma(\boldsymbol{u},\boldsymbol{v}) = 1 \wedge \exp\left(\frac{1}{2\sigma^2}\left[\sum_{j=1}^r \lambda_j(\mathrm{u}_j^\dagger - \mathrm{u}_j)^2 - \sum_{j=1}^r \lambda_j(\mathrm{u}_j^\dagger - \mathrm{v}_j)^2\right]\right) =: \alpha_\sigma^T(\mathbf{u},\mathbf{v}).$$

We proceed analogously to the proof for P_σ and construct two random vectors U_σ and V_σ such that with $\eta_{\Gamma_\sigma}^T(\mathrm{d}\mathbf{u},\mathrm{d}\mathbf{v}) := T_* P_{\Gamma_\sigma}(\mathbf{u},\mathrm{d}\mathbf{v})\, T_*\mu_\sigma(\mathrm{d}\mathbf{u})$ we have

$$\mathrm{U}_\sigma \sim T_*\mu_\sigma, \qquad (\mathrm{U}_\sigma, \mathrm{V}_\sigma) \sim \eta_{\Gamma_\sigma}^T.$$

Such random vectors are given by

$$\mathrm{U}_{j,\sigma} = \mathrm{m}_{j,\sigma} + \gamma_{j,\sigma}\mathrm{W}_j, \qquad \mathrm{V}_{j,\sigma} = a_{j,\sigma}\left(\mathrm{m}_{j,\sigma} + \gamma_{j,\sigma}\mathrm{W}_j\right) + s\gamma_{j,\sigma}\mathrm{Z}_j, \qquad 1 \le j \le n,$$

where again we assumed $\mathrm{W},\mathrm{Z} \sim N(0, I_n)$ independently. Thus, similar to Section 6.3.1 we get

$$\mathrm{ESJD}_i(T_* M_{\Gamma_\sigma}) = \mathbb{E}\left[|\mathrm{U}_{i,\sigma} - \mathrm{V}_{i,\sigma}|^2\, \alpha_\sigma^T(\mathrm{U}_\sigma, \mathrm{V}_\sigma)\right],$$

where now

$$|\mathrm{U}_{i,\sigma} - \mathrm{V}_{i,\sigma}|^2 = |(1-a_{i,\sigma})\mathrm{U}_{i,\sigma} - s\gamma_{i,\sigma}\mathrm{Z}_i|^2 = \gamma_{i,\sigma}^2\left|\frac{1-a_{i,\sigma}}{\gamma_{i,\sigma}}\mathrm{U}_{i,\sigma} - s\mathrm{Z}_i\right|^2.$$

Since $\mathrm{U}_\sigma = (\mathrm{U}_{1,\sigma},\ldots,\mathrm{U}_{n,\sigma}) \sim T_*\mu_\sigma$, i.e., $\mathrm{Var}_{T_*\mu_\sigma}(\mathrm{u}_i) = \mathrm{Var}(\mathrm{U}_{i,\sigma}) = \gamma_{i,\sigma}^2$, we study

$$\begin{aligned}\lim_{\sigma\to 0}\frac{\mathrm{ESJD}_i(T_* M_{\Gamma_\sigma})}{\mathrm{Var}(\mathrm{U}_{i,\sigma})} &= \lim_{\sigma\to 0}\frac{\mathrm{ESJD}_i(T_* M_{\Gamma_\sigma})}{\gamma_{i,\sigma}^2} \\ &= \lim_{\sigma\to 0}\mathbb{E}\left[\left|\frac{1-a_{i,\sigma}}{\gamma_{i,\sigma}}\mathrm{U}_{i,\sigma} - s\mathrm{Z}_i\right|^2 \alpha_\sigma^T(\mathrm{U}_\sigma,\mathrm{V}_\sigma)\right]\end{aligned}$$

in the following. Due to $\gamma_{i,\sigma}^2 = (1+\sigma^{-2}\lambda_i)^{-1}$ and $\lambda_i = 0$ iff $i > r$, we get $\mathbb{P}$-a.s.

$$\lim_{\sigma\to 0}\mathrm{U}_{i,\sigma} = \lim_{\sigma\to 0}\mathrm{m}_{i,\sigma} + \left(\lim_{\sigma\to 0}\gamma_{i,\sigma}\right)\mathrm{W}_i = \begin{cases}\mathrm{u}_i^\dagger, & i \le r,\\ \mathrm{W}_i, & i > r,\end{cases}$$

where we used $\lim_{\sigma\to 0}\mathrm{m}_{i,\sigma} = \mathrm{u}_i^\dagger$, see Proposition 6.16, and $\mathrm{u}_i^\dagger = 0$ for $i > r$. Further,

due to $a_{i,\sigma} = \sqrt{1 - s^2\gamma_{i,\sigma}^2}$ and by applying L'Hôpital's rule, we obtain for $1 \leq i \leq r$

$$\lim_{\sigma\to 0} \frac{1 - a_{i,\sigma}}{\gamma_{i,\sigma}} = \lim_{x\to 0} \frac{1 - \sqrt{1 - s^2x^2}}{x} = \lim_{x\to 0} \frac{\frac{s^2x}{2\sqrt{1-s^2x^2}}}{1} = 0,$$

and for $i > r$ there holds, since then $\gamma_{i,\sigma} = 1$, that

$$\frac{1 - a_{i,\sigma}}{\gamma_{i,\sigma}} = 1 - \sqrt{1 - s^2}.$$

Thus, we obtain $\mathbb{P}$-almost surely

$$\mathrm{S}_i^2 := \lim_{\sigma\to 0} \left| \frac{1 - a_{i,\sigma}}{\gamma_{i,\sigma}} \mathrm{U}_{i,\sigma} - s\gamma_{i,\sigma}\mathrm{Z}_i \right|^2 = \begin{cases} s^2\mathrm{Z}_i^2, & 1 \leq i \leq r, \\ \left((1 - \sqrt{1 - s^2})\mathrm{W}_i - s\mathrm{Z}_i\right)^2, & i > r. \end{cases}$$

Further, we investigate the limit of the acceptance probability and obtain $\mathbb{P}$-a.s.

$$\begin{aligned} \lim_{\sigma\to 0} \alpha_\sigma^T(\mathrm{U}_\sigma, \mathrm{V}_\sigma) &= 1 \wedge \lim_{\sigma\to 0} \exp\left(\frac{1}{2\sigma^2} \sum_{j=1}^r \lambda_j \left[(\mathrm{u}_j^\dagger - \mathrm{U}_{j,\sigma})^2 - (\mathrm{u}_j^\dagger - \mathrm{V}_{j,\sigma})^2 \right] \right) \\ &= 1 \wedge \exp\left(\frac{1}{2} \sum_{j=1}^r \lambda_j \lim_{\sigma\to 0} \left[\frac{(\mathrm{u}_j^\dagger - \mathrm{U}_{j,\sigma})^2}{\sigma^2} - \frac{(\mathrm{u}_j^\dagger - \mathrm{V}_{j,\sigma})^2}{\sigma^2} \right] \right). \end{aligned}$$

Moreover, there holds $\mathbb{P}$-a.s.

$$\lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - \mathrm{U}_{j,\sigma}}{\sigma} = \lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - \mathrm{m}_{j,\sigma} - \gamma_{j,\sigma}\mathrm{W}_j}{\sigma} = \lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - \mathrm{m}_{j,\sigma}}{\sigma} - \left(\lim_{\sigma\to 0} \frac{\gamma_{j,\sigma}}{\sigma} \right) \mathrm{W}_j,$$

where we obtain by Proposition 6.16 for $1 \leq j \leq r$

$$\lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - \mathrm{m}_{j,\sigma}}{\sigma} = \lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - \frac{\lambda_j \mathrm{u}_j^\dagger}{\sigma^2 + \lambda_j}}{\sigma} = \lim_{\sigma\to 0} \frac{\sigma^2 \mathrm{u}_j^\dagger}{\sigma(\sigma^2 + \lambda_j)} = 0$$

and

$$\lim_{\sigma\to 0} \frac{\gamma_{j,\sigma}}{\sigma} = \lim_{\sigma\to 0} \frac{(1 + \sigma^{-2}\lambda_j)^{-1/2}}{\sigma} = \lim_{\sigma\to 0} \frac{1}{\sqrt{\sigma^2 + \lambda_j}} = \lambda_j^{-1/2},$$

recalling that $\lambda_j > 0$ for $j = 1, \ldots, r$. We get that $\mathbb{P}$-a.s.

$$\lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - \mathrm{V}_{j,\sigma}}{\sigma} = \lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - a_{j,\sigma}\mathrm{m}_{j,\sigma}}{\sigma} - \left(\lim_{\sigma\to 0} \frac{a_{j,\sigma}\gamma_{j,\sigma}}{\sigma} \right) \mathrm{Z}_j - \left(\lim_{\sigma\to 0} \frac{\gamma_{j,\sigma}}{\sigma} \right) \mathrm{W}_j.$$

Since for $1 \leq j \leq r$

$$\lim_{\sigma\to 0} \frac{\gamma_{j,\sigma}}{\sigma} = \lambda^{-1/2}, \qquad \lim_{\sigma\to 0} \frac{a_{j,\sigma}\gamma_{j,\sigma}}{\sigma} = \lim_{\sigma\to 0} a_{j,\sigma} \lim_{\sigma\to 0} \frac{\gamma_{j,\sigma}}{\sigma} = \lambda^{-1/2},$$

where $\lim_{\sigma\to 0} a_{j,\sigma} = \lim_{\sigma\to 0} \sqrt{1 - s^2(1+\sigma^{-2}\lambda_j)^{-1}} = 1$, we only need to investigate further

$$\begin{aligned}
\lim_{\sigma\to 0} \frac{\mathrm{u}_j^\dagger - a_{j,\sigma}\mathrm{m}_{j,\sigma}}{\sigma} &= \lim_{\sigma\to 0} \frac{1 - \frac{\lambda_j a_{j,\sigma}}{\sigma^2+\lambda_j}}{\sigma}\mathrm{u}_j^\dagger = \lim_{\sigma\to 0} \frac{\sigma^2 + (1-a_{j,\sigma})\lambda_j}{\sigma(\sigma^2+\lambda_j)}\mathrm{u}_j^\dagger \\
&= \lim_{\sigma\to 0} \frac{1-a_{j,\sigma}}{\sigma(\sigma^2+\lambda_j)}\lambda_j\mathrm{u}_j^\dagger = \frac{\lim_{\sigma\to 0}\frac{1-a_{j,\sigma}}{\sigma}}{\lim_{\sigma\to 0}(\sigma^2+\lambda_j)}\lambda_j\mathrm{u}_j^\dagger \\
&= \left(\lim_{\sigma\to 0} \frac{1-a_{j,\sigma}}{\sigma}\right)\mathrm{u}_j^\dagger.
\end{aligned}$$

By another application of L'Hôpital's rule we obtain

$$\begin{aligned}
\lim_{\sigma\to 0} \frac{1-a_{j,\sigma}}{\sigma} &= \lim_{\sigma\to 0} \frac{1 - \sqrt{1 - s^2(1+\sigma^{-2}\lambda_j)^{-1}}}{\sigma} = \lim_{\sigma\to 0} \frac{\frac{2s^2\sigma^{-3}\lambda_j(1+\sigma^{-2}\lambda_j)^{-2}}{2\sqrt{1-s^2(1+\sigma^{-2}\lambda_j)^{-1}}}}{1} \\
&= \lim_{\sigma\to 0} \frac{s^2\lambda_j\sigma(\sigma^2+\lambda_j)^{-2}}{\sqrt{1-s^2(1+\sigma^{-2}\lambda_j)^{-1}}} = 0.
\end{aligned}$$

In summary, we have

$$\begin{aligned}
\lim_{\sigma\to 0} \alpha_\sigma^T(\mathrm{U}_\sigma, \mathrm{V}_\sigma) &= 1 \wedge \exp\left(\frac{1}{2}\sum_{j=1}^r \lambda_j \lim_{\sigma\to 0}\left[\frac{(\mathrm{u}_j^\dagger - \mathrm{U}_{j,\sigma})^2}{\sigma^2} - \frac{(\mathrm{u}_j^\dagger - \mathrm{V}_{j,\sigma})^2}{\sigma^2}\right]\right) \\
&= 1 \wedge \exp\left(\frac{1}{2}\sum_{j=1}^r (\mathrm{W}_j^2 - (\mathrm{Z}_j + \mathrm{W}_j)^2\right) \qquad \mathbb{P}\text{-a.s.},
\end{aligned}$$

and, hence, by dominated convergence

$$\lim_{\sigma\to 0} \frac{\mathrm{ESJD}_i(T_* M_{\Gamma_\sigma})}{\mathrm{Var}(\mathrm{U}_{i,\sigma})} = \mathbb{E}\left[\mathrm{S}_i^2 \left(1 \wedge \mathrm{e}^{\frac{1}{2}\sum_{j=1}^r (\mathrm{W}_j^2 - (\mathrm{Z}_j+\mathrm{W}_j)^2)}\right)\right],$$

i.e., we have shown that the Markov kernels M_{Γ_σ} have a variance independent ESJD according to Definition 6.5. The statements

$$\lim_{\sigma\to 0} \mathrm{ESJD}(M_{\Gamma_\sigma}) > 0, \qquad \lim_{\sigma\to 0} \mathrm{EAP}(M_{\Gamma_\sigma}) > 0$$

follow with the same reasoning as for M_σ. □

6.4. Numerical Illustrations

In the first part of this section, we choose a Bayesian inference setting which satisfies Assumption 6.11, i.e., the corresponding forward map G is linear. We apply MH algorithms for approximate sampling of the posterior where we choose the same four MH algorithms as in Section 5.6. We then investigate how the resulting acceptance rate, expected squared jump distance and effective sample size of these MH algorithms depend on the noise variance parameter σ. In particular, we will verify the statements of Theorem 6.12 and show that even in terms of the ESS the random walk and gpCN proposal P_σ and P_{Γ_σ}, respectively, seem to perform independently w.r.t. the noise variance.

In the second part we change the forward map G slightly such that it becomes nonlinear and compare again the acceptance rate and mean squared step size of the four MH algorithms. We will proceed as in Section 5.6 and linearize the forward map at the MAPE in order to get an approximation of the target covariance which we then employ for the random walk proposal P_σ and the gpCN proposal P_{Γ_σ}. We choose two stages of nonlinearity. At the first stage we observe simply a nonlinear functional of the observations which we made in the linear case. This construction preserves the existence of an informed and an uninformed subspace as in the first example, in particular, the manifold $\mathcal{M}$ on which the posterior concentrates is again a linear subspace. However, due to the nonlinearity the posterior is non longer Gaussian. At the second stage we will insert a nonlinear mapping before we apply the same linear forward map as in the first example. The resulting nonlinear forward is of such a kind that there exist no longer an uninformed subspace, i.e., a decomposition of Φ as in (6.3) no longer holds. In the nonlinear case we observe the following results:

- For the nonlinearity at the first stage the experimental results remain the same as for the linear forward, i.e., we observe a variance independent ESJD of those Metropolis algorithms which employ, this time, approximations of the posterior covariance in the proposal. This suggests that the Gaussianity of the target measure is not crucial for the statements of Theorem 6.12.

- In case of the second nonlinear example we do not observe a variance independent ESJD for any of the four MH algorithms. However, the two MH algorithms employing an approximation to the target covariance in the proposals show a more robust dependence of the ESJD w.r.t. noise variance than the simple random walk and pCN Metropolis algorithm.

6.4.1. Linear Forward Maps

We consider a convolution operator $A\colon C(0,1) \to C(0,1)$

$$Au(x) := \int_0^1 \exp\left(-\frac{(x-y)^2}{2w^2}\right) f(y)\,\mathrm{d}y, \qquad f \in C(0,1),$$

where we choose $w = 1/20$ and define the forward map $G\colon C(0,1) \to \mathbb{R}^4$ by

$$Gf := [Af(0.2), Af(0.4), Af(0.6), Af(0.8)]^\top.$$

We then would like to infer a random function U in $C(0,1)$ based on noisy observations of

$$Y = GU + \varepsilon_\sigma, \qquad \varepsilon_\sigma \sim N(0, \sigma^2 I_4),$$

where our prior for U is a truncated Brownian motion prior, i.e.,

$$U(x,\omega) = x\,\xi_1(\omega) + \sum_{k=1}^{n-1} \sqrt{2}\sin(k\pi x)\,\xi_{1+k}(\omega),$$

with $\xi_1 \sim N(0,1)$ and $\xi_k \sim N(0,(k-1)^{-2})$, $k > 1$, stochastically independent. Thus, as in Section 5.6 we actually infer $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ given the prior $\mu_0 = N(0,C)$ with covariance matrix $C = \mathrm{diag}(1,1,1/4,1/9,\ldots,1/(n-1)^2) \in \mathbb{R}^{n\times n}$, and the forward map

$$\boldsymbol{\xi} \mapsto Gu(\cdot,\boldsymbol{\xi}), \qquad u(x,\boldsymbol{\xi}) := x\,\xi_1 + \sum_{k=1}^{n-1} \sqrt{2}\sin(k\pi x)\,\xi_{1+k},$$

where G is as given above. In the following we will always use $n = 100$ and the data $y = Gu^\dagger$ resulting from $u^\dagger(x) = 5\mathrm{sinc}(5(x-0.5))$. Moreover, for numerical computation of Af we use a uniform discretization of $[0,1]$ with $\Delta x = 2^{-10}$ and apply the trapezoidal rule for quadrature.

For approximate sampling from the target distribution of $\boldsymbol{\xi}$ given $Y = y$ we employ the same four proposals in the MH algorithm as in Section 5.6, i.e.,

- RW: Gaussian random walk proposal $\quad P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C)$,
- pCN: pCN proposal $\quad P(\boldsymbol{\xi}) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C)$,
- GN-RW: Gauss-Newton random walk proposal $\quad P(\boldsymbol{\xi}) = P_\sigma(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C_\sigma)$,
- gpCN: gpCN proposal $\quad P(\boldsymbol{\xi}) = P_{\Gamma_\sigma}(\boldsymbol{\xi}) = N(A_{\Gamma_\sigma}\boldsymbol{\xi}, s^2 C_{\Gamma_\sigma})$,

where the posterior covariance matrix $C_\sigma = C_{\Gamma_\sigma}$ and the matrix Γ_σ — both depending on σ — are as given in Theorem 6.12. We vary $\sigma \in (0,1]$ and examine numerically the performance of the four MH algorithms in terms of the expected acceptance probability, the expected squared jump distances in each direction and the effective sample size for several functions of interest. In all experiments we have chosen a burn-in length of 10^5 iterations and let the chain run afterwards for 10^6 iterations. The run length seemed to be sufficiently long in terms of sufficiently small estimated sampling errors for the expected acceptance probability and expected squared jump distances.

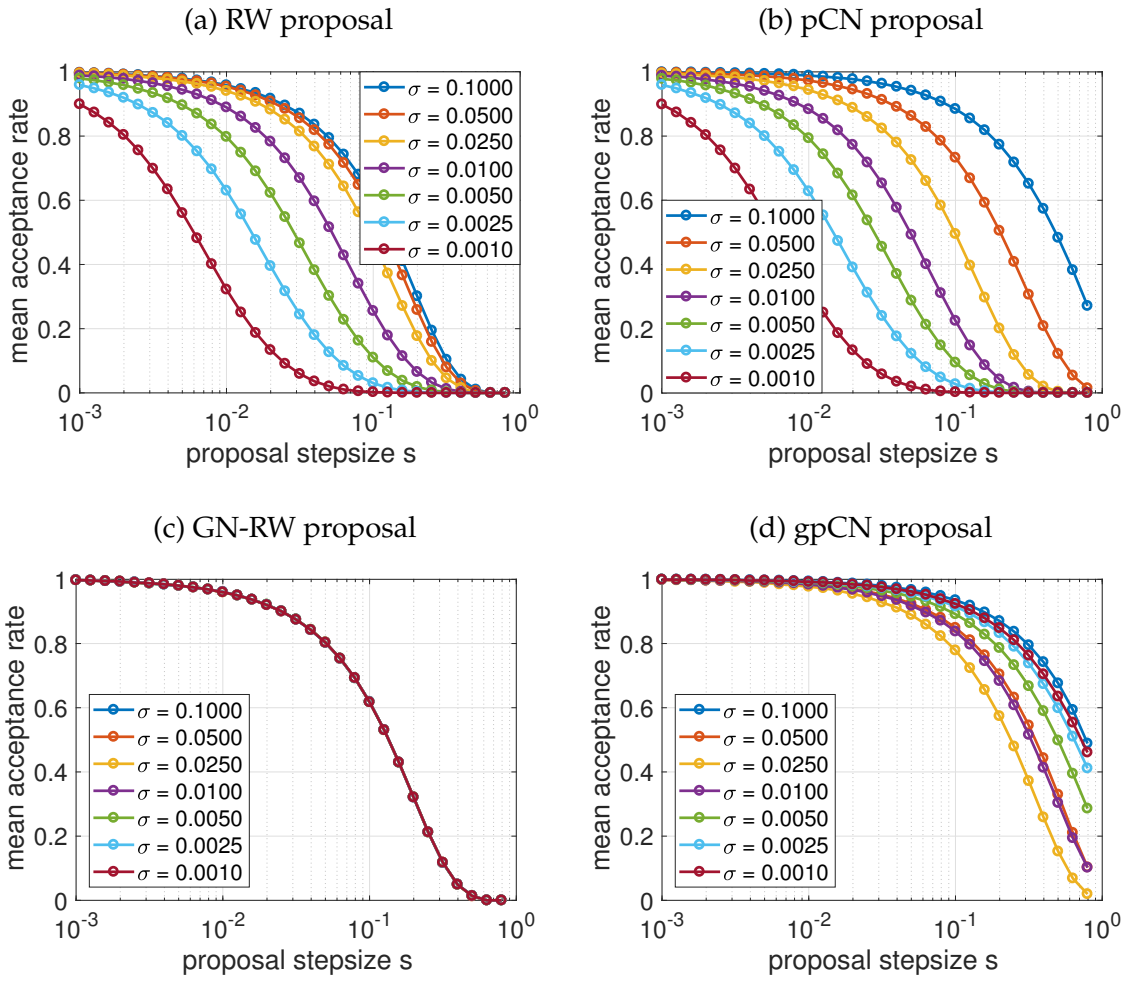

Figure 6.1.: Estimated dependence of expected acceptance probability w.r.t. proposal stepsize parameter s for several noise variance levels σ for the four Metropolis algorithms employing the RW, pCN, GN-RW and gpCN proposal.

Results for the expected acceptance probability. In Figure 6.1 we present the dependence of the expected acceptance probability w.r.t. the proposal stepsize parameter $s \in (0,1]$ for all four MH algorithms and certain values of σ. We observe for the RW and pCN proposal a deteriorating behaviour as $\sigma \to 0$, i.e., to obtain

the same expected acceptance probability α we would have to choose smaller and smaller proposal stepsizes s as σ tends to zero. On the other hand, the corresponding behaviour in case of the GN-RW proposal seems to be independent of σ. And for the gpCN proposal the behaviour even starts to improve when σ drops below 0.025, i.e., we even can allow for a slightly larger proposal stepsize and maintain the same expected acceptance probability. In summary, we observe for each chosen s a positive lower bound for the expected acceptance probability as $\sigma \to 0$ in case of the GN-RW and gpCN proposal — as predicted by Theorem 6.12 — whereas the same does not hold for the RW and pCN proposal.

Results for expected squared jump distances in each coordinate. Analogously to Figure 6.1, we display in Figure 6.2 the dependence of

$$\min_{i=1,\dots,n} \frac{\mathrm{ESJD}_i(M_\sigma)}{\mathrm{Var}_{\mu_\sigma}(\xi_i)},$$

which we will call *minimal relative ESJD (minimal RESJD)*. We recall that Theorem 6.12 just states that there exists a specific coordinate system w.r.t. which we should observe a positive non-zero bound for the minimal RESJD as $\sigma \to 0$. However, out of curiosity and for simplicity we consider the minimal RESJD w.r.t. canonical basis of $\mathbb{R}^n$. Indeed, we observe in case of the GN-RW and the gpCN proposal for each fixed s a positive lower bound the minimal RESJD as $\sigma \to 0$. However, for the RW and pCN proposal we obtain deteriorating minimal RESJD as $\sigma \to 0$. This clearly shows that the efficient exploration of the state space by MH algorithms based on the GN-RW or gpCN proposal is independent of the spread or concentration of the target measure.

Results for the effective sample size. In Figure 6.3 we present estimated ESS for four quantities of interest $f_j \colon \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, 4$, and their dependence on σ for each proposal. For details of the estimation of the effective sample size we refer to Section 5.6. Each time the corresponding MH algorithm was tuned to an expected acceptance probability of 0.25%. The solid lines correspond to the estimated ESS using initial monotone sequence estimators and the dashed lines to the estimated ESS using batch means. Again, we display both curves for validation. The four quantities of interest we are considering are given in Figure 6.3. There we denote by $\boldsymbol{v}_1 \in \mathbb{R}^n$ the first eigenvector of the generalized eigenproblem given in (6.17). Thus, quantity $f_4(\boldsymbol{\xi}) = \boldsymbol{v}_1^\top C^{-1} \boldsymbol{\xi}$ corresponds to the C^{-1}-orthogonal projection of $\boldsymbol{\xi}$ onto $\mathrm{span}(\boldsymbol{v}_1)$. We recall that the pushforward measure $(f_4)_* \mu_\sigma$ becomes more and more concentrated as $\sigma \to 0$, see Proposition 6.14.

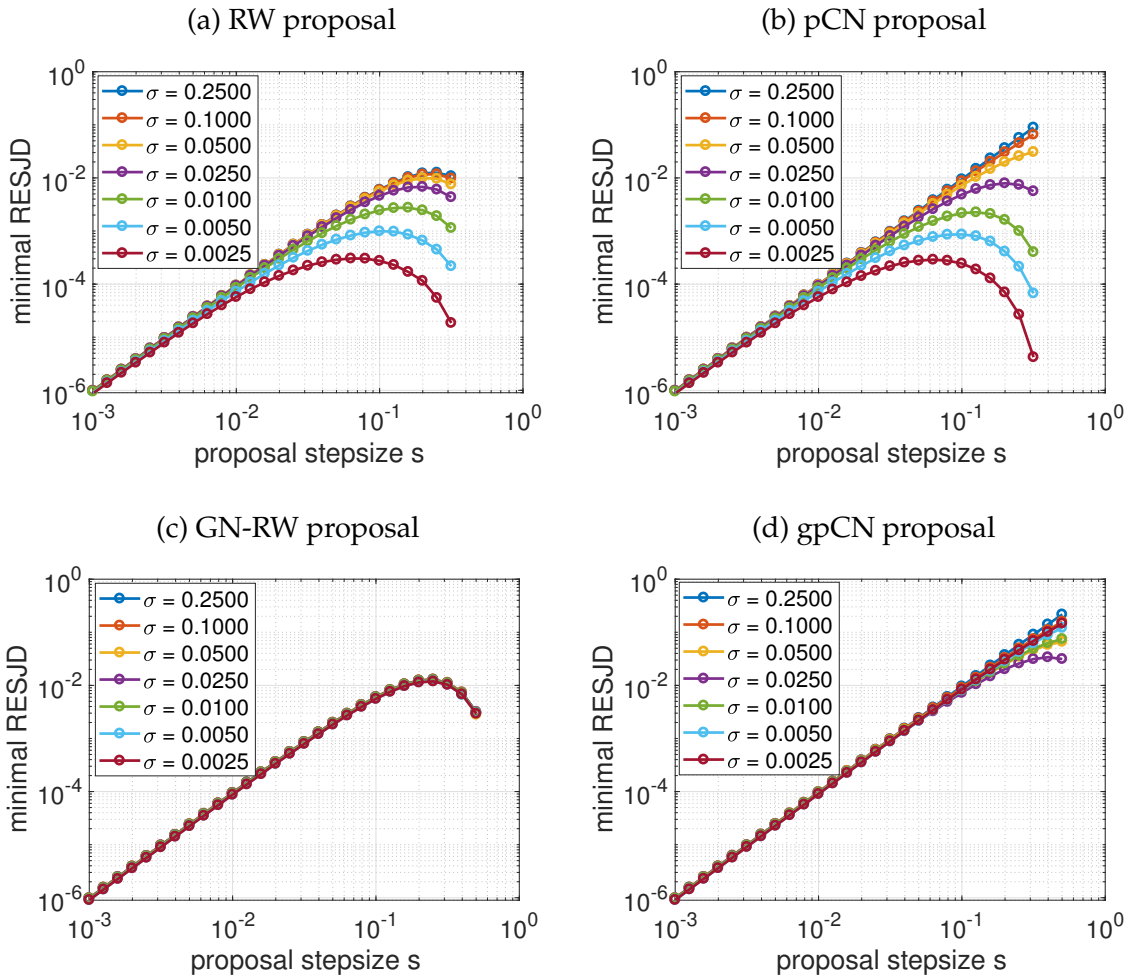

Figure 6.2.: Estimated dependence of the minimal RESJD $\min_i \mathrm{ESJD}_i(M_\sigma) / \mathrm{Var}_{\mu_\sigma}(\xi_i)$ w.r.t. stepsize parameter s for several noise variance levels σ for the four Metropolis algorithms employing the RW, pCN, GN-RW and gpCN proposal.

We observe even in terms of the ESS that the GN-RW and the gpCN perform independently w.r.t. the noise variance σ for all four quantities. In particular, the gpCN again seems to improve its ESS for decreasing σ after σ dropped below 0.025. The other two proposals show a deteriorating ESS for the first three quantities of interest — as we would expect given the previous numerical results. However, we notice an interesting behaviour of the ESS for f_4: here, also the RW and pCN proposal show a stable ESS which is apparently slightly larger than the corresponding ESS of the GN-RW and gpCN proposal. This is surprising at the first glance. At the second glance we might explain it as follows: since all proposal are tuned to the same expected acceptance probability of 25%, we only compare the resulting proposal step sizes in the direction of v_1 of the four MH algorithms. For the GN-RW and the gpCN proposal the proposal stepsize in the direction of v_1 is given by $\frac{s^2}{1+\sigma^{-2}\lambda_1} = \sigma^2 \frac{s^2}{\sigma^2+\lambda_1}$, see Proposition 6.17, i.e., it is of order $\mathcal{O}(\sigma^2)$. On the other hand, we know from Beskos et al. [16] that proposal stepsize of the RW proposal has to

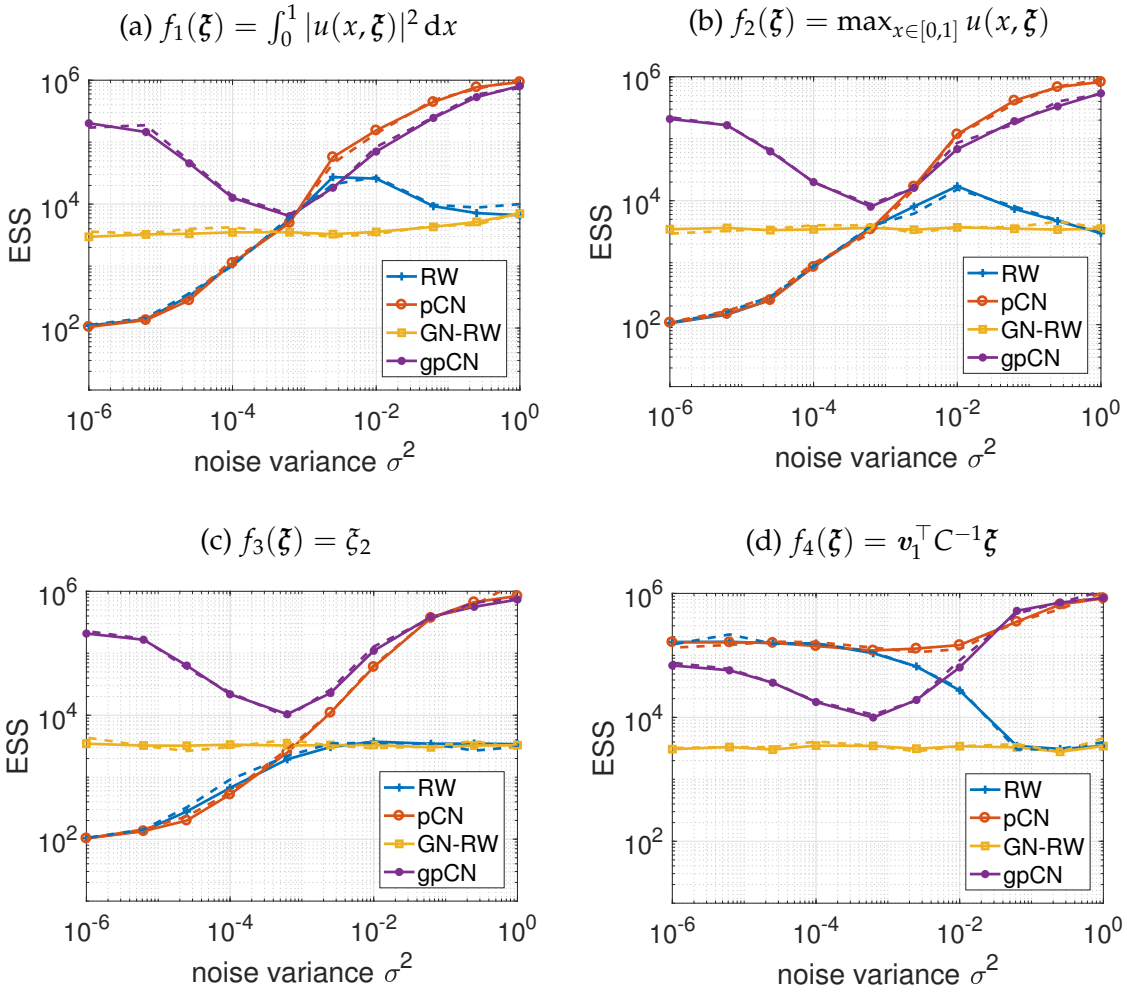

Figure 6.3.: Estimated dependence of the ESS w.r.t. σ given functions of interst f_i, $i = 1, \ldots, 4$, for the four Metropolis algorithms based on the RW, pCN, GN-RW and gpCN proposal (solid lines correspond to estimated ESS using IMSE, dashed lines to results using batch means).

scale w.r.t. σ like $\mathcal{O}(\sigma^2)$ in order to guarantee a fixed expected acceptance probability. Assuming the same holds for the pCN proposal, all four proposals apply the same scaling of the proposal step size in the direction of v_1 when tuned to a common fixed expected acceptance probability. This might provide an explanation for the observed ESS for f_4.

6.4.2. Nonlinear Forward Maps

We now slightly change the forward map G to obtain a nonlinear problem, but the other parts of the Bayesian inference setting such as prior and noise model and numerical discretization remain the same.

First stage. We consider the forward map $\mathbb{R}^n \ni \boldsymbol{\xi} \mapsto G\left(u(\cdot, \boldsymbol{\xi})\right) \in \mathbb{R}^4$ where the mapping $G\colon C(0,1) \to \mathbb{R}^4$ is given by

$$G(f) := [G_1(f), \dots, G_4(f)], \qquad G_j(f) := \exp\left(5\, Af(0.2j)\right), \quad j = 1, \dots, 4, \tag{6.27}$$

and where A denotes the linear convolution operator introduced in the previous subsection. Again, we set $n = 100$ and $y = G(u^\dagger)$ with $u^\dagger$ as in Subsection 6.4.1 and perform the same tests for the expected squared jump distance of the four proposals as before. We emphasize that the nonlinear forward is chosen in such a way that, although the posterior is non-Gaussian, there still exists a linear subspace on which the posterior marginal is not affected by changes in σ, namely,

$$\begin{aligned}\{\boldsymbol{\xi} \in \mathbb{R}^n : \Phi(\boldsymbol{\xi}) = 0\} &= \{\boldsymbol{\xi} \in \mathbb{R}^n : |y - G(u(\boldsymbol{\xi}))| = 0\} \\ &= \{\boldsymbol{\xi} \in \mathbb{R}^n : \ln(y_j) = \big(Au(\boldsymbol{\xi})\big)(0.2j)\ \forall j = 1, \dots, 4\}\end{aligned}$$

which is clearly a linear subspace of dimension $n - 4$.

For constructing the GN-RW and gpCN proposal we proceed as in 5.6 and linearize the forward G at the MAPE to obtain an approximation to the target covariance. For the partial derivatives of the forward map we get with $\phi_1(x) = x$ and $\phi_k(x) = \sqrt{2}\sin(k\pi x)$ for $k \geq 2$ that

$$\frac{\partial G_j(u(\cdot, \boldsymbol{\xi}))}{\partial \xi_k} = 5\, A\phi_k(0.2k)\, G_j(u(\cdot, \boldsymbol{\xi})).$$

The MAPE is computed as described in Section 5.6, i.e., we apply the MATLAB function `lsqnonlin` to solve the corresponding minimization problem.

We observe in Figure 6.4 again a variance independent expected acceptance probability and minimal RESJD for the GN-RW and gpCN proposals as in the linear example. The results for the RW and pCN proposal are not displayed, but they are basically also unchanged to the linear example.

Second stage. In the second nonlinear example we first compute the exponential $\exp(f)$ of a function $f \in C(0,1)$ before we apply the linear convolution operator A as in (6.4.1), i.e., we set for $f \in C(0,1)$

$$G(f) := [G_1(f), \dots, G_4(f)], \qquad G_j(f) := [A\exp(f)]\,(0.2j), \quad j = 1, \dots, 4. \tag{6.28}$$

The rest remains unchanged compared to the first stage example. Of course, the data y is modified, i.e., we take $y = G(u^\dagger)$ with $u^\dagger$ as before but with G as given

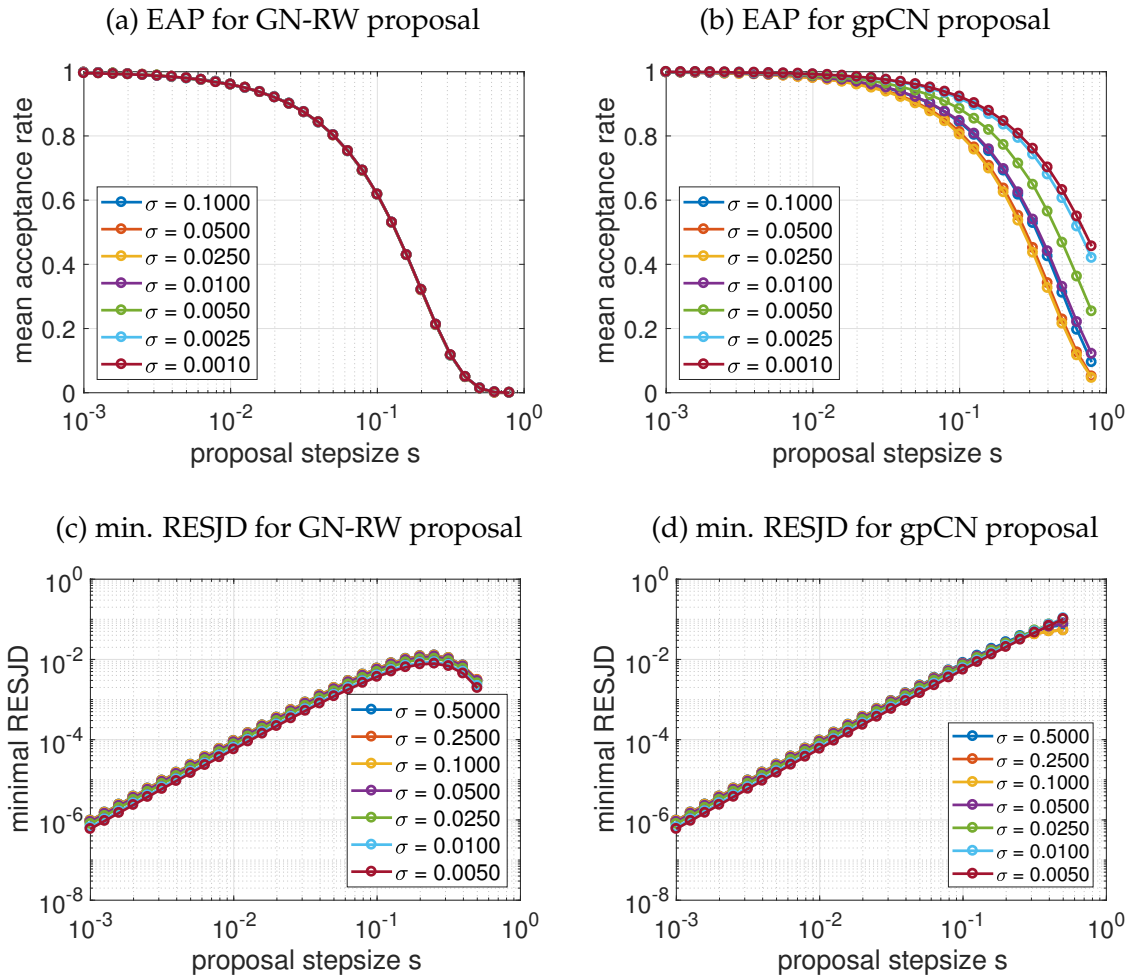

Figure 6.4.: Estimated dependence of expected acceptance probability and minimal RESJD w.r.t. stepsize parameter s for several noise variance levels σ in case of the first nonlinear forward map G as given in (6.27).

in (6.28). This forward map comes a bit closer to the boundary value problem with lognormal diffusion coefficient considered in Section 5.6, since we apply here the convolution operator A to a lognormal random function $u(\cdot, \xi)$ and evaluate the result at some points. Moreover, there exists no longer a linear subspace on which the posterior marginal is not affected by σ in contrast to the previous nonlinear example. We run the same test for the four proposals and linearize G again at the MAPE to get the proposal covariances for P_σ and P_{Γ_σ}. Here we have with ϕ_k as above

$$\frac{\partial G_j(u(\cdot,\xi))}{\partial \xi_k} = A\left[\phi_k \exp(u(\cdot,\xi))\right](0.2j).$$

This time, we do not observe a variance independence of the minimal RESJD for the GN-RW and gpCN proposals in Figure 6.5, but their minimal RESJD is significantly less affected by $\sigma \to 0$ than the minimal RESJD of the RW and pCN Metropolis. These results coincide with the observations made in Section 5.6 in terms of the

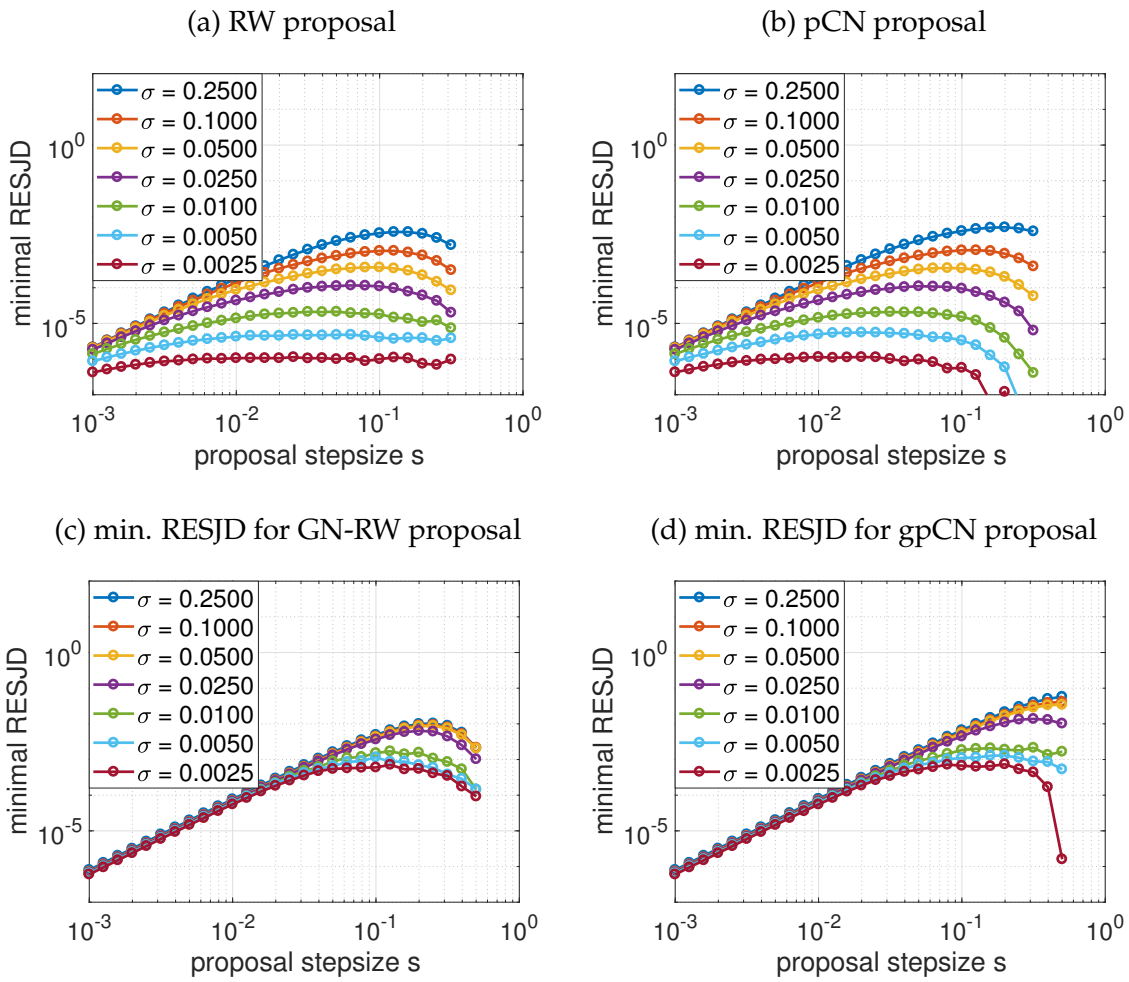

Figure 6.5.: Estimated dependence of minimal relative expected squared jump distance w.r.t. stepsize parameter s for several noise variance levels σ in case of the second nonlinear forward map G as given in (6.28).

ESS. Thus, also for this second nonlinear example Metropolis algorithms based on proposals using approximations to the posterior covariance outperform standard random walk Metropolis algorithms, although they do suffer from a decreasing variance.

Chapter 7

Case Study: Bayesian Inference for the WIPP Groundwater Flow Problem

In the following we will apply the numerical methods for uncertainty quantification and Bayesian inference established in the previous chapters to a real-world problem. The UQ problem we are considering is related to the waste isolation pilot plant (WIPP), a deep geological repository for radioactive waste located close to Carlsbad, New Mexico (USA), which is operating since 1999. The site of the repository has been investigated geologically for decades and several reports and compliance (re)certifications are publicly available on the WIPP website[1]. The repository itself lies approx. 655m below ground in a halite formation. However, above the repository there exists a groundwater transmissive layer called the Culebra Dolomite. One important scenario in the safety analysis and compliance recertification of the repository is that radionuclides are accidently released from the repository and transported by groundwater through the Culebra Dolomite. People are particularly interested in the chance that these released radionuclides exit a specific rectangular domain around the location of the repository, called the WIPP site, within $10,000$ years. This constitutes a UQ task par excellence, because (a) we only have limited knowledge about the porosity and hydraulic conductivity of the Culebra Dolomite layer provided by roughly 40 borehole measurements, and (b) experimental approaches via tracers to estimate the exit time are unfeasible due to the low conductivity. Thus, only numerical simulations and mathematical methods for UQ remain to provide answers.

In the following section we will explain our UQ approach to compute the probability that released radionuclides need less than 10,000 years to exit the WIPP site. Since we do not have access to the original computer models employed in the compliance recertification reports, our physical model for the groundwater flow will be much simpler: it is an elliptic PDE based on *Darcy's law*. The same model was

[1] http://www.wipp.energy.gov/index.htm

	m East	m North
Computational domain D		
Southwest corner	602 700	3 566 500
Southeast corner	624 000	3 566 500
Northeast corner	624 000	3 597 100
Northwest corner	602 700	3 597 100
WIPP site D_0		
Southwest corner	610 567	3 578 623
Southeast corner	617 015	3 578 681
Northeast corner	616 941	3 585 109
Northwest corner	610 495	3 585 068
Repository x_0	613 602	3 581 425

Table 7.1.: UTM coordinates of the four corners of the computational domain and the WIPP site as well as the location of the repository.

applied by Stone [165] who also considered UQ methods for the WIPP problem. Furthermore, we employ a two-dimensional PDE model (as also done in the recertification reports), since the thickness of the Culebra Dolomite layer is rather small (7.7m) compared to the computational domain (roughly 20km times 30km). As in [165] we use the measurement data provided in LaVenue et al. [103].

7.1. Problem Setting and General Approach

As the computational domain D we consider the same reactangular area given in LaVenue et al [103]. The UTM coordinates of D as well as of the WIPP site D_0 are presented in Table 7.1.

Employing Darcy's law and the law of mass conservation we obtain the following sytem of differential equations for the random groundwater pressure head p and the random groundwater flux $\boldsymbol{u}$:

$$\boldsymbol{u}(x,\omega) = -a(x,\omega)\nabla p(x,\omega) \qquad \text{in } D, \mathbb{P}\text{-a.s.}, \tag{7.1a}$$

$$\nabla\cdot \boldsymbol{u}(x,\omega) = 0 \qquad \text{in } D, \mathbb{P}\text{-a.s.}, \tag{7.1b}$$

$$p(x,\omega) = g(x) \qquad \text{on } \partial D, \mathbb{P}\text{-a.s.}, \tag{7.1c}$$

where $a(x,\omega)$ describes the random hydraulic conductivity. Moreover, we assume deterministic Dirichlet boundary data g. Of course, the groundwater head p is also unknown at the boundary ∂D, but for simplicity we assume g to be deterministic. In particular, we estimate g by geostatistical methods (see next section) applied to the pressure head measurements taken in the domain D.

Besides the equations (7.1) determing the groundwater flow we employ the following simple ordinary differential equation (ODE) for modelling the transport of radionuclides via groundwater when released at the repository location x_0:

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t,\omega) = \boldsymbol{u}(x,\omega), \quad x(0,\omega) = x_0 \in D_0, \qquad \mathbb{P}\text{-a.s.}, \tag{7.2}$$

i.e., we are neglecting molecular dispersion. The quantity of interest (QoI) is then the random exit time or travel time

$$t_{\text{exit}}(\omega) := \min\{t > 0 : x(t,\omega) \notin D_0 \text{ and } x \text{ solves (7.2)}\}. \tag{7.3}$$

Although the main interest for the compliance recertification of the WIPP site is the probability $\mathbb{P}(t_{\text{exit}} < 10,000 \text{ years})$, we will estimate the cumulative distribution function (CDF) of the real-valued random variable $\log_{10} t_{\text{exit}}$ in the subsequent simulations. Besides this, we sometimes also present kernel density estimates for the associated probability distribution function (PDF) of $\log_{10} t_{\text{exit}}$.

In order to establish a random field model for the uncertain conducticity a we will employ all available relevant observational data provided in LaVenue et al. [103]. This data consists of 38 measurements of the log transmissivity $\log T$ taken at locations $x_j \in D$, $j = 1\ldots,38$, and 33 noisy measurements of the groundwater pressure head p at locations $x_k \in D$, $k = 1\ldots,33$, both given in LaVenue et al. [103, Table 2.4 and Table 2.6]. Here, transmissivity is the rate at which the groundwater flows horizontally through the transmissive rock and relates to the conductivity a by $a(x,\omega) = \frac{1}{b\phi}T(x,\omega)$ where $b = 7.7$ denotes the constant layer thickness and $\phi = 0.16$ the constant rock porosity. That b and ϕ do not vary spatially is again a simplifying assumption. We build our random field model for a or $\log a$, respectively, following a two-step procedure:

1. Given some common assumptions on $\log a$, we apply standard geostatistical methods to build a *prior model* for $\log a$ based only on the transmissivity measurements $\log T(x_j) = \log a(x_j) + \log(b\phi)$, $j = 1,\ldots,38$, and compute the Karhunen-Loève expansion of this prior random field. The geostatistical approach itself consists of two steps:
   a) We estimate the parameters appearing in the assumed covariance function of $\log a$ by *maximum likelihood* methods.
   b) We condition the random field $\log a$ on the observations of $\log a$ taken at the measurement locations $x_j \in D$, $j = 1\ldots,38$, by *kriging* methods. This results again in an explicit random field model for $\log a$.

2. We condition the random coefficients appearing in the KLE of $\log a$ on the noisy pressure head measurements $p(x_k) + \varepsilon_k$, $k = 1, \dots, 33$ via Bayes' theorem. In particular, we will run a Markov chain Monte Carlo simulation to generate samples of these coefficients which are approximately distributed according to the resulting posterior measure. These samples yield (posterior) samples of t_{exit} which we use to estimate the (posterior) CDF of $\log_{10} t_{\text{exit}}$.

Numerical discretization. For solving (7.1) we apply the finite element method to the corresponding pathwise weak form:

$$\langle a^{-1}(\omega)\boldsymbol{u}(\omega), \boldsymbol{v}\rangle_{L^2(D)} - \langle p(\omega), \nabla\cdot \boldsymbol{v}\rangle_{L^2(D)} = -\int_{\partial D} g\, \boldsymbol{v}\cdot\boldsymbol{n}\,\mathrm{d}s \qquad \forall \boldsymbol{v} \in H(\mathrm{div}; D), \tag{7.4a}$$

$$\langle v, \nabla\cdot \boldsymbol{u}(\omega)\rangle_{L^2(D)} = 0 \qquad \forall v \in L^2(D), \tag{7.4b}$$

$\mathbb{P}$-almost surely, where $\boldsymbol{n}$ denotes the outward unit normal vector of ∂D. The finite elements employed are *Raviart-Thomas elements* of lowest order, i.e., given a triangulation of D of mesh size $h > 0$ the pressure head $p(\omega)$ is approximated by a piecewise constant function $p_h(\omega)$ and $\boldsymbol{u}(\omega)$ by a piecewise linear function $\boldsymbol{u}_h(\omega)$ with continuous normal component along the edges of the triangular mesh. For details about Raviart-Thomas elements, we refer to Brezzi and Fortin [22, Chapter III]. By the properties of the finite element space and due to $\langle v, \nabla\cdot \boldsymbol{u}(\omega)\rangle_{L^2(D)} = 0$ the approximation $\boldsymbol{u}_h(\omega)$ is $\mathbb{P}$-almost surely elementwise constant. This makes the elementwise solution of the resulting particle transport equation

$$\frac{\mathrm{d}}{\mathrm{d}t} x_h(t,\omega) = \boldsymbol{u}_h(x,\omega), \quad x_h(0,\omega) = x_0 \in D_0, \qquad \mathbb{P}-\text{a.s.}\,, \tag{7.5}$$

trivial and, thus, the main computational work is required for solving (7.4). We mention that the employed triangular mesh respects the boundary ∂D_0 of the WIPP site.

Remark 7.1 (On the wellposedness of (7.2) and (7.5))**.** The Picard-Lindelöf theorem ensures a unique pathwise solution $x(\cdot,\omega)$ and $x_h(\cdot,\omega)$ of the random ODEs (7.2) and (7.5), respectively, if $\boldsymbol{u}(\cdot,\omega)$ and $\boldsymbol{u}_h(\cdot,\omega)$ are $\mathbb{P}$-a.s. Lipschitz continuous. This condition is clearly satisfied for $\boldsymbol{u}_h(\cdot,\omega)$, since it is a piecewise linear function with continuous normal component along the edges. For the pathwise solution $\boldsymbol{u}(\cdot,\omega)$ of (7.1) the Lipschitz continuity is a more delicate issue: given that a has $\mathbb{P}$-a.s. Lipschitz continuous realizations, then it can be shown that also $\boldsymbol{u}(\cdot,\omega)$ is Lipschitz continuous, see the discussion in Graham et al. [77, Section 5.3]. However, our

case where we assume that a is a lognormal random field with a Matérn covariance function with smoothness parameter $\nu = 0.5$, see next section, we get only Hölder continuity with exponent less than $1/2$ for the realizations of a. Thus, in that case the existence of a unique solution of (7.2) is unclear. On the other hand, if we substitute $\log a$ by a finite truncation of its KLE, then the resulting realizations are Lipschitz and (7.2) well-posed, because the Karhunen-Loève eigenfunctions inherit the regularity of the covariance function, see Schwab and Todor [158, Propostion 2.23].

7.2. Prior Random Field Model for the Log Conductivity

Our general assumptions about $\log a$ are the following: $\log a$ is a Gaussian random field on D with

- a mean field of the form

$$\mathbb{E}[\log a(x)] = \sum_{i=1}^{I} \beta_i f_i(x) \tag{7.6}$$

  where $f_i \colon D \to \mathbb{R}$ denote known *regression functions* and β_i, $i = 0, \dots, I$, the corresponding unknown regression coefficients

- an isotropic *exponential covariance* function $c \colon [0,\infty) \to [0,\infty)$,

$$c_{\sigma^2,\rho}(r) := \sigma^2 \exp\left(-\frac{r}{\rho}\right), \qquad r \geq 0, \tag{7.7}$$

  with variance parameter σ^2 and correlation length parameter ρ.

Linear regression models for the mean and the exponential covariance are common models in geostatistics, see Stein [164] or Chilès and Delfiner [29]. A widely used method for estimating the unknown parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_I)$, σ^2 and ρ given observational data $\boldsymbol{z} := (\log a(x_j))_{j=1,\dots,38}$ is the maximum likelihood (ML) method. The ML approach is as follows: since the random field $\log a$ is assumed Gaussian, the joint probability density function of the random vector $Z(\omega) := (\log a(x_j, \omega))_{j=1,\dots,38}$ is a multivariate normal density

$$p(\boldsymbol{z}; \boldsymbol{\beta}, \sigma^2, \rho) := \frac{1}{\sqrt{2\pi \det(\boldsymbol{C}_{\sigma^2,\rho})}} \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{F}^\top \boldsymbol{\beta})^\top \boldsymbol{C}_{\sigma^2,\rho}^{-1} (\boldsymbol{z} - \boldsymbol{F}^\top \boldsymbol{\beta})\right),$$

where

$$\boldsymbol{C}_{\sigma^2,\rho} := \left[c_{\sigma^2,\rho}(|x_i - x_j|)\right]_{i,j=1}^{38} \in \mathbb{R}^{38\times 38}, \quad \boldsymbol{F} = \begin{pmatrix} f_1(x_1) & \dots & f_I(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_{38}) & \dots & f_I(x_{38}) \end{pmatrix} \in \mathbb{R}^{38\times I}.$$

The ML estimates for $\boldsymbol{\beta}$, σ^2 and ρ are then given by

$$(\hat{\boldsymbol{\beta}}_{\mathrm{ML}}, \hat{\sigma}^2_{\mathrm{ML}}, \hat{\rho}_{\mathrm{ML}}) = \operatorname*{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^I,\ \sigma^2>0,\ \rho>0} (\boldsymbol{z} - \boldsymbol{F}^\top\boldsymbol{\beta})^\top \boldsymbol{C}_{\sigma^2,\rho}^{-1}(\boldsymbol{z} - \boldsymbol{F}^\top\boldsymbol{\beta}) + \log\det(\boldsymbol{C}_{\sigma^2,\rho}). \tag{7.8}$$

The simultaneous estimation of the mean and covariance function parameters leads, in general to an underestimation of the variance parameter σ^2, see Stein [164, Section 6.4]. An alternative approach which avoids this problem is the *restricted maximum likelihood method* (ReML) which we will describe below. But first, we choose suitable regression functions f_i for the mean model (7.6).

Choosing the regression model for the mean field. Concerning the mean field model for $\log a$ there are several regression functions f_j conceivable, e.g., a linear trend in a coordinate direction, $f(x) = x_1$, or the thickness $d(x)$ of the overburden above the Culebra layer, $f(\boldsymbol{\xi}) = d(x)$, see, e.g., Stone [165]. However, each such regression function would imply a certain structure of the mean field which may or may not be true. Therefore, we follow the principle of parsimony and assume the simplest possible regression model

$$\mathbb{E}\left[\log a(x)\right] = \beta_1 f_1(x), \qquad f_1(x) \equiv 1.$$

Estimating the parameters of the Matérn covariance function. After we have chosen a linear regression model for the mean field of $\log a$, we can estimate the covariance function parameters independently of any "true" or estimated value of $\boldsymbol{\beta}$ by the ReML method. Here, the data vector $\boldsymbol{z} = (\log a(x_1), \dots, \log a(x_{38}))^\top \in \mathbb{R}^{38}$ is projected on the orthogonal complement of the span of $\boldsymbol{F}$, where $\boldsymbol{F}$ is as defined above, by

$$\boldsymbol{r} = (I_{38} - \boldsymbol{F}(\boldsymbol{F}^\top\boldsymbol{F})^{-1}\boldsymbol{F}^\top)\boldsymbol{z}.$$

The resulting *residual vector* $\boldsymbol{r}$, or, to be more precise, the random vector $(I_{38} - \boldsymbol{F}(\boldsymbol{F}^\top\boldsymbol{F})^{-1}\boldsymbol{F}^\top)Z$ with Z as above, follows again a multivariate normal distribution which depends now only on σ^2 and ρ. Thus, we can compute a ML estimate for σ^2 and ρ independently of $\boldsymbol{\beta}$ by employing the likelihood of the residuals. For

more details we refer again to Stein [164, Section 6.4].

Using the observational data of the log transmissivity given in LaVenue et al. [103] we obtain by ReML the following estimates

$$\hat{\sigma}^2 = 25.78, \qquad \hat{\rho} = 17,665.$$

Remark 7.2. We also checked if the assumed form of the covariance function (7.7) is justified: we allowed for a covariance function belonging to the Matérn, see Example 2.8, and performed ReML estimation for the resulting three parameters σ^2, ρ and ν. For the latter we obtained $\hat{\nu} = 0.5882$, i.e., the simpler choice (7.7), which corresponds to a Matérn covariance functions with $\nu = 0.5$, is justified.

Conditioning Gaussian random fields and universal kriging. Given the data $z_j = \log a(x_j)$, $j = 1, \dots, 38$, we can already apply Bayesian inference and condition our current Gaussian random field model for $\log a$ on these linear observations. This results, again, in a Gaussian random field for $\log a$. In particular, let us consider an $x \neq x_j$, then the random vector $(\log a(x), \log a(x_1), \dots, \log a(x_{38}))$ is jointly Gaussian and we can apply Theorem 4.3 to obtain the posterior distribution of $\log a(x)$ given the data $\boldsymbol{z} = (\log a(x_j))_{j=1,\dots,38}$. Clearly, Theorem 4.3 can also be applied to obtain the jointly Gaussian posterior distribution of $\log a$ at several distinct locations. Thus, the conditioned random field $\log a$ given $\boldsymbol{z} = (\log a(x_j))_{j=1,\dots,38}$ is again a Gaussian random field and its mean and covariance function are given by

$$\begin{aligned} m^{\boldsymbol{z}}(x) &:= m(x) + \boldsymbol{c}_{\hat{\sigma}^2,\hat{\rho}}^{\top}(x)\, \boldsymbol{C}_{\hat{\sigma}^2,\hat{\rho}}^{-1}\, (\boldsymbol{z} - \boldsymbol{m}), \qquad x \in D \\ c^{\boldsymbol{z}}(x,y) &:= c_{\hat{\sigma}^2,\hat{\rho}}(|x-y|) - \boldsymbol{c}_{\hat{\sigma}^2,\hat{\rho}}(x)^{\top}\, \boldsymbol{C}_{\hat{\sigma}^2,\hat{\rho}}^{-1} \boldsymbol{c}_{\hat{\sigma}^2,\hat{\rho}}(y), \qquad x, y \in D, \end{aligned}$$

where $c_{\hat{\sigma}^2,\hat{\rho}}$ and $\boldsymbol{C}_{\hat{\sigma}^2,\hat{\rho}}$ are as defined above employing the ReML estimates for σ^2 and ρ, and where m denotes the mean function of the unconditioned random field $\log a$ as well as

$$\boldsymbol{m} := (m(x_1), \dots, m(x_{38}))^{\top}, \quad \boldsymbol{c}_{\hat{\sigma}^2,\hat{\rho}}(x) := \Big(c_{\hat{\sigma}^2,\hat{\rho}}(|x-x_1|), \dots, c_{\hat{\sigma}^2,\hat{\rho}}(|x-x_{38}|)\Big)^{\top}.$$

We highlight that the functions $m^{\boldsymbol{z}}$ and $c^{\boldsymbol{z}}$ coincide with the prediction and corresponding prediction error covariance provided by a geostatistical method called *simple kriging* which we outline in detail in Appendix D.

However, since we do not know the exact mean field m of $\log a$, we can just use an estimate for it in the above formulas. This is, of course, a valid approach, but we could also apply another variant of kriging called *universal kriging*. For the latter

we just need to assume a linear regression model for the mean field such as in (7.6), in particular, we do not need to know the exact regression coefficients. Then, the resulting universal kriging prediction is given by

$$\log \hat{a}_{\mathrm{uk}}(x) = \begin{pmatrix} \boldsymbol{c}_{\hat{\sigma}^2,\hat{\rho}}(x) \\ \boldsymbol{f}(x) \end{pmatrix}^{\top} \begin{pmatrix} \boldsymbol{C}_{\hat{\sigma}^2,\hat{\rho}} & \boldsymbol{F} \\ \boldsymbol{F}^{\top} & 0 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{z} \\ 0 \end{pmatrix}, \tag{7.9}$$

with $\boldsymbol{c}_{\hat{\sigma}^2,\hat{\rho}}$, $\boldsymbol{F}$ and $\boldsymbol{C}_{\hat{\sigma}^2,\hat{\rho}}$ just as above and $\boldsymbol{f}(x) := (f_1(x), \ldots, f_I(x))^{\top}$. Moreover, the universal kriging error covariance takes the form

$$c_{\mathrm{uk}}(x,y) = c^{z}(x,y) + \boldsymbol{\gamma}^{T}(x)(\boldsymbol{F}^{T}\boldsymbol{C}^{-1}_{\hat{\sigma}^2,\hat{\rho}}\boldsymbol{F})^{-1}\boldsymbol{\gamma}(y) \tag{7.10}$$

where c^z is as above and $\boldsymbol{\gamma}(x) := \boldsymbol{f}(x) - \boldsymbol{F}^{\top}\boldsymbol{C}^{-1}_{\hat{\sigma}^2,\hat{\rho}}\boldsymbol{c}(x)$. As we explain in Appendix D the additional term in the universal kriging error covariance relates to the error in the predicition caused by the ML estimation of $\boldsymbol{\beta}$. Furthermore, it can be shown that the universal kriging prediction $\log \hat{a}_{\mathrm{uk}}(x)$ coincides with the simple kriging prediciton for $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ being the ML estimate for $\boldsymbol{\beta}$, see again Appendix D. Thus, by universal kriging we get almost the same random field model as by simple kriging, but we take into account our uncertainty about the unknown mean field of $\log a$.

Of course, we are also uncertain about the true parameters in the covariance function $c_{\sigma^2,\rho}$ of $\log a$, but considering and quantifying also this uncertainty is not as easy as for the mean function and out of the scope of this thesis. However, we refer to, e.g., Stone [165] and Remark (7.4) for an approach to quantify the uncertainty about the covariance function.

The (final) prior random field. As our (final) prior random field model for $\log a$ which was supposed to be based only on the observational data of $\log a(x)$, we take the Gaussian random field with mean and covariance function resulting from universal kriging, i.e.,

$$m_{\mathrm{prior}}(x) := \log \hat{a}_{\mathrm{uk}}(x), \qquad c_{\mathrm{prior}}(x,y) := c_{\mathrm{uk}}(x,y). \tag{7.11}$$

Contour plots of the resulting $m_{\mathrm{prior}}(x)$ and $c_{\mathrm{prior}}(x,x)$ given the data in LaVenue et al. [103] are provided in Figure 7.1. We note, that although we started with a stationary Gaussian random field with an isotropic covariance function, after the universal kriging we end up with a non-stationary random field.

Remark 7.3 (On the regularity of the resulting prior random field)**.** In our case, since

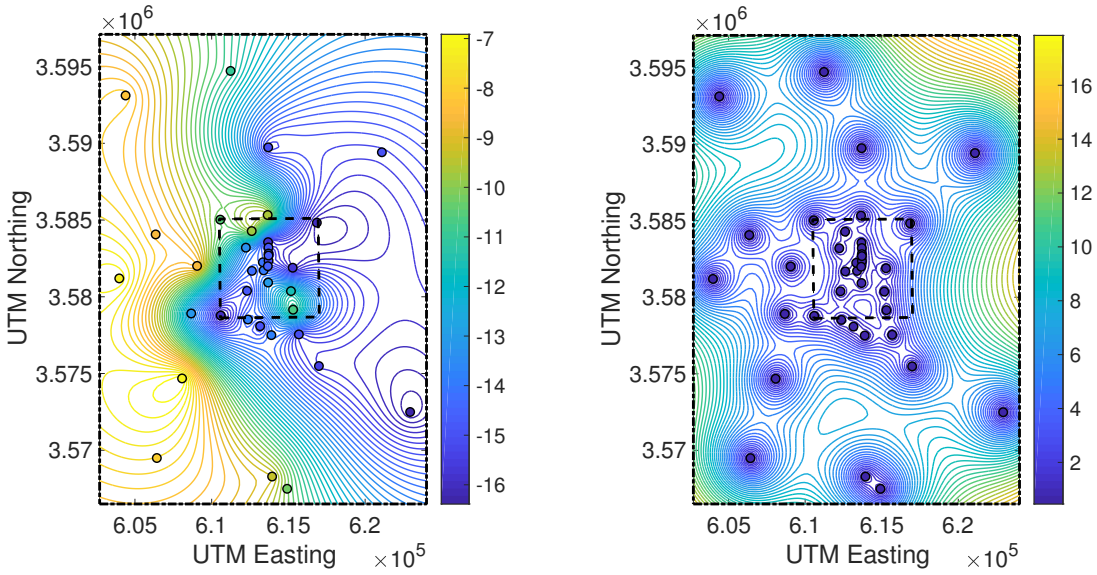

Figure 7.1.: Contour plots of the mean (left) and pointwise variance (right) resulting from universal kriging for the WIPP data. The filled dots correspond to the measurement locations and values.

we have chosen just a constant mean model, i.e., $\boldsymbol{f}(x) \equiv 1$, the covariance function resulting from universal kriging has the same regularity as the exponential covariance (7.7) we started with. This follows from (7.10). In general, the smoothness of the universal kriging covariance and, thus, also the smoothness of the associated Karhunen-Loève eigenfunctions, depends as well on the regularity of the regression functions f_i employed for the mean. However, the smoothness of the realizations of the *kriged* and *unkriged* random field is the same, since they are linear combinations of the correspoding mean and KL eigenfunctions.

Remark 7.4. One can argue that our procedure for building up the prior model for $\log a$ is not strictly Bayesian, since we used the available data of $\log a$ twice: the first time to estimate the parameters of the covariance function and a second time for conditioning the random field. A proper Bayesian procedure would have been to define also priors for the parameters $\boldsymbol{\beta}$, σ^2 and ρ and condition these on the data $\boldsymbol{z}$. However, as mentioned above this is out of the scope of this thesis, since we then could not work with just one KLE of $\log a$ but would have to compute for each posterior sample of $\boldsymbol{\beta}$, σ^2 and ρ a new KLE. For a recent work on this issue we refer to Sraj et al. [163].

Numerical computation of the Karhunen-Loève expansion. The eigenvalue problem

$$C_{\text{prior}}\phi(x) = \int_D c_{\text{prior}}(x,y)\,\phi(y)\,\mathrm{d}y = \lambda\phi(x), \qquad x \in D,$$

can be solved numerically again by the Galerkin method and finite elements. First, we transform it into a variational eigenvalue problem: find $\phi \in L^2(D)$ such that for all $v \in L^2(D)$

$$\begin{aligned}\langle C_{\text{prior}}\phi, v\rangle_{L^2(D)} &= \int_{D\times D} c_{\text{prior}}(x,y)\,\phi(y)\,v(x)\,\mathrm{d}y\,\mathrm{d}x = \lambda \int_D \phi(x)v(x)\,\mathrm{d}x \\ &= \lambda\langle C\phi, v\rangle_{L^2(D)}.\end{aligned}$$

Then, we choose finite subspaces $V_h \subset L^2(D)$ and seek for $\phi_h \in V_h$ such that

$$\langle C_{\text{prior}}\phi_h, v_h\rangle_{L^2(D)} = \lambda\langle C\phi_h, v_h\rangle_{L^2(D)}, \qquad \forall v_h \in V_h,$$

which yields a generalized eigenvalue problem for the coefficients $\boldsymbol{\phi}_h \in \mathbb{R}^N$ of ϕ_h w.r.t. a basis $\{v_1, \ldots, v_N\}$ of V_h:

$$A\boldsymbol{\phi}_h = \lambda_h B\boldsymbol{\phi}_h, \qquad A = \left[\langle Cv_i, v_j\rangle_{L^2(D)}\right]_{i,j=1}^N, \qquad B = \left[\langle v_i, v_j\rangle_{L^2(D)}\right]_{i,j=1}^N.$$

As finite subspace we employ piecewise constant functions based on a triangulation of D. The triangulation is chosen much finer than the corresponding one for solving (7.4), i.e., we have chosen a triangulation of $N = 142,872$ elements. The matrix B is diagonal due to our choice of V_h and the large matrix $A \in \mathbb{R}^{N\times N}$ is approximated by *hierarchical matrices*, see Hackbusch [79]. The eigenvalue problem is then solved via a restarted Lanczos method. We used MATLAB and C code provided by Ingolf Busch to compute the Karhunen-Loève expansion of our prior random field model for $\log a$. We computed the first $M = 2,500$ terms of the KLE, i.e., the M largest eigenvalues and their associated eigenfunctions which account for 95% of the variance of $\log a$ or, equivalently

$$\sum_{m>M} \lambda_m < 0.05 \sum_{m\geq 1} \lambda_m,$$

i.e., denoting the random field resulting from truncating the KLE of $\log a$ after M terms,

$$\log a_M(x,\omega) = \phi_0(x) + \sum_{m=1}^{M} \sqrt{\lambda_m}\phi_m(x)\xi_m(\omega), \qquad \xi_m \sim N(0,1) \text{ i.i.d.}, \tag{7.12}$$

there holds

$$\frac{\|\log a - \log a_M\|_{L^2(\Omega;L^2(D))}}{\|\log a\|_{L^2(\Omega;L^2(D))}} \leq 0.05.$$

The numerical decay of the eigenvalues is displayed in Figure 7.2 and coincides

with prediction by theory, i.e., $\lambda_m \leq C\, m^{-r}$ with $r = \frac{3}{2}$, cf. Remark 2.32. Moreover, we show in Figure 7.3 some numerically computed eigenfunctions of $\log a$. We observe that the eigenfunctions have value 0 at the measurement locations x_j of the transmissivity, since the covariance function c_{prior} resulting from universal kriging attains zero there, cf. Remark D.4. We also see the increasingly oscillating behaviour of ϕ_m as $m \to \infty$.

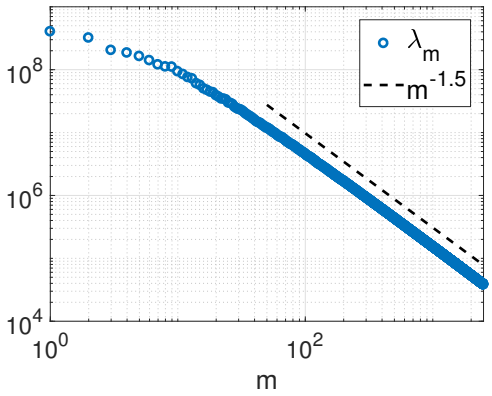

Figure 7.2.: Decay of the computed eigenvalues in the Karhunen-Loève expansion for the prior Gaussian random field.

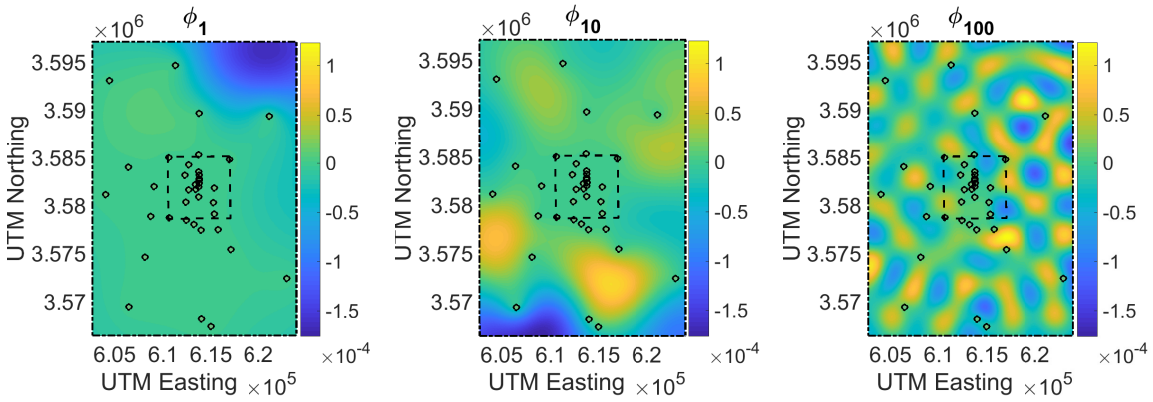

Figure 7.3.: Computed eigenfunctions ϕ_m in the Karhunen-Loève expansion for the prior Gaussian random field. The filled dots correspond to the measurement locations and values.

Prior mesh and truncation error. In this paragraph, we will quantify the error in estimating the CDF of $\log_{10} t_{\text{exit}}$ caused by the finite element method and by truncating the KLE of $\log a$ after M terms. We use the usual estimator for CDFs: given N (independent) samples t_j, $j = 1, \ldots, N$, of t_{exit} we use the simple function given by

$$\hat{F}_N(t) := \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{(-\infty, t_j]}(t), \qquad t \in \mathbb{R},$$

as an estimate for the true CDF of t_{exit}. The statistical error of the empirical estimator $\hat{F}$ can bounded due to Donsker's theorem, see Kallenberg [96, Theorem 14.15], which tells us that for

$$e_N := \sup_{t \in \mathbb{R}} \left| F(t) - \hat{F}_N(t) \right| \tag{7.13}$$

the term $\sqrt{N} e_N$ converges in distribution to $\sup_{t \in [0,1]} |BB(t)|$, where BB denotes a standard Brownian bridge. Hence, in order to obtain, for instance,

$$\mathbb{P}\left(e_N \leq 0.01\right) \leq 0.95, \tag{7.14}$$

we need roughly $N = 20,000$ independent samples.

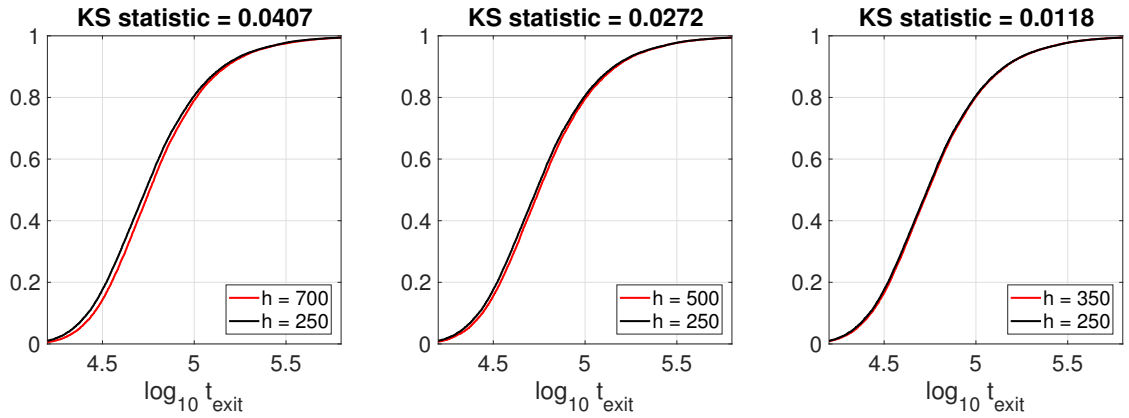

Figure 7.4.: Empirical cumulative distribution functions of $\log_{10} t_{\text{exit}}$ resulting from 20,000 samples of the truncated prior random field $\log a_{2500}$ evaluated on different meshes.

For quantifying the mesh error, we employed several meshes with different mesh sizes on which we solve (7.4) and (7.2). For each mesh we draw 20,000 independent realizations of $\log a_{2500}$ and compute the resulting 20,000 samples of t_{exit}. In Figure 7.4 we show the resulting empirical CDFs where we used a mesh with mesh size $h_{\text{ref}} = 200$ (yields 55,874 triangles) as reference and show results for $h = 350$ (18,142 triangles), $h = 500$ (8,894 triangles) and $h = 700$ (4,454 triangles). In the eyeball norm the lines for $h = 350$ and $h = 200$ are indistinguishable which is verified by a two-sample *Kolmogorov-Smirnov test*, see, e.g., Williams [176, Section 8.4]. This test computes the *Kolmogorov-Smirnov statistic* $e_{N,N} := \sup_{t \in \mathbb{R}} |\hat{F}_{N,h_{\text{ref}}}(t) - \hat{F}_{N,h}(t)|$ of the two empirical CDFs obtained by the different meshes, and tests the null hypothesis that both samples were drawn from the same distribution. In case of $h = 350$ and $h = 200$ the test did not reject the null hypothesis, i.e., the discretization error is smaller than the sampling error. Recall that for the latter (7.14) holds. Thus, it seems sufficient for our purposes to use a mesh with mesh size parameter $h = 350$ in the following.

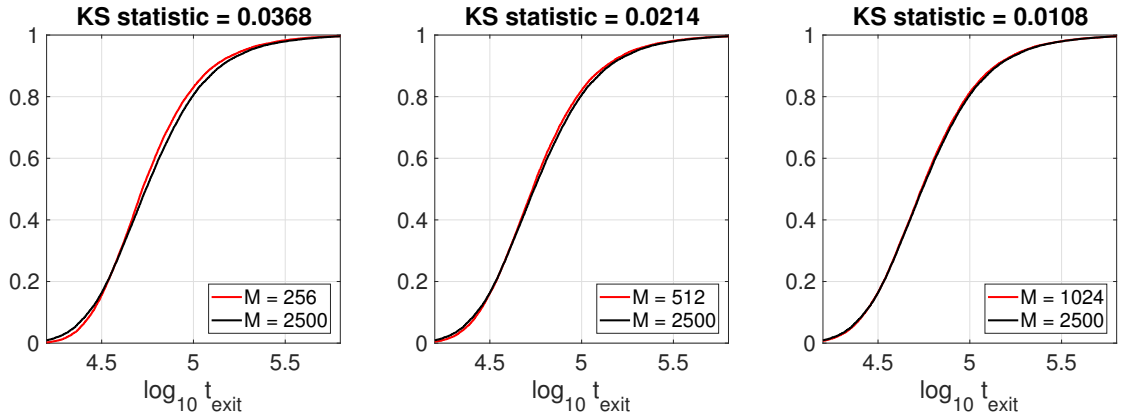

Figure 7.5.: Empirical cumulative distribution functions of $\log_{10} t_{\text{exit}}$ resulting from 20,000 samples of the truncated prior random field $\log a_M$ evaluated for several M.

We run an analogouos test for quantifying the effect of the length M for truncating the KLE of $\log a$ and display the empirical CDFs resulting from different truncation lengths M in Figure 7.5. We see $M = 256$ yields a statistically significant truncation error, but for $M = 1,024$ we pass again the two-sample Kolmogorov-Smirnov test. However, we will continue our posterior simulations with a truncated KLE of length $M = 2,500$.

Remark 7.5. Given the large number of relevant dimensions ($M > 256$) it seems challenging to provide sufficiently accurate approximations to the solution of (7.4) by surrogate methods for random and parametric PDEs as outlined in Section 2.3.2. Therefore, we continue our simulations by simply solving each time the weak formulation (7.4) by the finite element method.

7.3. Posterior Simulations

Now, we incorporate the 33 noisy groundwater pressure head measurements $y_k := p(x_k) + \varepsilon_k$, $k = 1, \dots, 33$, into our random field model for $\log a$ by Bayesian inference. Since we use a truncated Karhunen-Loève expansion $\log a_M$ as given in (7.12) of length M to approximate the prior random field $\log a$, we can apply a reparametrization, see Section 2.3, and infer the M-dimensional random vector consisting of the random coefficients appearing in the truncated KL

$$\xi(\omega) := (\xi_1(\omega), \dots, \xi_M(\omega))^\top, \qquad \xi \sim \mu_0 := \bigotimes_{m=1}^{M} N(0,1).$$

Thus, the observation model reads

$$y = G(\boldsymbol{\xi}) + \varepsilon, \qquad \varepsilon \sim N(0, I_{33}),$$

where $G\colon \mathbb{R}^M \to \mathbb{R}^{33}$ denotes the mapping

$$G\colon \boldsymbol{\xi} \mapsto \log a_M(x, \boldsymbol{\xi}) \mapsto p(x, \boldsymbol{\xi}) \mapsto (p(x_k, \boldsymbol{\xi}))_{k=1}^{33},$$

i.e., evaluating G includes solving the mixed variational problem (7.4).

Remark 7.6. Since $p(\cdot, \boldsymbol{\xi}) \in L^2(D)$ we could view the measurements as $L^2(D)$-functionals, i.e.,

$$y_k = \int_{B_\epsilon(x_k)} p(x, \boldsymbol{\xi})\, \mathrm{d}x, \qquad k = 1, \dots, 33$$

with $B_\epsilon(x) = \{y \in D\colon |x - y| < \epsilon\}$ and $\epsilon \ll 1$ which might be even more realistic, since the groundwater pressure head is always measured via a finite volume and not precisely at a single point.

Remark 7.7. (On measurement errors) The distribution of the measurement noise ε is not explicitly described in LaVenue et al. [103]. However, in [103, Table 2.6] the authors provide asymmetric uncertainties for the head measurements varying from ± 0.1 up to $+2.6/-2.2$ depending on the borehole. For simplicity we have chosen the model $\varepsilon \sim N(0, I_{33})$ for our simulations. Moreover, LaVenue et al. provide in [103, Table 2.4] also standard deviations for the transmissivity measurements. They are relatively small (from 0.25 to 0.5) and, again for simplicity, we neglected these errors in the previous section for deriving the prior random field model for $\log a$. However, we mention that these standard deviations can be easily included, in, e.g., the universal kriging procedure by adding the corresponding values to the diagonal of the matrix $\boldsymbol{C}_{\sigma^2,\rho}$ in (7.9) and (7.10), respectively, which represents then the covariance of the noisy observations $\tilde{Z}(\omega) := (\log a(x_j, \omega))_{j=1,\dots,38} + \tilde{\varepsilon}$.

We apply the Bayesian procedure outlined in Chapter 3 and use the Markov chain Monte Carlo methods established in Chapter 5 for approximate sampling of the resulting posterior distribution for $\boldsymbol{\xi}$ given $y = G(\boldsymbol{\xi}) + \varepsilon$. Particularly, we will apply the gpCN Metropolis algorithm with

$$\Gamma = \nabla G(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}) \nabla G(\hat{\boldsymbol{\xi}}_{\mathrm{CM}})^\top$$

where $\hat{\boldsymbol{\xi}}_{\mathrm{CM}}$ denotes an approximation to the posterior mean which was computed by a small preliminary run using the pCN Metropolis. An alternative would be to

compute the MAP estimate $\boldsymbol{\xi}_{\mathrm{MAPE}}$ by numerical optimization and linearize G there. In the next paragraph we explain how we evaluated the gradient ∇G.

Adjoint methods for mixed variational problems. Let us denote by $O\colon L^2(D) \to \mathbb{R}^{33}$, $O = (o_1, \dots, o_{33})$, the linear measurement operator which maps $p(\boldsymbol{\xi}) \in L^2(D)$ to the observable data. By linearity there holds $\partial_{\xi_m} G(\boldsymbol{\xi}) = O(\partial_{\xi_m} p(\boldsymbol{\xi}))$ and the partial derivatives $\partial_{\xi_m} p(\cdot, \boldsymbol{\xi})$ can be computed by adjoint methods. To this end, we formally differentiate the parametric variational equations

$$\langle a^{-1}(\boldsymbol{\xi})\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{v}\rangle_{L^2(D)} - \langle p(\boldsymbol{\xi}), \nabla\cdot \boldsymbol{v}\rangle_{L^2(D)} = -\int_{\partial D} g\, \boldsymbol{v}\cdot\boldsymbol{n}\,\mathrm{d}s \tag{7.15a}$$

$$\langle v, \nabla\cdot \boldsymbol{u}(\boldsymbol{\xi})\rangle_{L^2(D)} = 0, \tag{7.15b}$$

and get by interchanging differentiation w.r.t. ξ_m and integration w.r.t. x

$$\langle a^{-1}(\boldsymbol{\xi})\partial_{\xi_m}\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{v}\rangle_{L^2(D)} - \langle \partial_{\xi_m} p(\boldsymbol{\xi}), \nabla\cdot \boldsymbol{v}\rangle_{L^2(D)} = -\langle \partial_{\xi_m} a^{-1}(\boldsymbol{\xi})\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{v}\rangle_{L^2(D)} \tag{7.16a}$$

$$\langle v, \partial_{\xi_m}\nabla\cdot \boldsymbol{u}(\boldsymbol{\xi})\rangle_{L^2(D)} = 0, \tag{7.16b}$$

for all test functions $\boldsymbol{v} \in H(\mathrm{div}; D)$ and $v \in L^2(D)$. Let us now define $(\boldsymbol{\theta}_k(\boldsymbol{\xi}), q_k(\boldsymbol{\xi})) \in (H(\mathrm{div}; D), L^2(D))$, $k = 1, \dots, 33$, as the solution to

$$\langle a^{-1}(\boldsymbol{\xi})\boldsymbol{\theta}_k(\boldsymbol{\xi}), \boldsymbol{v}\rangle_{L^2(D)} - \langle q_k(\boldsymbol{\xi}), \nabla\cdot \boldsymbol{v}\rangle_{L^2(D)} = 0 \qquad \forall \boldsymbol{v} \in H(\mathrm{div}; D), \tag{7.17a}$$

$$\langle v, \nabla\cdot \boldsymbol{\theta}_k(\boldsymbol{\xi})\rangle_{L^2(D)} = o_k(v), \qquad \forall v \in H^1_0(D). \tag{7.17b}$$

Then we have for each $\boldsymbol{\xi} \in \mathbb{R}^{33}$ and eack $k = 1, \dots, 33$

$$\begin{aligned} o_k\left(\partial_{\xi_m} p(\boldsymbol{\xi})\right) &= \langle \partial_{\xi_m} p(\boldsymbol{\xi}), \nabla\cdot \boldsymbol{\theta}_k(\boldsymbol{\xi})\rangle_{L^2(D)} \\ &= \langle a^{-1}(\boldsymbol{\xi})\partial_{\xi_m}\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{\theta}_k(\boldsymbol{\xi})\rangle_{L^2(D)} + \langle \partial_{\xi_m} a^{-1}(\boldsymbol{\xi})\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{\theta}_k(\boldsymbol{\xi})\rangle_{L^2(D)} \\ &= \langle q_k(\boldsymbol{\xi}), \nabla\cdot \partial_{\xi_m}\boldsymbol{u}(\boldsymbol{\xi})\rangle_{L^2(D)} + \langle \partial_{\xi_m} a^{-1}(\boldsymbol{\xi})\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{\theta}_k(\boldsymbol{\xi})\rangle_{L^2(D)} \\ &= \langle \partial_{\xi_m} a^{-1}(\boldsymbol{\xi})\boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{\theta}_k(\boldsymbol{\xi})\rangle_{L^2(D)}. \end{aligned}$$

Thus, for computing $\nabla G(\hat{\boldsymbol{\xi}}_{\mathrm{CM}})$ we need to

1. solve (7.15) for $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}_{\mathrm{CM}}$,

2. solve for each $k = 1, \dots, 33$ the mixed problem (7.17) with $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}_{\mathrm{CM}}$,

3. evaluate for each $m = 1, \dots, M$ the inner product

$$\langle \partial_{\xi_m} a^{-1}(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}) \boldsymbol{u}(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}), \boldsymbol{\theta}_k(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}) \rangle_{L^2(D)}$$

where $\partial_{\xi_m} a^{-1}(x, \boldsymbol{\xi}) = -\sqrt{\lambda_m}\phi_m(x) a^{-1}(x, \boldsymbol{\xi})$.

Remark 7.8. In PDE-constrained optimization, one typically only needs to compute the action $\nabla G(\hat{\boldsymbol{\xi}}_{\mathrm{CM}})^\top \boldsymbol{\zeta}$ for $\boldsymbol{\zeta} \in \mathbb{R}^{33}$. This yields that we would only have to solve one adjoint problem, namely,

$$\begin{aligned} \langle a^{-1}(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}) \boldsymbol{\theta}_k(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}), \boldsymbol{v} \rangle_{L^2(D)} - \langle q_k(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}), \nabla \cdot \boldsymbol{v} \rangle_{L^2(D)} = 0 \qquad & \forall \boldsymbol{v} \in H(\mathrm{div}; D), \\ \langle v, \nabla \cdot \boldsymbol{\theta}_k(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}) \rangle_{L^2(D)} = \sum_{k=1}^{33} \zeta_k \, o_k(v), \qquad & \forall v \in H_0^1(D). \end{aligned}$$

However, for our purposes, i.e., computing once $\Gamma = \nabla G(\hat{\boldsymbol{\xi}}_{\mathrm{CM}}) \nabla G(\hat{\boldsymbol{\xi}}_{\mathrm{CM}})^\top$ and employing it for the gpCN Metropolis, we simply solve once the 33 corresponding adjoint problems.

The following numerical results were all obtained by applying the gpCN Metropolis algorithm if not stated otherwise. We always tuned the proposal stepsize to achieve an average acceptance rate of approximately 25%. Moreover, we allowed for a burn-in of 50,000 iterations and let the Markov chain run afterwards for 500,000 iterations. To check, if there is evidence that our Markov chain simulation of $\log_{10} t_{\mathrm{exit}}$ has not yet converged, we applied the *Geweke* and *Heidelberger and Welch* tests, see Remark 7.9. Both tests were passed for all presented simulations.

Remark 7.9 (On MCMC convergence diagnostics)**.** In general, we can not verify that a Markov chain "has converged", i.e., if it is sufficiently close to its stationary distribution, unless the latter is known. We refer to Geyer [69] for a discussion. However, there are several diagnostics available to verify that a Markov chain has not yet converged, see Cowles and Carlin [36] for an overview. The two diagnostics we apply in this thesis are designed for scalar-valued Markov chains. Geweke's diagnostic computes the empirical mean of the first $0.1N$ iterations and the last $0.5N$ iterations of a generated path of length N and tests if there is statistical evidence that both means differ. The Heidelberger and Welch diagnostic uses the Cramer-von Mises statistic to test if the simulated path can be considered as a realization of a stationary process which would be the case if the Markov chain would have already converged to its stationary distribution. Again, we refer to Cowles and Carlin [36] for details.

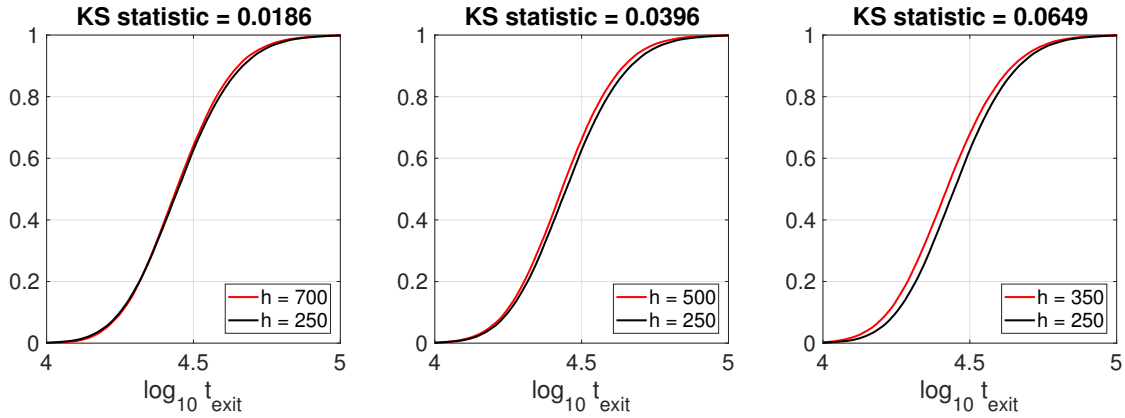

Figure 7.6.: Empirical cumulative distribution functions of $\log_{10} t_{\text{exit}}$ resulting from $500,000$ samples taken from MCMC simulations of the posterior random field $\log a_{2500}$ evaluated on different meshes.

Quantifying the mesh and truncation error. We quantify, again, the effect caused by numerical discretization and truncation of the KLE of $\log a$ for simulations of the posterior measure of $\log_{10} t_{\text{exit}}$. To this end, we choose again several meshes and several truncated KLEs and perform for each a MCMC simlulation. Analogously, to Figure 7.4 and 7.5 we plot the corresponding numerical results for the empirical posterior CDF in Figure 7.6 and 7.7, respectively. We observe a much more sensitive dependence on the mesh and the length of the truncated KLE than for the prior simulations. The behaviour w.r.t. the mesh size parameter h seems odd, i.e., we can not yet detect a convergence of the CDFs, in particular, the empirical CDF obtained on the coarsest mesh is closer to the reference CDF than the corresponding one obtained on a finer mesh. This calls for further investigation in the future. Moreover, we observe also a much stronger effect of the truncation of the KLE than for the simulations w.r.t. prior distribution.

However, in the subsequent simulations we will use a mesh with mesh size parameter $h = 350$ and a truncation length of $M = 2,500$.

Effects of conditioning. We illustrate the effects of conditioning the prior random field on the available groundwater head measurements. In Figure 7.8 and 7.9 we show the approximated mean and variance field for the prior and the posterior distribution of $\log a_M$. Concerning the variance fields we observe a significant decrease in the pointwise variance from prior to posterior across the whole computational domain. Also the posterior mean looks more detailed than the prior mean field, showing, for example, two subdomains of very low hydraulic conductivity on the right-hand part of the computational domain.

The corresponding posterior mean and variance of the ξ_m are displayed in Figure

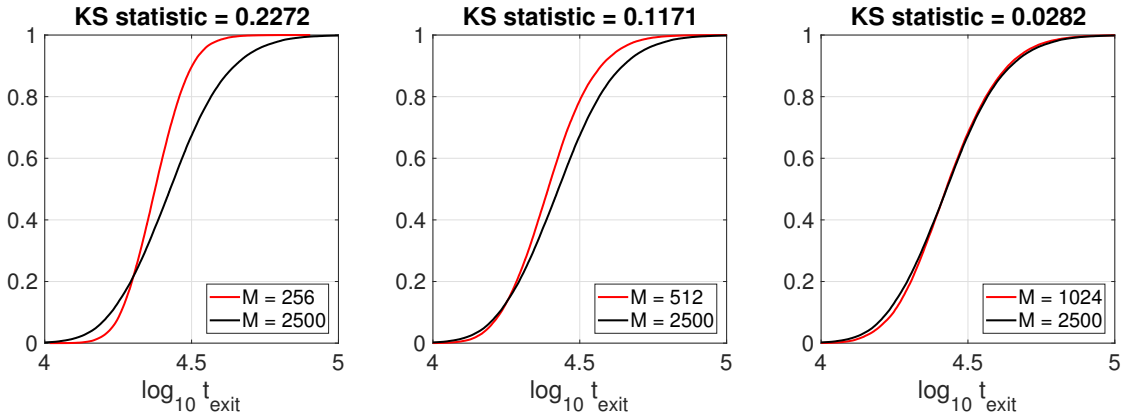

Figure 7.7.: Empirical cumulative distribution functions of $\log_{10} t_{\text{exit}}$ resulting from 500,000 samples taken from MCMC simulations of the posterior random field $\log a_M$ for several M.

7.10. We see that the largest changes in the mean $\mathbb{E}[\xi_m]$ from prior to posterior appears in the, say, first 100 dimensions $m = 1, \ldots, 100$, although we can also observe some significant changes for some higher dimensions. For the reduction in variance $\text{Var}(\xi_m)$ from prior to posterior we also observe the most dramatic changes in the first, roughly, 100 dimensions. For example, for ξ_8 we obtain a reduction of the variance from 1 (prior) to approx. 0.3732 (posterior). For some ξ_m it seems that the variance increased from prior to posterior, but these observations might be simply caused by the sampling error of the MCMC integration. That only the first variables ξ_m are informed or influenced by incorporating measurements of the pressure head is to be expected, because finitely many observations of the solution of an elliptic PDE carry usually less information about the high-frequency modes of the diffusion coefficient.

Moreover, concerning our quantity of interest, we present kernel density estimates of the prior and posterior probability density function of $\log_{10} t_{\text{exit}}$ as well as 1,000 particle paths resulting from prior and posterior simulations of $\log a_M$. Each kernel density estimate was based on 20,000 samples of $\boldsymbol{\xi}$. For the prior PDF we generated independent samples whereas for the posterior PDF we subsampled the Markov chain, i.e., we took only each 25th state of the Markov chain after burn-in. By the same procedure we obtained the 1,000 samples for the particle path simulations, i.e., this time we took only each 500th state of the Markov chain. By subsampling we reduce the correlation between the samples, i.e., we get only a few but less correlated samples. However, subsampling does, in general, not improve the MCMC estimate, see Yue and Chan [181]. We see that the posterior PDF of $\log_{10} t_{\text{exit}}$ is significantly more concentrated than the corresponding prior PDF. Moreover, also the location, i.e., the center of the probability mass and the peak of

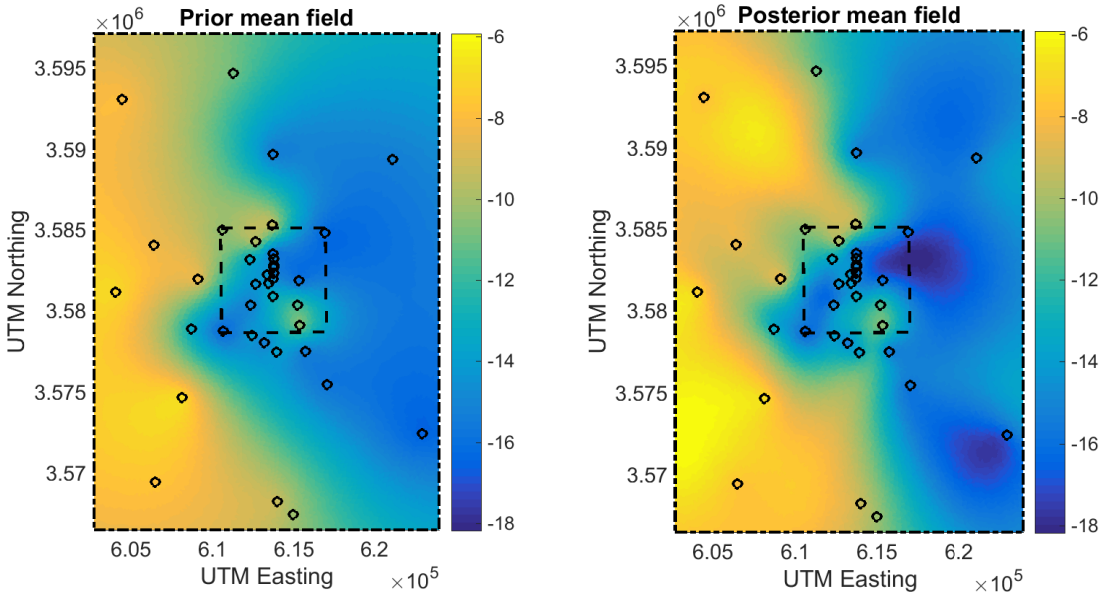

Figure 7.8.: Prior and posterior mean field of $\log a_M$ with $M = 2500$ obtained by solving the forward map on a triangular mesh with 18,142 elements. The filled dots correspond to the measurement locations and values.

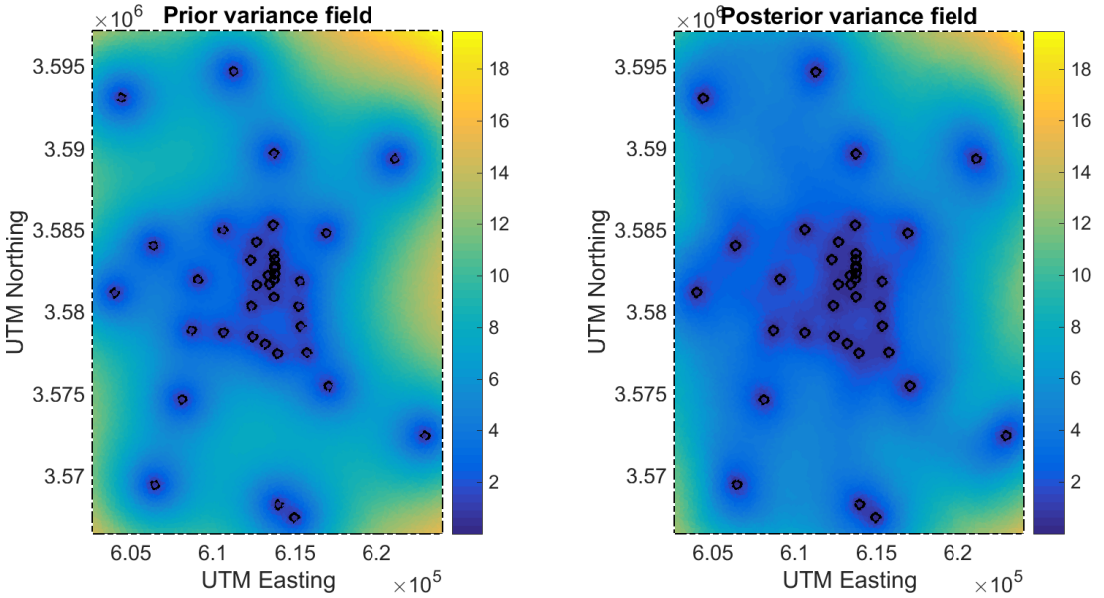

Figure 7.9.: Prior and posterior variance field of $\log a_M$ with $M = 2500$ obtained by solving the forward map on a triangular mesh with 18,142 elements. The dots correspond to the transmissivity measurement location.

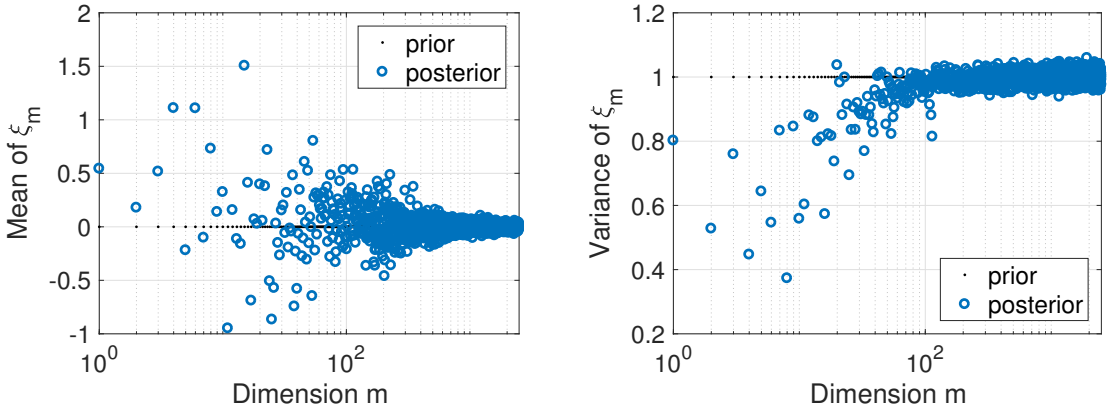

Figure 7.10.: Empirical posterior mean and variance of the random variables ξ_m for $m = 1, \ldots, 2500$ obtained by solving the forward map on a triangular mesh with 18,142 elements.

the PDF changed from prior to posterior, particularly, to smaller values. Besides that, we can also observe much more focused particle paths in case of the posterior simulations. All this illustrates the reduction in the uncertainty about $\log_{10} t_{\text{exit}}$ obtained by Bayesian inference using groundwater pressure head measurements.

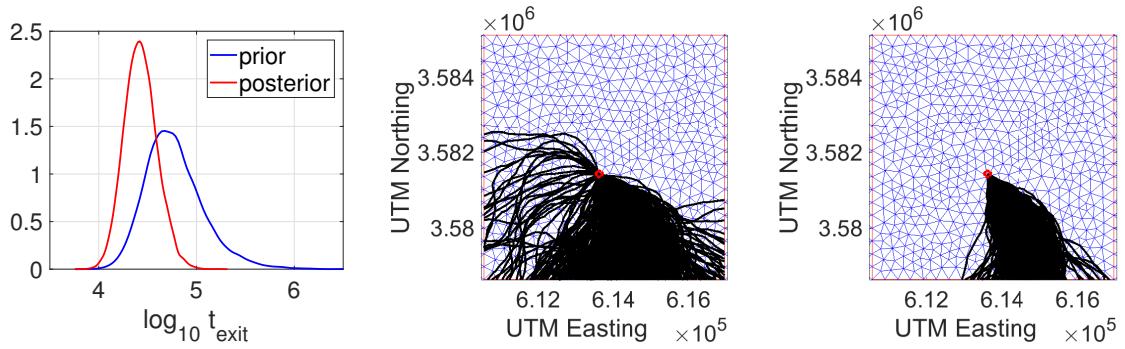

Figure 7.11.: Kernel density estimates of the probability density function of $\log_{10} t_{\text{exit}}$ resulting from prior and posterior simulations (left), 1,000 particle paths according to the prior (middle) and posterior (right) distribution of ξ or $\log a_M$, respectivly. All numerical simulations were based on a triangular mesh with 18,142 elements.

Performance of gpCN Metropolis. We also compare the performance of the gpCN Metropolis to the pCN Metropolis algorithm. To this end, we let both algorithms run for a length of 500,000 iterations after an initial burn-in phase of 50,000 iterations. Again, for both algorithms we tuned the proposal stepsize such that an average acceptance rate of 25% was achieved. In Figure 7.12 we present the resulting empirical autocorrelation function of $\log_{10} t_{\text{exit}}$ and observe a much faster decay in case of the gpCN Metropolis. In particular, an estimation of the corresponding

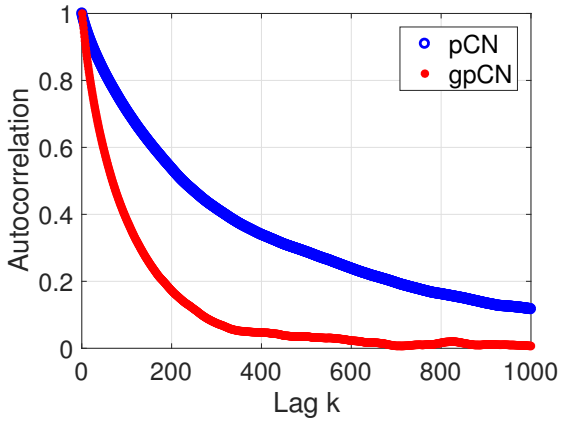

Figure 7.12.: Estimated autocorrelation times for the quantity of interest $\log_{10} t_{\text{exit}}$ resulting from simulations of the gpCN and pCN Metropolis algorithm.

integrated autocorrelation time (IACT) via initial monotone sequence estimators (IMSE) and batch means (BM) yields the results displayed in Table 7.2. Although the values obtained by the two estimators differ slightly, the IACT resulting from running the gpCN Metropolis is roughly one third of the corresponding IACT resulting from the pCN Metropolis. We recall that the computational work for both Metropolis algorithms is roughly the same: the only difference is the cost for sampling from the proposal, here:

$$P_{\text{pCN}}(\boldsymbol{\xi}, \cdot) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 I), \qquad P_{\text{gpCN}}(\boldsymbol{\xi}, \cdot) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma),$$

i.e., in case of the gpCN proposal we have to perform two additional matrix vector multiplications, namely, $A_\Gamma \boldsymbol{\xi}$ and $L_\Gamma \zeta$ where $L_\Gamma L_\Gamma^\top = C_\Gamma$ and $\zeta \sim N(0, I)$. However, compared to the work required for evaluating the forward map, i.e., solving a PDE by the finite element method, this additional work is neglegible. In particular, the required CPU time for running the Markov chain for 550,000 iterations was 63.94 hours in case of the gpCN Metropolis and 63.42 hours in case of pCN Metropolis. Thus, for estimating, e.g., the posterior mean of $\log_{10} t_{\text{exit}}$ the pCN Metropolis takes three times as long as the gpCN Metropolis to achieve the same accuracy.

	IACT via IMSE	IACT via BM
pCN Metropolis	789.36	720.20
gpCN Metropolis	231.58	188.63

Table 7.2.: Integrated autocorrelation times for the QoI $\log_{10} t_{\text{exit}}$ resulting from pCN and gpCN Metropolis simulations estimated by initial monotone sequence estimators and batch means.

Performance of EnKF. For comparison, we also apply the EnKF as presented in Section 4.1.2 to the WIPP Bayesian inference problem. As ensemble size we use $N = 100,000$. We know that the results obtained by the EnKF are in general different to those obtained by MCMC simulations, since the EnKF yields an analysis ensemble $(\boldsymbol{\xi}_1^a, \dots, \boldsymbol{\xi}_N^a)$ which is not distributed according to the posterior measure. In fact, we can observe this in Figure 7.13. There we compare the empirical mean and variances of the obtained analysis ensemble as well as the empirical PDF of the resulting ensemble $(\log_{10} t_{\text{exit}}(\boldsymbol{\xi}_1^a), \dots, \log_{10} t_{\text{exit}}(\boldsymbol{\xi}_N^a))$ to the corresponding results obtained by the gpCN Metropolis algorithm. Concerning the mean and the variance of the ξ_m, the EnKF performs rather poorly, i.e., the resulting empirical means and variances show significant mismatches to the true posterior moments. However, we observe the same general effect of conditioning, i.e., the first dimensions ξ_m are effected the most by incorporating the observational data and, surprisingly, the resulting PDF for the exit time shows a quite good match to the one resulting from the MCMC simulation — which might be pure coincidence.

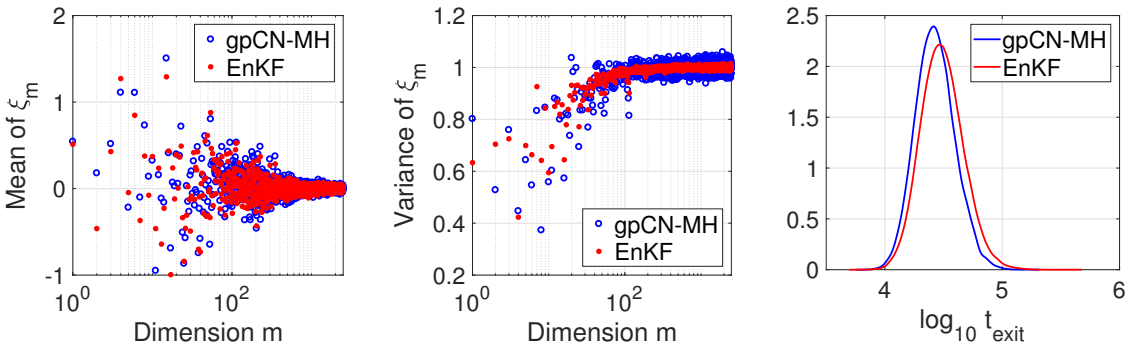

Figure 7.13.: Comparison between the results obtained by gpCN Metropolis and EnKF simulations for mean and variance of ξ_m (left and middle) and empirical PDF of $\log_{10} t_{\text{exit}}$ (right).

Test for synthetic data. Finally, we test how well we can recover a synthetic "true" log conductivity field $\log a^\dagger$ by Bayesian inference given corresponding synthetic observational data. We generate the true log conductivity field $\log a^\dagger$ by sampling a Gaussian random field with constant mean $m(x) \equiv -11$ and covariance function $c_{\sigma^2,\rho}$ as in (7.7) with parameters $\sigma^2 = \hat{\sigma}^2 = 25.78$ and $\rho = \hat{\rho} = 17,665$, i.e., the ReML estimates for σ^2 and ρ obtained from the original WIPP data. We generate $\log a^\dagger$ on a fine triangular mesh of mesh size $h = 200$ (55,874 elements) and solve also the PDE (7.4) and the ODE (7.2) on that mesh, resulting, e.g., in a true pressue head $p^\dagger$. As synthetic observational data we take the values of $\log a^\dagger$ at exactly the same 38 measurement locations of the original WIPP transmissivity data as well as

perturbed values of $p^\dagger$ at the same 33 measurement locations of the original WIPP pressure head data. The perturbations for the pressure head data were simulated by sampling from the assumed Gaussian noise model $\varepsilon \sim N(0, I_{33})$. Then, we apply the same procedure as for the original WIPP data: we perform universal kriging using the synthetic log conductivity measurements to obtain a prior mean and covariance function, compute and truncate the resulting KLE of the prior random field and perform Bayesian inference by conditioning the M random coefficients in the truncated KLE on the synthetic pressure head data. We omit another ReML estimation of the parameters σ^2 and ρ in the exponential covariance model (7.7) based on the synthetic log conductivity data and simply use the parameters $\sigma^2 = 25.78$ and $\rho = 17,665$. Moreover, since the universal kriging error does not depend on the actual observed data, but only on the measurement locations, the covariance function resulting from universal kriging given the synthetic data is the same as for the original WIPP data. Hence, also the KL eigenfunctions and eigenvalues remain the same and only the prior mean, i.e., the universal kriging prediction is different. For the Bayesian inference we run again a MCMC simulation emplying a mesh of mesh size $h = 350$ (18,142 elements) for evaluating the forward map and use as before a truncated KLE of length $M = 2,500$.

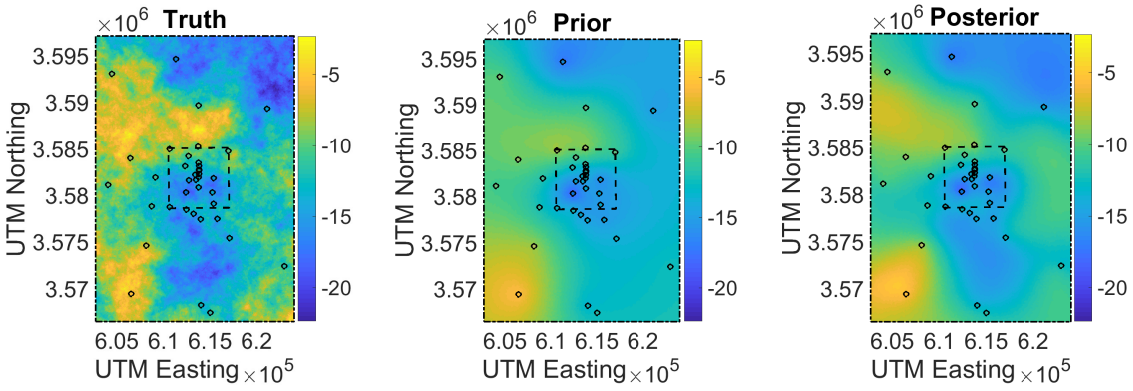

Figure 7.14.: Synthetic true log conductivity field, resulting prior and posterior mean field. The filled dots mark the location and value of the synthetic measurements.

We display the true log conductivity field, the corresponding prior mean field obtained by universal kriging and the resulting posterior mean field obtained by gpCN Metropolis simulation in Figure 7.14. We can observe an improvement from prior to posterior mean, i.e., some features of the true field are better resolved by the posterior mean field than by the prior mean field, e.g., the relatively low conductivity in the lower part.

Concerning the quantity of interest we present kernel density estimates for the PDF of $\log_{10} t_{\text{exit}}$ as well as particle paths resulting from prior and posterior sam-

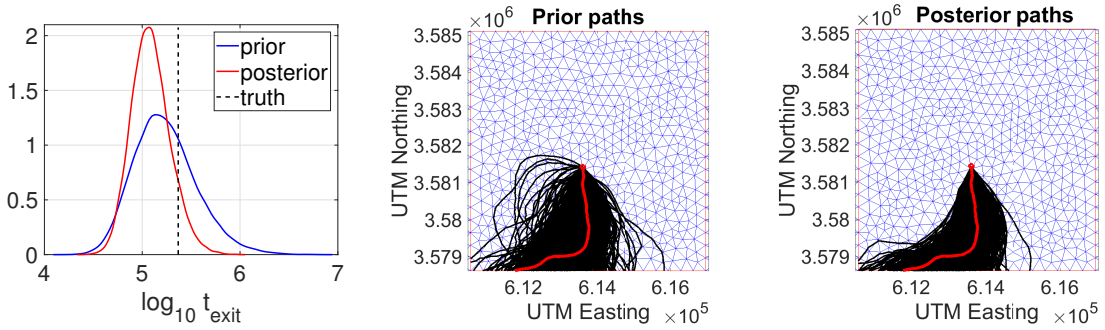

Figure 7.15.: Kernel density estimates of the prior and posterior PDF of $\log_{10} t_{\text{exit}}$ (left) and 1,000 particle paths resulting from prior (middle) and posterior (right) samples of $\log a_M$. The red curve in the particle path plots shows the particle path resulting from the true log conductivity field.

ples in Figure 7.15. In all plots of Figure 7.15 we also present the corresponding "truth", i.e., the value of t_{exit} and the associated particle path obtained for the true log conductivity field. Although the posterior yields a better prediction of the true particle path, the posterior distribution of $\log_{10} t_{\text{exit}}$ seems to concentrate on values smaller than the true exit time. If this is a coincidence or may have a particular reason, is an open question for future reseach.

Chapter 8
Conclusions and Outlook

In this thesis we provided a short introduction into uncertainty quantification for elliptic PDEs and made theoretical as well as algorithmic contributions to numerical methods for the associated inverse problem, i.e., for Bayesian inference in function spaces.

In particular, we presented in Chapter 2 the key points of the existing theory for the UQ forward problem, including random fields, function space-valued random variables and their relation as well as the basic approach to PDEs with random data and associated approximation methods. Then, we focused in Chapter 3 on Bayesian inference in general Hilbert spaces, explained how Bayes' rule and the posterior measure relates to conditional measures and presented some stability results for the posterior which were slight extension of the corresponding results given by Stuart [167]. We also recalled the concept of Bayes estimators in Chapter 3 which proved useful in the analysis and understanding of the EnKF and PCKF in Chapter 4. There we showed that in the general Bayesian inference setting, given mild assumptions on prior, noise and forward map,

- the random variable generated by one update of the PCKF converges in the large polynomial basis limit in the L^2-sense to the analysis variable associated to the Bayesian inference problem,

- the empirical measure associated to the analysis ensemble generated by one update of the EnKF converges in the large ensemble limit almost surely weakly to the distribution of the analysis variable,

- the analysis variable coincides with the linear conditional mean estimate for the given observational data plus the prior random error of the linear conditional mean estimator.

These facts imply that both methods, EnKF and PCKF, are not suitable methods for the UQ inverse problem, since the distribution of the analysis variable can differ

significantly from the posterior measure. For example, only the mean of the analysis variable – and, thus, also the mean of the analysis ensemble and of the random variable provided by the PCKF – depends on the observational data. Indeed, EnKF and PCKF can be viewed as numerical methods for the approximation of the posterior mean which are typically cheaper than MCMC integration. However, there are still several questions open for future research, e.g.,

- How good is the estimate for the posterior mean provided by the EnKF and PCKF? Is it possible to establish error bounds?
- How do the EnKF and PCKF behave if we perform several updates for the same data? Does an iterative procedure as presented by Iglesias et al. [92] and Schillings and Stuart [155] improve the estimate for the posterior mean?
- Can we derive reasonable bounds for the difference between the posterior measure and the distribution of the analysis variable under suitable assumptions?

The latter question aims at characterizing situations in which EnKF and PCKF may be applied to get cheap yet reasonable approximations of the posterior measure.

In Chapter 5 we outlined the MCMC method for approximate sampling of posterior measures and explained how Metropolis-Hastings algorithms can be defined in infinite-dimensional Hilbert spaces. We then presented a generalization of the pCN Metropolis algorithm which allows to use (approximations of) the posterior covariance for proposing new states, since the latter potentially yield a higher statistical efficiency. In particular, we showed

- the well-definedness of the resulting gpCN Metropolis algorithm in separable Hilbert spaces,
- the $L^2_{\mu_R}$-geometric ergodicity of the restricted gpCN Metropolis algorithm targeting arbitrarily close approximations μ_R of the posterior measure μ,
- the higher efficiency of the gpCN compared to the pCN Metropolis algorithm in case of a simple but common Bayesian inference problem.

Our approach to combine infinite-dimensional MCMC algorithms with the idea of "geometric MCMC", i.e., MCMC methods exploiting geometric information about the posterior measure such as covariance or local curvature, is one of several recent contributions to this topic [37, 104, 12]. In particular, we focus on the rather simple pCN Metropolis whereas other authors consider generalizations of MALA or HMC algorithms to infinite dimensions. However, we provide well-definedness of the

algorithm under very mild conditions which are easy to verify, i.e., the operator Γ in the definition of the gpCN proposal just has to be bounded, self-adjoint and positive. Moreover, due to the special structure of our gpCN proposal we were able to establish a geometric ergodicity result for our MH algorithm whereas convergence results for the proposed algorithms are missing in the above publications. Again, there are various open question we would like to answer in the future:

- Can we rigorously prove, given reasonable assumptions, that for suitable Γ the gpCN Metropolis algorithm has a higher (statistical) efficiency than the pCN Metropolis algorithm, e.g., in terms of spectral gaps or the ESS?
- Does there exist an "optimal" Γ, i.e., an operator Γ^* such that the resulting gpCN Metropolis performs better (again in terms of larger spectral gaps or ESS) than for any other admissible choice of Γ?
- Can we show geometric ergodicity of the local gpCN and local pCN Metropolis algorithm?
- How do these local variants of the gpCN and pCN Metropolis relate and perform w.r.t. the algorithms proposed by Beskos et al. [12]?
- Can we reduce the computational cost for the local gpCN and local pCN Metropolis algorithm by appyling Krylov methods and random determinant estimators? (See, e.g., Saibaba et al. [150] for the latter.)

Motivated by the numerical results presented in Chapter 5, we investigated in Chapter 6 how Metropolis-Hastings algorithms perform when the target measures becomes more concentrated. This is an important question, since, for example, a highly concentrated posterior measure means a small remaining uncertainty about the unknown or, equivalently, that we have incorporated highly informative data. Surprisingly, this question has not yet drawn much attention within the statistics or UQ community. In Chapter 6 we presented a first attempt to define and prove a variance independent performance of Metropolis-Hastings algorithms where the term variance refers to the (decreasing) variance of the noise corrupting the observational data used for Bayesian inference. In particular, we provided

- a discussion about several concepts for variance independent performance resulting in a definition of variance independent expected squared jump distance,
- a theorem about the variance independence of the expected squared jump distance and expected acceptance probability of the Gaussian random walk

and the gpCN Metropolis algorithm if both apply the covariance matrix of a Gaussian target measure for proposing new states.

Numerical experiments suggested that the latter statement holds also for non-Gaussian targets provided the target measure concentrates on a linear subspace. Open issues for future research are

- the extension of the established theorem about variance independent performance to non-Gaussian posterior measures which concentrate on linear subspaces for vanishing observational noise,
- a better understanding of this "subspace condition" and a resulting approach to analyze variance independent performance under more general conditions,
- and, based on that an analysis, guidelines on how to develop variance robust MCMC methods.

Finally, we applied the methods studied in the previous chapters to a real-world UQ problem in Chapter 7. We explained how geostatistical methods can be used to build prior random field models given measurement data and showed the effects of Bayesian inference for groundwater flow and related quantities of intest. We have seen that taking into account noisy pressure head measurements can yield a significantly reduced uncertainty about the transport of polutants by groundwater flow. Moreover, the proposed gpCN Metropolis provided also in this real-world application a substantial computational gain compared to the pCN Metropolis algorithm. Of course, we can think of further approaches to reduce the computational cost of the MCMC simulation, e.g.,

- by applying multilevel MCMC methods, see, e.g., Dodwell et al. [46],
- by applying reduced basis methods to solve the involved PDE, see, e.g., Quarteroni et al. [135],
- by determining the, hopefully small, subspace of those coefficients in the KLE of the random log conductivity which are significantly affected by the conditioning on the observed pressure head data, cf. the approach of Cui et al. [37], and inferring only those coefficents where we could then apply surrogate techniques to approximate the forward map – given the number of relevant coefficients is moderate.

Hence, this thesis provided several contributions to numerical methods for inverse problems in UQ or Bayesian inference in function spaces, respectively, which, in turn, also lead to interesting questions for future research.

Appendix A
Spaces of Linear Operators

We present the definitions of trace class and Hilbert-Schmidt operators on Hilbert spaces and state some related results. Subsequently, let $\mathcal{H}$ and $\mathcal{H}_i$, $i = 1,2$, be a separable Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}_i}$, $i = 1,2$, respectively. The norms induced by these inner products are denoted by $\| \cdot \|_{\mathcal{H}}$ and $\| \cdot \|_{\mathcal{H}_i}$, $i = 1,2$, respectively. We recall that the space of all *bounded linear operators* $A \colon \mathcal{H}_1 \to \mathcal{H}_2$ is denoted by $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and that $\mathcal{L}(\mathcal{H}) = \mathcal{L}(\mathcal{H}, \mathcal{H})$.

Definition A.1 (Positive operators)**.** A linear operator $A \colon \mathcal{H} \to \mathcal{H}$ is called *positive* if for each $x \in \mathcal{H}$ there holds $\langle x, Ax \rangle \geq 0$. We denote the space of all bounded, positive and self-adjoint linear operators by $\mathcal{L}_+(\mathcal{H})$.

Definition A.2 (Nuclear operators)**.** A linear operator $A \colon \mathcal{H}_1 \to \mathcal{H}_2$ is called *nuclear* if there exist $\{a_n : n \in \mathbb{N}\} \subset \mathcal{H}_1$ and $\{b_n : n \in \mathbb{N}\} \subset \mathcal{H}_2$ such that

$$Au = \sum_{n=1}^{\infty} \langle a_n, u \rangle_{\mathcal{H}_1} b_n \qquad \forall u \in \mathcal{H}_1,$$

and

$$\sum_{n=1}^{\infty} \|a_n\|_{\mathcal{H}_1} \|b_n\|_{\mathcal{H}_2} < \infty.$$

We denote the space of all nuclear operators $A \colon \mathcal{H}_1 \to \mathcal{H}_2$ by $\mathcal{L}^1(\mathcal{H}_1, \mathcal{H}_2)$ and define on $\mathcal{L}^1(\mathcal{H}_1, \mathcal{H}_2)$ the *nuclear norm*

$$\|A\|_1 := \inf \left\{ \sum_{n=1}^{\infty} \|a_n\|_{\mathcal{H}_1} \|b_n\|_{\mathcal{H}_2} \, : \, Au = \sum_{n=1}^{\infty} \langle a_n, u \rangle_{\mathcal{H}_1} b_n \quad \forall u \in \mathcal{H}_1 \right\}.$$

We state some properties of nuclear operators provided by Peszat and Zabczyk [132]:

Theorem A.3 ([132, Appendix A.2])**.** There holds:

1. The linear space $\mathcal{L}^1(\mathcal{H}_1, \mathcal{H}_2)$ equipped with the norm $\| \cdot \|_1$ is a Banach space.

2. If $A \in \mathcal{L}^1(\mathcal{H}, \mathcal{H}_1)$ and $B \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, then $AB \in \mathcal{L}^1(\mathcal{H}, \mathcal{H}_2)$. The same holds, if $A \in \mathcal{L}(\mathcal{H}, \mathcal{H}_1)$ and $B \in \mathcal{L}^1(\mathcal{H}_1, \mathcal{H}_2)$.

3. Every nuclear operator is compact.

Definition A.4 (Trace of an operator)**.** A nuclear operator $A\colon \mathcal{H} \to \mathcal{H}$ is also called *trace class* and its *trace* is defined by

$$\operatorname{tr}(A) := \sum_{n=1}^{\infty} \langle e_n, Ae_n \rangle_{\mathcal{H}}.$$

Moreover, we set $\mathcal{L}^1(\mathcal{H}) := \mathcal{L}^1(\mathcal{H}, \mathcal{H})$ and $\mathcal{L}^1_+(\mathcal{H}) := \mathcal{L}^1(\mathcal{H}) \cap \mathcal{L}_+(\mathcal{H})$.

The definition of $\operatorname{tr}(A)$ is indeed independent of the employed CONS, see [132, Appendix A.2]. Moreover, if $A \in \mathcal{L}(\mathcal{H})$ is self-adjoint and compact, then A is trace class iff its eigenvalues λ_n satisfy $(\lambda_n)_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$, see Reed and Simon [138, Theorem VI.22].

Definition A.5 (Fredholm determinant)**.** Let $A \in \mathcal{L}^1(\mathcal{H})$ be self-adjoint with eigenvalues λ_n, $n \in \mathbb{N}$. Then its *(Fredholm) determinant* is given by

$$\det(I + A) := \prod_{n=1}^{\infty} (1 + \lambda_n).$$

Due to the trace class property and $\ln|1 + x| \leq |x|$ for $x \neq -1$ the determinant $\det(I + A)$ is indeed finite: assume that $\lambda_n \neq -1$ for each $n \in \mathbb{N}$, otherwise we have $\det(I + A) = 0$, then

$$\begin{aligned} |\det(I + A)| &= \prod_{n=1}^{\infty} |1 + \lambda_n| = \exp\left(\sum_{n=1}^{\infty} \ln|1 + \lambda_n| \right) \leq \exp\left(\sum_{n=1}^{\infty} |\lambda_n| \right) \\ &= \exp(\operatorname{tr}(A)) < \infty. \end{aligned}$$

Definition A.6 (Hilbert-Schmidt operators)**.** A linear operator $A\colon \mathcal{H}_1 \to \mathcal{H}_2$ is *Hilbert-Schmidt* if

$$\sum_{n=1}^{\infty} \|Ae_n\|^2_{\mathcal{H}_2} < \infty,$$

where $\{e_n\}_{n\in\mathbb{N}}$ denotes an arbitrary CONS for $\mathcal{H}_1$. The space of all Hilbert-Schmidt operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ is denoted by $\mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2)$ and we set $\mathcal{L}^2(\mathcal{H}) := \mathcal{L}^2(\mathcal{H}, \mathcal{H})$ as well as $\mathcal{L}^2_+(\mathcal{H}) := \mathcal{L}^2(\mathcal{H}) \cap \mathcal{L}_+(\mathcal{H})$.

Again, we state some basic results about Hilbert-Schmidt operators.

Theorem A.7 ([132, Appendix A.2], [138, Theorem VI.22], [156, Proposition B.22])**.** There holds:

1. Nuclear operators are Hilbert-Schmidt operators.
2. The linear space $\mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2)$ equipped with the inner product

$$\langle A, B\rangle_{\mathrm{HS}} := \sum_{n=1}^{\infty} \langle Ae_n, Be_n\rangle_{\mathcal{H}_2},$$

   where $\{e_n\}_{n\in\mathbb{N}}$ denotes a CONS for $\mathcal{H}_1$, is a Hilbert space. In particular, $\langle A, B\rangle_{\mathrm{HS}}$ does not depend on the choice of the CONS.
3. $A \in \mathcal{L}^1(\mathcal{H})$ if and only if $A = BC$ with $B, C \in \mathcal{L}^2(\mathcal{H})$.
4. If $A \in \mathcal{L}^2(\mathcal{H})$ and $B \in \mathcal{L}(\mathcal{H})$, then $AB, BA \in \mathcal{L}^2(\mathcal{H})$
5. Every Hilbert-Schmidt operator is compact. Moreover, a self-adjoint compact operator $A \in \mathcal{L}(\mathcal{H})$ is Hilbert-Schmidt iff its eigenvalues λ_n satisfy $(\lambda_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$.

Appendix B

Tensor Products of Hilbert Spaces

We provide the basic details about tensor products of Hilbert spaces. The presentation mainly follows Reed and Simon [138] and several results are taken from Schwab and Gittelson [156]. For a more detailed introduction, particularly, to tensor products of Banach spaces, we refer to Light and Cheney [111]. In the following let $\mathcal{H}_1$ and $\mathcal{H}_2$ denote two Hilbert spaces with inner products and norms $\langle\cdot,\cdot,\rangle_{\mathcal{H}_i}$ and $\|\cdot\|_{\mathcal{H}_i}$, $i = 1, 2$, respectively.

Definition B.1 (Tensor product)**.** For each $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$ we define $x \otimes y$ to be a bilinear mapping $x \otimes y \colon \mathcal{H}_1 \times H_2 \to \mathbb{R}$ given by

$$(x \otimes y)(u, v) := \langle x, u\rangle_{\mathcal{H}_1} \langle y, v\rangle_{\mathcal{H}_2}, \qquad u \in \mathcal{H}_1,\, v \in \mathcal{H}_2.$$

The *tensor product* $\mathcal{H}_1 \otimes \mathcal{H}_2$ is then the completion of all finite linear combinations of such bilinear forms w.r.t. the inner product

$$\langle x_1 \otimes y_1, x_2 \otimes y_2\rangle_{\mathcal{H}_1\otimes\mathcal{H}_2} := \langle x_1, x_2\rangle_{\mathcal{H}_1} \langle y_1, y_2\rangle_{\mathcal{H}_2}, \qquad x_1, x_2 \in \mathcal{H}_1,\, y_1, y_2 \in \mathcal{H}_2.$$

We obtain for the norm $\|\cdot\|_{\mathcal{H}_1\otimes\mathcal{H}_2}$ induced by $\langle\cdot,\cdot\rangle_{\mathcal{H}_1\otimes\mathcal{H}_2}$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$ that

$$\|x \otimes y\|_{\mathcal{H}_1\otimes\mathcal{H}_2} = \|x\|_{\mathcal{H}_1} \|y\|_{\mathcal{H}_2}, \qquad x \otimes y \in \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{B.1}$$

As one might suppose, the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is again separable if $\mathcal{H}_1$ and $\mathcal{H}_2$ are:

Proposition B.2 ([138, Proposition II.4.2])**.** If $\{e_m : m \in \mathbb{N}\}$ and $\{f_n : n \in \mathbb{N}\}$ are CONS for $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, then $\{e_m \otimes f_n : m, n \in \mathbb{N}\}$ is a CONS for $\mathcal{H}_1 \otimes \mathcal{H}_2$.

As usual, we can identify the bilinear mapping $x \otimes y \in \mathcal{H}_1 \otimes \mathcal{H}_2$ with a bounded

linear operator from $\mathcal{H}_2$ to $\mathcal{H}_1$ which we will denote again by $x \otimes y$:

$$(x \otimes y)z := \langle y, z \rangle_{\mathcal{H}_2} \, x, \qquad z \in \mathcal{H}_2, \tag{B.2}$$

For the linear operator $x \otimes y \in \mathcal{L}(\mathcal{H})$ we obtain

$$\|x \otimes y\| = \|x\|_{\mathcal{H}_1} \, \|y\|_{\mathcal{H}_2} = \|x \otimes y\|_{\mathcal{H}_1 \otimes \mathcal{H}_2},$$

where $\|\cdot\|$ denotes the usual operator norm in $\mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$. Actually, there holds the following isomorphism.

Proposition B.3 ([156, Section B.4])**.** The tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ is isomorphic to the Hilbert space $\mathcal{L}^2(\mathcal{H}_2, \mathcal{H}_1)$ of Hilbert-Schmidt operators from $\mathcal{H}_2$ to $\mathcal{H}_1$ as defined in Appendix A, i.e.,

$$\mathcal{H}_1 \otimes \mathcal{H}_2 \simeq \mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2).$$

We now state some further isomorphisms which are useful to us.

Theorem B.4 ([156, Theorem B.17], [111, Theorem 1.39])**.** Let $(\Omega, \mathcal{A}, \mu)$ denote a measure space and $\mathcal{H}$ a Hilbert space, then there holds

$$L^2_\mu(\Omega; \mathcal{H}) \simeq L^2_\mu(\Omega; \mathbb{R}) \otimes \mathcal{H}.$$

Furthermore, let $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$, denote two measure spaces, then we have

$$L^2_{\mu_1 \otimes \mu_2}(\Omega_1 \times \Omega_2; \mathbb{R}) \simeq L^2_{\mu_1}(\Omega_1; \mathbb{R}) \otimes L^2_{\mu_2}(\Omega_2; \mathbb{R}).$$

A corollary of the above results is then

Corollary B.5 ([138, Theorem VI.23])**.** Let $(\Omega, \mathcal{A}, \mu)$ denote a measure space and set $\mathcal{H} = L^2_\mu(\Omega; \mathbb{R})$. Then $A \in \mathcal{L}(\mathcal{H})$ is Hilbert-Schmidt iff there exists a $k \in L^2(\mu \otimes \mu)(\Omega \times \Omega; \mathbb{R})$ such that

$$Af(x) = \int_\Omega k(x, y) \, f(y) \, \mu(\mathrm{d}y), \qquad f \in \mathcal{H}.$$

Appendix C
Equivalence of Gaussian Measures on Hilbert Spaces

The following is a collection of useful results about the equivalence of Gaussian measures $\mu_1 = N(m_1, C_1)$, $\mu_2 = N(m_2, C_2)$ on a separable Hilbert space $\mathcal{H}$ given in Da Prato and Zabczyk [39, Section 1]. As before, let $\|\cdot\|_{\mathcal{H}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ denote the norm and inner-product in $\mathcal{H}$.

Definition C.1 (Cameron-Martin space)**.** The *Cameron-Martin space* $\mathcal{H}_\mu$ of a Gaussian measure $\mu = N(m, C)$ on $\mathcal{H}$ is defined as the range of $C^{1/2}$, i.e., $\mathcal{H}_\mu := \mathrm{rg}(C^{1/2})$, equipped with the inner product

$$\langle u, v\rangle_{C^{-1}} := \langle C^{-1/2}u, C^{-1/2}v\rangle_{\mathcal{H}}, \qquad \forall u, v \in \mathrm{rg}(C^{1/2}).$$

It is easy to see, that the Cameron-Martin space is again Hilbert space and that $\mathcal{H}_\mu$ is dense in $\mathcal{H}$ if C is nonsingular. Moreover, $\mathcal{H}_\mu$ can be characterized as the intersection of all measurable linear subspaces $\mathcal{X} \subseteq \mathcal{H}$ with $\mu(\mathcal{X}) = 1$, but if $\mathcal{H}$ is infinite dimensional, then $\mu(\mathcal{H}_\mu) = 0$. We refer to Hairer [80, Proposition 3.42] for a proof. The space $\mathcal{H}_\mu$ plays a crucial role for the equivalence of Gaussian measures as rigorously expressed in the Cameron-Martin theorem below. Before stating the result we need some more notation.

Definition C.2. Let $\mu = N(0, C)$ be a Gaussian measure on $\mathcal{H}$. Then, for each $u \in \mathcal{H}_\mu$ we define a random variable $W_u \colon (\mathcal{H}, \mathcal{B}(\mathcal{H})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by

$$W_u(v) := \langle C^{-1/2}u, v\rangle_{\mathcal{H}}, \qquad v \in \mathcal{H}.$$

By applying the Cauchy-Schwarz inequality we can verify that $W_u \in L^2_\mu(\mathcal{H}; \mathbb{R})$. Moreover, the mapping $\mathcal{H}_\mu \ni u \mapsto W_u \in L^2_\mu(\mathcal{H}; \mathbb{R})$ is an isometry, see [39, Section 1.2.4]. This allows to state the following

Definition C.3. Given a Gaussian measure $\mu = N(0, C)$ on $\mathcal{H}$ and $u \in \mathcal{H}$, let

$(u_n)_{n\in\mathbb{N}}$ denote a sequence in $\mathcal{H}_\mu$ such that $\lim_{n\to\infty}\|u_n - u\|_{\mathcal{H}} = 0$. We then denote the limit of the random variables $(W_{u_n})_{n\in\mathbb{N}}$ in $L^2_\mu(\mathcal{H};\mathbb{R})$ by

$$\langle C^{-1/2}u, \cdot\rangle := L^2_\mu(\mathcal{H};\mathbb{R})\text{-}\lim_{n\to\infty} W_{u_n}.$$

Moreover, for each $u \in \mathcal{H}_\mu$ we set $\langle C^{-1}u, \cdot\rangle := \langle C^{-1/2}(C^{-1/2}u), \cdot\rangle$.

We note, that the definition of $\langle C^{-1/2}u, \cdot\rangle$ is independent of the choice of the sequence $(u_n)_{n\in\mathbb{N}}$ due to the isometry mentioned above. However, we highlight, that for $v \in \mathcal{H}$ there holds, in general, $\langle C^{-1/2}u, v\rangle \neq \lim_{n\to\infty}\langle C^{-1/2}u_n, v\rangle_{\mathcal{H}}$, since the pointwise limit does not necessarily exist.

Proposition C.4 ([39, Proposition 1.2.7])**.** Let $\mu = N(0,C)$ be a Gaussian measure on $\mathcal{H}$. Then there holds

$$\int_{\mathcal{H}} \mathrm{e}^{\langle C^{-1/2}u, v\rangle}\ \mu(\mathrm{d}v) = \mathrm{e}^{\frac{1}{2}\|u\|^2_{\mathcal{H}}} \qquad \forall u \in \mathcal{H}. \tag{C.1}$$

Theorem C.5 (Cameron-Martin formula, [39, Theorem 1.3.6])**.** Let $\mu = N(0,C)$ and $\mu_h = N(h,C)$ be Gaussian measures on $\mathcal{H}$. Then, μ and μ_h are equivalent iff $h \in \mathcal{H}_\mu = \mathrm{rg}(C^{1/2})$ in which case

$$\frac{\mathrm{d}\mu_h}{\mathrm{d}\mu}(v) = \exp\left(-\frac{1}{2}\|C^{-1/2}h\|^2_{\mathcal{H}} + \langle C^{-1}h, v\rangle\right).$$

Thus, two Gaussian measures $N(m,C)$ and $N(m+h,C)$ are equivalent only if $h \in \mathrm{rg}(C^{1/2})$. Next, we will consider the equivalence of two Gaussian measures $\mu = N(0,C)$ and $\nu = N(0,Q)$ with $C \neq Q$. Before we state the corresponding results, we introduce again some more notations.

Definition C.6. Let $\mu = N(0,C)$ be a Gaussian measure on $\mathcal{H}$ and $T \in \mathcal{L}^1(\mathcal{H})$ be a self-adjoint with eigenvalues t_n, $n \in \mathbb{N}$. Then, we define

$$\langle TC^{-1/2}u, C^{-1/2}u\rangle := \lim_{N\to\infty}\langle TC^{-1/2}\,\Pi_N u, C^{-1/2}\,\Pi_N u\rangle_{\mathcal{H}}, \qquad \mu\text{-a.e.},$$

where Π_N denotes the orthogonal projection to the span of first N eigenvectors $e_1, \ldots, e_N$ of C.

The existence of the μ-a.e.-limit in Definition C.6 is proven in Da Prato and Zabczyk [39, Proposition 1.2.10]. Moreover, there holds

Proposition C.7 ([39, Proposition 1.2.11])**.** Let $\mu = N(0,C)$ denote a Gaussian measure on $\mathcal{H}$ and $T \in \mathcal{L}^1(\mathcal{H})$ be self-adjoint and such that $\langle Tu, u\rangle_{\mathcal{H}} < \|u\|^2_{\mathcal{H}}$ for each

$u \in \mathcal{H}$. Then, we have

$$\int_{\mathcal{H}} \mathrm{e}^{\frac{1}{2}\langle TC^{-1/2}u, C^{-1/2}u\rangle} \, \mathrm{d}\mu(u) = \frac{1}{\sqrt{\det(I-T)}}. \tag{C.2}$$

For the definition of the determinant $\det(I-T)$ we refer to Appendix A and note, that $\det(I-T) \neq 0$ follows by the assumption $\langle Tu, u\rangle_{\mathcal{H}} < \|u\|_{\mathcal{H}}^2$, i.e., all eigenvalues of T are smaller than 1.

Theorem C.8 ([39, Proposition 1.3.11])**.** Let $\mu = N(0, C)$ and $\nu = N(0, Q)$ be Gaussian measures on $\mathcal{H}$. If $T := I - C^{-1/2}QC^{-1/2}$ is self-adjoint, trace class and satisfies $\langle Tu, u\rangle_{\mathcal{H}} < \|u\|_{\mathcal{H}}^2$ for each $u \in \mathcal{H}$, then μ and ν are equivalent with

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(u) = \frac{1}{\sqrt{\det(I-T)}} \exp\left(-\frac{1}{2}\langle T(I-T)^{-1} C^{-1/2}u, C^{-1/2}u\rangle\right).$$

The assumptions of Theorem C.8 can be relaxed to $I - C^{-1/2}QC^{-1/2}$ being Hilbert-Schmidt which is known as the *Feldman-Hajek theorem*. Also in this case explicit expression of the Radon-Nikodym derivative can be established, see Bogachev [18, Corollary 6.4.11].

Finally, we recall two simple but useful facts resulting from a change of variables.

Lemma C.9. Let $0 < s < \infty$ and $h \in \mathcal{H}$. Then, for $\mu = N(m, C)$ and $\nu = N(m + h, s^2C)$ there holds

$$\int_{\mathcal{H}} f(v)\mu(\mathrm{d}v) = \int_{\mathcal{H}} f\left(\frac{1}{s}(v-h)\right) \nu(\mathrm{d}v), \qquad f\colon \mathcal{H} \to \mathbb{R}.$$

Moreover, if $\mu_1 = N(m_1, C_1)$ and $\mu_2 = N(m_2, C_2)$ are equivalent with $\frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_1}(u) = \pi(u)$, then the measures $\nu_1 = N(m_1 + h, s^2C_1)$ and $\nu_2 = N(m_2 + h, s^2C_2)$ are also equivalent with

$$\frac{\mathrm{d}\nu_2}{\mathrm{d}\nu_1}(u) = \pi\left(\frac{u-h}{s}\right).$$

Appendix D

Kriging

The term *kriging* is due to the French founder of the field of geostatistics Georges Matheron. It refers to a technique for predicting the value of a random field a given its covariance function and a set of obervations of a at known spatial locations. The method was first introduced by the South African mining engineer D. G. Krige [101] and yields predictions which are linear in the observations, unbiased, and optimal in the sense of minimizing the mean squared error (MSE). Thus, the *kriging prediction* represents the best unbiased linear prediction (BLUP) or best unbiased linear estmimatior (BLUE). Moreover, the kriging prediction coincides with a suitable kernel interpolation given the observed data, see Scheurer et al. [152] for a discussion. However, we will follow the geostatistical point of view in the following and outline two common kriging variants, *simple* and *universal* kriging. We will also provide interpretations of both methods in the light of uncertainty quantification. For a general introduction to kriging we refer to Chilès and Delfiner [29] or Stein [164].

Simple kriging

Let a be a second-order random field on $D \subseteq \mathbb{R}^d$ with known mean $m(x) = \mathbb{E}[a(x)]$ and known covariance function $c(x,y) = \mathrm{Cov}(a(x), a(y))$. Further, assume that we can observe a at n spatial locations $x_j \in D$, $j = 1, \dots, n$. The goal is then to construct a linear predictor (or estimator)

$$\hat{a}(x,\omega) = \lambda_0(x) + \boldsymbol{\lambda}(x)^\top \boldsymbol{a}(\omega), \qquad \boldsymbol{a}(\omega) := (a(x_1,\omega), \dots, a(x_n,\omega))^\top, \tag{D.1}$$

where $\boldsymbol{a}$ denotes the random vector of the (yet unobserved) values of $a(x_j)$, $j = 1, \dots, n$, and $\lambda_0 \colon D \to \mathbb{R}$ as well as $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_n) \colon D \to \mathbb{R}^n$ denote weighing functions to be determined such that for each $x \in D$ and any other $\nu_0 \colon D \to \mathbb{R}^n$ and

$\boldsymbol{\nu}\colon D \to \mathbb{R}^n$ we have

$$\mathbb{E}\left[a(x)\right] = \mathbb{E}\left[\hat{a}(x)\right], \quad \mathbb{E}\left[\left(a(x) - \hat{a}(x)\right)^2\right] \leq \mathbb{E}\left[\left(a(x) - \nu_0(x) - \boldsymbol{\nu}(x)^\top \boldsymbol{a}\right)^2\right]. \quad \text{(D.2)}$$

The first equation expresses the unbiasedness of the predictor and yields

$$\lambda_0(x) = m(x) - \boldsymbol{\lambda}^T(x)\boldsymbol{m}, \qquad \boldsymbol{m} := \left(m(x_1), \ldots, m(x_n)\right)^\top.$$

The mean squared error then reads

$$\begin{aligned}
\mathbb{E}\left[\left(a(x) - \hat{a}(x)\right)^2\right] &= \operatorname{Var}\left(a(x) - \hat{a}(x)\right) = \operatorname{Var}\left(a(x)\right) + \operatorname{Var}\left(\hat{a}(x)\right) - 2\operatorname{Cov}\left(a(x), \hat{a}(x)\right) \\
&= \operatorname{Var}\left(a(x)\right) + \operatorname{Var}\left(\boldsymbol{\lambda}(x)^\top \boldsymbol{a}\right) - 2\operatorname{Cov}\left(a(x), \boldsymbol{\lambda}(x)^\top \boldsymbol{a}\right) \\
&= c(x,x) + \boldsymbol{\lambda}^T(x)\boldsymbol{C}\boldsymbol{\lambda}(x) - 2\boldsymbol{\lambda}^T(x)\boldsymbol{c}(x),
\end{aligned}$$

where

$$\boldsymbol{C} := \left[c(x_i, x_j)\right]_{i,j=1}^n, \qquad \boldsymbol{c}(x) := \left(c(x, x_1), \ldots, c(x, x_n)\right)^\top.$$

Thus, we have to minimize the quadratic form $\boldsymbol{\lambda}(x)^\top \boldsymbol{C}\boldsymbol{\lambda}(x) - 2\boldsymbol{\lambda}(x)^\top \boldsymbol{c}(x)$ which is obtained by

$$\boldsymbol{\lambda}(x) = \boldsymbol{C}^{-1}\boldsymbol{c}(x), \qquad x \in D.$$

Hence, we end up with

Definition D.1 (Simple Kriging prediciton and error)**.** For a second-order random field a on $D \subseteq \mathbb{R}^d$ with mean function $m\colon D \to \mathbb{R}$ and covariance function $c\colon D \times D \to \mathbb{R}$ the *simple kriging prediction* or *simple kriging mean* based on the observation $\boldsymbol{a} = \left(a(x_1), \ldots, a(x_n)\right)^\top$ is given by

$$\hat{a}_{\mathrm{sk}}(x) := m(x) + \boldsymbol{c}^\top(x)\boldsymbol{C}^{-1}\left(\boldsymbol{a} - \boldsymbol{m}\right) \quad \text{(D.3)}$$

with $\boldsymbol{m}$, $\boldsymbol{c}$ and $\boldsymbol{C}$ as above, and the corresponding *simple kriging (error) covariance* function is

$$c_{\mathrm{sk}}(x,y) := \operatorname{Cov}(a(x) - \hat{a}_{\mathrm{sk}}(x), a(y) - \hat{a}_{\mathrm{sk}}(y)) = c(x,y) - \boldsymbol{c}(x)^\top \boldsymbol{C}^{-1}\boldsymbol{c}(y). \quad \text{(D.4)}$$

Remark D.2 (On interpolation)**.** We easily see that for $x = x_j$, $j = 1, \ldots, n$, we get

$$\hat{a}_{\mathrm{sk}}(x_j) = m(x_j) + \boldsymbol{c}^\top(x_j)\boldsymbol{C}^{-1}\left(\boldsymbol{a} - \boldsymbol{m}\right) = m(x_j) + \boldsymbol{e}_j^\top\left(\boldsymbol{a} - \boldsymbol{m}\right) = a(x_j),$$

where $\boldsymbol{e}_j$ denotes the jth unit vector in $\mathbb{R}^n$ and $\boldsymbol{C}^{-1}\boldsymbol{c}(x_j) = \boldsymbol{e}_j$ by construction.

Thus, the simple kriging mean indeed interpolantes the values of a at the locations x_j. Again, for the relation of (simple) kriging to kernel interpolation we refer to Scheurer et al. [152] and the references therein. Analogously, we get

$$c_{\mathrm{sk}}(x_j, x_j) = c(x_j, x_j) - \boldsymbol{c}(x_j)^\top \boldsymbol{C}^{-1} \boldsymbol{c}(x_j) = c(x_j, x_j) - \boldsymbol{c}(x_j)^\top \boldsymbol{e}_j = 0, \qquad j = 1, \dots, n.$$

Simple kriging and conditioning. Obviously, the minimization in (D.2) is the same one as for the linear conditional mean of $U := a(x)$ given observations of $Y := \boldsymbol{a} = (a(x_1), \dots, a(x_n))^\top$, see Section 4.3.1. Thus, the simple kriging prediction for $a(x)$ based on $\boldsymbol{a}$ provides the linear conditional mean estimate for $a(x)$ given $\boldsymbol{a}$ and $c_{\mathrm{sk}}(x, x)$ is equal to the correspoding (prior) error variance of the linear conditional mean estimator.

If a is a Gaussian random field, we know that the linear conditional mean coicides with the conditional mean, i.e., the simple kriging prediciton $\hat{a}(x)$ yields the posterior mean of $a(x)$ given the observation of $\boldsymbol{a}$. Moreover, by virtue of Theorem 4.3 the conditional distribution of $a(x)$ given $\boldsymbol{a}$ is again Gaussian and the corresponding variance coincides with $c_{\mathrm{sk}}(x, x)$. Thus, the Gaussian random field a_{sk} determined by the mean function $m_{\mathrm{sk}}(x) = \hat{a}_{\mathrm{sk}}(x)$ with $\hat{a}_{\mathrm{sk}}$ as in (D.3) and the covariance function c_{sk} as in (D.4) coincides with the conditioned random field a given $\boldsymbol{a}$ and is also called the *kriged* random field. We observe that the pointwise variance of a the (simple) kriged random field a_{sk} is always equal or smaller than the pointwise variance of the *unkriged* random field, since $\boldsymbol{C}^{-1}$ is positive semi-definite.

Universal kriging

An extension of simple kriging is to allow for an unknown mean function $m \colon D \to \mathbb{R}$ of the second-order random field a on D. However, we require m to follow a linear regression model of the form

$$m(x) = \boldsymbol{f}(x)^\top \boldsymbol{\beta}, \qquad \boldsymbol{f}(x) = (f_1(x), \dots, f_k(x))^\top, \tag{D.5}$$

where $f_j \colon D \to \mathbb{R}$, $j = 1, \dots, k$ are known *regression functions* and $\boldsymbol{\beta} \in \mathbb{R}^k$ yet unknown *regression coefficients* to be determined. Given observations $\boldsymbol{a} = (a(x_i))_{i=1}^n$ one usually estimates $\boldsymbol{\beta}$ by a (weighted) least squares approach

$$\hat{\boldsymbol{\beta}} := \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^k} |\boldsymbol{a} - \boldsymbol{F}\boldsymbol{\beta}|^2_{\boldsymbol{C}^{-1}}$$

with $\boldsymbol{C}$ as in the previous section and

$$\boldsymbol{F} := \begin{pmatrix} f_1(x_1) & \dots & f_k(x_1) \\ \vdots & & \vdots \\ f_1(x_n) & \dots & f_k(x_n) \end{pmatrix} \in \mathbb{R}^{n \times k}.$$

The weight matrix $\boldsymbol{C}^{-1}$ appears naturally in the least squares problem for $\hat{\boldsymbol{\beta}}$ given the assumed correlation structure of a. The least squares approach yields

$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = \left(\boldsymbol{F}^\top \boldsymbol{C}^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{F}^\top \boldsymbol{C}^{-1} \boldsymbol{a}.$$

Of course, given the estimate $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ we can perform simple kriging assuming the estimated mean function $\hat{m}(x) = \boldsymbol{f}(x)^\top \hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ to be true. However, for universal kriging we ask again for the best unbiased prediction $\hat{a}(x)$ of $a(x)$ which is linear w.r.t. $\boldsymbol{a}$ as in (D.1). Since the unbiasedness, i.e.,

$$\mathbb{E}\left[\hat{a}(x)\right] = \lambda_0(x) + \boldsymbol{\lambda}(x)^\top \boldsymbol{F}\boldsymbol{\beta},$$

requires that

$$\lambda_0(x) + \boldsymbol{\lambda}(x)^\top \boldsymbol{F}\boldsymbol{\beta} = \boldsymbol{f}(x)^\top \boldsymbol{\beta} \qquad \forall \boldsymbol{\beta} \in \mathbb{R}^k,$$

it follows that

$$\lambda_0(x) = 0, \qquad \boldsymbol{F}^\top \boldsymbol{\lambda}(x) = \boldsymbol{f}(x), \qquad x \in D. \tag{D.6}$$

Thus, the best linear unbiased predictor minimizes $\mathbb{E}\left[(a(x) - \hat{a}(x))^2\right]$ subject to (D.6). By introducing a Lagrange multiplier $\boldsymbol{\mu}(x) \in \mathbb{R}^k$, $x \in D$, this results in the unconstrained minimization of the Lagrange function

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\mu}) &:= \mathbb{E}\left[(a(x) - \hat{a}(x))^2\right] - 2\boldsymbol{\mu}(x)^\top \left(\boldsymbol{F}^\top \boldsymbol{\lambda}(x) - \boldsymbol{f}(x)\right), \\ &= c(x,x) + \boldsymbol{\lambda}^T(x)\, \boldsymbol{C}\boldsymbol{\lambda}(x) - 2\boldsymbol{\lambda}^T(x)\, \boldsymbol{c}(x) - 2\boldsymbol{\mu}(x)^\top \left(\boldsymbol{F}^\top \boldsymbol{\lambda}(x) - \boldsymbol{f}(x)\right) \end{aligned}$$

with $\boldsymbol{c}$ and $\boldsymbol{C}$ as in the case of simple kriging. The minimizer of L is then given by

$$\begin{bmatrix} \boldsymbol{\lambda}(x) \\ \boldsymbol{\mu}(x) \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F}^\top & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{c}(x) \\ \boldsymbol{f}(x) \end{bmatrix},$$

which yields the predicitor

$$\hat{a}(x) = \boldsymbol{\lambda}^\top \boldsymbol{a} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{a} \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}(x) \\ \boldsymbol{f}(x) \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F}^\top & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{a} \\ 0 \end{bmatrix}.$$

For the resulting error covariance

$$\operatorname{Cov}(a(x) - \hat{a}(x), a(y) - \hat{a}(y)) = c(x,y) - \boldsymbol{c}(x)^\top \boldsymbol{\lambda}(y) - \boldsymbol{\lambda}(x)^\top \boldsymbol{c}(y) + \boldsymbol{\lambda}(x)^\top \boldsymbol{C} \boldsymbol{\lambda}(y)$$

we obtain by

$$\begin{aligned}\boldsymbol{\lambda}(x)^\top \boldsymbol{c}(y) &= \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{c}(y)\\ \boldsymbol{0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}I\\ \boldsymbol{0}\end{bmatrix} \begin{bmatrix}\boldsymbol{c}(y)\\ \boldsymbol{f}(y)\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}I\\ \boldsymbol{0}\end{bmatrix} \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix} \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{c}(y)\\ \boldsymbol{f}(y)\end{bmatrix}\end{aligned}$$

and

$$\boldsymbol{\lambda}(x)^\top \boldsymbol{C} \boldsymbol{\lambda}(y) = \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{C} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix} \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{c}(y)\\ \boldsymbol{f}(y)\end{bmatrix},$$

the form

$$\operatorname{Cov}(a(x) - \hat{a}(x), a(y) - \hat{a}(y)) = c(x,y) - \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^T \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^T & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{c}(y)\\ \boldsymbol{f}(y)\end{bmatrix}.$$

Definition D.3 (Universal Kriging prediciton and error)**.** For a second-order random field a on $D \subseteq \mathbb{R}^d$ with a mean function $m\colon D \to \mathbb{R}$ of the form (D.5) and a known covariance function $c\colon D \times D \to \mathbb{R}$ the *universal kriging prediction* or *universal kriging mean* based on the observations of $\boldsymbol{a} = (a(x_1), \dots, a(x_n))^\top$ is given by

$$\hat{a}_{\text{uk}}(x) := \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^\top & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{a}\\ 0\end{bmatrix}. \tag{D.7}$$

with $\boldsymbol{f}$, $\boldsymbol{c}$, $\boldsymbol{F}$Âğ and $\boldsymbol{C}$ as above. The *universal kriging (error) covariance* is defined as the covariance function

$$c_{\text{uk}}(x,y) := c(x,y) - \begin{bmatrix}\boldsymbol{c}(x)\\ \boldsymbol{f}(x)\end{bmatrix}^T \begin{bmatrix}\boldsymbol{C} & \boldsymbol{F}\\ \boldsymbol{F}^T & \boldsymbol{0}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{c}(y)\\ \boldsymbol{f}(y)\end{bmatrix}. \tag{D.8}$$

Comparing simple and universal kriging. We can employ the Schur complement $\boldsymbol{S} := -\boldsymbol{F}^\top \boldsymbol{C}^{-1} \boldsymbol{F}$ to obtain a more explicit expression of $\boldsymbol{\lambda}$ in the case of

universal kriging:

$$\boldsymbol{\lambda}(x) = \left[\boldsymbol{C}^{-1} + \boldsymbol{C}^{-1}\boldsymbol{F}\boldsymbol{S}^{-1}\boldsymbol{F}^\top\boldsymbol{C}^{-1}\right]\boldsymbol{c}(x) + \boldsymbol{C}^{-1}\boldsymbol{F}\boldsymbol{S}^{-1}\boldsymbol{f}(x).$$

This, however, leads to

$$\hat{a}_{\mathrm{uk}}(x) = \boldsymbol{f}(x)^\top\hat{\boldsymbol{\beta}}_{\mathrm{LS}} + \boldsymbol{c}^\top(x)\boldsymbol{C}^{-1}\left(\boldsymbol{a} - \boldsymbol{F}\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right),$$

i.e., the universal kriging prediciton coincides with the simple kriging prediction if we employ for the latter the least squares estimate $\boldsymbol{\beta}_{\mathrm{LS}}$ of $\boldsymbol{\beta}$ in the mean function model D.5. The main difference between both kriging methods is the resulting error covariance. Employing the Schur complement once more we obtain, in particular,

$$\begin{aligned} c_{\mathrm{uk}}(x,y) &= c(x,y) - \boldsymbol{c}(x)^\top\boldsymbol{C}^{-1}\boldsymbol{c}(y) + \boldsymbol{h}^\top(x)\boldsymbol{S}^{-1}\boldsymbol{h}(y) \\ &= c_{\mathrm{sk}}(x,y) + \boldsymbol{h}^\top(x)\boldsymbol{S}^{-1}\boldsymbol{h}(y), \end{aligned}$$

where we set $\boldsymbol{h}(x) := \boldsymbol{f}(x) - \boldsymbol{F}^\top\boldsymbol{C}^{-1}\boldsymbol{c}(x)$. Due to $\boldsymbol{S}^{-1}$ being positiv semi-definite, the pointwise simple kriging variance is always smaller or equal to the universal simple kriging variance. This is not surprising, since the simple kriging assumes a known mean function, i.e., it ignores the uncertainty about the mean of a. Universal kriging takes this uncertainty into account. In fact, the additional term in the universal kriging error covariance allows for the following interpretation: considering the mean squared error between the simple kriging prediction $\hat{a}_{\mathrm{sk}}(x)$ using the true $\boldsymbol{\beta}$ in the model (D.5) and the simple kriging prediction $\hat{a}_{\mathrm{sk,LS}}(x)$ using the least squares estimate $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ instead of $\boldsymbol{\beta}$, we obtain by the unbiasedness $\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right] = \boldsymbol{\beta}$ that

$$\begin{aligned} \mathbb{E}\left[\left(\hat{a}_{\mathrm{sk}}(x) - \hat{a}_{\mathrm{sk,LS}}(x)\right)^2\right] &= \mathbb{E}\Bigg[\Big(\Big[\boldsymbol{f}^\top(x)\boldsymbol{\beta} + \boldsymbol{c}^\top(x)\boldsymbol{C}^{-1}\left(\boldsymbol{a} - \boldsymbol{F}\boldsymbol{\beta}\right)\Big] - \\ &\qquad \Big[\boldsymbol{f}^\top(x)\hat{\boldsymbol{\beta}}_{\mathrm{LS}} + \boldsymbol{c}^\top(x)\boldsymbol{C}^{-1}\left(\boldsymbol{a} - \boldsymbol{F}\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right)\Big]\Big)^2\Bigg] \\ &= \mathbb{E}\left[\left(\left(\boldsymbol{f}^\top(x) - \boldsymbol{c}^\top(x)\boldsymbol{C}^{-1}\boldsymbol{F}\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right)\right)^2\right] \\ &= \mathbb{E}\left[\left(\boldsymbol{h}(x)^\top\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right)\right)^2\right] = \boldsymbol{h}(x)^\top\operatorname{Cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right)\boldsymbol{h}(x) \\ &= \boldsymbol{h}(x)^\top\underbrace{\boldsymbol{S}^{-1}\boldsymbol{F}^\top\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{C}^{-1}\boldsymbol{F}}_{=I}\boldsymbol{S}^{-1}\boldsymbol{h}(x) \\ &= \boldsymbol{h}^\top(x)\boldsymbol{S}^{-1}\boldsymbol{h}(x), \end{aligned}$$

where we have used

$$\operatorname{Cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right) = \operatorname{Cov}\left(\boldsymbol{S}\boldsymbol{F}^\top \boldsymbol{C}^{-1}\boldsymbol{a}\right) = \boldsymbol{S}\boldsymbol{F}^\top \boldsymbol{C}^{-1} \underbrace{\operatorname{Cov}(\boldsymbol{a})}_{=\boldsymbol{C}} (\boldsymbol{S}\boldsymbol{F}^\top \boldsymbol{C}^{-1})^\top.$$

Hence, the additional term in the universal kriging error can be viewed as the resulting mean squared error in simple kriging caused by the least squares estimate $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$. Thus, very loosely speaking we have

$$\operatorname{Var}(a - \hat{a}_{\mathrm{uk}}) = \operatorname{Var}(a - \hat{a}_{\mathrm{sk}}) + \operatorname{Var}(\hat{a}_{\mathrm{sk}}(x) - \hat{a}_{\mathrm{sk,LS}}(x)).$$

Remark D.4 (On interpolation)**.** By the above relation between the universal and simple kriging prediciton and Remark D.2 we obtain that also the universal kriging mean is an interpolant of a at the locations x_i, $i = 1, \ldots, n$. Moreover, by

$$\begin{aligned}\begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{F}^\top \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F}^\top & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{F}^\top \end{bmatrix} &= \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{F}^\top \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F}^\top & \boldsymbol{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F}^\top & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} I & \boldsymbol{0} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{F}^\top \end{bmatrix}^\top \begin{bmatrix} I & \boldsymbol{0} \end{bmatrix} = \boldsymbol{C}\end{aligned}$$

we also get that $c_{\mathrm{uk}}(x_i, x_i) = c(x_i, x_i) - c(x_i, x_i) = 0$ for $i = 1, \ldots, n$.

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, Washington, 10th edition, 1972.

[2] R. J. Adler. *The Geometry of Random Fields*. John Wiley & Sons, New York, 1981.

[3] A. Apte, M. Hairer, A. M. Stuart, and J. Voss. Sampling the posterior: An approach to non-Gaussian data assimilation. *Physica D: Nonlinear Phenomena*, 230:50–64, 2007.

[4] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptical partial differential eqautions with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[5] I. Babuška, R. Tempone, and G. Zouraris. Galerkin finite element approximations of stochastic elliptical partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[6] M. Bachmayr, A. Cohen, R. DeVore, and G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. part II: lognormal coefficients. *ESAIM Math. Model. Numer. Anal.*, 2016.

[7] M. Bachmayr, A. Cohen, and G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. part I: affine coefficients. *ESAIM Math. Model. Numer. Anal.*, 2016.

[8] H. Bandemer. *Mathematics of Uncertainty - Ideas, Methods, and Applications*. Springer-Verlag, Berlin, 2006.

[9] J. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 2nd edition, 1985.

[10] J. Bernardo. Bayesian statistics. In R. Viertl, editor, *Probability and Statistics*, Encyclopedia of Life Support Systems (EOLSS). UNESCO, Oxford, UK, 2003.

[11] P. Berti, L. Pratelli, and P. Rigo. Almost sure weak convergence of random probability measures. *Stochastics*, 78(2):91–97, 2006.

[12] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, A. M. Stuart, and J. Voss. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351, 2017.

[13] A. Beskos, N-Pillai, G. Roberts, J.-M. Sanz-Serna, and A. Stuart. Optimal tuning the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[14] A. Beskos, G. Roberts, and A. Stuart. Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.*, 19(3):863–898, 2009.

[15] A. Beskos, G. Roberts, A. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stoch. Dynam.*, 8(3):319–350, 2008.

[16] A. Beskos, G. Roberts, A. Thiery, and N. Pillai. Asymptotic analysis of the random-walk Metropolis algorithm on ridged densities. arXiv:1510.02577v1, 2015.

[17] E. D. Blanchard, A. Sandu, and C. Sandu. A polynomial chaos-based Kalman filter approach for parameter estimation of mechanical systems. *Journal of Dynamic Systems, Measurement, and Control*, 132(6):061404, 2010.

[18] V. Bogachev. *Gaussian Measures*. American Mathematical Society, Providence, 1998.

[19] N. Bou-Rabee and M. Hairer. Nonasymptotic mixing of the MALA algorithm. *IMA J. Numer. Anal.*, 33(1):80–110, 2013.

[20] J. P. Boyd. The rate of convergence of Hermite function series. *Mathematics of Computation*, 35(152):1309–1316, 1980.

[21] J. P. Boyd. Asymptotic coefficients of Hermite function series. *Journal of Computational Physics*, 54(3):382–410, 1984.

[22] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer Series in Computational Mathematics. Springer-Verlag, New-York, 1991.

[23] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[24] M. Burger and F. Lucka. Maximum-a-posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators. *Inverse Problems*, 30:114004, 2014.

[25] G. Burgers, P. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

[26] D. E. Catlin. *Estimation, Control, and the Discrete Kalman Filter*. Springer-Verlag, New York, 1989.

[27] J. Charrier. Strong and weak error estimates for elliptic partial differential equations with random coefficients. *SIAM Journal on Numerical Analysis*, 50(1):216–246, 2012.

[28] J. Charrier and A. Debussche. Weak truncation error estimates for elliptic PDEs with lognormal coefficients. *Stochastic Partial Differential Equations: Analysis and Computations*, 1(1):63–93, 2013.

[29] J.-P. Chilès and P. Delfiner. *Geostatistics - Modeling Spatial Uncertainty*. John Wiley and Sons, Inc., New York, second edition, 2012.

[30] A. Chkifa, A. Cohen, and Ch. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Foundations of Computational Mathematics*, 14(4):601–633, 2014.

[31] A. Chkifa, A. Cohen, and Ch. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *J. Math. Pures Appl.*, 103(2):400–428, 2015.

[32] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput. Visual Sci.*, 14(3):3–15, 2011.

[33] A. Cohen and R. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, 24:1–159, 2015.

[34] A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Analysis and Applications*, 9:11–47, 2011.

[35] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28(3):283 – 464, 2013.

[36] M. K. Cowles and P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.

[37] T. Cui, K. Law, and Y. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.

[38] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge, 1992.

[39] G. Da Prato and J. Zabczyk. *Second Order Partial Differential Equations in Hilbert Spaces*. Cambridge University Press, Cambridge, 2004.

[40] G. Da Prato and J. Zabczyk. *An Introduction to Infinite-Dimensional Analysis*. Springer-Verlag, Berlin Heidelberg, 2006.

[41] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017:1–27, 2013.

[42] M. Dashti and A. M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.*, 49(6):2524–2542, 2011.

[43] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 311–428. Springer International Publishing, Cham, 2017.

[44] R. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[45] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comput.*, 18(4):1088–1107, 1997.

[46] T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA J. Uncertainty Quantification*, 3(1):1075–1108, 2015.

[47] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.

[48] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.

[49] R. Douglas. On majorization, factorization, and range inclusion of operators on Hilbert space. *Proc. Amer. math. Soc.*, 17:413–415, 1966.

[50] N. Dunford and J. Schwartz. *Linear Operators, Part I: General Theory*. Wiley-Interscience, New York, 1958.

[51] T. A. El Moselhy and Y. M. Marzouk. Bayesian inference with optimal maps. *Journal of Computational Physics*, 231(23):7815 – 7850, 2012.

[52] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, 2000.

[53] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. Springer-Verlag, New York, LLC, 2004.

[54] O. Ernst, A. Mugler, H.-J Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(2):317–339, 2012.

[55] O. Ernst and B. Sprungk. Stochastic collocation for elliptic pdes with random data - the lognormal case. In J. Garcke and D. Pflüger, editors, *Sparse Grids and Applications - Munich 2012*, volume 97 of *Lecture Notes in Computational Science and Engineering*, pages 29–53. Springer International Publishing, Cham, 2014.

[56] O. Ernst, B. Sprungk, and H.-J. Starkloff. Bayesian inverse problems and Kalman filters. In S. Dahlke et al., editor, *Extraction of Quantifiable Information from Complex Systems*, volume 102 of *Lecture Notes in Computational Science and Engineering*, pages 133–159. Springer International Publishing, Cham, 2014.

[57] O. Ernst, B. Sprungk, and H.-J Starkloff. Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. *SIAM/ASA J. Uncertainty Quantification*, 3(1):823–851, 2015.

[58] O. Ernst, B. Sprungk, and L. Tamellini. Convergence of sparse collocation for functions of countably many Gaussian random variables - with application to lognormal elliptic diffusion problems. arXiv:1611.07239, 2016.

[59] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrohpic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162, 1994.

[60] G. Evensen. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

[61] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, New York, 2nd edition, 2009.

[62] G. Evensen. The ensemble Kalman filter for combined state and parameter estimation. *Control Systems Magazine*, 29(3):83–104, 2009.

[63] G. Evensen and P. J. van Leeuwen. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128:1852–1867, 2000.

[64] R. Allan Freeze. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resources Research*, 11(5):725–741, 1975.

[65] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity. *SIAM J. Numer. Anal.*, 47(5):3624–3651, 2009.

[66] W. Gautschi. *Orthogonal Polynomials – Computation and Approximation*. Oxford University Press, Oxford, 2004.

[67] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85:398–409, 1990.

[68] C. Geyer. Practical Markov chain Monte Carlo. *Stat. Sci.*, 7(4):473–483, 1992.

[69] C. J. Geyer. Introduction to Markov Chain Monte Carlo. In S. Brooks, A. Gelman, G. J. Jones, and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, Handbooks of Modern Statistical Methods, pages 3–48. CRC Press, Boca Raton, 2011.

[70] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York, 1991.

[71] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2001.

[72] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, Berlin Heidelberg, 2001.

[73] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B*, 73(2):123–214, 2010.

[74] C. J. Gittelson. Stochastic Galerkin discretization of the log-normal isotropic diffusion problem. *Math. Models Methods Appl. Sci.*, 20(2):237–263, 2010.

[75] F. Le Gland, V. Monbet, and V.-D. Tran. Large sample asymptotics for the ensemble Kalman filter. In D. Crisan and B. Rozovskii, editors, *Oxford Handbook of Nonlinear Filtering*, chapter 22, pages 598–631. Oxford University Press, Oxford, 2011.

[76] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, Ch. Schwab, and I. H. Sloan. Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficient. *Numer. Math.*, 4:41–75, 2016.

[77] I. G. Graham, R. Scheichl, and E. Ullmann. Mixed finite element analysis of lognormal diffusion and multilevel Monte Carlo methods. *Stoch PDE: Anal Comp*, 131:329–368, 2015.

[78] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[79] W Hackbusch. *Hierarchical Matrices: Algorithms and Analysis*. Springer-Verlag, Berlin Heidelberg, 2015.

[80] M. Hairer. An introduction to stochastic PDEs. Lecture notes, 2009.

[81] M. Hairer, A. Stuart, and S. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.

[82] M. Hairer, A. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling part II: The nonlinear case. *Ann. Appl. Probab.*, 17(5/6):1657–1706, 2007.

[83] P. R. Halmos. *Introduction to Hilbert Space*. Chelsea Publishing, New York, 1951.

[84] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[85] E. Hille. Contributions to the theory of Hermitian series. *Duke Mathematical Journal*, 5:875–936, 1939.

[86] E. Hille. Contributions to the theory of Hermitian series. II. The representation problem. *Transactions of the American Mathematical Society*, 47:80–94, 1940.

[87] M. Hladnik and M. Omladič. Spectrum of the product of operators. *Proc. Amer. math. Soc.*, 102(2):300–302, 1988.

[88] Viet Ha Hoang and Christoph Schwab. N-term Wiener chaos approximation rates for elliptic PDEs with lognormal Gaussian random inputs. *Mathematical Models and Methods in Applied Sciences*, 24(4):797–826, 2014.

[89] P. D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer-Verlag, New York, 2009.

[90] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990.

[91] Z. Hu, Z. Yao, and J. Li. On an adaptive preconditioned Crank-Nicolson algorithm for infinite dimensional Bayesian inferences. *Journal of Computational Physics*, 332:492–503, 2017.

[92] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001:1–20, 2013.

[93] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Evaluation of Gaussian approximations for data assimilation in reservoir models. *Computational Geosciences*, 17(5):851–885, 2013.

[94] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.

[95] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer-Verlag, New York, 2005.

[96] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, 2nd edition, 2002.

[97] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the AMSE – Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[98] C. Kipnis and S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.

[99] A. Klenke. *Probability Theory - A Comprehensive Course*. Springer-Verlag, London, 2008.

[100] Y. Kovchegov and N. Michalowski. A class of Markov chains with no spectral gap. *Proc. Amer. Math. Soc.*, 141(12):4317–4326, 2013.

[101] D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, 1951.

[102] S. Lan, T. Bui-Thanh, M. Christie, and M. Girolami. Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems. *Journal of Computational Physics*, 308:81–101, 2016.

[103] A. M. LaVenue, T. L. Cauffman, and J. F. Pickens. Ground-Water Flow Modeling of the Culebra Dolomite. Volume I: Model Calibration. Contractor Report SAND89-7068/1, Sandia National Laboratories, 1990.

[104] K. Law. Proposals which speed up function-space MCMC. *J. Comput. Appl. Math.*, 262:127–138, 2014.

[105] K. J. H. Law and A. M. Stuart. Evaluating data assimilation algorithms. *Monthly Weather Review*, 140:3757–3782, 2012.

[106] K. J. H. Law, H. Tembine, and R. Tempone. Deterministic mean-field ensemble Kalman filtering. *SIAM J. Sci. Comput.*, 38(3):A1251–A1279, 2016.

[107] G. Lawler and A. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988.

[108] O. P. Le Maitre and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer-Verlag, New York, 2010.

[109] A. Lee and K. Łatuszyński. Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika*, 101(3):655–671, 2014.

[110] J. M. Lewis, S. Lakshmivarahan, and S. Dhall. *Dynamic Data Assimilation – A Least Squares Appoach*. Cambridge University Press, Cambridge, 2006.

[111] W. A. Light and E. W. Cheney. *Approximation Theory in Tensor Product Spaces*, volume 1169 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin Heidelberg, 1985.

[112] D. V. Lindley. *Bayesian Statistics, A Review*. SIAM, Philadelphia, 1972.

[113] S. Livingstone. Geometric ergodicity of the Random Walk Metropolis with position-dependent proposal covariance. arXiv:1507.05780, 2015.

[114] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge University Press, New York, 2014.

[115] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley and Sons, Inc., New York, 1969.

[116] J. Mandel, L. Cobb, and J. D. Beezley. On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, 56(6):533–541, 2011.

[117] A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 65:385–397, 1984.

[118] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.

[119] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[120] S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, 2nd edition, 2009.

[121] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(5):1237–1263, 2013.

[122] I. Myrseth and H. Omre. The ensemble Kalman filter and related filters. In L. Biegler, editor, *Large-Scale Inverse Problems and Quantification of Uncertainty*, Wiley Series in Computational Statistics, pages 217–246. Wiley, Chichester, 2011.

[123] R. M. Neal. Regression and classification using Gaussian process priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, Wiley Series in Computational Statistics, pages 475–501. Oxford University Press, Oxford, 1999.

[124] F. Nobile, R. Tempone, and C. G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.

[125] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[126] R. A. Norton and C. Fox. Efficiency and computability of MCMC with Langevin, Hamiltonian, and other matrix-splitting proposals. arXiv:1501.03150v1, 2015.

[127] M. Oberguggenberger. The mathematics of uncertainty: models, methods and interpretations. In W. Fellin, H. Lessman, M. Oberguggenberger, and R. Vieider, editors, *Analyzing Uncertainty in Civil Engineering*. Springer-Verlag, Berlin, 2005.

[128] W. J. Padgett and R. L. Taylor. *Law of large numbers for normed linear spaces and certain Fréchet spaces*. Springer-Verlag, Berlin Heidelberg, 1973.

[129] O. Pajonk, B. V. Rosić, A. Litvinenko, and H. G. Matthies. A deterministic filter for non-Gaussian Bayesian estimation — applications to dynamical system estimation with noisy measurements. *Physica D: Nonlinear Phenomena*, 241(7):775–788, 2012.

[130] C. Pasarica and A. Gelman. Adaptively scaling the Metropolis algorithm using expected squared jumped distance. *Statistica Sinica*, 20:343–364, 2010.

[131] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60(3):607–612, 1973.

[132] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*. Cambridge University Press, Cambridge, UK, 2010.

[133] F. Pinski, G. Simpson, A. Stuart, and H. Weber. Algorithms for Kullback–Leibler approximation of probability measures in infinite dimensions. *SIAM J. Sci. Comput.*, 37(6):A2733–A2757, 2015.

[134] J. Potthoff. Sample properties of random fields III : Differentiability. *Communications on Stochastic Analysis*, 4(3):335–353, 2010.

[135] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations*. Springer International Publishing, Cham, 2016.

[136] B. S. Rajput and S. Cambanis. Gaussian processes and Gaussian measures. *Ann. Math. Statist.*, 43(6):1944–1952, 1972.

[137] M. M. Rao. *Conditional measures and applications*. Chapman and Hall/CRC, Boca Raton, 2010.

[138] M. Reed and B. Simon. *Functional Analysis*, volume 1 of *Methods of Modern Mathematical Physics*. Academic Press, San Diego, 2nd edition, 1980.

[139] G. Roberts and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2(2):13–25, 1997.

[140] G. Roberts and J. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B*, 60(1):255–268, 1998.

[141] G. Roberts and J. Rosenthal. Optimal scaling for various Metropolis–Hastings algorithms. *Stat. Sci.*, 16(4):351–367, 2001.

[142] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[143] B. V. Rosić, A. Kučerová, J. Sýkora, O. Pajonk, A. Litvinenko, and H. G. Matthies. Parameter identification in a probabilistic setting. *Engineering Structures*, 60:179–196, 2013.

[144] B. V. Rosić, A. Litvinenko, O. Pajonk, and H. G. Matthies. Sampling-free linear Bayesian update of polynomial chaos representations. *Journal of Computational Physics*, 231(17):5761–5787, 2012.

[145] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math. (Rozprawy Mat.)*, 485:1–93, 2012.

[146] D. Rudolf and B. Sprungk. On a generalization of the preconditioned Crank–Nicolson Metropolis algorithm. *Found. Comput. Math.*, 2016. doi:10.1007/s10208-016-9340-x.

[147] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. *Electron. Commun. Probab.*, 18:1–8, 2013.

[148] G. Saad and R. Ghanem. Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resources Research*, 45(4), 2009. doi:10.1029/2008WR007148.

[149] G. Saad, R. Ghanem, and S. Masri. Robust system identification of strongly non-linear dynamics using a polynomial chaos-based sequential data assimilation techique. In *Collection of Technical Papers–48th*

AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, volume 6, New York, 2007. Springer-Verlag.

[150] A. K. Saibaba, A. Alexanderian, and I. C. F. Ipsen. Randomized matrix-free trace and log-determinant estimators. *Numer. Math.*, 137:353–395, 2017.

[151] R. Scheichl, A. M. Stuart, and A. L. Teckentrup. Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems. *SIAM/ASA J. Uncertainty Quantification*, 5:493–518, 2017.

[152] M. Scheurer, R. Schaback, and M. Schlather. Interpolation of spatial data – a stochastic or a deterministic problem? *European Journal of Applied Mathematics*, 24(4):601–629, 2013.

[153] René L. Schilling. *Measures, Integrals and Martingales*. Cambridge University Press, Cambridge, 2005.

[154] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6):065011:1–28, 2013.

[155] C. Schillings and A. M. Stuart. Analysis of the ensemble Kalman filter for inverse problems. *SIAM J. Numer. Anal.*, 55(3):1264–1290, 2017.

[156] C. Schwab and C. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.

[157] C. Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4):045003:1–32, 2012.

[158] C. Schwab and R. A. Todor. Karhunen–Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.

[159] D. Simon. *Optimal state estimation: Kalman, H_∞, and nonlinear approaches*. Wiley, Hoboken, 2006.

[160] R. C. Smith. *Uncertainty Quantification: Theory, Implementation and Applications*. SIAM, Philadelhpia, 2014.

[161] A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk. Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM J. Sci. Comput.*, 37(6):A2451–A2487, 2015.

[162] J. Speyer and W. Chung. *Stochastic processes, estimation, and control*. SIAM, Philadelphia, 2008.

[163] I. Sraj, O. P. Le Maitre, O. M. Knio, and I. Hoteit. Coordinate transformation and polynomial chaos for the Bayesian inference of a Gaussian process with parametrized prior covariance function. *Comput. Methods Appl. Mech. Engrg.*, 298:205–228, 2016.

[164] M. L. Stein. *Interpolation of Spatial Data - Some Theory for Kriging*. Springer-Verlag, New York, 1999.

[165] Nicola Stone. *Gaussian process emulators for uncertainty analysis in groundwater flow*. PhD thesis, University of Nottingham, 2011.

[166] A. S. Stordal, H. A. Karlsen, G. Nærvdal, H. J. Skaug, and B. Vallès. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, 15(2):293–305, 2011.

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

[168] G. Szegő. *Orthogonal Polynomials*. American Mathematical Society, New York, fourth edition, 1975.

[169] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numer. Math.*, 125(3):569–600, 2013.

[170] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Stat.*, 22(4):1701–1762, 1994.

[171] L. Tierney. A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.

[172] E. Ullmann. *Solution Strategies for Stochastic Finite Element Discretizations*. PhD thesis, TU Bergakademie Freiberg, 2008.

[173] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.

[174] H. Widom. Asymptotic behavior of the eigenvalues of certain integral equations. *Trans. Amer. Math. Soc.*, 109:278–295, 1963.

[175] N. Wiener. The homogeneous chaos. *Amer. J. Math.*, 160:897–936, 1938.

[176] David Williams. *Weighing the Odds*. Cambridge University Press, Cambridge, second edition, 2004.

[177] D. Xiu and J. S. Hesthaven. High-order collocation methods differential equations with random inputs. *SIAM Journal on Scientific Computing*, 37(3):1118–1139, 2005.

[178] D. Xiu and G. E. Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Comput. Methods Appl. Mech. Engrg.*, 191(43):4927–4948, 2002.

[179] D. Xui. *Numerical Methods for Stochastic Computations*. Princeton University Press, Philadelphia, 2010.

[180] K. Yosida. *Functional Analysis*. Springer-Verlag, Berlin Heidelberg, sixth edition, 1995.

[181] H. Yue and K. S. Chan. Asymptotic efficiency of the sample mean in Markov chain Monte Carlo schemes. *J. R. Stat. Soc. Ser. B*, 58(3):525–539, 1996.