Nonparametric Bayesian Inverse Problems with Laplacian Noise

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

Doctor rerum naturalium

2019

von

Remo Kretschmann

bei der Fakultät für Mathematik der Universität Duisburg-Essen vorgelegt

Gutachter der Dissertation:

Prof. Dr. Christian Clason und *Prof. Tapio Helin, PhD*

Datum der Disputation:

20. August 2019

DOI: 10.17185/duepublico/70452
URN: urn:nbn:de:hbz:464-20190910-082111-3

Abstract

The focus of this work are Bayesian inverse problems in an infinite-dimensional setting with Gaussian prior and data corrupted by additive Laplacian noise. In particular, the connection between Tikhonov–Phillips regularisation with an ℓ^1-discrepancy term and Bayesian MAP estimation based upon a Laplacian noise model is investigated for linear problems as well as the consistency of MAP and CM estimator.

For Laplacian infinite product measures on a separable Hilbert space, a result similar to the Cameron–Martin theorem for Gaussian measures is shown, stating the space of admissible shifts and the relative density of shifted Laplacian measures.

Under certain conditions on the log-likelihood, MAP estimates in separable Hilbert spaces are characterised as minimisers of the Onsager–Machlup functional of the posterior distribution, which in this case has the form of a Tikhonov–Phillips functional with a discrepancy term given by the log-likehood and a squared norm penalty term.

The behaviour of MAP and CM estimator is studied for a severely ill-posed linear problem; a generalised form of the inverse heat equation, under the presence of additive Laplacian noise. The posterior distribution is derived via Bayesian inference and both MAP and CM estimator are computed explicitly. The MAP estimator is shown to be asymptotically unbiased in a frequentist setting. An estimate for the convergence rate of the bias is stated under an analytic source condition. Moreover, an estimate for the convergence rate of the mean squared error of the MAP estimator is proved under an analytic source condition and in conjunction with an a priori parameter choice. This rate is then compared to the minimax rate in the fully Gaussian case.

The behaviour and consistency of MAP and CM estimator is studied numerically for the classical inverse heat equation in one dimension with additive Laplacian noise. The empirical MSE of both estimators is observed to converge to zero in the small noise limit with the estimated rate if an analytic source condition is satisfied, whereas neither MAP nor CM estimator converge towards the true solution in mean square if only a Sobolev-type source condition is satisfied. Moreover, empirical confidence regions around both estimators are computed. Finally, a direct sampler for the posterior distribution is developed and used to compute credible regions around both estimators.

Acknowledgements

First and foremost, I want to thank Christian Clason for supervising and supporting me throughout the emergence of this thesis with extensive knowledge, helpful advice, fruitful open discussions and occasional nudges in the right direction. I want to thank Tapio Helin for agreeing to survey this thesis. I would also like to thank Tapio Helin and Markus Reiß for their feedback on my work, inspiring discussions and making me see my work from a different angle. Moreover, I want to thank Masoumeh Dashti for the illuminating correspondence about certain subtleties regarding Gaussian measures. Finally, I want to thank my family and friends for supporting me.

Contents

Introduction

The objective of an inverse problem, in general, is reconstructing an unknown quantity of interest from indirect measurements which are connected to the sought-after quantity by a mathematical model. This might be necessary, for example, because the unknown quantity itself cannot be measured directly. Both unknown and measured quantities are typically functions in a certain function space and as such infinite-dimensional. The model can usually be expressed as an operator equation

$$y = F(u) \tag{1}$$

with a linear or nonlinear mapping F between two Banach spaces describing the relation between the unknown quantity u and the measured quantity y. Now, predicting the measured data y for a given value of u is called the *direct problem*. Contrary to this is the *inverse problem*, which consists in finding the unknown u for given data y.

The term inverse problem usually refers to ill-posed inverse problems. According to Hadamard, a problem is called *well-posed* if the following three conditions hold.

(H1) A solution exists.

(H2) The solution is unique.

(H3) The solution depends continuously on the data.

If, on the other hand, one of these conditions is violated, the problem is called *ill-posed*.

An inverse problem may be ill-posed for several reasons: The solution might not be unique because the setup of the experiment limits the amount of information that can be obtained about the unknown quantity. Even if a unique solution exists for attainable data y, i.e., data in the range of the forward mapping F, there might not exist a solution for arbitrary noisy data. Or, most importantly, the forward mapping F might not be continuously invertible. A direct problem can be well-posed, even if the inverse problem is ill-posed. This is, for instance, the case for operator equations involving a compact linear forward operator F with infinite-dimensional range.

Model (1) is still highly idealised as it does not take measurement errors into account. A more realistic model

$$y = F(u) + \eta$$

is obtained by incorporating additive *noise* η present in the measurements. Classically, the noise is assumed to be a *deterministic* quantity whose norm is strictly bounded by the *noise level* $\delta > 0$, i.e.,

$$\|y - F(u)\|_Y = \|\eta\|_Y \le \delta,$$

but it can also be modelled as a *stochastic* quantity, that is as a random variable with a known probability distribution.

Often, the distribution of the noise can reasonably be assumed to be Gaussian. There are, however, cases where it is natural to assume a non-Gaussian noise distribution, for instance, inverse problems with impulsive noise such as salt-and-pepper noise or random valued impulsive noise, which arise in many applications in image and signal processing, see [Bovik 2005], e.g., image acquisition with faulty pixels in a sensor or faulty memory locations. Such impulsive noise functions (or vectors, respectively) take very large values on a small part of their domain, while being small or identical to zero elsewhere.

While the assignment of the terms direct and inverse problem can be considered somewhat arbitrary, it is the ill-posedness of a problem that makes solving it ad hoc in a stable way impossible. This is overcome by, instead of trying to compute the true solution, computing a *regularised solution* that is close to the true solution and depends continuously on the data. The nonexistent or discontinuous inverse F^{-1} of the forward mapping is approximated by a family $\{R_\alpha\}_{\alpha>0}$ of continuous mappings, called a *regularisation*, that converges pointwise towards F^{-1}. This means that for each value of α the regularised solution $u_\alpha := R_\alpha(y)$ depends continuously on the data y and for each value of the unknown u the regularised solution $u_\alpha = R_\alpha(F(u))$ for noise free data converges to u as α tends to zero. One of the advantages of modelling not only the problem but also the regularisation method on infinite-dimensional spaces is that desirable properties such as convergence and stability can be established inpependent of the chosen discretisation.

The choice of the *regularisation parameter α* plays a central role in obtaining a good reconstruction. On the one hand, it should be chosen small enough, so that F^{-1} is approximated well enough. On the other hand, choosing it too small results in an increased error due to the discontinuity of F^{-1}. If α is chosen only depending on the noise level we speak of an *a priori parameter choice rule*, whereas we speak of an *a posteriori parameter choice rule* if both the noise level and the measured data is taken into account.

A widely used class of regularisation methods is variational regularisation, where the regularised solution u_α is given as the solution of an optimisation problem

$$\min_u \{\Phi(u, y) + \alpha R(u)\}$$

with a *discrepancy term* (or *data fitting term*) $\Phi(u, y)$ and a *penalty term* (or *regularisation term*) $\alpha R(u)$. The most prominent example of such a regularisation method is Tikhonov–Phillips regularisation, where the objective functional has the form

$$u \mapsto \|y - F(u)\|_Y^q + \alpha \|u\|_X^p.$$

For inverse problems with impulsive noise, Tikhonov–Phillips regularisation with an L^1-data fitting term has been observed to provide better estimates than L^2-data fitting [Kärkkäinen, Kunisch, and Majava 2005; Clason, Jin, and Kunisch 2010], due to its higher robustness towards outliers. This remarkable difference in performance has been studied further in [Hohage and Werner 2014; König, Werner, and Hohage 2016] within a deterministic framework in case of finitely and infinitely smoothing forward operators, resulting in improved convergence rate estimates. Here, an impulsive noise function may be arbitrarily large on a small part of its domain, while being small in L^1-norm on the rest of its domain.

The objective of statistical inference is to estimate certain quantities of interest which are unknown or unobservable, called the *parameter*, given measurements of observable quantities (the *data*) that are linked to the unknown quantities by a *model*. For any fixed value of the parameter the data is assumed to be a random variable with a probability distribution specified by the model. Classical statistical inference takes a frequentist point of view, insofar as the parameter is assumed to be a deterministic quantity. If the parameter is infinite-dimensional, the setting is commonly called *nonparametric*. Inverse problems with stochastic noise fall in this framework. Here, the model is derived from the operator equation describing the relation between unknown, noise and data and the distribution of the noise.

Consider a basic denoising problem with direct measurements $y = u + \eta$ of an unknown image u corrupted by noise η. First, assume that observation and parameter are scalar, e.g., the value of a single pixel of the image. Given multiple measurements $y_1, \ldots, y_n$ of the observable quantity, an estimate $\hat{u}$ for the parameter can be defined to minimise a certain measure of scatter for the available samples. Such estimators are called *maximum likelihood estimators*. Classical choices for the objective functional are the *root mean square deviation*

$$s_n(u) = \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - u)^2\right)^{1/2}$$

or the *mean absolute deviation*

$$d_n(u) = \frac{1}{n}\sum_{i=1}^{n}|y_i - u|.$$

For the former choice, $\hat{u}$ is the mean of the samples, for the latter one, $\hat{u}$ is their median. If the errors present in the observations are identically distributed and Gaussian, an estimate minimising s_n has a higher asymptotic efficiency as the number of samples n tends to infinity, whereas one minimising d_n is asymptotically more efficient if even a small percentage of the samples is known to be outliers that have, e.g., a higher variance, see Example 1.1 in [Huber 2009].

The whole original image u can be modelled as a function on a bounded set $D \subset \mathbb{R}^2$, independent of the resolution or the layout of a specific sensor. Assume that a single measurement is taken (i.e., the sample size n is equal to 1) and that the value measured in every point $x \in D$ is independently afflicted by impulsive noise. The previous considerations suggest estimating the original image u by minimising

$$\|y - u\|_{L^1} = \int_D |y(x) - u(x)|\mathrm{d}x.$$

The approach of regression analysis is to do this under the assumption that u lies in a certain set of functions. A specific sensor can be modelled by assuming that the image is observed in m points $x_1, \ldots, x_m \in D$. In this case, estimating the original image u by minimising the mean absolute deviation

$$\frac{1}{m}\sum_{k=1}^{m}|y(x_k) - u(x_k)|$$

can be motivated in the same way. This approach is called method of *least absolute deviations*. In order to obtain reasonable estimates, the function u is typically assumed to be of a particular form and determined by a small number of parameters. Introducing positive weights w_k and minimising

$$\sum_{k=1}^{\infty} w_k |y(x_k) - u(x_k)|$$

can now be considered as the limit case, in which the number of points m tends to infinity.

In the context of statistical inference, the regularised solution can also be considered as an estimator for the sought-after quantity and its statistical properties can be studied, e.g., if it converges towards the true parameter in probability (*consistency*) and at which rate.

In the Bayesian approach to statistical inference, in contrast, the parameters are also treated as random variables. A probability distribution, called the *prior distribution*, is assigned to all relevant unknown quantities, formalising any assumptions about their distribution without (or before) taking the data into account. Here, the model links parameters and data by specifying the *conditional distribution* of the observed, given the unknown quantities. In our context, the density of the conditional distribution of the data, given the parameter, with respect to a reference measure is called *likelihood*. The objective of Bayesian inference is to then find the *posterior distribution*: the conditional distribution of the parameter given the data. Infering the posterior distribution from model and prior distribution involves changing the order of conditioning, which is achieved by some form of Bayes' formula.

For linear inverse problems with a Gaussian prior distribution and additive Gaussian noise the posterior distribution is again a Gaussian measure. The posterior contraction rate, which describes the concentration rate of the posterior distribution around a point in the small noise limit, has been studied in this setting under the frequentist assumption that a fixed data-generating value of the unknown exists [Agapiou and Mathé 2018].

On finite-dimensional spaces, *maximum a posteriori (MAP) estimates* are defined as modes of the posterior distribution. If a probability measure has a continuous density with respect to the Lebesgue measure, then its *modes* are defined as maximisers of this density. This way, variational regularisation with a continuous objective functional can be interpreted as maximum a posteriori estimation, where the spread of the prior plays the role of the regularisation parameter. This yields an analytic justification for the choice of the objective functional by Bayesian modelling and statistical inference based upon few and explicit assumptions about prior and noise distribution. The above definition of modes is, however, limited to measures on finite-dimensional spaces, since there exists no Lebesgue measure on infinite-dimensional spaces.

Another frequently used Bayesian estimator is the *conditional mean (CM) estimator*; it is defined as the mean of the posterior distribution. For linear inverse problems with a Gaussian prior distribution and additive Gaussian noise the posterior distribution is again a Gaussian measure, so that its mode and its mean, and hence also MAP and CM estimator, coincide and can be stated explicitly. In contrast, MAP and CM estimator can be distinct in case of non-Gaussian noise.

The consistency and convergence rate of the MAP estimator based upon a Gaussian prior has been investigated both in a Bayesian and in a frequentist framework for linear inverse problems with additive white Gaussian noise [Kekkonen, Lassas, and Siltanen 2016; Burger, Helin, and Kekkonen 2018].

On infinite-dimensional spaces, the posterior distribution does not have a canonical density due to the lack of a Lebesgue measure. Here, both maximum a posteriori estimates and modes are commonly defined via the limit of small ball probabilities, see [Dashti, Law, et al. 2013]. This approach can be generalised using bounded, convex, and open sets instead of balls, see [Lie and Sullivan 2018]. It is, in general, an open question if nonparametric MAP estimates are given as solutions of a canonical optimisation problem. The Onsager–Machlup functional of the posterior distribution, which is also defined via the limit of small ball probabilites, can be considered as its generalised negative logarithmic density. As such, it is a natural candidate for use as an objective functional. Under certain conditions on the likelihood, MAP estimates for nonlinear inverse problems with a Gaussian prior have been shown to coincide with minimisers of the Onsager–Machlup functional of the posterior distribution [Dashti, Law, et al. 2013], which in this case has the form of a Tikhonov–Phillips functional with a discrepancy term given by the negative log-likelihood and a squared norm penalty term. These conditions are, for example, satisfied for linear problems with additive Gaussian noise and finite-dimensional data. A similar variational characterisation of MAP estimates has been shown to hold true for nonlinear inverse problems with a B_1^s-Besov prior [Agapiou, Burger, et al. 2018], involving a Besov norm penalty term. This is of particular interest, because the B_1^s-Besov norm can be considered as a weighted ℓ^1-norm. If, on the other hand, the posterior distribution is discontinuous in a certain sense, e.g., if it is not quasi-invariant along any direction, the Onsager–Machlup functional is not defined and minimisers of a canonical Tikhonov–Phillips functional are, in general, no MAP estimates but only *generalised MAP estimates* [Clason, Helin, et al. 2019].

In the context of these results, the question arises if a similar variational characterisaton of nonparametric MAP estimates is possible in case of non-Gaussian noise. In particular, we are interested in the question if a statistical interpretation of Tikhonov–Phillips regularisation with an ℓ^1-discrepancy term as a Bayesian MAP estimator is possible and on which exact noise model it is based. A promising candidate for a noise model that might lead to such an estimator is Laplacian infinite product noise. In order to connect its MAP estimates to an optimisation problem we show that a variational characterisation of MAP estimates based upon a Gaussian prior is possible for a general class of models. Moreover, we investigate the behaviour and statistical properties of such a MAP estimator for a specific problem both analytically and numerically, and examine if the CM estimator behaves fundamentally different. To this end, we consider a severly ill-posed linear inverse problem with additive Laplacian noise. Here, we obtain an objective functional with a discrepancy term that coincides with the weighted ℓ^1-norm up to a constant on finite-dimensional subspaces.

This thesis is structured as follows: In Chapter 1, we discuss fundamental notions necessary to conduct nonparametric Bayesian inference, such as conditional probabilities and regular conditional distributions, introduce Bayes' formula and portray how inverse problems with stochastic noise fit into this framework. In Chapter 2, we review the definition of Gaussian measures on separable Hilbert spaces and their representation as an infinite product measure. Furthermore, we present the Cameron–Martin theorem, which gives a criterion for the equivalence of shifted Gaussian measures and states their relative density.

In Chapter 3, we construct Laplacian infinite product measures on separable Hilbert spaces based upon the Laplace distribution on $\mathbb{R}$. We then determine the directions in which translations of a Laplacian measure lead to an equivalent measure and state the density of shifted Laplacian measures with respect to centred ones.

In Chapter 4, we characterise MAP estimates for nonlinear Bayesian inverse problems with Gaussian prior distribution under certain assumptions on the log-likelihood as minimisers of the Onsager–Machlup functional of the posterior distribution, which in this case has the form of a Tikhonov–Phillips functional with a squared norm penalty term and a discrepancy term given by the negative log-likelihood.

In Chapter 5, we consider a severely ill-posed linear problem and explain how it can be understood as a generalisation of the inverse heat equation. We infer the posterior distribution in case of a Gaussian prior and additive Laplacian noise. Then, we express the MAP estimator in analytic form using its variational characterisation and study its behaviour in a frequentist setting: We show that it is asymptotically unbiased in conjunction with an a priori parameter choice rule and estimate the convergence rate of the bias under a source condition. Moreover, we estimate the convergence rate of its mean squared error under a source condition using an a priori parameter choice, and compare this rate with the minimax rate for Gaussian noise. Here, we also express the CM estimator in analytic form.

In Chapter 6, we study the behaviour of MAP and CM estimator numerically for the inverse heat equation in one dimension under the presence of Laplacian measurement noise. We assess the spread of Laplacian noise depending on its smoothness by considering empirical credible regions. We examine how the degree of smoothness of the noise and the prior affects the ability of both estimators to reconstruct the unknown. In a frequentist setting, we study the effect of the regularisation parameter on the mean squared error of MAP and CM estimator as well as the consistency and convergence rate of both estimators in terms of the mean squared error. Moreover, we consider empirical confidence regions for the true solution around both estimators for different values of the regularisation parameter. Eventually, we evaluate the spread of the posterior distribution depending on the prior variance in a Bayesian setting by computing empirical credible regions using a direct posterior sampler.

1 Nonparametric Bayesian Inference

In this chapter we review the definition of conditional probabilities and provide a proof of Bayes' formula in the required generality. For a broader introduction into nonparametric Bayesian inference we refer, e.g., to [Ghosal and van der Vaart 2017, Chapter 1].

1.1 Fundamental Notions

First we recapitulate some basic definitions. Let X be a set and let $\mathcal{P}(X)$ denote its power set. A set $\mathcal{X} \subseteq \mathcal{P}(X)$ is called *σ-algebra* (or *σ-field*) if

(i) $\varnothing \in \mathcal{X}$,

(ii) $X \setminus A \in \mathcal{X}$ for every $A \in \mathcal{X}$, and

(iii) $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{X}$ for every countable subset $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{X}$.

Let $\mathcal{E} \subseteq \mathcal{P}(X)$. Then the smallest σ-algebra $\sigma(\mathcal{E})$ with $\mathcal{E} \subseteq \sigma(\mathcal{E})$ is called the *σ-algebra generated by $\mathcal{E}$*. If X is a topological space, then the σ-algebra $\mathcal{B}(X)$ generated by the open sets in X is called *Borel σ-algebra* on X. A pair $(X, \mathcal{X})$ of a set X and a σ-algebra $\mathcal{X}$ is called *measureable space*. A function $\mu\colon \mathcal{X} \to [0, \infty]$ is called *measure* on $(X, \mathcal{X})$ if

(i) $\mu(\varnothing) = 0$, and

(ii) for every countable subset $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{X}$ of pairwise disjoint sets,

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A measure μ on $(X, \mathcal{X})$ that satisfies $\mu(X) = 1$ is called *probability measure* and the triple $(X, \mathcal{X}, \mu)$ *probability space*.

Let $\mathcal{X}$ be a σ-algebra on a set X and let μ and ν be two measures on $(X, \mathcal{X})$. Then μ is called *absolutely continuous* with respect to ν ($\mu \ll \nu$) if for every $A \in \mathcal{X}$, $\nu(A) = 0$ implies $\mu(A) = 0$. If $\mu \ll \nu$ then by the Radon–Nikodym theorem there exists $\rho \in L^1(X, \mathcal{X}, \nu)$, called *density* of μ with respect to ν, such that

$$\mu(A) = \int_A \rho \,\mathrm{d}\nu \quad \text{for all } A \in \mathcal{X}.$$

If $\mu \ll \nu$ and $\nu \ll \mu$ then we say that μ and ν are *equivalent*. If, on the other hand, there is an $A \in \mathcal{X}$ such that $\mu(A) = 0$ and $\nu(X \setminus A) = 0$, then μ and ν are called *singular*.

A function $f: X \to Y$ between two measurable spaces $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ is called *measurable* if $f^{-1}(A) \in \mathcal{X}$ for every $A \in \mathcal{Y}$. A measureable function $u: X \to Y$ between a probability space $(X, \mathcal{X}, \mu)$ and a measurable space $(Y, \mathcal{Y})$, in turn, is called *random variable with values in* Y. In this case, the probability measure $\mu \circ u^{-1}$ is called *distribution of* u. For $A \in \mathcal{Y}$, we denote the probability of $\{u \in A\} := u^{-1}(A)$ by

$$\mathbb{P}[u \in A] := \mu(u^{-1}(A)).$$

If a random variable u is Bochner integrable (with respect to μ), then

$$\mathbb{E}[u] := \int_X u(x)\mu(\mathrm{d}x)$$

is called *mean*, *expected value* or *expectation* of u.

1.2 Regular Conditional Distributions

Consider a pair of random variables (u, y) with values in the measurable space $(X \times Y, \mathcal{X} \times \mathcal{Y})$. In the context of Bayesian inference, we will denote the parameter by u and the data by y. For events $A \in \mathcal{Y}$ and $B \in \mathcal{X}$ with $\mathbb{P}[y \in A] > 0$ the conditional probability is defined as

$$\mathbb{P}[u \in B | y \in A] = \frac{\mathbb{P}[u \in B, y \in A]}{\mathbb{P}[y \in A]}.$$

However, we want to define conditional probabilities of the form $\mathbb{P}[u \in B | y = y_0]$ for all $y_0 \in Y$, that is to say we want to be able to condition on events with probability zero. We will do so in a consistent way by means of regular conditional distributions.

For a fixed $B \in \mathcal{X}$ the *conditional probability of* $\{u \in B\}$ *given* y is defined as the random variable $g(y)$ and denoted by $\mathbb{P}[u \in B | y]$, where $g: Y \to \mathbb{R}$ is a measurable function such that

$$\mathbb{E}[g(y)1_A(y)] = \mathbb{E}[1_B(u)1_A(y)] \quad \text{for every } A \in \mathcal{Y}. \tag{1.1}$$

The existence of such a function can be shown using the Radon–Nikodym theorem as follows. First note that $A \mapsto \mathbb{E}[1_B(u)1_A(y)] = \mathbb{P}[u \in B, y \in A]$ defines a finite measure on $(Y, \mathcal{Y})$. This measure is absolutely continuous with respect to the marginal distribution of y, as $\mathbb{P}[y \in A] = 0$ implies $\mathbb{P}[u \in B, y \in A] = 0$. Thus it has a density g with respect to the marginal distribution of y by the Radon–Nikodym theorem [Klenke 2014, Cor. 7.34], which means that (1.1) is satisfied. The function g is unique up to changes on a null set under the marginal distribution of y, as $g(y) = \tilde{g}(y)$ almost surely whenever $\mathbb{E}[g(y)1_A(y)] = \mathbb{E}[\tilde{g}(y)1_A(y)]$ for all $A \in \mathcal{Y}$. This null set does, however, depend on B.

In order to define a conditional distribution from these conditional probabilities in a consistent way we need additional requirements. A map $G: Y \times \mathcal{X} \to [0, \infty)$ is called *Markov kernel* (or *stochastic kernel*) from $(Y, \mathcal{Y})$ to $(X, \mathcal{X})$ if

(i) for any $B \in \mathcal{X}$ the map $y_0 \mapsto G(y_0, B)$ is $\mathcal{Y}$-measurable, and

(ii) for any $y_0 \in Y$ the map $B \mapsto G(y_0, B)$ is a probability measure on $(X, \mathcal{X})$.

Now, $G\colon Y \times \mathcal{X} \to [0,\infty)$ is called *regular conditional distribution of u given y*, if it is a Markov kernel from $(Y, \mathcal{Y})$ to $(X, \mathcal{X})$ and for every $B \in \mathcal{X}$ we have

$$G(y, B) = \mathbb{P}\left[u \in B | y\right] \quad \text{almost surely,}$$

i.e., if

$$\mathbb{E}\left[G(y,B)1_A(y)\right] = \mathbb{E}\left[1_B(u)1_A(y)\right] \quad \text{for all } B \in \mathcal{X} \text{ and all } A \in \mathcal{Y}.$$

In this case we define the *conditional probability of* $\{u \in B\}$ *given* $y = y_0$ as $\mathbb{P}\left[u \in B | y = y_0\right] := G(y_0, B)$.

A sufficient condition for the existence of a regular conditional distribution is that X is a Polish space and $\mathcal{X}$ its Borel-σ-algebra, see [Klenke 2014, Thm. 8.37]. It is in particular satisfied if X is a separable Banach space equipped with its Borel-σ-algebra. Note that in spite of its name the regular conditional distribution is not actually a probability distribution, but a family $\{G(y_0, \cdot)\}_{y_0 \in Y}$ of probability distributions.

Conditional probabilites can be defined in more generality, conditioning on an arbitrary σ-algebra instead of a random variable, see [Klenke 2014, Sections 8.2 and 8.3]. This is, however, not necessary for our purposes.

1.3 Bayes' Formula

Since there exists no Lebesgue-measure on infinite-dimensional separable Banach spaces, we will state the density of the posterior distribution with respect to the prior distribution.

Proposition 1.1. *On an infinite-dimensional, separable Banach space there exists no locally finite, translation-invariant Borel measure, except for the trivial measure.*

Proof. Let μ be a locally finite, translation-invariant measure on an infinite-dimensional, separable Banach space $(X, \mathcal{B}(X))$. Local finiteness assures that for some $\delta > 0$ the open ball $B_\delta(0)$ has finite μ-measure. Since X is infinite-dimensional, we can, using Riesz's lemma, construct a sequence $\{x_n\}_{n\in\mathbb{N}}$ of points in X such that $\|x_n\| = 1$ for all $n \in \mathbb{N}$ and

$$\|x_n - x\| \geq \frac{2}{3} \quad \text{for all } x \in \operatorname{span}\{x_1, \ldots, x_{n-1}\}.$$

Consequently, the balls $\{B_{\delta/4}(\frac{3}{4}x_n)\}_{n\in\mathbb{N}}$ are all contained in $B_\delta(0)$ and pairwise disjoint. By translation-invariance, all these balls have the same measure, and since

$$\sum_{n=1}^{\infty} \mu\left(B_{\frac{\delta}{4}}\left(\frac{3}{4}x_n\right)\right) \leq \mu(B_\delta(0)) < \infty,$$

the μ-measure of each ball $B_{\delta/4}(\frac{3}{4}x_n)$ must be zero. However, as X is separable, it can be covered by a countable collection of balls of radius $\delta/4$, which is why $\mu(X) = 0$ as well. □

Let μ_0 denote the prior distribution on $(X, \mathcal{X})$. If we assume that a regular conditional distribution $(u_0, A) \mapsto P_{u_0}(A)$ of y given u exists, then the joint distribution of (u, y) is given by

$$\mathbb{P}\left[y \in A, u \in B\right] = \mathbb{E}\left[1_A(y)1_B(u)\right] = \mathbb{E}\left[P_u(A)1_B(u)\right] = \int_B P_u(A)\mathrm{d}\mu_0(u) \tag{1.2}$$

for all $A \in \mathcal{Y}$ and $B \in \mathcal{X}$. This allows us to express the marginal distribution of y as

$$\mathbb{P}[y \in A] = \int_X P_u(A) \mathrm{d}\mu_0(u) \quad \text{for all } A \in \mathcal{Y}. \tag{1.3}$$

We derive Bayes' formula under the assumption that there is a probability measure ν on $(Y, \mathcal{Y})$ such that for every $u_0 \in X$ the measure P_{u_0} is absolutely continuous with respect to ν. Let p_{u_0} denote the density of P_{u_0} with respect to ν, i.e,

$$P_{u_0}(A) = \int_A p_{u_0}(y) \mathrm{d}\nu(y) \quad \text{for all } A \in \mathcal{Y}. \tag{1.4}$$

Then we can write (1.3) as

$$\mathbb{P}[y \in A] = \int_A \int_X p_u(y) \mathrm{d}\mu_0(u) \mathrm{d}\nu(y) \quad \text{for all } A \in \mathcal{Y} \tag{1.5}$$

using Fubini's theorem. This shows that the density of the marginal distribution of y with respect to ν is given by

$$y_0 \mapsto Z(y_0) := \int_X p_u(y_0) \mathrm{d}\mu_0(u).$$

Theorem 1.2. *Assume that there exists a probability measure ν on $(Y, \mathcal{Y})$ such that for every $u_0 \in X$ the measure P_{u_0} is absolutely continuous with respect to ν and let p_{u_0} denote the respective density. Moreover, assume that $Z(y)$ is ν-almost surely positive, where*

$$Z(y_0) := \int_X p_u(y_0) \mathrm{d}\mu_0(u) \quad \textit{for all } y_0 \in Y.$$

If the family of posterior distributions $\{\mu^{y_0}\}_{y_0 \in Y}$ exists in the form of a regular conditional distribution $(y_0, B) \mapsto \mu^{y_0}(B)$ of u given y, then μ^y is ν-almost surely absolutely continuous with respect to the prior distribution μ_0 and in this case the density is given by

$$\frac{\mathrm{d}\mu^{y_0}}{\mathrm{d}\mu_0}(u) = \frac{p_u(y_0)}{\int_X p_{\tilde{u}}(y_0) \mathrm{d}\mu_0(\tilde{u})} \quad \mu_0\textit{-almost surely.} \tag{1.6}$$

Proof. We can express (1.2) using the density p_u as

$$\mathbb{P}[y \in A, u \in B] = \int_A \int_B p_u(y) \mathrm{d}\mu_0(u) \mathrm{d}\nu(y)$$

by means of Fubini's theorem. By definition of the regular conditional distribution this probability is equal to

$$\mathbb{P}[y \in A, u \in B] = \mathbb{E}[1_B(u) 1_A(y)] = \mathbb{E}[\mu^y(B) 1_A(y)] = \int_A \mu^y(B) Z(y) \mathrm{d}\nu(y),$$

which implies that for all $B \in \mathcal{X}$ we have

$$\int_B p_u(y) \mathrm{d}\mu_0(u) = \mu^y(B) Z(y) \quad \nu\text{-almost surely.}$$

As $Z(y)$ is ν-almost surely positive we may divide by $Z(y)$, which yields

$$\mu^y(B) = \frac{\int_B p_u(y)\mathrm{d}\mu_0(u)}{\int_X p_u(y)\mathrm{d}\mu_0(u)} \quad \nu\text{-almost surely.}$$

In particular, μ^y is ν-almost surely absolutely continuous with respect to μ_0 and in this case the density is given by

$$u_0 \mapsto \frac{p_{u_0}(y)}{\int_X p_u(y)\mathrm{d}\mu_0(u)}. \qquad \square$$

In our context equation (1.6) is called *Bayes' formula*. If the densities $\{p_{u_0}\}_{u_0\in X}$ are chosen appropriately and $Z(y_0)$ is positive for all $y_0 \in Y$, then it holds for all $y_0 \in Y$ and defines a regular conditional distribution of u given y.

Theorem 1.3. *Assume that there exists a probability measure ν on $(Y, \mathcal{Y})$ such that for every $u_0 \in X$ the measure P_{u_0} is absolutely continuous with respect to ν and let p_{u_0} denote the respective density. If the densities $\{p_{u_0}\}_{u_0\in X}$ can be chosen in such a way that*

$$Z(y_0) := \int_X p_u(y_0)\mathrm{d}\mu_0(u) > 0 \quad \textit{for all } y_0 \in Y$$

and $(u_0, y_0) \mapsto p_{u_0}(y_0)$ is $\mathcal{X} \times \mathcal{Y}$-measurable, then

$$(y_0, B) \mapsto \mu^{y_0}(B) := \frac{\int_B p_u(y_0)\mathrm{d}\mu_0(u)}{\int_X p_{\tilde{u}}(y_0)\mathrm{d}\mu_0(\tilde{u})}$$

defines a regular conditional distribution of u given y. In particular, for every $y_0 \in Y$ the posterior distribution μ^{y_0} is absolutely continuous with respect to the prior distribution μ_0 and its density is given by

$$\frac{\mathrm{d}\mu^{y_0}}{\mathrm{d}\mu_0}(u) = \frac{p_u(y_0)}{\int_X p_{\tilde{u}}(y_0)\mathrm{d}\mu_0(\tilde{u})} \quad \mu_0\textit{-almost surely.}$$

Proof. We first show that $(y_0, B) \mapsto \mu^{y_0}(B)$ is a Markov kernel. On the one hand, $(u_0, y_0) \mapsto p_{u_0}(y_0)1_B(u_0)$ is $\mathcal{X} \times \mathcal{Y}$-measurable for every $B \in \mathcal{X}$. Therefore,

$$y_0 \mapsto \int_B p_u(y_0)\mathrm{d}\mu_0(u)$$

is $\mathcal{Y}$-measurable for every $B \in \mathcal{X}$ by Fubini's theorem. This implies in particular the $\mathcal{Y}$-measurability of

$$y_0 \mapsto Z(y_0) = \int_X p_u(y_0)\mathrm{d}\mu_0(u).$$

Consequently, $y_0 \mapsto \mu^{y_0}(B)$ is $\mathcal{Y}$-measurable as well for every $B \in \mathcal{X}$. On the other hand, $u_0 \mapsto p_{u_0}(y_0)$ is $\mathcal{X}$-measurable for every $y_0 \in Y$ by [Klenke 2014, Lemma 14.3] and $\mu^{y_0}(X) = 1$, so that μ^{y_0} is a probability measure, see Remark 4.14 in [Klenke 2014]. This shows that $\mu^{y_0}(B)$ is a Markov kernel.

Moreover, by (1.2), (1.4) and the definition of $\mu^y(B)$ we have

$$\mathbb{E}\left[1_B(u)1_A(y)\right] = \mathbb{P}\left[y \in A, u \in B\right] = \int_A \int_B p_u(y)\mathrm{d}\mu_0(u)\mathrm{d}\nu(y) = \int_A \mu^y(B)Z(y)\mathrm{d}\nu(y)$$

for all $B \in \mathcal{X}$ and $A \in \mathcal{Y}$. Since Z is the density of the marginal distribution of y with respect to ν by (1.5), this yields

$$\mathbb{E}\left[1_B(u)1_A(y)\right] = \int_A \mu^y(B)Z(y)\mathrm{d}\nu(y) = \mathbb{E}\left[\mu^y(B)1_A(y)\right].$$

So $(y_0, B) \mapsto \mu^{y_0}(B)$ is indeed a regular conditional distribution of u given y. □

In many cases ν is absolute continuous with respect to P_{u_0} for every $u_0 \in X$ as well. Then $p_{u_0}(y)$ is ν-almost surely positive, so that we can express the density p_{u_0} as

$$p_{u_0}(y) = \exp(-\Phi(u_0, y)) \quad \nu\text{-almost surely,}$$

using a measurable function $\Phi\colon X \times Y \to \mathbb{R}$, which we call *potential* (or *negative log-likelihood*). In this case Bayes' formula can be written in the form

$$\frac{\mathrm{d}\mu^{y_0}}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u, y_0))}{\int_X \exp(-\Phi(\tilde{u}, y_0))\mathrm{d}\mu_0(\tilde{u})} \quad \mu_0\text{-almost surely.}$$

1.4 Bayesian Inverse Problems

Here we briefly discuss the case when the model is defined by an operator equation with additive noise. For more information on the Bayesian approach to inverse problems see, e.g., [Dashti and Stuart 2017]. Let X and Y be separable Banach spaces, each equipped with its Borel σ-algebra. We assume that parameter and data follow the relation

$$y = F(u) + \eta,$$

where F is a (possibly nonlinear) operator from X to Y and η is stochastic noise, independent of u. In this case, the model describes both the behaviour of the forward operator F and the effect of the noise η. The distribution ν of the noise η on $(Y, \mathcal{B}(Y))$ plays the role of a reference measure. For every $y_0 \in Y$ we define the shifted measure

$$\nu_{y_0} := \nu(\cdot - y_0).$$

Proposition 1.4. *If $\nu_{F(u_0)}$ is absolutely continuous with respect to ν for every $u_0 \in X$, then*

$$(u_0, A) \mapsto P_{u_0}(A) := \nu_{F(u_0)}(A)$$

is a regular conditional distribution of y given u.

Proof. By definition, $\nu_{F(u_0)}$ is a probability measure for every $u_0 \in X$. Let

$$p_{u_0} := \frac{\mathrm{d}\nu_{F(u_0)}}{\mathrm{d}\nu}$$

for every $u_0 \in X$ denote the density of $\nu_{F(u_0)}$ with respect to ν. Since $\nu_{F(u_0)} \ll \nu$ for all $u_0 \in X$, the family $\{\nu_{F(u_0)}\}_{u_0 \in X}$ is separable with respect to the Hellinger distance by [Strasser 1985, Lemma 4.1]. Consequently, by [Strasser 1985, Lemma 4.6], the densities $\{p_{u_0}\}_{u_0 \in X}$ can be chosen such that $(u_0, y_0) \mapsto p_{u_0}(y_0)$ is $\mathcal{B}(X) \times \mathcal{B}(Y)$-measurable. Then $(u_0, y_0) \mapsto p_{u_0}(y_0)1_A(y_0)$ is measurable as well for any $A \in \mathcal{B}(Y)$, which in turn implies that

$$u_0 \mapsto \int_A p_{u_0}(y)\mathrm{d}\nu(y) = \int_A \mathrm{d}\nu_{F(u_0)}(y) = \nu_{F(u_0)}(A)$$

is measurable by Fubini's theorem. This shows that $(u_0, A) \mapsto \nu_{F(u_0)}(A)$ is a Markov kernel. Now for every $A \in \mathcal{B}(Y)$ the conditional probability of $\{y \in A\}$ given u is given by

$$\mathbb{P}\left[y \in A | u\right] = \nu_{F(u)}(A) \quad \mu_0\text{-almost surely,}$$

since

$$\begin{aligned}\mathbb{E}\left[\nu_{F(u)}(A)1_B(u)\right] &= \int_B \nu(A - F(u))\mathrm{d}\mu_0(u) = \int_B \mathbb{P}\left[\eta \in A - F(u)\right]\mathrm{d}\mu_0(u) \\ &= \int_B \mathbb{P}\left[\eta + F(u) \in A\right]\mathrm{d}\mu_0(u) = \mathbb{E}\left[1_A(F(u) + \eta)1_B(u)\right] = \mathbb{E}\left[1_A(y)1_B(u)\right]\end{aligned}$$

for all $B \in \mathcal{B}(X)$. This shows that $(u_0, A) \mapsto P_{u_0}(A) := \nu_{F(u_0)}(A)$ is a regular conditional distribution of y given u. □

Typically, $\nu_{F(u_0)}$ is not only absolutely continuous with respect to ν but even equivalent for all $u_0 \in X$ (see Theorems 2.4 and 3.10 below). Then the family $\{p_{u_0}\}_{u_0 \in X}$ of densities can be expressed as

$$\frac{\mathrm{d}\nu_{F(u_0)}}{\mathrm{d}\nu}(y) = p_{u_0}(y) = \exp(-\Phi(u_0, y)) \quad \nu\text{-almost surely,}$$

using a measurable function $\Phi\colon X \times Y \to \mathbb{R}$. As described in the proof of Proposition 1.4 the densities $(u_0, y_0) \mapsto p_{u_0}(y_0) = \exp(-\Phi(u_0, y_0))$ and therefore also Φ can always be chosen to be $\mathcal{B}(X) \times \mathcal{B}(Y)$-measurable.

2 Gaussian Measures on Hilbert Spaces

The way we will define Laplacian measures on infinite-dimensional Hilbert spaces is strongly influenced by the possibility of representing Gaussian measures on separable Hilbert spaces as a product measure of Gaussian measures on the real line. For this reason we recapitulate the definition of a Gaussian measure on a locally convex space, give a brief review of the construction of Gaussian measures on separable Hilbert spaces as infinite-dimensional product measures and present a result regarding the equivalence of Gaussian measures and their Radon–Nikodym derivative with respect to each other.

2.1 Definition on Locally Convex Spaces

The definition of a Gaussian measure on a locally convex space is based upon the one of a Gaussian measure on $\mathbb{R}$. Here we follow the way they are defined in [Bogachev 1998].

Definition 2.1 ([Bogachev 1998, Def. 1.1.1])**.** A Borel probability measure γ on $\mathbb{R}$ is called *Gaussian* if it is either the Dirac measure δ_a at a point $a \in \mathbb{R}$ or has density

$$p : t \mapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-a)^2}{2\sigma^2}\right)$$

with respect to the Lebesgue measure for some $a \in \mathbb{R}$ and $\sigma > 0$. In the latter case the measure γ is called nondegenerate.

The Dirac measure δ_a on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined by

$$\delta_a(B) = \begin{cases} 1 & \text{if } a \in B, \\ 0 & \text{if } a \notin B. \end{cases}$$

For any Dirac measure we put $\sigma = 0$. The mean and the variance of a Gaussian measure γ are given by

$$\int_{\mathbb{R}} t\,\gamma(\mathrm{d}t) = a, \qquad \int_{\mathbb{R}} (t-a)^2\,\gamma(\mathrm{d}t) = \sigma^2.$$

We will denote this measure by $\mathcal{N}_{a,\sigma^2}$. A measure with $a = 0$ and $\sigma = 1$ is called *standard*, a mean zero Gaussian measure is called *centred*. When $a = 0$ we write $\mathcal{N}_{\sigma^2}$ instead of $\mathcal{N}_{a,\sigma^2}$ for short. The Fourier transform (or characteristic function) of a Gaussian measure γ with parameters (a, σ^2) is given by

$$\widehat{\gamma}(y) := \int_{\mathbb{R}} \exp(iyx)\,\gamma(\mathrm{d}x) = \exp\left(iay - \frac{1}{2}\sigma^2 y^2\right) \quad \text{for all } y \in \mathbb{R}.$$

Now let X be a locally convex space and X^* its dual space. We call a set $C \subset X$ *cylindrical* if it has the form

$$C = \{x \in X : (l_1(x), \dots, l_n(x)) \in C_0\}, \quad l_i \in X^*.$$

Denote by $\mathcal{E}(X)$ the σ-field generated by all cylindrical subsets of X. This is the minimal σ-field, with respect to which all continuous linear functionals on X are measurable. While $\mathcal{E}(X)$ is always contained in the Borel σ-field $\mathcal{B}(X)$, it may not coincide with it. However, for separable Fréchet spaces, so in particular for separable Banach spaces, the equality $\mathcal{E}(X) = \mathcal{B}(X)$ does hold true, see [Bogachev 1998, Thm. A.3.7]. Gaussian measures on X are now defined via their pushforwards under continuous linear functionals.

Definition 2.2 ([Bogachev 1998, Def. 2.2.1 (ii)])**.** Let X be a locally convex space. A probability measure γ defined on the σ-field $\mathcal{E}(X)$, generated by X^*, is called *Gaussian* if, for any $f \in X^*$, the induced measure $\gamma \circ f^{-1}$ on $\mathbb{R}$ is Gaussian. The measure γ is called *centred* if all the measures $\gamma \circ f^{-1}$, $f \in X^*$, are centred.

A Gaussian measure on a separable Hilbert space can be expressed in terms of its mean and its covariance operator. Let μ be a probability measure on a separable Hilbert space X. If $x \mapsto x$ is Bochner integrable with respect to μ then μ is said to have *finite expectation* and its *expectation*, *expected value* or *mean* $a \in X$ is defined as

$$a := \int_X x\mu(\mathrm{d}x).$$

If the map $x \mapsto \|x\|_X^2$ is Bochner integrable with respect to μ then μ is said to have *finite variance* and the bounded linear operator $Q\colon X \to X$ defined by

$$Qh := \int_X (h, x-a)(x-a)\mu(\mathrm{d}x)$$

is called its *covariance operator*. Furthermore, the *Fourier transform* (or *characteristic function*) $\hat{\mu}$: $X \to \mathbb{R}$ of μ is defined by

$$\hat{\mu}(h) := \int_X e^{i(h,x)}\mu(\mathrm{d}x).$$

Theorem 2.3 ([Bogachev 1998, Thm. 2.3.1])**.** *Let γ be a Gaussian measure on a separable Hilbert space X and let X^* be identified with X by means of the Riesz representation. Then there exist a vector $a \in X$ and a symmetric nonnegative nuclear operator K such that the Fourier transform of the measure γ equals*

$$x \mapsto \exp\left(i(a,x) - \frac{1}{2}(Kx,x)\right). \tag{2.1}$$

Conversely, for every pair (a, K) of the aforementioned type, the function (2.1) *is the Fourier transform of a Gaussian measure on the space X. In addition, a is the mean of the measure γ and K is its covariance operator.*

We denote such a Gaussian measure by $\mathcal{N}_{a,Q}$, and by $\mathcal{N}_Q$ if $a = 0$.

2.2 Representation on Separable Hilbert Spaces

Now we take a look at the situation in infinite-dimensional separable Hilbert spaces. The separability permits the representation of Gaussian measures as an infinite product of Gaussian measures on $\mathbb{R}$. Here we briefly review the construction of Gaussian measures in [Da Prato 2001]. In Chapter 3, we will, however, construct Laplacian measures on separable Hilbert spaces in the same way and portray their construction in detail.

Let H be an infinite-dimensional separable Hilbert space. A continuous linear operator Q on H is called *symmetric* if

$$(Qx, y) = (x, Qy) \quad \text{for all } x, y \in H,$$

Q is called *positive* if

$$(Qx, x) \geq 0 \quad \text{for all } x \in H,$$

and Q is called *trace class* if

$$\operatorname{Tr} Q := \sum_{k=1}^{\infty} (Qe_k, e_k) < \infty$$

for any orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ in H.

For any $a \in H$ and any symmetric, positive trace class operator $Q \in L(H)$ the measure $\mathcal{N}_{a,Q}$ can be represented as follows. Since Q is of trace class, there exists an orthonormal basis $(e_k)_{k\in\mathbb{N}}$ in H and a sequence of nonnegative numbers $(\lambda_k)_{k\in\mathbb{N}}$ such that

$$Qe_k = \lambda_k e_k \quad \text{for all } k \in \mathbb{N}.$$

We identify H with ℓ^2 via the natural isomorphism γ, defined by

$$\gamma(x) = \sum_{k=1}^{\infty} x_k e_k \quad \text{for all } x \in \ell^2.$$

Then we define the product measure

$$\mu = \bigotimes_{k=1}^{\infty} \mathcal{N}_{a_k, \lambda_k}$$

on $\mathbb{R}^\infty := \prod_{k=1}^{\infty} \mathbb{R}$, where $a_k = (a, e_k)$ for all $k \in \mathbb{N}$. This measure is concentrated on ℓ^2, that is $\mu(\ell^2) = 1$, see [Da Prato 2001, Prop. 1.3.5]. Now the pushforward $\mu \circ \gamma^{-1}$ of the measure μ under γ is a Gaussian measure on H with mean a and covariance operator Q, because its Fourier transform is given by

$$\widehat{(\mu \circ \gamma^{-1})}(h) = e^{i(a,h) - \frac{1}{2}(Qh,h)} \quad \text{for all } h \in H$$

(see [Da Prato 2001, Prop. 1.3.7]). This results in the representation

$$\mathcal{N}_{a,Q} = \left(\bigotimes_{k=1}^{\infty} \mathcal{N}_{a_k, \lambda_k} \right) \circ \gamma^{-1}.$$

By [Da Prato 2001, Prop. 1.3.7], its mean and covariance operator satisfy

$$\int_H (a,h)\,\mathcal{N}_{a,Q}(\mathrm{d}x) = (a,h), \quad h \in H,$$

$$\int_H (x-a,h)(x-a,k)\,\mathcal{N}_{a,Q}(\mathrm{d}x) = (Qh,k), \quad h,k \in H.$$

2.3 The Cameron–Martin Theorem

Given an infinite-dimensional separable Hilbert space H and a symmetric, positive trace class operator $Q \in L(H)$, we present a criterion for the equivalence of the centred Gaussian measure $\mathcal{N}_Q$ and the Guassian measure $\mathcal{N}_{a,Q}$ with mean $a \in H$.

We assume that $\ker(Q) = \{0\}$. Let (e_k) again denote an orthonormal basis of H, such that $Qe_k = \lambda_k e_k$, $k \in \mathbb{N}$, where (λ_k) are the eigenvalues of Q. We introduce the operator $Q^{1/2}$ on H, which is defined by

$$Q^{\frac{1}{2}} x = \sum_{k=1}^{\infty} \sqrt{\lambda_k}(x,e_k)e_k.$$

Its range is given by

$$\mathcal{R}(Q^{\frac{1}{2}}) = \left\{ y \in H : \sum_{k=1}^{\infty} \frac{y_k^2}{\lambda_k} < \infty \right\}$$

and $\mathcal{R}(Q^{1/2})$ is a dense proper subspace of H. However,

$$\mathcal{N}_Q(\mathcal{R}(Q^{1/2})) = 0,$$

see [Da Prato 2001, Prop. 1.5.2].

Theorem 2.4 (Cameron–Martin). *(i) If $a \notin \mathcal{R}(Q^{1/2})$ then $\mathcal{N}_{a,Q}$ and $\mathcal{N}_Q$ are singular.*

(ii) If $a \in \mathcal{R}(Q^{1/2})$ then $\mathcal{N}_{a,Q}$ and $\mathcal{N}_Q$ are equivalent. Moreover, the density $\frac{\mathcal{N}_{a,Q}}{\mathcal{N}_Q}$ is given by

$$\frac{\mathcal{N}_{a,Q}}{\mathcal{N}_Q}(x) = \prod_{k=1}^{\infty} \exp\left(-\frac{1}{2}\frac{a_k^2}{\lambda_k} + \frac{a_k x_k}{\lambda_k} \right) \tag{2.2}$$

for $\mathcal{N}_Q$-almost all $x \in H$, where $a_k := (a,e_k)$ and $x_k := (x,e_k)$.

A proof of Theorem 2.4 can be found in [Da Prato 2001, Thm. 2.3.1]. Equation (2.2) is called *Cameron–Martin formula* and the space $\mathcal{R}(Q^{1/2})$ *Cameron–Martin space* of $\mathcal{N}_Q$. The Cameron–Martin theorem shows in particular that on an infinite-dimensional separable Hilbert space not every translate $\mathcal{N}_{a,Q}$, $a \in H$, of a centred Gaussian measure $\mathcal{N}_Q$ is equivalent to $\mathcal{N}_Q$.

The following proposition is a special case of Proposition 3 in Section 18 of [Lifšic 1995].

Proposition 2.5. *Let $A \subset X$ be a convex, symmetric about zero, weakly bounded Borel subset of X and $h \in \mathcal{R}(Q^{1/2})$. Then we have*

$$\lim_{r\to 0} \frac{\mathcal{N}_Q(h+rA)}{\mathcal{N}_Q(rA)} = \exp\left(-\frac{1}{2}\left\|Q^{-\frac{1}{2}}h\right\|_X^2 \right).$$

It can be used to describe the asymptotic probability of small balls around two points. Let $B_r(x) \subset X$ denote the open ball with radius r centred at $x \in X$.

Corollary 2.6. *For all $h_1, h_2 \in \mathcal{R}(Q^{1/2})$ we have*

$$\lim_{r\to 0} \frac{\mathcal{N}_Q(B_r(h_1))}{\mathcal{N}_Q(B_r(h_2))} = \exp\left(\frac{1}{2}\|Q^{-\frac{1}{2}}h_2\|_X^2 - \frac{1}{2}\|Q^{-\frac{1}{2}}h_1\|_X^2\right).$$

Proof. It follows from Proposition 2.5 that

$$\frac{\mathcal{N}_Q(B_r(h_1))}{\mathcal{N}_Q(B_r(h_2))} = \frac{\mathcal{N}_Q(rB_1(0))}{\mathcal{N}_Q(h_2 + rB_1(0))} \cdot \frac{\mathcal{N}_Q(h_1 + rB_1(0))}{\mathcal{N}_Q(rB_1(0))}$$

converges towards $\exp(\frac{1}{2}\|Q^{-\frac{1}{2}}h_2\|_X^2 - \frac{1}{2}\|Q^{-\frac{1}{2}}h_1\|_X^2)$ as $r \to 0$. □

3 Laplacian Measures on Hilbert Spaces

In this section we will define Laplacian measures on an infinite-dimensional separable real Hilbert space. First we will review the common definition of multivariate Laplace distributions and conclude that for our purposes, they lack some desirable properties. Therefore we will not try to generalise this definition to infinite-dimensional spaces, but instead construct a Laplacian measure as an infinite product measure, similar to the construction of Gaussian measures in [Da Prato 2001]. It will be based upon the Laplacian measures on $\mathbb{R}$ and $\mathbb{R}^d$, which we will study first.

3.1 Laplacian Measures on $\mathbb{R}$

First we introduce Laplacian measures on $\mathbb{R}$ and summarise some basic properties. For any $a \in \mathbb{R}$ and $\lambda \geq 0$ we define the *Laplacian measure* $\mathcal{L}_{a,\lambda}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as follows. If $\lambda = 0$ we set

$$\mathcal{L}_{a,0} = \delta_a,$$

where δ_a is the Dirac measure at a. If $\lambda > 0$ we set

$$\mathcal{L}_{a,\lambda}(B) = \frac{1}{\sqrt{2\lambda}} \int_B e^{-\frac{\sqrt{2}|x-a|}{\sqrt{\lambda}}} dx \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

$\mathcal{L}_{a,\lambda}$ is a probability measure, since

$$\mathcal{L}_{a,\lambda}(\mathbb{R}) = \frac{1}{\sqrt{2\lambda}} \int_{\mathbb{R}} e^{-\sqrt{\frac{2}{\lambda}}|x-a|} dx = \frac{1}{2} \int_{\mathbb{R}} e^{-|x|} dx = 1.$$

If $a = 0$, we write $\mathcal{L}_\lambda$ short hand for $\mathcal{L}_{0,\lambda}$. This measure has the following properties.

Proposition 3.1. *The probability measure $\mathcal{L}_{a,\lambda}$ has mean a and variance λ, i.e.,*

$$\int_{\mathbb{R}} x \mathcal{L}_{a,\lambda}(dx) = a,$$

$$\int_{\mathbb{R}} (x-a)^2 \mathcal{L}_{a,\lambda}(dx) = \lambda.$$

Moreover, its characteristic function $\widehat{\mathcal{L}_{a,\lambda}}$ *is given by*

$$\widehat{\mathcal{L}_{a,\lambda}}(h) := \int_{\mathbb{R}} e^{ihx} \mathcal{L}_{a,\lambda}(dx) = \frac{e^{iah}}{1 + \frac{1}{2}\lambda h^2} \quad \forall h \in \mathbb{R}.$$

We will call a real valued random variable *Laplacian* if its probability distribution is a Laplacian measure $\mathcal{L}_{a,\lambda}$ on $\mathbb{R}$ for some $a \in \mathbb{R}$ and $\lambda \geq 0$.

3.2 Elliptically Contoured Laplace Distributions on $\mathbb{R}^d$

Now we introduce and discuss the common definition of multivariate Laplace distributions. First of all, multivariate Laplace distributions are defined differently by different authors – for details we refer to the Introduction to Part II of [Kotz, Kozubowski, and Podgórski 2001]. Most often, however, they are defined as a class of elliptically contoured distributions, which means that their characteristic function and their density depend on its variables only through a quadratic form.

Let $Q \in \mathbb{R}^{d\times d}$ be a nonnegative definite symmetric matrix. Then a d-dimensional distribution is said to be *multivariate symmetric Laplace* with parameter Q, denoted $\mathcal{SL}_d(Q)$ if its characteristic function is of the form

$$\Psi(h) = \frac{1}{1 + \frac{1}{2}(Qh, h)} \quad \text{for all } h \in \mathbb{R}^d.$$

This distribution is centred at zero and its covariance matrix is given by Q. Moreover, its density function (for a nonsingular distribution) is given by

$$g(x) = (2\pi)^{-\frac{2}{d}} \det(Q)^{-\frac{1}{2}} \int_0^\infty \exp\left(-\frac{(Q^{-1}x, x)}{2z} - z\right) z^{-\frac{d}{2}} \mathrm{d}z \quad \text{for all } x \in \mathbb{R}^d \setminus \{0\},$$

see Corollary 6.5.1 in [Kotz, Kozubowski, and Podgórski 2001]. The density can also be expressed in terms of the modified Bessel function of the third kind K_ν as

$$g(x) = 2(2\pi)^{-\frac{2}{d}} \det(Q)^{-\frac{1}{2}} \left(\frac{(Q^{-1}x, x)}{2}\right)^{\frac{\nu}{2}} K_\nu\left(\sqrt{2(Q^{-1}x, x)}\right) \quad \text{for all } x \in \mathbb{R}^d \setminus \{0\},$$

where $\nu := (2-d)/2$, see Section 5.2.2 in [Kotz, Kozubowski, and Podgórski 2001]. Note that the density tends to infinity as $x \to 0$ unless $d = 1$.

A random variable $X \sim \mathcal{SL}_d(Q)$ has the following representation. Let Z be a Gaussian random variable on $\mathbb{R}^d$ with mean zero and covariance matrix Q and let W be exponentially distributed on $\mathbb{R}$ with mean 1, independent of Z. Then

$$X \overset{\mathrm{d}}{=} \sqrt{W} Z$$

This means that a multivariate symmetric Laplacian random variable can be thought of as a centred Gaussian random variable with stochastic variance which has an exponential distribution.

Let $Z = (Z_1, \dots, Z_d)$ be a multivariate Gaussian random variable with mean zero and covariance matrix Q. If Z is uncorrelated, i.e., if Q is diagonal, then its components $Z_1, \dots, Z_d$ are independent. In contrast, this is not the case for a multivariate symmetric Laplacian random variable $X \sim \mathcal{SL}_d(Q)$.

One way to generalise this definition to an infinite-dimensional Hilbert space H would be to seek a distribution whose characteristic function is given by

$$\Phi(h) = \frac{1}{1 + \frac{1}{2}(Qh, h)_H} \quad \text{for all } h \in H,$$

or to define an H valued random variable X via

$$X = \sqrt{W} Z,$$

where Z is an H valued Gaussian random variable with mean zero and W is exponentially distributed with mean 1, respectively.

We do, however, want to generalise the 1-dimensional Laplace distribution to a multivariate Laplace distribution whose density depends on its variable through a weighted 1-norm rather than a quadratic form. So we do not want the distribution to be elliptically contoured. Also, we want the components of a multivariate Laplacian random variable to be independent. This leads to the definition of a Laplacian measure on $\mathbb{R}^d$ as a product measure.

3.3 Laplacian Product Measures on $\mathbb{R}^d$

Finite Laplacian product measures are of interest not only for their own sake but also because the definition of Laplacian measures on infinite-dimensional Hilbert spaces will be based upon theirs. We want to define a probability measure $\mathcal{L}_{a,Q}$ on $\mathbb{R}^d$ for any $a \in \mathbb{R}^d$ and any self-adjoint positive $Q \in \mathcal{L}(\mathbb{R}^d)$ in such a way that it has mean a and covariance operator Q.

Let $Q \in \mathcal{L}(\mathbb{R}^d)$ be self-adjoint and positive definite and let $e_1, \dots, e_d$ be an orthonormal basis of H consisting of eigenvectors of Q, such that

$$Qe_k = \lambda_k e_k, \quad \text{for } k = 1, \dots, d,$$

with the associated eigenvalues $\lambda_k \geq 0$. Now we define the *Laplacian measure* $\mathcal{L}_{a,Q}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ as the pushforward

$$\mathcal{L}_{a,Q} = \mu \circ \gamma^{-1}$$

of the product measure

$$\mu = \bigotimes_{k=1}^{d} \mathcal{L}_{a_k, \lambda_k}.$$

under the isomorphism $\gamma\colon (x_1, \dots, x_d) \mapsto \sum_{k=1}^d x_k e_k$, where

$$a_k = (a, e_k) \quad \text{for } k = 1, \dots, d.$$

If $a = 0$ we again write $\mathcal{L}_Q$ short hand for $\mathcal{L}_{0,Q}$.

Remark 3.2. Note that although we do not include the basis $e_1, \dots, e_d$ into the notation $\mathcal{L}_{a,Q}$, the definition of the Laplacian measure does depend on its specific choice.

Let $n \leq d$, $k_1 < \cdots < k_n \leq d$ be positive integers. For every $A \in \mathcal{B}(\mathbb{R}^n)$ let $I_{k_1,\dots,k_n;A}$ denote the cylindrical subset

$$I_{k_1,\dots,k_n;A} := \{(x_1, \dots, x_d) \in \mathbb{R}^d : (x_{k_1}, \dots, x_{k_n}) \in A\}.$$

The product measure μ has the essential property that

$$\mu(I_{k_1,\dots,k_n;A}) = \big(\mathcal{L}_{(a,e_{k_1}),\lambda_{k_1}} \times \cdots \times \mathcal{L}_{(a,e_{k_d}),\lambda_{k_d}}\big)(A)$$

for all $A \in \mathcal{B}(\mathbb{R}^n)$. Consequently,

$$\begin{aligned}\mathcal{L}_{a,Q}\big(\{x \in \mathbb{R}^d : \big((x, e_{k_1}), \dots, (x, e_{k_d})\big) \in A\}\big) &= (\mu \circ \gamma^{-1})\big(\gamma(I_{k_1,\dots,k_n;A})\big) \\ &= \big(\mathcal{L}_{a_{k_1},\lambda_{k_1}} \times \cdots \times \mathcal{L}_{a_{k_d},\lambda_{k_d}}\big)(A)\end{aligned}$$

for all $A \in \mathcal{B}(\mathbb{R}^n)$. The measure $\mathcal{L}_{a,Q}$ has the following basic properties.

Proposition 3.3. *Let $a \in \mathbb{R}^d$ and $Q \in \mathcal{L}(\mathbb{R}^d)$ be self-adjoint and positive. Then $\mathcal{L}_{a,Q}$ has mean a and covariance operator Q, i.e.,*

$$\int_{\mathbb{R}^d} x\mathcal{L}_{a,Q}(dx) = a,$$

$$\int_{\mathbb{R}^d} (y, x-a)(z, x-a)\mathcal{L}_{a,Q}(dx) = (Qy, z) \quad \forall y, z \in \mathbb{R}^d.$$

Moreover, its characteristic function $\widehat{\mathcal{L}_{a,Q}}$ is given by

$$\widehat{\mathcal{L}_{a,Q}}(h) := \int_{\mathbb{R}^d} e^{i(h,x)}\mathcal{L}_{a,Q}(dx) = e^{i(a,h)} \prod_{k=1}^{d} \frac{1}{1 + \frac{1}{2}\lambda_k (h, e_k)^2} \quad \forall h \in \mathbb{R}^d.$$

Finally, if Q is injective, then we have

$$\begin{aligned}\mathcal{L}_{a,Q}(B) &= \frac{1}{\sqrt{2^d \det Q}} \int_B \exp\left(-\sum_{k=1}^{d} \frac{\sqrt{2}|(x-a, e_k)|}{\sqrt{\lambda_k}}\right) dx \\ &= \frac{1}{\sqrt{2^d \det Q}} \int_B \exp\left(-\sqrt{2}\sum_{k=1}^{d} \left|\left(Q^{-\frac{1}{2}}(x-a), e_k\right)\right|\right) dx\end{aligned}$$

for all $B \in \mathcal{B}(\mathbb{R}^d)$.

Proof. For every $y \in \mathbb{R}^d$ we have

$$\begin{aligned}\left(\int_{\mathbb{R}^d} x\mathcal{L}_{a,Q}(\mathrm{d}x), y\right) &= \int_{\mathbb{R}^d} (x, y)\mathcal{L}_{a,Q}(\mathrm{d}x) = \int_{\mathbb{R}^d} \sum_{k=1}^{n} (x, e_k)(e_k, y)\mathcal{L}_{a,Q}(\mathrm{d}x) \\ &= \sum_{k=1}^{n} (e_k, y) \int_{\mathbb{R}} \tilde{x}\mathcal{L}_{(a,e_k),\lambda_k}(\mathrm{d}\tilde{x}) = \sum_{k=1}^{n} (a, e_k)(e_k, y) = (a, y)\end{aligned}$$

by Proposition 3.1, which implies $\int_{\mathbb{R}^d} x\mathcal{L}_{a,Q}(\mathrm{d}x) = a$.

Next, we consider

$$\int_{\mathbb{R}^d} (x-a, y)(x-a, z)\mathcal{L}_{a,Q}(\mathrm{d}x) = \sum_{k=1}^{n}\sum_{j=1}^{n} (e_j, y)(e_k, z) \int_{\mathbb{R}^d} (x-a, e_j)(x-a, e_k)\mathcal{L}_{a,Q}(\mathrm{d}x)$$

for $y, z \in \mathbb{R}^d$. As

$$\begin{aligned}
&\int_{\mathbb{R}^d} (x-a, e_j)(x-a, e_k)\mathcal{L}_{a,Q}(\mathrm{d}x) \\
&= \int_{\mathbb{R}^d} \big((x, e_j) - (a, e_j)\big)\big((x, e_k) - (a, e_k)\big)\mathcal{L}_{a,Q}(\mathrm{d}x) \\
&= \int_{\mathbb{R}^2} \big(\tilde{x}_1 - (a, e_j)\big)\big(\tilde{x}_2 - (a, e_k)\big)\mathcal{L}_{(a,e_j),\lambda_j}(\mathrm{d}\tilde{x}_1)\mathcal{L}_{(a,e_k),\lambda_k}(\mathrm{d}\tilde{x}_2) \\
&= \left(\int_{\mathbb{R}} \tilde{x}_1\mathcal{L}_{(a,e_j),\lambda_j}(\mathrm{d}\tilde{x}_1) - (a, e_j)\right)\left(\int_{\mathbb{R}} \tilde{x}_2\mathcal{L}_{(a,e_k),\lambda_k}(\mathrm{d}\tilde{x}_2) - (a, e_k)\right) = 0
\end{aligned}$$

for $j \neq k$ and

$$\int_{\mathbb{R}^d} (x-a, e_k)^2\mathcal{L}_{a,Q}(\mathrm{d}x) = \int_{\mathbb{R}} (\tilde{x} - (a, e_k))^2\mathcal{L}_{(a,e_k),\lambda_k}(\mathrm{d}\tilde{x}) = \lambda_k$$

by Proposition 3.1, we obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} (x-a, y)(x-a, z)\mathcal{L}_{a,Q}(\mathrm{d}x) &= \sum_{k=1}^{n}\sum_{j=1}^{n}(e_j, y)(e_k, z)\int_{\mathbb{R}^d} (x-a, e_j)(x-a, e_k)\mathcal{L}_{a,Q}(\mathrm{d}x) \\
&= \sum_{k=1}^{n}(y, e_k)(e_k, z)\lambda_k = \sum_{k=1}^{n}(y, Qe_k)(e_k, z) \\
&= \sum_{k=1}^{n}(Qy, e_k)(e_k, z) = (Qy, z).
\end{aligned}$$

Concerning the characteristic function of $\mathcal{L}_{a,Q}$, Proposition 3.1 yields

$$\begin{aligned}
\int_{\mathbb{R}^d} e^{i(h,x)}\mathcal{L}_{a,Q}(\mathrm{d}x) &= \int_{\mathbb{R}^d} e^{i\sum_{k=1}^{n}(x,e_k)(e_k,h)}\mathcal{L}_{a,Q}(\mathrm{d}x) = \int_{\mathbb{R}^d}\prod_{k=1}^{n} e^{i(x,e_k)(e_k,h)}\mathcal{L}_{a,Q}(\mathrm{d}x) \\
&= \prod_{k=1}^{n}\int_{\mathbb{R}} e^{i\tilde{x}_k(e_k,h)}\mathcal{L}_{(a,e_k),\lambda_k}(\mathrm{d}\tilde{x}_k) = \prod_{k=1}^{n}\frac{e^{i(a,e_k)(e_k,h)}}{1+\frac{1}{2}\lambda_k(h,e_k)^2} \\
&= e^{i(a,h)}\prod_{k=1}^{n}\frac{1}{1+\frac{1}{2}\lambda_k(h,e_k)^2}
\end{aligned}$$

for all $h \in \mathbb{R}^d$.

In order to find the Lebesgue density of $\mathcal{L}_{a,Q}$, it is sufficient to consider measurable rectangles in $\mathbb{R}^d$, i.e., sets of the form $R = B_1 \times \cdots \times B_d$ with $B_1, \ldots, B_d \in \mathcal{B}(\mathbb{R})$, since they generate the product σ-algebra $\mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d$. Equivalently, we may consider the images of measurable rectangles under the isomorphism γ, which can be expressed as $\gamma(R) = \{x \in \mathbb{R}^d : (x, e_k)_k \in$

B_k for $k = 1, \dots, d\}$. Now, the definition of the product measure and a change of variables yields

$$\begin{aligned}\mathcal{L}_{a,Q}(\gamma(R)) &= \prod_{k=1}^{d} \mathcal{L}_{(a,e_k),\lambda_k}(B_k) = \prod_{k=1}^{d} \left(\frac{1}{\sqrt{2\lambda_k}} \int_{B_k} \exp\left(-\frac{\sqrt{2}|\tilde{x}_k - (a, e_k)|}{\sqrt{\lambda_k}} \right) \mathrm{d}\tilde{x}_k \right) \\ &= \frac{1}{\sqrt{2^d \det Q}} \int_R \exp\left(-\sum_{k=1}^{n} \frac{\sqrt{2}|\tilde{x}_k - (a, e_k)|}{\sqrt{\lambda_k}} \right) \mathrm{d}\tilde{x} \\ &= \frac{1}{\sqrt{2^d \det Q}} \int_{\gamma(R)} \exp\left(-\sum_{k=1}^{n} \frac{\sqrt{2}|(x, e_k) - (a, e_k)|}{\sqrt{\lambda_k}} \right) \mathrm{d}x. \qquad \square\end{aligned}$$

Again we call an $\mathbb{R}^d$ valued random variable *Laplacian*, if its probability distribution is a Laplacian measure $\mathcal{L}_{a,Q}$ on $\mathbb{R}^d$. We can write such a Laplacian random variable ξ as

$$\xi = \sum_{k=1}^{d} \xi_k e_k,$$

where $\xi_k \sim \mathcal{L}_{a_k,\lambda_k}$ are independent, real valued Laplacian random variables.

3.4 Laplacian Infinite Product Measures

Now we turn towards the case of an infinite-dimensional separable real Hilbert space H with norm $\|\cdot\|$ and inner product $(\cdot, \cdot)$. Let $a \in H$ and let $Q \in L(H)$ be a self-adjoint positive trace class operator. We want to define a Laplacian product measure on H that has mean a and covariance operator Q.

Since Q is trace class and self-adjoint, an orthonormal basis $(e_k)_{k\in\mathbb{N}}$ of H consisting of eigenvectors of Q exists. Let $(\lambda_k)_{k\in\mathbb{N}}$ be the associated nonnegative eigenvalues in descending order. Then we have

$$Qe_k = \lambda_k e_k \quad \text{for all } k \in \mathbb{N}.$$

We can identify H with the space ℓ^2 of all sequences $(x_k)_{k\in\mathbb{N}}$ of real numbers with $\sum_{k=1}^{\infty} |x_k|^2 < \infty$ via the natural isomorphism $\gamma\colon \ell^2 \to H$,

$$(x_k)_{k\in\mathbb{N}} \mapsto \sum_{k=1}^{\infty} x_k e_k. \tag{3.1}$$

We will first define a Laplacian product measure on the space $\mathbb{R}^\infty := \prod_{k=1}^{\infty} \mathbb{R}$ of all sequences of real numbers and then restrict it to ℓ^2. A subset of $\mathbb{R}^\infty$ is called *cylindrical* if it is of the form

$$I_{k_1,\dots,k_n;A} := \{(x_j)_j \in \mathbb{R}^\infty : (x_{k_1}, \dots, x_{k_n}) \in A\},$$

where $n, k_1 < \dots < k_n$ are positive integers and $A \in \mathcal{B}(\mathbb{R}^n)$. Let C denote the set of all cylindrical subsets of $\mathbb{R}^\infty$. By [Klenke 2014, Remark 14.10], the Borel σ-algebra generated by C coincides with $\prod_{k=1}^{\infty} \mathcal{B}(\mathbb{R})$, which in turn is equal to the Borel σ-algebra $\mathcal{B}(\mathbb{R}^\infty)$ induced by the product topology on $\mathbb{R}^\infty$ by [Klenke 2014, Thm. 14.8].

Now let $a_k := (a, e_k)$ for all $k \in \mathbb{N}$ and $\mu_k := \mathcal{L}_{a_k,\lambda_k}$. We define the product measure

$$\mu := \bigotimes_{k=1}^{\infty} \mu_k = \bigotimes_{k=1}^{\infty} \mathcal{L}_{a_k,\lambda_k}$$

for all $I_{k_1,\dots,k_n;A} \in C$ by

$$\mu(I_{k_1,\dots,k_n;A}) = (\mu_{k_1} \times \dots \times \mu_{k_n})(A).$$

Then, by [Da Prato 2001, Thm. 1.3.3], the function μ is σ-additive on C and has a unique extension to a probability measure on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$.

Proposition 3.4. *The measure μ is concentrated on ℓ^2, i.e., $\mu(\ell^2) = 1$.*

Proof. We have

$$\begin{aligned}\int_{\mathbb{R}} x_k^2 \mathcal{L}_{a_k,\lambda_k}(dx_k) &= \int_{\mathbb{R}} (x_k - a_k)^2 - a_k^2 + 2x_k a_k \, \mathcal{L}_{a_k,\lambda_k}(dx_k) \\ &= \lambda_k - a_k^2 + 2a_k^2 = \lambda_k + a_k^2\end{aligned}$$

for all $k \in \mathbb{N}$. Note that $\mu_k = \mathcal{L}_{a_k,\lambda_k}$ is the pushforward measure $\mu \circ p_k^{-1}$ of μ under the projection $p_k \colon \mathbb{R}^\infty \to \mathbb{R}$, $x \mapsto x_k$, since

$$(\mu \circ p_k^{-1})(A) = \mu(I_{k;A}) = \mathcal{L}_{a_k,\lambda_k}(A)$$

for every Borel measurable set $A \subseteq \mathbb{R}$. Together with the monotone convergence theorem this yields

$$\begin{aligned}\int_{\mathbb{R}^\infty} \sum_{k=1}^{\infty} x_k^2 \, \mu(dx) &= \lim_{n\to\infty} \int_{\mathbb{R}^\infty} \sum_{k=1}^{n} x_k^2 \, \mu(dx) = \lim_{n\to\infty} \sum_{k=1}^{n} \int_{I_{k;\mathbb{R}}} x_k^2 \, \mu(dx) \\ &= \sum_{k=1}^{\infty} \int_{\mathbb{R}} x_k^2 \mathcal{L}_{a_k,\lambda_k}(dx_k) = \sum_{k=1}^{\infty} \left(\lambda_k + a_k^2\right) = \operatorname{Tr} Q + \|a\|^2 < \infty.\end{aligned}$$

This implies that $\|\cdot\|_{\ell^2}$ is μ-almost surely finite on $\mathbb{R}^\infty$. In other words, $\mu(\mathbb{R}^\infty \setminus \ell^2) = 0$, and therefore $\mu(\ell^2) = 1$. $\square$

Proposition 3.5. *Let $\mathbb{R}^\infty$ be equipped with the product topology and let ℓ^2 be equipped with the $\|\cdot\|_2$ norm topology. Then $\mathcal{B}(\mathbb{R}^\infty) \cap \ell^2 = \mathcal{B}(\ell^2)$.*

Proof. We know that $\mathcal{B}(\mathbb{R}^\infty)$ is generated by the cylindrical sets in $\mathbb{R}^\infty$ and that $\mathcal{B}(\ell^2)$ is generated by open balls in ℓ^2.

To prove the inclusion $\mathcal{B}(\mathbb{R}^\infty) \cap \ell^2 \subseteq \mathcal{B}(\ell^2)$, consider the intersection $I_{1,\dots,n;M} \cap \ell^2$ of an arbitrary cylindrical subset $I_{1,\dots,n;M}$ of $\mathbb{R}^\infty$ with $n \in \mathbb{N}$ and $M \in \mathcal{B}(\mathbb{R}^n)$ and the space ℓ^2. The projection $P_n \colon \ell^2 \to \mathbb{R}^n$, $x \mapsto (x_1, \dots, x_n)^T$ is continuous and hence measurable, so that

$$I_{1,\dots,n;M} \cap \ell^2 = P_n^{-1}(M) \in \mathcal{B}(\ell^2).$$

To prove the reverse inclusion, consider an open ball

$$B_{\ell^2}(x,r) := \left\{ y \in \ell^2 : \sum_{k=1}^{\infty} |y_k - x_k|^2 < r^2 \right\}$$

in ℓ^2 with radius $r > 0$ around $x \in \ell^2$. It can be written as the union $B_{\ell^2}(x,r) = \bigcup_{n\in\mathbb{N}} U_{\ell^2}(x, r-\frac{1}{n})$ of closed balls

$$U_{\ell^2}\left(x, r - \frac{1}{n}\right) := \left\{ y \in \ell^2 : \sum_{k=1}^{\infty} |y_k - x_k|^2 \le \left(r - \frac{1}{n}\right)^2 \right\}.$$

Each closed ball, in turn, can be written as the intersection $U_{\ell^2}(x,s) = \bigcap_{m\in\mathbb{N}} U_m$ of closed sets

$$U_m := \left\{ y \in \ell^2 : \sum_{k=1}^{m} |y_k - x_k|^2 \le s^2 \right\}.$$

U_m is closed, since for every $y \in \ell^2 \setminus U_m$, the open ball $B_{\ell^2}(y,t)$ with radius

$$t := \left(\sum_{k=1}^{m} |y_k - x_k|^2 \right)^{\frac{1}{2}} - s$$

is contained in the complement of U_m. Furthermore, each $U_m = A_m \cap \ell^2$ is the intersection of a cylindrical subset

$$A_m := I_{1,\dots,m;U_{\mathbb{R}^m}((x_1,\dots,x_m)^T,s)} = \left\{ y \in \mathbb{R}^\infty : \sum_{k=1}^{m} |y_k - x_k|^2 \le s^2 \right\}$$

of $\mathbb{R}^\infty$ and the space ℓ^2, so that $U_m \in \mathcal{B}(\mathbb{R}^\infty) \cap \ell^2$. This implies that

$$U_{\ell^2}(x,s) = \bigcap_{m\in\mathbb{N}} U_m \in \mathcal{B}(\mathbb{R}^\infty) \cap \ell^2$$

and hence also

$$B_{\ell^2}(x,r) = \bigcup_{n\in\mathbb{N}} U_{\ell^2}\left(x, r - \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R}^\infty) \cap \ell^2. \qquad \square$$

Propositions 3.4 and 3.5 show that the restriction of μ to $(\ell^2, \mathcal{B}(\ell^2))$ is a probability measure. In a last step, we define the Laplacian measure $\mathcal{L}_{a,Q}$ on H as the pushforward

$$\mathcal{L}_{a,Q} := \mu \circ \gamma^{-1} = \left(\bigotimes_{k=1}^{\infty} \mathcal{L}_{a_k,\lambda_k} \right) \circ \gamma^{-1}$$

of μ under the natural isomorphism γ between ℓ^2 and H defined by (3.1). Here we can also state the inverse γ^{-1} explicitly as $x \mapsto ((x, e_k))_{k\in\mathbb{N}}$.

Remark 3.6. Please be aware that, as in the finite dimensional case, the definition of $\mathcal{L}_{a,Q}$ does depend on the specific choice of the basis $\{e_k\}_{k\in\mathbb{N}}$, even though we do not include it into the notation.

Proposition 3.7. *$\mathcal{L}_{a,Q}$ has mean a and covariance operator Q, i.e.,*

$$\int_H (x,y)\mathcal{L}_{a,Q}(dx) = (a,y) \quad \text{for all } y \in H,$$

$$\int_H (y,x-a)(z,x-a)\mathcal{L}_{a,Q}(dx) = (Qy,z) \quad \text{for all } y,z \in H.$$

Furthermore, its characteristic function is given by

$$\widehat{\mathcal{L}_{a,Q}}(h) = \exp\left(i(a,h)\prod_{k=1}^{\infty}\frac{1}{1+\frac{1}{2}\lambda_k(h,e_k)^2}\right) \quad \text{for all } h \in H.$$

Proof. For all $x \in H$ and $n \in \mathbb{N}$ we define

$$P_n x := \sum_{k=1}^{n}(x,e_k)e_k.$$

Then $P_n x \longrightarrow x$ for all $x \in H$,

$$|(P_n x, y)| \leq \|P_n x\|\|y\| \leq \|x\|\|y\|$$

for all $x, y \in H$ and $x \mapsto \|x\|\|y\|$ is $\mathcal{L}_{a,Q}$-integrable, since

$$\left(\int_H \|x\|\mathcal{L}_{a,Q}(dx)\right)^2 \leq \int_H \|x\|^2\mathcal{L}_{a,Q}(dx) = \operatorname{Tr} Q + \|a\|^2 < \infty.$$

This allows us to use Lebesgue's dominated convergence theorem to obtain

$$\int_H (x,y)\mathcal{L}_{a,Q}(dx) = \lim_{n\to\infty}\int_H (P_n x, y)\mathcal{L}_{a,Q}(dx)$$

for all $y \in H$. Furthermore, we compute

$$\begin{aligned}\int_H (P_n x,y)\mathcal{L}_{a,Q}(dx) &= \int_H \sum_{k=1}^{n}(x,e_k)(e_k,y)\mathcal{L}_{a,Q}(dx) = \sum_{k=1}^{n}(e_k,y)\int_{\mathbb{R}} \tilde{x}\mathcal{L}_{(a,e_k),\lambda_k}(d\tilde{x}) \\ &= \sum_{k=1}^{n}(a,e_k)(e_k,y) = (P_n a, y)\end{aligned}$$

using Proposition 3.1. Now passing on to limits on both sides yields the first proposition.

Similarly,

$$|(P_n(x-a),y)(P_n(x-a),z)| \leq \|x-a\|^2\|y\|\|z\|$$

for all $x, y, z \in H$ and $\|x-a\|^2$ is $\mathcal{L}_{a,Q}$-integrable with

$$\begin{aligned}\int_H \|x-a\|^2\mathcal{L}_{a,Q}(dx) &= \int_H \sum_{k=1}^{\infty} |(x-a,e_k)|^2\,\mathcal{L}_{a,Q}(dx) \\ &= \sum_{k=1}^{\infty}\int_{\mathbb{R}} |\tilde{x}-(a,e_k)|^2\,\mathcal{L}_{(a,e_k),\lambda_k}(d\tilde{x}) = \sum_{k=1}^{\infty}\lambda_k = \operatorname{Tr} Q\end{aligned}$$

by the monotone convergence theorem and Proposition 3.1, so that

$$\int_H (x-a,y)(x-a,z)\mathcal{L}_{a,Q}(dx) = \lim_{n\to\infty}\int_H (P_n(x-a),y)(P_n(x-a),z)\mathcal{L}_{a,Q}(dx)$$

for all $y, z \in H$ by Lebesgue's dominated convergence theorem. If $j \neq k$, we observe that

$$\begin{aligned}
&\int_H (x-a,e_j)(x-a,e_k)\mathcal{L}_{a,Q}(dx)\\
&= \int_{I_{j,k;\mathbb{R}^2}} \big((x,e_j)-(a,e_j)\big)\big((x,e_k)-(a,e_k)\big)\mathcal{L}_{a,Q}(dx)\\
&= \int_{\mathbb{R}^2} \big(\tilde{x}_1-(a,e_j)\big)\big(\tilde{x}_2-(a,e_k)\big)\mathcal{L}_{(a,e_j),\lambda_j}(d\tilde{x}_1)\mathcal{L}_{(a,e_k),\lambda_k}(d\tilde{x}_2)\\
&= \left(\int_{\mathbb{R}} \tilde{x}_1\mathcal{L}_{(a,e_j),\lambda_j}(d\tilde{x}_1)-(a,e_j)\right)\left(\int_{\mathbb{R}} \tilde{x}_2\mathcal{L}_{(a,e_k),\lambda_k}(d\tilde{x}_2)-(a,e_k)\right) = 0
\end{aligned}$$

by Proposition 3.1, whereas

$$\begin{aligned}
\int_H (x-a,e_k)^2\mathcal{L}_{a,Q}(\mathrm{d}x) &= \int_{I_{k;\mathbb{R}}} (x-a,e_k)^2\mathcal{L}_{a,Q}(\mathrm{d}x)\\
&= \int_{\mathbb{R}} (\tilde{x}-(a,e_k))^2\mathcal{L}_{(a,e_k),\lambda_k}(\mathrm{d}\tilde{x}) = \lambda_k.
\end{aligned}$$

This leads to

$$\begin{aligned}
&\int_H (P_n(x-a),y)(P_n(x-a),z)\mathcal{L}_{a,Q}(dx)\\
&= \sum_{k=1}^n\sum_{j=1}^n (e_j,y)(e_k,z)\int_H (x-a,e_j)(x-a,e_k)\mathcal{L}_{a,Q}(dx)\\
&= \sum_{k=1}^n (y,e_k)(e_k,z)\lambda_k = \sum_{k=1}^n (y,Qe_k)(e_k,z)\\
&= \sum_{k=1}^n (Qy,e_k)(e_k,z) = (P_nQy,z).
\end{aligned}$$

Letting n tend to infinity on both sides proves the second proposition.

Finally, we consider the characteristic function of $\mathcal{L}_{a,Q}$. As $|e^{i(h,P_nx)}| \leq 1$ for all $x, h \in H$ and $n \in \mathbb{N}$ and $\int_H 1\mathcal{L}_{a,Q}(dx) = 1$, it follows from Lebesgue's dominated convergence theorem that

$$\widehat{\mathcal{L}_{a,Q}}(h) = \int_H e^{i(h,x)}\mathcal{L}_{a,Q}(dx) = \lim_{n\to\infty}\int_H e^{i(h,P_nx)}\mathcal{L}_{a,Q}(dx)$$

for all $h \in H$, where

$$\begin{aligned}\int_H e^{i(h,P_n x)} \mathcal{L}_{a,Q}(dx) &= \int_H e^{i\sum_{k=1}^n (x,e_k)(e_k,h)} \mathcal{L}_{a,Q}(dx) = \int_H \prod_{k=1}^n e^{i(x,e_k)(e_k,h)} \mathcal{L}_{a,Q}(dx) \\ &= \prod_{k=1}^n \int_{\mathbb{R}} e^{i\tilde{x}(e_k,h)} \mathcal{L}_{(a,e_k),\lambda_k}(d\tilde{x}) = \prod_{k=1}^n \frac{e^{i(a,e_k)(e_k,h)}}{1+\frac{1}{2}\lambda_k(e_k,h)^2} \\ &= e^{i(P_n a,h)} \prod_{k=1}^n \frac{1}{1+\frac{1}{2}\lambda_k(e_k,h)^2}\end{aligned}$$

by Proposition 3.1. Forming limits on both sides yields

$$\widehat{\mathcal{L}_{a,Q}}(h) = e^{i(a,h)} \prod_{k=1}^\infty \frac{1}{1+\frac{1}{2}\lambda_k(h,e_k)^2} \quad \text{for all } h \in H. \qquad \square$$

Similar to the finite-dimensional case we call an H valued random variable *Laplacian*, if its probability distribution is a Laplacian measure $\mathcal{L}_{a,Q}$ on H. We can express such a Laplacian random variable ξ as a series

$$\xi = \sum_{k=1}^\infty \xi_k e_k,$$

where $\xi_k \sim \mathcal{L}_{a_k,\lambda_k}$ are independent, real valued Laplacian random variables.

Since there is no Lebesgue measure on infinite-dimensional Banach spaces, we cannot state a density in the same way as in the finite-dimensional case. We will, however, be able to state the density of $\mathcal{L}_{a,Q}$ with respect to $\mathcal{L}_Q$.

3.5 Admissible Shifts

Now we address the question for which $a \in H$ the measure $\mathcal{L}_{a,Q}$ with mean a is absolutely continuous with respect to the centred measure $\mathcal{L}_Q$.

If H is finite-dimensional, then $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$ are equivalent for every $a \in H$ and we can state the density

$$\begin{aligned}\frac{d\mathcal{L}_{a,Q}}{d\mathcal{L}_Q} &= \frac{\exp(-\sqrt{2}\sum_{k=1}^d |(Q^{-\frac{1}{2}}(x-a),e_k)|)}{\exp(-\sqrt{2}\sum_{k=1}^d |(Q^{-\frac{1}{2}}x,e_k)|)} \\ &= \exp\left(-\sqrt{2}\sum_{k=1}^d \frac{|(x-a,e_k)|-|(x,e_k)|}{\sqrt{\lambda_k}}\right).\end{aligned}$$

directly as the quotient of the densities of $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$. If, on the other hand, H is infinite dimensional, this does not have to be the case for all $a \in H$, as we have seen for the Gaussian measure in Section 2.3.

We will derive a criterion similar to Theorem 2.4 and show that, as in the Gaussian case, $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$ are equivalent if $a \in \mathcal{R}(Q^{1/2})$ and singular otherwise. We will do this using a result by Kakutani characterising the equivalence of infinite product measures using Hellinger integrals.

Let μ and ν be probability measures on $(H, \mathcal{B}(H))$. Then the *Hellinger integral* of μ and ν is defined by

$$H(\mu, \nu) = \int_H \sqrt{\frac{d\mu}{d\zeta}\frac{d\nu}{d\zeta}} d\zeta,$$

where ζ is a probability measure on $(H, \mathcal{B}(H))$, such that both μ and ν are absolutely continuous with respect to ζ. Such a reference measure is always given by $\zeta = \frac{1}{2}(\mu + \nu)$ and the integral does not depend on the choice of ζ. If μ and ν are equivalent, then we have

$$\frac{d\nu}{d\zeta} = \frac{d\nu}{d\mu}\frac{d\mu}{d\zeta}$$

and can therefore express the Hellinger integral of μ and ν as

$$H(\mu, \nu) = \int_H \sqrt{\frac{d\nu}{d\mu}\frac{d\mu}{d\zeta}} d\zeta = \int_H \sqrt{\frac{d\nu}{d\mu}} d\mu,$$

without the need for a reference measure ζ.

We take a look at the Hellinger-integral of Laplacian measures on $\mathbb{R}$.

Example 3.8. For $a \in \mathbb{R}$ and $\lambda > 0$, we have

$$\frac{d\mathcal{L}_{a,\lambda}}{d\mathcal{L}_\lambda}(x) = e^{-\sqrt{2}\frac{|x-a|-|x|}{\sqrt{\lambda}}},$$

and a straightforward computation yields

$$H(\mathcal{L}_\lambda, \mathcal{L}_{a,\lambda}) = \int_{\mathbb{R}} e^{-\frac{|x-a|-|x|}{\sqrt{2\lambda}}} \mathcal{L}_\lambda(dx) = \left(1 + \frac{|a|}{\sqrt{2\lambda}}\right) e^{-\frac{|a|}{\sqrt{2\lambda}}} > 0.$$

The following result by Kakutani allows us to use the Hellinger integral to draw conclusions about the equivalence of infinite product measures. Its proof can be found in [Kakutani 1948].

Theorem 3.9 (Kakutani). *Let $(\mu_k)_{k\in\mathbb{N}}$ and $(\nu_k)_{k\in\mathbb{N}}$ be sequences of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where μ_k and ν_k are equivalent for all $k \in \mathbb{N}$, and define the product measures*

$$\mu = \bigotimes_{k=1}^\infty \mu_k, \qquad \nu = \bigotimes_{k=1}^\infty \nu_k$$

on $\mathbb{R}^\infty$. Then,

$$H(\mu, \nu) = \prod_{k=1}^\infty H(\mu_k, \nu_k).$$

Moreover, the following alternative holds.

(i) If $H(\mu, \nu) > 0$, then μ and ν are equivalent and the Radon–Nikodym derivative is given by

$$\frac{d\nu}{d\mu}(x) = \prod_{k=1}^\infty \frac{d\nu_k}{d\mu_k}(x_k)$$

μ-almost everywhere.

(ii) *If $H(\mu, \nu) = 0$, then μ and ν are singular.*

We apply Kakutani's theorem to Laplacian measures on H. Note that the range of $Q^{1/2}$ is given by

$$\mathcal{R}(Q^{\frac{1}{2}}) = \left\{ a \in H : \sum_{k=1}^{\infty} \frac{a_k^2}{\lambda_j} < \infty \right\}.$$

Theorem 3.10. (i) *If $a \notin \mathcal{R}(Q^{\frac{1}{2}})$, then $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$ are singular.*

(ii) *If $a \in \mathcal{R}(Q^{\frac{1}{2}})$, then $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$ are equivalent and the density $\frac{d\mathcal{L}_{a,Q}}{d\mathcal{L}_Q}$ is given by*

$$\begin{aligned}
\frac{d\mathcal{L}_{a,Q}}{d\mathcal{L}_Q}(x) &= \exp\left(-\sqrt{2}\sum_{k=1}^{\infty}\left(\left|(Q^{-\frac{1}{2}}(x-a), e_k)\right| - \left|(Q^{-\frac{1}{2}}x, e_k)\right|\right)\right) \\
&= \exp\left(-\sqrt{2}\sum_{k=1}^{\infty}\frac{|x_k - a_k| - |x_k|}{\sqrt{\lambda_k}}\right)
\end{aligned}$$

$\mathcal{L}_Q$-almost everywhere, where $x_k := (x, e_k)$ and $a_k := (a, e_k)$ for all $k \in \mathbb{N}$.

Proof. We apply Theorem 3.9 with $\nu_k := \mathcal{L}_{a_k,\lambda_k}$ and $\mu_k := \mathcal{L}_{\lambda_k}$ for all $k \in \mathbb{N}$. Then $\mathcal{L}_{a,Q} = \nu \circ \gamma^{-1}$ and $\mathcal{L}_Q = \mu \circ \gamma^{-1}$, where $\gamma(x) := \sum_{k=1}^{\infty} x_k e_k$ for all $x \in \ell^2$. From Example 3.8 we know that

$$H(\mu, \nu) = \prod_{k=1}^{\infty} H(\mu_k, \nu_k) = \prod_{k=1}^{\infty}\left(\left(1 + \frac{|a_k|}{\sqrt{2\lambda_k}}\right) e^{-\frac{|a_k|}{\sqrt{2\lambda_k}}}\right).$$

We consider

$$-\ln H(\mu, \nu) = \sum_{k=1}^{\infty}\left(\frac{|a_k|}{\sqrt{2\lambda_k}} - \ln\left(1 + \frac{|a_k|}{\sqrt{2\lambda_k}}\right)\right) \tag{3.2}$$

and note that $H(\mu, \nu) > 0$ if and only if this series converges. By the alternating series test, the first partial sum $S_1 = t$ of the Mercator series

$$\ln(1+t) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{t^k}{k} \quad \forall t \in (-1, 1]$$

satisfies the error bound

$$t - \ln(1+t) \le \frac{t^2}{2} \quad \forall t \in [0, 1).$$

Also, we can show that

$$t - \ln(1+t) \ge \frac{t^2}{2} - \frac{t^3}{3} \ge \frac{t^2}{6} \quad \forall t \in [0, 1).$$

To this end, we note that we have equality for $t = 0$ and that the derivative computes as

$$\left(t - \ln(1+t) - \frac{t^2}{2} + \frac{t^3}{3}\right)' = \frac{t^3}{1+t} \ge 0 \quad \forall t \in [0, 1).$$

From these two estimates we obtain

$$\frac{a_k^2}{12\lambda_k} \leq \frac{|a_k|}{\sqrt{2\lambda_k}} - \ln\left(1 + \frac{|a_k|}{\sqrt{2\lambda_k}}\right) \leq \frac{a_k^2}{4\lambda_k} \tag{3.3}$$

for all $k \in \mathbb{N}$.

For $a \in \mathcal{R}(Q^{1/2})$ we have $\sum_{k=1}^{\infty} \frac{a_k^2}{\lambda_j} < \infty$ and thus in particular $\frac{a_k^2}{2\lambda_k} \longrightarrow 0$ as $k \longrightarrow \infty$. Now we can choose $N \in \mathbb{N}$ such that $\frac{|a_k|}{\sqrt{2\lambda_k}} < 1$ for all $k \geq N$. This yields

$$-\ln H(\mu, \nu) \leq \sum_{k=1}^{N-1} \left(\frac{|a_k|}{\sqrt{2\lambda_k}} - \ln\left(1 + \frac{|a_k|}{\sqrt{2\lambda_k}}\right)\right) + \frac{1}{4} \sum_{k=N}^{\infty} \frac{a_k^2}{\lambda_k} < \infty,$$

which implies $H(\mu, \nu) > 0$. Now Kakutani's theorem guarantees equivalence of μ and ν in case that $a \in \mathcal{R}(Q^{1/2})$ and, together with Example 3.8, states that the density of ν with respect to μ is given by

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) = \prod_{k=1}^{\infty} \frac{\mathrm{d}\nu_k}{\mathrm{d}\mu_k}(x_k) = \prod_{k=1}^{\infty} e^{-\sqrt{2}\frac{|x_k - a_k| - |x_k|}{\sqrt{\lambda_k}}} = e^{-\sqrt{2}\sum_{k=1}^{\infty} \frac{|x_k - a_k| - |x_k|}{\sqrt{\lambda_k}}}.$$

If, on the other hand, we assume that $-\ln H(\mu, \nu) < \infty$ for $a \notin \mathcal{R}(Q^{\frac{1}{2}})$, we have

$$\frac{|a_k|}{\sqrt{2\lambda_k}} - \ln\left(1 + \frac{|a_k|}{\sqrt{2\lambda_k}}\right) \longrightarrow 0$$

as $k \longrightarrow \infty$ by (3.2). Using (3.3), we obtain that $\frac{a_k^2}{12\lambda_k} \longrightarrow 0$ as well, which again allows us to choose $N \in \mathbb{N}$ such that $\frac{|a_k|}{\sqrt{2\lambda_k}} < 1$ for all $k \geq N$. However, since $a \notin \mathcal{R}(Q^{\frac{1}{2}})$ we have

$$-\ln H(\mu, \nu) \geq \sum_{k=1}^{N-1} \left(\frac{|a_k|}{\sqrt{2\lambda_k}} - \ln\left(1 + \frac{|a_k|}{\sqrt{2\lambda_k}}\right)\right) + \frac{1}{12} \sum_{k=N}^{\infty} \frac{a_k^2}{\lambda_k} = \infty,$$

which is a contradiction. So $H(\mu, \nu) = 0$, which implies that μ and ν are singular.

By Proposition 3.4, $\mathbb{R}^{\infty} \setminus \ell^2$ is a null set both under μ and ν. Therefore the equivalence or singularity of μ and ν on $\mathbb{R}^{\infty}$ transfers to their restrictions to ℓ^2. In a last step, we show that it further transfers to $\mathcal{L}_Q$ and $\mathcal{L}_{a,Q}$. Assume that $\nu \ll \mu$. Then for every $A \in \mathcal{B}(H)$ we have

$$\begin{aligned}\mathcal{L}_{a,Q}(A) = (\nu \circ \gamma^{-1})(A) &= \int_{\gamma^{-1}(A)} \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\mu(\mathrm{d}x) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\big(\gamma^{-1}(\tilde{x})\big)\mathcal{L}_Q(\mathrm{d}\tilde{x}) \\ &= \int_A \exp\left(-\sqrt{2}\sum_{k=1}^{\infty} \frac{|(\tilde{x}, e_k) - (a, e_k)| - |(\tilde{x}, e_k)|}{\sqrt{\lambda_k}}\right)\mathcal{L}_Q(\mathrm{d}\tilde{x}).\end{aligned}$$

This shows that $\mathcal{L}_{a,Q}$ is absolutely continuous with respect to $\mathcal{L}_Q$ and has the stated density. A similar computation shows that $\mu \ll \nu$ implies $\mathcal{L}_Q \ll \mathcal{L}_{a,Q}$.

Finally, assume that μ and ν are singular and let $A \in \mathcal{B}(\ell^2)$ such that $\mu(A) = 0$ and $\nu(\ell^2 \setminus A) = 0$. Then $\mathcal{L}_Q$ and $\mathcal{L}_{a,Q}$ are singular as well, because

$$\mathcal{L}_Q(\gamma(A)) = (\mu \circ \gamma^{-1})(\gamma(A)) = \mu(A) = 0$$

and
$$\mathcal{L}_{a,Q}(H \setminus \gamma(A)) = \mathcal{L}_{a,Q}(\gamma(\ell^2) \setminus \gamma(A)) = \mathcal{L}_{a,Q}(\gamma(\ell^2 \setminus A)) = \nu(\ell^2 \setminus A) = 0$$
by the surjectivity of γ. □

Now we show that the space of admissible shifts is a null set under $\mathcal{L}_Q$.

Lemma 3.11. *We have* $\mathcal{L}_Q(\mathcal{R}(Q^{\frac{1}{2}})) = 0$.

Proof. For any $n, k \in \mathbb{N}$ we define
$$U_n = \left\{ y \in H : \sum_{j=1}^{\infty} \frac{y_j^2}{\lambda_j} < n^2 \right\},$$
and
$$U_{n,k} = \left\{ y \in H : \sum_{j=1}^{2k} \frac{y_j^2}{\lambda_j} < n^2 \right\}.$$
Then $\mathcal{R}(Q^{\frac{1}{2}}) = \bigcup_{n \in \mathbb{N}} U_n$ and the sets U_n are ascending, which yields
$$\mathcal{L}_Q(\mathcal{R}(Q^{\frac{1}{2}})) = \lim_{n \to \infty} \mathcal{L}_Q(U_n).$$
Furthermore, $U_n = \bigcap_{k \in \mathbb{N}} U_{n,k}$ holds for all $n \in \mathbb{N}$ and the sets $U_{n,k}$ are descending in k. So it is enough to show that
$$\mathcal{L}_Q(U_n) = \lim_{k \to \infty} \mathcal{L}_Q(U_{n,k}) = 0.$$
We substitute and estimate,
$$\begin{aligned}
\mathcal{L}_Q(U_{n,k}) &= \int_{\left\{ y \in \mathbb{R}^{2k} : \sum_{j=1}^{2k} \frac{y_j^2}{\lambda_j} < n^2 \right\}} \prod_{j=1}^{2k} \mathcal{L}_{\lambda_j}(dy_j) = \int_{\left\{ y \in \mathbb{R}^{2k} : \sum_{j=1}^{2k} y_j^2 < n^2 \right\}} \prod_{j=1}^{2k} \mathcal{L}_1(dy_j) \\
&= \int_{\left\{ y \in \mathbb{R}^{2k} : \|y\|_2 < n \right\}} \mathcal{L}_{\mathrm{Id}_{2k}}(dy) = \int_{\left\{ y \in \mathbb{R}^{2k} : \|y\|_2 < n \right\}} \frac{1}{\sqrt{2^{2k}}} e^{-\sqrt{2} \sum_{j=1}^{2k} |y_j|} dy \\
&\leq \frac{1}{2^k} \int_{\left\{ y \in \mathbb{R}^{2k} : \|y\|_2 < n \right\}} dy
\end{aligned}$$
where Id_{2k} denotes the identity on $\mathbb{R}^{2k}$. This implies
$$\mathcal{L}_Q(U_{n,k}) \leq \frac{1}{k!} \left(\frac{\pi n^2}{2} \right)^k$$
because the Lebesgue measure of the Euclidean $2k$-ball with radius n is $n^{2k} \frac{\pi^k}{k!}$, see, e.g., Example 8.7.11 in [Benedetto 2009]. It follows, that $\lim_{k \to \infty} U_{n,k} = 0$. □

4 Variational Characterisation of MAP Estimates

We return to the setting described in Chapter 1. Let X and Y be separable Banach spaces, each equipped with its Borel σ-algebra. We assume that for $y \in Y$ the posterior distribution μ^y is absolutely continuous with respect to the prior distribution μ_0 and that its density is given by

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u, y))}{\int_X \exp(-\Phi(\tilde{u}, y))\mathrm{d}\mu_0(\tilde{u})} \quad \mu_0\text{-almost surely,} \tag{4.1}$$

where $\Phi\colon X \times Y \to \mathbb{R}$ is a measurable function. Throughout this chapter, we will consider the posterior for a fixed value $y \in Y$ and regard $\Phi(u) = \Phi(u, y)$ as a function of u exclusively.

4.1 Maximum A Posteriori Estimates

A common way to define an estimator for the posterior is by considering modes of the posterior distribution, i.e., points that maximise the posterior probability in an appropriate sense. For a separable Banach space X, the following definition of a MAP estimate has been introduced in [Dashti, Law, et al. 2013]. Let $B_\varepsilon(x) \subset X$ denote the open ball with radius ε centred at $x \in X$.

Definition 4.1 ([Dashti, Law, et al. 2013, Def. 3.1])**.** Let μ be a probability measure on X. A point $\hat{u} \in X$ is called *mode* of μ, if it satisfies

$$\lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(\hat{u}))}{\sup_{u \in X} \mu(B_\varepsilon(u))} = 1. \tag{4.2}$$

A mode of the posterior distribution μ^y is called *maximum a posteriori (MAP) estimate*.

If the space X is finite-dimensional and the prior distribution μ_0 has a density with respect to the Lebesgue measure, we can express this density in the form $\exp(-R(\cdot))$ with a function $R\colon X \to \mathbb{R} \cup \{\infty\}$. Consequently, the Lebesgue density of the posterior distribution μ^y is given by

$$u \mapsto \exp(-\Phi(u) - R(u)).$$

In a finite-dimensional setting, MAP estimates are usually defined directly as maximisers of the posterior density or, equivalently, as minimisers of

$$u \mapsto \Phi(u) + R(u). \tag{4.3}$$

If both Φ and R are continuous, then this definition coincides with Definition 4.1. Also note that Definition 4.1 is a global definition in the sense that it excludes points which maximize the posterior density only locally and not globally.

In an infinite-dimensional setting, in contrast, the posterior distribution does not have a canonical density due to the lack of a Lebesgue measure. We only have its density with respect to the prior distribution. This gives rise to the question how to define a generalised posterior density in a canonical way, such that its maximisers are precisely the modes of μ^y according to Definition 4.1. A natural candidate for a generalised logarithmic density is the Onsager–Machlup functional, which in our context is defined by the asymptotic probability of small balls around two points.

Definition 4.2. Let μ be a probability measure on X. Let $E \subset X$ denote the set of all admissible shifts for μ that yield an equivalent measure, i.e., all $h \in X$ for which the shifted measure

$$\mu_h := \mu(\cdot - h)$$

is equivalent with μ. A functional $I\colon E \to \mathbb{R}$ is called *Onsager–Machlup functional* of μ, if

$$\lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(h_1))}{\mu(B_\varepsilon(h_2))} = \exp\left(I(h_2) - I(h_1)\right) \quad \text{for all } h_1, h_2 \in E.$$

Note that this property is only required to hold for all points from a subset of X, whereas, in general, the limit does not exist for all $h_1, h_2 \in X$.

4.2 Bounded Potential

It was shown in [Dashti, Law, et al. 2013] that in case of a Gaussian prior and under certain conditions on the potential Φ the Onsager–Machlup functional of the posterior distribution μ^y is indeed of the form (4.3) and that MAP estimates can be characterised as its minimisers. We briefly recapitulate these results here for a separable Hilbert space X, even though they are valid for any separable Banach space.

Let the prior follow a centred Gaussian distribution $\mu_0 = \mathcal{N}_Q$ with covariance operator $Q \in \mathcal{L}(X)$. We assume that Q is self-adjoint, positive definite and trace class. Then, by Theorem 2.4, the space of admissible shifts is given by the Cameron–Martin space $E := \mathcal{R}(Q^{\frac{1}{2}})$ of μ_0, which we equip with the norm $\|h\|_E := \|Q^{-\frac{1}{2}} h\|_X$.

We make the following assumptions on the potential Φ.

Assumption 4.3. (i) For every $\varepsilon > 0$, there is an $M = M(\varepsilon) \in \mathbb{R}$, such that for all $u \in X$,

$$\Phi(u) \geq M - \varepsilon \|u\|_X^2 .$$

(ii) Φ is bounded from above on bounded sets, i.e., for every $r > 0$, there exists $K = K(r) > 0$, such that for all $u \in B_r(0)$, we have

$$\Phi(u) \leq K.$$

(iii) Φ is Lipschitz continuous on bounded sets, i.e., for every $r > 0$, there exists $L = L(r) > 0$, such that for all $u, v \in B_r(0)$, we have

$$|\Phi(u) - \Phi(v)| \leq L \|u - v\|_X .$$

These assumptions are, for instance, satisfied for the potential Φ_N resulting from the heat equation with finite-dimensional data and Gaussian noise, see Section 3.3 in [Dashti and Stuart 2017]. The following theorem yields the Onsager–Machlup functional of μ^y and show that it has a minimiser.

Theorem 4.4 ([Dashti, Law, et al. 2013, Thm. 3.2]). *If Φ satisfies Assumption 4.3, then $I\colon X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$,*

$$I(u) := \begin{cases} \Phi(u) + \frac{1}{2}\|u\|_E^2 & \text{if } u \in E, \\ \infty & \text{if } u \in X \setminus E, \end{cases}$$

is the Onsager–Machlup functional of μ^y.

The second term $R(u) := \frac{1}{2}\|u\|_E^2$ of the functional I can be interpreted as a generalised logarithmic prior density, since R is the Onsager–Machlup functional of the prior distribution μ_0 by Corollary 2.6.

Theorem 4.5 ([Dashti and Stuart 2017, Cor. 1]). *If Φ satisfies Assumption 4.3 (i) and (iii) and $\mu^y(X) = 1$, then there exists $\bar{u} \in E$ such that*

$$I(\bar{u}) = \inf\{I(u) : u \in E\}.$$

The main result of [Dashti, Law, et al. 2013] now shows that a point $\hat{u} \in X$ is a minimiser of the Onsager–Machlup functional of μ^y if and only if it is a MAP estimate.

Theorem 4.6 ([Dashti, Law, et al. 2013, Thm. 3.5]). *Suppose that Assumption 4.3 (ii) and (iii) hold. Assume also that there exists an $M \in \mathbb{R}$ such that $\Phi(u) \geq M$ for any $u \in X$.*

(i) Let $z^\delta = \arg\max_{z\in X} \mu^y(B_\delta(z))$. There is a $\bar{z} \in E$ and a subsequence of $\{z^\delta\}_{\delta>0}$ which converges to $\bar{z}$ strongly in X.

(ii) The limit $\bar{z}$ is a MAP estimate and a minimiser of I.

Corollary 4.7 ([Dashti, Law, et al. 2013, Cor. 3.10]). *Under the conditions of Theorem 4.6, we have the following.*

(i) Any MAP estimate, given by Definition 4.1, minimises the Onsager–Machlup functional I.

(ii) Any $z^* \in E$ which minimises the Onsager–Machlup functional I is a MAP estimate for the measure μ^y given by (4.1).

4.3 Unbounded Potential

As we will see in Section 5.4, the heat equation with Laplacian noise leads to a potential Φ that is even globally Lipschitz continuous, but not bounded from below, as required in Theorem 4.6. Motivated by this, we will show that a variational characterisation of MAP estimates is possible without this assumption in case that Φ is globally Lipschitz continuous.

Assumption 4.8. The function Φ is Lipschitz continuous, i.e., there exists an $L > 0$, such that

$$|\Phi(u) - \Phi(v)| \leq L\|u - v\|_X \quad \text{for all } u, v \in X.$$

We first note that this assumption implies the previous assumptions on Φ.

Lemma 4.9. *If Φ is Lipschitz continuous, then it satisfies Assumption 4.3.*

Proof. By the Lipschitz continuity of Φ,

$$\Phi(u) + \varepsilon \|u\|_X^2 \geq \Phi(0) - L\|u\| + \varepsilon\|u\|_X^2 = \Phi(0) + \varepsilon \left(\|u\|_X - \frac{L}{\varepsilon}\right) \|u\|_X$$

holds for all $\varepsilon > 0$ and $u \in X$. Now the minimum of the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(t) = \varepsilon(t - \frac{L}{\varepsilon})t$ is attained in $\frac{L}{2\varepsilon}$, so that for given $\varepsilon > 0$ condition (i) is satisfied with

$$M := \Phi(0) + f\left(\frac{L}{2\varepsilon}\right) = \Phi(0) - \frac{L^2}{4\varepsilon}.$$

Condition (ii) is satisfied with $K := Lr$ by the Lipschitz continuity of Φ, as

$$\Phi(u) \leq L\|u\|_X \leq Lr = K$$

for all $u \in B_r(0)$. Condition (iii) is trivially satisfied. □

We proceed similar as in the proof of Theorem 4.6 and Corollary 4.7. For this, we require a series of Lemmas about small ball probabilities under the Gaussian prior measure μ_0. The following two Lemmas are valid for centred Gaussian measures on a separable Banach space X.

Lemma 4.10 ([Dashti, Law, et al. 2013, Lemma 3.6])**.** *Let $\varepsilon > 0$. Then we have*

$$\frac{\mu_0(B_\varepsilon(u))}{\mu_0(B_\varepsilon(0))} \leq e^{\frac{a_1}{2}\varepsilon^2} e^{-\frac{a_1}{2}(\|u\|_X - \varepsilon)^2}$$

for all $u \in X$ and a constant a_1 independent of z and ε.

Lemma 4.11 ([Dashti, Law, et al. 2013, Lemma 3.7])**.** *Suppose that $\bar{u} \notin E$, $\{u_\varepsilon\}_{\varepsilon>0} \subset X$ and that u_{ε_n} converges weakly to $\bar{u}$ in X for $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset (0, \infty)$ with $\varepsilon_n \to 0$. Then for any $\delta > 0$, there exists $n \in \mathbb{N}$ such that*

$$\frac{\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu_0(B_{\varepsilon_n}(0))} < \delta.$$

Lemma 4.12. *Let $u \in X$ and $\varepsilon > 0$. For all $n \in \mathbb{N}$ and $x \in X$ define the projections $P_n\colon X \mapsto \mathbb{R}^n$,*

$$P_n x := ((x, \varphi_1)_X, \ldots, (x, \varphi_n))^T.$$

Moreover, for every $n \in \mathbb{N}$ let A_n be the cylindrical set

$$A_n := \{x \in X : P_n x \in B_\varepsilon(P_n u)\},$$

where $B_\varepsilon(P_n u) := \{x \in \mathbb{R}^n : \|x - P_n u\|_2 < \varepsilon\}$ denotes an open ball in $\mathbb{R}^n$. Then for every $\delta > 0$ there exists an $N \in \mathbb{N}$ such that

$$\mu_0(B_\varepsilon(u) \vartriangle A_n) \leq \delta \quad \textit{for } n \geq N,$$

where $\vartriangle$ denotes the symmetric difference.

Proof. First, we note that $\mu_0(\overline{B_\varepsilon(u)}) = \mu_0(B_\varepsilon(u))$ and

$$\mu_0(\{x \in X : P_n x \in \overline{B_\varepsilon(P_n u)}\}) = \mu_0(\overline{A_n}) = \mu_0(A_n) \quad \text{for all } n \in \mathbb{N}.$$

Next, we show that the sets $\overline{A_n}$ decrease to $\overline{B_\varepsilon(u)}$. It can easily be seen that $\overline{A_1} \supset \overline{A_2} \supset \dots$ holds and that $\overline{B_\varepsilon(u)} \subset \overline{A_n}$ for all $n \in \mathbb{N}$. In order to see that $\bigcap_{n=1}^\infty \overline{A_n} = \overline{B_\varepsilon(u)}$, we consider a point $x \in X \setminus \overline{B_\varepsilon(u)}$. Then, $\rho := \|x-u\|^2 - \varepsilon^2 > 0$ and we can choose a $K \in \mathbb{N}$ such that

$$\|P_K x - P_K u\|_2^2 = \sum_{k=1}^K |(x-u, \varphi_k)_X|^2 \geq \|x-u\|^2 - \frac{\rho}{2} = \varepsilon^2 + \frac{\rho}{2} > \varepsilon^2.$$

This shows that $P_K x \notin \overline{B_\varepsilon(P_K u)}$. Therefore, $x \notin \overline{A_K}$ and in particular $x \notin \bigcap_{n=1}^\infty \overline{A_n}$.

As a probability measure, μ_0 is upper semicontinuous by [Klenke 2014, Thm. 1.36], so that

$$\mu_0(\overline{A_n}) \longrightarrow \mu_0(\overline{B_\varepsilon(u)}) \quad \text{as } n \longrightarrow \infty.$$

For every $\delta > 0$, this allows us to choose an $N \in \mathbb{N}$ such that

$$\mu_0(B_\varepsilon(u) \bigtriangleup A_n) = \mu_0(A_n \setminus B_\varepsilon(u)) = \mu_0(A_n) - \mu_0(B_\varepsilon(u)) \leq \delta \quad \text{for } n \geq N. \qquad \square$$

The following statement is a generalisation of Lemma 3.9 in [Dashti, Law, et al. 2013].

Lemma 4.13. *Suppose that $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}} \subset X$ converges weakly but not strongly to $\bar{u} \in E$ for $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ with $\varepsilon_n \longrightarrow 0$. Then for every $\delta > 0$, there is an $n \in \mathbb{N}$ such that*

$$\frac{\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu_0(B_{\varepsilon_n}(0))} \leq \delta.$$

Proof. Let $\delta > 0$. Let $\{\varphi_k\}_{k\in\mathbb{N}}$ be an orthonormal basis of X consisting of eigenvectors of Q, let $\{\lambda_k\}_{k\in\mathbb{N}}$ be the associated eigenvalues in descending order and define $a_k := \frac{1}{\lambda_k}$ for all $k \in \mathbb{N}$. Furthermore, we define $u_k := (u, \varphi_k)_X$ and

$$P_k u := (u_1, u_2, \dots, u_k)^T$$

for all $u \in X$ and $k \in \mathbb{N}$. For the measure $\mu_{0,m} := \mu_0 \circ P_m^{-1}$ on $\mathbb{R}^m$,

$$\mu_{0,m}(M) = C_m \int_M e^{-\frac{1}{2}(a_1 x_1^2 + \dots + a_m x_m^2)} \mathrm{d}x$$

holds for all $M \in \mathcal{B}(\mathbb{R}^m)$ by [Bogachev 1998, Prop. 1.2.2], where $C_m = ((2\pi)^m \prod_{k=1}^m \lambda_k)^{-1/2}$. Since u_{ε_n} converges weakly and not strongly to $\bar{u}$, we have

$$\liminf_{n\to\infty} \|u_{\varepsilon_n}\|_X > \|\bar{u}\|_X.$$

Therefore, $n_0 \in \mathbb{N}$ and $c > 0$ exist such that

$$\|u_{\varepsilon_n}\|_X^2 \geq \|\bar{u}\|_X^2 + c \quad \text{for all } n \geq n_0.$$

We choose $A > 0$ such that $e^{-\frac{c^2}{36}A^2} \le \frac{\delta}{2}$ and $K \in \mathbb{N}$ such that $a_k > A^2$ for all $k \ge K$. For $m \ge K$ and $n \in \mathbb{N}$ we consider

$$\begin{aligned}
\mu_{0,m}(B_{\varepsilon_n}(0)) &= C_m \int_{B_{\varepsilon_n}(0)} e^{-\frac{1}{2}\left(a_1 x_1^2+\dots+a_m x_m^2\right)} \mathrm{d}x \\
&= C_m \int_{B_{\varepsilon_n}(0)} e^{-\frac{1}{2}A^2\left(x_{K+1}^2+\dots+x_m^2\right)} \\
&\qquad \cdot e^{-\frac{1}{2}\left(a_1 x_1^2+\dots+a_K x_K^2+\left(a_{K+1}-A^2\right)x_{K+1}^2+\dots+\left(a_m-A^2\right)x_m^2\right)} \mathrm{d}x \\
&= \frac{C_m}{\widehat{C}_m} \int_{B_{\varepsilon_n}(0)} e^{-\frac{1}{2}A^2\left(x_{K+1}^2+\dots+x_m^2\right)} \widehat{\mu}_{0,m}(\mathrm{d}x) \\
&\ge \frac{C_m}{\widehat{C}_m} e^{-\frac{1}{2}A^2{\varepsilon_n}^2} \widehat{\mu}_{0,m}(B_{\varepsilon_n}(0)),
\end{aligned}$$

where

$$\widehat{\mu}_{0,m}(M) := \widehat{C}_m \int_M e^{-\frac{1}{2}\left(a_1 x_1^2+\dots+a_K x_K^2+\left(a_{K+1}-A^2\right)x_{K+1}^2+\dots+\left(a_m-A^2\right)x_m^2\right)} \mathrm{d}x$$

for all $M \in \mathcal{B}(\mathbb{R}^m)$. Note that $\widehat{\mu}_{0,m}$ is a centred Gaussian measure. The weak convergence implies $(u_{\varepsilon_n}, \varphi_k)_X \to (\bar{u}, \varphi_k)_X$ for all $k \in \mathbb{N}$. Therefore, we can choose $n_1 \ge n_0$ such that for all $n \ge n_1$,

$$\sum_{k=1}^{K} \left((u_{\varepsilon_n}, \varphi_k)_X^2 - (\bar{u}, \varphi_k)_X^2\right) \le \frac{c}{3},$$

and consequently

$$\begin{aligned}
\sum_{k=K+1}^{\infty} (u_{\varepsilon_n}, \varphi_k)_X^2 &= \|u_{\varepsilon_n}\|_X^2 - \sum_{k=1}^{K} (u_{\varepsilon_n}, \varphi_k)_X^2 \ge \|\bar{u}\|_X^2 + c - \sum_{k=1}^{K} (u_{\varepsilon_n}, \varphi_k)_X^2 \\
&= c + \sum_{k=1}^{K} \left((\bar{u}, \varphi_k)_X^2 - (u_{\varepsilon_n}, \varphi_k)_X^2\right) + \sum_{k=K+1}^{\infty} (\bar{u}, \varphi_k)_X^2 \ge \frac{2}{3}c.
\end{aligned}$$

Finally, we choose $n \ge n_1$ such that $\varepsilon_n \le \frac{c^2}{36}$ and $\rho > 0$ such that

$$\left(\frac{\delta}{2} + 1\right)\rho \le \frac{\delta}{2}\mu_0(B_{\varepsilon_n}(0)).$$

By Lemma 4.12, there exists an $m_0 \ge K$ for the balls $B_{\varepsilon_n}(0)$ and $B_{\varepsilon_n}(u_{\varepsilon_n})$ such that the cylindrical sets

$$A_0 := P_m^{-1}(B_{\varepsilon_n}(P_m 0))) \quad \text{and} \quad A_u = P_m^{-1}(B_{\varepsilon_n}(P_m u_{\varepsilon_n}))$$

satisfy $\mu_0(B_{\varepsilon_n}(0) \triangle A_0) < \rho$ and $\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n}) \triangle A_u) < \rho$ for all $m > m_0$. Note that here, $B_\varepsilon(P_m u)$ denotes an open ball in $\mathbb{R}^m$ for $\varepsilon > 0$ and $u \in X$. It follows that

$$\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n})) \le \mu_{0,m}(B_{\varepsilon_n}(P_m u_{\varepsilon_n})) + \rho$$

and

$$\mu_{0,m}(B_{\varepsilon_n}(P_m 0)) \le \mu_0(B_{\varepsilon_n}(0)) + \rho.$$

In a last step, we choose $m \geq m_0$ such that

$$\sum_{k=K+1}^{m} (u_{\varepsilon_n}, \varphi_k)_X^2 \geq \frac{c}{3}.$$

By Andersons's inequality (see Theorem 2.8.10 in [Bogachev 1998]), we have

$$\widehat{\mu}_{0,m}(B_{\varepsilon_n}(P_m u_{\varepsilon_n})) \leq \widehat{\mu}_{0,m}(B_{\varepsilon_n}(0)).$$

By the choice of n and A, this leads to

$$\begin{aligned}
\mu_{0,m}(B_{\varepsilon_n}(P_m u_{\varepsilon_n})) &= \frac{C_m}{\widehat{C}_m} \int_{B_{\varepsilon_n}(P_m u_{\varepsilon_n})} e^{-\frac{1}{2}A^2(x_{K+1}^2+\dots+x_m^2)} \widehat{\mu}_{0,m}(\mathrm{d}x) \\
&\leq \frac{C_m}{\widehat{C}_m} e^{-\frac{1}{2}A^2\left(\frac{c}{3}-\varepsilon_n\right)^2} \widehat{\mu}_{0,m}(B_{\varepsilon_n}(P_m u_{\varepsilon_n})) \\
&\leq \frac{C_m}{\widehat{C}_m} e^{-\frac{1}{2}A^2\left(\frac{c^2}{9}-2\varepsilon_n\right)} e^{-\frac{1}{2}A^2\varepsilon_n^2} \widehat{\mu}_{0,m}(B_{\varepsilon_n}(0)) \\
&\leq e^{-\frac{c^2}{36}A^2} \mu_{0,m}(B_{\varepsilon_n}(0)) \leq \frac{\delta}{2}\mu_{0,m}(B_{\varepsilon_n}(0)).
\end{aligned}$$

Consequently, by the choice of m_0 and ρ we have

$$\begin{aligned}
\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n})) &\leq \frac{\delta}{2}\left(\mu_0(B_{\varepsilon_n}(0)) + \rho\right) + \rho = \frac{\delta}{2}\mu_0(B_{\varepsilon_n}(0)) + \left(\frac{\delta}{2}+1\right)\rho \\
&\leq \frac{\delta}{2}\mu_0(B_{\varepsilon_n}(0)) + \frac{\delta}{2}\mu_0(B_{\varepsilon_n}(0)) = \delta\mu_0(B_{\varepsilon_n}(0)).
\end{aligned}$$ □

The proof of the following lemma was kindly provided by Masoumeh Dashti (personal communication, 3 July 2017).

Lemma 4.14. *Let $\varepsilon_n > 0$ for all $n \in \mathbb{N}$ and $\varepsilon_n \to 0$ as $n \to \infty$. Assume that $\{u_n\}_{n\in\mathbb{N}} \subset X$ converges towards $\bar{u} \in E$ with respect to $\|\cdot\|_X$. Then*

$$\limsup_{n\to\infty} \frac{\mu_0(B_{\varepsilon_n}(u_n))}{\mu_0(B_{\varepsilon_n}(\bar{u}))} \leq 1.$$

Proof. First note that $Z = Q(X)$ is dense in $E = Q^{\frac{1}{2}}(X)$, and that for every $w \in Z$ the linear functional $W_{Q^{-1/2}w} = (Q^{-1}w, \cdot)_X$ is continuous. Now, by the Cameron-Martin theorem and Anderson's inequality,

$$\begin{aligned}
\mu_0(B_{\varepsilon_n}(u_n)) &= \int_{B_{\varepsilon_n}(u_n-w)} \exp\left(-\|w\|_E^2 + W_{Q^{-\frac{1}{2}}w}(v)\right) \mu_0(\mathrm{d}v) \\
&\leq e^{-\frac{1}{2}\|w\|_E^2} \sup_{v\in B_{\varepsilon_n}(u_n-w)} \left\{\exp((Q^{-1}w, v)_X)\right\} \mu_0(B_{\varepsilon_n}(u_n-w)) \\
&\leq e^{-\frac{1}{2}\|w\|_E^2} \sup_{v\in B_{\varepsilon_n}(u_n-w)} \left\{\exp((Q^{-1}w, v)_X)\right\} \mu_0(B_{\varepsilon_n}(0))
\end{aligned}$$

holds for all $w \in Z$ and $n \in \mathbb{N}$. On the other hand, the symmetry of $B_{\varepsilon_n}(0)$ implies

$$\begin{aligned}
\mu_0(B_{\varepsilon_n}(\bar{u})) &= e^{-\frac{1}{2}\|\bar{u}\|_E^2} \int_{B_{\varepsilon_n}(0)} \exp(W_{Q^{-\frac{1}{2}}\bar{u}}(v))\mu_0(\mathrm{d}v) \\
&= e^{-\frac{1}{2}\|\bar{u}\|_E^2} \int_{B_{\varepsilon_n}(0)} \frac{1}{2}\left(\exp(W_{Q^{-\frac{1}{2}}\bar{u}}(v)) + \exp(-W_{Q^{-\frac{1}{2}}\bar{u}}(v))\right)\mu_0(\mathrm{d}v) \\
&\geq e^{-\frac{1}{2}\|\bar{u}\|_E^2}\mu_0(B_{\varepsilon_n}(0)).
\end{aligned}$$

Using the continuity of $(Q^{-1}w, \cdot)_X$ we obtain

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\mu_0(B_{\varepsilon_n}(u_n))}{\mu_0(B_{\varepsilon_n}(\bar{u}))} &\leq e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|w\|_E^2} \limsup_{n\to\infty} \left\{ \sup_{v\in B_{\varepsilon_n}(u_n - w)} \exp((Q^{-1}w, v)_X) \right\} \\
&= e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|w\|_E^2} \exp((Q^{-1}w, \bar{u} - w)_X) \\
&= e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|w\|_E^2} \exp((w, \bar{u} - w)_E)
\end{aligned}$$

for all $w \in Z$. In particular, if we consider a sequence $\{w_j\}_{j\in\mathbb{N}} \subset Z$ with $w_j \longrightarrow \bar{u}$ in E as $j \longrightarrow \infty$, the previous estimate leads to

$$\limsup_{n\to\infty} \frac{\mu_0(B_{\varepsilon_n}(u_n))}{\mu_0(B_{\varepsilon_n}(\bar{u}))} \leq 1. \qquad \square$$

Now we are able to prove the main results of this chapter.

Theorem 4.15. *For every $\varepsilon > 0$ let $u_\varepsilon \in X$ be a maximiser of $u \mapsto \mu^y(B_\varepsilon(u))$, i.e.,*

$$\mu^y(B_\varepsilon(u_\varepsilon)) = \max_{u\in X} \mu^y(B_\varepsilon(u)).$$

If Φ is Lipschitz continuous, then the following holds true:

(i) For every positive sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ with $\varepsilon_n \longrightarrow 0$, the sequence $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ contains a subsequence that converges in X towards some $\bar{u} \in E$.

(ii) Every cluster point $\bar{u} \in X$ of $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ satisfies

$$\lim_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(\bar{u}))}{\mu^y(B_{\varepsilon_n}(u_{\varepsilon_n}))} = 1$$

and minimises the Onsager–Machlup functional of μ^y.

(iii) Every cluster point $\bar{u} \in X$ of $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ is a MAP estimate.

Proof. We proceed along the lines of the proof of Theorem 3.5 in [Dashti, Law, et al. 2013]. Without loss of generality, we may assume that $\Phi(0) = 0$, because adding a constant to Φ is absorbed by the normalisation constant Z, while the measure μ^y remains the same.

Ad (i): First of all, we show that $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ is bounded in X. Then

$$|\Phi(u)| = |\Phi(u) - \Phi(0)| \leq L\|u\|_X$$

for all $u \in X$, where L denotes the Lipschitz constant of Φ. From this we obtain, using Anderson's inequality, that

$$\begin{aligned}
\mu^y(B_\varepsilon(u_\varepsilon)) &= \max_{u\in E} \int_{B_\varepsilon(u)} \mu^y(dv) = \max_{u\in E} \int_{B_\varepsilon(u)} \frac{1}{Z} e^{-\Phi(v)} \mu_0(dv) \\
&\geq \frac{1}{Z} \int_{B_\varepsilon(0)} e^{-\Phi(v)} \mu_0(dv) \geq \frac{1}{Z} \int_{B_\varepsilon(0)} e^{-L\|v\|_X} \mu_0(dv) \\
&\geq \frac{1}{Z} e^{-L\varepsilon} \mu_0(B_\varepsilon(0))
\end{aligned}$$

with $Z = \int_X \exp(-\Phi(v))\mu_0(\mathrm{d}v)$ as before. On the other hand,

$$\begin{aligned}
\mu^y(B_\varepsilon(u)) &= \int_{B_\varepsilon(u)} \frac{1}{Z} e^{-\Phi(v)} \mu_0(dv) \leq \frac{1}{Z} \int_{B_\varepsilon(u)} e^{L\|v\|_X} \mu_0(dv) \\
&\leq \frac{1}{Z} e^{L(\|u\|_X+\varepsilon)} \mu_0(B_\varepsilon(u))
\end{aligned}$$

holds for all $u \in X$. Altogether, this yields

$$\mu_0(B_\varepsilon(u_\varepsilon)) \geq Z e^{-L(\|u_\varepsilon\|_X+\varepsilon)} \mu^y(B_\varepsilon(u_\varepsilon)) \geq e^{-L(\|u_\varepsilon\|_X+2\varepsilon)} \mu_0(B_\varepsilon(0))$$

for all $\varepsilon > 0$. However, Lemma 4.10 says that there is an $a_1 > 0$ such that

$$\frac{\mu_0(B_\varepsilon(u_\varepsilon))}{\mu_0(B_\varepsilon(0))} \leq e^{-\frac{a_1}{2}\left(\|u_\varepsilon\|_X^2 - 2\varepsilon\right)}$$

for all $\varepsilon > 0$. Assuming that $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ is unbounded, i.e., that there is a subsequence, again denoted by $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$, with $\|u_{\varepsilon_n}\|_X \to \infty$ as $n \to \infty$, leads to a contradiction, because

$$\begin{aligned}
&\frac{a_1}{2}\left(\|u_{\varepsilon_n}\|^2 - 2\varepsilon_n\right) - L\left(\|u_{\varepsilon_n}\| + 2\varepsilon_n\right) \\
&= \left(\frac{a_1}{2}\|u_{\varepsilon_n}\| - L\right)\|u_{\varepsilon_n}\| - 2\left(\frac{a_1}{2} + L\right)\varepsilon_n \to \infty
\end{aligned}$$

as $n \to \infty$, which implies

$$e^{-L\left(\|u_{\varepsilon_n}\|_X + 2\varepsilon_n\right)} > e^{-\frac{a_1}{2}\left(\|u_{\varepsilon_n}\|_X^2 - 2\varepsilon_n\right)}$$

for sufficiently large n. So $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ is bounded and therefore contains a subsequence, again denoted by $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$, which converges weakly towards some $\bar{u} \in X$ as $n \to \infty$.

Now, we show that $\bar{u} \in E$. By definition of u_ε and the boundedness of $\{u_\varepsilon\}_{\varepsilon>0}$ we have

$$\begin{aligned}
1 &\leq \frac{\mu^y(B_\varepsilon(u_\varepsilon))}{\mu^y(B_\varepsilon(0))} = \frac{\int_{B_\varepsilon(u_\varepsilon)} e^{-\Phi(v)}\mu_0(\mathrm{d}v)}{\int_{B_\varepsilon(0)} e^{-\Phi(v)}\mu_0(\mathrm{d}v)} \leq \frac{e^{L(\|u_\varepsilon\|_X+\varepsilon)}}{e^{-L\varepsilon}} \frac{\mu_0(B_\varepsilon(u_\varepsilon))}{\mu_0(B_\varepsilon(0))} \\
&= e^{L(\|u_\varepsilon\|_X+2\varepsilon)} \frac{\mu_0(B_\varepsilon(u_\varepsilon))}{\mu_0(B_\varepsilon(0))} \leq e^{L(R+2\varepsilon_1)} \frac{\mu_0(B_\varepsilon(u_\varepsilon))}{\mu_0(B_\varepsilon(0))} \qquad (4.4)
\end{aligned}$$

for an $R > 0$ and all $0 < \varepsilon \le \varepsilon_1$. If we assume that $\bar{u} \notin E$, then Lemma 4.11 tells us that there is an $n \in \mathbb{N}$ such that $\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n}))/\mu_0(B_{\varepsilon_n}(0)) \le \frac{1}{2}e^{-L(R+2\varepsilon_1)}$ and consequently

$$\frac{\mu^y(B_\varepsilon(u_{\varepsilon_n}))}{\mu^y(B_{\varepsilon_n}(0))} \le \frac{1}{2},$$

which poses a contradiction. So $\bar{u} \in E$.

Next, we show that u_{ε_n} converges strongly in X. Suppose it does not. Then Lemma 4.13 applies and yields

$$\liminf_{n\to\infty} \frac{\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu_0(B_{\varepsilon_n}(0))} = 0,$$

which is contradictory to (4.4). So the subsequence $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ does indeed converge strongly to $\bar{u} \in E$.

Ad (ii): Let $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ denote the subsequence that converges towards the cluster point $\bar{u} \in X$. First, we show that

$$\lim_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(\bar{u}))}{\mu^y(B_{\varepsilon_n}(u_{\varepsilon_n}))} = 1.$$

By definition of u_ε and the Lipschitz continuity of Φ we have

$$\begin{aligned} 1 \le \frac{\mu^y(B_\varepsilon(u_\varepsilon))}{\mu^y(B_\varepsilon(\bar{u}))} &= e^{\Phi(\bar{u})-\Phi(u_\varepsilon)} \frac{\int_{B_\varepsilon(u_\varepsilon)} e^{\Phi(u_\varepsilon)-\Phi(v)}\mu_0(\mathrm{d}v)}{\int_{B_\varepsilon(\bar{u})} e^{\Phi(\bar{u})-\Phi(v)}\mu_0(\mathrm{d}v)} \\ &\le e^{L\|u_\varepsilon-\bar{u}\|_X} e^{2L\varepsilon} \frac{\mu_0(B_\varepsilon(u_\varepsilon))}{\mu_0(B_\varepsilon(\bar{u}))}, \end{aligned}$$

for all $\varepsilon > 0$ and consequently, by the convergence $u_{\varepsilon_n} \to \bar{u}$ and Lemma 4.14,

$$\begin{aligned} 1 \le \liminf_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu^y(B_{\varepsilon_n}(\bar{u}))} &\le \limsup_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu^y(B_{\varepsilon_n}(\bar{u}))} \\ &\le \limsup_{n\to\infty} \frac{\mu_0(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu_0(B_{\varepsilon_n}(\bar{u}))} \le 1. \end{aligned}$$

Next we show that $\bar{u}$ minimises the Onsager–Machlup functional I of μ^y. By Theorem 4.5 a minimiser $u^* \in E$ of I exists. If we suppose that $\bar{u}$ was not a minimiser of I, then $I(\bar{u}) - I(u^*) > 0$, and thus

$$\begin{aligned} 1 \le \lim_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu^y(B_{\varepsilon_n}(u^*))} &= \lim_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(u_{\varepsilon_n}))}{\mu^y(B_{\varepsilon_n}(\bar{u}))} \lim_{n\to\infty} \frac{\mu^y(B_{\varepsilon_n}(\bar{u}))}{\mu^y(B_{\varepsilon_n}(u^*))} \\ &= 1 \exp(I(u^*) - I(\bar{u})) < 1, \end{aligned}$$

by the definition of u_ε and Theorem 4.4, which poses a contradiction.

Ad (iii): It remains to show that $\bar{u}$ is a mode of μ^y, i.e., that for every positive sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$ we have

$$\lim_{n\to\infty} \frac{\mu^y(B_{\delta_n}(\bar{u}))}{\mu^y(B_{\delta_n}(u_{\delta_n}))} = 1. \tag{4.5}$$

To this end, we choose an arbitrary subsequence of $\{\delta_n\}_{n\in\mathbb{N}}$, again denoted by $\{\delta_n\}_{n\in\mathbb{N}}$. Then, by (i), there exists a subsubsequence, again denoted by $\{\delta_n\}_{n\in\mathbb{N}}$, such that $u_{\delta_n} \to \tilde{u}$ for some $\tilde{u} \in E$. Moreover, by (ii), the limit $\tilde{u}$ minimises I and satisfies

$$\lim_{n\to\infty} \frac{\mu^y(B_{\delta_n}(\tilde{u}))}{\mu^y(B_{\delta_n}(u_{\delta_n}))} = 1.$$

Since $\bar{u}$ minimises I as well by (ii), this implies

$$\begin{aligned}\lim_{n\to\infty} \frac{\mu^y(B_{\delta_n}(\bar{u}))}{\mu^y(B_{\delta_n}(u_{\delta_n}))} &= \lim_{n\to\infty} \frac{\mu^y(B_{\delta_n}(\bar{u}))}{\mu^y(B_{\delta_n}(\tilde{u}))} \lim_{n\to\infty} \frac{\mu^y(B_{\delta_n}(\tilde{u}))}{\mu^y(B_{\delta_n}(u_{\delta_n}))} \\ &= \exp(I(\tilde{u}) - I(\bar{u}))1 = 1\end{aligned}$$

for the subsubsequence $\{\delta_n\}_{n\in\mathbb{N}}$ by Theorem 4.4. Now (4.5) follows for the original sequence $\{\delta_n\}_{n\in\mathbb{N}}$ from a subsequence-subsequence argument. □

Corollary 4.16. *If Φ is Lipschitz continuous, then $\bar{u} \in E$ is a MAP estimate if and only if it minimises the Onsager–Machlup functional I of μ^y.*

Proof. Let $\hat{u}$ be a mode of μ^y. By Theorem 4.15 $\bar{u}$ is also a mode, so that

$$\lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(\hat{u}))}{\mu^y(B_\varepsilon(\bar{u}))} = \lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(\hat{u}))}{\mu^y(B_\varepsilon(u_\varepsilon))} \lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(u_\varepsilon))}{\mu^y(B_\varepsilon(\bar{u}))} = 1.$$

And because Φ is Lipschitz continuous, we have

$$\frac{\mu^y(B_\varepsilon(\hat{u}))}{\mu^y(B_\varepsilon(\bar{u}))} = e^{\Phi(\bar{u})-\Phi(\hat{u})} \frac{\int_{B_\varepsilon(\hat{u})} e^{\Phi(\hat{u})-\Phi(v)}\mu_0(\mathrm{d}v)}{\int_{B_\varepsilon(\bar{u})} e^{\Phi(\bar{u})-\Phi(v)}\mu_0(\mathrm{d}v)} \le e^{L\|\hat{u}-\bar{u}\|_X} e^{2L\varepsilon} \frac{\int_{B_\varepsilon(\hat{u})} \mu_0(\mathrm{d}v)}{\int_{B_\varepsilon(\bar{u})} \mu_0(\mathrm{d}v)}.$$

This implies $\hat{u} \in E$, as otherwise Lemma 4.11 leads to

$$1 = \liminf_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(\hat{u}))}{\mu^y(B_\varepsilon(\bar{u}))} \le e^{L\|\hat{u}-\bar{u}\|_X} \liminf_{\varepsilon\to 0} \frac{\mu_0(B_\varepsilon(\hat{u}))}{\mu_0(B_\varepsilon(\bar{u}))} = 0,$$

a contradiction. Now Theorem 4.4 yields

$$1 = \lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(\hat{u}))}{\mu^y(B_\varepsilon(\bar{u}))} = \exp(I(\bar{u}) - I(\hat{u})),$$

and consequently $I(\bar{u}) = I(\hat{u})$.

Conversely, let u^* be a minimiser of I. Since $\bar{u}$ from Theorem 4.15 is also a minimiser and a mode, Theorem 4.4 tells us that

$$\begin{aligned}\lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(u^*))}{\mu^y(B_\varepsilon(u_\varepsilon))} &= \lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(u^*))}{\mu^y(B_\varepsilon(\bar{u}))} \lim_{\varepsilon\to 0} \frac{\mu^y(B_\varepsilon(\bar{u}))}{\mu^y(B_\varepsilon(u_\varepsilon))} \\ &= \exp(I(\bar{u}) - I(u^*))1 = 1.\end{aligned}$$

□

5 A Severely Ill-posed Linear Problem with Laplacian Noise

In this chapter we will study a generalised form of the inverse heat equation in a Bayesian setting. Roughly speaking, we will consider the operator equation

$$y = Ku + \eta$$

with additive Laplacian noise η, a centred Gaussian prior u and a linear operator K between two separable Hilbert spaces X and Y whose eigenvalues λ_k decay in the order of $\exp(-pk^{\frac{2}{d}})$ for some $p > 0$ and $d \in \mathbb{N}$. We will determine the conditional distribution of the unknown u given the data y via Bayesian inference as described in Chapter 1. This problem was studied in [Dashti and Stuart 2017] with Gaussian instead of Laplacian noise.

We will proceed as follows. First we will give some background on Laplace-like operators, their functional calculus and Hilbert scales. Then, we will state the considered problem setting in detail and point out the connection to the inverse heat equation. Subsequently, we will derive the posterior distribution, determine the CM and MAP estimators and study the consistency of the MAP estimator.

5.1 Laplace-like Operators

5.1.1 Definition and Basic Properties

We will use real powers of a Laplace-like operator to model the smoothness of both the noise η and the prior u in terms of the Hilbert scale it induces. Let A be a linear operator in a separable Hilbert space X. We make the following assumptions on A:

Assumption 5.1. (i) The operator A is densely defined, i.e., its domain $\mathcal{D}(A)$ is dense in X, and A is invertible.

(ii) A is self-adjoint, i.e.,

$$\mathcal{D}(A) = \mathcal{D}(A^*) := \{y \in X : x \mapsto (Ax, y)_X \text{ is continuous on } \mathcal{D}(A)\}$$

and

$$(Au, v)_X = (u, Av)_X \quad \text{for all } u, v \in \mathcal{D}(A).$$

(iii) There exists an orthonormal basis $\{\varphi_k\}_{k\in\mathbb{N}}$ of X consisting of eigenvectors of A, i.e.,

$$A\varphi_k = \alpha_k\varphi_k \quad \text{for all } k \in \mathbb{N}.$$

(iv) The eigenvalues $\{\alpha_k\}_{k\in\mathbb{N}}$ are positive, ordered to be non-decreasing and there are $C_+ \geq C_- > 0$ and $d \in \mathbb{N}$ such that

$$C_- k^{\frac{2}{d}} \leq \alpha_k \leq C_+ k^{\frac{2}{d}} \quad \text{for all } k \in \mathbb{N}.$$

We call operators that satisfy Assumption 5.1 *Laplace-like* operators. It follows from Assumption 5.1 (iv) that A is unbounded, but as a self-adjoint operator, A is at least closed, see [Werner 2007, Thm. VII.2.4]. In the following two lemmas, we show that, without loss of generality, we can weaken Assumptions 5.1 (i) and (ii). In Assumption 5.1 (ii), we may replace the self-adjointness of A by symmetry, i.e. we can drop the requirement $\mathcal{D}(A) = \mathcal{D}(A^*)$, and in Assumption 5.1 (i), the invertibility of A can be replaced by surjectivity (see Corollary 5.4 below).

Proposition 5.2. *Let A be a densely defined operator in X that satisfies Assumptions 5.1 (iii) and (iv). If, in addition, A is symmetric, i.e.,*

$$(Ax, y)_X = (x, Ay)_X \quad \textit{for all } x, y \in \mathcal{D}(A)$$

then the spectral decomposition

$$Ax = \sum_{k=1}^{\infty} \alpha_k (x, \varphi_k)_X \varphi_k$$

holds for all $x \in \mathcal{D}(A)$ and A is injective.

Proof. Since $\{\varphi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis of X by Assumption 5.1 (iii), we can decompose $Ax \in X$ for every $x \in \mathcal{D}(A)$ into

$$Ax = \sum_{k=1}^{\infty} (Ax, \varphi_k)_X \varphi_k.$$

Now the symmetry of A together with the fact that $\{\varphi_k\}_{k\in\mathbb{N}}$ are eigenvectors of A yields

$$Ax = \sum_{k=1}^{\infty} (x, A\varphi_k)_X \, \varphi_k = \sum_{k=1}^{\infty} \alpha_k \, (x, \varphi_k)_X \, \varphi_k \quad \text{for all } x \in \mathcal{D}(A).$$

We also obtain

$$\|Ax\|_X^2 = \sum_{k=1}^{\infty} \left|(Ax, \varphi_k)_X\right|^2 = \sum_{k=1}^{\infty} \alpha_k^2 \left|(x, \varphi_k)_X\right|^2 \quad \text{for all } x \in \mathcal{D}(A). \tag{5.1}$$

Now injectivity follows from Assumption 5.1 (iv), as

$$\|Ax\|_X^2 \geq \alpha_1^2 \sum_{k=1}^{\infty} \left|(x, \varphi_k)_X\right|^2 \geq C_-^2 \, \|x\|_X^2 \tag{5.2}$$

for all $x \in \mathcal{D}(A)$. □

Proposition 5.3. *Let A be a densely defined, symmetric operator in X that satisfies Assumptions 5.1 (iii) and (iv). If, in addition, A is surjective then A is continuously invertible,*

$$\mathcal{D}(A) = \mathcal{X}^2 := \left\{ x \in X : \sum_{k=1}^{\infty} \alpha_k^2 |(x, \varphi_k)_X|^2 < \infty \right\}$$

and A is self-adjoint.

Proof. By Proposition 5.2, A is injective and hence invertible. Moreover, the inverse A^{-1} is continuous by equation (5.2).

We have $\mathcal{D}(A) \subseteq \mathcal{X}^2$ by equation (5.1). Now we show that $\mathcal{D}(A) \supseteq \mathcal{X}^2$. For every $x \in \mathcal{X}^2$ we have

$$y := \sum_{k=1}^{\infty} \alpha_k (x, \varphi_k)_X \varphi_k \in X.$$

However,

$$y = Az = \sum_{k=1}^{\infty} \alpha_k (z, \varphi_k)_X \varphi_k$$

for some $z \in \mathcal{D}(A)$ by the surjectivity of A and Proposition 5.2. This implies $x = z \in \mathcal{D}(A)$, since $\{\varphi_k\}_{k \in \mathbb{N}}$ is an orthonormal basis of X and $\alpha_k > 0$ for all $k \in \mathbb{N}$.

Finally, we show that $\mathcal{D}(A) \supseteq \mathcal{D}(A^*)$ (the inclusion $\mathcal{D}(A) \subseteq \mathcal{D}(A^*)$) holds by the symmetry of A). Let $y \in \mathcal{D}(A^*)$ and set

$$P_n z := \sum_{k=1}^{n} (z, \varphi_k)_X \varphi_k$$

for all $n \in \mathbb{N}$ and $z \in X$. Then $P_n z \in \mathcal{X}^2 = \mathcal{D}(A)$ for all $n \in \mathbb{N}$ and $z \in X$ and $P_n z \to z$ as $n \to \infty$. Moreover, P_n is symmetric, because $\{\varphi_k\}_{k \in \mathbb{N}}$ is an orthonormal basis of X and $AP_n x = P_n A x$ for all $x \in \mathcal{D}(A)$ by Proposition 5.2. Now

$$\begin{aligned}(x, P_n A^* y)_X &= (P_n x, A^* y)_X = (AP_n x, y)_X \\ &= (P_n A x, y)_X = (Ax, P_n y)_X = (x, AP_n y)_X\end{aligned}$$

for all $x \in \mathcal{D}(A)$ by definition of the adjoint operator A^*. As $\mathcal{D}(A)$ is dense in X, this implies

$$P_n A^* y = AP_n y = \sum_{k=1}^{n} \alpha_k (y, \varphi_k)_X \varphi_k$$

for all $n \in \mathbb{N}$, and consequently

$$A^* y = \lim_{n \to \infty} P_n A^* y = \sum_{k=1}^{\infty} \alpha_k (y, \varphi_k)_X \varphi_k.$$

This shows that $y \in \mathcal{X}^2 = \mathcal{D}(A)$, as

$$\sum_{k=1}^{\infty} \alpha_k^2 |(y, \varphi_k)_X|^2 = \|A^* y\|^2 < \infty. \qquad \square$$

Corollary 5.4. *Every densely defined, symmetric, surjective operator A in X that satisfies Assumptions 5.1 (iii) and (iv) is Laplace-like.*

Proof. This follows immediatley from Propositions 5.2 and 5.3. □

Example 5.5. Let Ω be a bounded, open subset of $\mathbb{R}^d$ with C^∞ boundary $\partial\Omega$ and let

$$\Delta = \sum_{j=1}^{d} \frac{\partial}{\partial x_j}$$

denote the (weak) Laplace operator on $H^2(\Omega)$. We use Corollary 5.4 to show that $A = -\Delta$ is a Laplace-like operator in $L^2(\Omega)$ if its domain is chosen as $\mathcal{D}(A) := H^2(\Omega) \cap H_0^1(\Omega)$.

First of all, A is densely defined in $L^2(\Omega)$, since $C_c^\infty(\Omega)$, the space of smooth functions on Ω with compact support, is dense in $L^2(\Omega)$ and $C_c^\infty(\Omega) \subset H^2(\Omega) \cap H_0^1(\Omega)$. We show that, moreover, A is surjective. Theorem 2.2.2.3 in [Grisvard 2011] states that for every $f \in L^2(\Omega)$ there exists a unique $u \in H^2(\Omega)$ such that

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ \gamma u = 0 & \text{on } \partial\Omega, \end{cases}$$

where γ denotes the trace operator. Furthermore, for $u \in H^1(\Omega)$ we have $u \in H_0^1(\Omega)$ if and only if $\gamma u = 0$ by Corollary 1.5.1.6 in [Grisvard 2011]. This yields the existence of a $u \in H^2(\Omega) \cap H_0^1(\Omega)$ with $Au = f$. Note that surjectivity also holds for any convex bounded open set $\Omega \subset \mathbb{R}^d$, this follows from [Grisvard 2011, Thm. 3.2.1.2].

A is symmetric, since

$$\begin{aligned} (-\Delta u, v)_{L^2} &= \int_\Omega -\Delta u v \,\mathrm{d}x = \int_\Omega (\nabla u, \nabla v)_{\mathbb{R}^d} \,\mathrm{d}x \\ &= \int_\Omega -u \Delta v \,\mathrm{d}x = (u, -\Delta v)_{L^2} \end{aligned}$$

for all $u, v \in H^2(\Omega) \cap H_0^1(\Omega)$, where $\nabla u = (\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_d})^T$ for all $u \in H^1(\Omega)$.

Next, we show that A satisfies Assumptions 5.1 (iii) and (iv). By Theorem 6.5.1 in [Evans 1998], which holds for any bounded open set $\Omega \subset \mathbb{R}^d$, each eigenvalue of $-\Delta$ is real, the spectrum of $-\Delta$ is given by

$$\Sigma = \{\lambda_k\}_{k=1}^\infty,$$

where each eigenvalue is repeated according to its multiplicity,

$$0 < \lambda_1 \le \lambda_2 \le \lambda_3 \le \dots,$$

and

$$\lambda_k \to \infty \quad \text{as } k \to \infty.$$

Moreover, there exists an orthonormal basis $\{\varphi_k\}_{k=1}^\infty$ of $L^2(\Omega)$, where for all $k \in \mathbb{N}$, $\varphi_k \in H_0^1(U)$ is an eigenfunction corresponding to λ_k:

$$\begin{cases} -\Delta \varphi_k = \lambda_k \varphi_k & \text{in } U, \\ \varphi_k = 0 & \text{on } \partial U. \end{cases}$$

Finally, Weyl's asymptotic formula

$$\lambda_k \asymp 4\pi \left(\frac{1}{\Gamma(d/2+1)} \mathrm{vol}(\overline{\Omega}) \right)^{2/d} k^{\frac{2}{d}}$$

(see Theorem 8.16 and Remark 8.17 in [Roe 1998]) provides an estimate for the eigenvalues that yields $\lambda_k \asymp k^{2/d}$, i.e., there exist $C_+ \geq C_- > 0$ such that

$$C_- k^{\frac{2}{d}} \leq \lambda_k \leq C_+ k^{\frac{2}{d}} \quad \text{for all } k \in \mathbb{N}.$$

5.1.2 Functional Calculus

In order to define real powers of a Laplace-like operator A as well as $\exp(-tA)$, $t \geq 0$, we briefly introduce the functional calculus for self-adjoint operators as described in [Engl, Hanke, and Neubauer 1996, Section 2.2].

Proposition 5.6 ([Engl, Hanke, and Neubauer 1996, Prop. 2.14]). *Let A be a self-adjoint operator in a Hilbert space X. Then there exists a unique spectral family $\{E_\lambda\}_{\lambda \in \mathbb{R}}$, called the* spectral decomposition *of A, such that*

$$\mathcal{D}(A) = \left\{ x \in X : \int_{-\infty}^{\infty} \lambda^2 d\|E_\lambda x\|^2 < \infty \right\}$$

and

$$Ax = \int_{-\infty}^{\infty} \lambda dE_\lambda x \quad \textit{for all } x \in \mathcal{D}(A).$$

We have already seen that the spectral decomposition of A simplifies to

$$Au = \sum_{k=1}^{\infty} \alpha_k (u, \varphi_k)_X \varphi_k \quad \text{for all } u \in \mathcal{D}(A).$$

So, the unique spectral family of A simply consists of the orthogonal projections E_λ, given by

$$E_\lambda u = \sum_{\substack{k \in \mathbb{N} \\ \alpha_k < \lambda}} (u, \varphi_k)_X \varphi_k \quad \text{for all } u \in \mathcal{D}(A).$$

Definition 5.7 ([Engl, Hanke, and Neubauer 1996, Def. 2.15]). Let A be a self-adjoint operator in a Hilbert space X with spectral family $\{E_\lambda\}_{\lambda \in \mathbb{R}}$. Moreover, let $\mathcal{M}_0$ denote the set of all real functions measurable with respect to the measure $d\|E_\lambda x\|^2$ for all $x \in X$. Then for all $f \in \mathcal{M}_0$ the operator $f(A)$ is defined by

$$f(A)x = \int_{-\infty}^{\infty} f(\lambda) dE_\lambda x \quad \text{for all } x \in \mathcal{D}(f(A)),$$

where

$$\mathcal{D}(f(A)) = \left\{ x \in X : \int_{-\infty}^{\infty} f(\lambda)^2 d\|E_\lambda x\|^2 < \infty \right\}.$$

Note that in particular $\mathcal{M}_0$ contains all piecewise continuous functions. We can now express fractional powers of A with $s \in \mathbb{R}$ as

$$A^s u = \sum_{k=1}^{\infty} \alpha_k^s \, (u, \varphi_k)_X \, \varphi_k \quad \text{for all } u \in \mathcal{D}(A^s),$$

where

$$\mathcal{D}(A^s) = \Big\{ u \in X : \sum_{k=1}^{\infty} \alpha_k^{2s} \left| (u, \varphi_k)_X \right|^2 < \infty \Big\}.$$

Proposition 5.8. *For every $t \geq 0$, $\exp(-tA)$ is a continuous linear operator on X.*

Proof. For fixed $t \geq 0$ the real function $\exp(-t\cdot)$ is continuous and hence in $\mathcal{M}_0$. This allows us to define $\exp(-At)$ in X by means of the functional calculus as

$$\exp(-At)u = \sum_{k=1}^{\infty} e^{-\alpha_k t} (u, \varphi_k)_X \varphi_k$$

for all $u \in \mathcal{D}(\exp(-At))$. Now $\mathcal{D}(\exp(-At)) = X$, because

$$\sum_{k=1}^{\infty} e^{-2\alpha_k t} \left| (u, \varphi_k)_X \right|^2 \leq \sum_{k=1}^{\infty} \left| (u, \varphi_k)_X \right|^2 = \|u\|_X^2 < \infty$$

for all $u \in X$. Moreover, $\exp(-At)$ is continuous, because

$$\|\exp(-At)u\|_X^2 = \sum_{k=1}^{\infty} e^{-2\alpha_k t} |(u, \varphi_k)_X|^2 \leq \|u\|^2$$

for all $u \in X$. □

Lemma 5.9. *For every $t \geq 0$ and $h \geq -t$, $\mathcal{R}(\exp(-tA)) \subseteq \mathcal{D}(\exp(-hA))$.*

Proof. Let $u \in \mathcal{R}(\exp(-tA))$. Then there exists $w \in X$ with $u = e^{-tA} w$. Now $u \in \mathcal{D}(\exp(-hA))$, because

$$\sum_{k=1}^{\infty} e^{-2h\alpha_k} |(u, \varphi_k)_X|^2 = \sum_{k=1}^{\infty} e^{-2(t+h)\alpha_k} |(w, \varphi_k)_X|^2 \leq \|w\|_X^2 < \infty.$$ □

5.1.3 Hilbert Scales

Now we can define the Hilbert scale induced by a Laplace-like operator A as in [Engl, Hanke, and Neubauer 1996, Section 8.4]. To this end, we note that A is an unbounded densely defined self-adjoint strictly positive operator in X. We consider the set

$$\mathcal{M} = \bigcap_{k=0}^{\infty} \mathcal{D}(A^k).$$

$\mathcal{M}$ is dense in X and it follows by spectral theory that A^s is defined on $\mathcal{M}$ for all $s \in \mathbb{R}$ and that

$$\mathcal{M} = \bigcap_{s \in \mathbb{R}} \mathcal{D}(A^s).$$

In $\mathcal{M}$, we introduce for all $s \in \mathbb{R}$ the inner products and norms

$$(u, v)_{\mathcal{X}^s} := \left(A^{\frac{s}{2}} u, A^{\frac{s}{2}} v\right)_X,$$
$$\|u\|_{\mathcal{X}^s} := \left\|A^{\frac{s}{2}} u\right\|_X,$$

respectively, for all $u, v \in \mathcal{M}$. Now, $\mathcal{H}^s$ is defined as the completion of $\mathcal{M}$ with respect to the norm $\|\cdot\|_{\mathcal{X}^s}$ and $(\mathcal{H}^s)_{s\in\mathbb{R}}$ is called the *Hilbert scale induced by* A. For each $s \in \mathbb{R}$, $\mathcal{H}^s$ by definition is a Hilbert space. It is, however, by now only equipped with the norm

$$\left\|[\{u_n\}_{n\in\mathbb{N}}]_{\mathcal{X}^s}\right\|_{\mathcal{H}^s} := \lim_{n\to\infty} \|u_n\|_{\mathcal{X}^s},$$

for any Cauchy sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{M}$, where $[u]_{\mathcal{X}^s}$ denotes the equivalence class of u with respect to the equivalence relation defined by

$$\{u_n\}_{n\in\mathbb{N}} \sim \{v_n\}_{n\in\mathbb{N}} \quad \Leftrightarrow \quad \lim_{n\to\infty} \|u_n - v_n\|_{\mathcal{X}^s} = 0$$

for all Cauchy sequences $\{u_n\}_{n\in\mathbb{N}}, \{v_n\}_{n\in\mathbb{N}}$ in $\mathcal{M}$. We will fix this later and show that at least for positive s the $\mathcal{H}^s$-norm coincides with the original $\mathcal{X}^s$-norm. The following proposition summarises some key properties of Hilbert scales that we will use throughout the rest of this chapter.

Proposition 5.10 ([Engl, Hanke, and Neubauer 1996, Prop. 2.16]).

(i) The space $\mathcal{H}^t$ is densely and continuously embedded in $\mathcal{H}^s$ for all $s < t$.

(ii) The operator $A^{\frac{t-s}{2}}$, defined on $\mathcal{M}$, has a unique extension to $\mathcal{H}^t$ for all $s, t \in \mathbb{R}$. This extension, again denoted by $A^{\frac{t-s}{2}}$, is an isomorphism from $\mathcal{H}^t$ to $\mathcal{H}^s$. Moreover, $A^{\frac{t-s}{2}}$ is self-adjoint and strictly positive in $\mathcal{H}^s$ for $t > s$.

(iii) $A^{t-s} = A^t A^{-s}$ holds for all $s, t \in \mathbb{R}$, and in particular $(A^s)^{-1} = A^{-s}$.

(iv) For all $s \geq 0$, $\mathcal{H}^s$ is isometrically isomorphic to $\mathcal{D}(A^{\frac{s}{2}})$ and $\mathcal{H}^{-s}$ is isometrically isomorphic to $(\mathcal{H}^s)^*$, the dual space of $\mathcal{H}^s$.

In the following, we will identify $\mathcal{H}^s$ for $s \geq 0$ with the subspaces

$$\mathcal{X}^s := \mathcal{D}(A^{\frac{s}{2}}) = \left\{u \in X : \sum_{k=1}^{\infty} \alpha_k^s \left|(u, \varphi_k)_X\right|^2 < \infty\right\}$$

of X by assigning the equivalence class of the series

$$\{E_{\alpha_n} u\}_{n\in\mathbb{N}} = \left\{\sum_{k=1}^{n} (u, \varphi_k)_X \, \varphi_k\right\}_{n\in\mathbb{N}}$$

with limit u to any function $u \in \mathcal{D}(A^{\frac{s}{2}})$. This way, we can extend the definition of the $\mathcal{X}^s$-norm to $\mathcal{D}(A^{\frac{s}{2}})$ and it coincides with the $\mathcal{H}^s$-norm, since

$$\begin{aligned}
\left\| \left[\{E_{\alpha_n} u\}_{n\in\mathbb{N}} \right]_{\mathcal{X}^s} \right\|_{\mathcal{H}^s}^2 &= \lim_{n\to\infty} \left\| E_{\alpha_n} u \right\|_{\mathcal{X}^s}^2 = \lim_{n\to\infty} \left\| A^{\frac{s}{2}} E_{\alpha_n} u \right\|_X^2 \\
&= \lim_{n\to\infty} \left\| \sum_{k=1}^{n} \alpha_k^{\frac{s}{2}} (u,\varphi_k)_X \varphi_k \right\|_X^2 = \sum_{k=1}^{\infty} \alpha_k^s \left| (u,\varphi_k)_X \right|^2 \\
&= \sum_{k=1}^{\infty} \left| \left(u, \alpha_k^{\frac{s}{2}} \varphi_k \right)_X \right|^2 = \sum_{k=1}^{\infty} \left| \left(A^{\frac{s}{2}} u, \varphi_k \right)_X \right|^2 \\
&= \left\| A^{\frac{s}{2}} u \right\|_X^2 = \|u\|_{\mathcal{X}^s}^2
\end{aligned}$$

for all $u \in \mathcal{D}(A^{\frac{s}{2}})$.

Lemma 5.11. *The set $\{\alpha_k^{-s/2}\varphi_k\}_{k\in\mathbb{N}} = \{A^{-s/2}\varphi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis in $\mathcal{X}^s$.*

Proof. $\{A^{-s/2}\varphi_k\}_{k\in\mathbb{N}}$ is an orthonormal system in $\mathcal{X}^s$, because

$$\left(A^{-\frac{s}{2}}\varphi_k, A^{-\frac{s}{2}}\varphi_j \right)_{\mathcal{X}^s} = \left(\varphi_k, \varphi_j\right)_X \quad \text{for all } k, j \in \mathbb{N}$$

by definition of the $\mathcal{X}^s$-norm and the orthonormality of $\{\varphi_k\}_{k\in\mathbb{N}}$. Moreover,

$$\begin{aligned}
\sum_{k=1}^{\infty} (u, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s} \alpha_k^{-\frac{s}{2}}\varphi_k &= \sum_{k=1}^{\infty} (A^{\frac{s}{2}}u, \varphi_k)_X \alpha_k^{-\frac{s}{2}}\varphi_k = \sum_{k=1}^{\infty} (u, A^{\frac{s}{2}}\varphi_k)_X \alpha_k^{-\frac{s}{2}}\varphi_k \\
&= \sum_{k=1}^{\infty} (u,\varphi_k)_X \varphi_k = u
\end{aligned}$$

for all $u \in \mathcal{X}^s$, because $\{\varphi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis for X. □

Proposition 5.12. *For every $s \geq 0$ and $t \geq 0$, $A^{-t} \colon \mathcal{X}^s \to \mathcal{X}^s$ is continuous.*

Proof. By Proposition 5.10, A^{-t} is an isomorphism from $\mathcal{X}^s$ to $\mathcal{X}^{s+2t}$ and $\mathcal{X}^{s+2t}$ is continuously embedded in $\mathcal{X}^s$. So A^{-t} is well-defined from $\mathcal{X}^s$ to $\mathcal{X}^s$ and continuous. □

Lemma 5.13. *For $\beta, s \in \mathbb{R}$ with $0 \leq s < \beta - \frac{d}{2}$ the operator $A^{s-\beta} \colon \mathcal{X}^s \to \mathcal{X}^s$ is trace class.*

Proof. By Proposition 5.12, $A^{s-\beta}$ is continuous from $\mathcal{X}^s$ to $\mathcal{X}^s$. As $\{\alpha_k^{-s/2}\varphi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{X}^s$ by Lemma 5.11 and $\alpha_k \geq C_- k^{2/d}$ for all $k \in \mathbb{N}$ by Assumption 5.1, the trace of $A^{s-\beta}$ computes as

$$\operatorname{Tr} A^{s-\beta} = \sum_{k=1}^{\infty} \left(A^{s-\beta}\alpha_k^{-\frac{s}{2}}\varphi_k, \alpha_k^{-\frac{s}{2}}\varphi_k \right)_{\mathcal{X}^s} = \sum_{k=1}^{\infty} \alpha_k^{s-\beta} \leq C_-^{s-\beta} \sum_{k=1}^{\infty} k^{\frac{2(s-\beta)}{d}}.$$

This series converges because $\frac{2(s-\beta)}{d} < -1$ by assumption. □

Lemma 5.14. *For $\gamma > 0$ the function $g\colon (0,\infty) \to \mathbb{R}$, $t \mapsto g(t) = t^\gamma e^{-t}$ attains its maximum at $t = \gamma$, increases monotonically for $0 < t \leq \gamma$ and decreases monotonically for $t \geq \gamma$. Moreover,*

$$0 < t^\gamma e^{-t} \leq \gamma^\gamma e^{-\gamma} \quad \text{for all } t > 0.$$

For $\gamma \leq 0$, in contrast, g decreases monotonically.

Proof. First of all, $g(t) = t^\gamma e^{-t} > 0$ for all $t > 0$. Differentiating leads to

$$g'(t) = \gamma t^{\gamma-1} e^{-t} - t^\gamma e^{-t} = (\gamma - t)\, t^{\gamma-1} e^{-t}$$

for all $t > 0$. As $t^{\gamma-1}e^{-t} > 0$ for all $t > 0$, we have $g'(t) = 0$ if and only if $t = \gamma$, $g'(t) > 0$ for $0 < t < \gamma$ and $g'(t) < 0$ for $t > \gamma$. So, for $\gamma > 0$, g attains its maximum at $t = \gamma$, which yields the first estimate. For $\gamma \leq 0$, it follows that g decreases monotonically. □

Lemma 5.15. *Let $s > 0$. Then $\mathcal{R}(\exp(-tA)) \subseteq \mathcal{X}^s$ for all $t > 0$ and there is a $C = C(s) > 0$, such that*

$$\|\exp(-tA)u\|_{\mathcal{X}^s} \leq C t^{-\frac{s}{2}} \|u\|_X \quad \text{for all } t > 0.$$

Proof. First, we consider

$$\sum_{k=1}^\infty \alpha_k^s |(e^{-tA}u, \varphi_k)_X|^2 = t^{-s} \sum_{k=1}^\infty (\alpha_k t)^s e^{-2\alpha_k t} |(u, \varphi_k)_X|^2.$$

By Lemma 5.14, the sequence $(\alpha_k t)^s e^{-2\alpha_k t}$ is bounded from above by $C := s^s e^{-s}$, so that

$$\sum_{k=1}^\infty \alpha_k^s |(e^{-tA}u, \varphi_k)_X|^2 \leq t^{-s} C \|u\|_X^2 < \infty.$$

This implies $e^{-tA}u \in \mathcal{X}^s$ and proves the estimate. □

5.2 Problem Setting

Let A be a Laplace-like operator in X and let $(\mathcal{X}^t)_{t\in\mathbb{R}}$ denote the Hilbert scale induced by A. We consider the linear operator equation

$$y = e^{-A}u + \eta. \tag{5.3}$$

for the unknown u, the noise η and the data y. We assume that $u \in X$ and $\eta \in \mathcal{X}^s$ for some $s \geq 0$. More precisely, we make the following assumptions on the probability distributions of u and η:

- The prior $u \sim \mathcal{N}_{r^2 A^{-\tau}}$ has a centred Gaussian distribution on X with $r > 0$ and $\tau > \frac{d}{2}$.
- The noise $\eta \sim \mathcal{L}_{b^2 A^{s-\beta}}$ has a centred Laplacian distribution on $\mathcal{X}^s$ with $b > 0$ and $\beta > s + \frac{d}{2}$ and is independent of the prior u. Here we define $\mathcal{L}_{b^2 A^{s-\beta}}$ using the orthonormal basis $\{\alpha_k^{-s/2}\varphi_k\}_{k\in\mathbb{N}}$ of $\mathcal{X}^s$.

Now we want to determine the regular conditional distribution of u given y. Note that by the choice of τ and β and Lemma 5.13, both covariance operators are trace class.

The reason why we chose η to have the covariance operator $b^2 A^{s-\beta}$ rather than $b^2 A^{-\beta}$ is that we want the Laplacian distribution μ of η on $\mathcal{X}^s$ to satisfy

$$\int_{\mathcal{X}^s} (\eta, \varphi_k)_X (\eta, \varphi_j)_X \mu(\mathrm{d}\eta) = (b^2 A^{-\beta} \varphi_k, \varphi_j)_X$$

for all $k, j \in \mathbb{N}$, independent of s. This is achieved by choosing $\mu := \mathcal{L}_{b^2 A^{s-\beta}}$, because

$$(u, \varphi_k)_X = (A^{\frac{s}{2}} u, A^{-\frac{s}{2}} \varphi_k)_X = (u, A^{-s} \varphi_k)_{\mathcal{X}^s}$$

for all $u \in \mathcal{X}^s$, $s \geq 0$, and $k \in \mathbb{N}$, and consequently,

$$\begin{aligned}\int_{\mathcal{X}^s} (\eta, \varphi_k)_X (\eta, \varphi_j)_X \mathcal{L}_{b^2 A^{s-\beta}}(\mathrm{d}\eta) &= \int_{\mathcal{X}^s} (\eta, A^{-s} \varphi_k)_{\mathcal{X}^s} (\eta, A^{-s} \varphi_j)_{\mathcal{X}^s} \mathcal{L}_{b^2 A^{s-\beta}}(\mathrm{d}\eta) \\ &= (b^2 A^{s-\beta} A^{-s} \varphi_k, A^{-s} \varphi_j)_{\mathcal{X}^s} = (b^2 A^{-\beta} \varphi_k, \varphi_j)_X\end{aligned}$$

by definition of the covariance operator.

5.3 Derivation from the Heat Equation

Now we go into the connection between the operator equation (5.3) and the heat equation. More precisely, we will derive equation (5.3) from the following abstract Cauchy problem.

Let A be a Laplace-like operator in X. Given initial data $u \in X$ we want to find a solution v to the initial value problem

$$\begin{cases} \dfrac{dv(t)}{dt} = -Av(t) & \text{for } t > 0, \\ v(0) = u, \end{cases} \tag{5.4}$$

in the sense of a continuous function $v\colon [0, \infty) \to X$ that is continuously differentiable for all $t > 0$ and satisfies $v(t) \in \mathcal{D}(A)$ for all $t > 0$.

Example 5.16. We obtain the classical heat equation on a bounded, open domain $\Omega \subset \mathbb{R}^d$ with C^∞ boundary by choosing $A := -\Delta$ to be the (weak) Laplace operator in $X := L^2(\Omega)$ and $\mathcal{D}(A) := H^2(\Omega) \cap H^1_0(\Omega)$.

We will to show that $v(t) := e^{-tA} u$ is the unique solution to the initial value problem (5.4).

Proposition 5.17. *The family $T(t) := \exp(-tA)$, $t \geq 0$, forms a strongly continuous semigroup of bounded linear operators on X.*

Proof. By Proposition 5.8, $\exp(-tA)$ is a continuous linear operator on X for every $t \geq 0$. The semigroup properties can be shown using the spectral decomposition of $\exp(-At)$. We have

$$T(0) = \exp(-A \cdot 0) = \mathrm{Id}_X$$

and

$$T(t + s) = \exp(-A(t + s)) = \exp(-At) \exp(-As) = T(t)T(s)$$

for all $t, s \geq 0$. It remains to show that

$$\lim_{\substack{t \to 0 \\ t \geq 0}} T(t)x = x \quad \text{for all } x \in X.$$

Let $\varepsilon > 0$. For every $t \geq 0$ we have

$$T(t)x - x = e^{-tA}x - x = \sum_{k=1}^{\infty} \left(e^{-t\alpha_k} - 1\right)(x, \varphi_k)_X \varphi_k.$$

Now choose $N \in \mathbb{N}$, such that $\sum_{k=N+1}^{\infty} |(x, \varphi_k)_X|^2 \leq \frac{\varepsilon^2}{2}$. Next, choose $t_0 > 0$, such that

$$t_0 \alpha_k \leq \frac{\varepsilon}{\sqrt{2}\|x\|_X}.$$

Note that $1 - e^{-t} \leq \min\{t, 1\}$ for all $t \geq 0$. Consequently, for all $t \in [0, t_0]$ we have

$$\begin{aligned}
\left\|x - e^{-tA}x\right\|_X^2 &= \sum_{k=1}^{\infty} \left(1 - e^{-t\alpha_k}\right)^2 |(x, \varphi_k)_X|^2 \\
&\leq \frac{\varepsilon^2}{2\|x\|_X^2} \left(\sum_{k=1}^{N} |(x, \varphi_k)_X|^2\right) + \frac{\varepsilon^2}{2} \leq \varepsilon^2.
\end{aligned}$$

□

Proposition 5.18. *(i) For every $u \in X$ the function $(0, \infty) \to X$, $t \mapsto \exp(-tA)u$ is differentiable, and*

$$\lim_{\substack{h \to 0 \\ t+h>0,\, h \neq 0}} \frac{\exp(-(t+h)A)u - \exp(-tA)u}{h} = -A\exp(-tA)u.$$

(ii) For all $u \in \mathcal{D}(A)$, it is, moreover, differentiable in $t = 0$, and

$$\lim_{\substack{h \to 0 \\ h>0}} \frac{\exp(-hA)u - u}{h} = -Au.$$

Proof. We first show the second statement. Let $u \in \mathcal{D}(A)$ and consider for fixed $h > 0$ the difference

$$\begin{aligned}
\frac{e^{-Ah}u - u}{h} - (-Au) &= \sum_{k=1}^{\infty} \left(\frac{e^{-\alpha_k h} - 1}{h} + \alpha_k\right)(u, \varphi_k)_X \varphi_k \\
&= \sum_{k=1}^{\infty} \left(1 - \frac{e^{-\alpha_k h} - 1}{-\alpha_k h}\right) \alpha_k (u, \varphi_k)_X \varphi_k.
\end{aligned}$$

For a given $\varepsilon > 0$ we choose $N \in \mathbb{N}$, such that

$$\sum_{k=N+1}^{\infty} \alpha_k^2 |(u, \varphi_k)_X|^2 < \frac{\varepsilon^2}{2}$$

holds. By the defining property of the exponential function,

$$\lim_{x\to 0}\frac{e^x-1}{x}=1.$$

This allows us to choose $t_0>0$, such that for all $h\in(0,t_0]$ and $k\in\{1,\dots,N\}$

$$\left(1-\frac{e^{-\alpha_k h}-1}{-\alpha_k h}\right)^2<\frac{1}{\|Au\|_X^2}\frac{\varepsilon^2}{2}$$

is satisfied. Furthermore,

$$0\le\frac{e^{-x}-1}{-x}\le 1$$

for $x>0$. Consequently, we have

$$\begin{aligned}\left\|\frac{e^{-Ah}u-u}{h}-(-Au)\right\|_X^2&=\sum_{k=1}^{\infty}\left(1-\frac{e^{-\alpha_k h}-1}{-\alpha_k h}\right)^2\alpha_k^2|(u,\varphi_k)_X|^2\\&\le\frac{1}{\|Au\|_X^2}\frac{\varepsilon^2}{2}\left(\sum_{k=1}^{N}\alpha_k^2|(u,\varphi_k)_X|^2\right)+\frac{\varepsilon^2}{2}\le\varepsilon^2\end{aligned}$$

for all $h\in(0,t_0]$.

Now we show the first statement. Let $u\in X$ and $t>0$. Then $v:=e^{-tA}u\in\mathcal{X}^2=\mathcal{D}(A)$ by Lemma 5.15 and Proposition 5.10 and $v\in\mathcal{D}(e^{-hA})$ for all $h\ge -t$ by Lemma 5.9. We proceed similarly as before. We have

$$\begin{aligned}\frac{e^{-(t+h)A}u-e^{-tA}u}{h}-(-Ae^{-tA}u)&=\frac{e^{-hA}v-v}{h}-(-Av)\\&=\sum_{k=1}^{\infty}\left(1-\frac{e^{-\alpha_k h}-1}{-\alpha_k h}\right)\alpha_k(v,\varphi_k)_X\varphi_k\end{aligned}$$

for all $h\ge -t$, $h\ne 0$. For a given $\varepsilon>0$ we choose $N\in\mathbb{N}$, such that $\sum_{k=N+1}^{\infty}\alpha_k^2|(u,\varphi_k)_X|^2<\frac{\varepsilon^2}{2}$. Then we choose $h_0\in(0,t]$, such that

$$\left(1-\frac{e^{-\alpha_k h}-1}{-\alpha_k h}\right)^2<\frac{1}{\|Av\|_X^2}\frac{\varepsilon^2}{2}$$

holds for all $h\in[-h_0,h_0]\setminus\{0\}$ and $k\in\{1,\dots,N\}$. This yields

$$\begin{aligned}\left\|\frac{e^{-hA}v-v}{h}-(-Av)\right\|_X^2&=\sum_{k=1}^{\infty}\left(1-\frac{e^{-\alpha_k h}-1}{-\alpha_k h}\right)^2\alpha_k^2|(v,\varphi_k)_X|^2\\&\le\frac{1}{\|Av\|_X^2}\frac{\varepsilon^2}{2}\left(\sum_{k=1}^{N}\alpha_k^2|(v,\varphi_k)_X|^2\right)+\frac{\varepsilon^2}{2}\le\varepsilon^2.\end{aligned}$$

for all $h\in[-h_0,h_0]\setminus\{0\}$. □

Corollary 5.19. *The family $T(t) = \exp(-tA)$, $t \geq 0$, is a differentiable semigroup and $-A$ is its infinitesimal generator.*

Proposition 5.20. *For every $u \in X$ there is a unique solution v of the initial value problem* (5.4) *given by $v(t) = \exp(-tA)u$ for all $t \geq 0$.*

Proof. As $-A$ is the infinitesimal generator of the differentiable semigroup $T(t) = \exp(-tA)$, $t \geq 0$, by Corollary 5.19, Theorem 4.1.5 in [Pazy 1983] yields that (5.4) has a unique solution v for every $u \in X$ and that it is given by $v(t) = T(t)u$. □

Now we assume that for given $u \in X$ only a measurement y of the solution $v(t) = e^{-tA}u$ at some fixed time $t > 0$ is available and that, moreover, this measurement is perturbed by additive noise $\eta \in \mathcal{X}^s$. This leads to the operator equation

$$y = v(t) + \eta = e^{-tA}u + \eta.$$

By Lemma 5.15, $e^{-tA}u \in \mathcal{X}^s$, so that also $y \in \mathcal{X}^s$. Without loss of generality we may now assume that $t = 1$ by considering the initial value problem for the scaled operator tA instead of A. Clearly, if A is a Laplace-like operator, then so is tA. This way we arrive at equation (5.3).

5.4 Posterior Distribution

Now we determine the posterior distribution for the inverse problem described in Section 5.2. We will show that introducing a prior and considering the problem from a Bayesian point of view acts as a regularisation and resolves its initial ill-posedness. Subsequently, we will show that the posterior distribution is also stable with regard to approximations of the log-likelihood.

5.4.1 Derivation

We want to use Theorem 1.3 with the noise distribution $\nu = \mathcal{L}_{b^2A^{s-\beta}}$ as a reference measure to obtain the posterior distribution μ^y in terms of its density with respect to the prior distribution $\mu_0 = \mathcal{N}_{r^2A^{-\tau}}$. To this end we need to show that all translates $\nu_{e^{-A}u} = \mathcal{L}_{e^{-A}u,b^2A^{s-\beta}}$ of the noise distribution are absolutely continuous with respect to $\mathcal{L}_{b^2A^{s-\beta}}$. Then the regular conditional distribution of y, given u, is given by $(u, V) \mapsto \mathcal{L}_{e^{-A}u,b^2A^{s-\beta}}(V)$ as described in Section 1.4.

Proposition 5.21. *For all $u \in X$, the measures $\mathcal{L}_{e^{-A}u,b^2A^{s-\beta}}$ and $\mathcal{L}_{b^2A^{s-\beta}}$ are equivalent and*

$$\frac{\mathrm{d}\mathcal{L}_{e^{-A}u,b^2A^{s-\beta}}}{\mathrm{d}\mathcal{L}_{b^2A^{s-\beta}}}(y) = \exp(-\Phi(u, y))$$

for $\mathcal{L}_{b^2A^{s-\beta}}$-almost all $y \in \mathcal{X}^s$, where

$$\Phi(u, y) := \sqrt{2}\sum_{k=1}^{\infty} \frac{|(y, \varphi_k)_X - e^{-\alpha_k}(u, \varphi_k)_X| - |(y, \varphi_k)_X|}{b\alpha_k^{-\beta/2}} \tag{5.5}$$

for all $u \in X$ and $y \in \mathcal{X}^s$.

Proof. Theorem 3.10, applied with $H = \mathcal{X}^s$, $a = e^{-A}u$ and $Q = b^2 A^{s-\beta}$, tells us that $\mathcal{L}_{e^{-A}u, b^2 A^{s-\beta}}$ and $\mathcal{L}_{b^2 A^{s-\beta}}$ are equivalent if and only if $e^{-A}u \in bA^{\frac{s-\beta}{2}}(\mathcal{X}^s)$. This is indeed the case for all $u \in X$, since $A^{\frac{s-\beta}{2}}(\mathcal{X}^s) = \mathcal{X}^\beta$ by Proposition 5.10 and $e^{-A}u \in \mathcal{X}^\beta$ for all $u \in X$ by Lemma 5.15. From Theorem 3.10 we also obtain the density

$$\frac{\mathrm{d}\mathcal{L}_{e^{-A}u, b^2 A^{s-\beta}}}{\mathrm{d}\mathcal{L}_{b^2 A^{s-\beta}}}(y) = \exp\left(-\sqrt{2}\sum_{k=1}^{\infty}\left(\left|(b^{-1}A^{\frac{\beta-s}{2}}(y - e^{-A}u), \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s}\right| - \left|(b^{-1}A^{\frac{\beta-s}{2}}y, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s}\right|\right)\right)$$

for $\mathcal{L}_{b^2 A^{s-\beta}}$-almost all $y \in \mathcal{X}^s$. Note that here the orthonormal basis $e_k = \alpha_k^{-s/2}\varphi_k$, $k \in \mathbb{N}$, of $\mathcal{X}^s$ and the $\mathcal{X}^s$-inner product were used. From Proposition 5.10 and the self-adjointness of e^{-A} we obtain that for all $u \in X$,

$$\begin{aligned}\frac{\mathrm{d}\mathcal{L}_{e^{-A}u, b^2 A^{s-\beta}}}{\mathrm{d}\mathcal{L}_{b^2 A^{s-\beta}}}(y) &= \exp\left(-\sqrt{2}\sum_{k=1}^{\infty} b^{-1}\left(\left|(y - e^{-A}u, A^{\frac{\beta}{2}}\varphi_k)_X\right| - \left|(y, A^{\frac{\beta}{2}}\varphi_k)_X\right|\right)\right)\\ &= \exp\left(-\sqrt{2}\sum_{k=1}^{\infty}\frac{\left|(y,\varphi_k)_X - (u, e^{-A}\varphi_k)_X\right| - |(y,\varphi_k)_X|}{b\alpha_k^{-\beta/2}}\right)\\ &= \exp(-\Phi(u,y))\end{aligned}$$

$\mathcal{L}_{b^2 A^{s-\beta}}$-almost surely. □

Corollary 5.22. *The regular conditional distribution of y given u is given by*

$$(u, V) \mapsto \mathcal{L}_{e^{-A}u, b^2 A^{s-\beta}}(V).$$

Proof. This follows from Proposition 1.4. □

We put some basic properties of the function Φ on record.

Proposition 5.23. *The function $\Phi\colon X \times \mathcal{X}^s \to \mathbb{R}$ defined by* (5.5) *is continuous and for every $y \in \mathcal{X}^s$, $u \mapsto \Phi(u,y)$ is Lipschitz continuous with a Lipschitz constant independent of y.*

Proof. We first show the Lipschitz continuity of $\Phi(\cdot, y)$ with Lipschitz constant

$$L := \frac{\sqrt{2}}{b}\beta^\beta e^{-\beta}(\operatorname{Tr} A^{-\beta})^{\frac{1}{2}}$$

independent of $y \in \mathcal{X}^s$. Here we use the notation $x_k := (x, \varphi_k)_X$ for all $k \in \mathbb{N}$ and $x \in X$. For $u, v \in X$ we estimate

$$\begin{aligned}|\Phi(u,y) - \Phi(v,y)| &\le \frac{\sqrt{2}}{b}\sum_{k=1}^{\infty}\alpha_k^{\frac{\beta}{2}}\left||y_k - e^{-\alpha_k}u_k| - |y_k - e^{-\alpha_k}v_k|\right|\\ &\le \frac{\sqrt{2}}{b}\sum_{k=1}^{\infty}\alpha_k^{-\frac{\beta}{2}}\alpha_k^{\beta}e^{-\alpha_k}|u_k - v_k|\end{aligned}$$

using the triangle inequality. The sequence $\{\alpha_k^\beta e^{-\alpha_k}\}_{k\in\mathbb{N}}$ is bounded from above by $\beta^\beta e^{-\beta}$ by Lemma 5.14 and the operator $A^{-\beta}$ is trace class by Lemma 5.13. We further estimate

$$\begin{aligned}|\Phi(u,y)-\Phi(v,y)| &\le \frac{\sqrt{2}}{b}\beta^\beta e^{-\beta}\left(\sum_{k=1}^{\infty}\alpha_k^{-\beta}\right)^{\frac{1}{2}}\left(\sum_{k=1}^{\infty}|u_k-v_k|^2\right)^{\frac{1}{2}}\\ &= \frac{\sqrt{2}}{b}\beta^\beta e^{-\beta}(\operatorname{Tr}A^{-\beta})^{\frac{1}{2}}\|u-v\|_X = L\|u-v\|_X\end{aligned}$$

using the Cauchy–Schwarz inequality.

Now we show the continuity in y. Let $u \in X$ and $\varepsilon > 0$. Here we estimate

$$\begin{aligned}|\Phi(u,y)-\Phi(u,z)| &= \left|\frac{\sqrt{2}}{b}\sum_{k=1}^{\infty}\alpha_k^{\frac{\beta}{2}}\left(|y_k-e^{-\alpha_k}u_k|-|y_k|-|z_k-e^{-\alpha_k}u_k|+|z_k|\right)\right|\\ &\le \frac{\sqrt{2}}{b}\sum_{k=1}^{N}2\alpha_k^{\frac{\beta}{2}}|y_k-z_k| + \frac{\sqrt{2}}{b}\sum_{k=N+1}^{\infty}2\alpha_k^{\frac{\beta}{2}}|e^{-\alpha_k}u_k|\end{aligned}$$

for all $y, z \in \mathcal{X}^s$ and $N \in \mathbb{N}$. As the series $\sum_{k=1}^{\infty}\alpha_k^{\frac{\beta}{2}}e^{-\alpha_k}|u_k|$ converges, we can choose $N = N(u)$ such that

$$\frac{\sqrt{2}}{b}\sum_{k=N+1}^{\infty}2\alpha_k^{\frac{\beta}{2}}e^{-\alpha_k}|u_k| \le \frac{\varepsilon}{2}.$$

Next, we choose

$$\delta := \frac{b}{2\sqrt{2}}\left(\sum_{k=1}^{N}\alpha_k^{\beta-s}\right)^{-\frac{1}{2}}\frac{\varepsilon}{2}.$$

This way, we have

$$\begin{aligned}\frac{\sqrt{2}}{b}\sum_{k=1}^{N}2\alpha_k^{\frac{\beta}{2}}|y_k-z_k| &\le \frac{2\sqrt{2}}{b}\left(\sum_{k=1}^{N}\alpha_k^{\beta-s}\right)^{\frac{1}{2}}\left(\sum_{k=1}^{N}\alpha_k^{s}|y_k-z_k|^2\right)^{\frac{1}{2}}\\ &\le \frac{\varepsilon}{2\delta}\|y-z\|_{\mathcal{X}^s} \le \frac{\varepsilon}{2}\end{aligned}$$

for all $y, z \in \mathcal{X}^s$ with $\|y-z\|_{\mathcal{X}^s} \le \delta$, and consequently

$$|\Phi(u,y)-\Phi(u,z)| \le \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon.$$

The continuity of Φ now follows from the continuity in u and y and the triangle inequality. □

Corollary 5.24. *The function* $(u, y) \mapsto \exp(\Phi(u, y))$ *is* $\mathcal{B}(X)\times\mathcal{B}(\mathcal{X}^s)$*-measurable.*

Proof. By Proposition 5.23, $\exp(\Phi(u, y))$ is continuous and hence $\mathcal{B}(X)\times\mathcal{B}(\mathcal{X}^s)$-measurable. □

Moreover, $u \mapsto \Phi(u, y)$ is convex for every $y \in \mathcal{X}^s$, but not strictly convex. The following example shows that in our case $u \mapsto \Phi(u, y)$ is not necessarily bounded from below.

Example 5.25. We can construct $y \in \mathcal{X}^s$ and a sequence $\{u^n\}_{n\in\mathbb{N}} \in X$ such that $\Phi(u^n, y) \to -\infty$ as follows. Set

$$y := \sum_{k=1}^{\infty} \alpha_k^{-\frac{s}{2}} \frac{1}{k} \varphi_k \qquad \text{and} \qquad u^n := \sum_{k=1}^{n} e^{\alpha_k} \alpha_k^{-\frac{s}{2}} \frac{1}{k} \varphi_k$$

for all $n \in \mathbb{N}$. Now consider

$$\Phi(u^n, y) = \frac{\sqrt{2}}{b} \sum_{k=1}^{n} \alpha_k^{\frac{\beta}{2}} \left(0 - \alpha_k^{-\frac{s}{2}} \frac{1}{k} \right).$$

Let $m \in \mathbb{N}$ be large enough such that $\alpha_k \geq 1$ for all $k \geq m$. As $\beta - s > 0$, it follows that

$$\Phi(u^n, y) \leq -\frac{\sqrt{2}}{b} \sum_{k=m}^{n} \alpha_k^{\frac{\beta-s}{2}} \frac{1}{k} \leq -\frac{\sqrt{2}}{b} \sum_{k=m}^{n} \frac{1}{k}$$

for all $n \geq m$. However, $-\sum_{k=m}^{n} \frac{1}{k} \to -\infty$ as $n \to \infty$ and hence also $\Phi(u^n, y) \to -\infty$. As $\|u^n\|_X \to \infty$ this example also shows that $\Phi(\cdot, y)$ is not coercive.

In contrast, the potential $\Phi_{\mathcal{N}}$ resulting from the heat equation with Gaussian noise is bounded from below in u for fixed $y \in \mathcal{X}^s$. Replacing the noise distribution $\mathcal{L}_{b^2 A^{s-\beta}}$ on $\mathcal{X}^s$ by $\mathcal{N}_{A^{s-\beta}}$ in this setting leads to the potential $\Phi_{\mathcal{N}} \colon X \times \mathcal{X}^s \to \mathbb{R}$,

$$\Phi_{\mathcal{N}}(u, y) = \frac{1}{2} \left\| A^{\frac{\beta}{2}} e^{-A} u \right\|_X^2 - \left(A^{\frac{\beta}{2}} e^{-\frac{A}{2}} y, A^{\frac{\beta}{2}} e^{-\frac{A}{2}} u \right)_X,$$

see Section 3.3 in [Dashti and Stuart 2017]. However, due to the quadratic term, the function $u \mapsto \Phi_{\mathcal{N}}(u, y)$ is not globally Lipschitz continuous for any fixed $y \in \mathcal{X}^s$, but only Lipschitz continuous on bounded sets.

Lemma 5.26. *Let μ_0 be a centred Gaussian meausure on X. Then for every $C > 0$, the function $u \mapsto \exp(C\|u\|_X)$ defined on X is μ_0-integrable.*

Proof. By Fernique's theorem [Bogachev 1998, Thm 2.8.5], there exists $\alpha > 0$ such that the integral $\int_X \exp(\alpha \|u\|_X^2) \mu_0(\mathrm{d}u)$ is finite. Set $R := \frac{C}{\alpha}$. Then the integral

$$\int_X \exp\left(C\|u\|_X\right) \mu_0(\mathrm{d}u) \leq \int_{B_R(0)} \exp\left(\frac{C^2}{\alpha}\right) \mu_0(\mathrm{d}u) + \int_{X \setminus B_R(0)} \exp\left(\alpha \|u\|_X^2\right) \mu_0(\mathrm{d}u)$$

is finite as well. □

Proposition 5.27. *The function $u \mapsto \exp(-\Phi(u, y))$ is $\mathcal{N}_{r^2 A^{-\tau}}$-integrable for all $y \in \mathcal{X}^s$ and there exists a constant $C_Z > 0$ such that*

$$\int_X \exp(-\Phi(u, y)) \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}u) \geq C_Z \quad \textit{for all } y \in \mathcal{X}^s.$$

Proof. We first show the integrability. Let $y \in \mathcal{X}^s$ be arbitrary. We use the Lipschitz continuity of $\Phi(\cdot, y)$, which holds by Proposition 5.23, to estimate

$$\int_X \exp(-\Phi(u,y))\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u) \le \exp(-\Phi(0,y)) \int_X \exp(L\|u\|_X)\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u).$$

Now $\Phi(0, y) = 0$ for all $y \in \mathcal{X}^s$ by definition of Φ and the integral on the right hand side is finite by Lemma 5.26.

Now we address the lower bound. By the Lipschitz continuity of Φ in u, the estimate

$$\begin{aligned}\int_X \exp(-\Phi(u,y))\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u) &\ge \int_X \exp(-L\|u\|_X)\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u)\\ &\ge \int_{B_1(0)} e^{-L}\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u) = e^{-L}\mathcal{N}_{r^2A^{-\tau}}(B_1(0)) =: C_Z\end{aligned}$$

holds for all $y \in \mathcal{X}^s$. By Theorem 3.6.1 in [Bogachev 1998], the topological support of the Gaussian measure $\mathcal{N}_{r^2A^{-\tau}}$ is given by the closure of its Cameron–Martin space $\mathcal{R}(A^{-\tau/2}) = \mathcal{X}^\tau$. Since $\mathcal{X}^\tau$ is dense in X the topological support is the whole space X. As a consequence, all balls in X have positive measure under $\mathcal{N}_{r^2A^{-\tau}}$, which in turn implies that the constant C_Z is positive. □

With this knowledge we can apply Bayes' formula.

Theorem 5.28. *A regular conditional distribution $(y, B) \mapsto \mu^y(B)$ of u given y exists, the posterior distribution μ^y is absolutely continuous with respect to the prior distribution $\mathcal{N}_{r^2A^{-\tau}}$ for every $y \in \mathcal{X}^s$ and has the density*

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mathcal{N}_{r^2A^{-\tau}}}(u) = \frac{1}{Z(y)}\exp(-\Phi(u,y)) \quad \mathcal{N}_{r^2A^{-\tau}}\text{-almost surely,} \tag{5.6}$$

where

$$\Phi(u,y) := \sqrt{2}\sum_{k=1}^{\infty} \frac{|(y,\varphi_k)_X - e^{-\alpha_k}(u,\varphi_k)_X| - |(y,\varphi_k)_X|}{b\alpha_k^{-\beta/2}}$$

for all $u \in X$ and $y \in \mathcal{X}^s$ and

$$Z(y) := \int_X \exp(-\Phi(u,y))\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u) \quad \text{for all } y \in \mathcal{X}^s.$$

Proof. By Proposition 5.21 the measure $P_u := \mathcal{L}_{e^{-A}u,\, b^2A^{s-\beta}}$ is absolutely continuous with respect to $\nu := \mathcal{L}_{b^2A^{s-\beta}}$ for all $u \in X$ with the density $y \mapsto p_u(y) := \exp(-\Phi(u,y))$. The function $(u, y) \mapsto p_u(y)$ is measurable by Corollary 5.24 and $Z(y) > 0$ for all $y \in \mathcal{X}^s$ by Proposition 5.27. Therefore we may apply Theorem 1.3, which yields the proposition. □

5.4.2 Stability

Now we show that Bayesian inference acts as a regularisation for the ill-posed operator equation $y = e^{-A}u$ and stabilises the problem in the sense that small changes in the data y lead to small changes in the posterior distribution μ^y. This means that introducing a prior and considering the problem from a Bayesian point of view turns it into a well-posed problem.

We use the Hellinger distance as a metric to describe the difference between two probability measures μ and μ' on a Hilbert space X. Let ν be a reference measure, such that both μ and μ' are absolutely continuous with respect to ν. Then the *Hellinger distance* is defined as

$$d_{\mathrm{Hell}}(\mu, \mu') = \left(\frac{1}{2} \int_X \left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\nu}} \right)^2 \mathrm{d}\nu \right)^{\frac{1}{2}}.$$

Lemma 5.29. *Let $\{y^n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}^s$ that converges towards $y^\dagger \in \mathcal{X}^s$. Then both*

$$\int_X \left| \exp(-\Phi(u, y^n)) - \exp(-\Phi(u, y^\dagger)) \right| \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}u) \longrightarrow 0$$

and

$$\int_X \left| \exp\left(-\frac{1}{2}\Phi(u, y^n)\right) - \exp\left(-\frac{1}{2}\Phi(u, y^\dagger)\right) \right|^2 \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}u) \longrightarrow 0$$

as $n \longrightarrow \infty$.

Proof. We define $f_n(u) := \exp(-\Phi(u, y^n))$, $f(u) := \exp(-\Phi(u, y^\dagger))$ and $g(u) := \exp(L\|u\|_X)$ for all $u \in X$ and $n \in \mathbb{N}$, where $L > 0$ denotes the joint Lipschitz constant of $\{\Phi(\cdot, y)\}_{y\in\mathcal{X}^s}$. For the sake of brevity we moreover set $\mu_0 := \mathcal{N}_{r^2 A^{-\tau}}$. For $M \geq 0$, the restriction of the exponential function to the inverval $[-M, M]$ is Lipschitz continuous with Lipschitz constant $\exp(M)$. Together with Proposition 5.23 this yields

$$\begin{aligned} |f_n(u) - f(u)| &= \left| \exp(-\Phi(u, y^n)) - \exp(-\Phi(u, y^\dagger)) \right| \\ &\leq \exp\left(L\|u\|_X\right) \left| \Phi(u, y^n) - \Phi(u, y^\dagger) \right| \end{aligned}$$

for all $u \in X$. So $f_n \longrightarrow f$ almost surely as $n \longrightarrow \infty$ by Proposition 5.23, which implies convergence in probability. Furthermore, $\{f_n\}_{n\in\mathbb{N}}$ is dominated by g, because, by Proposition 5.23,

$$|f_n(u)| = \exp\left(-\Phi(u, y^n)\right) \leq \exp\left(L\|u\|_X\right) = g(u)$$

holds for all $u \in X$ and $n \in \mathbb{N}$. This also ensures that $f_n \in L^1(X, \mu_0)$ for all $n \in \mathbb{N}$, as $g \in L^1(X, \mu_0)$ by Lemma 5.26. Now Lebesgue's dominated convergence theorem [Klenke 2014, Cor. 6.26] yields that $f_n \longrightarrow f$ in $L^1(X, \mu_0)$ (convergence in mean), i.e.,

$$\int_X |f_n(u) - f(u)| \, \mu_0(\mathrm{d}u) \longrightarrow 0,$$

and that $\{f_n\}_{n\in\mathbb{N}} = \{|f_n^{\frac{1}{2}}|^2\}_{n\in\mathbb{N}}$ is uniformly integrable. As $f_n^{\frac{1}{2}} \in L^2(X, \mu_0)$ for all $n \in \mathbb{N}$ and $f_n^{\frac{1}{2}} \longrightarrow f^{\frac{1}{2}}$ almost surely, Theorem 7.3 in [Klenke 2014] yields $f_n^{\frac{1}{2}} \longrightarrow f^{\frac{1}{2}}$ in $L^2(X, \mu_0)$ (convergence in mean square), i.e.

$$\int_X \left| f_n(u)^{\frac{1}{2}} - f(u)^{\frac{1}{2}} \right|^2 \mu_0(\mathrm{d}u) \longrightarrow 0. \qquad \square$$

Theorem 5.30. *Let $\{y^n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}^s$ that converges towards $y^\dagger \in \mathcal{X}^s$. Then the associated posterior measures μ^{y^n} converge towards $\mu^{y^\dagger}$ with respect to the Hellinger distance, i.e.,*

$$d_{Hell}(\mu^{y^n}, \mu^{y^\dagger}) \to 0.$$

Proof. Again, we set $\mu_0 := \mathcal{N}_{r^2 A^{-\tau}}$. As both $\mu^{y^n} \ll \mu_0$ and $\mu^{y^\dagger} \ll \mu_0$, we can express the (squared) Hellinger distance as

$$\begin{aligned} d_{\mathrm{Hell}}(\mu^{y^n}, \mu^{y^\dagger})^2 &= \frac{1}{2}\int_X \left(\sqrt{\frac{\mathrm{d}\mu^{y^n}}{\mathrm{d}\mu_0}} - \sqrt{\frac{\mathrm{d}\mu^{y^\dagger}}{\mathrm{d}\mu_0}}\right)^2 \mathrm{d}\mu_0 \\ &= \frac{1}{2}\int_X \left(\frac{1}{Z(y^n)^{\frac{1}{2}}}\exp\left(-\Phi(u,y^n)\right)^{\frac{1}{2}} - \frac{1}{Z(y^\dagger)^{\frac{1}{2}}}\exp\left(-\Phi(u,y^\dagger)\right)^{\frac{1}{2}}\right)^2 \mu_0(\mathrm{d}u) \end{aligned}$$

Now we abbreviate $f_n(u) := \exp(-\Phi(u, y^n))$, $f(u) := \exp(-\Phi(u, y^\dagger))$ as before and set

$$Z_n := Z(y^n) = \int_X f_n \mathrm{d}\mu_0, \qquad Z := Z(y^\dagger) = \int_X f \mathrm{d}\mu_0.$$

We use $(a+b)^2 \le 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$ to obtain

$$\begin{aligned} \frac{1}{2}\left(\frac{f_n^{\frac{1}{2}}}{Z_n^{\frac{1}{2}}} - \frac{f^{\frac{1}{2}}}{Z^{\frac{1}{2}}}\right)^2 &= \frac{1}{2}\left(\frac{1}{Z_n^{\frac{1}{2}}}\left(f_n^{\frac{1}{2}} - f^{\frac{1}{2}}\right) - \left(\frac{1}{Z_n^{\frac{1}{2}}} + \frac{1}{Z^{\frac{1}{2}}}\right) f^{\frac{1}{2}}\right)^2 \\ &\le \frac{1}{Z_n}\left(f_n^{\frac{1}{2}} - f^{\frac{1}{2}}\right)^2 + \left(\frac{1}{Z_n^{\frac{1}{2}}} + \frac{1}{Z^{\frac{1}{2}}}\right)^2 f, \end{aligned}$$

which results in

$$\begin{aligned} d_{\mathrm{Hell}}(\mu^{y^n}, \mu^{y^\dagger})^2 = &\frac{1}{Z_n}\int_X \left(f_n(u)^{\frac{1}{2}} - f(u)^{\frac{1}{2}}\right)^2 \mu_0(\mathrm{d}u) \\ &+ \left(\frac{1}{Z_n^{\frac{1}{2}}} - \frac{1}{Z^{\frac{1}{2}}}\right)^2 \int_X f(u)\mu_0(\mathrm{d}u) =: I_1 + I_2. \end{aligned}$$

Now Lemma 5.29 implies that $Z_n - Z = \int_X (f_n - f)\mathrm{d}\mu_0 \to 0$ and consequently,

$$I_2 = \left(\frac{1}{Z_n^{\frac{1}{2}}} - \frac{1}{Z^{\frac{1}{2}}}\right)^2 Z \to 0.$$

Moreover, Lemma 5.29 shows that $I_1 \to 0$ as well. □

By interpreting Theorem 5.30 appropriately, we can conclude that approximating the log-likelihood Φ also results in small changes in the posterior distribution μ^y. For every $N \in \mathbb{N}$ let P_N denote the orthogonal projection onto the subspace $\operatorname{span}\{\varphi_1, \dots, \varphi_N\} \subset \mathcal{X}^s$, defined by

$$P_N y := \sum_{k=1}^N (y, \varphi_k)_X\, \varphi_k = \sum_{k=1}^N \left(y, \alpha_k^{-\frac{s}{2}}\varphi_k\right)_{\mathcal{X}^s} \alpha_k^{-\frac{s}{2}}\varphi_k \quad \text{for all } y \in \mathcal{X}^s.$$

We consider finite approximations Φ^N of the log-likelihood, defined by

$$\Phi^N(u,y) := \Phi(u, P_N y) = \frac{\sqrt{2}}{b} \sum_{k=1}^{N} \alpha_k^{\frac{\beta}{2}} \left(|(y,\varphi_k)_X - e^{-\alpha_k}(u,\varphi_k)_X| - |(y,\varphi_k)_X| \right).$$

for all $N \in \mathbb{N}$, $u \in X$ and $y \in \mathcal{X}^s$. For every $y \in \mathcal{X}^s$ the sequence $\{y^N\}_{N\in\mathbb{N}}$, defined by $y^N := P_N y$ for all $N \in \mathbb{N}$, converges towards y in $\mathcal{X}^s$. Therefore Theorem 5.30 tells us that $d_{\mathrm{Hell}}(\mu^{y^N}, \mu^y) \to 0$ as $N \to \infty$. We can, however, interpret μ^{y^N} as the posterior distribution μ^N resulting from an approximation of Φ instead of an approximation of y, since

$$\begin{aligned} \frac{\mathrm{d}\mu^{y^N}}{\mathrm{d}\mathcal{N}_{r^2 A^{-\tau}}}(u) &= \frac{\exp(-\Phi(u, y^N))}{\int_X \exp(-\Phi(\tilde{u}, y^N)) \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}\tilde{u})} \\ &= \frac{\exp(-\Phi^N(u, y))}{\int_X \exp(-\Phi^N(\tilde{u}, y)) \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}\tilde{u})} =: \frac{\mathrm{d}\mu^N}{\mathrm{d}\mathcal{N}_{r^2 A^{-\tau}}}(u) \end{aligned}$$

$\mathcal{N}_{r^2 A^{-\tau}}$-almost surely. This tells us that Bayesian inference is stable with respect to such changes in the model.

We can also use the Hellinger distance to make statements about the closeness of expectations. For any function $f \in L^2(X, \mu^y) \cap L^2(X, \mu^z)$, a small Hellinger distance implies that the expectations of f under μ^y and μ^z are close.

Lemma 5.31. *For all $f \in L^2(X, \mu^y) \cap L^2(X, \mu^z)$ and $y, z \in \mathcal{X}^s$ we have*

$$\left| \mathbb{E}^{\mu^y} f(u) - \mathbb{E}^{\mu^z} f(u) \right| \le 2\sqrt{2} \left(\mathbb{E}^{\mu^y} |f(u)|^2 + \mathbb{E}^{\mu^z} |f(u)|^2 \right)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu^y, \mu^z).$$

Proof. By means of the Cauchy-Schwarz inequality and $(a+b)^2 \le 2a^2 + 2b^2$, which holds for all $a, b \in \mathbb{R}$, we compute

$$\begin{aligned} \left| \mathbb{E}^{\mu^y} f(u) - \mathbb{E}^{\mu^z} f(u) \right| &= \left| \int_X f \, \mathrm{d}\mu^y - \int_X f \, \mathrm{d}\mu^z \right| = \left| \int_X f \left(\frac{\mathrm{d}\mu^y}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu^z}{\mathrm{d}\nu} \right) \mathrm{d}\nu \right| \\ &= \left| \int_X f \left(\sqrt{\frac{\mathrm{d}\mu^y}{\mathrm{d}\nu}} + \sqrt{\frac{\mathrm{d}\mu^z}{\mathrm{d}\nu}} \right) \left(\sqrt{\frac{\mathrm{d}\mu^y}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu^z}{\mathrm{d}\nu}} \right) \mathrm{d}\nu \right| \\ &\le \sqrt{2} \left(\int_X |f|^2 \left(\sqrt{\frac{\mathrm{d}\mu^y}{\mathrm{d}\nu}} + \sqrt{\frac{\mathrm{d}\mu^z}{\mathrm{d}\nu}} \right)^2 \mathrm{d}\nu \right)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu^y, \mu^z) \\ &\le 2\sqrt{2} \left(\mathbb{E}^{\mu^y} |f(u)|^2 + \mathbb{E}^{\mu^z} |f(u)|^2 \right)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu^y, \mu^z) \end{aligned}$$

□

5.5 Consistency of the Maximum A Posteriori Estimator

Here we determine the unique mode of the posterior distribution, use it in Subsection 5.5.1 to define an estimator for the posterior, the MAP estimator, and consider its consistency in a frequentist setting in Subsections 5.5.2 to 5.5.6. In Subsection 5.5.3, we establish an a priori parameter choice strategy for which the MAP estimator is asymptotically unbiased, i.e., under which its expectation converges towards a true solution, and, in Subsection 5.5.5, a convergence rate for the bias in case that a source condition is satisfied. Subsequently, we prove a convergence rate for the mean squared error under a source condition and compare it with the optimal rate in case of Gaussian noise in Subsection 5.5.6.

5.5.1 Derivation and Basic Properties

We apply the results from Chapter 4 to the posterior measure μ^y derived in Section 5.4. We determine the Onsager–Machlup functional of μ^y and obtain the MAP estimator $\hat{u}_{\mathrm{MAP}}$ by minimising it.

The Cameron–Martin space of the Gaussian measure $\mathcal{N}_{r^2A^{-\tau}}$ is given by $\frac{1}{r}A^{-\tau/2}(X)$, equipped with the norm $\frac{1}{r}\|A^{-\tau/2}\cdot\|_X$. Using Proposition 5.10, we find that $\frac{1}{r}A^{-\tau/2}(X) = \mathcal{X}^\tau$.

Theorem 5.32. *For every $y \in \mathcal{X}^s$ the functional $I^y : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$, defined by*

$$\begin{aligned} I^y(u) &:= \Phi(u, y) + \frac{1}{2r^2}\|u\|^2_{\mathcal{X}^\tau} \\ &= \frac{\sqrt{2}}{b}\sum_{k=1}^{\infty} \alpha_k^{\frac{\beta}{2}} \left(|(y, \varphi_k)_X - e^{-\alpha_k}(u, \varphi_k)_X| - |(y, \varphi_k)_X|\right) + \frac{1}{2r^2}\sum_{k=1}^{\infty} \alpha_k^\tau\, |u_k|^2 \end{aligned} \tag{5.7}$$

for all $u \in \mathcal{X}^\tau$ and $I^y(u) = \infty$ for $u \in X \setminus \mathcal{X}^\tau$ is the Onsager–Machlup functional of μ^y.

Proof. Let $y \in \mathcal{X}^s$. By Theorem 5.28, the density of the posterior distribution μ^y w.r.t. the prior distribution $\mathcal{N}_{r^2A^{-\tau}}$ is given by (5.6). Φ is Lipschitz continuous in u by Proposition 5.23, so that Assumption 4.3 is satisfied by Lemma 4.9. Consequently, $u \mapsto \Phi(u, y) + \frac{1}{2r^2}\|u\|^2_{\mathcal{X}^\tau}$ is the Onsager–Machlup functional of μ^y by Theorem 4.4. □

Corollary 5.33. *Let $y \in \mathcal{X}^s$. Then the functional I^y, defined by* (5.7), *has a minimiser $\bar{u} \in \mathcal{X}^\tau$ and every minimiser of I^y is a MAP estimate. Conversely, every MAP estimate minimises I^y.*

Proof. By Theorem 5.28, the density of μ^y w.r.t. $\mathcal{N}_{r^2A^{-\tau}}$ is given by (5.6). Now Corollary 4.16 tells us that the minimisers of I^y are precisely the MAP estimates for μ^y as Φ is Lipschitz continuous in u by Proposition 5.23. □

Corollary 5.33 tells us in particular that every MAP estimate lies in $\mathcal{X}^\tau$. We express the minimiser explicitly and show that it is unique.

Lemma 5.34. *Let $y \in \mathcal{X}^s$ and let $\bar{u} = \bar{u}(y) \in \mathcal{X}^\tau$ be a minimiser of I^y. Then $\bar{u} = \sum_{k=1}^\infty \bar{u}_k\varphi_k$, where*

$$\bar{u}_k = \max\left\{-\frac{r^2}{b}R_k, \min\left\{e^{\alpha_k}(y, \varphi_k)_X, \frac{r^2}{b}R_k\right\}\right\}$$

and

$$R_k := \sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau}e^{-\alpha_k}$$

for all $k \in \mathbb{N}$. In particular, the minimiser $\bar{u}$ of I^y is unique.

Proof. For all $k \in \mathbb{N}$ and $u_k \in \mathbb{R}$, we define

$$f_k(u_k) = \frac{1}{2r^2}\alpha_k^\tau |u_k|^2 + \frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}}\left(|e^{-\alpha_k}u_k - (y,\varphi_k)_X| - |(y,\varphi_k)_X|\right).$$

This way, $I^y(u) = \sum_{k=1}^\infty f_k((u,\varphi_k)_X)$. As a minimiser of I^y, $\bar{u}$ satisfies

$$0 \le I^y(\bar{u} + t\varphi_k) - I^y(\bar{u}) = f_k((\bar{u},\varphi_k)_X + t) - f_k((\bar{u},\varphi_k)_X)$$

for all $t \in \mathbb{R}$ and $k \in \mathbb{N}$. Hence $\bar{u}_k := (\bar{u},\varphi_k)_X$ minimises f_k for every $k \in \mathbb{N}$.

Consider an arbitrary, fixed $k \in \mathbb{N}$. The function f_k is continuous on $\mathbb{R}$ and continuously differentiable on $\mathbb{R} \setminus \{e^{\alpha_k}(y,\varphi_k)_X\}$ with

$$f_k'(u_k) = \begin{cases} \frac{1}{r^2}\alpha_k^\tau u_k - \frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}}e^{-\alpha_k} & \text{if } u_k < e^{\alpha_k}(y,\varphi_k)_X, \\ \frac{1}{r^2}\alpha_k^\tau u_k + \frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}}e^{-\alpha_k} & \text{if } u_k > e^{\alpha_k}(y,\varphi_k)_X. \end{cases}$$

Set

$$S_k := \frac{r^2}{b}R_k = \frac{r^2}{b}\sqrt{2}\alpha_k^{\beta/2-\tau}e^{-\alpha_k}.$$

In case $e^{\alpha_k}(y,\varphi_k)_X > S_k$, $\bar{u}_k = S_k$ is the unique minimiser of f_k, because $f_k'(S_k) = 0$,

$$\begin{aligned} f_k'(u_k) &< 0 && \text{for } u_k \in (-\infty, S_k), \text{ and} \\ f_k'(u_k) &> 0 && \text{for } u_k \in (S_k,\infty) \setminus \{e^{\alpha_k}(y,\varphi_k)\}. \end{aligned}$$

In case $e^{\alpha_k}(y,\varphi_k)_X < -S_k$, $\bar{u}_k = -S_k$ is the unique minimiser of f_k, because $f_k'(-S_k) = 0$,

$$\begin{aligned} f_k'(u_k) &< 0 && \text{for } u_k \in (-\infty, -S_k) \setminus \{e^{\alpha_k}(y,\varphi_k)\}, \text{ and} \\ f_k'(u_k) &> 0 && \text{for } u_k \in (-S_k,\infty). \end{aligned}$$

Finally, in case $e^{\alpha_k}(y,\varphi_k) \in [-S_k, S_k]$,

$$\begin{aligned} f_k'(u_k) &< 0 && \text{if } u_k < e^{\alpha_k}(y,\varphi_k)_X, \text{ and} \\ f_k'(u_k) &> 0 && \text{if } u_k > e^{\alpha_k}(y,\varphi_k)_X, \end{aligned}$$

so that the unique minimiser of f_k is given by $\bar{u}_k = e^{\alpha_k}(y,\varphi_k)_X$. □

By Corollary 5.33 and Lemma 5.34, I^y has a unique minimiser $\bar{u} = \bar{u}(y)$ for every $y \in \mathcal{X}^s$, which at the same time is a MAP estimate for μ^y. With this knowledge, we can define a unique MAP estimator.

Definition 5.35. We define the *maximum a posteriori (MAP) estimator* $\hat{u}_{\mathrm{MAP}}\colon \mathcal{X}^s \to X$ by assigning to every $y \in \mathcal{X}^s$ the respective MAP estimate $\hat{u}_{\mathrm{MAP}}(y) := \bar{u}(y)$ for μ^y.

Equivalently, we can express $\hat{u}_{\mathrm{MAP}}$ as

$$\hat{u}_{\mathrm{MAP}}(y) = \operatorname*{arg\,min}_{u \in \mathcal{X}^\tau} I^y(u) = \operatorname*{arg\,min}_{u \in \mathcal{X}^\tau}\left\{\Phi(u,y) + \frac{1}{2r^2}\|u\|_{\mathcal{X}^\tau}^2\right\}$$

for all $y \in \mathcal{X}^s$. This means that the MAP estimator can be viewed as Tikhonov–Phillips regularisation with disrepancy term $\Phi(u, y)$ and penalty term $\frac{1}{2r^2}\|u\|^2_{\mathcal{X}^s}$.

We can formally interpret Lemma 5.34 by reformulating the minimiser $\bar{u}$ of I^y as

$$\bar{u}(y) = e^A P_{\frac{r^2}{b}Q}(y) \quad \text{for all } y \in \mathcal{X}^s,$$

where P_S denotes the metric projection onto $S \subset \mathcal{X}^s$ characterised by

$$\|P_S(y) - y\|_{\mathcal{X}^s} = \inf_{z \in S}\|z - y\|_{\mathcal{X}^s},$$

and $Q \subset \mathcal{X}^s$ is the convex closed set

$$Q := \{y \in \mathcal{X}^s : (y, \varphi_k)_X \le e^{-\alpha_k} R_k \text{ for all } k \in \mathbb{N}\}.$$

So the MAP estimator acts on the data y by projecting it onto a hyperrectangle and then applying the inverse of the forward operator e^{-A}.

Next, we show that $\hat{u}_{\mathrm{MAP}}$ is continuous, so that in particular a finite dimensional approximation of the data leads to a close MAP estimate.

Lemma 5.36. *The series*

$$\sum_{k=1}^{\infty} R_k^2 = \sum_{k=1}^{\infty} 2\alpha_k^{\beta-2\tau} e^{-2\alpha_k}$$

converges.

Proof. By Lemma 5.14, $\alpha_k^{\beta-2\tau} e^{-\alpha_k} \le (\beta - 2\tau)^{\beta-2\tau} e^{2\tau-\beta} =: C_{\tau,\beta}$. We define the monotonically decreasing dominating function $\kappa \mapsto e^{-C_-\kappa^{\frac{2}{d}}}$ on $\mathbb{R}$ and use it to estimate

$$\begin{aligned}\sum_{k=1}^{\infty} e^{-\alpha_k} &\le \sum_{k=1}^{\infty} e^{-C_- k^{\frac{2}{d}}} \le \int_0^{\infty} e^{-C_-\kappa^{\frac{2}{d}}}\,\mathrm{d}\kappa \\ &= \int_0^{\infty} e^{-t} C_-^{-\frac{d}{2}} \frac{d}{2} t^{\frac{d}{2}-1}\mathrm{d}t = C_-^{-\frac{d}{2}}\frac{d}{2}\Gamma\left(\frac{d}{2}\right) =: C_d,\end{aligned}$$

where we substituted $t = C_-\kappa^{2/d}$. We conclude the proof by combining these estimates to

$$\sum_{k=1}^{\infty} 2\alpha_k^{\beta-2\tau} e^{-2\alpha_k} = 2\sum_{k=1}^{\infty}\left(\alpha_k^{\beta-2\tau} e^{-\alpha_k}\right) e^{-\alpha_k} \le 2C_{\tau,\beta}C_d < \infty. \qquad \square$$

Theorem 5.37. *The MAP estimator $\hat{u}_{\mathrm{MAP}} \colon \mathcal{X}^s \to X$ according to Definition 5.35 is continuous.*

Proof. Let $\varepsilon > 0$. By Lemma 5.34, we have

$$(\hat{u}_{\mathrm{MAP}}(y), \varphi_k)_X = \max\left\{-\frac{r^2}{b}R_k, \min\left\{\alpha_k^{-\frac{s}{2}} e^{\alpha_k}(y, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s}, \frac{r^2}{b}R_k\right\}\right\}$$

for all $k \in \mathbb{N}$, where $R_k = \sqrt{2}\alpha_k^{\beta/2-\tau} e^{-\alpha_k}$. We make the fundamental estimate

$$|(\hat{u}_{\mathrm{MAP}}(y), \varphi_k)_X - (\hat{u}_{\mathrm{MAP}}(z), \varphi_k)_X| \le \min\left\{\alpha_k^{-\frac{s}{2}} e^{\alpha_k}\left|(y, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s} - (z, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s}\right|, 2\frac{r^2}{b}R_k\right\}$$

for all $y, z \in \mathcal{X}^s$ and $k \in \mathbb{N}$, which leads to

$$\begin{aligned} \|\hat{u}_{\mathrm{MAP}}(y) - \hat{u}_{\mathrm{MAP}}(z)\|_X^2 &= \sum_{k=1}^{\infty} |(\hat{u}_{\mathrm{MAP}}(y) - \hat{u}_{\mathrm{MAP}}(z), \varphi_k)_X|^2 \\ &\leq \sum_{k=1}^{N} \alpha_k^{-s} e^{2\alpha_k} \left| (y, \alpha_k^{-\frac{s}{2}} \varphi_k)_{\mathcal{X}^s} - (z, \alpha_k^{-\frac{s}{2}} \varphi_k)_{\mathcal{X}^s} \right|^2 + \sum_{k=N+1}^{\infty} 2 \frac{r^2}{b} R_k^2. \end{aligned}$$

for all $N \in \mathbb{N}$.

As $\sum_{k=1}^{\infty} R_k^2 < \infty$ by Lemma 5.36, we can choose $N \in \mathbb{N}$ large enough, such that

$$\sum_{k=N+1}^{\infty} 2 \frac{r^2}{b} R_k^2 \leq \frac{\varepsilon^2}{2}.$$

Next, we set $M := \max_{k=1,\dots,N} \alpha_k^{-s/2} e^{\alpha_k}$ and choose $\delta := \frac{\varepsilon}{2M}$. Since $\{\alpha_k^{-s/2} \varphi_k\}_{k \in \mathbb{N}}$ is an orthonormal basis of $\mathcal{X}^s$ by Lemma 5.11, we arrive at

$$\|\hat{u}_{\mathrm{MAP}}(y) - \hat{u}_{\mathrm{MAP}}(z)\|_X^2 \leq M^2 \|y - z\|_{\mathcal{X}^s}^2 + \frac{\varepsilon^2}{2} \leq \frac{\varepsilon^2}{4} + \frac{\varepsilon^2}{2} \leq \varepsilon.$$

for all $y, z \in \mathcal{X}^s$ with $\|y - z\|_{\mathcal{X}^s} \leq \delta$. □

Corollary 5.38. *Let $y \in \mathcal{X}^s$ and set $y^N := \sum_{k=1}^{N} (y, \varphi_k)_X \varphi_k$ for all $N \in \mathbb{N}$. Then*

$$\hat{u}_{\mathrm{MAP}}(y^N) \to \hat{u}_{\mathrm{MAP}}(y) \quad \text{as } N \to \infty.$$

Proof. We verify that $y^N \to y$ in $\mathcal{X}^s$. This is indeed the case, because

$$y^N = \sum_{k=1}^{N} (y, \varphi_k)_X \varphi_k = \sum_{k=1}^{N} (y, \alpha_k^{-\frac{s}{2}} \varphi_k)_{\mathcal{X}^s} \alpha_k^{-\frac{s}{2}} \varphi_k$$

and $\{\alpha_k^{-s/2} \varphi_k\}_{k \in \mathbb{N}}$ is an orthonormal basis of $\mathcal{X}^s$ by Lemma 5.11. □

5.5.2 Frequentist Setting

Although we derived the MAP estimator from the Bayesian setting, i.e., from the posterior distribution μ^y, we will consider its consistency in a frequentist setting. Instead of a prior distribution we now assume that there is a deterministic true solution $u^\dagger \in X$ and only the noise η is stochastic with the same distribution as before. The data $y = e^{-A} u^\dagger + \eta$ is then a Laplacian random variable with distribution

$$y \sim \mathcal{L}_{e^{-A} u^\dagger, b^2 A^{s-\beta}}$$

on $\mathcal{X}^s$.

We will study the MAP estimator $\hat{u}_{\mathrm{MAP}}$ when it is applied to data with this distribution. In this setting, $\hat{u}_{\mathrm{MAP}}(y)$ itself is a random variable. In particular, we want to show that $\hat{u}_{\mathrm{MAP}}(y)$

converges towards the true solution $u^\dagger$ in some sense as b, and with it the variance of the noise, tends to zero.

For every $y \in \mathcal{X}^s$ the MAP estimate $\hat{u}_{\mathrm{MAP}}(y)$ is given as the minimiser of the functional $I^y(u) = \Phi(u, y) + \frac{1}{2r^2}\|u\|^2_{\mathcal{X}^\tau}$ by Corollary 5.33. We may, however, minimise a scaled and shifted version of I^y instead, since multiplying a functional by a positive number and adding a constant does not change its minimisers. We can consider $\Phi(u, y)$ as a discrepancy term and by multiplying I^y by b we obtain a functional whose discrepancy term is independent of b. Additionally, we add a constant to bI^y, chosen in such a way that the discrepancy term is equal to zero for $u = u^\dagger$ and $y = e^{-A}u^\dagger$. This leads to the new objective functional $J^y \colon X \to \mathbb{R}$,

$$\begin{aligned} J^y(u) &:= bI^y(u) - b\Phi(u^\dagger, e^{-A}u^\dagger) \\ &= b\Phi(u, y) - b\Phi(u^\dagger, e^{-A}u^\dagger) + \frac{b}{2r^2}\|u\|^2_{\mathcal{X}^\tau}. \end{aligned}$$

Again, J^y decomposes into a series,

$$\begin{aligned} J^y(u) = \sqrt{2}\sum_{k=1}^{\infty} \alpha_k^{\frac{\beta}{2}} \left(|(y, \varphi_k)_X - e^{-\alpha_k}(u, \varphi_k)_X| - |(y, \varphi_k)_X| + |e^{-\alpha_k}(u^\dagger, \varphi_k)_X|\right) \\ + \frac{b}{2r^2}\sum_{k=1}^{\infty} \alpha_k^\tau \, |u_k|^2 . \end{aligned}$$

Now we set

$$\Psi_k(u, y) := \sqrt{2}\alpha_k^{\frac{\beta}{2}} \Big(|(y, \varphi_k)_X - e^{-\alpha_k}(u, \varphi_k)_X| - |(y, \varphi_k)_X| + |e^{-\alpha_k}(u^\dagger, \varphi_k)_X|\Big)$$

and $\Psi(u, y) := \sum_{k=1}^\infty \Psi_k(u, y)$ for all $u \in X$, $y \in \mathcal{X}^s$ and $k \in \mathbb{N}$. This way,

$$J^y(u) = \Psi(u, y) + \frac{b}{2r^2}\|u\|^2_{\mathcal{X}^\tau},$$

and the new discrepancy term $\Psi(u, y)$ does not depend on b. Moreover, $\Psi(u^\dagger, \cdot)$ is nonnegative and $\Psi(u^\dagger, e^{-A}u^\dagger) = 0$.

The functional J^y can, up to a constant, be interpreted as Onsager–Machlup functional of the posterior distribution resulting from unscaled Laplacian noise $\tilde{\eta} \sim \mathcal{L}_{A^{-\beta}}$ in combination with a scaled Gaussian prior $\tilde{u} \sim \mathcal{N}_{\frac{r^2}{b}A^{-\tau}}$, and $\Psi(\tilde{u}, \tilde{y})$ as the negative log-likelihood of the resulting data $\tilde{y}$ given $\tilde{u}$.

5.5.3 Asymptotic Unbiasedness

First, we examine the convergence of the expectation $\mathbb{E}\left[\hat{u}_{\mathrm{MAP}}(y)\right]$ of the MAP estimator towards the true solution $u^\dagger$ as $b \to 0$. To this end, we show that the expectation of the discrepancy term $\Psi(u^\dagger, y)$ converges to 0 and prove an inequality for the expectations of $\Psi(\hat{u}_{\mathrm{MAP}}(y), e^{-A}u^\dagger)$ and $\Psi(\hat{u}_{\mathrm{MAP}}(y), y)$. In a first step, we compute the mean and the variance of $\Psi(u^\dagger, y)$.

Lemma 5.39. *Let $u^\dagger \in X$, $b > 0$ and $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{s-\beta}}$. Then*

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] = b \sum_{k=1}^{\infty} \left(1 - \exp\left(-\frac{c_k}{b}\right)\right)$$

and

$$\operatorname{Var}\left(\Psi(u^\dagger, y)\right) = b^2 \sum_{k=1}^{\infty} \left(4\left(1 - e^{-\frac{c_k}{b}} - \frac{c_k}{b} e^{-\frac{c_k}{b}}\right) - \left(1 - e^{-\frac{c_k}{b}}\right)^2\right),$$

where

$$c_k := \sqrt{2}\alpha_k^{\frac{\beta}{2}} e^{-\alpha_k} \left|(u^\dagger, \varphi_k)_X\right| \quad \text{for all } k \in \mathbb{N}.$$

Proof. Let $e_k := \alpha_k^{-s/2}\varphi_k$, $k \in \mathbb{N}$, denote the orthonormal basis of $\mathcal{X}^s$ from Lemma 5.11. First, we note that due to the independence of the components $(y, e_k)_{\mathcal{X}^s}$ of y, the summands $\Psi_k(u^\dagger, y)$, which can be written as

$$\Psi_k(u^\dagger, y) = \sqrt{2}\alpha_k^{\frac{\beta-s}{2}} \left(\left|(y, e_k)_{\mathcal{X}^s} - (e^{-A}u^\dagger, e_k)_{\mathcal{X}^s}\right| - |(y, e_k)_{\mathcal{X}^s}| + \left|(e^{-A}u^\dagger, e_k)_{\mathcal{X}^s}\right|\right),$$

are independent, too. This allow us to write

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \Psi_k(u^\dagger, y)\right] = \sum_{k=1}^{\infty} \mathbb{E}\left[\Psi_k(u^\dagger, y)\right].$$

By construction of the Laplacian measure, each component $(y, e_k)_{\mathcal{X}^s} = p_k \circ y$ is distributed according to

$$\mathcal{L}_{(e^{-A}u^\dagger, e_k)_{\mathcal{X}^s}, b^2\alpha_k^{s-\beta}} = \mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}} \circ p_k^{-1},$$

the pushforward of $\mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}$ under the projection $p_k = (\cdot, e_k)_{\mathcal{X}^s}$. We set

$$a := (e^{-A}u^\dagger, e_k)_{\mathcal{X}^s}, \quad \lambda := b^2\alpha_k^{s-\beta}$$

and substitute $x = (y, e_k)_{\mathcal{X}^s}$ in order to compute

$$\begin{aligned}
\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] &= \int_{\mathcal{X}^s} b\sqrt{\frac{2}{\lambda}} \left(\left|(y, e_k)_{\mathcal{X}^s} - a\right| - \left|(y, e_k)_{\mathcal{X}^s}\right| + |a|\right) \mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}(\mathrm{d}y) \\
&= \int_{\mathbb{R}} b\sqrt{\frac{2}{\lambda}} \left(|x - a| - |x| + |a|\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x)
\end{aligned}$$

By substituting a by $-a$ and x by $-x$ in case that $a < 0$, we obtain

$$\begin{aligned}
\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] &= \int_{\mathbb{R}} b\sqrt{\frac{2}{\lambda}} \left(|x - |a|| - |x| + |a|\right) \mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&= \int_{-\infty}^{0} b\sqrt{\frac{2}{\lambda}} 2\,|a|\,\mathcal{L}_{|a|,\lambda}(\mathrm{d}x) + \int_{|a|}^{\infty} 0\mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&\quad + \int_{0}^{|a|} b\sqrt{\frac{2}{\lambda}} 2\left(|a| - x\right) \mathcal{L}_{|a|,\lambda}(\mathrm{d}x).
\end{aligned}$$

The first integral computes as

$$\begin{aligned}
\int_{-\infty}^{0} b\sqrt{\frac{2}{\lambda}}2\,|a|\,\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x &= b\sqrt{\frac{2}{\lambda}}\,|a|\int_{-\infty}^{0}\sqrt{\frac{2}{\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x \\
&= b\sqrt{\frac{2}{\lambda}}\,|a|\left[e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]_{-\infty}^{0} \\
&= b\sqrt{\frac{2}{\lambda}}\,|a|\,e^{-\sqrt{\frac{2}{\lambda}}|a|}-0.
\end{aligned}$$

The last integral computes as

$$\begin{aligned}
&b\sqrt{\frac{2}{\lambda}}\int_{0}^{|a|}(|a|-x)\sqrt{\frac{2}{\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x \\
&= b\sqrt{\frac{2}{\lambda}}\left[(|a|-x)\,e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]_{0}^{|a|}-b\sqrt{\frac{2}{\lambda}}\int_{0}^{|a|}(-1)e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x \\
&= 0-b\sqrt{\frac{2}{\lambda}}\,|a|\,e^{-\sqrt{\frac{2}{\lambda}}|a|}+b\left[e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]_{0}^{|a|} \\
&= -b\sqrt{\frac{2}{\lambda}}\,|a|\,e^{-\sqrt{\frac{2}{\lambda}}|a|}+b\left(1-e^{-\sqrt{\frac{2}{\lambda}}|a|}\right).
\end{aligned}$$

We notice that

$$\sqrt{\frac{2}{\lambda}}\,|a| = \frac{1}{b}\sqrt{2}\alpha_k^{\frac{\beta-s}{2}}\left|\left(e^{-A}u^\dagger, e_k\right)_{X^s}\right| = \frac{1}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}}e^{-\alpha_k}\left|\left(u^\dagger, \varphi_k\right)_X\right| = \frac{c_k}{b}.$$

Summing up and resubstituting finally results in

$$\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] = b\left(1-e^{-\frac{c_k}{b}}\right).$$

Now we turn towards

$$\begin{aligned}
\mathbb{E}\left[\Psi_k(u^\dagger, y)^2\right] &= \int_{\mathbb{R}} b^2\frac{2}{\lambda}\left(|x-|a||-|x|+|a|\right)^2\mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&= \int_{-\infty}^{0} b^2\frac{2}{\lambda}4\,|a|^2\,\mathcal{L}_{|a|,\lambda}(\mathrm{d}x)+\int_{0}^{|a|} b^2\frac{2}{\lambda}4\,(x-|a|)^2\,\mathcal{L}_{|a|,\lambda}(\mathrm{d}x).
\end{aligned}$$

Here the first integral computes as

$$\int_{-\infty}^{0} b^2\frac{8}{\lambda}\,|a|^2\,\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x = b^2\frac{8}{\lambda}\,|a|^2\left[\frac{1}{2}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]\Big|_{-\infty}^{0} = b^2\frac{4}{\lambda}\,|a|^2\,e^{-\sqrt{\frac{2}{\lambda}}|a|},$$

and the second one as

$$
\begin{aligned}
&\int_0^{|a|} b^2 \frac{8}{\lambda} (x-|a|)^2 \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)} \mathrm{d}x \\
&= b^2 \frac{8}{\lambda} \left[(x-|a|)^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)} \right]\Bigg|_0^{|a|} - \int_0^{|a|} b^2 \frac{8}{\lambda} (x-|a|) \, e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)} \mathrm{d}x \\
&= -b^2 \frac{4}{\lambda} |a|^2 \, e^{-\sqrt{\frac{2}{\lambda}}|a|} - 4b^2 \sqrt{\frac{2}{\lambda}} \left[(x-|a|) \, e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)} \right]\Bigg|_0^{|a|} \\
&\quad + \int_0^{|a|} 4b^2 \sqrt{\frac{2}{\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)} \mathrm{d}x \\
&= -b^2 \frac{4}{\lambda} |a|^2 \, e^{-\sqrt{\frac{2}{\lambda}}|a|} - 4b^2 \sqrt{\frac{2}{\lambda}} \, |a| \, e^{-\sqrt{\frac{2}{\lambda}}|a|} + 4b^2 \left(1 - e^{-\sqrt{\frac{2}{\lambda}}|a|} \right).
\end{aligned}
$$

Summing up and resubstituting yields

$$
\begin{aligned}
\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] &= 4b^2 \left(1 - e^{-\sqrt{\frac{2}{\lambda}}|a|} - \sqrt{\frac{2}{\lambda}} \, |a| \, e^{-\sqrt{\frac{2}{\lambda}}|a|} \right) \\
&= 4b^2 \left(1 - e^{-\frac{c_k}{b}} - \frac{c_k}{b} e^{-\frac{c_k}{b}} \right).
\end{aligned}
$$

With this, we can compute

$$
\begin{aligned}
\operatorname{Var}\left(\Psi_k(u^\dagger, y)\right) &= \mathbb{E}\left[\Psi_k(u^\dagger, y)^2\right] - \mathbb{E}\left[\Psi_k(u^\dagger, y)\right]^2 \\
&= 4b^2 \left(1 - e^{-\frac{c_k}{b}} - \frac{c_k}{b} e^{-\frac{c_k}{b}} \right) - b^2 \left(1 - e^{-\frac{c_k}{b}} \right)^2
\end{aligned}
$$

Because the components $(y, e_k)_{X^s}$ of y are independent, we can write

$$
\begin{aligned}
\operatorname{Var}\left(\Psi(u^\dagger, y)\right) &= \operatorname{Var}\left(\sum_{k=1}^{\infty} \Psi_k(u^\dagger, y) \right) = \sum_{k=1}^{\infty} \operatorname{Var}\left(\Psi_k(u^\dagger, y)\right) \\
&= b^2 \sum_{k=1}^{\infty} \left(4 \left(1 - e^{-\frac{c_k}{b}} - \frac{c_k}{b} e^{-\frac{c_k}{b}} \right) - \left(1 - e^{-\frac{c_k}{b}} \right)^2 \right).
\end{aligned}
$$

□

Corollary 5.40. *Under the assumptions of Lemma 5.39, we have*

$$
\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] \le \min\{b, c_k\} \quad \textit{for all } k \in \mathbb{N},
$$

where $c_k = \sqrt{2}\alpha_k^{\beta/2} e^{-\alpha_k} |(u^\dagger, \varphi_k)_X|$.

Proof. In the proof of Lemma 5.39 we saw that

$$
\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] = b(1 - e^{-\frac{c_k}{b}}).
$$

Now on the one hand, the expected value of each $\Psi_k(u^\dagger, y)$ complies with the estimate

$$
\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] \le b,
$$

since $1 - e^{-c_k/b} \leq 1$ for all $b > 0$. On the other hand, the estimate $1 - e^{-x} \leq x$ for $x \geq 0$ yields

$$\mathbb{E}\left[\Psi_k(u^\dagger, y)\right] \leq b\frac{c_k}{b} = c_k. \qquad \square$$

Next, we show that the expectation of $\Psi(u^\dagger, y)$ converges in the order of δ if b tends to 0 in the order of $\delta^{1+\varepsilon}$ for some $\varepsilon > 0$.

Lemma 5.41. *Let $u^\dagger \in X$ with $\|u^\dagger\|_X \leq \rho$, $b > 0$, $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{s-\beta}}$ and $\delta > 0$. For every $\varepsilon > 0$ there is a constant $C_\varepsilon = C_\varepsilon(\varepsilon, \rho, d, \beta) > 0$ such that*

$$b \leq C_\varepsilon \min\left\{1, \delta^{1+\varepsilon}\right\}$$

implies

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] \leq \delta.$$

Proof. Let $\varepsilon > 0$. The idea is to use Lemma 5.39, split up the expected value

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] \leq \sum_{k=1}^{N} b\left(1 - e^{-\frac{c_k}{b}}\right) + \sum_{k=N+1}^{\infty} b\left(1 - e^{-\frac{c_k}{b}}\right) \leq Nb + \sum_{k=N+1}^{\infty} c_k \tag{5.8}$$

for some $N \in \mathbb{N}$ and estimate each summand either by b or by c_k. Now we show that we can let $\sum_{k=N+1}^{\infty} c_k$ become arbitrary small by choosing N large enough. We estimate using the Cauchy-Schwarz inequality and Assumption 5.1 (iv) that

$$\begin{aligned}
\sum_{k=N+1}^{\infty} c_k &= \sum_{k=N+1}^{\infty} \sqrt{2}\alpha_k^{\frac{\beta}{2}} e^{-\alpha_k} \left|(u^\dagger, \varphi_k)_X\right| \leq \left(\sum_{k=N+1}^{\infty} 2\alpha_k^{\beta} e^{-2\alpha_k}\right)^{\frac{1}{2}} \left(\sum_{k=N+1}^{\infty} \left|(u^\dagger, \varphi_k)_X\right|^2\right)^{\frac{1}{2}} \\
&\leq \left(\sum_{k=N+1}^{\infty} 2C_+^{\beta} k^{\frac{2\beta}{d}} e^{-2C_- k^{\frac{2}{d}}}\right)^{\frac{1}{2}} \left\|u^\dagger\right\|_X \\
&\leq \left(\sum_{k=N+1}^{\infty} 2C_+^{\beta} (2C_-)^{-\beta} \left(2C_- k^{\frac{2}{d}}\right)^{\beta} e^{-2C_- k^{\frac{2}{d}}}\right)^{\frac{1}{2}} \rho.
\end{aligned}$$

For N large enough we have $2C_- k^{2/d} \geq 1$ for $k \geq N+1$, so that

$$\sum_{k=N+1}^{\infty} c_k \leq \left(\sum_{k=N+1}^{\infty} 2C_+^{\beta} (2C_-)^{-\beta} \left(2C_- k^{\frac{2}{d}}\right)^{\lceil\beta\rceil} e^{-2C_- k^{\frac{2}{d}}}\right)^{\frac{1}{2}} \rho,$$

where $\lceil\beta\rceil := \min\{B \in \mathbb{Z} : B \geq \beta\}$. The exponential function satisfies the estimate $e^x > \frac{x^{M+B}}{(M+B)!}$ for all $x \in \mathbb{R}$ and $M, B \in \mathbb{N}$. Hence, $x^B e^{-x} \leq (M+B)! x^{-M}$, and consequently,

$$\begin{aligned}
\sum_{k=N+1}^{\infty} c_k &\leq \left(\sum_{k=N+1}^{\infty} 2C_+^{\beta} (2C_-)^{-\beta} (\lceil\beta\rceil + M)! (2C_-)^{-M} k^{-\frac{2M}{d}}\right)^{\frac{1}{2}} \rho \\
&= \left(2C_+^{\beta} (2C_-)^{-(\beta+M)} (\lceil\beta\rceil + M)! \sum_{k=N+1}^{\infty} k^{-\frac{2M}{d}}\right)^{\frac{1}{2}} \rho
\end{aligned}$$

for all $M \in \mathbb{N}$ and N large enough. Since $k \mapsto k^{-\frac{2M}{d}}$ is monotonically decreasing, we can estimate the remaining series by an integral,

$$\sum_{k=N+1}^{\infty} k^{-\frac{2M}{d}} \leq \int_N^\infty t^{-\frac{2M}{d}} \mathrm{d}t = \left(\frac{2M}{d} - 1\right) N^{1-\frac{2M}{d}},$$

which leads to

$$\sum_{k=N+1}^{\infty} c_k \leq \left(4C_+^\beta (2C_-)^{-(\beta+M)} (\lceil\beta\rceil + M)! \left(\frac{M}{d} - \frac{1}{2}\right)\right)^{\frac{1}{2}} \rho N^{-\left(\frac{M}{d}-\frac{1}{2}\right)} =: C_M \rho N^{-\left(\frac{M}{d}-\frac{1}{2}\right)}.$$

Now we choose $M \geq d\left(\frac{1}{\varepsilon} + \frac{1}{2}\right)$ to ensure

$$\frac{M}{d} - \frac{1}{2} \geq \frac{1}{\varepsilon} > 0.$$

Then we achieve $\sum_{k=N+1}^\infty c_k \leq \frac{\delta}{2}$ by choosing

$$\left(\frac{2}{\delta} C_M \rho\right)^{\frac{1}{\frac{M}{d}-\frac{1}{2}}} \leq N \leq 2\left(\frac{2}{\delta} C_M \rho\right)^{\frac{1}{\frac{M}{d}-\frac{1}{2}}}.$$

Finally, we want to bound the first term Nb in (5.8) by $\frac{\delta}{2}$ as well. By the choice of N and M, we have

$$\frac{\delta}{2N} \geq \frac{\delta}{4}\left(\frac{2}{\delta} C_M \rho\right)^{-\frac{1}{\frac{M}{d}-\frac{1}{2}}} = \frac{1}{4}(2C_M\rho)^{-\frac{1}{\frac{M}{d}-\frac{1}{2}}} \delta^{1+\frac{1}{\frac{M}{d}-\frac{1}{2}}} = C_\varepsilon \delta^{1+\frac{1}{\frac{M}{d}-\frac{1}{2}}},$$

where $C_\varepsilon := \frac{1}{4}(2C_M\rho)^{-\frac{1}{\frac{M}{d}-\frac{1}{2}}}$. If $\delta > 1$, then we use $b \leq C_\varepsilon$ to get

$$Nb \leq NC_\varepsilon \leq NC_\varepsilon \delta^{1+\frac{1}{\frac{M}{d}-\frac{1}{2}}} \leq \frac{\delta}{2}.$$

If, on the other hand, $\delta \in (0,1]$, we use $b \leq C_\varepsilon \delta^{1+\varepsilon}$ to get

$$Nb \leq NC_\varepsilon \delta^{1+\varepsilon} \leq NC_\varepsilon \delta^{1+\frac{1}{\frac{M}{d}-\frac{1}{2}}} \leq \frac{\delta}{2}.$$

Finally, summing up yields

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] \leq Nb + \sum_{k=N+1}^\infty c_k \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta. \qquad \square$$

We use the explicit representation of the MAP estimator from Lemma 5.34 to study the expected values of $\Psi(\hat{u}_{\mathrm{MAP}}(y), y)$ and $\Psi(\hat{u}_{\mathrm{MAP}}(y), e^{-A}u^\dagger)$.

Lemma 5.42. *Let $u^\dagger \in X$, $b > 0$ and $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{s-\beta}}$. Then*

$$0 \leq \mathbb{E}\left[\Psi(\hat{u}_{\mathrm{MAP}}(y), e^{-A}u^\dagger)\right] \leq \mathbb{E}\left[\Psi(\hat{u}_{\mathrm{MAP}}(y), y)\right].$$

Proof. Set $\bar{u} := \hat{u}_{\mathrm{MAP}}(y)$. As in the proof of Lemma 5.39, we write

$$\mathbb{E}\left[\Psi(\bar{u}, y)\right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \Psi_k(\bar{u}, y)\right] = \sum_{k=1}^{\infty} \mathbb{E}\left[\Psi_k(\bar{u}, y)\right]$$

and $\mathbb{E}[\Psi(\bar{u}, e^{-A}u^\dagger)] = \sum_{k=1}^{\infty} \mathbb{E}[\Psi_k(\bar{u}, e^{-A}u^\dagger)]$, respectively, and show that $0 \le \mathbb{E}[\Psi_k(\bar{u}, e^{-A}u^\dagger)] \le \mathbb{E}\left[\Psi_k(\bar{u}, y)\right]$ for all $k \in \mathbb{N}$. For a fixed $k \in \mathbb{N}$ we set

$$a := (e^{-A}u^\dagger, e_k)_{\mathcal{X}^s}, \quad \lambda := b^2 \alpha_k^{s-\beta},$$

where $e_k := \alpha_k^{-s/2}\varphi_k$. A change of variables results in

$$\begin{aligned}
\mathbb{E}[\Psi_k(\bar{u}, y)] &= \int_{\mathcal{X}^s} b\sqrt{\frac{2}{\lambda}}\left(\left|(y, e_k)_{\mathcal{X}^s} - (e^{-A}\bar{u}, e_k)_{\mathcal{X}^s}\right| - \left|(y, e_k)_{\mathcal{X}^s}\right| + |a|\right) \mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}(\mathrm{d}y) \\
&= \int_{\mathbb{R}} b\sqrt{\frac{2}{\lambda}}\left(\left|x - \alpha_k^{\frac{s}{2}} e^{-\alpha_k}\left(\bar{u}, \varphi_k\right)_X\right| - |x| + |a|\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x) \qquad (5.9)
\end{aligned}$$

By Lemma 5.34, we have

$$\begin{aligned}
\alpha_k^{\frac{s}{2}} e^{-\alpha_k}(\bar{u}, \varphi_k)_X &= \alpha_k^{\frac{s}{2}} e^{-\alpha_k} \max\left\{-\sqrt{2}\frac{r^2}{b}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}, \min\left\{e^{\alpha_k}(y, \varphi_k)_X, \sqrt{2}\frac{r^2}{b}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right\}\right\} \\
&= \max\left\{-R, \min\left\{(y, e_k)_{\mathcal{X}^s}, R\right\}\right\},
\end{aligned}$$

where $R := \sqrt{2}\frac{r^2}{b}\alpha_k^{s/2+\beta/2-\tau} e^{-2\alpha_k}$. Inserting this into (5.9) and splitting up the integral yields

$$\begin{aligned}
\mathbb{E}[\Psi_k(\bar{u}, y)] &= \int_{-\infty}^{-R} b\sqrt{\frac{2}{\lambda}}\left(|x - (-R)| - |x| + |a|\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&\quad + \int_{-R}^{R} b\sqrt{\frac{2}{\lambda}}\left(0 - |x| + |a|\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&\quad + \int_{R}^{\infty} b\sqrt{\frac{2}{\lambda}}\left(|x - R| - |x| + |a|\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&= b\sqrt{\frac{2}{\lambda}}\,|a| + \int_{-R}^{0} b\sqrt{\frac{2}{\lambda}} x \mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{0}^{R} b\sqrt{\frac{2}{\lambda}}\left(-x\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&\quad + \int_{R}^{\infty} b\sqrt{\frac{2}{\lambda}}\left(-R\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{-\infty}^{-R} b\sqrt{\frac{2}{\lambda}}\left(-R\right) \mathcal{L}_{a,\lambda}(\mathrm{d}x)
\end{aligned}$$

We use

$$\int_{-\infty}^{-R} \mathrm{d}\mathcal{L}_{a,\lambda} + \int_{R}^{\infty} \mathrm{d}\mathcal{L}_{a,\lambda} = \int_{R}^{\infty} \mathrm{d}\mathcal{L}_{-|a|,\lambda} + \int_{R}^{\infty} \mathrm{d}\mathcal{L}_{|a|,\lambda}$$

and

$$\int_{-R}^{0} x \mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{0}^{R} (-x)\, \mathcal{L}_{a,\lambda}(\mathrm{d}x) = \int_{0}^{R} (-x)\, \mathcal{L}_{-|a|,\lambda}(\mathrm{d}x) + \int_{0}^{R} (-x)\, \mathcal{L}_{|a|,\lambda}(\mathrm{d}x)$$

to arrive at

$$\begin{aligned}
\mathbb{E}[\Psi_k(\bar{u}, y)] &= b\sqrt{\frac{2}{\lambda}}\,|a| - \int_0^R b\sqrt{\frac{2}{\lambda}} x \mathcal{L}_{-|a|,\lambda}(\mathrm{d}x) - \int_0^R b\sqrt{\frac{2}{\lambda}} x \mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&\quad - \int_R^\infty b\sqrt{\frac{2}{\lambda}} R \mathcal{L}_{-|a|,\lambda}(\mathrm{d}x) - \int_R^\infty b\sqrt{\frac{2}{\lambda}} R \mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&=: b\sqrt{\frac{2}{\lambda}}\,|a| + I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

Here we must distinguish between $|a| < R$ and $|a| \geq R$. We begin with $|a| < R$, where the second integral is split up into

$$\begin{aligned}
I_2 &= -\int_0^{|a|} b\sqrt{\frac{2}{\lambda}} x \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x - \int_{|a|}^R b\sqrt{\frac{2}{\lambda}} x \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\mathrm{d}x \\
&= -b\frac{1}{\sqrt{2\lambda}}\left[x e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]_0^{|a|} + \int_0^{|a|} b\frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x \\
&\quad - b\frac{1}{\sqrt{2\lambda}}\left[-x e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\right]_{|a|}^R + \int_{|a|}^R b\frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\mathrm{d}x \\
&= -b\frac{1}{\sqrt{2\lambda}}\left(|a| - 0 - R e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)} + |a|\right) \\
&\quad + b\left[\frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]_0^{|a|} + b\left[-\frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\right]_{|a|}^R \\
&= b\frac{R}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)} - b\sqrt{\frac{2}{\lambda}}\,|a| \\
&\quad + \frac{b}{2}\left(1 - e^{-\sqrt{\frac{2}{\lambda}}|a|}\right) + \frac{b}{2}\left(1 - e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)}\right).
\end{aligned}$$

The fourth integral computes as

$$\begin{aligned}
I_4 &= -\int_R^\infty b\sqrt{\frac{2}{\lambda}} R \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\mathrm{d}x = -b\frac{R}{\sqrt{2\lambda}}\left[-e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\right]_R^\infty \\
&= -b\frac{R}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)}.
\end{aligned}$$

Similarly, the third integral equates to

$$I_3 = -\int_R^\infty b\sqrt{\frac{2}{\lambda}} R \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(x+|a|)}\mathrm{d}x = -b\frac{R}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)}.$$

Finally,

$$\begin{aligned}
I_1 &= -\int_0^R b\sqrt{\frac{2}{\lambda}}x\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(x+|a|)}\mathrm{d}x \\
&= -b\frac{1}{\sqrt{2\lambda}}\left[-xe^{-\sqrt{\frac{2}{\lambda}}(x+|a|)}\right]_0^R + \int_0^R b\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(x+|a|)}\mathrm{d}x \\
&= b\frac{R}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)} + b\left[-\frac{1}{2}e^{-\sqrt{\frac{2}{\lambda}}(x+|a|)}\right]_0^R \\
&= b\frac{R}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)} - \frac{b}{2}e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)} + \frac{b}{2}e^{-\sqrt{\frac{2}{\lambda}}|a|}.
\end{aligned}$$

For $|a| < R$, summing up yields

$$\mathbb{E}[\Psi_k(\bar{u}, y)] = \frac{b}{2}\left(1 - e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)}\right) + \frac{b}{2}\left(1 - e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)}\right).$$

The case $|a| \geq R$ remains. Here, the fourth integral is split up into

$$\begin{aligned}
I_4 &= -\int_R^{|a|} b\sqrt{\frac{2}{\lambda}}R\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x - \int_{|a|}^{\infty} b\sqrt{\frac{2}{\lambda}}R\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\mathrm{d}x \\
&= -b\frac{R}{\sqrt{2\lambda}}\left[e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]\Bigg|_R^{|a|} - b\frac{R}{\sqrt{2\lambda}}\left[-e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\right]\Bigg|_{|a|}^{\infty} \\
&= -b\frac{R}{\sqrt{2\lambda}}\left(1 - e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} - 0 + 1\right) \\
&= -b\sqrt{\frac{2}{\lambda}}R + b\frac{R}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)}.
\end{aligned}$$

The second integral computes as

$$\begin{aligned}
I_2 &= -\int_0^R b\sqrt{\frac{2}{\lambda}}x\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x \\
&= -b\frac{1}{\sqrt{2\lambda}}\left[xe^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]\Bigg|_0^R + \int_0^R b\frac{1}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x \\
&= -b\frac{R}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} + b\left[\frac{1}{2}e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\right]\Bigg|_0^R \\
&= -b\frac{R}{\sqrt{2\lambda}}e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} + \frac{b}{2}e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} - \frac{b}{2}e^{-\sqrt{\frac{2}{\lambda}}|a|}.
\end{aligned}$$

I_1 and I_3 remain unchanged. So for $|a| \geq R$, summing up results in

$$\mathbb{E}[\Psi_k(\bar{u}, y)] = b\sqrt{\frac{2}{\lambda}}(|a| - R) + \frac{b}{2}\left(e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} - e^{-\sqrt{\frac{2}{\lambda}}(|a|+R)}\right).$$

Next, we consider

$$
\begin{aligned}
\mathbb{E}[\Psi_k(\bar{u}, e^{-A}u^\dagger)] &= \int_{\mathcal{X}^s} \sqrt{2}\alpha_k^{\beta-s} 2 \left|(e^{-A}u^\dagger, e_k)_{\mathcal{X}^s} - (e^{-A}\bar{u}, e_k)_{\mathcal{X}^s}\right| \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{s-\beta}}(\mathrm{d}y) \\
&= \int_{-\infty}^{-R} b\sqrt{\frac{2}{\lambda}}\, |a+R|\, \mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{-R}^{R} b\sqrt{\frac{2}{\lambda}}\, |a-x|\, \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&\quad + \int_{R}^{\infty} b\sqrt{\frac{2}{\lambda}}\, |a-R|\, \mathcal{L}_{a,\lambda}(\mathrm{d}x)
\end{aligned}
$$

Here, we use

$$
\int_{-\infty}^{-R} |R+a| \mathrm{d}\mathcal{L}_{a,\lambda} + \int_{R}^{\infty} |R-a| \mathrm{d}\mathcal{L}_{a,\lambda} = \int_{R}^{\infty} |R+|a||\, \mathrm{d}\mathcal{L}_{-|a|,\lambda} + \int_{R}^{\infty} |R-|a||\, \mathrm{d}\mathcal{L}_{|a|,\lambda}
$$

to arrive at

$$
\begin{aligned}
\mathbb{E}[\Psi_k(\bar{u}, e^{-A}u^\dagger)] &= \int_{R}^{\infty} b\sqrt{\frac{2}{\lambda}}\, |R+|a||\, \mathcal{L}_{-|a|,\lambda}(\mathrm{d}x) + \int_{R}^{\infty} b\sqrt{\frac{2}{\lambda}}\, |R-|a||\, \mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&\quad + \int_{-R}^{R} b\sqrt{\frac{2}{\lambda}}\, |x-|a||\, \mathcal{L}_{|a|,\lambda}(\mathrm{d}x) \\
&=: I_5 + I_6 + I_7.
\end{aligned}
$$

The integral I_5 computes as

$$
I_5 = b\frac{1}{\sqrt{2\lambda}}\,(R+|a|)\, e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)}.
$$

In case that $|a| < R$, I_7 is split up into

$$
\begin{aligned}
I_7 &= \int_{-R}^{|a|} b\sqrt{\frac{2}{\lambda}}\,(|a|-x)\, e^{-\sqrt{\frac{2}{\lambda}}(|a|-x)}\mathrm{d}x + \int_{|a|}^{R} b\sqrt{\frac{2}{\lambda}}\,(x-|a|)\, e^{-\sqrt{\frac{2}{\lambda}}(x-|a|)}\mathrm{d}x \\
&= -b\frac{1}{\sqrt{2\lambda}}\,(R+|a|)\, e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)} + \frac{b}{2}e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)} \\
&\quad - b\frac{1}{\sqrt{2\lambda}}\,(R-|a|)\, e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)} + \frac{b}{2}e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)},
\end{aligned}
$$

and I_6 equals

$$
I_6 = b\frac{1}{\sqrt{2\lambda}}\,(R-|a|)\, e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)}.
$$

This leads to

$$
\mathbb{E}\left[\Psi_k(\bar{u}, e^{-A}u^\dagger)\right] = \frac{b}{2}\left(e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)} - e^{-\sqrt{\frac{2}{\lambda}}(R+|a|)}\right) \geq 0
$$

for $|a| < R$, as the exponential function increases monotonically. Hence, we obtain

$$
\mathbb{E}\left[\Psi_k(\bar{u}, y)\right] - \mathbb{E}\left[\Psi_k(\bar{u}, e^{-A}u^\dagger)\right] = b\left(1 - e^{-\sqrt{\frac{2}{\lambda}}(R-|a|)}\right) \geq 0
$$

for $|a| < R$.

In case that $|a| \geq R$,

$$I_6 = b\sqrt{\frac{2}{\lambda}}\,(|a| - R) - b\frac{1}{\sqrt{2\lambda}}\,(|a| - R)\,e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)},$$

and

$$\begin{aligned} I_7 = b\frac{1}{\sqrt{2\lambda}} &\left((|a| - R)\,e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} - (|a| + R)\,e^{-\sqrt{\frac{2}{\lambda}}(|a|+R)}\right) \\ &+ \frac{b}{2}\left(e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} - e^{-\sqrt{\frac{2}{\lambda}}(|a|+R)}\right), \end{aligned}$$

which results in

$$\mathbb{E}\left[\Psi_k(\bar{u}, e^{-A}u^\dagger)\right] = b\sqrt{\frac{2}{\lambda}}\,(|a| - R) + \frac{b}{2}\left(e^{-\sqrt{\frac{2}{\lambda}}(|a|-R)} - e^{-\sqrt{\frac{2}{\lambda}}(|a|+R)}\right) \geq 0$$

for $|a| \geq R$, as the exponential function increases monotonically. This yields $\mathbb{E}\left[\Psi_k(\bar{u}, e^{-A}u^\dagger)\right] = \mathbb{E}\left[\Psi_k(\bar{u}, y)\right]$ in case $|a| \geq R$. □

Now we state a first results regarding the weak convergence of the expectation of the MAP estimator.

Theorem 5.43. *Let $\{b_n\}_{n\in\mathbb{N}}$ and $\{r_n\}_{n\in\mathbb{N}}$ be positive sequences such that $b_n \to 0$ and $r_n \to 0$. Moreover, let $u^\dagger \in \mathcal{X}^\tau$ and $y^n \sim \mathcal{L}_{e^{-A}u^\dagger, b_n^2 A^{s-\beta}}$ for all $n \in \mathbb{N}$. If*

$$\frac{b_n}{r_n^2} \to 0 \quad \textit{and} \quad \frac{r_n^2}{b_n^\omega} \leq C$$

for some $\omega \in (0, 1)$ and $C > 0$, then $\{\mathbb{E}[\hat{u}_{\mathrm{MAP}}(y^n)]\}_{n\in\mathbb{N}}$ contains a subsequence that converges weakly towards $u^\dagger$ in $\mathcal{X}^\tau$.

Proof. For all $n \in \mathbb{N}$ set $u^n := \hat{u}_{\mathrm{MAP}}(y^n)$. Define $\delta_n := \left(C_\varepsilon^{-1} b_n\right)^{\frac{1}{1+\varepsilon}}$ for all $n \in \mathbb{N}$, where $\varepsilon := \frac{\omega}{1-\omega}$ and C_ε is the associated constant from Lemma 5.41. Then $\delta_n \to 0$,

$$\delta_n^{1+\varepsilon} = C_\varepsilon^{-1} b_n \quad \text{and} \quad \frac{\delta_n}{b_n} = C_\varepsilon^{-\frac{1}{1+\varepsilon}} b_n^{-\frac{\varepsilon}{1+\varepsilon}} = \frac{1}{C_\varepsilon^{1-\omega} b_n^\omega}.$$

As u^n minimises J_n, the inequality

$$\Psi(u^n, y^n) + \frac{b_n}{2r_n^2}\,\|u^n\|_{\mathcal{X}^\tau}^2 \leq \Psi(u^\dagger, y^n) + \frac{b_n}{2r_n^2}\|u^\dagger\|_{\mathcal{X}^\tau}^2$$

holds in particular. We pass on to expected values,

$$\mathbb{E}\left[\Psi(u^n, y^n)\right] + \frac{b_n}{2r_n^2}\mathbb{E}\left[\|u^n\|_{\mathcal{X}^\tau}^2\right] \leq \mathbb{E}[\Psi(u^\dagger, y^n)] + \frac{b_n}{2r_n^2}\|u^\dagger\|_{\mathcal{X}^\tau}^2. \tag{5.10}$$

Lemma 5.42 guarantees that $\mathbb{E}\left[\Psi(u^n, y^n)\right]$ is non-negative, so that we can estimate both terms on the left-hand side separately. The convexity of $\|\cdot\|^2_{\mathcal{X}^\tau}$ allows us to use Jensen's inequality and we use Lemma 5.41 as well as the boundedness of $\frac{r_n^2}{b_n^\omega}$ to obtain

$$\begin{aligned}\|\mathbb{E}[u^n]\|^2_{\mathcal{X}^\tau} &\le \mathbb{E}\left[\|u^n\|^2_{\mathcal{X}^\tau}\right] \le \frac{2r_n^2}{b_n}\mathbb{E}[\Psi(u^\dagger, y^n)] + \|u^\dagger\|^2_{\mathcal{X}^\tau} \\ &\le \frac{2r_n^2}{b_n}\delta_n + \|u^\dagger\|^2_{\mathcal{X}^\tau} = \frac{2r_n^2}{C_\varepsilon^{1-\omega} b_n^\omega} + \|u^\dagger\|^2_{\mathcal{X}^\tau} \le \frac{2C}{C_\varepsilon^{1-\omega}} + \|u^\dagger\|^2_{\mathcal{X}^\tau}\end{aligned}$$

from (5.10), where

$$\mathbb{E}[u^n] = \int_{\mathcal{X}^s} u^n(y^n)\mathcal{L}_{e^{-A}u^\dagger, b_n^2 A^{s-\beta}}(\mathrm{d}y^n) \in \mathcal{X}^\tau$$

in the sense of the Bochner integral. So $\{\mathbb{E}[u^n]\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{X}^\tau$. Thereby it contains a subsequence, again denoted by $\{\mathbb{E}[u^n]\}_{n\in\mathbb{N}}$, that converges weakly towards a $\bar{u} \in \mathcal{X}^\tau$. e^{-A} is compact, as it is the limit of bounded linear operators

$$u \mapsto \sum_{k=1}^{n} e^{-\alpha_k}\,(u, \varphi_k)_X\,\varphi_k$$

on X with finite-dimensional range. This implies that e^{-A} is completely continuous, since X is reflexive and both X and $\mathcal{X}^s$ are Hilbert spaces. Hence, $e^{-A}\mathbb{E}[u^n] \longrightarrow e^{-A}\bar{u}$ strongly in $\mathcal{X}^s$.

It remains to show that $\bar{u} = u^\dagger$. To this end we conclude

$$\mathbb{E}[\Psi(u^n, y^n)] \le \mathbb{E}[\Psi(u^\dagger, y^n)] + \frac{b_n}{2r_n^2}\|u^\dagger\|^2_{\mathcal{X}^\tau} \le \delta_n + \frac{b_n}{2r_n^2}\|u^\dagger\|^2_{\mathcal{X}^\tau}$$

from (5.10), again using Lemma 5.41. Now the assumptions on r_n and b_n ensure $\mathbb{E}[\Psi(u^n, y^n)] \longrightarrow 0$. This implies $\mathbb{E}[\Psi(u^n, e^{-A}u^\dagger)] \longrightarrow 0$, since

$$0 \le \lim_{n\to\infty} \mathbb{E}\left[\Psi(u^n, e^{-A}u^\dagger)\right] \le \lim_{n\to\infty} \mathbb{E}\left[\Psi(u^n, y^n)\right] = 0$$

as a consequence of Lemma 5.42. However,

$$\begin{aligned}\Psi(u^n, e^{-A}u^\dagger) &= \sqrt{2}\sum_{k=1}^{\infty} \alpha_k^{\frac{\beta-s}{2}} \left|\left(e^{-A}u^\dagger, e_k\right)_{\mathcal{X}^s} - \left(e^{-A}u^n, e_k\right)_{\mathcal{X}^s}\right| \\ &= \sqrt{2}\sum_{k=1}^{\infty} \alpha_k^{\frac{\beta}{2}} \left|\left(e^{-A}u^\dagger - e^{-A}u^n, \varphi_k\right)_X\right| \\ &\ge \sqrt{2}\left(\sum_{k=1}^{\infty} \alpha_k^{\beta} \left|\left(e^{-A}u^\dagger - e^{-A}u^n, \varphi_k\right)_X\right|^2\right)^{\frac{1}{2}} \\ &= \sqrt{2}\left\|e^{-A}u^n - e^{-A}u^\dagger\right\|_{\mathcal{X}^\beta},\end{aligned}$$

so that the continuity of e^{-A}, the convexity of $\|\cdot\|_{\mathcal{X}^\beta}$ and Jensen's inequality yield

$$\begin{aligned}\left\|e^{-A}\mathbb{E}[u^n] - e^{-A}u^\dagger\right\|_{\mathcal{X}^\beta} &= \left\|\mathbb{E}\left[e^{-A}u^n - e^{-A}u^\dagger\right]\right\|_{\mathcal{X}^\beta} \le \mathbb{E}\left[\left\|e^{-A}u^n - e^{-A}u^\dagger\right\|_{\mathcal{X}^\beta}\right] \\ &\le \frac{1}{\sqrt{2}}\mathbb{E}\left[\Psi(u^n, e^{-A}u^\dagger)\right].\end{aligned}$$

Consequently, $e^{-A}\mathbb{E}[u^n] \to e^{-A}u^\dagger$ in $\mathcal{X}^\beta$ and by the continuity of the embedding $\mathcal{X}^\beta \hookrightarrow \mathcal{X}^s$ according to Proposition 5.10 (i) also in $\mathcal{X}^s$. Now, the uniqueness of the limit implies $e^{-A}u^\dagger = e^{-A}\bar{u}$. In a last step, it follows from the injectivity of e^{-A} that $\bar{u} = u^\dagger$. □

We can achieve strong convergence if we assume that r_n^2/b_n^ω is not only bounded, but converges to 0.

Theorem 5.44. *Let $\{b_n\}_{n\in\mathbb{N}}$ and $\{r_n\}_{n\in\mathbb{N}}$ be positive sequences such that $b_n \to 0$ and $r_n \to 0$. Moreover, let $u^\dagger \in \mathcal{X}^\tau$ and $y^n \sim \mathcal{L}_{e^{-A}u^\dagger, b_n^2 A^{s-\beta}}$, $n \in \mathbb{N}$. If*

$$\frac{b_n}{r_n^2} \to 0 \quad \text{and} \quad \frac{r_n^2}{b_n^\omega} \to 0$$

for some $\omega \in (0,1)$, then $\mathbb{E}[\hat{u}_{\mathrm{MAP}}(y^n)] \to u^\dagger$ in $\mathcal{X}^\tau$.

Proof. We consider an arbitrary subsequence of $\{u^n\}_{n\in\mathbb{N}}$, again denoted by $\{u^n\}_{n\in\mathbb{N}}$. The minimisation property of u^n implies

$$\mathbb{E}\left[\Psi(u^n, y^n)\right] + \frac{b_n}{2r_n^2}\mathbb{E}\left[\|u^n\|_{\mathcal{X}^\tau}^2\right] \le \mathbb{E}[\Psi(u^\dagger, y^n)] + \frac{b_n}{2r_n^2}\|u^\dagger\|_{\mathcal{X}^\tau}^2.$$

The convexity of $\|\cdot\|_{\mathcal{X}^\tau}^2$ allows us to use Jensen's inequality, and by Lemma 5.42 and Lemma 5.41 with $\varepsilon := \frac{\omega}{1-\omega}$ we have

$$\begin{aligned}
\|\mathbb{E}[u^n]\|_{\mathcal{X}^\tau}^2 \le \mathbb{E}\left[\|u^n\|_{\mathcal{X}^\tau}^2\right] &\le \frac{2r_n^2}{b_n}\mathbb{E}[\Psi(u^n, y^n)] + \mathbb{E}\left[\|u^n\|_{\mathcal{X}^\tau}^2\right] \\
&\le \frac{2r_n^2}{b_n}\mathbb{E}[\Psi(u^\dagger, y^n)] + \|u^\dagger\|_{\mathcal{X}^\tau}^2 = \frac{2r_n^2}{C_\varepsilon^{1-\omega} b_n^\omega} + \|u^\dagger\|_{\mathcal{X}^\tau}^2
\end{aligned}$$

for all $n \in \mathbb{N}$ with $b_n \le C_\varepsilon$, that is for n large enough. As $\frac{r_n^2}{b_n^\omega} \to 0$ by assumption, this implies that

$$\limsup_{n\to\infty} \|\mathbb{E}[u^n]\|_{\mathcal{X}^\tau}^2 \le \|u^\dagger\|_{\mathcal{X}^\tau}^2$$

and that $\{\mathbb{E}[u^n]\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{X}^\tau$. It thereby contains a subsequence, again denoted by $\{\mathbb{E}[u^n]\}_{n\in\mathbb{N}}$, that converges weakly towards a $\bar{u} \in \mathcal{X}^\tau$, which implies

$$\|\bar{u}\|_{\mathcal{X}^\tau} \le \liminf_{n\to\infty} \|\mathbb{E}[u^n]\|_{\mathcal{X}^\tau} \le \limsup_{n\to\infty} \|\mathbb{E}[u^n]\|_{\mathcal{X}^\tau} \le \|u^\dagger\|_{\mathcal{X}^\tau}$$

because of the weak lower semi-continuity of $\|\cdot\|_{\mathcal{X}^\tau}$. As in the proof of Theorem 5.43, we show that $\bar{u} = u^\dagger$, so that

$$\mathbb{E}[u^n] \rightharpoonup u^\dagger \quad \text{and} \quad \lim_{n\to\infty} \|\mathbb{E}[u^n]\|_{\mathcal{X}^\tau} = \|u^\dagger\|_{\mathcal{X}^\tau}.$$

Since $\mathcal{X}^\tau$ is a Hilbert space, it follows that

$$\left\|\mathbb{E}[u^n - u^\dagger]\right\|_{\mathcal{X}^\tau} = \left\|\mathbb{E}[u^n] - u^\dagger\right\|_{\mathcal{X}^\tau} \to 0.$$

As the subsequence chosen in the beginning was arbitrary, the whole sequence $\{\mathbb{E}[u^n]\}_{n\in\mathbb{N}}$ converges towards $u^\dagger$. □

5.5.4 Convergence Rate in Mean of the Discrepancy Term

We show that for dimension $d \leq 4$, the expectation of the discrepancy term $\Psi(u^\dagger, y)$ converges at least in the order of $b(\ln C/b)^{d/2}$ as $b \longrightarrow 0$. To do so, we need the following inequality for the incomplete gamma function $\Gamma(a, x) := \int_x^\infty t^{a-1} e^{-t} \mathrm{d}t$.

Lemma 5.45. *For all $a \in (0, 2]$ and $x > 0$ we have*

$$\Gamma(a, x) \leq \left(x^{a-1} + |a-1|\, x^{a-2}\right) e^{-x}.$$

Proof. Integration by parts produces

$$\Gamma(a, x) = x^{a-1} e^{-x} + (a-1) \int_x^\infty t^{a-2} e^{-t} \mathrm{d}t.$$

Now $t^{a-2} \leq x^{a-2}$, because $a \leq 2$, which leads to

$$\Gamma(a, x) \leq x^{a-1} e^{-x} + |a-1| x^{a-2} e^{-x}.$$

□

Lemma 5.46. *Let $d \in \{1, 2, 3, 4\}$, $u^\dagger \in X$, $b > 0$ and $y \sim \mathcal{L}_{e^{-A} u^\dagger, b^2 A^{s-\beta}}$. For every $C > 1$ there is an $\varepsilon > 0$ such that*

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] \leq C \left(\frac{2}{C_-}\right)^{\frac{d}{2}} b \ln\left(\frac{\sqrt{2}}{b} \|u^\dagger\|_X \left(\frac{C_+ \beta}{C_- e}\right)^{\frac{\beta}{2}}\right)^{\frac{d}{2}} \quad \textit{for all } b \in (0, \varepsilon).$$

Proof. By Lemma 5.39,

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] = b \sum_{k=1}^\infty \left(1 - \exp\left(-\frac{c_k}{b}\right)\right)$$

holds with $c_k := \sqrt{2} \alpha_k^{\beta/2} e^{-\alpha_k} |(u^\dagger, \varphi_k)_X|$. We use Assumption 5.1 (iv) and Lemma 5.14 to estimate

$$\begin{aligned} \frac{c_k}{b} &\leq \frac{\sqrt{2}}{b} \left(C_+ k^{\frac{2}{d}}\right)^{\frac{\beta}{2}} e^{-C_- k^{\frac{2}{d}}} \left|(u^\dagger, \varphi_k)_X\right| \leq \frac{\sqrt{2}}{b} \left(\frac{2C_+}{C_-}\right)^{\frac{\beta}{2}} \left(\frac{C_-}{2} k^{\frac{2}{d}}\right)^{\frac{\beta}{2}} e^{-\frac{C_-}{2} k^{\frac{2}{d}}} e^{-\frac{C_-}{2} k^{\frac{2}{d}}} \|u^\dagger\|_X \\ &\leq \frac{\sqrt{2}}{b} \left(\frac{2C_+}{C_-}\right)^{\frac{\beta}{2}} \left(\frac{\beta}{2}\right)^{\frac{\beta}{2}} e^{-\frac{\beta}{2}} e^{-\frac{C_-}{2} k^{\frac{2}{d}}} \|u^\dagger\|_X = \frac{\sqrt{2}}{b} \|u^\dagger\|_X \left(\frac{C_+ \beta}{C_- e}\right)^{\frac{\beta}{2}} e^{-\frac{C_-}{2} k^{\frac{2}{d}}} \end{aligned}$$

for all $k \in \mathbb{N}$. We define a dominating function f by

$$f(\kappa) := \frac{\sqrt{2}}{b} \|u^\dagger\|_X \left(\frac{C_+ \beta}{C_- e}\right)^{\frac{\beta}{2}} e^{-\frac{C_-}{2} \kappa^{\frac{2}{d}}} \quad \text{for all } \kappa \in [0, \infty).$$

Note that the function f decreases monotonically. Together with the monotonicity of $t \mapsto 1 - \exp(-t)$, this allows us to estimate

$$\sum_{k=1}^\infty \left(1 - \exp\left(-\frac{c_k}{b}\right)\right) \leq \sum_{k=1}^\infty (1 - \exp(-f(k))) \leq \int_{\kappa=0}^\infty (1 - \exp(-f(\kappa))) \, \mathrm{d}\kappa.$$

Substituting $t := \frac{C_-}{2}\kappa^{\frac{2}{d}}$ and setting $C_1 := \frac{\sqrt{2}}{b}\|u^\dagger\|_X \left(\frac{C_+\beta}{C_- e}\right)^{\beta/2}$ leads to

$$\sum_{k=1}^{\infty}\left(1-\exp\left(-\frac{c_k}{b}\right)\right) \leq \int_0^\infty \left(1-\exp\left(-C_1 e^{-\frac{C_-}{2}\kappa^{\frac{2}{d}}}\right)\right) d\kappa$$

$$= \int_0^\infty \left(1-\exp\left(-C_1 e^{-t}\right)\right)\left(\frac{2}{C_-}\right)^{\frac{d}{2}} \frac{d}{2} t^{\frac{d}{2}-1} dt$$

We assume that b is small enough, split up the integral at

$$T = T(b) := \ln C_1,$$

and use the estimate $1-\exp(-t) \leq \min\{t,1\}$ for all $t \geq 0$ to arrive at

$$\sum_{k=1}^{\infty}\left(1-\exp\left(-\frac{c_k}{b}\right)\right) \leq \int_0^T \left(\frac{2}{C_-}\right)^{\frac{d}{2}} \frac{d}{2} t^{\frac{d}{2}-1} dt + \int_T^\infty C_1 \left(\frac{2}{C_-}\right)^{\frac{d}{2}} \frac{d}{2} t^{\frac{d}{2}-1} e^{-t} dt$$

$$= \left(\frac{2}{C_-}\right)^{\frac{d}{2}} T^{\frac{d}{2}} + \left(\frac{2}{C_-}\right)^{\frac{d}{2}} \frac{d}{2} e^T \Gamma\left(\frac{d}{2}, T\right).$$

Applying Lemma 5.45 with $a = d/2$ and $x = T$ results in

$$\sum_{k=1}^{\infty}\left(1-\exp\left(-\frac{c_k}{b}\right)\right) \leq \left(\frac{2}{C_-}\right)^{\frac{d}{2}} \left(T^{\frac{d}{2}} + \frac{d}{2} e^T \left(T^{\frac{d}{2}-1} + \left|\frac{d}{2}-1\right| T^{\frac{d}{2}-2}\right) e^{-T}\right)$$

$$= \left(\frac{2}{C_-}\right)^{\frac{d}{2}} T^{\frac{d}{2}} \left(1 + \frac{d}{2} T^{-1} + \frac{d}{2}\left|\frac{d}{2}-1\right| T^{-2}\right)$$

for $d \in \{1,2,3,4\}$. As b tends to 0, $T(b) \longrightarrow \infty$, so that for every $C > 1$ there is an $\varepsilon > 0$, such that

$$1 + \frac{d}{2} T^{-1} + \frac{d}{2}\left|\frac{d}{2}-1\right| T^{-2} \leq C \quad \text{for all } b \in (0,\varepsilon).$$

This implies

$$\mathbb{E}\left[\Psi(u^\dagger, y)\right] = b\sum_{k=1}^{\infty}\left(1-\exp\left(-\frac{c_k}{b}\right)\right) \leq \left(\frac{2}{C_-}\right)^{\frac{d}{2}} Cb(\ln C_1)^{\frac{d}{2}}$$

$$\leq C\left(\frac{2}{C_-}\right)^{\frac{d}{2}} b \ln\left(\frac{\sqrt{2}}{b}\|u^\dagger\|_X \left(\frac{C_+\beta}{C_- e}\right)^{\frac{\beta}{2}}\right)^{\frac{d}{2}}$$

for all $b \in (0,\varepsilon)$. $\square$

5.5.5 Convergence Rate of the Bias

We examine the rate of convergence of the bias under a source condition. To this end, we first compute the expectation of the components of the MAP estimator explicitly.

Lemma 5.47. *Let $u^\dagger \in X$, $b > 0$ and $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{s-\beta}}$. Then*

$$
\begin{aligned}
&\mathbb{E}\left[(\hat{u}_{\mathrm{MAP}}(y), \varphi_k)_X\right] - (u^\dagger, \varphi_k)_X \\
&= b\frac{1}{2c_k}\left(\exp\left(-\frac{1}{b}c_k\left|(u^\dagger, \varphi_k)_X + \frac{r^2}{b}c_k\alpha_k^{-\tau}\right|\right) - \exp\left(-\frac{1}{b}c_k\left|(u^\dagger, \varphi_k)_X - \frac{r^2}{b}c_k\alpha_k^{-\tau}\right|\right)\right) \\
&\quad + \chi_{(-\infty, -(u^\dagger, \varphi_k)_X)}\left(\frac{r^2}{b}c_k\alpha_k^{\tau}\right)\cdot\left|(u^\dagger, \varphi_k)_X + \frac{r^2}{b}c_k\alpha_k^{-\tau}\right| \\
&\quad - \chi_{(-\infty, (u^\dagger, \varphi_k)_X)}\left(\frac{r^2}{b}c_k\alpha_k^{\tau}\right)\cdot\left|(u^\dagger, \varphi_k)_X - \frac{r^2}{b}c_k\alpha_k^{-\tau}\right|
\end{aligned}
$$

for all $k \in \mathbb{N}$, where $c_k := \sqrt{2}\alpha_k^{\beta/2}e^{-\alpha_k}$.

Proof. Consider a fixed $k \in \mathbb{N}$. By Lemma 5.34 and the definition of the $\mathcal{X}^s$ norm the k-th component of the MAP estimator $\bar{u} := \hat{u}_{\mathrm{MAP}}(y)$ is given by

$$
\begin{aligned}
(\bar{u}(y), \varphi_k)_X &= \max\left\{-\sqrt{2}\frac{r^2}{b}\alpha_k^{\frac{\beta}{2}-\tau}e^{-\alpha_k}, \min\left\{e^{\alpha_k}(y, \varphi_k)_X, \sqrt{2}\frac{r^2}{b}\alpha_k^{\frac{\beta}{2}-\tau}e^{-\alpha_k}\right\}\right\} \\
&= \alpha_k^{-\frac{s}{2}}e^{\alpha_k}\max\left\{-\sqrt{2}\frac{r^2}{b}\alpha_k^{\frac{s}{2}+\frac{\beta}{2}-\tau}e^{-2\alpha_k}, \min\left\{(y, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s}, \sqrt{2}\frac{r^2}{b}\alpha_k^{\frac{s}{2}+\frac{\beta}{2}-\tau}e^{-2\alpha_k}\right\}\right\} \\
&= \gamma\max\left\{-R, \min\left\{x, R\right\}\right\}, \qquad (5.11)
\end{aligned}
$$

where

$$
x := (y, \alpha_k^{-\frac{s}{2}}\varphi_k)_{\mathcal{X}^s}, \quad \gamma := \alpha_k^{-\frac{s}{2}}e^{\alpha_k} \quad \text{and} \quad R := \frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{s}{2}+\frac{\beta}{2}-\tau}e^{-2\alpha_k}.
$$

Now $x = x(y) \sim \mathcal{L}_{a,\lambda}$ with

$$
a := (e^{-A}u^\dagger, e_k)_{\mathcal{X}^s} = \frac{1}{\gamma}(u^\dagger, \varphi_k)_X \quad \text{and} \quad \lambda := b^2\alpha_k^{s-\beta}.
$$

We compute

$$
\begin{aligned}
\mathbb{E}\left[(\bar{u}(y), \varphi_k)_X\right] &= \int_{\mathcal{X}^s}(\bar{u}(y), \varphi_k)_X\,\mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}(\mathrm{d}y) \\
&= \int_{\mathbb{R}}\gamma\max\{-R, \min\{x, R\}\}\mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&= \int_{-\infty}^{-R}(-\gamma R)\,\mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{-R}^{R}\gamma x\mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{R}^{\infty}\gamma R\mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&= \int_{\mathbb{R}}\gamma x\mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{-\infty}^{-R}\gamma\,(-R - x)\,\mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{R}^{\infty}\gamma\,(R - x)\,\mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&=: I_1 + I_2 + I_3.
\end{aligned}
$$

Note that $I_1 = \gamma \int_{\mathbb{R}} x \mathcal{L}_{a,\lambda}(\mathrm{d}x) = \gamma a = (u^\dagger, \varphi_k)_X$. For $R < a$, the last integral computes as

$$\begin{aligned}
I_3 &= \int_R^a \gamma\,(R-x)\,\frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(a-x)}\mathrm{d}x + \int_a^\infty \gamma\,(R-x)\,\frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\mathrm{d}x \\
&= \left[\gamma\,(R-x)\,\frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(a-x)}\right]\Bigg|_R^a - \int_R^a \left(-\gamma\frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(a-x)}\right)\mathrm{d}x \\
&\quad + \left[-\gamma\,(R-x)\,\frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\right]\Bigg|_a^\infty - \int_a^\infty \gamma\frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\mathrm{d}x \\
&= \frac{\gamma}{2}(R-a) - \left[-\frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}(a-x)}\right]\Bigg|_R^a + \frac{\gamma}{2}(R-a) - \left[-\frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\right]\Bigg|_a^\infty \\
&= -\gamma\,(a-R) - \frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}(a-R)}.
\end{aligned}$$

Otherwise, that is if $R \geq a$,

$$\begin{aligned}
I_3 &= \int_R^\infty \gamma\,(R-x)\,\frac{1}{\sqrt{2\lambda}} e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\mathrm{d}x \\
&= \left[-\frac{\gamma}{2}(R-x)\,e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\right]\Bigg|_R^\infty - \left[-\frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}(x-a)}\right]\Bigg|_R^\infty \\
&= -\frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}(R-a)}
\end{aligned}$$

We can combine these to

$$I_3 = -\frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}|R-a|} - \chi_{(-\infty,a)}(R)\gamma\,|R-a|\,,$$

where χ_M denotes the characteristic function of a set M. A similar computation yields

$$I_2 = \frac{\gamma}{2}\sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}|R-(-a)|} + \chi_{(-\infty,-a)}(R)\gamma\,|R-(-a)|\,.$$

By summing up, we obtain

$$\begin{aligned}
\mathbb{E}\left[(\bar{u},\varphi_k)_X\right] - (u^\dagger,\varphi_k)_X &= \frac{\gamma}{2}\sqrt{\frac{\lambda}{2}}\left(e^{-\sqrt{\frac{2}{\lambda}}|R-(-a)|} - e^{-\sqrt{\frac{2}{\lambda}}|R-a|}\right) \\
&\quad + \gamma\left(\chi_{(-\infty,-a)}(R)\,|R-(-a)| - \chi_{(-\infty,a)}(R)\,|R-a|\right).
\end{aligned}$$

Resubstituting a, λ, R and γ results in

$$\begin{aligned}
&\mathbb{E}\left[(\bar{u}, \varphi_k)_X\right] - (u^\dagger, \varphi_k)_X \\
&= b\frac{1}{2\sqrt{2}}\alpha_k^{-\frac{\beta}{2}} e^{\alpha_k}\left(\exp\left(-\frac{1}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}} e^{-\alpha_k}\left|(u^\dagger, \varphi_k)_X + \frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right|\right)\right. \\
&\quad - \exp\left(-\frac{1}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}} e^{-\alpha_k}\left|(u^\dagger, \varphi_k)_X - \frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right|\right)\Bigg) \\
&\quad + \chi_{(-\infty,-(u^\dagger,\varphi_k)_X)}\left(\frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right)\cdot\left|(u^\dagger, \varphi_k)_X + \frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right| \\
&\quad - \chi_{(-\infty,(u^\dagger,\varphi_k)_X)}\left(\frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right)\cdot\left|(u^\dagger, \varphi_k)_X - \frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}\right|.
\end{aligned}$$

Inserting $c_k = \sqrt{2}\alpha_k^{\frac{\beta}{2}} e^{-\alpha_k}$ finishes the proof. □

Under a source condition we can show that the bias converges at least in the order of the noise level.

Theorem 5.48. *Let $u^\dagger \in X$, $b, r > 0$ and $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}$. If a $w \in X$ exists, such that*

$$u^\dagger = A^{\frac{\beta}{2}-\tau} e^{-2A} w \quad \textit{and} \quad \sup_{k\in\mathbb{N}}|(w, \varphi_k)_X| \leq \rho,$$

and if

$$r^2 \geq \rho b,$$

then

$$\left\|\mathbb{E}\left[\hat{u}_{\mathrm{MAP}}(y)\right] - u^\dagger\right\|_X \leq \frac{1}{4}\left(\operatorname{Tr} A^{-\beta}\right)^{\frac{1}{2}} b.$$

Proof. We have

$$\mathbb{E}[\hat{u}_{\mathrm{MAP}}] = \int_{\mathcal{X}^s} \hat{u}_{\mathrm{MAP}}(y)\mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}(\mathrm{d}y) \in \mathcal{X}^\tau.$$

By Proposition 5.10, $\mathcal{X}^\tau$ is continuously embedded into X, so that $(\cdot, \varphi_k)_X$ is continuous on $\mathcal{X}^\tau$. This allows us to write

$$\left\|\mathbb{E}\left[\hat{u}_{\mathrm{MAP}}\right] - u^\dagger\right\|_X^2 = \sum_{k=1}^{\infty}\left|(\mathbb{E}\left[\hat{u}_{\mathrm{MAP}}\right], \varphi_k)_X - (u^\dagger, \varphi_k)_X\right|^2 = \sum_{k=1}^{\infty}\left|\mathbb{E}\left[(\hat{u}_{\mathrm{MAP}}, \varphi_k)_X\right] - (u^\dagger, \varphi_k)_X\right|^2.$$

We want to use Lemma 5.47. The assumptions on $u^\dagger$ and r ensure

$$\begin{aligned}
\left|(u^\dagger, \varphi_k)_X\right| &= \left|\left(A^{\frac{\beta}{2}-\tau} e^{-2A} w, \varphi_k\right)_X\right| = \left|\left(w, e^{-2A}A^{\frac{\beta}{2}-\tau}\varphi_k\right)_X\right| \\
&= e^{-2\alpha_k}\alpha_k^{\frac{\beta}{2}-\tau}\left|(w, \varphi_k)_X\right| \leq \alpha_k^{\frac{\beta}{2}-\tau} e^{-2\alpha_k}\rho \leq \frac{r^2}{b}\alpha_k^{\frac{\beta}{2}-\tau} e^{-2\alpha_k}
\end{aligned} \tag{5.12}$$

for all $k \in \mathbb{N}$. In particular, $|(u^\dagger, \varphi_k)_X| \leq \sqrt{2}\frac{r^2}{b}\alpha_k^{\beta/2-\tau} e^{-\alpha_k}$ holds. Thus the last two terms in the expression in Lemma 5.47 are equal to zero, that is

$$\begin{aligned}\mathbb{E}\left[(\hat{u}_{\mathrm{MAP}}, \varphi_k)_X\right] - (u^\dagger, \varphi_k)_X &= b\frac{1}{2c_k}\left(\exp\left(-\frac{1}{b}c_k\left((u^\dagger, \varphi_k)_X + \frac{r^2}{b}c_k\alpha_k^{-\tau}\right)\right)\right.\\ &\qquad \left. - \exp\left(-\frac{1}{b}c_k\left(\frac{r^2}{b}c_k\alpha_k^{-\tau} - (u^\dagger, \varphi_k)_X\right)\right)\right)\\ &= b\frac{1}{c_k}\exp\left(-\frac{r^2}{b^2}c_k^2\alpha_k^{-\tau}\right)\sinh\left(-\frac{1}{b}c_k(u^\dagger, \varphi_k)_X\right),\end{aligned}$$

where $c_k := \sqrt{2}\alpha_k^{\beta/2}e^{-\alpha_k}$. The hyperbolic sine is convex on $[0, \infty)$ and odd, which leads to the estimate

$$|\sinh(t)| = \sinh(|t|) \leq \frac{|t|}{T}\sinh(T) \leq \frac{|t|}{2T}e^T$$

for all $t \in \mathbb{R}$ and $T \geq |t|$. We apply this inequality with $t := -\frac{1}{b}c_k(u^\dagger, \varphi_k)_X$ and $T := \frac{r^2}{b^2}c_k^2\alpha_k^{-\tau}$ as well as (5.12) and obtain

$$\begin{aligned}\left|\mathbb{E}\left[(\hat{u}_{\mathrm{MAP}}, \varphi_k)_X\right] - (u^\dagger, \varphi_k)_X\right| &= b\frac{1}{c_k}\exp\left(-T\right)\left|\sinh\left(t\right)\right| \leq b\frac{1}{c_k}\frac{|t|}{2T}\\ &= \left|(u^\dagger, \varphi_k)_X\right|\left(\frac{b^2}{4r^2}\alpha_k^{\tau-\beta}e^{2\alpha_k}\right) \leq \frac{1}{4}\alpha_k^{-\frac{\beta}{2}}b.\end{aligned}$$

Squaring and summing up results in

$$\left\|\mathbb{E}\left[\hat{u}_{\mathrm{MAP}}\right] - u^\dagger\right\|_X^2 \leq \sum_{k=1}^{\infty}\frac{1}{16}\alpha_k^{-\beta}b^2 = \frac{1}{16}\left(\operatorname{Tr} A^{-\beta}\right)b^2,$$

which finishes the proof. □

Note that Theorem 5.48 yields convergence of the bias in the order of b as $b \longrightarrow 0$ even if r is chosen constant.

5.5.6 Convergence Rate of the Mean Squared Error

Now we consider the rate of convergence of the mean squared error under a source condition. By Lemma 5.34, the components of $\hat{u}_{\mathrm{MAP}}$ are independent, so that

$$\mathbb{E}\left[\|\hat{u}_{\mathrm{MAP}} - u^\dagger\|_X^2\right] = \mathbb{E}\left[\sum_{k=1}^{\infty}\left|(\hat{u}_{\mathrm{MAP}} - u^\dagger, \varphi_k)_X\right|^2\right] = \sum_{k=1}^{\infty}\mathbb{E}\left[\left|(\hat{u}_{\mathrm{MAP}} - u^\dagger, \varphi_k)_X\right|^2\right].$$

We first compute the componentwise expectations.

Lemma 5.49. *Let $u^\dagger \in X$, $b > 0$ and $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{s-\beta}}$. Then*

$$\begin{aligned}
&\mathbb{E}\left[\left|(\hat{u}_{\mathrm{MAP}}(y) - u^\dagger, \varphi_k)_X\right|^2\right] \\
&= \frac{b^2}{c_k^2} f\left(\frac{c_k}{b}\left|r^2\frac{c_k}{b}\alpha_k^{-\tau} + \left|(u^\dagger, \varphi_k)_X\right|\right|\right) + \frac{b^2}{c_k^2} f\left(\frac{c_k}{b}\left|r^2\frac{c_k}{b}\alpha_k^{-\tau} - \left|(u^\dagger, \varphi_k)_X\right|\right|\right) \\
&\quad + \chi_{(-\infty, |(u^\dagger, \varphi_k)_X|)}\left(r^2\frac{c_k}{b}\alpha_k^{-\tau}\right)\frac{b^2}{c_k^2} g\left(\frac{c_k}{b}\left|r^2\frac{c_k}{b}\alpha_k^{-\tau} - \left|(u^\dagger, \varphi_k)_X\right|\right|\right).
\end{aligned}$$

for all $k \in \mathbb{N}$, where $c_k := \sqrt{2}\alpha_k^{\beta/2} e^{-\alpha_k}$,

$$f(t) := 1 - e^{-t} - te^{-t} \quad \text{and} \quad g(t) := t^2 - 2f(t)$$

for all $t \geq 0$.

Proof. Let $k \in \mathbb{N}$ be arbitrary, but fixed, and set $\bar{u} := \hat{u}_{\mathrm{MAP}}(y)$. By Lemma 5.34 and (5.11) we have

$$(\bar{u}, \varphi_k)_X = \gamma \max\{-R, \min\{x, R\}\},$$

where $x = x(y) := (y, \alpha_k^{-s/2}\varphi_k)_{\mathcal{X}^s}$, $\gamma := \alpha_k^{-s/2} e^{\alpha_k}$ and $R := \frac{r^2}{b}\sqrt{2}\alpha_k^{s/2+b/2-\tau} e^{-2\alpha_k}$. As $x \sim \mathcal{L}_{a,\lambda}$ with $a := \frac{1}{\gamma}(u^\dagger, \varphi_k)_X$ and $\lambda := b^2\alpha_k^{s-\beta}$, we have

$$\begin{aligned}
\mathbb{E}\left[|(\bar{u} - u^\dagger, \varphi_k)_X|^2\right] &= \int_{\mathcal{X}^s} |(\bar{u}, \varphi_k)_X - (u^\dagger, \varphi_k)_X|^2 \mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}(\mathrm{d}y) \\
&= \int_{\mathbb{R}} |\gamma \max\{-R, \min\{x, R\}\} - \gamma a|^2 \mathcal{L}_{a,\lambda}(x) \\
&= \int_{-\infty}^{-R} \gamma^2 |-R - a|^2 \mathcal{L}_{a,\lambda}(\mathrm{d}x) + \int_{-R}^{R} \gamma^2 |x - a|^2 \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&\quad + \int_R^\infty \gamma^2 |R - a|^2 \mathcal{L}_{a,\lambda}(\mathrm{d}x) \\
&= \gamma^2|-R-a|^2 \int_{-\infty}^{-R-a} \mathcal{L}_\lambda(\mathrm{d}z) + \gamma^2 |R - a|^2 \int_{R-a}^\infty \mathcal{L}_\lambda(\mathrm{d}z) \\
&\quad + \gamma^2 \int_{-R-a}^{R-a} z^2 \mathcal{L}_\lambda(\mathrm{d}z) \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

Now

$$\begin{aligned}
I_1 + I_2 &= \gamma^2|-R-a|^2 \int_{-\infty}^{-R-a} \mathcal{L}_\lambda(\mathrm{d}z) + \gamma^2|-R+a|^2 \int_{-\infty}^{-R+a} \mathcal{L}_\lambda(\mathrm{d}z) \\
&= \gamma^2|-R-|a||^2 \int_{-\infty}^{-R-|a|} \mathcal{L}_\lambda(\mathrm{d}z) + \gamma^2|-R+|a||^2 \int_{-\infty}^{-R+|a|} \mathcal{L}_\lambda(\mathrm{d}z).
\end{aligned}$$

In case $R \geq |a|$ this equals

$$I_1 + I_2 = \gamma^2|-R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} + \gamma^2|-R+|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R+|a||},$$

and the remaining integral equates to

$$\begin{aligned}
I_3 &= \gamma^2 \int_{-R-a}^{0} z^2 \mathcal{L}_\lambda(\mathrm{d}z) + \gamma^2 \int_{-R+a}^{0} z^2 \mathcal{L}_\lambda(\mathrm{d}z) \\
&= \gamma^2 \int_{-R-|a|}^{0} z^2 \mathcal{L}_\lambda(\mathrm{d}z) + \gamma^2 \int_{-R+|a|}^{0} z^2 \mathcal{L}_\lambda(\mathrm{d}z) \\
&= -\gamma^2 |-R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} + \gamma^2 \frac{\lambda}{2} f\left(\sqrt{\frac{2}{\lambda}}|-R-|a||\right) \\
&\quad - \gamma^2 |-R+|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R+|a||} + \gamma^2 \frac{\lambda}{2} f\left(\sqrt{\frac{2}{\lambda}}|-R+|a||\right),
\end{aligned}$$

where $f(t) := 1 - e^{-t} - te^{-t}$ for all $t \geq 0$, which adds up to

$$\begin{aligned}
\mathbb{E}\left[\left|(\bar{u} - u^\dagger, \varphi_k)_X\right|^2\right] &= \gamma^2 \frac{\lambda}{2} f\left(\sqrt{\frac{2}{\lambda}}|-R-|a||\right) + \gamma^2 \frac{\lambda}{2} f\left(\sqrt{\frac{2}{\lambda}}|R-|a||\right) \\
&= \frac{b^2}{c_k^2} f\left(\frac{c_k}{b}\left|r^2 \frac{c_k}{b}\alpha_k^{-\tau} + \left|(u^\dagger, \varphi_k)_X\right|\right|\right) \\
&\quad + \frac{b^2}{c_k^2} f\left(\frac{c_k}{b}\left|r^2 \frac{c_k}{b}\alpha_k^{-\tau} - \left|(u^\dagger, \varphi_k)_X\right|\right|\right)
\end{aligned}$$

for $R \geq |a|$.

In case $R < |a|$ we compute

$$I_1 + I_2 = \gamma^2 |-R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} + \gamma^2 |R-|a||^2 \left(1 - \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|R-|a||}\right).$$

As

$$I_3 = \gamma^2 \int_{-R-a}^{R-a} z^2 \mathcal{L}_\lambda(\mathrm{d}z) = \gamma^2 \int_{-R+a}^{R+a} z^2 \mathcal{L}_\lambda(\mathrm{d}z),$$

we can compute this integral as

$$\begin{aligned}
I_3 &= \gamma^2 \int_{-R-|a|}^{R-|a|} z^2 \mathcal{L}_\lambda(\mathrm{d}z) \\
&= \gamma^2 |R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|R-|a||} - \gamma^2 |-R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} \\
&\quad + \gamma^2 |R-|a|| \sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}|R-|a||} + \gamma^2 \frac{\lambda}{2} e^{-\sqrt{\frac{2}{\lambda}}|R-|a||} \\
&\quad - \gamma^2 |-R-|a|| \sqrt{\frac{\lambda}{2}} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} - \gamma^2 \frac{\lambda}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} \\
&= \gamma^2 |R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|R-|a||} - \gamma^2 |-R-|a||^2 \frac{1}{2} e^{-\sqrt{\frac{2}{\lambda}}|-R-|a||} \\
&\quad + \gamma^2 \frac{\lambda}{2} f\left(\sqrt{\frac{2}{\lambda}}|-R-|a||\right) - \gamma^2 \frac{\lambda}{2} f\left(\sqrt{\frac{2}{\lambda}}|R-|a||\right).
\end{aligned}$$

This adds up to

$$\begin{aligned}\mathbb{E}\left[\left|(\bar u - u^\dagger, \varphi_k)_X\right|^2\right] &= \gamma^2|R-|a||^2 + \gamma^2\frac{\lambda}{2}f\left(\sqrt{\frac{2}{\lambda}}|-R-|a||\right) - \gamma^2\frac{\lambda}{2}f\left(\sqrt{\frac{2}{\lambda}}|R-|a||\right)\\ &= \frac{b^2}{c_k^2}f\left(\frac{c_k}{b}\left|r^2\frac{c_k}{b}\alpha_k^{-\alpha} + \left|(u^\dagger,\varphi_k)_X\right|\right|\right) + \frac{b^2}{c_k^2}\left(\frac{c_k}{b}\left|r^2\frac{c_k}{b}\alpha_k^{-\tau} - \left|(u^\dagger,\varphi_k)_X\right|\right|\right)^2\\ &\quad - \frac{b^2}{c_k^2}f\left(\frac{c_k}{b}\left|r^2\frac{c_k}{b}\alpha_k^{-\tau} - \left|(u^\dagger,\varphi_k)_X\right|\right|\right)\end{aligned}$$

in case that $R < |a|$. □

Now we can show that under a source condition the mean squared error converges at least in the order of the noise level.

Theorem 5.50. *Let $u^\dagger \in X$, $b, r > 0$ and $y \sim \mathcal{L}_{e^{-A}u^\dagger, b^2A^{s-\beta}}$. If a $w \in X$ exists, such that*

$$u^\dagger = A^{\frac{\beta}{2}-\tau}e^{-A}w \quad \textit{and} \quad \sup_{k\in\mathbb{N}}|(w,\varphi_k)_X| \le \rho,$$

and if there is a $C > 0$, such that

$$\frac{\rho}{\sqrt{2}}b \le r^2 \le Cb,$$

then

$$\mathbb{E}\left[\|\hat u_{\mathrm{MAP}}(y) - u^\dagger\|_X^2\right] \le 2C\,(\operatorname{Tr}A^{-\tau})\,b.$$

Proof. Since the components of $\hat u_{\mathrm{MAP}}$ are independent by Lemma 5.34, we have

$$\mathbb{E}\left[\|\hat u_{\mathrm{MAP}} - u^\dagger\|_X^2\right] = \mathbb{E}\left[\sum_{k=1}^\infty \left|(\hat u_{\mathrm{MAP}} - u^\dagger, \varphi_k)_X\right|^2\right] = \sum_{k=1}^\infty \mathbb{E}\left[\left|(\hat u_{\mathrm{MAP}} - u^\dagger,\varphi_k)_X\right|^2\right].$$

We apply Lemma 5.49. The requirements on $u^\dagger$ and r ensure

$$\begin{aligned}\left|(u^\dagger,\varphi_k)_X\right| &= \left|\left(A^{\frac{\beta}{2}-\tau}e^{-A}w,\varphi_k\right)_X\right| = \left|\left(w, e^{-A}A^{\frac{\beta}{2}-\tau}\varphi_k\right)_X\right|\\ &= e^{-\alpha_k}\alpha_k^{\frac{\beta}{2}-\tau}\left|(w,\varphi_k)_X\right| \le \alpha_k^{\frac{\beta}{2}-\tau}e^{-\alpha_k}\rho \le \frac{r^2}{b}\sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau}e^{-\alpha_k}\end{aligned}$$

for all $k \in \mathbb{N}$ and $n \ge N$. Thus the last term in the expression in Lemma 5.49 is equal to zero, that is,

$$\begin{aligned}\mathbb{E}\left[\left|(\hat u_{\mathrm{MAP}} - u^\dagger,\varphi_k)_X\right|^2\right] &= \frac{b^2}{c_k^2}f\left(\frac{c_k}{b}\left(r^2\frac{c_k}{b}\alpha_k^{-\tau} + \left|(u^\dagger,\varphi_k)_X\right|\right)\right)\\ &\quad + \frac{b^2}{c_k^2}f\left(\frac{c_k}{b}\left(r^2\frac{c_k}{b}\alpha_k^{-\tau} - \left|(u^\dagger,\varphi_k)_X\right|\right)\right)\end{aligned}$$

for all $k \in \mathbb{N}$, where $c_k := \sqrt{2}\alpha_k^{\beta/2} e^{-\alpha_k}$ and $f(t) := 1 - e^{-t} - te^{-t}$ for all $t \geq 0$. We use the estimate

$$f(t) \leq 1 - e^{-t} \leq t,$$

that holds for all $t \geq 0$, and obtain

$$\mathbb{E}\left[\left|(\hat{u}_{\mathrm{MAP}} - u^\dagger, \varphi_k)_X\right|^2\right] \leq 2r^2 \alpha_k^{-\tau}.$$

Consequently, by the choice of r, we have

$$\mathbb{E}\left[\|\hat{u}_{\mathrm{MAP}} - u^\dagger\|_X^2\right] \leq 2r^2 \sum_{k=1}^{\infty} \alpha_k^{-\tau} = 2\,\mathrm{Tr}\,(A^{-\tau})\, r^2 \leq 2C\,\mathrm{Tr}\,(A^{-\tau})\, b. \qquad \square$$

Theorem 5.50 shows in particular that the MAP estimator is consistent, since its convergence towards the true solution in mean square implies convergence in probability by Markov's inequality.

We classify the stated rate of convergence by comparing it with the optimal convergence rate of the minimax risk in the case of Gaussian noise. Here we consider the setting of [Ding and Mathé 2017], which provides a comparably general framework for deriving minimax rates. Minimax rates for the particular case of an exponentially ill-posed problem with analytic smoothness of the solution have been established in general in [Cavalier et al. 2004], and for a specific problem in [Golubev and Khasminskii 2001].

We fix the dimension $d = 2$ and assume that the eigenvalues of A associated with the eigenvectors φ_k are exactly

$$\alpha_k = pk^{\frac{2}{d}} = pk \quad \text{for all } k \in \mathbb{N}. \tag{5.13}$$

Moreover, we assume the presence of Gaussian noise

$$\eta \sim \mathcal{N}_{A^{s-\beta}}$$

instead of Laplacian noise.

Now we can restate the problem within the framework of [Ding and Mathé 2017]. The model considered therein is

$$y^\sigma = Tx + \sigma\xi, \tag{5.14}$$

where T is a compact linear operator between Hilbert spaces X and Y with singular system $\{(s_j, u_j, v_j)\}_{j\in\mathbb{N}}$, i.e.,

$$Tx = \sum_{j=1}^{\infty} s_j (x, v_j)_X u_j,$$

ξ is Gaussian white noise, $\sigma > 0$ is the noise level and y^σ is the noisy data. If $s_j \asymp e^{-pj}$ then the problem is called *severely ill-posed* or *exponentially ill-posed*. We can bring equation (5.3) into this form by choosing

$$T = A^{\frac{\beta}{2}} e^{-A},$$

$\sigma = b$, $x = u$, $s_j = (pj)^{\frac{\beta}{2}} e^{-pj} \asymp e^{-pj}$ and $v_j = u_j = \varphi_j$ for all $j \in \mathbb{N}$.

Model (5.14) in turn is equivalent to the sequence space model

$$z_j^\sigma = \theta_j + \sigma\sigma_j\xi_j, \quad \xi_j \sim \mathcal{N}_{0,1},$$

where $\theta_j = (x, v_j)_X = (u, \varphi_j)_X$ and $\sigma_j = s_j^{-1} = (pj)^{-\frac{\beta}{2}} e^{pj}$, if x is in the orthogonal complement of the kernel of T. This is the case because both e^{-A} and $A^{\frac{\beta}{2}}$ are injective.

The true solution is assumed to be an element of a *Sobolev-type ellipsoid*

$$\Theta_a(Q) = \left\{ \theta = (\theta_j)_{j=1}^\infty : \sum_{j=1}^\infty a_j^2 \theta_j^2 \le Q^2 \right\},$$

where $a = (a_j)_{j=1}^\infty$, $a_j > 0$, is a given increasing sequence and $Q \in \mathbb{R}$. If $a_j \asymp e^{\kappa j}$ for some $\kappa > 0$, then the solution is called *analytic*. We obtain the source condition from Theorem 5.50 by choosing $a_j = (pj)^{\tau - \frac{\beta}{2}} e^{pj} \asymp e^{pj}$, so that $\kappa = p$, and $Q = \rho^{\frac{1}{2}}$. In order for a to be increasing, we moreover assume that $\tau \ge \frac{\beta}{2}$.

The considered risk is the root-mean-square (RMS) error $(\mathbb{E}[\|\hat{\theta} - \theta\|_2^2])^{\frac{1}{2}}$ of an estimator $\hat{\theta} = \hat{\theta}(z^\sigma)$. The minimax risk on the class Θ_a is then defined as

$$e(\Theta_a, \sigma) := \inf_{\hat{u}} \sup_{\theta \in \Theta_a} \left(\mathbb{E}\left[\|\hat{\theta} - \theta\|_2^2 \right] \right)^{\frac{1}{2}},$$

where the infimum is again taken over all estimators $\hat{\theta} = \hat{\theta}(z^\sigma)$ that are based on the data z^σ. The main result of [Ding and Mathé 2017] now states that

$$e(\Theta_a, \sigma) \le \inf_{D \in \mathbb{N}} \left(\frac{Q^2}{a_{D+1}^2} + \sigma^2 \sum_{j=1}^D \frac{1}{s_j^2} \right)^{\frac{1}{2}} \le 2.2 e(\Theta_a, \sigma).$$

As a consequence of this, a minimax rate

$$e(\Theta_a, \sigma) \asymp \sigma^{\frac{\kappa}{p+\kappa}} = \sigma^{\frac{1}{2}} \quad \text{as } \sigma \longrightarrow 0$$

follows for severely ill-posed problems with analytically smooth solution. This translates into our notation as follows:

$$\inf_{\hat{u}} \sup_{u^\dagger \in \Theta_a} \mathbb{E}\left[\|\hat{u} - u^\dagger\|_2^2 \right] \asymp b \quad \text{as } b \longrightarrow 0.$$

Note that here the choice $\kappa = p$ results in a rate independent of p.

Considered in this context, Theorem 5.50 provides an upper bound for the minimax rate in case of Laplacian noise, namely

$$\inf_{\hat{u}} \sup_{u^\dagger \in \Theta_a} \mathbb{E}\left[\|\hat{u} - u^\dagger\|_2^2 \right] \le \sup_{u^\dagger \in \Theta_a} \mathbb{E}\left[\|\hat{u}_{\mathrm{MAP}} - u^\dagger\|_2^2 \right] \le 2C(\operatorname{Tr} A^{-\tau})b.$$

This shows that at least for $d = 2$ and $\alpha_k = pk$ the optimal rate of convergence of the minimax risk for Laplacian noise is not worse than for Gaussian noise.

5.6 Conditional Mean Estimator

Here we compute the conditional mean estimator $\hat{u}_{\mathrm{CM}}$ explicitly and use the results from Section 5.4.2 to show that small changes in the data $y \in \mathcal{X}^s$ cause only small changes in the CM estimator.

The *conditional mean (CM) estimator* $\hat{u}_{\mathrm{CM}}\colon \mathcal{X}^s \to X$ is defined by

$$\hat{u}_{\mathrm{CM}}(y) \coloneqq \mathbb{E}^{\mu^y}[u] = \int_X u\, \mu^y(\mathrm{d}u),$$

i.e., for every $y \in \mathcal{X}^s$, $\hat{u}_{\mathrm{CM}}(y)$ is the mean of the posterior distribution μ^y. We consider a single component

$$(\hat{u}_{\mathrm{CM}}(y), \varphi_m)_X = \left(\int_X u\, \mu^y(\mathrm{d}u), \varphi_m \right)_X .$$

As $(\cdot, \varphi_m)_X$ is a continuous linear functional, this equals

$$\begin{aligned}(\hat{u}_{\mathrm{CM}}(y), \varphi_m)_X &= \int_X (u, \varphi_m)_X \mu^y(\mathrm{d}u) = \int_X (u, \varphi_m)_X \frac{1}{Z(y)} \exp(-\Phi(u, y))\mu_0(\mathrm{d}u) \\ &= \frac{\int_X (u, \varphi_m)_X \exp(-\Phi(u, y))\mu_0(\mathrm{d}u)}{\int_X \exp(-\Phi(u, y))\mu_0(\mathrm{d}u)}.\end{aligned}$$

We work with the orthogonal projections onto finite dimensional subspaces P_n, defined by

$$P_n u \coloneqq \sum_{k=1}^{n} (u, \varphi_k)_X \varphi_k$$

for all $u \in X$ and $n \in \mathbb{N}$. Now

$$\exp(-\Phi(P_n u, y)) \to \exp(-\Phi(u, y))$$

for all $u \in X$ because $P_n u \to u$ as $n \to \infty$ and Φ is continuous in u. Moreover,

$$|\exp(-\Phi(P_n u, y))| \leq \exp(L\|P_n u\|_X) \leq \exp(L\|u\|_X)$$

by Proposition 5.23, where $L > 0$ is the Lipschitz constant of $\Phi(\cdot, y)$, and $\exp(L\|\cdot\|_X) \in L^1(X, \mu_0)$ by Lemma 5.26, so that $\exp(-\Phi(P_n \cdot, y)) \in L^1(X, \mu_0)$ for all $n \in \mathbb{N}$, too. We may apply Lebesgue's dominated convergence theorem [Klenke 2014, Cor. 6.26] and obtain

$$\int_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u) \to \int_X \exp(-\Phi(u, y))\mu_0(\mathrm{d}u).$$

Also,

$$(P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y)) \to (u, \varphi_m)_X \exp(-\Phi(u, y))$$

for all $u \in X$ and

$$\begin{aligned}|(P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y))| &\leq \|u\|_X \exp(L\|u\|_X) \\ &\leq \exp((1 + L)\|u\|_X) \in L^1(X, \mu_0),\end{aligned}$$

so that $(P_n\cdot, \varphi_m)_X \exp(-\Phi(P_n\cdot, y)) \in L^1(X, \mu_0)$ for all $n \in \mathbb{N}$. Here Lebesgue's dominated convergence theorem yields

$$\int_X (P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u) \longrightarrow \int_X (u, \varphi_m)_X \exp(-\Phi(u, y))\mu_0(\mathrm{d}u).$$

Note that for all $n \in \mathbb{N}$, $\bigotimes_{k=1}^n \mathcal{N}_{r^2\alpha_k^{-\tau}}$ is the pushforward measure of $\mathcal{N}_{r^2A^{-\tau}}$ under the projection $\gamma_n \colon X \longrightarrow \mathbb{R}^n$, $\gamma_n(x) = ((x, \varphi_1)_X, \ldots, (x, \varphi_n)_X)$. This allows us to write

$$\begin{aligned}
&\int_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u) \\
&= \int_X \prod_{k=1}^n \exp\left(-\frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}} \left(|(y,\varphi_k)_X - e^{-\alpha_k}(u,\varphi_k)_X| - |(y,\varphi_k)_X|\right)\right)\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u) \\
&= \int_{\mathbb{R}^n} \prod_{k=1}^n \exp\left(-\frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}} \left(|(y,\varphi_k)_X - e^{-\alpha_k}u_k| - |(y,\varphi_k)_X|\right)\right)\bigotimes_{k=1}^n \mathcal{N}_{r^2\alpha_k^{-\tau}}(\mathrm{d}u_k) \\
&= \prod_{k=1}^n \int_{\mathbb{R}} \exp\left(-\frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}} \left(|(y,\varphi_k)_X - e^{-\alpha_k}u_k| - |(y,\varphi_k)_X|\right)\right)\mathcal{N}_{r^2\alpha_k^{-\tau}}(\mathrm{d}u_k).
\end{aligned} \tag{5.15}$$

On the other hand, for $n \geq m$ we have

$$\begin{aligned}
&\int_X (P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u) \\
&= \int_X (u,\varphi_m)_X \prod_{k=1}^n \exp\left(-\frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}} \left(|(y,\varphi_k)_X - e^{-\alpha_k}(u,\varphi_k)_X| - |(y,\varphi_k)_X|\right)\right)\mathcal{N}_{r^2A^{-\tau}}(\mathrm{d}u) \\
&= \int_{\mathbb{R}} u_m \exp\left(-\frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}} \left(|(y,\varphi_m)_X - e^{-\alpha_m}u_m| - |(y,\varphi_m)_X|\right)\right)\mathcal{N}_{r^2\alpha_m^{-\tau}}(\mathrm{d}u_m) \\
&\quad \cdot \prod_{\substack{k=1\\k\neq m}}^n \int_{\mathbb{R}} \exp\left(-\frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}} \left(|(y,\varphi_k)_X - e^{-\alpha_k}u_k| - |(y,\varphi_k)_X|\right)\right)\mathcal{N}_{r^2\alpha_k^{-\tau}}(\mathrm{d}u_k).
\end{aligned} \tag{5.16}$$

So, if we divide (5.16) by (5.15) for $n \geq m$, all factors except for one cancel out and we obtain

$$\begin{aligned}
&\frac{\int_X (P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u)}{\int_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u)} \\
&= \frac{\int_{\mathbb{R}} u_m \exp\Big(-\frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}}|y_m - e^{-\alpha_m}u_m|\Big)\mathcal{N}_{r^2\alpha_m^{-\tau}}(\mathrm{d}u_m)}{\int_{\mathbb{R}} \exp\Big(-\frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}}|y_m - e^{-\alpha_m}u_m|\Big)\mathcal{N}_{r^2\alpha_m^{-\tau}}(\mathrm{d}u_m)} \\
&= \frac{\int_{\mathbb{R}} x \exp\Big(-\frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}}|y_m - e^{-\alpha_m}x| - \frac{1}{2r^2}\alpha_m^{\tau}x^2\Big)\mathrm{d}x}{\int_{\mathbb{R}} \exp\Big(-\frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}}|y_m - e^{-\alpha_m}x| - \frac{1}{2r^2}\alpha_m^{\tau}x^2\Big)\mathrm{d}x},
\end{aligned}$$

where $y_m := (y, \varphi_m)_X$. Consequently,

$$
\begin{aligned}
(\hat{u}_{\mathrm{CM}}(y), \varphi_m)_X &= \frac{\lim_{n\to\infty} \int_X (P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u)}{\lim_{n\to\infty} \int_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u)} \\
&= \lim_{n\to\infty} \frac{\int_X (P_n u, \varphi_m)_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u)}{\int_X \exp(-\Phi(P_n u, y))\mu_0(\mathrm{d}u)} \\
&= \frac{\displaystyle\int_{\mathbb{R}} x \exp\Big(- \frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}} |y_m - e^{-\alpha_m} x| - \frac{1}{2r^2}\alpha_m^{\tau} x^2\Big)\mathrm{d}x}{\displaystyle\int_{\mathbb{R}} \exp\Big(- \frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}} |y_m - e^{-\alpha_m} x| - \frac{1}{2r^2}\alpha_m^{\tau} x^2\Big)\mathrm{d}x}.
\end{aligned}
$$

We further rewrite this as

$$
(\hat{u}_{\mathrm{CM}}(y), \varphi_m)_X = \frac{\int_{\mathbb{R}} x \exp(-ax^2 - |2\tilde{b}x - c|)\mathrm{d}x}{\int_{\mathbb{R}} \exp(-ax^2 - |2\tilde{b}x - c|)\mathrm{d}x}
$$

with

$$
a := \frac{1}{2r^2}\alpha_m^{\tau}, \quad \tilde{b} := \frac{\sqrt{2}}{2b}\alpha_m^{\frac{\beta}{2}} e^{-\alpha_m} \quad \text{and} \quad c := \frac{\sqrt{2}}{b}\alpha_m^{\frac{\beta}{2}}(y, \varphi_m)_X.
$$

We will use the following lemma to compute these integrals. In it the *complementary error function* erfc appears, which is defined on $\mathbb{R}$ by

$$
\operatorname{erfc} x = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2}\,\mathrm{d}t.
$$

Lemma 5.51. *For all $a > 0$ and $b, c, x \in \mathbb{R}$ we have*

$$
\int_x^{\infty} \exp(-at^2 - 2bt - c)\mathrm{d}t = \frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{a} - c\right) \operatorname{erfc}\left(\sqrt{a}x + \frac{b}{\sqrt{a}}\right),
$$

and

$$
\begin{aligned}
\int_x^{\infty} t\exp(-at^2 - 2bt - c)\mathrm{d}t = &-\frac{b}{a}\frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{a} - c\right) \operatorname{erfc}\left(\sqrt{a}x + \frac{b}{\sqrt{a}}\right) \\
&+ \frac{1}{2a}\exp(-ax^2 - 2bx - c).
\end{aligned}
$$

Proof. We differentiate the right hand side of the first equation,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{a} - c\right) \operatorname{erfc}\left(\sqrt{a}x + \frac{b}{\sqrt{a}}\right)\right] \\
&= \frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{a} - c\right)\left[-\frac{2}{\sqrt{\pi}}\exp\left(-\left(\sqrt{a}x + \frac{b}{\sqrt{a}}\right)^2\right)\sqrt{a}\right] \\
&= -\exp\left(\frac{b^2}{a} - c - \left(ax^2 + 2bx + \frac{b^2}{a}\right)\right) \\
&= -\exp(-ax^2 - 2bx - c).
\end{aligned}
$$

Now the first equation follows from the fundamental theorem of calculus, as $\operatorname{erfc} x \to 0$ as $x \to \infty$. We proceed in the same way with the second equation,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\left[-\frac{b}{a}\frac{1}{2}\sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{a}-c\right)\operatorname{erfc}\left(\sqrt{a}x+\frac{b}{\sqrt{a}}\right)+\frac{1}{2a}\exp(-ax^2-2bx-c)\right]\\
&=\frac{b}{a}\exp(-ax^2-2bx-c)+\frac{1}{2a}\exp(-ax^2-2bx-c)(-2ax-2b)\\
&=-x\exp(-ax^2-2bx-c).
\end{aligned}$$

□

We split up the integral in the denominator into

$$\begin{aligned}
&\int_{\mathbb{R}}\exp(-ax^2-|2\tilde{b}x-c|)\mathrm{d}x\\
&=\int_{-\infty}^{\frac{c}{2\tilde{b}}}\exp(-ax^2+2\tilde{b}x-c)\mathrm{d}x+\int_{\frac{c}{2\tilde{b}}}^{\infty}\exp(-ax^2-2\tilde{b}x+c)\mathrm{d}x\\
&=\int_{-\frac{c}{2\tilde{b}}}^{\infty}\exp(-ax^2-2\tilde{b}x-c)\mathrm{d}x+\int_{\frac{c}{2\tilde{b}}}^{\infty}\exp(-ax^2-2\tilde{b}x+c)\mathrm{d}x.
\end{aligned}$$

By Lemma 5.51, this equals

$$\begin{aligned}
&\int_{\mathbb{R}}\exp(-ax^2-|2\tilde{b}x-c|)\mathrm{d}x\\
&=\frac{1}{2}\sqrt{\frac{\pi}{a}}\left[\exp\left(\frac{\tilde{b}^2}{a}-c\right)\operatorname{erfc}\left(-\frac{\sqrt{a}c}{2\tilde{b}}+\frac{\tilde{b}}{\sqrt{a}}\right)+\exp\left(\frac{\tilde{b}^2}{a}+c\right)\operatorname{erfc}\left(\frac{\sqrt{a}c}{2\tilde{b}}+\frac{\tilde{b}}{\sqrt{a}}\right)\right]\\
&=\chi\left[e^{-c}\operatorname{erfc}(\gamma_-)+e^{c}\operatorname{erfc}(\gamma_+)\right]
\end{aligned}$$

with

$$\chi:=\frac{1}{2}\sqrt{\frac{\pi}{a}}\exp\left(\frac{\tilde{b}^2}{a}\right),\quad \gamma_-:=\frac{\tilde{b}}{\sqrt{a}}-\frac{\sqrt{a}c}{2\tilde{b}}\quad\text{and}\quad\gamma_+:=\frac{\tilde{b}}{\sqrt{a}}+\frac{\sqrt{a}c}{2\tilde{b}}.$$

We split up the numerator in the same manner into

$$\begin{aligned}
&\int_{\mathbb{R}}x\exp(-ax^2-|2\tilde{b}x-c|)\mathrm{d}x\\
&=\int_{-\infty}^{\frac{c}{2\tilde{b}}}x\exp(-ax^2+2\tilde{b}x-c)\mathrm{d}x+\int_{\frac{c}{2\tilde{b}}}^{\infty}x\exp(-ax^2-2\tilde{b}x+c)\mathrm{d}x\\
&=-\int_{-\frac{c}{2\tilde{b}}}^{\infty}x\exp(-ax^2-2\tilde{b}x-c)\mathrm{d}x+\int_{\frac{c}{2\tilde{b}}}^{\infty}x\exp(-ax^2-2\tilde{b}x+c)\mathrm{d}x.
\end{aligned}$$

Now, by Lemma 5.51, we have

$$
\begin{aligned}
&\int_{\mathbb{R}} x \exp(-ax^2 - |2\tilde{b}x - c|)\mathrm{d}x \\
&= \frac{\tilde{b}}{a}\frac{1}{2}\sqrt{\frac{\pi}{a}}\left[\exp\left(\frac{\tilde{b}^2}{a} - c\right)\operatorname{erfc}\left(-\frac{\sqrt{a}c}{2\tilde{b}} + \frac{\tilde{b}}{\sqrt{a}}\right) - \exp\left(\frac{\tilde{b}^2}{a} + c\right)\operatorname{erfc}\left(\frac{\sqrt{a}c}{2\tilde{b}} + \frac{\tilde{b}}{\sqrt{a}}\right)\right] \\
&\quad + \frac{1}{2a}\left[-\exp\left(-\frac{ac^2}{4\tilde{b}^2} + c - c\right) + \exp\left(-\frac{ac^2}{4\tilde{b}^2} - c + c\right)\right] \\
&= R\chi\left[e^c \operatorname{erfc}(\gamma_+) - e^{-c}\operatorname{erfc}(\gamma_-)\right]
\end{aligned}
$$

with $R := \sqrt{2}\frac{r^2}{b}\alpha_m^{\beta/2-\tau}e^{-\alpha_m}$. This results in

$$
(\hat{u}_{\mathrm{CM}}(y), \varphi_m)_X = R\frac{e^{-c}\operatorname{erfc}(\gamma_-) - e^c\operatorname{erfc}(\gamma_+)}{e^{-c}\operatorname{erfc}(\gamma_-) + e^c\operatorname{erfc}(\gamma_+)}.
$$

Using

$$
\gamma_+^2 - \gamma_-^2 = (\gamma_+ + \gamma_-)(\gamma_+ - \gamma_-) = \frac{\tilde{b}}{\sqrt{a}}\frac{2\sqrt{a}c}{\tilde{b}} = 2c
$$

we can finally express the components in terms of the scaled complementary error function $\operatorname{erfcx}(x) = \exp(x^2)\operatorname{erfc}(x)$ as

$$
(\hat{u}_{\mathrm{CM}}(y), \varphi_m)_X = R\frac{e^{\gamma_-^2}\operatorname{erfc}(\gamma_-) - e^{\gamma_+^2}\operatorname{erfc}(\gamma_+)}{e^{\gamma_-^2}\operatorname{erfc}(\gamma_-) + e^{\gamma_+^2}\operatorname{erfc}(\gamma_+)} = R\frac{\operatorname{erfcx}(\gamma_-) - \operatorname{erfcx}(\gamma_+)}{\operatorname{erfcx}(\gamma_-) + \operatorname{erfcx}(\gamma_+)}. \tag{5.17}
$$

This formulation will be beneficial for the numerical computation of the CM estimator, since $\operatorname{erfcx}(x)$ decays slower than $\operatorname{erfc}(x)$ as x increases.

Now we consider the continuity of $\hat{u}_{\mathrm{CM}}$.

Theorem 5.52. *If $y^n \to y^\dagger$ in $\mathcal{X}^s$ then*

$$
\hat{u}_{CM}(y^n) \to \hat{u}_{CM}(y^\dagger)
$$

in X.

Proof. First, we write the conditional expectation of u for any $y \in \mathcal{X}^s$ as

$$
\mathbb{E}^{\mu^y} u = \int_X \sum_{k=1}^{\infty}(u, \varphi_k)_X\varphi_k\mu^y(du) = \sum_{k=1}^{\infty}\int_X (u, \varphi_k)_X\mu^y(du)\varphi_k = \sum_{k=1}^{\infty}\mathbb{E}^{\mu^y}(u, \varphi_k)_X\varphi_k.
$$

Parseval's identity yields

$$
\left\|\mathbb{E}^{\mu^y} u - \mathbb{E}^{\mu^z} u\right\|_X^2 = \sum_{k=1}^{\infty}\left|\mathbb{E}^{\mu^y}(u, \varphi_k)_X - \mathbb{E}^{\mu^z}(u, \varphi_k)_X\right|^2
$$

for all $y, z \in \mathcal{X}^s$. Now $(\cdot, \varphi_k)_X \in L^2(X, \mu^y)$ for any $y \in \mathcal{X}^s$, as

$$
|(u, \varphi_k)_X|^2\exp(-\Phi(u, y)) \le \|u\|_X^2\exp(L\|u\|_X) \le 2\exp\left((1 + L)\|u\|_X\right), \tag{5.18}
$$

and $\exp((1+L)\|u\|_X)$ is μ_0-integrable by Lemma 5.26. So, we may apply Lemma 5.31, and obtain

$$\left|\mathbb{E}^{\mu^y}(u,\varphi_k)_X - \mathbb{E}^{\mu^z}(u,\varphi_k)_X\right|^2 \leq 8\left(\mathbb{E}^{\mu^y}|(u,\varphi_k)_X|^2 + \mathbb{E}^{\mu^z}|(u,\varphi_k)_X|^2\right) d_{\mathrm{Hell}}(\mu^y,\mu^z)^2 \tag{5.19}$$

for all $y, z \in \mathcal{X}^s$. Summing up yields

$$\begin{aligned}\left\|\mathbb{E}^{\mu^{y^n}} u - \mathbb{E}^{\mu^{y^\dagger}} u\right\|_X^2 &\leq \sum_{k=1}^{\infty} 8\left(\mathbb{E}^{\mu^{y^n}}|(u,\varphi_k)_X|^2 + \mathbb{E}^{\mu^{y^\dagger}}|(u,\varphi_k)_X|^2\right) d_{\mathrm{Hell}}(\mu^{y^n},\mu^{y^\dagger})^2 \\ &= 8\left(\mathbb{E}^{\mu^{y^n}}\|u\|_X^2 + \mathbb{E}^{\mu^{y^\dagger}}\|u\|_X^2\right) d_{\mathrm{Hell}}(\mu^{y^n},\mu^{y^\dagger})^2 \\ &\leq 16 C d_{\mathrm{Hell}}(\mu^{y^n},\mu^{y^\dagger})^2,\end{aligned}$$

where

$$C := \int_X \|u\|_X^2 \exp(L\|u\|_X)\mu_0(\mathrm{d}u)$$

is finite due to the previous considerations. Now the proposition follows from the convergence $d_{\mathrm{Hell}}(\mu^{y^n},\mu^{y^\dagger}) \to 0$ according to Theorem 5.30. □

6 Numerical Study of the Inverse Heat Equation

In this chapter we consider the inverse heat equation with finite-dimensional data. We will compute both MAP and the CM estimator exactly and, moreover, derive a direct sampler for the resulting posterior distribution. Such a sampler can be used for example to compute integrals over the posterior by Monte Carlo integration or to approximate credible sets. Our goal is to illustrate the results from Chapter 5 and beyond their scope study the behaviour of both estimators numerically.

6.1 Problem Setting

We consider the classical one-dimensonal inverse heat equation on the interval $D := (0, 1)$. Given a noisy temperature measurement $y \in Y := L^2(D)$ at the time $t := 0.002$ we want to reconstruct the temperature $u \in X := L^2(D)$ at time 0. As described in Subsection 5.3, this corresponds to solving the operator equation

$$y = e^{-A}u + \eta, \tag{6.1}$$

where $A := -t\Delta = -t\frac{\partial^2}{\partial x^2}$ is the scaled weak Laplace operator in $L^2(D)$ with domain $\mathcal{D}(A) := H^2(D) \cap H_0^1(D)$ and η is the noise. Here we assume that $\eta \in \mathcal{X}^0 = L^2(D)$, i.e., we choose $s := 0$. Moreover, we assume that $u \sim \mathcal{N}_{r^2A^{-\tau}}$ and $\eta \sim \mathcal{L}_{b^2A^{-\beta}}$, independent of each other, where $\tau, \beta > \frac{d}{2}$ and $r, b > 0$, as in Subsection 5.2.

The operator A is Laplace-like, as pointed out in Example 5.5, so that we are in the setting of Chapter 5 and the posterior distribution μ^y on $L^2(D)$ is given by Theorem 5.28. For every $k \in \mathbb{N}$, the function $\varphi_k\colon (0, 1) \to \mathbb{R}$,

$$\varphi_k(x) := \sqrt{2}\sin(\pi k x) \qquad \text{for all } x \in (0, 1),$$

is an eigenfunction of A (in particular, $\varphi_k \in \mathcal{D}(A)$) with the associated eigenvalue $\alpha_k := t\pi^2k^2$. Moreover, $\{\varphi_k\}_{k\in\mathbb{N}}$ forms an orthonormal basis of $L^2(D)$ by Proposition 4.5.2 (iv) in [Zeidler 1995] and thereby satisfies Assumption 5.1 (iii). The eigenvalues $\{\alpha_k\}_{k\in\mathbb{N}}$ are positive, non-decreasing and quadratic in k, so that they satisfy Assumption 5.1 (iv) with $d = 1$ and $C_- = C_+ = t\pi^2$.

We want to numerically verify the behaviour of the CM and MAP estimators predicted in theory and study their behaviour beyond the scope of the theory. We will compute both estimates for data resulting from different values of the unknown u. Here we adopt a frequentist point of view and assume that there is a true solution $u^\dagger \in L^2(D)$. We consider three different scenarios:

- *Scenario 1*: There is a source element $w \in L^2(D)$ such that
$$u^\dagger = A^{\frac{\beta}{2}-\tau} e^{-A} w \quad \text{and} \quad \sup_{k\in\mathbb{N}} |(w, \varphi_k)_X| \leq \rho.$$
- *Scenario 2*: $u^\dagger \in \mathcal{X}^\tau$.
- *Scenario 3*: $u^\dagger \in L^2(D)$.

Scenario 1 is precisely the setting of Theorem 5.50. The motivation for Scenario 2 is that $\mathcal{R}(\hat{u}_{\text{MAP}}) \subset \mathcal{X}^\tau$ by Definition 5.35, which leads to the question if here, the MAP estimator performs better than in Scenario 3.

We will divide each scenario into two subscenarios, labeled *a* and *b*, one with rougher noise and one with smoother noise. In *Scenarios 1a, 2a, 3a* we choose $\beta = 0.65$, in *Scenarios 1b, 2b, 3b* we choose $\beta = 1.3$. In both cases, $\beta > \frac{1}{2} = s + \frac{d}{2}$. Note that in Scenario 1, $u^\dagger$ depends of β, whereas in Scenarios 2 and 3 it does not. Moreover, we choose $\tau = 0.55$ throughout. This way, $\tau > \frac{1}{2} = \frac{d}{2}$.

6.2 Discretisation

Now, we assume that instead of the exact data y we only have knowledge of its orthogonal projection
$$y^N := P_N y = \sum_{k=1}^{N} (y, \varphi_k)_{L^2} \varphi_k$$
to the finite dimensional subspace $U_N := \operatorname{span}\{\varphi_1, \dots, \varphi_N\} \subset L^2(D)$ for some $N \in \mathbb{N}$.

We saw that the resulting posterior measure μ^{y^N} will be close to the exact one in the sense of the Hellinger distance: By Theorem 5.30 the convergence of y^N towards y implies $d_{\text{Hell}}(\mu^{y^N}, \mu^y) \to 0$. The posterior estimates using y^N are also close to those using the exact data y: Theorem 5.38 tells us that $\hat{u}_{\text{MAP}}(y^N) \to \hat{u}_{\text{MAP}}(y)$ and Theorem 5.52 tells us that $\hat{u}_{\text{CM}}(y^N) \to \hat{u}_{\text{CM}}(y)$ in $L^2(D)$ as $N \to \infty$. Moreover, the componentwise representations of $\hat{u}_{\text{MAP}}$ from Lemma 5.34 and of $\hat{u}_{\text{CM}}$ from Section 5.6 show that only the first N components of $\hat{u}_{\text{CM}}(y^N)$ and $\hat{u}_{\text{MAP}}(y^N)$ are nonzero, i.e., both $\hat{u}_{\text{CM}}(y^N)$ and $\hat{u}_{\text{MAP}}(y^N)$ also belong to U_N.

Motivated by this observation, we discretise the problem by projecting both sides of the operator equation (6.1) to U_N. Using
$$P_N e^{-A} u = \sum_{k=1}^{N} e^{-\alpha_k} (u, \varphi_k)_{L^2} \varphi_k = e^{-A} P_N u,$$
this leads to
$$P_N y = P_N e^{-A} u + P_N \eta = e^{-A} P_N u + P_N \eta. \tag{6.2}$$
In a next step we identify U_N with $\mathbb{R}^N$: We can restate (6.2) as a linear equation
$$\tilde{y} = K\tilde{u} + \tilde{\eta} \tag{6.3}$$

in $\mathbb{R}^N$ with the $N \times N$ diagonal matrix

$$K := \operatorname{diag}(e^{-\alpha_1}, \dots, e^{-\alpha_N})$$

by setting $\tilde{y} = \gamma_N(y)$, $\tilde{u} = \gamma_N(u)$ and $\tilde{\eta} = \gamma_N(\eta)$, where $\gamma_N(u) := ((u, \varphi_1)_{L^2}, \dots, (u, \varphi_N)_{L^2})$ for all $u \in L^2(D)$. As $\varphi_1, \dots, \varphi_N$ is an orthonormal basis of U_N, we have

$$\|P_N u\|_{L^2}^2 = \sum_{k=1}^N |(u, \varphi_k)_{L^2}|^2 = \|\gamma_N(u)\|_2^2 \quad \text{for all } u \in L^2(D).$$

We also define the discretisation of the operator A, the $N \times N$ diagonal matrix

$$\tilde{A} := \operatorname{diag}(\alpha_1, \dots, \alpha_N).$$

Moreover, we denote MAP and CM estimator on $\mathbb{R}^N$ by

$$\hat{\tilde{u}}_{\mathrm{MAP}}(y_1, \dots, y_N) := \gamma_N \hat{u}_{\mathrm{MAP}} \left(\sum_{k=1}^N y_k \varphi_k \right)$$

and $\hat{\tilde{u}}_{\mathrm{CM}}(y_1, \dots, y_N) := \gamma_N \hat{u}_{\mathrm{CM}} \left(\sum_{k=1}^N y_k \varphi_k \right)$, respectively.

6.3 Numerical Implementation

In this section we explain in detail how we will create the true solution in the different scenarios numerically, how we will generate samples of the Laplacian noise and how we will compute MAP and CM estimates numerically.

In the frequentist setting we create a true solution in the following way: We first discretise a piecewise constant function $f_1 \in L^2(D)$ on an equidistant grid on $[0, 1]$ with grid size $\frac{1}{N+2}$, where the value of f_1 in the points 0 and 1 is bound to be 0. Then we represent f_1 using the discretised first N eigenvectors of A,

$$\left(0, \varphi_k\left(\frac{1}{N+2}\right), \dots, \varphi_k\left(\frac{N+1}{N+2}\right), 0\right), \quad k = 1, \dots, N,$$

as a basis. Subsequently, we either directly use this element $\tilde{w} \in \mathbb{R}^N$ as the true solution $\tilde{u}^\dagger$ (Scenario 3), or we apply $\tilde{A}^{-\tau/2}$ to $\tilde{w}$ to obtain an element corresponding to a function in $\mathcal{X}^\tau$ (Scenario 2), or we create the true solution by successively applying to $\tilde{w}$ the discrete forward operator K and $\tilde{A}^{\beta/2-\tau}$ (Scenario 1).

The discretised noise $\tilde{\eta}$ by definition has the distribution $\mathcal{L}_{b^2 A^{-\beta}} \circ \gamma_N^{-1}$, the pushforward of the noise measure $\mathcal{L}_{b^2 A^{-\beta}}$ under γ_N. And as $\gamma_N^{-1}(A) = I_{1,\dots,N;A}$ is a cylindrical set for every $A \in \mathcal{B}(\mathbb{R}^N)$, this measure is by definition a Laplacian product measure on $\mathbb{R}^N$, so that

$$\tilde{\eta} \sim \mathcal{L}_{b^2 A^{-\beta}} \circ \gamma_N^{-1} = \bigotimes_{k=1}^N \mathcal{L}_{b^2 \alpha_k^{-\beta}}.$$

The independence of the components η_k allows us to draw a sample $\tilde{\eta} \sim \bigotimes_{k=1}^{N} \mathcal{L}_{b^2\alpha_k^{-\beta}}$ of the noise by sampling each component individually. We do this by the inverse cumulative distribution method, which is described in Subsection 6.4 and will also be used for sampling the posterior. We generate $r_k, r_k' \sim \mathrm{unif}(0,1)$, $k = 1, \dots, N$, independently and then assigning

$$\eta_k = \frac{\sqrt{2}}{2} \operatorname{sign}(r_k - 0.5) b \alpha_k^{-\frac{\beta}{2}} \log(1 - r_k') \quad \text{for } k = 1, \dots, N.$$

In the frequentist setting, we then generate a data sample from $\tilde{y} := Ku^\dagger + \tilde{\eta}$.

By Lemma 5.34 the components of the MAP estimate $\hat{u}_{\mathrm{MAP}}(y^N)$ are given by

$$(\hat{u}_{\mathrm{MAP}}(y^N), \varphi_k)_{L^2} = \max\left\{-\frac{r^2}{b}R_k, \min\left\{e^{\alpha_k}(y^N, \varphi_k)_X, \frac{r^2}{b}R_k\right\}\right\} \quad \text{for all } k \in \mathbb{N},$$

where $R_k = \sqrt{2}\alpha_k^{\beta/2-\tau} e^{-\alpha_k}$. In particular,

$$(\hat{u}_{\mathrm{MAP}}(y^N), \varphi_k)_{L^2} = 0 \quad \text{for } k > N.$$

So $\hat{u}_{\mathrm{MAP}}(y^N) \in \operatorname{span}\{\varphi_1, \dots, \varphi_N\}$ and we only need to compute its first N components.

In Section 5.6 we found that components of the CM estimate $\hat{u}_{\mathrm{CM}}(y^N)$ are given by

$$(\hat{u}_{\mathrm{CM}}(y^N), \varphi_k)_X = R_k \frac{\operatorname{erfcx}(\gamma_k^-) - \operatorname{erfcx}(\gamma_k^-)}{\operatorname{erfcx}(\gamma_k^-) + \operatorname{erfcx}(\gamma_k^-)}. \tag{6.4}$$

for all $k \in \mathbb{N}$, where

$$R_k := \frac{\tilde{b}}{a}, \qquad \gamma_k^- := \frac{\tilde{b}}{\sqrt{a}} - \frac{\sqrt{ac}}{2\tilde{b}}, \qquad \gamma_k^+ := \frac{\tilde{b}}{\sqrt{a}} + \frac{\sqrt{ac}}{2\tilde{b}},$$

$$a := \frac{1}{2r^2}\alpha_k^\tau, \qquad \tilde{b} := \frac{\sqrt{2}}{2b}\alpha_k^{\frac{\beta}{2}} e^{-\alpha_k}, \qquad c := \frac{\sqrt{2}}{b}\alpha_k^{\frac{\beta}{2}}(y, \varphi_k)_X.$$

Again,

$$(\widehat{u}_{\mathrm{CM}}(y^N), \varphi_k)_{L^2} = 0 \quad \text{for } k > N,$$

so that $\widehat{u}_{\mathrm{CM}}(y^N) \in \operatorname{span}\{\varphi_1, \dots, \varphi_N\}$, too, and we only need to compute its first N components.

We have expressed (6.4) in this form because it allows us to evaluate $\operatorname{erfcx}(x) = \exp(x^2)\operatorname{erfc}(x)$ instead of $\operatorname{erfc}(x)$, which decays slower for $x \to +\infty$. However, $\operatorname{erfcx}(x)$ increases very fast for $x \to -\infty$, which is why we will only evaluate it for nonnegative values of x. We can do this by using the identity

$$\operatorname{erfcx}(-x) = 2\exp(x^2) - \operatorname{erfcx}(x),$$

which follows directly from the symmetry $\operatorname{erfc}(-x) = 2 - \operatorname{erfc}(x)$.

In case $\gamma_k^+ \geq 0$ and $\gamma_k^- \geq 0$, we can evaluate (6.4) as it is. In the remaining cases we perform some transformations in order to evaluate the k-th component of $\hat{u}_{\mathrm{CM}}(y^N)$ sufficiently numerically stably. Note that γ_k^+ and γ_k^- cannot both be negative, as $\frac{\tilde{b}}{\sqrt{a}} > 0$. In case $\gamma_k^+ \geq 0$ and $\gamma_k^- < 0$, we replace $\operatorname{erfcx}(\gamma_k^-) = 2\exp((\gamma_k^-)^2) - \operatorname{erfcx}(-\gamma_k^-)$, which leads to

$$
\begin{aligned}
(\hat{u}_{\mathrm{CM}}(y), \varphi_k)_{L^2} &= R_k \frac{2\exp((\gamma_k^-)^2) - \operatorname{erfcx}(-\gamma_k^-) - \operatorname{erfcx}(\gamma_k^+)}{2\exp((\gamma_k^-)^2) - \operatorname{erfcx}(-\gamma_k^-) + \operatorname{erfcx}(\gamma_k^+)}, \\
&= R_k \frac{2 - \exp(-(\gamma_k^-)^2)\left[\operatorname{erfcx}(-\gamma_k^-) + \operatorname{erfcx}(\gamma_k^+)\right]}{2 - \exp(-(\gamma_k^-)^2)\left[\operatorname{erfcx}(-\gamma_k^-) - \operatorname{erfcx}(\gamma_k^+)\right]}.
\end{aligned}
$$

In case $\gamma_k^+ < 0$ and $\gamma_k^- \geq 0$, we replace $\operatorname{erfcx}(\gamma_k^+) = 2\exp((\gamma_k^+)^2) - \operatorname{erfcx}(-\gamma_k^+)$, which results in

$$
\begin{aligned}
(\hat{u}_{\mathrm{CM}}(y), \varphi_k)_{L^2} &= R_k \frac{\operatorname{erfcx}(-\gamma_k^-) + \operatorname{erfcx}(\gamma_k^+) - 2\exp((\gamma_k^+)^2)}{\operatorname{erfcx}(-\gamma_k^-) - \operatorname{erfcx}(\gamma_k^+) + 2\exp((\gamma_k^+)^2)}, \\
&= R_k \frac{\exp(-(\gamma_k^+)^2)\left[\operatorname{erfcx}(-\gamma_k^-) + \operatorname{erfcx}(\gamma_k^+)\right] - 2}{\exp(-(\gamma_k^+)^2)\left[\operatorname{erfcx}(-\gamma_k^-) - \operatorname{erfcx}(\gamma_k^+)\right] + 2}.
\end{aligned}
$$

6.4 Direct Posterior Sampling

In order to implement a posterior sampler that samples the first N components of the posterior $u \sim \mu^{y^N}$, we need their marginal distribution, or, in other words, the distribution of the projection of the posterior to $U_N = \operatorname{span}\{\varphi_1, \dots, \varphi_N\}$. Here, we could also choose a subspace with a dimension different from N, but this would either result in a loss of part of the information contained in y^N, or in no gain of information, because additional components would be completely determined by the prior. If we identify U_N with $\mathbb{R}^N$, then this distribution is given by the pushforward

$$\mu^N := \mu^{y^N} \circ \gamma_N^{-1}$$

of μ^N under the projection $\gamma_N \colon X \to \mathbb{R}^N$,

$$\gamma_N(u) = ((u, \varphi_1)_{L^2}, \dots, (u, \varphi_N)_{L^2}).$$

Unlike μ^{y^N}, which is still a measure on $L^2(D)$, μ^N constitutes a measure on $\mathbb{R}^N$. By definition of $\mathcal{N}_{A^{-\tau}}$, we have

$$
\begin{aligned}
\mu^N(M) &= \mu^{y^N}(\{u \in X : \gamma_N u \in M\}) \\
&= \frac{1}{Z^N(y)} \int\limits_{\{u \in X : \gamma_N u \in M\}} \exp(-\Phi^N(u, y)) \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}u) \\
&= \frac{1}{Z^N(y)} \int\limits_M \prod_{k=1}^N \exp\left(-\frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} \left(|y_k - e^{-\alpha_k} x_k| - |y_k|\right)\right) \left(\bigotimes_{k=1}^N \mathcal{N}_{r^2 \alpha_k^{-\tau}}\right)(\mathrm{d}x) \\
&= \frac{C_N}{Z^N(y)} \int\limits_M \prod_{k=1}^N \exp\left(-\frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} \left(|y_k - e^{-\alpha_k} x_k| - |y_k|\right) - \frac{1}{2r^2} \alpha_k^{\tau} x_k^2\right) \mathrm{d}x
\end{aligned}
$$

for all $M \in \mathcal{B}(\mathbb{R}^N)$, where $y_k := (y, \varphi_k)_X$, $x = (x_1, \ldots, x_N)$ and

$$C_N := \left((2\pi)^d \prod_{k=1}^{N} \left(r^2 \alpha_k^{-\tau} \right) \right)^{-\frac{1}{2}}.$$

A similar computation yields the normalisation constant

$$\begin{aligned} Z^N(y) = Z(P_N y) &= \int_X \exp(-\Phi(u, P_N y)) \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}u) \\ &= \int_X \exp(-\Phi^N(u, y)) \mathcal{N}_{r^2 A^{-\tau}}(\mathrm{d}u) \\ &= \int_{\mathbb{R}^N} \prod_{k=1}^{N} \exp\left(-\frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} \left(|y_k - e^{-\alpha_k} x_k| - |y_k| \right) \right) \left(\bigotimes_{k=1}^{N} \mathcal{N}_{r^2 \alpha_k^{-\tau}} \right) (\mathrm{d}x) \\ &= C_N \prod_{k=1}^{N} \int_{\mathbb{R}} \exp\left(-\frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} \left(|y_k - e^{-\alpha_k} t| - |y_k| \right) - \frac{1}{2r^2} \alpha_k^{\tau} t^2 \right) \mathrm{d}t. \end{aligned}$$

Consequently,

$$\mu^N(M) = \int_M \prod_{k=1}^{N} \frac{\exp\left(-\frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} |y_k - e^{-\alpha_k} u_k| - \frac{1}{2r^2} \alpha_k^{\tau} u_k^2 \right)}{\int_{\mathbb{R}} \exp\left(-\frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} |y_k - e^{-\alpha_k} t| - \frac{1}{2r^2} \alpha_k^{\tau} t^2 \right) \mathrm{d}t} \mathrm{d}u$$

for all $M \in \mathcal{B}(\mathbb{R}^N)$. So, the approximated posterior μ^N has a probability density p_{post} with respect to the Lebesgue measure on $\mathbb{R}^N$ of the form

$$p_{\text{post}}(u_1, \ldots, u_N) = \prod_{k=1}^{N} p_k(u_k)$$

with

$$p_k(x) := \frac{\exp(-a_k x^2 - |2\tilde{b}_k x - c_k|)}{\int_{\mathbb{R}} \exp(-a_k t^2 - |2\tilde{b}_k t - c_k|) \mathrm{d}t}$$

for all $x \in \mathbb{R}$ and $k = 1, \ldots, N$. Here,

$$a_k := \frac{1}{2r^2} \alpha_k^{\tau}, \quad \tilde{b}_k := \frac{\sqrt{2}}{2b} \alpha_k^{\frac{\beta}{2}} e^{-\alpha_k} \quad \text{and} \quad c_k := \frac{\sqrt{2}}{b} \alpha_k^{\frac{\beta}{2}} (y, \varphi_k)_X.$$

Hence the components u_k of $u \sim \mu^N$ are by definition independent.

This independence allows us to draw a sample $u \sim \mu^N$ from the marginal posterior distribution by sampling each component u_k individually. We have seen that the probability density p of a single component u_k is of the form

$$p(x) = \frac{1}{\mathcal{Z}} \exp\left(-ax^2 - |bx - c| \right)$$

with $a, b > 0$, $c \in \mathbb{R}$ and $\mathcal{Z} \coloneqq \int_{\mathbb{R}} \exp(-at^2 - |bt - c|)\mathrm{d}t$,

Now we develop a direct sampler for a distribution ν with density p that uses the inverse cumulative distribution method. The *cumulative distribution function (cdf)* F of u_k is defined by

$$F(x) = \int_{-\infty}^{x} p(t)\mathrm{d}t$$

for all $x \in \mathbb{R}$. In general, $F^{-1}\colon [0,1] \to \mathbb{R}$ denotes the generalised inverse of F, in our case, however, the ordinary inverse of F exists, because p is positive and hence F is strictly monotonic. The basis for the sampler is the observation, that, if $r \sim \mathrm{unif}(0,1)$ is a uniformly distributed random variable on the interval $[0,1]$, then $F^{-1}(r)$ has the desired distribution ν. As F^{-1} is monotonic and the probability density of r is equal to 1, the probability density q of $F^{-1}(r)$ is given by the change of variable formula as

$$q(x) = \left|\frac{\mathrm{d}}{\mathrm{d}x}F(x)\right| \cdot 1 = p(x)$$

for all $x \in \mathbb{R}$.

Lemma 6.1. *If the probability density function p of a real random variable x is of the form*

$$p(x) = \frac{1}{\mathcal{Z}} \exp\left(-ax^2 - |bx - c|\right)$$

with $a, b, \mathcal{Z} > 0$, $c \in \mathbb{R}$, then its cumulative distribution function F is given by

$$F(x) = \begin{cases} \frac{1}{\mathcal{Z}}\frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} - c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} - \sqrt{a}x\right) & \text{if } x \leq \frac{c}{b}, \\ 1 - \frac{1}{\mathcal{Z}}\frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} + c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} + \sqrt{a}x\right) & \text{if } x > \frac{c}{b}. \end{cases}$$

Moreover, the normalisation factor $\mathcal{Z}$ is given by

$$\mathcal{Z} = \frac{1}{2}\sqrt{\frac{\pi}{a}} \left[\exp\left(\frac{b^2}{4a} - c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} - \frac{\sqrt{a}c}{b}\right) + \exp\left(\frac{b^2}{4a} + c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} + \frac{\sqrt{a}c}{b}\right)\right].$$

Proof. For $x \leq \frac{c}{b}$ we have

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} p(t)\mathrm{d}t = \frac{1}{\mathcal{Z}}\int_{-\infty}^{x} \exp(-at^2 + bt - c)\mathrm{d}t \\ &= \frac{1}{\mathcal{Z}}\int_{-x}^{\infty} \exp(-at^2 - 2\frac{b}{2}t - c)\mathrm{d}t \\ &= \frac{1}{\mathcal{Z}}\frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} - c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} - \sqrt{a}x\right) \end{aligned}$$

by Lemma 5.51. In contrast, for $x > \frac{c}{b}$ we compute, using $\int_{\mathbb{R}} p(t)\mathrm{d}t = 1$,

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} p(t)\mathrm{d}t = 1 - \frac{1}{\mathcal{Z}}\int_{x}^{\infty} \exp(-at^2 - bt + c)\mathrm{d}t \\ &= 1 - \frac{1}{\mathcal{Z}}\int_{x}^{\infty} \exp(-at^2 - 2\frac{b}{2}t - (-c))\mathrm{d}t \\ &= 1 - \frac{1}{\mathcal{Z}}\frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} + c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} + \sqrt{a}x\right) \end{aligned}$$

Finally, again by Lemma 5.51, we have

$$
\begin{aligned}
\mathcal{Z} &= \int_{\mathbb{R}} \exp\left(-ax^2 - |bx - c|\right) \mathrm{d}t \\
&= \int_{-\infty}^{\frac{c}{b}} \exp(-at^2 + bt - c)\mathrm{d}t + \int_{\frac{c}{b}}^{\infty} \exp(-at^2 - bt + c)\mathrm{d}t \\
&= \int_{-\frac{c}{b}}^{\infty} \exp(-at^2 - 2\frac{b}{2}t - c)\mathrm{d}t + \int_{\frac{c}{b}}^{\infty} \exp(-at^2 - 2\frac{b}{2}t - (-c))\mathrm{d}t \\
&= \frac{1}{2}\sqrt{\frac{\pi}{a}} \left[\exp\left(\frac{b^2}{4a} - c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} - \frac{\sqrt{a}c}{b}\right) + \exp\left(\frac{b^2}{4a} + c\right) \operatorname{erfc}\left(\frac{b}{2\sqrt{a}} + \frac{\sqrt{a}c}{b}\right)\right] \qquad \square
\end{aligned}
$$

We can express F and $\mathcal{Z}$ more concisely with $\chi := \frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a}\right)$, $\gamma_- := \frac{b}{2\sqrt{a}} - \frac{\sqrt{a}c}{b}$, $\gamma_+ := \frac{b}{2\sqrt{a}} + \frac{\sqrt{a}c}{b}$ and $\gamma := \frac{b}{2\sqrt{a}}$ as

$$\mathcal{Z} = \chi \left[\exp(-c) \operatorname{erfc}(\gamma_-) + \exp(c) \operatorname{erfc}(\gamma_+)\right]$$

and

$$F(x) = \begin{cases} \frac{1}{\mathcal{Z}} \chi \exp(-c) \operatorname{erfc}(\gamma - \sqrt{a}x) & \text{for } x \le \frac{c}{b}, \\ 1 - \frac{1}{\mathcal{Z}} \chi \exp(c) \operatorname{erfc}(\gamma + \sqrt{a}x) & \text{for } x > \frac{c}{b}. \end{cases}$$

We further simplify this to

$$F(x) = \frac{\operatorname{erfc}(\gamma - \sqrt{a}x)}{\operatorname{erfc}(\gamma_-) + \exp(2c) \operatorname{erfc}(\gamma_+)}$$

for $x \le \frac{c}{b}$ and

$$F(x) = 1 - \frac{\operatorname{erfc}(\gamma + \sqrt{a}x)}{\exp(-2c) \operatorname{erfc}(\gamma_-) + \operatorname{erfc}(\gamma_+)}$$

for $x > \frac{c}{b}$ respectively.

Now we can invert F. Because of the monotonicity of F, $x \le \frac{c}{b}$ holds if and only if $F(x) \le F(\frac{c}{b})$. Given $r \in [0, F(\frac{c}{b})]$ we want to find an $x \in \mathbb{R}$ with $F(x) = r$. We divide this task into two steps: First we find a $z \in \mathbb{R}$ such that

$$r = \frac{z}{\operatorname{erfc}(\gamma_-) + \exp(2c) \operatorname{erfc}(\gamma_+)},$$

then we find an $x \in \mathbb{R}$ such that

$$z = \operatorname{erfc}(\gamma - \sqrt{a}x).$$

The first condition leads to

$$z = r \left[\operatorname{erfc}(\gamma_-) + \exp(2c) \operatorname{erfc}(\gamma_+)\right], \tag{6.5}$$

and the second one results in

$$x = \frac{\gamma - \operatorname{erfcinv}(z)}{\sqrt{a}},$$

where erfcinv denotes the inverse of the complementary error function. This way, $F(x) = r$ holds by construction. We proceed in the same way for $r \in (F(\frac{c}{b}), 1]$. Here, we demand

$$r = 1 - \frac{z}{\exp(-2c)\,\mathrm{erfc}(\gamma_-) + \mathrm{erfc}(\gamma_+)} \quad \text{and} \quad z = \mathrm{erfc}(\gamma + \sqrt{a}x),$$

which leads to

$$z = (1-r)\left[\exp(-2c)\,\mathrm{erfc}(\gamma_-) + \mathrm{erfc}(\gamma_+)\right] \tag{6.6}$$

and

$$x = \frac{\mathrm{erfcinv}(z) - \gamma}{\sqrt{a}}.$$

We can combine this to

$$F^{-1}(r) = \begin{cases} \frac{1}{\sqrt{a}}\left(\gamma - \mathrm{erfcinv}\left(r\left[\mathrm{erfc}(\gamma_-) + \exp(2c)\,\mathrm{erfc}(\gamma_+)\right]\right)\right) & \text{if } r \in [0, F(\frac{c}{b})], \\ \frac{1}{\sqrt{a}}\left(\mathrm{erfcinv}\left((1-r)\left[\exp(-2c)\,\mathrm{erfc}(\gamma_-) + \mathrm{erfc}(\gamma_+)\right]\right) - \gamma\right) & \text{if } r \in (F(\frac{c}{b}), 1]. \end{cases}$$

The numerical evaluation of F^{-1} requires additional thought. We will utilise several of the techniques used in [Lucka 2012, Appendix B] (or [Lucka 2014, A.3]) for the implementation of an ℓ_1 sampler. In order to compute the arguments of erfcinv in equations (6.5) and (6.6) with the necessary precision, we express them in terms of the scaled complementary error function $\mathrm{erfcx}(x) = \exp(x^2)\,\mathrm{erfc}(x)$, which decays slower for $x \to +\infty$. We do this in such a way that $\mathrm{erfcx}(x)$ is only evaluated for nonnegative values of x, since $\mathrm{erfcx}(x)$ increases very fast for $x \to -\infty$. Additionally, we compute the logarithm of z instead of z in equations (6.5) and (6.6) in order to avoid multiplying very large numbers with very small ones. Furthermore, we use an asymptotic approximation of $\mathrm{erfcinv}(\exp(w))$ for $w \to -\infty$.

We first compute

$$\gamma_+^2 - \gamma_-^2 = (\gamma_+ + \gamma_-)(\gamma_+ - \gamma_-) = \frac{b}{\sqrt{a}}\frac{2\sqrt{a}c}{b} = 2c. \tag{6.7}$$

Note that γ_+ and γ_- by definition cannot both be negative, because $a, b > 0$, so that we only have to consider the following three cases:

If both $\gamma_+ \geq 0$ and $\gamma_- \geq 0$, then z in (6.5) is given by

$$\begin{aligned} z_1 &= r\left[\mathrm{erfc}(\gamma_-) + \exp(2c)\,\mathrm{erfc}(\gamma_+)\right] \\ &= r\exp(-\gamma_-^2)\left[\exp(\gamma_-^2)\,\mathrm{erfc}(\gamma_-) + \exp(\gamma_+^2)\,\mathrm{erfc}(\gamma_+)\right] \\ &= r\exp(-\gamma_-^2)\omega_{++} \end{aligned} \tag{6.8}$$

with $\omega_{++} := \mathrm{erfcx}(\gamma_+) + \mathrm{erfcx}(\gamma_-)$. In (6.6), z is given by

$$\begin{aligned} z_2 &= (1-r)\left[\exp(-2c)\,\mathrm{erfc}(\gamma_-) + \mathrm{erfc}(\gamma_+)\right] \\ &= (1-r)\exp(-\gamma_+^2)\left[\exp(\gamma_-^2)\,\mathrm{erfc}(\gamma_-) + \exp(\gamma_+^2)\,\mathrm{erfc}(\gamma_+)\right] \\ &= (1-r)\exp(-\gamma_+^2)\omega_{++}. \end{aligned} \tag{6.9}$$

In case $\gamma_+ \geq 0$ and $\gamma_- < 0$, we use the identity

$$\operatorname{erfcx}(-x) = 2\exp(x^2) - \operatorname{erfcx}(x),$$

which follows directly from $\operatorname{erfc}(-x) = 2 - \operatorname{erfc}(x)$. With

$$\omega_{+-} := \operatorname{erfcx}(\gamma_+) - \operatorname{erfcx}(-\gamma_-) = \operatorname{erfcx}(\gamma_+) + \operatorname{erfcx}(\gamma_-) - 2\exp(\gamma_-^2),$$

z in (6.5) is given by

$$\begin{aligned} z_3 &= r\exp(-\gamma_-^2)\left[\operatorname{erfcx}(\gamma_-) + \operatorname{erfcx}(\gamma_+)\right] \\ &= r\exp(-\gamma_-^2)\left[\omega_{+-} + 2\exp(\gamma_-^2)\right] \\ &= r\left[\exp(-\gamma_-^2)\omega_{+-} + 2\right], \end{aligned} \tag{6.10}$$

whereas z in (6.6) is given by

$$\begin{aligned} z_4 &= (1-r)\exp(-\gamma_+^2)\left[\operatorname{erfcx}(\gamma_-) + \operatorname{erfcx}(\gamma_+)\right] \\ &= (1-r)\exp(-\gamma_+^2)\left[\omega_{+-} + 2\exp(\gamma_-^2)\right] \\ &= (1-r)\left[\exp(-\gamma_+^2)\omega_{+-} + 2\exp(-2c)\right]. \end{aligned} \tag{6.11}$$

In case $\gamma_+ < 0$ and $\gamma_- \geq 0$, we set

$$\omega_{-+} := \operatorname{erfcx}(-\gamma_+) - \operatorname{erfcx}(\gamma_-) = 2\exp(\gamma_+^2) - \operatorname{erfcx}(\gamma_+) - \operatorname{erfcx}(\gamma_-).$$

Here, z in (6.5) is given by

$$\begin{aligned} z_5 &= r\exp(-\gamma_-^2)\left[\operatorname{erfcx}(\gamma_-) + \operatorname{erfcx}(\gamma_+)\right] \\ &= r\exp(-\gamma_-^2)\left[-\omega_{-+} + 2\exp(\gamma_+^2)\right] \\ &= r\left[-\exp(-\gamma_-^2)\omega_{-+} + 2\exp(2c)\right], \end{aligned} \tag{6.12}$$

and z in (6.6) is given by

$$\begin{aligned} z_6 &= (1-r)\exp(-\gamma_+^2)\left[\operatorname{erfcx}(\gamma_-) + \operatorname{erfcx}(\gamma_+)\right] \\ &= (1-r)\exp(-\gamma_+^2)\left[-\omega_{-+} + 2\exp(\gamma_+^2)\right] \\ &= (1-r)\left[-\exp(-\gamma_+^2)\omega_{-+} + 2)\right]. \end{aligned} \tag{6.13}$$

Now, we compute the logarithms of the above expressions. Here, we denote the natural logarithm by log. We have

$$\log z_1 = \log r - \gamma_-^2 + \log \omega_{++}, \tag{6.14}$$

$$\log z_2 = \log(1-r) - \gamma_+^2 + \log \omega_{++}. \tag{6.15}$$

For the remaining expressions, we use the following identity: For $x > 0$ and $y \in \mathbb{R} \setminus \{0\}$ with $x + y > 0$,

$$\log(x+y) = \log x + \log(1 + \operatorname{sign}(y)\exp(\log|y| - \log x)). \tag{6.16}$$

For $\omega_{+-} \neq 0$ we obtain

$$\begin{aligned}
\log z_3 &= \log r + \log[2 + \exp(-\gamma_-^2)\omega_{+-}] \qquad (6.17)\\
&= \log r + \log 2 + \log\left(1 + \operatorname{sign}(\omega_{+-}) \exp\left(\log|\omega_{+-}| - \gamma_-^2 - \log 2\right)\right),\\
\log z_4 &= \log(1-r) + \log[2\exp(-2c) + \exp(-\gamma_+^2)\omega_{+-}]\\
&= \log(1-r) + \log 2 - 2c\\
&\quad + \log\left(1 + \operatorname{sign}(\omega_{+-}) \exp\left(\log|\omega_{+-}| - \gamma_+^2 - \log 2 + 2c\right)\right),\\
&= \log(1-r) + \log 2 - 2c \qquad (6.18)\\
&\quad + \log\left(1 + \operatorname{sign}(\omega_{+-}) \exp\left(\log|\omega_{+-}| - \gamma_-^2 - \log 2\right)\right),
\end{aligned}$$

using $\gamma_+^2 - \gamma_-^2 = 2c$. For $\omega_{-+} \neq 0$ we compute

$$\begin{aligned}
\log z_5 &= \log r + \log[2\exp(2c) - \exp(-\gamma_-^2)\omega_{-+}]\\
&= \log r + \log 2 + 2c\\
&\quad + \log\left(1 - \operatorname{sign}(\omega_{-+}) \exp\left(\log|\omega_{-+}| - \gamma_-^2 - \log 2 - 2c\right)\right)\\
&= \log r + \log 2 + 2c \qquad (6.19)\\
&\quad + \log\left(1 - \operatorname{sign}(\omega_{-+}) \exp\left(\log|\omega_{-+}| - \gamma_+^2 - \log 2\right)\right)\\
\log z_6 &= \log(1-r) + \log[2 - \exp(-\gamma_+^2)\omega_{-+}]\\
&= \log(1-r) + \log 2 \qquad (6.20)\\
&\quad + \log\left(1 - \operatorname{sign}(\omega_{-+}) \exp\left(\log|\omega_{-+}| - \gamma_+^2 - \log 2\right)\right).
\end{aligned}$$

If $\omega_{+-} = 0$ (or $\omega_{-+} = 0$), the expressions for z_3 and z_4 (or z_5 and z_6) simplify considerably, so that we do not have to use (6.16).

Finally, we can compute $\operatorname{erfcinv}(z)$. To this end let $w = \log z$. If z is not too small, we can compute $z = \operatorname{erfcinv}(\exp(w))$ using the standard implementation of erfcinv. For $\omega < -680$, we use the following asymptotic approximation of $z \approx \operatorname{erfcinv}(\exp(w))$ for $w \to -\infty$ from [DLMF, §7.17(iii)], which follows from [Blair, Edwards, and Johnson 1976], after modifications:

$$z := s^{-\frac{1}{2}} + a_2 s^{\frac{3}{2}} + a_3 s^{\frac{5}{2}} + a_4 s^{\frac{7}{2}},$$

where

$$\begin{aligned}
\theta &:= -\log\pi - \log(-w),\\
s &:= \frac{2}{\theta - 2w},\\
v &:= -\theta - 2,\\
a_2 &:= \frac{1}{8}v,\\
a_3 &:= -\frac{1}{32}\left(v^2 + 6v - 6\right),\\
a_4 &:= \frac{1}{384}\left(4v^3 + 27v^2 + 108v - 300\right).
\end{aligned}$$

6.5 Numerical Results

6.5.1 Laplacian Noise

First, we will plot samples of the Laplacian noise for different values of β to get an impression of its smoothness and compute empirical confidence regions to illustrate its spread. Figure 6.1 shows samples

$$\tilde{\eta} \sim \bigotimes_{k=1}^{N} \mathcal{L}_{b^2 \alpha_k^{-\beta}}$$

of the noise for differenct values of β, where $N = 800$. Here, b is chosen as $(\sum_{k=1}^{N} \alpha_k^{-\beta})^{-\frac{1}{2}}$, so that the variance

$$\mathbb{E}\left[\|\tilde{\eta}\|_{R^N}^2\right] = b^2 \sum_{k=1}^{N} \alpha_k^{-\beta}$$

of $\tilde{\eta}$ is equal to 1. Note that only for $\beta > \frac{1}{2}$, $\tilde{\eta}$ corresponds to Laplacian noise in $L^2(D)$, because only in this case $A^{-\beta}$ is trace class and hence the Laplacian measure $\mathcal{L}_{b^2 A^{-\beta}}$ on $L^2(D)$ is covered by our definition.

Next, we consider empirical credible regions for the noise and a range of values of β. We generate $M = 10000$ samples of the noise for each value of β and plot the empirical inverse cumulative distribution of their norm $\|\tilde{\eta}\|_2$ in Figure 6.2, i.e., for every value $r \in [0, 1]$ the minimal radius of a ball in $L^2(D)$ around 0 that contains a fraction r of the samples. We can see that although the noise has variance 1 in all cases, for smaller values of β the norm of the samples is more concentrated around 1, whereas for larger values of β their norm is spread out more into small values closer to 0 and large values greater than 1.

6.5.2 Frequentist Setting

In this subsection, we will plot MAP and CM estimator for Scenarios 1-3, juxtaposed with the true solution, and inspect them visually. We will study how the choice of the regularisation parameter r affects the mean squared error of both estimators and if the mean squared error converges to zero in the small noise limit. Additionally, we will compare the spread of both estimators around the true solution in the different scenarios by computing empirical confidence regions around them. We will discretise with $N = 180$ throughout the rest of this section, unless otherwise stated.

In Figures 6.3, 6.4 and 6.5, we take a first look at the MAP and CM estimator in Scenarios 1 to 3 for a fixed ratio between the standard deviation $b(\sum_{k=1}^{N} \alpha_k^{-\beta})^{\frac{1}{2}}$ of the noise and the norm $\|y^N\|_{L^2} = \|\tilde{y}\|_2$ of the noise-free data of $\frac{1}{1000}$ by choosing

$$b = 0.001\|\tilde{y}\|_2 \left(\sum_{k=1}^{N} \alpha_k^{-\beta} \right)^{-\frac{1}{2}}. \tag{6.21}$$

In Scenario 1, we set $\rho = \sup_{k \in \mathbb{N}} |(w, \varphi_k)_{L^2}|$ and choose $r = 1.1 \cdot 2^{-\frac{1}{4}} \rho^{\frac{1}{2}} b^{\frac{1}{2}}$ a priori, motivated by Theorem 5.50. In Scenarios 2 and 3 we anticipate the results of the subsequent experiment and

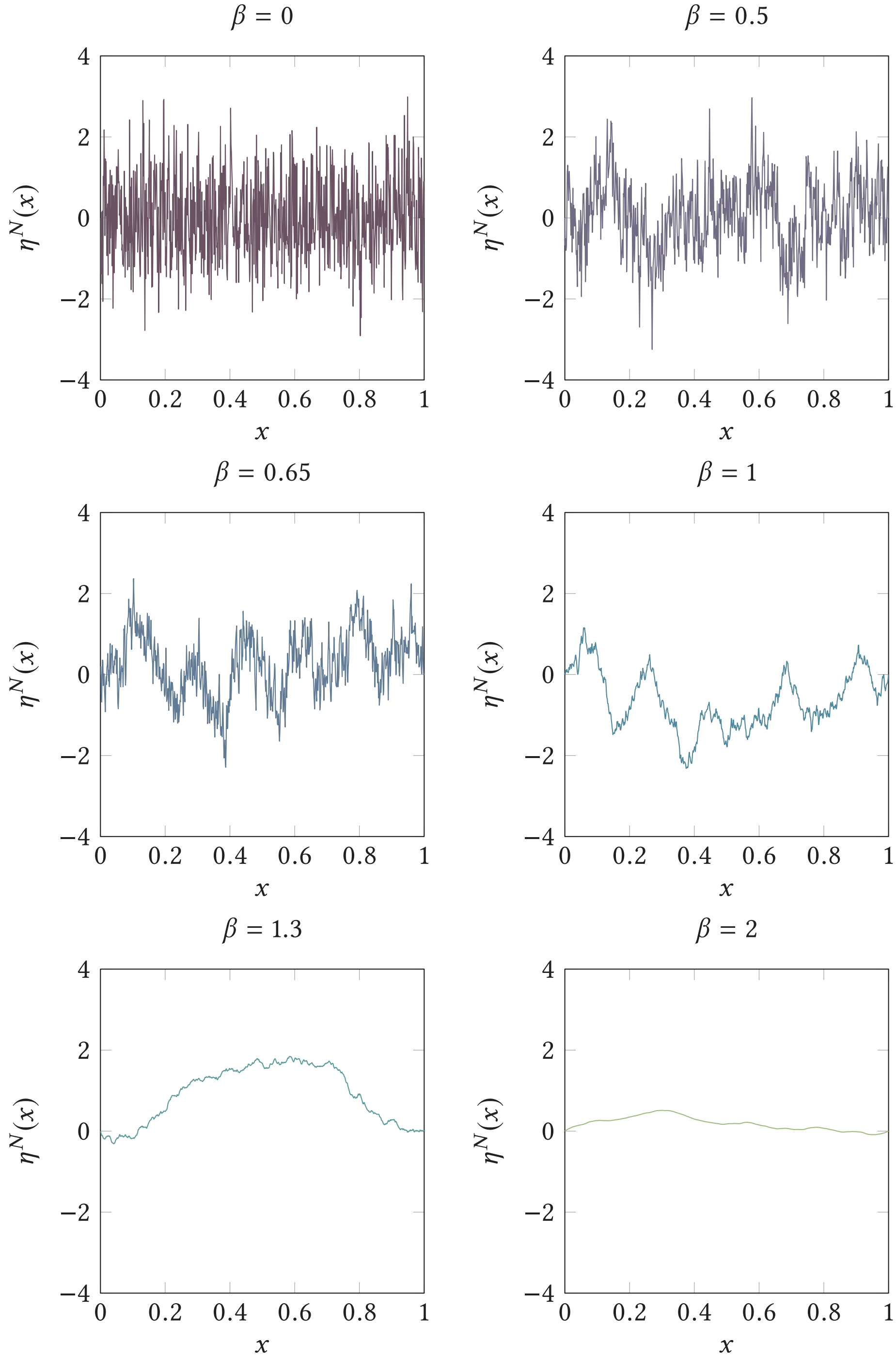

Figure 6.1: Samples of Laplacian noise $\tilde{\eta} \sim \bigotimes_{k=1}^{N} \mathcal{L}_{b^2 \alpha_k^{-\beta}}$ with $N = 800$.

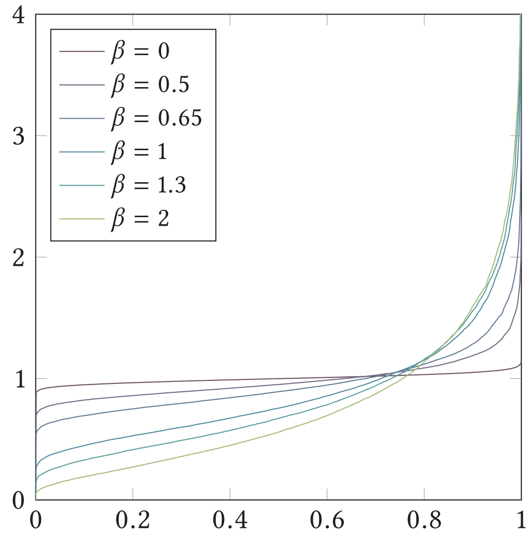

Figure 6.2: The empirical inverse cumulative distribution of $\|\tilde{\eta}\|_{R^N}$ for $M = 10000$ samples of Laplacian noise $\tilde{\eta} \sim \bigotimes_{k=1}^{N} \mathcal{L}_{b^2 \alpha_k^{-\beta}}$ with $N = 800$.

choose r optimal in the sense that it minimises the sum of the mean squared errors of MAP and CM estimator. Visually, both estimators appear very smooth, in particular smoother than the true solution $u^\dagger \in \mathcal{X}^\tau$ from Scenario 2.

Next, we consider the *mean squared error (MSE)*

$$\mathbb{E}\left[\|\hat{u} - u^\dagger\|_{L^2}^2\right] = \int_{L^2(D)} \|\hat{u}(y) - u^\dagger\|_{L^2}^2 \mathcal{L}_{e^{-A}u^\dagger, b^2 A^{-\beta}}(\mathrm{d}y)$$

of both estimators for different values of r, with b chosen according to (6.21) as before. Here $\hat{u}$ stands for $\hat{u}_{\mathrm{MAP}}$ or $\hat{u}_{\mathrm{CM}}$, respectively. We approximate the MSE of the respective estimator by its empirical mean squared error

$$\frac{1}{M} \sum_{m=1}^{M} \left\|\hat{\tilde{u}}(\tilde{y}_m) - \tilde{u}^\dagger\right\|_{\mathbb{R}^N}^2$$

for $M = 100$ realisations of the data $\tilde{y}_m = K\tilde{u}^\dagger + \tilde{\eta}_m$ determined by M samples $\tilde{\eta}_1, \ldots, \tilde{\eta}_M$ of the noise. In Figures 6.6, 6.7 and 6.8 we plot the empirical MSE of both estimators against r.

In Scenarios 1a and 1b we observe that the MSE of the MAP estimator decreases monotonically as r decreases up to the lower bound $r_0 := 2^{-\frac{1}{4}} \rho^{\frac{1}{2}} b^{\frac{1}{2}}$ from Theorem 5.50, where the MSE changes its behaviour abruptly and increases very rapidly. The MSE of the CM estimator displays roughly the same behaviour as the MAP estimator, it decreases monotonically as r decreases up to a point slightly above r_0 and then increases rapidly. In contrast to the MAP estimator, there is no abrupt change in its behaviour but a smooth transition. For the MAP estimator this suggests

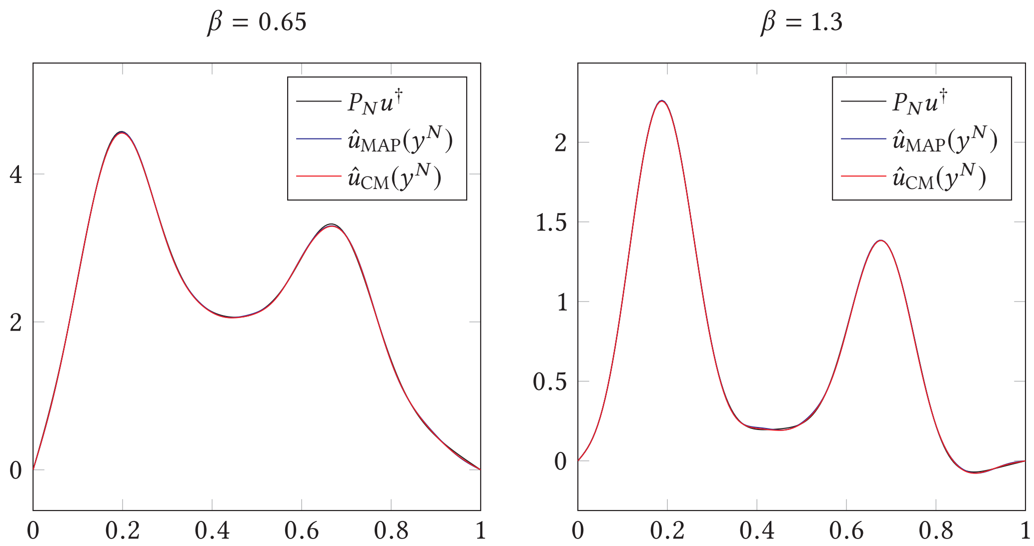

Figure 6.3: Scenarios 1a (left) and 1b (right) with $r = 1.1 \cdot 2^{-\frac{1}{4}} \rho^{\frac{1}{2}} b^{\frac{1}{2}}$ chosen a priori.

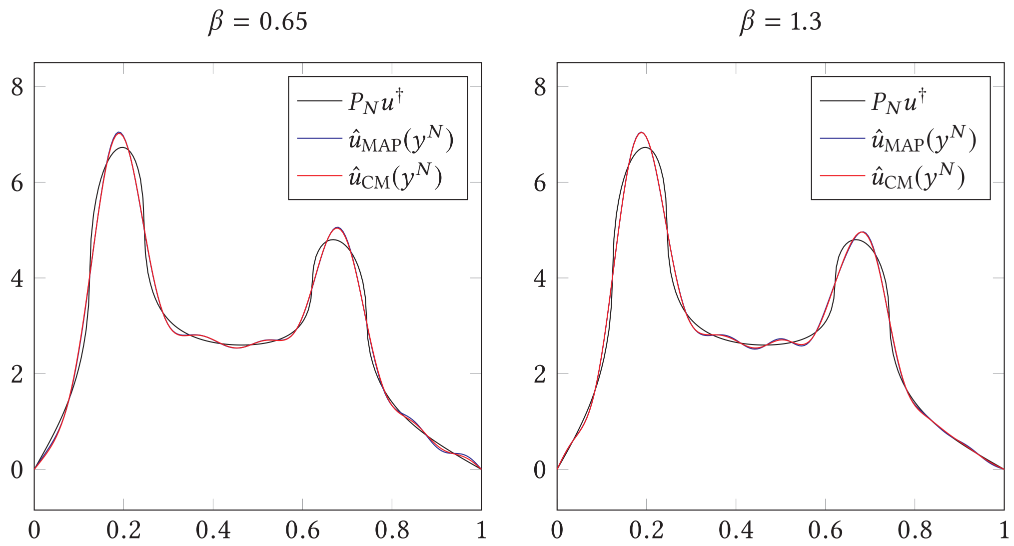

Figure 6.4: Scenarios 2a with $r = 0.0477$ (left) and 2b with $r = 0.0731$ (right).

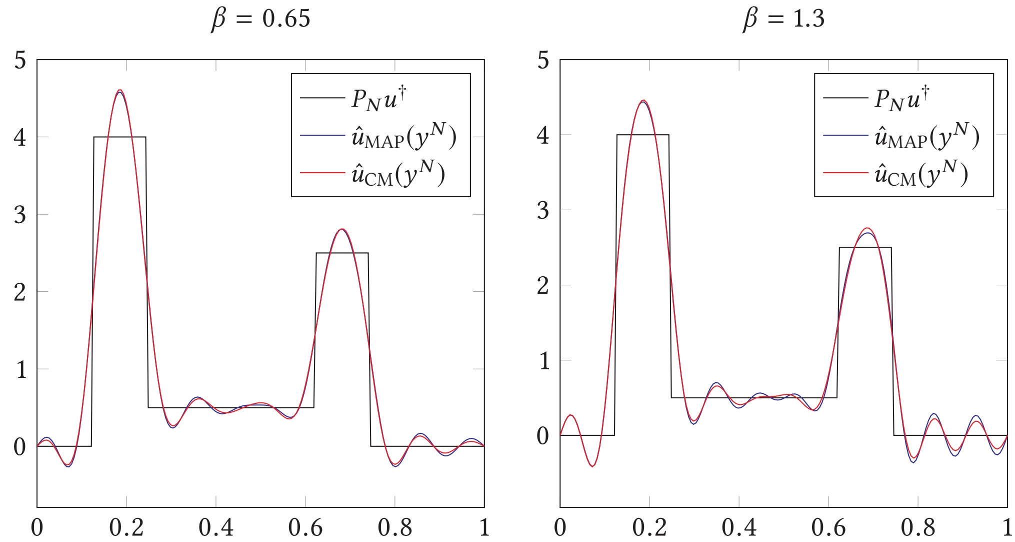

Figure 6.5: Scenarios 3a with $r = 0.1096$ (left) and 3b with $r = 0.1313$ (right).

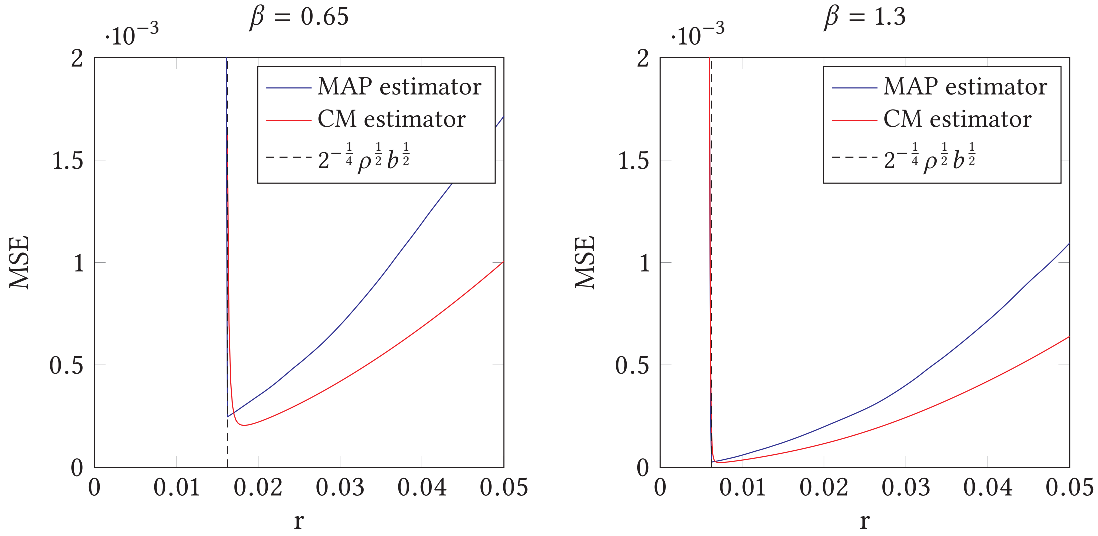

Figure 6.6: Mean squared error for different values of r for Scenarios 1a (left) and 1b (right) and the lower bound $2^{-\frac{1}{4}}\rho^{\frac{1}{2}}b^{\frac{1}{2}}$ for r from Theorem 5.50.

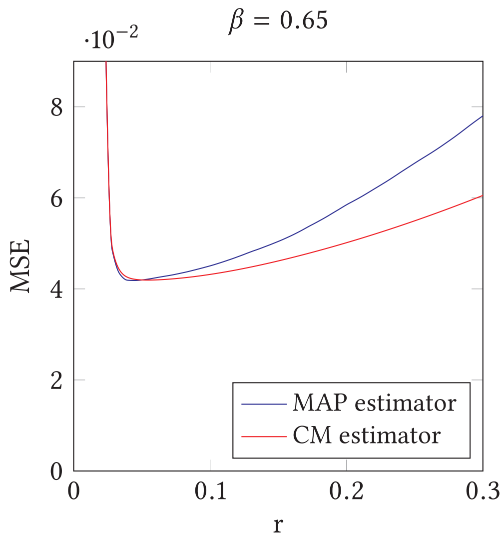

Figure 6.7: Mean squared error for different values of r for Scenarios 2a (left) and 2b (right).

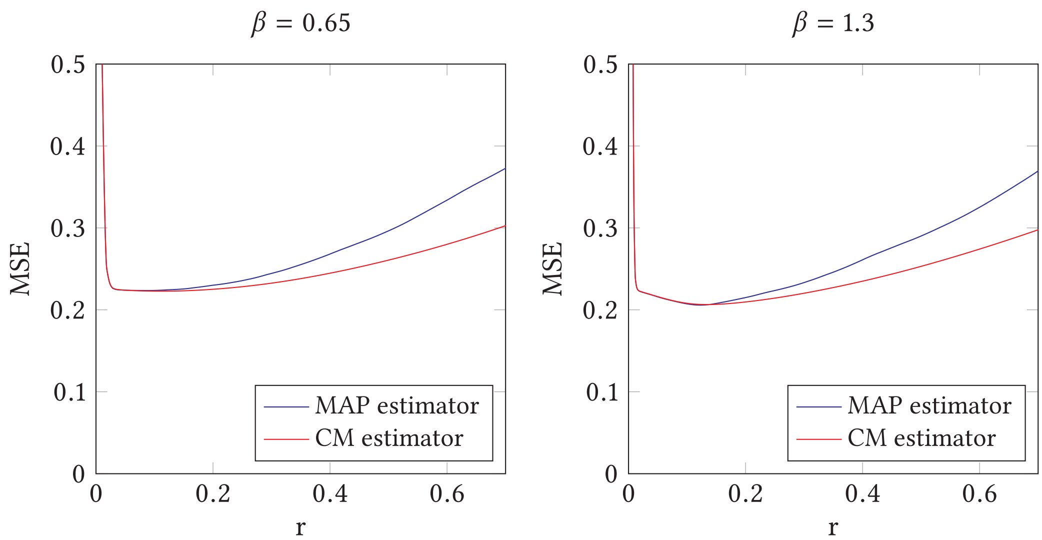

Figure 6.8: Mean squared error for different values of r for Scenarios 3a (left) and 3b (right).

choosing r as small as possible, but larger than r_0. Motivated by this, we will use the *a priori choice*

$$r = 1.1 \cdot 2^{-\frac{1}{4}} \rho^{\frac{1}{2}} b^{\frac{1}{2}}$$

in the following for Scenario 1.

In Scenarios 2 and 3 the mean squared errors of both estimators decrease monotonically as r decreases up to a certain point and then increase monotonically. In contrast to Scenario 1, there is a flat valley around the value of r that minimises the MSE, while for small r the MSE again grows very rapidly. In Scenarios 2a, 3a and 3b the MSE of both estimators transitions smoothly and so does the MSE of the CM estimator in Scenario 2b. The MAP estimator in Scenario 2b poses an exception. Here the MSE changes its behaviour abruptly at a small value of r. In contrast to Scenario 1 this sudden change does not occur in the value of r that minimises the MSE but outside of the valley around it.

In Figures 6.9 and 6.10 we study the frequentist consistency of both estimators numerically by considering the MSE and the *variance of the squared error (VSE)*

$$\mathrm{Var}\left(\|\hat{u} - u^\dagger\|_{L^2}^2\right) = \mathbb{E}\left[\left(\|\hat{u} - u^\dagger\|_{L^2}^2 - \mathbb{E}\left[\|\hat{u} - u^\dagger\|_{L^2}^2\right]\right)^2\right]$$

for values of b ranging roughly from 0.1 to $1.0 \cdot 10^{-10}$. Here, we use $M = 1000$ noise samples to approximate MSE and VSE.

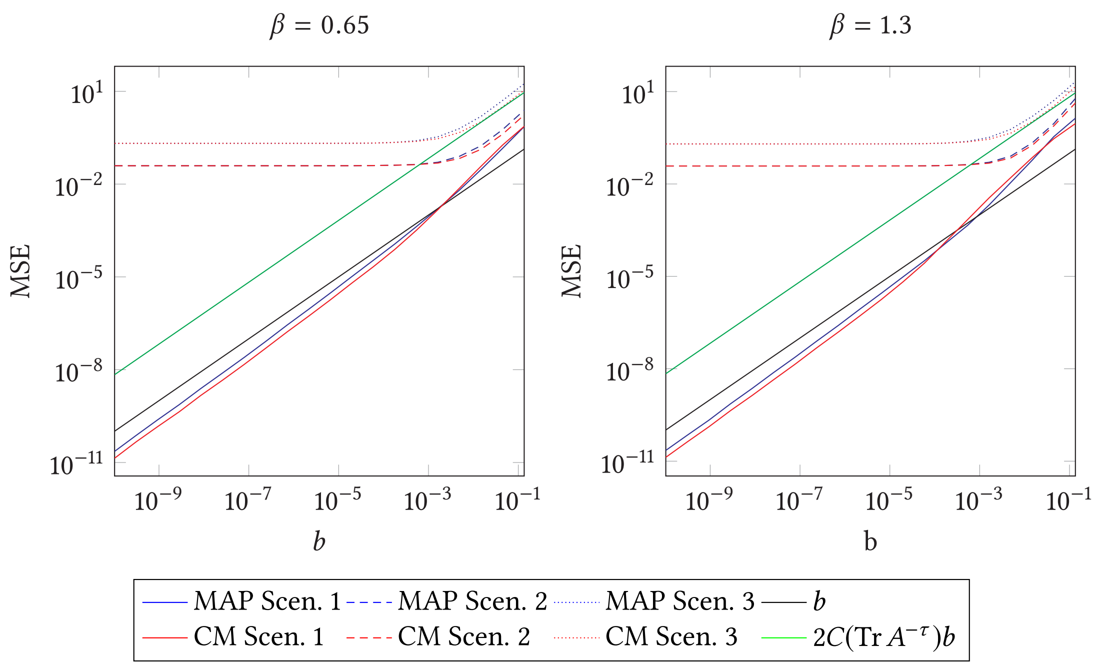

Figure 6.9: The mean squared error of both estimators for different values of b.

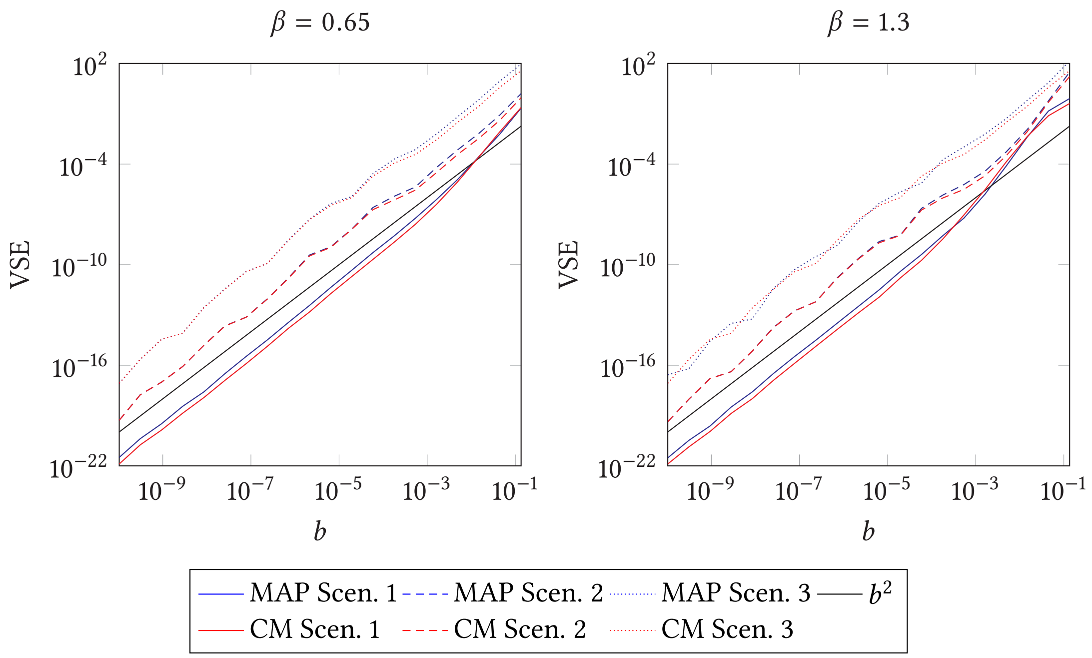

Figure 6.10: The variance of the squared error of both estimators for different values of b.

We observe, that in Scenarios 1a and 1b the mean squared error of both estimators converges to 0 as b tends to 0. The MAP estimator converges in the order of b, the rate estimated in Theorem 5.50. The CM estimator behaves, up to a constant, in the same way and also converges in the order of b. The MSE of the MAP estimator remains below the upper bound $2C(\operatorname{Tr} A^{-\tau})b$ from Theorem 5.50. The actual MSE is smaller than the bound by a factor between 12.2 ($b \approx 0.1$) and 298 ($b \approx 1.0 \cdot 10^{-10}$). In Scenarios 2 and 3 the MSE of neither MAP nor CM estimator converges to 0. It only decreases down to around 0.043 ($b \approx 5.4 \cdot 10^{-4}$) in Scenarios 2a and 2b and down to around 0.25 ($b \approx 5.4 \cdot 10^{-4}$) in Scenarios 3a and 3b. Then it remains practically constant for smaller values of b. In all considered scenarios the VSE of both estimators converges to 0 in the order of b^2.

Finally, we determine empirical confidence regions for the true solution $\tilde{u}^\dagger$. In particular, we consider balls in $\mathbb{R}^N$ around the MAP or CM estimate as confidence sets. We generate $M = 10000$ realisations of the data $\tilde{y} = K\tilde{u}^\dagger + \tilde{\eta}$, which in turn are determined by M samples of the noise $\tilde{\eta}$. For every value $p \in [0, 1]$ we seek the minimal radius, such that for a fraction p of the samples a ball around the respective estimate $\hat{\tilde{u}}(\tilde{y})$ with the same radius includes the true solution $\tilde{u}^\dagger$. This is equivalent to choosing for every $p \in [0, 1]$ the minimal radius of a ball around $\tilde{u}^\dagger$ that contains a fraction p of the respective estimates, or, in other words, finding the empirical inverse cumulative distribution of the error $\|\hat{\tilde{u}}(\tilde{y}) - \tilde{u}^\dagger\|_2$ of the respective estimates. In order to make the confidence sets from the different Scenarios quantitatively comparable, we normalise the true solution $\tilde{u}^\dagger$ in each scenario to satisfy $\|\tilde{u}^\dagger\|_2 = 1$. Moreover, we choose b such that the standard deviation $b(\sum_{k=1}^N \alpha_k^{-\beta})^{\frac{1}{2}}$ of the noise is equal to 10^{-3}. The results are plotted in Figure 6.11.

In Scenarios 2 and 3 there is a ball around $u^\dagger$ with radius 0.055 and 0.25, respectively, that contains none of the samples of the estimates, whereas virtually all samples are contained in a ball with a slightly larger radius of 0.07 and 0.3, respectively, suggesting that in these Scenarios for both estimators the bias is predominant and the effect of the variance is much smaller. Only in Scenario 1 the samples are spread out more smoothly and the effect of the bias is smaller in comparison. Here, the difference between a 0.05- and a 0.95-confidence region is more pronounced, while there still exists a ball with radius 0.002 around $u^\dagger$ that contains none of the samples.

6.5.3 Fully Bayesian Setting

Now, we consider CM and MAP estimator in a fully Bayesian setting. We will plot a number of samples of the posterior for different values of τ to visually inspect their smoothness and their spread and compare them to both estimators. Moreover, we will compare the spread of the posterior for different values of τ and β by computing empirical credible sets around MAP and CM estimator.

As before, we draw the noise $\tilde{\eta}$ from a Laplacian distribution

$$\bigotimes_{k=1}^{N} \mathcal{L}_{b^2\alpha_k^{-\beta}},$$

where b is chosen in such a way, that its standard deviation $b(\sum_{k=1}^{N}\alpha_k^{-\beta})^{\frac{1}{2}}$ is equal to 10^{-3}. The discretised prior $\tilde{u}$ by definition has the distribution $\mathcal{N}_{r^2A^{-\tau}} \circ \gamma_N^{-1}$, the pushforward of the prior distribution $\mathcal{N}_{r^2A^{-\tau}}$ under γ_N. And as $\gamma_N^{-1}(A) = I_{1,\dots,N;A}$ is a cylindrical set for every $A \in \mathcal{B}(\mathbb{R}^N)$, this measure is by definition a Gaussian product measure on $\mathbb{R}^N$, i.e.,

$$\mathcal{N}_{r^2A^{-\tau}} \circ \gamma_N^{-1} = \bigotimes_{k=1}^{N} \mathcal{N}_{r^2\alpha_k^{-\tau}}.$$

We thus draw the prior $\tilde{u}$ from a Gaussian distribution $\bigotimes_{k=1}^{N} \mathcal{N}_{r^2\alpha_k^{-\tau}}$, where r is chosen such that its standard deviation $r(\sum_{k=1}^{N}\alpha_k^{-\tau})^{\frac{1}{2}}$ is equal to 1. Then we set $\tilde{y} := K\tilde{u} + \tilde{\eta}$ and study the marginal posterior distribution μ^N.

We take a first look at $M = 8$ samples of the posterior $\tilde{u}_{\text{post}} := \tilde{u}|\tilde{y} \sim \mu^N$ for

$$\tau \in \{0.55, 1.55, 2.55\}$$

and $\beta = 0.65$. Here we choose $N = 1000$. In Figure 6.12 we plot the corresponding functions

$$u_{\text{post},m} = \sum_{k=1}^{N} (\tilde{u}_{\text{post}})_k \varphi_k \quad \text{for } m = 1, \dots, M.$$

The posterior samples resemble the prior sample in smoothness and distance to MAP and CM estimate. For all values of τ both estimates appear to be significantly smoother than the posterior samples.

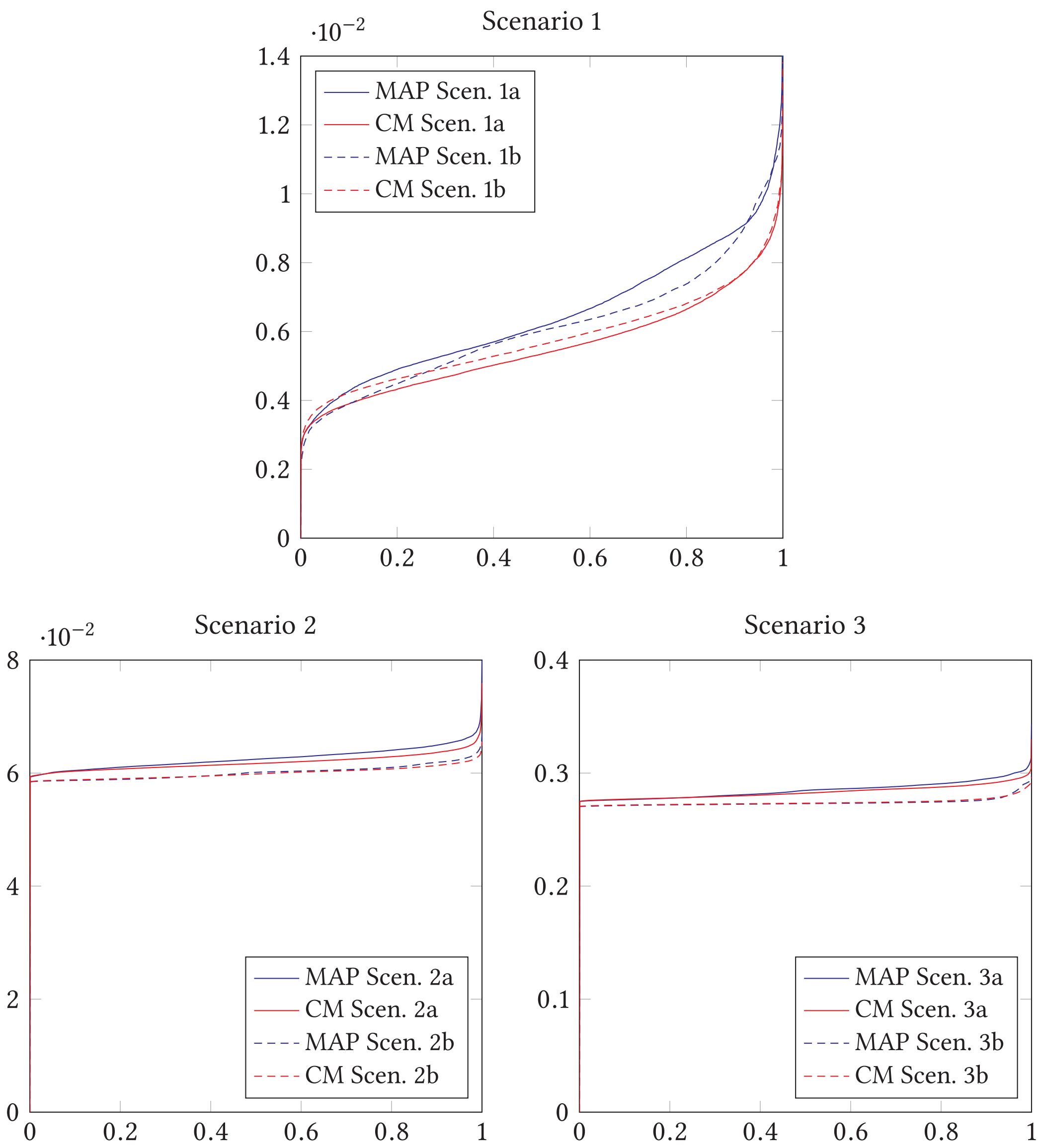

Figure 6.11: The empirical inverse cumulative distribution of $\|\hat{\tilde{u}}(\tilde{y}) - \tilde{u}^{\dagger}\|_{R^N}$ for $M = 10000$ samples.

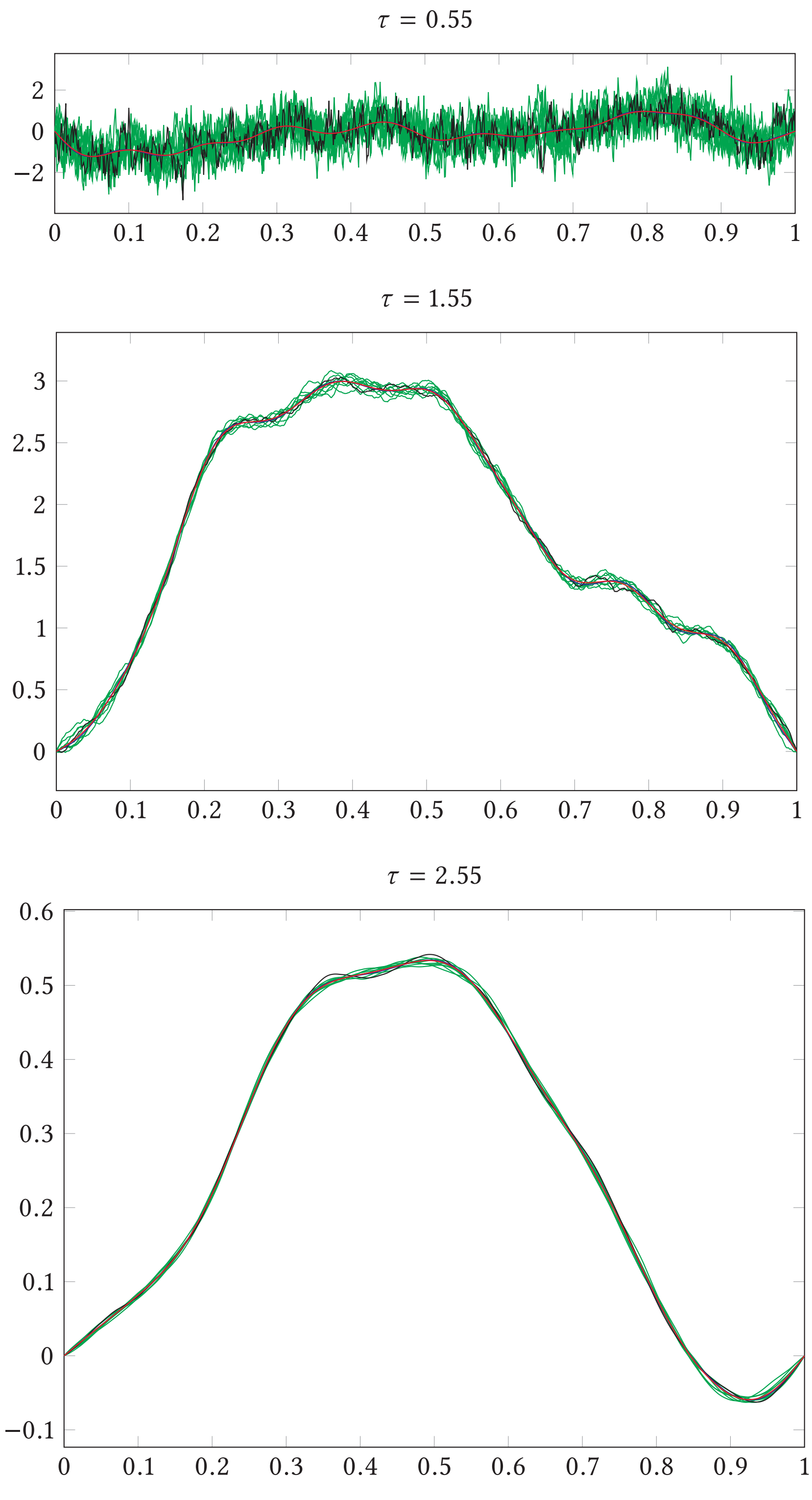

Figure 6.12: $M = 8$ samples —— of the posterior, the sample —— of the prior underlying the data, the MAP estimate —— and the CM estimate —— for different values of τ.

In Figure 6.13 we consider the convergence of the empirical posterior mean

$$\hat{\tilde{u}}_{\mathrm{CM},M} := \frac{1}{M}\sum_{m=1}^{M} \tilde{u}_{\mathrm{post},m}$$

of M samples of the posterior $\tilde{u}_{\mathrm{post}} \sim \mu^N$ towards the exact posterior mean, the CM estimate $\hat{\tilde{u}}_{\mathrm{CM}}(\tilde{y})$. Here we choose $\tau = 2.55$ and $\beta = 0.65$. We observe that the error $\|\hat{\tilde{u}}_{\mathrm{CM},M} - \hat{\tilde{u}}_{\mathrm{CM}}\|_2$

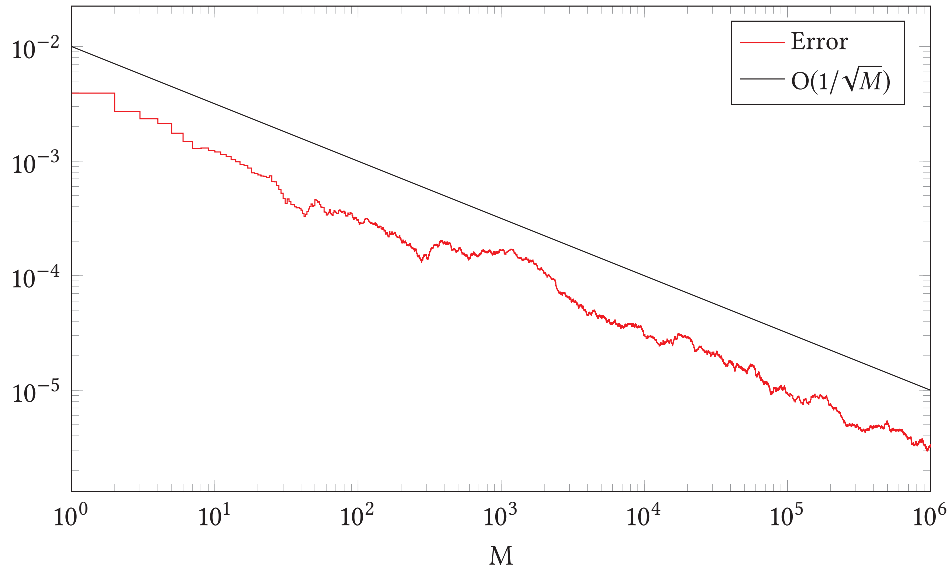

Figure 6.13: Error $\|\hat{\tilde{u}}_{\mathrm{CM},M} - \hat{\tilde{u}}_{\mathrm{CM}}\|_{\mathbb{R}^N}$ of the empirical posterior mean using M samples compared to the exact posterior mean.

converges to 0 in the order of $1/\sqrt{M}$ as M tends to infinity.

Finally, we consider empirical credible regions of the posterior around the MAP and the CM estimate. We generate $M = 100000$ samples of the posterior $\tilde{u}_{\mathrm{post}} \sim \mu^N$ and plot the empirical inverse cumulative distribution of their distance $\|\tilde{u}_{\mathrm{post}} - \hat{\tilde{u}}(\tilde{y})\|_2$ to the MAP or the CM estimate, respectively, in Figure 6.14. For every value $r \in [0,1]$ this is the minimal radius of a ball in $\mathbb{R}^N$ around $\hat{\tilde{u}}(\tilde{y})$ that contains a fraction r of the samples. Consequently, these balls can be considered as an approximation of credible regions for the marginal posterior $\tilde{u}_{\mathrm{post}}$, i.e., sets that contain $\tilde{u}_{\mathrm{post}}$ with a probability greater than or equal to r.

We do this for different values of τ and β. We consider a rougher prior with $\tau = 0.55$ (*Scenario 4*) and a smoother one with $\tau = 2.55$ (*Scenario 5*). Again, we divide each scenario into two subscenarios, one with $\beta = 0.65$, labeled *a*, and one with $\beta = 1.3$, labeled *b*. Scenario 5 corresponds to a prior u which belongs to $\mathcal{X}^t$ almost surely for every $t \in [0, 2.05)$, whereas for the prior corresponding to Scenario 4 this is the case only for $t \in [0, 0.05)$.

We observe that the smoother noise in Scenarios 4b and 5b leads to a slightly more concentrated posterior than the rougher noise in Scenarios 4a and 5a. The effect of a smoother prior, on the other hand, is tremendous; the posterior is much more concentrated than for a

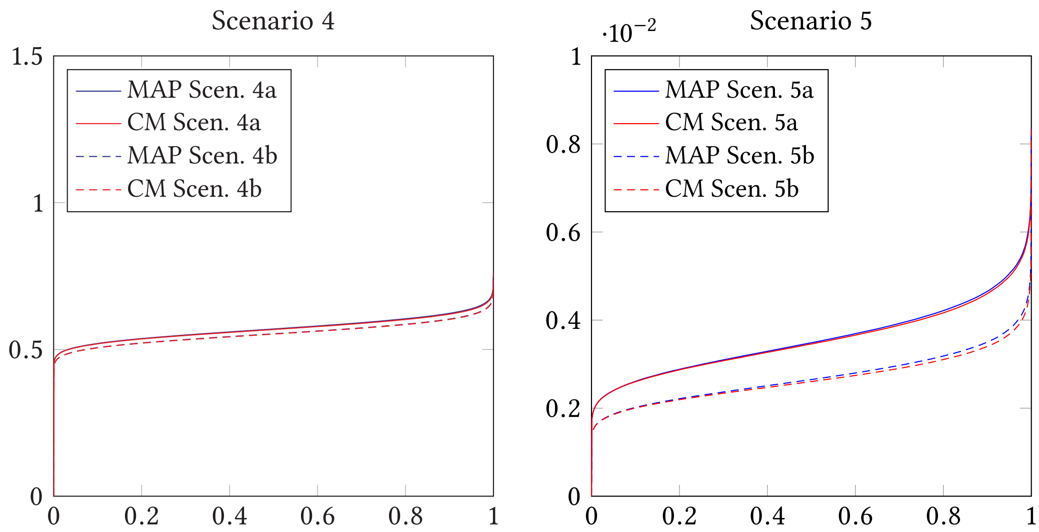

Figure 6.14: The empirical inverse cumulative distribution of $\|\tilde{u}_{\text{post}} - \hat{\tilde{u}}(\tilde{y})\|_{R^N}$ for $M = 100000$ samples of the posterior $\tilde{u}_{\text{post}}$.

rougher prior. In Scenario 5 the posterior belongs to a ball with radius $5.06 \cdot 10^{-3}$ around the MAP or the CM estimate with a probability of 0.95, compared to a ball with radius 0.639 in Scenario 4, although in both Scenarios the variance of the prior is the same. This also becomes apparent when we consider the empirical standard deviation of the posterior; in Scenario 5a it is approximately $3.39 \cdot 10^{-3}$, compared to 0.571 in Scenario 4a.

Conclusion

The equivalence of Laplacian infinite product measures under translations was studied. A Laplacian measure has the same admissible shifts as a Gaussian measure with the same covariance operator. Apart from that, the density of a shifted measure relative to a centred one displays similarities to the weighted ℓ^1-norm.

For nonlinear inverse problems, Tikhonov–Phillips regularisation with a quadratic norm penalty is equivalent to Bayesian MAP estimation with a Gaussian prior if the log-likelihood is Lipschitz continuous and chosen as a discrepancy term. This holds even if the log-likelihood is unbounded and thereby extends the main result from [Dashti, Law, et al. 2013].

This variational characterisation of MAP estimates was used to study consistency of the MAP estimator for a severely ill-posed linear problem with data corrupted by additive Laplacian noise. In this case, the posterior distribution has a density with respect to the prior distribution and the log-posterior density coincides with the weighted ℓ^1-norm up to a constant on finite subspaces. This yields a rigorous probability theoretical interpretation of variational regularisation with an ℓ^1-discrepancy term: The regularised solution can be understood as the mode of the posterior distribution. In a frequentist setting, the MAP estimator is asymptotically unbiased in the small noise limit if an a priori rule is employed to choose the regularisation parameter. Under an analytic source condition, the bias converges to zero at least in the order of the noise level, even if the regularisation parameter is chosen to be constant. The mean squared error of the MAP estimator converges towards the true solution at least in the order of the noise level if an analytic source condition holds and an a priori parameter choice rule is used. For an exponentially ill-posed linear problem, this rate coincides with the optimal asymptotic convergence rate in a minimax sense under the presence of Gaussian noise.

The behaviour and consistency of MAP and CM estimator were studied numerically for the classical inverse heat equation in one dimension with additive Laplacian noise. The lower bound from the a priori parameter choice rule was observed to be sharp, insofar as choosing the parameter below this bound results in a dramatic increase of the MSE. In a frequentist setting, the empirical MSE of both MAP and CM estimator converges to zero in the order of the noise level if an analytic source condition is satisfied. This means that here, the upper bound for the convergence rate of the MAP estimator is attained. In contrast, neither MAP nor CM estimator converge towards the true solution in mean square if only a Sobolev-type source condition is satisfied.

Future research could pursue the question if a rigorous statistical interpretation of minimisers of a Tikhonov–Phillips functional with an L^1-discrepancy term as Bayesian MAP estimates is possible, and if so, on which noise model it is based. More generally, it could be investigated if MAP estimates can be characterised as minimisers of an Onsager–Machlup functional when other heavy-tailed infinite-dimensional non-Gaussian noise models are utilised. The study of the considered linear problem could be continued by examining the posterior contraction rate in a frequentist setting or the consistency of the MAP estimator in a fully Bayesian setting. Also, sublevel sets of the Onsager–Machlup functional could be used as credible regions that capture the structure of the posterior distribution. Furthermore, one could examine if the MAP estimator minimises a cost functional involving the Bregman distance, i.e., if it is a Bayes estimator, for linear inverse problems with Laplacian infinite product noise, as it is the case for Gaussian noise in a finite-dimensional setting [Burger and Lucka 2014].

Bibliography

Agapiou, S., Burger, M., Dashti, M., and Helin, T. (2018). *Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems*. Inverse Problems 34.4, p. 045002. doi: 10.1088/1361-6420/aaacac.

Agapiou, S. and Mathé, P. (2018). "Posterior contraction in Bayesian inverse problems under Gaussian priors". In: New Trends in Parameter Identification for Mathematical Models. Springer International Publishing, Cham, pp. 1–29. doi: 10.1007/978-3-319-70824-9_1.

Benedetto, J. J. (2009). *Integration and Modern Analysis*. Birkhäuser Boston, Boston. doi: 10.1007/978-0-8176-4656-1.

Blair, J. M., Edwards, C. A., and Johnson, J. H. (1976). *Rational Chebyshev approximations for the inverse of the error function*. Math. Comp. 30.136, pp. 827–830. doi: 10.2307/2005402.

Bogachev, V. I. (1998). *Gaussian Measures*. Vol. 62. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI. doi: 10.1090/surv/062.

Bovik, A. (2005). *Handbook of Image and Video Processing*. Elsevier Inc.

Burger, M., Helin, T., and Kekkonen, H. (2018). *Large noise in variational regularization*. Transactions of Mathematics and Its Applications 2.1, pp. 1–45. doi: 10.1093/imatrm/tny002.

Burger, M. and Lucka, F. (2014). *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*. Inverse Problems 30, p. 114004. doi: 10.1088/0266-5611/30/11/114004.

Cavalier, L., Golubev, Y., Lepski, O., and Tsybakov, A. (2004). *Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems*. Theory of Probability & Its Applications 48 (3), pp. 426–446. doi: 10.1137/S0040585X97980555.

Clason, C., Helin, T., Kretschmann, R., and Piiroinen, P. (2019). *Generalized modes in Bayesian inverse problems*. SIAM/ASA Journal on Uncertainty Quantification 7.2, pp. 625–684. doi: 10.1137/18M1191804.

Clason, C., Jin, B., and Kunisch, K. (2010). *A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration*. SIAM Journal on Imaging Sciences 3.2, pp. 199–231. doi: 10.1137/090758003.

Da Prato, G. (2001). *An Introduction to Infinite-Dimensional Analysis*. Appunti/Scuola normale superiore. Pisa: Scuola Normale Superiore.

Dashti, M., Law, K. J. H., Stuart, A. M., and Voss, J. (2013). *MAP estimators and their consistency in Bayesian nonparametric inverse problems*. Inverse Problems 29.9, p. 095017. doi: 10.1088/0266-5611/29/9/095017.

Dashti, M. and Stuart, A. M. (2017). *The Bayesian approach to inverse problems*. In: Handbook of Uncertainty Quantification. Ed. by R. Ghanem, D. Higdon, and H. Owhadi. Springer International Publishing, Cham, pp. 311–428. doi: 10.1007/978-3-319-12385-1_7.

Ding, L. and Mathé, P. (2017). *Minimax rates for statistical inverse problems under general source conditions*. Computational Methods in Applied Mathematics. DOI: 10.1515/cmam-2017-0055.

Engl, H., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands.

Evans, L. C. (1998). *Partial Differential Equations*. Vol. 19. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. DOI: 10.1017/9781139029834.

Golubev, G. and Khasminskii, R. (2001). *A statistical approach to the Cauchy problem for the Laplace equation*. Lecture Notes-Monograph Series 36, pp. 419–433. URL: http://www.jstor.org/stable/4356124.

Grisvard, P. (2011). *Elliptic Problems in Nonsmooth Domains*. Vol. 69. Classics in Applied Mathematics. Reprint. Society for Industrial and Applied Mathematics, Philadelphia, PA. DOI: 10.1137/1.9781611972030.

Hohage, T. and Werner, F. (2014). *Convergence rates for inverse problems with impulsive noise*. SIAM Journal on Numerical Analysis 52.3, pp. 1203–1221. DOI: 10.1137/130932661.

Huber, P. J. (2009). *Robust Statistics*. 2nd ed. Wiley, Hoboken, N.J. DOI: 10.1002/9780470434697.

Kakutani, S. (1948). *On equivalence of infinite product measures*. Annals of Mathematics 49.1, pp. 214–224. URL: http://www.jstor.org/stable/1969123.

Kärkkäinen, T., Kunisch, K., and Majava, K. (2005). *Denoising of smooth images using L^1-fitting*. Computing 74.4, pp. 353–376. DOI: 10.1007/s00607-004-0097-8.

Kekkonen, H., Lassas, M., and Siltanen, S. (2016). *Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators*. Inverse Problems 32.8, p. 085005. DOI: 10.1088/0266-5611/32/8/085005.

Klenke, A. (2014). *Probability Theory*. 2nd ed. Universitext. Springer, London. DOI: 10.1007/978-1-4471-5361-0.

König, C., Werner, F., and Hohage, T. (2016). *Convergence rates for exponentially ill-posed inverse problems with impulsive noise*. SIAM Journal on Numerical Analysis 54.1, pp. 341–360. DOI: 10.1137/15M1022252.

Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Boston [u.a.] DOI: 10.1007/978-1-4612-0173-1.

Lie, H. C. and Sullivan, T. J. (2018). *Equivalence of weak and strong modes of measures on topological vector spaces*. Inverse Problems 34.11, p. 115013. DOI: 10.1088/1361-6420/aadef2.

Lifšic, M. A. (1995). *Gaussian Random Functions*. Kluwer Academic Publishers, Dordrecht [u.a.]

Lucka, F. (2012). *Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors*. Inverse Problems 28.12, p. 125012. DOI: 10.1088/0266-5611/28/12/125012.

Lucka, F. (2014). "Bayesian Inversion in Biomedical Imaging". PhD thesis. Westfälische Wilhelms-Universität Münster.

Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., Boisvert, R. F., Clark, C. W., Miller, B. R., and Saunders, B. V., eds. (2019). *NIST Digital Library of Mathematical Functions.* Release 1.0.22 of 2019-03-15. URL: `https://dlmf.nist.gov`.

Pazy, A. (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Vol. 44. Applied Mathematical Sciences. Springer New York. DOI: `10.1007/978-1-4612-5561-1`.

Roe, J. (1998). *Elliptic Operators, Topology and Asymptotic Methods.* 2nd ed. Vol. 395. Pitman Research Notes in Mathematics Series. Longman, Harlow.

Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory.* De Gruyter Studies in Mathematics. De Gruyter.

Werner, D. (2007). *Funktionalanalysis.* 6th ed. Springer, Berlin, Heidelberg. DOI: `10.1007/978-3-540-72536-7`.

Zeidler, E. (1995). *Applied Functional Analysis.* Vol. 108. Applied Mathematical Sciences. Springer-Verlag, New York. DOI: `10.1007/978-1-4612-0821-1`.