ON INFINITE-DIMENSIONAL HIERARCHICAL PROBABILITY MODELS IN STATISTICAL INVERSE PROBLEMS

TAPIO HELIN
Department of Mathematics and System Analysis
Helsinki University of Technology
P.O. Box 1100 (Otakaari 1 M), FI-02015 TKK, Finland

ABSTRACT. In this article, the solution of a statistical inverse problem \( M = AU + \mathcal{E} \) by the Bayesian approach is studied where \( U \) is a function on the unit circle \( T \), i.e., a periodic signal. The mapping \( A \) is a smoothing linear operator and \( \mathcal{E} \) a Gaussian noise. The connection to the solution of a finite-dimensional computational model \( M_{kn} = A_k U_n + \mathcal{E}_k \) is discussed. Furthermore, a novel hierarchical prior model for obtaining edge-preserving conditional mean estimates is introduced. The convergence of the method with respect to finer discretization is studied and the posterior distribution is shown to converge weakly. Finally, theoretical findings are illustrated by a numerical example with simulated data.

1. INTRODUCTION

Reconstruction methods with edge-preserving or -enhancing properties are widely studied topic in deterministic inverse problems. There exists a variety of different sophisticated approaches in the literature including functional regularization (e.g., the total variation approach [43]) or geometrical methods (e.g., the level set methods [45]). In the Bayesian inversion theory some methods have been introduced aiming for an edge-preserving point estimate in the finite-dimensional setting [19][8][12][42]. Especially the work by Calvetti and Somersalo [10][9] with hierarchical priors is closely related to this paper. In general it seems to be difficult to establish how the posterior distribution behaves asymptotically, i.e., as discretization of the problem gets finer. This is due to the fact that such methods usually require non-Gaussian prior modeling and the related infinite-dimensional Bayesian theory is not fully developed. This paper introduces a novel hierarchical structure leading to non-Gaussian prior modeling for signal segmenting problems. We show that the limiting behavior of our model can be analyzed.

Let us discuss the current perspectives in Bayesian modeling. Consider a linear inverse problem

\[ M = AU + \mathcal{E} \]

where \( U \) is the object of interest, \( \mathcal{E} \) a noise and \( M \) the measured data on some function spaces. In the Bayesian inversion these quantities are modeled as random variables and their probability distributions depict all information available prior to the measurement. With this information the goal is to make statistical inference on \( U \) given the model equation (1) and a realization \( M(\omega_0) \) of \( M \). Sometimes the prior distribution of the object of interest \( U \) depends on an unknown parameter which then becomes part of the modeling and inference problem. Such prior structures are often referred to as hierarchical models.

In practice the measurement is often produced by some finite-dimensional projection \( M_k = P_k M \). Furthermore, one also has to discretize \( U \) for computational purposes. This yields the computational model

\[ M_{kn} = P_k (AU_n + \mathcal{E}) = A_k U_n + \mathcal{E}_k. \]

Notice the two independent discretization levels \( n \) and \( k \). Solving the inverse problem with the Bayesian approach requires two steps: first, one translates all a priori information into the probability distributions of \( U_n \) and the noise \( \mathcal{E}_k \). The posterior probability \( P_{kn}(\cdot \mid m) \),
i.e., the probability distribution of \( U_n \) conditioned on the measurement \( m = M_k(\omega_0) \), is then obtained by using the Bayes formula and equation (2).

Usually the ultimate goal is to compute some information, e.g., point or spread estimates, from the posterior distribution. A point estimate that we discuss frequently in this paper is the conditional mean (CM) estimate which can be written for equation (2) in Euclidian spaces \( \mathbb{R}^n \) and \( \mathbb{R}^k \) as

\[
(3) \quad u_{kn}^{CM} = \int_{\mathbb{R}^n} u \, dP_{kn}(u \mid m).
\]

Now a natural question follows: what happens to the reconstructed information if \( U_n \) or \( \mathcal{E}_k \) is modeled on finer discretization, i.e., with a bigger \( n \) or \( k \)? Moreover, do the posterior probability distributions converge and how to guarantee that the reconstructed objects stay stable (e.g., CM estimate converges) as \( n \) and \( k \) increase?

The interplay between solutions of problems (1) and (2) in general situations is not fully understood. However, some partial results exist. In fact, if \( U_n \) and \( \mathcal{E}_k \) are obtained by projections from Gaussian distributions the convergence of posterior distribution has been proved in very general setting by Lasanen in [32]. To the author’s knowledge only convergence studies with non-Gaussian posterior distribution have been done from this point of view recently in [41] and [34]. These first positive results show some general conditions for obtaining weakly converging posterior distributions and in addition converging CM estimates. We emphasize that these results require Gaussian noise distribution.

Yet another non-trivial question is how to make sure that the crucial statistical properties of posterior distribution are not lost asymptotically? This is highly relevant to the edge-preserveness discussed above. Namely, in [33] it was shown that the usual modeling of TV prior carries an unpleasant defect such that the edge-preserving property is lost from the CM reconstructions as dimensionality of the problem increases. The reason behind this is that under different parameterization the prior distribution either converges to a Gaussian smoothness prior or diverges. In [34] a non-Gaussian prior structure is proposed for edge-preserving CM estimates. The estimates \( u_{kn}^{CM} \) are shown to converge to so-called reconstructors that generalize the concept of CM estimates in infinite-dimensional spaces. We discuss this in more details later. The work by Piiroinen in [41] contains results about the existence of a discretization leading to converging posterior information in general non-Gaussian setting.

Let us now review other related literature on the topic. First results on the Bayesian inversion in infinite-dimensional function spaces were introduced in [16] by Franklin. This research has then been continued and generalized by Mandelbaum [37], Lehtinen, Päivärinta and Somersalo [35], Fitzpatrick [15], and Luschgy [15]. Lastly, we want to stress that the convergence of posterior distributions has also been studied from different perspectives. Namely, in [26, 27, 40] such convergence is studied when the objective information becomes more accurate. Also, model reduction problems are considered in [29]. For a general presentation on the Bayesian inverse problems theory and computation see [28] and [11]. The topic of probability theory in Banach spaces is covered in [49].

This paper studies the problem of edge-preserving reconstructions in signal restoration problems with the emphasis on how to locate discontinuities. For technical reasons we concentrate on periodic signals, i.e., the domain for our study is a 1-dimensional sphere \( \mathbb{T} \). We model our prior beliefs of the unknown signal \( u \) with a hierarchical structure \((U, V)\) where the auxiliary random variable \( V \) models how the discontinuities are distributed. The conditional distribution of \( U \) given a sample of \( V \) then models our prior information about \( u \) if we know where the discontinuities are located. Such Bayesian modeling has close connection to previous hierarchical segmentation methods [19, 10, 9]. The method draws also a lot of inspiration from the celebrated Mumford–Shah image segmentation method [39] and its variational approximation introduced by Ambrosio and Tortorelli [2, 3].
In this paper we introduce a finite-dimensional prior structure \((U_n, V_n)\) that produces a weakly converging posterior structure in the presence of a Gaussian noise. The main theoretical results concerning the prior can be divided into three parts:

(i) There exists a well-defined random variable \((U, V) : \Omega \to L^2(\mathbb{T}) \times L^2(\mathbb{T})\) to which \((U_n, V_n)\) converges in distribution.

(ii) The posterior distributions \(P_{kn}\) converge weakly in \(L^2(\mathbb{T}) \times L^2(\mathbb{T})\) assuming that the measurements converge.

(iii) The CM estimate \((u_{kn}^{CM}, v_{kn}^{CM})\) converges to reconstructors of problem \((\Pi)\).

In addition we improve the results in [24] concerning the general theory. We implement our method in practice and include some numerical examples with computer generated data. The connection of maximum a posteriori (MAP) estimates to Ambrosio-Tortorelli minimizers that was presented in [25] is not studied here.

This paper is organized as follows. In Section 2 we introduce relevant concepts and main results concerning the general theory. The infinite-dimensional hierarchical prior model \((U, V)\) in \(L^2(\mathbb{T}) \times L^2(\mathbb{T})\) is defined in Section 3. We carefully show that such a construction is well-defined. Discretized prior distributions for \((U_n, V_n)\) are constructed in Section 4. It is important to note that we can explicitly write down the related density functions. This becomes highly valuable in numerical implementation as no more approximations need to be made. Section 5 is divided into three parts. First the theorems of Section 2 are proved. Secondly, we show here that \((U_n, V_n)\) converges to \((U, V)\) in distribution on \(L^2(\mathbb{T}) \times L^2(\mathbb{T})\). We conclude Section 5 by showing the important property of uniformly finite exponential moments for the introduced prior structure. Finally in Section 6 we illustrate with numerical examples how our method works in practice.

2. General setting

Next we define problem \((\Pi)\) rigorously. In order to do so let us introduce some notations. Below \(\langle \cdot, \cdot \rangle\) refers to pairing of generalized functions with test functions. In real Banach space \(B\) the dual pairing is denoted by \(\langle \cdot, \cdot \rangle_{B' \times B}\). In a real Hilbert space \(H\) we denote the inner product by \(\langle \cdot, \cdot \rangle_H\). We denote the Borel sets in \(B\) by \(B(B)\). Throughout this paper whenever not explicitly mentioned we assume the measurable structure of Borel sets. The notation \(\mathcal{L}(B_1, B_2)\) stands for the space of bounded linear operators between Banach spaces \(B_1\) and \(B_2\), and \(\mathcal{L}(B, B)\) is abbreviated as \(\mathcal{L}(B)\). If the operator \(T : B_1 \to B_2\) is a bounded linear operator, we denote the adjoint operator by \(T' : B_2' \to B_1'\). Recall also that a bounded linear operator \(T\) in a Hilbert space \(H\) is said to be in the trace class if

\[
\text{Tr}_H(T) := \sum_{j=1}^{\infty} \langle Te_j, e_j \rangle_H < \infty
\]

for some orthonormal basis \(\{e_j\}_{j=1}^{\infty} \subset H\). We want to point out that the definition is independent of the choice of the basis. Throughout the paper if not explicitly mentioned \(C\) denotes a positive constant. For two functions \(f, g : X \to \mathbb{R} \cup \{\infty\}\) we also write \(f \leq g\) if there exists a constant \(C > 0\) such that \(f \leq C g\) as functions. Finally, for any \(s \in \mathbb{R}\), let \(H^s(\mathbb{T})\) be the \(L^2\)-based Sobolev space equipped with Hilbert space inner product

\[
\langle \phi, \psi \rangle_{H^s} = \int_{\mathbb{T}} \langle (I - \Delta)^{s/2} \phi(x), (I - \Delta)^{s/2} \psi(x) \rangle dx
\]

for any \(\phi, \psi \in H^s(\mathbb{T})\).

Let us return to considering problem \((\Pi)\). Let \((\Omega, \Sigma, P)\) be a complete probability space with a product structure \(\Omega = \Omega_{pr} \times \Omega_{er}, \Sigma = \Sigma_{pr} \otimes \Sigma_{er}\) and \(P = P_{pr} \otimes P_{er}\). Throughout this section \(H\) will be fixed to denote a real separable Hilbert space. We assume the following conditions:

(i) The mapping \(U : \Omega_{pr} \to H\) is a random variable.

(ii) The mapping \(A : H \to H^1(\mathbb{T})\) is a bounded linear operator.
(iii) The random variable $\mathcal{E} : \Omega_{\mathcal{E}} \rightarrow H^{-1}(\mathbb{T})$ is Gaussian with expectation $E\mathcal{E} = 0$ and a covariance operator $C_{\mathcal{E}} : H^{-1}(\mathbb{T}) \rightarrow H^{-1}(\mathbb{T})$.

(iv) The range of $C_{\mathcal{E}}$ is dense in $H^{-1}(\mathbb{T})$.

The conditions (iii) and (iv) imply that $C_{\mathcal{E}}$ is one-to-one, self-adjoint and in the trace class and that we have a unique positive and self-adjoint power $C_{\mathcal{E}}^t$ for any $t \in \mathbb{R}$. Later in numerical examples $\mathcal{E}$ has a covariance operator $C_{\mathcal{E}} = (I - \Delta)^{-1} : H^{-1}(\mathbb{T}) \rightarrow H^1(\mathbb{T})$.

Such a random variable is white noise in the sense of generalized random variables [32].

**Definition 1.** Let $\mu$ be a centered Gaussian measure on $(H, \mathcal{B}(H))$ and its covariance operator $C : H \rightarrow H$ such that $\text{Ran}(C)$ is dense in $H$. We call the real separable Hilbert space

$$H(\mu) = \left\{ f \in H \mid \left\| C_X^{-1/2} f \right\|_H < \infty \right\}$$

equipped with inner product

$$\langle f, g \rangle_{H(\mu)} = \langle C_X^{-1/2} f, C_X^{-1/2} g \rangle_H$$

for any $f, g \in H(\mu)$ the Cameron-Martin space (or the reproducing kernel Hilbert space) of $\mu$.

This definition can be seen to coincide with the usual definition of Cameron-Martin spaces by Proposition 2.9 in [13]. The Cameron-Martin space structure is used later in Section 4. For an extensive presentation on the topic in locally convex spaces see [7].

If $U \in L^1(\Omega, \Sigma; H)$ and $\Sigma_0$ is a sub $\sigma$-algebra of $\Sigma$, we denote the conditional expectation of $U$ with respect to $\sigma$-algebra $\Sigma_0$ by $E(U|\Sigma_0)$. That is, $E(U|\Sigma_0) \in L^1(\Omega, \Sigma_0; H)$ and it satisfies

$$\int_D E(U|\Sigma_0)(\omega)\mathbb{P}(d\omega) = \int_D U(\omega)\mathbb{P}(d\omega) \quad \text{for all } D \in \Sigma_0.$$

All vector-valued integrals in this work are standard Bochner integrals. For more information on Bochner integrals see [14]. The operator $P_{\Sigma_0} : U \mapsto E(U|\Sigma_0)$ is a projection $P_{\Sigma_0} : L^1(\Omega, \Sigma; H) \rightarrow L^1(\Omega, \Sigma_0; H)$, where $L^1(\Omega, \Sigma_0; H)$ denotes the space of measurable functions from $(\Omega, \Sigma_0)$ to $(H, \mathcal{B}(H))$ which are Bochner integrable.

**Definition 2.** Denote by $\mathcal{M} \subset \Sigma$ the $\sigma$-algebra generated by the random variable $M$. We say that any deterministic function

$$R_{M}(U) : H^{-1}(\mathbb{T}) \rightarrow H, \quad m \mapsto R_{M}(U|m),$$

is a reconstructor of $U \in L^1(\Omega, \Sigma; H)$ with measurement $M$ if

$$R_{M}(U|M(\omega)) = E(U|M(\omega)) \quad \text{almost surely.}$$

If $\tilde{H}$ is a real separable Hilbert space, $g : (H, \mathcal{B}(H)) \rightarrow (\tilde{H}, \mathcal{B}(\tilde{H}))$ is a measurable function and $g(U) \in L^1(\Omega, \Sigma; \tilde{H})$, we define $R_{M}(g(U)|\cdot) : H^{-1}(\mathbb{T}) \rightarrow \tilde{H}$ to be any deterministic function satisfying

$$R_{M}(g(U)|M(\omega)) = E(g(U)|M(\omega)) \quad \text{almost surely.}$$

We refer to [34] for the existence of $R_{M}$. Note that although $R_{M}$ is not necessarily unique it was shown in [34] that in the presence of Gaussian noise the following choice can be made: Assume that the prior distribution $\lambda$ of $U$ has finite exponential moments, i.e.,

$$\int_H \exp(c \|u\|_H) d\lambda(u) < \infty$$

for any $c \in \mathbb{R}$, and assume $\tilde{H}$ is a real separable Hilbert space. Furthermore, let $g : (H, \mathcal{B}(H)) \rightarrow (\tilde{H}, \mathcal{B}(\tilde{H}))$ be a measurable function satisfying $E \|g(U)\|_{\tilde{H}} < \infty$. Then a
function $R_M(U|\cdot) : H^{-1}(\mathbb{T}) \to H$ defined by formula
\begin{equation}
R_M(g(U) | m) = \frac{\int_H g(u) \Xi(u, m) d\lambda(u)}{\int_H \Xi(u, m) d\lambda(u)}
\end{equation}
is a reconstructor, where $\Xi : H \times H^{-1}(\mathbb{T}) \to \mathbb{R}$ is the function
\[\Xi(u, m) = \exp\left(-\frac{1}{2} ||Au||_{L^2}^2 + \langle C^{-1}E^{-1}u, m \rangle_{H^{-1}} \right)\].

Throughout this paper we make the above choice of reconstructors.

As was discussed earlier the measurement is never infinite-dimensional in practice. Let us next explain how we assume the measurement to be obtained.

**Definition 3.** The finite-dimensional linear projections $P_k : H^{-1}(\mathbb{T}) \to H^{-1}(\mathbb{T})$, $k \in \mathbb{N}$, are called proper measurement projections when they satisfy the following conditions:

(i) We have $\text{Ran}(P_k) \subset H^1(\mathbb{T})$ and $\|P_k\|_{\mathcal{L}(H)} \leq C_0$ for some constant $C_0$ with all $k \in \mathbb{N}$.

(ii) For $t \in \{-1, 1\}$ we have
\[\lim_{k \to \infty} \|P_k f - f\|_{H^t} = 0\]
for all $f \in H^t(\mathbb{T})$.

(iii) For all $\phi, \psi \in L^2(\mathbb{T})$ it holds that
\[\langle P_k \phi, \psi \rangle_{L^2} = \langle \phi, P_k \psi \rangle_{L^2}\].

The conditions in Definition 3 are as in [34, Thm. 3] and are motivated there. We note that in this paper these assumptions are only used in the proof of Theorem 1.

In practical situations the measurement is a realization of a random variable
\begin{equation}
M_k = P_k M = A_k U + \mathcal{E}_k,
\end{equation}
where $A_k = P_k A$, $\mathcal{E}_k = P_k \mathcal{E}$. In order to be able to compute a numerical solution one has to discretize also the random variable $U$ (independently of $P_k$) in $H$. Denote the discretization by $U_n : \Omega \to H_n \subset H$ in a finite-dimensional subspace $H_n$. Now the two discretizations with respect to $n$ and $k$ lead to the computational model (2). We note that the reconstructor can be defined for all above models, for problem (1) on $H^{-1}(\mathbb{T})$ and for problems (2) and (9) on $\text{Ran}(P_k)$. Before next definition recall that probability measures $\mu_n, n \in \mathbb{N}$, converge weakly to $\mu$ in $(H, \mathcal{B}(H))$ if for every bounded and continuous function $f : H \to \mathbb{R}$ it holds that
\[\lim_{n \to \infty} \int_H f(u) d\mu_n(u) = \int_H f(u) d\mu(u)\].

In the following definition we characterize a condition that allows converging probability measures to have only very small tails.

**Definition 4.** We call measures $\mu$ and $\mu_n, n \in \mathbb{N}$, on $(H, \mathcal{B}(H))$ uniformly discretized with exponential weights if

(i) $\mu_n$ converges weakly to $\mu$ on $H$ and

(ii) for every $b > 0$ there exists a constant $0 < C(b) < \infty$ such that
\[\int_H \exp(b \|u\|_H) d\mu_n(u) \leq C(b) \quad \text{and} \quad \int_H \exp(b \|u\|_H) d\mu(u) \leq C(b)\]
for every $n \in \mathbb{N}$.

We are now ready to formulate our main theorem regarding the general theory. We postpone the proof to Section 5.1.

**Theorem 1.** Assume the following three conditions:

(i) The operators $P_k : H^{-1}(\mathbb{T}) \to H^{-1}(\mathbb{T})$, $k \in \mathbb{N}$, are proper measurement projections.
(ii) The probability distributions of $U_n, U : \Omega \to H, n \in \mathbb{N}$, are uniformly discretized with exponential weights.

(iii) A continuous function $g : H \to \widetilde{H}$ where $\widetilde{H}$ is a real separable Hilbert space, satisfies

$$
\|g(u)\|_{\widetilde{H}} \leq C \exp(C \|u\|_H)
$$

for all $u \in H$ with some constant $C$.

Now let $u = U(\omega_0)$ and $\epsilon = \mathcal{E}(\omega_0)$ be realizations of the random variables $U$ and $\mathcal{E}$, respectively, and let

$$
m = Au + \epsilon \quad \text{and} \quad m_k = A_k u + P_k \epsilon
$$

be the realizations of the random variables $M$ and $M_k$ in equations (1) and (2), respectively. Then the reconstructors defined by formula (3) for models (1) and (2) satisfy

$$
\lim_{k,n \to \infty} \mathcal{R}_{M_k n} (g(U_n) \mid m_k) = \mathcal{R}_M (g(U) \mid m)
$$

in $\widetilde{H}$.

Let $E \subset H$ be a Borel set and $1_E$ be the indicator function of $E$. Define probability measures

$$
\mathcal{P}(E \mid m) = \mathcal{R}_M (1_E(U) \mid m),
$$

$$
\mathcal{P}_k n (E \mid m_k) = \mathcal{R}_{M_k n} (1_E(U) \mid m_k)
$$

on $H$ with the same choices of reconstructors made in Theorem [1]. One notices that these measures correspond to the posterior distribution obtained from Bayes formula in the finite-dimensional case. An important corollary to Theorem [1] is shown in [34].

**Corollary 1.** Let the assumptions in Theorem [1] hold. Then the measures $\mathcal{P}_k n (\cdot \mid m_k)$ converge weakly to the measure $\mathcal{P}(\cdot \mid m)$ on $H$.

We conclude this section by discussing shortly how to solve reconstructors in practice. For the moment assume that all the conditions in Theorem [1] hold and $\dim \operatorname{Ran}(P_k) = K \in \mathbb{N}$. Moreover, assume $U_n : \Omega \to H_n \subset H$ where $\dim H_n = N \in \mathbb{N}$. Let $\mathcal{I}_n : H_n \to \mathbb{R}^N$ and $\mathcal{K}_k : \operatorname{Ran}(P_k) \to \mathbb{R}^K$ be isometries and let us use them to map the computational model (2) into a matrix equation. In the following we use bolded notation for vectors and matrices in Euclidian spaces. Denote $U_n = \mathcal{I}_n U_n = (u_1^N, \ldots, u_N^N)^T : \Omega \to \mathbb{R}^N$. This yields

$$
M_k n = \mathcal{K}_k M_k n = A_k n U_n + E_k
$$

where $A_k n \in \mathbb{R}^{K \times N}$ and $M_k n, E_k : \Omega \to \mathbb{R}^K$. The posterior density function $\pi_k n$ can now be easily obtained for problem (10) via the Bayes formula. In Section [6] assumptions on the noise $\mathcal{E}$ and the measurement projections imply that $E_k$ is white noise. In such a case $\pi_k n$ has the form

$$
\pi_k n (u_n \mid m_k) = \frac{\Pi_n (u_n) \exp(-\frac{1}{2} \|m_k - A_k n u_n\|^2)}{\Upsilon_k n (m_k n)},
$$

where $\Pi_n$ is the prior density and $\Upsilon_k n$ is the density function of $M_k n$. For a related discussion on the discretization of white noise see the Appendix B in [34]. The CM estimate corresponds to a reconstructor with $g = \text{id} : H \to H$ and it can be obtained by computing integral

$$
u^{CM} = \int_{\mathbb{R}^N} u \pi_k n (u \mid m_k) du
$$

since with the choice of reconstructors in equation (8) it holds that

$$
\mathcal{R}_{M_k n} (U_n \mid m_k) = \mathcal{I}_n^{-1} \left( \nu^{CM} \right)
$$

for any $k, n \in \mathbb{N}$.
3. The continuous prior model

In this section we introduce a hierarchical probability distribution in $L^2(\mathbb{T}) \times L^2(\mathbb{T})$ and prove that it is well-defined. Denote first by $D_q$ a perturbed derivation

\begin{equation}
D_q = D + \epsilon^q P : H^1(\mathbb{T}) \to L^2(\mathbb{T})
\end{equation}

with some $q > 1$ and a projection operator $Pf(x) = (\int_T f(t)dt)\mathbf{1}(x)$ for $f \in L^1(\mathbb{T})$ and $\mathbf{1}(x) = 1$ for every $x \in \mathbb{T}$. The reason for this perturbation is that the operator $D_q : H^1(\mathbb{T}) \to L^2(\mathbb{T})$ is invertible. Also denote $L = D_q^{-1} : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ and a multiplication operator $\Lambda : L^2(\mathbb{T}) \to \mathcal{L}(L^2(\mathbb{T}))$ by

\[\Lambda(v)f = (\epsilon^2 + v^2)^{-1}f\]

for any $v, f \in L^2(\mathbb{T})$. Define operators

\begin{equation}
C_V = \left(\frac{1}{4\epsilon}I - \epsilon\Delta \right)^{-1} \quad \text{and} \quad C_U(v) = L\Lambda(v)L^*
\end{equation}

on $L^2(\mathbb{T})$ with each $v \in L^2(\mathbb{T})$ where $L^*$ is the Hilbert-adjoint of $L$. It is straightforward to show that both operators $(C_U(v)$ with fixed $v)$ are positive self-adjoint trace class operators. This allows us to define the following Gaussian measures on $L^2(\mathbb{T})$ which we use in the construction of the prior probability distribution.

**Definition 5.** Let $\nu$ be the Gaussian measure on $L^2(\mathbb{T})$ centered at value $\mathbf{1}(x) \equiv 1$ with covariance operator $C_V$ and with given $v \in L^2(\mathbb{T})$ let $\nu^v$ be the Gaussian measure on $L^2(\mathbb{T})$ centered at $0$ with covariance operator $C_U(v)$.

**Remark 1.** Now a possible way to proceed is to define a probability measure $\lambda$ on $(L^2(\mathbb{T}) \times L^2(\mathbb{T}), \mathcal{B}(L^2(\mathbb{T}) \times L^2(\mathbb{T})))$ in such a way that with any measurable sets $E,F \subset \mathcal{B}(L^2(\mathbb{T}))$ we have

\begin{equation}
\lambda(E \times F) = \int_F \nu^v(E)dv(v)
\end{equation}

and assign $\lambda$ as a distribution to a random variable $(U,V) : \Omega \to L^2(\mathbb{T}) \times L^2(\mathbb{T})$. In fact, finding a unique extension to $\lambda$ for all Borel sets connects this problem to more general considerations of the existence of Markov chains with given transition operators \cite{20,17,6}. The unique extension can be shown to exist using results related to stochastic kernels \cite{30}. Also, in the framework of $M$-spaces and Markov operators the extension result here can be proved using Lemma 1.3 in \cite{41}.

However, in the rest of the paper the marginal distributions of $\lambda$ play a central role. We achieve more flexible framework especially for the analysis of the discretized distributions by constructing a suitable probability space and defining random variables $U$ and $V$ separately. Consequently, we exclude the extension proof at this stage since later the joint distribution of $(U,V)$ is shown to satisfy equation \eqref{15} as a byproduct of the construction.

**Remark 2.** Throughout the rest of the paper we keep $\epsilon > 0$ and $q > 1$ fixed. The role of $\epsilon$ is to control how sharp edges we will have in the reconstructions.

To simplify our notations we assume that the probability space has the additional structure $\Omega_{pr} = \Omega_1 \times \Omega_2$, $\Sigma_{pr} = \Sigma_1 \otimes \Sigma_2$ and $P_{pr} = P_1 \otimes P_2$.

**Definition 6.** Let $V : \Omega_2 \to L^2(\mathbb{T})$ be a random variable with distribution $\nu$. We note that $V$ has a very similar distribution with the so-called Gaussian smoothness prior. The smoothness prior is well-known to have realizations in $H^s(\mathbb{T})$ almost surely for any $s < 1/2$ and this can similarly be shown to $V$. In fact here the one-dimensional domain allows us to go further with the smoothness. Below the notation $C^{0,\alpha}$ refers to Hölder spaces with exponent $\alpha > 0$ and $W^{t,p}$ denotes the $L^p$-based Sobolev space with exponent $t \in \mathbb{R}$ (see \cite{11}).
Lemma 1. The random variable $V : \Omega_2 \rightarrow L^2(\mathbb{T})$ satisfies following two statements:

(i) For any $t < 1/2$ and $1 < p < \infty$ we have $V \in W^{t,p}(\mathbb{T})$ almost surely,
\[ \mathbb{E} \| V - 1 \|^p_{W^{t,p}} < \infty \]
and there exists a version $V'$ of $V$ such that $V' : \Omega_2 \rightarrow W^{t,p}(\mathbb{T})$ is measurable.

(ii) For any $0 < \alpha < 1/2$ we have $V \in C^{0,\alpha}(\mathbb{T})$ almost surely and
\[ \mathbb{E} \| V - 1 \|^p_{C^{0,\alpha}} < \infty. \]

Proof. Consider the centered variable $V' = V - 1$. By the Schwartz kernel theorem there exists a unique distribution $K_{V'} \in \mathcal{D}'(\mathbb{T} \times \mathbb{T})$ such that $\langle C_{V'}, \phi, \psi \rangle = \langle K_{V'}, \phi \otimes \psi \rangle$. It is straightforward to verify that $K_{V'}$ is the Green function of $\frac{1}{4\epsilon}I - \epsilon \Delta$. Such a function is known to be Lipschitz continuous, i.e., $K_{V'} \in C^{0,1}(\mathbb{T} \times \mathbb{T})$ and even in $C^\infty$ outside the diagonal. Let $t \in [0, \frac{1}{2})$ and define a new kernel $K$ on $\mathbb{T}^2$ as
\[ K(x, y) = (1 - \Delta_x)^{1/2}(1 - \Delta_y)^{1/2} K_{V'}(x, y). \]

Now by [46, Prop. 13.8.3] and [46, Sect. 13, (8.7)], we have $K(x, y) \in C^{0,1-2t}(\mathbb{T} \times \mathbb{T})$ and since $t < \frac{1}{2}$, we have in particular that $K$ is continuous and bounded. By [7, Prop. 3.11.15] we have that for any $1 < p < \infty$ there exists a random variable $V_p$ in $L^p(\mathbb{T})$ with covariance operator $C_p : L^p(\mathbb{T}) \rightarrow L^p(\mathbb{T})$, $\frac{1}{p} + \frac{1}{p'} = 1$, such that
\[ C_p f(x) = \int_{\mathbb{T}} K(x, y) f(y) dy. \]

Furthermore, $V_p$ satisfies
\[ \mathbb{E} \| V_p \|^p_{L^p} < \infty. \]

Due to [46, Prop. 13.8.3] and [46, Sect. 13, (8.7)] we can define for any $1 < p < \infty$ a Gaussian centered random variable $V'_p = (I - \Delta)^{-1/2} V_p$ in $W^{t,p}(\mathbb{T})$ with the property
\[ \mathbb{E} \| V'_p \|^p_{W^{t,p}} < \infty. \]

One notices that the covariance operator of $V'_p$ coincides with $C_{V'}$. The claim (i) follows from the two distributions being the same. Furthermore, the Sobolev embedding theorem states that the space $W^{t,p}(\mathbb{T})$ can be embedded compactly into $C^{0,1-t/p}(\mathbb{T})$ [11]. This proves the claim (ii). \[\square\]

Definition 7. From this moment on in all our analysis we replace $V$ with such a version $V'$ that $V'(\omega_2) \in W^{t_0,p_0}(\mathbb{T})$ for all $\omega \in \Omega$ with some fixed $t_0$ and $p_0$ and $V' : \Omega_2 \rightarrow W^{t_0,p_0}(\mathbb{T})$ is measurable. We keep denoting this new random variable by $V$.

Let $W : \Omega_1 \rightarrow H^s(\mathbb{T})$, $s < -1/2$, be a Gaussian random variable satisfying $\mathbb{E}W = 0$ and
\[ \mathbb{E} \langle \langle W, \phi \rangle_{H^s} \langle W, \psi \rangle_{H^s} \rangle = \langle C_s \phi, \psi \rangle_{H^s} \]
for any $\phi, \psi \in H^s$ where $C_s = (I - \Delta)^s$. The random variable $W$ is white noise in $H^s(\mathbb{T})$ in the sense discussed in Section 2.

In the following the idea is to define $U(\omega_1, \omega_2)$ by operating to $W(\omega_1)$ with a square root of the mapping $C_{U}(V(\omega_2))$. Since $C_{U}(V(\omega_2))$ was defined above on $L^2(\mathbb{T})$ we have to be careful how to define the square root.

Let us begin by defining an unbounded bilinear form $b_v : L^2(\mathbb{T}) \times L^2(\mathbb{T}) \rightarrow \mathbb{R}$,
\[ b_v(\phi, \psi) = \int_{\mathbb{T}} (\epsilon^2 + v^2) D_q \phi \cdot D_q \psi dx \]
for $\phi, \psi \in H^1(\mathbb{T})$ and $v \in C^{0,\alpha}(\mathbb{T})$ with $\alpha > 0$. Due to [31, Thm. VI.1.21, Thm.VI.2.1] there exists a unique linear self-adjoint operator $B_v : \mathcal{D}(B_v) \rightarrow L^2(\mathbb{T})$, $\mathcal{D}(B_v) = \{ \phi \in L^2(\mathbb{T}) \mid (\epsilon^2 + v^2) D_q \phi \in H^1(\mathbb{T}) \}$, such that
\[ b_v(\phi, \psi) = \langle B_v \phi, \psi \rangle \]
Combining inequalities (28) and (21) we get

\[ B_v = D_q^*(\epsilon^2 + v^2)D_q, \]

which is an invertible operator from \( D(B_v) \) to \( L^2(\mathbb{T}) \). The operator \( D_q^* \) denotes the \( L^2 \)-adjoint of \( D_q \). Clearly, \( B_v \) is the inverse of \( C_U(v) \) defined in equation (14) for any \( v \in C^{0,\alpha}(\mathbb{T}) \).

The operator \( B_v \) was constructed in such a way that its spectrum in \( L^2(\mathbb{T}) \) is strictly positive, i.e., \( \sigma(B_v) \subset [c, \infty) \) with \( c = c(\epsilon) > 0 \). Next let us study the mapping properties of \( B_v \) in \( H^1(\mathbb{T}) \). We notice that \( B_v : H^1(\mathbb{T}) \to H^{-1}(\mathbb{T}) \) is an invertible mapping and the pairing \( \langle B_vu, u \rangle_{H^{-1} \times H^1} \) can be estimated with the \( H^1 \)-norm of \( u \) from below. For later purposes choose \( \delta = \delta(\epsilon) > 0 \) such that it satisfies

\[ \langle B_vu, u \rangle_{H^{-1} \times H^1} \geq \delta \| u \|_{H^1}^2 \]

for \( u \in H^1(\mathbb{T}) \). It is important to note that both \( c \) and \( \delta \) are independent of \( v \). As the spectrum of \( B_v \) is positive we can define a square root of \( C_U(v) \) as a Dunford-Taylor integral

\[ \Gamma_v = \frac{1}{2\pi i} \int_{\gamma} z^{-1/2}(B_v - z)^{-1} \, dz : H^{-1}(\mathbb{T}) \to H^{-1}(\mathbb{T}) \]

where \( \gamma \) is the curve

\[ \gamma = \{ z \in \mathbb{C} : \text{dist}(z, \mathbb{R}_-) = \frac{\delta}{2} \} \]

oriented in such a way it turns around the origin in the positive direction. Furthermore, \( z \mapsto z^{-1/2} \) maps \( \mathbb{C} \setminus \mathbb{R}_- \to \mathbb{C} \) so that \( \mathbb{R}_+ \) maps to itself. By [31, Thm. V.3.35] the restriction of \( \Gamma_v \) to \( L^2(\mathbb{T}) \) is an unbounded self-adjoint operator and by [31, Lemma V.3.36] satisfies

\[ (\Gamma_v|_{L^2})^2 = B_v^{-1}|_{L^2} = C_U(v) \]

in \( L^2(\mathbb{T}) \). Next we prove a uniform bound for the norm of \( \Gamma_v \).

**Lemma 2.** There exists a constant \( C = C(s, \delta) \) such that for any \( \alpha > 0 \) and for all \( v \in C^{0,\alpha}(\mathbb{T}) \) we have

\[ \| \Gamma_v \|_{\mathcal{L}(H^s, L^2)} \leq C \]

with \( s > -1 \).

**Proof.** Let \( \alpha > 0 \) and \( v \in C^{0,\alpha}(\mathbb{T}) \). We prove the claim by interpolation arguments. First note that

\[ \| (B_v - z)^{-1} \|_{\mathcal{L}(L^2)} \leq \frac{1}{\text{dist}(z, \sigma(B_v))} \]

for any \( z \in \gamma \). Recall now that \( B_v - z \) with \( z \in \gamma \) is an invertible operator between spaces \( H^1(\mathbb{T}) \) and \( H^{-1}(\mathbb{T}) \). We assume that \( f \in H^{-1}(\mathbb{T}) \) and \( u \in H^1(\mathbb{T}) \) satisfy equation

\[ (B_v - z)u = f \]

in \( H^{-1}(\mathbb{T}) \) for some \( z \in \gamma \). Taking duality pairing of \( f \) with \( u \) in equation (26) yields then

\[ \langle B_vu, u \rangle_{H^{-1} \times H^1} = \| u \|_{L^2}^2 + \langle f, u \rangle_{H^{-1} \times H^1}. \]

For \( z \in \gamma \) we have \( \text{Re}(z) < \delta/2 \) and thus

\[ \langle B_vu, u \rangle_{H^{-1} \times H^1} \leq \frac{\delta}{2} \| u \|_{L^2}^2 + \text{Re} \langle f, u \rangle_{H^{-1} \times H^1}. \]

Combining inequalities (28) and (21) we get

\[ \delta \| u \|_{H^1}^2 \leq \frac{\delta}{2} \| u \|_{H^1}^2 + \| u \|_{H^1} \| f \|_{H^{-1}}. \]

This yields the bound

\[ \| (z - B_v)^{-1} \|_{\mathcal{L}(H^{-1}, H^1)} \leq \frac{2}{\delta} \]
when \( z \in \gamma \). The equation (27) implies
\[
\text{Re}(-z+\delta)\|u\|_{L^2}^2 = -\left(\langle B_vu,u\rangle - \delta \|u\|_{L^2}^2 \right) + \text{Re} \langle f, u \rangle_{H^{-1} \times H^1}
\]
where we have added the term \( \delta \|u\|_{L^2}^2 \) and taken the real part. Again due to inequality (21) the right hand side is less than \( \text{Re} \langle f, u \rangle_{H^{-1} \times H^1} \). Furthermore by applying the Cauchy-Schwarz inequality and inequality (29) we have
\[
\|u\|_{L^2}^2 \leq \frac{1}{\text{Re}(-z+\delta)} \frac{2}{\delta} \|f\|_{H^{-1}}^2
\]
which proves the estimate
\[
\|(z-B_v)^{-1}\|_{\mathcal{L}(H^s,L^2)} \leq |z|^{-1/2}
\]
with \( z \in \gamma \). Now we are ready to interpolate (see, e.g., [46, Prop. 13.6.2], [5] and [47]) equations (25) and (32) and get
\[
\|(z-B_v)^{-1}\|_{\mathcal{L}(H^s,L^2)} \leq \left(|z|^{-1/2}\right)^{-s} \left(\frac{1}{\text{dist}(z,\sigma(B_v))}\right)^{1+s} \leq |z|^{-1-\frac{s}{2}}
\]
for \(-1 \leq s \leq 0\). For \( s > -1 \) and \( z \in \gamma \) we see that
\[
|z|^{-1/2} \|(z-B_v)^{-1}\|_{\mathcal{L}(H^s,L^2)} \leq |z|^{-\frac{3}{2}-\frac{s}{2}}
\]
is an integrable function on \( \gamma \). Finally this yields
\[
\|\Gamma_v\|_{\mathcal{L}(H^s,L^2)} \leq C,
\]
for any \( s > -1 \) with some \( C = C(s,\delta) > 0 \) that is independent of \( v \).

**Definition 8.** Define the mapping \( U : \Omega \rightarrow L^2(\mathbb{T}) \) as
\[
U(\omega_1,\omega_2) = \Gamma_{V(\omega_2)}W(\omega_1)
\]
where \( W \) is the centered Gaussian random variable defined by equation (17) in \( H^s(\mathbb{T}) \) with some \(-1 < s < -1/2\).

Let us show that this mapping is measurable and hence a random variable. Recall that a function \( X : \Omega \rightarrow H \) is said to be strongly measurable if there exists a sequence \( \{X_j\}_{j=1}^\infty \) of simple functions converging pointwise to \( X \). In separable spaces such as \( H^s(\mathbb{T}) \), \( s \geq 0 \), the measurability is equivalent to the strong measurability. In addition, an operator valued function \( X : \Omega \rightarrow \mathcal{L}(H_1,H_2) \) is said to be strongly measurable if the vector valued function \( \omega \mapsto X(\omega)f \) is strongly measurable in \( H_2 \) in the sense presented above for all \( f \in H_1 \).

**Proposition 1.** The mapping \( \omega_2 \mapsto \Gamma_{V(\omega_2)} \in \mathcal{L}(H^s(\mathbb{T}),L^2(\mathbb{T})) \) is strongly measurable for all \(-1 < s < -1/2\).

**Proof.** Recall from Definition 7 that \( V \) is a \( W^{t_0,p_0} \)-valued random variable. As such a space is separable we have a sequence of simple random variables \( V_j \) converging pointwise to \( V \). Due to the Sobolev embedding theorem there exists \( 0 < \alpha < 1/2 \) and \( C > 0 \) such that
\[
\|V_j(\omega_2) - V(\omega_2)\|_{C^0,\alpha} \leq C \|V_j(\omega_2) - V(\omega_2)\|_{W^{t_0,p_0}}
\]
for all \( \omega_2 \in \Omega_2 \). Hence \( V_j \) converges pointwise also in \( C^0,\alpha(\mathbb{T}) \). Next fix \( \omega_2 \in \Omega_2 \) and set \( v_j = V_j(\omega_2) \) for all \( j \in \mathbb{N} \) and \( v = V(\omega_2) \). Let us factorize the operator
\[
(B_v - z)^{-1} - (B_{v_j} - z)^{-1} = (B_v - z)^{-1}(B_{v_j} - B_v)(B_{v_j} - z)^{-1} : H^s(\mathbb{T}) \rightarrow L^2(\mathbb{T})
\]
where the right hand side operators are considered as a sequence of mappings
\[
H^s(\mathbb{T}) \xrightarrow{(B_{v_j} - z)^{-1}} H^1(\mathbb{T}) \xrightarrow{B_{v_j} - B_v} H^{-1}(\mathbb{T}) \xrightarrow{(B_{v} - z)^{-1}} L^2(\mathbb{T}).
\]
An operator and its adjoint have the same norms and, since \( \{ z \mid z \in \gamma \} = \{ \bar{z} \mid z \in \gamma \} \), inequality (33) yields
\[
\left\| (B_v - z)^{-1} \right\|_{L^2(H^1)} = \left\| (B_v - \bar{z})^{-1} \right\|_{L^2(H^{-1}, L^2)} \leq |z|^{-1/2}.
\]
Interpolating inequalities (35) and (29) gives us
\[
\left\| (B_v - z)^{-1} \right\|_{L^2(H^s, H^1)} \leq |z|^{-\frac{1}{2}(1+s)}
\]
for \(-1 < s < -1/2\). In the same way as above we see how the operator \( B_v - B_{v_j} \) maps
\[
H^1(\mathbb{T}) \xrightarrow{D_0} L^2(\mathbb{T}) \xrightarrow{(\omega^2 - \nu^2)1d} L^2(\mathbb{T}) \xrightarrow{D_0'} H^{-1}(\mathbb{T}).
\]
In this framework the operators \( D_0 \) and \( D_0' \) are both bounded. The multiplication operator is also bounded and converges to zero in the norm topology due to (35) as \( j \) increases. Altogether this yields
\[
\lim_{j \to \infty} \left\| B_{v_j} - B_v \right\|_{L^2(H^1, H^{-1})} = 0.
\]
Now returning to random variables \( V_j \) and \( V \) and adding up inequality (37) with (32) we get
\[
\left\| (B_{V(\omega_j)} - z)^{-1} - (B_{V_j(\omega_j)} - z)^{-1} \right\|_{L^2(H^s, L^2)} \leq C \left\| B_{V(\omega_j)} - B_{V_j(\omega_j)} \right\|_{L^2(H^1, H^{-1})}|z|^{-1+\frac{1}{2}}
\]
for all \( \omega_j \in \Omega_2 \) and furthermore
\[
\left\| (\Gamma_{V(\omega_j)} - \Gamma_{V_j(\omega_j)})f \right\|_{L^2} \leq C \int |z|^{-\frac{1}{2}+\frac{1}{2}} \left\| B_{V(\omega_j)} - B_{V_j(\omega_j)} \right\|_{L^2(H^1, H^{-1})} \left\| f \right\|_{H^s} dz
\]
\[
\leq C \left\| B_{V(\omega_j)} - B_{V_j(\omega_j)} \right\|_{L^2(H^1, H^{-1})} \left\| f \right\|_{H^s}
\]
for all \( f \in H^s(\mathbb{T}) \) and \( \omega_j \in \Omega_2 \). Due to equation (38) this proves the claim. \( \square \)

**Corollary 2.** The mapping \( U : \Omega \to L^2(\mathbb{T}) \) in Definition 3 is strongly measurable.

**Proof.** According to the Proposition 1 we can take simple random variables \( \Gamma_{V_j} \) that converge pointwise to \( \Gamma_V \) in \( L^2(H^s(\mathbb{T}), L^2(\mathbb{T})) \) and simple random variables \( W_j \) that converge pointwise to \( W \) in \( H^s(\mathbb{T}) \) with \(-s < -1/2\). Now for any \( \omega = (\omega_1, \omega_2) \in \Omega \) we have that
\[
\left\| \Gamma_{V(\omega_j)}W(\omega_1) - \Gamma_{V_j(\omega_j)}W_j(\omega_1) \right\|_{L^2} \leq \left\| \Gamma_{V(\omega_j)} \right\|_{L^2(H^s, L^2)} \left\| W(\omega_1) - W_j(\omega_1) \right\|_{H^s}
\]
\[
+ \left\| \Gamma_{V(\omega_j)} \right\|_{L^2(H^s, L^2)} \left\| W_j(\omega_1) \right\|_{H^s}
\]
converges to zero for \(-1 < s < -1/2\). \( \square \)

Let us return to the discussion in Remark 1. Also let \(-1 < s < -1/2\) and fix \( \omega_j \in \Omega_2 \) and \( v = V(\omega_j) \). For any \( \phi, \psi \in L^2(\mathbb{T}) \) we have
\[
\mathbb{E}(U(\cdot, \omega_j), \phi)L^2(U(\cdot, \omega_j), \psi)L^2 = \mathbb{E}(W(\cdot), \Gamma_{\nu}^\phi)_{H^s \times H^{-s}}(W(\cdot), \Gamma_{\nu}^\psi)_{H^s \times H^{-s}}
\]
\[
= \mathbb{E}(W(\cdot), C_s\Gamma_{\nu}^\phi)_{H^s}(W(\cdot), C_s\Gamma_{\nu}^\psi)_{H^s}
\]
\[
= \langle C_s C_s C_s \Gamma_{\nu}^\phi, C_s C_s \Gamma_{\nu}^\psi \rangle_{H^s}
\]
\[
= \langle \Gamma_{\nu}^\phi, \psi \rangle_{L^2}
\]
\[
= \langle C_U(v) \phi, \psi \rangle_{L^2}
\]
where \( C_t = (I - \Delta)^t \) for \( t \in \mathbb{R} \). By the Fubini theorem we can deduce that the probability distribution of \((U, V)\) on \( L^2(\mathbb{T}) \times L^2(\mathbb{T}) \) is some extension of \( \lambda \) defined in equation (15).
4. The finite-dimensional prior model

We have two objectives in the construction of a finite-dimensional prior model for the discretized problem. Obviously it is necessary to have weakly converging probability measures. After defining $U_n$ and $V_n$ this property is proved later in Section 5. The second objective is to be able to compute the probability densities explicitly. For anyone applying such a method in practice it is valuable that no additional approximations are needed. The main difficulty in obtaining the explicit form is clearly then the nonlinear dependence of $C_U(V)$ with $V$.

The following definitions can be intuitively considered as truncated random series or projections of the original random variables $U$ and $V$. There is a well-known result [7, Prop. 3.5.1] about Gaussian series which states that Cameron-Martin space provides a natural framework for the basis of the series. Also, as we will see, this approach makes it easier to control the nonlinearity discussed above.

Notice that the Cameron-Martin spaces $H(\nu)$ and $H(\lambda^V(\omega))$ for all fixed $\omega \in \Omega_2$ have equivalent norms with the standard norm of $H^1(\mathbb{T})$. More precisely, the norms satisfy

$$\|\cdot\|_{H^1(\mathbb{T})} = \frac{1}{4\epsilon} \|\cdot\|_{L^2}^2 + \epsilon \|D\cdot\|_{L^2}^2$$

and

$$\|\cdot\|^2_{H^1(\mathbb{T})} = \langle (\epsilon^2 + \nu^2)D\cdot, D\cdot \rangle_{L^2}.$$

This can be shown by density arguments after the equalities are first established for functions in $C^\infty(\mathbb{T})$.

Inspired by this connection we show that the continuous and piecewise linear functions provide a suitable framework for the discretizations. For any $n \in \mathbb{N}$ define

$$PL(n) = \{f \in C(\mathbb{T}) \mid f \text{ is linear on each } K_j^n, j = 1, \ldots, N \} \subset H^1(\mathbb{T})$$

with $K_j^n = [(j - 1)/N, j/N)$, $j = 1, \ldots, N$. The value of $N$ depends on $n$ and for the rest of the paper we fix notation

$$N = N(n) = 2^n.$$

In addition, whenever needed we consider $\mathbb{T}$ as the closed interval $[0, 1]$ with the point 1 identified as 0. Notice that with the notation above $PL(n) \subset PL(n + 1)$ for all $n \in \mathbb{N}$. Define also piecewise constant functions on the same mesh

$$PC(n) = \{f \in L^2(\mathbb{T}) \mid f \text{ is constant on each } K_j^n, j = 1, \ldots, N \} \subset L^2(\mathbb{T}).$$

In the following we use frequently the fact that $D_q|_{PL(n)} : PL(n) \to PC(n)$ is an invertible mapping.

4.1. The definition of $V_n$. Let us consider for a while $H^1(\mathbb{T})$ equipped with the inner product $\langle \cdot, \cdot \rangle_{H^1(\mathbb{T})}$. Form an orthonormal basis $\{g_j\}_{j=1}^\infty$ with respect to this inner product so that for each $n \in \mathbb{N}$ the set $\{g_j\}_{j=1}^N$ spans $PL(n)$. Define an orthogonal projection

$$R_n : H^1(\mathbb{T}) \to PL(n) \subset H^1(\mathbb{T})$$

as

$$R_n g = \sum_{j=1}^N \langle g, g_j \rangle_{H^1(\mathbb{T})} g_j$$

with $g \in H^1(\mathbb{T})$. A short computation yields that the corresponding adjoint operator in $H^{-1}(\mathbb{T})$ is

$$R_n^* g' = \sum_{j=1}^N \langle g', g_j \rangle_{H^{-1}(\mathbb{T}) \times H^1(\mathbb{T})} C^{-1}_V g_j$$

for any $g' \in H^{-1}(\mathbb{T})$.

**Definition 9.** Define $V_n : \Omega_2 \to PL(n) \subset L^2(\mathbb{T})$ as

$$V_n(\omega_2) = \sum_{j=1}^N V_j^N(\omega_2) g_j + 1,$$

where
where $V^N_j : \Omega_2 \rightarrow \mathbb{R}$ are independent random variables with standard normal distribution, $1(x) \equiv 1$ and $g_j \in PL(n)$ are as chosen above. Denote the probability distribution of $V_n$ on $L^2(\mathbb{T})$ by $\nu_n$.

Let us shortly consider the covariance operator of $V_n$ in $L^2(\mathbb{T})$. Clearly, for any $\phi \in L^2(\mathbb{T})$ it holds that

$$C_V R_n^t \phi = R_n C_V \phi.$$  

Furthermore, we have that

$$\langle C_V \phi, \psi \rangle_{L^2} = \mathbb{E}(V_n - 1) \langle \phi \rangle_{L^2} \langle \psi \rangle_{L^2} = \sum_{j,k=1}^N \langle \mathbb{E}V^N_j \mathbb{E}V^N_k \rangle \langle g_j, \phi \rangle_{L^2} \langle g_k, \psi \rangle_{L^2} = \langle \sum_{j=1}^N (g_j, \phi)_{L^2} g_j, \psi \rangle_{L^2} = \langle R_n C_V \phi, \psi \rangle_{L^2}$$

for any $\phi, \psi \in L^2(\mathbb{T})$. Hence we can conclude that

$$C_{V_n} = R_n C_V R_n^t |_{L^2} : L^2(\mathbb{T}) \rightarrow L^2(\mathbb{T})$$

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

4.2. The definition of $U_n$. The discretization method applied to $V$ cannot be used with $U$ since we do not want the corresponding basis to depend on realizations of $V$. To avoid this consider now $H^1(\mathbb{T})$ equipped with the inner product

$$\langle f, g \rangle_{D_q} = \langle D_q f, D_q g \rangle_{L^2}$$

for $f, g \in H^1(\mathbb{T})$. In the same manner as above form an orthonormal basis $\{f_j\}_{j=1}^\infty \subset H^1(\mathbb{T})$ with respect to $\langle \cdot, \cdot \rangle_{D_q}$ so that for each $n \in \mathbb{N}$ the set $\{f_j\}_{j=1}^N$ spans $PL(n)$. Define then an orthogonal projection $S_n : H^1(\mathbb{T}) \rightarrow PL(n) \subset H^1(\mathbb{T})$ as

$$S_n f = \sum_{j=1}^N \langle f, f_j \rangle_{D_q} f_j$$

for any $f \in H^1(\mathbb{T})$. The dual operator $S_n' : H^{-1} \rightarrow H^{-1}$ can then be written

$$S_n' f' = \sum_{j=1}^N \langle f', f_j \rangle_{H^{-1} \times H^1} D_q f_j$$

for any $f' \in H^{-1}(\mathbb{T})$.

The functions $\{D_q f_j\}_{j=1}^\infty \subset L^2(\mathbb{T})$ form by definition an orthonormal basis to $L^2(\mathbb{T})$ with respect to the usual inner product of $L^2(\mathbb{T})$. Denote by $T_n$ the orthogonal projection

$$T_n \phi = \sum_{j=1}^N \langle \phi, D_q f_j \rangle_{L^2} D_q f_j$$

from $L^2(\mathbb{T})$ to $PC(n) \subset L^2(\mathbb{T})$. One notices that

$$D_q S_n D_q^{-1} \phi = D_q \sum_{j=1}^N (D_q^{-1} \phi, f_j) D_q f_j = T_n \phi$$

for any $\phi \in L^2(\mathbb{T})$. The projection $T_n$ is self-adjoint on $L^2(\mathbb{T})$ and hence we also have equality $T_n \phi = (D_q')^{-1} S_n'' D_q' \phi$ for any $\phi \in L^2(\mathbb{T})$. Let us next show an auxiliary lemma about the convergence of the projections $S_n$. 

Lemma 3. For the orthogonal projection $S_n$ defined in equation 13, it holds that
\[ \lim_{n \to \infty} \| I - S_n \|_{\mathcal{L}(H^1, H^1)} = 0 \]
for any $t < \frac{1}{2}$.

Proof. Let $t < 1/2$ and notice that $(D_q D_q')^{t-1}$ is a trace class operator in $L^2(\mathbb{T})$. Since trace is invariant with respect to the basis and norms $\| \cdot \|_{H^1}$ and $\| D_q \|_{L^2}$ are equivalent, we have that
\[ \sum_{j \in \mathbb{N}} \| f_j \|_{H^1}^2 \leq \sum_{j \in \mathbb{N}} \| D_q f_j \|_{L^2}^2 = \sum_{j \in \mathbb{N}} \langle D_q f_j, (D_q D_q')^{t-1} D_q f_j \rangle_{L^2} < \infty \]
since functions $\{D_q f_j\}_{j=1}^\infty$ are an orthonormal basis in $L^2(\mathbb{T})$. Let $\delta > 0$ and choose $N$ so that
\[ \sum_{j > N} \| f_j \|_{H^1}^2 < \delta. \]
Obviously the functions $\{D_q' f_j\}_{j=1}^\infty$ also form an orthonormal basis for $L^2(\mathbb{T})$ and we can write for each $f \in H^1(\mathbb{T})$
\[ \| f - S_n f \|_{H^1}^2 \leq C \sum_{j=1}^\infty \| (D_q' (f - S_n f), D_q' f_j)_{L^2}^2 \]
\[ = C \sum_{j=1}^\infty \| D_q f, (D_q')^{-1} (I - S_n)' D_q' (D_q') f_j \|_{L^2}^2 \]
\[ \leq C \sum_{j=1}^\infty \| D_q f \|_{L^2}^2 \| (I - T_n)' (D_q') f_j \|_{L^2}^2 \]
since $(D_q') f_j \in L^2(\mathbb{T})$. Hence we can estimate the sum as follows
\[ \| f - S_n f \|_{H^1}^2 \leq C \left( \sum_{j=1}^\infty \left( \sum_{k > n} \| (D_q') f_j, D_q f_k \|_{L^2}^2 \right) \right) \| D_q f \|_{L^2}^2 \]
\[ = C \left( \sum_{j=1}^\infty \sum_{k > n} \| (D_q') f_j, D_q f_k \|_{L^2}^2 \right) \| D_q f \|_{L^2}^2 \]
\[ = C \left( \sum_{k > n} \| (D_q')^2 f_k \|_{L^2}^2 \right) \| D_q f \|_{L^2}^2 \]
\[ \leq C \delta \| f \|_{H^1}^2 \]
when $n > N$. \( \square \)

Before defining $U_n$ let us still introduce one more notation. Let $\Lambda_n$ be the multiplication operator
\[ \Lambda_n(v) f = (e^2 + (Q_n v)^2)^{-1} f \]
for any $v \in L^2(\mathbb{T})$ and $f \in L^2(\mathbb{T})$ where
\[ Q_n v = N \sum_{j=1}^N \int_{K_j^N} v(x) dx \cdot 1_{K_j^N} \]
and $1_{K_j^N}$ is the indicator function of the set $K_j^N = [(j - 1)/N, j/N)$. 


Definition 10. Let \( U_n : \Omega \to L^2(\mathbb{T}) \) be the random variable

\[
U_n(\omega_1, \omega_2) = \sum_{j=1}^N U_j^N(\omega_1, \omega_2) f_j
\]

where the random vector \( U^N(\omega) = (U_j^N(\omega))_{j=1}^N \in \mathbb{R}^N \) is given the following structure: Denote by \( \omega_2 \mapsto C(\omega_2) \in \mathbb{R}^{N \times N} \) a random matrix such that

\[
C_{jk}(\omega_2) = \langle \Lambda_n(V_n(\omega_2)) D_q f_j, D_q f_k \rangle_{L^2}. \]

Due to the positive definiteness of \( C \) we can define

\[
U^N(\omega) = C(\omega_2)^{1/2} W_N(\omega_1)
\]

where \( W_N : \Omega_1 \to \mathbb{R}^N \) is centered Gaussian random variable with identity covariance matrix.

The measurability of \( U^N : \Omega \to \mathbb{R}^N \) is a consequence of the mapping \( \omega_2 \mapsto V_n(\omega_2) \) being measurable. Also it follows from Definition 10 that with fixed \( \omega_2 \) the probability distribution of \( \omega_1 \mapsto U_n(\omega_1, \omega_2) \) is centered Gaussian with covariance operator

\[
C_{U_n}(V_n(\omega_2)) = S_n D_q^{-1} \Lambda_n(V_n(\omega_2)) (D_q')^{-1} (S_n)' L^2(\mathbb{T}).
\]

This can be seen from the short computation

\[
\langle C_{U_n}, \phi, \psi \rangle_{L^2} = \sum_{j=1}^N \sum_{k=1}^N C_{jk}(f_j, \phi)_{L^2} (f_k, \psi)_{L^2}
\]

\[
= \left( \Lambda_n(V_n(\omega_2)) \left( \sum_{j=1}^N (f_j, \phi)_{L^2} D_q f_j \right), \sum_{k=1}^N (f_k, \psi)_{L^2} D_q f_k \right)_{L^2}
\]

\[
= \langle \Lambda_n(V_n(\omega_2)) (D_q')^{-1} (S_n)' \phi, (D_q')^{-1} (S_n)' \psi \rangle_{L^2}
\]

for all \( \psi, \phi \in L^2(\mathbb{T}) \). Denote the distribution of \( U_n(\cdot, \omega_2) \) on \( L^2(\mathbb{T}) \) by \( \lambda_n^V(\omega_2) \) and the joint distribution of \( (U_n, V_n) \) on \( L^2(\mathbb{T}) \times L^2(\mathbb{T}) \) by \( \lambda_n \).

4.3. Prior density. Let us show in this subsection how the prior density function of the random variable \((U_n, V_n)\) can be written down explicitly. Consider mappings \( \mathcal{I}_n, \mathcal{J}_n : PL(n) \to \mathbb{R}^N \) such that

\[
\mathcal{I}_n \left( \sum_{j=1}^N x_j f_j \right) = x \quad \text{and} \quad \mathcal{J}_n \left( \sum_{j=1}^N x_j g_j \right) = x.
\]

for any \( x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N \). Use the following notation for the density functions: let \( \Pi_{(U_n, V_n)}, \Pi_{V_n} \) and \( \Pi_{U_n|V_n}(\cdot | \mathcal{J}_n v) \) denote the densities of the probability measures \( \lambda_n \circ (\mathcal{I}_n^{-1}, \mathcal{J}_n^{-1}) \) on \( \mathbb{R}^{2N} \) and \( \nu_n \circ \mathcal{J}_n^{-1} \) and \( \lambda_n^{V_n} \circ \mathcal{I}_n^{-1} \) on \( \mathbb{R}^N \), respectively, with any \( v \in PL(n) \). Below \( \phi \sim \psi \) denotes relation \( \phi \equiv \psi \) with some constant \( c \).

Theorem 2. Let \( v \in PL(n) \) be arbitrary and \( v = \mathcal{J}_n v \in \mathbb{R}^N \). Then

\[
\Pi_{V_N}(v) \propto \exp \left( -\frac{1}{2} \left( \epsilon \| Dv \|^2_{L^2} + \frac{1}{4\epsilon} \| v - 1 \|^2_{L^2} \right) \right)
\]

with \( 1(x) = 1 \) for all \( x \in \mathbb{T} \).

Proof. We recall that by definition \( V_j^N \) are independent standard Gaussian random variables for all \( 1 \leq j \leq N \). It is easy to see that

\[
\| v - \mathcal{J}_n 1\|_{\mathbb{R}^N} = \| \mathcal{J}_n (v - 1) \|_{\mathbb{R}^N} = \| v - 1 \|_{H(\nu)}
\]

since \( \mathcal{J}_n \) is an isometry between \( PL(n) \subset H(\nu) \) and \( \mathbb{R}^N \). By equation (39) we now obtain the claim. \( \square \)
Theorem 3. Let \( u, v \in P \mathcal{L}(n) \) be arbitrary and \( u = I_n u, v = J_n v \in \mathbb{R}^N \). Then it holds that

\[
\Pi_{U_n|V_n}(u \mid v) \propto \exp \left( -\frac{1}{2} \left( \int_T -N \log(e^2 + (Q_n v)^2) + (e^2 + (Q_n v)^2) |D_q u|^2 dx \right) \right).
\]

Proof. The density function of a Gaussian random variable in \( \mathbb{R}^N \) can be written as

\[
\Pi_{U_n|V_n}(u \mid v) \propto \exp \left( -\frac{1}{2}(\log \det C + \langle u, C^{-1} u \rangle_{\mathbb{R}^N}) \right)
\]

where the matrix \( C \) depends on \( v \) and its elements satisfy

\[
C_{jk} = \langle \Lambda_n(v) D_q f_j, D_q f_k \rangle_{L^2}
\]

for \( 1 \leq j, k \leq N \). Our challenge is to compute explicitly \( \det C \) and the inverse matrix \( C^{-1} \).

Notice first how \( \Lambda_n(v) \) maps \( PC(n) \) to itself. Inspired by this let us consider \( C \) as a matrix representation of the linear operator \( \Lambda_n(v) : PC(n) \to PC(n) \) in the basis \( \{ D_q f_j \}_{k=1}^N \).

Next consider another \( L^2 \)-orthonormal basis for \( PC(n) \), namely, \( \{ \sqrt{N} 1_{K_j} \}_{j=1}^N \). Let the matrix \( S \in \mathbb{R}^{N \times N} \) be the matrix presentation of the change of the basis \( \{ D_q f_j \}_{j=1}^N \) to \( \{ \sqrt{N} 1_{K_j} \}_{j=1}^N \). The components of this matrix are given by the formula

\[
S_{jk} = \langle D_q f_j, \sqrt{N} 1_{K_j} \rangle_{L^2}
\]

for \( 1 \leq j, k \leq N \). Moreover, \( S \) is invertible and satisfies \( S^{-1} = S^T \).

Now the key notion is that since \( \Lambda_n(v) \) is diagonal in the basis \( \{ \sqrt{N} 1_{K_j} \}_{j=1}^N \), we can factorize matrix \( C \) as

\[
C = S^{-1} LS
\]

where the diagonal matrix \( L \) is the representation of the multiplication operator \( \Lambda_n(v) \) in the basis \( \{ \sqrt{N} 1_{K_j} \}_{j=1}^N \). One can show that the diagonal of the matrix \( L \) consists of elements \( \langle (e^2 + (Q_n v)^2)^{-1} , N 1_{K_j} \rangle_{L^2} \) for \( 1 \leq j \leq N \). This immediately yields that

\[
\det C = \det L = \prod_{j=1}^N \langle (e^2 + (Q_n v)^2)^{-1} , N 1_{K_j} \rangle_{L^2}.
\]

Now we have

\[
- \log \det C = \sum_{j=1}^N - \log \langle (e^2 + (Q_n v)^2)^{-1} , N 1_{K_j} \rangle_{L^2} = \int_T N \log(e^2 + (Q_n v)^2) dx,
\]

which yields the first part of the density function. Furthermore, a simple computation yields

\[
\langle u, C^{-1} u \rangle_{\mathbb{R}^N} = \langle Su, L^{-1} Su \rangle_{\mathbb{R}^N} = \sum_{j=1}^N \langle (e^2 + (Q_n v)^2)^{-1} , N 1_{K_j} \rangle_{L^2} \langle Su \rangle_j^2.
\]

Assume that \( u = \sum_{k=1}^N u_k f_k \) and \( u = (u_1, ..., u_N)^T \in \mathbb{R}^N \). Then by the equation (46) it holds that

\[
(Su)_j = \sum_{k=1}^N u_k \langle D_q f_k, \sqrt{N} 1_{K_j} \rangle_{L^2} = \langle D_q u, \sqrt{N} 1_{K_j} \rangle_{L^2}
\]

and finally

\[
\langle u, C^{-1} u \rangle_{\mathbb{R}^N} = \sum_{j=1}^N \frac{1}{N} \langle (e^2 + (Q_n v)^2)^{-1} , N 1_{K_j} \rangle_{L^2} \langle D_q u, N 1_{K_j} \rangle_{L^2}^2 = \int_T (e^2 + (Q_n v)^2) |D_q u|^2 dx,
\]

which proves the statement. \( \square \)
We conclude this section by pointing out that
\begin{equation}
\Pi_{(u_n,v_n)}(u,v) = \Pi_{v_n}(v) \cdot \Pi_{u_n|v_n}(u \mid v)
\end{equation}
for any \( u, v \in \mathbb{R}^N \). In consequence, the joint density is obtained from lemmata 2 and 3.

5. CONVERGENCE OF THE CM ESTIMATES

Two previous sections were devoted to the construction of the prior distributions. This is however only halfway in our search for a scalable reconstruction method. In order to show the convergence of conditional mean estimates one also has to consider the interplay between likelihoods, prior distributions and the measurement equation. We turn our attention to this in the following.

5.1. General conditions. Some general conditions under which reconstructors converge were given in [34]. We generalize these conditions in Theorem 1. The essential difference is that the finite-dimensional priors are not given by linear projections. Note that here we consider now a general prior random variable \( U : \Omega \to H \) with a real separable Hilbert space \( H \). Let us first prove a version of the Vitali convergence theorem for probability measures satisfying Definition 4.

Lemma 4. Assume that \( \mu_n \) and \( \mu \) are uniformly discretized probability measures on \( H \). Suppose that \( f : H \to \mathbb{R} \) is continuous and \( 0 \leq f(u) \leq \exp(b \|u\|_H) \) for some constant \( b > 0 \). Then we have the convergence
\begin{equation}
\lim_{n \to \infty} \int_H f(u) d\mu_n(u) = \int_H f(u) d\mu(u).
\end{equation}

Proof. Let us first denote \( B_j = \{ u \in H \mid f(u) > j \} \) and \( f_j(u) := \min(f(u), j) \) for any \( u \in H \). We get an upper bound for the probability of \( B_j \) by
\begin{equation}
\mu(B_j) \leq \frac{1}{j} \int_{B_j} f(u) d\mu(u) \leq \frac{1}{j} \int_H \exp(b \|u\|_H) d\mu(u) \leq \frac{C(b)}{j},
\end{equation}
where \( C(b) \) is given in Definition 4. Notice how the exactly same bound applies also for \( \mu_n(B_j) \). From equation (50) we can deduce
\begin{align*}
\int_H |f - f_j| d\mu &= \int_{B_j} |f - f_j| d\mu \\
&\leq 2 \int_{B_j} \exp(b \|u\|_H) d\mu(u) \\
&\leq 2\sqrt{C(2b)} \sqrt{\mu(B_j)} \\
&\leq \frac{C(b)}{\sqrt{j}}
\end{align*}
where \( \widetilde{C}(b) = 2\sqrt{C(2b)} \sqrt{C(b)} \). Again the same procedure applies for \( \mu_n \) yielding the same upper bound. Notice carefully that the bound does not depend on \( n \). Now the result follows by approximating
\begin{equation}
\left| \int_H f d\mu_n - \int_H f d\mu \right| \leq \int_H |f - f_j| d\mu + \int_H |f - f_j| d\mu_n + \int_H f_j(d\mu - d\mu_n)
\end{equation}
and using the weak convergence. Namely, for each \( \delta > 0 \) we can choose \( j \) so that we have \( \widetilde{C}(b)/\sqrt{j} \leq \delta/3 \). On the other hand for each \( j \) there exists \( n' \) so that \( \left| \int_H f_j(d\mu - d\mu_n) \right| < \delta/3 \) for each \( n > n' \). This results to
\begin{equation}
\left| \int_H f d\mu_n - \int_H f d\mu \right| < \delta
\end{equation}
when \( n > n' \). \( \square \)
Combining Lemma 4 and the formula (8) we can now prove Theorem 1.

Proof of Theorem 1. First, let us consider another measurement model
\begin{equation}
\Theta_{kn} = A_k U_n + \mathcal{E},
\end{equation}
where the noise is not discretized and is now infinite-dimensional. The reconstructor formula can be used for this equation giving
\begin{equation}
\mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) = \frac{\int_H g(u)\bar{\varphi}(u,m_k)d\lambda_n(u)}{\int_H \bar{\varphi}(u,m_k)d\lambda_n(u)}
\end{equation}
with
\begin{equation}
\bar{\varphi}(u,m_k) = \exp(-\frac{1}{2}\|A_k u\|_{L^2}^2 + \langle C^{-1}_E A_k u, m_k \rangle_{H^{-1}}) \leq \exp(b\|u\|_H)
\end{equation}
with some $b > 0$. Now Lemma 4 yields
\begin{equation}
\lim_{k,n \to \infty} \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) = \mathcal{R}_{\mu}(g(U)|m).
\end{equation}
The claim follows from [34] Lemma 1. \hfill \square

5.2. Weak convergence of the prior distribution. The Proposition 3.8.12. in [7] yields the weak convergence of measures $\nu_n$.

Lemma 5. The probability distributions $\nu_n$ converge weakly to $\nu$ on $L^2(\mathbb{T})$.

We want to show that with fixed $\omega_2 \in \Omega_2$ the distribution $\lambda_{V_n}^{\omega_2}$ converges weakly to $\lambda^{\omega_2}$. Since $\lambda_{V_n}^{\omega_2}$ is not obtained with a straight-forward projection as in the case of $\nu_n$, we recall conditions that are needed in the weak convergence of general Gaussian distributions. The following lemma is proved in [7] as Example 3.8.15.

Lemma 6. A sequence of Gaussian measures $\mu_n$ with means $a_n$ and covariance operators $C_n$ on a separable Hilbert space $H$ converges weakly to a Gaussian measure $\mu$ with mean $a$ and covariance operator $C$ if and only if the following conditions are satisfied:

(i) $\lim_{n \to \infty} \|a_n - a\|_H = 0$,
(ii) $\lim_{n \to \infty} \|C_n - C\|_{\mathcal{L}(H)} = 0$ and
(iii) $\lim_{n \to \infty} \text{Tr}_H(C_n) = \text{Tr}_H(C)$.

Let us prove an auxiliary lemma concerning the convergence of the multiplication operators.

Lemma 7. Let $v_n \to v$ in $W^{t_0,p_0}(\mathbb{T})$ as $n \to \infty$. Then we have
\begin{equation}
\lim_{n \to \infty} \|\Lambda(v) - \Lambda_n(v_n)\|_{\mathcal{L}(L^2)} = 0.
\end{equation}

Proof. First notice that for some $\alpha > 0$ we have by the Sobolev embedding theorem that $\|v - v_n\|_{C^0,\alpha} \to 0$. For any continuous $f : \mathbb{T} \to \mathbb{R}$ denote
\begin{equation}
\|f\|_{\infty} = \sup_{x \in \mathbb{T}} |f(x)|.
\end{equation}
Let us then compute an upper bound
\begin{align*}
\left| \frac{1}{\epsilon^2 + v^2} - \frac{1}{\epsilon^2 + (Q_n v_n)^2} \right| &\leq \frac{1}{\epsilon^4} \left( \|Q_n v_n\|^2 - v_n^2 \right) + \left| v_n^2 - v^2 \right| \\
&\leq \frac{1}{\epsilon^4} \left( 2 \|v_n\|_{\infty} |Q_n v_n - v_n| + (\|v_n\|_{\infty} + \|v\|_{\infty}) \|v_n - v\|_{\infty} \right).
\end{align*}
Lemma 8. For all condition (ii) follows from the suitable convergence of the operators weakly to $H^1$. Denote $L$ the first term. Let us study the three terms separately: a dual norm estimation yields an upper bound for $H^1$ we see that Lemma 3 provides the convergence of these terms. The convergence of the third term follows from Lemma 7.

We next consider condition (iii). Recall now the projection $T_n = D_q S_n D_q^{-1} : L^2(\mathbb{T}) \to L^2(\mathbb{T})$. In the following we consider $T_n$ from $L^2(\mathbb{T})$ to $H^s(\mathbb{T})$, $s < 0$, and hence the dual operators occur. Denote $e_j(x) = e^{2\pi j x}$ for all $j \in \mathbb{Z}$ and notice $|D_q^{-1} e_j| \leq \langle j \rangle^{-1}$ where $\langle j \rangle = |j| + 1$. We can then write

$$\langle (C_U - C_{u_n})(v), e_j \rangle_{L^2} = \langle (T_n \Lambda_n(v) - \Lambda(v)) (D_q' - 1) e_j, (D_q^{-1} e_j) \rangle_{L^2}$$

Let us study the three terms separately: a dual norm estimation yields an upper bound for the first term

$$\langle (T_n - I) \Lambda_n(v) (D_q')^{-1} e_j, (D_q^{-1} e_j) \rangle_{L^2} \leq \|T_n - I\|_{L(L^2,H^s)} \|\Lambda_n(v)\|_{L(L^2)} \|T_n\|_{L(L^2)} \|D_q' - 1\|_{L(L^2)} \|\Lambda(v)(D_q')^{-1} e_j\|_{L(L^2)} \|D_q - 1\|_{L(L^2,H^s)} \|e_j\|_{H^s} \leq C \|T_n - I\|_{L(L^2,H^s)} \langle j \rangle^{-2-s}.$$
for any $-1 < s < -1/2$. In the second term we can use Lemma 7 to get

\[
\langle (\Lambda_n(v_n) - \Lambda(v))T_n(D_q')^{-1}e_j, (D_q')^{-1}e_j \rangle_{L^2} \\
\leq \|\Lambda_n(v_n) - \Lambda(v)\|_{\mathcal{L}(L^2)} \|T_n\|_{\mathcal{L}(L^2)} \|(D_q')^{-1}e_j\|_{L^2}^2 \\
\leq o(n)j^{-2},
\]

where $o : \mathbb{N} \rightarrow [0, \infty)$ denotes a function that satisfies $\lim_{n \rightarrow \infty} o(n) = 0$. The third term yields similar upper estimate as the first term since

\[
\langle (\lambda_n(v_n)T_n - I)(D_q')^{-1}e_j, (D_q')^{-1}e_j \rangle_{L^2} \\
\leq \|\lambda_n(v_n)T_n - I\|_{\mathcal{L}(L^2)} \|T_n - I\|_{\mathcal{L}(H^{-s}, L^2)} \|(D_q')^{-1}e_j\|_{L^2} \|(D_q')^{-1}e_j\|_{L^2} \\
\leq C \|T_n - I\|_{\mathcal{L}(L^2, H^s)} j^{-2-s}.
\]

Due to Lemma 3 and the fact that $D_q$ is invertible between $H^1(\mathbb{T})$ and $H^{t-1}(\mathbb{T})$ for any $t \in \mathbb{R}$ we have $\|T_n - I\|_{\mathcal{L}(L^2, H^s)} = o(n)$. Combining these three bounds yields

\[
\langle (T_n\lambda_n(v)T_n' - \Lambda(v))(D_q')^{-1}e_j, (D_q')^{-1}e_j \rangle_{L^2} \leq o(n)j^{-2-s}.
\]

Since $\sum_{j=1}^{\infty} j^{-2-s}$ with $-s > 1$ is finite, we have shown that $\text{Tr}_{L^2}(C_U(v) - C_{U_n}(v))$ converges to zero. This concludes the proof. \qed

Let us recall the Skorohod coupling theorem.

**Theorem 4.** Suppose that a sequence of Borel probability measures $\mu_n$ on a complete separable metric space $B$ converges weakly to a Borel measure $\mu$. Then there exists a probability space $(\Omega, \mathbb{P})$ and measurable mappings $X, X_n : \Omega \rightarrow B$ such that $\mu_n = \mathbb{P} \circ X_n^{-1}$, $\mu = \mathbb{P} \circ X^{-1}$ and $X_n \rightarrow X$ a.s.

At this point we fix $\Omega$ according to Theorem 4 in such a way that $V_n \rightarrow V$ in $W^{t, p_0}(\mathbb{T})$ almost surely. This choice is made to achieve the final result. Before following theorem recall the definition of uniform tightness: A sequence $\{\mu_n\}_{n=1}^{\infty}$ Borel measures on Banach space $X$ is said to be uniformly tight if for every $\delta > 0$ there exists a compact set $K_\delta \subset X$ such that $\mu_n(X \setminus K_\delta) < \delta$ for every $n \in \mathbb{N}$.

**Theorem 5.** When $n$ goes to infinity the random variable $(U_n, V_n)$ converges in distribution to $(U, V)$ in $L^2(\mathbb{T}) \times L^2(\mathbb{T})$.

**Proof.** Let us first show the uniform tightness of the sequence $\{\lambda_n\}_{n=1}^{\infty}$ where $\lambda_n$ is the joint distribution of $(U_n, V_n)$ on $L^2(\mathbb{T}) \times L^2(\mathbb{T})$. The convergence of $V_n$ in distribution yields that probability measures $\{\nu_n\}_{n=1}^{\infty}$ are uniformly tight. Let $\delta > 0$ be given and choose a compact set $K_2 \subset L^2(\mathbb{T})$ in such a way that $\nu_n(K_2) > 1 - \frac{\delta}{2}$. Next we consider the tightness of a family $\{\lambda_n^v \mid v \in K_2, n \in \mathbb{N}\}$. By Lemma 8 the sequence $\{\lambda_n^0\}_{n=1}^{\infty}$ converges weakly and in consequence is uniformly tight. We choose $K_1 \subset L^2(\mathbb{T})$ so that $\lambda_n^0(K_1) > 1 - \frac{\delta}{2}$. We may also assume that $K_1$ is absolutely convex since by Proposition A.1.6 in [7] closed absolutely convex hulls of compact sets are compact. Recall the definition of the covariance $C_U(v) = L\Lambda(v)L^* \lambda_n^v$ in equation (14). For any fixed $v \in L^2(\mathbb{T})$ we know that

\[
\int_{L^2(\mathbb{T})} \langle u, \phi \rangle_{L^2}^2 d\lambda_n^v(u) = \langle \Lambda(v)L^* \phi, L^* \phi \rangle_{L^2} \leq \frac{1}{\epsilon^2} \|L^* \phi\|_{L_2}^2 = \int_{L^2(\mathbb{T})} \langle u, \phi \rangle_{L^2}^2 d\lambda_n^0(u).
\]

for all $\phi \in L^2(\mathbb{T})$. By Theorem 3.3.6 of [7] this yields

\[
1 - \frac{\delta}{2} < \lambda_n^0(K_1) \leq \lambda_n^v(K_1).
\]
Now we are able to deduce the uniform tightness of \( \{ \lambda_n \} \) by setting \( K_\delta = K_1 \times K_2 \), namely,

\[
\begin{align*}
\lambda_n((L^2(\mathbb{T}) \times L^2(\mathbb{T})) \setminus K_\delta) \\
= \lambda_n((L^2(\mathbb{T}) \setminus K_1) \times L^2(\mathbb{T}) \setminus K_2)) \\
\leq \int_{K_2} \lambda_n^u(L^2(\mathbb{T}) \setminus K_1)d\nu_n(v) + \int_{L^2(\mathbb{T}) \setminus K_2} \lambda_n^v(L^2(\mathbb{T}))d\nu_n(v) \\
\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta.
\end{align*}
\]

Moreover, by the Fubini theorem the characteristic function of \((U_n, V_n)\) can be written as

\[
\mathbb{E}\exp(i\langle U_n, \phi \rangle_{L^2} + i\langle V_n, \psi \rangle_{L^2}) = \mathbb{E}\left(\int_{L^2(\mathbb{T})} \exp(i\langle u, \phi \rangle_{L^2})d\lambda_n^\nu(u, \omega_2) \exp(i\langle V_n(\omega_2), \psi \rangle_{L^2})\right).
\]

The almost sure convergence of \( V_n \) and Lemma 8 together imply

\[
\lim_{n \to \infty} \int_{L^2(\mathbb{T})} \exp(i\langle u, \phi \rangle_{L^2})d\lambda_n^\nu(u, \omega_2) = \int_{L^2(\mathbb{T})} \exp(i\langle u, \phi \rangle_{L^2})d\lambda^\nu(u, \omega_2)
\]

and furthermore

\[
\lim_{n \to \infty} \exp(i\langle V_n(\omega_2), \psi \rangle_{L^2}) = \exp(i\langle V(\omega_2), \psi \rangle_{L^2})
\]

for almost every \( \omega_2 \in \Omega_2 \). In consequence, we see by the Lebesgue dominated convergence theorem that the characteristic functions of \((U_n, V_n)\) converge to the characteristic function of \((U, V)\) pointwise.

By Corollary 3.8.5 in [7] the uniform tightness and pointwise converging characteristic functions yield that the random variables \((\hat{U}_n, \hat{V}_n)\) converge in distribution. Since two measures on \( \mathcal{B}(L^2(\mathbb{T}) \times L^2(\mathbb{T})) \) with equal characteristic functionals coincide we conclude that \((U, V)\) is a limit.

**5.3. Uniformly finite exponential moments.** In this section we establish the uniform exponential boundedness of \((U_n, V_n), n \in \mathbb{N}\) and \((U, V)\). Here we denote

\[
\|(f, g)\|_{L^2 \times L^2} := \sqrt{\|f\|_{L^2}^2 + \|g\|_{L^2}^2}
\]

for all \( f, g \in L^2(\mathbb{T}) \).

**Lemma 9.** For every \( b > 0 \) there exists a constant \( C(b) > 0 \) such that

\[
\mathbb{E}\exp(b \|(U_n, V_n)\|_{L^2 \times L^2}) < C(b) \quad \text{and} \quad \mathbb{E}\exp(b \|(U, V)\|_{L^2 \times L^2}) < C(b)
\]

for every \( n \in \mathbb{N} \).

**Proof.** Let us first show the boundedness of the exponential moments of \((U, V)\). By using the inequality \( \|(f, g)\|_{L^2 \times L^2} \leq \|f\|_{L^2} + \|g\|_{L^2} \) and Lemma 2 we have

\[
\begin{align*}
\mathbb{E}\exp(b \|(U, V)\|_{L^2 \times L^2}) & \leq \mathbb{E}\exp(b \|(\Gamma V(\omega_2) W(\omega_1) \|_{L^2} + b \|V(\omega_2)\|_{L^2}) \\
& \leq \mathbb{E}\exp(b \|\Gamma W(\omega_1)\|_{H^s}) \cdot \mathbb{E}\exp(b \|V(\omega_2)\|_{L^2})
\end{align*}
\]

with some constant \( \tilde{b} > 0 \) and some \(-1 < s < -\frac{1}{2}\). Moreover, the Fernique theorem [13, Thm. 2.6] states that for every Gaussian random variable \( X \) in Banach space \((B, B(B))\) there exists a constant \( a > 0 \) such that

\[
\mathbb{E}\exp(a \|X - \mathbb{E}X\|_B^2) < \infty.
\]

Let \( b \in \mathbb{R} \) be arbitrary. The trivial estimate \( 0 \leq a(x - b/2a)^2 \) for any \( x \in \mathbb{R} \) yields

\[
\mathbb{E}\exp(b \|X\|_B) \leq \exp(b \|\mathbb{E}X\|_B) \cdot \exp(b^2/4a) \cdot \mathbb{E}\exp(a \|X - \mathbb{E}X\|_B^2) < \infty.
\]
Now the claim for \((U, V)\) follows by applying inequality \((57)\) to the right-hand side of inequality \((56)\).

The uniform bound for \((U_n, V_n), n \in \mathbb{N}\), requires more careful analysis. Consider for the moment a Gaussian random variable \(X\) in \(L^2(\mathbb{T})\) with covariance operator \(C_X : L^2(\mathbb{T}) \to L^2(\mathbb{T})\) such that \(\text{dim Ran}(C_X) = \ell < \infty\). Let \(\{\rho_j\}_{j=1}^\ell\) be the non-zero eigenvalues and \(\{\phi_j\}_{j=1}^\ell\) be the corresponding \(L^2\)-normalized eigenvectors of \(C_X\). Notice that the normal random variables \((X - \mathbb{E}X, \phi_j)_{L^2}\) and \((X - \mathbb{E}X, \phi_k)_{L^2}\) are independent when \(j \neq k\). For any \(a < 1/(2\rho_j)\), \(1 \leq j \leq \ell\), we have

\[
\mathbb{E}(\exp(a(X - \mathbb{E}X, \phi_j)_{L^2}^2)) = (1 - 2a\rho_j)^{-\frac{1}{2}}.
\]

The operator \(C_X\) is positive definite and hence

\[
\max_{j \in \{1, \ldots, \ell\}} \rho_j \leq \text{Tr}_{L^2}(C_X).
\]

Notice now that \((1 - s)^{-1/2} < 1 + s\) with \(0 < s < 1/2\). In consequence, if \(a\) satisfies

\[
a < \frac{1}{4\text{Tr}_{L^2}(C_X)}
\]

then for every \(j = 1, \ldots, \ell\) it follows that

\[
\mathbb{E}(\exp(a(X - \mathbb{E}X, \phi_j)^2)) \leq 1 + 2a\rho_j \leq \exp(2a\rho_j).
\]

Due to the independence of random variables \((X - \mathbb{E}X, \phi_j)_{L^2}\) and \((58)\) we have

\[
\mathbb{E} \exp(a \|X - \mathbb{E}X\|_{L^2}^2) = \prod_{j=1}^\ell \mathbb{E} \exp(a(X - \mathbb{E}X, \phi_j)_{L^2}^2)
\]

\[
\leq \exp(2a \sum_{j=1}^\ell \rho_j) \leq \exp(2a\text{Tr}_{L^2}(C_X)) < \infty
\]

where we have used the inequality \((58)\). Combining inequalities \((57)\) and \((59)\) in the case \(B = L^2(\mathbb{T})\) yields

\[
\mathbb{E} \exp(b \|X\|_{L^2}) \leq \exp(b \|\mathbb{E}X\|_{L^2}) \cdot \exp(b^2/4a) \cdot \exp(2a\text{Tr}_{L^2}(C_X)).
\]

Let us next show that the trace of \(C_{U_n}(V_n(\omega_2))\) is bounded uniformly with respect to \(n \in \mathbb{N}\) and \(\omega_2 \in \Omega_2\). Denote \(e_j(x) = \exp(-2\pi i j x)\) for \(x \in \mathbb{T}\) and \(j \in \mathbb{Z}\). A straightforward computation yields

\[
\text{Tr}_{L^2}(C_{U_n}(V_n(\omega_2))) = \sum_{j \in \mathbb{Z}} \langle \Lambda_n(V_n(\omega_2)) T_n D_q^{-1} e_j, T_n D_q^{-1} e_j \rangle_{L^2}
\]

\[
\leq \frac{1}{c^2} \sum_{j \in \mathbb{Z}} \|T_n D_q^{-1} e_j\|_{L^2}
\]

\[
\leq \frac{1}{c^2} \sum_{j \in \mathbb{Z}} \|D_q^{-1} e_j\|_{L^2}^2 = C' < \infty
\]

for some constant \(C' < \infty\). Clearly \(C'\) does not depend on \(n\) or \(\omega_2\). With similar arguments we can show that

\[
\text{Tr}_{L^2}(C_{V_n}) \leq C''
\]

where constant \(C''\) does not depend on \(n\). By the Fubini theorem we have

\[
\mathbb{E} \exp(b \|(U_n, V_n)\|_{L^2 \times L^2}) \leq \int_{L^2(\mathbb{T})} \left( \int_{L^2(\mathbb{T})} \exp(b \|u\|_{L^2}) d\lambda_n^u(u) \cdot \exp(b \|v\|_{L^2}) \right) d\nu_n(v)
\]
and finally due to inequalities (60), (61), and (62) we obtain
\[ \mathbb{E} \exp(b \| (U_n, V_n) \|_{L^2 \times L^2}) \leq \exp(b^2/4a) \exp(2aC') \mathbb{E} \exp(b \| V_n(\omega_2) \|_{L^2}) \leq \exp(b^2/2a + b + 2a(C' + C'')) \]
for any \( a < \frac{1}{T} \min(\frac{1}{C'}, \frac{1}{C''}) \). The claim follows by taking the maximum of the bounds on \((U, V)\) and \((U_n, V_n), n \in \mathbb{N}\). \( \square \)

6. Computational example

In this section we illustrate by a numerical example how the method produces reconstructions with similar properties as Ambrosio-Tortorelli minimization \([2, 3]\) in deterministic case. We show how the choice of \( \epsilon \) controls the edge-preserving property of our reconstruction method. Moreover, we compute reconstructions with different choices of \( n \) to convince the reader that the estimates stay stable.

6.1. The model problem. Let us consider a Bayesian deblurring problem \( M = AU + \mathcal{E} \) on \( \mathbb{T} \) where \( A : L^2(\mathbb{T}) \to C^\infty(\mathbb{T}) \) is the operator

\[ Au(x) = \int_{\mathbb{T}} K(x, y)u(y)dy \]
with a priori known smooth kernel \( K \) satisfying \( \int_{\mathbb{T}} K(x', y)dx' = \int_{\mathbb{T}} K(x, y')dy' = 1 \) for all \( x, y \in \mathbb{T} \). Assume also the following two properties:

(i) the noise \( \mathcal{E} \) can be modeled by white noise statistics and
(ii) the measurement projection \( \hat{P}_k : L^2(\mathbb{T}) \to PL(k) \) is proper in the sense of Definition [3]

As we have earlier discussed the assumptions above are related to the measurement situation. Let us then implement the prior distributions and discretization introduced in previous sections. Recall mappings \( I_n, J_n : PL(n) \to \mathbb{R}^N \) with \( N = 2^n \) from Section 3.3. Using Theorems [2] and [3] we see that the posterior density for computational model (2) has the following form: let \( u = I_n(u) \) and \( v = J_n(v) \). Then we have

\[
\pi_{kn}(u, v | m) \propto \exp\left(-\frac{1}{2}F_{\epsilon,k,n}(u, v | m)\right)
\]
where \( u, v \in \mathbb{R}^N, m \in \mathbb{R}^K \) and

\[
F_{\epsilon,k,n}(u, v | m) = \int_{\mathbb{T}} \left( -N \log(e^2 + (Q_nv)^2) + \left( e^2 + (Q_nv)^2 \right) |Du|^2 \\
+ \epsilon |Dv|^2 + \frac{1}{4\epsilon} (1 - v)^2 + |A_{kn}u - m|^2 \right) dx
\]
where \( u, v \in PL(n), m \in PL(k), N = 2^n \) and \( K = 2^k \). Due to equation (12) the computational task is then to evaluate integrals

\[
u_{kn}^{CM} = \int_{\mathbb{R}^N \times \mathbb{R}^N} u \cdot \pi_{kn}(u, v | m) du dv \quad \text{and}
\]

\[
v_{kn}^{CM} = \int_{\mathbb{R}^N \times \mathbb{R}^N} v \cdot \pi_{kn}(u, v | m) du dv.
\]

6.2. Computation of the CM estimates. The integrals in equation (65) are taken over a very large dimensional space and for that reason it is impossible to implement efficiently any quadrature rule. Usually in such situations different types of Markov Chain Monte Carlo (MCMC) methods are used to obtain a solution. In the following let us ease our presentation by denoting

\[
w = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^{2N}.
\]
The idea of MCMC algorithms is to generate a collection \( w^1, \ldots, w^L \in \mathbb{R}^{2N} \) of samples according to the posterior distribution. When \( L \) is large we can approximate the CM estimates in (63) by

\[
\left( u^{CM}_{kn}, v^{CM}_{kn} \right) = w^{CM}_{kn} = \int_{\mathbb{R}^{2N}} w \cdot \pi_{kn}(w | m) \, dw \approx \frac{1}{L - \ell_0} \sum_{\ell = \ell_0 + 1}^{L} w^\ell
\]

where \( \ell_0 \) stands for the number of samples in a burn-in period, i.e., the samples that do not explore the posterior distribution representatively and are discarded.

The algorithm used here for generating the ensemble is an adaptive version of the Metropolis–Hastings (MH) algorithm [24, 38, 21, 22, 44], namely single component adaptive Metropolis (SCAM) algorithm introduced in [23]. The SCAM algorithm is similar to the basic single component Metropolis algorithm in the sense that a sample state, say, \( w^\ell \) is attained by updating the coordinates separately. When deciding the \( j^{th} \) coordinate \( w^\ell_j \) a sample is drawn from the normal distribution \( \mathcal{N}(w^{\ell-1}_j, \sigma^\ell_j) \) centered at the previous point with variance \( \sigma^\ell_j \). The difference is to update variances \( \sigma^\ell_j \) according to the rule

\[
\sigma^\ell_j = \begin{cases} 
\sigma^0_j, & \ell \leq \ell_0, \\
\text{sVar} \left( w^0_j, w^1_j, \ldots, w^{\ell-1}_j \right) + \delta, & \ell > \ell_0.
\end{cases}
\]

Here \( \delta \) denotes the scaling factor for which the value \( s = 2.4 \) (see [23, 18]) is used here. The role of \( \delta \) is to prevent the variance from shrinking to zero and a small constant (\( \delta = 10^{-3} \)) is used as its value. We close this section by showing in pseudo-code how the SCAM algorithm can be implemented.

1. Initialize \( w^0 \in \mathbb{R}^{2N} \) and variances \( (\sigma^0_j)_{j=1}^{2N} \). Set \( \ell := 1 \) and \( j := 1 \).
2. Update \( \sigma^\ell_j \) from formula (67).
3. Sample \( \tau_j \in \mathbb{R} \) from \( \mathcal{N}(0, \sigma^\ell_j) \) and set

\[
w^\text{new} = (w^\ell_1, \ldots, w^\ell_{j-1}, w^\ell_j + \tau_j, w^\ell_{j+1}, \ldots, w^{\ell-1}_{2N})^T \quad \text{and} \quad w^\text{odd} = (w^\ell_1, \ldots, w^\ell_{j-1}, w^\ell_{j+1}, \ldots, w^{\ell-1}_{2N})^T.
\]

4. If

\[
\pi_{kn}(w^\text{new} | m) \geq \pi_{kn}(w^\text{odd} | m),
\]

set \( w^\ell_j := w^{\ell-1}_j + \tau_j \); and go to 6.
5. Draw a random number \( t \) from the uniform distribution on \([0, 1]\). If

\[
t \leq \frac{\pi_{kn}(w^\text{odd} | m)}{\pi_{kn}(w^\text{new} | m)},
\]

set \( w^\ell_j := w^{\ell-1}_j + \tau_j \); else set \( w^\ell_j := w^{\ell-1}_j \).
6. If \( j < 2N \), set \( j := j + 1 \) and go to 2; else if \( j = 2N \) and \( \ell < L \), set \( \ell := \ell + 1 \) and \( j := 1 \) and go to 2; else if \( j = 2N \) and \( \ell = L \) then stop.

6.3. Results. All computations were done using the interval \([0, 1]\) with point 1 identified as 0. Here the parameter for measurement nodes is kept fixed and is chosen to be \( k = 7 \), i.e., we have \( K = 2^k = 128 \) measurement nodes. The number of nodes for the estimates varies between 64 and 256, i.e., \( n \) varies between 6 and 8. See Figure 1 for the exact solution \( u \in L^2(\mathbb{T}) \) and the measured data \( m_k \in PL(k) \). The noise in the measurement was produced from a white noise distribution. Parameters of the MCMC computations are given in Table 1; in each case we take initial values that correspond zero function for \( u \) and \( 1(x) \equiv 1 \) function for \( v \). Both Figures 2 and 3 illustrate how the results look when \( n \) is increased. The difference between the two figures is the choice of \( \epsilon \); in Figure 2 we have chosen \( \epsilon = 10^{-3} \) and in Figure 3 the corresponding value is \( 3 \times 10^{-4} \). Moreover, parameter \( q \) in (13) was chosen large enough in order to get quantity \( e^q \) negligible.
TABLE 1. Parameters of MCMC computations. The number $N$ is the dimension of reconstruction, $\epsilon$ is the prior parameter, $L - \ell_0$ is the number of samples used for computing the CM estimate, $r$ is the total acceptance ratio, i.e., all samples accepted vs. samples tested and the last column indicates the amount of CPU time used for computations.

| $N$  | $\epsilon$ | $L - \ell_0$ | $r$ | Time (h) |
|------|-------------|--------------|-----|----------|
| 64   | $10^{-3}$   | $10^9$       | 0.35| 6.6      |
| 64   | $3 \times 10^{-4}$ | $10^6$ | 0.36| 7.3      |
| 128  | $10^{-3}$   | $2 \times 10^6$ | 0.27| 25.3     |
| 128  | $3 \times 10^{-4}$ | $2 \times 10^6$ | 0.33| 26.9     |
| 256  | $10^{-3}$   | $2 \times 10^6$ | 0.18| 50.6     |
| 256  | $3 \times 10^{-4}$ | $2 \times 10^6$ | 0.25| 53.7     |

We perform all the computations with Matlab 7.6 running in a desktop PC computer with an AMD Opteron 265 dual-dual processor and 8 GB of RAM. Note that the algorithm is not parallelized and thus only one of the processors running at 1.8GHz was in full use at a time.

![Figure 1](image1.png)

**Figure 1.** Left: exact solution $u$, Right: measurement $m_k = M_k(\omega_0)$.

![Figure 2](image2.png)

**Figure 2.** All the plots in this figure are obtained with the choice $\epsilon = 10^{-3}$ and $k = 7$. Top row: the CM estimates $u_{kn}^{CM}$ with $n = 6, 7, 8$ (thick line) and the true signal (thin line) Bottom row: the CM estimates $v_{kn}^{CM}$.

6.4. **Discussion.** We have computed the CM estimates in relatively low dimensions (highest being $N = 256$). This is due to the long computational times of MCMC algorithms. The computational times can be improved with more sophisticated algorithm design, e.g.,
All the plots in this figure are obtained with the choice $\epsilon = 3 \cdot 10^{-4}$ and $k = 7$. Top row: the CM estimates $u_{kn}^{CM}$ with $n = 6, 7, 8$ (thick line) and the true signal (thin line) Bottom row: the CM estimates $v_{kn}^{CM}$.

parallelization. Furthermore, we expect MCMC methods to become much feasible in the future due to evolution of computers.

It is evident from Figures 2 and 3 that the sharpness of edges in the CM estimates can be controlled via $\epsilon$ and the CM estimates $u_{kn}^{CM}$ seem stable with respect to $n$. The results concerning $u_{kn}^{CM}$ fit well to our expectations of the true CM estimate being a slightly smoothened approximation of the real signal represented in Figure 1. Considering the relatively large noise in the measurement we conclude that the method estimates the true signal $u$ well. However, one can notice changes in functions $v_{kn}^{CM}$. First of all, given larger value of $N$ the functions $v_{kn}^{CM}$ become smoother. This phenomena is less visible with smaller values of $\epsilon$ but note that we have not proved what the limiting estimates are exactly. The author expects this phenomena to stabilize with higher values of $N$ but it should be checked in the future studies. Second, given smaller value of $\epsilon$ the maximum of $|v_{kn}^{CM} - 1|$ becomes smaller. Although the asymptotic analysis of taking $\epsilon$ to zero was not considered in this paper we expect that some coupling of $N$ and $\epsilon$ need to be made for algorithm to work properly asymptotically with respect to $\epsilon$. In the deterministic minimization problems of discrete Ambrosio–Tortorelli functionals one typically needs to assume that $N(\epsilon)\epsilon^2 \to \infty$ when $\epsilon$ goes to zero (see e.g. [4]).

We conclude this discussion by pointing out that we have not used any ad-hoc weighting of the prior or likelihood information. This additional flexibility of the algorithm can be achieved by scaling the covariances of $U$ or $V$ with a constant.

Acknowledgements: This work was supported by Emil Aaltonen foundation, Graduate School of Inverse problems (Academy of Finland) and Finnish Centre of Excellence in Inverse Problems Research (Academy of Finland). The author would like to thank Petteri Piiroinen, Hanna Pikkarainen and Samuli Siltanen for various useful discussions and is grateful to the anonymous referees for careful examination of the manuscript.

REFERENCES

[1] R. Adams, "Sobolev spaces". Academic Press, New York, 1975.
[2] L. Ambrosio and V.M. Tortorelli, Approximation of functionals depending on jumps by elliptic functional via $\Gamma$-convergence, Comm. Pure Appl. Math., 43 (1990), 999–1036.
[3] L. Ambrosio and V.M. Tortorelli, On the approximation of free discontinuity problem. Boll. Un. Mat. Ital. B, 6 (1992), 105–123.
[35] M. Lassila, L. Päävääntö, and E. Somersalo, Linear inverse problems for generalised random variables, Inverse Problems, 5 (1989), 599–612.
[36] H. Luschg, Linear estimators and radonifying operators, Teor. Veroyatnost. i Primenen 40, translation in Theory Probab. Appl., 40 (1995), 167–175.
[37] A. Mandelbaum, Linear Estimators and Measurable Linear Transformations on a Hilbert Space, Z. Wahrsch. Verw. Gebiete, 65 (1984), 385–397.
[38] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equations of state calculations by fast computing machine, J. Chem. Phys., 21 (1953), 1087–1091.
[39] D. Mumford and J. Shah, Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems, Commun. Pure Appl. Math., 42 (1989), 577–685.
[40] A. Neubauer and H.K. Pikkarainen, Convergence results for the Bayesian inversion theory, J. Inverse Ill-Posed Prob., 16 2008, 601–613.
[41] P. Piiroinen, Statistical Measurements, Experiments and Applications, Ann. Acad. Sci. Fenn. Math. Diss., 143 (2005).
[42] A. Raj, G. Singh, R. Zabih, B. Kressler, Y. Wang, N. Schuff and M. Weiner, Bayesian parallel imaging with edge-preserving priors, Magn. Reson. Med., 57 (2006), 8–21.
[43] L.I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), 259–268.
[44] E. Saksman and M. Vihola, On the ergodicity of the adaptive Metropolis algorithm on unbounded domains, [arXiv:0806.2933] 2008.
[45] J.A. Sethian, “Level Set Methods and Fast Marching Methods”, Cambridge University Press, Cambridge, 1999.
[46] M.E. Taylor, “Partial differential equations III: Nonlinear equations” Applied Mathematical Sciences 117, Springer-Verlag, New York, 1997.
[47] H. Triebel, “Interpolation theory, function spaces, differential operators”, North-Holland Mathematical Library 18, North-Holland Publishing Co., Amsterdam-New York, 1978.
[48] N. N. Vakhania, Träger des Gausschen Masses im Hilbertraum. Math. Nachr. 64 (1974), 319–322.
[49] N. N. Vakhania, V. I. Tarieladze and S. A. Chobanyan, “Probability distributions on Banach spaces,” Mathematics and its Applications (Soviet Series) 14, D. Reidel Publishing Co., Dordrecht, 1987.

E-mail address: Tapio.Helin@tkk.fi