

DISCRETIZATION-INVARIANT BAYESIAN INVERSION AND BESOV SPACE PRIORS

MATTI LASSAS AND EERO SAKSMAN

Department of Mathematics and Statistics
University of Helsinki
P.O. Box 68 (Gustaf Hallstromin katu 2b) FI-00014, Finland

SAMULI SILTANEN

Institute of Mathematics, Tampere University of Technology
P.O. Box 553, 33101 Tampere, Finland

(Communicated by Erkki Somersalo)

ABSTRACT. Bayesian solution of an inverse problem for indirect measurement $M = AU + \mathcal{E}$ is considered, where U is a function on a domain of \mathbb{R}^d . Here A is a smoothing linear operator and \mathcal{E} is Gaussian white noise. The data is a realization m_k of the random variable $M_k = P_k AU + P_k \mathcal{E}$, where P_k is a linear, finite dimensional operator related to measurement device. To allow computerized inversion, the unknown is discretized as $U_n = T_n U$, where T_n is a finite dimensional projection, leading to the computational measurement model $M_{kn} = P_k A U_n + P_k \mathcal{E}$. Bayes formula gives then the posterior distribution $\pi_{kn}(u_n | m_{kn}) \sim \Pi_n(u_n) \exp(-\frac{1}{2} \|m_{kn} - P_k A u_n\|_2^2)$ in \mathbb{R}^d , and the mean $\mathbf{u}_{kn} := \int u_n \pi_{kn}(u_n | m_k) du_n$ is considered as the reconstruction of U . We discuss a systematic way of choosing prior distributions Π_n for all $n \geq n_0 > 0$ by achieving them as projections of a distribution in a infinite-dimensional limit case. Such choice of prior distributions is *discretization-invariant* in the sense that Π_n represent the same *a priori* information for all n and that the mean \mathbf{u}_{kn} converges to a limit estimate as $k, n \rightarrow \infty$. Gaussian smoothness priors and wavelet-based Besov space priors are shown to be discretization invariant. In particular, Bayesian inversion in dimension two with B_{11}^1 prior is related to penalizing the ℓ^1 norm of the wavelet coefficients of U .

1. Introduction. Consider a quantity U that can be observed indirectly through a relation

$$(1) \quad M = AU + \mathcal{E},$$

where A is a smoothing linear operator and \mathcal{E} is white noise. Here U and M are considered as continuum objects, or functions defined on subsets of \mathbb{R}^d , so that our discussion applies to classical models of mathematical physics such as Laplace, Maxwell, Helmholtz or Schrödinger equation. We are interested in the use of Bayesian inversion to find information about U from measurement data concerning M . Let $U(x, \omega)$, $M(y, \omega)$ and $\mathcal{E}(y, \omega)$ be random functions where $\omega \in \Omega$ is an element of a complete probability space $(\Omega, \Sigma, \mathbb{P})$ and x and y denote the variables

2000 *Mathematics Subject Classification.* 60F17, 65C20, 42C40.

Key words and phrases. Inverse problem, statistical inversion, Bayesian inversion, discretization invariance, reconstruction, wavelet, Besov space.

in domains of Euclidean spaces. We analyze Bayesian estimates of U when a continuum model of the form (1) is approximated by finite-dimensional models to allow computerized inversion.

Assume that a measurement device provides us with a realization of the random variable

$$(2) \quad M_k = P_k M = A_k U + \mathcal{E}_k,$$

where $A_k = P_k A$ and $\mathcal{E}_k = P_k \mathcal{E}$. Here P_k is a linear operator related to the device; for simplicity we assume that P_k is an orthogonal projection with k -dimensional range. We call (2) the *practical measurement model* and (1) the *continuum model*. Realizations of measurements are denoted by $m_k = M_k(\omega_0)$ and $m = M(\omega_0)$, respectively, where $\omega_0 \in \Omega$ is a specific element in the probability space.

This study concentrates on the inverse problem

$$(3) \quad \text{given a realization } M_k(\omega_0), \text{ estimate } U,$$

where the estimates in question are means and confidence intervals related to a Bayesian posterior probability distribution.

For example, consider the brain imaging method called magnetoencephalography (MEG), see e.g. [23]. Maxwell's equations describe how synchronized neuronal currents U in the cerebral cortex produce a magnetic field AU that can be measured at the surface of the head. Let \mathcal{E} denote the magnetic fields produced by all external sources; then the continuum model (1) describes the total magnetic field $M = AU + \mathcal{E}$. In practice one measures the inner products $\langle M, \phi_j \rangle$, $j = 1, 2, \dots, k$, where ϕ_j are linearly independent device functions corresponding to measuring the flux of the magnetic field M through a small surface determined by the j th measurement unit (SQUID). As an idealization, let us assume that ϕ_j are orthonormal so that $P_k v = \sum_{j=1}^k \langle v, \phi_j \rangle \phi_j$ is an orthogonal projection. Then MEG data is modelled by $P_k M$.

Computational solution of (3) using Bayesian inversion involves discretization of the unknown quantity U . We assume that U is *a priori* known to take values in a function space Y . Choose a linear projection $T_n : Y \rightarrow Y$ with n -dimensional range Y_n , and define a random variable $U_n := T_n U$ taking values in Y_n . This leads to the *computational model*

$$(4) \quad M_{kn} = A_k U_n + \mathcal{E}_k$$

involving two *independent* discretizations: P_k is related to the measurement device and T_n to finite representation of the unknown.

In the above application related to MEG, the projection T_n corresponds to an approximate representation of the electromagnetic sources in the brain using a finite set of basis functions (defined for instance according to a finite element method).

Note that the model (4) is virtual in the sense that U_n appears neither in the continuum model (1) nor in the practical measurement model (2). In particular, measurement $M_{kn}(\omega_0)$ is related to the computational model but not to the practical measurement model. This is why we use $m_k = M_k(\omega_0)$ as the given data.

Denote the probability density function of the random variable M_{kn} by $\Upsilon_{kn}(m_{kn})$. The *posterior density* for U_n is given by the Bayes formula:

$$(5) \quad \pi_{kn}(u_n | m_{kn}) = \frac{\Pi_n(u_n) \exp(-\frac{1}{2} \|m_{kn} - A_k u_n\|_2^2)}{\Upsilon_{kn}(m_{kn})},$$

where the exponential function corresponds to (4) with white noise statistics with identity variance, and *a priori* information about U is expressed in the form of a *prior density* Π_n for the random variable U_n . The density Π_n assigns high probability to functions that are typical in light of *a priori* information, and low probability to atypical functions.

We can now state the inverse problem (3) more specifically:

$$(6) \quad \text{given a realization } m_k = M_k(\omega_0), \text{ estimate } U \text{ by } \mathbf{u}_{kn},$$

where the conditional mean (CM) estimate (or posterior mean estimate) \mathbf{u}_{kn} is

$$(7) \quad \mathbf{u}_{kn} := \int_{Y_n} u_n \pi_{kn}(u_n | m_k) du_n.$$

Note that formula (7) differs from the conventional definition of posterior mean estimate since it involves $m_k = M_k(\omega_0)$ instead of $m_{kn} = M_{kn}(\omega_0)$.

The estimate \mathbf{u}_{kn} and confidence intervals for it are typically computed approximately with Markov chain Monte Carlo (MCMC) methods involving simulation software for the finite model (4). The solution strategy (6) has been applied to image restoration [7], geological prospecting [50, 4], atmospheric and ionospheric remote sensing [3, 62, 52, 46, 30], medical X-ray tomography [60, 36, 6] and electrical impedance imaging [31, 5]. For general reference on Bayesian inversion see [33, 48, 27, 63, 13].

We remark that applying the MCMC solution strategy requires also practical implementation of the operator A_k . In the case of MEG imaging A_k corresponds to computing the electromagnetic field $A_k U_n$ using the discrete current U_n and an approximate numerical solution to Maxwell's equations. We neglect the effects of numerical error in the implementation of A_k in this paper. (One possibility to take this error into account is to use the approximation error model of Kaipio and Somersalo [33, Section 5.8], but this leads to non-Gaussian noise statistics in general and thus falls outside the scope of this discussion.)

Summarizing, our starting point is the infinite-dimensional continuum model (1). A measurement instrument provides us with finite-dimensional noisy data $m_k = M_k(\omega_0)$ described by the practical measurement model (2). Our aim is to use m_k to find information about the unknown U . To allow computerized inversion we construct the fully discrete computational model (4) involving *a priori* information about U , and we write down the posterior distribution (5). Finally, we use formula (7) and numerical methods to estimate U with \mathbf{u}_{kn} . However, in definition (7) the data m_k comes from the practical measurement model (2), while π_{kn} is related to the computational model (4). Taking this incompatibility into account is one of the central novelties in this work.

Constructing T_n and Π_n is the core difficulty in Bayesian inversion. Often there is no natural discretization for the continuum quantity U , so n can be freely chosen. Consequently, T_n and Π_n should in principle be described for all $n > 0$, or at least for an infinite sequence of increasing values of n . Also, updating our measurement device may increase k independently of n . This work is motivated by the need to avoid the following unwanted phenomena:

- (a) **The estimates \mathbf{u}_{kn} diverge as $n \rightarrow \infty$.** In this case investing more computational resources to modeling our unknown does not necessarily result in improved reconstructions.

- (b) **The estimates \mathbf{u}_{kn} diverge as $k \rightarrow \infty$.** In this case performing more measurements may lead to worse reconstructions.
- (c) **Representation of *a priori* knowledge is incompatible with discretization.** It is reported in [39] that discrete (non-Gaussian) total variation priors converge to a Gaussian smoothness prior as $n \rightarrow \infty$. In this case one makes the mistake of specifying different *a priori* information for different values of n . See Appendix B below for more details.

A choice of T_n and Π_n is called *discretization-invariant* if it avoids (a)–(c).

We construct prior distributions for U in the infinite-dimensional space Y . Then the random variable $U_n = T_n U$ takes values in the finite-dimensional subspace $Y_n \subset Y$ and represents approximately the same *a priori* knowledge as U . Further, we analyze convergence of \mathbf{u}_{kn} using a deterministic function called *reconstructor* that almost surely maps a given measurement to the conditional mean estimate. For example, the reconstructor $\mathcal{R}_{M_{kn}}(U_n | \cdot)$ corresponding to the computational model (4) takes the measurement data $m_k = M_k(\omega_0)$ to the mean \mathbf{u}_{kn} defined in (7). The infinite-dimensional model $M = AU + \mathcal{E}$ has a reconstructor $\mathcal{R}_M(U | \cdot)$ as well. Theorem 3 below states under suitable assumptions on U, T_n and P_k that

$$(8) \quad \lim_{n, k \rightarrow \infty} \mathcal{R}_{M_{kn}}(U_n | m_k) = \mathcal{R}_M(U | m),$$

where the realization $m_k = P_k m$ comes from a realization $m = M(\omega_0)$ of the random variable M .

Our proving strategy involves another measurement model analogous to (4):

$$(9) \quad \Theta_{kn} = A_k U_n + \mathcal{E},$$

where the noise is not finite-dimensional. The noise models (1) and (9) now contain the same (continuum) white noise process. This allows us to prove convergence results in a fixed function space. Theorem 2 below states under suitable assumptions on U, T_n and P_k that if $\lim_{k \rightarrow \infty} m_k = m$ then

$$\lim_{n, k \rightarrow \infty} \mathcal{R}_{\Theta_{kn}}(U_n | m_k) = \mathcal{R}_M(U | m).$$

Formula (8) follows by showing that the reconstructors coincide: $\mathcal{R}_{\Theta_{kn}}(U_n | m_k) = \mathcal{R}_{M_{kn}}(U_n | m_k)$ for $m_k \in \text{Ran}(P_k)$. We will consider also more general reconstructors $\mathcal{R}_M(g(U) | m)$ that can be used to analyze convergence of confidence intervals.

Our way of using infinite-dimensional limit processes is one possibility to achieve discretization-invariance since we avoid problems (a)–(c).

We are especially interested in discretization-invariant and edge-preserving Bayesian inversion. Total variation regularization of Rudin, Osher and Fatemi [57] penalizes the L^1 norm of derivatives and yields edge-preserving reconstructions in practice [20, 64, 31, 60, 36]. These results are equivalent to computing *maximum a posteriori* estimates using a total variation prior and a Gaussian likelihood. However, Bayesian inversion with discretized total variation prior distribution,

$$(10) \quad \Pi_{U_n}(u) = c_n \exp(-\alpha_n \|\nabla u\|_{L^1(0,1)}), \quad u \in Y_n,$$

where $Y_n \subset W_0^{1,1}(0,1)$ are finite dimensional subspaces, is not discretization-invariant [39]; in particular, conditional mean estimates (i.e. posterior mean estimates) lose their edge-preserving quality as $n \rightarrow \infty$.

Wavelet-based Bayesian inversion using Besov space priors is used in applications with results similar to total variation regularization, see [8, 55, 51, 34]. The Besov space B_{11}^1 that bounds L^1 norms of (band-filtered) first derivatives similarly to (10)

is useful for image processing, see Meyer [47]. One of the main results of this paper is that Besov priors are discretization-invariant; we emphasize that this gives us non-Gaussian discretization-invariant priors. The proof is based on the analysis of the infinite-dimensional limit case and includes a quantitative estimate on the speed of convergence of reconstructors. Further, we show in Section 2.2 for the special case of two-dimensional deblurring that B_{11}^1 inversion using \mathbf{u}_{kn} reduces to applying ℓ^1 -type prior on the wavelet coefficients of U_n —a combination of two well-known computational methods. See Section 2.2 for more details. There is an interesting parallel to algorithms computing MAP estimates with penalty on the ℓ^1 norm of Fourier or wavelet coefficients [21, 61, 17, 42].

Let us review previous literature on the topic. The study of Bayesian inversion in infinite-dimensional function spaces was initiated by Franklin [29] and continued by Mandelbaum [45], Lehtinen, Päivärinta and Somersalo [40], Fitzpatrick [28], and Luschgy [44]. The concept of discretization invariance was formulated by Markku Lehtinen in the 1990's and has been studied by Lasanen [37], Piiroinen [53], D'Ambrogi, Mäenpää and Markkanen [3], and Lasanen and Roininen [38]. A definition of discretization invariance similar to the above was given in [39]. For other kinds of discretization of continuum objects in the Bayesian framework, see [6, 50]. The use of wavelets and Besov spaces in statistical algorithms is discussed in [1, 2, 8, 41, 14, 49]. For regularization based approaches for statistical inverse problems, see [9, 25, 26, 54]. The relationship between continuous and discrete (non-statistical) inversion is studied in Hilbert spaces in [66]. See [11] for specialized discretizations for inverse problems.

We remark that working entirely within the computational model (4) would require using a realization $M_{kn}(\omega_0)$ of the random variable M_{kn} instead of $m_k = M_k(\omega_0)$ in formula (7). The starting point of inversion in earlier studies of discretization invariance is indeed the random variable M_{kn} or its realization. However, the appropriate model of data given by the measurement device is a realization m_k related to model (2). Rigorous analysis of this incompatibility is a central novelty in this paper. To emphasize this aspect we show that inversion from m_k using Gaussian smoothness priors or Besov space priors is discretization-invariant. We do not discuss non-Gaussian noise models in this paper.

This paper is organized as follows. In Section 2 we discuss discretization-invariant Bayesian inversion using Gaussian and Besov priors. In Section 3 we present the general theory of discretization invariance. In Section 4 we discuss random variables with values in Besov spaces. After proving some technical results in Section 5 we apply them in Section 6 to prove convergence of reconstructors arising from Besov priors. In Appendix B we consider examples.

Below $\langle \cdot, \cdot \rangle$ refers to pairing of either generalized functions with test functions, or a Banach space with its dual. We denote by $\langle \cdot, \cdot \rangle_X$ the inner product in a Hilbert space X . The notation $L(X, Y)$ stands for the space of bounded linear operators between Banach spaces X and Y , and $L(X, X)$ is abbreviated as $L(X)$. We occasionally denote the norm $\| \cdot \|_{L(X, Y)}$ by $\| \cdot \|_{X \rightarrow Y}$. The specific element $\omega_0 \in \Omega$ of the probability space denotes realizations of measurements throughout the paper.

Acknowledgements: the authors thank Markku Lehtinen, Heikki Haario, Tapio Helin, and Marko Laine for useful discussions and the anonymous referees whose thoughtful comments helped us to improve the paper. ML and SS were supported by National Technology Agency of Finland (TEKES) under Contract 206/03, Finnish

Centre of Excellence in Inverse Problems Research and PaloDEX Group. ES was supported by the Academy of Finland, projects no. 113826 and 118765, and by the Finnish Center of Excellence in Analysis and Dynamics.

2. Example: Discretization-invariant deblurring. In this section we give examples of discretization-invariant prior distributions and consider a simple inverse problem to give a flavour of our results to the reader. The precise definitions in a more general setting are postponed to the later sections.

We discuss Bayesian deconvolution using a Gaussian smoothness prior and a Besov space prior. In both cases we define the prior distribution in the continuous context and then marginalize it to discrete cases to allow practical computation. The Gaussian case is shown to be related to deterministic Tikhonov regularization with derivative penalty.

For simplicity, we ignore boundary effects by considering periodic functions. The loss of generality is not too bad from the practical point of view since the periodic analysis covers compactly supported non-periodic cases.

Let \mathbb{T}^2 be the two-dimensional torus constructed by identifying parallel sides of the square $D = (0, 1)^2 \subset \mathbb{R}^2$; we model periodic images as elements of function spaces over \mathbb{T}^2 . The continuum model is $M = AU + \mathcal{E}$ with convolution operator A defined by

$$(11) \quad Au(x) = \int_{\mathbb{T}^2} \Phi(x-y)u(y) dy,$$

where $\Phi \in C^\infty(\mathbb{T}^2)$ is a point spread function.

2.1. Gaussian smoothness prior. For any $s \in \mathbb{R}$, let $H^s(\mathbb{T}^2)$ be the L^2 -based Sobolev space equipped with Hilbert space inner product

$$(12) \quad \langle \phi, \psi \rangle_{H^s(\mathbb{T}^2)} = \int_{\mathbb{T}^2} ((I - \Delta)^{s/2} \phi)(x) ((I - \Delta)^{s/2} \psi)(x) dx.$$

Note that $H^0(\mathbb{T}^2) = L^2(\mathbb{T}^2)$.

Recall that a generalized Gaussian random variable V takes values in the space of generalized functions, and the pairing $\langle V, \phi \rangle$ with any test function $\phi \in C^\infty(\mathbb{T}^2)$ is a Gaussian random variable taking values in \mathbb{R} , see [56]. The Gaussian random variables we will consider below are assumed to take values in some Hilbert space, typically in a Sobolev space $H^s(\mathbb{T}^2)$, where the smoothness index $s \in \mathbb{R}$ may also be negative. Now, if V takes values in a Hilbert space X we say that V has the covariance operator $C_V : X \rightarrow X$ if

$$(13) \quad \mathbb{E}(\langle V - \mathbb{E}V, \phi \rangle_X \langle V - \mathbb{E}V, \psi \rangle_X) = \langle C_V \phi, \psi \rangle_X,$$

with any $\phi, \psi \in X$. Here $\langle \cdot, \cdot \rangle_X$ stands for the inner product in X .

Next we analyze a simple measurement model as an example. Consider the continuum model $M = AU + \mathcal{E}$ where the convolution operator (11) is now viewed as a smoothing map $A : H^s(\mathbb{T}^2) \rightarrow C^\infty(\mathbb{T}^2)$.

We construct the smoothness prior by choosing U to be a generalized Gaussian random variable taking values in $H^{-1}(\mathbb{T}^2)$ and having expectation $\mathbb{E}U = 0$ and covariance operator $C_U = \alpha^{-1}(I - \Delta)^{-2}$. Here $\alpha > 0$ is a regularization parameter. The operator C_U corresponds formally to the prior

$$\pi_U(u) \underset{\text{formally}}{=} c \exp\left(-\frac{\alpha}{2} \|u\|_{H^1(\mathbb{T}^2)}^2\right)$$

and generates the discrete smoothness priors widely used in practice. However, it is curious that in spite of the term *smoothness prior* the realizations of U are almost surely not even $L^2(\mathbb{T}^2)$ functions, let alone differentiable. This is why we need to consider U as taking values in some space $H^s(\mathbb{T}^2)$ with negative smoothness index s ; the value $s = -1$ is chosen just for convenience.

White noise \mathcal{E} is a generalized Gaussian random variable with expectation $\mathbb{E} \mathcal{E} = 0$. Let us discuss the choice of an appropriate covariance operator: the standard definition of white noise as a generalized random variable is that

$$(14) \quad \mathbb{E}(\langle \mathcal{E}, \phi \rangle \langle \mathcal{E}, \psi \rangle) = \langle \phi, \psi \rangle, \quad \text{for all } \phi, \psi \in C^\infty(\mathbb{T}^2),$$

where $\langle \cdot, \cdot \rangle$ denotes the pairing between generalized functions and test functions. To consider the white noise \mathcal{E} as a Hilbert-space-valued random variable, we can choose the Hilbert space to be any Sobolev space $H^s(\mathbb{T}^2)$, $s < -1$. One possible choice is to consider \mathcal{E} as taking values in $H^{-2}(\mathbb{T}^2)$ and choose the covariance operator $C_{\mathcal{E}} = (I - \Delta)^{-2} : H^{-2}(\mathbb{T}^2) \rightarrow H^{-2}(\mathbb{T}^2)$ as defined in (13). Note that realizations of \mathcal{E} belong to $L^2(\mathbb{T}^2)$ only with probability zero, and this is why we need to use Sobolev spaces with negative smoothness indices s .

Now the continuous framework for inversion is in place. Let $m = M(\omega_0)$ be a realization of the measurement $M = AU + \mathcal{E}$. Since both the prior and noise statistics are Gaussian, the conditional mean estimate coincides with the location of the maximum of the posterior density. Thus we can evaluate the CM estimate \mathbf{u} as

$$(15) \quad \mathbf{u} = \arg \max_{u \in H^1(\mathbb{T}^2)} \left\{ \exp\left(-\frac{1}{2}\|Au\|_{L^2(\mathbb{T}^2)}^2 + \langle m, Au \rangle - \langle C_U^{-1}u, u \rangle_{H^{-1}(\mathbb{T}^2)}\right) \right\}.$$

We omitted in formula (15) the constant term $\|m\|_{L^2(\mathbb{T}^2)}^2$ in the formal expansion

$$(16) \quad \|m - Au\|_{L^2(\mathbb{T}^2)}^2 = \|Au\|_{L^2(\mathbb{T}^2)}^2 - 2\langle m, Au \rangle_{L^2(\mathbb{T}^2)} + \|m\|_{L^2(\mathbb{T}^2)}^2,$$

which is well-defined only when $m \in L^2(\mathbb{T}^2)$, and this happens with probability zero. Also, the smoothing properties of the operator A make it possible to replace the formal quantity $\langle m, Au \rangle_{L^2(\mathbb{T}^2)}$ in (16) by the rigorously defined pairing $\langle m, Au \rangle$ in (15). Note that we can write (15) in the form

$$\mathbf{u} = \arg \min_{u \in H^1(\mathbb{T}^2)} \left\{ \frac{1}{2}\|Au\|_{L^2(\mathbb{T}^2)}^2 - \langle m, Au \rangle + \frac{\alpha}{2}\|(I - \Delta)^{1/2}u\|_{L^2(\mathbb{T}^2)}^2 \right\}.$$

The practical measurement model is $M_k = P_k AU + P_k \mathcal{E}$, where $P_k : L^2(\mathbb{T}^2) \rightarrow L^2(\mathbb{T}^2)$ is an orthogonal projection with k -dimensional range. We require that the sequence P_k converges strongly to the identity operator on $L^2(\mathbb{T}^2)$ as $k \rightarrow \infty$. For example, P_k may measure averages of AU over k pixels on \mathbb{T}^2 and construct a piecewise function or compute a truncated Fourier series expansion to get an element of $L^2(\mathbb{T}^2)$. Here k can be arbitrarily large, enabling models of imaging devices with any resolution.

Let us now turn to practical inversion where all appearing quantities are finite dimensional. We need to discretize the unknown. Take $T_n : H^{-1}(\mathbb{T}^2) \rightarrow H^{-1}(\mathbb{T}^2)$ to be truncations of Fourier series expansions to n lowest frequency terms. Then T_n are linear orthogonal projections with n -dimensional range Y_n converging strongly to the identity operator on $H^{-1}(\mathbb{T}^2)$ as $n \rightarrow \infty$. Let U be as in the continuum case above and define a random variable $U_n := T_n U$ taking values in Y_n . The conditional mean estimate for U_n (determined using the posterior distribution corresponding

to model $M_{kn} = P_k A U_n + \mathcal{E}_k$) is

$$(17) \quad \mathbf{u}_{kn} := \arg \min_{u \in Y_n \cap H^1(\mathbb{T}^2)} \left\{ \frac{1}{2} \|P_k A T_n u\|_{L^2(\mathbb{T}^2)}^2 - \langle m_k, P_k A T_n u \rangle + \frac{\alpha}{2} \langle (T_n C_U T_n)^{-1} u, u \rangle_{H^{-1}(\mathbb{T}^2)} \right\},$$

where $(T_n C_U T_n)^{-1}$ is the inverse of $T_n C_U T_n : Y_n \rightarrow Y_n$. Here, $m_k = M_k(\omega_0)$ is the realization of measurement M_k in the practical measurement model.

Theorem 2 below implies in particular that $\mathbf{u}_{kn} \rightarrow \mathbf{u}$ as $k, n \rightarrow \infty$, showing that Bayesian deblurring with Gaussian smoothness prior is discretization-invariant.

Much of the material in this Subsection 2.1 is due to Lasanen [37] and Piironen [53]. We emphasize that we assume given a realization $m_k = M_k(\omega_0)$ from practical measurement model (2) and that we do not have available any realization of M_{kn} . Because of this, m_k is used in formula (17). This is the main novelty of our Gaussian example compared to [37].

We close this section by discussing a connection to generalized Tikhonov regularization, here defined as finding the element in $u \in Y_n \cap H^1(\mathbb{T}^2)$ that minimizes the functional

$$\frac{1}{2} \|m_k - P_k A u\|_{L^2(\mathbb{T}^2)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\mathbb{T}^2)}^2 + \frac{\alpha}{2} \|\nabla u\|_{L^2(\mathbb{T}^2)}^2.$$

After similar modification as above we can define the Tikhonov solution by

$$(18) \quad \mathbf{u}_{kn}^T = \arg \min_{u \in Y_n \cap H^1(\mathbb{T}^2)} \left\{ \frac{1}{2} \|P_k A u\|_{L^2(\mathbb{T}^2)}^2 - \langle m_k, P_k A u \rangle + \frac{\alpha}{2} \langle (I - \Delta) u, u \rangle_{L^2(\mathbb{T}^2)} \right\}.$$

The quadratic forms defined in (17) and (18) are the same, so $\mathbf{u}_{kn} = \mathbf{u}_{kn}^T$. Thus our results apply to the convergence analysis of Tikhonov regularization as well.

2.2. Besov prior. Gaussian smoothness priors are designed for representing the prior information that the unknown physical quantity U does not vary sharply. However, sometimes we know *a priori* that U is piecewise regular, and other kind of priors are needed. One could use Gaussian hyperpriors as in [32, 12] or geometric priors as in [50]. We discuss a different approach that allows analysis of an infinite-dimensional limit case and consequently leads to discretization-invariant inversion.

We replace the discretization-dependent [39] total variation prior

$$\Pi(u) \underset{\text{formally}}{=} c \exp(-\alpha \|\nabla u\|_{L^1})$$

by the discretization-invariant Besov space prior

$$\Pi(u) \underset{\text{formally}}{=} c \exp(-\alpha \|u\|_{B_{11}^1(\mathbb{T}^2)}).$$

Since the norm of $B_{11}^1(\mathbb{T}^2)$ bounds the L^1 norms of (band-limited) first derivatives of u , we expect that the B_{11}^1 prior can be used for edge-preserving inversion. It is computationally convenient that B_{11}^1 functions can be written using a compactly supported wavelet bases and the B_{11}^1 norm can be computed as a weighted sum of wavelet coefficients, see Appendix A.

The continuum measurement model is $M = AU + \mathcal{E}$ where the linear convolution operator (11) is considered as a smoothing map $A : B_{11}^1(\mathbb{T}^2) \rightarrow C^\infty(\mathbb{T}^2)$.

We construct the $B_{11}^1(\mathbb{T}^2)$ prior by expanding U in the wavelet basis as the sum

$$U = \sum_{\ell=1}^{\infty} X_{\ell} \psi_{\ell}$$

with each random coefficient X_{ℓ} distributed independently according to $\pi_X(x) = c \exp(-|x|)$, where $c = 2e^{-1}$ is the normalization constant. This distribution arises naturally due to the wavelet characterization of Besov spaces. The scale of the wavelet basis functions becomes finer when ℓ increases; for the exact bookkeeping of scales and locations of ψ_{ℓ} as function of ℓ we refer to Appendix A.

We take $T_n : B_{11}^1(\mathbb{T}^2) \rightarrow B_{11}^1(\mathbb{T}^2)$ to be the finite-dimensional projections

$$(19) \quad T_n \left(\sum_{\ell=1}^{\infty} c_{\ell} \psi_{\ell} \right) = \sum_{\ell=1}^n c_{\ell} \psi_{\ell}$$

that simply truncate the wavelet expansion to n first terms. These operators T_n converge strongly to the identity as $n \rightarrow \infty$. For each n , define a random variable $U_n := T_n U$ taking values in $Y_n := \text{span}(\psi_1, \dots, \psi_n)$ and consider the model $M_{kn} = P_k A T_n U + \mathcal{E}_k$ where $P_k : L^2(\mathbb{T}^2) \rightarrow L^2(\mathbb{T}^2)$ is an orthogonal projection with k -dimensional range.

With the above choices the posterior distribution of U_n takes the computationally efficient form in terms of the wavelet coefficients $X^n := (X_1, X_2, \dots, X_n)$ of U_n , namely, the probability distribution of X^n conditioned with M_{kn} is

$$(20) \quad \pi_{X^n | M_{kn}}(x_1, \dots, x_n, m_k) = C \exp \left(-\frac{1}{2} \left\| m_k - A \sum_{\ell=1}^n x_{\ell} \psi_{\ell} \right\|_{L^2(\mathbb{T}^2)}^2 - \alpha \sum_{\ell=1}^n |x_{\ell}| \right).$$

The conditional mean estimate can be computed approximately e.g. using a Markov chain Monte Carlo algorithm such as the Gibbs sampler or the Metropolis-Hastings method. Sampling from L^1 distributions is explained e.g. in [43, 31] and [33, Section 3.3.2], and modifying those methods for computing the mean of (20) involves only adding the fast wavelet transform appropriately.

Our new results below show that Besov priors are discretization-invariant. First, by Theorems 3 and 5 and Corollary 2 the estimates \mathbf{u}_{kn} converge when $n, k \rightarrow \infty$ either separately or simultaneously. Thus Besov priors do not involve the possible difficulties (a) and (b) mentioned in the introduction. Second, the prior distributions defined in Y_n converge to a limit distribution in Y as $n \rightarrow \infty$ (Proposition 1). Thus we do not have the unwanted phenomenon (c) of the introduction.

3. General theory of discretization invariance. Consider two independent random variables U and \mathcal{E} and the measurement model $M = AU + \mathcal{E}$. We construct a rigorous stochastic framework for discretization-invariant Bayesian inversion using the following diagram of spaces and maps:

$$(21) \quad \begin{array}{ccccccc} Y & \xrightarrow{A} & S^1 & \subset & S^{1/2} = Z & \subset & S \\ \Psi & & & & & & \Psi \\ U(\omega_1) & & & & & & \mathcal{E}(\omega_2) \end{array}$$

Our immediate task is to define all the objects in (21).

Let $(\Omega, \Sigma, \mathbb{P})$ a complete probability space with product structure: $\Omega = \Omega_1 \times \Omega_2$ and $\Sigma = \overline{\Sigma_1 \otimes \Sigma_2}$ and $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$.

Recall the notion of a random variable taking values in a Banach space. For any given separable Banach space X we denote the dual of X by X' . Let \mathcal{B}_X be the Borel σ -algebra of (X, τ^w) with τ^w the weak topology of X . Note that the separability of X verifies that \mathcal{B}_X coincides with the Borel σ -algebra of (X, τ^n) , where τ^n is the norm topology of X . An X -valued random variable V is simply any measurable map $V : (\Omega, \Sigma) \rightarrow (X, \mathcal{B}_X)$. In this paper we consider only random variables with values in separable Banach spaces.

We now assume that the space Y in the diagram (21) is a separable Banach space and that $U = U(\omega_1)$ is a Y -valued random variable.

The measurement noise $\mathcal{E} = \mathcal{E}(\omega_2)$ with $\omega_2 \in \Omega_2$ is a Gaussian random variable taking values in a separable Hilbert space S ; the expectation satisfies $\mathbb{E}\mathcal{E} = 0$, and the covariance operator $C_{\mathcal{E}} : S \rightarrow S$, is defined by requiring

$$\mathbb{E}(\langle s_1, \mathcal{E} \rangle_S \langle \mathcal{E}, s_2 \rangle_S) = \langle s_1, C_{\mathcal{E}} s_2 \rangle_S \quad \text{for } s_1, s_2 \in S.$$

We assume that the essential range of \mathcal{E} is dense in S . Then $C_{\mathcal{E}}$ is one-to-one, self-adjoint and in the trace class, and we may define the unique positive and self-adjoint power $C_{\mathcal{E}}^t$ for any $t \in \mathbb{R}$. For $t \geq 0$ we denote $S^t := C_{\mathcal{E}}^t S$; henceforth the spaces $S^{1/2}$ and S^1 in (21) are well defined.

The space $Z = S^{1/2} = C_{\mathcal{E}}^{1/2} S$ is called the *Cameron-Martin space* of \mathcal{E} , and \mathcal{E} is called the white noise process in Z . We remark that the realizations of \mathcal{E} belong to Z only with probability zero. Note that $C_{\mathcal{E}}^t : S \rightarrow S^t$ is an isomorphism when the norm of the dense subspace $S^t \subset S$ is defined as $\|C_{\mathcal{E}}^t u\|_{S^t} := \|u\|_S$. The domain of definition of $C_{\mathcal{E}}^{-t}$ is S^t , and the map $C_{\mathcal{E}}^{-t} : S^t \rightarrow S$ is an isomorphism. Then

$$\langle z_1, z_2 \rangle_Z = \langle C_{\mathcal{E}}^{-1/2} z_1, C_{\mathcal{E}}^{-1/2} z_2 \rangle_S.$$

In the Gaussian example in Section 2.1 we have $S^t = H^{-2+4t}(\mathbb{T}^2)$ so that $S = S^0 = H^{-2}(\mathbb{T}^2)$ and $Z = L^2(\mathbb{T}^2)$ and we may choose $Y = H^{-1}(\mathbb{T}^2)$. Considering $S^{t+\frac{1}{2}} \subset Z \subset S^{-t+\frac{1}{2}}$ as a Gel'fand triple, where $S^{t+\frac{1}{2}}$ and $S^{-t+\frac{1}{2}}$ are considered as dual spaces with the pairing $\langle w, z \rangle_{S^{t+\frac{1}{2}} \times S^{-t+\frac{1}{2}}} = \langle C_{\mathcal{E}}^{-t} w, C_{\mathcal{E}}^t z \rangle_Z$, it then holds that

$$\mathbb{E}(\langle \mathcal{E}, z_1 \rangle_{S^0 \times S^1} \langle \mathcal{E}, z_2 \rangle_{S^0 \times S^1}) = \langle z_1, z_2 \rangle_Z \quad \text{for } z_1, z_2 \in S^1.$$

Finally, we assume that $A : Y \rightarrow S^1$ is a bounded linear operator. The definition of the objects in diagram (21) is now complete, and we turn to discussing reconstructors.

We analyze finite- and infinite-dimensional Bayesian inversion simultaneously, so let us introduce a rigorous setting for discrete approximations to the random variable U above. We say that Y -valued random variables U_n tend *weakly in distribution* (*w.i.d.*) to U if

$$\lim_{n \rightarrow \infty} \langle U_n, y' \rangle = \langle U, y' \rangle \quad \text{in distribution}$$

for every $y' \in Y'$. Note that here “weakly” refers to the weak topology used in space Y , and “distribution” to the convergence of scalar-valued random variables.

Definition 1. (Linear discretization of random functions) Let Y be a separable Banach space and $Y_n \subset Y$ be finite-dimensional subspaces. The spaces Y_n need not be nested. Let $U_n = U_n(\omega)$ be Y_n -valued random variables with $n = 1, 2, 3, \dots$. Assume that

1. There is a Y -valued random variable $U = U(\omega)$ such that

$$\lim_{n \rightarrow \infty} U_n = U \quad \text{w.i.d.}$$

2. There are bounded linear operators $T_n \in L(Y)$ such that

$$U_n = T_n U.$$

Then we say that the U_n , $n \geq 1$, are *proper linear discretizations* of U in Y .

Examples of random variables that form or do not form proper linear discretizations are given in Appendix B.

In following, we mainly consider cases where T_n are projection operators.

Recall the following definition from [59, II.7].

Definition 2. Let $U \in L^1(\Omega, \Sigma; Y)$ be Y -valued random variable and Σ_0 a sub σ -algebra of Σ . Then the conditional expectation $\mathbb{E}(U|\Sigma_0)$ of U exists with respect to Σ_0 . That is, $\mathbb{E}(U|\Sigma_0) \in L^1(\Omega, \Sigma_0; Y)$ and it satisfies

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

Note that all vector-valued integrals in this work are standard Bochner integrals. We refer the reader to [18] for definition and basic facts on Bochner integral and vector-valued conditional expectations. The operator $P : U \mapsto \mathbb{E}(U|\Sigma_0)$ is a projection $P : L^1(\Omega, \Sigma; Y) \rightarrow L^1(\Omega, \Sigma_0; Y)$, where $L^1(\Omega, \Sigma_0; Y)$ denotes the space of measurable functions $V : (\Omega, \Sigma_0) \rightarrow (Y, \tau^n)$ which are Bochner integrable.

Now we are ready to give the definition of a (non-unique) reconstructor.

Definition 3. Denote by $\mathcal{M} \subset \Sigma$ be the σ -algebra generated by the random variable $M(\omega)$. We say that any deterministic function

$$\mathcal{R}_M(U|\cdot) : S \rightarrow Y, \quad m \mapsto \mathcal{R}_M(U|m),$$

is a *reconstructor* of U with measurement M if

$$\mathcal{R}_M(U|M(\omega)) = \mathbb{E}(U|\mathcal{M})(\omega) \quad \text{almost surely.}$$

If \tilde{Y} is a separable Banach space and $g : Y \rightarrow \tilde{Y}$ is a measurable function, we define $\mathcal{R}_M(g(U)|\cdot) : S \rightarrow \tilde{Y}$ to be any deterministic function satisfying

$$\mathcal{R}_M(g(U)|M(\omega)) = \mathbb{E}(g(U)|\mathcal{M})(\omega) \quad \text{almost surely.}$$

Note that reconstructor is a deterministic function. Also, if a realization of the measurement in the computational model, $M_{kn}(\omega_0)$, is substituted in the reconstructor $\mathcal{R}_M(U|\cdot)$, then the obtained result $\mathcal{R}_M(U|M_{kn}(\omega_0))$ coincides with conditional expectation. Thus, reconstructor, considered as a deterministic function, is just the functional representation of conditional expectation (see e.g. [59], section II.7, formula (43)). However, we assume that we are not given a realization of the measurement in computational model, $M_{kn}(\omega_0)$, as data, but instead $M_k(\omega_0) = P_k M(\omega_0)$, a realization of the measurement in the practical measurement model (2). An essential feature of the reconstructor is that it is defined for all elements in S . Thus, even though the reconstructor is related to the computational model (4), it is possible to substitute into the reconstructor the realization of the practical measurement model (2). This is the reason why the reconstructor, which has the same functional representation as conditional expectation, is defined as a new concept.

One may generalize the well-known scalar-valued result on the existence of reconstructor, see [59, II.3 Theorem 3 and II.7.5], to Bochner-valued conditional expectations. However, we will not need this since we next establish a specific formula for a reconstructor in our situation. Note that the following result is close to the usual functional representation of the conditional expectation, c.f. [59]. As we need to define reconstructors as a deterministic function of the space S and also introduce notations for later use, we present the proof of the result for completeness.

Theorem 1. *Assume that the exponential moments of U satisfy*

$$(22) \quad \mathbb{E}(\exp(\lambda\|U\|_Y)) < \infty \quad \text{for all } \lambda > 0.$$

Denote by $\mu : \mathcal{B}_S \rightarrow [0, 1]$ the distribution of \mathcal{E} in S , and set $\mu_a(E) = \mu(E - a)$ for $a \in S$, $E \in \mathcal{B}_S$. Let $g : Y \rightarrow Y$ be a measurable function satisfying $\mathbb{E}\|g(U)\|_Y < \infty$. Set for $m \in S$

$$(23) \quad \begin{aligned} H_{U,M}^g(m) &= \int_Y g(u) \exp\left(-\frac{1}{2}\|Au\|_Z^2 + \langle C_{\mathcal{E}}^{-1}Au, m \rangle_S\right) d\lambda(u), \\ H_{U,M}^1(m) &= \int_Y \exp\left(-\frac{1}{2}\|Au\|_Z^2 + \langle C_{\mathcal{E}}^{-1}Au, m \rangle_S\right) d\lambda(u), \end{aligned}$$

where λ stands for the distribution of U in Y . Then the function

$$(24) \quad \mathcal{R}_M(g(U)|m) = \frac{H_{U,M}^g(m)}{H_{U,M}^1(m)}, \quad m \in S,$$

is well-defined and satisfies $\mathcal{R}_M(g(U)|M(\omega)) = \mathbb{E}(g(U)|\mathcal{M})(\omega)$ almost surely, that is, formula (24) defines a reconstructor.

Proof. Using the equality $M = AU + \mathcal{E}$, and Fubini theorem, we have for any measurable $D = E \times F \subset S \times Y$

$$\begin{aligned} \mathbb{P}(\{(M, U) \in D\}) &= \int_{\Omega_1} \left[\int_S \chi_E(\varepsilon + AU(\omega_1)) \chi_F(U(\omega_1)) d\mu(\varepsilon) \right] d\mathbb{P}_1(\omega_1) \\ &= \int_{\Omega_1} \left[\int_S \chi_E(m') \chi_F(U(\omega_1)) d\mu_{AU(\omega_1)}(m') \right] d\mathbb{P}_1(\omega_1). \end{aligned}$$

Above, m' is an integration variable running over the space S where M is taking values. Thus we have for any integrable function $f : S \times Y \rightarrow \mathbb{C}$

$$(25) \quad \mathbb{E}(f(M, U)) = \int_{\Omega_1} \left[\int_S f(m', U(\omega_1)) d\mu_{AU(\omega_1)}(m') \right] d\mathbb{P}_1(\omega_1).$$

One checks that the same holds for any Bochner integrable function $f : S \times Y \rightarrow Y$ by simply using the fact that such an f is an almost sure limit of simple functions f_k that satisfy the pointwise inequality $\|f_k\|_Y \leq \|f\|_Y$.

Since Z is the Cameron-Martin space of \mathcal{E} , we have for any $a \in S^1$ by [10, Cor. 2.4.3] the Radon-Nikodym¹ derivative

$$(26) \quad \frac{d\mu_a}{d\mu}(m) = \exp\left(-\frac{1}{2}\|a\|_Z^2 + \langle C_{\mathcal{E}}^{-1}a, m \rangle_S\right).$$

¹To motivate formula (26), we note that if Z would be finite-dimensional, we could write

$$\frac{d\mu_a}{d\mu}(m) = \exp\left(-\frac{1}{2}\|m - a\|_Z^2 + \frac{1}{2}\|m\|_Z^2\right) = \exp\left(-\frac{1}{2}\|a\|_Z^2 + \langle a, m \rangle_Z\right).$$

The formula (26) is a generalization of this for the infinite-dimensional case.

The latter formula has the advantage of being well-defined for every $m \in S$. In particular, we have

$$(27) \quad \frac{d\mu_{AU(\omega_1)}}{d\mu}(m) = \exp\left(-\frac{1}{2}\|AU(\omega_1)\|_Z^2 + \langle C_\varepsilon^{-1}AU(\omega_1), m \rangle_S\right).$$

Using formula (27) we see that formula (23) can be written as

$$\begin{aligned} H_{U,M}^g(m) &= \int_{\Omega_1} g(U(\omega_1)) \frac{d\mu_{AU(\omega_1)}}{d\mu}(m) d\mathbb{P}_1(\omega_1), \\ H_{U,M}^1(m) &= \int_{\Omega_1} \frac{d\mu_{AU(\omega_1)}}{d\mu}(m) d\mathbb{P}_1(\omega_1). \end{aligned}$$

By Fubini theorem we may continue from (25) to obtain

$$\mathbb{E}(f(M, U)) = \int_S \left[\int_{\Omega_1} f(m', U(\omega_1)) \frac{d\mu_{AU(\omega_1)}}{d\mu}(m') d\mathbb{P}_1(\omega_1) \right] d\mu(m').$$

Especially for $E \subset \mathcal{B}_S$ it holds that

$$\mathbb{E}(\chi_E(M)g(U)) = \int_E \left[\int_{\Omega_1} g(U(\omega_1)) \frac{d\mu_{AU(\omega_1)}}{d\mu}(m') d\mathbb{P}_1(\omega_1) \right] d\mu(m').$$

Let $\nu : \mathcal{B}_S \rightarrow [0, 1]$ be the measure $\nu(E) = \mathbb{P}(M^{-1}(E))$, that is, ν is the distribution of M . Now

$$\int_E d\nu(m') = \mathbb{E}(\chi_E(M)) = \int_E \left[\int_{\Omega_1} \frac{d\mu_{AU(\omega_1)}}{d\mu}(m') d\mathbb{P}_1(\omega_1) \right] d\mu(m')$$

and thus $\nu \ll \mu$ and $\frac{d\nu}{d\mu}(m) = H_{U,M}^1(m)$ for almost every m with respect to μ , where

$$H_{U,M}^1(m) = \int_{\Omega_1} \frac{d\mu_{AU(\omega_1)}}{d\mu}(m) d\mathbb{P}_1(\omega_1).$$

By assumption (22), $H_{U,M}^1(m) \in (0, \infty)$ for all $m \in S$. Observe that $H_{U,M}^1(m) > 0$ for all $m \in S$ and thus $\nu(E) = 0$ if and only if $\mu(E) = 0$. Hence also $\mu \ll \nu$ and

$$\frac{d\mu}{d\nu}(m) = \left(\frac{d\nu}{d\mu}(m)\right)^{-1}$$

is well defined. Now by (25)

$$\int_E \int_{\Omega_1} \|g(U(\omega_1))\|_Y \frac{d\mu_{AU(\omega_1)}}{d\mu}(m') \left(\frac{d\nu}{d\mu}(m')\right)^{-1} d\mathbb{P}_1(\omega_1) d\nu(m') \leq \mathbb{E}(\|g(U)\|_Y) < \infty.$$

Hence by Fubini theorem for $E \in \mathcal{B}_S$

$$\begin{aligned} &\mathbb{E}(\chi_E(M)g(U)) \\ &= \int_E \left[\int_{\Omega_1} g(U(\omega_1)) \frac{d\mu_{AU(\omega_1)}}{d\mu}(m') d\mathbb{P}_1(\omega_1) \right] \left(\frac{d\nu}{d\mu}(m')\right)^{-1} d\nu(m') \\ &= \int_{\Omega} \chi_E(M(\omega)) \left[\int_{\Omega_1} g(U(\omega_1)) \frac{d\mu_{AU(\omega_1)}}{d\mu}(M(\omega)) d\mathbb{P}_1(\omega_1) \right] \left(\frac{d\nu}{d\mu}(M(\omega))\right)^{-1} d\mathbb{P}(\omega). \end{aligned}$$

Thus we have the almost sure equality

$$\mathbb{E}(g(U)|\mathcal{M})(\omega) = \left[\int_{\Omega_1} g(U(\omega_1)) \frac{d\mu_{AU(\omega_1)}}{d\mu}(M(\omega)) d\mathbb{P}_1(\omega_1) \right] \left(\frac{d\nu}{d\mu}(M(\omega))\right)^{-1}.$$

This verifies the formula for $\mathcal{R}_M(g(U)|m)$ given in the assertion. \square

For convenience, let us look at Theorem 1 in the finite-dimensional case. Then $Y = \mathbb{R}^n$ and $Z = S = \mathbb{R}^m$, \mathcal{E} is the Gaussian white noise with the identity covariance matrix, and U and M have smooth everywhere positive probability density functions $\pi_U(u)$ and $\pi_M(m)$, correspondingly. It follows that

$$(28) \quad \mathcal{R}_M(U|m) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} u \exp\left(-\frac{1}{2}\|Au - m\|_Z^2\right) \frac{\pi_U(u)}{\pi_M(m)} du$$

satisfies the assumptions of Definition 3. Compare formulas (24) and (28) and see Appendix C for further discussion. Note that (28) is widely used in practical Bayesian inversion [63, 33, 13].

Stability of the reconstructor with respect to data m is important from the point of view of practical inversion. The following theorem yields a non-quantitative convergence result for reconstructors in general. We provide sharper results for the special case of Besov priors later in sections 5 and 6.

Theorem 2. *Assume that U satisfies (22) and let $g : Y \rightarrow \mathbb{R}$ be such a continuous function that*

$$(29) \quad |g(u)| \leq a \exp(a\|u\|_Y) \quad \text{for } u \in Y$$

with some constant a . Assume that $T_n : Y \rightarrow Y$, $n > 0$, and $P_k : S^1 \rightarrow S^1$, $k > 0$, are linear projections satisfying

$$(30) \quad \lim_{n \rightarrow \infty} \|T_n y - y\|_Y = 0 \quad \text{for all } y \in Y,$$

$$(31) \quad \lim_{k \rightarrow \infty} \|P_k z - z\|_{S^1} = 0 \quad \text{for all } z \in \text{Ran}(A),$$

$$(32) \quad \|T_n\|_{L(Y)} \leq C_0, \quad \|P_k\|_{L(S^1)} \leq C_0 \quad \text{for all } n, k$$

with some $C_0 > 0$. Finally, let $m_k, m \in S$ satisfy

$$(33) \quad \lim_{k \rightarrow \infty} m_k = m \quad \text{in } S.$$

Then we have the convergence

$$\lim_{n, k \rightarrow \infty} \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) = \mathcal{R}_M(g(U)|m),$$

where the reconstructors are defined using formula (24) for models (9) and (1), respectively. Moreover, the limits

$$\lim_{n \rightarrow \infty} \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k)$$

exist for a fixed value of k (resp. n).

Proof. We have

$$H_{U_n, \Theta_{kn}}^1(m_k) = \int_{\Omega_1} \frac{d\mu_{P_k A T_n U(\omega_1)}}{d\mu}(m_k) d\mathbb{P}_1(\omega_1)$$

and since (22) and (29) imply $\mathbb{E}|g(T_n U)| < \infty$ we may also write

$$H_{U_n, \Theta_{kn}}^g(m_k) = \int_{\Omega_1} g(T_n U(\omega_1)) \frac{d\mu_{P_k A T_n U(\omega_1)}}{d\mu}(m_k) d\mathbb{P}_1(\omega_1).$$

Above

$$\begin{aligned} \frac{d\mu_{P_k A T_n U(\omega_1)}}{d\mu}(m_k) &= \exp\left(-\frac{1}{2}\|P_k A T_n U(\omega_1)\|_Z^2 + \langle C_\varepsilon^{-1} P_k A T_n U(\omega_1), m_k \rangle_S\right) \\ &\leq \exp(c\|U(\omega_1)\|_Y) \quad \text{for all } k, n, \end{aligned}$$

by our assumptions. The claims follows now from the Lebesgue dominated convergence theorem by applying the majorant $a \exp((c + aC_0)\|U(\omega_1)\|_Y)$. \square

Now the general theory of discretization-invariant Bayesian inversion is in place for the case of measurement models $M = AU + \mathcal{E}$ and $\Theta_{kn} = P_k AT_n U + \mathcal{E}$ concerning infinite-dimensional noise \mathcal{E} . Using the continuum noise \mathcal{E} is convenient above because we can work in the same function space regardless of k .

Assume given data $M_k(\omega_0)$ corresponding to the practical measurement model (2) and consider the computational solution of the inversion problem using the computational model (4), where random error is finite-dimensional white noise $\mathcal{E}_k = P_k \mathcal{E}$. It remains to discuss the implications of our general theory for these practical models. To do this, assume that P_k are projections $P_k : S \rightarrow S$ having the following properties:

$$(34) \quad \text{Ran}(P_k) \text{ is a finite-dimensional subset of } S^1,$$

$$(35) \quad \langle P_k \phi, \psi \rangle_Z = \langle \phi, P_k \psi \rangle_Z \quad \text{for } \phi, \psi \in Z.$$

First we show that reconstructors corresponding to the measurement models $\Theta_{kn} = P_k AU_n + \mathcal{E}$ and $M_{kn} = P_k AU_n + P_k \mathcal{E}$ actually coincide.

Lemma 1. *Assume $P_k : S \rightarrow S$ is a projection satisfying (34) and (35). Then the reconstructors defined in Theorem 1 for the measurement models $\Theta_{kn} = P_k AU_n + \mathcal{E}$ and $M_{kn} = P_k AU_n + P_k \mathcal{E}$ satisfy*

$$\mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) = \mathcal{R}_{M_{kn}}(g(U_n)|m_k)$$

for $m_k \in \text{Ran}(P_k)$.

Proof. Consider first $\mathcal{E}_k = P_k \mathcal{E}$ as a Gaussian random variable taking values in the space $\text{Ran}(P_k)$ that has the inner product inherited from S . The random variable \mathcal{E}_k has zero expectation.

Consider S and S^1 as dual Hilbert spaces. The corresponding pairing is

$$\langle \eta, \phi \rangle_{S \times S^1} = \langle \eta, C_{\mathcal{E}}^{-1} \phi \rangle_S, \quad \eta \in S, \quad \phi \in S^1,$$

which is an extension of the pairing

$$\langle \eta, \phi \rangle_{S \times S^1} = \langle \eta, \phi \rangle_Z$$

defined for $\eta \in Z \subset S$ and $\phi \in S^1$. Moreover, for such η and ϕ

$$\langle P_k \eta, \phi \rangle_{S \times S^1} = \langle P_k \eta, \phi \rangle_Z = \langle \eta, P_k \phi \rangle_Z = \langle \eta, P_k \phi \rangle_{S \times S^1}.$$

As the finite-dimensional projection $P_k : S \rightarrow S$ is bounded, the density of Z in S implies that $\langle P_k \eta, \phi \rangle_{S \times S^1} = \langle \eta, P_k \phi \rangle_{S \times S^1}$ for all $\eta \in S$ and $\phi \in S^1$. Using this, we see for $\phi, \psi \in \text{Ran}(P_k)$

$$\begin{aligned} \mathbb{E}(\langle \mathcal{E}_k, \phi \rangle_Z \langle \mathcal{E}_k, \psi \rangle_Z) &= \mathbb{E}(\langle P_k \mathcal{E}, \phi \rangle_{S \times S^1} \langle P_k \mathcal{E}, \psi \rangle_{S \times S^1}) \\ &= \mathbb{E}(\langle \mathcal{E}, P_k \phi \rangle_{S \times S^1} \langle \mathcal{E}, P_k \psi \rangle_{S \times S^1}) \\ &= \mathbb{E}(\langle \mathcal{E}, C_{\mathcal{E}}^{-1} P_k \phi \rangle_S \langle \mathcal{E}, C_{\mathcal{E}}^{-1} P_k \psi \rangle_S) \\ &= \langle C_{\mathcal{E}}^{-1} \phi, C_{\mathcal{E}} C_{\mathcal{E}}^{-1} \psi \rangle_S \\ &= \langle \phi, \psi \rangle_Z. \end{aligned}$$

This implies that the covariance operator of \mathcal{E}_k , considered now as a Gaussian random variable taking values in $\text{Ran}(P_k)$ endowed with the inner product inherited

from Z , is the identity operator. Using this we see that the reconstructor defined in Theorem 1 for the measurement M_{kn} has the form

$$\begin{aligned}\mathcal{R}_{M_{kn}}(g(U_n)|m_k) &= \frac{\mathbb{E}(g(U) \exp(-\frac{1}{2}\|AU_k - m_k\|_Z^2))}{\mathbb{E}(\exp(-\frac{1}{2}\|AU_k - m_k\|_Z^2))} \\ &= \frac{\mathbb{E}(g(U) \exp(-\frac{1}{2}\|AU_k\|_Z^2 + \langle AU_k, m_k \rangle_Z))}{\mathbb{E}(\exp(-\frac{1}{2}\|AU_k\|_Z^2 + \langle AU_k, m_k \rangle_Z))}\end{aligned}$$

for $m_k \in \text{Ran}(P_k)$. The reconstructor $\mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k)$ defined in Theorem 1 for the measurement Θ_{kn} has the same form, and the assertion follows. \square

Finally, we prove the convergence of reconstructors for models with discrete noise.

Theorem 3. *Assume that in addition to conditions (22)-(32), the projections $P_k : S \rightarrow S$ satisfy (34), (35), together with*

$$(36) \quad \lim_{k \rightarrow \infty} \|P_k z - z\|_S = 0 \quad \text{for all } z \in S.$$

Let $u = U(\omega_0)$, $\varepsilon = \mathcal{E}(\omega_0)$, $\omega_0 \in \Omega$ be realizations of the random variables U and \mathcal{E} , and let

$$m = Au + \varepsilon, \quad m_k = A_k u + P_k \varepsilon,$$

be the realizations of the measurements (1) and (2), respectively. Then the reconstructors defined in Theorem 1 for the measurement models $M_{kn} = P_k A T_n U + P_k \mathcal{E}$ and $M = AU + \mathcal{E}$ satisfy

$$\lim_{n, k \rightarrow \infty} \mathcal{R}_{M_{kn}}(g(U_n)|m_k) = \mathcal{R}_M(g(U)|m).$$

Moreover, the limits

$$\lim_{n \rightarrow \infty} \mathcal{R}_{M_{kn}}(g(U_n)|m_k) \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathcal{R}_{M_{kn}}(g(U_n)|m_k)$$

exist for a fixed value of k (resp. n).

Proof. By (36), $\lim_{k \rightarrow \infty} m_k = m$ in S , and hence the assertion follows by Theorem 2 and Lemma 1. \square

Theorem 3 concerns the convergence of practical inversion methods: m_k is data provided by an actual measurement device, and the computational model $M_{kn} = P_k A T_n U + \mathcal{E}_k$ allows computer implementation. For instance, most Markov chain Monte Carlo inversion algorithms are programmed to evaluate $\mathcal{R}_{M_{kn}}(U_n|m_k)$.

Let $E \subset Y$ be a Borel set and χ_E be the indicator function of E . Using reconstructors, we define

$$\begin{aligned}\mathcal{P}(E|m) &= \mathcal{R}_M(\chi_E(U)|m), \\ \mathcal{P}_{kn}(E|m_k) &= \mathcal{R}_{M_{kn}}(\chi_E(U_n)|m_k).\end{aligned}$$

For a given $m \in S$, the map $E \mapsto \mathcal{P}(E|m)$ is a probability measure on Y by equation (24). Next, let E be a fixed Borel set of Y . If we substitute $M(\omega)$ in the function $m \mapsto \mathcal{P}(E|m)$, we obtain by Definition 3

$$\mathcal{P}(E|M(\omega)) = \mathbb{E}(\chi_E(U)|\mathcal{M})(\omega) = \mathbb{P}(\{U \in E\}|\mathcal{M})(\omega) \quad \text{almost surely,}$$

where $\mathbb{P}(\{U \in E\}|\mathcal{M})(\omega)$ is the conditional probability for the event $\{U \in E\}$ with respect to the σ -algebra \mathcal{M} (see [35, Sec. 6]). Thus, roughly speaking, $m \mapsto \mathcal{P}(E|m)$ can be considered as the posterior probability for the event $U \in E$ when

the measurement M gets value m , and $\mathcal{P}_{kn}(E|m_k)$ the posterior probability for the event $U_n \in E$ when the measurement M_{kn} gets value m_k .

Recall that measures μ_j on Y convergence weakly to measure μ as $j \rightarrow \infty$ if $\lim_{j \rightarrow \infty} \int_Y g d\mu_j = \int_Y g d\mu$ for all bounded continuous functions $g : Y \rightarrow \mathbb{R}$. Thus Theorem 3 yields the following corollary.

Corollary 1. *Let the assumptions of Theorem 3 hold. Then the Borel measures $E \mapsto \mathcal{P}_{kn}(E|m_k)$ converge weakly to the measure $E \mapsto \mathcal{P}(E|m)$ as $n, k \rightarrow \infty$.*

4. Random variables in Besov spaces. We wish to use priors of the form

$$\Pi(u) \underset{\text{formally}}{=} c \exp(-\|u\|_{B_{pp}^s(\mathbb{T}^d)}^p)$$

for any integrability parameter $1 \leq p < \infty$ and some chosen smoothness $s \in \mathbb{R}$. Note that we consider now functions on a general d -dimensional torus \mathbb{T}^d .

Following Appendix A we use a compactly supported wavelet $\tilde{\psi}(x)$, $x \in \mathbb{R}^d$ and scaling function $\tilde{\phi}(x)$ suitable for multi-resolution analysis of smoothness C^r in $L^2(\mathbb{R}^d)$ with large enough r . Then we can expand functions as

$$f(x) = \sum_{\ell=1}^{\infty} c_{\ell} \psi_{\ell}(x), \quad x \in \mathbb{T}^d$$

and $f \in B_{pp}^s(\mathbb{T}^d)$ if the norm [47]

$$(37) \quad \left\| \sum_{\ell=1}^{\infty} c_{\ell} \psi_{\ell}(x) \right\|_{B_{pp}^s(\mathbb{T}^d)} := \left(\sum_{\ell=1}^{\infty} \ell^{(ps/d+p/2-1)} |c_{\ell}|^p \right)^{1/p}$$

is finite. For the exact bookkeeping of scales and locations of ψ_{ℓ} as function of ℓ we refer to Appendix A.

Definition 4. Let $1 \leq p < \infty$ and $s \in \mathbb{R}$. Let $(X_{\ell})_{\ell=1}^{\infty}$ be independent identically distributed real-valued random variables with probability density function

$$(38) \quad \pi_X(x) = c_p \exp(-|x|^p), \quad \text{with } c_p = \left(\int_{\mathbb{R}} \exp(-|x|^p) dx \right)^{-1}.$$

Let U be the random function

$$U(x) = \sum_{\ell=1}^{\infty} \ell^{-(s/d+1/2-1/p)} X_{\ell} \psi_{\ell}(x), \quad x \in \mathbb{T}^d.$$

Then we say that U is distributed according to a B_{pp}^s prior.

Next we show that random variable U in Definition 4 is a well defined object.

Lemma 2. *Let U be as in Definition 4 and take $t \in \mathbb{R}$. The following three conditions are equivalent:*

- (i) $\|U\|_{B_{pp}^t} < \infty$ almost surely.
- (ii) $\mathbb{E} \exp\left(\frac{1}{2}\|U\|_{B_{pp}^t}^p\right) < \infty$.
- (iii) $t < s - \frac{d}{p}$.

Proof. Denote by $(X_\ell)_{\ell=1}^\infty$ a sequence of independent random variables with density (38). Assume (iii). In order to deduce (ii) we need to show the finiteness of the quantity

$$\begin{aligned}
 \mathbb{E} \exp\left(\frac{1}{2} \left\| \sum_{\ell=1}^{\infty} \ell^{-s/d-(1/2-1/p)} X_\ell \psi_\ell \right\|_{B_{pp}^t}^p\right) &\simeq \mathbb{E} \exp\left(\sum_{\ell=1}^{\infty} \frac{1}{2} \ell^{(t-s)p/d} |X_\ell|^p\right) \\
 &= \prod_{\ell=1}^{\infty} \mathbb{E} \exp\left(\frac{1}{2} \ell^{-(s-t)p/d} |X_\ell|^p\right) \\
 (39) \qquad \qquad \qquad &= \prod_{\ell=1}^{\infty} (1 - \ell^{-(s-t)p/d} / 2)^{-1/p}.
 \end{aligned}$$

Above we used independence and the observation that for $k \in (0, 1)$ one may compute $\mathbb{E} \exp(k|X_\ell|^p) = (1 - k)^{-1/p}$. Clearly the product in (39) converges if $t < s - d/p$. The notation $a \simeq b$ stands for the existence of a positive constant $c < \infty$ such that $a/c \leq b \leq ca$.

Observe next that obviously (ii) implies (i). Finally, assume that (i) is true. Then

$$\sum_{\ell=1}^{\infty} \ell^{(t-s)p/d} |X_\ell|^p < \infty$$

almost surely. Since the random variables $|X_\ell|^p$ are non-negative and identically distributed, an easy application of truncation and [35, Thm. 4.17] shows that almost sure finiteness of the sum implies finiteness of the expectation. Hence (i) implies (iii). \square

Now we easily see that $B_{pp}^s(\mathbb{T}^d)$ distributions generate discretization-invariant priors.

Proposition 1. *Let U be distributed according to B_{pp}^s prior as in Definition 4, and let $t < s - \frac{d}{p}$. Take $T_n : Y \rightarrow Y$ to be the sequence of linear operators defined by*

$$T_n \left(\sum_{\ell=1}^{\infty} c_\ell \psi_\ell \right) = \sum_{\ell=1}^n c_\ell \psi_\ell;$$

then $\lim_{n \rightarrow \infty} T_n y = y$ for all $y \in Y := B_{pp}^t(\mathbb{T}^d)$. Then $U_n = T_n U$, $n = 1, 2, \dots$ are proper linear discretizations of U in the sense of Definition 1.

Proof. The random variables $U_n = T_n U$ converge almost surely in norm topology of $B_{pp}^t(\mathbb{T}^d)$. Since almost sure convergence implies convergence weakly in distribution, the assertion follows. \square

Note that if above $t > \frac{d}{p}$, then the continuous embedding $B_{pp}^t(\mathbb{T}^d) \rightarrow C(\mathbb{T}^d)$ implies that realizations of U and U_n are almost surely continuous.

5. Quantitative estimates for reconstructors. This section studies the case that U is distributed according to B_{pp}^s -prior. We provide quantitative stability estimates for reconstructors. For $p > 1$ qualitative results are described by Theorem 2, thanks to Lemma 2(ii). However, the case $p = 1$ is more difficult since condition (22) fails.

We collect a set of assumptions together for later reference:

Assumption A. Let $s \in \mathbb{R}$ and $1 \leq p < \infty$ be arbitrary. Fix $t, \tilde{t}, r \in \mathbb{R}$ such that

$$t < \tilde{t} < s - d/p \quad \text{and} \quad r > d/2.$$

Let A be a linear operator satisfying

$$(40) \quad A : B_{pp}^t(\mathbb{T}^d) \rightarrow B_{11}^r(\mathbb{T}^d).$$

Assume that $g : B_{pp}^t \rightarrow B_{pp}^t$ is a map such that for some $q < \infty$ we have

$$\begin{aligned} \|g(u)\|_{B_{pp}^t} &\leq c(1 + \|u\|_{B_{pp}^t})^q \quad \text{and} \\ \|g(u_1) - g(u_2)\|_{B_{pp}^t} &\leq c\|u_1 - u_2\|_{B_{pp}^t} (1 + \max(\|u_1\|_{B_{pp}^t}, \|u_2\|_{B_{pp}^t}))^q. \end{aligned}$$

Finally, let $G : B_{pp}^{\tilde{t}} \rightarrow B_{pp}^t$ be a bounded operator.

How does the diagram (21) look for the B_{pp}^s -prior? Take $t < s - \frac{d}{p}$ and set

$$\begin{array}{ccc} Y = B_{pp}^t(\mathbb{T}^d) & \xrightarrow{A} & B_{11}^r(\mathbb{T}^d) \subset B_{\infty\infty}^{-r}(\mathbb{T}^d). \\ \Psi & & \Psi \\ U(\omega_1) & & \mathcal{E}(\omega_2) \end{array}$$

Observe that by Lemma 2 the random variable U takes values in the Besov space $Y = B_{pp}^t(\mathbb{T}^d)$. Above \mathcal{E} is standard Gaussian white noise: $\mathbb{E} \mathcal{E} = 0$ and

$$\mathbb{E} (\langle \mathcal{E}, \phi \rangle \langle \mathcal{E}, \psi \rangle) = \langle \phi, \psi \rangle$$

for all $\phi, \psi \in C^\infty(\mathbb{T}^d)$, where $\langle \cdot, \cdot \rangle$ is the distribution duality. To make our results more precise, we will consider the white noise as a random variable taking values in a Besov space instead of Sobolev spaces, and consider \mathcal{E} here as a random variable taking values in the Besov space $B_{\infty\infty}^{-r}(\mathbb{T}^d)$.

In assumption A we are particularly interested in g being a characteristic function $g(u) = \chi_E(u)$ of some set E (corresponding to confidence intervals), the identity map $g(u) = u$ (corresponding to the mean), and $g(u) = \|u\|_{B_{pp}^t}^q$. We denote by $H^g(m, A, G)$ the quantity

$$H^g(m, A, G) := \mathbb{E} \left(g(GU) \exp \left(-\frac{1}{2} \|AU\|_{L^2}^2 + \langle AU, m \rangle \right) \right).$$

In the case $g \equiv 1$ the operator G plays no role, and the corresponding real-valued quantity is denoted simply by $H^1(m, A)$. In general, the above integral is defined as a B_{pp}^t -valued Bochner integral. The weak measurability of the integrand is obvious, whence the strong measurability follows by the separability of the spaces involved. The following auxiliary result will be used to verify integrability and to estimate the sensitivity of the quantity $H^g(m, A, G)$ with respect to changes in the variables. Below, for $1 \leq p < \infty$ we denote $p' = p/(p - 1)$.

Proposition 2. Let U be distributed according to B_{pp}^s prior as in Definition 4. Let $q \geq 0$ and $m \in B_{\infty\infty}^{-r}$. Denote by $S_q(m, A)$ the random variable

$$S_q(m, A) := (1 + \|U\|_{B_{pp}^{\tilde{t}}})^q \exp \left(-\frac{1}{2} \|AU\|_{L^2}^2 + \langle AU, m \rangle \right).$$

Let $w \geq r$ be an arbitrary index, which in case $p = 1$ satisfies $w > r$. Assume that $A : B_{pp}^t(\mathbb{T}^d) \rightarrow B_{11}^w(\mathbb{T}^d)$ is bounded. Then there is a constant $c = c(p, r, t, w, d, q)$ such that

$$(41) \quad \mathbb{E} S_q(m, A) \leq c \exp \left(c(\|A\|_{B_{pp}^t \rightarrow B_{11}^w})^{2r/(w-r+2r/p')} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r+2r/p')} \right)$$

with the understanding that in the case $p = 1$ one sets $2r/p' = 0$.

Proof. In order to prove (41) we first observe that by Lemma 2(ii) and Cauchy-Schwarz inequality it is enough to estimate the expectation

$$(42) \quad I' := \mathbb{E} \exp \left(-\|AU\|_{L^2}^2 + 2\langle AU, m \rangle \right).$$

By Lemma 2 we have that $\mathbb{E} \exp(\frac{1}{2}\|U\|_{B_{pp}^t}^p) < \infty$. The decomposition

$$I' = \mathbb{E} \exp \left(\frac{1}{2}\|U\|_{B_{pp}^t}^p + \left(-\frac{1}{2}\|U\|_{B_{pp}^t}^p - \|AU\|_{L^2}^2 + 2\langle AU, m \rangle \right) \right)$$

shows that

$$(43) \quad I' \leq c \exp \left(\sup_U S(U) \right),$$

where

$$S(U) := -\frac{1}{2}\|U\|_{B_{pp}^t}^p - \|AU\|_{L^2}^2 + 2\langle AU, m \rangle.$$

For $\ell \geq 0$ denote by R_ℓ the standard projection to the first ℓ coordinates in the wavelet basis, $\ell \geq 0$, with $R_0 = 0$ and $Q_\ell = I - R_\ell$. Fix an integer $\ell_0 \geq 0$ and divide further above $m = R_{\ell_0}m + Q_{\ell_0}m$.

We now assume that $p > 1$. By completing the square, and applying Young's inequality in the form $-x^p/2 + 2xy \leq c_p y^{p'}$ we obtain

$$(44) \quad \begin{aligned} S(U) &= -\frac{1}{2}\|U\|_{B_{pp}^t}^p + 2\langle AU, Q_{\ell_0}m \rangle - \|AU\|_{L^2}^2 + 2\langle AU, R_{\ell_0}m \rangle \\ &\leq -\frac{1}{2}\|U\|_{B_{pp}^t}^p + 2\|U\|_{B_{pp}^t} \|A^*Q_{\ell_0}m\|_{B_{p'p'}^{-t}} - \|AU - R_{\ell_0}m\|_{L^2}^2 + \|R_{\ell_0}m\|_{L^2}^2 \\ &\leq c_p \|A^*Q_{\ell_0}m\|_{B_{p'p'}^{-t}}^{p'} + \|R_{\ell_0}m\|_{L^2}^2. \end{aligned}$$

If $m = 0$ there is nothing to prove, so assume that $m \neq 0$. Recall that $w \geq r$ and consider first the situation where

$$(45) \quad \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}}^{p'} \geq \left(\|m\|_{B_{\infty\infty}^{-r}} \right)^{2-p'},$$

where $1/p + 1/p' = 1$. In this case we choose

$$\begin{aligned} \ell_0 &:= \left[\left(\|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}}^{p'} \|m\|_{B_{\infty\infty}^{-r}}^{p'-2} \right)^{d/(2r+p'(w-r))} \right] + 1 \\ &\simeq \left(\|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}}^{p'} \|m\|_{B_{\infty\infty}^{-r}}^{p'-2} \right)^{d/(2r+p'(w-r))}. \end{aligned}$$

We may estimate

$$(46) \quad \begin{aligned} \|A^*Q_{\ell_0}m\|_{B_{p'p'}^{-t}} &\leq \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}} \|Q_{\ell_0}m\|_{B_{\infty\infty}^{-w}} \\ &\leq \ell_0^{(r-w)/d} \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}} \|m\|_{B_{\infty\infty}^{-r}}. \end{aligned}$$

Moreover, one easily verifies that

$$(47) \quad \|R_{\ell_0}m\|_{L^2}^2 \leq (2\ell_0)^{2r/d} \|m\|_{B_{\infty\infty}^{-r}}^2.$$

By invoking the choice of ℓ_0 and applying the auxiliary estimates (46) and (47) we compute from (44) that

$$S(U) \leq c(\|A\|_{B_{pp}^t \rightarrow B_{11}^w})^{2r/(w-r+2r/p')} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r+2r/p')}$$

which finishes the proof in the situation (45). Above, we also used the observation that $\|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}} = \|A\|_{B_{pp}^t \rightarrow B_{11}^w}$. On the other hand, if (45) is not valid, that is,

$$(48) \quad \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{p'p'}^{-t}} < \left(\|m\|_{B_{\infty\infty}^{-r}}\right)^{2-p'},$$

we apply the choice $\ell_0 = 0$ in (44) to estimate

$$(49) \quad \begin{aligned} S(U) &\leq c_p \|A^* m\|_{B_{p'p'}^{-t}}^{p'} \\ &\leq c \|A^*\|_{B_{\infty\infty}^{-r} \rightarrow B_{p'p'}^{-t}}^{p'} \|m\|_{B_{\infty\infty}^{-r}}^{p'} \\ &\leq c(\|A\|_{B_{pp}^r \rightarrow B_{11}^w})^{2r/(w-r+2r/p')} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r+2r/p')}. \end{aligned}$$

In case $w > r$ the last inequality above followed from inequality (48).

Consider next the case $p = 1$. Assuming first that

$$(50) \quad \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{\infty\infty}^{-t}} \|m\|_{B_{\infty\infty}^{-r}}^{p'} \geq 1/4$$

we utilize the fact that $w > r$ and choose

$$\begin{aligned} \ell_0 &:= \left[(4 \|m\|_{B_{\infty\infty}^{-r}} \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{\infty\infty}^{-t}})^{d/(w-r)} \right] + 1 \\ &\simeq (4 \|m\|_{B_{\infty\infty}^{-r}} \|A^*\|_{B_{\infty\infty}^{-w} \rightarrow B_{\infty\infty}^{-t}})^{d/(w-r)} \end{aligned}$$

and observe that then (46) verifies that $\|A^* Q_{\ell_0} m\|_{B_{\infty\infty}^{-t}} \leq 1/4$, which in turn yields

$$\begin{aligned} S(U) &= -\frac{1}{2} \|U\|_{B_{11}^t} + 2 \langle AU, Q_{\ell_0} m \rangle - \|AU\|_{L^2}^2 + 2 \langle AU, R_{\ell_0} m \rangle_{L^2} \\ &\leq -\|AU - R_{\ell_0} m\|_{L^2}^2 + \|R_{\ell_0} m\|_{L^2}^2 \\ &\leq \|R_{\ell_0} m\|_{L^2}^2 \\ &\leq (2\ell_0)^{2r/d} \|m\|_{B_{\infty\infty}^{-r}}^2 \\ &\leq c(4\|A\|_{B_{11}^t \rightarrow B_{11}^w})^{2r/(w-r)} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r)}, \end{aligned}$$

where we also applied the estimate (47). Finally, in case (50) does not hold we may choose $\ell_0 = 0$ above and obtain that

$$S(U) \leq c(\|A\|_{B_{pp}^t \rightarrow B_{11}^w})^{2r/(w-r+2r/p')} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r+2r/p')}$$

for all U . Summarizing, we have shown that in all the above cases

$$(51) \quad \sup_U S(U) \leq c_1 (\|A\|_{B_{pp}^t \rightarrow B_{11}^w})^{2r/(w-r+2r/p')} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r+2r/p')},$$

with some $c_1, c_2 > 0$, which finishes the proof of the Proposition. \square

Our first application of the above result will be a local Lipschitz continuity estimate for the quantity $H^g(m, A, G)$.

Proposition 3. *Denote*

$$\begin{aligned} K &:= \max(\|m\|_{B_{\infty\infty}^{-r}}, \|m'\|_{B_{\infty\infty}^{-r}}), \\ a &:= \max(\|A\|_{B_{pp}^t \rightarrow B_{11}^w}, \|A'\|_{B_{pp}^t \rightarrow B_{11}^w}), \\ v &:= \max(\|G\|_{B_{pp}^i \rightarrow B_{pp}^t}, \|G'\|_{B_{pp}^i \rightarrow B_{pp}^t}). \end{aligned}$$

Then it holds that

$$(52) \quad \begin{aligned} &\|H^g(m, A, G) - H^g(m', A', G')\|_{B_{pp}^t} \\ &\leq (\|m - m'\|_{B_{\infty\infty}^{-r}} + \|A - A'\|_{B_{pp}^t \rightarrow B_{11}^w} + \|G - G'\|_{B_{11}^i \rightarrow B_{11}^t})h(K, a, v), \end{aligned}$$

where

$$h(K, a, v) := c(K + a + (1 + v)^q) \exp\left(a^{(2r/(w-r+2r/p'))} K^{(2w/(w-r+2r/p'))}\right).$$

Here w and other indices are as in Proposition 2.

Proof. Choose arbitrary $h' \in B_{p'p'}^{-t}$ with $\|h'\|_{B_{p'p'}^{-t}} = 1$. It is enough to estimate the difference

$$\mathbb{E}|f(2, U) - f(1, U)|,$$

where for $x \in [1, 2]$ and $U \in B_{pp}^t$ we have

$$(53) \quad \begin{aligned} f(x, U) = & \left\langle h', g((G + (x-1)(G' - G))U) \exp\left(-\frac{1}{2}\|(A + (x-1)(A' - A))U\|_{L^2}^2 + \right. \right. \\ & \left. \left. + \langle (A + (x-1)(A' - A))U, m + (x-1)(m' - m) \rangle\right) \right\rangle \end{aligned}$$

It will be convenient to introduce for each $x \in [1, 2]$ the notation $G_x = G + (x-1)(G' - G)$, $A_x = A + (x-1)(A' - A)$, and $m_x = m + (x-1)(m' - m)$. As g was assumed to be Lipschitz it follows that for each fixed U the function $x \mapsto f(x, U)$ is also Lipschitz. Hence we may compute for almost every $x \in [1, 2]$

$$\begin{aligned} D_x f(x, U) &= \left(\frac{d}{dx} \langle h', g(G_x U) \rangle + \langle h', g(G_x U) \rangle (-\langle (A' - A)U, A_x U \rangle_{L^2} + \right. \\ &\quad \left. + \langle A_x U, m' - m \rangle + \langle (A' - A)U, m_x \rangle) \right) \cdot \\ &\quad \cdot \exp\left(-\frac{1}{2}\|A_x U\|_{L^2}^2 + \langle A_x U, m_x \rangle\right). \end{aligned}$$

In order to estimate the right hand side we observe first that our assumption on g yields for almost every $x \in [1, 2]$

$$\begin{aligned} \left| \frac{d}{dx} (\langle h', g(G_x U) \rangle) \right| &\leq \|(G' - G)U\|_{B_{pp}^t} (1 + v\|U\|_{B_{pp}^i})^q \\ &\leq \|G' - G\|_{B_{pp}^i \rightarrow B_{pp}^t} \|U\|_{B_{pp}^i} (1 + v\|U\|_{B_{pp}^i})^q \\ &\leq \|G' - G\|_{B_{pp}^i \rightarrow B_{pp}^t} (1 + v)^q (1 + \|U\|_{B_{pp}^i})^{q+1}. \end{aligned}$$

Moreover, by using the observation that $B_{11}^r(\mathbb{T}^d) \subset L^2(\mathbb{T}^d)$ it follows that

$$\begin{aligned} & | - \langle (A' - A)U, A_x U \rangle_{L^2} + \langle A_x U, m' - m \rangle + \langle (A' - A)U, m_x \rangle | \\ \leq & \|A_x\|_{B_{pp}^t \rightarrow L^2} \|A' - A\|_{B_{pp}^t \rightarrow L^2} \|U\|_{B_{pp}^t}^2 + \|m' - m\|_{B_{\infty\infty}^{-r}} \|A_x\|_{B_{pp}^t \rightarrow B_{11}^r} \|U\|_{B_{pp}^t} \\ & + \|m_x\|_{B_{\infty\infty}^{-r}} \|A' - A\|_{B_{pp}^t \rightarrow B_{11}^r} \|U\|_{B_{pp}^t} \\ \leq & c(K + a) (\|m' - m\|_{B_{\infty\infty}^{-r}} + \|A' - A\|_{B_{p,p}^t \rightarrow B_{11}^r}) (1 + \|U\|_{B_{pp}^t})^2. \end{aligned}$$

By combining the previous bounds and the obvious bound for $g(G_x U)$ we obtain

$$(54) \quad |D_x f(x, U)| \leq c(K + a + (1 + v)^q) (\|m' - m\|_{B_{\infty\infty}^{-r}} + \|A' - A\|_{B_{p,p}^t \rightarrow B_{11}^r} + \|G' - G\|_{B_{pp}^i \rightarrow B_{pp}^t}) \cdot S_{q+2}(m_x, A_x).$$

The Fubini theorem allows us to compute

$$(55) \quad \begin{aligned} \mathbb{E} |f(2, U) - f(1, U)| & \leq \int_1^2 \left(\mathbb{E} \left| \frac{d}{dx} (f(x, U)) \right| \right) dx \\ & \leq c(K + a + (1 + v)^q) (\|m' - m\|_{B_{\infty\infty}^{-r}} + \|A' - A\|_{B_{p,p}^t \rightarrow B_{11}^r} + \|G' - G\|_{B_{pp}^i \rightarrow B_{pp}^t}) \\ & \quad \cdot \int_1^2 \mathbb{E} S_{q+2}(m_x, A_x) dx. \end{aligned}$$

According to Proposition 2 there is the uniform bound

$$(56) \quad \mathbb{E} S_{q+2}(m_x, A_x) \leq c \exp \left(a^{(2r/(w-r+2r/p'))} K^{(2w/(w-r+2r/p'))} \right),$$

and the (52) follows immediately by combining this with (55). \square

The local Lipschitz constant of a given map $r : E \rightarrow F$ between the Banach spaces E and F is defined as

$$\text{Lip}(r)(x) := \limsup_{y \rightarrow x} \frac{\|r(y) - r(x)\|_F}{\|y - x\|_E} \quad \text{for } x \in E.$$

In terms of this quantity the previous proposition states that

$$\text{Lip}(H^g)(m, A, G) \leq c(1 + \|G\|_{B_{pp}^i \rightarrow B_{pp}^t})^q \exp \left(c(\|A\|_{B_{pp}^t \rightarrow B_{11}^w})^{2r/(w-r+2r/p')} (\|m\|_{B_{\infty\infty}^{-r}})^{2w/(w-r+2r/p')} \right),$$

where the linear factors containing norms of A and m were absorbed in the exponential term. Above we consider H^g as the map

$$H^g : \left(B_{\infty\infty}^{-r} \oplus L(B_{pp}^t, B_{11}^w) \oplus L(B_{pp}^i, B_{pp}^t) \right) \longrightarrow B_{pp}^t.$$

Before we are able to give good estimates for the Lipschitz constant of the ratio H^g/H^1 we need a couple of auxiliary results.

Lemma 3. *Let $a > 0$ and assume that f and F are non-negative random variables with $\mathbb{E} F < \infty$. Then there is a constant c depending only on a so that*

$$(57) \quad \frac{\mathbb{E}(fF)}{\mathbb{E} F} \leq c \left(c + \log(\mathbb{E} \exp(f^a/2)) + \log(\mathbb{E} F^2) - 2 \log(\mathbb{E} F) \right)^{1/a}.$$

Proof. Let us consider the change of probability measure \mathbb{P}_F that is simply obtained by using F as a weight and normalizing. Then the left hand side of (57) equals $\mathbb{E}_F f$. By invoking the convex function $\phi_a(x) := \exp((1+|x|)^a/a)$ and applying the Jensen inequality we obtain

$$\phi_a(\mathbb{E}_F f) \leq \mathbb{E}_F \phi_a(f) \leq \frac{(\mathbb{E}(\phi_a(f))^2)^{1/2} (\mathbb{E} F^2)^{1/2}}{\mathbb{E} F},$$

where the last inequality followed from the Cauchy-Schwartz inequality. By using the inequality $\frac{2}{a}(1 + |x|)^a \leq c + c|x|^a$ and applying the inverse $\phi_a^{-1}(x) = (a \log(x))^{1/a} - 1$ a direct computation yields the inequality

$$(58) \quad \mathbb{E}(fF)/\mathbb{E}F \leq c(c + \log(\mathbb{E} \exp(cf^a)) + \log(\mathbb{E} F^2) - 2 \log(\mathbb{E} F))^{1/a}.$$

Then (57) follows by applying (58) on $(2c)^{-1/a}f$ in place of f . \square

Lemma 4. *For any $t < s - n/p$ there is a constant $C = C(t) > 0$ such that*

$$H^1(m, A) \geq C(t) \exp\left(-\frac{1}{2}\|A\|_{B_{pp}^t \rightarrow L^2}^2\right),$$

independently of m .

Proof. Since the prior measure is invariant under the change of variables $U \rightarrow -U$ we obtain that

$$\begin{aligned} & \mathbb{E} H^1(m, A) \\ &= \frac{1}{2} \mathbb{E} \left(\exp\left(-\frac{1}{2}\|AU\|_{L^2}^2 + \langle AU, m \rangle\right) + \exp\left(-\frac{1}{2}\|AU\|_{L^2}^2 - \langle AU, m \rangle\right) \right) \\ &\geq \mathbb{E} \exp\left(-\frac{1}{2}\|AU\|_{L^2}^2\right). \end{aligned}$$

We may clearly take $C(t) := \mathbb{P}(\{\|U\|_{B_{pp}^t} \leq 1\})$. \square

Lemma 5. *Let $\lambda > 0$. Under assumption A and with w as in Proposition 2 there is the estimate*

$$\frac{\mathbb{E} S_\lambda(m, A)}{\mathbb{E} S_0(m, A)} \leq c(1 + \|m\|_{B_{\infty}^{-r}})^{\beta_1} (1 + \|A\|_{B_{pp}^t \rightarrow B_{11}^w})^{\beta_2}$$

where $\beta_1 = \frac{2\lambda w}{p(w-r+2r/p')}$ and $\beta_2 = \max\left(\frac{2\lambda}{p}, \frac{2r\lambda}{p(w-r+2r/p')}\right)$.

Proof. We apply Lemma 3 with the choice $a = p/\lambda$, $f = (\frac{1}{2}(1 + \|U\|_{B_{pp}^i}))^\lambda$ and $F = \exp\left(-\frac{1}{2}\|AU\|_{L^2}^2 + \langle AU, m \rangle\right)$. One inserts the estimate for the quantity $\mathbb{E} F^2 = I'$, see (42), obtained in (43) and (51) in the proof of Proposition 2. Moreover, a lower bound for $\mathbb{E} F$ is given by Lemma 4 and we recall that $\mathbb{E} \exp(f^a/2) < \infty$ thanks to Lemma 2. \square

We are ready for the main result of this section.

Theorem 4. *Consider the ratio $H^g(m, A, G)/H^1(m, A, G)$ as a map*

$$\frac{H^g}{H^1} : \left(B_{\infty}^{-r} \oplus L(B_{pp}^t, B_{11}^w) \oplus L(B_{pp}^{\tilde{t}}, B_{pp}^t) \right) \longrightarrow B_{pp}^t.$$

Under assumption A, and with w as in Proposition 2, the local Lipschitz constant of this function satisfies

(59)

$$\text{Lip}\left(\frac{H^g}{H^1}\right)(m, A, G) \leq c(1 + \|G\|_{B_{pp}^i \rightarrow B_{pp}^t})^q (1 + \|m\|_{B_{\infty}^{-r}})^\gamma (1 + \|A\|_{B_{pp}^t \rightarrow B_{11}^w})^\alpha,$$

where $\alpha = 1 + \left(\frac{2q+8}{p}\right) \max\left(1, \left(\frac{r}{w-r+2r/p'}\right)\right)$ and $\gamma = 1 + \left(\frac{2q+8}{p}\right) \left(\frac{w}{w-r+2r/p'}\right)$.

Proof. Denote $K = \|m\|_{B_{\infty}^{-r}}$, $a = \|A\|_{B_{pp}^t \rightarrow B_{11}^w}$ and $v = \|G\|_{B_{pp}^i \rightarrow B_{pp}^t}$. Proposition 3 verifies that both H^g and H^1 are locally Lipschitz. As a special case we obtain that

$\mathbb{E}S_{q+2}(m, A)$ is continuous with respect to variables A and m . Hence inequality (55) yields the estimate

$$\text{Lip}(H^g)(m, A, G) \leq c(K + a + (1 + v)^q)\mathbb{E}S_{q+2}(m, A).$$

The simple inequality (here x_1, x_2 are vectors and $y_1, y_2 > 0$ are scalars)

$$\left\| \frac{x_1}{y_1} - \frac{x_2}{y_2} \right\| \leq \frac{1}{y_2} \|x_2 - x_1\| + \frac{|y_2 - y_1|}{y_1 y_2} \|x_1\|$$

verifies that point-wise

$$\text{Lip}\left(\frac{H^g}{H^1}\right) \leq \frac{\text{Lip}(H^g)}{H^1} + \frac{\|H^g\|_{B_{pp}^t} \text{Lip}(H^1)}{(H^1)^2}.$$

Observe that by assumption A we have $\|H^g\|_{B_{pp}^t} \leq c\mathbb{E}S_q$. By combining the previous estimates and remembering that $H^1 = \mathbb{E}S_0$ we thus obtain

$$\text{Lip}\left(\frac{H^g}{H^1}\right) \leq c(K + a + (1 + v)^q) \left(\frac{\mathbb{E}S_{q+2}}{\mathbb{E}S_0}\right) + c\left(\frac{\mathbb{E}S_q}{\mathbb{E}S_0}\right) c(K + a) \left(\frac{\mathbb{E}S_2}{\mathbb{E}S_0}\right).$$

The statement then follows by three applications of Lemma 5. □

Remark 1. The main content of Theorem 4 is a stability estimate that grows polynomially in the norm of the measurement m . We have not striven for the optimal exponents in the above computations. Moreover, one should observe that in the case $p > 1$ it is possible to choose $w = r$ in Theorem 4, which corresponds to the weakest condition on smoothness of A . If $p \in (1, 2)$ this yields the exponents $\alpha = \gamma = 1 + p'(q + 4)/p$. In the important special case of $p = 1$ one is forced to choose $w > r$. On the other hand, the choices $w > r$ yield considerably smaller exponents for all small values $p \geq 1$: in the limit $w \rightarrow \infty$ one has $\alpha \rightarrow 1 + \frac{(2q+8)}{p}$, and $\gamma \rightarrow 1 + \frac{(2q+8)}{p}$.

6. Convergence results for Besov priors. In the present section we assume that indices t, \tilde{t}, r and the quantities m, A , and g satisfy Assumption A. We take $T_n : B_{pp}^{\tilde{t}}(\mathbb{T}^d) \rightarrow B_{pp}^{\tilde{t}}(\mathbb{T}^d)$, $n \geq 1$, defined by the familiar truncation

$$(60) \quad T_n \left(\sum_{\ell=1}^{\infty} c_{\ell} \psi_{\ell} \right) = \sum_{\ell=1}^n c_{\ell} \psi_{\ell}$$

discussed above and in Appendix A. Then $T_n \rightarrow I$ strongly in $L(B_{pp}^{\tilde{t}})$ as $n \rightarrow \infty$. We consider proper linear discretizations $U_n = T_n U$ of the random variable U .

By standard compact imbedding results it follows that $(I - T_n) \rightarrow 0$ as $n \rightarrow \infty$ in the operator norm topology of $L(B_{pp}^{\tilde{t}}, B_{pp}^t)$, and $(I - P_k) \rightarrow 0$ as $k \rightarrow \infty$ in the operator norm topology of $L(B_{11}^{\tilde{r}}, B_{11}^r)$ with $\tilde{r} > r$, see Appendix A. Next we formulate the assumptions in quantitative terms.

Assumption B: *Let $\tilde{r} > r$ and assume that $A : B_{pp}^t(\mathbb{T}^d) \rightarrow B_{11}^{\tilde{r}}(\mathbb{T}^d)$. Define T_n by (60); then we have*

$$(61) \quad \|I - T_n\|_{L(B_{pp}^{\tilde{t}}, B_{pp}^t)} < \eta_2(n) \quad \text{for } n \geq 1$$

with $\lim_{n \rightarrow \infty} \eta_2(n) = 0$. Further, let $P_k : B_{pp}^{\tilde{r}}(\mathbb{T}^d) \rightarrow B_{pp}^{\tilde{r}}(\mathbb{T}^d)$ be bounded linear projections with k -dimensional range satisfying

$$(62) \quad \|I - P_k\|_{L(B_{11}^{\tilde{r}}, B_{11}^r)} < \eta_1(k), \quad \text{for } k \geq 1$$

with $\lim_{k \rightarrow \infty} \eta_1(k) = 0$. Assume that $\|T_n\|_{L(B_{pp}^i)} \leq C$ and $\|P_k\|_{L(B_{pp}^r)} \leq C$ for all $k, n \geq 1$. The measurements are assumed to be uniformly bounded by a constant K :

$$(63) \quad \|m\|_{B_{\infty\infty}^{-r}} \leq K \quad \text{and} \quad \|m_k\|_{B_{\infty\infty}^{-r}} \leq K \quad \text{for all } k \geq 1.$$

Finally, it is assumed that

$$(64) \quad \eta_3(k) := \|m_k - m\|_{B_{\infty\infty}^{-r}} \text{ satisfies } \lim_{k \rightarrow \infty} \eta_3(k) = 0.$$

Observe that the condition (22) is valid for Besov prior only when $p > 1$. However, under Assumption A, Prop. 2 and Lemma 4 verify for $p \in [1, \infty)$ that the function

$$(65) \quad \mathcal{R}_M(g(U)|m) = \frac{H_{U,M}^g(m)}{H_{U,M}^1(m)}$$

is well defined for all $m \in S$. The proof of Theorem 1 now applies unchanged yielding that (65) is a reconstructor.

Let $\Theta_{kn} = P_k A U_n + \mathcal{E}$. As before we define the reconstructor $\mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k)$ corresponding to measurement Θ_{kn} at m_k as

$$(66) \quad \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) = \frac{H_{U_n, \Theta_{kn}}^g(m_k)}{H_{U_n, \Theta_{kn}}^1(m_k)}.$$

The following result yields a quantitative convergence result for Besov priors:

Theorem 5. *Let Assumptions A and B hold. Denote $\gamma = 1 + (\frac{2q+8}{p})(\frac{w}{w-r+2r/p'})$. There is a constant $C' = C'(A, g, t, r, w, p)$ such that*

$$\|\mathcal{R}_{\Theta_{kn}}(U_n|m_k) - \mathcal{R}_M(U|m)\|_{B_{pp}^t} \leq C'(1+K)^\gamma [\eta_1(k) + \eta_2(n) + \eta_3(k)].$$

Moreover, the limits

$$\lim_{n \rightarrow \infty} \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k) \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathcal{R}_{\Theta_{kn}}(g(U_n)|m_k)$$

exist for a fixed value of k (resp. n).

Proof. The statement is an immediate consequence of the estimates obtained in the previous section: observe that in our notation

$$\frac{H_{U_n, \Theta_{kn}}^g(m_k)}{H_{U_n, \Theta_{kn}}^1(m_k)} = \frac{H^g(m_k, P_k A T_n, T_n)}{H^1(m_k, P_k A T_n, T_n)}.$$

Moreover, apart from a possible change of the uninteresting constants, the Lipschitz bound given by Theorem 4 on the quantity H^g/H^1 is uniform on any bounded neighbourhood of m , A and G , where G is the identical embedding $\text{Id} : B_{pp}^t \rightarrow B_{pp}^t$. Note also that by our assumptions $(A - P_k A T_n) \rightarrow 0$ in the operator norm topology of $L(B_{pp}^t, B_{11}^t)$. The first claim now follows from definitions and Theorem 4. The last statements follow immediately by the same reasoning. \square

We emphasize that Theorem 5 covers the highly interesting value $p = 1$ despite the failure of assumption (22) of the general theory in that case.

Let us revisit the deblurring example of Section 2.2, where $p = 1$, $s = 1$, $d = 2$ and A has a smooth kernel. Often it is possible to take the projection P_k related to the measurement device to be truncation of the wavelet expansion to the first k terms analogously to (60). (For instance, P_k might measure local averages at a grid

of points and then compute discrete wavelet transform by convolutions with finite filters.)

Then Theorem 5 yields the convergence of reconstructors and a result analogous to Theorem 3 with an explicit convergence speed.

Corollary 2. *Let Assumptions A and B hold. Let $p = s = 1$ and assume that $A : \mathcal{D}'(\mathbb{T}^2) \rightarrow C^\infty(\mathbb{T}^2)$ is a bounded linear operator. Let P_k and T_n be truncations to k and n first terms in wavelet expansion, respectively. Moreover, let $t < \tilde{t} < -1$, $r > r_1 > 1$, $\lambda > 11$, and $\tau > 0$. Then*

(i) *There is a constant $c_0 = c_0(A, r, t, \lambda) > 0$ such that the reconstructors satisfy*

$$\|\mathcal{R}_{\Theta_{kn}}(U_n|m_k) - \mathcal{R}_M(U|m)\|_{B_{11}^t(\mathbb{T}^2)} \leq c_0(1 + K)^\lambda[\eta_1(k) + \eta_2(n) + \eta_3(k)].$$

(ii) *Let $u = U(\omega_0)$, $\varepsilon = \mathcal{E}(\omega_0)$, $\omega_0 \in \Omega$ be realizations of the random variables U and \mathcal{E} , and*

$$m = Au + \varepsilon, \quad m_k = A_k u + P_k \varepsilon,$$

be the realizations of the measurements (1) and (2), respectively. Moreover, assume that $m \in B_{11}^{-r_1}(\mathbb{T}^2)$. Then there is $C > 0$ independent of n and k so that the reconstructors defined in Theorem 1 for measurements (2) and (1) satisfy

$$(67) \quad \|\mathcal{R}_{M_{kn}}(U_n|m_k) - \mathcal{R}_M(U|m)\|_{B_{11}^t(\mathbb{T}^2)} \leq C[k^{-\tau} + n^{-(\tilde{t}-t)/2}].$$

Note that in (ii) we have $m \in B_{11}^{-r_1}(\mathbb{T}^2)$ for \mathbb{P} -a.e. ω .

Proof. Claim (i) follows directly from Theorem 5 when we use $q = 1$ and so large w that $\lambda \geq \gamma$. As A is an infinitely smoothing operator, we can choose above any $-\infty < t < \tilde{t} < -1$ and $\tilde{r} > r + 2\tau > r_1 + 4\tau$. Moreover, since $m \in B_{pp}^{-r_1}(\mathbb{T}^2)$, we can take above $\eta_1(k) = c_1 k^{-(\tilde{r}-r)/2}$, $\eta_2(n) = c_2 n^{-(\tilde{t}-t)/2}$, and $\eta_3(k) = c_3 k^{-(r-r_1)/2}$ where c_1, c_2 and c_3 depend on ω_0 and the parameters $\tilde{r}, r, r_1, \tilde{t}, t$, but not on k or n . As the projections P_k satisfy the conditions (34) and (35), the assertion (ii) follows from (i) and Lemma 1. \square

Appendix A. Besov spaces and wavelets. Let $\tilde{\psi}$ and $\tilde{\phi}$ be compactly supported wavelet and scaling function suitable for multi-resolution analysis of smoothness C^r in $L^2(\mathbb{R})$.

Following Daubechies [16, section 9.3] we construct a wavelet representation for periodic functions in \mathbb{R} with period 1; in other words, for functions on the one-dimensional torus \mathbb{T}^1 . Set

$$\begin{aligned} \phi_{j,k}(x) &= \sum_{\ell \in \mathbb{Z}} \tilde{\phi}(2^j(x + \ell) - k), \\ \psi_{j,k}(x) &= \sum_{\ell \in \mathbb{Z}} \tilde{\psi}(2^j(x + \ell) - k). \end{aligned}$$

We use in the following the subspaces of $L^2(\mathbb{T}^1)$,

$$V_j := \overline{\text{span}\{\phi_{j,k} \mid k \in \mathbb{Z}\}}, \quad W_j := \overline{\text{span}\{\psi_{j,k} \mid k \in \mathbb{Z}\}}.$$

It turns out that V_j are spaces of constant functions for $j \leq 0$. Thus we have a ladder $V_0 \subset V_1 \subset V_2 \subset \dots$ of multiresolution spaces satisfying

$$\bigcup_{j \geq 0} V_j = L^2(\mathbb{T}^1).$$

Further, we denote the successive orthogonal complements of V_j in V_{j+1} by W_j for $j \geq 0$. Then we have orthonormal bases

$$\begin{aligned} &\{\phi_{j,k} \mid k = 0, \dots, 2^j - 1\} \text{ in } V_j, \\ &\{\psi_{j,k} \mid k = 0, \dots, 2^j - 1\} \text{ in } W_j. \end{aligned}$$

Following Meyer [47, section 3.9] we define a wavelet basis for periodic functions in \mathbb{R}^d ; in other words, for functions on the torus \mathbb{T}^d . Let E denote the set of $2^d - 1$ sequences $\nu = (\nu_1, \nu_2, \dots, \nu_d)$ of zeroes and ones, excluding the sequence $(0, 0, \dots, 0)$. Define for $\nu \in E$ and $j \geq 0$ the wavelets

$$\psi_{j,k}^\nu(x) := 2^{dj/2} \psi^{\nu_1}(2^j x_1 - k_1) \dots \psi^{\nu_d}(2^j x_d - k_d)$$

with the convention that $\psi^0 = \phi$ and $\psi^1 = \psi$, and integer-valued components of vector k ranging over

$$0 \leq k_i \leq 2^j - 1 \quad \text{for all } i = 1, 2, \dots, d.$$

The functions $\psi_{j,k}^\nu(x)$ constitute an orthonormal basis for $L^2(\mathbb{T}^d)$. Let us renumber the above basis functions using just one integer $\ell = 1, 2, \dots$. First, $\ell = 1$ corresponds to the scaling function $\phi(x_1) \dots \phi(x_n)$. The remaining numbering is done scale by scale; that is, we first number wavelets with $j = 0$, then wavelets with $j = 1$, and so on. The $2^d - 1$ indices $\nu \in E$ are naturally numbered by thinking them as binary representation of integers. The exact ordering of all 2^{jd} translations corresponding to a fixed j can be chosen arbitrarily. This leads to a numbering of the following type:

$$\begin{aligned} \text{scale } j = 0 : & \quad \ell = 2, \dots, 2^d, \\ \text{scale } j = 1 : & \quad \ell = 2^d + 1, \dots, 2^{2d}, \\ \text{scale } j = 2 : & \quad \ell = 2^{2d} + 1, \dots, 2^{3d}, \\ & \quad \vdots \quad \quad \quad \vdots \end{aligned}$$

According to Meyer [47, Section 6.10], we can characterize periodic Besov space functions using these wavelets. Namely, the series

$$f(x) = \sum_{\ell=1}^{\infty} c_\ell \psi_\ell(x)$$

belongs to $B_{pq}^s(\mathbb{T}^d)$ if and only if

$$2^{js} 2^{dj(\frac{1}{2} - \frac{1}{p})} \left(\sum_{\ell=2^{jd}}^{2^{(j+1)d} - 1} |c_\ell|^p \right)^{1/p} \in \ell^q(\mathbb{N})$$

We always assume that r is large enough for providing bases for Besov spaces with smoothness s . The case $q = p$ is especially relevant to us, and we obtain the equivalent norm

$$(68) \quad \left\| \sum_{\ell=1}^{\infty} c_\ell \psi_\ell(x) \right\|_{B_{pp}^s(\mathbb{T}^d)} := \left(\sum_{\ell=1}^{\infty} \ell^{(ps/d + p/2 - 1)} |c_\ell|^p \right)^{1/p}.$$

We use the above quantity as the definition of the Besov norm $\|\cdot\|_{B_{pp}^s(\mathbb{T}^d)}$ for generalized functions on \mathbb{T}^d , $s \in \mathbb{R}$ and $p \in [1, \infty]$. In case $p = \infty$ the definition

must be understood as follows:

$$\left\| \sum_{\ell=1}^{\infty} c_{\ell} \psi_{\ell}(x) \right\|_{B_{pp}^s(\mathbb{T}^d)} := \sup_{\ell \geq 1} \ell^{(s/d+1/2)} |c_{\ell}|.$$

It follows that $B_{pp}^s(\mathbb{T}^d)$ is isometrically isomorphic to the sequence space ℓ^p , and by the simplicity of the norm it is easy to control the basic properties of the spaces. Especially, there is an embedding (an easy corollary of the Hölder inequality),

$$(69) \quad B_{p_1 p_1}^{s_1}(\mathbb{T}^d) \subset B_{p_2 p_2}^{s_2}(\mathbb{T}^d) \quad \text{if and only if} \quad s_1 - d/p_1 \geq s_2 - d/p_2,$$

and it is easy to verify that this embedding is compact if $s_1 - d/p_1 > s_2 - d/p_2$. The dual of $B_{pp}^s(\mathbb{T}^d)$ is $B_{p'p'}^{-s}(\mathbb{T}^d)$ if $p \in (1, \infty)$, and p' stands for the dual index: $1/p + 1/p' = 1$. The duality is with respect to the standard duality bracket: if $f = \sum_{\ell=1}^{\infty} f_{\ell} \psi_{\ell}$ and $g = \sum_{\ell=1}^{\infty} g_{\ell} \psi_{\ell}$ are finite sums, then

$$\langle f, g \rangle = \int_{\mathbb{T}^d} f(x)g(x) dx = \sum_{\ell=1}^{\infty} f_{\ell} g_{\ell}.$$

We finally observe that natural bounded linear projection operators on $B_{pp}^s(\mathbb{T}^d)$ are obtained by setting

$$T_n f(x) = \sum_{\ell=1}^n c_{\ell} \psi_{\ell}(x).$$

These projections work at the same time for all the spaces and are contractions. Moreover, if $p < \infty$, then $\lim_{n \rightarrow \infty} \|f - T_n f\|_{B_{pp}^s(\mathbb{T}^d)} = 0$ for all $f \in B_{pp}^s(\mathbb{T}^d)$.

Appendix B. Examples of limits of finite-dimensional random variables.

Here we illustrate difficulties related to finite-dimensional models and their possible convergence to an infinite-dimensional continuum model. Unless otherwise stated, we work on a circle \mathbb{T}^1 , or equivalently, the interval $[0, 1]$ with end points identified. Let $u \in L^2(\mathbb{T}^1)$. We consider two measurements

$$A^{(1)}u = \int_{\mathbb{T}^1} u(t) dt, \quad A^{(2)}u = u\left(\frac{1}{2}\right)$$

and the corresponding measurement models

$$(70) \quad M_n^{(1)} = A^{(1)}U_n + \mathcal{E}, \quad M^{(1)} = A^{(1)}U + \mathcal{E},$$

$$(71) \quad M_n^{(2)} = A^{(2)}U_n + \mathcal{E}, \quad M^{(2)} = A^{(2)}U + \mathcal{E}$$

where U_n is a finite-dimensional random variable, U is a random variable in an infinite-dimensional function space, and \mathcal{E} is a normalized Gaussian random variable independent of U and U_n . We consider various examples of U_n and study whether some random variable U could be considered as a limit of U_n as $n \rightarrow \infty$, and whether models (70) or (71) make sense in the limit.

In all examples below, X_j^n , $j, n \in \mathbb{N}$ are independent real-valued normalized Gaussian random variables.

Example 1. (“Non-proper discretization of white noise”) Let $I(n, j) = (\frac{j-1}{n}, \frac{j}{n}]$, $j = 1, \dots, n$ and $\chi_j^n(t) = \chi_{I(n, j)}(t)$ be the indicator functions of intervals $I(n, j)$. Let

$$(72) \quad U_n(t) = \sum_{j=1}^n a_j^n X_j^n \chi_j^n(t), \quad t \in \mathbb{T}^1$$

where $a_j^n > 0$ are parameters.

For a fixed n , with an *ad hoc* choice $a_j^n = 1$, the functions U_n could be considered as an interesting random signal. However, for any function $\phi \in L^2(\mathbb{T}^1)$

$$\lim_{n \rightarrow \infty} \int_{\mathbb{T}^1} U_n(t) \phi(t) dt = 0 \quad \text{in distribution,}$$

and thus U_n , considered as $L^2(\mathbb{T}^1)$ valued random variables, converge to zero weakly in distribution as $n \rightarrow \infty$. Taking $U = 0$ we see that the measurements $M_n^{(1)}$ converge in distribution to $M^{(1)}$. Concerning measurement (71) we notice the $U_n(\frac{1}{2}) \sim N(0, 1)$, but $U(\frac{1}{2}) \sim N(0, 0)$. Thus $M_n^{(2)} \sim N(0, 2)$ for all n , but for $U = 0$ we have $M^{(2)} \sim N(0, 1)$. This shows that measurement models (71) do not behave nicely with the choice $a_j^n = 1$. In this example U_n are not proper linear discretizations of U in any Banach space, since the second condition in Definition 1 is violated.

Example 2. (“Proper discretization of white noise”) Consider random variables U_n defined in (72) with parameters $a_j^n = n^{1/2}$. This choice is motivated by the fact that the functions $n^{1/2} \chi_{I(n,j)}$ are orthonormal in $L^2(\mathbb{T}^1)$. Then, for $\phi \in C^\infty(\mathbb{T}^1)$

$$(73) \quad \lim_{n \rightarrow \infty} \int_{\mathbb{T}^1} U_n(t) \phi(t) dt = \langle U, \phi \rangle \quad \text{in distribution,}$$

where U is Gaussian white noise in $L^2(\mathbb{T}^1)$. This actually holds also when $\phi \in H^1(\mathbb{T}^1)$. Let Q_n be the $L^2(\mathbb{T}^1)$ -orthogonal projection on the subspace of functions that have constant value on the intervals $I(n, j)$. Then $\langle U_n, \phi \rangle$ appearing on the left hand side of (73) is a real-valued Gaussian random variable with covariance

$$\sum_{j=1}^n |\langle \phi, n^{1/2} \chi_{I(n,j)} \rangle|^2 = \|Q_n \phi\|_2^2.$$

As $n \rightarrow \infty$ this tends to the value $\|\phi\|_2^2$, as is easily seen by first approximating ϕ by smooth functions. Hence $U_n \rightarrow U$ weakly in distribution in the space $H^{-1}(\mathbb{T}^1)$.

Moreover, we note that U_n has the same distribution as the random variable $T_n U$, where $T_n : H^{-1}(\mathbb{T}^1) \rightarrow H^{-1}(\mathbb{T}^1)$ is the linear operator $T_n v = \sum_{j=1}^n \langle v, \phi_j^n \rangle n^{1/2} \chi_{I(n,j)}$, where $(\phi_j^n)_{j=1}^n$ is any orthonormal sequence in $L^2(\mathbb{T}^1)$ consisting of elements of $H^1(\mathbb{T}^1)$. We have thus verified that the variables U_n are proper linear discretizations of U in the space $H^{-1}(\mathbb{T}^1)$, according to Definition 1.

Now the measurement $M^{(1)}$ is well defined and $M_n^{(1)}$ converge in distribution to $M^{(1)}$ as $n \rightarrow \infty$. However, we have $M_n^{(2)} \sim N(0, n)$ and thus measurements $M^{(2)}$ do not converge in distribution as $n \rightarrow \infty$. This is related to the fact that the white noise U is a well defined $H^{-1}(\mathbb{T}^1)$ valued Gaussian random variable, and the constant function 1 is in the dual of the space $H^{-1}(\mathbb{T}^1)$ but the point evaluation $u \mapsto u(\frac{1}{2})$ does not define a bounded linear operator in $H^{-1}(\mathbb{T}^1)$.

Remark 2. In Example 2 above one may verify that it is possible to choose the $L^2(\mathbb{T}^1)$ -orthonormal sequence $(\phi_j^n)_{j=1}^n$ so that the norm $\|T_n\|_{H^{-1}(\mathbb{T}^1) \rightarrow H^{-1}(\mathbb{T}^1)}$ remains uniformly bounded for all n . However, this choice is somewhat complicated. A way to construct discretizations with this property (and such that they fall in the scope of the basic theory developed in Section 3) is to apply suitable finite dimensional approximations of identity that are uniformly bounded simultaneously on both $H^{-1}(\mathbb{T}^1)$ and $L^2(\mathbb{T}^1)$. E.g., one may truncate Fourier series or apply basis

projections corresponding to a wavelet basis (set $p = 2$ in Section 6). Details of these comments will be considered elsewhere. Similar remarks apply to Example 3 below.

Example 3. (“Discretization of the Gaussian smoothness prior”) Choose continuous functions $\eta_j^n : \mathbb{T}^1 \rightarrow \mathbb{R}$ so that they are affine on intervals $I(n, j)$, $j = 1, \dots, n$ and that $(\eta_j^n)_{j=1}^n$ are orthonormal in $H^1(\mathbb{T}^1)$. Let

$$(74) \quad U_n(t) = \sum_{j=1}^n b_j^n X_j^n \eta_j^n(t),$$

where $b_j^n > 0$ are parameters. We choose $b_j^n = 1$. Then, for $\phi \in L^2(\mathbb{T}^1)$

$$(75) \quad \lim_{n \rightarrow \infty} \int_{\mathbb{T}^1} U_n(t) \phi(t) dt = \int_{\mathbb{T}^1} U(t) \phi(t) dt \quad \text{in distribution,}$$

where U is a Gaussian random variable in $L^2(\mathbb{T}^1)$ having zero expectation and covariance operator $(I - \Delta)^{-1}$, that is, U is the one-dimensional Gaussian smoothness prior. Let Q_n be the $H^1(\mathbb{T}^1)$ -orthogonal projection onto the subspace Y_n spanned by η_j^n , $j = 1, 2, \dots, n$. Analogously to Example 2, one can see that U_n have the same distribution as random variables $T_n U$ where where $T_n : L^2(\mathbb{T}^1) \rightarrow L^2(\mathbb{T}^1)$ is the linear operator

$$T_n v = \sum_{j=1}^n \langle v, (I - \Delta) \phi_j^n \rangle \eta_j^n,$$

where $(\phi_j^n)_{j=1}^n$ is any orthonormal sequence in $H^1(\mathbb{T}^1)$ consisting of elements of $H^2(\mathbb{T}^1)$. Thus U_n are proper linear discretizations of U in $L^2(\mathbb{T}^1)$.

Let us next consider $\phi \in H^{-1}(\mathbb{T}^1)$. Due to the formula (74), the $(H^1(\mathbb{T}^1) \times H^{-1}(\mathbb{T}^1))$ -duality $\langle U_n(\omega), \phi \rangle$ defines a real-valued Gaussian random variable with covariance

$$(76) \quad \begin{aligned} \mathbb{E}(\langle U_n, \phi \rangle^2) &= \mathbb{E} \left(\sum_{j,k=1}^n b_j^n X_j b_k^n X_k \langle \eta_j^n, \phi \rangle \langle \eta_k^n, \phi \rangle \right) \\ &= \sum_{j=1}^n \langle \eta_j^n, \phi \rangle^2 \\ &= \sum_{j=1}^n \langle \eta_j^n, (I - \Delta)^{-1} \phi \rangle_{H^1(\mathbb{T}^1)}^2 \\ &= \|Q_n (I - \Delta)^{-1} \phi\|_{H^1(\mathbb{T}^1)}^2. \end{aligned}$$

The kernel of the covariance of operator of U in $L^2(\mathbb{T}^1)$, that is, the function

$$G(t, t') = \mathbb{E}(U(t)U(t')), \quad t, t' \in \mathbb{T}^1$$

is Green’s function of the differential operator $-\frac{d^2}{dt^2} + 1$,

$$\left(-\frac{d^2}{dt^2} + 1\right)G(t, t') = \delta(t - t').$$

This implies that $\mathbb{E}(U(t)U(t'))$ is Lipschitz smooth on $\mathbb{T}^1 \times \mathbb{T}^1$. As U is Gaussian, one can see using e.g. [35, Theorem 3.23] that the values of the Gaussian smoothness prior U are almost surely in any Hölder space $C^\alpha(\mathbb{T}^1)$ with $\alpha < 1/2$. Thus, after

fixing $\alpha \in (0, \frac{1}{2})$, we can consider U also as a Gaussian $C^\alpha(\mathbb{T}^1)$ -valued random variable satisfying

$$(77) \quad \mathbb{E}(\langle U, \phi \rangle^2) = \|(I - \Delta)^{-1} \phi\|_{H^1(\mathbb{T}^1)}^2$$

for every ϕ in the dual space of $C^\alpha(\mathbb{T}^1)$. Note that as $H^1(\mathbb{T}^1) \subset C^\alpha(\mathbb{T}^1)$, we have $(C^\alpha(\mathbb{T}^1))' \subset H^{-1}(\mathbb{T}^1)$. Thus, since the projectors Q_n converge strongly to identity in $H^1(\mathbb{T}^1)$ as $n \rightarrow \infty$, we can use (76), (77), and the fact that the ranges of the operator T_n used above are in $H^1(\mathbb{T}^1) \subset C^\alpha(\mathbb{T}^1)$, and infer that that U_n are proper linear discretizations of U in $C^\alpha(\mathbb{T}^1)$, too. Because of this the measurements $M^{(1)}$ and $M^{(2)}$ are well defined and one can see that the measurements $M_n^{(1)}$ and $M_n^{(2)}$ converge in distribution to $M^{(1)}$ and $M^{(2)}$, respectively, as $n \rightarrow \infty$.

Example 4. (“Discrete total variation priors”) Let us next consider an example on interval $I = [0, 1]$ (i.e., the end points are not identified) and let $\theta_j^n : [0, 1] \rightarrow \mathbb{R}$ be continuous functions with are affine functions on intervals $I(n, j)$, $j = 1, \dots, n$, vanishing at $t = 0$ and $t = 1$ such that θ_j^n , $j = 1, 2, \dots, n - 1$ are orthonormal in $H^1(I)$. Let

$$(78) \quad U_n(t) = \sum_{j=1}^{n-1} Z_j^n \eta_j^n(t),$$

where $Z^n = (Z_1^n, Z_2^n, \dots, Z_{n-1}^n)$ is a \mathbb{R}^n valued random variable having the probability density function

$$\pi_{Z^n}(z_0, \dots, z_{n-1}) = c_n \exp \left(-a_n \left\| \frac{d}{dt} \left(\sum_{j=1}^{n-1} z_j \eta_j^n(t) \right) \right\|_{L^1(I)} \right)$$

where $a_n > 0$ is a parameter and c_n is a normalization constant of the probability density function. The distribution of the random variables U_n are sometimes called the discrete total variation prior. By [39], the random variables U_n converge in distribution if $a_n = n^{1/2}$ but then the limit U is a *Gaussian* random variable. As a Gaussian distribution stays Gaussian in a linear transformation, we see that in this example U_n are not proper linear discretizations of U in any Banach space.

Example 5. (“Discretization of the two-dimensional Gaussian smoothness prior”) Let us consider also higher dimensional example analogous to Example 3. Let $L(n, j, k, 1), L(n, j, k, 2) \subset \mathbb{T}^2$ be disjoint triangles so that their union is the square $I(n, j) \times I(n, k) \subset \mathbb{T}^2$, $j, k = 1, \dots, n$. Let $\zeta_{j,k,l}^n : \mathbb{T}^2 \rightarrow \mathbb{R}$, $j, k = 1 \dots, n$, $l = 1, 2$ be continuous functions on \mathbb{T}^2 which are affine functions on triangles $L(n, j, k, l)$ such that ζ_{jkl}^n , $j, k = 1, 2, \dots, n$, $l = 1, 2$ are orthonormal in $H^1(\mathbb{T}^2)$. Let

$$(79) \quad U_n(t_1, t_2) = \sum_{j,k=1}^n \sum_{l=1}^2 b_{jkl}^n X_{jkl}^n \zeta_{jkl}^n(t_1, t_2), \quad (t_1, t_2) \in \mathbb{T}^2$$

where $X_{jkl}^n \sim N(0, 1)$ are independent and $b_{jkl}^n > 0$ are parameters. Let $b_{jkl}^n = 1$. Then the U_n can be considered as Gaussian random variables in $H^{-1}(\mathbb{T}^1)$ that converge weakly in distribution to a $H^{-1}(\mathbb{T}^1)$ valued Gaussian random variable U having zero expectation and the covariance operator $(I - \Delta)^{-2}$ in $H^{-1}(\mathbb{T}^1)$. This means that U is the two-dimensional Gaussian smoothness prior.

Let now $u \in L^2(\mathbb{T}^1)$. We consider two measurements

$$A^{(1)}u = \int_{\mathbb{T}^2} u(t) dt, \quad A^{(2)}u = u\left(\frac{1}{2}, \frac{1}{2}\right)$$

and define models $M_n^{(1)}, M^{(1)}$ and $M_n^{(2)}, M^{(2)}$ as in (70) where \mathcal{E} is Gaussian white noise in $L^2(\mathbb{T}^2)$. Then, the measurement $M^{(1)}$ is well defined and the measurements $M_n^{(1)}$ converge in distribution to $M^{(1)}$. However, one can show that $M_n^{(2)} \sim N(0, \sigma_n^2)$, where $\sigma_n \rightarrow \infty$ as $n \rightarrow \infty$ and we see that the measurements $M_n^{(2)}$ do not converge in distribution as $n \rightarrow \infty$. In this example one can verify that the U_n are proper linear discretizations of U in $H^{-1}(\mathbb{T}^2)$.

The fact the point value measurements $M_n^{(2)}$ do not converge is related to the fact that the two-dimensional Gaussian smoothness prior has, formally speaking, the covariance function $\mathbb{E}(U(t)U(s)) = G(t, s)$, $t, s \in \mathbb{T}^2$ is the Green's function of the operator $-\Delta + I$ and has thus on the diagonal $t = s$ a logarithmic singularity. This fact is extensively used in quantum field theory in the study of the free Gaussian field, a random field that is very similar to the Gaussian smoothness prior [58].

Appendix C. On the domain of reconstructors. Considering formula (28) in the infinite-dimensional case, we meet the difficulty that realizations of M belong to Z only with probability zero. Therefore the function $m \mapsto \mathcal{R}_M(U|m)$ should be defined in some larger set than Z .

A generalized definition of a reconstructor is as follows:

Definition 5. The deterministic function $\mathcal{R}_M(U|\cdot) : S_0 \rightarrow Y$, where $m \mapsto \mathcal{R}_M(U|m)$, defined in a Borel-measurable subspace $S_0 \subset S$ is *reconstructor* of U with measurement M if $M(\omega) \in S_0$ almost surely and

$$(80) \quad \mathbb{E}(U|\mathcal{M})(\omega) = \mathcal{R}_M(U|M(\omega)) \quad \text{almost surely.}$$

The quantity $E(g(U)|\cdot) : S_0 \rightarrow \tilde{Y}$ is defined analogously.

For the domain of $\mathcal{R}_M(U|m)$ we could consider any of the non-trivial Borel-measurable subspaces $L \subset S$, such that the realization $M(\omega)$ belongs to L almost surely. Given any two such subspaces L_1 and L_2 , the value of the function $\mathcal{R}_M(U|m)$ can be changed in $L_1 \setminus L_2$, without contradicting the property (80). It is tempting to choose the domain to be the intersection of all such subspaces $L \subset S$, but unfortunately, this intersection is the set Z where the realizations of M lie only with probability zero. It appears to be hard to pick a candidate for a smallest space S_0 where the reconstructor $\mathcal{R}_M(U|m)$ should be defined.

Because of the above difficulties, we restricted ourselves to the case where the operator A maps $A : S \rightarrow S^1$ and a condition analogous to (22) is valid so that $m \mapsto \mathcal{R}_M(U|m)$ can be defined in the whole space S .

REFERENCES

- [1] F. Abramovich, T. Sapatinas and B. W. Silverman, *Wavelet thresholding via a Bayesian approach*, J. R. Stat. Soc. Ser. B Stat. Methodol., **60** (1998), 725–749.
- [2] F. Abramovich and T. B. Sapatinas, *Bayesian approach to wavelet decomposition and shrinkage*, in “Bayesian Inference in Wavelet-Based Models,” 33–50, Lecture Notes in Statist., **141**, Springer, New York, 1999.
- [3] B. D’Ambrogio, S. Mäenpää and M. Markkanen, *Discretization independent retrieval of atmospheric ozone profile*, Geophysica, **35** (1999), 87–99.
- [4] K. E. Andersen, S. P. Brooks and M. B. Hansen, *Bayesian inversion of geoelectrical resistivity data*, J. R. Statist. Soc. B, **65** (2003), 619–642.

- [5] K. E. Andersen, S. P. Brooks and M. B. Hansen, *A Bayesian approach to crack detection in electrically conducting media*, Inverse Problems, **17** (2001), 121–136.
- [6] X. L. Battle, G. S. Cunningham and K. M. Hanson, *3D Tomographic reconstruction using geometrical models*, in “Medical Imaging: Image Processing,” K. M. Hanson, ed., Proc. SPIE, **3034** (1997), 346–357.
- [7] J. Besag, J. York and A. Mollié, *Bayesian image restoration, with two applications in spatial statistics*, Ann. Inst. Statist. Math., **43** (1991), 1–59.
- [8] J. M. Bioucas-Dias, *Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of Heavy-Tailed priors*, IEEE Tr Image Processing, **15** (2006), 937–951.
- [9] N. Bissantz, T. Hohage and A. Munk, *Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise*, Inverse Problems, **20** (2004), 1773–1789.
- [10] V. I. Bogachev, “Gaussian Measures,” Mathematical Surveys and Monographs, **62**, American Mathematical Society, Providence, RI, 1998.
- [11] L. Borcea, V. Druskin and L. Knizhnerman, *On the continuum limit of a discrete inverse spectral problem on optimal finite difference grids*, Communications on Pure and Applied Mathematics, **58** (2005), 1231–1279.
- [12] D. Calvetti and E. Somersalo, *A Gaussian hypermodel to recover blocky objects*, Inverse Problems, **23** (2007), 733–754.
- [13] D. Calvetti and E. Somersalo, “Introduction to Bayesian Scientific Computing — Ten Lectures on Subjective Computing,” Springer, New York, 2007.
- [14] H. Choi and R. Baraniuk, *Multiple basis wavelet denoising using Besov projections*, Proceedings of ICIP 99, Kobe, Japan, (1999), 595–599.
- [15] J. A. Christen and C. Fox, *Markov Chain Monte Carlo using an approximation*, Journal of Computational & Graphical Statistics, **14** (2005), 795–810.
- [16] I. Daubechies, “Ten Lectures on Wavelets,” SIAM 2006.
- [17] I. Daubechies, M. Defrise and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., **57** (2004), 1413–1457.
- [18] J. Diestel and J. J. Uhl, Jr., “Vector Measures,” Mathematical Surveys, **15**, American Mathematical Society, 1977.
- [19] N. Dinculeanu, “Vector Integration and Stochastic Integration in Banach Spaces,” Wiley-Interscience, New York, 2000.
- [20] D. C. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., **56** (1996), 1181–1198.
- [21] D. L. Donoho and Xiaoming Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE transactions on information theory, **47** (2001), 2845–2862.
- [22] R. M. Dudley, “Real Analysis and Probability,” The Wadsworth & Brooks/Cole Mathematics Series, Pacific Grove, CA, 1989.
- [23] R. Hari and O. Lounasmaa, *Recording and interpretation of cerebral magnetic fields*, Science, **244** (1989), 432–436.
- [24] Y. Efendiev, T. Hou and W. Luo, *Preconditioning Markov chain Monte Carlo simulations using coarse-scale models*, SIAM J. Sci. Comput. **28** (2006), 776–803
- [25] H. Engl, A. Hofinger and S. Kindermann, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems, **21** (2005), 399–412.
- [26] H. W. Engl and M. Z. Nashed, *Generalized inverses of random linear operators in Banach spaces*, J. Math. Anal. Appl., **83** (1981), 582–610.
- [27] S. N. Evans and P. B. Stark, *Inverse problems as statistics*, Inverse Problems, **18** (2002), R55–R97.
- [28] B. G. Fitzpatrick, *Bayesian analysis in inverse problems*, Inverse Problems, **7** (1991), 675–702.
- [29] J. N. Franklin, *Well-posed stochastic extensions of ill-posed linear problems*, Journal of Mathematical Analysis and Applications, **31** (1970), 682–716.
- [30] H. Haario, M. Laine, M. Lehtinen, S. Saksman and J. Tamminen, *MCMC methods for high dimensional inversion in remote sensing*, J. Roy. Statist. Soc. Ser. B., **66** (2004), 591–607.
- [31] J. P. Kaipio, V. Kolehmainen, E. Somersalo and M. Vauhkonen, *Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography*, Inverse Problems, **16** (2000), 1487–1522.
- [32] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen and E. Somersalo, *Inverse problems with structural prior information*, Inverse Problems, **15** (1999), 713–729.

- [33] J. P. Kaipio and E. Somersalo, “Statistical and Computational Inverse Problems,” Applied Mathematical Sciences **160**, Springer-Verlag, New York, 2005.
- [34] M. Kalke, S. Siltanen, S. Vänskä, M. Lassas and M. Rantala, “Method and Arrangement for Multiresolution Reconstruction for Medical X-ray Imaging,” United States patent 7215730, 2007.
- [35] O. Kallenberg, “Foundations of Modern Probability Theory”, Springer, 1997.
- iv
- [36] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä and E. Somersalo, *Statistical inversion for X-ray tomography with few radiographs II: Application to dental radiology*, Physics in Medicine and Biology, **48** (2003), 1465–1490.
- [37] S. Lasanen, “Discretizations of Generalized Random Variables with Applications to Inverse Problems,” Dissertation, University of Oulu. Ann. Acad. Sci. Fenn. Math. Diss. No. 130, 2002.
- [38] S. Lasanen and L. Roininen, *Statistical inversion with Green’s priors*, Proceedings of the 5th International Conference on Inverse Problems in Engineering: Theory and Practice, Cambridge, UK, 11–15th July 2005.
- [39] M. Lassas and S. Siltanen, *Can one use total variation prior for edge preserving Bayesian inversion?* Inverse Problems, **20** (2004), 1537–1564.
- [40] M. Lehtinen, L. Päärinta and E. Somersalo, *Linear inverse problems for generalised random variables*, Inverse Problems, **5** (1989), 599–612.
- [41] D. Leporini and J.-C. Pesquet, *Bayesian wavelet denoising: Besov priors and non-Gaussian noises*, Signal Processing, **81** (2001), 55–67.
- [42] I. Loris, G. Nolet, I. Daubechies and F. A. Dahlen, *Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients*, (2006), [arXiv:physics/0608094v1](https://arxiv.org/abs/physics/0608094v1).
- [43] T. Lunttila, J. Nenonen and E. Somersalo, *Regularization in cardiac source imaging*, in I. Magnin et al. (Eds.): FIMH 2003, LNCS 2674, 101–111, 2003, Springer-Verlag Berlin Heidelberg, 2003.
- [44] H. Luschgy, *Linear estimators and radonifying operators*, Teor. Veroyatnost. i Primenen **40**, translation in Theory Probab. Appl., **40** (1995), 167–175.
- [45] A. Mandelbaum, *Linear estimators and measurable linear transformations on a Hilbert space*, Zeitung für Wahrscheinlichkeitstheorie und verwandte Gebiete, **65** (1984), 385–397.
- [46] M. Markkanen, M. Lehtinen, T. Nygrén, J. Pirttilä, P. Henelius, E. Vilenius, E. D. Tereshchenko and B. Z. Khudukon, *Bayesian approach to satellite radiotomography with applications in the Scandinavian sector*, Ann. Geophysicae **13** (1995), 1277–1287.
- [47] Y. Meyer, “Oscillating Patterns in Image Processing and Nonlinear Evolution Equations,” University lecture series, **22**, AMS, 2001.
- [48] K. Mosegaard and M. Sambridge, *Monte Carlo analysis of inverse problems*, Inverse Problems, **18** (2002), R29–R54.
- [49] P. Mueller and B. Vidakovic (Eds.), “Bayesian Inference in Wavelet-Based Models,” Lecture Notes in Statistics, Springer, New York, 1999.
- [50] G. Nicholls, *Bayesian image analysis with Markov chain Monte Carlo and coloured continuum triangulation models*, J. R. Statist. Soc. B, **60** (1998), 643–659.
- [51] K. Niinimäki, S. Siltanen and V. Kolehmainen, *Bayesian multiresolution method for local tomography in dental x-ray imaging*, Phys. Med. Biol., **52** (2007), 6663–6678.
- [52] T. Nygrén, M. Markkanen, M. Lehtinen, E. D. Tereshchenko and B. Z. Khudukon, *Stochastic inversion in ionospheric radiotomography*, Radio Science, **32** (1997), 2359–2372.
- [53] P. Piiroinen, *Statistical measurements, experiments and applications*, Ann. Acad. Sci. Fenn. Math. Diss., (2005).
- [54] H. K. Pikkariainen, *Stable estimation approach to nonstationary inverse problems: discretization error and filtering problem*, Inverse Problems, **22** (2006), 365–379.
- [55] M. Rantala, S. Vänskä, S. Järvenpää, M. Kalke, M. Lassas, J. Moberg and S. Siltanen, *Wavelet-based reconstruction for limited-angle X-ray tomography*, IEEE Tr. Medical Imaging, **25** (2006), 210–217.
- [56] Yu. A. Rozanov, “Markov Random Fields,” Applications of Mathematics, Springer-Verlag, 1982.
- [57] L. I. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, **60** (1992), 259–268.
- [58] S. Sheffield, *Gaussian free fields for mathematicians*, Probab. Theory Related Fields, **139** (2007), 521–541.
- [59] A. N. Shiryaev, “Probability,” Springer-Verlag, 1996.

- [60] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä and E. Somersalo, *Statistical inversion for X-ray tomography with few radiographs I: General theory*, Physics in Medicine and Biology, **48** (2003), 1437–1463.
- [61] A. S. Stern, D. L. Donoho and J. C. Hoch, *NMR data processing using iterative thresholding and minimum ℓ_1 -norm reconstruction*, Journal of Magnetic Resonance, **188** (2007), 295–300.
- [62] J. Tamminen, *MCMC methods for inverse problems*, Geophysical publications, **48** (1999), Finnish Meteorological Institute.
- [63] A. Tarantola, “Inverse Problem Theory and Methods for Model Parameter Estimation,” SIAM 2005.
- [64] C. R. Vogel and M. Oman, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Trans. Image Process, **7** (1998), 813–824.
- [65] 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.
- [66] C. R. Vogel, *Stochastic inversion of linear first kind equations II. Discrete theory and convergence results*, Journal of Integral Equations, **7** (1984), 73–92.
- [67] M. Wilhelm, *Fubini-type theorems for integrable spaces*, Bull. Acad. Polon. Sci. Ser. Sci. Math-Astronom. Phys., **22** (1974), 257–261.

Received February 2008; revised November 2008.

E-mail address: Matti.Lassas@helsinki.fi

E-mail address: Eero.Saksman@helsinki.fi

E-mail address: Samuli.Siltanen@tut.fi