SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 5, pp. 1024–1060

Well-Posed Bayesian Inverse Problems with Infinitely Divisible and Heavy-Tailed Prior Measures*

Bamdad Hosseini†

Abstract. We present a new class of prior measures based on the generalized Gamma distribution that are closely related to ℓ_p-regularization techniques when $p \in (0, 1)$. Furthermore, we use the laws of pure jump Lévy processes in order to define new classes of prior measures that are concentrated on the space of functions with bounded variation. These priors serve as an alternative to the classic total variation prior and result in well-defined inverse problems. Some of these prior measures are heavy-tailed, nonconvex, and infinitely divisible. Motivated by this observation we study the class of infinitely divisible prior measures and draw a connection between their tail behavior and the tail behavior of their Lévy measures. We then study the well-posedness of Bayesian inverse problems in a general enough setting that encompasses the above-mentioned classes of prior measures. We establish that well-posedness relies on a balance between the growth of the log-likelihood function and the tail behavior of the prior and apply our results to special cases such as additive noise models and linear problems. Finally, we discuss some of the practical aspects of Bayesian inverse problems such as their consistent approximation and present three concrete examples of well-posed Bayesian inverse problems with heavy-tailed or stochastic process prior measures.

Key words. inverse problems, Bayesian, infinitely divisible, non-Gaussian, bounded variation

AMS subject classifications. 35R30, 62F99, 60B11

DOI. 10.1137/16M1096372

1. Introduction.

Gaussian prior measures are perhaps the most commonly used class of priors in infinite-dimensional Bayesian inverse problems. While the Gaussian class is very convenient to use in both theory and practice, it has serious shortcomings in certain types of inverse problems, such as estimation of sparse parameters (see [54, Chap. 4] or [42]). In this article we introduce some non-Gaussian prior measures that can overcome some of these shortcomings of the Gaussian class. We will discuss our goals in more detail after a brief introduction to the Bayesian framework for solution of inverse problems.

Consider the problem of estimating a parameter $u \in X$ from a set of measurements $y \in Y$ where both X and Y are Banach spaces and y is associated with u through a model of the form

$$y = \tilde{\mathcal{G}}(u). \tag{1.1}$$

Here $\tilde{\mathcal{G}}$ is a generic stochastic mapping called the *forward map* that models the relationship between the parameter and the observed data by taking the measurement noise into account

*Received by the editors September 29, 2016; accepted for publication (in revised form) July 27, 2017; published electronically October 10, 2017.
http://www.siam.org/journals/juq/5/M109637.html
Funding: This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.
†Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada (bhossein@sfu.ca).

(be it additive, multiplicative, etc.). As an example, if the measurement noise is additive, then we can write

$$\tilde{\mathcal{G}}(u) = \mathcal{G}(u) + \eta,$$

where $\mathcal{G} : X \mapsto Y$ is the (deterministic) *forward model* and η is the (random) measurement noise which is independent of u. We want to estimate the parameter u given a realization of y. Since the map $\tilde{\mathcal{G}}$ may not be stably invertible, this problem is in general ill-posed.

In this paper, we consider the Bayesian framework for solution of such ill-posed problems. Recall the infinite-dimensional version of *Bayes' rule* [51] which is understood in the sense of the Radon–Nikodym theorem [8, Thm. 3.2.2]:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\Phi(u; y)\right), \quad \text{where} \quad Z(y) = \int_X \exp(-\Phi(u; y))\mathrm{d}\mu_0(u). \tag{1.2}$$

Here μ_0 is the *prior measure* which reflects our prior knowledge of the parameter u, $\Phi(u; y)$ is the *likelihood potential* that can be thought of as the negative log of the density of the data conditioned on the parameter u, and μ^y is the posterior measure on u. The posterior μ^y is an updated version of the prior μ_0 that is informed by the data y.

The Bayesian approach has attracted a lot of attention in the last two decades [13, 36, 51]. Here, the unknown parameter u is modelled as a random variable, and our goal is to obtain a probability distribution μ^y on u that is informed by the data y and our prior knowledge about u (modelled by the measure μ_0). We can generate samples from the posterior μ^y, and if this measure is concentrated around the true value of the parameter, the sample mean or median will be a good estimator of the true value of the parameter.

The Bayesian approach is well established in the statistics literature [14, 5], where it is often applied in the setting where X, Y are finite-dimensional spaces. Here we take X to be an infinite-dimensional Banach space which is not necessarily separable, motivated by applications where the parameter u belongs to a function space such as L^2 or BV (the space of functions with bounded variation). Such problems arise when the forward map involves the solution of a partial differential equation (PDE) or an integral equation such as the examples in section 5.

In practice we solve these problems by discretizing the forward model and approximating the infinite-dimensional posterior measure with a finite-dimensional one. An important task is to ensure that the finite-dimensional approximation to the posterior measure remains consistent with the infinite-dimensional posterior measure. For example, we require that the finite-dimensional posterior converge to the (true) infinite-dimensional measure in the limit when the discretization is infinitely fine. Ensuring this consistency is a delicate task. An example of an inconsistent discretization of an infinite-dimensional inverse problem was studied in [39], where the authors demonstrated that the total variation prior loses its edge preserving properties in the limit of fine discretizations. To resolve this issue we study the infinite-dimensional inverse problem before constructing the discrete approximations.

In this paper we set out to achieve the following goals:

G1. *Present a systematic study of the class of infinitely divisible prior measures, and construct a new class of infinitely divisible prior measures for recovery of compressible parameters.*

G2. *Introduce an alternative to the classic total variation prior using the laws of pure jump Lévy processes that are well defined in infinite dimensions.*

G3. *Present a theory of well-posedness for Bayesian inverse problems that encompasses the prior measures introduced under* G1 *and* G2.

Let us motivate some of these goals with an example.

Example 1.1. Suppose $u \in \mathbb{R}^n$ and the data $y \in \mathbb{R}^m$ is generated via the model

$$y = \mathbf{A}u + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma^2 \mathbf{I}),$$

where $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\sigma > 0$ is fixed, and $\mathbf{I}$ is the $m \times m$ identity matrix. We wish to estimate u given y. Here we are taking $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, and the forward map has the form $\mathcal{G}(u) = \mathbf{A}u$. Since η has a Gaussian density, we can write the likelihood potential $\Phi(u; y)$ as

$$\Phi(u; y) = \frac{1}{2\sigma^2} \|\mathbf{A}u - y\|_2^2.$$

Then, Bayes' rule gives

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{A}u - y\|_2^2 \right).$$

Now define the prior measure via

$$\frac{\mathrm{d}\mu_0}{\mathrm{d}\Lambda}(u) = \frac{1}{U} \exp\left(-\|u\|_p^p \right), \tag{1.3}$$

where $\mathrm{d}\Lambda$ denotes the Lebesgue measure on $\mathbb{R}^n$, $\|\cdot\|_p$ denotes the usual ℓ_p (quasi-)norm in $\mathbb{R}^n$ for $p > 0$, and U is the appropriate normalizing constant. Then the posterior μ^y can be identified via its Lebesgue density as

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\Lambda}(u) = \frac{1}{Z(y)} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{A}u - y\|_2^2 - \|u\|_p^p \right). \tag{1.4}$$

The maximizer of the posterior density is referred to as the maximum a posteriori (MAP) estimate. Formally, the MAP estimate of the posterior in (1.4) is given by

$$u_{\mathrm{MAP}} = \operatorname*{arg\,min}_{z \in \mathbb{R}^n} \left\{ \frac{1}{2\sigma^2} \|\mathbf{A}z - y\|_2^2 + \|z\|_p^p \right\}.$$

For $p \geq 1$ this optimization problem is convex and can be solved efficiently. Taking $p = 1$ results in the well-known ℓ_1-regularization technique which is often used in the recovery of sparse parameters. For values of $p \in (0, 1)$ the resulting optimization problem is no longer convex, but it is a good model for compressibility [26, 42].

It is straightforward to check that the prior distribution (1.3) for $p \in (0, 1)$ is nonconvex and heavy-tailed (i.e., its tails are not exponentially bounded). However, we will see that this measure belongs to the much larger class of infinitely divisible measures. A random variable ξ is infinitely divisible if for every $n \in \mathbb{N}$ its law coincides with the law of $\sum_{k=1}^n \xi_k^{1/n}$, where $\{\xi_k^{1/n}\}$ are i.i.d. random variables. Thus, the above example motivates our interest in infinitely divisible prior measures (goal G1) that are introduced in section 2. The connection between sparse recovery and heavy-tailed or infinitely divisible priors has been observed in the literature. Unser and Tafti [54] and Unser and colleagues [56, 55] study the sparse behavior of stochastic processes that are driven by infinitely divisible force terms and advocate their use in solution of inverse problems. A detailed discussion of some heavy-tailed prior distributions such as generalizations of the Student's t-distribution and the ℓ_p-priors can also be found in the dissertation [42]. Moreover, Polson and Scott [47] and Carvalho, Polson, and Scott [16] propose a class of hierarchical horseshoe priors that are tailored to the recovery of sparse signals.

Later on we see that the well-posedness of a Bayesian inverse problem relies on the type of prior measure μ_0 that is chosen during the modelling step, as well as certain properties of the potential Φ. Well-posed Bayesian inverse problems were studied in [51, 20] with Gaussian prior measures, in [22] with Besov priors, in [34] with convex prior measures, and in [23, 52] with heavy-tailed priors on separable Banach spaces. We note that our well-posedness results in this paper are closely related to those of [23]. The main difference is that our theory does not rely on the assumption that the parameter space X is separable, and we impose slightly different conditions on the potential Φ. The nonseparability condition is particularly interesting when one takes X to be C^α (the space of Hölder continuous functions) or BV, neither of which is separable. In section 3 we introduce a class of prior measures that are concentrated on BV and have piecewise constant samples (goal G2). This example is later used in section 5 as a prior measure in a deblurring problem.

1.1. Key definitions and notation. We gather here some key definitions and assumptions that are used in the remainder of the article. We let $\mathbb{R}_+$ denote the positive real line $[0,\infty)$, and we use the shorthand notation $a \lesssim b$ when a and b are real valued functions and there exists an independent constant $C > 0$ such that $a \le Cb$. Given two random variables ξ and ζ, we use the notation $\xi \overset{d}{=} \zeta$ to denote that they have the same laws (or distributions).

We use the shorthand notation $\{\gamma_k\}$ to denote a sequence of elements $\{\gamma_k\}_{k=1}^\infty$ in a vector space. The usual ℓ_p sequence spaces for $p \in [1,\infty]$ are defined as the space of real valued sequences $\{\gamma_k\}$ such that $\|\{\gamma_k\}\|_p < \infty$, where

$$\|\{\gamma_k\}\|_p := \left(\sum_{k=1}^\infty |\gamma_k|^p\right)^{1/p} \quad \text{if} \quad p \in [1,\infty) \qquad \text{and} \qquad \|\{\gamma_k\}\|_\infty := \sup_k |\gamma_k|.$$

Similarly, we define the $\|\cdot\|_p$ norms of finite-dimensional vectors. In particular, $\|\cdot\|_2$ will denote the usual Euclidean norm. Given a positive definite matrix $\mathbf{\Sigma}$ of size $m \times m$, we define the norm

$$\|x\|_{\mathbf{\Sigma}} := \|\Sigma^{-1/2} x\|_2 \quad \text{for} \quad x \in \mathbb{R}^m.$$

Throughout the paper we use Λ to denote the Lebesgue measure in finite dimensions. Given a Borel measure μ on a Banach space X, we define the spaces $L^p(X,\mu)$ for $p \in [1,\infty)$ as the space of μ-equivalent classes of functions $h : X \mapsto \mathbb{R}$ such that $|h|^p$ is μ-integrable. We also use the shorthand notation $L^p(X)$ instead of $L^p(X,\Lambda)$ whenever we are working with the

Lebesgue measure. Finally, if X is a Banach space, we use X^* to denote the topological dual of X, and $B_X(r)$ to denote the open ball of radius $r > 0$ in X that is centered at the origin. The shorthand notation B_X denotes the unit ball.

We shall consider the prior probability measure μ_0 to be in the class of Borel probability measures on X. In some instances we take the prior μ_0 to be an inner regular probability measure on the Borel sets of X, i.e., a Radon measure. This assumption often simplifies our analysis since Radon measures on Banach spaces are automatically concentrated on a separable and reflexive subspace (see Theorem 4.1). Whenever we say that μ is a probability measure on X we automatically mean that $\mu(X) = 1$. Throughout this paper we consider only complete probability measures. Recall that a probability measure μ on X is called complete if every subset of every μ-measure zero set is measurable. We recall that given a Borel probability measure μ on a Banach space X, its characteristic function $\hat{\mu} : X^* \mapsto \mathbb{C}$ is defined as

$$\hat{\mu}(\varrho) = \int_X \exp(i\varrho(u))\mathrm{d}\mu(u) \qquad \forall \varrho \in X^*.$$

Given a Borel probability measure μ, we use μ^{*n} to denote the n-fold convolution of μ with itself. Finally, we say a probability measure μ is heavy-tailed if the integral $\int_X \exp(\kappa\|u\|_X)\mathrm{d}\mu(u)$ is not finite for any value of $\kappa > 0$.

In this paper, we focus on the following notion of a well-posed Bayesian inverse problem.

Definition 1.2 (well-posedness). *Suppose that X is a Banach space and $d(\cdot,\cdot) \mapsto \mathbb{R}$ is a metric on the space of Borel probability measures on X. Then for a choice of the prior measure μ_0 and the likelihood potential Φ, the Bayesian inverse problem given by (1.2) is well-posed with respect to d if the following hold:*

1. *(Existence and uniqueness) There exists a unique posterior probability measure $\mu^y \ll \mu_0$ given by Bayes' rule (1.2).*
2. *(Stability) For every choice of $\epsilon > 0$ there exists a $\delta > 0$ so that $d(\mu^y, \mu^{y'}) \le \epsilon$ for all $y, y' \in Y$ so that $\|y - y'\|_Y \le \delta$.*

We will study the convergence of probability measures using the Hellinger and total variation metrics on the space of probability measures on X. For two probability measures μ_1 and μ_2 that are absolutely continuous with respect to a third measure ν on X, the total variation and Hellinger metrics are defined as

(1.5)
$$d_{TV}(\mu_1,\mu_2) := \frac{1}{2}\int_X \left|\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu} - \frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}\right| \mathrm{d}\nu \quad \text{and} \quad d_H(\mu_1,\mu_2) := \left(\frac{1}{2}\int_X \left(\sqrt{\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}}\right)^2 \mathrm{d}\nu\right)^{1/2}.$$

Both metrics are independent of the choice of the measure ν [8, Lem. 4.7.35]. Furthermore, convergence in one of these metrics implies convergence in the other, due to the following inequalities (see [8, Lem. 4.7.37] for a proof):

$$2d_H^2(\mu_1,\mu_2) \le d_{TV}(\mu_1,\mu_2) \le \sqrt{8}d_H(\mu_1,\mu_2). \tag{1.6}$$

However, one might prefer to work with the Hellinger metric as it relates directly to the error in expectation of certain functions. Suppose that $h \in L^2(X,\mu_1) \cap L^2(X,\mu_2)$. Then using the

Radon–Nikodym theorem and Hölder's inequality, one can show (see [34, sect. 1] for details)

$$\left| \int_X h(u) \mathrm{d}\mu_1(u) - \int_X h(u) \mathrm{d}\mu_2(u) \right| \le 2 \left(\int_X h^2(u) \mathrm{d}\mu_1 + \int_X h^2(u) \mathrm{d}\mu_2 \right)^{1/2} d_H(\mu_1, \mu_2). \tag{1.7}$$

For reasons that will become clear in section 4, we prefer to study the well-posedness of inverse problems using both the Hellinger and total variation metrics. The main difference is in the restrictions that we need to impose on the prior μ_0 to obtain a certain rate of convergence for each metric.

2. Infinitely divisible prior measures.

We start by presenting a generalization of the prior distribution (1.3) that was considered in Example 1.1. We show that this generalization belongs to a larger class of distributions that are closely related to ℓ_p-regularization techniques. We shall extend these distributions to measures on Banach spaces with an unconditional Schauder basis and observe that they belong to the much larger class of infinitely divisible (ID) measures (see Definition 2.8). Motivated by this connection between ℓ_p-regularization and ID priors, we turn our attention to the ID class and discuss some of its properties. In particular, we study the tail behavior of ID priors with respect to their Lévy measures (see Definition A.2).

2.1. A class of shrinkage priors with compressible samples.

When faced with the problem of recovering a sparse or compressible parameter we require the prior measure to reflect the intuition that the solution to the inverse problem is likely to have only a few large modes in some basis and the rest of the modes are negligible (see [42, sect. 6.1]). Such prior distributions are often referred to as "shrinkage priors," and they have been the subject of extensive research [47, 16, 28, 18, 17]. In this section we consider a few examples of shrinkage priors that are closely related to ℓ_p-regularization techniques.

Most of the existing literature on shrinkage priors is focused on finite-dimensional problems. We present an extension of these priors to infinite-dimensional Banach spaces. Since compressibility is often considered with respect to a basis, it makes sense for us to consider a parameter space X that has a basis.

Given a parameter space X, or at least a subspace $\tilde{X} \subseteq X$ that has an unconditional Schauder basis $\{x_k\}$, we construct random variables of the form

$$u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k, \tag{2.1}$$

where $\{\gamma_k\}$ is a fixed sequence of real valued coefficients that decay sufficiently fast and the $\{\xi_k\}$ are a sequence of independent real valued random variables that need not be identically distributed. We will take the prior measure μ_0 to be the law of the random variable u in (2.1). We refer to such a prior measure μ_0 as the *product prior* obtained from $\{\gamma_k\}$ and $\{\xi_k\}$. This construction of the prior is reminiscent of the Karhunen–Loéve expansion of Gaussian measures [7, Thm. 3.5.1]. The following theorem gives sufficient conditions that ensure $\|u\|_X < \infty$ a.s.

Theorem 2.1 (see [34, Thm. 3.9]). *Suppose that X is a Banach space with an unconditional Schauder basis, and let u be defined as in (2.1). If $\{\gamma_k^2\} \in \ell_p$ and $\{\mathbf{Var}\xi_k\} \in \ell^q$ for $1 < p, q <$*

∞ so that $1/p + 1/q = 1$ (with $p = 1$ for the limiting case when $q = \infty$), then $\|u\|_X < \infty$ a.s. In particular, if the $\{\xi_k\}$ are i.i.d., $\mathbf{Var}\xi_1 < \infty$*, and $\{\gamma_k\} \in \ell^2$, then $\|u\|_X < \infty$ a.s.*

We can also show that the product prior μ_0 that is induced by (2.1) is Radon. Proof of the next theorem follows the same approach as [34, Thm. 3.10(ii)] and is hence omitted.

Theorem 2.2. *Let μ be the probability measure that is induced by the random variable u given by (2.1), where $\{\gamma_k\}$ and $\{\xi_k\}$ satisfy the conditions of Theorem 2.1. Then μ is a Radon probability measure on X if the random variables $\{\xi_k\}$ are distributed according to Radon probability measures on $\mathbb{R}$.*

Before going further we present a result on the second raw moment of product priors which will be useful throughout the remainder of the paper.

Theorem 2.3. *Suppose that X is a Banach space with an unconditional Schauder basis $\{x_k\}$, and let μ be the product prior obtained from $\{\gamma_k\} \in \ell^2$ and $\{\xi_k\}$, where ξ_k are i.i.d. and* $\mathbf{Var}\xi_k < \infty$*. Then $\|\cdot\|_X \in L^2(X, \mu)$.*

Proof. Let $u_N = \sum_{k=1}^N \gamma_k \xi_k x_k$; then for $M > N > 0$ we have

$$\left| \int_X \|u_M\|_X^2 \, \mathrm{d}\mu - \int_X \|u_N\|_X^2 \, \mathrm{d}\mu \right| = \left| \int_X (\|u_M\|_X - \|u_N\|_X)(\|u_M\|_X + \|u_N\|_X) \, \mathrm{d}\mu \right|.$$

By Theorem 2.1 we know that $\|u\|_X < \infty$ a.s., and so in the limit as $M, N \to \infty$, $|(\|u_M\|_X + \|u_N\|_X)| \to 2\|u\|_X$ and $|(\|u_M\|_X - \|u_N\|_X)| \to 0$, and so $\{\|u_N\|_X^2\}$ is Cauchy in $L^2(X, \mu)$. ∎

We are now in a position to discuss a few examples of shrinkage priors. Motivated by Example 1.1, we define the class of ℓ_p-priors as follows.

Definition 2.4 (ℓ_p-prior). *Let X be a Banach space with an unconditional Schauder basis $\{x_k\}$; then we say that a Radon probability measure μ is an ℓ_p-prior on X if its samples can be expressed as $u = \sum_{k=1}^\infty \gamma_k \xi_k x_k$, where $\{\gamma_k\} \in \ell^2$ and $\{\xi_k\}$ is an i.i.d. sequence of real valued random variables with Lebesgue density*

$$\xi_k \sim \frac{p}{2\alpha\Gamma(1/p)} \exp\left(-\frac{|t|^p}{\alpha^p} \right) d\Lambda(t), \tag{2.2}$$

where $p \in (0, \infty)$ and $\alpha = \sqrt{\Gamma(1/p)/\Gamma(3/p)}$.

Here Γ denotes the usual Gamma function. The distribution in (2.2) belongs to the larger class of generalized normal distributions [44]. This class is also referred to as a Kotz-type distribution [43] or a generalized Laplace distribution [37]. Here we will use neither of these terms and simply refer to this distribution as the ℓ_p-distribution to emphasize its connection to ℓ_p-regularization techniques. The random variables ξ_k have bounded moments of all orders (see [44] or the discussions following the definition of the $G_{p,q}$-prior below); in fact,

$$\mathbb{E}\, \xi_k^s = \frac{\alpha^s(1 + (-1)^s)}{2\Gamma(1/p)} \Gamma\left(\frac{s+1}{p} \right) \quad \text{for} \quad s \in \mathbb{N}.$$

In particular, we have that $\mathbf{Var}\xi_k = 1$, and so it follows from Theorem 2.3 that the ℓ_p-prior has bounded second moments.

Another class of priors, closely related to the ℓ_p-priors, can be obtained by a symmetrization of the Weibull distribution.

Definition 2.5 (W_p-prior). *Let X be a Banach space with an unconditional Schauder basis $\{x_k\}$; then we say that a Radon probability measure μ is a W_p-prior on X if its samples can be expressed as $u = \sum_{k=1}^{\infty} \gamma_k \xi_k x_k$, where $\{\gamma_k\} \in \ell^2$ and $\{\xi_k\}$ is an i.i.d. sequence of real valued random variables with Lebesgue density,*

$$\xi_k \sim \frac{p}{\alpha}\left(\frac{|t|}{\alpha}\right)^{p-1} \exp\left(-\frac{|t|^p}{\alpha^p}\right) d\Lambda(t), \tag{2.3}$$

where $p \in (0,\infty)$ and $\alpha = (2\Gamma(1+2/p))^{-1/2}$.

The distribution of ξ_k is simply a symmetric version of the well-known Weibull distribution [35], hence the name W_p. A straightforward calculation shows that $\mathbf{Var}\xi_k = 1$, and once again it follows from Theorem 2.3 that the W_p-priors have bounded second moments.

Both the W_p- and ℓ_p-distributions reduce to the Laplace distribution when $p = 1$. For $p < 1$ the ℓ_p-distribution has nonconvex level sets and puts a large portion of its mass close to the axes (see Figure 1). This behavior becomes stronger for smaller p and suggests that the ℓ_p-prior will incorporate sparse behavior as $p \to 0$.

When $p < 1$, the W_p-distribution blows up at the origin and so is very different in comparison to the ℓ_p- or Laplace distributions (see Figure 1). This means that the W_p-distribution puts more of its mass at the origin, which leads us to believe that it must incorporate stronger compressibility than the ℓ_p-distribution.

Further intuition into the behavior of the W_p-prior can be obtained by considering its MAP point estimate in finite dimensions. Formally, using this prior in Example 1.1 gives rise to an optimization problem of the form

$$u_{\mathrm{MAP}} = \underset{z\in\mathbb{R}^n}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{A}z - y\|_2^2 + \|z\|_p^p + (1-p)\sum_{k=1}^{n} \log(|z_k|) \right\}.$$

Of course, the log term on the right-hand side is not bounded from below, and so we cannot solve this problem. However, we can consider a slightly modified version of this optimization problem by introducing a small parameter $\epsilon > 0$:

$$u_{\epsilon} = \underset{z\in\mathbb{R}^n}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{A}z - y\|_2^2 + \|z\|_p^p + (1-p)\sum_{k=1}^{n} \log(\epsilon + |z_k|) \right\}.$$

If ϵ is small, then the log term will heavily penalize any modes of the solution that are on a larger scale than that of ϵ, and so we expect that most of the modes of the solution u_ϵ will be on the scale of the small parameter ϵ. The stronger shrinkage of the posterior due to the W_p-prior is also evident in Figure 2, where we compare a prototypical example of posteriors that arise from the W_p- and ℓ_p-priors for solution of Example 1.1 in 2D. Here, we clearly see that the $W_{1/2}$-prior results in a posterior that is highly concentrated around the axes compared to the posterior that arises from the $\ell_{1/2}$-prior, which is more spread out. Note that

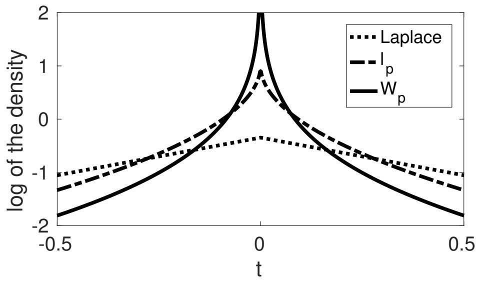

Figure 1. *Log of ℓ_p and W_p densities in 1D compared to the Laplace density.*

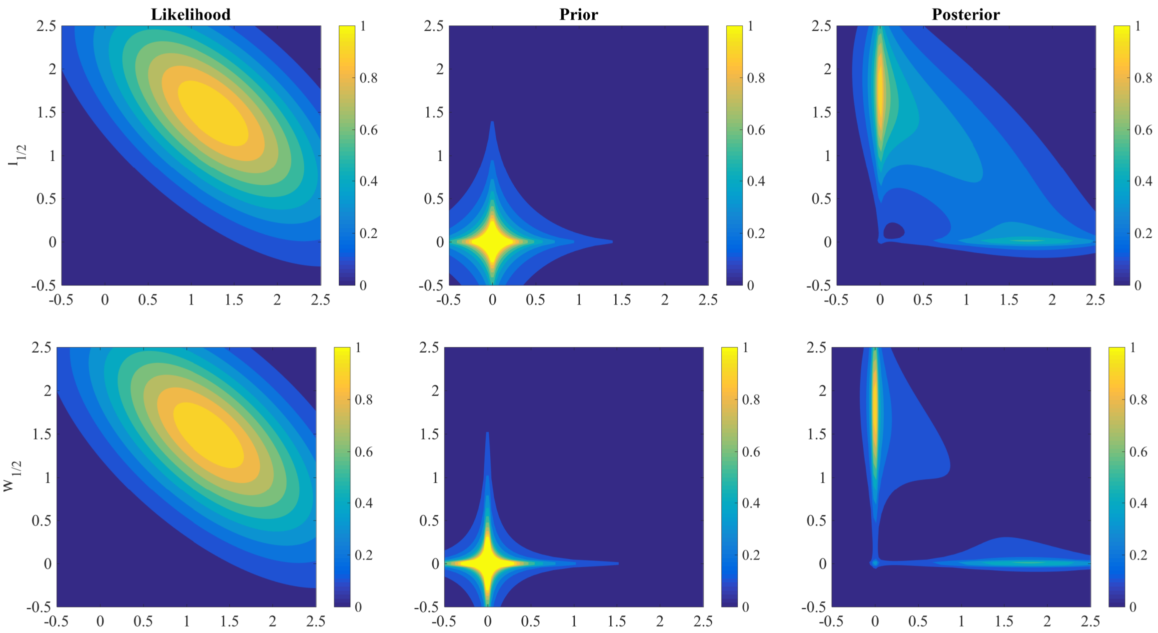

Figure 2. *A prototypical example of densities that arise in the solution of Example 1.1 in 2D with the $\ell_{1/2}$ (top row) and $W_{1/2}$-priors (bottom row). From left to right columns: The likelihood that arises from the additive Gaussian noise model, the prior densities, and the resulting posteriors. The densities are rescaled for better visualization.*

in either case, the posteriors are highly concentrated around the axes, meaning that the MAP estimates, as well as most of the samples from these posteriors, will incorporate sparsity.

Comparing the distributions (2.2) and (2.3) suggests the definition of a larger class of priors that can interpolate between the ℓ_p- and W_p-priors. To this end, we introduce a new class of prior measures called the $G_{p,q}$-priors. The letter G is chosen due to the connection of the one-dimensional version of these measures to the generalized Gamma distribution [10].

Definition 2.6 ($G_{p,q}$-prior). *Let X be a Banach space with an unconditional Schauder basis $\{x_k\}$; then we say that a Radon probability measure μ is a $G_{p,q}$-prior on X if its samples can be expressed as $u = \sum_{k=1}^{\infty} \gamma_k \xi_k x_k$ with $\{\gamma_k\} \in \ell^2$ and $\{\xi_k\}$ an i.i.d. sequence of real valued* *random variables with Lebesgue density,*

$$\xi_1 \sim \frac{p}{2\alpha\Gamma(q/p)} \left|\frac{t}{\alpha}\right|^{q-1} \exp\left(-\left|\frac{t}{\alpha}\right|^{p}\right) d\Lambda(t), \tag{2.4}$$

where $p \in (0,\infty)$ *and* $\alpha = (\Gamma(q/p)/\Gamma((2+q)/p))^{1/2}$.

Using the change of variables $s = \frac{t^p}{\beta^p}$, we see that for $k, \beta \geq 0$

$$\int_0^\infty t^k \left(\frac{t}{\beta}\right)^{q-1} \exp\left(-\left(\frac{t^p}{\beta^p}\right)\right) d\Lambda(t) = \frac{\beta}{p}\int_0^\infty s^{\frac{k+q}{p}-1} \exp(-s)\, d\Lambda(s) = \frac{\beta^{k+1}}{p}\Gamma\left(\frac{k+q}{p}\right).$$

Setting $k = 0$ leads us to the normalizing constant in the definition of the distribution in (2.4). Furthermore, we obtain the following expression for the moments of the $G_{p,q}$-distributions:

$$\mathbb{E}\,|\xi_1|^s = \frac{\alpha^s(1+(-1)^s)\Gamma\left((s+q)/p\right)}{2\Gamma\left(q/p\right)}, \qquad s \in \mathbb{N}.$$

In particular, $\mathbf{Var}\xi_1 = \alpha^2\Gamma((2+q)/p)/\Gamma(q/p) = 1$. Clearly, the ℓ_p-prior is equivalent to $G_{p,1}$, and W_p is equivalent to $G_{p,p}$. Furthermore, the $G_{1,q}$-distribution coincides with a symmetrization of the Gamma distribution. For $q < 1$ the distribution (2.4) will blow up at the origin, and so it will put a lot of its mass at zero. The $G_{p,q}$-distributions belong to the class of ID measures by the following theorem of Bondesson.

Theorem 2.7 (see [10, Cor. 2]). *All probability density functions on* $(0,\infty)$ *of the form*

$$\pi(t) = \frac{p}{\alpha\Gamma(q/p)}\left(\frac{t}{\alpha}\right)^{q-1} \exp\left(-\left(\frac{t}{\alpha}\right)^{p}\right)$$

are ID for $q, \alpha > 0$ *and* $0 < p \leq 1$.

Later on we show that if the ξ_k are distributed according to an ID distribution, then the corresponding product prior on X will also be an ID probability measure. Then the $G_{p,q}$-priors are also ID. This fact suggests the question, What other types of ID measures are good models for compressibility? We know that heavy-tailed distributions such as the Cauchy or Student's t-distributions are ID and that they incorporate compressible samples as well. Then there is much to be gained from the study of ID prior measures in Bayesian inverse problems. To the best of our knowledge, a thorough study of the compressible behavior of ID distributions is still missing in the literature. The closest reference in this direction is the work of Unser and colleagues [54, 56, 55]. While we do not study the modelling of compressible parameters, we recognize the potential impact of ID priors in this subject, and so we dedicate the remainder of this section to the study of ID priors.

2.2. ID priors. We collect some useful results from the theory of ID measures in Appendix A. We present only the results that are needed in our exposition and refer the reader to [41] for a detailed introduction to ID measures on Banach spaces. Further reading can be found in the monograph [54], which contains a modern treatment of ID probability measures on nuclear spaces, and the books [3, 49, 50], which are good references on the theory of ID measures in finite dimensions.

Definition 2.8 (ID measures [41]). *A Radon probability measure μ on a Banach space X is called an ID measure if for each $n \in \mathbb{N}$ there exists a Radon probability measure $\mu_{1/n}$ so that $\mu = (\mu_{1/n})^{*n}$. Equivalently, the probability measure μ is ID if $\hat{\mu}(\varrho) = (\hat{\mu}_{1/n}(\varrho))^n$ for all $\varrho \in X^*$.*

Well-known distributions such as the Gaussian, Laplace, Gamma, log-normal, Cauchy, and Student's t belong to the ID class. More examples can be found in the monograph [50], where ID distributions on $\mathbb{R}$ are studied in detail. We note that an equivalent definition of an ID measure is given as the law of a Lévy process terminated at unit time. However, we will not use this definition in order to avoid the technicalities of dealing with Lévy processes and refer the interested reader to the monographs [46, 19] for further reading.

It follows from the Lévy–Khintchine representation (Theorem A.1) that every ID measure is uniquely identified by the triple $(m, \mathcal{R}, \lambda)$, where $m \in X$ is a fixed element, $\mathcal{R} : X^* \to X$ is a covariance operator, and λ is a Lévy measure on X (see Definition A.2). Thus, we use the shorthand notation $\mu = \mathrm{ID}(m, \mathcal{R}, \lambda)$. In this paper we will restrict our attention to the case of ID measures with finite Lévy measures, i.e., $\lambda(X) < \infty$. Under this assumption, the Lévy–Khintchine representation implies that if $\mu = \mathrm{ID}(m, \mathcal{R}, \lambda)$, then it can be decomposed as (see the discussion leading to (A.5))

$$\mu = \delta_{\tilde{m}} * \mathcal{N}(0, \mathcal{R}) * \mathrm{CPois}\left(\lambda(X), \frac{1}{\lambda(X)}\lambda\right),$$

where $\tilde{m} \in X$ is a fixed element that is different from m and $\mathrm{CPois}(c, \nu)$ is a compound Poisson measure with rate c and compounded measure ν (see Definition A.4). In other words, draws from μ can be written as the sum of a fixed element with independent Gaussian and compound Poisson random variables on X.

The tail behavior and the number of moments are the most important features of a prior measure in the context of our well-posedness results in section 4. With this in mind, we now present some results concerning the tail behavior of ID measures. We begin with the notion of a submultiplicative function.

Definition 2.9 (submultiplicative function). *A nonnegative, nondecreasing, locally bounded function $h : \mathbb{R} \mapsto \mathbb{R}^+$ is called submultiplicative if it satisfies*

$$h(t+s) \le Ch(t)h(s) \qquad \forall t, s \in \mathbb{R}$$

with an independent constant $C > 0$.

Our interest in the class of submultiplicative functions arises from the next theorem, which describes some of the properties of this class.

Theorem 2.10 (see [49, Prop. 25.4]).
(i) *The product of two submultiplicative functions is also submultiplicative.*
(ii) *If h is submultiplicative, then so is $(h(at+b))^\alpha$ for constants $a, b \in \mathbb{R}$ and $\alpha > 0$.*
(iii) *The functions $\max\{1, |t|\}$ and $\exp(|t|^\beta)$ for $\beta \in (0, 1]$ are submultiplicative.*

We now present a theorem that relates the tail behavior of an ID measure to that of its Lévy measure. This result was originally proved by Kruglov [38] for ID random variables on

$\mathbb{R}$ with Lévy measures that are not necessarily finite. Different generalizations of Kruglov's result to larger spaces are also available in the literature. For example, see [49, Thm. 25.3] for extension to $\mathbb{R}^n$ and [46, Prop. 6.9] for Hilbert space valued Lévy processes. For the reader's convenience we briefly prove this result for Banach space valued ID random variables with finite Lévy measures.

Theorem 2.11. *Let X be a Banach space, and let λ be a Lévy measure so that $0 < \lambda(X) < \infty$. Suppose that $u \sim \mu = ID(m, \mathcal{R}, \lambda)$, $\mu(X) = 1$, and $\| \cdot \|_X < \infty$ μ-a.s. Then, given a submultiplicative function h, we have that $h(\| \cdot \|_X) \in L^1(X, \mu)$ if $h(\| \cdot \|_X) \in L^1(X, \lambda)$.*

Proof. Let $u \sim \mu$. Then following Theorem A.3 and the decomposition (A.5), we know that there exist an element $\tilde{m} \in X$ and independent random variables $w \sim \mathcal{N}(0, R)$ and $v \sim \mathrm{CPois}(\lambda(X), \frac{1}{\lambda(X)}\lambda)$ so that

$$\mathbb{E}\, h(\|u\|_X) = \mathbb{E}\, h(\|\tilde{m} + w + v\|_X) \leq C^3 h(\|\tilde{m}\|_X)(\mathbb{E}\, h(\|w\|_X))(\mathbb{E}\, h(\|v\|_X)),$$

where the inequality follows because of the triangle inequality and the fact that h is nondecreasing and locally bounded. Now by [49, Lem. 25.5] we know that there exist constants $a, b > 0$ such that $h(x) \leq b \exp(a|x|)$. Using this bound with the assumption that $\|u\|_X < \infty$ μ-a.s. along with Fernique's theorem [7, Thm. 2.8.5] for Gaussian measures on Banach spaces implies that $\mathbb{E}\, h(\|w\|_X) < \infty$. Now suppose that $\frac{1}{\lambda(X)} \int_X h(\|u\|_X) \mathrm{d}\lambda(u) = U < \infty$. Then using the law of total expectation [6, Thm. 34.4], the fact that h is submultiplicative, and v is a compound Poisson random variable, we get

$$\begin{aligned}
\mathbb{E}\, h(\|v\|_X) &= \mathbb{E}\left(\mathbb{E}\, h\left(\left\| \sum_{k=0}^{N} v_k \right\|_X \right) \middle| N \right) \leq \mathbb{E}\left(\mathbb{E}\, h\left(\sum_{k=0}^{N} \|v_k\|_X \right) \middle| N \right) \\
&\leq \mathbb{E}\left(C^N \mathbb{E}\left(\prod_{k=0}^{N} h\, \|v_k\|_X \right) \middle| N \right) = \mathbb{E}\left(C^N \left(\prod_{k=0}^{N} \mathbb{E}\, h\, \|v_k\|_X \right) \middle| N \right) \\
&= \mathbb{E}\left((UC)^N | N \right) = \sum_{k=0}^{\infty} \frac{e^{-\lambda(X)}(UC\lambda(X))^k}{k!} < \infty. \qquad \blacksquare
\end{aligned}$$

By putting Theorems 2.11 and 2.10 together we immediately obtain the following corollary concerning the moments of ID measures.

Corollary 2.12. *Suppose that X is a Banach space and $\mu = ID(m, \mathcal{R}, \lambda)$. If λ is a Lévy measure on X so that $0 < \lambda(X) < \infty$, $\mu(X) = 1$, and $\| \cdot \|_X < \infty$ μ-a.s., then $\| \cdot \|_X \in L^p(X, \mu)$ whenever $\| \cdot \|_X \in L^p(X, \lambda)$ for $p \in [1, \infty)$.*

Another interesting case is when the Lévy measure λ is convex. Recall that a Radon probability measure ν on X is said to be convex whenever it satisfies

$$\nu(\beta A + (1 - \beta) B) \geq \nu(A)^{\beta} \nu(B)^{1-\beta}$$

for $\beta \in [0, 1]$ and all Borel sets A and B (see [34, 11] for more details about convex measures). We are interested in convex measures since they have exponential tails under mild assumptions [34, Thm. 3.6]. More precisely, if ν is a convex probability measure on X and $\| \cdot \|_X < \infty$

ν-a.s., then there exists a constant $0 < b < \infty$ so that $\exp(b\|\cdot\|_X) \in L^1(X,\nu)$. Since the exponential is a submultiplicative function, we immediately obtain the following corollary.

Corollary 2.13. *Suppose that X is a Banach space and $\mu = ID(m, \mathcal{R}, \nu)$. If ν is a convex probability measure on X, $\mu(X) = 1$, and $\|\cdot\|_X < \infty$ a.s. under both ν and μ, then there exists a constant $b > 0$ so that $\exp(b\|\cdot\|_X) \in L^1(X,\mu)$.*

At the end of this section we ask whether we would obtain an ID measure if we used a sequence of ID random variables to generate a product prior. The answer to this question is affirmative and serves as the proof of our claim that $G_{p,q}$-priors that were introduced earlier belong to the class of ID probability measures.

Theorem 2.14. *Let X be a Banach space with an unconditional Schauder basis $\{x_k\}$, and let μ be the product prior that is obtained from $\{\gamma_k\} \in \ell^2$ and an i.i.d. sequence $\{\xi_k\}$ of real valued random variables. Suppose that $\xi_k \sim ID(0, \sigma^2, \lambda)$, where $\sigma > 0$ and λ is a symmetric and finite Lévy measure on $\mathbb{R}$ such that $\max\{1, |\cdot|^2\} \in L^1(\mathbb{R}, \lambda)$. Then μ is a Radon ID probability measure on X with characteristic function*

$$\hat{\mu}(\varrho) = \exp\left[-\frac{1}{2}\sum_{k=1}^{\infty} \sigma^2\gamma_k^2\varrho(x_k)^2 + \sum_{k=1}^{\infty}\int_{\mathbb{R}} (\cos(\gamma_k\varrho(x_k)t_k) - 1)\, d\lambda(t_k)\right] \qquad \forall \varrho \in X^*.$$

Proof. Since $\max\{1, |\cdot|^2\} \in L^1(\mathbb{R},\lambda)$, the Lévy measure of the ξ_k has bounded second moment, and so by Corollary 2.12 we see that $\mathbf{Var}\xi_k < \infty$. Now it follows from Theorem 2.1 that $\|\cdot\|_X < \infty$ μ-a.s. since $\{\gamma_k\} \in \ell^2$. The fact that μ is Radon follows from Theorem 2.2. Now we consider the characteristic function of μ. Using the definition of the characteristic function of μ and the fact that $\hat{\xi}_k(z) = \exp(-\frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}}[\cos(tz) - 1]\,\mathrm{d}\lambda(z))$, we can write

$$\hat{\mu}(\varrho) = \prod_{k=1}^{\infty} \mathbb{E}\,\exp\left(i\gamma_k\xi_k\varrho(x_k)\right) = \exp\left[-\frac{1}{2}\sum_{k=1}^{\infty}\sigma^2\gamma_k^2\varrho(x_k)^2 + \sum_{k=1}^{\infty}\int_{\mathbb{R}}[\cos(\gamma_k\varrho(x_k)t_k) - 1]\mathrm{d}\lambda(t_k)\right].$$

Now consider the sequence of measures $\{\mu_N\}_{N=1}^{\infty}$ that are defined via

$$\hat{\mu}_N(\varrho) = \exp\left[-\frac{1}{2}\sum_{k=1}^{N}\sigma^2\gamma_k^2\varrho(x_k)^2 + \sum_{k=1}^{N}\int_{\mathbb{R}}[\cos(\gamma_k\varrho(x_k)t_k) - 1]\mathrm{d}\lambda(t_k)\right].$$

Each μ_N is ID given the fact that a finite sum of ID random variables is ID. Since the $\{x_k\}$ are normalized and $\{\gamma_k\} \in \ell^2$, then $\sum_{k=1}^{\infty}\gamma_k^2\varrho(x_k)^2 < \infty$. Furthermore, using the inequality $|\cos(t) - 1| \le t^2$ we can write

$$\sum_{k=1}^{\infty}\int_{\mathbb{R}}(\cos(\gamma_k\varrho(x_k)t_k) - 1)\mathrm{d}\lambda(t_k) \le \sum_{k=1}^{\infty}\int_{\mathbb{R}}\gamma_k^2\varrho(x_k)^2 t_k^2\mathrm{d}\lambda(t_k) = \sum_{k=1}^{\infty}\gamma_k^2\varrho(x_k)^2\int_{\mathbb{R}} t_k^2\mathrm{d}\lambda(t_k).$$

But this term is also bounded since $\{\gamma_k\} \in \ell^2$, $\{x_k\}$ are normalized, and $\max\{1, |x|^2\} \in L^1(\mathbb{R},\lambda)$. Then, $\hat{\mu}_N(\ell) \mapsto \hat{\mu}(\ell)$ for all $\ell \in X^*$, and so the sequence μ_N converges weakly to μ. Therefore, μ is also ID by [41, Thm. 5.6.2]. Observe that the Lévy measure of μ is concentrated along the coordinate axes of the basis $\{x_k\}$. ■

3. Stochastic process priors on BV.

Total variation regularization is a classic technique for recovery of blocky images in the variational approach to inverse problems [57, Chap. 8]. As we mentioned earlier, it was shown in [39] that the TV-prior is not discretization invariant and converges to a Gaussian measure in the limit of fine discretizations. In this section we consider prior measures that are defined as laws of stochastic processes with jump discontinuities in order to model discontinuous functions with bounded variation. The resulting prior measures are defined directly on the function space BV and are hence discretization invariant. Therefore, our definition can get around the inconsistency that was observed in [39]. We emphasize that our construction of a BV-prior does not disprove the result of [39] but provides a well-defined alternative to the classic TV-prior. We also note that our approach is not the only way to construct a discretization invariant prior on the space of functions with bounded total variation. For example, [58] also presents an alternative to the TV-prior that is absolutely continuous with respect to an underlying Gaussian measure and results in a well-posed inverse problem.

Following [40, Chap. 13] we define the space $BV(\Omega)$ of functions of bounded variation on an open set $\Omega \subset \mathbb{R}^n$ as the space of functions $u \in L^1(\Omega)$ whose first order partial derivatives are finite signed Radon measures; i.e., for $j = 1, 2, \dots, n$ there exist finite signed measures $\mathcal{U}_j : \mathcal{B}(\Omega) \mapsto \mathbb{R}$ so that

$$\int_\Omega u(t) \frac{\partial \phi}{\partial t_j}(t) \mathrm{d}\Lambda(t) = - \int_\Omega \phi(t) \mathrm{d}\mathcal{U}_j(t) \qquad \forall \phi \in C_c^\infty(\Omega).$$

We define the *variation* of u as

$$V(u) := \sup \left\{ \sum_{k=1}^n \int_\Omega \phi_k(t) \mathrm{d}\mathcal{U}_k(t) \;\middle|\; \phi \in C_c^\infty(\Omega; \mathbb{R}^n),\ |\phi(t)| \le 1 \ \forall t \in \Omega \right\},$$

where ϕ_k denotes the kth coordinate of ϕ and $C_c^\infty(\Omega; \mathbb{R}^n)$ denotes the space of smooth functions with compact support in Ω that take their values in $\mathbb{R}^n$. The space $BV(\Omega)$ is a Banach space when equipped with the norm

$$\|u\|_{BV(\Omega)} := \|u\|_{L^1(\Omega)} + V(u),$$

but it is not separable [12, Prop. 2.3]. There is a correspondence between the space $BV([0,1])$ and the space of functions with finite total variation in 1D. Recall that the *total variation* of a function $u : [0,1] \mapsto \mathbb{R}$ is defined as

$$TV(u) := \sup \left\{ \sum_{k=1}^K |u(t_k) - u(t_{k-1})| \right\},$$

where the supremum is taken over all finite partitions $0 = t_0 < t_1 < t_2 < \cdots < t_K = 1$ of the interval $[0,1]$ with $K \in \mathbb{N}$. It is known that if $TV(u) < \infty$, then $V(u) \le TV(u)$ and every $u \in BV([0,1])$ has a right continuous representative with bounded total variation [40, Thm. 7.2]. We prefer to work with the space BV and its corresponding norm rather than the total variation functional since the former is readily defined in higher dimensions. We start by constructing a prior measure on $BV([0,1])$ as the law of a pure jump Lévy process. We shall consider priors on $BV(\Omega)$ in section 3.2. The theory of Lévy processes is the basis of our approach here (see [19] for an extensive introduction). Using the Lévy–Khintchine formula for Lévy processes [19, Thm. 3.1], we identify a Lévy process $u(t)$ via its characteristic function

$$\mathbb{E}\exp(isu(t)) = \exp(t\psi(s)) \quad \text{for} \quad s \in \mathbb{R},$$

where

$$\psi(s) = ims - \frac{1}{2}(\sigma s)^2 + \int_{\mathbb{R}} \exp(i\xi s) - 1 - is\xi \mathbf{1}_{\{|\xi|\le 1\}}(\xi)\mathrm{d}\lambda(\xi).$$

Here, the constants $m \in \mathbb{R}$ and $\sigma \ge 0$ are fixed, and λ is a Lévy measure on $\mathbb{R}$ (see Definition A.2). Similar to the case of ID measures, the *characteristic triplet* (m, σ, λ) uniquely identifies the stochastic process $u(t)$. Certain pathwise properties of $u(t)$ can be inferred from its characteristic triplet.

Theorem 3.1 (see [19, Prop. 3.9]). *Let $u(t)$ be a Lévy process with characteristic triplet (m, σ, λ). Then $u(t) \in BV([0,1])$ a.s. and $\mathbb{E}\,\|u\|_{BV([0,1])} < \infty$ if*

$$\sigma = 0 \quad \textit{and} \quad \int_{\{|\xi|\le 1\}} |\xi| d\lambda(\xi) < \infty. \tag{3.1}$$

Such a process is of the pure jump type. If, in addition, $\lambda(\mathbb{R}) < \infty$, then $u(t)$ is a compound Poisson process with piecewise constant sample paths.

Thus, the law of a Lévy process $u(t)$ that satisfies (3.1) coincides with a probability measure that is supported on $BV([0,1])$. Let us denote this measure by μ. We wish to use this measure as a prior within the Bayesian framework and achieve a well-posed inverse problem. We show a few draws from a compound Poisson process with characteristic triplet $(0, 0, \mathcal{N}(0,1))$ in Figure 3(a).

An important question at this point is whether μ is a Radon measure on $BV([0,1])$ since the Radon property can often simplify the well-posedness analysis. We will now show that in the compound Poisson case, i.e., when $\lambda(\mathbb{R}) < \infty$, the measure μ is tight and hence Radon [2, Lem. 12.6].

Lemma 3.2. *Let μ be the law of a pure jump Lévy process $u(t)$ with characteristic triplet $(0, 0, \lambda)$. Then μ is tight if*

$$\lambda(\mathbb{R}) < \infty \qquad \textit{and} \qquad \int_{\mathbb{R}} s d\lambda(s) < \infty. \tag{3.2}$$

Proof. Recall Helly's selection principle [29, Thm. 12] stating that a set $A \subset BV([0,1])$ is relatively compact if there exists a constant $M > 0$ so that

$$\|w\|_{L^\infty([0,1])} < M \quad \text{and} \quad TV(w) < M \qquad \forall w \in A.$$

Thus, to show that μ is a tight measure on $BV([0,1])$, we need to argue that for every $\epsilon > 0$ there is an $M > 0$ so that

$$\mu(\{w \in BV([0,1]) : \|w\|_{L^\infty([0,1])} \ge M \text{ or } TV(w) \ge M\}) < \epsilon.$$

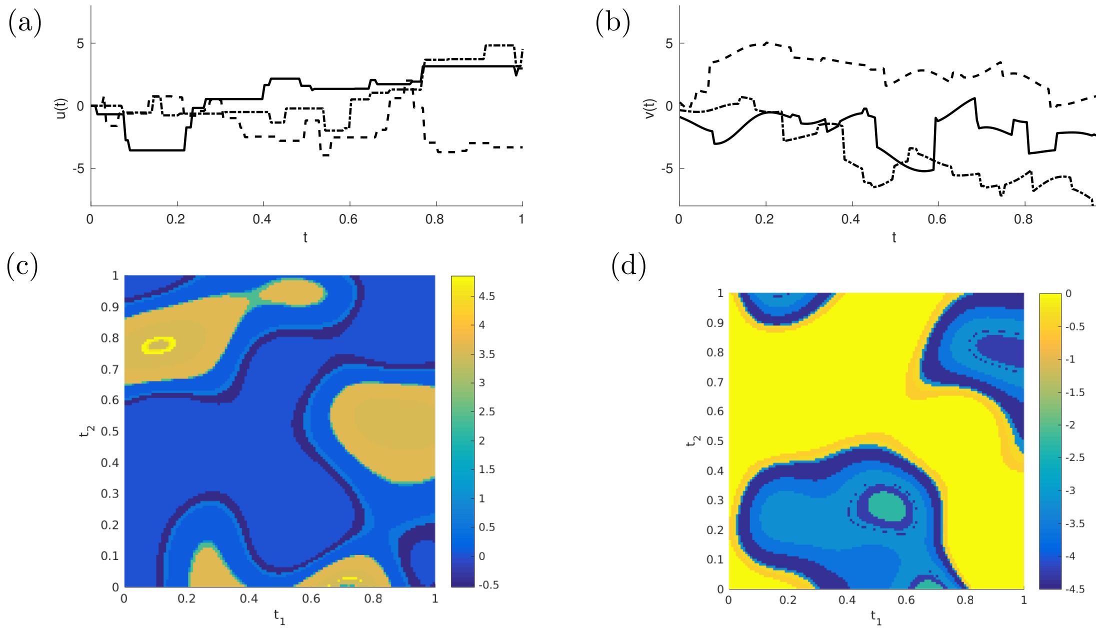

Figure 3. (a) *Three samples from the compound Poisson process* (3.3) *with normal jumps depicting the piecewise constant sample paths.* (b) *Samples from the* $v(t)$ *process generated using* (3.6) *by combining a smooth Gaussian process and an independent compound Poisson process with normal jumps.* (c), (d) *Two draws from the random field* $u(\mathbf{t})$ *generated using* (3.7) *by combining a smooth Gaussian field with a Poisson process.*

Condition (3.2) implies that $u(t)$ is a compound Poisson process, and so we can write

$$u(0) = 0, \quad u(t) = \sum_{k=1}^{\tau(t)} \xi_k \quad \text{for} \quad t \in (0,1]. \tag{3.3}$$

Here $\tau(t)$ is a Poisson process with intensity $c = \lambda(\mathbb{R})$, i.e.,

$$\mathbb{P}(\tau(t) = k) = \frac{(ct)^k}{k!} \exp(-ct),$$

and $\{\xi_k\}$ is an i.i.d. sequence of random variables distributed according to the measure $c^{-1}\lambda$. Using this representation of $u(t)$ and the law of total expectation [6, Thm. 34.4], we can write

$$\mathbb{E}\, \|u\|_{L^\infty([0,1])} = \mathbb{E} \sup_{t\in[0,1]} \left| \sum_{k=1}^{\tau(t)} \xi_k \right| \le \mathbb{E} \left(\mathbb{E} \sum_{k=1}^{N} |\xi_k| \middle| \tau(1) = N \right) = c\mathbb{E}\, |\xi_1| < \infty. \tag{3.4}$$

Furthermore, since the total variation of a piecewise constant function is simply the sum of the jump sizes, we have

$$\mathbb{E}\, TV(u) = \mathbb{E} \sum_{k=1}^{\tau(1)} |\xi_k| = c\mathbb{E}\, |\xi_1| < \infty. \tag{3.5}$$

Now it follows from Markov's inequality that for any $M > 0$

$$\mathbb{P}(\|u\|_{L^\infty([0,1])} \geq M) \leq \frac{\mathbb{E}\,\|u\|_{L^\infty([0,1])}}{M}, \qquad \mathbb{P}(TV(u) \geq M) \leq \frac{\mathbb{E}\,TV(u)}{M}.$$

A straightforward argument yields that for any choice of $\epsilon > 0$ we can choose $M > 0$ large enough so that $\mathbb{P}(\|u\|_{L^\infty([0,1])} > M \text{ or } TV(u) > M) \leq \epsilon$ and so the measure μ, the law of $u(t)$, is tight (Radon) on $BV([0,1])$. ■

To this end, Lemma 3.2 states that under mild conditions, the law of a compound Poisson process is a Radon probability measure on BV. It then follows from Theorem 4.1 that if μ is the law of a compound Poisson process, then it is concentrated on a separable subspace of BV. However, identification of this separable subspace is not always possible, and so in many applications such as the well-posedness results of section 4, one might prefer to work with μ as a measure on the nonseparable space BV. We further emphasize that Lemma 3.2 concerns only the laws of compound Poisson processes, and to the best of our knowledge this result does not hold for general choices of the Lévy measure λ.

3.1. Combination with Gaussian processes. The compound Poisson process is a convenient model for functions with jump discontinuities. However, the fact that its sample paths are piecewise constant can be too restrictive. In order to achieve a more flexible prior, which can model piecewise continuous functions, we combine our compound Poisson processes with a Gaussian process. If the sample paths of the Gaussian process are sufficiently regular, then the resulting prior measure will still be concentrated on $BV([0,1])$. The theory of Gaussian processes is well developed, and a detailed introduction can be found in the monograph [48]. Here we recall some basic results and consider only the case of a Gaussian process with C^∞ sample paths. Our approach can easily be generalized to less regular Gaussian processes by choosing a different kernel [48, sect. 4.2].

Let $g(t)$ be a random function on $[0,1]$ so that for any finite collection of points $\{t_k\}_{k=1}^n$ the random variables $\{g(t_k)\}_{k=1}^n$ are jointly Gaussian. Furthermore, suppose that

$$(g(t_1), \ldots, g(t_n))^T \sim \mathcal{N}(0, \mathbf{K}), \quad \text{where} \quad \mathbf{K}_{k,j} = \kappa(t_k, t_j).$$

Here $\kappa(r,s) := \exp(-b|r-s|^2)$ is the *covariance kernel* of $g(t)$, and $b > 0$ is a fixed constant. Under these assumptions $g(t)$ is a mean zero Gaussian process, and its samples are almost surely in $C^\infty([0,1])$ (see [45, sect. 2.5.4]). By definition, the law of $g(t)$ is a Gaussian measure, and since the kernel κ is positive definite and continuous, it follows from the Karhunen–Loéve theorem (see [27, sect. 2.3]) that the law of $g(t)$ is supported on a Hilbert space and so it is a Radon measure on $BV([0,1])$.

Now let us consider a compound Poisson process $u(t)$ as in (3.3) in addition to the Gaussian process $g(t)$. Then the new process

$$v(t) = g(t) + u(t) \tag{3.6}$$

will have sample paths that are piecewise C^∞ with finitely many jumps. Furthermore, since the laws of $u(t)$ and $g(t)$ are both Radon, the law of $v(t)$ is also Radon. Examples of draws from the process $v(t)$ are given in Figure 3(b).

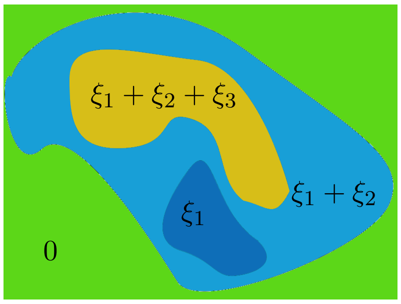

Figure 4. *Schematic representation of the piecewise constant random field of* (3.7)*. The shape of the disjoint sets is defined via the level sets of the field* g*.*

3.2. Extension to higher dimensions. At the end of this section we discuss some possibilities for extension of the compound Poisson process priors to random fields in $BV(\Omega)$ for compact domains $\Omega \subset \mathbb{R}^n$ for $n = 2, 3$, with Lipschitz boundary. Let g be a Gaussian process on Ω with kernel $\kappa(r, s) = \exp(-|r - s|^2)$, i.e.,

$$(g(t_1), \dots, g(t_n))^T \sim \mathcal{N}(0, \mathbf{K}), \quad \mathbf{K}_{k,j} = \kappa(t_k, t_j) \quad \text{for any collection} \quad t_1, \dots t_n \in \Omega.$$

Under these assumptions $g \in C^\infty(\Omega)$ a.s. Now consider the random field

$$u(t) = \sum_{k=1}^{\tau(g^+(t))} \xi_k \qquad \text{for } t \in \Omega, \qquad \text{where } g^+(t) := \max\{0, g(t)\}. \tag{3.7}$$

See Figure 4 for a schematic representation of how this field is defined. As before, τ is an independent Poisson random variable with rate $c > 0$, and $\{\xi_k\}$ is a sequence of i.i.d. random variables distributed according to the probability measure $c^{-1}\lambda$. It is straightforward to check that samples from u are piecewise constant functions on Ω that jump across a finite number of the positive level sets of the field g. These level sets are chosen by the Poisson random variable τ. Since we assumed that g is in $C^\infty(\Omega)$ a.s., we expect that the level sets of g are also smooth, and so u is a piecewise constant function that jumps across finitely many smooth curves (surfaces). We will now check whether the law of the field u is indeed supported on $BV(\Omega)$.

In what follows we will occasionally suppress the dependence of different functions on t to make the expressions more readable. Consider a test function $\phi \in C_c^\infty(\Omega; \mathbb{R}^n)$ so that $|\phi(t)| \le 1$. We can write

$$\int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}\Lambda(t) = \int_\Omega \left(\sum_{k=1}^{\tau(g^+)} \xi_k \right) \mathrm{div}\phi \,\mathrm{d}\Lambda(t) = \sum_{j=1}^{\tau(\max g^+)} \int_{A_j} \left(\sum_{k=1}^{j} \xi_k \right) \mathrm{div}\phi \,\mathrm{d}\Lambda(t),$$

with the convention that the sum is set to zero whenever $\tau(\max g^+) = 0$. The sets $A_j := \{t \in \Omega \,|\, \tau(g^+(t)) = j\}$ are the subsets of Ω on which u is constant. In what follows Θ^{n-1} denotes $(n-1)$-dimensional Hausdorff measure [25, Chap. 2] on $\mathbb{R}^n$. Recall that the one-dimensional

Hausdorff measure of a simple curve in 2D coincides with its length [25, sect. 3.4]. Since τ is almost surely finite, we can integrate by parts [25, Thm. 5.6] to get

$$\int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}\Lambda(t) = \sum_{j=1}^{\tau(\max g^+)} \int_{\partial A_j} \mathcal{J}_j(u)\langle \phi, \vartheta_j\rangle \mathrm{d}\Theta^{n-1}(s).$$

Here ∂A_j is the boundary of A_j, ϑ_j is the unit outward normal on ∂A_j, and $\mathcal{J}_j(u)$ is the jump of u across ∂A_j going from A_{j-1} to A_j. The reason for using the Hausdorff measure rather than the usual arclength element is that the level sets of the field g are random, and we do not yet know if they have finite length. Since the size of the jumps of u are a.s. finite, we can write

$$\begin{aligned}\int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}\Lambda(t) &\le \left(\sum_{k=1}^{\tau(\max g^+)} |\xi_k|\right)\left(\sum_{j=1}^{\tau(\max g^+)} \int_{\partial A_j} \langle \phi, \vartheta_j\rangle \mathrm{d}\Theta^{n-1}(s)\right)\\ &\le \left(\sum_{k=1}^{\tau(\max g^+)} |\xi_k|\right)\sum_{j=1}^{\tau(\max g^+)} \Theta^{n-1}(\partial A_j).\end{aligned}$$

The main question now is whether or not $\Theta^{n-1}(\partial A_j)$ are finite a.s. This is solely a property of the field g. We need a generalization of Rice's formula in order to respond to this question.

Theorem 3.3 (see [4, Thm. 6.8 and Prop. 6.12]). *Let Ω be a compact set in $\mathbb{R}^n$ with $n \ge 1$. Let $z : \Omega \mapsto \mathbb{R}$ be a Gaussian random field so that $z \in C^2(\Omega)$ a.s. and $\mathbf{Var} z(t) > 0$ for all $t \in \Omega$. For a fixed constant $b \in \mathbb{R}$, $\mathbb{E}\,\Theta^{n-1}(\{t \in \Omega | z(t) = b\}) < \infty$ if the pair $(z(t), \nabla z(t))$ have a joint density on $\Omega \times \mathbb{R}^n$ that is locally bounded.*

It is well known [1, Thm. 2.2.2] that for $j \in \{1, \dots, n\}$ the processes $g_j(t) := \frac{\partial g}{\partial t_j}(t)$ are themselves Gaussian processes with kernels $\kappa_j(r, s) = \frac{\partial^2}{\partial r_j \partial s_j}\kappa(r, s)$ whenever the second order partial derivative of the kernel exists. Using this fact it is straightforward to check that our choice of the field g satisfies the assumptions of the above theorem. Therefore, $\Theta^{n-1}(\partial A_j)$ are finite a.s., and we conclude that $V(u) < \infty$ a.s.

We show two samples from the random field of (3.7) on the box $[0, 1]^2$ in Figures 3(c) and 3(d) with standard normal jumps. The choice of the field g influences the shape of the discontinuity curves of u. The main difficulty in proving $u \in BV(\Omega)$ is in showing that the level sets of the underlying Gaussian field have finite length. Theorem 3.3 allows us to relax the regularity assumptions on g and take Gaussian fields that are in $C^2(\Omega)$ rather than $C^\infty(\Omega)$, but for less regular fields it is not clear whether the level sets have finite length.

4. Well-posed Bayesian inverse problems. Recall from section 1 that we are interested in the problem of inferring a parameter $u \in X$ from data $y \in Y$ that is related via a generic stochastic mapping $\tilde{\mathcal{G}}$ that models the physical process that generates the data as well as the measurement noise:

$$y = \tilde{\mathcal{G}}(u).$$

To solve this problem we employ Bayes' rule (1.2). In this section we collect certain conditions on the prior measure μ_0 and the likelihood potential Φ that result in well-posed inverse problems. We consider a general enough setting that encompasses the heavy-tailed priors of section 2 and the stochastic process priors of section 3. We assume that the parameter space X is a Banach space that is not necessarily separable (such as BV) and the prior measure μ_0 is possibly heavy-tailed (such as the $G_{p,q}$-priors) and not necessarily Radon (such as the law of the pure jump Lévy processes when λ is not finite).

The main results of this section are Theorems 4.3 and 4.4, which establish the existence, uniqueness, and stability of the posterior measure. We acknowledge that these theorems are very similar to the results in [23, sect. 4.1] and [52]. In comparison to these articles, we impose slightly different assumptions on the potential Φ and assume that the space X is not necessarily separable. We also note that under the assumption that the prior measure μ_0 is a Radon measure one can immediately generalize the result of [23] to nonseparable parameter spaces X using the fact that Radon measures on a Banach space are automatically concentrated on a separable subspace.

Theorem 4.1 (see [9, Thm. 7.12.4]). *Let μ be a Radon probability measure on a Banach space X. Then, there exists a reflexive and separable Banach space E embedded in X such that $\mu(X \setminus E) = 0$ and the closed balls of E are compact in X.*

However, it is important to note that while this theorem guarantees the existence of the separable space E, it does not provide us with a method for identifying E or its norm. In the case of the product priors of section 2 one can argue that the measures are concentrated on a separable Hilbert space, but for the stochastic process priors of section 3 it is no longer clear what the space E is, and so it is more convenient for us to analyze the inverse problem on the ambient space X rather than passing to the space E.

We will present our well-posedness results using the total variation metric, since this metric is less often used in previous works, and we refer the reader to [23] for proofs using the Hellinger metric that can easily be generalized to our setting by comparison to the proofs using the total variation metric. Given the inequalities (1.6), we immediately see that well-posedness in one of these metrics implies well-posedness in the other, but the convergence rates will differ.

We start by presenting minimal assumptions on the likelihood potential and the forward map and make our way to more specific cases of inverse problems such as problems with linear forward maps. In a nutshell, as we put more restrictions on Φ, we are able to relax our assumptions on the prior μ_0. In order to help with navigation through this section we present Table 1, which collects our main results and the key underlying assumptions.

We begin by identifying some conditions on Φ that allow us to use a very large class of prior measures including those that are heavy-tailed.

Assumption 4.2. *Suppose that X and Y are Banach spaces and the likelihood potential $\Phi : X \times Y \mapsto \mathbb{R}$ satisfies the following properties:*

(i) *(Lower bound in u) There is a positive and nondecreasing function $f_1 : \mathbb{R}_+ \mapsto [1, \infty)$ so that for all $r > 0$, there is a constant $M(r) \in \mathbb{R}$ such that*

$$\Phi(u; y) \geq M(r) - \log\left(f_1(\|u\|_X)\right) \qquad \forall u \in X \text{ and } \forall y \in B_Y(r).$$

Table 1

Summary of the key theorems and corollaries of section 4. In each case we identify the key underlying assumptions as well as the type of final result.

Theorem/corollary	Main assumptions	Type of result
Theorem 4.3	Φ is locally bounded and Lipschitz in u	μ^y is well-defined
Theorem 4.4	Φ satisfies Assumption 4.2	μ^y depends continuously on y
Corollary 4.5	Φ has polynomial growth in u, and μ_0 has finitely many moments	well-posedness
Corollary 4.7	$\Phi \geq 0$ in addition to Assumption 4.2	well-posedness
Corollary 4.9	$Y = \mathbb{R}^m$, measurement noise is additive and Gaussian, prior is ID	well-posedness
Corollary 4.10	$Y = \mathbb{R}^m$, forward map is linear and bounded, measurement noise is additive and Gaussian	well-posedness
Corollary 4.12	$Y = \mathbb{R}^m$, forward map is linear and bounded, measurement noise is additive and Gaussian, $G_{p,q}$-prior	well-posedness

(ii) *(Boundedness above) For all $r > 0$, there is a constant $K(r) > 0$ such that*

$$\Phi(u; y) \leq K(r) \qquad \forall u \in B_X(r) \text{ and } \forall y \in B_Y(r).$$

(iii) *(Continuity in u) For all $r > 0$, there exists a constant $L(r) > 0$ such that*

$$|\Phi(u_1; y) - \Phi(u_2, y)| \leq L(r)\|u_1 - u_2\|_X \qquad \forall u_1, u_2 \in B_X(r) \text{ and } y \in B_Y(r).$$

(iv) *(Continuity in y) There is a positive and nondecreasing function $f_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ so that for all $r > 0$, there is a constant $C(r) \in \mathbb{R}$ such that*

$$|\Phi(u; y_1) - \Phi(u, y_2)| \leq C(r) f_2(\|u\|_X)\|y_1 - y_2\|_Y \qquad \forall y_1, y_2 \in B_Y(r) \text{ and } \forall u \in X.$$

Our first task is to establish the existence and uniqueness of the posterior measure.

Theorem 4.3. *Suppose X is a Banach space, μ_0 is a Borel probability measure on X, and Φ satisfies Assumptions 4.2(i)–(iii) with a function $f_1 \geq 1$. If $f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$, then the posterior μ^y given by (1.2) is a well-defined Borel probability measure on X. If μ_0 is Radon, then so is μ^y.*

Proof. Our proof will closely follow the approach of [52, Thm. 4.3] and [34, Thm. 2.2]. Assumption 4.2(iii) implies the continuity of Φ on X, which in turn implies that $\Phi(\cdot, y) : X \mapsto \mathbb{R}$ is μ_0-measurable. We will now show that the normalizing constant satisfies $0 < Z(y) < \infty$, which proves that μ^y is well defined. The assertion that μ^y inherits the Radon property of μ_0 will then follow from the absolute continuity of μ^y with respect to μ_0 [9, Lem. 7.1.11].

Following Assumption 4.2(i), we can write

$$Z(y) \leq \int_X \exp(\log(f_1(\|u\|_X)) - M)\mathrm{d}\mu_0(u) = \exp(-M)\int_X f_1(\|u\|_X)\mathrm{d}\mu_0(u) < \infty.$$

We now need to show that the normalizing constant $Z(y)$ does not vanish. It follows from Assumption 4.2 that for $R > 0$,

$$Z(y) \geq \int_{B_X(R)} \exp(-K)\mathrm{d}\mu_0(u) = \exp(-K)\mu_0(B_X(R)).$$

However, $\mu_0(B_X(R)) > 0$ for large enough R. To see this consider the disjoint sets $A_k := \{u | k-1 \leq \|u\|_X < k\}$ for $k \in \mathbb{N}$. The A_k are open and hence measurable, and $\sum_{k=1}^{\infty} \mu_0(A_k) = \mu_0(\bigcup_{k=1}^{\infty} A_k) = \mu(X) = 1$. Then the measure of at least one of the A_k has to be nonzero. ∎

We now establish the stability of Bayesian inverse problems with respect to perturbations in the data. Similar versions of the following theorems are available for Gaussian priors in [51], for Besov priors in [22], for convex priors in [34], and for heavy-tailed priors on separable Banach spaces in [23, 52].

Theorem 4.4. *Suppose that X is a Banach space, μ_0 is a Borel probability measure on X, and Φ satisfies Assumptions 4.2(i), (ii), and (iv) with functions f_1, f_2. Let μ^y and $\mu^{y'}$ be two measures defined via (1.2) for any y and $y' \in Y$, both absolutely continuous with respect to μ_0.*

(i) *If $f_2(\|\cdot\|_X) f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$, then for all $r > 0$, there exists a constant $C(r) > 0$ so that $d_{TV}(\mu^y, \mu^{y'}) \leq C\|y - y'\|_Y$ for all $y, y' \in B_Y(r)$.*

(ii) *If instead $(f_2(\|\cdot\|_X))^2 f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$, then for all $r > 0$, there exists a constant $C'(r) > 0$ so that $d_H(\mu^y, \mu^{y'}) \leq C'\|y - y'\|_Y$.*

Proof. We will prove only (i) and refer the reader to [23, sect. 4.1] for the proof of (ii) that will readily generalize to our setting. Consider the normalizing constants $Z(y)$ and $Z(y')$. We have already established in the proof of Theorem 4.3 that neither of these constants will vanish and that they are both bounded. Thus the measures μ^y and $\mu^{y'}$ are well defined. Applying the mean value theorem to the exponential function and using Assumptions 4.2(i) and (iv) and the assumption that $f_2(\|\cdot\|_X) f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$, we obtain

$$\begin{aligned}
|Z(y) - Z(y')| &\leq \int_X \exp(-\Phi(u; y)) |\Phi(u; y) - \Phi(u; y')| \mathrm{d}\mu_0(u) \\
&\leq \left(\int_X \exp(\log(f_1(\|u\|_X)) - M) C f_2(\|u\|_X) \mathrm{d}\mu_0(u) \right) \|y - y'\|_Y \\
&\leq C \exp(-M) \left(\int_X f_1(\|u\|_X) f_2(\|u\|_X) \mathrm{d}\mu_0(u) \right) \|y - y'\|_Y \lesssim \|y - y'\|_Y.
\end{aligned} \tag{4.1}$$

Following the definition of the total variation distance, we have

$$\begin{aligned}
2d_{TV}(\mu^y, \mu^{y'}) &= \int_X \left| Z(y)^{-1} \exp(-\Phi(u; y)) - Z(y')^{-1} \exp(-\Phi(u, y')) \right| \mathrm{d}\mu_0(u) \\
&\leq \int_X \left| Z(y)^{-1} \exp(-\Phi(u; y)) - Z(y')^{-1} \exp(-\Phi(u, y)) \right| \mathrm{d}\mu_0(u) \\
&\quad + Z(y')^{-1} \int_X \left| \exp(-\Phi(u; y)) - \exp(-\Phi(u, y')) \right| \mathrm{d}\mu_0(u) =: I_1 + I_2.
\end{aligned}$$

Now using (4.1) we have

$$I_1 = |Z(y)^{-1} - Z(y')^{-1}| Z(y) = \frac{Z(y)}{Z(y')} |Z(y') - Z(y)| \lesssim \|y - y'\|_X.$$

Furthermore, using the mean value theorem and Assumptions 4.2(i) and (iv), we can write

$$\begin{aligned} Z(y')I_2 &= \int_X \left|\exp(-\Phi(u;y)) - \exp(-\Phi(u,y'))\right| \mathrm{d}\mu_0(u) \\ &\le \int_X \exp(-\Phi(u;y)) \left|\Phi(u;y') - \Phi(u,y)\right| \mathrm{d}\mu_0(u) \\ &\le C\exp(-M)\left(\int_X \exp(\log(f_1(\|u\|_X)))f_2(\|u\|_X)\mathrm{d}\mu_0(u)\right)\|y-y'\|_Y \lesssim \|y-y'\|_Y. \quad \blacksquare \end{aligned}$$

The main distinction between the choice of the metrics in Theorem 4.4 is that in order to obtain the same rate of convergence in the Hellinger metric we need a (possibly) stronger assumption regarding the integrability of $f_1(\|u\|_X)$ and $f_2(\|u\|_X)$. So far we have encountered conditions of the form $(f_2(\|u\|_X))^p f_1(\|u\|_X) \in L^1(X,\mu_0)$ for $p \in \{0,1,2\}$. Intuitively, these conditions identify the interplay between the growth of $\Phi(u;y)$ as a function of $\|u\|_X$ and the tail behavior of the prior μ_0.

Corollary 4.5. *Suppose that Φ satisfies the conditions of Assumption 4.2 with $f_1(t) = f_2(t) = \max\{1, |t|^p\}$ for $p \ge 0$ and μ_0 is a Borel probability measure on X. If μ_0 has bounded raw moments of degree up to $\lceil 2p \rceil$, then the Bayesian inverse problem (1.2) is well-posed in both the total variation and Hellinger metrics.*

Corollary 4.6. *Suppose that μ_0 is a Borel probability measure on X and $\exp(b\|\cdot\|_X) \in L^1(X,\mu_0)$ for some constant $b > 0$. Then the Bayesian inverse problem (1.2) is well-posed in both the total variation and Hellinger metrics whenever Φ satisfies the conditions of Assumption 4.2 with functions f_1, f_2 that are polynomially bounded.*

For the remainder of this section we will focus on specific classes of likelihood potentials Φ which allow us to further relax our assumption regarding the tail behavior of μ_0. The rest of our results follow from Theorems 4.3 and 4.4, but they are of great interest in practical applications. We start with the case of additive noise models and consider linear inverse problems afterwards.

4.1. The case of additive noise models. Additive noise models have a special place in practical applications due to their convenience and flexibility [36]. Suppose that the data is finite-dimensional and, without loss of generality, take $Y = \mathbb{R}^m$, $m \in \mathbb{N}$. Now suppose that $y \in Y$ is related to the parameter $u \in X$ via the model

$$y = \mathcal{G}(u) + \eta, \qquad \eta \sim \pi(y)\mathrm{d}\Lambda(y), \tag{4.2}$$

where $\pi(y)$ is the Lebesgue density of the measurement noise η and $\mathcal{G}: X \mapsto \mathbb{R}^m$ is the forward map. It is straightforward to check that under these assumptions

$$\Phi(u;y) = -\log \pi(\mathcal{G}(u) - y). \tag{4.3}$$

In particular, if $\eta \sim \mathcal{N}(0, \mathbf{\Sigma})$ with an $m \times m$ positive definite matrix $\mathbf{\Sigma}$, then

$$\Phi(u;y) = \frac{1}{2}\|(\mathcal{G}(u) - y)\|^2_{\mathbf{\Sigma}}. \tag{4.4}$$

Now if $\log \pi(y) \le 0$ (which is clearly the case when η is Gaussian or Laplace), then $\Phi(u; y)$ will satisfy Assumption 4.2(i) with the constant $M = 0$ and $f_1(x) = 1$. This observation will allow us to relax our assumption on the tail behavior of the prior whenever the measurement noise is additive.

Corollary 4.7. *Suppose $Y = \mathbb{R}^m$, X is a Banach space, and $\Phi(u; y) \ge 0$ and satisfies Assumptions* 4.2(ii) *and* (iv) *with a function f_2. Suppose that the prior measure μ_0 is a Borel probability measure on X, and let μ^y and $\mu^{y'}$ be two measures defined via* (1.2) *for y and $y' \in Y$. Then the posterior measure μ^y is well defined, and the following hold:*

(i) *If $f_2(\|\cdot\|_X) \in L^1(X, \mu_0)$, then for all $r > 0$, there exists $C(r) > 0$ so that $d_{TV}(\mu^y, \mu^{y'}) \le C(r)\|y - y'\|_Y$ for all $y, y' \in B_Y(r)$.*

(ii) *If $f_2(\| \cdot \|_X) \in L^2(X, \mu_0)$, then there exists $C'(r) > 0$ $d_H(\mu^y, \mu^{y'}) \le C'(r)\|y - y'\|_Y$.*

At this point it is natural to identify conditions on the distribution of the noise and the forward operator that guarantee that the likelihood potential of (4.3) satisfies the conditions of Assumption 4.2. We will address this when η is Gaussian, but our approach can be generalized to other types of additive noise models.

Theorem 4.8. *Consider the additive noise model of* (4.2) *when $\eta \sim \mathcal{N}(0, \mathbf{\Sigma})$ and $\mathbf{\Sigma}$ is a positive definite matrix. Then the corresponding likelihood potential $\Phi(u; y) \ge 0$. Furthermore, Φ satisfies the conditions of Assumption* 4.2(iv) *with $f_2(x) = 1 + \tilde{f}(x)$ if there is a positive, nondecreasing, and locally bounded function $\tilde{f} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ so that the following hold:*

(i) *There exists $C > 0$ for which $\|\mathcal{G}(u)\|_2 \le C\tilde{f}(\|u\|_X)$ for all $u \in X$.*

(ii) *For all $r > 0$, there exists $K(r) > 0$ so that $\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_2 \le K(r)\|u_1 - u_2\|_X$ for all $u_1, u_2 \in B_X(r)$.*

Proof. Since we assumed that η is Gaussian, the likelihood potential is of the form (4.4). Then it is clear that $\Phi(u; y) \ge 0$, which immediately implies that Φ satisfies Assumption 4.2(i) with $M = 0$ and $f_1(x) = 0$. Now fix $r > 0$ and suppose that $u \in B_X(r)$ and $y \in B_Y(r)$. Define $\tilde{r} = \max\{r, C\tilde{f}(r)\}$ and note that $\tilde{r}$ is bounded since we assumed that $\tilde{f}$ is locally bounded. Therefore, we have $\Phi(u; y) \le \|\mathcal{G}(u)\|^2_{\mathbf{\Sigma}} + \|y\|^2_{\mathbf{\Sigma}} \lesssim \tilde{r}^2$, and so Φ satisfies Assumption 4.2(ii).

Now we will show that Φ satisfies Assumption 4.2(iii) as well. Let r and $\tilde{r}$ be defined as above and consider $u_1, u_2 \in B_X(r)$ and $y \in B_Y(r)$. Using the identity $\|a\|_2^2 - \|b\|_2^2 = \langle a-b, a+b\rangle$ for $a, b \in \mathbb{R}^m$ and conditions (i) and (ii) of the theorem, we obtain

$$\begin{aligned} 2|\Phi(u_1; y) - \Phi(u_2; y)| &= \left| \langle \mathbf{\Sigma}^{-1/2}(\mathcal{G}(u_1) - \mathcal{G}(u_2)), \mathbf{\Sigma}^{-1/2}(\mathcal{G}(u_1) + \mathcal{G}(u_2) - 2y)\rangle \right| \\ &\le \left(\|\mathcal{G}(u_1)\|_{\mathbf{\Sigma}} + \|\mathcal{G}(u_2)\|_{\mathbf{\Sigma}} + 2\|y\|^2_{\mathbf{\Sigma}} \right) \|(\mathcal{G}(u_1) - \mathcal{G}(u_2))\|_{\mathbf{\Sigma}} \\ &\le C(\tilde{r})\|(\mathcal{G}(u_1) - \mathcal{G}(u_2))\|_{\mathbf{\Sigma}} \le 2K(r)\|u_1 - u_2\|_X. \end{aligned}$$

Finally, fix $r > 0$ and consider $y_1, y_2 \in B_Y(r)$. Then using the same line of reasoning as above, for any $u \in X$ we can write

$$\begin{aligned} 2|\Phi(u; y_1) - \Phi(u; y_2)| &= \left| \langle \mathbf{\Sigma}^{-1/2}(y_2 - y_1), \mathbf{\Sigma}^{-1/2}(2\mathcal{G}(u) - y_1 - y_2)\rangle \right| \\ &\le (\|y_2\|_{\mathbf{\Sigma}} - \|y_1\|_{\mathbf{\Sigma}} + 2\|\mathcal{G}(u)\|_{\mathbf{\Sigma}}) \|(y_2 - y_1)\|_{\mathbf{\Sigma}} \\ &\le C(r)(1 + \tilde{f}(\|u\|_X))\|y_1 - y_2\|_2. \end{aligned}$$

∎

By putting this result together with Theorem 2.11 and Corollary 4.7 we deduce the following corollary concerning the well-posedness of Bayesian inverse problems with ID priors.

Corollary 4.9. *Let X be a Banach space, and let $Y = \mathbb{R}^m$. Consider the additive noise model*

$$y = \mathcal{G}(u) + \eta, \qquad \eta \sim \mathcal{N}(0, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma}$ is a positive definite matrix and $\mathcal{G} : X \mapsto \mathbb{R}^m$ satisfies the conditions of Theorem 4.8 with a submultiplicative function $\tilde{f}$. Also, suppose that $\mu_0 = ID(m, \mathcal{R}, \lambda)$, where λ is a Lévy measure such that $\lambda(X) < \infty$, $\mu_0(X) = 1$, and $\| \cdot \|_X < \infty$ μ_0-a.s. Then the Bayesian inverse problem (1.2) is well-posed if $1 + \tilde{f}(\| \cdot \|_X) \in L^1(X, \lambda)$.

4.2. The case of linear inverse problems. We now assume that the likelihood potential Φ has the form

$$\Phi(u; y) : X \times \mathbb{R}^m \mapsto \mathbb{R}, \qquad \Phi(u; y) = \frac{1}{2} \|\mathcal{G}(u) - y\|^2_{\mathbf{\Sigma}},$$

where $\mathbf{\Sigma}$ is a positive definite matrix and $\mathcal{G} : X \mapsto \mathbb{R}^m$ is bounded and linear. This case is of particular importance due to its occurrence in the compressed sensing literature [26] and estimation of sparse parameters. In this case, we can further relax our conditions on the prior measure μ_0 and achieve well-posedness as long as the prior μ_0 has bounded first moment.

Corollary 4.10. *Let X be a Banach space, and let $Y = \mathbb{R}^m$. Suppose that the forward map $\mathcal{G} : X \mapsto \mathbb{R}^m$ is bounded and linear, and consider the additive noise model*

$$y = \mathcal{G}(u) + \eta, \quad \text{where} \quad \eta \sim \mathcal{N}(0, \mathbf{\Sigma}) \quad \text{and} \quad \mathbf{\Sigma} \text{ is positive definite.}$$

Then the Bayesian inverse problem of identifying the posterior μ^y via (1.2) is well-posed in both the Hellinger and total variation metrics if the prior μ_0 is a Borel probability measure on X and $\| \cdot \|_X \in L^1(X, \mu_0)$.

Proof. The proof follows directly from Theorems 4.8 and 4.3 and Corollary 4.7(i). ∎

Let us now return to the product priors of section 2.1 and show that those measures result in well-posed Bayesian inverse problems under the linear and additive noise assumptions.

Theorem 4.11. *Let X be a Banach space with an unconditional Schauder basis $\{x_k\}$, and take $Y = \mathbb{R}^m$. Suppose that the measurement noise is additive and Gaussian and the forward map $\mathcal{G}$ is bounded and linear. Furthermore, suppose that μ_0 is a product prior with sample paths $u = \sum_{k=1}^\infty \gamma_k \xi_k x_k$, where $\{\gamma_k\} \in \ell^2$ and $\{\xi_k\}$ are i.i.d. and $\mathbf{Var} \xi_k < \infty$. Then the inverse problem (1.2) is well-posed in both the total variation and Hellinger metrics.*

Proof. The fact that μ_0 is a Radon probability measures on X follows from Theorems 2.2 and 2.1. Now if $\mathbf{Var}\xi_k < \infty$, then $\mathbb{E}\, \xi_k^2 < \infty$ as well, and so it follows from Theorem 2.3 that $\| \cdot \|_X \in L^2(X, \mu)$. Then the assertion follows from Theorems 4.7 and 4.8. ∎

Finally we turn our attention to the $G_{p,q}$-priors of section 2. The proof of the following corollary follows directly from Theorem 2.3 and the fact that the $G_{p,q}$-distributions in 1D have bounded variance for $0 < p, q \leq 1$.

Corollary 4.12. *Let X be a Banach space with an unconditional Schauder basis $\{x_k\}$ and $Y = \mathbb{R}^m$. Suppose that the measurement noise is additive and Gaussian and that the forward* *map $\mathcal{G}$ is bounded and linear. Then the Bayesian inverse problem* (1.2) *is well-posed in both the Hellinger and total variation metrics if μ_0 is a $G_{p,q}$-prior with $0 < p, q < 1$.*

5. Practical considerations and examples.

We now turn our attention to practical aspects of solving an inverse problem within the Bayesian framework. In the first part of this section we discuss the problem of approximating the posterior measure via approximation of the likelihood potential. Afterwards, we will present three concrete examples of Bayesian inverse problems with heavy-tailed priors that arise from practical problems in image deblurring and ultrasound therapy.

5.1. Consistent approximation of the posterior.

Up to this point we were concerned with identifying prior measures μ_0 that result in a well-posed Bayesian inverse problem for a given likelihood potential Φ. However, in practice we cannot solve the inverse problem directly on the infinite-dimensional Banach space. Therefore, we need to obtain a finite-dimensional approximation to the posterior measure μ^y which is, in some sense, consistent with the infinite-dimensional limit.

To this end, we will define the notion of *consistent approximation* of a Bayesian inverse problem in the context of applications where one would discretize (1.2) by approximating the likelihood potential Φ with a discretized version $\Phi_N : X \times Y \mapsto \mathbb{R}$, akin to a finite element discretization. We define the approximation μ_N^y to μ^y via

$$\frac{\mathrm{d}\mu_N^y}{\mathrm{d}\mu_0} = \frac{1}{Z_N(y)} \exp(-\Phi_N(u; y)), \quad \text{where} \quad Z_N(y) = \int_X \exp(-\Phi_N(u; y))\mathrm{d}\mu_0(u). \tag{5.1}$$

Definition 5.1 (consistent approximation [34]). *The approximate Bayesian inverse problem* (5.1) *is a consistent approximation to* (1.2) *for a choice of μ_0, Φ, and Φ_N if $d(\mu^y, \mu_N^y) \to 0$ as $|\Phi(u; y) - \Phi_N(u; y)| \to 0$ pointwise everywhere. Here, d is either the total variation or the Hellinger metric.*

This notion of a consistent approximation relates directly to practical applications. Suppose, for example, that we are interested in computing the expected value of a quantity $h(u)$ under the posterior μ^y, but we can compute the expectation only under the approximation μ_N^y. If μ_N^y is a consistent approximation in the Hellinger metric, then we have, by the bound (1.7), that if $h \in L^2(X, \mu^y) \cap L^2(X, \mu_N^y)$, then

$$\left| \int_X h(u)\mathrm{d}\mu^y(u) - \int_X h(u)\mathrm{d}\mu_N^y(u) \right| \le C d_H(\mu^y, \mu_N^y).$$

In what follows, we will provide sharper bounds on the rate of convergence of the distances between μ^y and μ_N^y under mild conditions.

Theorem 5.2. *Assume that the measures μ^y and μ_N^y are defined via* (1.2) *and* (5.1)*, for a fixed $y \in Y$ and all values of N, and are absolutely continuous with respect to the prior μ_0 which is a Borel probability measure on X. Also assume that both Φ and Φ_N satisfy Assumptions* 4.2(i) *and* (ii) *with an appropriate function f_1, uniformly for all N, and that there exists a positive and nondecreasing function $f_3 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ so that*

$$|\Phi(u; y) - \Phi_N(u; y)| \le f_3(\|u\|_X)\rho(N), \tag{5.2}$$

where $\rho(N) \to 0$ *as* $N \to \infty$.

(i) *If* $f_3(\|\cdot\|_X)f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$*, then there exists a constant* $D > 0$*, independent of* N *such that* $d_{TV}(\mu^y, \mu^y_N) \le D\rho(N)$.

(ii) *If* $(f_3(\|\cdot\|_X))^2 f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$*, then there exists a constant* $D' > 0$*, independent of* N *such that* $d_H(\mu^y, \mu^y_N) \le D'\rho(N)$.

Proof. Our method of proof uses arguments similar to those in Theorem 4.4, and so we will present it briefly only for the total variation distance. Proof of part (ii) can also be found in [23] for separable Banach spaces. The existence and uniqueness of the measures μ^y and μ^y_N follows from Theorem 4.3 for all values of N. Next, the mean value theorem, Assumption 4.2(i), (5.2), and the assumption that $f_3(\|\cdot\|_X)f_1(\|\cdot\|_X)$ is μ_0-integrable give

$$\begin{aligned}|Z(y) - Z_N(y)| &\le \int_X \exp(-\Phi(u;y))|\Phi(u;y) - \Phi_N(u;y)|\mathrm{d}\mu_0(u) \\ &\le \left(\int_X \exp(\log(f_1(\|u\|_X)) - M)Cf_3(\|u\|_X)\mathrm{d}\mu_0(u)\right)\rho(N) \lesssim \rho(N).\end{aligned}$$

Furthermore, we have

$$\begin{aligned}2d_{TV}(\mu^y, \mu^{y'}) \le &\int_X \left|Z(y)^{-1}\exp(-\Phi(u;y)) - Z_N(y)^{-1}\exp(-\Phi(u,y))\right| \mathrm{d}\mu_0(u) \\ &+ Z_N(y)^{-1}\int_X |\exp(-\Phi(u;y)) - \exp(-\Phi_N(u,y))|\,\mathrm{d}\mu_0(u) =: I_1 + I_2.\end{aligned}$$

Then, similar to the proof of Theorem 4.4, it follows that $I_1 \lesssim \rho(N)$ and $I_2 \lesssim \rho(N)$, which gives the desired result. ∎

We now consider a more specific setting where the prior measure μ_0 has a product structure. Suppose that the likelihood potential Φ satisfies Assumption 4.2 with some functions f_1, f_2. Also, assume that the space X has an unconditional Schauder basis $\{x_k\}$, and consider the sequence of spaces $(X_N, \|\cdot\|_X)$, where $X_N = \operatorname{span}\{x_k\}_{k=1}^N$. These are linear subspaces of X, and for each $N \in \mathbb{N}$ we have $X = X_N \oplus X_N^\perp$, meaning that every $u \in X$ can be written as $u = u_N + u_N^\perp$, where $u_N \in X_N$ and $u_N^\perp \in X_N^\perp$.

Suppose that the prior measure μ_0 has the product structure of (2.1) and assume that it has sufficiently fast decaying tails so that the posterior measure μ^y is well defined. Observe that for every value of N the product prior can be factored as

$$\mu_0 = \mu_N \otimes \mu_N^\perp, \tag{5.3}$$

where μ_N and $\mu_N^\perp$ are Radon measures on X_N and $X_N^\perp$. It is natural for us to discretize the potential Φ using a projection operator:

$$\Phi_N(u;y) := \Phi(P_N u; y), \tag{5.4}$$

where $P_N : X \mapsto X_N$ is defined by $P_N(u) = u_N$. Next, define the approximate posterior measures μ^y_N as in (5.1) using the above definition of Φ_N. Under these assumptions, the μ^y_N will factor as (see [34, sect. 4.1] for the details)

$$\mu^y_N = \nu_N \otimes \mu_N^\perp, \tag{5.5}$$

where

$$\frac{\mathrm{d}\nu_N}{\mathrm{d}\mu_N} = \frac{1}{Z_N(y)} \exp(-\Phi(P_N u; y)).$$

In other words, the likelihood potential is informative only on the subspace X_N, and so by comparing (5.3) and (5.5) we see that the approximate posterior μ_N^y differs from the prior only on this subspace and is identical to the prior on $X_N^\perp$. As an example, we now check whether this method for discretization of the posterior results in a consistent approximation to μ^y in the additive Gaussian noise case.

Theorem 5.3. *Consider the above setting where the posterior and the prior have the prescribed product structures and the X_N are linear subspaces of X. Suppose that $\Phi$and Φ_N are given by*

$$\Phi(u; y) = \frac{1}{2}\|\mathcal{G}(u) - y\|_2^2, \qquad \Phi_N(u; y) = \frac{1}{2}\|\mathcal{G}(P_N u) - y\|_2^2,$$

where $P_N : X \mapsto X_N$ is the projection operator that was defined before. Assume that the following conditions are satisfied:

(a) $\|u - P_N u\|_X \leq \|u\|_X \rho(N)$.
(b) $\|\mathcal{G}(u)\|_2 \leq C\tilde{f}_1(\|u\|_X)$ *for all* $u \in X$.
(c) $\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_2 \leq \tilde{f}_2(\max\{\|u_1\|_X, \|u_2\|_X\})\|u_1 - u_2\|_X$ *for all* $u_1, u_2 \in X$.

Here ρ is a positive function such that $\rho(N) \to 0$ as $N \to \infty$, and the functions $\tilde{f}_1, \tilde{f}_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ are nondecreasing and locally bounded and $\tilde{f}_1 \geq 1$. Then

(i) *If $\tilde{f}_1(\|\cdot\|_X)\tilde{f}_2(\|\cdot\|_X) \in L^1(X, \mu_0)$, then there exists $D > 0$ independent of N so that $d_{TV}(\mu^y, \mu_N^y) \leq D\rho(N)$.*
(ii) *If $\tilde{f}_1(\|\cdot\|_X)\tilde{f}_2(\|\cdot\|_X) \in L^2(X, \mu_0)$, then there exists $D' > 0$ independent of N so that $d_H(\mu^y, \mu_N^y) \leq D'\rho(N)$.*

Proof. It follows from Theorem 4.8 that Φ and Φ_N satisfy Assumption 4.2 uniformly in N with $M = 0$, $f_1(x) = 1$. Then the measures μ^y and μ_N^y are well defined for all values of $N \in \mathbb{N}$ by Theorem 4.3. Now it follows from our assumptions on $\mathcal{G}$ that

$$\begin{aligned} 2|\Phi(u; y) - \Phi_N(u; y)| &= \left|\|(\mathcal{G}(u) - y)\|_{\mathbf{\Sigma}}^2 - \|(\mathcal{G}(P_N u) - y)\|_{\mathbf{\Sigma}}^2\right| \\ &= \left|\langle \mathbf{\Sigma}^{-1/2}(\mathcal{G}(u) - \mathcal{G}(P_N u)), \mathbf{\Sigma}^{-1/2}(\mathcal{G}(u) + \mathcal{G}(P_N u) - 2y)\rangle\right| \\ &\leq \left(\|\mathcal{G}(u)\|_{\mathbf{\Sigma}} + \|\mathcal{G}(P_N u)\|_{\mathbf{\Sigma}} + 2\|y\|_{\mathbf{\Sigma}}^2\right) \|(\mathcal{G}(u) - \mathcal{G}(P_N u))\|_{\mathbf{\Sigma}} \\ &\leq C\tilde{f}_1(\|u\|_X)\tilde{f}_2(\|u\|_X)\|u - P_N u\|_X. \end{aligned}$$

The claim will now follow by taking $f_3(x) = \tilde{f}_1(x)\tilde{f}_2(x)$ and applying Theorem 5.2. ■

A few comments are in order concerning the previous theorem. First, the function $\rho(N)$ is independent of the forward map and the prior and depends solely on the topology of X. Then the rate of convergence of μ_N^y to μ^y depends directly on the rate of convergence of P_N to the identity map in the operator norm. Also, observe that to achieve the same rate of convergence in the Hellinger metric as in the total variation metric, we need to impose stronger tail assumptions on the prior μ_0. Finally, we emphasize that the result of Theorem 5.3 relies on the convergence of the P_N to the identity operator as $N \to \infty$. Thus, it is only applicable to discretization by truncation of the basis expansions and does not necessarily hold for other types of discretization. For example, if we use finite difference approximations, then the discretization does not necessarily correspond to truncations of the basis expansions and the discretization operator may not converge to the identity operator.

5.2. Example 1: Deconvolution. We now turn our attention to a few concrete examples of inverse problems with heavy-tailed or non-Gaussian prior measures. We begin with a problem in deconvolution which is a classic example of a linear inverse problem with wide applications in optics and imaging [57, 30]. This problem was also considered in [34] as an example problem with a convex prior measure.

Let $X = L^2(\mathbb{T})$, where $\mathbb{T}$ is the circle of radius $(2\pi)^{-1}$, and let $Y = \mathbb{R}^m$ for a fixed integer m. Suppose that $\eta \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, where $\sigma \in \mathbb{R}$ is a fixed constant and $\mathbf{I}$ is the $m \times m$ identity matrix. Let $S : C(\mathbb{T}) \mapsto \mathbb{R}^m$ be a bounded linear operator that collects point values of a continuous function on a set of m points over $\mathbb{T}$. Given a fixed kernel $\varphi \in C^1(\mathbb{T})$, define the forward map $\mathcal{G} : X \to Y$ as

$$\mathcal{G}(u) = S(\varphi * u), \quad \text{where} \quad (\varphi * u)(t) := \int_{\mathbb{T}} \varphi(t-s)u(t)d\Lambda(s). \tag{5.6}$$

Now suppose that the data y is generated via $y = \mathcal{G}(u) + \eta$ and our goal is to estimate the original image u given noisy point values of its blurred version. Note that our assumptions so far imply a quadratic likelihood potential of the form (4.4).

It follows from Young's inequality [32, Thm. 13.8] that $(\varphi *\cdot) : L^2(\mathbb{T}) \mapsto L^2(\mathbb{T})$ is a bounded linear operator and furthermore, that $(\varphi * u) \in C^1(\mathbb{T})$ for all $u \in L^2(\mathbb{T})$. Since pointwise evaluation is a bounded linear functional on $C^1(\mathbb{T})$, then the forward map $\mathcal{G} : L^2(\mathbb{T}) \mapsto \mathbb{R}^m$ is bounded and linear. We will use the results of section 4.2 to show that this problem is well-posed.

We take our prior measure to be in the class of the product priors of section 2.1. Consider the functions

$$\tilde{w}(t) = \begin{cases} 1 & 0 \le t \le 1, \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \tilde{v}(t) = \begin{cases} 1 & 0 \le t \le 1/2, \\ 1 & 1/2 \le t \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The function $\tilde{v}$ is the Haar wavelet, and $\tilde{w}$ is its corresponding scaling function. Following [24, sect. 9.3], we can define the periodic functions

$$w_{jn}(t) := \sum_{l \in \mathbb{Z}} \phi(2^j(t+l) - n), \qquad v_{jn}(t) := \sum_{l \in \mathbb{Z}} \psi(2^j(t+l) - n),$$

as well as the functions

$$x_1(t) = w_{0,0}(t), \quad x_2(t) = v_{0,0}(t), \quad x_{2^j+n_j+1}(t) = v_{j,n_j}(t)$$

for $j \in \mathbb{Z}_+$ and $n_j = \{0, 1, \ldots, 2^j - 1\}$. The $\{x_k\}$ form an orthonormal basis for $L^2(\mathbb{T})$, and so they can be used in the construction of a $G_{p,q}$-prior.

Now choose $p, q \in (0,1)$ and take the prior μ_0 to be the $G_{p,q}$-prior generated by the wavelet basis $\{x_k\}$ and the fixed sequence $\{\gamma_k\}$, where

$$\gamma_{2^j+n+1} = (1 + |2^{j+1}|^2)^{-1/2} \qquad \forall n \in \mathbb{Z}_+.$$

Clearly, $\{\gamma_k\} \in \ell^2$, and so it follows from Theorem 2.1 that $\|u\|_{L^2(\mathbb{T})} < \infty$ a.s. Furthermore, we know that the $G_{p,q}$-priors have bounded moments of order two. Putting this together with the fact that the forward map $\mathcal{G}$ is bounded and linear, we immediately obtain the well-posedness of this inverse problem using Theorem 4.11.

5.3. Example 2: Deconvolution with a BV-prior. We now formulate the deconvolution problem of Example 1 with a prior measure that is supported on $BV(\mathbb{T})$ using the stochastic process priors of section 3. Let $u(t)$ for $t \in [0,1]$ denote a stochastic process such that

$$u(0) = 0, \qquad \hat{u}_t(s) \exp\left(t \int_{\mathbb{R}} \exp(i\xi s) - 1 \, \mathrm{d}\nu(\xi) \right), \quad s \in \mathbb{R},$$

where the measure $\nu = c\mathcal{N}(0,1)$ with a fixed constant $c \in (0,\infty)$. Then u is a compound Poisson process with piecewise constant sample paths and normal jumps. We can write $u(t) = \sum_{k=1}^{\tau(t)} \xi_k$, where $\{\xi_k\}$ are an i.i.d. sequence of standard normal random variables and $\tau(t)$ is a Poisson process with intensity c. In section 3 we saw that this process has piecewise constant sample paths and its law is a Radon measure on $BV([0,1])$.

Let us denote the law of this process by $\tilde{\mu}_0$. The next step is to use this measure to define a new measure μ_0 on $BV(\mathbb{T})$. Take μ_0 to be the law of the periodic versions of the sample paths of the above compound Poisson process u on the interval $[0,1]$. We can write $\mu_0 = \tilde{\mu}_0 \circ T^{-1}$, where $T : BV([0,1]) \mapsto BV(\mathbb{T})$ is a bounded and linear operator. Thus, μ_0 is a Radon measure on $BV(\mathbb{T})$. With an abuse of notation we use u to denote the corresponding periodic processes on $\mathbb{T}$. Since the convolution kernel $\varphi \in C^1(\mathbb{T})$ and $BV(\mathbb{T}) \subset L^1(\mathbb{T})$, the forward map $\mathcal{G} : BV(\mathbb{T}) \mapsto \mathbb{R}^m$ (given by(5.6)) is well defined, bounded, and linear, and so the likelihood potential has the form (4.4) once more. We have shown in section 3 that $\mathbb{E}\,\|u\|_{BV(\mathbb{T})} < \infty$. Putting this together with the fact that $\mathcal{G} : BV(\mathbb{T}) \mapsto \mathbb{R}^m$ is bounded and linear, we immediately obtain the well-posedness of this inverse problem via Corollary 4.10.

5.4. Example 3: Quadratic measurements of a continuous field. As our final example, we will consider a problem with a nonlinear forward map. Our goal is to estimate a continuous field from quadratic measurements of its point values. This inverse problem was encountered in [33] in recovery of aberrations in high intensity focused ultrasound treatment, and it is closely related to the phase retrieval problem [21, 31, 15]. Let $X = C(\mathbb{T})$, and let $\{t_k\}_{k=1}^n$ be a collection of distinct points in $\mathbb{T}$. Now define the operator

$$S : C(\mathbb{T}) \mapsto \mathbb{R}^n, \qquad (S(u))_j = u(t_j), \quad j = 1, 2, \ldots, n.$$

This operator collects point values of functions in $C(\mathbb{T})$. Let $\{z_k\}_{k=1}^m$ be a fixed collection of vectors $z_k \in \mathbb{R}^n$, and define the forward map

$$\mathcal{G} : C(\mathbb{T}) \mapsto \mathbb{R}^m, \qquad (\mathcal{G}(u))_j := |z_j^T S(u)|^2 \quad \text{for} \quad j = 1, 2, \ldots, m,$$

which collects quadratic measurements of the point values of a continuous function. We complete our model of the measurements with an additive layer of Gaussian noise

$$y = \mathcal{G}(u) + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma^2 \mathbf{I}),$$

where $\sigma > 0$. Our goal in this problem is to infer the function $u \in C(\mathbb{T})$ from the quadratic measurements y.

A straightforward calculation shows that

$$\|\mathcal{G}(u)\|_2 \le \tilde{K} \|S(u)\|_2^2 \le K \|u\|_{C(\mathbb{T})}^2, \tag{5.7}$$

where $\tilde{K}, K > 0$ are constants that are independent of u but depend on the z_k. The last inequality follows because pointwise evaluation is a bounded linear operator on $C(\mathbb{T})$.

Furthermore, we have that for $u_1, u_2 \in C(\mathbb{T})$

$$\begin{aligned} (\mathcal{G}(u_1) - \mathcal{G}(u_2))_j &= (z_j^T(S(u_1 - u_2)))(z_j^T(S(u_1) + S(u_2))) \\ &\le D_j(\max\{\|u_1\|_{C(\mathbb{T})}, \|u_2\|_{C(\mathbb{T})}\})\|u_1 - u_2\|_{C(\mathbb{T})}. \end{aligned}$$

Here, the constant $D_j > 0$ depends on z_j. We can now use this bound to obtain

$$\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_2 \le D(\max\{\|u_1\|_{C(\mathbb{T})}, \|u_2\|_{C(\mathbb{T})}\})\|u_1 - u_2\|_{C(\mathbb{T})}, \tag{5.8}$$

where the constant $D > 0$ will depend only on the D_j. Observe that the above bounds in (5.7) and (5.8) imply that $\mathcal{G}$ satisfies the conditions of Theorem 4.8 with a function $\tilde{f}(x) = x^2$. Therefore, that theorem implies that the likelihood function Φ for our problem will satisfy Assumption 4.2(iv) with $f_2(x) = 1 + x^2$. Now we use Corollary 4.7 to infer that well-posedness can be achieved if we choose a prior measure μ_0 for which $f_2(\|\cdot\|) = 1 + \|\cdot\|_{C(\mathbb{T})}^2 \in L^1(C(\mathbb{T}), \mu_0)$.

To construct such a prior measure μ_0, we will consider a product prior with samples of the form

$$u \sim \sum_{k \in \mathbb{Z}} \gamma_k \xi_k w_k, \quad \text{where} \quad w_k(t) = (2\pi)^{-1/2} \exp(2\pi i k t).$$

The $\{w_k\}$ are simply the Fourier basis on $\mathbb{T}$. Our plan is to construct the prior measure to be supported on a sufficiently regular Sobolev space that is embedded in $C(\mathbb{T})$. The reason for going through the Sobolev space is the fact that $C(\mathbb{T})$ does not have an unconditional Schauder basis and so we cannot directly apply the methodology of section 2.1.

To this end, we choose

$$\gamma_k = (1 + |k|^2)^{-3/2}, \quad k \in \mathbb{Z},$$

and suppose that the $\{\xi_k\}$ are i.i.d. and $\xi_1 \sim \mathrm{CPois}(0, \mathrm{Lap}(0, 1))$ (recall Definition A.4), where $\mathrm{Lap}(0, 1)$ is the standard Laplace distribution on the real line with Lebesgue density $\pi(x) = \frac{1}{2}\exp(-|x|)$ which clearly has exponential tails, and this, in turn, implies that $\mathbf{Var}\xi_1 < \infty$. Note that the random variables ξ_k have a positive probability of being zero, and hence draws from this prior will incorporate a certain level of sparsity. Observe that this is a different type of sparsity in comparison to the $G_{p,q}$-prior. Samples from this compound Poisson prior have a nonzero probability of having modes that are exactly zero. The samples from the $G_{p,q}$-prior have a zero probability of having modes that are exactly zero, and instead most of their modes will concentrate in a neighborhood of zero.

The Sobolev space $H^1(\mathbb{T})$ is defined as

$$H^1(\mathbb{T}) := \left\{ v \in L^2(\mathbb{T}) : \|v\|^2_{H^1(\mathbb{T})} := \sum_{k\in\mathbb{Z}} (1+|k|^2)|\langle v, w_k\rangle|^2 < \infty \right\},$$

where $\langle\cdot,\cdot\rangle$ is the usual $L^2(\mathbb{T})$ inner product. Now consider $u \sim \mu_0$; then

$$\|u\|^2_{H^1(\mathbb{T})} = \sum_{k\in\mathbb{Z}} (1+|k|^2)^{-1}|\xi_k|^2.$$

But $\{(1+|k|^2)^{-1}\} \in \ell^1$ and $\mathbf{Var}|\xi_k|^2 < \infty$, and so it follows from Theorem 2.1 that $\|u\|^2_{H^1(\mathbb{T})} < \infty$ a.s. Now the Sobolev embedding theorem [53, Prop. 3.3] guarantees that $\|u\|_{C(\mathbb{T})} < \infty$ a.s., and it follows from Theorem 2.3 that $\|u\|^2_{C(\mathbb{T})} \in L^1(C(\mathbb{T}), \mu_0)$.

6. Closing remarks.

At the beginning of this paper we set out to achieve three goals. We introduced the new classes $G_{p,q}$-, W_p-, and ℓ_p-prior measures in connection with ℓ_p-regularization techniques and showed that these prior measures belong to the larger class of ID measures. This motivated our study of the class of ID priors (G1). Afterwards, we introduced another class of prior measures based on the laws of pure jump Lévy processes (G2). Our goal here was to construct a well-defined alternative to the classic total variation prior. Finally, we presented a theory of well-posedness for Bayesian inverse problems that was general enough to include the new classes of prior measures that were introduced previously (G3). Our approach to well-posedness theory was to identify the minimal restrictions on the prior measure given a choice of the likelihood potential Φ. A common theme in our results was the trade-off between the tail decay of the prior and the growth of the likelihood potential. As an example, we considered the setting where the likelihood had a quadratic form and the forward map was linear. This example corresponds to linear inverse problems with additive Gaussian noise that are of great interest in practice. We showed that in this simple setting well-posedness can be achieved if the prior has moments of order one.

Finally, we considered some practical aspects of solving inverse problems with heavy-tailed or ID priors. We discussed consistent discretization of inverse problems and the use of projections in discretization of the likelihood. Afterwards, we presented three concrete examples of inverse problems that used heavy-tailed or ID prior measures. In particular, we studied the well-posedness of a deconvolution problem with a Lévy process prior that was cast on the nonseparable space $BV(\mathbb{T})$.

The results of this paper open the door for the use of large classes of prior measures in inverse problems, and they can be extended in several directions. For example, we showed that if the forward problem is linear and the measurement noise is Gaussian, then one can achieve well-posedness for priors that have poor tail behavior. Then many of the common heavy-tailed priors can be used to model sparsity in the linear case. But it is not clear which prior is the optimal choice and in what sense. Furthermore, given that the compressed sensing literature is focused mainly on recovery of sparse signals from linear measurements, it is interesting to study the implications of the compressed sensing theory in the setting of Bayesian inverse problems. Throughout the paper we mentioned the issue of sparsity on several occasions, but this is not the only setting where non-Gaussian priors can be useful. For example, non-Gaussian priors can be used in modelling of constraints or in construction of hierarchical models. Finally, a major issue when it comes to using non-Gaussian priors in practice is that of sampling. For example, even in finite dimensions, the $G_{p,q}$-priors are far from a Gaussian measure. Then we expect that Metropolis–Hastings algorithms that utilize a Gaussian proposal will have poor performance in sampling from posteriors that arise from $G_{p,q}$-priors. This issue will become worse as the dimension of the parameter space grows. Therefore, new sampling techniques that are tailor made to these non-Gaussian priors are needed if we wish to apply them in real world situations.

Appendix A. ID measures. Here we briefly discuss some of the useful properties of ID measures. The main reference for this material is the monograph [41]. The proof of the next theorem can be found in [41, sect. 5.1].

Theorem A.1. *Let μ be an ID probability measure on a Banach space X. Then the following hold:*

(i) $\hat{\mu}(\varrho) \neq 0$ *for all* $\varrho \in X^*$.
(ii) *There exists a unique and continuous (in the dual norm) function* $\psi : X^* \mapsto X$ *so that* $\hat{\mu}(\varrho) = \exp(\psi(\varrho))$ *and* $\psi(0) = 0$.
(iii) *If μ is symmetric, i.e., $\mu(A) = \mu(-A)$ for all Borel subsets A of X, then $\hat{\mu}$ is real valued and positive.*
(iv) *For every $n \in \mathbb{N}$, the measures $\mu_{1/n}$ (see Definition 2.8) are uniquely determined, and* $\hat{\mu}_{1/n}(\varrho) = \exp(n^{-1}\psi(\varrho))$ *for all* $\varrho \in X^*$.

Furthermore, we define the function

$$\Psi(u, \varrho) := \exp(i\varrho(u)) - 1 - i\varrho(u)\mathbf{1}_{B_X}(u) \qquad \forall u \in X,\ \varrho \in X^*,$$

where B_X is the unit ball in X and $\mathbf{1}_{B_X}$ is the characteristic function of the unit ball. We recall the definition of a Lévy measure on a Banach space.

Definition A.2 (Lévy measure). *A positive σ-finite Radon measure λ on X is called a Lévy measure if and only if*

1. $\lambda(\{0\}) = 0$,
2. $\int_X |\Psi(u, \varrho)| d\lambda(u) < \infty$ *for every* $\varrho \in X^*$,
3. $\exp(\int_X \Psi(u, \varrho) d\lambda(u))$ *is the characteristic function of a Radon probability measure on* $\mathbb{R}$ *for every* $\varrho \in X^*$.

We are now in a position to recall the celebrated Lévy–Khintchine representation theorem (see [41, sect. 5.7] for a proof).

Theorem A.3 (Lévy–Khintchine representation). *A Radon probability measure on a Banach space X is ID if and only if there exist an element $m \in X$, a (positive definite) covariance operator $\mathcal{R} : X^* \mapsto X$, and a Lévy measure λ so that*

$$\hat{\mu}(\varrho) = \exp(\psi(\varrho)), \quad \text{where} \quad \psi(\varrho) = i\varrho(m) - \frac{1}{2}\varrho(\mathcal{R}(\varrho)) + \int_X \Psi(u, \varrho) d\lambda(u). \tag{A.1}$$

Equivalently, μ is ID precisely when there exist a point mass δ_m, a Gaussian measure $\mathcal{N}(0,\mathcal{R})$, and a Radon measure ν identified via $\hat{\nu}(\varrho) = \int_X \Psi(u,\rho) d\lambda(u)$ so that

$$\mu = \delta_m * \mathcal{N}(0,\mathcal{R}) * \nu. \tag{A.2}$$

To gain more insight into the implications of the Lévy–Khintchine representation, we recall the class of compound Poisson random variables and their corresponding probability measures.

Definition A.4 (compound Poisson probability measure [41, sect. 5.3]). *Let η be a Radon probability measure on a Banach space X, and suppose that $\{u_k\}$ are a sequence of i.i.d. random variables so that $u_k \sim \eta$. Also, let τ be an independent Poisson random variable with rate $c > 0$ taking values in $\mathbb{Z}_+$. Then $u = \sum_{k=0}^{\tau} u_k$ is distributed according to a compound Poisson probability measure denoted by* $\mathrm{CPois}(c,\eta)$.

It is straightforward to check that the characteristic function of a compound Poisson measure has the form

$$\widehat{\mathrm{CPois}(c,\eta)}(\varrho) = \exp\left(c \int_X (\exp(i\varrho(u)) - 1)\, \mathrm{d}\eta(u) \right) \quad \forall \varrho \in X^*.$$

See [41, Prop. 5.3.1] for a proof of this formula along with the fact that $\mathrm{CPois}(c,\eta)$ is a Radon measure on X.

Now let us return to the characteristic function of the probability measure ν that was introduced in the Lévy–Khintchine representation (A.2):

$$\hat{\nu}(\varrho) = \exp\left(\int_X (\exp(i\varrho(u)) - 1)\, \mathrm{d}\lambda(u) \right) \exp\left(\int_{B_X} -i\varrho(u)\, \mathrm{d}\lambda(u) \right). \tag{A.3}$$

If $0 < \lambda(X) < \infty$, then λ can be renormalized to define a probability measure $\tilde{\lambda} := \frac{1}{\lambda(X)}\lambda$. Furthermore, we can define an element $u_\lambda \in X$ so that

$$\varrho(u_\lambda) = -\int_X \varrho(u) \mathbf{1}_{B_X}(u) \mathrm{d}\lambda(u) \quad \forall \varrho \in X^*.$$

Putting these observations together with (A.3) gives the decomposition

$$\nu = \mathrm{CPois}(\lambda(X), \tilde{\lambda}) * \delta_{u_\lambda}. \tag{A.4}$$

Therefore, from (A.2) we deduce that any measure $\mu = \mathrm{ID}(m,\mathcal{R},\lambda)$ with $\lambda(X) < \infty$ can be decomposed as

$$\mu = (\delta_{m+u_\lambda}) * \mathcal{N}(0,\mathcal{R}) * \mathrm{CPois}(\lambda(X), \tilde{\lambda}). \tag{A.5}$$

Acknowledgments. The author owes a debt of gratitude to Prof. Nilima Nigam for many useful discussions and comments. We are also thankful to the anonymous reviewers for their questions and suggestions that helped significantly improve this manuscript.

REFERENCES

[1] R. J. Adler, *The Geometry of Random Fields*, Classics Appl. Math. 62, SIAM, Philadelphia, 2010, https://doi.org/10.1137/1.9780898718980.

[2] C. D. Aliprantis and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed., Springer Science & Business Media, New York, 2006, https://doi.org/10.1007/3-540-29587-9.

[3] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge Stud. Adv. Math. 93, Cambridge University Press, Cambridge, UK, 2009, https://doi.org/10.1017/CBO9780511809781.

[4] J.-M. Azaïs and M. Wschebor, *Level Sets and Extrema of Random Processes and Fields*, John Wiley & Sons, Hoboken, NJ, 2009, https://doi.org/10.1002/9780470434642.

[5] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, Wiley Ser. Probab. Statist., John Wiley & Sons, Chichester, UK, 2000, https://doi.org/10.1002/9780470316870.

[6] P. Billingsley, *Probability and Measure*, Wiley Ser. Probab. Math. Statist., 3rd ed., John Wiley & Sons, New York, 1995.

[7] V. I. Bogachev, *Gaussian Measures*, Math. Surveys Monogr. 62, American Mathematical Society, Providence, RI, 1998, https://doi.org/10.1090/surv/062.

[8] V. I. Bogachev, *Measure Theory*, Vol. 1, Springer-Verlag, Berlin, 2007, https://doi.org/10.1007/978-3-540-34514-5.

[9] V. I. Bogachev, *Measure Theory*, Vol. 2, Springer-Verlag, Berlin, 2007.

[10] L. Bondesson, *A general result on infinite divisibility*, Ann. Probab., 7 (1979), pp. 965–979.

[11] C. Borell, *Convex measures on locally convex spaces*, Ark. Mat., 12 (1974), pp. 239–252, https://doi.org/10.1007/BF02384761.

[12] G. Buttazzo, M. Giaquinta, and S. Hildebrandt, *One-Dimensional Variational Problems: An Introduction*, Oxford Lecture Ser. Math. Appl. 15, Oxford University Press, Oxford, UK, 1998.

[13] D. Calvetti, J. P. Kaipio, and E. Somersalo, *Inverse problems in the Bayesian framework*, Inverse Problems, 30 (2014), 110301.

[14] D. Calvetti and E. Somersalo, *An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, Surv. Tutor. Appl. Math. Sci. 2, Springer Science+Business Media, New York, 2007, https://doi.org/10.1007/978-0-387-73394-4.

[15] E. J. Candes, T. Strohmer, and V. Voroninski, *Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming*, Comm. Pure Appl. Math., 66 (2013), pp. 1241–1274, https://doi.org/10.1002/cpa.21432.

[16] C. M. Carvalho, N. G. Polson, and J. G. Scott, *The horseshoe estimator for sparse signals*, Biometrika, 97 (2010), pp. 465–480, https://doi.org/10.1093/biomet/asq017.

[17] I. Castillo, J. Schmidt-Hieber, and A. van der Vaart, *Bayesian linear regression with sparse priors*, Ann. Statist., 43 (2015), pp. 1986–2018, https://doi.org/10.1214/15-AOS1334.

[18] I. Castillo and A. van der Vaart, *Needles and straw in a haystack: Posterior concentration for possibly sparse sequences*, Ann. Statist., 40 (2012), pp. 2069–2101, https://doi.org/10.1214/12-AOS1029.

[19] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC Financ. Math. Ser., CRC Press, Boca Raton, FL, 2004, https://doi.org/10.1201/9780203485217.

[20] S. L. Cotter, M. Dashti, and A. M. Stuart, *Approximation of Bayesian inverse problems for PDEs*, SIAM J. Numer. Anal., 48 (2010), pp. 322–345, https://doi.org/10.1137/090770734.

[21] J. C. Dainty and J. R. Fienup, *Phase retrieval and image reconstruction for astronomy*, in Image Recovery: Theory and Application, Academic Press, Orlando, FL, 1987, pp. 231–275.

[22] M. Dashti, S. Harris, and A. M. Stuart, *Besov priors for Bayesian inverse problems*, Inverse Probl. Imaging, 6 (2012), pp. 183–200, https://doi.org/10.3934/ipi.2012.6.183.

[23] M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, in Handbook of Uncertainty Quantification, R. Ghanem, D. Higdon, and H. Owhadi, eds., Springer International, Cham, Switzerland, 2016, pp. 1–118, https://doi.org/10.1007/978-3-319-11259-6_7-1.

[24] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 61, SIAM, Philadelphia, 1992, https://doi.org/10.1137/1.9781611970104.

[25] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Textb. Math., revised ed., CRC Press, Boca Raton, FL, 2015.

[26] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Appl. Numer. Harmon. Anal., Birkhäuser/Springer, New York, 2013, https://doi.org/10.1007/978-0-8176-4948-7.
[27] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Dover, New York, 2003, https://doi.org/10.1007/978-1-4612-3094-6.
[28] P. Ghosh and A. Chakrabarti, *Posterior Concentration Properties of a General Class of Shrinkage Priors around Nearly Black Vectors*, preprint, 2014, https://arxiv.org/abs/1412.8161.
[29] H. Hanche-Olsen and H. Holden, *The Kolmogorov–Riesz compactness theorem*, Expo. Math., 28 (2010), pp. 385–394, https://doi.org/10.1016/j.exmath.2010.03.001.
[30] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006, https://doi.org/10.1137/1.9780898718874.
[31] R. W. Harrison, *Phase problem in crystallography*, J. Opt. Soc. Amer. A, 10 (1993), pp. 1046–1055.
[32] C. Heil, *A Basis Theory Primer: Expanded Edition*, Appl. Numer. Harmon. Anal., Springer Science+Business Media, New York, 2010, https://doi.org/10.1007/978-0-8176-4687-5.
[33] B. Hosseini, C. Mougenot, S. Pichardo, E. Constanciel, J. M. Drake, and J. M. Stockie, *A Bayesian Approach for Energy-Based Estimation of Acoustic Aberrations in High Intensity Focused Ultrasound Treatment*, preprint, 2016, https://arxiv.org/abs/1602.08080.
[34] B. Hosseini and N. Nigam, *Well-posed Bayesian inverse problems: Priors with exponential tails*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 436–465, https://doi.org/10.1137/16M1076824.
[35] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume* 1, 2nd ed., John Wiley & Sons, New York, 1994.
[36] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Appl. Math. Sci. 160, Springer Science+Business Media, New York, 2005, https://doi.org/10.1007/b138659.
[37] S. Kotz, T. J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Springer Science+Business Media, New York, 2012, https://doi.org/10.1007/978-1-4612-0173-1.
[38] V. Kruglov, *A note on infinitely divisible distributions*, Theory Probab. Appl., 15 (1970), pp. 319–324, https://doi.org/10.1137/1115038.
[39] M. Lassas and S. Siltanen, *Can one use total variation prior for edge-preserving Bayesian inversion?*, Inverse Problems, 20 (2004), pp. 1537–1563, https://doi.org/10.1088/0266-5611/20/5/013.
[40] G. Leoni, *A First Course in Sobolev Spaces*, Grad. Stud. Math. 105, American Mathematical Society, Providence, RI, 2009, http://dx.doi.org/10.1090/gsm/105.
[41] W. Linde, *Probability in Banach Spaces: Stable and Infinitely Divisible Distributions*, John Wiley & Sons, New York, 1986.
[42] F. Lucka, *Bayesian Inversion in Biomedical Imaging*, Ph.D. thesis, University of Münster, Münster, Germany, 2014; available online from http://nbn-resolving.de/urn:nbn:de:hbz:6-80359613770.
[43] S. Nadarajah, *The Kotz-type distribution with applications*, Statistics, 37 (2003), pp. 341–358, https://doi.org/10.1080/0233188031000078060.
[44] S. Nadarajah, *A generalized normal distribution*, J. Appl. Stat., 32 (2005), pp. 685–694, https://doi.org/10.1080/02664760500079464.
[45] C. J. Paciorek, *Nonstationary Gaussian Processes for Regression and Spatial Modelling*, Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, 2003; available online from http://www.stat.berkeley.edu/~paciorek/diss/paciorek-thesis.pdf.
[46] S. Peszat and J. Zabczyk, *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*, Encyclopedia Math. Appl. 113, Cambridge University Press, Cambridge, UK, 2007.
[47] N. G. Polson and J. G. Scott, *Shrink globally, act locally: Sparse Bayesian regularization and prediction*, in Bayesian Statistics 9 (Benidorm, 2010), Oxford University Press, Oxford, UK, 2011, pp. 501–538, https://doi.org/10.1093/acprof:oso/9780199694587.003.0017.
[48] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
[49] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge Stud. Adv. Math. 68, Cambridge University Press, Cambridge, UK, 1999.
[50] F. W. Steutel and K. van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, Monogr. Textbooks Pure Appl. Math. 259, Marcel Dekker, New York, 2004, https://doi.org/10.1201/9780203014127.

[51] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061.

[52] T. J. Sullivan, *Well-Posed Bayesian Inverse Problems and Heavy-Tailed Stable Banach Space Priors*, preprint, 2016, https://arxiv.org/abs/1605.05898.

[53] M. E. Taylor, *Partial Differential Equations* I. *Basic Theory*, Appl. Math. Sci. 115, 2nd ed., Springer Science+Business Media, New York, 2011, https://doi.org/10.1007/978-1-4419-7055-8.

[54] M. Unser and P. Tafti, *An Introduction to Sparse Stochastic Processes*, Cambridge University Press, Cambridge, UK, 2013.

[55] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, *A unified formulation of Gaussian versus sparse stochastic processes. Part* II*: Discrete-domain theory*, IEEE Trans. Inform. Theory, 60 (2014), pp. 3036–3051, https://doi.org/10.1109/TIT.2014.2311903.

[56] M. Unser, P. D. Tafti, and Q. Sun, *A unified formulation of Gaussian versus sparse stochastic processes. Part* I*: Continuous-domain theory*, IEEE Trans. Inform. Theory, 60 (2014), pp. 1945–1962, https://doi.org/10.1109/TIT.2014.2298453.

[57] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898717570.

[58] Z. Yao, Z. Hu, and J. Li, *A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations*, Inverse Problems, 32 (2016), 075006, https://doi.org/10.1088/0266-5611/32/7/075006.