T. J. Sullivan

Well-posed Bayesian inverse problems and heavy-tailed stable Banach space priors

T. J. Sullivan, Free University of Berlin and Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany. sullivan@zib.de

Well-posed Bayesian inverse problems and heavy-tailed stable Banach space priors

T. J. Sullivan

30. Mai 2016

Zusammenfassung

This article extends the framework of Bayesian inverse problems in infinite-dimensional parameter spaces, as advocated by Stuart (*Acta Numer.* 19:451–559, 2010) and others, to the case of a heavy-tailed prior measure in the family of stable distributions, such as an infinite-dimensional Cauchy distribution, for which polynomial moments are infinite or undefined. It is shown that analogues of the Karhunen–Loève expansion for square-integrable random variables can be used to sample such measures. Furthermore, under weaker regularity assumptions than those used to date, the Bayesian posterior measure is shown to depend Lipschitz continuously in the Hellinger metric upon perturbations of the misfit function and observed data.

1 Introduction

The Bayesian perspective on inverse problems has attracted much mathematical attention in recent years (Kaipio and Somersalo, 2005; Stuart, 2010). Particular attention has been paid to Bayesian inverse problems (BIPs) in which the parameter to be inferred lies in an infinite-dimensional space, a typical example being a scalar or tensor field coupled to some observed data via an ordinary or partial differential equation. Numerical solution of such infinite-dimensional BIPs must necessarily be performed in an approximate manner on a finite-dimensional subspace, but it is profitable to delay discretisation to the last possible moment and consider the original infinite-dimensional problem as the primary object of study, since infinite-dimensional well-posedness results and algorithms descend to any finite-dimensional subspace in a discretisation-independent way, whereas careless early discretisation may lead to a sequence of well-posed finite-dimensional BIPs or algorithms whose stability properties degenerate as the discretisation dimension increases. Well-posedness results have been established for infinite-dimensional Gaussian priors by Stuart (2010), for Besov priors by Dashti et al. (2012), and for log-concave priors by Hosseini and Nigam (2016). There is a parallel approach of discretisation invariance, introduced by Lehtinen in the 1990s and advanced by e.g. Lassas et al. (2009), in which the finite-dimensional BIP is the primary object, but care is taken to ensure the existence of a well-defined continuum limit independent of the discretisation. A common assumption in these works is some exponential integrability of the prior, and one purpose of

this article is to relax this by permitting the prior to be heavy-tailed in the sense of only having finite polynomial moments of order $0 \leq p < \alpha$ for some $\alpha < \infty$.

A prototypical heavy-tailed prior on $\mathbb{R}$ is the Cauchy distribution with location $\delta \in \mathbb{R}$ and width $\gamma > 0$, here denoted $\mathcal{C}(\delta, \gamma)$, which has the Lebesgue density

$$\frac{\mathrm{d}\mathcal{C}(\delta, \gamma)}{\mathrm{d}u}(u) = \frac{1}{\gamma\pi}\frac{1}{1 + ((u - \delta)/\gamma)^2}. \tag{1.1}$$

$\mathcal{C}(\delta, \gamma)$ arises straightforwardly as the distribution of the ratio of two independent Gaussian random variables:

$$\delta + \frac{x}{z} \sim \mathcal{C}(\delta, \gamma) \text{ when } x \sim \mathcal{N}(0, \gamma^2), z \sim \mathcal{N}(0, 1) \text{ are independent.} \tag{1.2}$$

$\mathcal{C}(\delta, \gamma)$ has no well-defined mean, even though it is 'obviously' centred on δ, nor indeed polynomial moments of any order greater than $\alpha = 1$. Despite this, the Cauchy distribution arises naturally in even quite elementary applications. For example, Cauchy distributions arise naturally from quotients of Gaussian random variables, as in (1.2). More geometrically, if uniform measure on a circle is projected radially onto any line not passing through the centre of the circle, as in Figure 1.1, then the image measure is Cauchy. Markkanen et al. (2016) have recently proposed the use of heavy-tailed priors for edge-preserving Bayesian inversion in X-ray tomography, where the seemingly natural choice of a total variation regularisation term cannot be interpreted as a discretisation-invariant Bayesian prior (Lassas and Siltanen, 2004).

In a Bayesian context, the use of a heavy-tailed prior in preference to one with exponentially small tails corresponds to a prior belief that large deviations are not exponentially rare events. More concretely, in e.g. a wavelet basis of $L^2([0, 1], \mathrm{d}x)$, it is not rare to draw samples with localised very large deviations, as shown in Figure 1.2; physically, these might correspond to inclusions in an otherwise relatively homogeneous material matrix. The difference between the two modelling assumptions can be starkly illustrated the following information-theoretic calculation: the Kullback–Leibler divergences (relative entropy distances) between a standard normal and a standard Cauchy distribution on $\mathbb{R}$ are

$$D_{\mathrm{KL}}(\mathcal{N}(0, 1)\|\mathcal{C}(0, 1)) \approx 0.2592 < \infty \quad \text{but} \quad D_{\mathrm{KL}}(\mathcal{C}(0, 1)\|\mathcal{N}(0, 1)) = \infty.$$

Thus, the approximation of a heavy-tailed Cauchy prior by a thin-tailed Gaussian prior represents an infinite loss of information. However, asymmetrically, the 'defensive' adoption of a Cauchy prior in place of a Gaussian one represents a mild loss of information, with which one gains access to large deviations that would be exponentially rare in the Gaussian model.

The family of stable distributions generalises both the Cauchy and Gaussian examples. Because the stable family is, by definition, closed under linear combinations of independent members, it is an attractive model for spatial or temporal phenomena that decompose in an additive way over disjoint subsets of space or time. So, for example, a stable distribution is a natural modelling choice for the net external forces imparted on a passive tracer particle in some medium over a time interval: a Gaussian model leads to Brownian motion, whereas other stable models lead to Lévy flights.

Thus, after establishing some background and notation in Section 2, the purpose of this article is twofold:

Section 3 shows how to define Banach space analogues of heavy-tailed stable distributions via Karhunen–Loève-like random series, and studies their convergence and integrability properties. The usual variance-based arguments cannot be applied directly, but the situation can be rescued using Kolmogorov's three series theorem, and notably the same conditions on the decay of the coefficients suffice for the heavy-tailed stable case as in the Gaussian case.

Section 4 shows that the usual results on the Hellinger well-posedness of BIPs with respect to perturbations of the observed data and the misfit functional (negative log-likelihood) hold even in the case of a heavy-tailed prior, under weaker continuity assumptions than those used to date. Non-trivial growth of lower bounds on the misfit functional, which is typically enjoyed in applications, can and should be used to offset growth in other errors and retain well-posedness of the BIP.

2 Background and Notation

2.1 General Notation

The setting for the inference problems in this paper will be a real and separable Banach space $\mathcal{U}$. Observed data will take values in another real and separable Banach space $\mathcal{Y}$. Occasionally, we will need to make reference to an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as a common domain of definition for all the $\mathbb{R}$-, $\mathcal{U}$, and $\mathcal{Y}$-valued random variables of interest. $\mathbf{1}[P]$ denotes the indicator function of a measurable set or logical predicate P, e.g.

$$\mathbf{1}[x \in E] := \begin{cases} 1, & \text{if } x \in E, \\ 0, & \text{if } x \notin E. \end{cases}$$

A property will be said to hold almost surely if it fails only on a subset of a measurable set of measure zero, and this will be abbreviated to "a.s." If $f \colon \mathcal{U} \to \mathbb{R}$ is measurable, then $\mathbb{E}_{u \sim \mu}[f(u)]$ or simply $\mathbb{E}[f]$ denotes the expected value (Lebesgue integral) of f with respect to μ:

$$\mathbb{E}[f] \equiv \mathbb{E}_{u \sim \mu}[f(u)] := \int_{\mathcal{U}} f(u) \, \mathrm{d}\mu(u).$$

The set of all Borel probability measures on $\mathcal{U}$ will be denoted $\mathcal{M}_1(\mathcal{U})$, and d_{H} denotes the Hellinger metric on $\mathcal{M}_1(\mathcal{U})$, defined by

$$d_{\mathrm{H}}(\mu, \nu)^2 = \int_{\mathcal{U}} \left| \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(u)} - \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(u)} \right|^2 \, \mathrm{d}\lambda(u), \qquad (2.1)$$

where λ is any σ-finite Borel measure on $\mathcal{U}$ with respect to which both μ and ν are absolutely continuous, e.g. $\lambda := \mu + \nu$. By Kraft's inequality, the Hellinger topology coincides with the total variation topology; by Pinsker's inequality, the Hellinger topology is strictly weaker than that generated by the Kullback–Leibler divergence (the relative entropy distance); all these topologies are strictly stronger than the topology of weak convergence of measures. Expected values of square-integrable functions are Lipschitz continuous with respect to the Hellinger metric:

$$\left| \mathbb{E}_{\mu}[f] - \mathbb{E}_{\nu}[f] \right| \leq \sqrt{2} \sqrt{\mathbb{E}_{\mu}\big[|f|^2\big] + \mathbb{E}_{\nu}\big[|f|^2\big]} \, d_{\mathrm{H}}(\mu, \nu) \qquad (2.2)$$
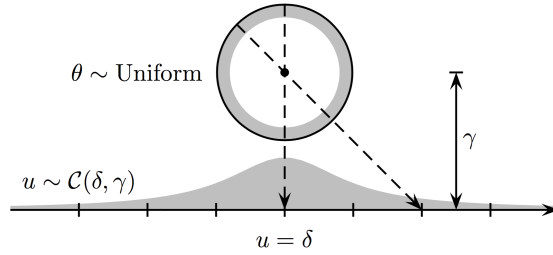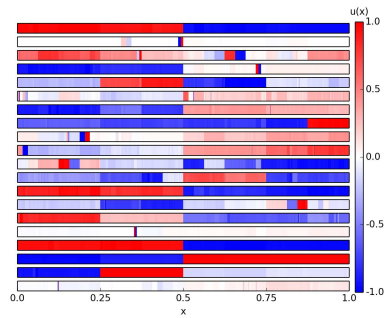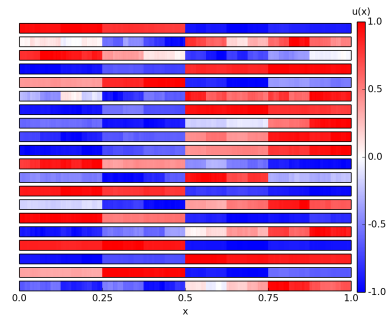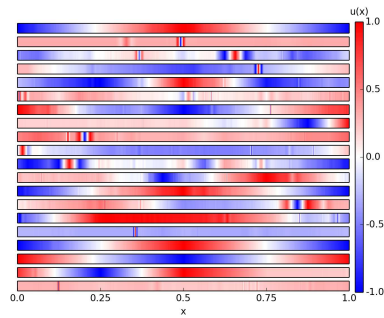
Abbildung 1.1: Uniform angular measure on a circle projects radially to give Cauchy measure with width parameter γ on any line at distance γ from the centre of the circle.
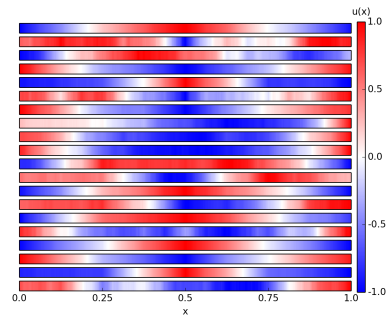

(a) Haar wavelets, Cauchy coefficients

(b) Haar wavelets, Gaussian coefficients

(c) Linear splines, Cauchy coefficients

(d) Linear splines, Gaussian coefficients

Abbildung 1.2: Cauchy and Gaussian wavelet expansions in the Haar and linear spline orthonormal bases of $L^2([0,1], \mathrm{d}x)$. Each horizontal stripe shows a random function $u(x) = \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} u_{j,k} 2^{j/2} \psi(2^j x - k)$, where each $u_{j,k} = (j+1)^{-2} 2^{-j}$ times a standard Cauchy or normal draw, and ψ denotes the mother wavelet. The plots show 20 i.i.d. samples with $J = 10$. Theorem 3.3 ensures a.s. convergence in $L^2([0,1])$ as $J \to \infty$. To enable easy comparisons between plots, the ensemble has been translated and linearly scaled to take values $u(x) \in [0,1]$, and the same random seed is used in each case. Note well the large local deviations in the Cauchy case.

when $f \in L^2(\mathcal{U}, \mu) \cap L^2(\mathcal{U}, \nu)$. In particular, $|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]| \le 2\|f\|_\infty d_{\mathrm{H}}(\mu, \nu)$.

2.2 Bayesian Inverse Problems

This paper is concerned with inverse problems of the following form: given spaces $\mathcal{U}$ and $\mathcal{Y}$, and a known forward operator $G \colon \mathcal{U} \to \mathcal{Y}$, recover $u \in \mathcal{U}$ from a randomly corrupted observation $y \in \mathcal{Y}$ of $G(u)$. A simple example is an inverse problem with additive noise, e.g.

$$y = G(u) + \eta, \tag{2.3}$$

where η is a draw from a $\mathcal{Y}$-valued random variable; crucially, we assume knowledge of the probability distribution of η, but not its exact value.

Inverse problems are typically ill-posed in the sense of having no solution, or multiple solutions, or solutions that depend sensitively upon the observed data y. While there is a long tradition dating back to Tikhonov (1963) and others of addressing such problems using regularisation, the Bayesian approach (Kaipio and Somersalo, 2005; Stuart, 2010) is to interpret both u and y as random variables, and relations such as (2.3) as defining the conditional distribution of y given u. First, one must posit prior beliefs about u independent of y in the form of a prior distribution $\mu_0 \in \mathcal{M}_1(\mathcal{U})$. Then, the Bayesian inverse problem (BIP) is to compute the posterior distribution $\mu^y \in \mathcal{M}_1(\mathcal{U})$, i.e. the conditional distribution of u given y. Naturally, one hopes to do this through an appropriate version of the Bayes formula, e.g. for probability densities with respect to Lebesgue measure on $\mathbb{R}^n$,

$$\rho^y(u) \equiv \rho(u|y) \propto \rho(y|u)\rho_0(u);$$

in the case $\dim \mathcal{U} = \infty$, in which there is no canonical choice of reference measure such as Lebesgue measure, this formula must be treated with some care.

As observed by Stuart (2010, Section 6.6), the correct statement of the Bayes formula when μ_0 is supported on an infinite-dimensional parameter space $\mathcal{U}$ is that the posterior μ^y has a probability density (Radon–Nikodým derivative) with respect to μ_0, and this density is proportional to the conditional probability density of $y|u$. It is both mathematically and computationally convenient to express this relationship in exponential form. That is, $\Phi \colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ will denote the misfit or negative log-likelihood, meaning that, under the hypothesis that $u \in \mathcal{U}$ is 'correct', the probability distribution of $y|u$ is

$$\mathbb{P}[y \in E | u] = \int_E \exp(-\Phi(u; y)) \, \mathrm{d}\varrho(y) \Big/ \int_{\mathcal{Y}} \exp(-\Phi(u; y)) \, \mathrm{d}\varrho(y) \,,$$

where ϱ is some σ-finite reference measure on $\mathcal{Y}$; it is implicitly assumed that $y|u$ is absolutely continuous with respect to ϱ for every $u \in \mathcal{U}$.

In this setting, the generalised Bayes formula is

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u; y))}{Z(y)}, \tag{2.4}$$
$$Z(y) = \mathbb{E}_{u \sim \mu_0}\big[\exp(-\Phi(u; y))\big].$$

However, care must still be taken to check that this formula does define a probability measure μ^y on $\mathcal{U}$; in particular, the normalisation constant $Z(y)$ must

be strictly positive and finite, and verifying this property for the stable priors μ_0 of interest in this paper is the business of Theorem 4.3.

Example 2.1. In the additive case (2.3) with $\eta \sim \mathcal{N}(0, \Sigma)$ independently of u, on $\mathcal{Y} = \mathbb{R}^n$, with $\varrho = n$-dimensional Lebesgue measure,

$$\Phi(u; y) = \tfrac{1}{2}\big\|\Sigma^{-1/2}(y - G(u))\big\|_2^2.$$

If $\dim \mathcal{Y}$ is infinite and $\eta \sim \mathcal{N}(0, \Sigma)$ is a Gaussian random variable on $\mathcal{Y}$ with Cameron–Martin space $\mathrm{ran}(\Sigma^{1/2})$, then this Φ is a.s. infinite since $y \notin \mathrm{ran}(\Sigma^{1/2})$ a.s. It is then necessary to 'subtract off the infinite part of Φ' by using the Cameron–Martin formula for translations of η (Stuart, 2010, Remark 3.8).

2.3 Stable Distributions on the Real Line

The family of stable distributions has been studied extensively in the statistical and probabilistic literature. A random variable u is **stable** if, whenever u_1 and u_2 are independent copies of u and $a, b > 0$, $au_1 + bu_2$ is equal in distribution to $cu + d$ for some $c > 0$ and $d \in \mathbb{R}$. Stability is a particularly appealing property if the aim is to construct prior measures for BIPs that are 'physically consistent' in the sense of remaining in the same model class regardless of discretisation or coordinate choices, at least when the 'physical quantity' obeys an additive law.[1]

Example 2.2. Suppose that the aim is to model (and later infer, in a Bayesian fashion) the distribution of electrical charge in some domain $\Omega \subseteq \mathbb{R}^3$. For computational purposes, Ω is approximated by a triangulation $\mathcal{T}$. Consider two elements $T_1, T_2 \in \mathcal{T}$. If $\mathrm{charge}(T_i)$ is stably distributed, then so too is

$$\mathrm{charge}(T_1 \cup T_2) = \mathrm{charge}(T_1) + \mathrm{charge}(T_2).$$

The charge density $\mathrm{charge}(T_i)/\mathrm{volume}(T_i)$ behaves similarly. Thus, we remain in the same stable model class if we coarsen or refine the mesh $\mathcal{T}$; this would not be true for an unstable random model of the charge, and this would complicate computational modelling in an undesirable fashion.

Real-valued stable random variables are completely classified by four parameters, and this article will follow "Parametrisation 0" of Nolan (2015): a random variable u will be said to be **stably distributed** with index of stability $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale parameter $\gamma \geq 0$, and location parameter $\delta \in \mathbb{R}$, denoted $u \sim \mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$, if the characteristic function (inverse Fourier transform) of u satisfies

$$\mathbb{E}\big[\exp(itu)\big] = \begin{cases} \exp\big(i\delta t - \gamma^\alpha |t|^\alpha [1 + i\beta(\tan\tfrac{\pi\alpha}{2})(\mathrm{sgn}\, t)(|\gamma t|^{1-\alpha} - 1)]\big) & \text{if } \alpha \neq 1, \\ \exp\big(i\delta t - \gamma |t|[1 + i\beta\tfrac{2}{\pi}(\mathrm{sgn}\, t)\log\gamma|t|]\big) & \text{if } \alpha = 1. \end{cases}$$

(The convention here is that $0\log 0 \coloneqq \lim_{s \searrow 0} s\log s = 0$.) If $\gamma = 1$ and $\delta = 0$, then u is said to be **standardised** and we write $u \sim \mathcal{S}(\alpha, \beta; 0)$.

Although a stable random variable u can easily be shown to have a smooth Lebesgue density, exact formulae for this density are not available except in

---

[1]There are situations where such an additive decomposition is not appropriate: e.g., the average homogenised permeability for Darcy flow given the permeabilities of smaller grid cells is obtained as the harmonic rather than arithmetic mean.

special cases. In particular, the normal distribution with mean m and standard deviation σ is $\mathcal{S}(2, 0, \sigma/\sqrt{2}, m; 0)$, and $\mathcal{C}(\delta, \gamma) = \mathcal{S}(1, 0, \gamma, \delta; 0)$. The stability properties of $\mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$ distributions are summarised by the following result:

Proposition 2.3 (Nolan, 2015, Proposition 1.16)**.** *If $u \sim \mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$, then, for $a \neq 0$ and $b \in \mathbb{R}$,*

$$au + b \sim \mathcal{S}(\alpha, (\operatorname{sgn} a)\beta, |a|\gamma, a\delta + b; 0).$$

Also, if $u_i \sim \mathcal{S}(\alpha, \beta_i, \gamma_i, \delta_i; 0)$ and $u_2 \sim \mathcal{S}(\alpha, \beta_2, \gamma_2, \delta_2; 0)$ are independent, then $u_1 + u_2 \sim \mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$ with

$$\beta := \frac{\beta_1 \gamma_1^\alpha + \beta_2 \gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha},$$

$$\gamma^\alpha := \gamma_1^\alpha + \gamma_2^\alpha,$$

$$\delta := \begin{cases} \delta_1 + \delta_2 + (\tan \frac{\pi\alpha}{2})(\beta\gamma - \beta_1\gamma_1 - \beta_2\gamma_2), & \textit{if } \alpha \neq 1, \\ \delta_1 + \delta_2 + \frac{2}{\pi}(\beta\gamma \log \gamma - \beta_1\gamma_1 \log \gamma_1 - \beta_2\gamma_2 \log \gamma_2), & \textit{if } \alpha = 1. \end{cases}$$

The stable distributions with $\alpha = 2$ are exactly the Gaussian measures (the skewness parameter β has no effect, and is conventionally set to 0): by Fernique's theorem, Gaussian measures are exponentially integrable, and in particular have polynomial moments of all orders. Conversely, for $\alpha \in (0, 2)$, the stable distributions are all heavy-tailed: when $u \sim \mathcal{S}(\alpha, \beta, \gamma, 0; 0)$ with $0 < \alpha < 2$,

$$\mathbb{E}\big[|u|^p\big] = \begin{cases} C_{\alpha,\beta} \gamma^\alpha < \infty, & \text{for } 0 < p < \alpha, \\ \infty, & \text{for } p \geq \alpha. \end{cases} \tag{2.5}$$

The asymptotic behaviour of the cumulative distribution and probability density functions of $u \sim \mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$, with $0 < \alpha < 2$ and $-1 < \beta \leq 1$, is that of a power law (Nolan, 2015, Theorem 1.12):

$$\mathbb{P}[u > x] \sim c_\alpha \gamma^\alpha (1 + \beta) x^{-\alpha} \qquad \text{as } x \to \infty, \tag{2.6}$$

$$\rho_u(x) \sim c_\alpha \alpha \gamma^\alpha (1 + \beta) x^{-(\alpha+1)} \qquad \text{as } x \to \infty. \tag{2.7}$$

Similar expressions hold for the behaviour as $x \to -\infty$. Henceforth, to avoid some technical complications, we assume that $-1 < \beta < 1$, so that $u \sim \mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$ is supported on the whole of $\mathbb{R}$.

One further theoretical argument in favour of modelling using stable random variables, particularly from a limiting mesh refinement point of view, is that the stable distributions are precisely the central limits of independent and identically distributed random variables:

Theorem 2.4 (Generalised central limit theorem: Nolan, 2015, Theorem 1.20)**.** *A non-degenerate random variable u is $\mathcal{S}(\alpha, \beta, \gamma, \delta; 0)$ if and only if there is a sequence of i.i.d. random variables $x_1, x_2, \ldots$ and constants $a_n > 0$, $b_n \in \mathbb{R}$ such that $b_n + a_n \sum_{i=1}^n x_i$ converges in distribution to u as $n \to \infty$.*

The literature contains many further application-specific arguments for or against the use of stable distributions in optimisation and inference. O'Hagan (1988) gives a general perspective on modelling with heavy-tailed distributions. Shao and Nikias (1993) treat applications to physics, biology, and electrical engineering, particularly for the modelling of signals and noises with occasional sharp spikes or bursts, as in Figure 1.2. Tsakalides et al. (2000) and

Achim et al. (2003) treat applications to communications and image processing, while (Tsionas, 1999) discusses applications to economics. Hansen et al. (2006) discuss applications to optimisation. Recently, Markkanen et al. (2016) have considered the use of heavy-tailed priors for edge-preserving (i.e. non-smoothing) Bayesian inversion in X-ray tomography; in essence, a Cauchy prior is placed on the gradient of the image to be reconstructed, thereby allowing for jump discontinuities in the image.

3 Karhunen–Loève Expansions for Stable Distributions on Banach Spaces

Now consider the problem of constructing and sampling heavy-tailed stable probability measures on a real Banach space $\mathcal{U}$, for example a vector space of summable sequences or a Sobolev space of fields of specified smoothness. Supposing that one already has access to a generator of real-valued stable random variables (Chambers et al., 1976), it is natural to try to define a $\mathcal{U}$-valued stable random variable via an infinite random series of the form

$$u := \sum_{n \in \mathbb{N}} u_n \psi_n, \tag{3.1}$$

where the ψ_n are a basis for $\mathcal{U}$ and the u_n are $\mathbb{R}$-valued stable random variables; this is the strategy used to generate the examples shown in Figure 1.2. The natural question is, when does (3.1) define a bona fide $\mathcal{U}$-valued random variable?

The Gaussian case is a useful reference point. Suppose that C is a positive-semi-definite and self-adjoint operator on a Hilbert space $\mathcal{U}$ with an eigensystem $(\lambda_n, \psi_n)_{n \in \mathbb{N}}$, and that $(\lambda_n)_{n \in \mathbb{N}} \in \ell^1$, i.e. C is a trace-class operator. Then the series (3.1) with $u_n \sim \mathcal{N}(0, \lambda_n)$ — i.e. with $u_n = \lambda_n^{1/2} \hat{u}_n$ with $\hat{u}_n \sim \mathcal{N}(0,1)$ — converges a.s., and is a draw from the Gaussian measure $\mathcal{N}(0, C)$ on $\mathcal{U}$ with covariance operator C. Similar expansions with different powers of λ_n and $\hat{u}_n$ having density proportional to $\exp(-|\hat{u}_n|^p)$ are used to define draws from Besov measures (Dashti et al., 2012).

However, the focus here is on u_n with heavy tails, so the usual variance-based arguments that are used to prove a.s. convergence of the series (3.1) will not be applicable. However, Theorem 3.3 below shows that the series (3.1) indeed converges almost surely in $\mathcal{U}$ under the assumption that the scale parameters γ_n of the stable random coefficients u_n are α-summable, modulo a logarithmic correction term in the case $\alpha = 1$. The proof of this rests on the following result, which is a synthesis of two classical results from probability theory, and gives a necessary and sufficient condition for the convergence of random series:

Theorem 3.1 (Kolmogorov's zero-one law and three series theorem)**.** *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of independent $\mathbb{R}$-valued random variables. Then the series $\sum_{n \in \mathbb{N}} x_n$ either converges a.s. or diverges a.s, and a.s. convergence holds if and only if, for some $A > 0$, the following series are all finite:*

$$\sum_{n \in \mathbb{N}} \mathbb{P}[|x_n| > A], \quad \sum_{n \in \mathbb{N}} \mathbb{E}[x_n \mathbf{1}[|x_n| \le A]], \quad \text{and} \quad \sum_{n \in \mathbb{N}} \mathbb{E}[x_n^2 \mathbf{1}[|x_n| \le A]].$$

Definition 3.2. Let $\mathcal{U}$ be a real Banach space with countable, unconditional, normalised, Schauder basis $(\psi_n)_{n\in\mathbb{N}}$. Let $\alpha \in (0, 2]$, $\underline{\beta} = (\beta_n)_{n\in\mathbb{N}} \subset (-1, 1)$, $\underline{\gamma} = (\gamma_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$, and $\underline{\delta} = (\delta_n)_{n\in\mathbb{N}} \subset \mathbb{R}$. Let $u_n \sim \mathcal{S}(\alpha, \beta_n, \gamma_n, \delta_n; 0)$ be independent for each $n \in \mathbb{N}$. Then we shall say that $u := \sum_{n\in\mathbb{N}} u_n\psi_n$ is a **stable $\mathcal{U}$-valued random variable** and write $u \sim \mathcal{S}(\alpha, \underline{\beta}, \underline{\gamma}, \underline{\delta}; 0)$.

Theorem 3.3 (Well-definedness of $\mathcal{U}$-valued stable random variables)**.** *Let $u \sim \mathcal{S}(\alpha, \underline{\beta}, \underline{\gamma}, \underline{\delta}; 0)$ with $\alpha \in (0, 2)$, $\underline{\beta} \subset (-1, 1)$, $\underline{\gamma} \in \ell^\alpha$, $\underline{\delta} \in \ell^1$ and, in addition,*

$$\|\underline{\gamma}\|_{\ell\log\ell} := \sum_{n\in\mathbb{N}} |\gamma_n \log \gamma_n| < \infty, \qquad\qquad \text{if } \alpha = 1. \qquad (3.2)$$

Then $u \in \mathcal{U}$ a.s.

Beweis. For each $n \in \mathbb{N}$, let $\hat{u}_n \sim \mathcal{S}(\alpha, \beta_n; 0)$, so that u_n is equal in distribution to $\delta_n + \gamma_n \hat{u}_n$. For $M, N \in \mathbb{N}$ with $N > M$,

$$\left\|\sum_{n=1}^{N} u_n\psi_n - \sum_{n=1}^{M} u_n\psi_n\right\|_{\mathcal{U}} = \left\|\sum_{n=M+1}^{N} (\delta_n + \gamma_n \hat{u}_n)\psi_n\right\|_{\mathcal{U}}$$

$$\leq \sum_{n=M+1}^{N} \|(\delta_n + \gamma_n \hat{u}_n)\psi_n\|_{\mathcal{U}}$$

$$= \sum_{n=M+1}^{N} |\delta_n| + \sum_{n=M+1}^{N} |\gamma_n \hat{u}_n|.$$

If $\underline{\delta} \in \ell^1$, then the (deterministic) first sum on the right-hand side converges to 0 as $N, M \to \infty$. Therefore, it remains only to show that the assumptions on $\underline{\gamma}$ are sufficient to ensure that the (random) second sum on the right-hand side converges a.s. to 0 as $N, M \to \infty$; it will then follow that the partial sums of u are a.s. Cauchy in the Banach space $\mathcal{U}$, and hence a.s. convergent to a well-defined limit in $\mathcal{U}$.

To that end, it will be shown that $\sum_{n\in\mathbb{N}} |\gamma_n \hat{u}_n|$ converges a.s. in $\mathbb{R}$. Let $A > 0$ be large enough that the asymptotic properties (2.6) and (2.7) hold true for $|x| > A$. Then, by (2.6),

$$\mathbb{P}\big[|\gamma_n \hat{u}_n| > A\big] \sim C \left|\frac{\gamma_n}{A}\right|^\alpha,$$

where C depends only on α and β_n. Since $\underline{\gamma} \in \ell^\alpha$, the series $\sum_{n\in\mathbb{N}} \mathbb{P}\big[|\gamma_n \hat{u}_n| > A\big]$ is convergent. By (2.7), for $p = 1, 2$, the truncated p^{th} moments of $|\gamma_n \hat{u}_n|$ satisfy

$$\mathbb{E}\big[|\gamma_n \hat{u}_n|^p \mathbf{1}[|\gamma_n \hat{u}_n| < A]\big] = |\gamma_n|^p \int_{-A/\gamma_n}^{A/\gamma_n} |s|^p \rho_{\hat{u}_n}(s)\,\mathrm{d}s$$

$$\leq \begin{cases} C|\gamma_n \log \gamma_n| & \text{if } \alpha = p = 1, \\ C|\gamma_n|^\alpha, & \text{otherwise,} \end{cases}$$

where C depends on α, β_n, p, and A but is independent of γ_n. Thus, since $\underline{\gamma} \in \ell^\alpha$, the sequence of truncated second moments ($p = 2$) is summable. For $\alpha \neq 1$, the assumption that $\underline{\gamma} \in \ell^\alpha$ also ensures that the sequence of truncated first absolute moments ($p = 1$) is summable; in the case $\alpha = 1$, summability follows from the additional assumption (3.2). Therefore, Theorem 3.1 implies that $\sum_{n\in\mathbb{N}} |\gamma_n \hat{u}_n|$ converges a.s. in $\mathbb{R}$, and so (3.1) converges a.s. in $\mathcal{U}$. $\qquad \square$

Remark 3.4. Note that condition (3.2) for the case $\alpha = 1$, requiring that the Orlicz norm $\|\underline{\gamma}\|_{\ell \log \ell}$ be finite, cannot be weakened to just requiring that $\underline{\gamma} \in \ell^1$. For example, for $\gamma_n := n^{-1}(\log n)^{-2}$, the integral test reveals that $\sum_{n \geq 2} |\gamma_n| < \infty$ but $\sum_{n \geq 2} |\gamma_n \log \gamma_n| = \infty$; in this situation, summability of the truncated first absolute moments of the coefficients $\gamma_n \hat{u}_n$ is no longer assured. However, for polynomial $\underline{\gamma}$, the ℓ^1 and Orlicz criteria *do* coincide: for $\gamma_n = Cn^{-p}$, $\|\underline{\gamma}\|_{\ell^1}$ is finite once $p > 1$, and then $\|\underline{\gamma}\|_{\ell \log \ell}$ is also finite.

Example 3.5. If the coefficients u_n in (3.1) are independent Cauchy random variables, $u_n \sim \mathcal{C}(0, \gamma_n) = \mathcal{S}(1, 0, \gamma_n, 0; 0)$, then the truncated moments are

$$\mathbb{P}\big[|\gamma_n u_n| \geq A\big] = 1 - \frac{2}{\pi} \arctan \frac{A}{\gamma_n},$$

$$\mathbb{E}\big[|\gamma_n u_n| \mathbf{1}[|\gamma_n u_n| < A]\big] = \frac{\gamma_n}{\pi} \log \left(1 + \frac{A^2}{\gamma_n^2}\right),$$

$$\mathbb{E}\big[|\gamma_n u_n|^2 \mathbf{1}[|\gamma_n u_n| < A]\big] = \frac{2A\gamma_n}{\pi} + \frac{2\gamma_n^2}{\pi} \arctan \frac{A}{\gamma_n}.$$

Consistent with Theorem 3.3, these three series all converge if $\|\underline{\gamma}\|_{\ell \log \ell}$ is finite, and in particular if $\gamma_n = \mathrm{O}(n^{-p})$ for some $p > 1$. When this convergence holds, the random series (3.1) converges a.s. in $\mathcal{U}$, and thereby defines a $\mathcal{U}$-valued Cauchy random variable.

Remark 3.6. An immediate consequence of the stability of each of the coefficients u_n in the basis $\{\psi_n\}_{n \in \mathbb{N}}$ is that $\mathcal{U}$-valued random variables in the sense of Definition 3.2 are stable under addition.

Remark 3.7. Let Γ be a self-adjoint positive definite operator on a separable Hilbert space $\mathcal{H}$, with eigensystem $(\gamma_n, \psi_n)_{n \in \mathbb{N}}$. If $(\gamma_n)_{n \in \mathbb{N}} \in \ell^\alpha$ for some $\alpha \in (0, 2)$, and satisfies (3.2) if $\alpha = 1$, and $\underline{\beta} \subset (-1, 1)$, then Theorem 3.3 shows that $u := \sum_{n \in \mathbb{N}} \gamma_n \hat{u}_n \psi_n$ with independent $\hat{u}_n \sim \mathcal{S}(\alpha, \beta_n; 0)$ is a well-defined random variable a.s. taking values in

$$\mathcal{U} := \left\{ \sum_{n \in \mathbb{N}} v_n \psi_n \,\middle|\, (v_n)_{n \in \mathbb{N}} \in \ell^1 \right\} \subseteq \mathcal{H}.$$

The final objective of this section is to show that $u \sim \mathcal{S}(\alpha, \underline{\beta}, \underline{\gamma}, \underline{\delta}; 0)$ has finite fractional lower-order moments $\mathbb{E}\big[\|u\|_{\mathcal{U}}^p\big]$ for $0 < p < \alpha$, as in the real-valued case. This requires an assumption that the synthesis operator

$$S_\psi : \underline{v} := (v_n)_{n \in \mathbb{N}} \mapsto \sum_{n \in \mathbb{N}} v_n \psi_n$$

is a continuous embedding of the sequence space ℓ^p of coefficients into $\mathcal{U}$, i.e.

$$\left\| \sum_{n \in \mathbb{N}} v_n \psi_n \right\|_{\mathcal{U}} \leq C \|\underline{v}\|_{\ell^p}. \tag{3.3}$$

If the reverse inequality also holds, possibly with a different constant, then the basis $(\psi_n)_{n \in \mathbb{N}}$ is called a p-**frame** for $\mathcal{U}$ (Christensen and Stoeva, 2003). In the case $p = 2$, $(\psi_n)_{n \in \mathbb{N}}$ is called a **Riesz basis** for $\mathcal{U}$, and this case arises often in applications e.g. when $(\psi_n)_{n \in \mathbb{N}}$ is a wavelet basis. When (3.3) holds for some $p = p_* > 0$, it also holds for all $p \in (0, p_*)$, since $\| \cdot \|_{\ell^{p_*}} \leq \| \cdot \|_{\ell^p}$.

Theorem 3.8 (p^{th}-mean convergence and fractional lower-order moments)**.** *Let* $u \sim \mathcal{S}(\alpha, \underline{\beta}, \underline{\gamma}, \underline{\delta}; 0)$ *satisfy the assumptions of Theorem 3.3, and suppose that* $(\psi_n)_{n \in \mathbb{N}}$ *satisfies (3.3) for some* $p = p_* > 0$. *Let* $0 < p \le p_*$ *and* $p < \alpha$. *Then* $\sum_{n=1}^{N} u_n \psi_n \to u$ *in* $L^p(\Omega, \mathbb{P}; \mathcal{U})$ *as* $N \to \infty$ *and, in particular,* $\mathbb{E}\big[\|u\|_{\mathcal{U}}^p\big] \le \|\underline{\gamma}\|_{\ell^\alpha} + \|\underline{\delta}\|_{\ell^1} < \infty$.

Beweis. For $M, N \in \mathbb{N}$ with $N > M$,

$$
\left\| \sum_{n=1}^{N} u_n \psi_n - \sum_{n=1}^{M} u_n \psi_n \right\|_{L^p(\Omega, \mathbb{P}; \mathcal{U})}
$$

$$
\le \left\| \sum_{n=N+1}^{M} \delta_n \psi_n \right\|_{\mathcal{U}} + \left(\mathbb{E}\left[\left\| \sum_{n=N+1}^{M} \gamma_n \hat{u}_n \psi_n \right\|_{\mathcal{U}}^p \right] \right)^{1/p} \quad \text{by the triangle inequality}
$$

$$
\le \sum_{n=N+1}^{M} |\delta_n| + C\left(\mathbb{E}\left[\left(\sum_{n=M+1}^{N} |\gamma_n \hat{u}_n|^{p_*} \right)^{p/p_*} \right] \right)^{1/p} \quad \text{by (3.3)}
$$

$$
\le \sum_{n=N+1}^{M} |\delta_n| + C\left(\mathbb{E}\left[\left(\sum_{n=M+1}^{N} |\gamma_n \hat{u}_n|^{p} \right)^{p/p} \right] \right)^{1/p} \quad \text{since } \|\cdot\|_{\ell^{p_*}} \le \|\cdot\|_{\ell^p}
$$

$$
\le \sum_{n=N+1}^{M} |\delta_n| + C\left(\sum_{n=M+1}^{N} \gamma_n^\alpha \right)^{1/p} \quad \text{by (2.5)}
$$

$$
\to 0 \text{ as } M, N \to \infty \quad \text{since } \underline{\gamma} \in \ell^\alpha, \, \underline{\delta} \in \ell^1.
$$

Thus, the partial sums of the series $\sum_{n \in \mathbb{N}} u_n \psi_n$ are Cauchy in $L^p(\Omega, \mathbb{P}; \mathcal{U})$, which by completeness[2] implies that they converge to $u \in L^p(\Omega, \mathbb{P}; \mathcal{U})$. In particular, $\|u\|_{L^p(\Omega, \mathbb{P}; \mathcal{U})} = \mathbb{E}\big[\|u\|_{\mathcal{U}}^p\big]$ is finite, and the estimate $\mathbb{E}\big[\|u\|_{\mathcal{U}}^p\big] \le \|\underline{\gamma}\|_{\ell^\alpha} + \|\underline{\delta}\|_{\ell^1}$ follows from the above and e.g. Fatou's lemma. (Note that we do not apply (3.3) to the δ_n terms because we do not know that $\underline{\delta} \in \ell^p$ for $p < 1$.) $\qquad \square$

The next section uses Theorem 3.8 in the form that, when μ_0 is the law of an α-stable u, $\exp(p \log \| \cdot \|_{\mathcal{U}})$ is μ_0-integrable for $0 < p < \alpha$.

4 Well-Posedness of Bayesian Inverse Problems with Heavy-Tailed Priors

This section establishes conditions for the BIP with an arbitrary prior μ_0 to be well-posed in sense that, for each $y \in \mathcal{Y}$, the posterior distribution μ^y of $u|y$ is a well-defined probability measure on $\mathcal{U}$ (Theorem 4.3), which changes continuously when either the observed data is perturbed to $y' \approx y$ (Theorem 4.4) or the misfit function is perturbed to $\Phi_N \approx \Phi$ (Theorem 4.5). It is natural to seek robustness of the BIP to such perturbations: a perturbation of y to y' may arise through observational error, whereas a perturbation of Φ to Φ_N may arise through a numerical approximation of the forward model (e.g. a PDE solution operator) G by a numerical solution operator G_N. As in the earlier

---

[2]Recall that though the L^p spaces are not normed for $0 < p < 1$, since the triangle inequality fails, they are nonetheless complete *quasinormed* spaces.

works following Stuart (2010), the mapping $y \mapsto \mu^y$ is shown to be $(\|\cdot\|_{\mathcal{Y}}, d_{\mathrm{H}})$-Lipschitz, and the convergence $\mu_N^y \to \mu^y$ in d_{H} inherits the same convergence rate as the convergence $\Phi_N \to \Phi$, so that the numerical analysis of the forward problem transfers to the BIP.

A notable feature of the results presented in this section is that, lacking any exponential integrability of μ_0, a careful tradeoff of growth rates of Φ is necessary in order to ensure well-definedness and well-posedness of the Bayesian posterior measure μ^y. Indeed, this tradeoff is a desirable feature, since 'good' behaviour of one growth rate can be used to compensate for 'bad' behaviour of another.

It bears emphasis that the results of this section are not particular to *stable* heavy-tailed priors, and the relaxed regularity assumptions used here provide additional understanding of the previously-studied Gaussian and Besov cases, in which well-posedness holds even when $M_{1,r}(t) \to -\infty$ at a polynomial rate as $t \to \infty$. The proof strategies used here are very similar to those of Stuart (2010, Section 4) and Dashti and Stuart (2015, Section 4).

Assumption 4.1. $\mathcal{U}$ and $\mathcal{Y}$ are separable Banach spaces over $\mathbb{R}$ and the misfit function $\Phi \colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ satisfies the following:

(A0) Φ is a locally bounded Carathéodory function, i.e. $\Phi(u; \cdot)$ is continuous for each $u \in \mathcal{U}$, $\Phi(\cdot; y)$ is measurable for each $y \in \mathcal{Y}$, and for every $r > 0$, there exists $M_{0,r} \in \mathbb{R}$ such that, for all $(u, y) \in \mathcal{U} \times \mathcal{Y}$ with $\|u\|_{\mathcal{U}} < r$ and $\|y\|_{\mathcal{Y}} < r$,
$$|\Phi(u; y)| \leq M_{0,r}.$$

(A1) For every $r > 0$, there exists a measurable $M_{1,r} \colon \mathbb{R}_+ \to \mathbb{R}$ such that, for all $(u, y) \in \mathcal{U} \times \mathcal{Y}$ with $\|y\|_{\mathcal{Y}} < r$ and $\|u\|_{\mathcal{U}}$ large enough,
$$\Phi(u; y) \geq M_{1,r}\big(\|u\|_{\mathcal{U}}\big).$$

(A2) For every $r > 0$, there exists a measurable $M_{2,r} \colon \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all $(u, y_1, y_2) \in \mathcal{U} \times \mathcal{Y} \times \mathcal{Y}$ with $\|y_1\|_{\mathcal{Y}} < r$, $\|y_2\|_{\mathcal{Y}} < r$,
$$|\Phi(u; y_1) - \Phi(u; y_2)| \leq \exp\big(M_{2,r}\big(\|u\|_{\mathcal{U}}\big)\big)\|y_1 - y_2\|_{\mathcal{Y}}.$$

Furthermore, for each $N \in \mathbb{N}$, $\Phi_N \colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ is an approximation to Φ that satisfies (A0)–(A2) with $M_{i,r}$ and $\rho_{i,r}$ independent of N, and such that

(A3) $\psi \colon \mathbb{N} \to \mathbb{R}_+$ is such that, for every $r > 0$, there exists a measurable $M_{3,r} \colon \mathbb{R}_+ \to \mathbb{R}_+$, such that, for all $(u, y) \in \mathcal{U} \times \mathcal{Y}$ with $\|y\|_{\mathcal{Y}} < r$,
$$|\Phi_N(u; y) - \Phi(u; y)| \leq \exp\big(M_{3,r}\big(\|u\|_{\mathcal{U}}\big)\big)\psi(N).$$

Remark 4.2. Assumptions (A0)–(A3) have been re-ordered relative to their counterparts in earlier works, such as those of Stuart (2010) and Dashti et al. (2012). The numbering and placement of (A0) (usually assumptions 2 and 3 in the previous works) highlights its role as a mild measurability assumption, so that (A1)–(A3) (usually assumptions 1, 4, and 5) form a natural sequence of statements about the growth rates $\rho_{i,r}$.

(A0) is significantly weaker than the corresponding assumptions in previous works, in which it is assumed that $\Phi(\cdot; y)$ is locally Lipschitz continuous. However, close inspection of the proofs in those works reveals that local Lipschitz continuity is used only in order to ensure that $e^{-\Phi(\cdot; y)}$ is locally μ_0-integrable, so that it can serve as a density of the non-normalised posterior with respect to

the prior. The above assumptions imply that, $\Phi(\,\cdot\,;y)$ and $e^{-\Phi(\,\cdot\,;y)}$ are locally bounded measurable functions; since μ_0 is a probability measure, this yields the desired local integrability. Furthermore, the separability assumptions on $\mathcal{U}$ and $\mathcal{Y}$ and (A0) imply, by Aliprantis and Border (2006, Lemma 4.51), that $\Phi(u;y)$ and $e^{-\Phi(u;y)}$ are jointly measurable in (u,y).

However, (A2) remains as a continuity assumption, since this is necessary in order to establish Hellinger continuity of the posterior with respect to y.

Theorem 4.3 (Well-definedness of the Bayesian posterior). *Let μ_0 be a Borel probability measure on $\mathcal{U}$, and let $y \in \mathcal{Y}$. If (A0) and (A1) hold with*

$$S_{1,r} := \mathbb{E}_{u\sim\mu_0}\big[\exp(-M_{1,r}(\|u\|_{\mathcal{U}}))\big] < \infty, \tag{4.1}$$

then $Z(y) := \mathbb{E}_{u\sim\mu_0}\big[\exp(-\Phi(u;y))\big]$ is strictly positive and finite, and setting

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u;y))}{Z(y)} \tag{4.2}$$

defines a Borel probability measure μ^y on $\mathcal{U}$, which is tight in the sense that

$$\mu^y(E) = \sup\{\mu^y(K) \mid K \subseteq E \text{ and } K \text{ is compact}\} \quad \text{for all measurable } E \subseteq \mathcal{U}.$$

Beweis. As discussed in Remark 4.2, $\exp(-\Phi(\,\cdot\,;y))$ is locally integrable with respect to μ_0. Therefore, by the Radon–Nikodým theorem, setting

$$\nu(E) := \mathbb{E}_{u\sim\mu_0}\big[\exp(-\Phi(u;y))\mathbf{1}[u \in E]\big]$$

for each measurable set $E \subseteq \mathcal{U}$ defines a countably additive measure ν on $\mathcal{U}$; what remains is to check that ν can be normalised to yield the probability measure μ^y, i.e. it is necessary to show that $0 < Z(y) \equiv \nu(\mathcal{U}) < \infty$. Let $r > \|y\|_{\mathcal{Y}}$. Then

$$Z(y) \leq \mathbb{E}_{u\sim\mu_0}\big[\exp\big(-M_{1,r}\big(\|u\|_{\mathcal{U}}\big)\big)\big] \leq S_{1,r} < \infty$$

by (A1) and (4.1); and

$$Z(y) \geq \mathbb{E}_{u\sim\mu_0}\big[\exp(-\Phi(u;y))\mathbf{1}\big[\|u\|_{\mathcal{U}} < r\big]\big] \geq \exp(-M_{0,r})\mu_0\big(\mathbb{B}_r(0;\|\cdot\|_{\mathcal{U}})\big)$$

by (A0). Since μ_0 is a countably additive Borel probability measure,

$$1 = \mu_0(\mathcal{U}) = \sum_{n\in\mathbb{N}} \mu_0\big(\mathbb{B}_{n+1}(0;\|\cdot\|_{\mathcal{U}}) \setminus \mathbb{B}_n(0;\|\cdot\|_{\mathcal{U}})\big),$$

and so it is impossible for all the summands on the right-hand side to vanish. Since at least one of the annuli $\mathbb{B}_{n+1}(0;\|\cdot\|_{\mathcal{U}}) \setminus \mathbb{B}_n(0;\|\cdot\|_{\mathcal{U}})$ has strictly positive measure, it follows that $\mu_0\big(\mathbb{B}_r(0;\|\cdot\|_{\mathcal{U}})\big) > 0$ once $r > 0$ is large enough. Hence, $Z(y) > 0$, and so μ^y is a well-defined Borel probability measure on $\mathcal{U}$, with Radon–Nikodým derivative with respect to μ_0 given by (4.2).

In any Polish space, and hence in the separable Banach space $\mathcal{U}$, every finite-mass measure is tight (Aliprantis and Border, 2006, Theorem 12.7). Hence, μ_0 and μ^y are both tight. $\qquad\square$

Theorem 4.4 (Perturbation of observed data). *Suppose that $r > 0$ is such that* (A0)–(A2) *hold with*

$$S_{1,2,r} := \mathbb{E}_{u \sim \mu_0}\big[\exp(2M_{2,r}(\|u\|_{\mathcal{U}}) - M_{1,r}(\|u\|_{\mathcal{U}}))\big] < \infty. \qquad (4.3)$$

Then there exists a constant C, which may depend on r, $S_{1,2,r}$, and the constants and functions in (A0)–(A2), *such that, whenever $\|y\|_{\mathcal{Y}}, \|y'\|_{\mathcal{Y}} < r$,*

$$|Z(y) - Z(y')| \leq C\|y - y'\|_{\mathcal{Y}} \qquad (4.4)$$

$$and \qquad d_{\mathrm{H}}\big(\mu^y, \mu^{y'}\big) \leq C\|y - y'\|_{\mathcal{Y}}. \qquad (4.5)$$

Beweis. First, consider the normalising constant $Z(y)$ as a function of y. Note that (4.3) implies (4.1), so $0 < Z(y) < \infty$. Furthermore, whenever $\|y\|_{\mathcal{Y}}, \|y'\|_{\mathcal{Y}} < r$,

$$|Z(y) - Z(y')| \leq \mathbb{E}_{u \sim \mu_0}\Big[\big|\exp(-\Phi(u; y)) - \exp(-\Phi(u; y'))\big|\Big]$$

$$\leq \mathbb{E}\Big[e^{-M_{1,r}(\|u\|_{\mathcal{U}})}\big|\Phi(u; y) - \Phi(u; y')\big|\Big] \qquad \text{by (A1)}$$

$$\leq \mathbb{E}\Big[e^{-M_{1,r}(\|u\|_{\mathcal{U}})}e^{M_{2,r}(\|u\|_{\mathcal{U}})}\|y - y'\|_{\mathcal{Y}}\Big] \qquad \text{by (A2)}$$

$$\leq C\|y - y'\|_{\mathcal{Y}} \qquad \text{by (4.3)},$$

which establishes (4.4). Now, from the definition (2.1) of d_{H},

$$d_{\mathrm{H}}\big(\mu^y, \mu^{y'}\big)^2$$

$$= \mathbb{E}_{u \sim \mu_0}\left[\left|\frac{e^{-\Phi(u;y)/2}}{\sqrt{Z(y)}} - \frac{e^{-\Phi(u;y')/2}}{\sqrt{Z(y')}}\right|^2\right]$$

$$= \mathbb{E}\left[\left|\frac{e^{-\Phi(u;y)/2}}{\sqrt{Z(y)}} - \frac{e^{-\Phi(u;y')/2}}{\sqrt{Z(y)}} + \frac{e^{-\Phi(u;y')/2}}{\sqrt{Z(y)}} - \frac{e^{-\Phi(u;y')/2}}{\sqrt{Z(y')}}\right|^2\right]$$

$$\leq \frac{2}{Z(y)}\mathbb{E}\Big[\big|e^{-\Phi(u;y)/2} - e^{-\Phi(u;y')/2}\big|^2\Big] + 2\mathbb{E}\left[e^{-\Phi(u;y')}\left|\frac{1}{\sqrt{Z(y)}} - \frac{1}{\sqrt{Z(y')}}\right|^2\right]$$

$$= \frac{2}{Z(y)}\underbrace{\mathbb{E}\Big[\big|e^{-\Phi(u;y)/2} - e^{-\Phi(u;y')/2}\big|^2\Big]}_{I_1} + 2Z(y')\underbrace{\left|\frac{1}{\sqrt{Z(y)}} - \frac{1}{\sqrt{Z(y')}}\right|^2}_{I_2},$$

where the inequality follows from the algebraic inequality $(a + b)^2 \leq 2a^2 + 2b^2$. For the first term,

$$I_1 \leq \mathbb{E}\big[e^{-M_{1,r}(\|u\|_{\mathcal{U}})}e^{2M_{2,r}(\|u\|_{\mathcal{U}})}\|y - y'\|_{\mathcal{Y}}^2\big] \qquad \text{by (A1)–(A2)}$$

$$\leq C\|y - y'\|_{\mathcal{Y}}^2 \qquad \text{by (4.3)}.$$

For the second term, (4.4) implies that

$$I_2 \leq \max\{Z(y)^{-3}, Z(y')^{-3}\}\big|Z(y) - Z(y')\big|^2 \leq C\|y - y'\|_{\mathcal{Y}}^2$$

Thus, $d_{\mathrm{H}}\big(\mu^y, \mu^{y'}\big)^2 \leq C\|y - y'\|_{\mathcal{Y}}^2$, and taking square roots completes the proof. $\square$

Theorem 4.5 (Perturbation of likelihood). *Let Φ and Φ_N satisfy* (A0)–(A3), *and suppose that, for some $r > 0$,*

$$S_{1,3,r} := \mathbb{E}_{u \sim \mu_0}\big[\exp(2M_{3,r}(\|u\|_{\mathcal{U}}) - M_{1,r}(\|u\|_{\mathcal{U}}))\big] < \infty. \qquad (4.6)$$

Then there exists a constant C, which may depend on r, $S_{1,3,r}$, and the constants and functions in (A0)–(A3) *but is independent of N, such that the posteriors μ^y and μ_N^y, arrived at using the same data y with $\|y\|_{\mathcal{Y}} < r$ but the misfit functions Φ and Φ_N respectively, satisfy*

$$d_{\mathrm{H}}\big(\mu^y, \mu_N^y\big) \leq C\psi(N).$$

Beweis. The proof is very similar to that of Theorem 4.4, and is omitted. $\qquad\square$

Remark 4.6. It is interesting to note the range of applicability of Theorems 4.3, 4.4, and 4.5 when the prior μ_0 is the probability law of a $\mathcal{U}$-valued α-stable random variable. Under the assumption (3.3), Theorem 3.8 implies that (4.1) is satisfied if $M_{1,r}(t) \geq C - p \log t$ for some constant C and some $0 < p < \alpha$, i.e. $\Phi(\,\cdot\,; y)$ is permitted to diverge to $-\infty$ at a logarithmic rate controlled by the index of stability of μ_0. Similarly, (4.3) is satisfied if $2M_{2,r}(t) - M_{1,r}(t) \leq C + p \log t$, and (4.6) is satisfied if $2M_{3,r}(t) - M_{1,r}(t) \leq C + p \log t$.

Example 4.7. Consider the additive noise model (2.3) with Gaussian noise $\eta \sim \mathcal{N}(0, \Sigma)$. Suppose that the following growth conditions hold for some strictly positive constants $\sigma^{\pm} > 0$ and some continuous $g^{\pm} \colon \mathbb{R}_+ \to \mathbb{R}_+$:

$$g^-(\|u\|_{\mathcal{U}}) \leq \|G(u)\|_{\mathcal{Y}} \leq g^+(\|u\|_{\mathcal{U}}) \qquad \text{for all } u \in \mathcal{U},$$
$$\sigma^-\|y\|_{\mathcal{Y}} \leq \|\Sigma^{-1}y\|_{\mathcal{Y}} \leq \sigma^+\|y\|_{\mathcal{Y}} \qquad \text{for all } y \in \mathcal{Y}.$$

Note that $\Phi(u; y)$ is smooth in (u, y), so (A0) holds. In fact, for a better lower bound on $\Phi(u; y)$ when $\|y\|_{\mathcal{Y}} < r$, observe that, for a suitable choice of additive constant $M_{1,r} \in \mathbb{R}$,

$$\Phi(u; y) \geq M_{1,r} + \sigma^- g^-(\|u\|_{\mathcal{U}})^2,$$

so (A1) holds with $M_{1,r}(t) = \sigma^- g^-(t)^2$. Thus, Theorem 4.3 gives well-definedness of μ^y even in the presence of the trivial lower bound $g^-(t) \equiv 0$, for any prior μ_0.

However, the analysis of well-posedness is more involved:

$$\frac{\partial \Phi}{\partial y}(u; y) = \Sigma^{-1}(y - G(u))$$

and so $\sigma^+(r + g^+(\|u\|_{\mathcal{U}}))$ is a Lipschitz constant for $\Phi(u; \,\cdot\,)$ on $\mathbb{B}_r(0; \|\cdot\|_{\mathcal{Y}}) \subseteq \mathcal{Y}$. Thus, (A2) holds with $M_{2,r}(t) = \log(r + g^+(t))$. By Theorem 4.4, the posterior μ^y depends in a Lipschitz fashion on $y \in \mathbb{B}_r(0; \|\cdot\|_{\mathcal{Y}}) \subseteq \mathcal{Y}$ if

$$u \mapsto 2\log(r + g^+(\|u\|_{\mathcal{U}})) - \sigma^- g^-(\|u\|_{\mathcal{U}})^2$$

is exponentially integrable with respect to μ_0. Suppose that the prior μ_0 is the law of an $\mathcal{S}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \underline{\delta}; 0)$ random vector satisfying the assumptions of Theorem 3.8. Then the hypotheses of Theorem 4.4 are satisfied if, for some $p < \alpha$,

$$2\log(r + g^+(t)) - \sigma^- g^-(t)^2 \leq C + p \log t. \qquad (4.7)$$

15

Since $p < 2$, the satisfaction of condition (4.7) depends crucially on the behaviour of $g^{\pm}(t)$ as $t \to \infty$. For example, suppose that the following slowly-growing lower bound and power-law upper bound on $\|G(u)\|_{\mathcal{Y}}$ hold:

$$\sqrt{c^- \log \|u\|_{\mathcal{U}}} \leq \|G(u)\|_{\mathcal{Y}} \leq c^+ \|u\|_{\mathcal{U}}^{\kappa}.$$

Then the BIP is well-posed with respect to y if $2\kappa - \sigma^- c^- \leq p$. Informally, this holds if the lower bounds on Σ^{-1} and G are far enough from zero compared to the upper bound growth rate κ.

As usual, similar arguments apply to approximation of Φ by Φ_N.

Remark 4.8. The question of whether or not μ^y depends continuously upon the prior measure μ_0 is a delicate one. First, probability measures on infinite-dimensional spaces are highly prone to mutual singularity even when they are related by surprisingly simple operations such as translation or dilation, cf. the Cameron–Martin and Feldman–Hájek theorems. Secondly, it is known that small perturbations of μ_0 in the weak, total variation, or Hellinger topologies can lead to discontinuous changes in posterior expected values of pre-chosen integrands. On the other hand, at least for finite-dimensional $\mathcal{U}$, with respect to the Kullback–Leibler topology, small perturbations in μ_0 lead to small perturbations in μ^y. For a more thorough treatment of this highly involved topic, see e.g. Owhadi and Scovel (2016, Section 1) and the references cited therein.

Acknowledgements

Literatur

A. Achim, P. Tsakalides, and A. Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote*, 41(8):1773–1784, 2003. doi: 10.1109/TGRS.2003.813488.

C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin, third edition, 2006. ISBN 978-3-540-32696-0; 3-540-32696-0.

J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *J. Amer. Statist. Assoc.*, 71(354):340–344, 1976.

O. Christensen and D. T. Stoeva. p-frames in separable Banach spaces. *Adv. Comput. Math.*, 18(2-4):117–126, 2003. doi: 10.1023/A:1021364413257.

M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems, 2015. arXiv:1302.6989v4.

M. Dashti, S. Harris, and A. Stuart. Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging*, 6(2):183–200, 2012. doi: 10.3934/ipi.2012.6.183.

N. Hansen, F. Gemperle, A. Auger, and P. Koumoutsakos. When do heavy-tail distributions help? In T. P. Runarsson, H.-G. Beyer, E. Burke, J. J. Merelo-Guervós, L. D. Whitley, and X. Yao, editors, *Parallel Problem Solving from Nature — PPSN IX: 9th International Conference, Reykjavik, Iceland, September 9–13, 2006, Proceedings*, pages 62–71. Springer, Berlin, Heidelberg, 2006. doi: 10.1007/11844297_7.

B. Hosseini and N. Nigam. Well-posed Bayesian inverse problems: Priors with exponential tails, 2016. arXiv:1604.02575v2.

J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005. doi: 10.1007/b138659.

M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems*, 20(5):1537–1563, 2004. doi: 10.1088/0266-5611/20/5/013.

M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging*, 3(1):87–122, 2009. doi: 10.3934/ipi.2009.3.87.

M. Markkanen, L. Roininen, J. M. J. Huttunen, and S. Lasanen. Cauchy difference priors for edge-preserving Bayesian inversion with an application to X-ray tomography, 2016. arXiv:1603.06135v1.

J. P. Nolan. *Stable Distributions — Models for Heavy Tailed Data*. Birkhauser, Boston, 2015. In progress, Chapter 1 online at `http://academic2.american.edu/~jpnolan`.

A. O'Hagan. Modelling with heavy tails. In *Bayesian Statistics, 3 (Valencia, 1987)*, Oxford Sci. Publ., pages 345–359. Oxford Univ. Press, New York, 1988.

H. Owhadi and C. Scovel. Qualitative robustness in Bayesian inference, 2016. arXiv:1411.3984v3.

M. Shao and C. Nikias. Signal processing with fractional lower order moments: Stable processes and their application. *Proc. IEEE*, 81(7):986–1010, August 1993. doi: 10.1109/5.231338.

A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19: 451–559, 2010. doi: 10.1017/S0962492910000061.

A. N. Tikhonov. On the solution of incorrectly put problems and the regularisation method. In *Outlines Joint Sympos. Partial Differential Equations (Novosibirsk, 1963)*, pages 261–265. Acad. Sci. USSR Siberian Branch, Moscow, 1963.

P. Tsakalides, P. Reveliotis, and C. L. Nikias. Scalar quantisation of heavy-tailed signals. *IEE Proc. - Vis. Image Sign.*, 147(5):475–484, October 2000. doi: 10.1049/ip-vis:20000470.

E. Tsionas. Monte Carlo inference in econometric models with symmetric stable distributions. *J. Economet.*, 88:365–401, 1999.