Well-posed Bayesian Inverse Problems: Beyond Gaussian Priors

Bamdad Hosseini
Department of Mathematics
Simon Fraser University, Canada

The Institute for Computational Engineering and Sciences
University of Texas at Austin
September 28, 2016
The Bayesian approach

- A model for indirect measurements $y \in Y$ of a parameter $u \in X$
  \[ y = \tilde{G}(u). \]

- $X, Y$ are Banach spaces.
- $\tilde{G}$ encompasses measurement noise.
- Simple example, additive noise model
  \[ y = G(u) + \eta. \]

- $G$–deterministic forward map
- $\eta$ – independent random variable.
- **Find $u$ given a realization of $y$.**
Application 1: atmospheric source inversion

\[
\begin{aligned}
(\partial_t - \mathcal{L})c &= u & \text{in } & D \times (0, T], \\
\quad c(x, t) &= 0 & \text{on } & \partial D \times (0, T), \\
\quad c(x, 0) &= 0.
\end{aligned}
\]

Advection-diffusion PDE.

Estimate \( u \) from accumulated deposition measurements\(^1\).

Application 2: high intensity focused ultrasound treatment

- Acoustic waves converge.
- Ablate diseased tissue.
- Phase shift due to skull bone.
- Defocused beam.

- Compensate for phase shift to focus the beam.

Estimate phase shift from MR-ARFI data\(^2\).

Example: Deconvolution

Let $X = L^2(\mathbb{T})$ and assume $G(u) = S(\varphi \ast u)$. Here $\varphi \in C^\infty(\mathbb{T})$ and $S : C(\mathbb{T}) \to \mathbb{R}^m$ collects point values of a function at $m$ distinct points $\{t_k\}_{k=1}^m$. Noise $\eta$ is additive and Gaussian.

We want to find $u$ given noisy pointwise observations of the blurred image.
The Bayesian approach

- Bayes’ rule\(^3\) in the sense of Radon-Nikodym theorem,
  \[
  \frac{d\mu_y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)).
  \]
  \(\mu_0\) – prior measure.
  \(\Phi\) – likelihood potential \(\leftarrow y = \tilde{G}(u)\).
  \(Z(y) = \int_X \exp(-\Phi(u; y))d\mu_0(u)\) – normalizing constant.
  \(\mu_y\) – posterior measure.

---

Why non-Gaussian priors?

\[
\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)).
\]

- \( \text{supp}\mu^y \subseteq \text{supp}\mu_0 \) since \( \mu^y \ll \mu_0 \).
- The prior has a major influence on the posterior.
Application 1: atmospheric source inversion

- \( \Omega := D \times (0, T] \)
- Measurement operators
  \[
  M_i : L^2(\Omega) \to \mathbb{R}, \quad M_i(c) = \int_{J_i \times (0, T]} c \, dx \, dt, \quad i = 1, \ldots, m.
  \]
- Forward map
  \[
  \mathcal{G} : L^2(\Omega) \to \mathbb{R}^m, \quad \mathcal{G}(u) = (M_1(c(u)), \ldots, M_m(c(u))^T, \quad c = (\partial_t - \mathcal{L})^{-1} u.
  \]
- Linear in \( u \).
- \( \|c\|_{L^2(\Omega)} \leq C\|u\|_{L^2(\Omega)} \).
- \( \mathcal{G} \) is bounded and linear.
Application 1: atmospheric source inversion

- Assume $y = G(u) + \eta$ where $\eta \sim \mathcal{N}(0, \sigma^2 I)$.
- $\Phi(u; y) = \frac{1}{2\sigma^2} \|G(u) - y\|_2^2$.

- Positivity constraint on source $u$.
- Sources are likely to be localized.
Application 2: high intensity focused ultrasound treatment

- Underlying aberration field $u$.
- Pointwise evaluation map for points $\{t_1, \cdots, t_d\}$ in $\mathbb{T}^2$
  \[S : C(\mathbb{T}^2) \to \mathbb{R}^m \quad (S(u))_j = u(t_j).\]
- (Experiments) A collection of vectors $\{z_j\}_{j=1}^m$ in $\mathbb{R}^d$.
- Quadratic forward map
  \[G : C(\mathbb{T}^2) \to \mathbb{R}^m \quad (G(u))_j := |z_j^T S(u)|^2.\]
- Phase retrieval in essence
Application 2: high intensity focused ultrasound treatment

- Assume \( y = G(u) + \eta \) where \( \eta \sim \mathcal{N}(0, \sigma^2 I) \).
- \( \Phi(u; y) = \frac{1}{2\sigma^2} \| G(u) - y \|_2^2 \).
- \( \| G(u) \|_2 \leq C \| u \|_{C^1(\mathbb{T}^2)}^2 \).
- Nonlinear forward map.

- Hydrophone experiments show sharp interfaces.
- Gaussian priors are too smooth.
We need to go beyond Gaussian priors!
Key questions

$$\frac{d\mu_y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)).$$

- Is $\mu^y$ well-defined?
- What happens if $y$ is perturbed?
- Easier to address when $X = \mathbb{R}^n$.
- More delicate when $X$ is infinite dimensional.
Outline

(i) General theory of well-posed Bayesian inverse problems.
(ii) Convex prior measures.
(iii) Models for compressible parameters.
(iv) Infinitely divisible prior measures.
Well-posedness

\[
\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y))
\]

**Definition:** Well-posed Bayesian inverse problem

Suppose \( X \) is a Banach space and \( d(\cdot, \cdot) \to \mathbb{R} \) is a probability metric. Given a prior \( \mu_0 \) and likelihood potential \( \Phi \), the problem of finding \( \mu^y \) is well-posed if:

(i) (Existence and uniqueness) There exists a unique posterior probability measure \( \mu^y \ll \mu_0 \) given by Bayes’ rule.

(ii) (Stability) For every choice of \( \epsilon > 0 \) there exists a \( \delta > 0 \) so that

\[
d(\mu^y, \mu^{y'}) \leq \epsilon \text{ for all } y, y' \in Y \text{ so that } \|y - y'\|_Y \leq \delta.
\]
The total variation and Hellinger metrics

\[ d_{TV}(\mu_1, \mu_2) := \frac{1}{2} \int_X \left| \frac{d\mu_1}{d\nu} - \frac{d\mu_2}{d\nu} \right| d\nu \]

\[ d_H(\mu_1, \mu_2) := \left( \frac{1}{2} \int_X \left( \sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}} \right)^2 d\nu \right)^{1/2}. \]

Note:

\[ 2d_H^2(\mu_1, \mu_2) \leq d_{TV}(\mu_1, \mu_2) \leq \sqrt{8}d_H(\mu_1, \mu_2). \]

Hellinger is more attractive in practice. For \( h \in L^2(X, \mu_1) \cap L^2(X, \mu_2) \)

\[ \left| \int_X h(u) d\mu_1(u) - \int_X h(u) d\mu_2(u) \right| \leq C(h)d_H(\mu_1, \mu_2). \]

Different convergence rates.
Well-posedness: analogy

- The likelihood $\Phi$ depends on the map $\tilde{G}$.
- Given $\Phi$ what classes of priors can be used?

**PDE analogy**

- A PDE where $g \in H^{-s}$ and $\mathcal{L} : H^p \to H^{-s}$ is a differential operator.
  
  \[ \mathcal{L}u = g \]

- Seek a solution $u = \mathcal{L}^{-1}g \in H^p$.
- Well-posedness depends on the smoothing behavior of $\mathcal{L}^{-1}$ and regularity of $g$.
- In the Bayesian approach we seek $\mu^y$ that satisfies
  
  \[ \mathcal{P} \mu^y = \mu_0. \]

- The mapping $\mathcal{P}^{-1}$ depends on $\Phi$.
- Well-posedness depends on behavior of $\mathcal{P}^{-1}$ and tail behavior of $\mu_0$.

**In a nutshell**, if $\Phi$ grows at a certain rate we have well-posedness if $\mu_0$ has sufficient tail decay.
Assumptions on likelihood

Minimal assumptions on $\Phi$ (BH, 2016)

The potential $\Phi : X \times Y \to \mathbb{R}$ satisfies:\n
(L1) (Locally bounded from below): There is a positive and non-decreasing function $f_1 : \mathbb{R}_+ \to [1, \infty)$ so that
\[
\Phi(u; y) \geq M - \log(f_1(\|u\|_X)).
\]

(L2) (Locally bounded from above):
\[
\Phi(u; y) \leq K.
\]

(L3) (Locally Lipschitz in $u$):
\[
|\Phi(u_1; y) - \Phi(u_2, y)| \leq L\|u_1 - u_2\|_X.
\]

(L4) (Continuity in $y$): There is a positive and non-decreasing function $f_2 : \mathbb{R}_+ \to \mathbb{R}_+$ so that
\[
|\Phi(u; y_1) - \Phi(u, y_2)| \leq C f_2(\|u\|_X)\|y_1 - y_2\|_Y.
\]

\textsuperscript{a}Stuart, “Inverse problems: a Bayesian perspective”.

Well-posedness: existence and uniqueness

- (L1) (Bounded from below) $\Phi(u; y) \geq M - \log(f_1(\|u\|_X))$.
- (L2) (Locally bounded from above) $\Phi(u; y) \leq K$.
- (L3) (Locally Lipschitz) $|\Phi(u_1; y) - \Phi(u_2, y)| \leq L\|u_1 - u_2\|_X$.

Existence and uniqueness (BH,2016)

Let $\Phi$ satisfy Assumptions L1–L3 with a function $f_1 \geq 1$, then the posterior $\mu^y$ is well-defined if $f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$.

Example:

If $y = G(u) + \eta, \eta \sim \mathcal{N}(0, \Sigma)$ then $\Phi(u; y) = \frac{1}{2}\|G(u) - y\|_\Sigma^2$ and so $M = 0$ and $f_1 = 1$ since $\Phi \geq 0$. 
Well-posedness: stability

- (L1) (Lower bound) $\Phi(u; y) \geq M - \log(f_1(\|u\|_X))$.
- (L2) (Locally bounded from above) $\Phi(u; y) \leq K$.
- (L4) (Continuity in $y$) $|\Phi(u; y_1) - \Phi(u, y_2)| \leq C f_2(\|u\|_X) \|y_1 - y_2\|_Y$.

Total variation stability (BH,2016)

Let $\Phi$ satisfy Assumptions L1, L2 and L4 with functions $f_1, f_2$ and let $\mu^y$ and $\mu^{y'}$ be two posterior measures for $y$ and $y' \in Y$. If $f_2(\|\cdot\|_X)f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$ then there is $C > 0$ such that $d_{TV}(\mu^y, \mu^{y'}) \leq C \|y - y'\|_Y$.

Hellinger stability (BH,2016)

If the stronger condition $(f_2(\|\cdot\|_X))^2 f_1(\|\cdot\|_X) \in L^1(X, \mu_0)$ is satisfied then there is $C > 0$ so that $d_H(\mu^y, \mu^{y'}) \leq C \|y - y'\|_Y$. 
The case of additive noise models

- let $Y = \mathbb{R}^m$, $\eta \sim \mathcal{N}(0, \Sigma)$ and suppose $y = G(u) + \eta$.
- $\Phi(u; y) = \frac{1}{2} \|G(u) - y\|_\Sigma^2$.
- $\Phi(u; y) \geq 0$ thus (L1) is satisfied with $f_1 = 1$ and $M = 0$.

Well-posedness with additive noise models (BH,2016)

Let the forward map $G$ satisfy:

(i) (Bounded) There is a positive and non-decreasing function $\tilde{f} \geq 1$ so that

$$\|G(u)\|_\Sigma \leq C\tilde{f}(\|u\|_X) \quad \forall u \in X.$$ 

(ii) (Locally Lipschitz)

$$\|G(u_1) - G(u_2)\|_\Sigma \leq K\|u_1 - u_2\|_X.$$ 

Then the problem of finding $\mu^y$ is well-posed if $\tilde{f}(\|\cdot\|_X) \in L^1(X, \mu_0)$. 

The case of additive noise models

**Example:** polynomialsly bounded forward map

Consider the additive noise model when $Y = \mathbb{R}^m$, $\eta \sim \mathcal{N}(0, I)$. Then

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

If $G$ is locally Lipschitz, $\|G(u)\|_2 \leq C \max\{1, \|u\|_X^p\}$ and $p \in \mathbb{N}$ then we have well-posedness if $\mu_0$ has bounded moments of degree $p$.

In particular, if $G$ is bounded and linear then it suffices for $\mu_0$ to have bounded moment of degree one. Recall the deconvolution example!

**Example:** Gaussian priors

In the setting of the above example, if $\mu_0$ is a centered Gaussian then it follows from Fernique’s theorem that we have well-posedness if $\|G(u)\|_2 \leq C \exp(\alpha \|u\|_X)$ for any $\alpha > 0$. 

(i) General theory of well-posed Bayesian inverse problems.
(ii) Convex prior measures ($\mu_0$ has exponential tails).
(iii) Models for compressible parameters.
(iv) Infinitely divisible prior measures.
From convex regularization to convex priors

- Let $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$.
- Common variational formulation for inverse problems

$$u^* = \arg\min_{v \in \mathbb{R}^n} \left\{ \frac{1}{2} \|G(v) - y\|_2^2 + \mathcal{R}(v) \right\}$$

$$\mathcal{R}(v) = \frac{\theta}{2} \|Lv\|_2^2 \quad \text{(Tikhonov)}, \quad \mathcal{R}(v) = \theta \|Lv\|_1 \quad \text{(Sparsity)}.$$  

- Bayesian analog

$$\frac{d\mu_y}{d\Lambda}(v) \propto \exp \left( -\frac{1}{2} \|G(v) - y\|_2^2 \right) \exp \left( -\mathcal{R}(v) \right).$$

- $\Lambda$ – Lebesgue measure.

A random variable with a log-concave Lebesgue density is convex.
Convex priors

- Gaussian, Laplace, Logistic, etc.
- $\ell_1$ regularization corresponds to Laplace priors.

$$\frac{d\mu^y}{d\Lambda}(v) \propto \exp\left(-\frac{1}{2}\|G(v) - y\|^2_\Sigma\right) \exp(-\|v\|_1).$$

$$\propto \exp\left(-\frac{1}{2}\|G(v) - y\|^2_\Sigma\right) \prod_{j=1}^n \exp(-|v_j|).$$

**Definition:** Convex measure

A Radon probability measure $\nu$ on $X$ is called convex whenever it satisfies the following inequality for $\beta \in [0, 1]$ and Borel sets $A, B \subset X$.

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

---

Convex priors

Convex measures have exponential tails\(^5\)

Let \( \nu \) be a convex measure on \( X \). If \( \| \cdot \|_X < \infty \) \( \nu \)-a.s. then there exists a constant \( \kappa > 0 \) so that \( \int_X \exp(\kappa \|u\|_X) d\nu(u) < \infty \).

Well-posedness with convex priors (BH & NN, 2016)

Let the prior \( \mu_0 \) be a convex measure assume

\[
\Phi(u; y) = \frac{1}{2} \|G(u) - y\|^2_{\Sigma}
\]

where \( G \) is locally Lipschitz and

\[
\|G(u)\|_{\Sigma} \leq C \max\{1, \|u\|_X^p\}, \quad \text{for} \quad p \in \mathbb{N}.
\]

Then we have a well-posed Bayesian inverse problem.

\(^5\)Borell, “Convex measures on locally convex spaces”.
Constructing convex priors

Product prior (BH & NN, 2016)

Suppose $X$ has an unconditional and normalized Schauder basis $\{x_k\}$.

(a) Pick a fixed sequence $\{\gamma_k\} \in \ell^2$.

(b) Pick a sequence of centered, real valued and convex random variables $\{\xi_k\}$ so that $\text{Var} \xi_k < \infty$ uniformly.

(c) Take $\mu_0$ to be the law of

$$u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k.$$ 

- $\| \cdot \|_X < \infty$, $\mu_0$-a.s. and $\| \cdot \|_X \in L^2(X, \mu_0)$.
- The $\xi_k$ are convex then so is $\mu_0$.
- Reminiscent of Karhunen-Loève expansion of Gaussians.

$$u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k, \quad \xi_k \sim \mathcal{N}(0, 1).$$

- $\{\gamma_k, x_k\}$ –eigenpairs of covariance operator.
Returning to deconvolution

\[ \varphi \ast u \]

**Example:** Deconvolution

Let \( X = L^2(\mathbb{T}) \) and assume \( \Phi(u; y) = \frac{1}{2} \| G(u) - y \|_2^2 \) where \( G(u) = S(\varphi \ast u) \).

Here \( \varphi \in C^\infty(\mathbb{T}) \) and \( S : C(\mathbb{T}) \to \mathbb{R}^m \) collects point values of a function at \( m \) distinct points \( \{t_j\} \).

We will construct a convex prior that is supported on \( B_{pp}^s(\mathbb{T}) \).
Example: deconvolution with a Besov type prior

- Let \( \{x_k\} \) be an \( r \)-regular wavelet basis for \( L^2(\mathbb{T}) \).
- For \( s < r, p \geq 1 \) define the Besov space \( B^s_{pp}(\mathbb{T}) \)

\[
B^s_{pp}(\mathbb{T}) := \left\{ w \in L^2(\mathbb{T}) : \sum_{k=1}^{\infty} k^{(sp-1/2)} |\langle w, x_k \rangle|^p < \infty \right\}
\]

- The prior \( \mu_0 \) is the law of \( u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k \).
- \( \xi_k \) are Laplace random variables with Lebesgue density \( \frac{1}{2} \exp(-|t|) \).
- \( \gamma_k = k^{-(\frac{1}{2p} + s)} \).

---


Example: deconvolution with a Besov type prior

- $\| \cdot \|_{B^{s}_{pp}(\mathbb{T})} < \infty \mu_0$-a.s. and $\mu_0$ is a convex measure.
- Forward map is bounded and linear.
- Problem is well-posed.\(^8\)

---

(i) General theory of well-posed Bayesian inverse problems.
(ii) Convex prior measures.
(iii) Models for compressible parameters.
(iv) Infinitely divisible prior measures.
A common problem in compressed sensing

\[ u^* = \arg \min_{v \in \mathbb{R}^n} \frac{1}{2} \| A v - y \|_2^2 + \theta \| v \|_p^p. \]

- \( p = 1 \), problem is convex.
- \( p < 1 \), no longer convex but a good model for compressibility.
- Bayesian analog

\[
\frac{d\mu_y}{d\Lambda}(v) \propto \exp \left( -\frac{1}{2} \| A v - y \|_2^2 \right) \prod_{j=1}^n \exp \left( -\theta |v_j|^p \right). 
\]
Models for compressibility

- $p = 1$.

- $p = 1/2$. 
Models for compressibility

- Symmetric generalized gamma prior for $0 < p, q \leq 1$

$$\frac{d\mu_0}{d\Lambda}(v) \propto \prod_{j=1}^{n} |v_j|^{p-1} \exp (-|v_j|^q).$$

- Corresponding posterior

$$\frac{d\mu_y}{d\Lambda}(v) \propto \exp \left(-\frac{1}{2} \|Av - y\|_2^2 - \|v\|_q^q + \sum_{j=1}^{n} (p - 1) \ln(|v_j|) \right)$$

- Maximizer is no longer well-defined.

- Perturbed variational analog for $\epsilon > 0$

$$u^*_\epsilon = \arg \min_{v \in \mathbb{R}^n} \frac{1}{2} \|Av - y\|_2^2 + \|v\|_q^q - \sum_{j=1}^{n} (p - 1) \ln(\epsilon + |v_j|)$$
Models for compressibility

\[ p = \frac{1}{2}, q = 1 \]

\[ p = q = \frac{1}{2} \]
Models for compressibility

- $SG(p, q, \alpha)$ density on the real line.

$$\frac{p}{2\alpha \Gamma(q/p)} \left| \frac{t}{\alpha} \right|^{p-1} \exp \left( - \left| \frac{t}{\alpha} \right|^q \right) d\Lambda(t).$$

- Has bounded moments of all order.

$SG(p, q, \alpha)$ prior: extension to infinite dimensions (BH,2016)

Suppose $X$ has an unconditional and normalized Schauder basis $\{x_k\}$.

(a) Pick a fixed sequence $\{\gamma_k\} \in \ell^2$.

(b) $\{\xi_k\}$ is an i.i.d sequence of $SG(p, q, \alpha)$ random variables.

(c) Take $\mu_0$ to be the law of $u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k$. 
Returning to deconvolution

**Example:** deconvolution with a \( SG(p, q, \alpha) \) prior

- Let \( \{x_k\} \) be the Fourier basis in \( L^2(\mathbb{T}) \).
- Define the Sobolev space \( H^1(\mathbb{T}) \)

\[
H^1(\mathbb{T}) := \left\{ w \in L^2(\mathbb{T}) : \sum_{k=1}^{\infty} (1 + k^2) |\langle w, x_k \rangle|^2 < \infty \right\}
\]

- The prior \( \mu_0 \) is the law of \( u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k \).
- \( \xi_k \) are i.i.d. \( SG(p, q, \alpha) \) random variables.
- \( \gamma_k = (1 + k^2)^{-3/4} \).
Example: deconvolution with a $SG(p, q, \alpha)$ prior

- $\| \cdot \|_{H^1(\mathbb{T})} < \infty \mu_0$-a.s.
- Forward map is bounded and linear.
- Problem is well-posed.
Outline

(i) General theory of well-posed Bayesian inverse problems.
(ii) Convex prior measures.
(iii) Models for compressible parameters.
(iv) **Infinitely divisible prior measures.**
Definition: infinitely divisible measure (ID)
A Radon probability measure \( \nu \) on \( X \) is infinitely divisible (ID) if for each \( n \in \mathbb{N} \) there exists a Radon probability measure \( \nu^{1/n} \) so that \( \nu = (\nu^{1/n})^* n \).

- \( \xi \) is ID if for any \( n \in \mathbb{N} \) there exist i.i.d random variables \( \{\xi_{1/n}^{1/n}\}_{k=1}^n \) so that \( \xi \overset{d}{=} \sum_{k=1}^n \xi_{k}^{1/n} \).
- \( SG(p, q, \alpha) \) priors are ID.
- Gaussian, Laplace, compound Poisson, Cauchy, student’s-t, etc.
- ID measures have an interesting compressible behavior\(^9\).

---

Deconvolution

**Example:** deconvolution with a compound Poisson prior

- Let \( \{x_k\} \) be the Fourier basis in \( L^2(\mathbb{T}) \).
- \( \mu_0 \) is the law of \( u \sim \sum_{k=1}^{\infty} \gamma_k \xi_k x_k \).
- \( \xi_k \) are i.i.d. compound Poisson random variables
  \[
  \xi_k \sim \sum_{j=0}^{\nu_k} \eta_{jk}.
  \]
- \( \nu_k \) are i.i.d Poisson random variables with rate \( b > 0 \).
- \( \eta_{jk} \) are i.i.d unit normals.
- \( \gamma_k = (1 + k^2)^{-3/4} \).
- \( \xi_k = 0 \) with probability \( e^{-b} \).
Example: deconvolution with a compound Poisson prior

- Truncations are sparse in the strict sense.
- $\| \cdot \|_{H^1(\mathbb{T})} < \infty$ a.s.
- We have well-posedness.
Lévy-Khintchine

- Recall the characteristic function of a measure $\mu$ on $X$

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

Lévy-Khintchine representation of ID measures

A Radon probability measure on $X$ is infinitely divisible if and only if there exists an element $m \in X$, a (positive definite) covariance operator $Q : X^* \rightarrow X$ and a Lévy measure $\lambda$, so that

$$\hat{\mu}(\varrho) = \exp(\psi(\varrho))$$

$$\psi(\varrho) = \underbrace{i\varrho(m)}_{\text{point mass}} - \frac{1}{2}\varrho(Q(\varrho)) + \int_X \exp(i\varrho(u)) - 1 - i\varrho(u)\mathbf{1}_{B_X}(u)d\lambda(u).$$

- $ID(m, Q, \lambda)$.
- If $\lambda$ is a symmetric probability measure on $X$

$$ID(m, Q, \lambda) = \delta_m * \mathcal{N}(0, Q) * \text{compound Poisson}.$$
Tail behavior of ID measures and well-posedness

- Tail behavior of ID is tied to the tail behavior of the Lévy measure $\lambda$

Moments of ID measures

Suppose $\mu = \text{ID}(m, Q, \lambda)$. If $0 < \lambda(X) < \infty$ and $\| \cdot \|_X < \infty \text{ } \mu\text{-a.s.}$ then

$$\| \cdot \|_X \in L^p(X, \mu) \text{ whenever } \| \cdot \|_X \in L^p(X, \lambda) \text{ for } p \in [1, \infty).$$

Well-posedness with ID priors \textit{(BH,2016)}

Suppose $\mu_0 = \text{ID}(m, Q, \lambda)$, $0 < \lambda(X) < \infty$ and take $\Phi(u; y) = \frac{1}{2} \| G(u) - y \|_\Sigma^2$. If $\max\{1, \| \cdot \|_X^p\} \in L^1(X, \lambda)$ for $p \in \mathbb{N}$ and $G$ is locally Lipschitz so that

$$\| G(u) \|_X \leq C \max\{1, \| u \|_X^p\},$$

then we have a well-posed Bayesian inverse problem.
Deconvolution once more

**Example:** deconvolution with a BV prior

- Consider the deconvolution problem on $\mathbb{T}$.
- Stochastic process $u(t)$ for $t \in (0, 1)$ defined via

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

- $\nu$ is a symmetric measure and $\int_{|\xi| \leq 1} |\xi| \, d\nu(\xi) < \infty$.
- Pure jump Lévy process.
- Similar to the Cauchy difference prior\(^\text{10}\).

---

Deconvolution once more

**Example:** deconvolution with a BV prior

- $u$ has countably many jump discontinuities.
- $\|u\|_{BV(T)} < \infty$ a.s.$^{11}$
- $\mu_0$ is the measure induced by $u(t)$.
- BV is non-separable.
- Forward map is bounded and linear.
- Well-posed problem.

---

Closing remarks

- Well-posedness can be achieved with relaxed conditions.
- Gaussians have serious limitations in terms of modelling.
- Many different priors to choose from.
Closing remarks

- Sampling.
- Random walk Metropolis-Hastings for self-decomposable priors.
- Randomize-then-optimize\(^\text{12}\).
- Fast Gibbs sampler\(^\text{13}\).


Closing remarks

- Analysis of priors:
  - What constitutes compressibility?
  - What is the support of the prior?
- Hierarchical priors.
- Modelling constraints.
Thank you


References


### Well-posedness

#### Minimal assumptions on $\Phi$ (BH, 2016)

The potential $\Phi : X \times Y \to \mathbb{R}$ satisfies:

1. **(L1) (Lower bound in $u$):** There is a positive and non-decreasing function $f_1 : \mathbb{R}_+ \to [1, \infty)$ so that for all $r > 0$, there is a constant $M(r) \in \mathbb{R}$ such that $\forall u \in X$ and $\forall y \in Y$ with $\|y\|_Y < r$,

   $$\Phi(u; y) \geq M - \log (f_1(\|u\|_X)).$$

2. **(L2) (Boundedness above):** For all $r > 0$ there is a constant $K(r) > 0$ such that $\forall u \in X$ and $\forall y \in Y$ with $\max\{\|u\|_X, \|y\|_Y\} < r$,

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

3. **(L3) (Continuity in $u$):** For all $r > 0$ there exists a constant $L(r) > 0$ such that $\forall u_1, u_2 \in X$ and $y \in Y$ with $\max\{\|u_1\|_X, \|u_2\|_X, \|y\|_Y\} < r$,

   $$|\Phi(u_1; y) - \Phi(u_2, y)| \leq L\|u_1 - u_2\|_X.$$

4. **(L4) (Continuity in $y$):** There is a positive and non-decreasing function $f_2 : \mathbb{R}_+ \to \mathbb{R}_+$ so that for all $r > 0$, there is a constant $C(r) \in \mathbb{R}$ such that $\forall y_1, y_2 \in Y$ with $\max\{\|y_1\|_Y, \|y_2\|_Y\} < r$ and $\forall u \in X$,

   $$|\Phi(u; y_1) - \Phi(u, y_2)| \leq Cf_2(\|u\|_X)\|y_1 - y_2\|_Y.$$
The case of additive noise models

Well-posedness with additive noise models

Consider the above additive noise model. In addition, let the forward map $G$ satisfy the following conditions with a positive, non-decreasing and locally bounded function $\tilde{f} \geq 1$:

(i) (Bounded) There is a constant $C > 0$ for which

$$
\|G(u)\|_\Sigma \leq C \tilde{f}(\|u\|_X) \quad \forall u \in X.
$$

(ii) (Locally Lipschitz) \( \forall r > 0 \) there is a constant $K(r) > 0$ so that for all $u_1, u_2 \in X$ and

$$
\max\{\|u_1\|_X, \|u_2\|_X\} < r
$$

$$
\|G(u_1) - G(u_2)\|_\Sigma \leq K \|u_1 - u_2\|_X.
$$

Then the problem of finding $\mu^y$ is well-posed if $\mu_0$ is a Radon probability measure on $X$ such that $\tilde{f}(\|\cdot\|_X) \in L^1(X, \mu_0)$. 