Bayesian methods for inverse problems

Masoumeh Dashti

Department of Mathematics
University of Sussex

CEMRACS 2021 – Summer school

19 July 2021

Inverse problems

Encountered when indirect measurements, y, of quantity of interest, u, is available

$$y \approx \mathcal{G}(u)$$

with $\mathcal{G}$ well-posed.

- $\mathcal{G}^{-1}$ might not be well-posed
- measurements are typically inaccurate, hence y might lie outside of range of $\mathcal{G}$

Example. Inverse heat equation

$$v_t - v_{xx} = 0, \quad x \in (0,1), \quad t > 0$$
$$v(0,t) = v(1,t) = 0, \quad t \geq 0.$$

Find $u(x) := v(x,0)$, given measurements $y(x) \approx v(x,T)$.

We have

$$v(x,t) = \sum_{n=1}^{\infty} c_n \, e^{-(n\pi)^2 t} \sin n\pi x$$

with c_n satisfying $u(x) = \sum_{n=1}^{\infty} c_n \sin n\pi x$.

Instability in data: For

$$y(x) - v(x,T) = \alpha \sin 5\pi x,$$

$$\tilde{u}(x) - u(x) = \alpha \, e^{25\pi^2 T} \sin 5\pi x$$

Example. Inverse heat equation

$$v_t - v_{xx} = 0, \quad x \in (0,1), \quad t > 0$$
$$v(0,t) = v(1,t) = 0, \quad t \geq 0.$$

Find $u(x) := v(x,0)$, given measurements $y(x) \approx v(x,T)$.

We have

$$v(x,t) = \sum_{n=1}^{\infty} c_n \, \mathrm{e}^{-(n\pi)^2 t} \sin n\pi x$$

with c_n satisfying $u(x) = \sum_{n=1}^{\infty} c_n \sin n\pi x$.

Instability in data: For

$$y(x) - v(x,T) = \alpha \sin 5\pi x,$$

$$\tilde{u}(x) - u(x) = \alpha \, \mathrm{e}^{25\pi^2 T} \sin 5\pi x$$

Example. Inverse heat equation

$$v_t - v_{xx} = 0, \quad x \in (0,1), \quad t > 0$$
$$v(0,t) = v(1,t) = 0, \quad t \geq 0.$$

Find $u(x) := v(x,0)$, given measurements $y(x) \approx v(x,T)$.

We have

$$v(x,t) = \sum_{n=1}^{\infty} c_n \, \mathrm{e}^{-(n\pi)^2 t} \sin n\pi x$$

with c_n satisfying $u(x) = \sum_{n=1}^{\infty} c_n \sin n\pi x$.

Instability in data: For

$$y(x) - v(x,T) = \alpha \sin 5\pi x,$$

$$\tilde{u}(x) - u(x) = \alpha \, \mathrm{e}^{25\pi^2 T} \sin 5\pi x$$

Example. Inverse problem for an elliptic PDE:

$$-\nabla \cdot \big(u(x)\,\nabla p(x)\big) = f, \quad x \in \Omega \subset \mathbb{R}^d$$
$$p(x) = 0, \quad x \in \partial\Omega.$$

Find u, given $y_j = p(x_j) + \eta_j$, $j = 1, \ldots, J$, $\eta_j \sim \mathcal{N}(0, \gamma_j)$, i.i.d.

$y = \mathcal{G}(u) + \eta$ with

- forward map $\mathcal{G} : u \mapsto \begin{pmatrix} p(x_1) \\ \vdots \\ p(x_J) \end{pmatrix}$

  data $y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix}$ and the noise vector $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_J \end{pmatrix}$

Example. Inverse problem for an elliptic PDE:

$$-\nabla \cdot \big(u(x)\,\nabla p(x)\big) = f, \quad x \in \Omega \subset \mathbb{R}^d$$
$$p(x) = 0, \quad x \in \partial\Omega.$$

Find u, given $y_j = p(x_j) + \eta_j$, $j = 1, \ldots, J$, $\eta_j \sim \mathcal{N}(0, \gamma_j)$, i.i.d.

$y = \mathcal{G}(u) + \eta$ with

- forward map $\mathcal{G} : u \mapsto \begin{pmatrix} p(x_1) \\ \vdots \\ p(x_J) \end{pmatrix}$

  data $y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix}$ and the noise vector $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_J \end{pmatrix}$

Classical approach

aiming to find a reasonable estimate of the quantity of interest

- Least-squares solution

$$u^* = \arg\min_{u \in X} \|y - \mathcal{G}(u)\|_Y$$

- Regularised solution

$$u^* = \arg\min_{u \in X} \|y - \mathcal{G}(u)\|_Y^2 + \alpha \mathcal{J}(u)$$

(typically $\mathcal{J}(u) = \|u\|_E^p$ for $E \subseteq X$ and $p \geq 1$)

Bayesian approach

For estimating u from

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

- u, y, and η, are modelled as random variables:

  Degree of information about values of these variables are expressed in the form of probability distributions

- The solution is the posterior probability distribution given via the Bayes rule

Bayesian approach

For estimating u from

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

- u, y, and η, are modelled as random variables:

  Degree of information about values of these variables are expressed in the form of probability distributions

- The solution is the posterior probability distribution given via the Bayes rule

Bayesian approach

Let

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

with $y \in \mathbb{R}^J$, $u \in X$, (X separable Banach spaces), and suppose

- prior $u \sim \mu_0$
- statistics of noise is known: $\eta \sim \rho_\eta$
- u and η are independent, $u \perp \eta$

Bayes' rule: $\mathrm{Prob}(u \,|\, \text{data}) \propto \mathrm{Prob}(\text{data} \,|\, u)\, \mathrm{Prob}(u)$

Noting that $\mathrm{data}|u \sim \rho_\eta(\cdot - \mathcal{G}(u))$ we have

$$\mu^y(\mathrm{d}u) \propto \rho_\eta(y - \mathcal{G}(u))\, \mu_0(\mathrm{d}u).$$

(when well-defined)

Bayesian approach

Let

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

with $y \in \mathbb{R}^J$, $u \in X$, (X separable Banach spaces), and suppose

- prior $u \sim \mu_0$
- statistics of noise is known: $\eta \sim \rho_\eta$
- u and η are independent, $u \perp \eta$

Bayes' rule: $\mathrm{Prob}(u \,|\, \text{data}) \propto \mathrm{Prob}(\text{data} \,|\, u)\, \mathrm{Prob}(u)$

Noting that $\mathrm{data}|u \sim \rho_\eta(\cdot - \mathcal{G}(u))$ we have

$$\mu^y(\mathrm{d}u) \propto \rho_\eta(y - \mathcal{G}(u))\, \mu_0(\mathrm{d}u).$$

(when well-defined)

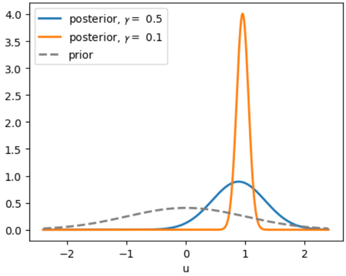
Example. Let $X = Y = \mathbb{R}$

$$y = g(u) + \eta, \quad \eta \sim \mathcal{N}(0, \gamma^2)$$
$$\mu_0 \sim \mathcal{N}(0, 1)$$

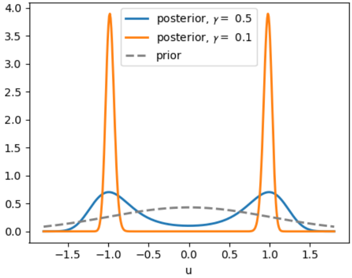
Then

$$\mu^y(\mathrm{d}u) \propto \mathrm{e}^{-\frac{1}{2\gamma^2}|y-g(u)|^2}\, \mathrm{e}^{-\frac{1}{2}u^2}\, \mathrm{d}u$$

- $g(u) = u$,
  $\mu^y \sim \mathcal{N}\left(\frac{y}{1+\gamma^2}, \frac{\gamma^2}{1+\gamma^2}\right)$



- $g(u) = u^2$,
  $\rho_{\mu^y} = \mathrm{e}^{-\frac{1}{2\gamma^2}|y-u^2|^2}\, \mathrm{e}^{-\frac{1}{2}u^2}$



Outline

Outline

Well-definedness of postrior

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \rho_\eta(y - \mathcal{G}(u)) =: \mathrm{e}^{-\Phi(u,y)}, \quad Z := \int_X \mathrm{e}^{-\Phi(u,y)}\,\mu_0(\mathrm{d}u)$$

ρ_η density of measure $\mathbb{Q}$ on $Y = \mathbb{R}^J$,
X separable Banach space

Assume that

- $\mu_0(X) = 1$
- $\Phi : X \times \mathbb{R}^J \to \mathbb{R}$ is $\mu_0 \otimes \mathbb{Q}$ measurable
- $0 < Z < \infty$.

Then $\mu^y \ll \mu_0$ and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u,y))$$

Well-definedness of postrior

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \rho_\eta(y - \mathcal{G}(u)) =: \mathrm{e}^{-\Phi(u,y)}, \quad Z := \int_X \mathrm{e}^{-\Phi(u,y)}\, \mu_0(\mathrm{d}u)$$

ρ_η density of measure $\mathbb{Q}$ on $Y = \mathbb{R}^J$,
X separable Banach space

Assume that

- $\mu_0(X) = 1$
- $\Phi : X \times \mathbb{R}^J \to \mathbb{R}$ is $\mu_0 \otimes \mathbb{Q}$ measurable
- $0 < Z < \infty$.

Then $\mu^y \ll \mu_0$ and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u, y))$$

Stability in data

If $y, y' \in \mathbb{R}^J$ are close, μ^y and $\mu^{y'}$ remain 'close':

To measure the distance between probability measures we use

$$\textit{Hellinger metric: } d^2_{\mathrm{Hell}}(\mu, \mu') = \frac{1}{2}\int_X \Big(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\Big)^2 d\nu$$

Suppose also that for $\max(|y_1|, |y_2|) < r$

$$|\Phi(u, y_1) - \Phi(u, y_2)| \leq M(r, \|u\|_X)\, |y_1 - y_2|$$

$$\text{with} \quad \int_X M(r, \|u\|_X)\, \mathrm{d}\mu_0 < \infty$$

Then

$$d_{\mathrm{Hell}}(\mu^{y_1}, \mu^{y_2}) \leq C|y_1 - y_2|$$

Stability in data

If $y, y' \in \mathbb{R}^J$ are close, μ^y and $\mu^{y'}$ remain 'close':

To measure the distance between probability measures we use

$$\textit{Hellinger metric:} \quad d^2_{\mathrm{Hell}}(\mu, \mu') = \frac{1}{2}\int_X \Big(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\,\Big)^2 d\nu$$

Suppose also that for $\max(|y_1|, |y_2|) < r$

$$|\Phi(u, y_1) - \Phi(u, y_2)| \;\leq\; M(r, \|u\|_X)\,|y_1 - y_2|$$

$$\text{with} \quad \int_X M(r, \|u\|_X)\,\mathrm{d}\mu_0 < \infty$$

Then

$$d_{\mathrm{Hell}}(\mu^{y_1}, \mu^{y_2}) \leq C|y_1 - y_2|$$

Stability in data

If $y, y' \in \mathbb{R}^J$ are close, μ^y and $\mu^{y'}$ remain 'close':

To measure the distance between probability measures we use

$$\textit{Hellinger metric:}\;\; d^2_{\mathrm{Hell}}(\mu, \mu') = \frac{1}{2}\int_X \Big(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\,\Big)^2 d\nu$$

Suppose also that for $\max(|y_1|, |y_2|) < r$

$$|\Phi(u, y_1) - \Phi(u, y_2)| \;\le\; M(r, \|u\|_X)\,|y_1 - y_2|$$

$$\text{with} \quad \int_X M(r, \|u\|_X)\,\mathrm{d}\mu_0 < \infty$$

Then

$$d_{\mathrm{Hell}}(\mu^{y_1}, \mu^{y_2}) \le C|y_1 - y_2|$$

- With Hellinger metric (for $f : X \to E$ in $L^2_\mu \cap L^2_\nu$)

$$\|\mathbb{E}^\mu f(u) - \mathbb{E}^{\mu'} f(u)\|_E \le C\, d_{\mathrm{Hell}}(\mu, \mu')$$

We have

$$\mathbb{E}^\mu f - \mathbb{E}^{\mu'} f = \int_X f d\mu - \int_X f d\mu' = \int f \left(\frac{d\mu}{d\nu} - \frac{d\mu'}{d\nu}\right) d\nu$$
$$= \int f(u) \left(\sqrt{\frac{d\mu}{d\nu}} + \sqrt{\frac{d\mu'}{d\nu}}\right)\left(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right) d\nu$$

Hence

$$\|\mathbb{E}^\mu f - \mathbb{E}^{\mu'} f\|_E \le \int \|f\|_E \sqrt{\frac{d\mu}{d\nu}} \left|\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right| d\nu$$
$$+ \int \|f\|_E \sqrt{\frac{d\mu'}{d\nu}} \left|\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right| d\nu$$
$$\le 2(\mathbb{E}^\mu \|f\|_E^2 + \mathbb{E}^{\mu'} \|f\|_E^2)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu, \mu')$$

- With Hellinger metric (for $f : X \to E$ in $L^2_\mu \cap L^2_\nu$)

$$\|\mathbb{E}^\mu f(u) - \mathbb{E}^{\mu'} f(u)\|_E \le C\, d_{\mathrm{Hell}}(\mu, \mu')$$

We have

$$\mathbb{E}^\mu f - \mathbb{E}^{\mu'} f = \int_X f d\mu - \int_X f d\mu' = \int f \left(\frac{d\mu}{d\nu} - \frac{d\mu'}{d\nu}\right) d\nu$$

$$= \int f(u) \left(\sqrt{\frac{d\mu}{d\nu}} + \sqrt{\frac{d\mu'}{d\nu}}\right)\left(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right) d\nu$$

Hence

$$\|\mathbb{E}^\mu f - \mathbb{E}^{\mu'} f\|_E \le \int \|f\|_E \sqrt{\frac{d\mu}{d\nu}} \left|\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right| d\nu$$

$$+ \int \|f\|_E \sqrt{\frac{d\mu'}{d\nu}} \left|\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right| d\nu$$

$$\le 2(\mathbb{E}^\mu \|f\|_E^2 + \mathbb{E}^{\mu'} \|f\|_E^2)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu, \mu')$$

- With Hellinger metric (for $f : X \to E$ in $L^2_\mu \cap L^2_\nu$)

$$\|\mathbb{E}^{\mu} f(u) - \mathbb{E}^{\mu'} f(u)\|_E \leq C\, d_{\mathrm{Hell}}(\mu, \mu')$$

We have

$$\mathbb{E}^{\mu} f - \mathbb{E}^{\mu'} f = \int_X f d\mu - \int_X f d\mu' = \int f \,(\frac{d\mu}{d\nu} - \frac{d\mu'}{d\nu}) d\nu$$

$$= \int f(u) \Big(\sqrt{\frac{d\mu}{d\nu}} + \sqrt{\frac{d\mu'}{d\nu}}\Big)\Big(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\Big) d\nu$$

Hence

$$\|\mathbb{E}^{\mu} f - \mathbb{E}^{\mu'} f\|_E \leq \int \|f\|_E \sqrt{\frac{d\mu}{d\nu}} \left|\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right| d\nu$$

$$+ \int \|f\|_E \sqrt{\frac{d\mu'}{d\nu}} \left|\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\right| d\nu$$

$$\leq 2(\mathbb{E}^{\mu} \|f\|_E^2 + \mathbb{E}^{\mu'} \|f\|_E^2)^{\frac{1}{2}}\, d_{\mathrm{Hell}}(\mu, \mu')$$

Well-definedness. Assume that

- $\mu_0(X) = 1$
- $\Phi : X \times \mathbb{R}^J \to \mathbb{R}$ is $\mu_0 \otimes \mathbb{Q}$ measurable
- $0 < Z < \infty$.

Then $\mu^y \ll \mu_0$ and
$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u, y))$$
(Stuart '10, D. Stuart '17)

Stability in data. Suppose also that for $\max(|y_1|, |y_2|) < r$

$$|\Phi(u, y_1) - \Phi(u, y_2)| \leq M(r, \|u\|_X)\, |y_1 - y_2|$$

$$\text{with} \quad \int_X M(r, \|u\|_X)\, \mathrm{d}\mu_0 < \infty$$

Then $d_{\mathrm{Hell}}(\mu^{y_1}, \mu^{y_2}) \leq C|y_1 - y_2|$.

(Stuart '10)

If ρ_η is Gaussian, blue conditions are enough for continuity in d_{Hell} (Latz '20)

The prior

- $(X, \|\cdot\|)$ separable Banach space with basis $\{\psi_j\}_{j\in\mathbb{N}}$
- *Define μ_0 through Karhunen-Loève expansion of its draws:*

$$u(x) = m_0 + \sum_{j\in\mathbb{N}} \alpha_j \, \xi_j \, \psi_j(x)$$

$m_0 \in X$ given

ξ_j i.i.d random variables,

$\{\alpha_j\}$ decreasing sequence determining smoothness of u

Example: $X = L^2(0,1), \quad \psi_j(x) = \sqrt{2}\,\sin(2j\pi x), \quad m_0 = 0$

$\alpha_j = j^{-s}$, with $s > 1/2, \quad \mathbb{E}(\xi_1^2) < \infty$

$$\mathbb{E}\|u\|_{L^2}^2 = \mathbb{E}(\xi_1^2) \sum_j \alpha_j^2 < \infty \quad \text{hence} \quad \mu(L^2) = 1$$

The prior

- $(X, \|\cdot\|)$ separable Banach space with basis $\{\psi_j\}_{j\in\mathbb{N}}$
- *Define μ_0 through Karhunen-Loève expansion of its draws:*

$$u(x) = m_0 + \sum_{j\in\mathbb{N}} \alpha_j\,\xi_j\,\psi_j(x)$$

$m_0 \in X$ given

ξ_j i.i.d random variables,

$\{\alpha_j\}$ decreasing sequence determining smoothness of u

Example: $X = L^2(0,1)$, $\psi_j(x) = \sqrt{2}\,\sin(2j\pi x)$, $m_0 = 0$

$\alpha_j = j^{-s}$, with $s > 1/2$, $\mathbb{E}(\xi_1^2) < \infty$

$$\mathbb{E}\|u\|_{L^2}^2 = \mathbb{E}(\xi_1^2)\sum_j \alpha_j^2 < \infty \quad \text{hence} \quad \mu(L^2) = 1$$

The prior

- $(X, \|\cdot\|)$ separable Banach space with basis $\{\psi_j\}_{j\in\mathbb{N}}$
- *Define μ_0 through Karhunen-Loève expansion of its draws:*

$$u(x) = m_0 + \sum_{j\in\mathbb{N}} \alpha_j \, \xi_j \, \psi_j(x)$$

$m_0 \in X$ given

ξ_j i.i.d random variables,

$\{\alpha_j\}$ decreasing sequence determining smoothness of u

Example: $X = L^2(0,1)$, $\psi_j(x) = \sqrt{2}\,\sin(2j\pi x)$, $m_0 = 0$

$\alpha_j = j^{-s}$, with $s > 1/2$, $\mathbb{E}(\xi_1^2) < \infty$

$$\mathbb{E}\|u\|_{L^2}^2 = \mathbb{E}(\xi_1^2) \sum_j \alpha_j^2 < \infty \quad \text{hence} \quad \mu(L^2) = 1$$

$u(x) = m_0 + \sum_{j \in \mathbb{N}} \alpha_j \, \xi_j \, \psi_j(x)$

- Gaussian μ_0:

  $X \subseteq L^2(D)$, $D \subset \mathbb{R}^d$,

  $\{\psi_j\}$ an orthonormal basis (regular enough),

  $\xi_j \sim c \, \exp(-\frac{1}{2}|x|^2), \quad \alpha_j = j^{-\frac{s}{d}}$

  $\mu_0(C^t) = 1$, for $t < s - \frac{d}{2}$

- Besov μ_0:

  $X \subseteq L^2(D)$, $D \subset \mathbb{R}^d$,

  $\{\psi_j\}$ wavelet basis (regular enough),

  $\xi_j \sim c \, \exp(-\frac{1}{2}|x|^p)$, $p \geq 1 \quad \alpha_j = j^{-(\frac{s}{d} + \frac{1}{2} - \frac{1}{p})}$

  $\mu_0(C^t) = 1$, for $t < s - \frac{d}{p}$

$u(x) = m_0 + \sum_{j \in \mathbb{N}} \alpha_j \, \xi_j \, \psi_j(x)$

- Gaussian μ_0:

  $X \subseteq L^2(D)$, $D \subset \mathbb{R}^d$,

  $\{\psi_j\}$ an orthonormal basis (regular enough),

  $\xi_j \sim c \exp(-\frac{1}{2}|x|^2), \quad \alpha_j = j^{-\frac{s}{d}}$

  $\mu_0(C^t) = 1$, for $t < s - \frac{d}{2}$

- Besov μ_0:

  $X \subseteq L^2(D)$, $D \subset \mathbb{R}^d$,

  $\{\psi_j\}$ wavelet basis (regular enough),

  $\xi_j \sim c \exp(-\frac{1}{2}|x|^p)$, $p \geq 1 \quad \alpha_j = j^{-(\frac{s}{d} + \frac{1}{2} - \frac{1}{p})}$

  $\mu_0(C^t) = 1$, for $t < s - \frac{d}{p}$

Inverse elliptic problem – wellposedness

Example. Inverse problem for an elliptic PDE:

$$-\nabla \cdot \big(u(x)\,\nabla p(x)\big) = f, \quad x \in \Omega \subset \mathbb{R}^d$$
$$p(x) = 0, \quad x \in \partial\Omega.$$

Find u, given $y_j = p(x_j) + \eta_j,\ j = 1, \ldots, J,\ \ \eta_j \sim \mathcal{N}(0, \gamma_j)$

$\mathcal{G}(u) = (p(x_1), \cdots, p(x_J))^T \qquad \Phi(u, y) = \frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2$

To show wellposedness we need to:

- Find X so that $\mathcal{G} : X \to \mathbb{R}^J$ is continuous
- Choose μ_0 such that $\mu_0(X) = 1$:

Inverse elliptic problem – wellposedness

Example. Inverse problem for an elliptic PDE:

$$-\nabla \cdot \big(u(x)\,\nabla p(x)\big) = f, \quad x \in \Omega \subset \mathbb{R}^d$$
$$p(x) = 0, \quad x \in \partial\Omega.$$

Find u, given $y_j = p(x_j) + \eta_j,\ j = 1, \ldots, J,\ \ \eta_j \sim \mathcal{N}(0, \gamma_j)$

$\mathcal{G}(u) = (p(x_1), \cdots, p(x_J))^T \qquad \Phi(u, y) = \frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2$

To show wellposedness we need to:

- Find X so that $\mathcal{G} : X \to \mathbb{R}^J$ is continuous
- Choose μ_0 such that $\mu_0(X) = 1$:

- $\mathcal{G} \in C^{0,1}(X)$:

  Let $0 < m := \inf_{x\in D}\{u_1(x), u_2(x)\}$ and $L = L(\|u_1\|_{L^\infty}, \|u_2\|_{L^\infty})$

  $$|\mathcal{G}(u_1) - \mathcal{G}(u_2)| \leq \frac{L}{m} \|u_1 - u_2\|_{L^\infty},$$

  hence $X \subset \left\{w \in L^\infty : \inf_x w(x) > 0\right\}$

- Choose μ_0 such that $\mu_0(X) = 1$:
  * Let $\xi_j \sim c \exp(-\frac{1}{2}|x|^2)$, $\alpha_j = j^{-\frac{s}{d}}$ with $s > \frac{d}{2}$

    define $\tilde{\mu}_0$ through

    $$\tilde{u}(x) = \sum_{j\in\mathbb{N}} \alpha_j \xi_j \psi_j(x)$$

    then $\tilde{\mu}_0(C^t(D)) = 1$ for $t < s - \frac{d}{2}$,
  * Set $u(x) = g(\tilde{u}(x)) := e^{\tilde{u}(x)}$,

    then $u \sim \mu_0$ with $\mu_0(X) = 1$ where $X = \left\{w \in C^t : \inf_x w(x) > 0\right\}$

- Then μ with $$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0} \propto \exp(-\Phi(u, y))$$ is wellposed.

- $\mathcal{G} \in C^{0,1}(X)$:

  Let $0 < m := \inf_{x \in D}\{u_1(x), u_2(x)\}$ and $L = L(\|u_1\|_{L^\infty}, \|u_2\|_{L^\infty})$

$$|\mathcal{G}(u_1) - \mathcal{G}(u_2)| \leq \frac{L}{m}\, \|u_1 - u_2\|_{L^\infty},$$

  hence $X \subset \{w \in L^\infty : \inf_x w(x) > 0\}$

- Choose μ_0 such that $\mu_0(X) = 1$:
  * Let $\xi_j \sim c\, \exp(-\frac{1}{2}|x|^2)$, $\alpha_j = j^{-\frac{s}{d}}$ with $s > \frac{d}{2}$

    define $\tilde{\mu}_0$ through

$$\tilde{u}(x) = \sum_{j \in \mathbb{N}} \alpha_j\, \xi_j\, \psi_j(x)$$

    then $\tilde{\mu}_0(C^t(D)) = 1$ for $t < s - \frac{d}{2}$,

  * Set $u(x) = g(\tilde{u}(x)) := e^{\tilde{u}(x)}$,

    then $u \sim \mu_0$ with $\mu_0(X) = 1$ where $X = \{w \in C^t : \inf_x w(x) > 0\}$

- Then μ with $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0} \propto \exp(-\Phi(u, y))$ is wellposed.

- $\mathcal{G} \in C^{0,1}(X)$:

  Let $0 < m := \inf_{x\in D}\{u_1(x), u_2(x)\}$ and $L = L(\|u_1\|_{L^\infty}, \|u_2\|_{L^\infty})$

  $$|\mathcal{G}(u_1) - \mathcal{G}(u_2)| \leq \frac{L}{m} \|u_1 - u_2\|_{L^\infty},$$

  hence $X \subset \{w \in L^\infty : \inf_x w(x) > 0\}$

- Choose μ_0 such that $\mu_0(X) = 1$:
  * Let $\xi_j \sim c \exp(-\frac{1}{2}|x|^2)$, $\alpha_j = j^{-\frac{s}{d}}$ with $s > \frac{d}{2}$

    define $\tilde{\mu}_0$ through

    $$\tilde{u}(x) = \sum_{j\in\mathbb{N}} \alpha_j \, \xi_j \, \psi_j(x)$$

    then $\tilde{\mu}_0(C^t(D)) = 1$ for $t < s - \frac{d}{2}$,
  * Set $u(x) = g(\tilde{u}(x)) := e^{\tilde{u}(x)}$,

    then $u \sim \mu_0$ with $\mu_0(X) = 1$ where $X = \{w \in C^t : \inf_x w(x) > 0\}$

- Then μ with $$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0} \propto \exp(-\Phi(u, y))$$ is wellposed.

Inverse elliptic problem

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0} \propto \exp(-\tfrac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2)$$

Summary statistics

- point estimators: conditional mean $\int_X u \,\mathrm{d}\mu$

  most likely functions under μ in X
- $\int_X F(u) \,\mathrm{d}\mu$

Need to approximate $\mathcal{G}$ and μ

Outline

Consider μ and μ^N:

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u)), \qquad \frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi^N(u))$$

Theorem. (Stuart 2010)
Assume that

$$|\Phi(u) - \Phi^N(u)| \leq M(\|u\|_X)\,\psi(N)$$

where $\int_X M(r, \|u\|_X)\,\mathrm{d}\mu_0 < \infty$ *and* $\psi(N) \to 0$ *as* $N \to \infty$

Then

$$d_{\mathrm{Hell}}(\mu, \mu^N) \leq C\,\psi(N).$$

Consider μ and μ^N:

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u)), \qquad \frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi^N(u))$$

Theorem. (Stuart 2010)

Assume that

$$|\Phi(u) - \Phi^N(u)| \le M(\|u\|_X)\,\psi(N)$$

where $\int_X M(r, \|u\|_X)\,\mathrm{d}\mu_0 < \infty$ *and* $\psi(N) \to 0$ *as* $N \to \infty$

Then

$$d_{\mathrm{Hell}}(\mu, \mu^N) \le C\,\psi(N).$$

$$\begin{aligned}\Big(d_{\mathrm{Hell}}(\mu,\mu^N)\Big)^2 &= \frac{1}{2}\int_X \Big(\sqrt{\frac{d\mu}{d\mu_0}} - \sqrt{\frac{d\mu^N}{d\mu_0}}\ \Big)^2 d\mu_0 \\ &= \frac{1}{2}\int_X \Big(\frac{1}{\sqrt{Z}}\mathrm{e}^{-\frac{1}{2}\Phi} - \frac{1}{\sqrt{Z^N}}\mathrm{e}^{-\frac{1}{2}\Phi^N}\Big)^2 d\mu_0 \\ &\le \frac{1}{Z}\int_X \big(\mathrm{e}^{-\frac{1}{2}\Phi} - \mathrm{e}^{-\frac{1}{2}\Phi^N}\big)^2 d\mu_0 \\ &\quad + \big|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}\big|^2 \int_X \mathrm{e}^{-\Phi^N} d\mu_0(u)\end{aligned}$$

$$\begin{aligned}\big|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}\big|^2 &= \frac{|Z - Z^N|^2}{Z\,Z^N\,(Z^{-\frac{1}{2}} + (Z^N)^{-\frac{1}{2}})^2} \\ &\le C\Big(\int_X (\mathrm{e}^{-\Phi} - \mathrm{e}^{-\Phi^N}) d\mu_0\Big)^2\end{aligned}$$

$$
\begin{aligned}
\Big(d_{\mathrm{Hell}}(\mu,\mu^N)\Big)^2 &= \frac{1}{2}\int_X \Big(\sqrt{\frac{d\mu}{d\mu_0}} - \sqrt{\frac{d\mu^N}{d\mu_0}}\ \Big)^2 d\mu_0 \\
&= \frac{1}{2}\int_X \Big(\frac{1}{\sqrt{Z}}\mathrm{e}^{-\frac{1}{2}\Phi} - \frac{1}{\sqrt{Z^N}}\mathrm{e}^{-\frac{1}{2}\Phi^N}\Big)^2 d\mu_0 \\
&\le \frac{1}{Z}\int_X \big(\mathrm{e}^{-\frac{1}{2}\Phi} - \mathrm{e}^{-\frac{1}{2}\Phi^N}\big)^2 d\mu_0 \\
&\quad + \big|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}\big|^2 \int_X \mathrm{e}^{-\Phi^N} d\mu_0(u)
\end{aligned}
$$

$$
\begin{aligned}
\big|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}\big|^2 &= \frac{|Z - Z^N|^2}{Z\,Z^N\,(Z^{-\frac{1}{2}} + (Z^N)^{-\frac{1}{2}})^2} \\
&\le C\Big(\int_X (\mathrm{e}^{-\Phi} - \mathrm{e}^{-\Phi^N}) d\mu_0\Big)^2
\end{aligned}
$$

Finite dimensional approximation

- $(X, \|\cdot\|)$ Hilbert space with orthonormal basis $\{\psi_j\}_{j\in\mathbb{N}}$
- $P^N : X \to \mathbb{R}^n$ orthogonal projection onto $W = \mathrm{span}\{\{\psi_j\}_{j=1}^N\}$
- $\Phi^N(u) = \Phi(P^N u)$
- Constructing μ_0 from KL expansion using $\{\psi_j\}$ we get

$$\mu_0 = \mu_0^N \otimes \mu_0^\perp$$

- Then $\mu^N = \nu^N \otimes \mu_0^\perp$ with

$$\frac{\mathrm{d}\nu^N}{\mathrm{d}\mu_0^N}(u) \propto \exp\left(-\Phi^N(u)\right), \quad u \in W$$

$$\|\mathbb{E}^\mu F(u) - \mathbb{E}^{\nu^N} F(u^N, 0)\|_S \leq C\,\psi(N), \qquad (F : X \to S)$$

Finite dimensional approximation

- $(X, \|\cdot\|)$ Hilbert space with orthonormal basis $\{\psi_j\}_{j\in\mathbb{N}}$
- $P^N : X \to \mathbb{R}^n$ orthogonal projection onto $W = \mathrm{span}\{\{\psi_j\}_{j=1}^N\}$
- $\Phi^N(u) = \Phi(P^N u)$
- Constructing μ_0 from KL expansion using $\{\psi_j\}$ we get

$$\mu_0 = \mu_0^N \otimes \mu_0^\perp$$

- Then $\mu^N = \nu^N \otimes \mu_0^\perp$ with

$$\frac{\mathrm{d}\nu^N}{\mathrm{d}\mu_0^N}(u) \propto \exp\left(-\Phi^N(u)\right), \quad u \in W$$

$$\|\mathbb{E}^\mu F(u) - \mathbb{E}^{\nu^N} F(u^N, 0)\|_S \leq C\,\psi(N), \qquad (F : X \to S)$$

Finite dimensional approximation

- $(X, \|\cdot\|)$ Hilbert space with orthonormal basis $\{\psi_j\}_{j\in\mathbb{N}}$
- $P^N : X \to \mathbb{R}^n$ orthogonal projection onto $W = \mathrm{span}\{\{\psi_j\}_{j=1}^N\}$
- $\Phi^N(u) = \Phi(P^N u)$
- Constructing μ_0 from KL expansion using $\{\psi_j\}$ we get

$$\mu_0 = \mu_0^N \otimes \mu_0^\perp$$

- Then $\mu^N = \nu^N \otimes \mu_0^\perp$ with

$$\frac{\mathrm{d}\nu^N}{\mathrm{d}\mu_0^N}(u) \propto \exp\left(-\Phi^N(u)\right), \quad u \in W$$

$$\|\mathbb{E}^\mu F(u) - \mathbb{E}^{\nu^N} F(u^N, 0)\|_S \leq C\,\psi(N), \qquad (F : X \to S)$$

Finite dimensional approximation

- $(X, \|\cdot\|)$ Hilbert space with orthonormal basis $\{\psi_j\}_{j\in\mathbb{N}}$
- $P^N : X \to \mathbb{R}^n$ orthogonal projection onto $W = \mathrm{span}\{\{\psi_j\}_{j=1}^N\}$
- $\Phi^N(u) = \Phi(P^N u)$
- Constructing μ_0 from KL expansion using $\{\psi_j\}$ we get

$$\mu_0 = \mu_0^N \otimes \mu_0^\perp$$

- Then $\mu^N = \nu^N \otimes \mu_0^\perp$ with

$$\frac{\mathrm{d}\nu^N}{\mathrm{d}\mu_0^N}(u) \propto \exp\left(-\Phi^N(u)\right), \quad u \in W$$

$$\|\mathbb{E}^\mu F(u) - \mathbb{E}^{\nu^N} F(u^N, 0)\|_S \leq C\,\psi(N), \qquad (F : X \to S)$$

F can be soln operator of the PDE (or some function of it), e.g. $u \mapsto p$ in

Inverse elliptic PDE:

$$-\nabla \cdot \big(u(x)\,\nabla p(x)\big) = f, \quad x \in \Omega \subset \mathbb{R}^d$$
$$p(x) = 0, \quad x \in \partial\Omega.$$

Find u, given $y_j = p(x_j) + \eta_j,\; j = 1, \ldots, J,\;\; \eta_j \sim \mathcal{N}(0, \gamma_j)$

$\bar{p} = \mathbb{E}^\mu p$ or $\mathbb{E}^\mu (p - \bar{p}) \otimes (p - \bar{p})$ give an indication of how uncertainty in u is affecting p

Posterior measure: updated prior (using data)

- improved estimation of our degree of info./uncertainty in input results in improved quantification of uncertainty

Outline

Posterior consistency

We look for an estimate of the underlying true $u^\dagger$ from:

$$y_j = \mathcal{G}(u^\dagger) + \eta_j, \quad j = 1, \ldots, n$$

with $y_j \in \mathbb{R}^K$,

$$\mathcal{G} : X \to \mathbb{R}^K, \quad \text{and} \quad \eta_j \sim \mathcal{N}(0, \Gamma), \quad \text{i.i.d.}$$

Given $\mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$ we have

$$\frac{\mathrm{d}\mu^{y_1,\ldots,y_n}}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left|y_j - \mathcal{G}(u)\right|_\Gamma^2\right).$$

Does μ^y concentrate on arbitrarily small neighbourhoods of $u^\dagger$ as $n \to \infty$ and how fast?

We look at convergence of *most likely functions* under the posterior (*weak consistency*)

Posterior consistency

We look for an estimate of the underlying true $u^\dagger$ from:

$$y_j = \mathcal{G}(u^\dagger) + \eta_j, \quad j = 1, \dots, n$$

with $y_j \in \mathbb{R}^K$,

$$\mathcal{G} : X \to \mathbb{R}^K, \quad \text{and} \quad \eta_j \sim \mathcal{N}(0, \Gamma), \quad \text{i.i.d.}$$

Given $\mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$ we have

$$\frac{\mathrm{d}\mu^{y_1,\dots,y_n}}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n} \left|y_j - \mathcal{G}(u)\right|_\Gamma^2\right).$$

Does μ^y concentrate on arbitrarily small neighbourhoods of $u^\dagger$ as $n \to \infty$ and how fast?

We look at convergence of *most likely functions* under the posterior (*weak consistency*)

Posterior consistency

We look for an estimate of the underlying true $u^\dagger$ from:

$$y_j = \mathcal{G}(u^\dagger) + \eta_j, \quad j = 1, \ldots, n$$

with $y_j \in \mathbb{R}^K$,

$$\mathcal{G} : X \to \mathbb{R}^K, \quad \text{and} \quad \eta_j \sim \mathcal{N}(0, \Gamma), \quad \text{i.i.d.}$$

Given $\mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$ we have

$$\frac{\mathrm{d}\mu^{y_1,\ldots,y_n}}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n} \left|y_j - \mathcal{G}(u)\right|_\Gamma^2\right).$$

Does μ^y concentrate on arbitrarily small neighbourhoods of $u^\dagger$ as $n \to \infty$ and how fast?

We look at convergence of *most likely functions* under the posterior (*weak consistency*)

MAP estimators

$$\frac{d\mu}{d\mu_0}(u) \propto \exp(-\Phi(u)), \quad \mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$$

- If $X = \mathbb{R}^d$, MAP estimators are maximisers of the density function

$$\rho_\mu = c \exp\left(-\Phi(u) - \frac{1}{2}|\mathcal{C}_0^{-1/2}u|^2\right)$$

which is maximised at minimisers of

$$I(u) := \Phi(u) + \frac{1}{2}|\mathcal{C}_0^{-1/2}u|^2$$

- When X is a function space

There is no density, modes can be define topologically

MAP estimators

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u)), \quad \mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$$

- If $X = \mathbb{R}^d$, MAP estimators are maximisers of the density function

$$\rho_\mu = c \exp\left(-\Phi(u) - \frac{1}{2}|\mathcal{C}_0^{-1/2} u|^2\right)$$

  which is maximised at minimisers of

$$I(u) := \Phi(u) + \frac{1}{2}|\mathcal{C}_0^{-1/2} u|^2$$

- When X is a function space

  There is no density, modes can be define topologically

MAP estimates – topological definition

$(X, \|\cdot\|)$ *a separable Banach space;* $\mu(X) = 1$

$B_\epsilon(u)$ ball of radius ϵ and centre u in X.

Definition (D., Law, Stuart, Voss 2013)

Any point $\tilde{u} \in X$ *satisfying*

$$\lim_{\epsilon \to 0} \frac{\sup_{u \in X} \mu(B_\epsilon(u))}{\mu(B_\epsilon(\tilde{u}))} = 1,$$

is a MAP estimator.

or equivalently $\tilde{u}$ satisfies

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\tilde{u} - h))}{\mu(B_\epsilon(\tilde{u}))} \leq 1, \quad \text{for any } h \in X$$

MAP estimates – topological definition

$(X, \|\cdot\|)$ *a separable Banach space;* $\mu(X) = 1$

$B_\epsilon(u)$ ball of radius ϵ and centre u in X.

Definition (D., Law, Stuart, Voss 2013)

Any point $\tilde{u} \in X$ *satisfying*

$$\lim_{\epsilon \to 0} \frac{\sup_{u \in X} \mu(B_\epsilon(u))}{\mu(B_\epsilon(\tilde{u}))} = 1,$$

is a MAP estimator.

or equivalently $\tilde{u}$ satisfies

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\tilde{u} - h))}{\mu(B_\epsilon(\tilde{u}))} \leq 1, \quad \text{for any } h \in X$$

$$\lim_{\epsilon\to 0} \frac{\mu(B_\epsilon(v))}{\mu(B_\epsilon(u))} \stackrel{?}{=} e^{-f(u,v)} \qquad (*)$$

- If $X = \mathbb{R}^n$

  $$f(u,v) = -I(u) + I(v) \text{ with } I(u) = -\log \rho_\mu(u), \quad \forall u \in \mathbb{R}^n$$

- For X function space
  - $\exists Z \subset X$ s.t. $f(u,v) = I(u) - I(v)$ for $u, v \in Z$
  - But $\mu(Z) = 0$

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(v))}{\mu(B_\epsilon(u))} \stackrel{?}{=} e^{-f(u,v)} \qquad (*)$$

- If $X = \mathbb{R}^n$

  $$f(u,v) = -I(u) + I(v) \text{ with } I(u) = -\log \rho_\mu(u), \quad \forall u \in \mathbb{R}^n$$

- For X function space
  - $\exists Z \subset X$ s.t. $f(u,v) = I(u) - I(v)$ for $u, v \in Z$
  - But $\mu(Z) = 0$

Theorem. (D., Law, Stuart, Voss 2013)

Let μ satisfy

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0} \propto \exp(-\Phi(u)),$$

with

$\mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$ *and* $\Phi : X \to \mathbb{R}$ *bounded below and locally Lipschitz.*

Then MAP estimators are characterised by the minimisers of

$$I(u) = \begin{cases} \Phi(u) + \frac{1}{2}\|u\|_E^2 & \textit{if } u \in Z, \\ +\infty & \textit{otherwise}, \end{cases}$$

where E is the space of admissible shifts of μ_0.

For $\Phi : X \to \mathbb{R}$ locally Lipschitz $\exists$ a minimizer $\bar{u} \in Z$ of I.

Theorem. (D., Law, Stuart, Voss 2013)

Let μ satisfy

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0} \propto \exp(-\Phi(u)),$$

with

$\mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$ *and* $\Phi : X \to \mathbb{R}$ *bounded below and locally Lipschitz.*

Then MAP estimators are characterised by the minimisers of

$$I(u) = \begin{cases} \Phi(u) + \frac{1}{2}\|u\|_E^2 & \text{if } u \in Z, \\ +\infty & \text{otherwise,} \end{cases}$$

where E is the space of admissible shifts of μ_0.

For $\Phi : X \to \mathbb{R}$ locally Lipschitz $\exists$ a minimizer $\bar{u} \in Z$ of I.

Posterier weak consistency – large data

$$y_j = \mathcal{G}(u^\dagger) + \eta_j, \quad j = 1, \dots, n$$

$$\frac{\mathrm{d}\mu^{y_1,\dots,y_n}}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left|y_j - \mathcal{G}(u)\right|_\Gamma^2\right); \quad \mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$$

MAP estimators: $u_n := \underset{u \in Z}{\operatorname{argmin}} \left\{ \sum_{j=1}^{n} |y_j - \mathcal{G}(u)|_\Gamma^2 + \frac{1}{2}\|u\|_Z^2 \right\}.$

$$u_n := \operatorname*{argmin}_{u \in Z} \left\{ \Phi(u, y) + \tfrac{1}{2} \|u\|_Z^2 \right\}.$$

Theorem.

Assume that

$G \colon X \to \mathbb{R}_+$ *is locally Lipschitz and*

$u^\dagger \in Z$.

Then

- $G(u_n) \to G(u^\dagger)$ *in probability.*
- *If* $\mathcal{G}$ *is injective* $\|u_n - u^\dagger\|_X \to 0$ *in probability.*

  Otherwise, $\exists u^* \in Z$ *and a subseq of* $\{u_n\}_{n \in \mathbb{N}}$ *such that* $\|u_n - u^\dagger\|_X \to 0$ *in probability. For any such* u^*, $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$.

If $u^\dagger \in X$, $\mathcal{G}(u_n) \to \mathcal{G}(u^\dagger)$ in probability.

$$u_n := \underset{u \in Z}{\operatorname{argmin}} \left\{ \Phi(u,y) + \frac{1}{2}\|u\|_Z^2 \right\}.$$

Theorem.

Assume that

$G\colon X \to \mathbb{R}_+$ *is locally Lipschitz and*

$u^\dagger \in Z$.

Then

- $G(u_n) \to G(u^\dagger)$ *in probability.*
- *If* $\mathcal{G}$ *is injective* $\|u_n - u^\dagger\|_X \to 0$ *in probability.*

  Otherwise, $\exists u^* \in Z$ *and a subseq of* $\{u_n\}_{n\in\mathbb{N}}$ *such that* $\|u_n - u^\dagger\|_X \to 0$ *in probability. For any such* u^*, $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$.

If $u^\dagger \in X$, $\mathcal{G}(u_n) \to \mathcal{G}(u^\dagger)$ in probability.

Outline

Markov Chain Monte Carlo methods

Approximate μ given by $\frac{d\mu}{d\mu_0}(u) \propto \mathrm{e}^{-\phi(u)}$, using a *sum of Dirac measures*:

$$\mathbb{E}_\mu f = \int_X f(u)\,\mu(\mathrm{d}u) \approx \frac{1}{N}\sum_{n=1}^{N} f(u^{(n)})$$

where $\{u^{(n)}\}_{n=1}^{N}$ is a Markov chain.

Recall: The sequence $\{u^{(n)}\}_{n=1}^{\infty}$ is called a Markov chain if

$$\Pr\left[u^{(j)} = v_j | u^{(1)} = v_1, \cdots, u^{(j-1)} = v_{j-1}\right] = \Pr\left[u^{(j)} = v_j | u^{(j-1)} = v_{j-1}\right]$$

Markov/transition Kernel $P : X \times \mathcal{B}(X) \to [0, 1]$

$$\int_B P(v, \mathrm{d}u) = P(v, B) := \Pr\left[u^{(j)} \in B | u^{(j-1)} = v\right]$$

Markov Chain Monte Carlo methods

Approximate μ given by $\dfrac{d\mu}{d\mu_0}(u) \propto \mathrm{e}^{-\phi(u)}$, using a *sum of Dirac measures*:

$$\mathbb{E}_\mu f = \int_X f(u)\,\mu(\mathrm{d}u) \approx \frac{1}{N}\sum_{n=1}^{N} f(u^{(n)})$$

where $\{u^{(n)}\}_{n=1}^N$ is a Markov chain.

Recall: The sequence $\{u^{(n)}\}_{n=1}^\infty$ is called a Markov chain if

$$\Pr\big[u^{(j)} = v_j | u^{(1)} = v_1, \cdots, u^{(j-1)} = v_{j-1}\big] = \Pr\big[u^{(j)} = v_j | u^{(j-1)} = v_{j-1}\big]$$

Markov/transition Kernel $P : X \times \mathcal{B}(X) \to [0, 1]$

$$\int_B P(v, \mathrm{d}u) = P(v, B) := \Pr\big[u^{(j)} \in B | u^{(j-1)} = v\big]$$

Markov Chain Monte Carlo methods

- construct a Markov kernel, P, that

  - can be sampled from
  - has μ as its invariant distribution

- draw iterative samples $\{u^{(n)}\}_{n=1}^{N}$ from P to approximate $\mathbb{E}_\mu f$ as

$$\int_X f(u)\,\mu(\mathrm{d}u) \approx \frac{1}{N}\sum_{n=1}^{N} f(u^{(n)})$$

- The Markov chain with kernel P is invariant wrt μ if

$$\int_X \mu(du)P(u,\cdot) = \mu(\cdot)$$

- The Markov kernel is said to satisfy detailed balance wrt μ if

$$\mu(du)P(u,dv) = \mu(dv)P(v,du)$$

  The resulting Markov chain is then said to be *reversible* wrt μ.

reversible wrt $\mu \implies$ invariant wrt μ

Metropolis-Hastings Markov chain Monte Carlo

M-H MCMC algorithm

1. Set $k = 0$ and pick $u^{(0)} \in X$.
2. Propose $v^{(k)} \sim Q(u^{(k)}, dv)$
3. Calculate acceptance probability $a_k = a(u^{(k)}, v^{(k)})$
4. Set $u^{(k+1)} = \begin{cases} v^{(k)} & \text{with prob } a_k \\ u^{(k)} & \text{otherwise} \end{cases}$
5. $k \to k+1$ and return to 2.

transition kernel for $\{u^{(k)}\}_{k \in \mathbb{N}}$:

$$P(u, A) = \int_A Q(u, dv) a(u, v) + \int_X \delta_u(A)(1 - a(u, w)) Q(u, dw)$$

$$\implies P(u, dv) = Q(u, dv) a(u, v) + \delta_u(dv) \int_X (1 - a(u, w)) Q(u, dw)$$

Metropolis-Hastings Markov chain Monte Carlo

M-H MCMC algorithm

1. Set $k = 0$ and pick $u^{(0)} \in X$.
2. Propose $v^{(k)} \sim Q(u^{(k)}, dv)$
3. Calculate acceptance probability $a_k = a(u^{(k)}, v^{(k)})$
4. Set $u^{(k+1)} = \begin{cases} v^{(k)} & \text{with prob } a_k \\ u^{(k)} & \text{otherwise} \end{cases}$
5. $k \to k + 1$ and return to 2.

transition kernel for $\{u^{(k)}\}_{k\in\mathbb{N}}$:

$$P(u, A) = \int_A Q(u, dv)a(u, v) + \int_X \delta_u(A)(1 - a(u, w))Q(u, dw)$$

$$\implies P(u, dv) = Q(u, dv)a(u, v) + \delta_u(dv) \int_X (1 - a(u, w))Q(u, dw)$$

Metropolis-Hastings Markov chain Monte Carlo

M-H MCMC algorithm

1. Set $k = 0$ and pick $u^{(0)} \in X$.
2. Propose $v^{(k)} \sim Q(u^{(k)}, dv)$
3. Calculate acceptance probability $a_k = a(u^{(k)}, v^{(k)})$
4. Set $u^{(k+1)} = \begin{cases} v^{(k)} & \text{with prob } a_k \\ u^{(k)} & \text{otherwise} \end{cases}$
5. $k \to k + 1$ and return to 2.

transition kernel for $\{u^{(k)}\}_{k\in\mathbb{N}}$:

$$P(u, A) = \int_A Q(u, dv)a(u, v) + \int_X \delta_u(A)(1 - a(u, w))Q(u, dw)$$

$$\implies P(u, dv) = Q(u, dv)a(u, v) + \delta_u(dv) \int_X (1 - a(u, w))Q(u, dw)$$

Let Q be reversible wrt μ_0, i.e.

$$\mu_0(du)Q(u,dv) = \mu_0(dv)Q(v,du)$$

Then P is reversible wrt μ if $a(u,v) = \min\left\{1, \exp\left(\Phi(u) - \Phi(v)\right)\right\}$.

Cotter, D. Stuart '12; D. Stuart '17

- $Q(u,dv) = \mu_0(dv)$ gives *Independent sampler*
- when μ_0 is $\mathcal{N}(0,\mathcal{C})$,

  setting $Q(u,\cdot) = \mathcal{N}((1-\beta^2)u, \beta^2\mathcal{C})$ gives

  preconditioned Crank-Nicolson (pCN) method

Let Q be reversible wrt μ_0, i.e.

$$\mu_0(du)Q(u,dv) = \mu_0(dv)Q(v,du)$$

Then P is reversible wrt μ if $a(u,v) = \min\left\{1, \exp\left(\Phi(u) - \Phi(v)\right)\right\}$.

Cotter, D. Stuart '12; D. Stuart '17

- $Q(u,dv) = \mu_0(dv)$ gives *Independent sampler*
- when μ_0 is $\mathcal{N}(0,\mathcal{C})$,

  setting $Q(u,\cdot) = \mathcal{N}((1-\beta^2)u, \beta^2\mathcal{C})$ gives

  preconditioned Crank-Nicolson (pCN) method

Let Q be reversible wrt μ_0, i.e.

$$\mu_0(du)Q(u,dv) = \mu_0(dv)Q(v,du)$$

Then P is reversible wrt μ if $a(u,v) = \min\left\{1, \exp\left(\Phi(u) - \Phi(v)\right)\right\}$.

Cotter, D. Stuart '12; D. Stuart '17

- $Q(u,dv) = \mu_0(dv)$ gives *Independent sampler*
- when μ_0 is $\mathcal{N}(0,\mathcal{C})$,

  setting $Q(u,\cdot) = \mathcal{N}((1-\beta^2)u, \beta^2\mathcal{C})$ gives

  preconditioned Crank-Nicolson (pCN) method

$Q(u,\cdot) = \mathcal{N}((1-\beta^2)u, \beta^2\mathcal{C})$ is reversible wrt $\mathcal{N}(0,\mathcal{C})$:

$\nu(du,dv) := \mu(du)P(u,dv)$ measure on $X \times X$;

$$\begin{aligned}
\hat{\nu}(\xi,\eta) &= \int_{X\times X} \mathrm{e}^{i\langle u,\xi\rangle + i\langle v,\eta\rangle}\mu(du)P(u,dv) \\
&= \exp\left(-\frac{1}{2}\|\mathcal{C}^{\frac{1}{2}}\eta\|^2 - \frac{1}{2}\|\mathcal{C}^{\frac{1}{2}}\xi\|^2 - (1-\beta^2)^{\frac{1}{2}}\langle \mathcal{C}^{\frac{1}{2}}\xi, \mathcal{C}^{\frac{1}{2}}\eta\rangle\right) \\
&= \hat{\nu}(\eta,\xi)
\end{aligned}$$

$Q(u,\cdot) = \mathcal{N}((1-\beta^2)u, \beta^2\mathcal{C})$ is reversible wrt $\mathcal{N}(0,\mathcal{C})$:

$\nu(du,dv) := \mu(du)P(u,dv)$ measure on $X \times X$;

$$\begin{aligned}
\hat{\nu}(\xi,\eta) &= \int_{X\times X} \mathrm{e}^{i\langle u,\xi\rangle + i\langle v,\eta\rangle} \mu(du)P(u,dv) \\
&= \exp\left(-\frac{1}{2}\|\mathcal{C}^{\frac{1}{2}}\eta\|^2 - \frac{1}{2}\|\mathcal{C}^{\frac{1}{2}}\xi\|^2 - (1-\beta^2)^{\frac{1}{2}}\langle \mathcal{C}^{\frac{1}{2}}\xi, \mathcal{C}^{\frac{1}{2}}\eta\rangle\right) \\
&= \hat{\nu}(\eta,\xi)
\end{aligned}$$

pCN algorithm

1. Set $k = 0$ and pick $u^{(0)} \in X$.
2. Propose $v^{(k)} = \sqrt{1-\beta^2}\, u^{(k)} + \beta \xi^{(k)}$, $\xi^{(k)} \sim \mu_0 = \mathcal{N}(0, \mathcal{C})$
3. Calculate acceptance probability $a_k = \min\{1, \mathrm{e}^{\Phi(u^{(k)}) - \Phi(v^{(k)})}\}$
4. Set $u^{(k+1)} = \begin{cases} v^{(k)} & \text{with prob } a_k \\ u^{(k)} & \text{otherwise} \end{cases}$
5. $k \to k+1$ and return to 2.

β parameter controlling the degree of locality of proposals

pCN algorithm

1. Set $k = 0$ and pick $u^{(0)} \in X$.
2. Propose $v^{(k)} = \sqrt{1-\beta^2}\, u^{(k)} + \beta \xi^{(k)}, \;\; \xi^{(k)} \sim \mu_0 = \mathcal{N}(0, \mathcal{C})$
3. Calculate acceptance probability $a_k = \min\{1, \mathrm{e}^{\Phi(u^{(k)}) - \Phi(v^{(k)})}\}$
4. Set $u^{(k+1)} = \begin{cases} v^{(k)} & \text{with prob } \; a_k \\ u^{(k)} & \text{otherwise} \end{cases}$
5. $k \to k+1$ and return to 2.

β parameter controlling the degree of locality of proposals

Convergence to μ

- If the density of $Q(u,\cdot)$ is positive for all u, then

$$d_{\mathrm{Hell}}(P^k(u^{(1)},\cdot),\nu^N) \to 0, \quad \text{as} \;\; k \to \infty$$

- ν^N finite-d approximation of μ; for $f : X \to E$

$$\|\mathbb{E}^{\mu} F - \mathbb{E}^{\nu^y} F\|_E \to 0 \;\; \text{as} \;\; N, k \to \infty$$

Convergence to μ

- If the density of $Q(u, \cdot)$ is positive for all u, then

$$d_{\mathrm{Hell}}(P^k(u^{(1)}, \cdot), \nu^N) \to 0, \quad \text{as} \;\; k \to \infty$$

- ν^N finite-d approximation of μ; for $f : X \to E$

$$\|\mathbb{E}^{\mu} F - \mathbb{E}^{\nu^y} F\|_E \to 0 \;\; \text{as} \;\; N, k \to \infty$$

Summary

- Wellposedness of Bayesian inverse problems for functions

  well-definedness of posterior and its stability in data

  construction of appropriate prior
- Effect of approximation of forward map on the posterior

  $\longrightarrow$ finite-d approximation of the posterior
- Weak consistency of the posterior at the level of the MAP estimators
- Metropolis-Hastings Markov chain Monte Carlo methods (pCN algorithm)

The problem for the practical session

Given

$$y_j = p(x_j) + \xi_j \quad x_j \in (0,1), \;\; j \in \{1, \ldots, J\}$$
$$\xi_j \sim \mathcal{N}(0, \gamma^2) \;\; \text{i.i.d.}$$

with p satisfying

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(u \frac{\mathrm{d}p}{\mathrm{d}x}\right) = 1, \quad x \in (0,1)$$
$$p(0) = 0, \quad p(1) = 1,$$

estimate $u \in X = \{w \in C^1(0,1) : \inf_{(0,1)} w > 0\}$.

- Let $G_j : u \mapsto p(x_j)$. Then μ satisfying

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp\Big(-\frac{1}{2\gamma^2}\sum_{j=1}^{J}|y - G_j(u)|^2\Big)$$

  is well-defined for $\mu_0(X) = 1$.

- **Prior**: μ_0 defined through

$$u = g\big(\sum_{j\in\mathbb{N}} \alpha_j\, \xi_j\, \psi_j(x)\big)$$

  for $u \sim \mu_0$ and with

  - $\{\psi_j(x)\}_{j\in\mathbb{N}} = \{1, \sqrt{2}\,\cos(2k\pi x), \sqrt{2}\,\sin(2k\pi x)\}_{k\in\mathbb{N}}$
  - $c_1 j^{-s} \leq \alpha_j \leq c_2 j^{-s}$, $s > 3/2$, (to ensure $u \in C^1(0,1)$)
  - $\xi_j \sim \mathcal{N}(0,1)$ i.i.d.
  - $g(\cdot) = c\,\exp(\cdot)$, (to impose positive lower bound on $u(x)$)

Forward solution of the PDE: Integrating with respect to x gives

$$u(x)\frac{\mathrm{d}p(x)}{\mathrm{d}x} = x + C$$

with $C \in \mathbb{R}$ constant. Integrating once more and noting that $p(0) = 0$ gives

$$p(x) = \int_0^x \frac{y + C}{u(y)}\,\mathrm{d}y$$

where by $p(1) = 1$ we have

$$C = \frac{1}{\int_0^1 \frac{1}{u(y)}\,\mathrm{d}y}\Big(1 - \int_0^1 \frac{y}{u(y)}\,\mathrm{d}y\Big)$$

References

- S. Cotter, M. Dashti, and A. Stuart, *Variational data assimilation using targetted random walks*, International Journal for Numerical Methods in Fluids, 68 (2012), pp. 403–421.
- M. Dashti and A. M. Stuart. *The Bayesian Approach to Inverse Problems*. In Handbook of Uncertainty Quantification, pages 311–428. 2017
- M.Dashti, K.J.H.Law, A.M.Stuart, and J.Voss. *MAP estimators and their consistency in Bayesian nonparametric inverse problems*. Inverse Problems, 29(9):095017, 27, 2013.
- J. Latz. *On the well-posedness of Bayesian inverse problems*. SIAM/ASA Journal on Uncertainty Quantification, 8(1):451?482, 2020.
- D. Sanz-Alonso, A. M. Stuart and A. Taeb, *Inverse Problems and Data Assimilation*, arXiv:1810.06191
- A. Stuart, *Inverse Problems: A Bayesian Perspective*, Acta Numerica, 19 (2010), pp. 451–559.