Bayesian approach to linear ill-posed inverse problems

Inverse problems

Let us consider a model

$$y = G(u),$$

where $u \in X_1$ and $G : X_1 \to X_2$. The inverse problem consists of finding u given y. Problems that may occur -

1) There maybe many solutions u corresponding to a single observation y.
2) The solution maybe very sensitive to the observation. Small error in observing y may cause large change in estimated value of u.
3) The error in observation may throw the observation y out of the range of G.

Our model

- We restrict ourselves to $X_1 = \mathcal{H}_1$, $X_2 = \mathcal{H}_2$ where $\mathcal{H}_1$ and $\mathcal{H}_2$ are separable Hilbert spaces.
- G is compact and injective. Only the second and third problem are relevant here.

Our model

- We restrict ourselves to $X_1 = \mathcal{H}_1$, $X_2 = \mathcal{H}_2$ where $\mathcal{H}_1$ and $\mathcal{H}_2$ are separable Hilbert spaces.
- G is compact and injective. Only the second and third problem are relevant here.

Our focus: Statistical, mainly Bayesian approaches to the inverse problem.

Our model

- We restrict ourselves to $X_1 = \mathcal{H}_1$, $X_2 = \mathcal{H}_2$ where $\mathcal{H}_1$ and $\mathcal{H}_2$ are separable Hilbert spaces.
- G is compact and injective. Only the second and third problem are relevant here.

Our focus: Statistical, mainly Bayesian approaches to the inverse problem.
Bayesian approach to inverse problems is still new for infinite dimensions and even some basic questions in this setup are unanswered.

Statistical version of classical inverse problem

Let us consider the case of noisy observations

$$y = G(u) + \frac{1}{\sqrt{n}}\eta,$$

where $\eta \sim N(0, \zeta)$ is the Gaussian noise and n is the parameter controlling the intensity of noise.

- The noise may throw the observation y out of the range of G almost surely. Infact, the commonly used white noise throws the observation out of $\mathcal{H}_2$ almost surely.
- y maybe seen as element of a Banach space, which is possibly an extension of $\mathcal{H}_2$.

Statistical inverse problem

Definition

Estimators of u are functions of the observation y with values in $\mathcal{H}_1$.

Statistical inverse problem

Definition

Estimators of u are functions of the observation y with values in $\mathcal{H}_1$.

Main concerns about estimators are **well-posedness** and **consistency**.

Statistical inverse problem

Definition

Estimators of u are functions of the observation y with values in $\mathcal{H}_1$.

Main concerns about estimators are **well-posedness** and **consistency**.

Setting u_0 as the **true solution**, define

- $y|u_0 \sim N(G(u_0), \frac{\zeta}{n}) \equiv \mathbb{Q}_{u_0,n}$.
- $\xi_n^{\hat{u},u_0} \equiv \|\hat{u}(y) - u_0\|_{L^2(\mathbb{Q}_{u_0,n})}$.

Definition

An estimator $\hat{u}(y)$ is said to be consistent if

$$\xi_n^{\hat{u},u_0} \to 0 \qquad \text{as } n \to \infty$$

Statistical inverse problem

Definition

Estimators of u are functions of the observation y with values in $\mathcal{H}_1$.

Main concerns about estimators are **well-posedness** and **consistency**.

Setting u_0 as the **true solution**, define

- $y|u_0 \sim N(G(u_0), \frac{\zeta}{n}) \equiv \mathbb{Q}_{u_0,n}$.
- $\xi_n^{\hat{u},u_0} \equiv \|\hat{u}(y) - u_0\|_{L^2(\mathbb{Q}_{u_0,n})}$.

Definition

An estimator $\hat{u}(y)$ is said to be consistent if

$$\xi_n^{\hat{u},u_0} \to 0 \qquad \text{as } n \to \infty$$

But at what rate?

Minimax rates

Definition

Minimax rates for the model over a set S is given by

$$\xi_n(S) \equiv \min_{\hat{u}} \max_{u_0 \in S} \xi_n^{\hat{u},u_0}$$

NOTE: Minimax rates put a bound on how quickly a posterior can approximate the true solution as noise goes to zero.

To get exact rates, we need more information concerning the operator G and the set S.

Minimax rates

Let $\{e_i, \rho_i^2\}$ be the eigenpair of $G^T G$. The ill-posedness of the model is characterised as

- when $\rho_i^2 \approx i^{-2\alpha}$, the problem is said to be mildly ill posed, e.g - Deconvolution problems.
- when $\rho_i^2 \approx \exp(-i^\beta)$, the problem is said to be severely ill posed, e.g - Heat equation.

Minimax rates

The sets on which the minimax rate is estimated are Sobolev balls in the basis $\{e_i\}$. That is,

$$\mathcal{H}^\gamma(R) \equiv \{u : \sum (i^\gamma \langle u, e_i \rangle)^2 \leq R\}$$

The minimax rates then are given as -

- Mildly ill posed problem: $\xi_n = n^{-\frac{\gamma}{1+2\alpha+2\gamma}}$
- Severely ill posed problem: $\xi_n = (\log n)^{-\frac{\gamma}{\beta}}$

Cavalier(2007)

Inverse problems in Bayesian setup: preliminaries

Bayesian Inverse problem

Recall the model:

$$y = G(u) + \frac{1}{\sqrt{n}}\eta$$

Bayesian Inverse problem

Recall the model:

$$y = G(u) + \frac{1}{\sqrt{n}}\eta$$

Introduce **Prior** - $u \sim N(0, \frac{\mathcal{C}}{R_n^2}) \equiv \mu_n$

Bayesian Inverse problem

Recall the model:

$$y = G(u) + \frac{1}{\sqrt{n}}\eta$$

Introduce **Prior** - $u \sim N(0, \frac{\mathcal{C}}{R_n^2}) \equiv \mu_n$

- The solution in the bayesian setup is given by the conditional random variable $u|y \sim \mu_n^y$.
- The prior allows us to incorporate any prior notions we might have about the behaviour of the true solution u_0.
- Functionals of posterior can serve as point estimators.

Our model

Assumption - $G(u)$ lies almost surely (w.r.t. the prior) in the Cameron-Martin space of the noise.

Our model

Assumption - $G(u)$ lies almost surely (w.r.t. the prior) in the Cameron-Martin space of the noise.
As a consequence, the likelihood $y|u \sim \mathbb{Q}_{u,n}$ *is absolutely continuous with respect to the noise measure* $\mathbb{Q}_{0,n}$ *almost surely* u.

Our model

Assumption - $G(u)$ lies almost surely (w.r.t. the prior) in the Cameron-Martin space of the noise.
As a consequence, the likelihood $y|u \sim \mathbb{Q}_{u,n}$ *is absolutely continuous with respect to the noise measure* $\mathbb{Q}_{0,n}$ *almost surely* u.

Likelihood density - Cameron-Martin theorem gives the density as

$$\exp\left(-\Phi(y,u)\right) = \exp\left(-\frac{n}{2}\left\langle G(u), G(u)\right\rangle_{\zeta} + n\left\langle y, G(u)\right\rangle_{\zeta}\right)$$

where $\langle ., . \rangle_{\zeta}$ is the Cameron-Martin norm. The expression is defined for almost all y.

Posterior measure

Posterior density - Bayes' theorem now gives the posterior density as

$$\frac{d\mu_n^y}{d\mu_n} = \frac{\exp\left(-\Phi(y,u)\right)}{\int_{\mathcal{H}_1} \exp\left(-\Phi(y,u)\right) d\mu_n}$$

- *The denominator is positive and finite for almost all y (Tonelli's theorem).*

Posterior measure

Posterior density - Bayes' theorem now gives the posterior density as

$$\frac{d\mu_n^y}{d\mu_n} = \frac{\exp\left(-\Phi(y,u)\right)}{\int_{\mathcal{H}_1} \exp\left(-\Phi(y,u)\right) d\mu_n}$$

- *The denominator is positive and finite for almost all y (Tonelli's theorem).*

Wellposedness captures the notion that the solution(posterior in this case) varies continuously with observation. This requires metrics on the relevant spaces.

Well-posedness

Definition

Given two probability measures μ and ν and a third probability measure λ such that μ and ν has densities with respect to λ, then the **Hellinger distance** is

$$d(\mu, \nu) \equiv \sqrt{\int \left(\frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right)^2 d\lambda}$$

Well-posedness

Definition

Given two probability measures μ and ν and a third probability measure λ such that μ and ν has densities with respect to λ, then the **Hellinger distance** is

$$d(\mu, \nu) \equiv \sqrt{\int \left(\frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right)^2 d\lambda}$$

Definition

Posterior for a model is said to be **wellposed** if there exists a Banach space $\{Y, \|.\|_Y\}$ such that observations $y \in Y$ almost surely and $y \to \mu_n^y$ is a continuous function from Y to the space of probability measures on $\mathcal{H}_1$.

Well-posedness

- Stuart(2010) : Well-posedness for Bayesian models on separable Banach spaces under certain sufficient technical conditions on the potential $\Phi(u, y)$ and the gaussian prior.
- Agapiou, Larsson and Stuart(2013) : The above result is used in context of our model to show its wellposedness. To satisfy the conditions, the authors have put extra conditions on the operators involved.

Well-posedness

Our first result shows that well-posedness for our model follows without any technical assumptions.

Theorem

For the model

$$y = G(u) + \frac{1}{\sqrt{n}}\eta,$$

with the terms defined as before, the posterior is well posed if $G(u)$ lies in the Cameron-Martin space of the noise measure almost surely with respect to the prior measure.

Note that the only assumption used in the theorem is also the assumption needed for the Bayesian procedure to work.

Consistency

- The posterior μ_n^y is a random measure on $\mathcal{H}_1$ with randomness coming from y.
- Assuming a true solution u_0, the distribution of y is $N(G(u_0), \frac{\zeta}{n})$.
- Posterior is a good representation of the solution only if it concentrates around the true solution in some appropriate fashion as the noise goes to 0.

Consistency

- The posterior μ_n^y is a random measure on $\mathcal{H}_1$ with randomness coming from y.
- Assuming a true solution u_0, the distribution of y is $N(G(u_0), \frac{\zeta}{n})$.
- Posterior is a good representation of the solution only if it concentrates around the true solution in some appropriate fashion as the noise goes to 0.

The random variable $X_n(\xi, y) \equiv \mu_n^y\{u : \|u - u_0\| > \xi\}$ quantifies the measure that the posterior assigns outside a ξ-ball of the true solution u_0.

Consistency

- The posterior μ_n^y is a random measure on $\mathcal{H}_1$ with randomness coming from y.
- Assuming a true solution u_0, the distribution of y is $N(G(u_0), \frac{\zeta}{n})$.
- Posterior is a good representation of the solution only if it concentrates around the true solution in some appropriate fashion as the noise goes to 0.

The random variable $X_n(\xi, y) \equiv \mu_n^y\{u : \|u - u_0\| > \xi\}$ quantifies the measure that the posterior assigns outside a ξ-ball of the true solution u_0.

Definition

The posterior is said to be **consistent** when

$$X_n(\xi, y) \to 0$$

in probability as $n \to \infty$ for all $\xi > 0$ and $u_0 \in \mathcal{H}_1$.

Contraction rates

Contraction rates quantify how quickly the posterior converges to the true solution.

Contraction rates

Contraction rates quantify how quickly the posterior converges to the true solution.

Definition

$\xi_n \to 0$ (as $n \to \infty$) is said to be a **contraction rate** for the posterior at u_0 if

$$X_n(\xi_n, y) \to 0$$

in probability as $n \to \infty$

As with minimax rates, contraction rates which are common over Sobolev balls are of interest. The discussion on contraction rates takes two main directions.

- Try to get contraction rates for a large class of priors
- Try to improve the rates by changing the parameters of the prior depending on the level of noise and even the observation y.

The parameters

- **Operator** : Let $\{e_i, \rho_i^2\}$ be the eigenpair of $G^T G$
  - Mildly ill-posed :$\rho_i^2 \approx i^{-2\alpha}$
  - severely ill posed : $\rho_i^2 \approx \exp(-i^{\beta})$
- **True solution** : Let the Sobolev space for the true solution be defined on the basis $\{\phi_i^1\}$.

$$\sum i^{2\gamma} \langle u_0, \phi_i^1 \rangle^2 < \infty$$

- **Prior** : Let the eigenpair for the covariance operator of prior be $\{\phi_i^2, \frac{\lambda_i}{R_n^2}\}$ with $\lambda_i = i^{-1-2\delta}$.

The specifics of the noise occur in conjunction with the operator, the details of which we will specify later.

The parameters

- **Operator** : Let $\{e_i, \rho_i^2\}$ be the eigenpair of $G^T G$
  - Mildly ill-posed :$\rho_i^2 \approx i^{-2\alpha}$
  - severely ill posed : $\rho_i^2 \approx \exp(-i^\beta)$
- **True solution** : Let the Sobolev space for the true solution be defined on the basis $\{\phi_i^1\}$.

$$\sum i^{2\gamma} \langle u_0, \phi_i^1 \rangle^2 < \infty$$

- **Prior** : Let the eigenpair for the covariance operator of prior be $\{\phi_i^2, \frac{\lambda_i}{R_n^2}\}$ with $\lambda_i = i^{-1-2\delta}$.

The specifics of the noise occur in conjunction with the operator, the details of which we will specify later.
Very little is known about contraction rates when the eigenbasis involved are not simultaneously diagonalizable.

Mildly ill-posed problems in the diagonal case

Assume that $\phi_i^1 = \phi_i^2 \equiv \phi_i = e_i$.
Knapik, van der Vaart, Zanten(2011)

- $$\xi_n = n^{-\frac{\gamma \wedge \delta}{1+2\alpha+2\delta}}$$

  when $R_n = 1$.

- $$\xi_n = n^{-\frac{\gamma}{1+2\alpha+2\gamma}}$$

  when $\gamma \leq 1 + 2\delta + 2\alpha$ and $R_n = n^{\frac{\gamma-\delta}{1+2\alpha+2\gamma}}$

Mildly ill-posed problems in the diagonal case

Assume that $\phi_i^1 = \phi_i^2 \equiv \phi_i = e_i$.
Knapik, van der Vaart, Zanten(2011)

- $$\xi_n = n^{-\frac{\gamma \wedge \delta}{1+2\alpha+2\delta}}$$

  when $R_n = 1$.
- $$\xi_n = n^{-\frac{\gamma}{1+2\alpha+2\gamma}}$$

  when $\gamma \leq 1 + 2\delta + 2\alpha$ and $R_n = n^{\frac{\gamma-\delta}{1+2\alpha+2\gamma}}$
- The rate is sub optimal when $\gamma > 1 + 2\delta + 2\alpha$. We note here that even when $\gamma \leq 1 + 2\delta + 2\alpha$, the scale depends on the smoothness of true solution.
- The paper also deals with credible sets and contraction rates for linear functionals of the posterior.

Mildly ill-posed problems in the diagonal case

Knapik, Szabo, van der Vaart, Zanten(2013) :

- Maximum likelihood estimator $\hat{\gamma}(y)$ is used for the parameter δ in attempt to improve the rates.
- $\hat{\gamma}(y)$ approximates γ if the true solution is regular.
- Contraction rate achieved is

$$\xi_n = n^{-\frac{\gamma}{1+2\alpha+2\gamma}}(\log n)^2(\log\log n)^{\frac{1}{2}}$$

- The paper also uses this method to get optimal contraction rates for analytic priors.

Severely ill-posed problems in the diagonal case

$\phi_i = e_i$

- Knapik, Van der vaart, Zanten(2013) : Deals with $\beta = 2$ and white noise.
- Agapiou, Stuart, Zhang(2013) : Deals with general β and coloured noise with same eigenbasis as e_i.

The contraction rates are -

- $$\xi_n = (\log n)^{-\frac{\gamma \wedge \delta}{\beta}}$$
  when $R_n = 1$.
- $$\xi_n = (\log n)^{-\frac{\gamma}{\beta}}$$
  when $(\log n)^{\frac{\gamma-\delta}{\beta}} \leq R_n \leq n^{\frac{1}{2}-\sigma}$ for some $0 < \sigma < \frac{1}{2}$.

Minimax rates are achieved using scalable priors with scales independent of smoothness of true solution.

Minimax rates in non-diagonal case

- It is enough to calculate minimax rates for linear estimators in case of additive Gaussian noise.
- Standard minimax rates are calculated when $e_i = \phi_i$ and depends on $\frac{1}{\rho_i}$ and γ. However, it can be shown that minimax rates depend on $\|(G^{-1})^T \phi_i\|$ and γ whenever $(G^{-1})^T \phi_i$ exists.
- Bayesian contraction rates shall be calculated under the same assumptions.

Contraction rates in non-diagonal case

Ray(2013) :

- A test function approach based on Ghosh *et.al.*(2000) is used to prove a general lemma about contraction rates assuming that for each i, $\langle \phi_i, e_j \rangle$ is 0 for all except finitely many j.
- The lemma applies to non-Gaussian priors as well. In effect, prior should be the distribution of random element $\sum \kappa_i \phi_i$ where κ_i are real valued random variables.

Contraction rates in non-diagonal case

Ray(2013) :

- A test function approach based on Ghosh *et.al.*(2000) is used to prove a general lemma about contraction rates assuming that for each i, $\langle \phi_i, e_j \rangle$ is 0 for all except finitely many j.
- The lemma applies to non-Gaussian priors as well. In effect, prior should be the distribution of random element $\sum \kappa_i \phi_i$ where κ_i are real valued random variables.
- The lemma reduces finding contraction rates to verifying certain technical conditions.
- In case of Gaussian priors, the author verifies the conditions only for the diagonal case for both mildly and severely-ill posed problems for non scalable priors.

Contraction rates in non-diagonal case

Ray(2013)(continued)

- The rate for mildly ill-posed problem matches that of Knapik *et.al.* (non-empirical, non scalable prior). The method gives sub optimal rates for severely ill posed problems -

$$\xi_n = (\log n)^{-\frac{\gamma \wedge (\delta - \frac{\beta}{2})}{\beta}}$$

Contraction rates for mildly-ill posed problems in non-diagonal case

Agapiou, Larsson, Stuart(2013) :

- Using discretization and tools from functional analysis, authors have found contraction rates for mildly ill-posed problems in certain class of non diagonal problems.
- The authors work with the more relaxed assumption that $G(u)$ lies almost surely in the Cameron-Martin space of the noise. The authors also allow for coloured noise.
- The operators G, $\mathcal{C}$ and ζ are related via several technical assumptions. Heuristically, they reflect the idea that the operators are equivalent to powers of each other on certain spaces. The method does not apply to severely ill-posed problems.
- The rate was not found for true solutions lying outside the Cameron-Martin space of prior that is, when $\gamma < \frac{1}{2} + \delta$.

Contraction rates for mildly-ill posed problems in non-diagonal case

Agapiou, Larsson, Stuart(continued) The class of problems for which the contraction rates were found are of the kind

- All the operators are defined on $L^2(\Omega)$
- $G = (\mathcal{C}^{-l} + \mathcal{M}_q)^{-1}$
- $\zeta = (\mathcal{C}^{-\frac{\beta}{2}} + \mathcal{M}_r)^{-2}$ where $\mathcal{C} \equiv \Delta^{-\theta}$ is the covariance operator of prior and $\mathcal{M}_q$ and $\mathcal{M}_r$ are multiplication operators with bounded, positive functions q and r which are smooth enough $(q, r \in W^{\theta,\infty}(\Omega))$.
- $2l - \beta > 0 - 1$.

In this setup, the authors show that effectively, the contraction rates are for an operator with ill-posedness of order $l - \frac{\beta}{2}$.

- Contraction rates are obtained for $\gamma \geq \theta$.
- Rates arbitrarily close to minimax rates are obtained using scalable priors (which depend on γ) upto an for the range $\theta + \kappa \geq \gamma \geq \theta$ for some κ with $\kappa < (2l - \beta + 1)\theta$.

Our work

We weaken the assumptions in the general lemma proved by Ray and further generalise it.

- We require only that $\|(G^{-1})^T \phi_i\|$ is finite for all ϕ_i. Alternatively,
$$\sum \left(\frac{\langle e_j, \phi_i\rangle}{\rho_j}\right)^2 < \infty.$$
- We assume that $G(u)$ lies in the Cameron-Martin space of the noise almost surely with respect to the prior. We also allow for coloured noise.
- We apply the lemma to non-diagonal problems and get contraction rates for all γ. Our class of examples is strictly larger than the one used by Agapiou *et.al.*.

Our work

Consider the following models

-
  - Let the noise be white and the eigenpairs of $G^T G$ and $\mathcal{C}$ be $\{e_i, \rho_i^2\}$ and $\{\phi_i, i^{-1-2\delta}\}$ respectively.
  - $\langle \phi_i, e_j \rangle = 0$ for all i, j such that $j \notin [k_1 i, k_2 i]$ for some $k_1, k_2 > 0$
  - $\rho_i \approx i^{-\alpha}$
-
  - $G = (\mathcal{C}^{-l} + \mathcal{M}_q)^{-1}$
  - $\zeta = (\mathcal{C}^{-\frac{\beta}{2}} + \mathcal{M}_r)^{-2}$ where $\mathcal{C} \equiv \Delta^{-\theta}$ is the covariance operator of prior and $\mathcal{M}_q$ and $\mathcal{M}_r$ are multiplication operators with bounded, positive functions q and r such that $q \in W^{\theta((l-\frac{\beta}{2})\wedge 0),\infty}$.
  - $2l - \beta > 0 - 1$
  - Put $2\theta = 1 + 2\delta$ and $\theta(l - \frac{\beta}{2}) = \alpha$

Theorem

The contraction rates ξ_n for the above models are given by the following expressions.

$$\xi_n = \begin{cases} n^{-\frac{\gamma}{1+2\alpha+2\gamma}} & 2\gamma \leq 1+2\delta, R_n = n^{\frac{\gamma-\delta}{1+2\alpha+2\gamma}} \\ n^{-\frac{2\delta+1}{4(1+\alpha+\delta)}} & 2\gamma > 1+2\delta, R_n = \frac{1}{4(1+\alpha+\delta)} \end{cases}$$

For severely ill-posed problems defined in similar fashion, we get the minimax rates $\xi_n = (\log n)^{-\frac{\gamma}{\beta}}$ for scalable priors with scales which are independent of the smoothness of true solution.

Future problems

- To extend the Bayesian empirical method to non-diagonal setting in the mildly ill-posed case.
- To get minimax rates for true solutions with smoothness defined on arbitrary bases.
- To get contraction rates for true solutions with smoothness defined on arbitrary bases. In particular, find the orientation of the prior for which the best rates are achieved.
    - We can get some contraction rates using the fact that ϕ_i allowed as eigenbases are dense on the unit ball but we dont know how they relate to optimal rates.
- Find an empirical method to estimate the basis of the prior for which we get optimal rates.