Reading Group UQ: A.M. Stuart (2010): *Inverse problems: a Bayesian perspective* : Chapters 4-5

Jonas Latz, M.Sc.

Technical University of Munich
Department of Mathematics
Chair of Numerical Mathematics (M2)

27.01.2017

Outline

- Chapter 4: Common structure
  - Ill- and wellposed problems
- Chapter 5: Algorithms
  - Importance Sampling
  - Markov Chain Monte Carlo
  - Filtering

A well-posed problem

Let $f : A \times B \mapsto T \ni \underline{0}$ and $\bar{y} \in B$. We refer to A by *parameter space*, and to B by *data space*.

$$\text{Find } u \in A : f(u, \bar{y}) = \underline{0}. \qquad \text{(Problem)}$$

(Problem) is well-posed[1], if the following statements are true:

- Existence: $\exists u \in A : f(u, \bar{y}) = \underline{0}$
- Uniqueness: $\#\{u \in A : f(u, \bar{y}) = \underline{0}\} = 1$
- Continuity wrt the data: The function $B \supseteq B_{(\bar{y},\varepsilon)} \ni y \mapsto u_y \in A$, such that $f(u_y, y) = \underline{0}$, is continuous.

[1] J. Hadamard (1902): *Sur les problemes aux derives partielles et leur signi cation physique.* Princeton University Bulletin, 13:(49-52)

Inverse Problems

Let $A := X, B := T := Y$. We defined an Inverse Problem by

$$\mathcal{G}(u) - y + \eta = 0, \tag{IP}$$

where η is a realisation of $(Y, \mathcal{B}Y, \mathrm{N}(0, \Gamma))$ and $\mathcal{G}$ fulfils previously discussed Assumptions[2].

[2] Assumptions 2.6 in Stuart(2010)

Inverse Problems

Let $A := X, B := T := Y$. We defined an Inverse Problem by

$$\mathcal{G}(u) - y + \eta = 0, \quad \text{(IP)}$$

where η is a realisation of $(Y, \mathcal{B}Y, \mathrm{N}(0, \Gamma))$ and $\mathcal{G}$ fulfils previously discussed Assumptions[2].

Check conditions:

- Existence: Noise can lead to $y \notin \mathrm{Img}(\mathcal{G})$
- Uniqueness: Typically, $\dim A \gg \dim B$, so the problem is underdetermined

[2] Assumptions 2.6 in Stuart(2010)

(Regularised) least-squares approach to IP

Define, $f : X \times Y \to \mathbb{R}$,

$$(u, y) \mapsto \|\Gamma^{-\frac{1}{2}}(\mathcal{G}(u) - y)\|_A^2 + \lambda R(u),$$

Assume, that f is well-defined and $0 \in \mathrm{Img}(f)$.

(Regularised) least-squares approach to IP

Define, $f : X \times Y \to \mathbb{R}$,

$$(u, y) \mapsto \|\Gamma^{-\frac{1}{2}}(\mathcal{G}(u) - y)\|_A^2 + \lambda R(u),$$

Assume, that f is well-defined and $0 \in \mathrm{Img}(f)$.

- Existence is given, if $R : X \to \mathbb{R}$ is a suitable regularisation and
- Uniqueness: Not necessarily.
- Continuity: Not necessarily.

Bayesian approach to IP

Reminder: Let $\mu_0 \in \mathrm{P}(X, \mathcal{F})$ be a prior distribution and $u \sim \mu_0$.

$$\text{Find } \mu^y = \mathbb{P}(u \in \cdot | \mathcal{G}(u) + \eta = y) \in \mathrm{P}(X, \mathcal{F}). \tag{BIP}$$

Bayesian approach to IP

Reminder: Let $\mu_0 \in \mathrm{P}(X, \mathcal{F})$ be a prior distribution and $u \sim \mu_0$.

$$\text{Find } \mu^y = \mathbb{P}(u \in \cdot | \mathcal{G}(u) + \eta = y) \in \mathrm{P}(X, \mathcal{F}). \qquad \text{(BIP)}$$

This can be rewritten in terms of (Problem) resp. A, B and f.

$$\begin{aligned} A &:= \{K : \mathcal{F} \times Y \mapsto [0,1] : K \text{ is a Markov kernel}\}, \\ B &:= Y, \\ T &:= \mathbb{R}^{\mathcal{B}Y \times \mathcal{F}}, \\ f &: A \times B \to T, \\ (K, y) &\mapsto [\mathbb{P}(\{\mathcal{G}(u) + \eta \in E\} \cap \{u \in F\}) \\ &- \int_E K(F|y)\mathbb{P}(\mathcal{G}(u) + \eta \in \mathrm{d}y) : E \in \mathcal{B}Y, F \in \mathcal{F}]. \end{aligned}$$

Bayesian approach to IP

- Existence: Yes.[3]
- Uniqueness: Yes, by the Theorem of Radon-Nikodym.
- Continuity: Lipschitz-continuous wrt. to the data in the Hellinger distance.[4]

[3] Theorem 4.1 in Stuart (2010)
[4] Theorem 4.2 ibid.

Deriving a posterior distribution in Bayesian statistics

Let μ_0 be the prior distribution, $L(y|u)$ be the likelihood, Z_y be the normalising constant and μ^y be the posterior.

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0} = \frac{L(y|u)}{Z_y} \qquad \text{(Bayes' Theorem)}$$

- Analytically: Only feasible, if prior + likelihood are a *conjugate pair*.
- Computationally: Importance Sampling, Markov Chain Monte Carlo, Sequential Monte Carlo

Importance Sampling

Let $Q : X \to \mathbb{R}$ be a quantity of interest. Consider the following identity

$$\mathbb{E}_{\mu^y}[Q] = \int_X Q(u)\mathrm{d}\mu^y(u) = \int_X Q(u)\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u)\mathrm{d}\mu_0(u)$$
$$= \int_X Q(u)\frac{L(y|u)}{Z_y}\mathrm{d}\mu_0(u) = \frac{1}{Z_y}\mathbb{E}_{\mu_0}[Q \cdot L(y|\cdot)].$$

Using Bayes' Theorem, we can derive the expected value of Q with respect to the posterior μ^y based on an expected value wrt. to the prior μ_0.

Importance Sampling

Given $(u^{(j)} : 1 \leq j \leq J) \sim \mu_0^{\otimes J}$, the Importance Sampling Estimator is given by:

$$\widehat{\mathbb{E}}_{\mu^y}[Q] = \frac{1}{\sum_{j=1}^J L(y|u^{(j)})} \sum_{j=1}^{J} Q(u^{(j)}) L(y|u^{(j)}).$$

Importance Sampling

Advantages

- Good convergence rate
- independent samples can be used to approximate the full posterior distribution

Importance Sampling

Advantages

- Good convergence rate
- independent samples can be used to approximate the full posterior distribution

Disadvantages

- Requires the estimation of the normalising constant Z_y
- Requires a high amount of samples
- Variance of the estimator can be infinite
- biased

Markov Chain Monte Carlo

Sampling independently from the posterior distribution μ^y is often infeasible. However, it is possible, to sample from a Markov Chain $(u^{(j)} : j \geq 1)$, which is stationary with respect to μ^y.

[5] J. Liu(2004): *Monte Carlo Strategies for Scientific Computing*, Springer or Stuart(2010)

Markov Chain Monte Carlo

Sampling independently from the posterior distribution μ^y is often infeasible. However, it is possible, to sample from a Markov Chain $(u^{(j)} : j \geq 1)$, which is stationary with respect to μ^y.

Theorem (Erdogic Markov Chain)

Given a Markov Chain $(u^{(j)} : j \geq 0)$ *that is stationary wrt.* μ^y *and that fulfils further assumptions*[5]

$$\frac{1}{J}\sum_{j=1}^{J} Q(u^{(j)}) \to \mathbb{E}_{\mu^y}[Q] \quad (J \to \infty, \mathbb{P}\text{-a.s.}) \tag{1}$$

[5] J. Liu(2004): *Monte Carlo Strategies for Scientific Computing*, Springer or Stuart(2010)

Metropolis-Hastings MCMC

The (probably) most common Markov Chain Monte Carlo method is Metropolis-Hastings MCMC. First $X := \mathbb{R}^k$, $\mathcal{F} := \mathcal{B}\mathbb{R}^k$ and μ^y is given by Lebesgue-density $f(u|y)$.
Let $V(\cdot|\cdot)$ be a Markov Kernel on $\mathcal{F} \times X$, that is absolutely continues, with Lebesgue-density $v(\cdot|\cdot)$. Furthermore, let $j = 1$ and $u^{(0)} \in X$ be the Markov Chains starting point.

1. Sample $u^* \sim v(\cdot|u^{(j-1)})$
2. $a := \min\{1, \frac{f(u^*|y)v(u^{(j-1)}|u^*)}{f(u^{(j-1)}|y)v(u^*|u^{(j-1)})}\}$
3. With probability a set $u^{(j)} := u^*$, otherwise set $u^{(j)} := u^{(j-1)}$
4. If $j \leq$ Max_steps go to [1].

Typical choices of the proposal Markov Kernel V

- Independence Sampler : V does not depend on the prior state u.
- Random Walk Metropolis: $V(\cdot|u)$ is symmetric at u.
- Preconditioned Crank Nicholson MCMC:
  $V(\cdot|u) = \mathrm{N}(\sqrt{1-\beta^2}u, \beta^2 C)$ (given a Gaussian prior with covariance operator C)

Importance Sampling

Advantages

- Does not require the estimation of the normalising constant
- unbiased

Importance Sampling

Advantages

- Does not require the estimation of the normalising constant
- unbiased

Disadvantages

- dependent samples, therefore worse convergence rate
- Problematic to sample from multimodal distributions.
- Convergence of the Markov Chain is hard to assess.

Filtering

Consider a time-dependent sequence of Inverse Problems:

$$(\mathcal{G}_t(u) + \eta_t = y_t : t \in \mathbb{N}).$$

Define $\mathbf{G}_t := (\mathcal{G}_1 + \eta_1, ..., \mathcal{G}_t + \eta_t), \mathbf{Y}_t := (y_1, ..., y_t)$ and the prior distribution μ_0. Furthermore, define

$$\mu_t := \mathbb{P}(u \in \cdot | \mathbf{G}_t = \mathbf{Y}_t) \ \ (t \in \mathbb{N})$$

Filter: $F : Y \times \mathrm{P}(X, \mathcal{F}) \to \mathrm{P}(X, \mathcal{F})$,

$$(y_{t+1}, \mu_t) \mapsto \mu_{t+1}$$

Helpful (?!) References

- Dashti, Stuart (2014) - The Bayesian Approach to Inverse Problems (Lecture Notes, `https://arxiv.org/pdf/1302.6989.pdf`)
- Short Review and Example code in `https://github.com/latz-io/bayesian_inversion`
- Liu (2004) - Monte Carlo Strategies for Scientific Computing