Stochastic Simulation

Autumn Semester 2024

Prof. Fabio Nobile *Assistant: Matteo Raviola*

Project - 4

Submission deadline: 16 January 2025

Bayesian inverse problems in large dimensions

1 Introduction and background

Consider the problem of finding a set of parameters $\xi \in \mathbb{R}^P$ from some measured data $y \in \mathbb{R}^J$, such that the following relation holds:

$$y = G(\xi) + \eta. \tag{1}$$

In the previous equation, G is a given "observation operator" (i.e a smooth, non-necessarily linear, map from ξ to y), and η is some additive random noise polluting the observations. For instance, G may involve some differential equation and return the solution observed at certain locations in space and time, whereas ξ may represent some parameter in the equation we want to infer from observed data. For simplicity we will assume that the noise is Gaussian of the form $\eta \sim \mathcal{N}(0, C_\eta)$. Adopting a Bayesian perspective, we treat ξ as random variables. Using Bayes theorem, the probability distribution of ξ given y, $\pi(\xi|y)$ is given by

$$\pi(\xi) = \pi(\xi|y) = \frac{\pi(y|\xi)\pi_0(\xi)}{\pi(y)}. \tag{2}$$

In the previous equation, $\pi(\xi)$ is called the *posterior distribution*, $\pi(y|\xi)$ is called the *likelihood*, $\pi_0(\xi)$ is called the *prior* and $\pi(y)$ is called the evidence. In the particular case of model (1) with η an additive Gaussian noise, we can write the likelihood as:

$$\pi(y|\xi) = \mathcal{N}(G(\xi), C_\eta). \tag{3}$$

The Bayesian approach allows us to include any prior information on ξ into the prior distribution $\pi_0(\xi)$. Since the evidence $\pi(y)$ is usually not known and the posterior $\pi(\xi|y)$ may have a complex form, we resort to Markov Chain Monte Carlo methods, which do not need knowledge of the normalizing constant $\pi(y)$.

1.1 Bayesian inference for log permeability

In this project, we consider the following model

$$\frac{d}{dx}\left(e^{u(x,\xi)}\frac{d}{dx}p(x;\xi)\right)=0, \quad p(0;\xi)=0, \quad p(1;\xi)=2, \quad x\in[0,1], \tag{4}$$

$$u(x,\xi)=\frac{\sqrt{2}}{\pi}\sum_{k=1}^{P}\xi_k\sin(k\pi x), \tag{5}$$

describing the pressure (hydraulic head) distribution in a 1D underground aquifer with spatially varying permeability field $k(x)=e^{u(x,\xi)}$. The log-permeability $u(x,\xi)$ is parametrized by the set of coefficients $\xi=(\xi_1,\dots,\xi_P)\in\mathbb{R}^P$, which we want to infer from available noisy measurements of the pressure in 4 locations along the aquifer:

$$y_j=p(0.2j;\xi)+\eta_j, \quad j=1,\dots,4, \quad \eta_j\sim N(0,\sigma^2) \quad i.i.d, \tag{6}$$

in a Bayesian perspective. The model (5) can be thought of as a sine-series expansion of the log-permeability, thus recasting the inference problem on the Fourier coefficients $\xi=(\xi_1,\dots,\xi_P)$. For computational purposes, the series needs to be truncated at the P-th term. The goal of this project is to understand the influence of the truncation level P in the performance of MCMC algorithms to explore the posterior distribution, particularly when $P\to\infty$. The likelihood of the data is given by

$$\pi(y|\xi)=\frac{1}{(\sqrt{2\pi}\sigma)^4}\exp\left(-\frac{|y-G(\xi)|^2}{2\sigma^2}\right), \tag{7}$$

where $|\cdot|$ is the Euclidean norm, $G(\xi)=(p(0.2;\xi),\dots,p(0.8;\xi))$ and we consider as prior distribution a multivariate normal:

$$\pi_0=\mathcal{N}(0,C),\ C:=\mathrm{diag}\{k^{-2},\ k\in\mathbb{N}\},\ \text{ i.e, } \pi_0(\xi_1,\dots,\xi_P)=\frac{1}{\prod_{k=1}^{P}\sqrt{2\pi k^{-2}}}\exp\left(-\sum_{k=1}^{P}\frac{\xi_k^2k^2}{2}\right). \tag{8}$$

The data y for this project is given in Table 1 and the noise level and distribution is known and given by $\eta\sim\mathcal{N}(0,\sigma^2I)$, with $\sigma=0.04$.

x	0.2	0.4	0.6	0.8
$y(x;\xi)$	0.5041	0.8505	1.2257	1.4113

Table 1: Measured data.

In this project we focus on Metropolis-Hastings algorithms and will consider and compare different proposals, studying in particular their performance when P increases. We will use the effective sample size (ESS) as a metric for comparison. Recall that for a scalar function $\xi\mapsto f(\xi)\in\mathbb{R}$ and a stationary process $\{\xi_n\}_{n\in\mathbb{Z}}$, the effective sample size is given by

$$\mathrm{ESS}=\mathrm{ESS}(N,f,\{\xi_n\}_{n=0}^{N})=N\left[1+2\sum_{n>0}\mathrm{Corr}(f(\xi_0),f(\xi_n))\right]^{-1}, \tag{9}$$

where N is the length of the generated Markov chain.

Remarks on implementation

Equation (4) has a closed form solution given by

$$p(x) = 2\frac{S_x(e^{-u})}{S_1(e^{-u})}, \quad \text{where } S_x(f) = \int_0^x f(y)dy. \tag{10}$$

To evaluate it, discretize the interval $[0, 1]$ into M sub-intervals of size h. The numerical integration can be done using the composite trapezoidal rule (you can use for instance Scipy's `scipy.integrate.trapz` or `scipy.integrate.cumtrapz`). For a given chosen spatial resolution it makes sense to truncate the series expansion in (5) to $P = M/2$. This implies that the higher the spatial resolution, the higher will be the size of the vector ξ, i.e, the dimension of the state space.

2 Goals of the project

(a) **Random walk Metropolis.** Implement the random walk Metropolis (RWM) algorithm for the problem discretized in the previous section using a proposal of the form $Q(\xi, \cdot) = \mathcal{N}\left(\xi, s^2 C\right)$ for different values of $s < 1$, where C is the same covariance matrix as in (8). Run a chain of length $N = 10^4$ for different values of M, and $P = M/2$. Evaluate your results in terms of mixing of the chains and ESS when taking $f = \xi_1$ in Eq. (9). Include traceplots of f and plots of the autocorrelation for each value of P. Explain your results.

(b) **Improving on RWM: preconditioned Crank-Nicholson (pCN).** An improvement over the standard random walk metropolis algorithm in large dimensions is the preconditioned Crank-Nicholson (pCN). In this case, the proposal distribution is

$$Q(\xi, \cdot) = \mathcal{N}\left(\sqrt{1 - s^2}\xi, s^2 C\right), \tag{11}$$

for some $s < 1$. Implement a Metropolis-Hastings MCMC algorithm using this proposal. Repeat the experiments in the previous point and compare the performances of pCN and RWM.

(c) **On some theoretical understanding.** Prove the following points:

(c1) Let p_M denote the numerical approximation of (10) on a grid $\{x_k = k/M, \ k = 0, \ldots, M\}$ and $\xi \in \mathbb{R}^P$. Deduce that the likelihood $\pi_M(y|\xi)$ is bounded from above and from below away from zero, uniformly with respect to M and ξ.

(c2) Show that the proposal $Q(\xi, \cdot)$ in Eq. (11) is in detailed balance with the prior π_0. Exploit this fact to deduce that the Metropolis-Hastings acceptance rate for the pCN algorithm can be written as

$$\alpha(\xi, \eta) = \min\left\{1, \frac{\pi_M(y|\eta)}{\pi_M(y|\xi)}\right\}$$

and there exists $\beta > 0$ independent of M such that

$$\alpha(\xi, \eta) \geq \beta \quad \forall \xi, \eta \in \mathbb{R}^P.$$

(c3) (Optional) Consider the Random Walk Metropolis algorithm and the total acceptance rate when being at a given state $\xi \in \mathbb{R}^P$

$$\alpha^*(\xi) = \int_{\mathbb{R}^P} \alpha(\xi, \eta) Q(\xi, d\eta)$$

with $Q(\xi, \cdot) = \mathcal{N}(\xi, s^2 C)$. You can try to prove that

$$\int_{\mathbb{R}^P} \alpha^*(\xi) \pi_0(\xi) d\xi \to 0 \quad \text{as } P \to \infty.$$

(*Hints*: for $\xi = (\xi_1, \ldots, \xi_P)$ and $\eta = (\eta_1, \ldots, \eta_P)$, split the integration domain into $\Omega = \{\sum_k \xi_k^2 k^2 > \sum_k \eta_k^2 k^2\}$ and its complement. Perform then a change of variables $x_k = k(\xi_k - \eta_k)$, $y_k = k\xi_k$. Exploit the fact that x_k, y_k, thought as random variables, are all iid standard normals.)

(d) **Laplace's approximation.** Another idea to construct a good proposal for MCMC is to use Laplace's approximation, i.e, to set the proposal distribution Q to be a normal $\mathcal{N}(\xi_{\text{map}}, H^{-1})$, where ξ_{map} is the maximum a posteriori point, i.e,

$$\xi_{\text{map}} = \arg\min_{\xi \in \mathbb{R}^P} \left(-\log \pi(\xi)\right)$$

and H^{-1} is the inverse of the Hessian H of $-\log \pi(\xi)$ evaluated at ξ_{map}. An approximation of these quantities can be obtained by standard `Python` optimization libraries (see BFGS on the Scipy optimization reference). Notice that the BFGS algorithm provides a low-rank approximation $\tilde{H}$ of the Hessian H. In practice, we set $Q = \mathcal{N}(\xi_{\text{map}}, (\tilde{H} + \alpha^2 I)^{-1})$, for some $\alpha > 0$. Once such a proposal is constructed, an independent sampler Metropolis algorithm can be implemented. Run the same experiments as in points (a) and (b) and compare your results in terms of ESS vs dimensionality. Explain your results.

(e) Use any of the previously discussed MCMC methods to estimate the posterior expectation $\mathbb{E}_\pi[q]$ of the following scalar quantity of interest $q(\xi) = \int_0^1 e^{u(x,\xi)} dx$. Monitor the convergence of the Markov chain and estimate the error of your MCMC estimator.

(f) Suppose now many more pressure observations are available

$$y_j = p\left(\frac{j}{40}, \xi\right) + \eta_j \quad j = 1, \ldots, 39.$$

You can generate the observations yourself by taking one realization of ξ from the prior π_0 and one realization of the noise $\eta_j \sim \mathcal{N}(0, \sigma^2 I_{39\times 39})$, with $\sigma = 0.04$. How do the performances of the previous algorithms change in this case? Comment the results you obtain.

References

1. Cotter, Simon L., et al. "MCMC methods for functions: modifying old algorithms to make them faster." Statistical Science (2013): 424-446.