Ensemble sampler for infinite-dimensional inverse problems

Jeremie Coullon[1] and Robert J. Webber[2]

[1]Mathematics and Statistics, Lancaster University
[2]New York University

AABI 2021

Abstract

We introduce a new Markov chain Monte Carlo sampler for infinite-dimensional Bayesian inverse problems. The new sampler is based on the affine invariant ensemble sampler, which we extend for the first time to function spaces. The new sampler is more efficient than preconditioned Crank-Nicolson, yet it requires no gradient information or posterior covariance information, making the sampler broadly applicable.

1 Introduction

In Bayesian inverse problems, Markov chain Monte Carlo (MCMC) methods are needed to approximate distributions on infinite-dimensional function spaces. Yet designing an efficient and broadly applicable MCMC method for this context has proved challenging.

The first MCMC sampler for function spaces was preconditioned Crank-Nicolson (pCN, [2]). This early approach can be used for a wide variety of Bayesian inverse problems, yet pCN is inefficient, requiring a very large number of iterations to estimate statistics of the posterior distribution [3]. In contrast, newly developed MCMC methods sample the posterior distribution more efficiently [3, 4, 1, 8, 7], yet the new samplers require gradient or posterior covariance information which is challenging to obtain for many inverse problems.

This raises the question: what function space sampler is more efficient than pCN without requiring gradient or posterior covariance information? A gradient-free sampler is needed when the forward model has a legacy code base and obtaining derivatives is difficult or impossible. A covariance-free sampler is needed when estimating the posterior covariance matrix is a challenging procedure that would use up excessive computational resources.

In the literature on finite-dimensional MCMC, there is a gradient-free sampler that quickly "learns" the posterior covariance structure through an ensemble of interacting walkers. This sampler, called the affine invariant ensemble sampler (AIES, [6]), is easy to tune, easy to parallelize, and efficient at sampling spaces of moderate dimensionality ($d \leq 20$). AIES is used extensively due to its implementation in the well-known emcee package for python [5].

The main contribution of this work is to extend the AIES to the setting of infinite-dimensional inverse problems with a Gaussian prior distribution. We first identify the

Karhunen–Loève (KL) expansion for the prior distribution. Then, we alternate between AIES and pCN sampling steps. We use AIES to sample the posterior distribution on the first few KL coefficients, while we use pCN to sample the posterior distribution on the higher KL coefficients. By combining AIES with pCN, we obtain an ensemble sampler that is stable, efficient, and broadly applicable across a large class of Bayesian inverse problems.

2 Background

In this section we briefly review the pCN and AIES samplers.

Firstly, pCN is a dimension-independent random walk sampler and has the following proposal: starting from a position U, the pCN update takes the form

$$\tilde{U} = \sqrt{1-\omega^2}U + \omega\xi, \tag{1}$$

where $\xi \sim \mathcal{N}(0, C)$ is a random draw from the Gaussian prior distribution and $\omega \in (0, 1]$ is the step size parameter. This proposal is independent of dimension and has acceptance probability

$$\min\left\{1, \exp\left(\phi(\tilde{U}) - \phi(U)\right)\right\}, \tag{2}$$

In contrast, AIES is a finite-dimensional sampler is defined by an ensemble of walkers $\overrightarrow{X} = (X_1, ...X_L)$ to sample from a product density $\pi(x_1)\cdots\pi(x_L)$ with each walker taking values in $\mathbb{R}^M$. The proposal for walker X_i to the new position is:

$$\tilde{X}_i = X_i + (1-Z)(X_j - X_i), \tag{3}$$

where $Z \in [1/a, a]$ is a random number with density $g(Z) \propto 1/\sqrt{z}$,.

This proposal has acceptance probability

$$\min\left\{1, Z^{M-1}\frac{\pi(\tilde{X}_i)}{\pi(X_i)}\right\}. \tag{4}$$

3 New ensemble sampler

Under the posterior distribution $\pi(du) \propto \exp(\phi(u))\pi_0(du)$, the KL coordinates have an unknown distribution, which must be approximated through sampling. However, the prior distribution restricts the posterior distribution in substantial ways. The higher KL coefficients (ie: the coefficients of the high frequency basis elements) in particular, are already highly constrained under the the prior distribution, so the prior and posterior distributions must be similar. As a consequence, the pCN approach, which is designed for efficient sampling from the prior distribution, works well for sampling the posterior distribution on the higher KL coordinates.

The most difficult part of sampling the Bayesian posterior distribution is sampling the first few KL coefficients. These coordinates are barely constrained under the prior distribution, so the posterior distribution can take an arbitrary shape, which may be poorly scaled or multimodal. As a consequence, the pCN sampler may work very slowly on these coordinates, in which case an improved sampling strategy is required.

To address the sampling bottlenecks in Bayesian inverse problems, we propose a Metropolis-within-Gibbs sampler that alternates between using pCN on the higher KL coefficients and using the AIES sampler on the first few KL coefficients.

The proposed function space ensemble sampler robustly improves the sampling performance of pCN, as we show through two examples. Remarkably, this is the first known sampler that improves on the performance of pCN without requiring gradients and without requiring posterior covariance information.

Algorithm 1: Functional ensemble sampler (FES).

To sample a Bayesian posterior distribution $\pi(du) \propto \exp(\phi(u))\,\pi_0(du)$ with a Gaussian prior $\pi(0, C)$, we perform the following steps:

1. Identify the matrix J whose columns are the first M eigenvectors of the covariance operator. Set $P = JJ^T$ and $Q = I - JJ^T$.
2. Initialise an ensemble of walkers $\overrightarrow{X} = (X_1, ...X_L)$.
3. Repeat the following sampling steps:
   (a) For each walker X_i, propose the update
   $$\tilde{X}_i = PX_i + Q\left(\sqrt{1-\omega^2}X_i + \omega\xi\right), \tag{5}$$
   where $\xi \sim \mathcal{N}(0, C)$. Accept the proposal with probability
   $$\min\left\{1, \exp\left(\phi\left(\tilde{X}_i\right) - \phi\left(X_i\right)\right)\right\}. \tag{6}$$
   (b) For each walker X_i, randomly choose a different walker $X_j \neq X_i$. Propose the update
   $$\tilde{X}_i = X_i + (1-Z)\,P\,(X_j - X_i)\,, \tag{7}$$
   where $Z \in [1/a, a]$ has density $g(z) \propto 1/\sqrt{z}$. Accept with probability
   $$\min\left\{1, Z^{M-1}\frac{\pi\left(\tilde{X}_i\right)}{\pi\left(X_i\right)}\right\}. \tag{8}$$

4 Numerical experiments

In this section we will consider two examples to illustrate the performance of the FES sampler on problems that have both scalar and functional parameters.

The first example - the advection equation - shows how the FES sampler can increase mixing speed by two orders of magnitude over pCN. We also illustrate the effect of varying the main tuning parameter M.

The second example - path reconstruction for langevin dynamics - shows the robustness of FES to mild multimodality and model misfit. We compare FES to pCN as well as to another gradient-free sampler (which is main competitor of our method) and find that FES performs better.

For both examples we fix the stretch move step size to $a = 2$, as recommended in [5] and tune the pCN proposal ω to give an acceptance rate of 20%.

4.1 Advection equation

Our first example involves parameter estimation for the advection equation, which is linear hyperbolic partial differential equation (PDE). This simple model is the special case of more complicated nonlinear hyperbolic PDEs arising in fluid dynamics, acoustics, motorway traffic flow, and many other applications.

$$\begin{cases} \frac{\partial \rho}{\partial t} + c\frac{\partial \rho}{\partial x} = 0, & t > 0. \\ \rho = \rho_0, & t = 0. \end{cases} \tag{9}$$

We simultaneously estimate the advection speed c and the initial condition ρ_0 from a set of noisy observations.

In this example we set a uniform prior on the wave speed $\pi_0(c) \sim \mathcal{U}(0, 1.4)$, and set a Gaussian process prior on the initial condition on $[0, 10]$. We use a squared exponential kernel, with $l = 1$ and $\sigma_0 = 130$:

$$k(x, x') = \sigma_0 \exp\{-\frac{1}{2l}(x - x')^2\}$$

We discretise the initial condition using 200 points. We generate observations we set $c_{true} = 0.5$ and use a sample from the GP prior as initial condition. We consider 9 equally spaced observation in the x-t plane and add Gaussian observational noise with standard deviation $\sigma_{obs} = 0.1$.

We run FES with $L = 100$ walkers and $M = 0, 1, 5, 10, 20$ for $20, 20, 10, 2, 2$ million iterations respectively. The sampler uses Metropolis-with-Gibbs updates as in algorithm (1); namely it alternates proposing the finite space and the complementary subspace.

We also run a "vanilla" sampler that proposes a joint update for the wave speed and initial condition. The proposals are a Gaussian random walk for the wave speed and pCN for the initial condition.

We show the ACF plots in figure (1): these illustrate how FES can show a two orders of magnitude speedup compared to pCN. We also see how $M = 10$ is optimal, and that choosing M too large causes mixing to slow down again.

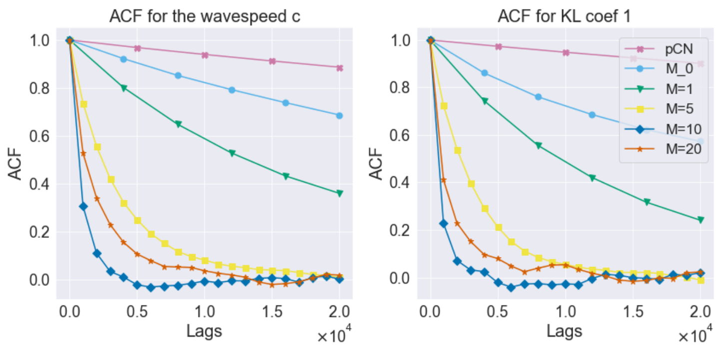


Figure 1: FES and pCN for the advection equation: ACF curves for c and the first basis coefficients. Choosing a reasonable value for M brings large perfomance gains

4.2 Path reconstruction for Langevin dynamics

In this example we compare the performance of pCN, FES, and the hybrid sampler developed in [8]. This hybrid sampler also considers a truncation of the KL expansion, but has adaptive Gaussian proposals for the finite-dimensional space.

We also compare the performance of FES for different number of walkers as well as joint and MwG updates. We find that MwG updates are preferable, and that using a large number of walkers is better (as recommended in [5]).

We consider Langevin dynamics on a single well potential giving the evolution of position X_t and momentum P_t:

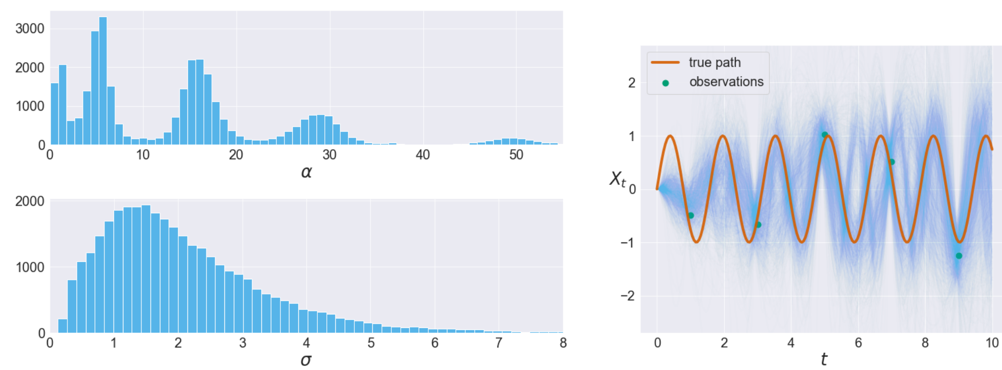
Our second example involves parameter estimation for the noisy harmonic oscillator

$$\begin{cases} dX_t = P\,dt, \\ dP_t = -\alpha X_t\,dt + \sigma\,dW_t, \\ X = P = 0, \quad t = 0. \end{cases} \tag{10}$$

We use as true path $X(t) = \sin(4t)$ and generate noisy observations from it. Note that this model misfit makes the sampling problem more challenging.

We simultaneously estimate the drift parameter α, the diffusion parameter σ, and the driving Brownian motion W from noisy observations. We compare the performance of pCN, the new ensemble sampler, and another gradient-free sampler in the literature.

This posterior exhibits multimodality in the alpha parameter which change the frequency of the X_t paths (as seen in figure 2).



(a) Posterior pdf

(b) Posterior paths

Figure 2: (a) Posterior pdf for α and σ parameter for the joint ensemble sampler (b) Posterior paths with multimodality for the joint ensemble sampler with observations

We show the ACF plots of the different samplers in figure (3). We find that pCN works slowly because of the multimodality of the α parameter. However, the new ensemble sampler freely traverses the modes, leading to faster sampling than pCN and the other alternative approach that explicitly estimates the posterior covariance matrix. FES is therefore more robust than the other samplers to mild multimodality as well as model misfit (recall that the data was generated from $X(t) = \sin(4t)$).

We also find that as the hybrid sampler takes a large number of iterations for the adaptation of the Gaussian proposal to stabilise (around 1 million iterations). This long adaptation can be a problem when the forward model is computationally expensive, as is often the case with ODE and PDE models.

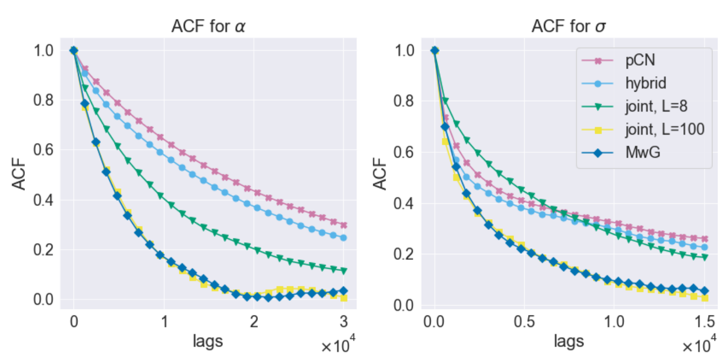


Figure 3: ACF curves for the two scalar parameters in langevin dynamics

5 Conclusion

We introduce a new gradient-free sampler for function spaces which is shown to outperform a random walk proposal (pCN) as well as a specialised gradient-free sampler in the literature. In terms of limitations, there are two aspects to our work. Both work well but both can potentially be improved further. First, we approximate the subspace where pCN is not good enough for sampling using the KL expansion and keeping the first few basis elements. It would beneficial to find a better way to estimate this low-dimensional subspace without gradients. However, estimating the full covariance is difficult, especially for high discretisation. After isolating the subspace where pCN is not good enough for sampling, the AIES is a good gradient-free sampler, but there might be other better methods that can be proposed, specifically, in the case of extreme multimodality, which is acknowledged [6] to be the main limitation of AIES.

References

[1] Alexandros Beskos, Mark Girolami, Shiwei Lan, Patrick E Farrell, and Andrew M Stuart. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351, 2017.

[2] Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350, 2008.

[3] Simon L Cotter, Gareth O Roberts, Andrew M Stuart, and David White. MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, pages 424–446, 2013.

[4] Tiangang Cui, Kody JH Law, and Youssef M Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.

[5] Daniel Foreman-Mackey, David W Hogg, Dustin Lang, and Jonathan Goodman. emcee: the MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306, 2013.

[6] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, 2010.

[7] Daniel Rudolf and Björn Sprungk. On a generalization of the preconditioned Crank–Nicolson Metropolis algorithm. *Foundations of Computational Mathematics*, 18(2):309–343, 2018.

[8] Qingping Zhou, Zixi Hu, Zhewei Yao, and Jinglai Li. A hybrid adaptive mcmc algorithm in function spaces. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):621–639, 2017.