A hybrid adaptive MCMC algorithm in function spaces

Qingping Zhou   Zixi Hu   Zhewei Yao   Jinglai Li

School of Mathematical Sciences and Institute of Natural Sciences

Shanghai Jiao Tong University

November 13, 2017

Outline

1. Background,motivation and goals
   - Bayesian inverse problem
   - Benefits of the hybrid adaptive pCN
2. Hybrid adaptive preconditioned Crank-Nicolson(Hybrid adaptive pCN)
   - Adaptive MCMC
   - Preconditioned Crank-Nicolson(pCN)
   - Structure of hybrid adaptive pCN
3. Numerical experiments
   - Prior covariance
   - Ordinary differential equation
   - One-dimensional heat conduction equation
4. Interesting conclusions

Bayesian inverse problem: infer input parameters from observations using MCMC

X is a separable Hilbert space equipped with inner product $< \cdot, \cdot >$.
The norm in this space $\| \cdot \|_C^2$ is derived by $< C^{-1/2} \cdot, C^{-1/2} \cdot >$

A typical inverse problem assumes that the unknown u is mapped to the data y via a forward model:

$$y = G(u) + \zeta \tag{1}$$

where $G : X \rightarrow \mathbb{R}^d$. ζ is the observational noise and is usually defined as a d-dimensional centered Gaussian measure $N(0, C_\zeta)$.

The Bayesian framework of inverse problem

Given the prior $\mu_0(u)$ of u,the solution can be obtained by sampling from the posterior probability measure $\mu^y(u)$,for u given y.

Bayesian inverse problem: infer input parameters from observations using MCMC

The posterior measure μ^y of u conditional on data y is provided by the Radon-Nikodym derivative:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp(-\Phi^y(u)) \quad (2)$$

where $Z = \int exp(-\Phi^y(u))\mu_0(du)$, and $\mu_0 \sim N(0, C)$.
The posterior probability measure $\mu^y(du)$ is given by:

$$\mu^y(du) \propto \exp(-\Phi^y(u)\mu_0(du) \quad (3)$$

Remind the forward model,we can obtain $\Phi(u) = \frac{1}{2}|C_\zeta^{-1/2}(G(u) - y)|_2^2$.
Without causing any ambiguity, we can drop the superscript y in Φ^y and μ^y for simplicity.

Review of a classical MCMC algorithm

Suppose we want to get samples from the target measure $\mu(u)$, the general MCMC procedure is as following:

1. Initialize $u^{(k)} = u^0$ and set $k = 0$
2. Propose $v^{(k)}$ from the proposal density $q(u^{(k)}, \cdot)$
3. Compute $a(u^{(k)}, v^{(k)}) = \min\{1, \frac{\mu(v^{(k)})}{\mu_0(u^{(k)})} \frac{q(u^{(k)}, v^{(k)})}{q(v^{(k)}, u^{(k)})}\}$.
4. If $a(u^{(k)}, v^{(k)}) > rand([0, 1])$,then

   Accept: $u^{(k+1)} = v^{(k)}$

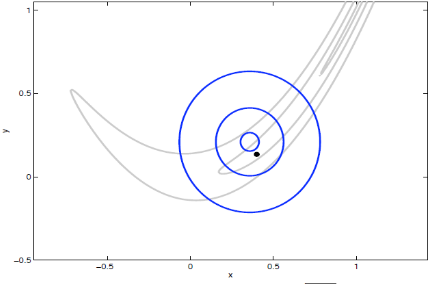
   Otherwise

   Reject: $u^{(k+1)} = u^{(k)}$

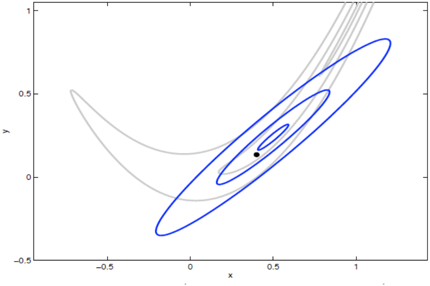
   End If
5. $k \leftarrow k + 1$

Benefits of hybrid adaptive pCN

- pCN is dimensional independent,which make it possible to infer the unknown in high-dimensional or infinite space like function spaces.
- Hybrid adaptive pCN can further improve the efficiency of pCN via selecting the suitable way of which updates the posterior covariance Σ of the projection space.



(a) normal Σ



(b) adaptive Σ

Adaptive MCMC

Effectiveness of MCMC

The performance of MCMC heavily depends on how the proposal distribution fits the target distribution.

Adaptive MCMC starts with an initial guess of the post covariance Σ and then updates it based on the sample path.Given a set of samples $\{x_1, \dots, x_n, \dots\}$. We can update Σ with

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4a}$$

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x})(x_i - \hat{x})^T + \delta I, \tag{4b}$$

where δ is a small positive constant and I is the identity matrix. The term δI is introduced to stabilize the iteration.

Preconditioned Crank-Nicolson(pCN)

The pCN method proposes new sample according to the equation:

$$v = (1-\beta^2)^{\frac{1}{2}} u + \beta\omega, \tag{5}$$

where $\beta \in [0, 1]$ and $\omega \sim N(0, C)$.
The accept probability is

$$a(u, v) = \min\{1, \exp(\Phi(u) - \Phi(v))\}. \tag{6}$$

what we do in reality?

We can proof that there exists a complete orthonormal basis $\{e_j\}_{j \in N}$ and a sequence of non-negative numbers $\{\alpha_j\}_{j \in N}$ such that $Ce_j = \alpha_j e_j$.
The space expanded by $\{e_j\}_{j=1}^{N}$ is chosen to be the projection space.

Structure of hybrid adaptive pCN

The hybrid adaptive pCN performs an adaptive Metropolis scheme in a chosen finite dimensional subspace and a standard pCN algorithm in the complement space of the chosen subspace.
Define $u_i = \langle u, e_i \rangle$ and $v_i = \langle v, e_i \rangle$.The basic idea of hybrid adaptive pCN is as following:

$$v_i = \begin{cases} u_i + \beta w_i & \text{for } i \leq J, \\ (1-\beta^2)^{\frac{1}{2}} u_i + \beta w_i & \text{for } i > J, \end{cases} \tag{7}$$

where $\beta \in [0, 1]$, $(w_1, ..., w_J)^T \sim N(0, \Sigma)$ and $w_i \sim N(0, \alpha_i)$ for $i > J$.
The accept probability of hybrid pCN is

$$a(u, v) = \min\{1, \exp[\Phi(u) - \Phi(v) + \frac{1}{2} \sum_{i=1}^{J} \frac{u_i^2 - v_i^2}{\alpha_i}]\}. \tag{8}$$

Compared method:Adaptive pCN(ApCN)

The basic idea of ApCN is as following:

$$v_i = \begin{cases} (1-\beta^2\lambda_i/\alpha_i)^{\frac{1}{2}} u_i + \beta w_i \quad \text{where} \quad w_i \sim N(0,\lambda_i) \quad \text{for } i \leq J \\ (1-\beta^2)^{\frac{1}{2}} u_i + \beta w_i \qquad \text{where} \quad w_i \sim N(0,\alpha_i) \qquad \text{for } i > J \end{cases} \tag{9}$$

where $\beta \in [0,1]$ and $\lambda_i = \langle Ce_i, e_i \rangle^{-1}$.

The accept probability of ApCN is

$$a(u,v) = \min\{1, \exp[\Phi(u) - \Phi(v)]\}. \tag{10}$$

- Hu, Z., Yao, Z., Li, J. (2015). On an adaptive preconditioned crank-nicolson algorithm for infinite dimensional bayesian inferences. Statistics, 82(3), 79-88.

Numerical experiments: Prior covariance

The prior is taken to be a zero mean Gaussian with Matérn covariance:

$$K(t_1, t_2) = \sigma^2 \frac{2^{1-\nu}}{\mathrm{Gam}(\nu)} (\sqrt{2\nu}\frac{d}{l})^{\nu} B_{\nu}(\sqrt{2\nu}\frac{d}{l}) \tag{11}$$

where $d = |t_1 - t_2|$, $\mathrm{Gam}(\cdot)$ is the Gamma function, and $B_\nu(\cdot)$ is the modified Bessel function.

Specification of σ and l for the following numerical experiments:

tests	σ	l
ODE-test1(J=14)	1	1
ODE-test2(J=5,10,20)	1	0.2
PDE(J=14)	1	1

Performance indicators:ACF and ESS

Autocorrelation function:ACF

Given the sample chain X_t,ACF at lag k is defined as

$$\rho(k) = \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} \tag{12}$$

The smaller $\rho(k)$, the better the performance.

Effective sample size:ESS

ESS is defined as

$$ESS = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho(k)} \tag{13}$$

where N is the total sample size.Usually, $\rho(k) < 0.05$ will be discarded.
The bigger ESS, the better the performance.

ODE

The first example is an inverse problem where the forward model is governed by an ordinary differential equation:

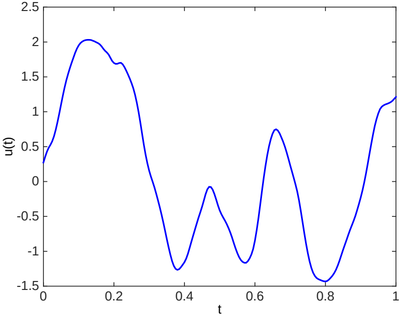
$$\frac{dx(t)}{dt} = -u(t)x(t)$$

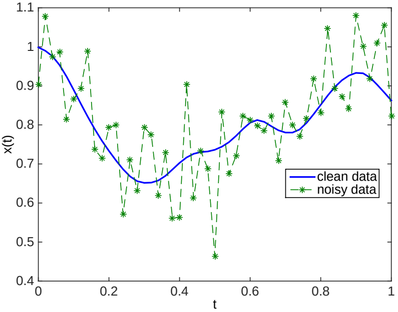
with a prescribed the initial condition be $x(0) = 1$.

The solution $x(t)$ is measured every 0.02 time unitin $[0, 1]$ and the error is assumed to be an independent Gaussian $N(0, 0.1^2)$.

We aim to infer the unknown coefficient $u(t)$ from the observed data.

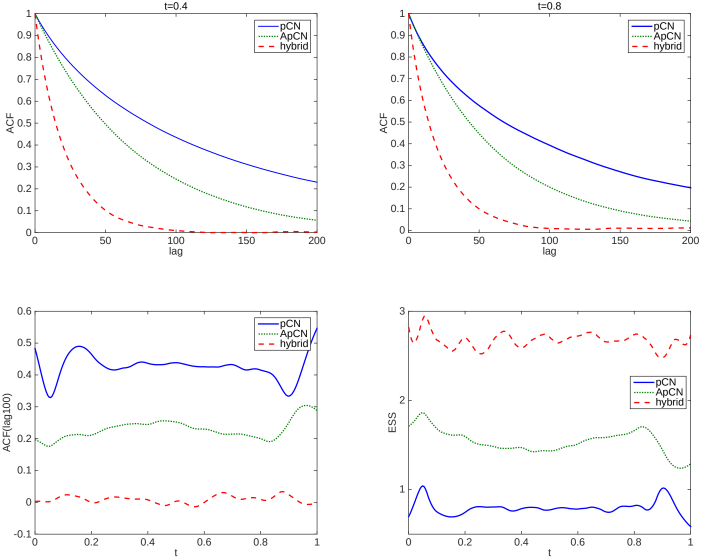
Test1(J=14): simulated data and sample size



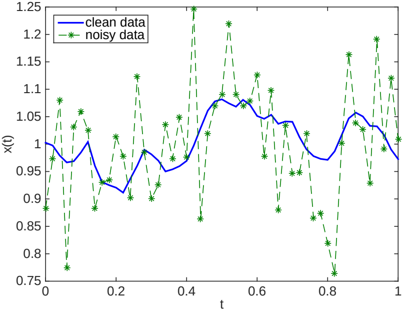


methods	pre-run	other
pCN	0	5.5×10^5
ApCN	0.5×10^5	5×10^5
hybrid algorithm	0.5×10^5	5×10^5

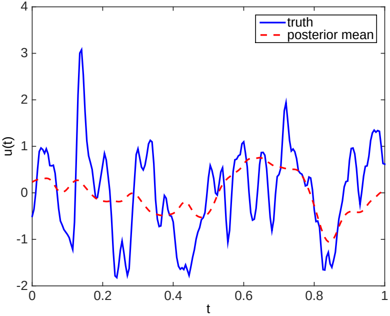
Test1(J=14): ACF and ESS



Test2(J=5,10,20):simulated data and posterior mean

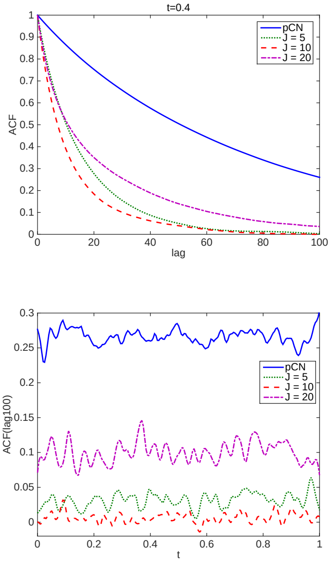


(g) the simulated data



(h) the truth and the posterior mean

Test2(J=5,10,20): ACF and ESS



One-dimensional heat conduction equation

The one-dimensional heat conduction equation in the region $x \in [0,1]$ is defined as:

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t), \tag{14a}$$

$$u(x,0) = g(x), \tag{14b}$$

with the following Robin boundary conditions:

$$-\frac{\partial u}{\partial x}(0,t) + \rho(t)u(0,t) = h_0(t), \tag{14c}$$

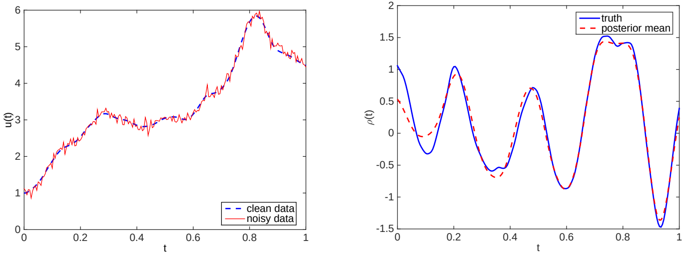
$$\frac{\partial u}{\partial x}(L,t) + \rho(t)u(L,t) = h_1(t). \tag{14d}$$

Here we choose $t \in [0,1]$ and the functions to be

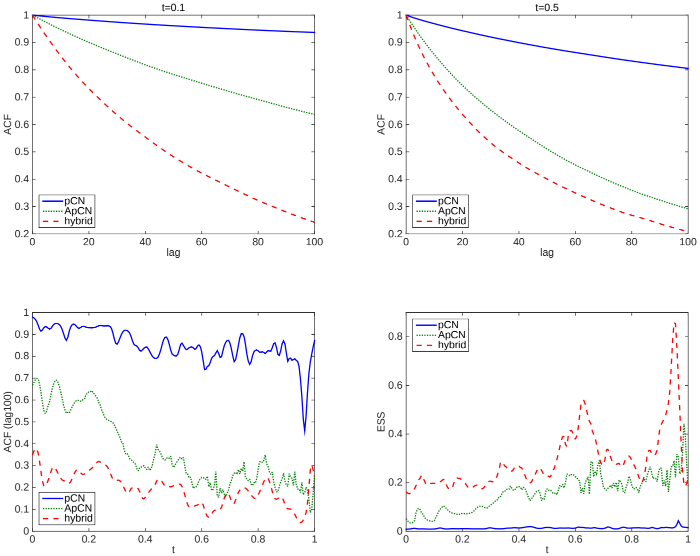
$$g(x) = x^2 + 1, \quad h_0 = t(2t+1), \quad h_1 = 2 + t(2t+2).$$

PDE setting

1. A temperature sensor is place at $x = 0$.
2. The solution is measured every $1/200$ time unit and the error in each measurement is an independent Gaussian $N(0, 0.1^2)$.
3. $\mathrm{J} = 14$
4. the number of samples is the same as ODE test,5.5×10^5 with 0.5×10^5 of pre-run.
5. simulated data and posterior mean



PDE test(J=14):ACF and ESS



Conclusions

Comparison of performance

- The experimental results the ODE and heat conduction equation, with a relatively high correlation between eigen-functions,show that the proposed adaptive method outperformed both the standard pCN and the ApCN methods.
- Hybrid method may not improve the efficiency much over the ApCN when the correlations between eigen-functions are weak.

Within hybrid pCN

- Tuning the number of adaptive eigenvalues, J, is a key part for the best performance of the hybrid adaptive method.

Some references

1. Hu, Z., Yao, Z., Li, J. (2015). On an adaptive preconditioned crank-nicolson algorithm for infinite dimensional bayesian inferences. Statistics, 82(3), 79-88.
2. Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D. (2013). MCMC methods for functions: modifying old algorithms to make them faster. Statistical Science, 28(3), pgs. 424-446.
3. Kaipio, J. P., Somersalo, E. (2015). Statistical and computational inverse problems. 16(2), xvi,339.
4. Stuart, A. M. (2010). Inverse problems: a Bayesian perspective. Acta Numerica, 19, 451-559.

Thank you