Journal of Mathematical Research with Applications
May, 2012, Vol. 32, No. 3, pp. 288–296
DOI:10.3770/j.issn:2095-2651.2012.03.003
Http://jmre.dlut.edu.cn

An Efficient Bayesian Iterative Method for Solving Linear Systems

Deng DING[1,*], Kin Sio FONG[2], Ka Hou CHAN[3]

1. *Department of Mathematics, University of Macau, Macao, P. R. China;*
2. *Faculty of Business Administration, University of Macau, Macao, P. R. China;*
3. *Keang Peng School, Macao, P. R. China*

Abstract This paper concerns with the statistical methods for solving general linear systems. After a brief review of Bayesian perspective for inverse problems, a new and efficient iterative method for general linear systems from a Bayesian perspective is proposed. The convergence of this iterative method is proved, and the corresponding error analysis is studied. Finally, numerical experiments are given to support the efficiency of this iterative method, and some conclusions are obtained.

Keywords linear system; matrix; iterative method; Bayesian perspective.

MR(2010) Subject Classification 15A06; 15A09; 62F15; 65P10

1. Introduction

Consider the linear system of equations:

$$Ax = b,\ A \in \mathbf{R}^{m\times n} \ \text{ and } \ b \in \mathbf{R}^m,$$

where $m \geq n$. There are many numerical methods to solve the system, such as direct methods (Gaussian elimination, LU and Cholesky decompositions) and iterative methods (Jacobi, Gauss-Seidel, and Krylov subspace methods), etc. (e.g. see [1]). Besides these methods based on matrix analysis, there are also some numerical methods from statistical ideas for solving the linear system (1).

By constructing a discrete Markovian chain, Forsythe and Leibler in [2] used Monte Carlo methods to find the inverse of a matrix in 1950. Since then, some different numerical algorithms based on Monte Carlo methods have been proposed to find the inverse of diagonally dominant matrices. For instance, Halton in [3] presented the sequential Monte Carlo techniques, Dimov et al proposed a Monte Carlo almost optimal algorithm in [4], etc. Recently, some authors presented new ideas based on Monte Carlo methods to find the inverse of a general matrix. A brief survey with numerical comparison for these methods can be found in [5].

Received May 31, 2011; Accepted October 31, 2011
Supported by the Research Grants from University of Macau (Grant Nos. UL020/08-Y4/MAT/JXQ01/FST; MYRG136(Y1-L2)-FST11-DD).
* Corresponding author
E-mail address: dding@umac.mo (Deng DING); yb17007@umac.mo (Kin Sio FONG); kahou0201@yahoo.com.hk (Ka Hou CHAN)

The idea of another statistical method is from the Bayesian inversion perspective. The Kaipoi and Somersalo's book [6] provides a good introduction to the Bayesian approach to inverse problems. Furthermore, the Calvetti and Somersalo's book [7] gives a useful introduction to the Bayesian perspective in scientific computing. Recently, Stuart's paper [8] surveys the newest development of the subject of Bayesian inverse problems. As a classical inverse problem, solving the linear system (1) from a Bayesian inversion perspective is also considered by several researchers (e.g. see [9–11] and references therein).

In this paper, we study the numerical approximation of the linear system (1) from a Bayesian inversion perspective. After this introduction, the Bayesian framework for solving the linear system (1), including the conditional mean estimate and the maximum a posterior estimate, is introduced in Section 2. In Section 3, we first propose a theoretical iterative scheme, and show some sufficient conditions of convergence for such scheme, as well as we discuss the corresponding error analysis. Then, based on an idea from the paper [9], we modify the iterative scheme into a numerical iterative scheme in practice, which is very efficient for the system with the general matrix A. The corresponding error analysis is also considered. Finally, we give some results of numerical experiments to support the presented numerical scheme, and we get some conclusions about this scheme in Section 4.

2. The Bayesian framework

In Bayesian inversion problems, the system (1) with additive noise term is replaced by its stochastic extension:

$$Y = AX + E, \tag{2}$$

where X is an $\mathbf{R}^n$-valued random vector, Y and E are $\mathbf{R}^m$-valued random vectors, and $A \in \mathbf{R}^{m\times n}$ is a given matrix. Denote by $\pi_{\mathrm{pr}}(x)$ the prior density of X, and by $\pi_{\mathrm{noise}}(e)$ the noise density of E. From (2), the conditional density of Y given $X = x$, which is called the likelihood density, is given by

$$\pi(y \mid x) = \pi_{\mathrm{noise}}\big(y - Ax\big). \tag{3}$$

Then, the conditional density of X given $Y = b$, which is called the posterior density of X, is provided by Bayes' formula:

$$\pi_{\mathrm{post}}(x) = \pi(x \mid b) \propto \pi_{\mathrm{pr}}(x)\pi(b \mid x) = \pi_{\mathrm{pr}}(x)\pi_{\mathrm{noise}}\big(b - Ax\big). \tag{4}$$

Here the notation '$\propto$' means to ignore the normalizing constant.

Based on the posterior density there are different estimates of the random variable X given $Y = b$, which are used to approximate the solution of system (1). The most commonly used statistical estimates are the conditional mean (CM):

$$x_{\mathrm{CM}} = \int_{\mathbf{R}^n} x\pi(x \mid b)\mathrm{d}x,$$

and the maximum a posterior (MAP):

$$x_{\mathrm{MAP}} = \arg\max_{x\in\mathbf{R}^n} \pi(x \mid b).$$

Assume that X and E are mutually independent Gaussian random vectors such that

$$X \sim \mathcal{N}(x_0, \Gamma_{\text{pr}}) \ \text{ and } \ E \sim \mathcal{N}(e_0, \Gamma_{\text{noise}}), \tag{5}$$

where covariance matrices $\Gamma_{\text{pr}} \in \mathbf{R}^{n\times n}$ and $\Gamma_{\text{noise}} \in \mathbf{R}^{m\times m}$ are positive definite. Then, we have the following result:

Theorem 1 (Theorem 3.7 in [6]) *Under the Bayesian setting (5), the posterior density of X given $Y = b$ is Gaussian and is given by*

$$\pi(x \mid b) \propto \exp\Big\{ -\frac{1}{2}(x-\bar{x})^{\text{T}}\Gamma_{\text{post}}^{-1}(x-\bar{x})\Big\},$$

where

$$\bar{x} = x_0 + \Gamma_{\text{pr}}A^{\text{T}}\big(A\Gamma_{\text{pr}}A^{\text{T}} + \Gamma_{\text{noise}}\big)^{-1}(b - Ax_0 - e_0), \tag{6}$$

and

$$\Gamma_{\text{post}} = \Gamma_{\text{pr}} - \Gamma_{\text{pr}}A^{\text{T}}\big(A\Gamma_{\text{pr}}A^{\text{T}} + \Gamma_{\text{noise}}\big)^{-1}A\Gamma_{\text{pr}}. \tag{7}$$

Moreover, $\bar{x} = x_{\text{CM}} = x_{\text{MAP}}$.

As mentioned in [6], the posterior covariance matrix Γ_{post} has an alternative representation:

$$\Gamma_{\text{post}} = \big(\Gamma_{\text{pr}}^{-1} + A^{\text{T}}\Gamma_{\text{noise}}^{-1}A\big)^{-1}. \tag{8}$$

Furthermore, the posterior mean $\bar{x}$ can be written as

$$\bar{x} = \Gamma_{\text{post}}\big(A^{\text{T}}\Gamma_{\text{noise}}^{-1}(b - e_0) + \Gamma_{\text{pr}}^{-1}x_0\big). \tag{9}$$

However, there are some other methods in practice to find the estimates x_{CM} and x_{MAP}.

The first method is the application of Markov chain Monte Carlo (MCMC) methods to calculate the CM estimate x_{CM}:

$$x_{\text{CM}} \approx \frac{1}{K}\sum_{k=1}^{K} x_k, \tag{10}$$

where x_k, $k = 1, \ldots, K$, are the random numbers sampled under the posterior density $\pi_{\text{post}}(x)$ in (4) by some MCMC algorithms. For instance, the well-known Metropolis-Hastings algorithm provides a foundational and central role in MCMC computation [6, 12].

The main idea of another method is to consider the MAP estimate x_{MAP}. Let $e_0 = 0$ and take the Cholesky factorizations:

$$\Gamma_{\text{pr}}^{-1} = L^{\text{T}}L \ \text{ and } \ \Gamma_{\text{noise}}^{-1} = S^{\text{T}}S.$$

Then, the posterior density in (4) can be rewritten as

$$\pi(x \mid b) \propto \exp\Big\{ -\frac{1}{2}\Big(\big\|L(x-x_0)\big\|_2^2 + \big\|S(b-Ax)\big\|_2^2\Big)\Big\},$$

where $\|\cdot\|_2$ is the vector 2-norm in $\mathbf{R}^n$ or $\mathbf{R}^m$. Therefore, the MAP estimate x_{MAP} is also the Tikhonov solution:

$$x_{\text{MAP}} = L^{-1}w + x_0,$$

where w is the minimizer of the functional:

$$G(w) = \left\|S(b - Ax_0) - SAL^{-1}w\right\|_2^2 + \|w\|_2^2.$$

The observations above suggest that the $x_{\rm MAP}$ can be approximated by solving iteratively the following preconditioning system:

$$SAL^{-1}y = S(b - Ax_0), \quad L(x - x_0) = y.$$

Hence, this method is also called the preconditioned iterative method from a Bayesian inversion perspective [10, 11].

Each of these methods has its advantages and disadvantages and specific applications. For example, according to our numerical experiments, the Metropolis-Hastings algorithm suffers from the local-trap problem in the most cases in the linear system (1), which has been pointed out by many researchers [6, 12]. On the other hand, the second method seriously depends on the selection of prior covariance $\Gamma_{\rm pr}$ and noise covariance $\Gamma_{\rm noise}$.

Recently, Akçelik et al in [9] presented a fast method for computation of an approximation to the posterior covariance $\Gamma_{\rm post}$, which is given in (7). In the next section, we will propose an iterative method based on their idea.

3. The Bayesian iterative scheme

In this section, based on Theorem 1 we present an iterative scheme for solving the linear system (1). First, we randomly select an initial vector $x_0 \in \mathbf{R}^n$ and set $e_0 = 0$. We also appropriately select positive numbers $\gamma_{\rm pr}$ and $\gamma_{\rm noise}$ and set

$$\Gamma_{\rm pr} = \gamma_{\rm pr} I_{n\times n} \quad \text{and} \quad \Gamma_{\rm noise} = \gamma_{\rm noise} I_{m\times m}, \tag{11}$$

where $I_{n\times n}$ and $I_{m\times m}$ are identity matrices. We consider the first iterative formula:

$$x_k = \Gamma_{\rm post} A^{\rm T} \Gamma_{\rm noise}^{-1} b + \Gamma_{\rm post} \Gamma_{\rm pr}^{-1} x_{k-1}, \tag{12}$$

where $\Gamma_{\rm post}$ is given in (7) or (8). In the following, we need the concept of matrix 2-norm, which is defined by

$$\left\|M\right\|_2 = \sqrt{\lambda_{\max}(M^{\rm T}M)} = \sigma_{\max}(M), \quad M \in \mathbf{R}^{m\times n},$$

where $\lambda_{\max}(M^{\rm T}M)$ is the largest eigenvalue of $M^{\rm T}M$, and $\sigma_{\max}(M)$ is the largest singular value of M. We also assume the system (1) has an exact solution x_b.

Theorem 2 *Let $\{x_k\}$ be a sequence generated by the iterative formula* (12)*. Then, under the Bayesian setting (5), for every k,*

$$\|x_k - x_b\|_2 \le \left\|\Gamma_{\rm post}\Gamma_{\rm pr}^{-1}\right\|_2^k \|x_0 - x_b\|_2, \tag{13}$$

i.e. the sequence $\{x_k\}$ converges to x_b if $\left\|\Gamma_{\rm post}\Gamma_{\rm pr}^{-1}\right\|_2 < 1$.

Proof From (8) and (12) we have

$$x_k - x_b = \Gamma_{\rm post} A^{\rm T} \Gamma_{\rm noise}^{-1} b + \Gamma_{\rm post} \Gamma_{\rm pr}^{-1} x_{k-1} - x_b$$

$$\begin{aligned}
&= \Gamma_{\text{post}}\big(A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}Ax_b + \Gamma_{\text{pr}}^{-1}x_{k-1} - \Gamma_{\text{post}}^{-1}x_b\big)\\
&= \Gamma_{\text{post}}\big[A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}Ax_b + \Gamma_{\text{pr}}^{-1}x_{k-1} - \big(\Gamma_{\text{pr}}^{-1} + A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}A\big)x_b\big]\\
&= \Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}(x_{k-1} - x_b).
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\|x_k - x_b\|_2 = \big\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}(x_{k-1} - x_b)\big\|_2 \;&\le\; \big\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2\|x_{k-1} - x_b\|_2\\
&\le \cdots \le \big\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2^k\|x_0 - x_b\|_2.
\end{aligned}$$

The theorem is proved. □

Theorem 3 *Under the Bayesian setting (5) and (11),* $\big\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2 < 1$ *if and only if* $A^{\mathrm{T}}A$ *is nonsingular. Moreover,*

$$\big\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2 = \Big(1 + \frac{\gamma_{\text{pr}}}{\gamma_{\text{noise}}}\lambda_{\min}(A^{\mathrm{T}}A)\Big)^{-1}, \tag{14}$$

where $\lambda_{\min}(A^{\mathrm{T}}A)$ *is the smallest eigenvalue of* $A^{\mathrm{T}}A$.

Proof Note that $\Gamma_{\text{pr}} = \gamma_{\text{pr}}I$ and $\Gamma_{\text{noise}} = \gamma_{\text{noise}}I$. From (8) we have

$$\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1} = \big(\Gamma_{\text{pr}}^{-1} + A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}A\big)^{-1}\Gamma_{\text{pr}}^{-1} = \Big(I + \frac{\gamma_{\text{pr}}}{\gamma_{\text{noise}}}A^{\mathrm{T}}A\Big)^{-1}.$$

Denote by $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ the largest and smallest singular values of matrix M, respectively. Then we have

$$\begin{aligned}
\big\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2 \;&=\; \sigma_{\max}\Big[\Big(I + \frac{\gamma_{\text{pr}}}{\gamma_{\text{noise}}}A^{\mathrm{T}}A\Big)^{-1}\Big]\\
&= \Big[\sigma_{\min}\Big(I + \frac{\gamma_{\text{pr}}}{\gamma_{\text{noise}}}A^{\mathrm{T}}A\Big)\Big]^{-1} \;=\; \Big(1 + \frac{\gamma_{\text{pr}}}{\gamma_{\text{noise}}}\lambda_{\min}(A^{\mathrm{T}}A)\Big)^{-1}.
\end{aligned}$$

That is the expression (14), and hence, $\|\Gamma_{\text{post}}\Gamma_{\text{pr}}^{-1}\|_2 < 1$ if and only if $\lambda_{\min}(A^{\mathrm{T}}A) > 0$, i.e., $A^{\mathrm{T}}A$ is nonsingular. The theorem is proved. □

From Theorems 2 and 3, we know that the iterative scheme defined by (12) is convergent. Moreover, from the expression (14) we can select an appropriate ratio $\gamma_{\text{pr}}/\gamma_{\text{noise}}$ to improve the convergence rate of the iterative scheme. However, we must mention here, it is not easy to directly find the posterior covariance matrix Γ_{post}. In the following, based on the idea given by Akçelik et al in [9] to calculate an approximation of Γ_{post}, we further improve the scheme (12).

We first take the spectral decomposition of the prior-preconditioned Hessian of the data misfit:

$$\tilde{H}_{\text{misfit}} \triangleq \Gamma_{\text{pr}}^{1/2}A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}A\Gamma_{\text{pr}}^{1/2} = V\Lambda V^{\mathrm{T}},$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues of $\tilde{H}_{\text{misfit}}$ and V is the matrix whose columns are the corresponding eigenvectors. When the eigenvalues of $\tilde{H}_{\text{misfit}}$ decay rapidly, we select a small number $\delta > 0$, and then we take the low-rank approximation:

$$\Gamma_{\text{pr}}^{1/2}A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}A\Gamma_{\text{pr}}^{1/2} \approx V_r\Lambda_r V_r^{\mathrm{T}},$$

under the cutoff criterion:

$$\lambda_i \geq \delta, \quad i = 1, \ldots, r \quad \text{and} \quad \lambda_{r+j} < \delta, \quad j = 1, \ldots, n-r, \tag{15}$$

where $\Lambda_r = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ is the truncated eigenvalue matrix of r largest eigenvalues, and $V_r \in \mathbf{R}^{n\times r}$ is the corresponding eigenvector matrix. Then, we set

$$D_r = \operatorname{diag}\Big(\frac{\lambda_1}{1+\lambda_1}, \ldots, \frac{\lambda_r}{1+\lambda_r}\Big),$$

and define

$$\tilde{\Gamma}_{\text{post}} = \Gamma_{\text{pr}} - \Gamma_{\text{pr}}^{1/2} V_r D_r V_r^{\mathrm{T}} \Gamma_{\text{pr}}^{1/2}, \tag{16}$$

which is used as an approximation of the posterior covariance matrix Γ_{post} under the Bayesian setting (5). According to the error analysis given in [9], we have the following estimation for such an approximation:

$$\begin{aligned} \big\|\Gamma_{\text{post}} - \tilde{\Gamma}_{\text{post}}\big\|_F &= \Big\|\Gamma_{\text{pr}}^{1/2}\big(\tilde{H}_{\text{misfit}} + I\big)^{-1}\Gamma_{\text{pr}}^{1/2} - \Gamma_{\text{pr}}^{1/2}\big(I - V_r D_r V_r^{\mathrm{T}}\big)\Gamma_{\text{pr}}^{1/2}\Big\|_F \\ &\leq \gamma_{\text{pr}}\Big\|\big(\tilde{H}_{\text{misfit}} + I\big)^{-1} - \big(I - V_r D_r V_r^{\mathrm{T}}\big)\Big\|_F \\ &< \gamma_{\text{pr}}\sqrt{\sum_{j=r+1}^{n}\lambda_j^2} \leq \gamma_{\text{pr}}\delta\sqrt{(n-r)}. \end{aligned} \tag{17}$$

Here $\|\cdot\|_F$ is the Frobenius norm of matrix.

Now, using the approximation (16) to modify the formula (12), we obtain a Bayesian iterative scheme:

$$\tilde{x}_k = \tilde{\Gamma}_{\text{post}} A^{\mathrm{T}} \Gamma_{\text{noise}}^{-1} b + \tilde{\Gamma}_{\text{post}} \Gamma_{\text{pr}}^{-1} \tilde{x}_{k-1} \tag{18}$$

with $\tilde{x}_0 = x_0$.

Theorem 4 *Let $\{\tilde{x}_k\}$ be a sequence generated by the iterative formula (18). Then, under the Bayesian setting (5) and (11), and the cutoff criterion (15),*

$$\big\|\tilde{x}_k - \tilde{x}_{k-1}\big\|_2 \leq \Big(\frac{1}{1+\delta}\Big)^{k-1}\big\|\tilde{x}_1 - \tilde{x}_0\big\|_2 \tag{19}$$

holds for every $k > 1$, i.e., the sequence $\{\tilde{x}_k\}$ is convergent.

Proof First, according to the formula (18), we have

$$\begin{aligned} \big\|\tilde{x}_k - \tilde{x}_{k-1}\big\|_2 = \big\|\tilde{\Gamma}_{\text{post}}\Gamma_{\text{pr}}^{-1}(\tilde{x}_{k-1} - \tilde{x}_{k-2})\big\|_2 \ &\leq \ \big\|\tilde{\Gamma}_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2\big\|\tilde{x}_{k-1} - \tilde{x}_{k-2}\big\|_2 \\ &\leq \cdots \leq \big\|\tilde{\Gamma}_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2^{k-1}\big\|\tilde{x}_1 - \tilde{x}_0\big\|_2. \end{aligned}$$

On the other hand, from (16) we have

$$\begin{aligned} \big\|\tilde{\Gamma}_{\text{post}}\Gamma_{\text{pr}}^{-1}\big\|_2 \leq \big\|\tilde{\Gamma}_{\text{post}}\big\|_2\big\|\Gamma_{\text{pr}}^{-1}\big\|_2 \ &= \ \lambda_{\max}\big(\Gamma_{\text{pr}} - \Gamma_{\text{pr}}^{1/2}V_r D_r V_r^{\mathrm{T}}\Gamma_{\text{pr}}^{1/2}\big)\gamma_{\text{pr}}^{-1} \\ &= \lambda_{\max}\Big(\Gamma_{\text{pr}}^{1/2}\big(I - V_r D_r V_r^{\mathrm{T}}\big)\Gamma_{\text{pr}}^{1/2}\Big)\gamma_{\text{pr}}^{-1} \ \leq \ \big(1 - \lambda_{\min}(D_r)\big) \ = \ \frac{1}{1+\lambda_r}, \end{aligned}$$

where λ_r is the r-th largest eigenvalue of $\Gamma_{\text{pr}}^{1/2}A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}A\Gamma_{\text{pr}}^{1/2}$ such that $\lambda_r \geq \delta$ and $\lambda_{r+j} < \delta$, $j = 1, \ldots, n-r$, under the cutoff criterion (15). Thus, we get the estimation (19). The theorem is proved. □

From the estimation (19), we see that the increase of the cutoff criterion δ can improve the convergence rate of the iterative scheme (18). However, from the approximation (17), we also see that the increase of δ influences the accuracy of the scheme. Hence, appropriately to select the ratio $\gamma_{\rm pr}/\gamma_{\rm noise}$ and the cutoff criterion δ can make the scheme more accurate and efficient. In the next section, we will illustrate how these parameters influence the scheme.

4. Numerical experiments

In this section, we employ the Bayesian iterative scheme (18) to do numerical tests for some practical examples of system (1). In all experiments, we use the random matrix generator in Matlab to select a matrix A such that its condition number $\kappa(A)$ is greater than a given bound, and randomly select a vector b. Then, we compute $\tilde{x}_k$ by (18) with the initial value $\tilde{x}_0 = 0$ until the error is less than 0.1 or the the iteration number k reaches 1000. In this section, the error of a numerical solution $\tilde{x}$ is given by

$$\text{Error} = \left\|A\tilde{x} - b\right\|_\infty, \tag{20}$$

where $\|v\|_\infty = \max_{1\le i\le m}|v_i|$ for all $v = (v_1, \ldots, v_m) \in \mathbf{R}^m$. On the other hand, the computer used for all numerical experiments has an Intel(R) Core(TM)2 Duo CPU E8400@3.00GHz 2.99GHz, 3.21GB RAM, and all codes in these numerical experiments are written in Matlab 7.11.

	$\gamma_{\rm pr} = 10$		$\gamma_{\rm pr} = 100$	
$\gamma_{\rm pr}/\gamma_{\rm noise}$	Iter. Num.	Error	Iter. Num.	Error
1.00E+1	1000	7.657379	1000	7.657385
1.00E+2	1000	2.338975	1000	2.338899
1.00E+3	341	0.099650	342	0.099253
1.00E+4	37	0.094159	37	0.094051
5.00E+4	11	0.096541	10	0.091853
1.00E+5	1000	0.174353	1000	0.139130

Table 1 Errors for matrix A_1 with $\delta = 0.1$

We first test the different values of $\gamma_{\rm pr}$ and different ratios of $\gamma_{\rm pr}/\gamma_{\rm noise}$. We use a 200×200 matrix A_1 whose condition number $\kappa(A_1) = 219017.1$. Tables 1–2 give the iteration numbers and errors against different ratios of $\gamma_{\rm pr}/\gamma_{\rm noise}$ for different value of $\gamma_{\rm pr}$ and different cutoff criterions δ, respectively.

From these two tables, we can see that the effect of the ratio of $\gamma_{\rm pr}/\gamma_{\rm noise}$ to the iteration numbers and the accuracy of the scheme is very evident and significant, while the values of $\gamma_{\rm pr}$ and δ have no evident effect on them. From these tables, we also see that $\gamma_{\rm pr}/\gamma_{\rm noise} = 50000$ is optimal in this case.

	$\delta = 0.01$		$\delta = 1$	
$\gamma_{\rm pr}/\gamma_{\rm noise}$	Iter. Num.	Error	Iter. Num.	Error
1.00E+1	1000	7.657381	1000	7.657381
1.00E+2	1000	2.338859	1000	2.338859
1.00E+3	341	0.099828	341	0.099828
1.00E+4	37	0.093121	37	0.093121
5.00E+4	10	0.085442	10	0.085442
1.00E+5	1000	0.181236	1000	0.181236

Table 2 Errors for matrix A_1 with $\gamma_{\rm pr} = 1000$

	A_2		A_3	
$\gamma_{\rm pr}/\gamma_{\rm noise}$	Iter. Num.	Error	Iter. Num.	Error
1.00E+1	1000	10.85551	1000	2.337134
1.00E+2	561	0.099649	206	0.098344
1.00E+3	59	0.095485	24	0.089019
1.00E+4	9	0.071523	1000	0.371761
1.00E+5	1000	0.399251	1000	3.206323

Table 3 Errors for matrices A_2 and A_3 with $\gamma_{\rm pr} = 100$ and $\delta = 0.1$

We then use a 1000×1000 matrix A_2 whose condition number $\kappa(A_2) = 18118.85$, and a 2000×2000 matrix A_3 whose condition number $\kappa(A_3) = 13091.33$. Table 3 gives the iterative numbers and errors against different ratios of $\gamma_{\rm pr}/\gamma_{\rm noise}$ for $\gamma_{\rm pr} = 100$ and $\delta = 0.1$.

From this table, we see that if the ratio $\gamma_{\rm pr}/\gamma_{\rm noise}$ is appropriately selected (in these two cases, we may take $\gamma_{\rm pr}/\gamma_{\rm noise} = 10000$ for A_2 and take $\gamma_{\rm pr}/\gamma_{\rm noise} = 1000$ for A_3), then the scheme can still reach the satisfied accuracy though the sizes of matrices are quite large.

Summarizing from Tables 1–3, we see that the iteration number decreases evidently and significantly as the ratio $\gamma_{\rm pr}/\gamma_{\rm noise}$ increases, which coincides with the estimates (13) and (14). However, from (17) we know that the approximation of $\Gamma_{\rm post}$ is effected by the numbers $\gamma_{\rm pr}$ and δ. Meanwhile, if we fix $\gamma_{\rm pr}$, then $\gamma_{\rm noise}$ will tend to zero as the ratio $\gamma_{\rm pr}/\gamma_{\rm noise}$ tends to infinite, which leads to the invalidation of our Bayesian framework. That explains why the iteration number becomes quickly worse when the ratio $\gamma_{\rm pr}/\gamma_{\rm noise}$ is too large in our numerical experiments. Also, we find that the appropriate value of $\gamma_{\rm pr}/\gamma_{\rm noise}$ has a light increase with the condition number becoming large, and it does not depend on the matrix size.

Finally, we compare the Bayesian iterative scheme (18) to the classical Bayesian method: the Monte Carlo method for CM estimate, which is given in (10), via the Metropolis-Hastings (MH) algorithm. The detail algorithm for this classical Bayesian method can be found in the book [6].

Table 4 respectively gives errors and CPU times (in second) of the Bayesian iterative scheme and the classical Bayesian method for solving the system (1) with matrix A_1. Here we set $\gamma_{\rm pr} = 100$ and $\delta = 0.1$. From this table we see that the Bayesian iterative scheme is evidently

and significantly much more efficient and accurate than the classical one. In fact, from our experiments it seems that the classical Bayesian method would be invalid when the condition number $\kappa(A)$ is quite large. Contrarily, we conclude that the Bayesian iterative scheme is very efficient even the matrix A is dense with large condition number, and has not any construction.

The Bayesian iterative scheme			
$\gamma_{\rm pr}/\gamma_{\rm noise}$	Iteration Num.	Error	CPU time(s)
50000	10	0.091853	0.093601
The Monte Carlo method via the MH algorithm			
$\gamma_{\rm pr}/\gamma_{\rm noise}$	Simulation Num.	Error	CPU time(s)
1000	10000	6660.822	3.619223
10000	10000	6574.869	5.525623
50000	50000	13266.50	18.43932

Table 4 Errors and CPU times for matrix A_1 via using different methods

Acknowledgments The authors are very grateful to the referee for his very valuable suggestions which helped in improving our paper.

References

[1] G. H. GOLUB, C. F. VAN LOAN. *Matrix Computations (III)*. Johns Hopkins University Press, Baltimore, MD, 1996.
[2] G. E. FORSYTHE, R. A. LEIBLER. *Matrix inversion by a Monte Carlo method*. Math. Tables and Other Aids to Computation, 1950, **4**: 127–129.
[3] J. H. HALTON. *Sequential Monte Carlo techniques for the solution of linear systems*. J. Sci. Comput., 1994, **9**(2): 213–257.
[4] I. T. DIMOV, T. T. DIMOV, T. V. GUROV. *A new iterative Monte Carlo approach for inverse matrix problem*. J. Comput. Appl. Math., 1998, **92**(1): 15–35.
[5] Deng DING, Xiaoqing JIN, Yingying ZHANG. *Numerical Comparison of Monte Carlo Methods for Linear Systems*. In Recent Advances in Computational Mathematics. Int. Press, Boston, MA, 2008.
[6] J. KAIPIO, E. SOMERSALO. *Statistical and Computational Inverse Problems*. Springer-Verlag, New York, 2005.
[7] D. CALVETTI, E. SOMERSALO. *Introduction to Bayesian Scientific Computing*. Springer, New York, 2007.
[8] A. M. STUART. *Inverse problems: a Bayesian perspective*. Acta Numer., 2010, **19**: 451–559.
[9] H. P. FLATH, L. C. WILCOX, V. AKÇELIK, et al. *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*. SIAM J. Sci. Comput., 2011, **33**(1): 407–432.
[10] D. CALVETTI. *Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective*. J. Comput. Appl. Math., 2007, **198**(2): 378–395.
[11] D. CALVETTI, E. SOMERSALO. *Priorconditioners for linear systems*. Inverse Problems, 2005, **21**(4): 1397–1418.
[12] Faming LIANG, Chuanhai LIU, R. J. CARROLL. *Advanced Markov Chain Monte Carlo Methods*. John Wiley & Sons, Ltd., Chichester, 2010.