Improved generic acceptance function for Multi-point Metropolis algorithm

Yinghua Zhang[1, a], Wensheng Zhang[1,b]

[1]State key Laboratory of Management and Control of Complex Systems

Institute of Automation, Chinese Academy of Science, BeiJing, China

[a]yinghua.zhang@ia.ac.cn, [b]wensheng.zhang@ia.ac.cn

Keywords: MCMC; Metropolis Hastings; Multi-point; Detailed Balance Condition;

Abstract. The key of designing MCMC algorithm is the choice of acceptance function. In this work, Selection criteria of acceptance function is given, and an improved Multi-point Metropolis algorithm with generic acceptance function is proposed, which is called GAF-MPM. Then GAF-MPM is showed to satisfy Detailed Balance Condition to ensure its convergence, the strict proof is given in this work. Further, several different acceptance functions are given, and we discuss the effect on the convergence speed, acceptance rate of the samples and the correlation due to the choice of different acceptance functions. Finally, its correctness and effectiveness is proven through numerical experiments.

Introduction

Bayesian inference obtains estimates of the parameters by sampling the posterior probability ofthe unknown parameters, and it is an effective probabilistic inversion method to deal with the inverse problems of non-linear, uncertainty systems [1, 2]. The Markov chain Monte Carlo (MCMC) method is one of the commonly used sampling methods for Bayesian inference. Currently, one of the most commonly used MCMC method is the Metropolis Hastings (M-H) algorithm [3, 4]. It gets the next point by random walking near to the current point, and then constructs a Markov chain. Specifically, a new sample points is produced by the proposal transition function, which is a definite function of the old sample point, then we choose whether to accept the new sample with a certain probability, otherwise, the sample turned to continue to use old samples of the previous step. The biggest flaw of the above algorithm is: samples taken are often highly correlated, resulting in estimated based on the calculation of the sample generated expectations tend to have larger variance. Furthermore, the classic M-H algorithm is often trapped into local optimal solution, convergence speed is slow.

A modified method is to augment sampling step [5, 6]. Multiple-Try Metropolis (MTM) method [7] is an extended form of the MH algorithm. It chooses the next sample point from some independent point, which is generated by the current point, then constructs a Markov chain. The advantage of MTM algorithm is that it is capable of exploring the probability space better through increasing step size, however, it did not reduce the sample probability of acceptance. The efficiency of MTM algorithm is better than the classical M-H algorithm. Pandolfi [6] proposed an improved algorithm of MTM by improving the Weight Function form. Martino [8] proposed an improved MTM algorithm with universal acceptance function on the basis of the above algorithm. Qin [9] gave an improved MCMC sampling method named Multi-point Metropolis (MPM) algorithm by increasing the associated relation of samples. Martino [10] proposed an MPM algorithm with generic weight function by improving the form of a weighting function.

Under the framework of MPM algorithm with generic weight function, this paper analyses the characteristics of acceptance function, and proposes an improved Multi-point Metropolis algorithm named GAF-MPM with generic acceptance function. To ensure its convergence, Detailed Balance Condition proof is given. Further, several different acceptance functions are given, and we prove its correctness and effectiveness through numerical experiments and discuss their effect on the experiment results. Finally, we conclude.

Published by Atlantis Press, Paris, France.

Improved MPM Algorithm

In M-H and its various variant algorithms, weight function and acceptance function is the key to design a Markov chain. To ensure the Markov chain constructed by algorithm has stationary distribution and converges to the target distribution, weight function and acceptance function used must satisfy certain conditions [11, 12]. Under theframework of MPM algorithm with generic weight function, we analyses the characteristics of acceptance function, and proposes an improved Multi-point Metropolis algorithm with generic acceptance function named GAF-MPM in this section. Finally, several different acceptance function forms are given. The proof of Detailed Balance Condition will be shown in next section.

GAR-MPM. Consider accepting function $\alpha(x,y)$ has the following form (where $k\in\{1,2,...,n\}$):

$$\begin{cases}\alpha(x,y)=\beta(x,y)\gamma(x,y\mid x_{-k}^{*},y_{-k})\\(0\le\alpha(x,y)\le 1)\end{cases}\qquad(0.1)$$

In the above equation, functions and parameter need to satisfy the following equations:

$$\begin{cases}p(x)\pi_k(y\mid x,y_{1:k-1})\beta(x,y)\\=p(y)\pi_k(x\mid y,x_{k:1}^{*})\beta(y,x)\\(0\le\beta(x,y)\le 1)\end{cases}\qquad(0.2)$$

$$\begin{cases}\varpi_y\gamma(x,y\mid x_{-k}^{*},y_{-k})=\varpi_x\gamma(y,x\mid y_{-k},x_{-k}^{*})\\(0\le\gamma(x,y\mid x_{-k}^{*},y_{-k})\le 1)\end{cases}\qquad(0.3)$$

Under the condition of function $\alpha(x,y)$ satisfy equation (2.1-2.3), we can select arbitrary acceptance. We will give some examples of acceptance function $\alpha(x,y)$ later in this section. The proof will be given in the next section. Before the detailed describe of algorithm GAF-MPM, some equations are necessary to be given. The joint proposal distribution T is

$$T(y_{[1:j]}\mid x)=T(y_1\mid x)\times...\times T(y_j\mid x,y_{[1:j-1]})\quad(0.4)$$

Where, $\begin{cases}y_{[1:j]}=(y_1,...y_j)\\y_{[j:1]}=(y_j,...y_1)\\y_{-k}=(y_1,...,y_{k-1},y_{k+1},...,y_n)\end{cases}$,

The weight function is given in the following: $\omega_j(x,y_{[1:j]})=\pi(x)T(y_{[1:j]}\mid x)\lambda_j(x,y_{[1:j]})$ (0.5)

λ_j meets the nature of the bounded, non-negative, sequence symmetric. Specially, the sequentially symmetric function satisfy the property $\lambda_j(x,y_{[1:j]})=\lambda_j(y_{[1:j]},x)$. The most simple case is $\lambda_j=1$.

After definite conditions given above, pseudo code of our proposed algorithm GAF-MPM is shown, where, assuming that the t-th iteration generates sample $x^{(t)}$. The detailed steps are shown below.

Examples of acceptance function $\alpha(x,y)$. Now, we consider to design acceptance function $\alpha(x,y)=D\times D\to[0,1]$, which satisfy the conditions (2.1-2.3). One way to get $\alpha(x,y)$, is to design functions β and γ, which meet the conditions, respectively.

1) Examples of function β. Consider to design function β, which satisfies condition (2.2). Firstly, define $R(x,y)=p(y)\pi_k(x\mid y,x_{-k}^{*})/p(x)\pi_k(y\mid x,y_{-k})$ (0.6).

Consider function $F(\upsilon):\mathbb{R}^{+}\to[0,1]$, which satisfies $F(\upsilon)=\upsilon F(1/\upsilon)$.

GAR-MPM algorithm:

step1: Assume the last iteratio n generates samples $x^{(t)}$, let $x = x^{(t)}$. Then generat e n candidate samples $y_1,...,y_n$ from the joint proposal distribution T (2.4);

step2: Compute weight function $\omega_j(x, y_{[1:j]})$, then normalize them to get $\varpi_j(x, y_{[1:j]})$, $j = 1,...,n$;

step3: According weight function $\varpi_j(x, y_{[1:j]})$, $j = 1,...,n$, choose t he sample point y from the candidate samples $y_1,...,y_n$. Assume $y = y_k$ is chosen, let $\varpi_y = \varpi_k = \frac{\omega_k(x, y_{k:1})}{\sum_{i=1}^{n} \omega_i(x, y_{i:1})}$;

step4: Get samples $\{x^*_{k+1},...,x^*_n\}$, by sampling x^*_{j+1} $j = k,...,n-1$, from $T(\cdot \mid y_{[k:1]}, x^*_{[k+1:j]})$, specifically, in T, $x^*_j = y_{k-j}$, $j = 1,...,k-1$, $x^*_k = x$;

step5: Compute $\omega_j(y, x^*_{j:1})$ $j = 1,...,n$, normalize them to get $\varpi_j(y, x^*_{j:1})$,

specifically, $\varpi_x = \frac{\omega_k(y, x_{k:1})}{\sum_{i=1}^{n} \omega_i(y, x_{i:1})}$;

step6: Accept $x^{(t+1)} = y$ with probability $\alpha(x, y)$,

specifically, $\alpha(x, y) = \min[1, \frac{p(y)\pi(x^*_{1:k} \mid y)}{p(x)\pi(y_{1:k} \mid x)} \cdot \frac{\varpi_x}{\varpi_y}]$, otherwise, accept $x^{(t+1)} = x$ with probability $1-\alpha$; $t = t+1$, turn into 2. to do the next iteration.

Further, we define universal function such as $\beta_g(x, y) = (F \circ R)(x, y) = F(R(x, y))$. When $F(\upsilon) = \min[1, \upsilon]$, classical M-H algorithm is gotten: $\beta_1(x, y) = \min[1, \frac{p(y)\pi_k(x \mid y, x^*_{-k})}{p(x)\pi_k(y \mid x, y_{-k})}]$ (0.7)

When $F(\upsilon) = \upsilon / (1+\upsilon)$, we can get [13] $\beta_2(x, y) = \frac{p(y)\pi_k(x \mid y, x^*_{-k})}{p(x)\pi_k(y \mid x, y_{-k}) + p(y)\pi_k(x \mid y, x^*_{-k})}$ (0.8)

Table 1 lists several other alternative functions β. (specifically, $\lambda(x, y)$ is a Non-negative symmetric function. That is to say, $\lambda(x, y) \geq 0$ and $\lambda(x, y) = \lambda(y, x)$ for all $(x, y) \in D \times D$.)

2) Examples of function γ. Consider to design function $\gamma(x, y \mid x^*_{-k}, y_{-k})$, which satisfies condition (2.3). Apparently, we can choose $\gamma_1(x, y \mid x^*_{-k}, y_{-k}) = \varpi_x$, then get $\varpi_y \varpi_x = \varpi_x \varpi_y$ which satisfies (2.3). Table 1 lists several other γ functions.

Table 1 Examples of function $\beta(x, y)$ and $\gamma(x, y \mid x^*_{-k}, y_{-k})$

function $\beta(x, y)$	*function* $\gamma(x, y \mid x^*_{-k}, y_{-k})$
$\beta_1(x, y) = \min[1, \frac{p(y)\pi_k(x \mid y, x^*_{-k})}{p(x)\pi_k(y \mid x, y_{-k})}]$ [14]	$\gamma_1(x, y \mid x^*_{-k}, y_{-k}) = \varpi_x$
$\beta_2(x, y) = \frac{p(y)\pi_k(x \mid y, x^*_{-k})}{p(x)\pi_k(y \mid x, y_{-k}) + p(y)\pi_k(x \mid y, x^*_{-k})}$ [13]	$\gamma_2(x, y \mid x^*_{-k}, y_{-k}) = \frac{\varpi_x}{\varpi_x + \varpi_y}$
$\beta_3(x, y) = \lambda(x, y) \Big/ (1 + \frac{p(y)\pi_k(x \mid y, x^*_{-k})}{p(x)\pi_k(y \mid x, y_{-k})})$ [4, 15]	$\gamma_3(x, y \mid x^*_{-k}, y_{-k}) = \min[1, \frac{\varpi_x}{\varpi_y}]$

Proof of Detailed Balance Condition

The following theorem shows that the Markov chain constructed by GAF-MPM algorithm is reversible, and its target distribution $h(x)$ is stationary distribution.

Theorem 1[16]. If a Markov chain has reversible initial distribution $h(x)$, with proposal transition function $P(x, y)$, its stationary distribution is $h(x)$. Theorem 2[15, 17]. The Convergence of Markov chain is equivalent to its detailed balance condition. That is if a Markov chain has stationary distribution $h(x)$, then detailed balance condition $h(y)P(y,x)=h(x)P(x,y)$ is satisfied. Specifically, $P(x,y)$ is proposal transition function, and $h(x)$ is reversible initial distribution.

According to Theorem 2, in order to ensure that the MCMC Markov chain converges, we should prove that its transition f unction satisfies the following equation $p(x)A(y\mid x)=p(y)A(x\mid y)$ (1.1)

Markov chain constructed by GAR-MPM with di fferent acceptance function, satisfies the Detailed Balance Condition. the following is the proof. when $x=y$, $A(y\mid x)$ is constant times of delta function $\delta(y-x)$. Satisfied. When $x\neq y$, we can get the proposal transition function $A(y\mid x)$, $A(y=y_k\mid x)=\sum_{i=1}^{n}h(y=y_k\mid x,k=i)$ (1.2). Specifically, $h(y=y_k\mid x,k=i)$ is acceptance ratio of $x_{t+1}=y_k$, when $x_t=x$. Because $A(y\mid x)$ is sum of symmetric function $h(\cdot)$, so if $h(y=y_k\mid x,k=i)$ satisfies detailed balance condition, then $A(y\mid x)$ satisfies, too. Therefore, we only need to prove the following equation satisfies, for arbitrary $k\in\{1,...,n\}$. $p(x)h(y\mid x,k)=p(y)h(x\mid y,k)$ (1.3)

According GAP-MPM algorithm step by step, every integration factor corresponds to one step in the algorithm. Consider when $y=y_k$, and rearrange the order of factor, then do further calculations on the above equation, we can get,

$$
\begin{aligned}
&p(x)h(y\mid x,k)\\
&=\int\!\ldots\!\int[\prod_{j=1;j\neq k}^{n}\pi_j(y_j\mid x,y_{1:j-1})]\cdot[\prod_{j=1;j\neq k}^{n}\pi_j(x_j^*\mid y_{k:1},x_{k+1:j-1}^*)]\cdot\\
&\cdot p(x)\pi_k(y\mid x,k_{1:k-1})\beta(x,y)\varpi_y\cdot\\
&\cdot\gamma(x,y\mid x_{-k}^*,y_{-k})dy_{1:k-1}dy_{k+1:n}dx_{1:k-1}^*dx_{k+1:n}^*
\end{aligned}
$$

It can be easily seen that the above formula is symmetrical for x and y. Because equations (2.2, 2.3) satisfy, after making change between x and y, also x_j^* and y_j, we get see that the right side of the above equation is an extended form of $p(y)h(x\mid y,k)$.

So, $p(x)h(y\mid x,k)=p(y)h(x\mid y,k)$. Thus, we can conclude that $p(x)A(y\mid x)=p(y)A(x\mid y)$.

Experiment

Consider random variable $x\in\mathbb{R}$, and its Probability density f unction is $p(x)$ [18]. We try to obtain target distribution $p_o(x)$. $p_o(x)\propto p(x)=0.3\exp(-0.2x^2)++0.7\exp(-0.2(x-10)^2)$ (1.4)

We define a Gaussian proposal transition function is $\pi(y\mid x_1,...,x_j)\propto \mathrm{N}(\overline{x},5^2)$, and $\lambda(x,y)=(\frac{\pi(x,y)+\pi(y,x)}{2})^{-1}$ .Consider different acceptance $\alpha(x,y)$ composed of different β,γ (Detail in section 2). $\alpha_{i,j}(x,y)=\beta_i(x,y)\gamma_j(x,y)(i,j=1,2,3)$ (1.5)

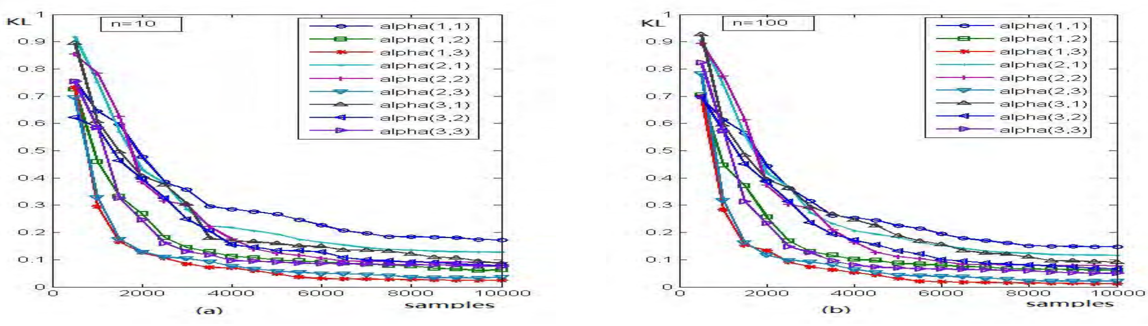

Figure 1 The convergence speed comparison of GAF-MPM algorithm taking different acceptance functions $\alpha(x,y)$, when n=10 and n=100 respectively.

$n=10$ and $n=100$ is selected respectively. Fig. 1 shows the convergence speed com parison of GAF-MPM algorithm taking different acceptance func tions. The horizontal axis represents the sample number, the vertical axis represents the KL distance of the sample and the objective function. It can be seen, when the acceptance function m eets the given conditions, all GAF-MPM algorithm s have a relatively stable convergence. Specially, the algorithm converges fastest with $\alpha_{1,3}(x,y)$.

Furthermore, table 2 shows the experim ent of MAR-MPM with 10000 iterations. The m ean of acceptance rate is the average probability, that new sam ple is accepted, am ong 10000 tim es of sampling. The correlation coefficient of the sam ple points is the linear correlation coefficient for a sample point variable in the Markov chain with the nextsample point variable. It can be seen that the value of n does not have an big effect on the co rrelation coefficient value. Specially, the algorithm take the low and high average acceptance rate with $\alpha_{1,3}(x,y)$ as its acceptance function.

Table 2 The results of GAF-MPM algorithm taking different acceptance functions $\alpha(x,y)$, when n=10 and n=100 respectively. (AR represent Acceptance ratio, and CC is short for Correlation coefficient)

$\alpha(x,y)$	n=10		n=100	
	Mean of AR	CC	Mean of AR	CC
$\alpha_{1,1}(x,y)$	0.11	0.99	0.01	0.99
$\alpha_{1,2}(x,y)$	0.32	0.98	0.33	0.98
$\alpha_{1,3}(x,y)$	**0.55**	**0.97**	**0.59**	**0.97**
$\alpha_{2,1}(x,y)$	0.13	0.99	0.07	0.99
$\alpha_{2,2}(x,y)$	0.23	0.98	0.25	0.98
$\alpha_{2,3}(x,y)$	0.33	0.98	0.35	0.98
$\alpha_{3,1}(x,y)$	0.17	0.98	0.14	0.98
$\alpha_{3,2}(x,y)$	0.21	0.99	0.19	0.99
$\alpha_{3,3}(x,y)$	0.31	0.99	0.37	0.99

Conclusion

In this work, an im proved Multi-point Metropo lis algorithm with generic acceptance function named GAF-MPM, is proposed. Then we show thatGAF-MPM satisfies Detailed Balance Condition to ensure its convergence, the strict proof is give n in this work. Also, we give several different acceptance functions and discuss the effect on the convergence speed, acceptance rate of the samples and the correlation due to the choice of differe nt acceptance functions. Finally, we prove its correctness and effectiveness through num erical experiments. In future, we will try to apply GAF-MPM algorithm to a higher dimensional distribution function of sampling problems and a wide range of practical problem areas.

Acknowledgement

This paper was supported by National Science Foundation of China No.U1135005 and No.90924026, the Fundamental Research of 12th five years Plan Project of China No.0101050302, the Advanced Research Projects under Grant No.5130101, and the Key Projects of Knowledge Innovation Program at CAS under Grant No. Y1S6051S31.

References

[1].Pearl, J., Probabilistic reasoning in intelligent systems: networks of plausible inference1988: Morgan Kaufmann.

[2].Pearl, J., Causality: models, reasoning, and inference. Vol. 47. 2000: Cambridge Univ Press.

[3].MacKay, D.J.C., Introduction to monte carlo methods. NATO ASI SERIES D BEHAVIOURAL AND SOCIAL SCIENCES, 1998. 89: p. 175-204.

[4].Hastings, W.K., Monte Carlo sam pling methods using Markov chains and their applications. Biometrika, 1970. 57(1): p. 97-109.

[5].Casarin, R., R. Craiu, and F. Leisen, Inter acting multiple try algorithms with different proposal distributions. Statistics and Computing, 2011p1-16.

[6].Pandolfi, S., F. Bartolucci , A generalization of the Multiple-try Metropolis algorithm for Bayesian estimation and model selection. 2010.

[7].Liu, J.S., The multiple-try method and local optimization in Metropolis sampling. Journal of the American Statistical Association, 2000: p 121-134.

[8].Martino, L. and J. Read, Different acceptance functions for Multiple Try Metropolis schem es. Arxiv preprint arXiv:1201.0646 2012

[9].Qin, Z.S. and J.S. Liu, Multipoint m etropolis method with application to hybrid Monte Carlo. J. Comput. Phys., 2001. 172(2): p. 827-840.

[10].Martino, L., V.P. Del Olmo, and J. Read, A multi-point Metropolis scheme with generic weight functions. Arxiv preprint arXiv:1112.4048, 2011.

[11].Neal, R.M., An im proved acceptance procedure for the hybrid Monte Carlo algorithm . Arxiv preprint hep-lat/9208011, 1992.

[12].Kennedy, A. Acceptances and autocorrelati ons in hybrid Monte Carlo. Nuclear Physics B-Proceedings Supplements, 1991. 20: p. 118-121.

[13].Barker, A., Monte Carlo Calculations of the Radial Distribution Functions for a Proton-Electron Plasma. Australian Journal of Physics, 1965.

[14].Metropolis, N., et al., Equation of state calculations by fast computing machines. The journal of chemical physics, 1953. 21: p. 1087.

[15].Robert, C.P. and G. Casella, Monte Carlo statistical methods2004: Springer Verlag.

[16].Gardiner, C.W., Stochastic methods1985: Springer.

[17].Liu, J.S., Monte Carlo strategies in scientific computing2008: Springer Verlag.

[18].Andrieu, C., et al., An introduction to MCMC for machine learning. Machine learning, 2003. 50(1): p. 5-43.