Far East Journal of Applied Mathematics
Volume 30, Number 3, March 2008, Pages 285-290
Published online: March 30, 2008
This paper is available online at http://www.pphmj.com

INCREASING THE EFFICIENCY OF THE MONTE CARLO METHOD AND REDUCING THE LENGTH OF MARKOV CHAINS FOR SOLVING LINEAR SYSTEMS

BEHROUZ FATHI VAJARGAH

Department of Mathematics
University of Guilan, Iran
e-mail: fathi@guilan.ac.ir

Abstract

The numerical Monte Carlo method for finding the solution of a linear system has the cost of computations based on the necessary length of applied Markov chains. In this paper, by resuming the previous work [4], we show under the covering property in our Markov chain that we can increase the efficiency of the Monte Carlo method in comparison to the standard Monte Carlo method introduced previously.

1. Introduction

As in the previous paper [4], we consider the linear system

$$Bx = f, \tag{1}$$

where B is a given nonsingular matrix, $f = (f_1, f_2, \ldots, f_n)^t$ is a known vector and $x = (x_1, x_2, \ldots, x_n)^t$ is the solution vector that we are looking for. If we consider matrix $M^{n\times n}$ such that $MB = I - T$, $Mf = c$, then

2000 Mathematics Subject Classification: 34A30.

Keywords and phrases: Monte Carlo, covering Markov chain, efficiency, length of Markov chain.

Received June 10, 2007

the linear system (1) is converted to

$$x^{(k+1)} = Tx^{(k)} + c, \tag{2}$$

where $T^{n\times n}$ is a given nonsingular matrix. It is well known that

$$\rho(T) \leq \| T \| < 1, \tag{3}$$

where $\rho(T)$ is the spectral radius of matrix T and $\| T \|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} | t_{ij} |$, the $x^{(k)}$ tends to the exact solution $x = (I - T)^{-1}c$ [1]. The inner product of x and $h = (h_1, h_2, \ldots, h_n) \in R^n$ is defined by $\langle h, x\rangle = h_1x_1 + h_2x_2 + \cdots + h_nx_n$. This inner product when $h = (\underbrace{0, 0, \ldots, 0, 1}_{i}, 0, \ldots, 0)$ is x_j, i.e., $\langle h, x\rangle = x_j$, which is the ith element of the solution vector. With Markov chain $i_0 \to i_1 \to \cdots \to i_k$ of the sample state $S = \{1, 2, \ldots, n\}$ which it will select the rows of columns indices of the matrix A, with initial distribution $p_{i_0} = P$ (Markov chain starts at $i_0 \in S$), and for $m \in \{1, 2, \ldots, n-1\}$ one step transition probability function is $p_{ij} = P(i_{m+1} = j \,|\, i_m = i)$ [2]. For solving the linear system (1) by Monte Carlo method, we consider transition probability matrix $P = [p_{ij}]_{i,j=1}^{n}$. Under the following conditions:

$$\begin{cases} \sum_{i=1}^{n} p_i = 1, & \\ \sum_{j=1}^{n} p_{ij} = 1, & i = 1, 2, \ldots, n \\ p_i \geq 0,\ p_{ij} \geq 0, & i, j = 1, 2, \ldots, n \\ h_i \neq 0 \to p_i > 0, & i = 1, 2, \ldots, n \\ T_{ij} \neq 0 \to p_{ij} > 0, & i, j = 1, 2, \ldots, n. \end{cases}$$

We define $W_m = \dfrac{T_{i_0 i_1} T_{i_1 i_2} \cdots T_{i_{m-1} i_m}}{p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m}}$ and $W_m = W_{m-1} \dfrac{T_{i_{m-1} i_m}}{p_{i_{m-1} i_m}}$, $W_0 = 1$, $\eta_k(h) = \dfrac{h_{i_0}}{p_{i_0}} \sum_{m=0}^{k} W_m b_{i_m}$. It has been proved that $E[\eta_k(h)] = \left\langle h, \sum_{m=0}^{k} T^m b \right\rangle = \langle h, x^{(k+1)} \rangle$ [4]. Since $\eta_k(h)$ is an unbiased estimator of $\langle h, x^{(k+1)} \rangle$, we can introduce the Monte Carlo estimator based on $i_0^{(s)} \to i_1^{(s)} \to i_2^{(s)} \to \cdots \to i_k^{(s)}$, $s = 1, 2, \ldots, N$. We have $\Theta_k(h) = \dfrac{1}{N} \sum_{s=1}^{N} \eta_k^{(s)}(h) \approx \langle h, x^{(k+1)} \rangle$, where for $s = 1, 2, \ldots, N$, $\eta_k^{(s)}(h) = \dfrac{h_{i_0^{(s)}}}{p_{i_0^{(s)}}} \sum_{m=0}^{k} W_m^{(s)} b_{i_m}$, with $W_m^{(s)} = \dfrac{T_{i_0^{(s)} i_1^{(s)}} T_{i_1^{(s)} i_2^{(s)}} \cdots T_{i_{m-1}^{(s)} i_m^{(s)}}}{p_{i_0^{(s)} i_1^{(s)}} p_{i_1^{(s)} i_2^{(s)}} \cdots p_{i_{m-1}^{(s)} i_m^{(s)}}}$, $W_0^{(s)} = 1$ and $\Theta_k(h) = \dfrac{1}{N} \sum_{s=1}^{N} \eta_k^{(s)}(h)$ [3]. $E[\eta_\infty(h)] = \lim_{k \to \infty} E[\eta_k(h)] = \lim_{k \to \infty} \langle h, x^{(k+1)} \rangle = \langle h, x \rangle$. $\Theta_k(e_j) = \dfrac{1}{N} \sum_{s=1}^{N} \eta_k^{(s)}(e_j) \approx x_j^{(k+1)} \approx x_j$. For obtaining the element x_j of solution vector x, we simulate N random paths $j \to i_1^{(s)} \to i_2^{(s)} \to \cdots \to i_k^{(s)}$, $s = 1, 2, \ldots, N$. Then we have $\eta_k^{(s)}(e_j) = \sum_{m=0}^{k} W_m^{(s)} b_{i_m}^{(s)}$ with $W_m^{(s)} = \dfrac{T_{j i_1^{(s)}} T_{i_1^{(s)} i_2^{(s)}} \cdots T_{i_{m-1}^{(s)} i_m^{(s)}}}{p_{j i_1^{(s)}} p_{i_1^{(s)} i_2^{(s)}} \cdots p_{i_{m-1}^{(s)} i_m^{(s)}}}$, $W_0^{(s)} = 1$. The above Monte Carlo method is called *standard* (or *basic*) *Monte Carlo method*.

2. Optimized Monte Carlo Method

It has been discussed in [4] that how we can reduce the nN random paths to N paths. We just have a quick review of the concept of this aim.

Definition 1. For a Markov chain with sample state $S = \{1, 2, \ldots, n\}$, the random path $i_0^{(s)} \to i_1^{(s)} \to \cdots \to i_R^{(s)}$ is called *covering* if each state $j \in S$ can be visited at least once.

Definition 2. Let $j \in S$ be an arbitrary state of a Markov chain. Then the first time that Markov chain visit j, i.e., $R_j = \min_t\{t : i_t = j\}$, is called the *hitting time to state j*. In this way, for $j = 1, 2, \ldots, n$, we can reach to the all unbiased estimators $\eta_k(e_j)$. We first consider the covering path: $i_0 \to i_1 \to \cdots \to i_{R_j} \to \cdots \to i_R \to \cdots \to i_{R+k}$, where $R = \max\limits_{1\le j\le n}\{R_j\} = \max\limits_{1\le j\le n}\min\{t : i_t = j\}$ and k is an integer number. For every $j = 1, 2, \ldots, n$, we consider a subpath with length k as $i_{R_j} \to i_{R_j+1} \to \cdots \to i_{R_j+k}$, now we set $\tilde{\eta}_k(e_j) = \sum\limits_{m=R_j}^{R_j+k} W_m b_{i_m}$, where $W_m = \dfrac{T_{i_{R_j}i_{R_j+1}} T_{i_{R_j+1}i_{R_j+2}} \cdots T_{i_{m-1}i_m}}{p_{i_{R_j}i_{R_j+1}} p_{i_{R_j+1}i_{R_j+2}} \cdots p_{i_{m-1}i_m}}$, $W_{R_j} = 1$.

Theorem 1. *Under the above conditions, we have* $E[\tilde{\eta}_k(e_j)] = x_j^{(k+1)}$ [4].

Theorem 2. $\operatorname{var}[\Theta_k(e_j)] = \operatorname{var}[\tilde{\Theta}_k(e_j)]$, *where* $\Theta_k(e_j) = \dfrac{1}{N}\sum\limits_{s=1}^{N} \eta_k^{(s)}(e_j)$ *and* $\tilde{\Theta}_k(e_j) = \dfrac{1}{N}\sum\limits_{s=1}^{N} \tilde{\eta}_k^{(s)}(e_j)$ *for* $j = 1, 2, \ldots, n$ [4].

3. Increasing the Efficiency of the Method

In this section, we compare the efficiency of two methods and discuss here based on [3, 4]. For simplicity of this comparison and bringing them to our analysis, the method of the Monte Carlo without covering property (usual Monte Carlo method) is considered as method (1) and the Monte Carlo method based on covering property is considered as method (2). We remember that the method (1) has $\Theta_k(e_j) = \dfrac{1}{N}\sum\limits_{s=1}^{N} \eta_k^{(s)}(e_j)$, as the Monte Carlo estimator and the method (2) has $\tilde{\Theta}_k(e_j) = \dfrac{1}{N}\sum\limits_{s=1}^{N} E[(\tilde{\eta}_k^{(s)}(e_j))^2(e_j)$ as the Monte Carlo estimator for x_j of the solution vector x.

Definition 1. Let $\Theta_k(e_j)$ and $\widetilde{\Theta}_k(e_j)$ are two Monte Carlo estimators for the parameter θ. Then the *efficiency* of $\widetilde{\Theta}_k(e_j)$ respect to $\Theta_k(e_j)$ is defined by $\varepsilon = \dfrac{t\, MSE[\Theta_k(e_j)]}{\widetilde{t}\, MSE[\widetilde{\Theta}_k(e_j)]}$, where t, $\widetilde{t}$ show the necessary length of Markov chains to reach the estimation for x_j. We recall that for any estimator Y of θ, we have $MSE[Y] = E[Y-\theta]^2 = \text{var}[Y] + [E[Y]-\theta]^2$. Then we have

Definition 2. For two unbiased estimators $\Theta_k(e_j)$ and $\widetilde{\Theta}_k(e_j)$ of x_j, the *efficiency of* $\widetilde{\Theta}_k(e_j)$ respect to $\Theta_k(e_j)$ is defined by $\varepsilon = \dfrac{t\,\text{var}(\Theta_k(e_j))}{\widetilde{t}\,\text{var}(\widetilde{\Theta}_k(e_j))}$, where t and $\widetilde{t}$ are the necessary lengths of Markov chains to reach the Monte Carlo estimation for x_j using methods (1) and (2), respectively.

Without loss of generality, we consider $N = 1$, then with regards to Theorem 1, $\text{var}(\Theta_k(e_j)) = \text{var}(\widetilde{\Theta}_k(e_j))$ therefore, we have $\varepsilon = \dfrac{t}{\widetilde{t}}$. In this case in method (1), we use n paths with length k, then the total lengths used in these paths is equal to nk. But, in method (2), we use only one path with length $(\max\limits_{1\le j\le n}\{R_j\}) + k$ with average $E[\max\limits_{1\le j\le n}\{R_j\}] + k$. For $n > 1$, we prove that the inequality $nk > E[\max\limits_{1\le j\le n}\{R_j\}] + k$ is valid. To prove this inequality we prove that $nk - k > E[\max\limits_{1\le j\le n}\{R_j\}]$ or $k > E[\max\limits_{1\le j\le n}\{R_j\}]/n-1$. It proves that whenever k is bigger than $E[\max\limits_{1\le j\le n}\{R_j\}]/n-1$, the method (2) is more efficient than the method (1). Selecting a suitable starting point of the Markov chain and using in method (2), we can reduce the cost of computation. In needs just, we select the starting point of chain $i_0 = l$ such that $R = E[\max\limits_{1\le j\le n}\{R_j\} | i_0 = l]$ $= \min\limits_{1\le i_0\le n} E[\max\{R_j\} | i_0 = l]$. In this manner, we set $p_{i_0} = \begin{cases} 1 & i_0 = l \\ 0 & i_0 \ne l \end{cases}$. Then we prove the following theorem:

Theorem 3. *In a covering Markov chain* (*as discussed above*), *if* $k > E[\max_{1\le j\le n}\{R_j\}]/n - 1$, *then the covering method is more efficient than the standard Monte Carlo method.*

4. Conclusion

The method (2) in Theorem 3 is more efficient than the basic Monte Carlo method. Then to reduce the Monte Carlo computation using covering Markov chain we use the second algorithm discussed in [4] and this paper.

References

[1] R. L. Burden, J. D. Faires and A. C. Reynold, Numerical Analysis, Prindle, Weber and Schmidt, Boston, Massachusetts, USA, 1981.

[2] E. Cinlar, Introduction to Stochastic Processes, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[3] I. M. Sobol, Monte Carlo Numerical Methods, Nauka, Moscow, 1973 (in Russian).

[4] B. Fathi Vajargah and M. Saraji, Significant reduction of the number of Markov chains in Monte Carlo computation, Far East J. Appl. Math. 26(1) (2007), 73-83.

■