Old Dominion University
ODU Digital Commons

Computer Science Faculty Publications | Computer Science

2013

Convergence Analysis of Markov Chain Monte Carlo Linear Solvers Using Ulam--von Neumann Algorithm

Hao Ji
Old Dominion University

Michael Mascagni

Yaohang Li
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs

 Part of the Computer Sciences Commons, and the Mathematics Commons

Repository Citation

Ji, Hao; Mascagni, Michael; and Li, Yaohang, "Convergence Analysis of Markov Chain Monte Carlo Linear Solvers Using Ulam--von Neumann Algorithm" (2013). *Computer Science Faculty Publications*. 66.
https://digitalcommons.odu.edu/computerscience_fac_pubs/66

Original Publication Citation

Ji, H., Mascagni, M., & Li, Y. (2013). Convergence analysis of Markov chain Monte Carlo linear solvers using Ulam--von Neumann Algorithm. *SIAM Journal on Numerical Analysis, 51*(4), 2107-2122. doi: 10.1137/130904867

This Article is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

SIAM J. NUMER. ANAL.
Vol. 51, No. 4, pp. 2107–2122
© 2013 Society for Industrial and Applied Mathematics

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

CONVERGENCE ANALYSIS OF MARKOV CHAIN MONTE CARLO LINEAR SOLVERS USING ULAM–VON NEUMANN ALGORITHM*

HAO JI†, MICHAEL MASCAGNI‡, AND YAOHANG LI†

Abstract. The convergence of Markov chain–based Monte Carlo linear solvers using the Ulam–von Neumann algorithm for a linear system of the form $x = Hx + b$ is investigated in this paper. We analyze the convergence of the Monte Carlo solver based on the original Ulam–von Neumann algorithm under the conditions that $\|H\| < 1$ as well as $\rho(H) < 1$, where $\rho(H)$ is the spectral radius of H. We find that although the Monte Carlo solver is based on sampling the Neumann series, the convergence of Neumann series is not a sufficient condition for the convergence of the Monte Carlo solver. Actually, properties of H are not the only factors determining the convergence of the Monte Carlo solver; the underlying transition probability matrix plays an important role. An improper selection of the transition matrix may result in divergence even though the condition $\|H\| < 1$ holds. However, if the condition $\|H\| < 1$ is satisfied, we show that there always exist certain transition matrices that guarantee convergence of the Monte Carlo solver. On the other hand, if $\rho(H) < 1$ but $\|H\| \geq 1$, the Monte Carlo linear solver may or may not converge. In particular, if the row sum $\sum_{j=1}^{N} |H_{ij}| > 1$ for every row in H or, more generally, $\rho(H^+) > 1$, where H^+ is the nonnegative matrix where $H^+_{ij} = |H_{ij}|$, we show that transition matrices leading to convergence of the Monte Carlo solver do not exist. Finally, given H and a transition matrix P, denoting the matrix H^* via $H^*_{ij} = H^2_{ij}/P_{ij}$, we find that $\rho(H^*) < 1$ is a necessary and sufficient condition for convergence of the Markov chain–based Monte Carlo linear solvers using the Ulam–von Neumann algorithm.

Key words. Markov chain Monte Carlo, linear solver, Ulam–von Neumann algorithm, Neumann series, convergence analysis, transition probability matrix

AMS subject classifications. 65C05, 65Y20, 68Q25, 68W40

DOI. 10.1137/130904867

1. Introduction.

Applying Monte Carlo methods to estimate solutions of linear systems was originally proposed by Ulam and von Neumann and later was described by Forsythe and Leibler in [1]. Consider a linear system in a somewhat suggestive form

$$x = Hx + b, \tag{1.1}$$

where H is an $N \times N$ nonsingular matrix, b is the given constant vector, and x is vector of unknowns. The fundamental idea of the Ulam–von Neumann algorithm is to construct a discrete Markov chain by using random walks on the indices of the matrix with an extra absorbing state in order to sample the solution to (1.1) as developed via the Neumann series. The transition probabilities of the random walks are defined by a transition probability matrix P satisfying the conditions

*Received by the editors January 7, 2013; accepted for publication (in revised form) May 21, 2013; published electronically July 16, 2013. This work was partially supported by NSF grant 1066471, an ODU 2011 SEECR grant, and an ODU 2013 multidisciplinary seed grant.
http://www.siam.org/journals/sinum/51-4/90486.html

†Department of Computer Science, Old Dominion University, Norfolk, VA 23529 (hji@cs.odu.edu, yaohang@cs.odu.edu).

‡Departments of Computer Science, Mathematics, and Scientific Computing, Florida State University, Tallahassee, FL 32306 (mascagni@fsu.edu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\begin{aligned} P_{ij} &\geq 0, \\ \sum_j P_{ij} &\leq 1, \\ H_{ij} \neq 0 &\rightarrow P_{ij} \neq 0 \end{aligned} \tag{1.2}$$

with the absorption (termination) probability T_i at row i is defined as

$$T_i = 1 - \sum_j P_{ij}. \tag{1.3}$$

For a random walk, γ_k,

$$\gamma_k : r_0 \rightarrow r_1 \rightarrow r_2 \rightarrow \cdots \rightarrow r_k \tag{1.4}$$

starting at state r_0 and terminating at r_k, define the estimator

$$X(\gamma_k) = \frac{H_{r_0 r_1} H_{r_1 r_2} \cdots H_{r_{k-1} r_k}}{P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k}} b_{r_k} / T_{r_k}. \tag{1.5}$$

The fundamental idea of the Ulam–von Neumann algorithm is to statistically sample the Neumann series representation

$$I + H + H^2 + H^3 + \cdots \tag{1.6}$$

of the linear system. Denoting $\|.\|$ as the L-∞ norm, as specified in the Monte Carlo linear solver literature [9], if $\|H\| < 1$, the Neumann series converges to $(I - H)^{-1}$ and hence $X(\gamma_k)$ is an unbiased estimator of $\left(H^k b\right)_{r_0}$, the r_0 element of the product of the kth power of H with the right-hand-side vector b, while $\sum_{k=1}^{\infty} X(\gamma_k) P(\gamma_k)$ equals the solution $(x)_{r_0}$.

As pointed out in [9, 10], if $\|H\| > 1$, the Monte Carlo method breaks down. Nevertheless, it is well known that the necessary and sufficient condition for the Neumann series to converge is $\rho(H) < 1$, where $\rho(H)$ is the spectral radius of H. As shown in Proposition 1.1, $\|H\| < 1$ is a stricter condition than $\rho(H) < 1$. Therefore, there exists a family of matrices whose corresponding Neumann series converge ($\rho(H) < 1$) but that the Monte Carlo linear solver will fail. To the best of our knowledge, in the literature of Monte Carlo methods for linear systems, there is a lack of in-depth analysis on the behavior of Markov chain–based Monte Carlo linear solvers under the condition $\rho(H) < 1$.

PROPOSITION 1.1. *For an $N \times N$, nonsingular matrix H, $\rho(H) \leq \|H\|$.*

Proof. Let λ be an eigenvalue of H and y the corresponding eigenvector. Thus $\lambda y = Hy$, and $\|\lambda y\| = \|\lambda\| \|y\| = \|Hy\| \leq \|H\| \|y\|$. Finally, $\|\lambda\| \leq \|H\|$ for all eigenvalues of H and therefore $\rho(H) \leq \|H\|$, since $\rho(H)$ is the largest absolute value of the eigenvalues of H. ∎

In this paper, we investigate the conditions for convergence of the Monte Carlo linear solver using the original Ulam–von Neumann algorithm. We start by considering a set of suggestive examples with 2×2 matrices in section 2 to study the behavior of the Monte Carlo linear solver. Then, in section 3 we analyze the role of transition matrix P in the Monte Carlo linear solver for H under various conditions. In section 4, we derive a necessary and sufficient condition that determines convergence of the Monte Carlo linear solver. Moreover, robustness, balancing, applicability, and potential advantages of Monte Carlo linear solvers are discussed in section 5. Finally, section 6 summarizes our conclusions.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

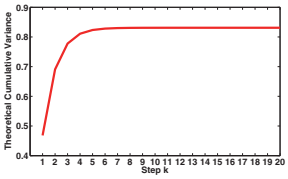
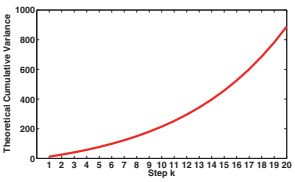
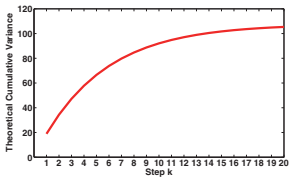
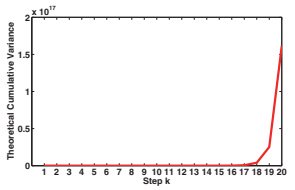
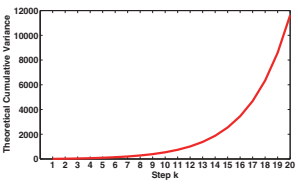
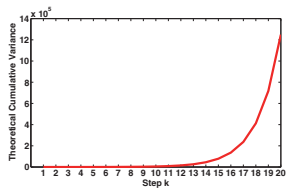
Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. Suggestive examples. Table 2.1 shows the behavior of the Monte Carlo linear solvers using the Ulam–von Neumann algorithm in six separate cases of 2×2 matrices under different conditions and different transition matrices. In all these cases, the H matrices satisfy the spectral radius condition of $\rho(H) < 1$; however, the Monte Carlo linear solver does not converge in all these cases. Hence, it is clear that

TABLE 2.1

Behavior of the Monte Carlo linear solvers using the Ulam–von Neumann algorithm in six cases of 2×2 matrices under different conditions and transition matrices. H^+ is an $N \times N$ matrix where $H^+_{ij} = |H_{ij}|$ and H^* is an $N \times N$ matrix where $H^*_{ij} = H^2_{ij}/P_{ij}$ given H and P.

Case	H and P	Conditions	Converged?	$Var\left(\sum_k X(\gamma_k)\right)$
1	$H = \begin{bmatrix} 0.1 & 0.3 \\ 0.3 & -0.05 \end{bmatrix}$ $P = \begin{bmatrix} 0.1 & 0.3 \\ 0.3 & 0.05 \end{bmatrix}$	$\|H\| < 1$ $\rho(H) < 1$ $\rho(H^+) < 1$ $\rho(H^*) < 1$	Yes	
2	$H = \begin{bmatrix} 0.1 & 0.3 \\ 0.3 & -0.05 \end{bmatrix}$ $P = \begin{bmatrix} 0.009 & 0.891 \\ 0.8 & 0.1 \end{bmatrix}$	$\|H\| < 1$ $\rho(H) < 1$ $\rho(H^+) < 1$ $\rho(H^*) > 1$	No	
3	$H = \begin{bmatrix} 0.8 & 0.35 \\ 0.1 & -0.01 \end{bmatrix}$ $P = \begin{bmatrix} 0.8 & 0.1 \\ 0.7 & 0.2 \end{bmatrix}$	$\|H\| > 1$ $\sum_{j=1}^N \|H_{ij}\| > 1$ for some but not all i $\rho(H) < 1$ $\rho(H^+) < 1$ $\rho(H^*) < 1$	Yes	
4	$H = \begin{bmatrix} 0.8 & 0.35 \\ 0.1 & -0.01 \end{bmatrix}$ $P = \begin{bmatrix} 0.1 & 0.8 \\ 0.7 & 0.2 \end{bmatrix}$	$\|H\| > 1$ $\sum_{j=1}^N \|H_{ij}\| > 1$ for some but not all i $\rho(H) < 1$ $\rho(H^+) < 1$ $\rho(H^*) > 1$	No	
5	$H = \begin{bmatrix} 0.4012 & 0.5305 \\ 0.5305 & -0.7023 \end{bmatrix}$ $P = \begin{bmatrix} 0.3306 & 0.5694 \\ 0.3303 & 0.5697 \end{bmatrix}$	$\|H\| > 1$ $\sum_{j=1}^N \|H_{ij}\| > 1$ for some but not all i $\rho(H) < 1$ $\rho(H^+) > 1$ $\rho(H^*) > 1$	No	
6	$H = \begin{bmatrix} 0.3968 & -0.7162 \\ -0.7162 & -0.6226 \end{bmatrix}$ $P = \begin{bmatrix} 0.2565 & 0.6435 \\ 0.5350 & 0.3650 \end{bmatrix}$	$\|H\| > 1$ $\sum_{j=1}^N \|H_{ij}\| > 1$ for all i $\rho(H) < 1$ $\rho(H^+) > 1$ $\rho(H^*) > 1$	No	

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

the convergence of the underlying Neumann series is not a sufficient condition for the Monte Carlo linear solver to converge. More interestingly, cases 1 and 2 use the same H matrix, where $\|H\| < 1$, but different transition matrices, P. The Monte Carlo linear solver converges in case 1 but diverges in case 2, indicating that the selection of transition matrix P is very important. If P is selected improperly, the Monte Carlo linear solver may diverge even if $\|H\| < 1$ holds. Furthermore, the H matrix in case 3 does not satisfy condition $\|H\| < 1$, but the Monte Carlo linear solver does not break down, which disagrees with the analysis in [9, 10] that if $\|H\| > 1$, the Monte Carlo method breaks down. The phenomenon in case 3 suggests that there are some situations where $\|H\| > 1$ but $\rho(H) < 1$ that permit the Monte Carlo linear solver to still converge, i.e., $\|H\| < 1$ is not a necessary condition for convergence in the Monte Carlo linear solver. Similar to the situation in cases 1 and 2, case 4 has the same H matrix as case 3 but a different transition matrix P, which results in divergence. Cases 5 and 6 show the behavior of the Monte Carlo linear solver under $\rho(H^+) > 1$, when $\sum_{j=1}^{N} |H_{ij}| > 1$, for some but not all i and $\sum_{j=1}^{N} |H_{ij}| > 1$ for all i, respectively. We provide theoretical analysis for the convergence of the Monte Carlo linear solver of these six example cases in the next two sections.

3. Convergence of the Monte Carlo linear solver. We consider a Monte Carlo linear solver as converging if the variance of the estimator $\sum_k X(\gamma_k)$,

$$Var\left(\sum_k X(\gamma_k)\right) = \sum_k Var\left(X(\gamma_k)\right),$$

is bounded as $k \to \infty$, provided that every random walk γ_k is independent. We first investigate the impact of selecting a transition matrix P on the convergence of the Monte Carlo linear solver. For convenience, we state what mathematical results are needed as lemmas. Also note that $Var(X(\gamma_k))$ diverging as $k \to \infty$, implies the same of $Var(\sum_k X(\gamma_k))$. Hence, when we study the convergence/divergence behavior of the Monte Carlo linear solver in the theorems in this section, we only consider $Var(X(\gamma_k))$ instead of $Var(\sum_k X(\gamma_k))$. Without loss of generality and for simplicity, we also assume that the Markov chains in the Monte Carlo linear solver are ergodic and that every element in the constant vector b in the linear system satisfies $b_i \neq 0$ for all i.

LEMMA 3.1. *Consider a vector* $a = (a_1, a_2, \ldots, a_N)^T$ *where at least one element is nonzero,* $a_k \neq 0$.

(i) *For a probability vector* $p = (p_1, p_2, \ldots, p_N)^T$ *satisfying the transition conditions* (1.2)*, the lower bound of* $\sum_k a_k^2/p_k$ *is* $\left(\sum_k |a_k|\right)^2$.

(ii) *There always exists a probability vector such that* $\sum_k a_k^2/p_k \geq c > 1$ *for all* $c > 1$.

Proof.

(i) Because $\sum_k p_k \leq 1$,

$$\begin{aligned}\sum_k \frac{a_k^2}{p_k} = \sum_k \left(\frac{|a_k|}{\sqrt{p_k}}\right)^2 &\geq \left(\sum_k p_k\right)\left(\sum_k \left(\frac{|a_k|}{\sqrt{p_k}}\right)^2\right) \\ &= \left(\sum_k (\sqrt{p_k})^2\right)\left(\sum_k \left(\frac{|a_k|}{\sqrt{p_k}}\right)^2\right) \\ &\geq \left(\sum_k \left(\frac{|a_k|}{\sqrt{p_k}}\sqrt{p_k}\right)\right)^2\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

by the Cauchy–Schwarz inequality, and so we have our result,

$$\sum_k \frac{a_k^2}{p_k} \geq \left(\sum_k \left(\frac{|a_k|}{\sqrt{p_k}} \sqrt{p_k} \right) \right)^2 = \left(\sum_k |a_k| \right)^2 .$$

.

(ii) Select α such that $\alpha \geq \max\left(1, c/\left(N \sum_k a_k^2\right)\right)$ and set $p_k = a_k^2/\left(\alpha \sum_k a_k^2\right)$. Then it is easy to show that $a_k \neq 0 \rightarrow p_k \neq 0$ and $p_k \leq 1$ for each p_k. Also, $\sum_k p_k = 1/\alpha \leq 1$. Moreover, $\sum_k a_k^2/p_k = \alpha N \sum_k a_k^2 \geq c$. ∎

In Lemma 3.1, the lower bound of $\sum_k a_k^2/p_k$ can be achieved by assigning $p_k = |a_k| / \left(\sum_k |a_k|\right)$. The transition probability, P, adopting this probability assignment is called the almost optimal density matrix, which has been used in the Monte Carlo almost optimal (MAO) algorithm [11, 12, 13]. Nevertheless, there is no upper bound for $\sum_k a_k^2/p_k$ over the space of the probability vectors. Theorem 3.2 shows that the divergence of the Monte Carlo linear solver could be ensured by an improper probability scheme.

THEOREM 3.2. *Let H be an $N \times N$, nonsingular matrix with spectral radius $\rho(H) < 1$, b a nonzero vector, and γ_k a random walk starting from r_0 and terminating after k transition steps. Then, there always exists a transition matrix P satisfying the transition conditions* (1.2) *such that $Var(X(\gamma_k))$ diverges as $k \rightarrow \infty$.*

Proof. Since $X(\gamma_k)$ is an unbiased estimator of $\left(H^k b\right)_{r_0}$ and thus $E(X(\gamma_k)) = \left(H^k b\right)_{r_0}$,

$$\begin{aligned} Var(X(\gamma_k)) &= E\left((X(\gamma_k))^2 \right) - (E(X(\gamma_k)))^2 \\ &= \sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_k=1}^{N} P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k} T_{r_k} \\ &\quad \times \left(\frac{H_{r_0 r_1} H_{r_1 r_2} \cdots H_{r_{k-1} r_k} b_{r_k}}{P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k} T_{r_k}} \right)^2 - \left(H^k b\right)_{r_0}^2 \\ &= \sum_{r_1=1}^{N} \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-1}=1}^{N} \frac{H_{r_{k-2} r_{k-1}}^2}{P_{r_{k-2} r_{k-1}}} \sum_{r_k=1}^{N} \frac{H_{r_{k-1} r_k}^2}{P_{r_{k-1} r_k}} \frac{b_{r_k}^2}{T_{r_k}} - \left(H^k b\right)_{r_0}^2 . \end{aligned}$$

Based on Lemma 3.1, for any row $(H_{i,1}, H_{i,2}, \ldots, H_{i,N})$ in H, there exists a probability vector $(P_{i,1}, P_{i,2}, \ldots, P_{i,N})$ and a constant c_i such that the transition probability conditions are satisfied and

$$\sum_{j=1}^{N} \frac{H_{ij}^2}{P_{ij}} \geq c_i > 1.$$

Denoting $c_{min} = \min_i (c_i) > 1$ and $b_{min} = \min_i \left(b_i^2/T_i\right) \neq 0$, then

$$\begin{aligned} Var(X(\gamma_k)) &\geq b_{min} \sum_{r_1=1}^{N} \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-1}=1}^{N} \frac{H_{r_{k-2} r_{k-1}}^2}{P_{r_{k-2} r_{k-1}}} c_{min} - \left(H^k b\right)_{r_0}^2 \\ &\cdots \\ &\geq b_{min} c_{min}^k - \left(H^k b\right)_{r_0}^2 . \end{aligned}$$

Here, $\left(H^k b\right)_{r_0}^2 \rightarrow 0$ when $k \rightarrow \infty$ because $\rho(H) < 1$, and b_{min} is a constant. However, $c_{min}^k \rightarrow \infty$ as $k \rightarrow \infty$ since $c_{min} > 1$. As a result, there exists a transition matrix P such that the $Var(X(\gamma_k))$ diverges as $k \rightarrow \infty$. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The random walk variance $Var\left(X\left(\gamma_k\right)\right)$ is normally used as an indicator for analyzing the convergence and robustness of Monte Carlo linear solvers. (More detailed discussions on the robustness of Monte Carlo solvers can be found in [11, 12].) Based on $Var\left(X\left(\gamma_k\right)\right)$, Theorem 3.2 indicates that selection of the transition matrix P is important for the convergence of the Monte Carlo linear solver using the Ulam–von Neumann algorithm. If P is selected improperly, even when $\|H\| < 1$ holds, the Monte Carlo linear solver can still diverge. This is demonstrated in case 2 in section 2. Fortunately, Theorem 3.4 shows that if $\|H\| < 1$, transition matrices enabling the convergence of the Monte Carlo linear solver are always readily available.

LEMMA 3.3. *Given a vector* $a = (a_1, a_2, \ldots, a_N)^T$ *where* $\sum_k |a_k| < 1$, *there exists a probability vector* $p = (p_1, p_2, \ldots, p_N)^T$ *satisfying the transition probability conditions* (1.2) *and* $0 \leq \sum_k a_k^2/p_k < 1$.

Proof. Simply select $p_k = |a_k|$. Then, $a_k \neq 0 \rightarrow p_k \neq 0$, $p_k = |a_k| \leq \sum_k |a_k| < 1$ for each p_k, and $\sum_k p_k = \sum_k |a_k| < 1$. Also, $\sum_k a_k^2/p_k = \sum_k |a_k| < 1$. ∎

THEOREM 3.4. *Let* H *be an* $N \times N$, *nonsingular matrix where* $\|H\| < 1$, b *a nonzero vector, and* γ_k *a random walk starting from* r_0 *and terminating after* k *transition steps. Then, there always exists a transition matrix* P *satisfying the transition conditions* (1.2) *such that the* $Var\left(X\left(\gamma_k\right)\right) \rightarrow 0$ *and* $Var\left(\sum_k X\left(\gamma_k\right)\right)$ *is bounded as* $k \rightarrow \infty$.

Proof. From Lemma 3.3, since $\|H\| < 1$, i.e., every row $(H_{i,1}, H_{i,2}, \ldots, H_{i,N})$ in H satisfying $\sum_{j=1}^{N} |H_{ij}| < 1$, there exists a probability vector $(P_{i,1}, P_{i,2}, \ldots, P_{i,N})$ for each row i in H that

$$\sum_{j=1}^{N} \frac{H_{ij}^2}{P_{ij}} < 1.$$

Denoting $b_{max} = \max_i \left(b_i^2/T_i\right)$ and $c_{max} = \max_i \left(\sum_{j=1}^{N} H_{ij}^2/P_{ij}\right)$, it is easy to show that $c_{max} < 1$.

Then,

$$\begin{aligned}
&Var\left(X\left(\gamma_k\right)\right) \\
&\quad = E\left(\left(X\left(\gamma_k\right)\right)^2\right) - \left(E\left(X\left(\gamma_k\right)\right)\right)^2 \\
&\quad \leq E\left(\left(X\left(\gamma_k\right)\right)^2\right) \\
&\quad = \sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_k=1}^{N} P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k} T_{r_k} \left(\frac{H_{r_0 r_1} H_{r_1 r_2} \cdots H_{r_{k-1} r_k} b_{r_k}}{P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k} T_{r_k}}\right)^2 \\
&\quad = \sum_{r_1=1}^{N} \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-1}=1}^{N} \frac{H_{r_{k-2} r_{k-1}}^2}{P_{r_{k-2} r_{k-1}}} \sum_{r_k=1}^{N} \frac{H_{r_{k-1} r_k}^2}{P_{r_{k-1} r_k}} \frac{b_{r_k}^2}{T_{r_k}} \\
&\quad \leq b_{max} \sum_{r_1=1}^{N} \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-1}=1}^{N} \frac{H_{r_{k-2} r_{k-1}}^2}{P_{r_{k-2} r_{k-1}}} c_{max} \\
&\quad \leq b_{max} \sum_{r_1=1}^{N} \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-2}=1}^{N} \frac{H_{r_{k-3} r_{k-2}}^2}{P_{r_{k-3} r_{k-2}}} c_{max}^2 \\
&\quad \cdots \\
&\quad \leq b_{max} c_{max}^k.
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Here, b_{max} is a constant, and $c_{max}^k \to 0$ as $k \to \infty$ since $c_{max} < 1$. Therefore, there exists a transition matrix P such that the $Var\left(X\left(\gamma_k\right)\right) \to 0$ as $k \to \infty$.

Furthermore,

$$\lim_{k\to\infty} Var\left(\sum_k X\left(\gamma_k\right)\right) \le \lim_{k\to\infty} \sum_k b_{max} c_{max}^k = \frac{b_{max}}{\left(1 - c_{max}\right)}.$$

Since $b_{max}/\left(1 - c_{max}\right)$ is a constant, $Var\left(\sum_k X\left(\gamma_k\right)\right)$ is bounded as $k \to \infty$, i.e., the Monte Carlo linear solver converges. ∎

Based on Theorem 3.4, a more precise version of "the Monte Carlo method converges if $\|H\| < 1$" is "there is always a transition matrix P enabling the convergence of the Monte Carlo method if $\|H\| < 1$." Such a transition matrix P is also easy to find. Clearly, when $\|H\| < 1$, both the Ulam–von Neumann original selection of P, where $P_{ij} = |H_{ij}|$ [9, 10], and a more popular selection of P, where $P_{ij} = |H_{ij}| / \sum_j |H_{ij}|$ [4, 11, 12, 13], satisfy the condition of $\sum_{j=1}^N H_{ij}^2 / P_{ij} < 1$ for every row i, according to the above proof of Theorem 3.4, which can result in the convergence of the Monte Carlo linear solver.

We now move forward to investigate the behavior of the Monte Carlo linear solver under the conditions where $\rho\left(H\right) < 1$ but $\|H\| \ge 1$. We first study the situation where $\rho\left(H\right) < 1$ but every row sum in H is greater than 1. Clearly, the underlying Neumann series converges due to $\rho\left(H\right) < 1$ in this situation. However, according to Theorem 3.5, the Monte Carlo linear solver cannot converge, regardless of how the transition probability matrix P is chosen. This explains why case 6 described in the previous section diverges.

THEOREM 3.5. *Let H be an $N \times N$, nonsingular matrix with spectral radius $\rho\left(H\right) < 1$ and $\sum_{j=1}^N |H_{ij}| > 1$ for every row i, b a nonzero vector, and γ_k a random walk starting from r_0 and terminating after k transition steps. Then, there does not exist a transition matrix P satisfying the transition conditions* (1.2) *such that the variance $Var\left(X\left(\gamma_k\right)\right)$ converges to zero as $k \to \infty$.*

Proof. Since $\forall i, \sum_{j=1}^N |H_{ij}| > 1$, denoting $c_{min} = \min_i \left(\sum_{j=1}^N |H_{ij}|\right) > 1$ and $b_{min} = \min_i \left(b_i^2 / T_i\right) \ne 0$, then, according to Lemma 3.1,

$$\begin{aligned}
Var\left(X\left(\gamma_k\right)\right) &= E\left(\left(X\left(\gamma_k\right)\right)^2\right) - \left(E\left(X\left(\gamma_k\right)\right)\right)^2 \\
&= \sum_{r_1=1}^N \sum_{r_2=1}^N \cdots \sum_{r_k=1}^N P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k} T_{r_k} \\
&\quad \times \left(\frac{H_{r_0 r_1} H_{r_1 r_2} \cdots H_{r_{k-1} r_k} b_{r_k}}{P_{r_0 r_1} P_{r_1 r_2} \cdots P_{r_{k-1} r_k} T_{r_k}}\right)^2 - \left(H^k b\right)_{r_0}^2 \\
&= \sum_{r_1=1}^N \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-1}=1}^N \frac{H_{r_{k-2} r_{k-1}}^2}{P_{r_{k-2} r_{k-1}}} \sum_{r_k=1}^N \frac{H_{r_{k-1} r_k}^2}{P_{r_{k-1} r_k}} \frac{b_{r_k}^2}{T_{r_k}} - \left(H^k b\right)_{r_0}^2 \\
&\ge b_{min} \sum_{r_1=1}^N \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-1}=1}^N \frac{H_{r_{k-2} r_{k-1}}^2}{P_{r_{k-2} r_{k-1}}} c_{min}^2 - \left(H^k b\right)_{r_0}^2
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\geq b_{min} \sum_{r_1=1}^{N} \frac{H_{r_0 r_1}^2}{P_{r_0 r_1}} \cdots \sum_{r_{k-2}=1}^{N} \frac{H_{r_{k-3} r_{k-2}}^2}{P_{r_{k-3} r_{k-2}}} c_{min}^4 - \left(H^k b\right)_{r_0}^2$$
$$\cdots$$
$$\geq b_{min} c_{min}^{2k} - \left(H^k b\right)_{r_0}^2 .$$

Here, $\left(H^k b\right)_{r_0}^2 \to 0$ when $k \to \infty$ because $\rho(H) < 1$, and b_{min} is a constant. However, $c_{min}^{2k} \to \infty$ as $k \to \infty$ since $c_{min} > 1$. In conclusion, regardless of how the transition matrix P is set up, if $\sum_{j=1}^{N} |H_{ij}| > 1$ for every row i, $Var\left(X\left(\gamma_k\right)\right) \to \infty$ as $k \to \infty$, even though $\rho(H) < 1$. ∎

Theorem 3.7 extends Theorem 3.5 to the more general situation where $\rho(H) < 1$ but $\rho(H^+) > 1$, with H^+ the $N \times N$ nonnegative matrix related to H as $H_{ij}^+ = |H_{ij}|$. According to Lemma 3.6, $\rho(H^+) > 1$ is a weaker condition than $\sum_{j=1}^{N} |H_{ij}| > 1$ for every row i in H. Theorem 3.7 can be used to explain the divergence of case 5 described in section 2. For more general situations where $\rho(H) < 1$ but $\|H\| \geq 1$, the Monte Carlo linear solver may converge or diverge. We will provide further analysis in next section.

LEMMA 3.6. *Suppose that H is an $N \times N$ nonsingular matrix and H^+ is the nonnegative matrix related to H by $H_{ij}^+ = |H_{ij}|$. If $\sum_{j=1}^{N} |H_{ij}| > 1$ for every row i in H, then $\rho(H^+) > 1$.*

Proof. Since $\sum_{j=1}^{N} |H_{ij}| > 1$ for every row i, denoting $\|\cdot\|_F$ as the Frobenius norm and $c_{min} = \min_i \left(\sum_{j=1}^{N} |H_{ij}|\right) > 1$, then

$$\|H^{+k}\|_F = \sqrt{\sum_{r_0=1}^{N} \sum_{r_k=1}^{N} \left(\sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_{k-1}=1}^{N} |H_{r_0 r_1}| |H_{r_1 r_2}| \cdots \left|H_{r_{k-1} r_k}\right| \right)^2}.$$

According to the quadratic mean–arithmetic mean inequality,

$$\begin{aligned}
\|H^{+k}\|_F &\geq \frac{1}{N} \sum_{r_0=1}^{N} \sum_{r_k=1}^{N} \left(\sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_{k-1}=1}^{N} |H_{r_0 r_1}| |H_{r_1 r_2}| \cdots \left|H_{r_{k-1} r_k}\right| \right) \\
&= \frac{1}{N} \sum_{r_0=1}^{N} \sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_k=1}^{N} |H_{r_0 r_1}| |H_{r_1 r_2}| \cdots \left|H_{r_{k-1} r_k}\right| \\
&= \frac{1}{N} \sum_{r_0=1}^{N} \sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_{k-1}=1}^{N} |H_{r_0 r_1}| |H_{r_1 r_2}| \cdots \sum_{r_k=1}^{N} \left|H_{r_{k-1} r_k}\right| \\
&\geq \frac{c_{min}}{N} \sum_{r_0=1}^{N} \sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_{k-1}=1}^{N} |H_{r_0 r_1}| |H_{r_1 r_2}| \cdots \left|H_{r_{k-2} r_{k-1}}\right| \\
&\cdots \\
&\geq \frac{c_{min}^k}{N}.
\end{aligned}$$

Based on the Gelfand's formula [20], for any matrix norm,

$$\rho\left(H^+\right) = \lim_{k \to \infty} \|H^{+k}\|^{\frac{1}{k}}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Hence,

$$\begin{aligned}\rho\left(H^{+}\right) &= \lim_{k\to\infty} \|H^{+k}\|_F^{\frac{1}{k}} \\ &\geq \lim_{k\to\infty} \frac{c_{min}}{N^{\frac{1}{k}}} \\ &= \frac{c_{min}}{\lim_{k\to\infty} N^{\frac{1}{k}}} \\ &= c_{min} > 1. \qquad \square\end{aligned}$$

THEOREM 3.7. *Let H be an $N \times N$, nonsingular matrix with spectral radius $\rho(H) < 1$. Let H^{+} be the $N \times N$ matrix where $H_{ij}^{+} = |H_{ij}|$. If $\rho(H^{+}) > 1$, there does not exist a transition matrix P satisfying the transition conditions* (1.2) *such that the variance $Var(X(\gamma_k))$ converges to zero as $k \to \infty$.*

Proof. Denoting $b_{min} = \min_i \left(b_i^2/T_i\right) \neq 0$,

$$\begin{aligned}Var(X(\gamma_k)) &= E\left((X(\gamma_k))^2\right) - (E(X(\gamma_k)))^2 \\ &= \sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N} P_{r_0r_1}P_{r_1r_2}\cdots P_{r_{k-1}r_k}T_{r_k} \\ &\quad \times \left(\frac{H_{r_0r_1}H_{r_1r_2}\cdots H_{r_{k-1}r_k}b_{r_k}}{P_{r_0r_1}P_{r_1r_2}\cdots P_{r_{k-1}r_k}T_{r_k}}\right)^2 - \left(H^k b\right)_{r_0}^2 \\ &\geq b_{min}\left(\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N}\left(\sqrt{P_{r_0r_1}P_{r_1r_2}\cdots P_{r_{k-1}r_k}}\right)^2\right) \\ &\quad \times \left(\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N}\left(\frac{H_{r_0r_1}H_{r_1r_2}\cdots H_{r_{k-1}r_k}}{\sqrt{P_{r_0r_1}P_{r_1r_2}\cdots P_{r_{k-1}r_k}}}\right)^2\right) - \left(H^k b\right)_{r_0}^2.\end{aligned}$$

According to the Cauchy–Schwarz inequality,

$$\begin{aligned}&\geq b_{min}\left(\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N}\left(\sqrt{P_{r_0r_1}P_{r_1r_2}\cdots P_{r_{k-1}r_k}}\frac{|H_{r_0r_1}||H_{r_1r_2}|\cdots\left|H_{r_{k-1}r_k}\right|}{\sqrt{P_{r_0r_1}P_{r_1r_2}\cdots P_{r_{k-1}r_k}}}\right)\right)^2 \\ &\quad - \left(H^k b\right)_{r_0}^2 \\ &= b_{min}\left(\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N}|H_{r_0r_1}||H_{r_1r_2}|\cdots\left|H_{r_{k-1}r_k}\right|\right)^2 - \left(H^k b\right)_{r_0}^2.\end{aligned}$$

Here, $\left(H^k b\right)_{r_0}^2 \to 0$ when $k \to \infty$ because $\rho(H) < 1$, and b_{min} is a constant. Since $\rho(H^{+}) > 1$, $\left(\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N}|H_{r_0r_1}||H_{r_1r_2}|\cdots\left|H_{r_{k-1}r_k}\right|\right)^2 \to \infty$ as $k \to \infty$ for random walks starting at r_0. Therefore, $Var(X(\gamma_k))$ diverges regardless of how the transition matrix P is selected. $\square$

4. A necessary and sufficient condition. In section 3, we discussed the importance of the transition matrix P. By taking both H and P into consideration, we derive a necessary and sufficient condition for convergence in the Monte Carlo linear solver using the Ulam–von Neumann algorithm in Theorem 4.2.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LEMMA 4.1. *Let H be an $N \times N$ nonsingular matrix and b be a nonzero vector. If $\rho(H) < 1$, $\sum_{k=0}^{\infty} \left(H^k b\right)_{r_0}^2$ is bounded.*

Proof. For any $\epsilon > 0$, a matrix R is generated such that

$$R = \frac{H}{\rho(H) + \epsilon}.$$

Due to $0 < \rho(H) < 1$, it is easy to show that $\rho(R) = \rho(H) / (\rho(H) + \epsilon) < 1$. Then,

$$\lim_{k \to \infty} R^k = 0.$$

Or, equivalently, this indicates that a natural number K exists such that

$$\forall k > K, \|R^k\| < 1.$$

Accordingly,

$$\forall k > K, \|R^k\| = \| \left(\frac{H}{\rho(H) + \epsilon} \right)^k \| = \frac{\|H^k\|}{(\rho(H) + \epsilon)^k} < 1.$$

That is,

$$\forall k > K, \|H^k\| < (\rho(H) + \epsilon)^k.$$

Therefore, $\forall k > K$,

$$\left| \left(H^k b\right)_{r_0} \right| \leq \|H^k b\| \leq \|H^k\| \|b\| < (\rho(H) + \epsilon)^k \|b\|$$

and

$$\left(H^k b\right)_{r_0}^2 \leq \|H^k b\|^2 < (\rho(H) + \epsilon)^{2k} \|b\|^2.$$

In particular, since ϵ can be any positive number, we can set $\epsilon = c^{\frac{1}{2}} - \rho(H) > 0$, where c is a positive number such that $\rho(H)^2 < c < 1$. Then

$$\left(H^k b\right)_{r_0}^2 < c^k \|b\|^2 \forall k > K.$$

Hence,

$$\begin{aligned}
\sum_{k=0}^{\infty} \left(H^k b\right)_{r_0}^2 &= \sum_{k=0}^{K} \left(H^k b\right)_{r_0}^2 + \sum_{k=K+1}^{\infty} \left(H^k b\right)_{r_0}^2 \\
&\leq \sum_{k=0}^{K} \left(H^k b\right)_{r_0}^2 + \sum_{k=K+1}^{\infty} c^k \|b\|^2 \\
&= \sum_{k=0}^{K} \left(H^k b\right)_{r_0}^2 + \|b\|^2 \sum_{k=K+1}^{\infty} c^k \\
&= \sum_{k=0}^{K} \left(H^k b\right)_{r_0}^2 + \frac{\|b\|^2 c^{K+1}}{1 - c}.
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Since $\sum_{k=0}^{K}\left(H^k b\right)_{r_0}^2$ has finite number of terms, and $\|b\|^2 c^{K+1}/(1-c)$ is a constant, $\sum_{k=0}^{\infty}\left(H^k b\right)_{r_0}^2$ is bounded. ∎

THEOREM 4.2. *Given an $N \times N$ nonsingular matrix H such that $\rho(H) < 1$, a nonzero vector b, and a transition matrix P, the necessary and sufficient condition for convergence of the Monte Carlo linear solver using the Ulam–von Neumann algorithm is $\rho(H^*) < 1$, where H^* is an $N \times N$ matrix such that $H^*_{ij} = H^2_{ij}/P_{ij}$.*

Proof. Since

$$\begin{aligned} Var\left(X\left(\gamma_k\right)\right) &= E\left(\left(X\left(\gamma_k\right)\right)^2\right) - \left(E\left(X\left(\gamma_k\right)\right)\right)^2 \\ &= \sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N} P_{r_0 r_1}P_{r_1 r_2}\cdots P_{r_{k-1}r_k}T_{r_k} \\ &\quad \times\left(\frac{H_{r_0 r_1}H_{r_1 r_2}\cdots H_{r_{k-1}r_k}b_{r_k}}{P_{r_0 r_1}P_{r_1 r_2}\cdots P_{r_{k-1}r_k}T_{r_k}}\right)^2 - \left(H^k b\right)_{r_0}^2 \\ &= \sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\cdots\sum_{r_k=1}^{N}\frac{H^2_{r_0 r_1}H^2_{r_1 r_2}\cdots H^2_{r_{k-1}r_k}b^2_{r_k}}{P_{r_0 r_1}P_{r_1 r_2}\cdots P_{r_{k-1}r_k}T_{r_k}} - \left(H^k b\right)_{r_0}^2 \\ &= \left(H^{*k}b^*\right)_{r_0} - \left(H^k b\right)_{r_0}^2, \end{aligned}$$

where b^* is a nonzero vector such that $b^*_i = b_i^2/T_i$, and T_i is the termination probability at row i, in the Ulam–von Neumann algorithm. If the k random walks are independent, it follows that

$$\begin{aligned} Var\left(\sum_{k=0}^{\infty}X\left(\gamma_k\right)\right) &= \sum_{k=0}^{\infty}Var\left(X\left(\gamma_k\right)\right) \\ &= \sum_{k=0}^{\infty}\left(\left(H^{*k}b^*\right)_{r_0} - \left(H^k b\right)_{r_0}^2\right) \\ &= \sum_{k=0}^{\infty}\left(H^{*k}b^*\right)_{r_0} - \sum_{k=0}^{\infty}\left(H^k b\right)_{r_0}^2. \end{aligned}$$

Since $\rho(H) < 1$, Lemma 4.1 implies the second term $\sum_{k=0}^{\infty}\left(H^k b\right)_{r_0}^2$ is bounded. Therefore, whether $Var\left(\sum_{k=0}^{\infty}X\left(\gamma_k\right)\right)$ is bounded depends solely on the first term, $\sum_{k=0}^{\infty}\left(H^{*k}b^*\right)_{r_0}$, which is bounded if and only if $\rho(H^*) < 1$. In conclusion, $\rho(H^*) < 1$ is the necessary and sufficient condition for convergence of the Monte Carlo linear solver. ∎

Denoting H^+ as the $N \times N$ matrix with $H^+_{ij} = |H_{ij}|$ and H^* as the $N \times N$ matrix where $H^*_{ij} = H^2_{ij}/P_{ij}$ given H and transition matrix P, Figure 4.1 summarizes the relationship between matrix H and the convergence of the Monte Carlo linear solver using the Ulam–von Neumann algorithm. According to Theorem 4.2, the key to convergence of the Monte Carlo linear solver is finding a transition matrix P such that $\rho(H^*) < 1$. Theorem 3.4 proves that such transition matrices always exist and are easy to find when $\|H\| < 1$. In contrast, Theorem 3.7 indicates that such transition matrices do not exist when $\rho(H^+) > 1$. For matrices where $\|H\| \geq 1$ and $\rho(H^+) \leq 1$, finding a transition matrix P becomes a constraints satisfaction problem defined as follows:

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

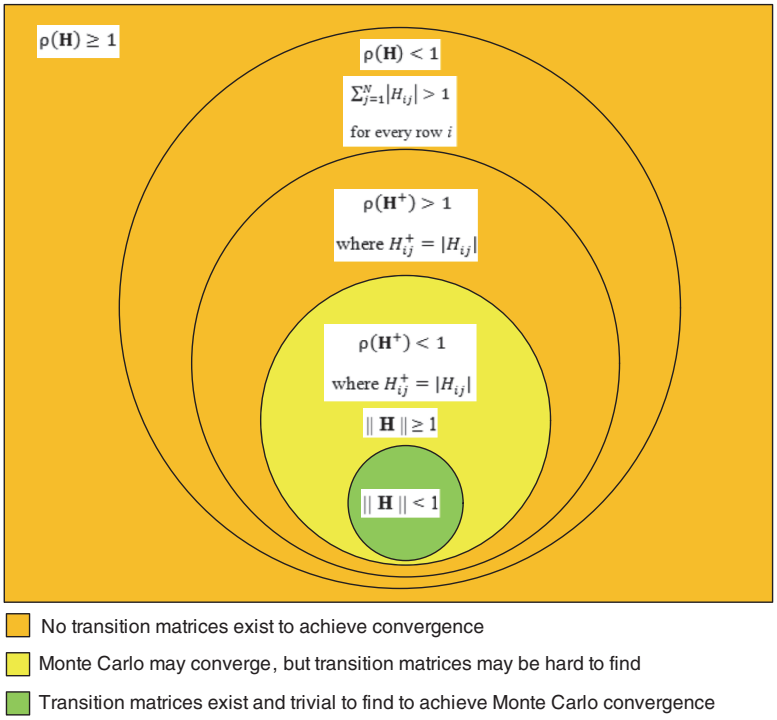


FIG. 4.1. *Summary of relationship between matrix H and convergence in Monte Carlo linear solver using Ulam–von Neumann algorithm.*

Variables: $\{P_{ij}|i = 1 \ldots N, j = 1 \ldots N\}$;
Domain: $[0, 1]$;
Constraints: $P_{ij} \geq 0; \sum_j P_{ij} \leq 1; H_{ij} \neq 0 \rightarrow P_{ij} \neq 0; \rho(H^*) < 1$.
Unfortunately, solving this constraint satisfaction problem can be at least as hard as solving the original problem of $x = Hx + b$.

5. Discussions.

5.1. Robustness. Generally, robustness in Monte Carlo methods requires that the variance of $X(\gamma_k)$ for $H^k b$ is bounded as $k \rightarrow \infty$ [11, 12]. Provided that the necessary and sufficient condition in Theorem 4.2 is satisfied, the Monte Carlo linear solver can automatically satisfy the robustness requirement. This is because the guarantee of the variance of $\sum_k X(\gamma_k)$ is bounded when $k \rightarrow \infty$ and implies the same for the variance of $X(\gamma_k)$ for each k.

5.2. Balancing. Recent studies [12] in Monte Carlo methods for numerical linear algebra show that balancing the matrix H is important to the speed of Monte Carlo convergence, because imbalance results in increased stochastic errors and thus slower convergence. We thus use the necessary and sufficient condition in Theorem 4.2 as an analysis tool to study the behavior of the Monte Carlo linear solver for unbalanced matrices under various transition matrices.

We start with a 100×100 perfectly balanced matrix H where each element of H is $\frac{1}{101}$ and then add noise to each matrix entry by introducing random perturbations to generate unbalanced matrices. At each perturbation percentage, we generate 100

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

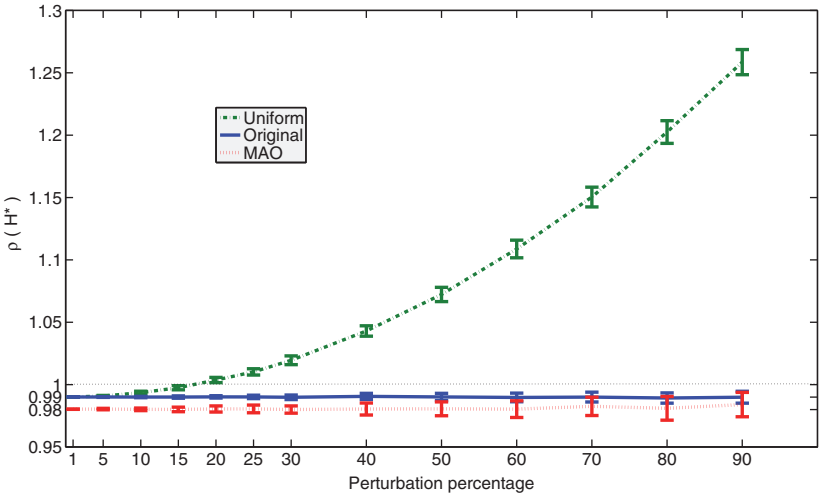


FIG. 5.1. *Variation of $\rho(H^*)$ values when different percentages of perturbation are imposed to H and transition matrix P adopts uniform, original (suggested by Ulam and von Neumann), or MAO schemes. $\rho(H^*) > 1$ suggests Monte Carlo divergence (Theorem 4.2).*

nonbalanced matrices. The higher perturbation percentage results in a more unbalanced H. We consider the transition matrices P adopting one of three schemes:

(i) Uniform: $P_{ij} = \frac{1}{101}$;
(ii) Original (suggested by Ulam and von Neumann): $P_{ij} = |H_{ij}|$;
(iii) MAO [13]: $P_{ij} = |H_{ij}| / \sum_j |H_{ij}|$.

We calculate the corresponding H^* matrix and $\rho(H^*)$ for each P and H at different perturbation percentages, as shown in Figure 5.1. According to Theorem 4.2, $\rho(H^*) < 1$ is an indicator for convergence of the Monte Carlo linear solver. Clearly, the selection of transition matrices P has a significant impact on Monte Carlo convergence. When a uniform transition matrix P is employed, the Monte Carlo solver diverges when as little as 20% disturbance is imposed. In contrast, the Monte Carlo solver can tolerate significantly higher disturbance (>90%) when the transition matrix P with the original or MAO scheme is adopted. Compared to the original transition scheme suggested by Ulam and von Neumann, the MAO scheme is even more stable, because it yields nearly optimal variance in Monte Carlo. (For further discussion of MAO and optimal schemes see [11].) Moreover, balancing plays an important role for Monte Carlo convergence. Even when the MAO scheme is used, higher disturbance on H leads to larger variation of $\rho(H^*)$. The larger value of $\rho(H^*)$ not only increases the number of iteration steps to convergence, as pointed out in [11], but also leads to divergence in the Monte Carlo solver once the $\rho(H^*) > 1$.

5.3. Applicability. The original Ulam–von Neumann algorithm is not efficient in practice [9]. Other techniques have been developed to improve the Monte Carlo algorithm for estimating the solutions of a linear system. Wasow [2] modified the scheme by designing another unbiased estimator, which has been shown to have smaller variance under special conditions. Halton [3] proposed a sequential Monte Carlo method to accelerate the Monte Carlo process by taking advantage of iterative refinement to transform the original linear system $x = Hx + b$ to a new system $y = Hy + d$, where $\|d\| < \|b\|$. Dimov et al. [4, 11] developed an accelerating Monte Carlo scheme to

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

control the convergence of the Monte Carlo algorithm for different unknown elements with different relaxation parameters, which can increase the efficiency of the random walk estimators. This iterative scheme is also used to approximately evaluate the matrix inverse. Tan [5] studied the antithetic variates techniques for variance reduction in Monte Carlo linear solvers. Srinivasan and Aggarwal [6] used nondiagonal splitting to improve Monte Carlo linear solvers. Moreover, for applications with large linear systems, Sabelfeld and Mozartova [7] designed a sparsified randomization algorithm by using a sparse, random matrix G, which is an unbiased estimator of H, to replace the original matrix H during the sampling process. Furthermore, Mascagni and Karaivanova [8] investigated the usage of quasi-random numbers in the Monte Carlo solver. Nevertheless, the fundamental mechanism of these Monte Carlo solvers, i.e., constructing Markov chains based on random walks to estimate the underlying Neumann series to evaluate solutions of the linear systems, remains the same. Therefore, provided that the random walks are based on Markov chains and the estimation is for the Neumann series, our analysis in this paper is applicable.

As shown in this paper, the limit of the Ulam–von Neumann algorithm lies in constructing the transition matrix P, which either does not exist when $\rho(H^+) > 1$ or is difficult to find when $\|H\| > 1$. Considering a more general form of a linear system

$$Ax = b,$$

only when A is strictly diagonally dominant can the linear system be easily converted to $x = Hx + b$ satisfying $\|H\| < 1$. If the convergence condition of the Monte Carlo linear solver can be extended to $\rho(H) < 1$, a much wider collection of matrices can be solved by the Monte Carlo linear solver.

In our subsequent paper, we will present a new Monte Carlo algorithm where the conditions of constructing the transition matrix P can be loosened and hence the Monte Carlo linear solver based on this new algorithm can converge when $\rho(H) < 1$.

5.4. Advantages of Monte Carlo linear solvers. Usually, Monte Carlo linear solvers are not as efficient as modern linear solvers in solving a general $Ax = b$ problem. Therefore, compared to the modern deterministic linear solvers, Monte Carlo linear solvers were not as widely used in practical applications in the past. However, due to the recent emergence of the "big data" problem [15], which is characterized by huge volume, rapid growth, and geometrically wide distribution, algorithms based on Monte Carlo sampling [16, 17, 18, 19] have become effective for handling various operations for large matrices. At the same time, Monte Carlo linear solvers have regained visibility. Compared to deterministic linear solvers, Monte Carlo linear solvers have several uniquely attractive advantages in handling extremely big matrices. First, Monte Carlo linear solvers are based on sampling, which does not need to access all elements of the matrix A. This is particularly suitable for applications such as large-scale sensor networks, where every element in the matrix A is available for access, but getting the complete picture of the matrix A is costly or infeasible. This is also helpful for handling incomplete or imperfect data. Second, random walks in Monte Carlo linear solvers can be carried out independently in a distributed manner, which is favorable for the large-scale parallel processing platforms of today as exemplified in grid and cloud computing [14] or general purpose graphics process units [21]. Third, Monte Carlo linear solvers can quickly obtain low-accuracy approximations to solutions. These approximate solutions may directly satisfy the accuracy requirements of big data applications or can be further refined to obtain highly accurate solutions. Fourth, random walks in Monte Carlo linear solvers have modest

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

memory requirements, and the random walk algorithm is scalable with the size of the matrices. Finally, for applications interested in only a few elements in the unknown vector, using Monte Carlo linear solvers based on the Ulam–von Neumann algorithm can eliminate unnecessary computations for other elements in the unknown vector.

6. Conclusions. We summarize our conclusions for the convergence of the Monte Carlo linear solver using the Ulam–von Neumann algorithm as follows:

(i) The convergence of the Neumann series is not a sufficient condition for the convergence of the Ulam–von Neumann algorithm.

(ii) The transition matrix P plays an important role. An improper selection of the transition matrix may result in divergence even though the condition $\|H\| < 1$ holds.

(iii) If $\|H\| < 1$ is satisfied, there always exist certain transition matrices that guarantee convergence of the Monte Carlo linear solver. These transition matrices are easy to find.

(iv) The Monte Carlo linear solver may or may not converge if $\|H\| < 1$ and $\rho(H) < 1$. If $\sum_{j=1}^{N} |H_{ij}| > 1$ for every row i in H or, more generally, $\rho(H^+) > 1$, where H^+ is the nonnegative matrix with $H_{ij}^+ = |H_{ij}|$, the Monte Carlo linear solver cannot converge, regardless of how the transition matrix P is selected.

(v) The necessary and sufficient condition for the Monte Carlo linear solver to converge is $\rho(H^*) < 1$,where $H_{ij}^* = H_{ij}^2/P_{ij}$ given H and a transition matrix P.

REFERENCES

[1] G. E. Forsythe and R. A. Leibler, *Matrix inversion by a Monte Carlo method*, Math. Tables Other Aids Comput., 4 (1950), pp. 127–129.

[2] W. Wasow, *A note on the inversion of matrices by random walks*, Math. Tables Other Aids Comput., 6 (1952), pp. 78–81.

[3] J. H. Halton, *Sequential Monte Carlo techniques for the solution of linear systems*, J. Sci. Comput., 9 (1994), pp. 213–257.

[4] I. T. Dimov, T. T. Dimov, and T. V. Gurov, *A new iterative Monte Carlo Approach for Inverse Matrix Problem*, J. Comput. Appl. Math., 92 (1998), pp. 15–35.

[5] C. J. K. Tan, *Antithetic Monte Carlo Linear Solver*, in Proceedings of the International Conference on Computational Science, Springer-Verlag, London, 2002, pp. 383–392.

[6] A. Srinivasan and V. Aggarwal, *Improved Monte Carlo linear solvers through non-diagonal splitting*, in Proceedings of the 2003 International Conference on Computational Science and Its Applications, Springer-Verlag, Berlin, 2003, pp. 168–177.

[7] K. Sabelfeld and N. Mozartova, *Sparsified randomization algorithms for large systems of linear equations and a new version of the random walk on boundary method*, Monte Carlo Methods Appl., 15 (2009), pp. 257–284.

[8] M. Mascagni and A. Karaivanova, *A Parallel Quasi-Monte Carlo Method for Solving Systems of Linear Equations*, in Proceedings of the International Conference on Computational Science, Springer-Verlag, London, 2002, pp. 598–608.

[9] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Chapman and Hall, London, 1964.

[10] J. H. Curtiss, *A theoretical comparison of the efficiencies of two classical methods and a Monte Carlo method for computing one component of the solution of a set of linear algebraic equations*, in Proceedings of Symposium on Monte Carlo Methods, University of Florida, 1954, John Wiley and Sons, New York, 1956, pp. 191–233.

[11] I. Dimov, *Monte Carlo Methods for Applied Scientists*, World Scientific Publishing, Singapore, 2008.

[12] I. Dimov, B. Philippe, A. Karaivanova, and C. Weihrauch, *Robustness and applicability of Markov chain Monte Carlo algorithms for eigenvalue problems*, Appl. Math. Model., 32 (2008), pp. 1511–1529.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[13] I. Dimov, *Minimization of the probable error for some Monte Carlo methods*, in Proceedings of the Summer School on Mathematical Modelling and Scientific Computations, Bulgarian Academy of Sciences, Sofia, 1991, pp. 159–170.
[14] Y. Li and M. Mascagni, *Analysis of Large-scale Grid-based Monte Carlo Applications*, Int. J. High Perform. Comput. Appl., 17 (2003), pp. 369–382.
[15] M. Hilbert and P. Lopez, *The Worlds Technological Capacity to Store, Communicate, and Compute Information*, Science, 332 (2011), pp. 60–65.
[16] P. Drineas, R. Kannan and M. W. Mahoney, *Fast Monte Carlo algorithms for matrices* I*: Approximating matrix multiplication*, SIAM J. Comput., 36 (2006), pp. 132–157.
[17] P. Drineas, R. Kannan and M. W. Mahoney, *Fast Monte Carlo algorithms for matrices* II*: Computing a low-rank approximation to a matrix*, SIAM J. Comput., 36 (2006), pp. 158–183.
[18] P. Drineas, R. Kannan and M. W. Mahoney, *Fast Monte Carlo algorithms for matrices* III*: Computing a compressed approximate matrix decompostion*, SIAM J. Comput., 36 (2006), pp. 184–206.
[19] S. Eriksson-Bique, M. Solbrig, M. Stefanelli, S. Warkentin, R. Abbey, and I. C. F. IPSEN, *Importance sampling for a Monte Carlo matrix multiplication algorithm, with application to information retrieval*, SIAM J. Sci. Comput., 33 (2011), pp. 1689–1706.
[20] I. Gelfand, *Normierte ringe*, Mat. Sb., 9 (1941), pp. 3–24.
[21] C. C. Su, M. R. Smith, F. A. Kuo, J. S. Wu, C. W. Hsieh, and K. C. Tseng, *Large-scale simulations on multiple graphics processing units (GPUs) for the direct simulation Monte Carlo method*, J. Comput. Phys., 231 (2012), pp. 7932–7958.

Downloaded 10/30/17 to 128.82.253.35. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.