General Sequential Sampling Techniques for Monte Carlo Simulations: Part I - Matrix Problems

Jerome Spanier and Liming Li

Abstract

We study sequential sampling methods based principally on ideas of Halton. Such methods are designed to build information drawn from early batches of random walk histories into the random walk process used to generate later histories in order to accelerate convergence. In previously published work, such methods have been applied within the pseudorandom, rather than the quasirandom, context and have been applied only to matrix problems. In this paper, more general sequential techniques are formulated in an abstract space, such as Banach space. The more general formulation enables applications to linear algebraic equations and to integral equations to be obtained as special cases through specification of the Banach space and the operator defined on it. In this initial paper we outline the ideas needed for consideration of the more general problem and exhibit greatly accelerated convergence for a simple matrix problem. In a companion paper in which similar ideas are applied to the more important class of integral equations, the need for quasirandom implementation is stressed.

1 Introduction

A promising direction for improvement of Monte Carlo convergence properties is the use of sequential sampling techniques (see, for example, [1], [2], [3] and [4]). Such techniques are designed to build information drawn from early batches of random walk histories into the random walk process used to generate later histories. When implemented properly, it should be possible to achieve much faster convergence of the statistical sampling error to zero than the asymptotic \( O(N^{-\frac{1}{2}}) \) rate promised by the central limit theorem.

It is to be expected, however, that implementation of any such sequential strategy will incur additional costs. Clearly, then, the error reduction achieved must more than offset these additional costs if real improvement is to be the result. If one defines, as is usual, the efficiency

\[ E = \frac{1}{VT} \]
where $V$ = appropriate measure of error, such as the relative variance or the relative variation of the estimator used, and $T$ = total computation time, then the extra computational burden imposed by sequential sampling will be reflected in increased values of $T$. The reductions in $V$ per iteration that result from the use of sequential sampling may or may not offset such additional costs by enough to yield substantial gains. That is, it is not a priori clear that anticipated exponential increases in convergence (reductions in $V$) will result in exponential increases in overall efficiency. This is one of the central issues that arises when sequential sampling methods are attempted.

In [1], Halton discussed sequential sampling methods in the context of the matrix problem

$$x = Ax + a$$

(1)

where $A$ is an $s \times s$ matrix, $x$ and $a$ are $s$-vectors. Briefly, [1] suggests the possibility of utilizing both biased and unbiased estimators for solving Equation (1) sequentially but no specifics are offered in [1] for their use. However, in [1] and in a recent paper [2], Halton reports that the use, instead, of a sequential form of correlated sampling produces convergence superior to that achieved by using either biased or unbiased estimators, and he bases his conclusions in these papers on this latter strategy exclusively.

We summarize Halton’s sequential mechanism by observing that if the solution of Equation (1) is represented as

$$x = \hat{x} + y$$

where $\hat{x}$ is an approximation to $x$, then the needed correction $y$ must satisfy

$$y = Ay + d$$

where

$$d = a + A\hat{x} - \hat{x}$$

is the equation error, or residual (see [5, Chapter 5, Section 5.1]), associated with using $\hat{x}$ in place of $x$ to solve Equation (1). Then, as $\hat{x}$ approaches $x$, the vectors $d$ and $y$ approach 0 and the effect of an error in estimating $y$ is correspondingly reduced in terms of its relative impact on the solution $x$.

To apply this idea sequentially, Halton begins with an initial approximation $\hat{y}^{(0)}$ to the solution of Equation (1) and defines $x^{(1)} = \hat{y}^{(0)}$. Then

$$d^{(1)} = a + Ax^{(1)} - x^{(1)}$$

denote by $\hat{y}^{(1)}$ an approximate solution of

$$y = Ay + d^{(1)}$$

and define

$$x^{(2)} = x^{(1)} + \hat{y}^{(1)} = \hat{y}^{(0)} + \hat{y}^{(1)}$$

$$d^{(2)} = a + Ax^{(2)} - x^{(2)}.$$
Continuing in this fashion produces the recursions

\[
\begin{align*}
\{ & x^{(k)} = x^{(k-1)} + \tilde{y}^{(k-1)} \\
& d^{(k)} = a + Ax^{(k)} - x^{(k)} \} \quad x^{(0)} = 0, \quad k = 1, 2, \ldots
\end{align*}
\]

(2)

where \( \tilde{y}^{(k-1)} \) is an approximate solution to

\[ y = Ay + d^{(k-1)}. \]

It follows easily that

\[ d^{(k)} = d^{(k-1)} + A\tilde{y}^{(k-1)} - \tilde{y}^{(k-1)} \]

and

\[ x^{(k)} = \sum_{i=0}^{k-1} \tilde{y}^{(i)}. \]

The main result of Halton [1], [2] is that, under appropriate restrictions, geometric convergence of \( x^{(k)} \) to \( x \) is possible for this sequential method. Since Halton's implementation uses pseudorandom numbers, this geometric convergence should presumably be regarded as holding in a statistical, not a deterministic sense.

We observe that implementation of Halton's sequential algorithm requires the approximate solution of the original problem (with a reduced source, \( d^{(k)} \)), and the computation of this reduced source itself requires application of the operator \( A \) to such an approximate solution, for each iteration. These operations impose a burden of additional computation but, at least in the matrix case (which appears to have been Halton's main concern), no intrinsic difficulty. In the integral equation applications which primarily interest us, however, more fundamental difficulties arise, as we explore in [6].

These problems notwithstanding, Halton reports in both [1] and [2] the effectiveness of his sequential correlated sampling algorithm for solving very large linear systems, especially ones whose matrices are full. He provides some analytic support for the claim that sequential correlated sampling is more effective even than the best currently available deterministic methods for such matrices. We will compare his method with the one developed in this paper for very simple matrices in section 3.

2 Generalization to Operator Equations

The sequential matrix method described above can be extended to include very important integral transport equation problems by considering the operator transport equation

\[ \phi = \mathcal{K}\phi + \sigma \]

(3)

where \( \mathcal{K} \) is a linear operator from \( B \) to \( B \) (a Banach space). We assume that the norm of the operator \( \mathcal{K} \) is less than one (\( \| \mathcal{K} \| < 1 \)) to guarantee that
the Equation (3) has exactly one solution (although the weaker assumption \( \| K^{n_0} \| < 1 \) for some integer \( n_0 \geq 1 \) will suffice; this follows from rather elementary fixed point theorems in \( L \)).

A quite general operator formulation for solving Equation (3) iteratively results by introducing a sequential dependence in both the source \( \sigma \) and kernel \( K \) via the operator equation

\[
\phi = K^{(k)} \phi + \sigma^{(k)}, \quad \sigma^{(0)} = \sigma, \quad K^{(0)} = K.
\]

(4)

In this way, Halton's sequential correlated sampling results by selecting \( K^{(k)} = K \) for all \( k \) and

\[
\sigma^{(k+1)} = d^{(k+1)} = d^{(k)} + K \hat{y}^{(k)} - \hat{y}^{(k)},
\]

(5)

where \( \hat{y}^{(k)} \) is an approximate solution to Equation (4), while various more general biased and unbiased importance sampling sequential methods result by modifying the kernel \( K \) suitably from one iteration to the next. Our interest in this paper is predominantly on the latter family of sequential methods, with a view to applying these ultimately to integral equations, thus avoiding the computation of the reduced source required by Halton's correlated sampling method.

A comprehensive theory of convergence for this more general operator formulation will likely not be simple since one can expect quite different convergence characteristics depending on the specifics of the problem and the details of the implementation method. For example, if we mimic Halton's treatment for matrix problems, we would compute an initial approximation \( \hat{y}^{(0)} \) to \( \phi \) and define

\[
\phi^{(1)} = \hat{y}^{(0)}.
\]

Then with

\[
d^{(1)} = \sigma + K \phi^{(1)} - \phi^{(1)},
\]

we let \( \hat{y}^{(1)} \) denote an approximate solution to

\[
y = K y + d^{(1)}
\]

and define

\[
\phi^{(2)} = \phi^{(1)} + \hat{y}^{(1)} = \hat{y}^{(0)} + \hat{y}^{(1)}.
\]

Continuing produces the recursions

\[
\begin{align*}
\phi^{(k)} &= \phi^{(k-1)} + \hat{y}^{(k-1)} \\
d^{(k)} &= \sigma + K \phi^{(k)} - \phi^{(k)}
\end{align*}
\]

(6)

where \( \hat{y}^{(k-1)} \) is an approximate solution of

\[
y = K y + d^{(k-1)}.
\]

It follows that

\[
d^{(k)} = d^{(k-1)} + K \hat{y}^{(k-1)} - \hat{y}^{(k-1)}.
\]
and

$$\phi^{(k)} = \sum_{i=0}^{k-1} \phi^{(i)}.$$  

The convergence of this generalized sequential method will depend on how the reduced equations

$$y = Ky + d^{(k)}$$

are approximately solved to produce \( \hat{y}^{(k)} \). If, for example, conventional (Picard) iteration

$$\phi^{(n+1)} = \mathcal{K} \phi^{(n)} + d^{(k)} \quad \phi^{(0)} = d^{(k)};$$

is employed and restricted to \( n = 0 \) (i.e., a single iteration is performed to obtain \( \hat{y}^{(k)} \) from \( d^{(k)} \)), we get

$$\hat{y}^{(k)} = \phi^{(1)} = \mathcal{K} d^{(k)} + d^{(k)}$$

with the result that

$$\phi^{(k)} = \sum_{i=0}^{k-1} \phi^{(i)} = (I + \mathcal{K}) \sigma + \mathcal{K}^2 \phi^{(k-1)}$$

and we have arrived back at the partial sums of the Neumann series for the solution \( \phi \). This series converges provided \( \|\mathcal{K}\|^2 < 1 \). Subject to this same assumption, it is easy to show that the error \( e^{(k)} = \phi^{(k)} - \phi \) can be estimated practically using the formula

$$\|e^{(k)}\| \leq \|\phi^{(k+1)} - \phi^{(k)}\| \frac{\|\mathcal{K}\|}{1 - \|\mathcal{K}\|^2}$$

which depends on the error upon successive iterations and the norm of \( \mathcal{K} \). While this leads to a satisfactory theory of convergence, effective sequential methods are designed to do more than advance each random walk a single step beyond the source (as in Picard iterations) and thus, properly implemented, should do substantially better than this rather naive iterative procedure. Also, permitting iterative redefinition of the operator \( \mathcal{K} \) vastly complicates the analysis of convergence in this very general setting. A different convergence theory will be derived in [6] for the integral equation applications by utilizing numerical quadrature effectively to replace the integral equation by a matrix equation and applying the sequential iterations to the matrix problem.

The more general operator formulation described above has two important applications, which otherwise have to be treated separately.

1. If \( \mathcal{B} = \mathbb{R}^s \) or \( \mathcal{C}^s \), regarded as \( s \)-dimensional real or complex vector spaces, \( \mathcal{K} = A \) is an \( s \times s \) real or complex matrix, \( \phi = x \), \( \sigma = a \) are \( s \)-vectors, then Equation (3) reduces to Equation (1) and may be interpreted as describing transport in an infinite homogeneous medium with \( s \) discrete energy states available for each particle.
II. If \( B = L^p(\mu) \) (or \( B = C(X) \), the space of continuous functions on a normed space \( X \)), \( K \) an integral operator, Equation (3) becomes the continuous transport equation

\[
\psi(P) = \int_{\Gamma} K(P, P')\psi(P')dP' + S(P)
\]

(8)

where

\[
K\psi(P) = \int_{\Gamma} K(P, P')\psi(P')dP'.
\]

This integral equation may be shown to be the appropriate equation to describe the directional particle collision density, \( \psi(P) = \Sigma(P)\phi(P) \), where \( \Sigma(P) \) is the total macroscopic cross section\(^1\) at \( P \) and \( \phi(P) \) is the particle flux. In Equation (8), \( P, P' \) denote generic points in a euclidean phase space \( \Gamma \), \( S(P) \) is the density of first collisions, and \( K(P, P') \) is the transport kernel describing transitions from \( P' \) to \( P \).

In the context of Equation (8), it is usual to apply Monte Carlo methods to the estimation of integrals of the form

\[
I = \int_{\Gamma} g(P)\psi(P)dP
\]

(9)

where \( g \) is some known bounded nonnegative function (usually a ratio of nuclear cross sections) and \( \psi \) satisfies Equation (8). In the more abstract setting of Equation (3), the integral (9) is realized as either as a pairing of the Banach space \( B \) and its dual space \( B^* \), or, if \( B = H \) is a Hilbert space, as an inner product on \( H \). That is, the function \( g \) can be regarded as defining a bounded linear functional on the space \( B \) via the integral (9).

Reflecting on the recursions (6) in the integral equations setting, we now see a possibly serious drawback to their practical use. Not only is a continuous approximation \( \bar{\psi}^{(k)} \) to the solution of Equation (8) required at each stage, but computation of \( K\bar{\psi}^{(k)} \) as a component of the reduced source is also needed. Such a computation is tantamount to advancing by exactly one collision each term in a Neumann series representation of the problem solution.

In an effort to avoid these difficulties, and to keep computational costs of sequential methods at a minimum, we have chosen to explore other sequential algorithms for solving Equation (3) iteratively. Indeed, since progressively improved knowledge of at least an approximation to the solution of Equation (3) seems a prerequisite for the design of an effective sequential algorithm, our thought was to see if such information alone could be used to build appropriate knowledge systematically into a rapidly converging (unbiased or biased) importance sampling scheme based on the well established theory for such that can be found, for example, in \cite{7}, Chapter 3, Section

\(^1\)The inverse of this function is the average distance travelled between successive collisions in an infinite homogeneous medium of material with cross section \( \Sigma(P) \) (the cross section \( \Sigma(P) \) has dimensions inverse length).
This theory establishes a duality between the abstract formulation based on the pair of Equations (3) and the equation

\[ I = \langle g, \phi \rangle \]  

(10)

where \( \phi \in \mathcal{B} \), \( g \in \mathcal{B}^* \) and the equation pair

\[ \phi^* = \mathcal{K}^* \phi + g \]  

(11)

and

\[ I^* = \langle \phi^*, \sigma \rangle . \]  

(12)

where \( \phi^* \in \mathcal{B}^* \). Here, \( \mathcal{K}^* \) is the operator adjoint to \( \mathcal{K} \), as defined by the identity \( \langle \mathcal{K}^* \theta^*, \theta \rangle = \langle \theta^*, \mathcal{K} \theta \rangle \) for all \( \theta \in \mathcal{B} \) and all \( \theta^* \in \mathcal{B}^* \). This duality permits the estimation of the inner product \( I = \langle g, \phi \rangle \) as \( I^* = \langle \phi^*, \sigma \rangle \), where \( \phi, \phi^* \) satisfy Equations (3), (11), respectively. Further, through this duality, problems may be formulated either in a direct or an equivalent adjoint mode. Through duality, random variables developed with respect to a given (discrete or continuous) equation have counterpart definitions with respect to the dual equation.

The theory of importance sampling based on this duality and applied to continuous problems can be found in [7, Chapter 3, Section 3.7], and a completely parallel theory for discrete problems can easily be developed. This theory reveals that the solution \( \phi^* \) of Equation (11) acts as a perfect importance function for the estimation of the quantity (10) and conversely, that the solution \( \phi \) of Equation (3) acts as a perfect importance function for the estimation of the quantity (12). While exact knowledge of \( \phi \) or \( \phi^* \) may not be assumed in designing Monte Carlo or quasi-Monte Carlo random walk simulations, approximations such as those generated in a sequential method should play a useful role in achieving reduced error sampling.

In the present paper we shall restrict our attention to applications of type I above. That is, we will apply the conventional theory of Monte Carlo simulation of random walks for the purpose of solving matrix equations as established for example, in [7, Chapter 2, Section 2.3]. Applications to integral equations will be treated in [6]. For the solution of matrix problems the random walks take place on a discrete phase space \( \{1, ..., s\} \), where \( s \) is the order of the linear algebraic system under consideration.

Random variables that are useful in the discrete case originated, as far as we are aware, with von Neumann and Ulam (published initially in [8]; see also [9]). For our study we chose discrete versions of the family introduced in [10], a family that includes both biased and unbiased importance sampling estimators. Our first experiments, reported here, were performed with unbiased versions of such estimators which coincide with more conventional random variables produced by importance sampling transformations.

Specifically, for the solution of the matrix problem defined by Equation (1), we will construct a new transition matrix, \( (q_{ij}^{(m)})_{s \times s} \) by using the approximate solution \( \hat{x}^{(m-1)} \) as an approximate (dual) importance function from iteration \( m - 1 \). The new transition matrix is then used to generate
the next batch of random walk histories. For the $m$th iterative stage, discrete random walks $\omega = (i_1, i_2, \ldots, i_k)$ are then generated using the matrix $(q_{ij}^{(m)})_{i\times j}$ instead of the original matrix $(A_{i,j})_{i\times j}$. To construct this matrix, dual importance sampling theory suggests that we should set

$$q_{ij}^{(m)} = \frac{A_{ij}}{\sum_{l=1}^{N} A_{il} \tilde{x}_l^{(m-1)} + a_i}$$  \hspace{1cm} (13)$$

where $\tilde{x}_j^{(m-1)}$ is the $j$th component of the approximate solution $\tilde{x}^{(m-1)}$ at the $(m-1)$st iterative stage. Then random walks $\omega = (i_0 = j, i_1, i_2, \ldots, i_k)$ originating in the fixed state $j$ are generated using $q_{ij}^{(m)}$ to determine successor states $j$ from the current state $i$ and weights $W_n$ are defined by:

$$W_0 = 1, \quad W_n = W_{n-1} \frac{A_{i_{n-1},i_n}^{(m)}}{q_{i_{n-1},i_n}^{(m)}}.$$  

It follows [7, Chapter 2, Section 2.3] that the random variable

$$\xi_j(\omega) = W_k a_{i_k} / p_{i_k}^{(m)},$$  \hspace{1cm} (14)$$

where $p_{i_k}^{(m)} = 1 - \sum_{l=1}^{s} q_{il}^{(m)}$, is an unbiased estimator of the $j$th component of $x$. When the approximate solution $\tilde{x}$ is close to the exact solution $x$, we expect the variance of the estimator (14) to be small. By originating random walks in each of the $s$ discrete states at each stage, we are able to estimate all $s$ components of the unknown vector $x$ in each stage $m$.

We recognize that the estimator $\xi$ of Equation (14) is a terminal-type estimator\footnote{Collision-type estimators [7, Chapter 2, Section 2.3] were also used in the successful construction of sequential importance sampling algorithms. These results will be reported more fully elsewhere.} ([7, Chapter 2, Section 2.3]) implemented in the adjoint mode simulation process. This may be regarded as a random walk process in which each random walk starts at index $j$ and moves backwards from state to state until arriving at its terminus. In such a backwards simulation, the original source vector $a$ is used as a scoring function at the final collision point, as in Equation (14).

In the next section we will report some numerical experience demonstrating that the suggested sequential method produces very rapid convergence when the number of (pseudorandomly generated) random walk histories per iteration is chosen properly.

3 A Ten State Transport Problem

To test the efficiency of our sequential sampling method, we considered an inherently finite dimensional problem first: namely, a transport problem.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2.53E-03</td>
<td>2.53E-03</td>
<td>2.97E-01</td>
</tr>
<tr>
<td>200</td>
<td>2.90E-03</td>
<td>1.08E-03</td>
<td>7.03E-02</td>
</tr>
<tr>
<td>300</td>
<td>2.17E-03</td>
<td>3.44E-04</td>
<td>4.39E-02</td>
</tr>
<tr>
<td>400</td>
<td>1.23E-02</td>
<td>1.18E-04</td>
<td>1.25E-02</td>
</tr>
<tr>
<td>500</td>
<td>6.95E-03</td>
<td>7.95E-05</td>
<td>6.82E-03</td>
</tr>
<tr>
<td>600</td>
<td>1.58E-03</td>
<td>1.78E-05</td>
<td>1.99E-03</td>
</tr>
<tr>
<td>700</td>
<td>1.40E-03</td>
<td>9.57E-06</td>
<td>6.07E-04</td>
</tr>
<tr>
<td>800</td>
<td>4.02E-03</td>
<td>3.74E-06</td>
<td>2.67E-04</td>
</tr>
<tr>
<td>900</td>
<td>5.08E-03</td>
<td>9.21E-07</td>
<td>1.59E-04</td>
</tr>
<tr>
<td>1000</td>
<td>4.14E-03</td>
<td>2.46E-07</td>
<td>8.97E-05</td>
</tr>
</tbody>
</table>

Table 1: Comparison of sequential & nonsequential errors

in an infinite, homogeneous material in which particles can occupy only a finite number $s$ of energy states. Particle collisions then can result either in absorption or in transition from the current energy state to any other, but spatial variation can be ignored.

Suppose the transition matrix in this material is $A = (A_{ij})_{s \times s}$, and the source vector $a = (a_1, a_2, ..., a_s)$ satisfies $a_i \geq 0$ and $\sum_{i=1}^{s} a_i = 1$, where $a_i$ is interpreted as the density of particles that undergo a first collision in the $i$th energy state. For our test we set

\[ A_{ij} = \frac{s - |i - j|}{s^2} \]  \tag{15}

Then the probability that a random walk moves to the $j$th state from the $i$th state depends only on the distance between the indices $i$ and $j$. The goal here is to find $x_i$, the density of particles that undergo collision in the $i$th energy state. It is not difficult to see [7, Chapter 2, Section 2.2] that the above transport problem can be exactly described by a linear matrix system. To keep the problem manageable, we look at a transport problem involving only ten energy states: $s = 10$.

The transition matrix, $A$, is symmetric and its $L^\infty$-norm,

\[ ||A||_\infty = \max_{i,j=1}^{10} A_{ij} = 0.75 \]

This represents a problem of moderate difficulty in terms of the rate of convergence of the matrix Neumann series.

The source vector is also chosen symmetrically:

\[ a = (0, 0, 0, 2, 3, 3, 2, 0, 0, 0)^T \]

so that the transport problem is symmetric in its ten states.

Using the sequential algorithm described above, we solved this problem approximately by generating (using pseudorandom numbers) 100 random walk histories for each iteration and iterating 10 times. For comparison, we also solved the problem using Halton's sequential correlated sampling method, also employing 10 sequential stages of 100 random walks each, and
Figure 1: Ten state problem without sequential sampling

Figure 2: Ten state problem with sequential importance sampling and correlated sampling
using plain Monte Carlo (i.e., no sequential sampling), for which we generated 10,000 random walks. The numerical results are graphed in Figures 1 and 2 in which plots of the errors in solving this 10×10 problem without and with the use of sequential sampling, respectively, are shown. A portion of the output data on which these graphs are based is displayed in Table 1. In order to emphasize the tremendous improvement obtained by using sequential sampling, Table 1 exhibits the errors obtained with and without sequential sampling for the first 1,000 random walks only, even though a total of 10,000 random walks were processed for the nonsequential run.

The actual error in Figure 1, and its logarithm in Figure 2, is a global measure of the error defined by

\[
\text{actual error} = \frac{1}{10} \sum_{j=1}^{10} |\hat{x}_j - x_j|^2
\]  

(16)

where \(\hat{x}_j\) is the estimated solution of the transport problem using sequential sampling, and \(x_j\) is the exact solution of the transport problem.

The procedure above solves a linear system of equations whose exact solution can be found by other techniques, which makes it sensible to base our assessment of the approximation error on the actual error described above. However, for non-model problems, when \(x_j\) is unknown, a different criterion is clearly needed. Instead of the actual error, we can use the
average error (= norm of residual) defined by

\[
\text{average error} = \frac{1}{10} \sum_{i=1}^{10} |\hat{x}_i - \sum_{j=1}^{10} A_{ij} \hat{x}_j - a_i|^2
\]  

(17)

to examine the efficiency. The following theorem shows that the actual error (16) and the average error (17) provide equivalent measures of the accuracy of the solution.

Theorem: The actual error \(\to 0\) if and only if the average error \(\to 0\).

Proof: Because

\[
|\hat{x}_i - x_i| = |\hat{x}_i - \sum_{j=1}^{10} A_{ij} \hat{x}_j - a_i + \sum_{j=1}^{10} A_{ij} \hat{x}_j + a_i - x_i|
\]

\[
\leq |\hat{x}_i - \sum_{j=1}^{10} A_{ij} \hat{x}_j - a_i| + \sum_{j=1}^{10} A_{ij} |\hat{x}_j - x_j|,
\]

and the norm of the transition matrix \(\|A\|_\infty = \max_{i,j} A_{ij} < 1\),

\[
\max_i |\hat{x}_i - x_i| \leq \max_i |\hat{x}_i - \sum_{j=1}^{10} A_{ij} \hat{x}_j - a_i|/(1 - \max_{j=1}^{10} A_{ij})
\]

On the other hand,

\[
|\hat{x}_i - \sum_{j=1}^{10} A_{ij} \hat{x}_j - a_i| = |\hat{x}_i - \sum_{j=1}^{10} A_{ij} \hat{x}_j - (x_i - \sum_{j=1}^{10} A_{ij} x_j)|
\]

\[
= |(\hat{x}_i - x_i) - \sum_{j=1}^{10} A_{ij} (\hat{x}_j - x_j)|
\]

\[
\leq \sum_{j=1}^{10} |A_{ij}| \max_i |\hat{x}_j - x_j|
\]

The results in Figures 1 and 2 show that convergence is greatly accelerated when sequential sampling is used. For example, to obtain absolute accuracy \(10^{-4}\), our sequential sampling technique required fewer than 100 \(\times 5 = 500\) random walk histories (per component of solution), but without sequential sampling, more than 10,000 random walk histories (per component) are required to achieve this same accuracy. Using Halton's correlated sampling method, this accuracy was attained after approximately 900 random walks were processed. Similar results were obtained on other sample problems using matrices of various orders; a fairly consistent advantage was seen for our algorithm over Halton's method, and both significantly outperformed plain Monte Carlo on each problem tried.

A closer analysis of our sequential sampling method for solving Equation (3) suggests that the amount of error reduction per iteration achieved depends, in general, on the operator \(K\), and in particular, its norm. However, to ensure that each iteration actually reduces, rather than increases the error, the number of random walk histories per iteration must be sufficiently large. In the above model problem, \(\|K\| = \|A\|_\infty = 0.75\) and the
Table 2: Comparison of slopes and intercepts for different N values: norm(A)=0.75

<table>
<thead>
<tr>
<th></th>
<th>N = 50</th>
<th>N = 20</th>
<th>N = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>-1.08E+00</td>
<td>-9.73E-01</td>
<td>-5.23E-01</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.01E+00</td>
<td>5.19E-01</td>
<td>1.61E-01</td>
</tr>
</tbody>
</table>

Table 3: Comparison of slopes and intercepts for different N values: norm(A)=0.99

<table>
<thead>
<tr>
<th></th>
<th>N = 50</th>
<th>N = 20</th>
<th>N = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>-1.55E+00</td>
<td>-1.32E+00</td>
<td>-8.09E-01</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.01E+00</td>
<td>3.27E+00</td>
<td>2.97E+00</td>
</tr>
</tbody>
</table>

Figure 4: Influence on sequential convergence of the number of random walks per iteration: norm(A)=0.99
iterative process converges reasonably quickly. If we alter the transition matrix $A$ so that $\|A\|_{\infty}$ is closer to 1, then the average number of collisions per random walk is increased and the convergence of the matrix Neumann series is correspondingly slowed. For example, if the matrix $A = (A_{ij})_{0 \times 10}$ above is replaced by

$$\tilde{A}_{ij} = \frac{s - |i - j| + 2.4}{s^2},$$

the norm of the transition matrix becomes $\|\tilde{A}\|_{\infty} = \max_{i} \sum_{j=1}^{10} \tilde{A}_{ij} = 0.99$.

Figures 3 and 4 show the influence on convergence rate of the number $N$ of random walk histories per iteration as a function of $\|A\|_{\infty}$. In Figure 3 ($\|A\|_{\infty} = 0.75$), when $N$ is 50 (which means that only 50 random walks are generated per iteration), the factor $\lambda$ that multiplies the error may actually exceed 1 and the error will then increase from one iteration to the next. Choosing $N$ to be larger assures that $\lambda < 1$ and produces a steady reduction in the error for this easier of the two problems studied here. In Figure 4 ($\|A\|_{\infty} = 0.99$) we see that the factor $\lambda$ can exceed 1 both for $N = 6$ and $N = 20$. However, setting $N = 50$ resulted in the steady geometric reduction in error per iteration that characterizes a successful implementation of the sequential method, for this harder of the two problems studied. Determining the minimum number of random walks per sequential stage needed to guarantee monotone error reduction as a function of the problem input $A$, appears to be a formidable problem.

In Figures 5 and 6 we have graphed three lines that were obtained by

Figure 5: Best linear fit to data: ln error vs. number of iterations: 
norm($A$) = 0.75
fitting the best (in the sense of least squares) straight lines to data obtained by plotting logarithms of errors versus iteration numbers for the three runs shown in Figures 3 and 4, respectively. The slopes and intercepts of these three lines are given in Table 2 and 3; the slopes provide estimates of the error reduction per iteration\(^3\) achieved by the three different applications of the sequential strategy. Again, the importance of generating sufficiently many random walks in each iteration to assure strict error reduction is emphasized by these data.

Experimentation with the ideas of this paper is continuing.

4 Acknowledgement

The authors gratefully acknowledge the help of Ms. Jinhong Jiang who provided invaluable assistance with the development and improvement of several computer algorithms used in this work. We also acknowledge the useful suggestions of a referee.

References


\(^3\)In fact, the slope is an estimate of \(\ln \lambda\), where \(\lambda\) is the average error reduction (or multiplication) factor per iteration.


