Comparison of Monte Carlo Algorithms for Obtaining Geometric Convergence for Model Transport Problems

Carole Hayakawa and Jerome Spanier

Department of Mathematics
Claremont Graduate University
925 North Dartmouth Avenue
Claremont, CA 91711, USA
Email: carole.hayakawa@cgu.edu & jerome.spanier@cgu.edu

Abstract. Two quite different methods for accelerating the convergence of global Monte Carlo solutions of continuous transport problems have been developed recently in Claremont. One of these is based on a sequential form of correlated sampling, first proposed for matrix problems by Halton[1]. The second method makes use of importance sampling transformations applied adaptively[2]. These two methods are contrasted and compared for a family of model transport problems in one dimension.

1 Introduction

Recently, new adaptive methods for estimating the global solution of transport problems have been developed in Claremont [2–5]. These methods, based either on correlated sampling or importance sampling, are capable of learning as they proceed in such a way that geometric convergence to the correct solution is obtained after 30 or so adaptive stages have been processed. Both methods work by representing the global solution of the transport problem as an infinite sum of basis functions in an appropriate function space, truncating the sum to a finite number of terms, and using Monte Carlo methods to estimate each coefficient. Further details may be found in the papers [2–5] appearing elsewhere in this volume.

The learning mechanism occurs as a result of collecting information in groups of histories, called stages, and making use of the information from a given stage to improve the solution for the next stage. The sequential correlated sampling algorithm uses the current stage information to reduce the source for the next stage. The adaptive importance sampling method uses the solution from the current stage to modify both source and kernel operator (which describes how random walks move from point to point) for the next stage. The question naturally arises: for what class of transport problems would one method be preferred over the other? In this paper we study this problem numerically and draw tentative conclusions from our results.
In our numerical study, we describe a parametrized family of problems to serve as a testbed for the comparison. Results are then averaged over a representative set of problems selected randomly from the file. The intent is to draw valid conclusions about the rates of convergence of two methods, one depending on the predominantly additive nature of the sequential correlated sampling method and the other on the predominantly multiplicative nature of the adaptive importance sampling method.

2 Monodirectional Transport Problem

Consider a simple neutron transport problem in a slab in which on scattering or absorption is permitted. Figure 1 illustrates such a slab, which a source emits particles from the left into a slab of thickness $T$ and then absorbs or scattered directly forward on every collision. The equations governing this transport problem are the following:

\[
\frac{d\Phi}{dx} + \Sigma_t \Phi = \Sigma_s \Phi, \quad 0 < x < T
\]

\[
\Phi(0) = Q_0
\]

where
- $\Phi(x)$ is the expected number of particles passing $x$ per unit time;
- $\Sigma_a = \Sigma_t - \Sigma_s$ is the absorption cross section;
- $\Sigma_t = \Sigma_t(x)$ is the total macroscopic cross section; \(^1\)
- $\Sigma_s = \Sigma_s(x)$ is the scattering cross section;
- $T$ is the total thickness of the slab;
- $Q_0$ is the source strength at $x = 0$.

This problem can be written in an equivalent integral equation for $\Phi(x)$:

\[
\Phi(x) = Q_0 e^{-\Sigma_t x} + \int_0^x \Sigma_s e^{-\Sigma_t (x-y)} \Phi(y) dy, \quad 0 \leq x \leq T
\]

or

\[
\Phi = K\Phi + S
\]

where

\[
K\Phi(x) = \int_0^x k(x, y) \Phi(y) dy
\]

\[
S(x) = Q_0 e^{-\Sigma_t x}
\]

Here the integral operator $K : B \to B$ where $B$ is the Banach space $L^1(\Gamma) = \{ f : \Gamma \to \mathbb{R} \mid \int_0^T |f(x)| dx < \infty \}$. For effective solutions, we require $\|K\| < 1$ (or the weaker condition $\|K^n\| < 1$ for some $n \geq 1$) to ensure existence of a unique $\Phi \in B$ and convergence of the Neumann series.

\(^1\) The inverse of this function is the mean distance traveled between collisions in an infinite homogeneous material whose cross section is interpreted as the scattering and absorption cross sections.
Fig. 1. Depiction of monodirectional model problem, finite slab

\[(I - K)^{-1}S = (I + K + K^2 + \ldots)S\]

to \(\Phi\).

This problem does not need to be solved using Monte Carlo methods at all. It can be solved in closed form very easily since Equation (1) is a simple ordinary differential equation. However, our purpose is to use the knowledge of the exact solution

\[\Phi(x) = Q_0e^{-Kx^2}, 0 \leq x \leq T\]

to accurately contrast and compare the two adaptive methods and their effectiveness in solving this problem and more complex transport problems.

3 Solution Algorithm

We assume that the approximate solution of Equation (1) can be written as a finite linear combination of Legendre polynomials and we represent the approximate solution \(\hat{\Phi}\) as

\[\hat{\Phi}(x) = \sum_{i=0}^{N} a_i P_i\left(\frac{2x}{T} - 1\right).\]  \(\text{(4)}\)

Then the coefficients \(a_i\) can be found by

\[a_i = \frac{2i + 1}{T} \int_0^T \hat{\Phi}(x) P_i\left(\frac{2x}{T} - 1\right)dx\]  \(\text{(5)}\)

or

\[a_i = <\hat{\Phi}, S_i^*>,\]  \(\text{(6)}\)
where
\[ S'_i(x) = \frac{2i+1}{T} P_i \left( \frac{2x}{T} - 1 \right) \]

and the inner product indicated in Equation (6) is the integral of the functions appearing inside the brackets. The coefficients are found using either of two adaptive strategies: sequential correlation and adaptive importance sampling described in the next section.

4 General Adaptive Strategies

Making use of the general model described by Equation (3), we have
\[ \Phi = K^{(\nu)} \Phi + S^{(\nu)}. \]

The superscript \( \nu \) refers to the adaptive stage number and both correlated sampling and adaptive importance sampling use information from the current stage \( \nu \) to improve the solution at stage \( \nu + 1 \). Spanier [2] has formulated this two methods.

4.1 Sequential Correlated Sampling

Halton described a method to solve matrix problems based on a form of correlated sampling. This method has been extended to continuous transport problems (see [3]). In Equation (8), choosing \( S^0 = S, S^{(\nu)} = d^{(\nu)} \), with
\[ d^{(\nu)} = d^{(\nu-1)} + K^{(\nu)} \hat{\Phi}^{(\nu-1)} - \hat{\Phi}^{(\nu-1)} \]
\[ d^{(0)} = S \]

where \( \hat{\Phi}^{(\nu-1)} \) is an approximate solution to
\[ \Phi = K \Phi + d^{(\nu-1)} \]

produces Halton’s sequential correlated sampling. This reduced source generates a correction term at each stage that is added to the solution at the previous stage.

4.2 Adaptive Importance Sampling

Unlike the sequential correlated sampling method, this method uses a kernel, \( K \), and the source term, \( S \), at every stage as a function of \( \Phi \) at the previous stage and the original kernel and source, respectively
\[ K^{(\nu)} = K^{(\nu)}(\Phi^{(\nu)}, K) \]
\[ S^{(\nu)} = S^{(\nu)}(\Phi^{(\nu)}, S) \]
The method uses knowledge of the solution obtained in stage \( \nu \) as an importance function to modify the construction of random walks in stage \( \nu + 1 \) and generates the full solution at each stage, rather than merely an additive correction. For the details of this method, refer to [4].

Certain non-standard features of the adaptive importance sampling method facilitate its implementation. These are:

- To solve the transport problem, the need arises to sample source points \( x_1 \) that are distributed according to the following probability density functions:

\[
\hat{S}_i(x_1) = \frac{\hat{\phi}(x_1) S_i^*(x_1)}{\int_0^\infty \hat{\phi}(x) S_i^*(x) \, dx}
\]

where \( S_i^* \) is given by Equation (7). These probability density functions are not of constant sign and require either sampling from a signed measure (see [4,7]) or conversion to a related problem with a nonnegative probability density function. We have used the latter strategy, solving a related problem with a positive source and recovering the coefficients algebraically.

- Notice also that the source \( \hat{S}_i \) in Equation (9) depends on the coefficient index \( i \). This means that there is a different probability density function for each coefficient, which would normally require separate and independent sets of random walks for each coefficient. To circumvent the need to generate separate sets of random walks for each coefficient, we use rejection sampling methods to achieve correlation in these sets of random walks (see explanation of this in [4]).

- To increase the efficiency of our rejection sampling scheme, we sample uniformly under the curve

\[
B(x) = \max(f_1(x), \cdots, f_N(x))
\]

where the \( f_i(x) \) are the probability density functions derived from the sources \( \hat{S}_i \). If the points are uniformly sampled under \( B(x) \), then they also will be uniform under each probability density function, but the rejection rate will be lowered by enclosing the graphs of all the \( f_i(x) \) in the smallest possible subset of the plane. This improvement increases the efficiency from approximately 46\% to approximately 97\% for the problems surveyed in this paper.

5 Parametrization of Model Transport Problems for Comparison

When comparing two algorithms, it is difficult to judge the overall effectiveness of one algorithm over the other using a small set of isolated test problems. The test problems may not cover a wide range of possibilities and/or
may inadvertently be prescribed in such a way that one method is better suited to solve them than the other. To create a testbed to fairly compare adaptive importance sampling and sequential correp-
pling, a family of representative test problems was identified. Ten
of model problems was constructed to include problems of varying
One aspect of the degree of difficulty in solving the monodirectional
problem can be described by the Neumann series solution to the
The solution to (3) can be written in the following form:

\[ \Phi = (I - K)^{-1}S = S + KS + K^2S + \ldots \]

As we have consistently done, we require \( ||K|| < 1 \) to ensure convergence this series and existence of a unique solution. The series will converge (and in this sense, represent an easier problem) for smaller \( ||K|| \), when the amount of scattering in the problem is small and the absorption is large. Conversely, a more difficult problem results when the amount of scattering is large and absorption is small. The conditional probability of scattering, given either a scattering or absorbing event, is a parameter that affects the difficulty of the problem.

We have defined affective and ineffective parameters relative to the monodirectional problem. The affective parameters, those parameters that presumably determine the difficulty of the problem, were defined as follows:

- Principal affective parameter:

\[ B_1 = \frac{\Sigma_s}{\Sigma_t} \]

where \( \Sigma_s \) is the scattering cross section and \( \Sigma_t \) is the total cross section.

- Secondary affective parameter:

\[ B_2 = T \ast \Sigma_t \]

where \( T \) is the total width of the slab.

Since the probability per collision of absorption is defined as \( 1 - B_1 \), the parameter \( B_1 \) determines the amount of absorption in the problem. Higher absorption defines an easier problem, lower absorption defines a harder problem. Conversely, lower scattering defines an easier problem, higher scattering a harder problem. Thus setting \( B_1 \) to larger values will generate easier problems.

The secondary affective parameter, \( B_2 \), is the length of the slab multiplied by the total cross section, or the optical thickness of the slab. Thick

\(^1\) These terms are meant to describe qualitative features and are taken from the work of Genz \( [8] \) where they were used, as far as we are aware, for the first time in such a context.
indicate harder problems because fewer particles penetrate deeply into the slab and consequently, less information is available to estimate the particle density, or flux \( \Phi \), there. Thus setting \( B_2 \) to higher values should generate harder problems.

An unaffection parameter is one that can presumably be chosen at random from a specified range of values without having more than a small effect on the difficulty of the problem. The single unaffection parameter is the total cross section of the slab, \( \Sigma_t \). The average distance between collisions in this monodirectional problem is \( \frac{1}{\Sigma_t} \), but this value alone does not determine the difficulty of the problem because the length of the slab, \( T \), has to be taken into consideration as well. This is captured in the parameter \( B_2 \). Both affective parameters chosen in this way are dimensionless. With fixed \( B_1 \) and \( B_2 \), choosing \( \Sigma_t \) randomly from a specified range creates a representative set of problems whose difficulties are then approximately the same and whose results can then be averaged to obtain meaningful comparisons of the two methods used to solve them.

For the numerical study, we examined a wide range of parameters \( B_1, B_2 \). Here we only report results using two different combinations of \( B_1 \) and \( B_2 \) for the comparison: \( B_1 = 0.1, B_2 = 1 \), and \( B_1 = 0.9, B_2 = 10 \), corresponding to problems of modest and intermediate difficulty, respectively. The additional combinations that were examined, however, support the same general conclusions reported in the next section.

6 Numerical Comparisons

The first comparison we performed used the terminal estimator (see [6] for a full discussion of transport theory estimators) for each of the two methods. We used this estimator to provide a consistent comparison and because the theory of importance sampling underlying zero variance choices for the terminal estimator is well known. For our second comparison we allowed the sequential correlation sampling method to make use of the generalized track length estimator [6]. This is a more powerful estimator but when this study was carried out, we had not yet developed an algorithm for accomplishing zero variance making use of this estimator so we did not attempt to incorporate this more complex estimator into the adaptive importance sampling code.

In all of the resulting graphs, \( \log_{10} |\text{error}| \) versus the number of adaptive stages is plotted for the two combinations of \( B_1 \) and \( B_2 \): \( B_1 = 0.1, B_2 = 1 \), and \( B_1 = 0.9, B_2 = 10 \). Each curve was obtained by averaging the errors from 10 different test problems, for each of which \( \Sigma_t \) was selected randomly from a uniform distribution on \([1, 10]\). For all problems run, eight Legendre coefficients were used in representing the approximate solution \( \hat{\Phi} \). The test point at which the approximate solution is evaluated is a point three-fourths of the way through the slab; i.e., at \( x = 0.75 \times B_2 \). Of course, the code we use is capable of producing the solution at any value of the independent variable
\( x, 0 \leq x \leq T \), since the adaptive algorithms actually produce a solution represented as a finite linear combination of Legendre polynomials.

In our comparison study we were primarily interested in comparing the rates of convergence of the two methods. To ensure that we had a consistent testbed for both methods, the number of utilized random walks was chosen to be the same for the two methods. The descriptor utilized was needed because, in contrast with the sequential correlated sampling method, which uses all random walks generated, the adaptive importance sampling algorithm uses the rejection method in its sampling scheme and retains only a subset of all random walks generated (see [4]).

6.1 First Comparison Results

Figures 2 and 3 show the results from the first comparison using the adaptive importance sampling estimator for both methods for the combinations \( B_1 = 0.1, B_2 = 0.9, B_1 = 0.9, B_2 = 10 \), respectively. In the first plot shown in both figures, the threshold number of random walks needed for either method to obtain geometric convergence and used this number of random walks per stage. The figures show that with only 200 random walks, the adaptive importance sampling attains stable geometric convergence with this same number of random walks per stage, the sequential correlated sampling iterations diverge. We increased the number of random walks for the sequential correlated sampling until it attained an equivalent convergence rate and found that 40,000 random walks per stage were needed for the case \( B_1 = 0.1, B_2 = 1 \), and 20,000 random walks per stage were needed for the case \( B_1 = 0.9, B_2 = 10 \).

In the second plot shown in both figures, we used 20,000 random walks per stage for each method and compared the results. In both cases, the adaptive importance sampling attains a faster rate of convergence than the sequential correlated sampling. These results show the advantage of using adaptive importance sampling in terms of the rate of geometric convergence and the number of random walks invested.

The computing time it took each method to run the 10 test problems, however, varied greatly. Overall, the sequential correlated sampling required less computing time than the adaptive importance sampling algorithm. For the easiest problem, for which \( B_1 = 0.1, B_2 = 1 \) with 10,000 random walks per stage, the adaptive importance sampling took to complete 10 test problems, whereas to achieve an equivalent convergence rate, the sequential correlated sampling method needed 300,000 random walks per stage but took only \( \frac{1}{2} \) hour. For the hardest problem where \( B_1 = 0.9 \) with 20,000 random walks, the adaptive importance sampling took to achieve an equivalent convergence rate, the sequential correlated sampling method needed 30,000 random walks per stage but took only \( \frac{1}{4} \) hour.
Fig. 2. Sequential correlated sampling (terminal estimator) vs. adaptive importance sampling (terminal estimator) for modest problem difficulty.
sampling method needed 800,000 random walks but took only 1\(\frac{1}{4}\) hours. This is primarily because the adaptive importance sampling algorithm determines the next collision point using a time-intensive numerical integral inversion procedure that the sequential correlated sampling does not need to use. These timing results reveal the advantage of using sequential correlated sampling with respect to computer time invested.

6.2 Second Comparison Results

Figure 4 shows the results from the second comparison using the terminal estimator for the adaptive importance sampling method and the track length estimator for the sequential correlated sampling method. The combinations \(B_1 = 0.1, B_2 = 1\), and \(B_1 = 0.9, B_2 = 10\) are again used. In this case, the results of 100 test problems were averaged by varying the unafffective parameter \(\Sigma_i\) uniformly in the interval \([1, 10]\).

The first figure illustrating the results for the easier test case shows that the sequential correlated sampling algorithm now attains a faster rate of convergence than the adaptive importance sampling method. In the second figure, however, illustrating the harder test case, just the reverse is indicated. Thus, when the more “information rich” track length estimator is used for sequential correlated sampling but the adaptive importance sampling method is restricted to use the terminal estimator, no easy conclusion can be reached concerning relative rates of convergence.

For this second series of tests the sequential correlated sampling algorithm again used much less computer time than the importance sampling-based method. For the easier problem, sequential correlated sampling took 2 hours and adaptive importance sampling required 3 hours. For the harder test case, these figures rose to 2 1/4 hours and 60 hours, respectively.

7 Conclusions

The results of this paper may be summarized as follows:

1. When both the correlated sampling and importance sampling algorithms employ the terminal estimator, the adaptive importance sampling methods seems always to attain more rapid geometric convergence (error reduction per adaptive stage) than sequential correlated sampling when the same numbers of random walks per stage are used.

2. The correlated sampling method is considerably easier to program and involves much less computation per collision; therefore, it processes random walks considerably faster than the importance sampling method.

3. Since the correlated sampling method is basically an analog method (i.e., the kernel used to move random walks from collision to collision is the analog kernel), it can be programmed to make use of any unbiased transport estimator, whereas the importance sampling method must use an estimator
Fig. 4. Sequential correlated sampling (track estimator) vs. adaptive sampling (terminal estimator) for varying problem difficulty.
for which a zero variance theory exists. Taking advantage of this fact, one can use the generalized track length estimator with the sequential correlated sampling algorithm. When this is done, the improvement in rate of convergence achieved with this method enables it to overtake the importance sampling method (still based on the terminal estimator) in terms of rate of geometric convergence for “easy” problems, but importance sampling seems to retain an advantage in convergence rate for “hard” problems.

These results suggest that neither method is always better to use than the other and that the preference for one method over the other will be strongly problem-dependent. Work is currently underway to develop theoretically-based descriptions of those transport problems for which a strong preference of one method over the other could be assumed.

References

2. Spanier, J.: Geometrically Convergent Learning Algorithms for Global Solutions of Transport Problems. this volume