Adaptive Importance Sampling Algorithms for Transport Problems

Yongzeng Lai & Jerome Spanier
Department of Mathematics
Claremont Graduate University
925 N. Dartmouth Ave.
Claremont, CA 91711, USA
Email: yongzeng.lai@cgu.edu & jerome.spanier@cgu.edu

Abstract. We describe how importance sampling methods may be applied adaptively to the solution of particle transport problems. While the methods apply quite generally, we have so far studied in detail only problems involving planar geometry.

The technique used is to represent the global solution of the transport equation as a linear combination of appropriately chosen basis functions and estimate a finite number of the resulting coefficients in stages. Each stage processes a fixed number of random walks making use of an importance function that has been determined from the previous stage. Special methods have been developed for importance sampling the resulting source and kernel, and some of these will be described. Numerical results exhibiting geometric convergence for the resulting algorithm will be presented.

1 Introduction

Monte Carlo methods are very useful in obtaining local information (e.g., reaction rates in subregions of a given problem phase space) about transport problems. However, their use in obtaining global solutions of transport problems is of very recent origin. The limitations of conventional probabilistic convergence for (pseudorandom) Monte Carlo methods mean that processing millions of random walks to obtain solutions accurate to perhaps 10% or more might require hours or even days on the fastest current computer platforms. For this reason, learning algorithms, in which information is steadily being improved from one adaptive stage to the next, are required to obtain sufficiently accurate global solutions.

Recently, in [2] and [5], such learning algorithms have been successfully applied to model transport problems, and the goal of achieving geometric convergence has been reported there. For example, in [2], a sequential correlated sampling method is responsible for the monotone error reduction from stage to stage characteristic of geometrically convergent algorithms. In this paper we apply an importance sampling method adaptively to the same problems that were described in [2], and we obtain similar indications of geometric convergence. Our description here will focus on some of the specifics of the rather complex algorithm and numerical results obtained with its use.
2 Derivation of \((x, \mu)\) Transport Problem

A one dimensional, full angle, planar particle transport problem can be described by the following integro-differential equation (see [1]):

\[
\frac{\partial \Phi}{\partial x} + \sigma \Phi = \frac{c}{2} \int_{-1}^{1} \Phi(x, \nu) d\nu + Q,
\]

where \(\Phi(x, \mu)\) is the angular flux, \(Q(x, \mu)\) is the internal source, \(\sigma\) is the total macroscopic section and \(c\) is the mean number of secondary particles per primary particle. If we define the scalar flux as usual by

\[
\varphi(x) = \int_{-1}^{1} \Phi(x, \nu) d\nu,
\]

then equation (1) can be transformed into

\[
\begin{align*}
\Phi(x, \mu) = \left\{
\begin{array}{ll}
  e^{-\sigma \frac{x}{\mu}} \Phi(0, \mu) + \int_{0}^{x} e^{\sigma \frac{y-x}{\mu}} \frac{Q(y, \mu)}{\mu} dy + \int_{0}^{x} e^{\sigma \frac{y-x}{\mu}} \frac{c}{2\mu} \varphi(y) dy, & \mu > 0, \\
  e^{\sigma \frac{x}{\mu}} \Phi(T, \mu) + \int_{x}^{T} e^{\sigma \frac{y-x}{\mu}} \left(-\frac{Q(y, \mu)}{\mu}\right) dy + \int_{x}^{T} e^{\sigma \frac{y-x}{\mu}} \left(-\frac{c}{2\mu} \varphi(y)\right) dy, & \mu < 0.
\end{array}
\right.
\end{align*}
\]

(2)

Converting to a vector formulation (for simplicity, we take \(Q \equiv 0, \Phi(0, \mu) = q_0, \Phi(T, \mu) = q_1\)), we obtain an equivalent integral equation form

\[
\Phi(X) = \int_{0}^{T} \int_{-1}^{1} K(X, Y) \Phi(Y) dY + S(X),
\]

(3)

where \(X = (x, \mu), Y = (y, \nu), dY = dy d\nu,\)

\[
K(X, Y) = K(x, \mu, y, \nu) = \left\{
\begin{array}{ll}
  \frac{c}{2\mu} e^{-\sigma \frac{x-y}{\mu}}, & x \geq y, \mu > 0, \\
  \frac{c}{2\mu} e^{\sigma \frac{y-x}{\mu}}, & x < y, \mu < 0, \\
  0, & \text{otherwise}.
\end{array}
\right.
\]

(4)

and

\[
S(X) = S(x, \mu) = \left\{
\begin{array}{ll}
  e^{-\sigma \frac{x}{\mu}} \Phi(0, \mu), & \mu > 0 \\
  e^{\sigma \frac{x}{\mu}} \Phi(T, \mu), & \mu < 0
\end{array}
\right. x \in [0, T].
\]

(5)

If we know \(\varphi(\cdot)\), then we can find \(\Phi(x, \mu)\) by substitution of \(\varphi(\cdot)\) into the right hand side of equation (2) (see [5] for a more general description of this technique).

Let
\[ \varphi(x) = \sum_{i=0}^{\infty} a_i P_i \left( \frac{2x}{T} - 1 \right), \]  

where \( P_i \) is the Legendre polynomial of degree \( i \) on \([-1,1]\). Then,

\[ a_i = \frac{2i + 1}{T} \int_{0}^{T} \varphi(x) P_i \left( \frac{2x}{T} - 1 \right) dx = \frac{2i + 1}{T} \int_{0}^{T} \int_{-1}^{1} \Phi(x, \mu) P_i \left( \frac{2x}{T} - 1 \right) dx d\mu. \]

With

\[ \tilde{P}_i(x, \mu) = \frac{1 + P_i \left( \frac{2x}{T} - 1 \right)}{2T}, \]

\[ \tilde{a}_i = (2i + 1) \int_{0}^{T} \int_{-1}^{1} \Phi(x, \mu) \tilde{P}_i(x, \mu) dx d\mu, \]

\( \tilde{P}_i \) becomes a probability density function on \([0,T] \times [-1,1]\):

\[ \int_{0}^{T} \int_{-1}^{1} \tilde{P}_i(x, \mu) dx d\mu = 1, \]

and

\[ \tilde{a}_i = \frac{2i + 1}{2T} \int_{0}^{T} \int_{-1}^{1} \Phi(x, \mu) \left[ 1 + P_i \left( \frac{2x}{T} - 1 \right) \right] dx d\mu \]

\[ = \frac{2i + 1}{2} \int_{0}^{T} \int_{-1}^{1} \Phi(x, \mu) dx d\mu + \frac{12i + 1}{2T} \int_{0}^{T} \int_{-1}^{1} \Phi(x, \mu) P_i \left( \frac{2x}{T} - 1 \right) dx d\mu \]

\[ = \frac{2i + 1}{2} a_0 + \frac{1}{2} a_i. \]

From this we find

\[ \tilde{a}_0 = a_0, \quad a_i = 2(\tilde{a}_i - \frac{2i + 1}{2} a_0) = 2\tilde{a}_i - (2i + 1)a_0 \]

so that the coefficients \( a_i \) may be recovered from \( \tilde{a}_i \), and vice versa.

To estimate \( \tilde{a}_i \), we use the equation dual to (3)

\[ \Phi^*(X) = \int_{0}^{T} \int_{-1}^{1} K^*(X,Y) \Phi^*(Y) dY + S^*(X), \]

where \( K^*(X,Y) = K(Y,X) \) and \( S^* \) is an adjoint source to be determined by the problem to be solved. For example, setting \( S^*_i(X) = (2i + 1)\tilde{P}_i(X) \), we find
\[ \tilde{a}_i = \langle \Phi, S^*_i \rangle = \langle S, \Phi^*_i \rangle, \]

and the theory of zero variance importance sampling [3] can be applied to the estimation of \( \tilde{a}_i \). In particular, we can use the following terminal type estimator [3] to estimate \( \tilde{a}_i \):

\[ \xi_i(\alpha) = \frac{S^*(X_1)}{p^*_1(X_1)} \frac{K^*(X_2, X_1)}{p(X_2, X_1)} \ldots \frac{K^*(X_k, X_{k-1}) S(X_k)}{p(X_k, X_{k-1}) p(X_k)}, \]

where

\[ p^*_1(X) = \frac{S^*_i(X) \Phi(X)}{\int_0^T \int_{-1}^1 S^*_i(X) \Phi(X) dX} \]

is the probability density function to sample initial points \( X_1 \) of random walks \( \alpha = (X_1, \ldots, X_k) \);

\[ p(X, Y) = \frac{K^*(X, Y) \Phi(X)}{\int_0^T \int_{-1}^1 K^*(X, Y) \Phi(X) dX + S(Y)} = \frac{\int_0^T \int_{-1}^1 K^*(Y, X) \Phi(X) dX + S(Y)}{\int_0^T \int_{-1}^1 S^*_i(X) \Phi(X) dX + S(Y)} \]

is the transition probability from \( Y \) to \( X \); and

\[ p(Y) = 1 - \int_0^T \int_{-1}^1 p(X, Y) dX = \frac{S(Y)}{\int_0^T \int_{-1}^1 K^*(Y, X) \Phi(X) dX + S(Y)} \]

is the absorption probability at \( Y \).

If we now assume that we have estimates of each coefficient \( \tilde{a}_i^{(0)} \) at the 0-th stage, beginning with stage \( \lambda \geq 1 \), we can use a rejection sampling method to sample first collision points \( X_1 \). To this end, introduce

\[ M = \max_{0 \leq i \leq n-1} \sup_X p^*_1(X), X = (x, \mu) \in [0, T] \times [-1, 1]. \]

Then

\[ p^*_1(X) = \frac{S^*_i(X) \Phi(X)}{\int_0^T \int_{-1}^1 S^*_i(X) \Phi(X) dX} = \frac{S^*_i(X) \Phi(X)}{\tilde{a}_i} = \frac{(2i + 1) \tilde{a}_i(X) \Phi(X)}{\tilde{a}_i} \]

\[ = \frac{(2i + 1) [1 + P_i(\frac{T}{2} \tilde{a}_i - 1)] \Phi(X)}{2T \tilde{a}_i} \leq \frac{(2i + 1) \Phi(X)}{T \tilde{a}_i}. \]

Now we define

\[ f_i(X) = \frac{p^*_1(X)}{M}. \]
The functions \( f_i(X) \) may then be used to generate very efficiently the initial collision points \( X_1 \) of the needed random walks. The random walk originating at \( X_1 \) can be used to estimate the coefficients \( \tilde{a}_i \) for every index \( i \) that is accepted as a result of the rejection sampling method. Details of this procedure can be found in [4].

3 Importance Sampling Algorithm

In order to estimate \( \tilde{a}_i^{(\lambda)} \) by \( \xi_i(\alpha) \), i.e., by (15): generate \( W \) random walks, and for each random walk, carry out the following steps.

1. Apply the rejection method to sample the source (first) point \( X_1 \) of the random walk.
2. Test for absorption at \( X_1 \) : generate a pseudorandom number \( \rho \in (0, 1) \) uniformly. If

\[
\rho \geq \frac{\int_0^T \int_{-1}^1 p(X, X_1) dX}{\int_0^T \int_{-1}^1 K(X_1, X) \Phi^{(\lambda-1)}(X) dX + S(X_1)}
\]

then the random walk terminates with absorption at \( X_1 \) and we compute

\[
\xi_i \leftarrow \xi_i \times \frac{S(X_1)}{p(X_1)} = \xi_i \times \left( \int_0^T \int_{-1}^1 K(X_1, X) \Phi^{(\lambda-1)}(X) dX + S(X_1) \right)
\]

and

\[
\tilde{a}_i^{(\lambda)} \leftarrow \tilde{a}_i^{(\lambda)} + \xi_i.
\]

However, if

\[
\rho < \frac{\int_0^T \int_{-1}^1 p(X, X_1) dX}{\int_0^T \int_{-1}^1 p(X, X_1) dX + S(X_1)}
\]

then the random walk suffers a scattering collision at \( X_1 \) and the next event is determined at step 3.

3. Find a new angle \( \arccos(\mu_2) \) and distance traveled \( x_2 \). To this end, define

\[
p_{\bar{p}}(x, \mu; x_1, \mu_1) = \frac{p(x, \mu; x_1, \mu_1)}{\int_0^T \int_{-1}^1 p(x, \mu; x_1, \mu_1) dx d\mu}.
\]

\[
p_{\mu}(\mu; x_1, \mu_1) = \int_0^T \bar{p}(x, \mu; x_1, \mu_1) dx,
\]

and
\[ p_x(x; z_1, \mu_1 | \mu) = \frac{\tilde{p}(x; \mu; z_1, \mu_1)}{p_\mu(\mu; z_1, \mu_1)} = \frac{\tilde{p}(x; \mu; z_1, \mu_1)}{\int_0^1 \tilde{p}(x; \mu; z_1, \mu_1) \, dz}. \]  

(28)

Sample independent pseudorandom random numbers \( \rho_\mu \) and \( \rho_x \) in \((0, 1)\) uniformly. Solve

\[
\rho_\mu = \int_{-1}^{\mu_2} \int_{-1}^{\mu_2} \int_0^T p(x; \mu; z_1, \mu_1) \, dz \, d\mu \\
= \int_{-1}^{\mu_2} \int_0^T \int_{-1}^{\mu_2} p(x; \mu; z_1, \mu_1) \, dz \, d\mu
= \int_{-1}^{\mu_2} \int_0^T K(X_1, X) \Phi(X) \, dX
= \int_{-1}^{\mu_2} \int_0^T K(X_1, X) \Phi(X) \, dX
\]

for \( \mu_2 \). Then solve

\[
\rho_x = \left\{
\begin{array}{ll}
\int_0^{\mu_2} p_x(x; z_1, \mu_1 | \mu_2) \, dx, & 0 < x_2 < x_1, \text{if } \mu_2 > 0, \\
\int_{z_2}^{\mu_2} p_x(x; z_1, \mu_1 | \mu_2) \, dx, & x_1 < x_2 < T, \text{if } \mu_2 < 0,
\end{array}
\right.

= \begin{cases}
\int_0^{\mu_2} K(x_1, x_2; \mu_1, \mu_2) \Phi(x_1, x_2) \, dx, & 0 < x_2 < x_1, \text{if } \mu_2 > 0, \\
\int_{x_2}^{\mu_2} K(x_1, x_2; \mu_1, \mu_2) \Phi(x_1, x_2) \, dx, & x_1 < x_2 < T, \text{if } \mu_2 < 0,
\end{cases}
\]

(29)

for \( x_2 \). Compute

\[
\xi_i \leftarrow \xi_i \star \frac{K^*(X_2, X_1)}{p(2, 1)} = \xi_i \star \frac{\int_0^T \int_{-1}^1 K(X_1, X) \Phi(x_2) \, dX + S(X_1)}{\Phi(x_2) \Phi(x_1)}
\]

\[
= \left\{
\begin{array}{ll}
\xi_i \star \int_0^1 \int_{-1}^1 \Phi(x_1, x_2) \, d\mu_1 \Phi(x_2, \mu_2) + S(x_1, \mu_1), & \mu_1 > 0 \\
\xi_i \star \int_{-1}^1 \int_{-1}^1 \Phi(x_1, x_2) \, d\mu_1 \Phi(x_2, \mu_2) + S(x_1, \mu_1), & \mu_1 < 0
\end{array}
\right.
\]

(31)

set \( X_1 = X_2 \), go to step 2.
After completing all $W$ random walks, we take the average of $\xi_i$, use it as the estimate of $a_i^{(\lambda)}$, and recover $a_i^{(\lambda)}$ from (12).

4 Difficulties

The algorithm described above encounters substantial difficulties in implementation. For example, in Step 3 of the above algorithm, after we have determined that the particle scatters at $X_1 = (x_1, \mu_1)$, we need to find a new angle $\arccos(\mu_2)$ and distance traveled $x_2$. Most of the computation time is spent on inverting the two kinds of integrals (29) and (30), which involve the following integrals that have removable singularities:

1. 

$$\int_0^{\mu_2} \frac{\mu}{\mu - \mu_1} (e^{\frac{-x_1}{\mu}} - e^{\frac{-x_1}{\mu_1}}) d\mu, \quad x_1, \mu_1 > 0. \tag{32}$$

If we let

$$f_1(\mu, \mu_1, x_1, \sigma) = \frac{\mu}{\mu - \mu_1} (e^{\frac{-x_1}{\mu}} - e^{\frac{-x_1}{\mu_1}}), \mu \neq 0, \mu_1,$$

then

$$\lim_{\mu \to 0^+} f_1(\mu, \mu_1, x_1, \sigma) = 0,$$

and

$$\lim_{\mu \to \mu_1} f_1(\mu, \mu_1, x_1, \sigma) = \frac{\sigma x_1}{\mu_1} e^{-\frac{x_1}{\mu_1}}.$$

2. 

$$\int_0^{\mu_2} \frac{d\mu}{\mu - \mu_1} \int_0^{x_1} \left( e^{\frac{x_1}{\mu} e^{\frac{y}{\mu_1}} e^{\frac{y-x_1}{\mu_1}} - e^{\frac{x_1}{\mu_1}}} \right) P_i \left( \frac{2y}{T} - 1 \right) dy, \quad x_1, \mu_1 > 0. \tag{33}$$

where $P_i(x)$ is the $i$th Legendre polynomial. If we let

$$f_2(\mu, \mu_1, x_1, \sigma) = \frac{1}{\mu - \mu_1} \int_0^{x_1} \left( e^{\frac{x_1}{\mu} e^{\frac{y}{\mu_1}} e^{\frac{y-x_1}{\mu_1}} - e^{\frac{x_1}{\mu_1}}} \right) P_i \left( \frac{2y}{T} - 1 \right) dy,$$

then

$$\lim_{\mu \to 0^+} f_2(\mu, \mu_1, x_1, \sigma) = \frac{1}{\mu_1} \int_0^{x_1} e^{\frac{x_1}{\mu_1}} P_i \left( \frac{2y}{T} - 1 \right) dy,$$

and

$$\lim_{\mu \to \mu_1} f_2(\mu, \mu_1, x_1, \sigma) = \frac{\sigma}{\mu_1^2} \int_0^{x_1} (y - x_1) e^{\frac{x_1}{\mu_1}} P_i \left( \frac{2y}{T} - 1 \right) dy.$$
The integral (32) cannot be evaluated in closed form exactly. Thus approximate formulas must be used. We used the following composite Weddle's rule with error of order 8.

**Lemma:** If \( f \in C^8[a, b] \), \( b > a \), then

\[
\int_a^b f(x) \, dx \approx \frac{h}{140} \left[ w(0) f(a) + w(6) f(b) + 2w(6) \sum_{i=1}^{n-1} f(a + 6i h) \right. \\
\left. + \sum_{j=1}^{5} w(j) \sum_{i=0}^{n-1} f(a + (6i + j) h) \right],
\]

where \( h = \frac{b-a}{N} \), \( N = 6n \), \( w(0) = w(6) = 41 \), \( w(1) = w(5) = 216 \), \( w(2) = w(4) = 27 \), and \( w(3) = 272 \). The error of the approximation is \( \frac{9(b-a)}{3400} h^8 f^{(8)}(c) \) for some \( c \in (a, b) \).

The lemma can be proved based on Weddle's rule (see [7]). For the integral (33), we can use a recursive formula to evaluate the integral with respect to \( x \) for fixed \( \mu \), and then use the composite Weddle's rule to find the numerical value of the integral (33). The recursive formula we used is

\[
\int_a^b e^{f+cz} P_i(x) \, dx = \frac{1}{c} \left[ e^{f+cz} (P_i(x) - P_{i-2}(x)) \right]_a^b - \frac{2i - 1}{c} \int_a^b e^{f+cz} P_{i-1}(x) \, dx \\
+ \int_a^b e^{f+cz} P_{i-2}(x) \, dx, \quad i \geq 2,
\]

which is precise for \( i \leq 15 \). However, for values of \( i \) larger than this, the recursion suffers serious loss of precision and cannot be used unless prohibitively many significant digits are carried in the algorithm. This is very similar to the numerical difficulties reported in [2].

5 Numerical Results

The numerical results we obtained (see Figure 1) when implementing the algorithm described above demonstrate quite convincingly that geometric convergence is obtained for the full \((x, \mu)\) problem (1), even with fewer than 11 Legendre polynomials. That is, the algorithm is designed to produce geometric convergence to the best approximation to the full solution available within the limitations of the (finite) basis function expansion used to obtain it. For example, Figure 2 provides three graphs of the scalar flux \( \phi(\cdot) \) obtained with our algorithm when four, six, and eleven Legendre polynomials, respectively, are used in the flux representation. Only minor differences are observable near \( x = 0 \) and \( 1 \) since, away from these boundaries, the solution \( \phi(\cdot) \) is very well approximated by a low degree polynomial.
The approximate numerical solutions that we obtain using importance sampling are also very close (agreeing to about 1 part in $10^4$) to those reported in [2] where the sequential correlated sampling method was used to obtain them. The data used in our test are: $T = 1.0, \varphi_0 = 1.0, \varphi_1 = 0.0, \sigma = 1.0, c = 0.5$, number of Legendre polynomials = 11, test point (where errors are estimated) $x_0 = 0.75, \mu_0 = 0.75$, number of adaptive stages = 30, number of random walks per stage (abbreviated as RW.stage in the Figure 1) = 200, 400, respectively. In Figure 1, the residual is defined as the absolute value of the difference between the left hand side and the right hand side of the equation (1) when the terms are evaluated at the approximate solution. In this example, for which no exact solution is easily available, we use the residual as a measure of the error.

We also obtained excellent results (not reported here) for a variety of monodirectional transport problems using the same method.

6 Summary and Conclusions

We have applied importance sampling to the estimation of global solutions of transport problems described by certain kinds of integral equations. The key idea is to expand the solution of the problem as a linear combination of appropriately defined basis functions, truncate the sum, and estimate each coefficient by a learning algorithm. Clear evidence of geometric convergence is obtained when sufficient care is taken to avoid some of the numerical difficulties that were encountered when a naive choice of basis functions was made. A completely different method, based on a sequential application of correlated sampling, can be found in [2]. In that reference, the correlated sampling method is applied to the same transport problem treated here, with equally good results in terms of geometric convergence. However, the inherently more complicated nature of the importance sampling implementation that we describe here requires more computing time per random walk than does the correlated sampling algorithm. On the other hand, there is evidence [6], that the importance sampling algorithm produces greater rates of convergence than the correlated sampling algorithm, at least for sufficiently difficult problems.

Work continues on extending these ideas to more general transport problems, on making more efficient choices of basis functions, and on developing a satisfactory error analysis.

References

2. Kong, R., Spanier, J.: Sequential Correlated Sampling Method for Particle Transport Problems. this volume
Fig. 1. Geometric convergence for slab geometry transport problem

Fig. 2. Graphs of Scalar Flux Based on 4, 8, and 11 Legendre Polynomials

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