Monte Carlo Methods for Flux Expansion Solutions of Transport Problems

Jerome Spanier*
Claremont Graduate University, Department of Mathematics
Claremont, California 91711-3988

Received April 6, 1998
Accepted December 8, 1998

Abstract—Adaptive Monte Carlo methods, based on the use of either correlated sampling or importance sampling, to obtain global solutions to certain transport problems have recently been described. The resulting "learning algorithms" are capable of achieving geometric convergence when applied to the estimation of a finite number of coefficients in a flux expansion representation of the global solution. However, because of the nonphysical nature of the random walk simulations needed to perform importance sampling, conventional transport estimators and source sampling techniques require modification to be used successfully in conjunction with such flux expansion methods. It is shown how these problems can be overcome.

First, the traditional path length estimators in wide use in particle transport simulations are generalized to include rather general "detector" functions (which, in this application, are the individual basis functions chosen for the flux expansion). Second, it is shown how to sample from the signed probabilities that arise as source density functions in these applications, without destroying the zero variance property needed to ensure geometric convergence to zero error.

I. INTRODUCTION

An unconventional application of Monte Carlo is to use adaptive Monte Carlo methods to estimate the transport flux or collision density everywhere in phase-space. This may be accomplished by representing the flux, for example, as an infinite linear combination of suitably chosen basis functions and observing that the coefficients of combination can then be represented as functionals of the flux. Thus, if

\[ F(P) = \sum_{i=1}^{\infty} a_i B_i(P) \tag{1} \]

represents the flux as a linear combination of basis functions \( B_i(P) \) (here \( P \in \Gamma \) is a phase-space vector) and if these basis functions are chosen to be orthogonal to each other,\(^\star\) then the combination coefficients \( a_i \) are themselves expressible as weighted integrals of the flux, with weights that depend on the normalization of the \( B_i \):

\[ a_i = K_i \int_{\Gamma} B_i(P) F(P) \, dP \tag{2} \]

and the constants \( K_i \) are determined by the exact choice of the functions \( B_i \). Under these circumstances, the functions \( B_i(P) \) may be chosen for mathematical convenience and may have little to do with the physics. For example, if the \( B_i(P) \) are represented as a product of one-dimensional functions, each of which is chosen to be, for example, a Legendre polynomial on a suitably defined interval, then the \( B_i(P) \) will not in general be of constant sign, and it will no longer be possible to interpret the coefficients \( a_i \) as conventional reaction rates.

Several complications arise out of this unconventional application. One complication occurs because of the need to extend the traditional transport estimators to this unfamiliar setting. We show in this paper how this is to be done by developing appropriate unbiased generalizations of the conventional estimators that reduce to the familiar formulas when physically defined reaction rates are the objects of estimation.

A second issue requiring special consideration results from the appearance of the basis functions as part
of the source distribution called for by the importance sampling algorithm. The fact that the basis functions may well change sign frequently over their intervals of definition raises questions about whether this problem could be avoided and, if not, how to do the sampling correctly. We prove that the zero variance prescription is essentially unique (and thus, sampling from the source densities in which the basis functions occur cannot be avoided), and we show how the nonstandard sampling could be carried out without destroying the zero variance property.

II. GENERALIZED PATH LENGTH ESTIMATION

It has long been appreciated that a variety of quite natural random variables play useful roles in Monte Carlo solutions of the transport equation. For example, in Sections 2.5–2.7 of Ref. 1, a rather complete description is given of the three most basic such random variables—the terminal, or absorption, estimator; the collision estimator; and the track length, or path length, estimator—used in estimating reaction rates:

$$I = \int g(P) \Psi(P) \, dP$$.

where the collision density \( \Psi(P) \) satisfies the transport equation

$$\Psi(P) = \int K(P,Q) \Psi(Q) \, dQ + S(P)$$

and the detector function \( g \) is essentially the probability that a random walk that has progressed to the state \( P \in \Gamma \) will cause a reaction (e.g., absorption) at \( P \) of the type described by the function \( g \). For example, estimating the absorption rate in a homogeneous subregion \( V \) of \( \Gamma \) results in the definition

$$g(P) = \chi_V(P) \frac{\Sigma_a(P)}{\Sigma_t(P)}$$

where

$$\chi_V(P) = \begin{cases} 1 & \text{if } P \in V \\ 0 & \text{if } P \notin V \end{cases}$$

is the characteristic, or indicator, function of \( V \) and \( \Sigma_a(P) \), \( \Sigma_t(P) \) are the macroscopic absorption and total cross sections, respectively.

The terminal estimator provides nonzero information only upon the final, or terminating, collision of each random walk, while the collision estimator potentially provides nonzero information on every collision. In Ref. 2, the path length estimator is rigorously derived from the collision estimator by a limiting argument that reveals it to be a continuously scoring version of the collision estimator. It is also well known by those working with

Monte Carlo transport calculations that the path length estimator is indispensable in estimating reaction rates in optically “thin” regions—regions for which the terminal and collision estimators are nearly useless.

These natural random variables may be defined as follows for such transport problems:

A. Terminal estimator:

$$\xi_T(\omega) = \frac{g(P_k)}{p(P_k)}$$.

where \( p(P_k) = 1 - \int \Gamma K(Q,P_k) \, dQ \) = probability of termination\(^b\) in state \( P_k \).

B. Collision estimator:

$$\xi_C(\omega) = \sum_{j=1}^k g(P_j)$$.

C. Track length estimator:

$$\xi_{TL}(\omega) = \sum_{i=1}^k d_i g_i$$

where \( d_i \) = length of \( i \)th track, \( g_i \) = value of \( g \) (assumed here to be unique) along \( i \)th track of the random walk \( \omega = (P_1, P_2, \ldots, P_k) \) (here, \( P_k \) denotes the state in which the random walk terminates by absorption). The collision density \( \Psi \) and flux \( \Phi \) are related by the identity

$$\Psi(P) = \Sigma_t(P) \Phi(P)$$.

where \( \Sigma_t(P) = \) total macroscopic cross section\(^c\) at \( P \). We note here that the detector functions \( g \) that occur most frequently in these formulations are defined in terms of cross sections for the reactions of interest. Hence, they are nonnegative everywhere in phase-space, and they are uniquely defined along each path length segment within a spatially homogeneous subregion of the geometry.

III. ILLUSTRATION: CONTINUOUS ANGLE, ONE-ENERGY TRANSPORT

To illustrate the main ideas, we describe the problem of transport in a finite slab. Assume that we have a slab of thickness \( T \) and that particles can enter the slab either from the left boundary \( (x = 0) \) or from the right boundary \( (x = T) \). The source-free homogeneous material inside the slab can cause the particles either to scatter or to

\(^b\)Here, we have assumed for simplicity that the transport equation describes particle motion in a nonmultiplying, steady-state system. This assumption is not essential, but it greatly simplifies our discussion.

\(^c\)The inverse of this function is the average distance traveled between successive collisions in an infinite homogeneous medium of material with cross section \( \Sigma_r(P) \) (which has dimensions inverse length).
be absorbed. We assume further that there is no energy dependence and for simplicity that scattering is isotropic. The particles here can move in infinitely many directions, each of which is prescribed by a pair of angles $(\theta, \phi)$ with $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$, as shown in Fig. 1. Let $\mu = \cos \theta$ and

$$F(x, \mu) = \int_{0}^{2\pi} \xi(x, \mu, \phi) d\phi,$$  \hspace{1cm} (8)

where $\xi(x, \mu, \phi)$ (the angular flux) is the expected number of particles per second passing $x$ along the direction $(\theta, \phi)$. The function $F(x, \mu)$ then satisfies

$$\frac{\partial F}{\partial x} + \Sigma_i F = \frac{\Sigma_i}{2} \int_{-1}^{1} F(x, \mu') d\mu' ,$$

$$0 < x < T, -1 \leq \mu \leq 1 ,$$

$$F(0, \mu) = Q_0(\mu) , \hspace{1cm} 0 \leq \mu \leq 1 ,$$

$$F(T, \mu) = Q_T(\mu) , \hspace{1cm} -1 \leq \mu \leq 0 ,$$  \hspace{1cm} (9)

where $Q_0(\mu)$ and $Q_T(\mu)$ define the source of particles impinging on the slab from the left and right boundaries, respectively. Without essential loss of generality, we assume $Q_T(\mu) = 0$.

Now, according to the flux expansion technique, we might assume that the particle flux $F(x, \mu)$ has the approximate form:

$$\hat{F}(x, \mu) = \sum_{i=0}^{N_x} \sum_{j=0}^{M_x} a_{ij} P_i \left( \frac{2x}{T} - 1 \right) P_j(\mu) ,$$  \hspace{1cm} (10)

and the functions $P_i$ and $P_j$ denote Legendre polynomials of degrees $i$ and $j$, respectively. Then, the coefficients can be expressed by

$$a_{ij} = \frac{(2i+1)(2j+1)}{2T} \int_{-1}^{1} \hat{F}(x, \mu) P_i(\mu) d\mu \times \int_{0}^{T} P_j \left( \frac{2x}{T} - 1 \right) dx .$$  \hspace{1cm} (11)

In other papers we explore how to estimate the coefficients $a_{ij}$ by adaptive Monte Carlo algorithms that converge geometrically. Here, however, we are concerned only with the nonphysical nature of the weighted integrals [Eq. (11)] that must be estimated to produce the coefficients $a_{ij}$. Each of these is of the form

$$I = \int_{r} G(x, \mu) \hat{F}(x, \mu) ,$$  \hspace{1cm} (12)

where

$$G(x, \mu) = L(x) L(\mu)$$  \hspace{1cm} (13)

and where $L(x), L(\mu)$ are appropriate orthogonal polynomials (e.g., Legendre polynomials) being used to represent the $x$ and $\mu$ dependences of the flux $\hat{F}$. Note that the functions $g$ and $G$ that occur in formulas (3) and (12) are related by $G = \Sigma_i g$ because of Eq. (7).

There is no completely obvious way to use random variables such as formula (6) when the function $G$ varies along track segments, as it obviously does when $G$ takes the form of Eq. (13). To see how to generalize formula (6) to accommodate more general functions $g$, we observe that every random walk in a continuous problem is generated by alternately sampling distances $d_i$ and directions $\mu_i$ until each walk terminates, either because of absorption at $(x_i, \mu_i)$ or because the walk escapes from the (finite) slab geometry. We also observe that use of formula (6) to estimate $I$ requires that the quantity $g$, be uniquely defined along each track segment, while use of formula (5) requires that the quantity $g(x, \mu_j)$ be uniquely defined at each collision point. Since the function $G$ of Eq. (13) depends continuously on both $x$ and $\mu$, while the values of $\mu$ are changing discontinuously at the collision points of $\omega_i$ in practice, ambiguities arise in attempting to use either the collision or the track length estimator. These ambiguities can be avoided by a suitable generalization of the classical transport random variables. The following theorem shows how to generalize the track length estimator to meet these needs.

**Theorem:** Let $\omega = (x_0, \mu_0, x_1, \mu_1, \ldots, x_k, \mu_k)$ designate a typical random walk generated in the one-speed, continuous angle transport problem, where subscripts zero refer to the (physical) source states, subscripts one to the first collision point, and subscripts $k$ to the terminating collision point, i.e., either the last collision point occurring upon absorption in the $x$ region of interest or the

**Note:** Other choices of basis functions are possible and might in fact be preferred for some problems. Legendre polynomials are used here simply to illustrate the main ideas of this paper, which are not dependent on the basis functions employed to represent the flux.

**Note:** Modulo appropriate constants.

---

NUCLEAR SCIENCE AND ENGINEERING  \hspace{1cm} VOL. 133  \hspace{1cm} SEP. 1999
collision just prior to escape from that region. Then, the random variable

$$
\xi_{ij}(\omega) = \sum_{\text{tracks } t} \frac{L_i(\mu_{t-1})}{\Sigma_{it}} \int_{x_{t-1}}^{x_i} L_t(x) \, dx ,
$$

where $\Sigma_{it}$ denotes the value of $\Sigma_t$ along the $l$th track, is an unbiased estimator of the integral

$$
I_{ij} = \int_{x_i}^{x_j} \hat{F}(x, \mu) L_i(x) L_j(\mu) \, dx \, d\mu .
$$

**Proof:** We generate each random walk $\omega = (P_0, P_1, \ldots, P_k) = (x_0, \mu_0, x_1, \mu_1, \ldots, x_k, \mu_k)$ in the manner dictated by the source $S$ and kernel $K$ of the transport problem (9) being simulated. We observe that the value of $\mu, \mu_{t-1}$ is constant along the $l$th track segment stretching from $x_{t-1}$ to $x_t$. Thus, the value of the function $L_i(\mu)$ along the $l$th track is unique and is equal to $L_i(\mu_{t-1})$. However, the value of $L_i(x)$ varies continuously with $x$ along each track segment.

Now, introduce a uniform mesh along each track segment: $x_{t-1} = t_{1,1} \leq t_{1,2} \leq \cdots \leq t_{1,n} = x_t$, and let $y_{l,m}$ designate the midpoint of each subinterval $(t_{l,m-1}, t_{l,m})$. If we approximate the value of $L_i(x)$ along the tiny track between $t_{l,m-1}$ and $t_{l,m}$ by $L_i(y_{l,m})$ and apply the path length estimator [formula (6)] to this small track segment, it contributes

$$
\frac{L_i(\mu)}{\Sigma_{it}} \frac{L_i(y_{l,m})}{\Delta_{l,m}} \Delta_{l,m} ,
$$

where $\Delta_{l,m} = t_{l,m} - t_{l,m-1}$, to the estimate of $I_{ij}$. If we now pass to the limit as $\Delta_{l,m} \to 0$ and sum over all track segments, we obtain the estimator of the theorem.

**Remarks:**

1. For the more general transport problem with two or three spatial dimensions and full angular dependence, the integrals that arise in the theorem become line integrals along each track.

2. The theorem generalizes in an obvious way to the choice of any basis set $B_i(P)$ in Eq. (1).

3. Because integrals in opposite directions over a given track segment cancel each other in the problem with only a single spatial variable, the final formula of the theorem can be simplified by reducing it to a sum of integrals over track segments that are traversed an odd number of times in each direction. This should help to minimize the effects of loss of significance due to cancellation in such problems.

4. If the functions $g, G$ that occur in the theory simplify to ones that are uniquely prescribed along each track segment (as is the case with the physical cross sections that often occur in such problems), the usual path length estimator results in the theorem.

5. The terminal and collision estimators are both trivially extended to unbiased estimators in this more general setting. The terminal estimator is already uniquely specified at the final collision point (where the random walk terminates by absorption), while the collision estimator needs to be interpreted as changing direction (and energy, in problems with energy-dependence) just subsequent to each collision. In other words, a unique value of the collision estimator at each collision point can be obtained by associating the change in direction and energy to occur immediately following each collision instead of immediately preceding it. This is fully consistent with the physics as well.

**IV. IMPORTANCE SOURCE SAMPLING**

When an adaptive zero variance-based importance sampling algorithm is used to estimate a variety of integrals such as formula (12) that arise in such problems, the weighting functions $G$ must be used to generate the initial states of the various random walks. But, the functions $G$ are not in general of constant sign over the $x$ and $\mu$ regions of the problem, so it is unclear how to create such source states. This is an especially critical issue in such adaptive algorithms because it is the source sampling, and in particular the normalization of the source, that plays a vital role in ultimately producing the zero variance estimates.

Our adaptive importance sampling algorithm requires sampling from a source function

$$
\hat{S}(P_0) = \frac{\int \hat{F}(P_0) G(P_0) \, dP_0}{\int \hat{F}(P_0) G(P_0) \, dP_0} ,
$$

where the function $G$ is the weighting function in an integral such as formula (12) and the function $\hat{F}$ has been determined in a previous calculation as an estimate of the solution $\hat{F}$ being sought. In general, there will be many such source functions required in the problem, one for each basis function occurring in the flux representation. Also, direct sampling from a source function such as Eq. (16) cannot be avoided because the zero variance procedure is essentially unique, as the following demonstrates.

**V. UNIQUENESS OF ZERO VARIANCE IMPORATANCE SAMPLING**

To understand that the zero variance importance sampling prescription is unique, one must think about the essential characteristics of importance sampling. The key ideas can be understood by studying a simple one-dimensional integration problem [indeed, each probability density function occurring in a zero variance...
importance sampling algorithm is ultimately written as a product of one-dimensional (conditional) probability density functions.

To this end, suppose, for example, that \( f(x) \) is an integrable function on \([0,1]\) whose integral

\[
I = \int_0^1 f(t) \, dt
\]

we want to estimate using importance sampling. The model we have in mind for \( f(x) \) is any of the one-dimensional factors of the function \( G \) occurring in Eq. (13). Now, let \( g(x) \) be an arbitrary integrable function on \([0,1]\) that vanishes only on a set of measure zero (i.e., only on isolated points of \([0,1]\)). Define

\[
J = \int_0^1 |g(t)| \, dt .
\]

Now, if we sample \( \gamma \in [0,1] \) from the probability density function

\[
p(x) = \frac{|g(x)|}{J}
\]

and if we form the importance sampling estimator

\[
IS(\gamma) = \begin{cases} 
\frac{Jf(\gamma)}{|g(\gamma)|} & g(\gamma) \neq 0 \\
0 & g(\gamma) = 0
\end{cases}
\]

then the expected value of \( IS \) is

\[
E[IS] = \int_0^1 IS(t)p(t) \, dt
\]

\[
= \int_0^1 f(t) \, dt = I
\]

so \( IS \) is an unbiased estimator of \( I \). Also, the variance of \( IS \) is

\[
Var[IS] = \int_0^1 [IS(t) - I]^2 p(t) \, dt
\]

\[
= \int_0^1 [IS(t)]^2 p(t) \, dt - I^2
\]

\[
= \int_0^1 \frac{f^2(t)}{|g(t)|} \, dt - I^2
\]

\[
= \int_0^1 \frac{f^2(t)}{|g(t)|} \, dt - I^2
\]

\[
= \int_0^1 |g(t)| \, dt \int_0^1 \frac{f^2(t)}{|g(t)|} \, dt - I^2 .
\]

At this point one could use a variational argument to bound the first term on the right side of Eq. (22) from below. But, it is simpler and more direct to apply the Cauchy-Schwartz inequality:

\[
|\langle f_1, f_2 \rangle| \leq \|f_1\| \|f_2\| .
\]

valid in an inner product (i.e., a Hilbert) space. With restrictions on \( f, g \) so that \( f_1, f_2 \) are square-integrable over \([0,1]\) \{i.e., \( \int_0^1 |g(x)| \, dx < \infty \) and \( \int_0^1 |f(x)|^2/|g(x)| \, dx < \infty \}, \) the choices

\[
f_1 = |g(x)|^{1/2}
\]

and

\[
f_2 = \frac{|f(x)|}{|g(x)|^{1/2}}
\]

then produce the inequality

\[
Var[IS] \equiv \left[ \int_0^1 |f(x)| \, dx \right]^2 - \left[ \int_0^1 f(x) \, dx \right]^2 ,
\]

and equality occurs only when \( f_1 \) and \( f_2 \) are proportional, so when

\[
|f(x)| = C |g(x)| ,
\]

where \( C \) is a nonzero constant. Thus, the condition for minimum variance is Eq. (26).

Notice that if \( f(x) \) is of constant sign over \([0,1]\), the minimum variance is zero and is achieved by the choices

\[
g(x) = f(x) ,
\]

and

\[
p(x) = \frac{g(x)}{J} = \frac{g(x)}{I} .
\]

However, if no assumptions are made about \( f(x) \) being of constant sign, then the condition of Eq. (26), required for minimum variance, shows that it will be necessary to sample from signed probability density functions to achieve zero variance by importance sampling. The details follow.

VI. SAMPLING A SIGNED PROBABILITY DENSITY

We consider \( g(x) \) as before, integrable on \([0,1]\) and vanishing at most on a set of measure zero, and let \( J \) be defined by Eq. (18). We now extract the positive and negative parts of the function \( g(x) \) over \([0,1]\) and construct (conventional) probability density functions from each to use in our sampling scheme. (A similar treatment can be found in Ref. 6.) It will then remain to be seen whether this still achieves zero variance.

Define

\[
g^+(x) = \max\{g(x), 0\}
\]
and
\[ g^-(x) = -\min\{g(x), 0\} , \]

and let
\[ S^+ = \{ x \in [0,1] | g^+ (x) = 0 \} , \]
\[ S^- = \{ x \in [0,1] | g^- (x) = 0 \} , \]
\[ G^+ = [0,1] \setminus S^+ , \]
and
\[ G^- = [0,1] \setminus S^- . \]  \hfill (29)

Now, we define \( J^+ \) and \( J^- \) by
\[ 0 < J^+ = \int_0^1 g^+(x) \, dx = \int_{G^+} g^+(x) \, dx \]
\[ = \int_{G^+} g(x) \, dx , \]
and
\[ 0 < J^- = \int_0^1 g^-(x) \, dx = \int_{G^-} g^+(x) \, dx \]
\[ = - \int_{G^-} g(x) \, dx , \] \hfill (30)

and we define two probability density functions:
\[ p^+(x) = \frac{g^+(x)}{J^+} , \]
and
\[ p^-(x) = \frac{g^-(x)}{J^-} . \] \hfill (31)

Then, we define two related importance sampling random variables by

\[ IS^+[\gamma] = \begin{cases} 
J^+ f(\gamma) & \text{if } \gamma \in G^+ \\
0 & \gamma \notin G^+ 
\end{cases} \] \hfill (32)

and

\[ IS^-[\delta] = \begin{cases} 
J^- f(\delta) & \text{if } \delta \in G^- \\
0 & \delta \notin G^- 
\end{cases} . \] \hfill (33)

In implementing our importance sampling strategy, we will sample \( \gamma, \delta \) independently from \( p^+(x), p^-(x) \) and add the two importance sampling estimators. Thus, with
\[ \bar{\xi}(\gamma, \delta) = IS^+(\gamma) + IS^-(\delta) \] \hfill (34)

and with \( \gamma, \delta \) sampled as just described,
\[ E[\bar{\xi}] = E[IS^+] + E[IS^-] \]
\[ = \int_{G^+} \frac{J^+ f(x)}{g^+(x)} \, dx + \int_{G^-} \frac{J^- f(x)}{g^-(x)} \, dx \]
\[ = \int_0^1 f(x) \, dx = 1 , \] \hfill (35)

so \( \bar{\xi} \) is certainly unbiased.

Also, since \( \gamma \) and \( \delta \) are independently chosen,
\[ Var[\bar{\xi}] = Var[IS^+] + Var[IS^-] \]
\[ = \int_{G^+} \frac{(J^+)^2 f^2(x) \, dx}{[g^+(x)]^2} \frac{g^+(x)}{J^+} \, dx - \left[ \int_{G^+} f(x) \, dx \right]^2 \]
\[ + \int_{G^-} \frac{(J^-)^2 f^2(x) \, dx}{[g^-(x)]^2} \frac{g^-(x)}{J^-} \, dx - \left[ \int_{G^-} f(x) \, dx \right]^2 \]
\[ = J^+ \int_{G^+} \frac{f^2(x)}{g^+(x)} \, dx - \left[ \int_{G^+} f(x) \, dx \right]^2 \]
\[ + J^- \int_{G^-} \frac{f^2(x)}{g^-(x)} \, dx - \left[ \int_{G^-} f(x) \, dx \right]^2 . \] \hfill (36)

Now, arguing as before, we see that the choices
\[ g^+(x) = |f(x)| \quad x \in G^+ \]
and
\[ g^-(x) = |f(x)| \quad x \in G^- \] \hfill (37)

produce zero variance.

VII. SUMMARY AND CONCLUSIONS

In implementing flux expansion methods by Monte Carlo, the nonphysical nature of the random walks needed to achieve good results requires special attention. This is especially the case if the goal of achieving extremely great accuracy through the use of adaptive, learning algorithms is to succeed. In this paper we have shown how to accommodate two of the problems that arise in such applications: first, the definition of an appropriate unbiased generalization of the conventional track length estimator so useful in transport problems, particularly when optically thin regions are encountered, and second, the need to sample source density functions that take both positive and negative values in the regions of their definition.

With respect to the first problem, we have shown that the proper extension of track length estimators to these more complex applications involves the computation of line integrals along each track. With respect to the second problem, we have shown in this paper that the source sampling problem cannot be avoided in general (i.e., the
zero variance prescription is essentially unique), and we have indicated how to preserve zero variance by decomposing the signed probability density function into a pair of conventional probability density functions and sampling each independently.

The real need for the theory described in this paper arises in applying adaptive importance sampling to the estimation of the global solution of transport problems, as described in Refs. 2 and 5.

REFERENCES


