International Journal of Pure and Applied Mathematics
Volume 59 No. 4 2010, 435-455

GEOMETRIC CONVERGENCE OF SECOND GENERATION ADAPTIVE MONTE CARLO ALGORITHMS FOR GENERAL TRANSPORT PROBLEMS BASED ON CORRELATED SAMPLING

Rong Kong[1], Jerome Spanier[2] §
[1]Claremont Graduate University
150 E. 10-th St., Claremont, California, 91711, USA
e-mail: kongr413@yahoo.com
[2]Beckman Laser Institute and Medical Clinic
University of California
1002, Health Science Road E., Irvine, California 92612, USA
e-mail: jspanier@uci.edu

Abstract: If standard variance reduction methods are applied sequentially in a Monte Carlo simulation, it is possible to achieve geometric reduction of the variance (with probability one) to arbitrarily low error levels. We previously applied this idea to the solution of general radiation transport equations (RTE) by estimating a finite number of coefficients of expansion of the solution in a fixed set of basis functions. However, this first generation adaptive method proves to be impractical because of growth in the number of expansion coefficients and because of the accumulation of computational errors needed in the general case. If the goal of *arbitrary* accuracy *everywhere* in phase space is replaced by the more practical goal of estimating a *small number of linear functionals* of the RTE solution (the measurements, or observables of the system), real gains in computational efficiency can be obtained. We recently introduced a second generation adaptive algorithm based on this principle and illustrated its rapid convergence and increased computational efficiency when compared with conventional and first generation Monte Carlo algorithms. In this paper we establish rigorously the geometric convergence of the new algorithm and exhibit its convergence characteristics when solving two dimensional

Received: February 2, 2010

§Correspondence author

transport problems The new algorithm shows promise of making possible the efficient solution of many continuous transport problems not amenable to either conventional or earlier adaptive methods.

AMS Subject Classification: 82D75
Key Words: transport equation, geometrically convergent Monte Carlo algorithms

1. Introduction

Monte Carlo (MC) simulations have provided a "gold standard"of computational support for many important problems of science and engineering that are modeled using the radiative transport equation (RTE). This situation has persisted in spite of the availability of many faster methods based on approximations to the RTE (e.g., the diffusion approximation) and even though MC is very slowly convergent when implemented in conventional ways. No doubt one of the reasons for MC's lasting prominence in the field is that it is capable of incorporating a description of the transport medium with arbitrary fidelity to the physics — at least in principle. That is, when the sources of radiation, the material properties and physical characteristics of the medium are described as input variables and parameters, the MC simulation produces an RTE solution with a precision that is only limited by the total number, W, of independent random walks generated. Thus, if a method could be devised to dramatically accelerate the convergence of MC simulations, its practical utility would greatly increase.

John Halton introduced the idea of sequentially applying variance reduction schemes to the MC solution of matrix problems in [6], [7] and this device has been successfully applied to continuous transport problems by estimating the coefficients of expansion of the RTE solution [1]-[9] and [11]-[15].

In [11], [12] we described a sequential correlated sampling algorithm (SCS) for solving continuous transport problems in this way (we refer to these as first generation adaptive algorithms) and we proved that the algorithm converges geometrically for a special family of such problems. In [14] we strengthened this result, proving that this method converges geometrically for quite general transport problems. Provided that sufficently many expansion coefficients are estimated and that the computation can be carried out with sufficiently high precision, these first generation adaptive methods are capable of finding RTE solutions with arbitrary precision. However, in practice it is not possible to

determine how many coefficients will be needed nor is it easy to guarantee the high precision needed in the numerical computations that must be performed.

For these reasons, in [10] we introduced a second generation sequential correlated sampling method (ASCS) for overcoming most of the shortcomings of first generation adaptive algorithms, and the present paper establishes rigorously the geometric convergence of this new algorithm. We also compare the performance of ASCS with that of SCS for members of the same two dimensional family described in [14].

We consider general transport problems that are formulated as integral equations of the form

$$\Psi(P) = \mathcal{K}\Psi(P) + S(P), \tag{1}$$

where

$$\mathcal{K}\Psi(P) \equiv \int_{\Gamma} K(P,Q)\Psi(Q)dQ, \quad P \in \Gamma \text{ the phase space,} \tag{2}$$

subject to the restrictions

$$S(P) \geq 0, \text{ and } |S(P)|, \ |K(P,Q)| \leq M \tag{3}$$

for some positive constant M. We also assume that the kernel $K(P,Q)$ is nonnegative and satisfies either

$$\kappa_0 = \max_{P\in\Gamma} \int_{\Gamma} K(P,Q)dQ < 1 \tag{4}$$

or

$$\kappa_0^* = \max_{Q\in\Gamma} \int_{\Gamma} K(P,Q)dP < 1. \tag{5}$$

Our assumptions ensure the existence and uniqueness of a nonnegative and bounded solution $\Psi(P)$; details may be found in [20] or [13].

For any function $S^*(P)$, the general problem we study here is to estimate the weighted integrals

$$\alpha_i = \int_{\Gamma_i} S^*(P)\Psi(P)dP, \ i = 1, ..., R \tag{6}$$

of the RTE solution with respect to a fixed decomposition of the underlying phase space of the problem. Let $\Pi = \{\Gamma_i\}_1^R$ denote this decomposition, for which we assume that $\Gamma = \cup_{i=1}^R \Gamma_i$ and $\Gamma_i \cap \Gamma_j = \emptyset$, $i \neq j$. Although the ASCS algorithm will estimate α_i for every $i = 1, .., R$, for the sake of illustrating the main ideas of this paper, we assume that our *primary* task is to estimate a small number, D, of radiation measurements as determined by the quantities $\alpha_{i_1}, ..., \alpha_{i_D}$ that correspond to D radiation "detectors" located inside or on the boundary of Γ. The idea behind our new algorithm is to aim for very accurate

estimates of these D quantities and to estimate the remaining α_i only with sufficient accuracy to ensure the needed precision in $\alpha_{i_1}, ..., \alpha_{i_D}$.

To this end, we assume that $S^*(P)$ is a globally defined "detector" function that incorporates the description of each physical or simulated detector in any region Γ_i defined by such a detector and is defined arbitrarily in all other regions. For example, for the sake of illustrating the main ideas involved, we set

$$S^*(P) = \sum_{i=1}^{R} \frac{1}{\text{vol}(\Gamma_i)} \Xi_i(P), \tag{7}$$

where

$$\Xi_i(P) = \begin{cases} 1 \text{ if } P \in \Gamma_i, \\ 0 \text{ if } P \notin \Gamma_i, \end{cases} \tag{8}$$

is the characteristic function of Γ_i. This choice produces the averages

$$\alpha_i = \Psi_{ai} = \frac{1}{\text{vol}(\Gamma_i)} \int_{\Gamma_i} \Psi(P) dP \tag{9}$$

of the solution over region Γ_i as the quantities to be estimated. The vector of unknown quantities to be estimated is thus

$$\Psi_a(P) = \Psi_{ai}, \ P \in \Gamma_i \text{ and } i = 1, 2, ..., R. \tag{10}$$

The second generation (G2) adaptive algorithm we have developed aims to replace the *global* expansion of the RTE solution $\Psi(P)$ by *local* representations in terms of histograms $\{\Psi_{ai}\}_{i=1}^{R}$ that are componentwise only as accurate as they need to be to provide the needed precision in estimating $\{\alpha_{i_j}\}_{j=1}^{D}$.

2. ASCS Strategy

In this section, we describe the sequential strategy used for estimating the integrals (6). While we will mimic the SCS algorithm described in detail in [14], there is one important difference. This is a consequence of the fact that the SCS algorithm strives to produce ever more accurate values of the solution of equation (1) at *every* point of phase space, while the ASCS algorithm that is the subject of this paper has the more limited objective of estimating *only* the solution *averages* in prespecified critical locations (at the detectors), rather than pointwise. By taking advantage of this more limited goal we find that real gains in computational efficiency can be realized — especially for higher dimensional and heterogeneous transport problems.

We assume the zero-th stage of the adaptive G2 algorithm is described by

the original transport equation:

$$\psi^0(P) = \int_\Gamma K(P,Q)\psi^0(Q)\,dQ + S^0(P), \tag{11}$$

where

$$S^0(P) = S(P). \tag{12}$$

Denote the estimated regionwise averaged approximate solution of (11) by $\widetilde{\psi}_a^0(P)$. Conventional (nonadaptive) Monte Carlo methods are used to produce $\widetilde{\psi}_a^0(P)$. With this initial estimate in hand, the SCS-based equation for the first stage is

$$\psi^1(P) = \int_\Gamma K(P,Q)\psi^1(Q)\,dQ + S^1(P), \tag{13}$$

where

$$\begin{aligned} S^1(P) &= S^0(P) - \widetilde{\psi}_a^0(P) + \int_\Gamma K(P,Q)\widetilde{\psi}_a^0(Q)\,dQ \\ &= S(P) - \widetilde{\psi}_a^0(P) + \int_\Gamma K(P,Q)\widetilde{\psi}_a^0(Q)\,dQ. \end{aligned} \tag{14}$$

Conventional Monte Carlo methods will again produce the first stage additive correction $\widetilde{\psi}_a^1(P)$ (an approximation to the solution of (13)) to $\widetilde{\psi}_a^0(P)$. This process can be repeated to produce $\widetilde{\psi}_a^1(P), \widetilde{\psi}_a^2(P), ..., \widetilde{\psi}_a^{m-1}(P)$. Then the equation for the m-th stage is

$$\psi^m(P) = \int_\Gamma K(P,Q)\psi^m(Q)\,dQ + S^m(P), \tag{15}$$

where

$$\begin{aligned} S^m(P) &= S^{m-1}(P) - \widetilde{\psi}_a^{m-1}(P) + \int_\Gamma K(P,Q)\widetilde{\psi}_a^{m-1}(Q)\,dQ \\ &= S(P) - \widetilde{\Psi}_a^{m-1}(P) + \int_\Gamma K(P,Q)\widetilde{\Psi}_a^{m-1}(Q)\,dQ, \end{aligned} \tag{16}$$

and where the complete RTE solution is estimated by

$$\widetilde{\Psi}_a^{m-1}(P) = \sum_{i=0}^{m-1} \widetilde{\psi}_a^i(P). \tag{17}$$

After obtaining a final correction $\widetilde{\psi}_a^m(P)$ by the same process, we can form the reconstructed estimate of the regionwise constant solution by

$$\widetilde{\Psi}_a^m(P) = \sum_{i=0}^{m} \widetilde{\psi}_a^i(P). \tag{18}$$

3. Estimator Construction

As formulated above, our goal is to estimate the integrals

$$\alpha_i^m \equiv \int_{\Gamma_i} S^*(P)\psi^m(Q)\,dP, \tag{19}$$

where $\psi^m(P)$ satisfies (15). Since our procedure will be the same for all stages m, we omit the superscript m temporarily and simply describe a method for estimating the integral

$$\alpha \equiv \int_{\Gamma} S^*(P)\psi(P)\,dP, \tag{20}$$

where

$$\psi(P) = \int_{\Gamma} K(P,Q)\psi(Q)\,dQ + S(P). \tag{21}$$

Our immediate task is to describe a random walk process (a method for generating random walks) and an (unbiased) estimating random variable defined on the space Ω of all random walks for each choice of S^*. We will then demonstrate how the sequential application of such estimators produces exponentially accelerated Monte Carlo convergence. As described in Section 3.7 of [20], a quite general random walk process may be defined by selecting a pair of nonnegative functions $\left\{\widehat{S}(P), \widehat{K}(P,Q)\right\}$ subject to the conditions:

$$\begin{aligned} &\int_{\Gamma} \widehat{S}(P)dP = 1, \quad \int_{\Gamma} \widehat{K}(P,Q)dQ = 1 - \widehat{p}(P) > 0, \\ &\widehat{S}(P) = 0 \Longrightarrow S(P) = 0, \quad \widehat{K}(P,Q) = 0 \Longrightarrow K(P,Q) = 0. \end{aligned} \tag{22}$$

The function $\widehat{S}(P)$ will be used to generate initial states P in the phase space Γ, while $\widehat{K}(P,Q)$ will be used to produce successor states Q conditioned by the current state P, and $\widehat{p}(P)$ describes the probability of terminating a random walk in state P. These functions determine random variables on the phase space Γ by means of

$$\begin{aligned} &\xi_0 \sim \widehat{S}, \quad \xi_i|\xi_{i-1} \sim \frac{\widehat{K}(\xi_{i-1},\cdot)}{1-\widehat{p}(\xi_{i-1})}, \quad i = 2,3,..., \\ &\eta_Q \sim B(\widehat{p}(Q)) : P(\eta_Q = 1) = \widehat{p}(Q),\ P(\eta_Q = 0) = 1 - \widehat{p}(Q), \end{aligned} \tag{23}$$

where $B(\widehat{p}(Q))$ is a Bernoulli random variable with parameter $\widehat{p}$ and where the symbol $\sim f$ means that the random variable is sampled making use of the probability density function $f(Q), Q \in \Gamma$.

Then a random variable ζ is defined on the space Ω of all random walk biographies by

$$\zeta = \frac{S^*(\xi_0)}{\widehat{S}(\xi_0)} S(\xi_0) + \frac{S^*(\xi_0)}{\widehat{S}(\xi_0)} \frac{K(\xi_0,\xi_1|\xi_0)}{\widehat{K}(\xi_0,\xi_1|\xi_0)} S(\xi_1)(1-\eta_{\xi_0}) + \cdots$$
$$+ \frac{S^*(\xi_0)}{\widehat{S}(\xi_0)} \frac{K(\xi_0,\xi_1|\xi_0)}{\widehat{K}(\xi_0,\xi_1|\xi_0)} \cdots \frac{K(\xi_{k-1}|\xi_{k-2},\xi_k|\xi_{k-1})}{\widehat{K}(\xi_{k-1}|\xi_{k-2},\xi_k|\xi_{k-1})} S(\xi_k)(1-\eta_{\xi_0}) \cdots (1-\eta_{\xi_{k-1}}) + \cdots$$
$$= \frac{S^*(\xi_0)}{\widehat{S}(\xi_0)} \omega(\xi_0), \quad (24)$$

where

$$\omega(\xi_0) = S(\xi_0) + \frac{K(\xi_0,\xi_1|\xi_0)}{\widehat{K}(\xi_0,\xi_1|\xi_0)} S(\xi_1)(1-\eta_{\xi_0}) + \cdots$$
$$+ \frac{K(\xi_0,\xi_1|\xi_0)}{\widehat{K}(\xi_0,\xi_1|\xi_0)} \cdots \frac{K(\xi_{k-1}|\xi_{k-2},\xi_k|\xi_{k-1})}{\widehat{K}(\xi_{k-1}|\xi_{k-2},\xi_k|\xi_{k-1})} S(\xi_k)(1-\eta_{\xi_0}) \cdots (1-\eta_{\xi_{k-1}})$$
$$+ \cdots. \quad (25)$$

As in [14], the two random variables defined above are the primary ones on which our adaptive methods are based.

4. Geometric Convergence Theorem

In this section, we make use of (24) and (25) to construct an estimator $\omega^m(P)$ of the solution $\psi^m(P)$ of (15). Making use of (25), we obtain:

$$\omega^m(P) = S^m(P) + \frac{K(P,\xi_1)}{\widehat{K}(P,\xi_1)} S^m(\xi_1)(1-\eta_P) + \cdots \quad (26)$$
$$+ \frac{K(P,\xi_1)}{\widehat{K}(P,\xi_1)} \cdots \frac{K(\xi_{k-1},\xi_k|\xi_{k-1})}{\widehat{K}(\xi_{k-1},\xi_k|\xi_{k-1})} S^m(\xi_k)(1-\eta_P) \cdots (1-\eta_{\xi_{k-1}}) + \cdots,$$

where $\widehat{K}(P,Q)$ satisfies (22). Then we can define an estimator of each integral (6), $\alpha_i^m = \int_\Gamma S_i^*(P)\psi^m(P)dP$, by

$$\zeta_i^m = \frac{S_i^*(\xi_0)}{\hat{S}(\xi_0)} \omega^m(\xi_0), \quad (27)$$

where $\hat{S}(P)$ is defined by (22).

Theorem 1. *Assume that $\omega^m(P)$ and ζ_i^m are defined by (26) and (27), respectively. Then $\omega^m(P)$ and ζ_i^m are unbiased estimators; i.e,*

$$E[\omega^m(P)] = \psi^m(P), \quad E[\zeta_i^m] = \int_\Gamma S_i^*(P)\psi^m(P)dP. \quad (28)$$

Furthermore, the variance of $\omega^m(P)$ satisfies

$$Var_\omega\left[\omega^m(P)\right] = \int_\Gamma \left(\frac{K(P,Q)}{\widehat{K}(P,Q)}\right)^2 \left(Var_\omega\left[\omega^m(Q)\right] + \left(\psi^m(Q)\right)^2\right) \widehat{K}(P,Q)dQ - \left(\psi^m(P) - S^m(P)\right)^2, \quad (29)$$

and the variance of ζ_i^m *can be obtained by*

$$Var_\zeta\left[\zeta_i^m\right] + <S_i^*, \psi^m>^2 = \int_\Gamma \left(\frac{S_i^*(P)}{\hat{S}(P)}\right)^2 \left(Var_\zeta\left[\omega^m(P)\right] + \left(\psi^m(P)\right)^2\right) \hat{S}(P)dP. \quad (30)$$

The proofs of these results are the same as those provided in [14], so we omit them here.

We now state and prove our main convergence theorem. First, for the phase space decomposition $\Pi = \{\Gamma_j\}_{j=1}^R$ of Γ set

$$r_\Pi = \sup_P \left|\Psi(P) - \Psi_a(P)\right|, \quad (31)$$

and notice that

$$r_{\Pi_1} \le r_{\Pi_2} \quad (32)$$

if Π_1 is a decomposition that is a refinement of Π_2. The following theorem describes the geometric convergence of the ASCS algorithm.

Theorem 2. *Consider equation (1) and assume that the kernel satisfies (4). Then, for any* $\varepsilon > 0$ *and* $\lambda < 1$*, there must be a critical number of independent random walks* $W_0 \equiv W(\lambda, \varepsilon, \kappa_1, \delta)$*, such that when* $W > W_0$

$$P\left\{\sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^m(P)\right| < \lambda\left(\sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^{m-1}(P)\right| + r_\Pi\right)\right\} \ge 1 - \varepsilon, \quad (33)$$

where r_Π *is defined by (31).*

The constant κ_1 *appearing in this theorem is easily computed from knowledge of the kernel* K *(see (35) below).*

Proof. From (4), $\exists\ \delta > 0$ such that

$$\frac{1}{1-\delta}\sup_P \int_\Gamma K(P,Q)dQ < 1. \quad (34)$$

Therefore, we can find a pair of functions $\left\{\widehat{S}(P), \widehat{K}(P,Q)\right\}$ satisfying (22), such that

$$\kappa_1 \equiv \sup_P \frac{1}{1-\delta}\int_\Gamma \left(\frac{K(P,Q)}{\widehat{K}(P,Q)}\right)^2 \widehat{K}(P,Q)dQ < 1. \quad (35)$$

Indeed, the simple choice $\widehat{K}(P,Q) = K(P,Q)$ would suffice owing to (34). Based on this pair $\left\{\widehat{S}(P), \widehat{K}(P,Q)\right\}$, we can construct estimators $\{\omega^m(P), \zeta_i^m\}$ for $m = 0, 1, 2, ...$

From Theorem 1, using equation (29), we have

$$\begin{aligned}\mathrm{Var}_\omega\left[\omega^m(P)\right] + \left(\psi^m(P)\right)^2 \\ = \int_\Gamma \left(\frac{K(P,Q)}{\widehat{K}(P,Q)}\right)^2 \left(\mathrm{Var}_\omega\left[\omega^m(Q)\right] + \left(\psi^m(Q)\right)^2\right) \widehat{K}(P,Q)dQ \\ + \left(\psi^m(P)\right)^2 - \left(\psi^m(P) - S^m(P)\right)^2. \quad (36)\end{aligned}$$

We first estimate the term

$$H^m \equiv \left(\psi^m(P)\right)^2 - \left(\psi^m(P) - S^m(P)\right)^2. \quad (37)$$

We have

$$H^m = \left(2\psi^m(P) - S^m(P)\right) S^m(P), \quad (38)$$

and, for the number δ introduced in (34), we have

$$\begin{aligned}|H^m| &= \left|\left(2\psi^m(P) - S^m(P)\right) S^m(P)\right| \\ &\leq \delta\left(\psi^m(P)\right)^2 + \left(1 + \frac{1}{\delta}\right)\left(S^m(P)\right)^2.\end{aligned}$$

Thus, from (36), we obtain

$$\begin{aligned}\mathrm{Var}_\omega\left[\omega^m(P)\right] + \left(\psi^m(P)\right)^2 \\ \leq \int_\Gamma \left(\frac{K(P,Q)}{\widehat{K}(P,Q)}\right)^2 \left(\mathrm{Var}_\omega\left[\omega^m(Q)\right] + \left(\psi^m(Q)\right)^2\right) \widehat{K}(P,Q)dQ \\ + \delta\left(\psi^m(P)\right)^2 + \left(1 + \frac{1}{\delta}\right)\left(S^m(P)\right)^2 \quad (39)\end{aligned}$$

which means

$$\begin{aligned}\mathrm{Var}_\omega\left[\omega^m(P)\right] + \left(\psi^m(P)\right)^2 \\ \leq \frac{1}{1-\delta}\int_\Gamma \left(\frac{K(P,Q)}{\widehat{K}(P,Q)}\right)^2 \left(\mathrm{Var}_\omega\left[\omega^m(Q)\right] + \left(\psi^m(Q)\right)^2\right) \widehat{K}(P,Q)dQ \\ + \frac{1}{1-\delta}\left(1 + \frac{1}{\delta}\right)\left(S^m(P)\right)^2. \quad (40)\end{aligned}$$

According to (35), the kernel of equation (36) is contractive and (see Appendix of [14]) so we obtain

$$\mathrm{Var}_\omega\left[\omega^m(P)\right] + \left(\psi^m(P)\right)^2 \quad (41)$$

$$\leq \frac{1}{1-\kappa_1}\sup_P\left|\frac{1}{1-\delta}\left(1+\frac{1}{\delta}\right)\left(S^m(P)\right)^2\right|,$$

or

$$\mathrm{Var}_\omega\left[\omega^m(P)\right]+\left(\psi^m(P)\right)^2\leq C_1\sup_P\left|\left(S^m(P)\right)^2\right|, \tag{42}$$

where C_1 depends only on δ and κ_1.

Now we examine the reduced source (defined by (16)) more carefully. We have

$$\begin{aligned} S^m(P) &= S^{m-1}(P)-\widetilde{\psi}_a^{m-1}(P)+\int_\Gamma K(P,Q)\widetilde{\psi}_a^{m-1}(Q)\,dQ \qquad (43)\\ &= S(P)-\widetilde{\Psi}_a^{m-1}(P)+\int_\Gamma K(P,Q)\widetilde{\Psi}_a^{m-1}(Q)\,dQ \\ &= \left(\Psi(P)-\widetilde{\Psi}_a^{m-1}(P)\right)-\int_\Gamma K(P,Q)\left(\Psi(Q)-\widetilde{\Psi}_a^{m-1}(Q)\right)dQ. \end{aligned}$$

Therefore,

$$S^m(P)\leq(1+\kappa_0)\sup_P\left|\Psi(P)-\widetilde{\Psi}_a^{m-1}(P)\right|, \tag{44}$$

where the constant κ_0 is defined in (4). From (42) and (44) we obtain

$$\mathrm{Var}_\omega\left[\omega^m(P)\right]+\left(\psi^m(P)\right)^2\leq C_1(1+\kappa_0)\sup_P\left|\Psi(P)-\widetilde{\Psi}_a^{m-1}(P)\right|^2. \tag{45}$$

Now, from (30), we have

$$\begin{aligned} &\mathrm{Var}_\zeta\left[\zeta_i^m\right]+<S_i^*,\psi^m>^2 \\ &\quad=\int_\Gamma\left(\frac{S_i^*(P)}{\hat{S}(P)}\right)^2\left(\mathrm{Var}_\omega\left[\omega^m(P)\right]+\left(\psi^m(P)\right)^2\right)\hat{S}(P)dP \\ &\quad\leq\sup_{P\in\Gamma}\left(\mathrm{Var}_\omega\left[\omega^m(P)\right]+\left(\psi^m(P)\right)^2\right)\int_\Gamma\left(\frac{S_i^*(P)}{\hat{S}(P)}\right)^2\hat{S}(P)dP \\ &\quad\leq M_2\sup_P\left|\Psi(P)-\widetilde{\Psi}_a^{m-1}(P)\right|^2, \qquad (46) \end{aligned}$$

where

$$M_2\equiv C_1(1+\kappa_0)\max_i\int_\Gamma\left(\frac{S_i^*(P)}{\hat{S}(P)}\right)^2\hat{S}(P)dP. \tag{47}$$

From (46) and (31), we obtain

$$\begin{aligned}
&\mathrm{Var}_\zeta\left[\zeta_i^m\right] + <S_i^*, \psi^m>^2 \\
&\leq M_2 \sup_P \left|\Psi(P) - \widetilde{\Psi}_a^{m-1}(P)\right|^2 \\
&= M_2 \sup_P \left|\Psi(P) - \Psi_a(P) + \Psi_a(P) - \widetilde{\Psi}_a^{m-1}(P)\right|^2 \\
&\leq M_2 \left(\sup_P |\Psi(P) - \Psi_a(P)| + \sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^{m-1}(P)\right|\right)^2,
\end{aligned} \tag{48}$$

which means that, for any i,

$$\begin{aligned}
\sqrt{\mathrm{Var}_\zeta\left[\zeta_i^m\right]} &\leq \sqrt{M_2}\left(\sup_P |\Psi(P) - \Psi_a(P)| + \sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^{m-1}(P)\right|\right) \\
&= \sqrt{M_2}\left(r_\Pi + \sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^{m-1}(P)\right|\right).
\end{aligned} \tag{49}$$

By appealing to Chebyshev's inequality, for any $\varepsilon > 0$, we have

$$\Pr\left\{\sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^m(P)\right| < \sqrt{\frac{\mathrm{Var}_\zeta\left[\zeta_i^m\right]}{\varepsilon W}}\right\} \geq 1 - \varepsilon. \tag{50}$$

Now, we substitute (49) into (50) to obtain

$$\begin{aligned}
\Pr\left\{\sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^m(P)\right| < \sqrt{\frac{M_2}{\varepsilon W}} r_\Pi + \sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^{m-1}(P)\right|\right\} \\
\geq 1 - \varepsilon.
\end{aligned} \tag{51}$$

The only remaining step is to choose W so large that

$$\lambda \equiv \sqrt{\frac{M_2}{\varepsilon W}} < 1 \tag{52}$$

which then completes the proof. □

Remark 1. The choice of $\widehat{K}(P,Q)$ that controls the simulated transport is quite arbitrary as long as (22) and (35) are satisfied. In practice, a good choice of $\widehat{K}(P,Q)$ would be one that produces small variances of $\omega^m(P)$ and ζ_i^m, and facilitates easy (and rapid!) sampling procedures for generation of the random walks.

Remark 2. Notice the term r_Π (defined by (31)) in Theorem 2. This term reflects the influence of the decomposition Π on the accuracy of the G2 solution. From Theorem 2, by iteration we can investigate how the error after the m-th stage, $\sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^m(P)\right|$, is related to the error from the 0-th stage, $\sup_P \left|\Psi_a(P) - \widetilde{\Psi}_a^0(P)\right|$. This results in the inequality,

$$\Pr\left\{\sup_P\left|\Psi_a(P)-\widetilde{\Psi}_a^0(P)\right|<\lambda^m\sup_P\left|\Psi_a(P)-\widetilde{\Psi}_a^0(P)\right|+\frac{1-\lambda^m}{1-\lambda}\lambda r_\Pi\right\} \geq 1-m\epsilon$$

which means that the error after m stages is the same order as the one induced by a single iteration. As a consequence we are assured that the error reduction factor from stage to stage is not much larger than λ.

5. Numerical Results

In this section, we apply the G2 algorithm to the same family of model transport problems treated in [14]. These are characterized by the coupled partial differential equations

$$\begin{cases} \dfrac{\partial\psi_1}{\partial x}+\Sigma_t(x,y)\,\psi_1(x,y)=\Sigma_s(x,y)\displaystyle\sum_{j=1}^{4}p_{1j}\psi_j(x,y)+s_1(x,y), \\ -\dfrac{\partial\psi_2}{\partial x}+\Sigma_t(x,y)\,\psi_2(x,y)=\Sigma_s(x,y)\displaystyle\sum_{j=1}^{4}p_{2j}\psi_j(x,y)+s_2(x,y), \\ \dfrac{\partial\psi_3}{\partial y}+\Sigma_t(x,y)\,\psi_3(x,y)=\Sigma_s(x,y)\displaystyle\sum_{j=1}^{4}p_{3j}\psi_j(x,y)+s_3(x,y), \\ -\dfrac{\partial\psi_4}{\partial y}+\Sigma_t(x,y)\,\psi_4(x,y)=\Sigma_s(x,y)\displaystyle\sum_{j=1}^{4}p_{4j}\psi_j(x,y)+s_4(x,y), \end{cases} \tag{53}$$

where

$$0<x<a,\quad 0<y<b, \tag{54}$$

with boundary sources

$$\begin{aligned} \psi_1(0,y) &= Q_1(y), \quad \psi_2(a,y)=Q_2(y), \\ \psi_3(x,0) &= Q_3(x), \quad \psi_4(x,b)=Q_4(y), \end{aligned}$$

and where we assume

$$\sum_{j=1}^{4}p_{ji}=1,\ i=1,2,3,4. \tag{55}$$

This system describes transport problems in the rectangle $0<x<a$, $0<y<b$ in which transport only in four discrete directions is allowed. The solution functions $\psi_1(x,y)$, $\psi_2(x,y)$, $\psi_3(x,y)$ and $\psi_4(x,y)$ are the right-, left-, up- and down-moving fluxes, respectively. The equivalent formulation as

coupled integral equations is

$$\begin{cases} \psi_1(x,y) = \Sigma_s \sum_{j=1}^{4} p_{1j} \int_0^x e^{-\Sigma_t(x-x')} \psi_j(x',y)\, dx' + S_1(x,y), \\ \psi_2(x,y) = \Sigma_s \sum_{j=1}^{4} p_{2j} \int_x^a e^{-\Sigma_t(x'-x)} \psi_j(x',y)\, dx' + S_2(x,y), \\ \psi_3(x,y) = \Sigma_s \sum_{j=1}^{4} p_{3j} \int_0^y e^{-\Sigma_t(y-y')} \psi_j(x,y')\, dy' + S_3(x,y), \\ \psi_4(x,y) = \Sigma_s \sum_{j=1}^{4} p_{4j} \int_y^b e^{-\Sigma_t(y'-y)} \psi_j(x,y')\, dy' + S_4(x,y), \\ S_1(x,y) \equiv \int_0^x e^{-\Sigma_t(x-x')} s_1(x',y)\, dx' + e^{-\Sigma_t x} Q_1(y), \\ S_2(x,y) \equiv \int_x^a e^{-\Sigma_t(x'-x)} s_2(x',y)\, dx' + e^{-\Sigma_t(a-x)} Q_2(y), \\ S_3(x,y) \equiv \int_0^y e^{-\Sigma_t(y-y')} s_3(x,y')\, dy' + e^{-\Sigma_t y} Q_3(x), \\ S_4(x,y) \equiv \int_y^b e^{-\Sigma_t(y'-y)} s_4(x,y')\, dy' + e^{-\Sigma_t(b-y)} Q_4(x). \end{cases} \tag{56}$$

We simulate a 4x5 rectangle (measured in mean free paths) of "tissue" with typical optical properties with the following input data:

$$a = 5,\ b = 4,\ \Sigma_a = 0.01,\ \Sigma_s = 0.99,$$

$$p_{ij} = \begin{pmatrix} 0.9025 & 0.0025 & 0.0475 & 0.0475 \\ 0.0025 & 0.9025 & 0.0475 & 0.0475 \\ 0.0475 & 0.0475 & 0.9025 & 0.0025 \\ 0.0475 & 0.0475 & 0.0025 & 0.9025 \end{pmatrix}$$

For this problem, scattering dominates absorption by a factor of 99/1, and the scattering phase function simulates Henyey-Greenstein (H-G) scattering [8] in two dimensions allowing only the four directions left, right, up, down. The matrix elements p_{ij} are here chosen to correctly reproduce the first two angular moments of H-G scattering with an anisotropy factor (average cosine of scattering angle), g, of 0.9, which is typical of many types of tissue. If the units chosen for the optical properties are mm^{-1}, these data describe a small, 4mm×5mm "patch" of tissue with optical properties characteristic of epithelial tissue.

We set up two problems making use of this input data by introducing two different incoming sources of radiation on the boundary of the rectangle. The

first set of boundary conditions (B1) is

$$Q_1(y) = \frac{y}{b},\ 0 < y < b, \qquad Q_2(y) = 0,\ 0 < y < b,$$
$$Q_3(x) = 0,\ 0 < x < a, \qquad Q_4(x) = 1 - \frac{x}{a},\ 0 < x < a,$$

which creates a source of incoming radiation that is distributed linearly along the left and top edges of the rectangle, falling from the value 1 at the upper left corner $(0, 1)$ to the value 0 along both the $x = 0$ and $y = 1$ edges.

The second set of boundary conditions (B2) is

$$Q_1(y) = \begin{cases} 1,\ 0.99b < y < b, \\ 0,\ \text{otherwise}, \end{cases}$$
$$Q_2(y) = 0,\ 0 < y < b, \qquad Q_3(x) = 0,\ 0 < x < a,$$
$$Q_4(x) = \begin{cases} 1,\ 0 < x < 0.01a, \\ 0,\ \text{otherwise}, \end{cases}$$

which creates an incoming source of radiation that is concentrated in a very small rectangle $[0, 0.01a] \times [0.99b, 1.0]$ at the upper left corner of the rectangle and is zero everywhere else.

In both of these problems, our goal is to estimate the integral of the solution over the subrectangle $[4.95, 5.0] \times [0, 0.04]$ at the lower right corner of the 4×5 rectangle that defines the problem geometry.

We use three different algorithms to solve these two problems: conventional Monte Carlo (CMC), our earlier G1 SCS algorithm and the G2 ASCS algorithm described in this paper. Because of the difficulty of transporting radiation from the small source through the tissue to a relatively small "detector", the second set of boundary conditions poses a much greater challenge for each algorithm than the first set of boundary conditions does. We are interested in comparing the computational efficiencies of our methods, and especially in examining how the G1 and G2 adaptive methods perform relative to conventional MC simulation.

First we plot the convergence characteristics of the G1 and G2 algorithms for the two problems treated.

In each of Figures 1, 2 we show the decline of the $\log_{10}$ of the variance of the G1 and G2 estimators with the number of adaptive stages run. Figure 1 plots the output from the G1 and G2 algorithms, respectively, for the problem with boundary source B1. Notice that the G1 SCS algorithm performs flawlessly for this problem, producing machine precision after 16 stages. To confirm the stability we actually ran more adaptive stages and observed no discernible change either in the estimates of the mean or of the variance.

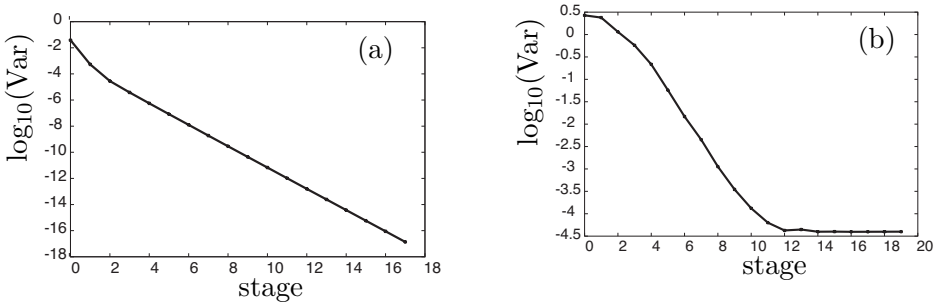

Figure 1: Continuous source case. Geometric convergence of: (a) G1 algorithm, (b) G2 algorithm

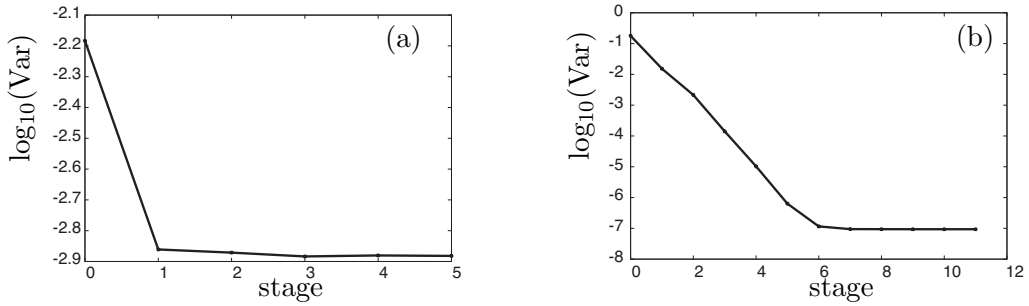

Figure 2: Discrete source case. Geometric convergence of: (a) G1 algorithm, (b) G2 algorithm

For the same problem, the G2 ASCS algorithm produced a geometric decrease in error that stabilized after 15 adaptive stages. Because of the additional error term r_Π that appears in Theorem 2, this algorithm is only able to achieve roughly 4 orders of magnitude error reduction based on the phase space decomposition Π we chose, which was determined by 500 uniform subdivisions in x and 400 uniform subdivisions in y.

For the second problem the comparative behaviors of the two algorithms is sharply different as seen in Figure 2. By concentrating the boundary source of radiation at the upper left corner, we have created an RTE solution that is not at all well represented by a polynomial expansion throughout the rectangle. Numerical evidence of this can clearly be seen in Figure 2(a) by the sharp fall of the variance of the G1 estimator only between the zeroth and the first adaptive stage of the G1 algorithm, stabilizing immediately after a very small small gain in accuracy. The G2 ASCS algorithm, on the other hand, behaves for this

problem much as it did for the first one. The numerical evidence indicates the clear geometric decline of the error over 7 decades in 15 adaptive stages, with stabilization thereafter.

Figures 3, 4 that plot the scalar fluxes produced by each algorithm for both problems underscore these contrasting behaviors.

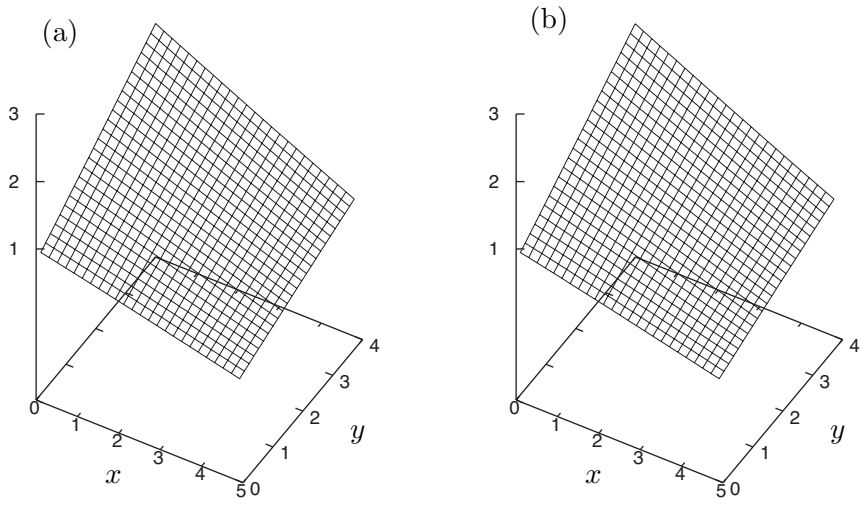

Figure 3: Continuous source case. Scalar flux produced by: (a) G1 algorithm and (b) G2 algorithm

The steep decline in the RTE solution at small distances from the radiation source is captured very well in Figures 3(b), 4(b) by the G2 histogram approximation, while the breakdown in the ability of the polynomial approximation produced by G1 to capture any meaningful qualitative shape is equally evident from Figure 4(a).

We are interested in comparing the computational efficiencies of the G1 and G2 adaptive methods with the efficiency of CMC. A useful indicator of efficiency in conventional MC implementations is

$$\text{Eff} = \frac{1}{\text{Var} \cdot t}, \tag{57}$$

where Var is the variance of the estimating random variable and t is the total computer time required to achieve this variance. For conventional MC sim-

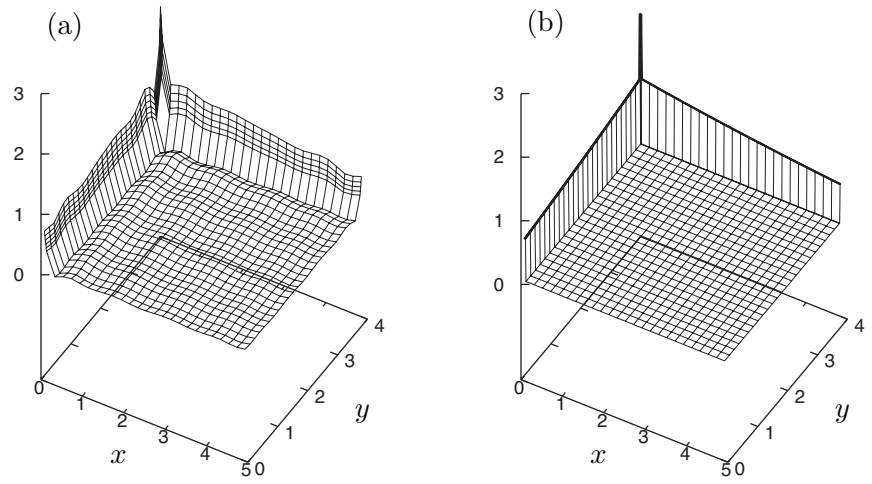

Figure 4: Discrete source case. Scalar flux produced by: (a) G1 algorithm and (b) G2 algorithm

ulations, this measure[1] is roughly independent of the number N of random samples processed since t is linear in N and Var is inversely proportional to N, and therefore to t. However, our adaptive algorithms are designed to produce variances that decrease *exponentially* with time. Thus, to compare computational efficiency of these new methods with conventional Monte Carlo requires that we examine the amount of error reduction attained by each of the more powerful algorithms and compute the run time that would be necessary for conventional Monte Carlo to achieve this amount of additional error reduction. This last computation is predicated on the assumption that the conventional simulations converge at the rate predicted by the central limit theorem. To apply these ideas, the time t required for each computation was obtained by careful timing of the runs performed for each of our methods on the same computer: a 1.6GHz Xeon E5310 10 quad core processor.

In Tables 1 and 2 we compare the efficiencies of our G1 and G2 algorithms with that of conventional Monte Carlo (CMC). Thus, for example, to achieve a variance of 5.775389×10^{-15} with conventional Monte Carlo when the variance

[1]In MCNP [4] the term "Figure of Merit"(FOM) is used for this quantitative estimate of efficiency.

it achieves in 134 seconds is 1.85702949×10^2 would require

$$\frac{1.85702949 \times 10^2}{5.775389 \times 10^{-15}} \times 134 \text{ seconds} = 4.30866 \times 10^{18} \text{ seconds}, \tag{58}$$

whereas this was achieved with the G1 algorithm at a cost of 16×1280 seconds. The advantage factor of G1 compared with conventional Monte Carlo is thus

$$\frac{4.30866 \times 10^{18}}{16 \times 1280} = 2.10342 \times 10^{14}. \tag{59}$$

The G2 entry in the first table was computed similarly, as were the entries in the second table.

Table 1 compares the performances of the three methods for the easier boundary source condition B1. In Tables 1 and 2 S indicates the number

Method	S	Est	Var	t	Eff
CMC	1	$1.27288066e-1$	$1.857029e+2$	134	1
SCS	16	$1.283232e-1$	$5.775387e-15$	16×1280	2.10342×10^{14}
ASCS	15	$1.286094e-1$	$3.971712e-5$	15×1000	4.1769×10^4

Table 1: Comparison of efficiencies of each method, easier boundary source condition, B1

of stages processed for each method, Est is the sample mean of the random variable employed and Var is the variance of that random variable, t is the total computation time in seconds, and Eff is the computational efficiency of each method relative to the efficiency of CMC.

Method	S	Est	Var	t	Eff
CMC	1	$2.10850e-3$	$4.9951255e-2$	126	1
SCS	2	$3.917029e-3$	$1.377152e-3$	2×1278	1.788025
ASCS	15	$2.047106e-3$	$1.232968e-3$	15×980	3.4725×10^3

Table 2: Comparison of efficiencies of each method, harder boundary source condition, B2

6. Summary, Conclusions and Future Work

We have established rigorously the geometric convergence obtained with our second generation G2 ASCS algorithm based on correlated sampling variance

reduction that is applied recursively. In the numerical examples that we studied here, we were interested in providing persuasive evidence to support typical convergence characteristics of this algorithm for two members of a simple family of RTEs that models two dimensional transport in a homogeneous rectangle of simulated tissue. While the convergence characteristics and the qualitative shapes of the RTE solution produced by the G2 algorithm for the two problems are very similar, the G1 algorithm performs very differently for these two problems. Indeed, the input was deliberately selected to reveal the breakdown in the G1 results that one should expect when the basis set chosen for the expansion of the RTE solution is unrelated to the eigensystem that would be produced by spectral analysis of the integral operator corresponding to the specific transport kernel for each problem. Unfortunately, this means that our goal of efficient and automated adaptive algorithms cannot be achieved using only the G1 strategy.

Preliminary evidence presented here shows the potential that the G2 algorithm will perform more reliably over a broader class of RTE problems because it does not depend on identifying a complete eigensystem for the integral operator $\mathcal{K}$ that characterizes each RTE problem. Work now in progress should help to understand how well the G2 algorithms might perform for more complex, higher dimensional, and fully heterogeneous transport problems.

Acknowledgments

The authors gratefully acknowledge support for this work from the National Science Foundation grant NSF/DMS 0712853 and the Laser Microbeam and Medical Program (LAMMP) grant NIH RR0192. We also appreciate assistance given by Dr. Katherine Bhan in preparing this manuscript.

References

[1] T. Booth, Exponential convergence for Monte Carlo particle transport, *Transactions of the American Nuclear Society*, **50** (1985), 267-268.

[2] T. Booth, Zero-variance solutions for linear Monte Carlo, *Nuclear Science and Engineering*, **102** (1989), 332-340.

[3] T. Booth, Exponential convergence on a continuous Monte Carlo transport problem, *Nuclear Science and Engineering*, **127** (1997), 338-345.

[4] J. Breismeister, Ed., *MCNP4C, A Monte Carlo N-particle Transport Code*, LANL Technical Report (1999).

[5] G.B. Folland, *Real Analysis*, A Wiley-Interscience Publication (1984).

[6] J. Halton, Sequential Monte Carlo, *Proc. Comb. Phil. Soc.*, **58** (1962).

[7] J. Halton, Sequential Monte Carlo techniques for the solution of linear systems, *J. Sci. Comp.*, **9** (1994), 213-257.

[8] L. Henyey, J. Greenstein, Diffuse radiation in the galaxy, *Astrophys. Journal*, **93** (1941), 70-83.

[9] C. Kollman, *Rare Event Simulation in Radiation Transport*, Unpublished Ph.D. Thesis, Dept. of Statistics, Stanford Univ. (1993).

[10] R. Kong, M. Ambrose, J. Spanier, Efficient, automated Monte Carlo methods for radiation transport, *J. Comp. Phys.*, **227** (2008), 9463-9476.

[11] R. Kong, J. Spanier, Sequential correlated sampling methods for some transport problems, In: *Monte-Carlo and Quasi Monte-Carlo Methods – 1998* (Ed-s: H. Niederreiter, J. Spanier), Springer-Verlag (1999), 238-251.

[12] R. Kong, J. Spanier, Error analysis of sequential Monte Carlo methods for transport problems, In: *Monte-Carlo and Quasi Monte-Carlo Methods – 1998* (Ed-s: H. Niederreiter, J. Spanier), Springer-Verlag (1999), 252-272.

[13] R. Kong, J. Spanier, Residual versus error in transport problems, In: *Monte-Carlo and Quasi Monte-Carlo Methods – 2000* (Ed-s: K.-T. Fang, F.J. Hickernell, H. Niederreiter) (2002), 306-317.

[14] R. Kong, J. Spanier, A new proof of geometric convergence for general transport problems based on sequential correlated sampling methods, *J. Comp. Phys.*, **227** (2008), 9762-9777.

[15] Y. Lai, J. Spanier, The adaptive importance sampling algorithm for transport problems, In: *Monte-Carlo and Quasi Monte-Carlo Methods – 1998* (Ed-s: H. Niederreiter, J. Spanier), Springer-Verlag (2000), 273-283.

[16] I. Lux, L. Koblinger, *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*, CRC Press (1990).

[17] W. Pogorzelski, *Integral Equations and their Applications*, Volume I, Pergamon Press (1966).

[18] S. Ross, *A First Course in Probability*, Prentice-Hall, 5-th Edition (1998).

[19] J. Spanier, Monte Carlo methods for flux expansion solutions of transport problems, *Nucl. Sci. Eng.*, **133** (1999), 1-7.

[20] J. Spanier, E.M. Gelbard, *Monte Carlo Principles and Neutron Transport Problems*, Addison-Wesley Pub. Co (1969); Reprinted by Dover Publications, Inc. (2008).