

UC Irvine

UC Irvine Previously Published Works

Title

A new proof of geometric convergence for the adaptive generalized weighted analog sampling (GWAS) method

Permalink

<https://escholarship.org/uc/item/6qq238kf>

Journal

Monte Carlo Methods and Applications, 22(3)

ISSN

0929-9629

Authors

Kong, Rong
Spanier, Jerome

Publication Date

2016-09-01

DOI

10.1515/mcma-2016-0110

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Published in final edited form as:

Monte Carlo Methods Appl. 2016 September ; 22(3): 161–196. doi:10.1515/mcma-2016-0110.

A new proof of geometric convergence for the adaptive generalized weighted analog sampling (GWAS) method

Rong Kong and

Hyundai Capital America, 3161 Michelson Drive, Suite 1900, Irvine, CA 92612, USA

Jerome Spanier*

Beckman Laser Institute and Medical Clinic, 1002 Health Science Road E., University of California, Irvine, California 92612, USA

Abstract

Generalized Weighted Analog Sampling is a variance-reducing method for solving radiative transport problems that makes use of a biased (though asymptotically unbiased) estimator. The introduction of bias provides a mechanism for combining the best features of unbiased estimators while avoiding their limitations. In this paper we present a new proof that adaptive GWAS estimation based on combining the variance-reducing power of importance sampling with the sampling simplicity of correlated sampling yields geometrically convergent estimates of radiative transport solutions. The new proof establishes a stronger and more general theory of geometric convergence for GWAS.

Keywords

Monte Carlo methods; radiative transport; generalized weighted analog; geometric convergence

1 Introduction and background

Our previous research (see [5, 6, 8–10]) on adaptive Monte Carlo algorithms for radiative transport problems resulted in the development of several geometrically convergent Monte Carlo algorithms for global transport solutions Φ . By geometric convergence, we mean

$$E_s < \lambda E_{s-1} < \lambda^s E_0, \quad 0 < \lambda < 1, \quad s = \text{stage number},$$

where $E_s =$ s th stage error; e.g.,

$$E_s = \|\Phi(P) - \tilde{\Phi}^s(P)\|_\infty$$

*Corresponding author: Jerome Spanier: Beckman Laser Institute and Medical Clinic, 1002 Health Science Road E., University of California, Irvine, California 92612, USA, jspanier@uci.edu.

and $\tilde{\Phi}^s(P)$ is an approximation obtained in the s th stage to $\Phi(P)$, the solution of the radiative transport equation (RTE). The geometric convergence means that the *rate of convergence* of the approximate solution $\tilde{\Phi}^s(P)$ to the solution $\Phi(P)$ is exponentially greater than the central limit theorem-constrained rate of non-adaptive methods. However, taking into account *both* variance *and* time, our true goal for adaptive methods is to exponentially increase the computational efficiency

$$\text{Eff} = \frac{1}{\text{Var} \times T}$$

when compared with non-adaptive Monte Carlo, where Var is the estimator variance and T is the total computer processing time.

We have demonstrated geometric convergence using both correlated sampling and importance sampling as the stage-to-stage variance reduction mechanisms. Our algorithms, as well as others developed at Los Alamos [1–3], also achieve geometric convergence but each faces implementation challenges and limitations. For example, for Sequential Correlated Sampling (SCS), the evaluation of the residual (i.e., the RTE equation error) and its use in generating a distributed source for each new adaptive stage creates unavoidable new sources of approximation errors. However, SCS is fast and very robust because each adaptive stage produces a correction to the estimate of the solution obtained from all of the previous stages. For Adaptive Importance Sampling (AIS), there is both a cost and loss of precision involved in sampling from the complex importance-modified expressions that result from altering the kernel K at each adaptive stage. On the plus side, AIS is very powerful and seems to produce the most rapid error reduction per adaptive stage of those adaptive methods we know.

In [14] we introduced a new adaptive Monte Carlo method – Generalized Weighted Analog Sampling (GWAS) – for the solution of RTEs. The idea behind GWAS is to combine the power of importance sampling with strategies that loosen the restrictions associated with sampling from importance-modified transport kernels. In this way, we hope to combine rapid error reduction with fast algorithm execution in order to exponentially increase the computational efficiency. The price we pay for the flexibility of GWAS is that it is biased. The fact that GWAS is biased (though asymptotically unbiased) greatly complicates the proof that it produces geometrically convergent estimates of RTE solutions.

If we adopt the mean integrated square error (MISE) to measure the overall quality of a global estimator of an unknown function (as is done frequently for biased estimators in the literature [15]), then

$$\text{MISE} = E \int (\hat{\Phi}(x) - \Phi(x))^2 dx = \int (E \hat{\Phi}(x) - \Phi(x))^2 dx + \int \text{Var}[\hat{\Phi}(x)] dx. \quad (1.1)$$

In (1.1), $\hat{\Phi}(x)$ denotes an estimator of $\Phi(x)$.

The first term of the right-hand side of (1.1) is the integral of the squared bias, while the second term is the integrated variance. Thus, for biased estimators it is necessary to control *both* the bias *and* the variance to exhibit geometric convergence. The proof that we outlined in [14] for GWAS and provided fully in an internal report [16] does that, but it was based on overly restrictive assumptions and rested on a conjecture that is likely to be true but had not yet been proved. In addition, because the report [16] is not readily available, we were motivated to provide a proof free of restrictive assumptions and unproven conjectures. Such a proof is provided in this paper and establishes a stronger and more comprehensive theory of geometric convergence for GWAS estimators. In essence, our convergence theorem states that, with probability one, the mean square error of GWAS estimation can be made as small as we like if we generate sufficiently many, W_0 , random walks for each adaptive stage. We illustrate this refined theory with new numerical results that indicate that W_0 can be chosen of moderate size.

In the next section we establish some notation and conventions we will use throughout the paper. Then, in Section 3, we construct the basic estimating random variables and in Section 4 we introduce the GWAS adaptive strategy. In Section 5 we establish results needed to obtain geometric convergence, which is studied in Sections 8 and 9, while Section 6 provides estimates of the bias and Section 7 estimates second moments of key estimators. Our numerical results are found in Section 10 which also guides the reader through the steps of the adaptive simulation. Detailed proofs are placed in the Appendices to avoid distractions from the primary flow of ideas.

2 Mathematical preliminaries

The methods we will describe here can be applied quite generally to RTE problems involving full spatial, angular, energy and time dependence. However, to simplify both the notation and the exposition, we specialize here to time-independent, single-speed radiation transport for which an accepted model is the integral equation (2.1)

$$\Phi(r, \Omega) = \int_L \int_{S^2} K(r, \Omega; r', \Omega') \Phi(r', \Omega') d\Omega' dl' + S(r, \Omega), \tag{2.1}$$

whose solution Φ is the radiance (or vector flux). The outer integral in (2.1) is a line integral along L , which is a ray beginning at r along the direction $-\Omega$ and terminating at an interface or boundary of the (spatial) region. Analytically, $L = \{r - \rho\Omega : 0 \leq \rho \leq R\}$, where $\rho = R$ indicates the nearest interface or boundary along the direction $-\Omega$. In (2.1) the spatial vector r ranges over the interior of a closed, bounded subregion V of \mathbb{R}^3 and the directional vector Ω ranges over the unit sphere S^2 . Thus the phase space $\Gamma = V \times S^2$ is a 5-dimensional space in general. The solution, $\Phi(r, \Omega)$, describes the radiation intensity at any point (r, Ω) in the phase space Γ due to a radiation source Q inside or on the boundary of V .

The source term in (2.1) is

$$S(r, \Omega) = \int_D Q(r', \Omega) T(r, r'; \Omega) dr',$$

where the function $Q(r, \Omega)$ is the physical source density function. For example, for a fiber optic source, it is essentially the characteristic function that describes the locations and angular apertures of the fibers that produce the sources of radiation. The function T is an exponential density function

$$T(r, r'; \Omega) = \mu_t(r') \exp \left[- \int_0^{\Omega \cdot (r-r')} \mu_t(r' + s\Omega) ds \right]$$

that can account for variation in the total attenuation coefficient along each track arising in the simulation. The kernel K is

$$K(r, \Omega; r', \Omega') = \mu_s(r') p(r; \Omega, \Omega') \exp \left[- \int_0^{\Omega \cdot (\Omega - r')} \mu_t(r' + s\Omega) ds \right]. \tag{2.2}$$

In (2.2) the functions $\mu_s(r)$ and $\mu_t(r)$ are respectively the scattering and the total attenuation coefficients and $p(r; \Omega, \Omega')$ is the single scattering phase function at r ; it is the probability density function for transforming the unit direction vector Ω' to Ω at collisions at r that result in scattering. A unique solution $\Phi(r, \Omega)$ of (2.1) is assured for all $r \in V, \Omega \in \mathcal{S}^2$ when the flux of radiation $\Phi_{\text{inc}}(r, \Omega)$ incident on V from *outside of* V is specified; that is, for unit directions Ω for which $\Omega \cdot n_V < 0$, where n_V is the unit outward normal vector on V ; full details may be found in [4].

To simplify our notation, we set $P = (r, \Omega)$ and we rewrite (2.1) as

$$\Phi(P) = \int_{\Gamma} K(P, Q) \Phi(Q) dQ + S(P), \tag{2.3}$$

where the phase space Γ is 5-dimensional in general and the integration with respect to Q takes into account all unit directions Ω' while the spatial component r' of Q consists only of those spatial locations that are scattered into the direction that points to r and are transported to r , the spatial component of P , along Ω' .

The basic assumption we make about equation (2.3) is the following. There is a constant $0 < \kappa < 1$ such that

$$\max_{P \in \Gamma} \int_{\Gamma} K(P, Q) dQ = \kappa. \tag{2.4}$$

The astute reader might have expected the assumption (2.4) to be

$$\max_{P \in \Gamma} \int_{\Gamma} K(Q, P) dQ = \kappa \tag{2.5}$$

rather than (2.4), since it is clear from Equation (2.3) that the kernel $K(P, Q)$ represents scattering AT Q FROM Q to P rather than in the opposite direction. Nevertheless, the condition we adopt is (2.4) because our Monte Carlo simulation follows *backward* trajectories, which makes use of the *adjoint* kernel $K^*(P, Q) = K(Q, P)$ rather than $K(P, Q)$.

A second condition we will need is that there are two positive numbers M_S and δ_S such that

$$M_S \geq S(P) \geq \delta_S. \tag{2.6}$$

The boundedness of $S(P)$ is reasonable to assume, but in case $S(P)$ is *not* bounded away from zero but only nonnegative, we can represent the source as

$$S(P) = S_1(P) + \Delta - \Delta = S_1(P) - S_2(P),$$

where $S_1(P) = S(P) + \Delta > 0$ and $S_2(P) = \Delta$ is a small positive constant. Then $\Phi(P) = \Phi_1(P) - \Phi_2(P)$ where $\Phi_i(P)$ is the RTE solution with source $S_i(P)$ with the same kernel as the original RTE. Using the linearity of the RTE, this implies that $\Phi(P) > 0$. Essentially, the solution $\Phi(P)$ can always be treated as a difference of the solutions of two equations with identical kernels $K(P, Q)$ but with two different but positive sources $S_1(P)$ and $S_2(P)$.

Based on conditions (2.4) and (2.6), the solution of equation (2.3) satisfies

$$M_\Phi \geq \Phi(P) \geq \delta_\Phi, \tag{2.7}$$

where M_Φ and δ_Φ are positive constants. The inequality (2.7) is a consequence of the assumptions (2.4) and (2.6). First, $\Phi(P)$ is bounded away from zero since each term of the convergent Neumann series for $\Phi(P)$

$$\Phi(P) = S(P) + \int_{\Gamma} K(P, Q) S(Q) dQ + \int_{\Gamma} \int_{\Gamma} K(P, Q_1) K(Q_1, Q_2) S(Q_2) dQ_1 dQ_2 + \dots$$

is nonnegative, and the first term is $S(P)$. On the other hand, from

$$\sup_P |\Phi(P)| \leq \frac{\sup_P |S(P)|}{1 - \kappa}$$

the proof of which can be found, e.g., in [7] and [13], we see that $\Phi(P)$ is also bounded.

The *analytic model* that derives from the source $S(P)$ and kernel $K(P, Q)$ of the RTE is completed by specifying the function $S^*(P)$ that characterizes the weighted integral

$$I = \int_{\Gamma} S^*(P) \Phi(P) dP. \tag{2.8}$$

of the solution of (2.3) to be estimated.

The description of the *probabilistic model* that provides the foundation for our Monte Carlo simulations follows the treatment in [13]. We introduce the space \mathcal{B} of all random walk biographies b by

$$\mathcal{B} = \left(\bigcup_{k=1}^{\infty} \mathcal{B}_k \right) \cup \mathcal{B}_{\infty},$$

where \mathcal{B}_k denotes the biographies terminating with the k -th collision point and \mathcal{B}_{∞} is the set of biographies that never terminate. In all of our work, conditions are imposed (such as (2.4) or (2.5); see, for example, the discussion in [13, Chapter 3]) that guarantee that the set \mathcal{B}_{∞} has measure 0 so that it can be safely ignored in our probabilistic computational model. This means that any biography will either terminate or exit the region of interest after a finite number of collisions with probability one. Therefore, the subsets \mathcal{B}_k form a non-overlapping decomposition of \mathcal{B} in the sense of probability. To define a measure on \mathcal{B} we will exhibit the measure of each \mathcal{B}_k .

In general, a probability measure on the space \mathcal{B} is constructed using a pair of nonnegative functions $(p^1(P), p(P, Q))$ (called a *random walk process*, see [13]) that satisfy

$$\int_{\Gamma} p^1(P) dP = 1, \\ p(P) \equiv 1 - \int_{\Gamma} p(P, Q) dQ \geq 0. \tag{2.9}$$

These random walk functions determine random variables on the phase space Γ :

$$\xi_1 \sim p^1(P), \quad P \in \Gamma, \quad \xi_P \sim \frac{p(P, Q)}{\int_{\Gamma} p(P, Q) dQ}, \quad P \in \Gamma, \\ \eta_P \sim B(p(P)) \text{ such that } \mathcal{P}(\eta_P=1) = p(P), \quad \mathcal{P}(\eta_P=0) = 1 - p(P),$$

where $B(\hat{p}(P))$ is a binomial variable and where the symbol $\sim f(P)$ means that the random variable is sampled making use of the probability density function $f(P)$. The random variable ξ_1 generates an initial collision point P_1 (direction and location), the random variable η_P terminates the random walk at P_1 with probability $p(P_1) \equiv 1 - \int_{\Gamma} p(P_1, Q) dQ$, and with probability $\int_{\Gamma} p(P_1, Q) dQ$ generates a next collision point P_2 from the pdf $p(P_1, Q) / \int_{\Gamma} p(P_1, Q) dQ$. This process continues until the biography $b = (P_1, P_2, \dots)$ has been terminated or has escaped from the physical region V of interest. The space \mathcal{B} defines the sample space for our probability model.

This process of generating biographies $b \in \mathcal{B}$ by sampling $p^1(P)$ for initial collisions and using $p(P, Q)$ to sample for all remaining collisions induces a probability measure \mathcal{M} on \mathcal{B} , the restriction of which to \mathcal{B}_k is

$$\mathcal{M}(\mathcal{B}_k) = \int_{\mathcal{B}_k} p^1(P_1)p(P_1, P_2) \cdots K(P_{k-1}, P_k)p(P_k)dP_1dP_2 \cdots dP_k. \tag{2.10}$$

It is not difficult to show that equation (2.10) does define a probability measure; we omit the proof (but see [13, Lemma 3.1] for a similar proof). The probabilistic counterpart of (2.8) is a random variable $\xi: \mathcal{B} \rightarrow \mathcal{R}$ that associates with each biography $b \in \mathcal{B}$ a number (sometimes called a *tally*), $\xi(b)$, that is used to estimate (2.8). If ξ is an unbiased estimator of I with respect to a measure \mathcal{M} on \mathcal{B} , then

$$E[\xi] = \int_{\mathcal{B}} \xi(b)d\mathcal{M}(b) = I = \int_{\Gamma} S^*(P)\Phi(P)dP.$$

However, as we stated earlier, the GWAS estimator is biased:

$$E[\xi_{\text{GWAS}}] \neq \int_{\Gamma} S^*(P)\Phi(P)dP,$$

which significantly complicates our analysis of it.

An important special case of this general formulation is the *analog random walk process* which mimics the physics described by the transport equation (2.3).

Example 2.1

To obtain the analog random walk process, choose

$$p^1(P) = \frac{S(P)}{\int_{\Gamma} S(P)dP},$$

$$p(P, Q) = K(Q, P),$$

and therefore, because of (2.6), $0 < p(P) = 1 - \int_{\Gamma} K(P, Q) dQ < 1$, where $p(P)$ is the probability of termination at P .

The analog random walk process induces the analog probability measure on \mathcal{B} , the restriction of which to \mathcal{B}_k is

$$\mathcal{M}_A(\mathcal{B}_k) = \int_{\mathcal{B}_k} \frac{S(P_1)}{\int_{\Gamma} S(P)dP} K(P_2, P_1) \cdots K(P_k, P_{k-1})p(P_k)dP_1dP_2 \cdots dP_k.$$

The general random walk process ($p^1(P), p(P, Q)$) offers a wide variety of possibilities for Monte Carlo simulations and, corresponding to each, a probability measure on \mathcal{B} . Here our interest focuses mainly on the analog process and measure, and those that stem from the

sequence of approximate solutions $\tilde{\Phi}^s(P)$ of (2.3) that are generated by our GWAS adaptive algorithm.

The random walk process $(\hat{S}(P), \hat{K}(P, Q))$ that incorporates an approximate solution $\tilde{\Phi}(P)$ is defined by

$$\begin{aligned} \hat{S}(P) &\equiv \frac{\tilde{\Phi}(P)S^*(P)}{\int_{\Gamma} \tilde{\Phi}(P)S^*(P)dP}, \\ \hat{K}(P, Q) &\equiv \frac{K(P, Q)\tilde{\Phi}(Q)}{\int_{\Gamma} K(P, Q)\tilde{\Phi}(Q)dQ + S(P)}, \\ \hat{p}(P) &= 1 - \int_{\Gamma} \hat{K}(P, Q)dQ. \end{aligned} \tag{2.11}$$

This random walk process $(\hat{S}(P), \hat{K}(P, Q))$ induces a measure $\tilde{\mathcal{N}}$ on \mathcal{B} by defining its restriction to \mathcal{B}_k :

$$\tilde{\mathcal{N}}(\mathcal{B}_k) = \int_{\mathcal{B}_k} \hat{p}(P_k)\hat{K}(P_k, P_{k-1}) \cdots \hat{K}(P_2, P_1)\hat{S}(P_1)dP_1 \cdots dP_k.$$

It is again easy to verify that $\tilde{\mathcal{N}}$ is a probability measure.

If we now specialize $p^1(P_1)$ in (2.10) by setting $p^1(P_1) = \delta(P - P_1)$ and also set $\hat{S}(P_1) = \delta(P - P_1)$ in (2.11), we obtain measures \mathcal{N}_P and $\tilde{\mathcal{N}}_P$ defined by

$$\mathcal{N}_P(\mathcal{B}_k) = \int_{\mathcal{B}_k} p(P_k)p(P_{k-1}, P_k) \cdots p(P_1, P_2)\delta(P_1 - P)dP_1 \cdots dP_k \tag{2.12}$$

and

$$\tilde{\mathcal{N}}_P(\mathcal{B}_k) = \int_{\mathcal{B}_k} \delta(P_1 - P)\hat{K}(P_1, P_2) \cdots \hat{K}(P_{k-1}, P_k)\hat{p}(P_k)dP_1 \cdots dP_k. \tag{2.13}$$

These special probability measures are used for estimating the solution at P_1 , $\Phi(P_1)$, for any $P_1 \in \Gamma$.

3 Construction of GWAS estimators

We turn next to a discussion of estimating random variables $\xi: \mathcal{B} \rightarrow \mathbb{R}$ where \mathbb{R} designates the real numbers. It is helpful to examine the *non-adaptive* GWAS estimator first.

To estimate the expected value

$$E_{\mathcal{M}_A}(\xi) = \int_{\mathcal{B}} \xi(b)d\mathcal{M}_A(b),$$

where \mathcal{M}_A =analog measure on \mathcal{B} , by analogy with the case of definite integrals [11], we define the non-adaptive GWAS estimator

$$\xi_{\text{GWAS}} = \frac{\sum_{i=1}^N \frac{\xi(b_i)d\mathcal{M}(b)}{d\mathcal{M}_1(b)}}{\sum_{i=1}^N \frac{d\mathcal{M}_2(b)}{d\mathcal{M}_1(b)}}, \quad b_i \sim \mathcal{M}_1, \tag{3.1}$$

and \mathcal{M}_2 and \mathcal{M} are assumed to be absolutely continuous with respect to \mathcal{M}_1 , which is otherwise arbitrary. We observe that the choices $\mathcal{M}_1 = \mathcal{M}_2 = \mathcal{N} \neq \mathcal{M}$ in (3.1) reduces ξ_{GWAS} to

$$\xi_{\text{imp}} = \frac{1}{N} \sum_{i=1}^N \frac{\xi(b_i)d\mathcal{M}(b)}{d\mathcal{N}(b)}, \quad b_i \sim \mathcal{N},$$

which is an *unbiased* importance sampling estimator with importance measure \mathcal{N} . It is useful to think of (3.1) as a *biased importance sampling estimator* with the flexibility to combine easy-to-implement sampling measures (choices of \mathcal{M}_1) with good approximations of the importance “weighting” measure (choices of \mathcal{M}_2).

We can now introduce the basic components of the GWAS family of estimators. We first define

$$h(P) = \frac{\hat{p}(P)}{p(P)}\eta_P + \frac{\hat{K}(P, \xi_1)\hat{p}(\xi_1)}{p(P, \xi_1)p(\xi_1)}(1 - \eta_P)\eta_{\xi_1} + \dots + \frac{\hat{K}(P, \xi_1)}{p(P, \xi_1)} \dots \frac{\hat{K}(\xi_{k-1}, \xi_k)\hat{p}(\xi_k)}{p(\xi_{k-1}, \xi_k)p(\xi_k)}(1 - \eta_P) \dots (1 - \eta_{\xi_{k-1}})\eta_{\xi_k} + \dots \tag{3.2}$$

and

$$\omega(P) = \frac{S(P)}{p(P)}\eta_P + \frac{K(P, \xi_1)S(\xi_1)}{p(P, \xi_1)p(\xi_1)}(1 - \eta_P)\eta_{\xi_1} + \dots \tag{3.3}$$

$$+ \frac{K(P, \xi_1)}{p(P, \xi_1)} \dots \frac{K(\xi_{k-1}, \xi_k)S(\xi_k)}{p(\xi_{k-1}, \xi_k)p(\xi_k)}(1 - \eta_P) \dots (1 - \eta_{\xi_{k-1}})\eta_{\xi_k} + \dots, \tag{3.4}$$

where

$$\xi_1 \sim \frac{p(P, Q)}{\int_{\Gamma} p(P, Q)dQ}, \quad P \in \Gamma, \quad \xi_i \sim \frac{p(\xi_{i-1}, Q)}{\int_{\Gamma} p(\xi_{i-1}, Q)dQ}, \quad Q \in \Gamma, \quad i=2, 3, \dots, \\ \eta_{\xi_i} \sim B(p(\xi_i)) \text{ such that } P(\eta_{\xi_i}=1) = p(\xi_i), \quad \mathcal{P}(\eta_{\xi_i}=0) = 1 - p(\xi_i).$$

Based on (2.12), (2.13) and (3.2) and assuming that the measure $\tilde{\mathcal{N}}_P$ is absolutely continuous with respect to \mathcal{N}_P (i.e., sets $E_k \subseteq \mathcal{B}_k$ of $\tilde{\mathcal{N}}_P$ measure 0 also have \mathcal{N}_P measure

0), the Radon–Nikodym derivative $\frac{d\tilde{\mathcal{N}}_P}{d\mathcal{N}_P}$ exists and

$$\frac{d\tilde{\mathcal{N}}_P}{d\mathcal{N}_P} = h(P). \tag{3.5}$$

We make use of $h(P)$ and $\omega(P)$ defined by (3.2) and (3.4), respectively, to define, for an integer $W > 0$, the GWAS estimator

$$\tau_W(P) = \frac{\sum_{i=1}^W \omega_i(P)}{\sum_{i=1}^W h_i(P)}, \tag{3.6}$$

where $\omega_i(P)$ and $h_i(P)$ are the i -th sample values of $\omega(P)$ and $h(P)$, respectively. The random variable $\tau_W(P)$ will be used to estimate the RTE solution at a specific point P . If we choose $p(P, Q) = \hat{K}(P, Q)$, then $\mathcal{N}_P = \tilde{\mathcal{N}}_P = \mathcal{N}$, $h(P) \equiv 1$ and, if $\mathcal{N} \neq \mathcal{M}_A$, where $\mathcal{M}_A =$ analog measure on \mathcal{B} , then

$$\tau_W(P) = \frac{1}{W} \sum_{i=1}^W \omega_i(P) \frac{d\mathcal{M}_A}{d\mathcal{N}}$$

reduces to *unbiased* importance sampling with importance measure \mathcal{N} . This means that the zero variance estimators developed in [9] are special cases of (3.6). It is also apparent that if $\mathcal{N} = \mathcal{M}_A$, (3.6) reduces to crude Monte Carlo based on the estimator ω_i :

$$\tau_W(P) = \frac{1}{W} \sum_{i=1}^W \omega_i(P).$$

Now we only need to make a slight modification of the estimator $\tau_W(P)$ in order to construct the random variable to estimate the weighted integral (2.8). Assume

$$\xi_0 \sim p^1(P)$$

and define

$$\begin{aligned} g &\equiv h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)}, \\ \zeta &\equiv \omega(\xi_0) \frac{S^*(\xi_0)}{p^1(\xi_0)}. \end{aligned} \tag{3.7}$$

Then for an integer $W > 0$, define

$$\tau_W = \frac{\sum_{i=1}^W \zeta_i}{\sum_{i=1}^W g_i}, \quad (3.8)$$

where ζ_i and g_i are the i -th sample values of ζ and g , respectively. The random variable (3.8) will be used to estimate weighted integrals of the RTE solution. Notice that, as in (3.5), we also have

$$\frac{d_{\mathcal{N}} \tilde{\mathcal{N}}}{d_{\mathcal{N}}} = g. \quad (3.9)$$

If we choose $(p^1(P), p(P, Q)) = (\hat{S}(P), \hat{K}(P, Q))$, then $g \equiv 1$ and (3.8) becomes

$$\tau_W = \frac{1}{W} \sum_{i=1}^W \zeta_i, \quad (3.10)$$

which is also what we used in [9] to estimate integrals like (2.8).

In our analysis of estimator variance, the bound

$$\kappa_p \equiv \max_{P \in \Gamma} \int_{\Gamma} \frac{K^2(P, Q)}{p(P, Q)} dQ < 1 \quad (3.11)$$

plays an important role. This inequality imposes restrictions on the function $p(P, Q)$ in the construction of the random variable τ_W . Evidently, the choice $p(P, Q) = K(P, Q)$ would satisfy (3.11) owing to condition (2.4). In addition there are restrictions that are imposed on the functions $p(P)$ and $p^1(P)$; namely, there exists a positive constant δ_p such that

$$p(P), p^1(P) \geq \delta_p > 0. \quad (3.12)$$

Bounding $p(P)$ and $p^1(P)$ away from 0 is a needed assumption because both functions usually appear as factors in the denominators, as in (3.2), (3.4) and (3.7).

With the basic GWAS estimators just developed, we are ready to discuss the adaptive strategy we will use to obtain geometric convergence.

4 Adaptive strategy

It is well known that importance sampling does not necessarily reduce the variance with probability one. A poor choice of importance function can actually increase the variance.

This means that we must expect to impose constraints on the initial approximation $\tilde{\Phi}^0(\mathbf{P})$ of

the solution of the RTE (2.3) in the adaptive algorithm for GWAS, since GWAS includes importance sampling as a special case. For geometric convergence to obtain, it is also necessary to establish that all approximations $\tilde{\Phi}^s(\mathbf{P})$ beyond the initial approximation satisfy such constraints.

We therefore begin with an initial approximation $\tilde{\Phi}^0(\mathbf{P})$ of the solution of the RTE (2.3) that we assume is bounded and bounded away from 0: i.e., assume that there exist a pair of *positive* numbers $M_{\tilde{\Phi}}$ and $\delta_{\tilde{\Phi}}$ such that

$$M_{\tilde{\Phi}} \geq \tilde{\Phi}^0(\mathbf{P}) \geq \delta_{\tilde{\Phi}}. \quad (4.1)$$

This constraint follows in the same way as (2.7) when $\tilde{\Phi}^0(\mathbf{P}) = \Phi(\mathbf{P})$. Without loss of generality, we may assume that

$$M_{\tilde{\Phi}^0} > M_{\Phi} \geq \delta_{\Phi} > \delta_{\tilde{\Phi}^0}, \quad (4.2)$$

where M_{Φ} and δ_{Φ} appear in (2.7).

Based on this initial guess, we construct estimators using either (3.5) or (3.8) and the corresponding random walk process (2.11) for the first stage, which will produce a new estimated solution $\tilde{\Phi}^1(P)$ for equation (2.3). We then construct estimators and the corresponding random walk process for the next stage based on $\tilde{\Phi}^1(P)$. By induction, we assume that we have obtained the approximate solutions $\tilde{\Phi}^0(P), \tilde{\Phi}^1(P), \dots, \tilde{\Phi}^{s-1}(P)$ for the first s stages. We then construct an estimator $\tilde{\Phi}^s(P)$ of the solution for the next stage based on $h(P)$ and $\omega(P)$ defined by (3.2) and (3.4), respectively, which, in turn, are based on $\tilde{\Phi}^{s-1}(P)$. In order to be specific about which stage we are talking about, we use the notations $h^s(P)$ and $\omega^s(P)$. In the following, we will see that $\omega^s(P)$ is actually independent of s , but we will continue to use the superscript s to indicate that it is used in the s -th stage. We define

$$h^s(P) = \frac{\hat{p}^s(P)}{p(P)} \eta_P + \frac{\hat{K}^s(P, \xi_1)}{p(P, \xi_1)} \frac{\hat{p}^s(\xi_1)}{p(\xi_1)} (1 - \eta_P) \eta_{\xi_1} + \dots \\ + \frac{\hat{K}^s(P, \xi_1)}{p(P, \xi_1)} \dots \frac{\hat{K}^s(\xi_{s-1}, \xi_s)}{p_1(\xi_{s-1}, \xi_s)} \frac{\hat{p}^s(\xi_s)}{p(\xi_s)} (1 - \eta_P) \dots (1 - \eta_{\xi_{s-1}}) \eta_{\xi_s} + \dots,$$

where

$$\hat{K}^s(P, Q) = \frac{K(P, Q) \tilde{\Phi}^{s-1}(Q)}{\int_{\Gamma} K(P, Q) \tilde{\Phi}^{s-1}(Q) dQ + S(P)}, \\ \hat{p}^s(P) = \frac{S(P)}{\int_{\Gamma} K(P, Q) \tilde{\Phi}^{s-1}(Q) dQ + S(P)}, \quad (4.3)$$

so that $h^s(P)$ depends on the $(s - 1)$ -st approximate solution $\tilde{\Phi}^{s-1}(P)$. Also we define

$$\omega^s(P) = \frac{S(P)}{p(P)}\eta_P + \frac{K(P, \xi_1)}{p(P, \xi_1)} \frac{S(\xi_1)}{p(\xi_1)}(1-\eta_P)\eta_{\xi_1} + \dots + \frac{K(P, \xi_1)}{p(P, \xi_1)} \dots \frac{K(\xi_{s-1}, \xi_s)}{p(\xi_{s-1}, \xi_s)} \frac{S(\xi_s)}{p(\xi_s)}(1-\eta_P) \dots (1-\eta_{\xi_{s-1}})\eta_{\xi_s} + \dots,$$

where

$$\xi_1 \sim \frac{p(P, Q)}{\int_{\Gamma} p(P, Q) dQ}, \quad P \in \Gamma, \quad \xi_i \sim \frac{p(\xi_{i-1}, Q)}{\int_{\Gamma} p(\xi_{i-1}, Q) dQ}, \quad Q \in \Gamma \quad i=2, 3, \dots,$$

$$\mathcal{P}(\eta_{\xi_i}=1)=p(\xi_i), \quad \mathcal{P}(\eta_{\xi_i}=0)=1 - p(\xi_i).$$

We can define $\tau_W^s(P)$ for the s -th stage in a way similar to the definition of $\tau_W(P)$ in (3.6). For an integer $W > 0$, define

$$\tau_W^s(P) = \frac{\sum_{i=1}^W \omega_i^s(P)}{\sum_{i=1}^W h_i^s(P)}, \tag{4.4}$$

where $\omega_i^s(P)$ and $h_i^s(P)$ are the i -th sample values of $\omega^s(P)$ and $h^s(P)$, respectively. Our main focus from this point on is to analyze the random variable $\tau_W^s(P)$ and to establish geometric convergence with respect to s for a fixed number W (that depends on the specific problem being solved) of biographies in each stage.

As we stated above, once the initial guess $\tilde{\Phi}^0(P)$ is chosen and satisfies the constraints we need, all of the solutions $\tilde{\Phi}^s(P)$ ($s > 0$) obtained from the succeeding stages must satisfy the same requirements so that the same adaptive process can be carried out through as many stages as needed to achieve a preset estimated accuracy. To address this need, we will obtain a series of estimates of the random variables $\tau_W^s(P)$ that are independent of the stage index s .

First we define two functions, $\chi^s(P)$ and $D^s(P)$, that are of interest in terms of error estimation. Let

$$\chi^s(P) \equiv \int_{\Gamma} K(P, Q)(\tilde{\Phi}^s(Q) - \Phi(Q))dQ + (\Phi(P) - \tilde{\Phi}^s(P)) \tag{4.5}$$

$$= S(P) + \int_{\Gamma} K(P, Q)\tilde{\Phi}^s(Q)dQ - \tilde{\Phi}^s(P), \tag{4.6}$$

and

$$D^s(P) \equiv \frac{\chi^s(P)}{\tilde{\Phi}^s(p)}. \quad (4.7)$$

Equation (4.6) shows that $\chi^s(P)$ is the residual for the s -th stage; that is, it is the *equation error* when $\tilde{\Phi}^s(P)$ replaces the exact solution $\Phi(P)$ in (2.3). It follows that $D^s(P)$ can be treated a relative residual.

We also define a random variable

$$X^s \equiv \left[\zeta^s - \left(\int_{\Gamma} \Phi(P) S^*(P) dP \right) \frac{d\mathcal{N}}{d\mathcal{N}} \right] = \left[\zeta^s - \left(\int_{\Gamma} \Phi(P) S^*(P) dP \right) g^s \right], \quad (4.8)$$

and its special case

$$X^s(P) \equiv \left[\omega^s(P) - \Phi(P) \frac{d\mathcal{N}}{d\mathcal{N}} \right] = [\omega^s(P) - \Phi(P) h^s(P)] \quad (4.9)$$

when we set

$$S^*(Q) \equiv \delta(Q - P), \quad p^1(Q) = \delta(Q - P).$$

These random variables play an important role in the proof of geometric convergence.

In the following sections (until Section 8), we will omit the superscript of the function $\tilde{\Phi}^s(P)$ to avoid notational complications.

Having defined the key random variables needed for the adaptive algorithm, we are able to address the convergence behavior for the GWAS estimators.

5 Theory of geometric convergence for GWAS

5.1 Preliminary results

Our immediate purpose in this section is to state a number of technical lemmas (proofs are in Appendix A) that will be used in our final proofs of convergence. In so doing, we will also identify the restrictions that seem to be required in order to carry out the proofs. These restrictions will involve, among others, some assumptions (mostly quite benign) on the source and kernel of the transport equation and others that assure that the approximate solution obtained from each stage of the adaptive algorithm is sufficiently accurate.

We first construct equations satisfied by the mean and variance of the random variables $h(P)$ and g defined in (3.2) and (3.7), respectively.

Lemma 5.1—Assume that $h(P)$ and g are defined by (3.2) and (3.7), respectively. Then the mean of $h(P)$, $E[h(P)]$, satisfies

$$E[h(P)] = \hat{p}(P) + \int_{\Gamma} \hat{K}(P, Q) E[h(Q)] dQ \tag{5.1}$$

and the mean $E[g]$ can be expressed using $E[h(P)]$ by

$$E[g] = \int_{\Gamma} E[h(P)] \hat{S}(P) dP. \tag{5.2}$$

The variance of $h(P)$, $V[h(P)]$, satisfies

$$V[h(P)] + (E[h(P)])^2 = \int_{\Gamma} (V[h(Q)] + (E[h(Q)])^2) \frac{\hat{K}^2(P, Q)}{p(P, Q)} dQ + \frac{\hat{p}^2(P)}{p(P)}, \tag{5.3}$$

and the variance of g , $V[g]$, can be expressed using $E[h(P)]$ and $V[h(P)]$ by

$$V[g] + (E[g])^2 = \int_{\Gamma} (V[h(P)] + (E[h(P)])^2) \frac{(\hat{S}(P))^2}{p^1(P)} dP. \tag{5.4}$$

The proof is given in Appendix A.

The definition (3.6) of the random variable $\tau_w(P)$ suggests that the function $h_w(P)$ and, therefore $h(P)$, should be bounded away from zero in some sense. The next lemma establishes that $h(P)$ is bounded away from zero probabilistically.

Lemma 5.2—Assume that equation (2.3) satisfies conditions (2.4) and (4.1), and functions $p(P, Q)$ and $p^1(P)$ satisfy (2.9). Assume that $\tilde{\Phi}(P)$ is an approximate solution of equation (2.3) and let positive constants ε_1 , ε_2 and α exist such that

$$\varepsilon_1 + \varepsilon_2 < 1, \quad \alpha < 1,$$

and

$$\mathcal{P} \left\{ \begin{aligned} &M_{\tilde{\Phi}} \geq \tilde{\Phi}(P) \geq \delta_{\tilde{\Phi}} > 1 - \varepsilon_1, \\ &\max_{P \in \Gamma} \frac{|\tilde{\Phi}(P) - \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ - S(P)|}{|\tilde{\Phi}(P)|} \leq \frac{\alpha(1 - \kappa_p) \delta_{\tilde{\Phi}}^2}{(M_{\tilde{\Phi}}(1 + \kappa) + M_S) M_{\tilde{\Phi}}} \end{aligned} \right\} > 1 - \varepsilon_2. \tag{5.5}$$

Then we have

$$\min_{P \in \Gamma} E[h(P)] \geq \delta_h \quad (5.6)$$

and

$$\max_{P \in \Gamma} (V_h[h(P)] + (E_h[h(P)])^2) \leq M_V,$$

where

$$\delta_h \equiv \frac{\delta_s}{M_{\Phi} \kappa + \delta_s}, \quad M_V \equiv \frac{M_s^2}{(1 - \alpha)(1 - \kappa_p) \delta_{\Phi}^2 \delta_P}. \quad (5.7)$$

The proof is given in Appendix A.

We continue our analysis of lower bounds for $h(P)$. Lemma 5.2 discusses lower bounds for the expectation of $h(P)$. In this Corollary the lower bound is established for the sample mean

$$\frac{1}{W} \sum_{w=1}^W h_w(P)$$

for sufficiently large W .

Corollary 5.3—Assume that the conditions of Lemmas 5.1 and 5.2 are satisfied. Then for any $\varepsilon_3 > 0$, there exists an integer W_h such that when $W > W_h$

$$\mathcal{P} \left\{ \frac{1}{W} \sum_{w=1}^W h_w(P) \geq \frac{1}{2} \delta_h \right\} \geq 1 - \varepsilon_3,$$

where δ_h is defined by (5.7) and $h_w(P)$ is the w -th sample of $h(P)$.

The proof is in Appendix A.

The following two lemmas concern X^s defined in (4.8). As indicated before, we omit the superscript s to avoid notational confusion. The next lemma will be used in the estimation of the bias, but not in the final theorem about the geometric convergence.

Lemma 5.4—Assume that equation (2.3) satisfies conditions (2.4) and (2.6). Assume that $\tilde{\Phi}(P) > 0$ is an approximate solution of equation (2.3), and there exist constants $0 < \beta, \varepsilon_1 < 1$ such that ($D(P)$ is defined in (4.7))

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |D(P)| \leq 1 - \frac{\kappa}{\beta} \right\} > 1 - \varepsilon_1. \quad (5.8)$$

Then we have

$$\mathcal{P} \left\{ E[|X|] \leq c_1 \max_{P \in \Gamma} |D(P)| + c_2 \frac{|\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q)) S^*(Q) dQ|}{(\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ)} \right\} > 1 - \varepsilon_1,$$

where

$$c_1 \equiv \frac{\beta \max_{P \in \Gamma} S(P) \int_{\Gamma} |S^*(P)| dP}{(1-\beta)\kappa(1-\kappa)}, \quad c_2 \equiv \frac{\beta \max_{P \in \Gamma} S(P) \int_{\Gamma} |S^*(P)| dP}{(1-\beta)\kappa}.$$

A very important special case is obtained if we set

$$S^*(P_1) \equiv \delta(P_1 - P), \quad p^1(P_1) = \delta(P_1 - P). \quad (5.9)$$

In this case, we obtain

$$E[|X(P)|] \leq \tilde{c}_1 \max_{P \in \Gamma} |D(P)| + \tilde{c}_2 \max_{P \in \Gamma} \left| \frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)} \right|,$$

where

$$X(P) \equiv [\omega(P) - \Phi(P)h(P)] = \left[\omega(P) - \Phi(P) \frac{d\tilde{\nu}_P}{d\nu_P} \right]$$

and

$$\tilde{c}_1 \equiv \frac{\beta \max_{P \in \Gamma} S(P)}{(1-\beta)\kappa(1-\kappa)}, \quad \tilde{c}_2 \equiv \frac{\beta \max_{P \in \Gamma} S(P)}{(1-\beta)\kappa}.$$

Remark 5.5—From (5.8), we can see that, if $\kappa = 1$, we would not be able to find a β ($0 < \beta < 1$) to make (5.8) happen.

The proof of Lemma 5.4 is given in Appendix A.

The next lemma deals with the second moments of the random variables $X(P)$ and X .

Lemma 5.6—Assume that equation (2.3) satisfies conditions (2.4) and (2.6), and functions $p(P, Q)$ and $p^1(P)$ satisfy conditions (3.11) and (3.12). Assume that $\tilde{\Phi}(P) > 0$ is an approximate solution of equation (2.3), and there exist constants $0 < \gamma, \varepsilon_1 < 1$ such that

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |D(P)| \leq 1 - \sqrt{\frac{\kappa p}{\gamma}} \right\} > 1 - \varepsilon_1. \quad (5.10)$$

Then we have

$$\mathcal{P} \left\{ E[X^2] \leq c_1 \max_{P \in \Gamma} |D(P)|^2 + c_2 \left(\frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q)) S^*(Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ} \right)^2 \right\} > 1 - \varepsilon_1, \tag{5.11}$$

where

$$c_1 \equiv \frac{2 \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) M_S^2 (1 + \kappa_p) \gamma}{\kappa_p (1 - \gamma) (1 - \kappa_p) \delta_P \left(\sqrt{\frac{\kappa_p}{\gamma}} - \kappa_p \right)}, \quad c_2 \equiv \frac{2 \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) \gamma M_S^2}{\kappa_p (1 - \gamma) \delta_P}.$$

A very important special case is obtained when we assume (5.9). In this case we have

$$\mathcal{P} \left\{ \max_{P \in \Gamma} E[X^2(P)] \leq \tilde{c}_1 \max_{P \in \Gamma} |D(P)|^2 + \tilde{c}_2 \max_{P \in \Gamma} \left| \frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)} \right|^2 \right\} > 1 - \varepsilon_2, \tag{5.12}$$

where

$$\tilde{c}_1 \equiv \frac{2M_S^2(1+\kappa_p)\gamma}{\kappa_p(1-\gamma)(1-\kappa_p)\left(\sqrt{\frac{\kappa_p}{\gamma}}-\kappa_p\right)\delta_P}, \quad \tilde{c}_2 \equiv \frac{2M_S^2\gamma}{\kappa_p(1-\gamma)\delta_P}.$$

The proof is given in Appendix A.

6 Estimation of the bias

In this section, we discuss how the bias changes when we use estimators τ_W and $\tau_W(P)$ to estimate the solution $\Phi(P)$ and the integral I (defined by (2.8)), respectively. We define the bias by

$$Z(P) \equiv |E[\tau_W(P)] - \Phi(P)|, \quad Z \equiv |E[\tau_W] - I|,$$

where I is defined by (2.8).

The following theorem is a combination of Corollary 5.3 and Lemma 5.4. This theorem will not be used in our final theorem about geometric convergence, but it shows how the bias is controlled by the errors of the approximation.

Theorem 6.1

Assume that equation (2.3) satisfies conditions (2.4) and (2.6). Assume that $\tilde{\Phi}(P) > 0$ is an approximate solution of equation (2.3). Then for $\varepsilon > 0$, there exists a constant $0 < \beta < 1$ such that

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |D(P)| \leq 1 - \frac{\kappa}{\beta} \right\} > 1 - \varepsilon.$$

Also, there exist constants δ_h , c_1 , c_2 , \tilde{c}_1 and \tilde{c}_2 such that

$$\begin{aligned} \mathcal{P} \left\{ Z \leq \frac{c_1}{\delta_h} \max_{P \in \Gamma} |D(P)| + \frac{c_2}{\delta_h} \frac{|\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q)) S^*(Q) dQ|}{(\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ)} \right\} &\geq 1 \\ - \varepsilon, \mathcal{P} \left\{ Z(P) \leq \frac{c_1}{\delta_h} \max_{P \in \Gamma} |D(P)| + \frac{\tilde{c}_2}{\delta_h} \max_{P \in \Gamma} \left| \frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)} \right| \right\} &\geq 1 - \varepsilon, \end{aligned} \tag{6.1}$$

where δ_h is defined by (5.7) and c_1 , c_2 , \tilde{c}_1 and \tilde{c}_2 are independent of $\tilde{\Phi}(P)$.

The proof is given in Appendix B.

A special case results from the choice $\mathcal{N} = \tilde{\mathcal{N}}$ which is made for adaptive importance sampling methods; please refer to [9]. Then

$$\frac{d_{\tilde{\mathcal{N}}_P}}{d_{\mathcal{N}_P}} = 1$$

and consequently,

$$Z(P) = \frac{1}{W} \left| E \left[\sum_{i=1}^W (\omega(P) - \Phi(P) \frac{d_{\tilde{\mathcal{N}}_P}}{d_{\mathcal{N}_P}}) \right] \right| = 0.$$

This shows once again that the adaptive importance sampling methods are produced as a special case of the GWAS method for which the bias is zero as a result of using the special sampling functions that are used to generate biographies in AIS [9].

7 Estimation of the second moments

In this section, we will derive an estimate of the second moments, which makes use of a combination of Corollary 5.3 and Lemma 5.6.

Theorem 7.1

Assume that equation (2.3) satisfies conditions (2.4) and (2.6), and functions $p(P, Q)$ and $p^1(P)$ satisfy conditions (3.11) and (3.12). Assume that $\tilde{\Phi}(P)$ is an approximate solution of equation (2.3) and there exist positive constants ε_1 , ε_2 and γ such that

$$2\varepsilon_1 + \varepsilon_2 < 1, \quad \gamma < 1, \tag{7.1}$$

and

$$\mathcal{P}\{M_{\tilde{\Phi}} \geq \tilde{\Phi}(P) \geq \delta_{\tilde{\Phi}}\} > 1 - \varepsilon_1, \quad (7.2)$$

$$\mathcal{P}\left\{\max_{P \in \Gamma} |D(P)| \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}}\right\} > 1 - \varepsilon_2. \quad (7.3)$$

Then there exist constants c_1, c_2, \tilde{c}_1 and \tilde{c}_2 , that are independent of $\tilde{\Phi}(P)$, such that

$$\begin{aligned} \mathcal{P}\left\{V \leq \frac{c_1}{W} \max_{P \in \Gamma} |D(P)|^2 + \frac{c_2}{W} \left(\frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q)) S^*(Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ}\right)^2\right\} &\geq 1 - \varepsilon_1 - \varepsilon_2, \\ \mathcal{P}\left\{V(P) \leq \frac{\tilde{c}_1}{W} \max_{P \in \Gamma} |D(P)|^2 + \frac{\tilde{c}_2}{W} \max_{P \in \Gamma} \left|\frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)}\right|^2\right\} &\geq 1 - \varepsilon_1 - \varepsilon_2, \end{aligned} \quad (7.4)$$

and

$$V(P) \leq \frac{C_1}{W}, \quad V \leq \frac{C_2}{W} \quad (7.5)$$

where

$$V(P) \equiv E[\tau_w(P) - \Phi(P)]^2, \quad V \equiv E[\tau_w - I]^2.$$

The proof is given in Appendix C.

8 Geometric convergence

Now we turn to our theory of geometric convergence. Recall in Section 4, starting from an initial guess $\tilde{\Phi}^0(P)$, we generate a series of approximate solutions of equation (2.3):

$$\tilde{\Phi}^0(P), \tilde{\Phi}^1(P), \dots, \tilde{\Phi}^s(P), \dots$$

Our main theorem states that this series converges in a probabilistic sense.

Theorem 8.1

Assume that equation (2.3) satisfies conditions (2.4) and (2.6), and functions $p(P, Q)$ and $p^1(P)$ satisfy conditions (3.11) and (3.12). We choose an initial approximation $\tilde{\Phi}^0(P)$ of equation (2.3) that satisfies

$$M_{\tilde{\Phi}} \geq \tilde{\Phi}^0(P) \geq \delta_{\tilde{\Phi}} \quad (8.1)$$

and

$$\max_{P \in \Gamma} |D^0(P)| \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}} \quad (8.2)$$

for a constant $1 > \gamma > 0$. We construct a series of approximate solutions

$\tilde{\Phi}^0(P), \tilde{\Phi}^1(P), \tilde{\Phi}^2(P), \dots, \tilde{\Phi}^s(P)$ following the algorithm described in Section 4. Then for any $\epsilon > 0$ and $0 < \lambda < 1$, there must be a $W_0 > 0$ such that when $W = W_0$,

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^s(P)| \leq \lambda \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^{s-1}(P)| \right\} \geq 1 - \epsilon.$$

The proof is given in Appendix D.

9 Flux expansion

The discussion in Section 8 concerns using the random variable $\tau_W(P)$ to estimate $\Phi(P)$ for any P . Our development is general, and the method can be used to produce an approximate solution at any particular point $P \in \Gamma$. However, as we observed earlier, examination of the formulas in Section 8 shows that in order to carry out the procedure prescribed there, we require a series of approximations $\tilde{\Phi}^s(Q)$ to the solution $\Phi(Q)$ that can be evaluated at every point $Q \in \Gamma$. We next provide an algorithm for accomplishing this, making use of the GWAS algorithm. The method depends on expanding the RTE solution in a set of basis functions and using Monte Carlo methods to estimate a finite number of the expansion coefficients adaptively. We have used this idea previously in deriving adaptive methods based on correlated sampling and importance sampling [8, 9].

Assume that $\{f_i(P)\}_{i=0}^\infty$ is a complete orthonormal system of basis functions for the space \mathcal{S} of RTE solutions in the sense that such $\Phi(P)$ can be expanded as a convergent series

$$\Phi(P) = \sum_{i=0}^\infty a_i f_i(P), \quad (9.1)$$

in the space \mathcal{S} , where we assume $\{f_i(P)\}_{i=0}^\infty$ to be bounded by a constant M_f :

$$\max |f_i(P)| \leq M_f.$$

In our work we have used the Legendre polynomials in each independent variable (more generally the spherical harmonics) to provide this basis set. We denote the norm in this space by $\|\cdot\|$; in practice, we have chosen this norm to be the strongest possible norm:

$$\|\Phi\| = \max_{P \in \Gamma} |\Phi(P)|.$$

The convergence of the infinite series in (9.1) implies that for any δ , there exists an $N = N(\delta, \Phi)$ such that

$$r_N \equiv \max_{P \in \Gamma} \left| \sum_{i=N+1}^{\infty} a_i f_i(P) \right| < \delta. \tag{9.2}$$

Under this assumption, our task is to estimate the a_j for $i = 0, \dots, N$. Most importantly, we only have to deal with those functions expressed as a finite linear combination of the basis functions, which are then continuous functions – in fact, polynomials.

Specifically, we will find approximate solutions in the form

$$\Phi(P) \approx \sum_{i=0}^N a_i f_i(P),$$

which amounts to estimating each of the coefficients a_j ,

$$a_i = \int_{\Gamma} \Phi(P) f_i(P) dP$$

for $i = 0, 1, 2, \dots, N$. The estimator τ_W defined by (3.8) in Section 3 will be used for this purpose with the function $f_i(P)$ playing the role of $S^*(P)$.

The procedure we use is very similar to the one described earlier using the random variable $\tau_W(P)$, but for clarity and completeness, we write down the algorithm.

First, find an initial approximation $\tilde{\Phi}^0(P)$ of the solution, which can be obtained using any method. We usually process two short stages of our SCS algorithm to obtain an approximate solution, $\tilde{\Phi}^0(P)$, that satisfies (8.1) and (8.2). Then, using the estimator τ_W^1 , we can obtain approximate values for the first $N + 1$ coefficients, $\tilde{a}_0^1, \tilde{a}_1^1, \dots, \tilde{a}_N^1$, and these define the approximate solution for the first adaptive stage

$$\tilde{\Phi}^1(P) = \sum_{i=0}^N \tilde{a}_i^1 f_i(P).$$

Note that the construction of τ_W^1 is based on the approximate solution $\tilde{\Phi}^0(P)$ (the initial approximation).

Suppose that we have obtained the approximate solutions up to the $(s - 1)$ -st stage,

$\tilde{\Phi}^0(P), \tilde{\Phi}^1(P), \dots, \tilde{\Phi}^{s-1}(P)$, each having the form

$$\tilde{\Phi}^k(P) = \sum_{i=0}^N \tilde{a}_i^k f_i(P), \quad k=0, 1, \dots, s-1.$$

Based on $\tilde{\Phi}^{s-1}(P)$, we can again construct estimators τ_W^s for the first $N+1$ coefficients a_i to obtain an approximate solution for the s -th stage

$$\tilde{\Phi}^s(P) = \sum_{i=0}^N \tilde{a}_i^s f_i(P). \tag{9.3}$$

We now prove that the series $\tilde{\Phi}^k(P)$ so constructed converges geometrically. The theorem below differs from Theorem 8.1 owing to the extra term r_N defined by (9.2).

Theorem 9.1

Assume that equation (2.3) satisfies conditions (2.4) and (2.6), and functions $p(P, Q)$ and $p^1(P)$ satisfy conditions (3.11) and (3.12). Furthermore, we assume that an initial approximation $\tilde{\Phi}^0(P)$ of equation (2.3) that satisfies

$$M_{\tilde{\Phi}} \geq \tilde{\Phi}^0(P) \geq \delta_{\tilde{\Phi}} \tag{9.4}$$

and

$$\max_{P \in \Gamma} |D^0(P)| \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}}, \tag{9.5}$$

for a constant $1 > \gamma > 0$. Then for any $\epsilon > 0$ and $0 < \lambda < 1$, there must be a $W_0 > 0$ such that when $W > W_0$,

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^s(P)| \leq \lambda \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^{s-1}(P)| + r_N \right\} \geq 1 - \epsilon,$$

where r_N is defined by (9.2).

The proof is given in Appendix E.

10 Numerical experimentation

In this section, we will apply GWAS to a model transport problem – the bidirectional transport problem

$$\begin{cases} \frac{d\varphi}{dx} = -\sum_t \varphi + \sum_s (p_{11}\varphi + p_{12}\psi) + R_1(x), & 0 < x \leq T, \\ -\frac{d\psi}{dx} = -\sum_t \psi + \sum_s (p_{21}\varphi + p_{22}\psi) + R_2(x), & 0 \leq x < T, \\ \varphi(0) = Q_0, \quad \psi(T) = Q_T. \end{cases} \quad (10.1)$$

where

$$Q_0 \geq 0, \quad Q_T \geq 0, \quad R_1(x) \geq 0, \quad R_2(x) \geq 0, \quad p_{ij} \geq 0, \quad p_{11} + p_{21} = p_{12} + p_{22} = 1.$$

In [9], we have examined the behavior of AIS on several cases of this problem (for different scattering rates). Here we also consider the same cases and compare the results with the ones obtained in [9].

First, as we did in [9], we convert (10.1) to integral form

$$\begin{cases} \varphi(x) = p_{11} \int_0^x e^{-\sum_t(x-y)} \varphi(y) dy + p_{12} \int_0^x e^{-\sum_t(x-y)} \psi(y) dy + S_1(x), \\ \psi(x) = p_{21} \int_x^T e^{-\sum_t(y-x)} \varphi(y) dy + p_{22} \int_x^T e^{-\sum_t(y-x)} \psi(y) dy + S_2(x), \end{cases} \quad (10.2)$$

where

$$S_1(x) \equiv Q_0 e^{-\sum_t x} + \int_0^x e^{-\sum_t(x-y)} R_1(y) dy, \quad S_2(x) \equiv Q_T e^{-\sum_t(T-x)} + \int_x^T e^{-\sum_t(y-x)} R_2(y) dy.$$

Or, using the scattering kernel functions, we can write (10.2) in the form

$$\begin{cases} \varphi(x) = \int_0^T K_{11}(x, y) \varphi(y) dy + \int_0^T K_{12}(x, y) \psi(y) dy + S_1(x), \\ \psi(x) = \int_0^T K_{21}(x, y) \varphi(y) dy + \int_0^T K_{22}(x, y) \psi(y) dy + S_2(x), \end{cases} \quad (10.3)$$

where

$$\begin{aligned} K_{11}(x, y) &\equiv p_{11} \sum_s e^{-\sum_t(x-y)}, & x \geq y, & \quad K_{12}(x, y) \equiv p_{12} \sum_s e^{-\sum_t(x-y)}, & x \geq y, \\ K_{21}(x, y) &\equiv p_{21} \sum_s e^{-\sum_t(y-x)}, & x \leq y, & \quad K_{22}(x, y) \equiv p_{22} \sum_s e^{-\sum_t(y-x)}, & x \leq y. \end{aligned}$$

We will solve equation (10.3) by expanding the solution in Legendre polynomials, but we will keep the notation general to describe the process. Assuming that we have obtained the solution $(\tilde{\varphi}^i(x), \tilde{\psi}^i(x))$ for $i = 0, 1, 2, \dots, m - 1$, we want to expand $\varphi^m(x)$ (the process is the same for $\psi^m(x)$) using the orthonormal basis functions $\{f_i(x)\}_0^\infty$:

$$\varphi^m(x) = \sum_{i=0}^{\infty} a_i^m f_i(x).$$

Then by orthonormality, we have

$$a_i^m = \int_0^T \varphi^m(x) f_i(x) dx. \tag{10.4}$$

The algorithm developed in previous sections will be used to estimate the integrals defined by (10.4) for $i = 0, \dots, I$. To simplify matters, we describe how to estimate the integrals

$$I_1 \equiv \int_0^T \varphi^m(x) S_1^*(x) dx, \tag{10.5}$$

$$I_2 \equiv \int_0^T \psi^m(x) S_2^*(x) dx, \tag{10.6}$$

where $S_j^*(x)$ is any nonnegative function (not identically zero), and the subscript is used to distinguish (10.5) from (10.6). We use the estimator τ_W defined by (3.10) to estimate the integrals $I_j, j = 1, 2$.

Our simulation (to estimate I) is carried out through the following steps.

10.1 Construct estimator

The first step is to choose a pair of nonnegative functions $(p^1(P), p(P, Q))$ satisfying (2.9), (3.11) and (3.12). In our case, we have $(p_1^1(x), p_2^1(x), p_{11}(x, y), p_{12}(x, y), p_{21}(x, y), p_{22}(x, y))$ satisfying

$$\begin{aligned} \int_0^T p_1^1(x) dx = 1, \quad p_1(x) &\equiv 1 - \int_0^T p_{11}(x, y) dy - \int_0^T p_{12}(x, y) dy, \\ \int_0^T p_2^1(x) dx = 1, \quad p_2(x) &\equiv 1 - \int_0^T p_{21}(x, y) dy - \int_0^T p_{22}(x, y) dy, \end{aligned} \tag{10.7}$$

and condition (3.11) will then be

$$\kappa_{ip} \equiv \max_{x \in \Gamma} \int_0^T \frac{K_{i1}^2(x, y)}{p_{i1}(x, y)} dy + \max_{x \in \Gamma} \int_0^T \frac{K_{i2}^2(x, y)}{p_{i2}(x, y)} dy \leq 1, \tag{10.8}$$

where at least one $\kappa_{ip} < 1$.

In practice it is easy to choose $(p_1^1(x), p_2^1(x), p_{11}(x, y), p_{12}(x, y), p_{21}(x, y), p_{22}(x, y))$ even though equations (10.7) and (10.8) look complicated. According to equation (4.3), we need to construct a function pair $(\hat{S}^m(P), \hat{K}^m(P, Q))$ which, in our case, will be $(\hat{S}_1^m(x), \hat{S}_2^m(x), \hat{K}_{11}^m(x, y), \hat{K}_{12}^m(x, y), \hat{K}_{21}^m(x, y), \hat{K}_{22}^m(x, y))$, where

$$\begin{aligned} \hat{S}_1^m(x) &= \frac{\tilde{\varphi}^{m-1}(x)S_1^*(x)}{\int_0^T \tilde{\varphi}^{m-1}(x)S_1^*(x)dx}, & \hat{S}_2^m(x) &= \frac{\tilde{\psi}^{m-1}(x)S_1^*(x)}{\int_0^T \tilde{\psi}^{m-1}(x)S_2^*(x)dx}, \\ \hat{K}_{11}^m(x, y) &= \frac{K_{11}(x, y)\tilde{\varphi}^{m-1}(y)}{\int_0^T K_{11}(x, y)\tilde{\varphi}^{m-1}(y)dy + \int_0^T K_{12}(x, y)\tilde{\psi}^{m-1}(y)dy + S_1(x)}, \\ \hat{K}_{12}^m(x, y) &= \frac{K_{12}(x, y)\tilde{\psi}^{m-1}(y)}{\int_0^T K_{11}(x, y)\tilde{\varphi}^{m-1}(y)dy + \int_0^T K_{12}(x, y)\tilde{\psi}^{m-1}(y)dy + S_1(x)}, \\ \hat{K}_{21}^m(x, y) &= \frac{K_{21}(x, y)\tilde{\varphi}^{m-1}(y)}{\int_0^T K_{21}(x, y)\tilde{\varphi}^{m-1}(y)dy + \int_0^T K_{22}(x, y)\tilde{\psi}^{m-1}(y)dy + S_2(x)}, \\ \hat{K}_{22}^m(x, y) &= \frac{K_{22}(x, y)\tilde{\psi}^{m-1}(y)}{\int_0^T K_{21}(x, y)\tilde{\varphi}^{m-1}(y)dy + \int_0^T K_{22}(x, y)\tilde{\psi}^{m-1}(y)dy + S_2(x)}, \end{aligned}$$

and then define

$$\begin{aligned} \hat{p}_1^m(x) &\equiv \frac{S_1(x)}{\int_0^T K_{11}(x, y)\tilde{\varphi}^{m-1}(y)dy + \int_0^T K_{12}(x, y)\tilde{\psi}^{m-1}(y)dy + S_1(x)}, \\ \hat{p}_2^m(x) &\equiv \frac{S_2(x)}{\int_0^T K_{21}(x, y)\tilde{\varphi}^{m-1}(y)dy + \int_0^T K_{22}(x, y)\tilde{\psi}^{m-1}(y)dy + S_2(x)}. \end{aligned}$$

Now we can formally write down the estimator τ_w^m of the integral I defined by (10.5). All we do is to follow equations (3.2), (3.4), (3.7) and (3.8). First, for each random walk labeled by w , we define g_w^m (note, $j_0 = 1$ because we are estimating $\varphi^m(x)$, not $\psi^m(x)$)

$$\begin{aligned} g_w^m &= \frac{\hat{S}_{j_0}(\xi_0) \hat{p}_{j_0}^m(\xi_0)}{p_{j_0}^1(\xi_0) p_{j_0}(\xi_0)} \eta_{\xi_0} \\ &+ \frac{\hat{S}_{j_0}(\xi_0) \hat{K}_{j_0 j_1}^m(\xi_0, \xi_1) \hat{p}_{j_1}^m(\xi_1)}{p_{j_0}^1(\xi_0) p_{j_0 j_1}(\xi_0, \xi_1) p_{j_1}(\xi_1)} (1 \\ &\quad - \eta_{\xi_0}) \eta_{\xi_1} \\ &+ \dots + \frac{\hat{S}_{j_0}(\xi_0) \hat{K}_{j_0 j_1}^m(\xi_0, \xi_1) \dots \hat{K}_{j_{k-1} j_k}^m(\xi_{k-1}, \xi_k) \hat{p}_{j_k}^m(\xi_k)}{p_{j_0}^1(\xi_0) p_{j_0 j_1}(\xi_0, \xi_1) \dots p_{j_{k-1} j_k}(\xi_{k-1}, \xi_k) p_{j_k}(\xi_k)} (1 - \eta_{\xi_0}) \dots (1 \\ &\quad - \eta_{\xi_{k-1}}) \eta_{\xi_k} + \dots \end{aligned}$$

and then ζ_w^m

$$\begin{aligned} \zeta_w^m &= \frac{S_{j_0}^*(\xi_0) S_{j_0}(\xi_0)}{p_{j_0}^1(\xi_0) p_{j_0}(\xi_0)} \eta_{\xi_0} + \frac{S_{j_0}^*(\xi_0) K_{j_0 j_1}(\xi_0, \xi_1) S_{j_1}(\xi_1)}{p_{j_0}^1(\xi_0) p_{j_0 j_1}(\xi_0, \xi_1) p_{j_1}(\xi_1)} (1 - \eta_{\xi_0}) \eta_{\xi_1} + \dots \\ &+ \frac{S_{j_0}^*(\xi_0) K_{j_0 j_1}(\xi_0, \xi_1) \dots K_{j_{k-1} j_k}(\xi_{k-1}, \xi_k) S_{j_1}(\xi_k)}{p_{j_0}^1(\xi_0) p_{j_0 j_1}(\xi_0, \xi_1) \dots p_{j_{k-1} j_k}(\xi_{k-1}, \xi_k) p_{j_1}(\xi_k)} (1 - \eta_{\xi_0}) \dots (1 - \eta_{\xi_{k-1}}) \eta_{\xi_k} + \dots, \end{aligned}$$

where

$$\xi_0 \sim \frac{\tilde{\varphi}^{m-1}(x)S_1^*(x)}{\int_0^T \tilde{\varphi}^{m-1}(x)S_1^*(x)dx}, \quad \xi_i \sim \frac{p_{j_{i-1}j_i}(\xi_{i-1}, x)}{\int_0^T p_{j_{i-1}j_i}(\xi_{i-1}, x)dx}, \quad i=1, 2, 3, \dots,$$

$$\mathcal{P}(\eta_{\xi_0}=1)=p_{p_{j_0}}(\xi_0), \quad \mathcal{P}(\eta_{\xi_1}=0)=1 - p_{j_1}(\xi_1).$$

The estimator τ_w^m is then defined by

$$\tau_w^m = \frac{\sum_{w=1}^W \zeta_w^m}{\sum_{w=1}^W g_w^m}. \tag{10.9}$$

10.2 Simulation

To clarify the generation of the photon biographies and their tallies, we follow one random walk through estimator τ_w^m . We first sample the starting point x_0 (note, $j_0 = 1$ as indicated before)

$$x_0 \sim p_{j_0}^1(x).$$

We then generate a pseudorandom number r (uniformly distributed over $(0, 1)$) and check which of the following inequalities is true (see (10.7)):

$$\begin{aligned} & r < p_{j_0}(x_0), \\ p_{j_0}(x_0) & \leq r < \int_0^T p_{j_01}(x_0, y)dy + p_{j_0}(x_0), \\ & r \geq \int_0^T p_{j_01}(x_0, y)dy + p_{j_0}(x_0). \end{aligned} \tag{10.10}$$

Let us consider each of these possibilities.

Case 1—If $r < p_{j_0}(x_0)$, then the random walk is terminated (absorbed) with the contribution

$$\frac{\hat{S}_{j_0}(x_0) \hat{p}_{j_0}^m(x_0)}{p_{j_0}^1(x_0) p_{j_0}(x_0)}$$

to g_w^m and the contribution

$$\frac{S_{j_0}^*(x_0) S_{j_0}(x_0)}{p_{j_0}^1(x_0) p_{j_0}(x_0)}$$

to ζ_w^m . We record these contributions and then go on to the next random walk.

Case 2—If $p_{j_0 1}(x_0) \leq r < \int_0^T p_{j_0}(x_0, y) dy + p_{j_0}(x_0)$, then the random walk is scattered and the next collision point x_1 is sampled from

$$x_1 \sim \frac{p_{j_0 1}(x_0, x)}{\int_0^T p_{j_0 1}(x_0, x) dy}$$

Case 3—If $r \geq \int_0^T p_{j_0 1}(x_0, y) dy + p_{j_0}(x_0)$, then the random walk is scattered (in a different direction) and the next collision point x_1 is sampled from

$$x_1 \sim \frac{p_{j_0 2}(x_0, x)}{\int_0^T p_{j_0 2}(x_0, x) dy}$$

We combine Cases 2 and 3 of (10.10) into one. That is, the next collision point x_1 is sampled from

$$x_1 \sim \frac{p_{j_0 j_1}(x_0, x)}{\int_0^T p_{j_0 j_1}(x_0, x) dy}$$

Suppose the random walk is scattered and we go one step further. We sample a random number r (uniformly distributed over $(0, 1)$) and check which of the following inequalities is true:

$$\begin{aligned} & r < p_{j_1}(x_1), \\ p_{j_1}(x_1) & \leq r < \int_0^T p_{j_1 1}(x_1, y) dy + p_{j_1}(x_1), \\ & r \geq \int_0^T p_{j_1 1}(x_1, y) dy + p_{j_1}(x_1). \end{aligned} \tag{10.11}$$

Again, we consider them one by one.

Case 1—If $r < p_{j_1}(x_1)$, then the random walk is terminated (absorbed) with the contribution

$$\frac{\hat{S}_{j_0}^*(x_0)}{p_{j_0}^1(x_0)} \frac{\hat{K}_{j_0 j_1}^m(x_0, x_1)}{p_{j_0 j_1}(x_0, x_1)} \frac{\hat{p}_{j_1}^m(x_1)}{p_{j_1}(x_1)}$$

to g_w^m and the contribution

$$\frac{S_{j_0}^*(x_0)}{p_{j_0}^1(x_0)} \frac{K_{j_0 j_1}(x_0, x_1)}{p_{j_0 j_1}(x_0, x_1)} \frac{S_{j_1}(x_1)}{p_{j_1}(x_1)}$$

to ζ_w^m . We record these contributions and then go on to the next random walk.

Case 2—If $p_{j_1}(x_1) \leq r < \int_0^T p_{j_1 1}(x_1, y) dy + p_{j_1}(x_1)$, then the random walk is scattered and the next collision point x_2 is sampled from

$$x_2 \sim \frac{p_{j_1 1}(x_1, x)}{\int_0^T p_{j_1 1}(x_1, y) dy}.$$

Case 3—If $r \geq \int_0^T p_{j_1 1}(x_1, y) dy + p_{j_1}(x_1)$, then the random walk is scattered (in a different direction) and the next collision point x_2 is sampled from

$$x_2 \sim \frac{p_{j_1 2}(x_1, x)}{\int_0^T p_{j_1 2}(x_1, y) dy}.$$

We again combine Cases 2 and 3 of (10.11) into one. That is, the next collision point x_2 is sampled from

$$x_2 \sim \frac{p_{j_1 j_2}(x_1, x)}{\int_0^T p_{j_1 j_2}(x_1, y) dy}.$$

This process can be continued until the random walk is terminated (absorbed). As we finish all the random walks, simply using (10.9), we obtain the desired estimate τ_W^m .

We will test this algorithm using input data typical of tissue problems:

$$\begin{aligned} T &= 10, \\ \sum_a &= 0.01, \quad \sum_s = 0.99, \\ p_{11} &= 0.985, \quad p_{12} = 0.015, \\ p_{12} &= 0.015, \quad p_{22} = 0.985. \end{aligned} \quad (10.12)$$

We solve the problem by expanding the solution in terms of eleven Legendre components, and the mean square error is estimated and shown in Figure 1, which also indicates how the base ten log of the relative (mean square) errors change with stage numbers. The different curves correspond to different numbers of random walks used together with the running time (seconds) for each stage. According to our theorems, geometric convergence can be reached if the number of random walks W is larger than a threshold number W_0 . We proved the existence of this threshold number, but we did not provide a way to obtain it. From Figure 1, we can see that, for our problem (10.2) with data (10.12), we can choose the number $W_0 \sim 8240$. When $W > W_0$, we see clear geometric convergence.

Table 1 provides a comparison of GWAS and AIS. We note that GWAS has a much higher computational efficiency than AIS because of its speed of execution. Note, too, that even though the variance of GWAS is more than 100 times as large as that of AIS, the efficiency of GWAS is more than 5,000 times that of AIS.

Now we examine a second problem with input data that is more like radiation “shielding” problems:

$$\begin{aligned}
 T &= 5, \\
 \sum_a &= 0.5, & \sum_s &= 0.5, \\
 p_{11} &= 0.5, & p_{12} &= 0.5, \\
 p_{12} &= 0.5, & p_{22} &= 0.5.
 \end{aligned} \tag{10.13}$$

This time we used sixteen Legendre components for expansion of the solution. The results are shown in Figure 2, whose graphs show how the base ten log of the relative (mean square) errors change with respect to stage numbers. Likewise, the legends show the number of random walks together with the running time used for each stage. It also shows clear geometric convergence when the number of random walks $W \approx W_0 \approx 3100$.

Again GWAS is more efficient than AIS in spite of the greater error reduction per adaptive stage with AIS (see Table 2). The computational efficiency of GWAS is approximately 400 times that of AIS for this example.

11 Summary and future work

In this paper we have developed a general class of estimators for solving radiative transport problems. The GWAS estimators have two degrees of freedom, one to use in generating the Monte Carlo biographies, the other to reweight the biography tallies. Under very general conditions, we have proved that the adaptive application of the GWAS estimators produces geometric convergence of the estimates it generates. Numerical results confirm these rates of convergence. They also show that a relatively modest number W_0 of biographies per adaptive stage is sufficient to trigger the geometric convergence and that the rate of convergence increases as W_0 is increased. Of course, the family of transport problems to which our theory applies is enormous, and the examples that we presented are very limited. Nevertheless, we believe that there is a role to be played by GWAS estimation because of the fact that it includes unbiased importance sampling as a special case, yet it offers opportunities to sidestep the computational issues that degrade the performance of “perfect” importance sampling. Interesting (and no doubt difficult!) optimization questions remain to be explored.

Acknowledgments

Funding: The second author gratefully acknowledges partial support from NIH NIBIB Laser Microbeam and Medical Program (LAMMP, P41EB015890) and from NIH R25GM103818.

A Preliminary estimates

The following formula from probability theory can be found in [12] and is used in many of our proofs.

Lemma A.1

For any random variables X and Y ,

$$\begin{aligned} E_X[X] &= E_Y[E_X[X|Y]], \\ V_X[X] &= E_Y[V_X[X|Y]] + V_Y[E_X[X|Y]]. \end{aligned}$$

As well, the following set-theoretic result arises in estimating probabilities:

Proposition A.2

Assume that A and B are two subsets of a probability space and, for two small positive numbers ε_1 and ε_2 , satisfy

$$\mathcal{P}(A) \geq 1 - \varepsilon_1, \quad \mathcal{P}(B) \geq 1 - \varepsilon_2.$$

Then

$$\mathcal{P}(A \cap B) \geq 1 - (\varepsilon_1 + \varepsilon_2).$$

Proof of Lemma 5.1

We will use the following formulas, for any random variables X and Y :

$$E_X[X] = E_Y[E_X[X|Y]], \quad V_X[X] = E_Y[V_X[X|Y]] + V_Y[E_X[X|Y]]. \quad (\text{A.1})$$

For $E[h(P)]$, by first conditioning on η_P and then on ξ_1 , we have

$$\begin{aligned} E[h(P)] &= E_{\eta_P}[E[h(P)|\eta_P]] \\ &= E[h(P)|\eta_P=1]p(P) + E[h(P)|\eta_P=0](1-p(P)) \\ &= \hat{p}(P) + (1-p(P))E_{\xi_1}[E[h(P)|\eta_P=0, \xi_1]] \\ &= \hat{p}(P) + (1-p(P)) \int_{\Gamma} E[h(P)|\eta_P=0, \xi_1=P_1] \frac{p(P, P_1)}{(1-p(P))} dP_1 \\ &= \hat{p}(P) + \int_{\Gamma} \hat{K}(P, P_1) E[h(P_1)] dP_1, \end{aligned}$$

where we have used

$$E[h(P)|\eta_P=0, \xi_1=P_1] = \frac{\hat{K}(P, P_1)}{p(P, P_1)} E[h(P_1)].$$

As for $E[g]$, taking averages of the both sides of equation (3.7) and conditioning on ξ_0 , we obtain

$$E[g] = E \left[h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)} \right] = E_{\xi_0} \left[E \left[h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)} \mid \xi_0 \right] \right] = \int_{\Gamma} E[h(P)\hat{S}(P)]dP = \int_{\Gamma} E[h(P)]\hat{S}(P)dP,$$

which is (5.2).

To calculate the variance of $h(P)$, by conditioning on η_P , we obtain

$$V_h[h(P)] = V_h[h(P) \mid \eta_P=1]p(P) + V_h[h(P) \mid \eta_P=0](1-p(P)) + E_{\eta_P} [E_h[h(P) \mid \eta_P]]^2 - \{E_{\eta_P} [E_h[h(P) \mid \eta_P]]\}^2.$$

The first term of the right-hand side is equal to zero because, under the condition $\eta_P = 1$, $h(P)$ is deterministic, while the last term is equal to $(E_h[h(P)])^2$ owing to (A.1). Again, applying (A.1) on the second term by conditioning $(h(P) \mid \eta_P = 0)$ on ξ_P we obtain

$$\begin{aligned} V_h[h(P)] &= E_{\xi_1} [V_h[h(P) \mid \eta_P=0, \xi_1]](1-p(P)) + V_{\xi_1} [E_h[h(P) \mid \eta_P=0, \xi_1]](1-p(P)) \\ &\quad + [E_h[h(P) \mid \eta_P=1]]^2 p(P) + [E_h[h(P) \mid \eta_P=0]]^2 (1-p(P)) - (E_h[h(P)])^2 \\ &= \int_{\Gamma} V_h[h(P) \mid \eta_P=0, \xi_1=Q]p(P, Q)dQ + E_{\xi_1} [E_h[h(P) \mid \eta_P=0, \xi_1]]^2 (1-p(P)) \\ &\quad - \{E_{\xi_1} [E_h[h(P) \mid \eta_P=0, \xi_1]]\}^2 (1-p(P)) \\ &\quad + \left(\frac{\hat{p}(P)}{p(P)}\right)^2 p(P) + [E_h[h(P) \mid \eta_P=0]]^2 (1-p(P)) - (E_h[h(P)])^2, \end{aligned}$$

where it can be easily verified that $E_h[h(P) \mid \eta_P=1] = \frac{\hat{p}(P)}{p(P)}$. According to (A.1), the third term and the fifth term cancel out. We then have

$$\begin{aligned} V_h[h(P)] &= \int_{\Gamma} V_h[h(Q)] \left(\frac{\hat{K}(P, Q)}{p(P, Q)}\right)^2 p(P, Q)dQ + \int_{\Gamma} (E_h[h(Q)])^2 \left(\frac{\hat{K}(P, Q)}{p(P, Q)}\right)^2 p(P, Q)dQ \\ &\quad + \left(\frac{\hat{p}(P)}{p(P)}\right)^2 p(P) - (E_h[h(P)])^2, \end{aligned}$$

or

$$V_h[h(P)] + (E_h[h(P)])^2 = \int_{\Gamma} (V_h[h(Q)] + (E_h[h(Q)])^2) \frac{\hat{K}^2(P, Q)}{p(P, Q)} dQ + \frac{\hat{p}^2(P)}{p(P)},$$

which is (5.3). To calculate $V[g]$, we have

$$\begin{aligned}
 V[g] &= V \left[h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)} \right] \\
 &= E_{\xi_0} \left[V \left[h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)} \mid \xi_0 \right] \right] + V_{\xi_0} \left[E \left[h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)} \mid \xi_0 \right] \right] \\
 &= \int_{\Gamma} V[h(P)\hat{S}(P)] \frac{dP}{p^1(P)} + E_{\xi_0} \left[E \left[h(\xi_0) \frac{\hat{S}(\xi_0)}{p^1(\xi_0)} \mid \xi_0 \right] \right]^2 - (E[g])^2 \\
 &= \int_{\Gamma} V[h(P)] \frac{(\hat{S}(P))^2}{p^1(P)} dP + \int_{\Gamma} (E[h(P)])^2 \frac{(\hat{S}(P))^2}{p^1(P)} dP - (E[g])^2,
 \end{aligned}$$

which is (5.4). The proof of Lemma 5.1 is completed. □

Proof of Lemma 5.2

From (2.11) and (5.1), we obtain

$$E[h(P)] = \frac{S(P) + \int_{\Gamma} K(P, P_1) \tilde{\Phi}(P_1) E[h(P_1)] dP_1}{\int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ + S(P)}.$$

We then have

$$E[h(P)] \geq \frac{S(P) + \min_{P \in \Gamma} \tilde{\Phi}(P) \int_{\Gamma} K(P, P_1) E[h(P_1)] dP_1}{\max_{P \in \Gamma} \tilde{\Phi}(P) \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ + S(P)}.$$

Using the first inequality of (5.5) and also applying Proposition A.2, we obtain

$$\mathcal{P} \left\{ E[h(P)] \geq \frac{\delta_s}{M_{\tilde{\Phi}} \max_{P \in \Gamma} \int_{\Gamma} K(P, Q) dQ + \delta_s} \right\} > 1 - \varepsilon_1,$$

which means

$$E[h(P)] \geq \frac{\delta_s}{M_{\tilde{\Phi}} \max_{P \in \Gamma} \int_{\Gamma} K(P, Q) dQ + \delta_s}$$

as both sides of the inequality are deterministic. Estimate (5.6) is proved.

Now, we derive an upper bound of $V_h[h(P)]$. Formula (5.3) can be written as

$$\begin{aligned}
 &(V_h[h(P)] + (E_h[h(P)])^2) \cdot \left(\int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ + S(P) \right)^2 \\
 &= \int_{\Gamma} (V_h[h(Q)] + (E_h[h(Q)])^2 (\tilde{\Phi}(Q))^2) \frac{K^2(P, Q)}{p(P, Q)} dQ + \frac{S^2(P)}{p(P)},
 \end{aligned}$$

or

$$\begin{aligned}
 & (V_h[h(P)]+(E_h[h(P)])^2) \cdot (\tilde{\Phi}(P))^2 \\
 & = \int_{\Gamma} (V_h[h(Q)]+(E_h[h(Q)])^2) (\tilde{\Phi}(Q))^2 \frac{K^2(P,Q)}{p(P,Q)} dQ \\
 & + \frac{S^2(P)}{p(P)} + (V_h[h(P)]+(E_h[h(P)])^2) \left((\tilde{\Phi}(P))^2 - \left(\int_{\Gamma} K(P,Q)\tilde{\Phi}(Q)dQ+S(P) \right)^2 \right). \tag{A.2}
 \end{aligned}$$

Since $p(P, Q)$ satisfies (3.11), as an equation of $(V_h[h(P)]+(E_h[h(P)])^2) \cdot (\tilde{\Phi}(P))^2$, (A.2) can be estimated as follows,

$$\begin{aligned}
 & (V_h[h(P)]+(E_h[h(P)])^2) \cdot (\tilde{\Phi}(P))^2 \\
 & \leq \frac{1}{1-\kappa_p} \max_{P \in \Gamma} \left((V_h[h(P)]+(E_h[h(P)])^2) \left| (\tilde{\Phi}(P))^2 - \left(\int_{\Gamma} K(P,Q)\tilde{\Phi}(Q)dQ+S(P) \right)^2 \right| \right) \\
 & \quad + \frac{1}{1-\kappa_p} \max_{P \in \Gamma} \frac{S^2(P)}{p(P)},
 \end{aligned}$$

or

$$\begin{aligned}
 & V_h[h(P)]+(E_h[h(P)])^2 \\
 & \leq \frac{1}{1-\kappa_p} \frac{1}{(\tilde{\Phi}(P))^2} \max_{P \in \Gamma} \left((V_h[h(P)]+(E_h[h(P)])^2) \left| (\tilde{\Phi}(P))^2 - \left(\int_{\Gamma} K(P,Q)\tilde{\Phi}(Q)dQ+S(P) \right)^2 \right| \right) \\
 & \quad + \frac{1}{1-\kappa_p} \frac{1}{(\tilde{\Phi}(P))^2} \max_{P \in \Gamma} \frac{S^2(P)}{p(P)}.
 \end{aligned}$$

Taking the maximum value of both sides, we obtain

$$\begin{aligned}
 & V_h[h(P)]+(E_h[h(P)])^2 \\
 & \leq \frac{1}{1-\kappa_p} \frac{1}{\min_{P \in \Gamma} (\tilde{\Phi}(P))^2} \max_{P \in \Gamma} (V_h[h(P)]+(E_h[h(P)])^2) \max_{P \in \Gamma} \left| (\tilde{\Phi}(P))^2 - \left(\int_{\Gamma} K(P,Q)\tilde{\Phi}(Q)dQ+S(P) \right)^2 \right| \\
 & \quad + \frac{1}{1-\kappa_p} \frac{1}{\min_{P \in \Gamma} (\tilde{\Phi}(P))^2} \max_{P \in \Gamma} \frac{S^2(P)}{p(P)}. \tag{A.3}
 \end{aligned}$$

In (A.3), notice that

$$\begin{aligned} & \left| (\tilde{\Phi}(P))^2 - \left(\int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ + S(P) \right)^2 \right| \\ &= \left| \left(\tilde{\Phi}(P) + \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ + S(P) \right) \left(\tilde{\Phi}(P) - \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ - S(P) \right) \right| \\ &\leq \left(\max_{P \in \Gamma} |\tilde{\Phi}(P)| (1 + \kappa) + M_s \right) \max_{P \in \Gamma} |\tilde{\Phi}(P)| \max_{P \in \Gamma} \frac{|\tilde{\Phi}(P) - \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ - S(P)|}{|\tilde{\Phi}(P)|} \\ &\leq \left(\max_{P \in \Gamma} |\tilde{\Phi}(P)| (1 + \kappa) + M_s \right) \max_{P \in \Gamma} |\tilde{\Phi}(P)| \max_{P \in \Gamma} \frac{|\tilde{\Phi}(P) - \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ - S(P)|}{|\tilde{\Phi}(P)|}, \end{aligned}$$

where κ is defined by (2.4). Therefore, from (A.3), we have

$$\mathcal{P} \left\{ V_h[h(P)] + (E_h[h(P)])^2 \leq \frac{(M_{\tilde{\Phi}}(1 + \kappa) + M_s) M_{\tilde{\Phi}}}{(1 - \kappa_p) \delta_{\tilde{\Phi}}^2} \max_{P \in \Gamma} \frac{|\tilde{\Phi}(P) - \int_{\Gamma} K(P, Q) \tilde{\Phi}(Q) dQ - S(P)|}{|\tilde{\Phi}(P)|} \cdot \max_{P \in \Gamma} (V_h[h(P)] + (E_h[h(P)])^2) \right\}$$

Now using the second inequality of (5.5) and Proposition A.2, we have

$$\mathcal{P} \left\{ V_h[h(P)] + (E_h[h(P)])^2 \leq \alpha \max_{P \in \Gamma} (V_h[h(P)] + (E_h[h(P)])^2) + \frac{M_s^2}{(1 - \kappa_p) \delta_{\tilde{\Phi}}^2 \delta_P} \right\} > 1 - \varepsilon_1 - \varepsilon_2,$$

or

$$\mathcal{P} \left\{ V_h[h(P)] + (E_h[h(P)])^2 \leq \frac{M_s^2}{(1 - \alpha)(1 - \kappa_p) \delta_{\tilde{\Phi}}^2 \delta_P} \right\} > 1 - \varepsilon_1 - \varepsilon_2,$$

which means

$$V_h[h(P)] + (E_h[h(P)])^2 \leq \frac{M_s^2}{(1 - \alpha)(1 - \kappa_p) \delta_{\tilde{\Phi}}^2 \delta_P},$$

because both sides of the inequality are deterministic and $\varepsilon_1 + \varepsilon_2 < 1$. □

The proof of Lemma 5.2 is completed.

Proof of Corollary 5.3

According to Chebyshev's inequality, for any W and $\varepsilon_3 > 0$, we have

$$\mathcal{P} \left\{ \left| E_h[h(P)] - \frac{1}{W} \sum_{w=1}^W h_w(P) \right| < \frac{\sqrt{V_h[h(P)]}}{\sqrt{W} \varepsilon_3} \right\} \geq 1 - \varepsilon_3,$$

which means

$$\mathcal{P} \left\{ \frac{1}{W} \sum_{w=1}^W h_w(P) > E_h[h(P)] - \frac{\sqrt{V_h[h(P)]}}{\sqrt{W}\varepsilon_3} \right\} \geq 1 - \varepsilon_3.$$

Using (5.6), we have

$$\mathcal{P} \left\{ \frac{1}{W} \sum_{w=1}^W h_w(P) > \delta_h - \frac{\sqrt{M_V}}{\sqrt{W}\varepsilon_3} \right\} \geq 1 - \varepsilon_3.$$

Now we just choose

$$W_h = \frac{4M_V}{\varepsilon_3 \delta_h^2}$$

to complete the proof of Corollary 5.3.

Proof of Lemma 5.4

From the definition (3.7) of ζ and equation (3.9) about $\frac{d\mathcal{N}}{d\mathcal{N}}$, we obtain

$$\begin{aligned} E[|X|] &= E \left[\left| \zeta - \left(\int_{\Gamma} \Phi(P) S * (P) dP \right) \frac{d\mathcal{N}}{d\mathcal{N}} \right| \right] \\ &= \sum_{k=1}^{\infty} \int_{\Lambda_k} \left| \zeta - \left(\int_{\Gamma} \Phi(P) S * (P) dP \right) \frac{d\mathcal{N}}{d\mathcal{N}} \right| d\nu \\ &= \sum_{k=1}^{\infty} \int_{\Lambda_k} |S * (P_1) K(P_1, P_2) \cdots K(P_{k-1}, P_k) S(P_k) \\ &\quad - \left(\int_{\Gamma} \Phi(P) S * (P) dP \right) \hat{S}(P_1) \hat{K}(P_1, P_2) \cdots \hat{K}(P_{k-1}, P_k) \hat{p}(P_k)| dP_1 \cdots dP_k \\ &= \sum_{k=1}^{\infty} \int_{\Lambda_k} |S * (P_1) K(P_1, P_2) \cdots K(P_{k-1}, P_k) S(P_k) \\ &\quad - \frac{\tilde{\Phi}(P_1) S * (P_1) \int_{\Gamma} \Phi(Q) S * (Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S * (Q) dQ} \frac{K(P_1, P_2) \tilde{\Phi}(P_2)}{\int_{\Gamma} K(P_1, Q) \tilde{\Phi}(Q) dQ + S(P_1)} \\ &\quad \cdots \frac{K(P_{k-1}, P_k) \tilde{\Phi}(P_k)}{K(P_{k-1}, P_k) \tilde{\Phi}(P_k)} \frac{S(P_k)}{\int_{\Gamma} K(P_k, Q) \tilde{\Phi}(Q) dQ + S(P_k)}| dP_1 \cdots dP_k. \end{aligned}$$

Noticing (4.5) and (4.7) about the definition of $D(P)$ (keep in mind that we ignore the superscript in this section), we can simply the factors after the minus sign in the integral by

$$\frac{\tilde{\Phi}(P_i)}{\int_{\Gamma} K(P_i, Q) \tilde{\Phi}(Q) dQ + S(P_i)} = \frac{1}{\frac{\int_{\Gamma} K(P_i, Q) \tilde{\Phi}(Q) dQ + S(P_i)}{\tilde{\Phi}(P_i)}} = \frac{1}{1 + D(P_i)}. \tag{A.4}$$

Therefore,

$$\begin{aligned}
 E[|X|] &= \sum_{k=1}^{\infty} \int S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \left| 1 - \frac{\tilde{\Phi}(P_1) \int_{\Gamma} \Phi(Q)S^*(Q)dQ}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \frac{\tilde{\Phi}(P_2)}{\int_{\Gamma} K(P_1, Q)\tilde{\Phi}(Q)dQ + S(P_1)} \right. \\
 &\quad \left. \cdots \frac{\tilde{\Phi}(P_k)}{\int_{\Gamma} K(P_{k-1}, Q)\tilde{\Phi}(Q)dQ + S(P_{k-1})} \frac{1}{\int_{\Gamma} K(P_k, Q)\tilde{\Phi}(Q)dQ + S(P_k)} \right| dP_1 \cdots dP_k \\
 &= \frac{1}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \sum_{k=1}^{\infty} \int S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \\
 &\quad \cdot \left| \int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ - \frac{\int_{\Gamma} \Phi(Q)S^*(Q)dQ}{(1+D(P_1)) \cdots (1+D(P_k))} \right| dP_1 \cdots dP_k \\
 &= \frac{1}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \sum_{k=1}^{\infty} \int S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \\
 &\quad \cdot \left| \int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ - \frac{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ}{(1+D(P_1)) \cdots (1+D(P_k))} - \frac{\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q))S^*(Q)dQ}{(1+D(P_1)) \cdots (1+D(P_k))} \right| dP_1 \cdots dP_k,
 \end{aligned}$$

which can be estimated by

$$\begin{aligned}
 E[|X|] &\leq \frac{1}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \sum_{k=1}^{\infty} \int S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \\
 &\cdot \left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ \left| 1 - \frac{1}{(1+D(P_1)) \cdots (1+D(P_k))} \right| + \left| \frac{\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q))S^*(Q)dQ}{(1+D(P_1)) \cdots (1+D(P_k))} \right| \right) dP_1 \cdots dP_k.
 \end{aligned}$$

Obviously, if $|D(P)| < 1$, then

$$\frac{1}{(1+D(P_1)) \cdots (1+D(P_k))} \leq \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k}, \tag{A.5}$$

and, therefore,

$$\begin{aligned}
 E[|X|] &\leq \frac{1}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \sum_{k=1}^{\infty} \int S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \\
 &\cdot \left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ \left| 1 - \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right| + \frac{|\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q))S^*(Q)dQ|}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right) dP_1 \cdots dP_k.
 \end{aligned}$$

(A.6)

On the other hand, using (2.4),

$$\begin{aligned} & \left| \int_{\Lambda_k} S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k)dP_1 \cdots dP_k \right| \\ & \leq \max_{P_k \in \Gamma} S(P_k) \cdot \int_{\Gamma} |S^*(P_1)| dP_1 \left(\max_{P_k \in \Gamma} \int_{\Gamma} K(P_1, P_2)dP_2 \right) \cdots \left(\max_{P_k \in \Gamma} \int_{\Gamma} K(P_{k-1}, P_k)dP_k \right) \\ & \quad \kappa^{k-1} \max_{P \in \Gamma} S(P) \int_{\Gamma} |S^*(P)| dP. \end{aligned} \tag{A.7}$$

Applying (A.7) to (A.6), we obtain

$$\begin{aligned} E[|X|] & \leq \max_{P \in \Gamma} S(P) \int_{\Gamma} |S^*(P)| dP \left(\sum_{k=1}^{\infty} \left(\frac{\kappa^{k-1}}{(1 - \max_{P \in \Gamma} |D(P)|)^k} - \kappa^{k-1} \right) \right. \\ & \quad \left. + \frac{|\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q)) S^*(Q) dQ|}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ} \sum_{k=1}^{\infty} \frac{\kappa^{k-1}}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right) \\ & = \frac{\max_{P \in \Gamma} S(P) \int_{\Gamma} |S^*(P)| dP}{1 - \max_{P \in \Gamma} |D(P)| - \kappa} \left(\frac{\max_{P \in \Gamma} |D(P)|}{1 - \kappa} + \frac{|\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q)) S^*(Q) dQ|}{(\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ)} \right). \end{aligned}$$

Using condition (5.8) and Proposition A.2, we obtain

$$\mathcal{P} \left\{ E[|X|] \leq \frac{\beta \max_{P \in \Gamma} S(P) \int_{\Gamma} |S^*(P)| dP}{(1 - \beta)\kappa} \cdot \left(\frac{\max_{P \in \Gamma} |D(P)|}{1 - \kappa} + \frac{|\int_{\Gamma} (\Phi(Q) - \tilde{\Phi}(Q)) S^*(Q) dQ|}{(\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ)} \right) \right\} > 1 - \varepsilon_1.$$

The proof of Lemma 5.4 is completed. □

Proof of Theorem 5.6

Noticing the definition (3.7) of ζ and the equation (3.9) satisfied by $\frac{d\tilde{v}}{dv}$, we obtain

$$\begin{aligned} E[X^2] & = E \left[\left(\zeta - \left(\int_{\Gamma} \Phi(P) S^*(P) dP \right) \frac{d\tilde{v}}{dv} \right)^2 \right] \\ & = \sum_{k=1}^{\infty} \int_{\Lambda_k} \left(\zeta - \left(\int_{\Gamma} \Phi(P) S^*(P) dP \right) \frac{d\tilde{v}}{dv} \right)^2 dv \\ & = \sum_{k=1}^{\infty} \int_{\Lambda_k} (S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \\ & \quad - \left(\int_{\Gamma} \Phi(P) S^*(P) dP \right) \hat{S}(P_1)\hat{K}(P_1, P_2) \cdots \hat{K}(P_{k-1}, P_k)\hat{p}(P_k))^2 \\ & \quad \cdot \frac{dP_1 \cdots dP_k}{p^1(P_1)p(P_1, P_2) \cdots p(P_{k-1}, P_k)p(P_k)}. \end{aligned}$$

Using (2.11), we obtain

$$\begin{aligned}
 E[X^2] &= \sum_{k=1}^{\infty} \int_{\Lambda_k} (S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k) \\
 &\quad \frac{\tilde{\Phi}(P_1)S^*(P_1) \int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \frac{K(P_1, P_2)\tilde{\Phi}(P_2)}{\int_{\Gamma} K(P_1, Q)\tilde{\Phi}(Q)dQ + S(P_1)} \\
 &\quad \cdots \frac{K(P_{k-1}, P_k)\tilde{\Phi}(P_k)}{\int_{\Gamma} K(P_{k-1}, Q)\tilde{\Phi}(Q)dQ + S(P_{k-1})} \frac{S(P_k)}{\int_{\Gamma} K(P_k, Q)\tilde{\Phi}(Q)dQ + S(P_k)})^2 \\
 &= \frac{1}{\left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ\right)^2} \sum_{k=1}^{\infty} \int_{\Lambda_k} (S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k))^2 \\
 &\quad \cdot \left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ - \int_{\Gamma} \Phi(Q)S^*(Q)dQ \frac{\tilde{\Phi}(P_1)}{\int_{\Gamma} K(P_1, Q)\tilde{\Phi}(Q)dQ + S(P_1)} \right. \\
 &\quad \left. \cdots \frac{\tilde{\Phi}(P_{k-1})}{\int_{\Gamma} K(P_{k-1}, Q)\tilde{\Phi}(Q)dQ + S(P_{k-1})} \frac{\tilde{\Phi}(P_k)}{\int_{\Gamma} K(P_k, Q)\tilde{\Phi}(Q)dQ + S(P_k)} \right)^2 \\
 &\quad \cdot \frac{dP_1 \cdots dP_k}{p^1(P_1)p(P_1, P_2) \cdots p(P_{k-1}, P_k)p(P_k)}.
 \end{aligned}$$

Noticing the notations $D(P)$ (ignore the superscript) defined in (4.7) or doing exactly what we have done in Lemma 5.4, equation (A.4), we obtain

$$\begin{aligned}
 E[X^2] &= \frac{1}{\left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ\right)^2} \sum_{k=1}^{\infty} \int_{\Lambda_k} \frac{(S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k))^2}{p^1(P_1)p(P_1, P_2) \cdots p(P_{k-1}, P_k)p(P_k)} \\
 &\quad \cdot \left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ - \int_{\Gamma} \Phi(Q)S^*(Q)dQ \frac{1}{1+D(P_1)} \cdots \frac{1}{1+D(P_{k-1})} \frac{1}{1+D(P_k)} \right)^2 dP_1 \cdots dP_k \\
 &= \frac{1}{\left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ\right)^2} \sum_{k=1}^{\infty} \int_{\Lambda_k} \frac{(S^*(P_1)K(P_1, P_2) \cdots K(P_{k-1}, P_k)S(P_k))^2}{p^1(P_1)p(P_1, P_2) \cdots p(P_{k-1}, P_k)p(P_k)} \\
 &\quad \cdot \left(\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ \left(1 - \frac{1}{1+D(P_1)} \cdots \frac{1}{1+D(P_{k-1})} \frac{1}{1+D(P_k)}\right) \right. \\
 &\quad \left. + \frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q))S^*(Q)dQ}{(1+D(P_1)) \cdots (1+D(P_k))} \right)^2 dP_1 \cdots dP_k.
 \end{aligned}$$

Using (A.5) in the proof of Lemma 5.4, we obtain

$$\begin{aligned}
 E[X^2] &\leq \sum_{k=1}^{\infty} \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) \left(\max_{P \in \Gamma} \int_{\Gamma} \frac{(K(P, Q))^2}{p(P, Q)} dQ \right)^{k-1} \max_{P \in \Gamma} \frac{(S(P))^2}{p(P)} \\
 &\quad \cdot \left(1 - \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} + \frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q))S^*(Q)dQ}{\int_{\Gamma} \tilde{\Phi}(Q)S^*(Q)dQ} \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right)^2. \tag{A.8}
 \end{aligned}$$

Since, for any a and b , $(a + b)^2 \leq 2(a^2 + b^2)$, we obtain

$$\begin{aligned} & \left(1 - \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} + \frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q)) S^*(Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ} \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right)^2 \\ & \leq 2 \left(\frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} - 1 \right)^2 + 2 \left(\frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q)) S^*(Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ} \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right)^2. \end{aligned}$$

(A.9)

Noticing the definition of κ_p (3.11), and using (A.9), from (A.8) we obtain

$$\begin{aligned} E[X^2] & \leq 2 \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) \left(\max_{P \in \Gamma} \frac{(S(P))^2}{p(P)} \right) \sum_{k=0}^{\infty} \kappa_p^{k-1} \left(\frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} - 1 \right)^2 \\ & \quad + 2 \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) \left(\max_{P \in \Gamma} \frac{(S(P))^2}{p(P)} \right) \\ & \quad \cdot \sum_{k=0}^{\infty} \kappa_p^{k-1} \left(\frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q)) S^*(Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ} \frac{1}{(1 - \max_{P \in \Gamma} |D(P)|)^k} \right)^2 \\ & \leq \frac{2 \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) (\max_{P \in \Gamma} \frac{(S^*(P))^2}{p(P)}) (1 - \max_{P \in \Gamma} |D(P)| + \kappa_p) \cdot \max_{P \in \Gamma} |D(P)|^2}{((1 - \max_{P \in \Gamma} |D(P)|)^2 - \kappa_p)(1 - \kappa_p)(1 - \max_{P \in \Gamma} |D(P)| - \kappa_p)} \\ & \quad + \frac{2 \left(\int_{\Gamma} \frac{(S^*(P_1))^2}{p^1(P_1)} dP_1 \right) (\max_{P \in \Gamma} \frac{(S^*(P))^2}{p(P)})}{(1 - \max_{P \in \Gamma} |D(P)|)^2 - \kappa_p} \left(\frac{\int_{\Gamma} (\tilde{\Phi}(Q) - \Phi(Q)) S^*(Q) dQ}{\int_{\Gamma} \tilde{\Phi}(Q) S^*(Q) dQ} \right)^2. \end{aligned}$$

Using (5.10) together with Proposition A.2, we obtain (5.11). □

The proof is completed.

B Estimation of the bias

Proof of Theorem 6.1

Recalling the definition of $\tau_W(P)$, (4.4), we obtain

$$Z(P) \equiv |E[\tau_W(P)] - \Phi(P)| = \left| E \left[\frac{\sum_{w=1}^W \omega_w(P)}{\sum_{w=1}^W h_w(P)} \right] - \Phi(P) \right| = \left| E \left[\frac{\frac{1}{W} \sum_{w=1}^W (\omega_w(P) - \Phi(P) h_w(P))}{\frac{1}{W} \sum_{w=1}^W h_w(P)} \right] \right|.$$

(B.1)

Note that $h_w(P)$ and $\omega_w(P)$ are the w -th samples of random variables $h(P) \left(= \frac{d\tilde{v}_p}{dv_p} \right)$ and $\omega(P)$, respectively.

Now, from Corollary 5.3, for $\varepsilon > 0$, there must be a positive number δ_h (defined by (5.7)) such that

$$\mathcal{P} \left\{ \frac{1}{W} \sum_{i=1}^W h_w(P) \geq \frac{1}{2} \delta_h \right\} \geq 1 - \varepsilon. \tag{B.2}$$

Therefore, from (B.1),

$$\mathcal{P} \left\{ Z(P) \leq \frac{E \left[\sum_{w=1}^W |\omega_w(P) - \Phi(P)h_w(P)| \right]}{W\delta_h} \right\} \geq 1 - \varepsilon,$$

or

$$\mathcal{P} \left\{ Z(P) \leq \sum_{w=1}^W \frac{E [|\omega_w(P) - \Phi(P)h_w(P)|]}{W\delta_h} \right\} \geq 1 - \varepsilon. \tag{B.3}$$

According to (4.9), the definition of $X(P)$ (ignore the superscript as indicated before), R (B.3) is actually

$$\mathcal{P} \left\{ Z(P) \leq \sum_{w=1}^W \frac{E [|X_w(P)|]}{W\delta_h} \right\} \geq 1 - \varepsilon,$$

where the subscript w indicates that $X_w(P)$ is the w -th sample of $X(P)$. Now, appealing to Lemma 5.4, we obtain the second inequality of (6.1).

As for the random variable Z , we can prove it similarly. Notice

$$Z \equiv |E[\tau_W] - I| = \left| E \left[\frac{\sum_{i=1}^W \zeta_w}{\sum_{i=1}^W g_w} \right] - I \right| = \left| E \left[\frac{\frac{1}{W} \sum_{i=1}^W (\zeta_w - (\int_{\Gamma} \Phi(Q)S^*(Q)dQ)g_w)}{\frac{1}{W} \sum_{i=1}^W g_w} \right] \right|.$$

Then using (B.2), we obtain the first inequality of (6.1).

This completes the proof of Theorem 6.1.

C Estimation of second moments

Proof of Theorem 7.1

We first estimate $V(P)$. Note that this is not the variance of the random variable τ_W because, in general,

$$E[\tau_W(P)] \neq \Phi(P),$$

i.e., $\tau_W(P)$ is not unbiased. However, the estimation of $V(P)$ will help us prove the geometric convergence of the solution through the random variable $\tau_W(P)$.

We have

$$V(P) = E[\tau_W(P) - \Phi(P)]^2 = E \left[\frac{\sum_{w=1}^W \omega_w(P)}{\sum_{w=1}^W h_w(P)} - \Phi(P) \right]^2 = E \left[\frac{\frac{1}{W} \sum_{w=1}^W (\omega_w(P) - \Phi(P) \cdot h_w(P))}{\frac{1}{W} \sum_{w=1}^W h_w(P)} \right]^2. \tag{C.1}$$

Now, from Corollary 5.3, for $\epsilon_1 > 0$, there must be a positive number δ_h and an integer $W_h > 0$ such that when $W > W_h$,

$$\mathcal{P} \left\{ \frac{1}{W} \sum_{w=1}^W h_w(P) \geq \frac{1}{2} \delta_h \right\} \geq 1 - \epsilon_1.$$

Therefore, from (C.1),

$$\mathcal{P} \left\{ V(P) \leq \frac{4E \left[\sum_{w=1}^W (\omega_w(P) - \Phi(P) \cdot h_w(P)) \right]^2}{W^2 \delta_h^2} \right\} \geq 1 - \epsilon_1,$$

or

$$\mathcal{P} \left\{ V(P) \leq \frac{4}{W^2 \delta_h^2} \sum_{w=1}^W E[(\omega_w(P) - \Phi(P) \cdot h_w(P))^2] + \frac{8}{W^2 \delta_h^2} \sum_{i < j} E[(\omega_i(P) - \Phi(P) \cdot h_i(P))(\omega_j(P) - \Phi(P) \cdot h_j(P))] \right\} \geq 1 - \epsilon_1. \tag{C.2}$$

Because of the independence of the random walks for different i and j , the second sum in the braces of (C.2) vanishes. We then have

$$\mathcal{P} \left\{ V(P) \leq \frac{4}{W^2 \delta_h^2} \sum_{w=1}^W E \left[(\omega_w(P) - \Phi(P) \cdot h_w(P))^2 \right] \right\} \geq 1 - \varepsilon_1,$$

or

$$\mathcal{P} \left\{ V(P) \leq \frac{4}{W \delta_h^2} E \left[(\omega(P) - \Phi(P) \cdot h(P))^2 \right] \right\} \geq 1 - \varepsilon_1. \tag{C.3}$$

According to condition (7.3), we can apply Lemma 5.6, specifically inequality (5.12):

$$\mathcal{P} \left\{ \max_{P \in \Gamma} E \left[X^2(P) \right] \leq \tilde{c}_1 \max_{P \in \Gamma} |D(P)|^2 + \tilde{c}_2 \max_{P \in \Gamma} \left| \frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)} \right|^2 \right\} > 1 - \varepsilon_2, \tag{C.4}$$

Now combining (C.3) and (C.4), and using Proposition A.2, we obtain

$$\mathcal{P} \left\{ V(P) \leq \frac{4\tilde{c}_1}{W \delta_h^2} \max_{P \in \Gamma} |D(P)|^2 + \frac{4\tilde{c}_2}{W \delta_h^2} \max_{P \in \Gamma} \left| \frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)} \right|^2 \right\} \geq 1 - \varepsilon_1 - \varepsilon_2, \tag{C.5}$$

which is the second inequality of (7.4) (after redefining the constants \tilde{c}_1 and \tilde{c}_2).

As for V we note

$$V = E[\tau_W - I]^2 = E \left[\frac{\sum_{i=1}^W \zeta_w}{\sum_{w=1}^W g_w} - I \right]^2 = E \left[\frac{\frac{1}{W} \sum_{w=1}^W (\zeta_w - (\int_{\Gamma} \Phi(Q) S^*(Q) dQ) \cdot g_w)}{\sum_{w=1}^W g_w} \right]^2.$$

Similarly, we can obtain the first inequality of (7.4).

In order to derive (7.5), we notice that

$$\begin{aligned} \max_{P \in \Gamma} |D(P)| R = \max_{P \in \Gamma} & \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}(P))}{\tilde{\Phi}(P)} \right| \\ & \leq \frac{(1+\kappa)(\max_{P \in \Gamma} \tilde{\Phi}(P) + \max_{P \in \Gamma} \Phi(P))}{\max_{P \in \Gamma} \tilde{\Phi}(P)} \\ & \leq \frac{(1+\kappa)(\max_{P \in \Gamma} \tilde{\Phi}(P) + M_{\Phi})}{\min_{P \in \Gamma} \tilde{\Phi}(P)} \end{aligned} \tag{C.6}$$

and

$$\max_{P \in \Gamma} \left| \frac{\tilde{\Phi}(P) - \Phi(P)}{\tilde{\Phi}(P)} \right| \leq \frac{(\max_{P \in \Gamma} \tilde{\Phi}(P) + M_\Phi)}{\min_{P \in \Gamma} \tilde{\Phi}(P)}. \tag{C.7}$$

Combining (C.5), (C.6) and (C.7), we obtain

$$\mathcal{P} \left\{ V(P) \leq \frac{4\tilde{c}_1}{W\delta_h^2} \left(\frac{(1+\kappa)(\max_{P \in \Gamma} \tilde{\Phi}(P) + M_\Phi)}{\min_{P \in \Gamma} \tilde{\Phi}(P)} \right)^2 + \frac{4\tilde{c}_2}{W\delta_h^2} \left(\frac{(\max_{P \in \Gamma} \tilde{\Phi}(P) + M_\Phi)}{\min_{P \in \Gamma} \tilde{\Phi}(P)} \right)^2 \right\} \geq 1 - \varepsilon_1 - \varepsilon_2.$$

Now using (7.2) and Proposition A.2, we obtain

$$\mathcal{P} \left\{ V(P) \leq \frac{C_1}{W} \right\} \geq 1 - 2\varepsilon_1 - \varepsilon_2,$$

where

$$C_1 = \frac{4(\tilde{c}_1(1+\kappa)^2 + \tilde{c}_2)(M_{\tilde{\Phi}} + M_\Phi)^2}{\delta_h^2 \delta_{\tilde{\Phi}}^2}.$$

Noticing the second inequality of (7.1), since $V(P)$ is deterministic, we obtain

$$V(P) \leq \frac{C_1}{W},$$

which is the first inequality of (7.5). The second inequality can be obtained similarly.

This completes the proof.

D General geometric convergence

Proof of Theorem 8.1

We will prove by induction that, for any $m \geq 1$,

$$\left\{ \begin{array}{l} \mathcal{P} \left\{ \delta_{\tilde{\Phi}} \leq \tilde{\Phi}^{m-1}(P) \leq M_{\tilde{\Phi}} \right\} \geq 1 - \frac{\varepsilon}{5}, \\ \mathcal{P} \left\{ \max_{P \in \Gamma} |D^{m-1}(P)| \leq 1 - \sqrt{\frac{\kappa_P}{\gamma}} \right\} \geq 1 - \frac{2\varepsilon}{5}, \\ \mathcal{P} \left\{ \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^m(P)| < \lambda \max_{P \in \Gamma} |\tilde{\Phi}^{m-1}(P) - \Phi(P)| \right\} \geq 1 - \varepsilon. \end{array} \right. \tag{D.1}$$

We have added two additional inequalities to the list (first two inequalities of (D.1)). The reason for doing so is that, to prove the third inequality, we need the first two inequalities.

For $m = 1$, the first two inequalities are trivial, as conditions (8.1) and (8.2) imply them. To prove the third one, we appeal to Theorem 7.1, the second inequality of (7.4),

$$\mathcal{P} \left\{ V^1(P) \leq \frac{\tilde{c}_1}{W} \max_{P \in \Gamma} |D^0(P)|^2 + \frac{\tilde{c}_2}{W} \max_{P \in \Gamma} \left| \frac{\tilde{\Phi}^0(P) - \Phi(P)}{\tilde{\Phi}^0(P)} \right|^2 \right\} \geq 1 - \frac{3\varepsilon}{5}, \tag{D.2}$$

where we have taken $\varepsilon_1 = \frac{\varepsilon}{5}$ and $\varepsilon_2 = \frac{2\varepsilon}{5}$.

We need to estimate $\max_{P \in \Gamma} |D^0(P)|^2$ and $\max_{P \in \Gamma} \left| \frac{\tilde{\Phi}^0(P) - \Phi(P)}{\tilde{\Phi}^0(P)} \right|^2$ by $\max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2$. We have

$$\begin{aligned} \max_{P \in \Gamma} |D^0(P)|^2 &= \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^0(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^0(P))}{\tilde{\Phi}^0(P)} \right|^2 \\ &\leq \frac{(1+\kappa)^2}{(\min_{P \in \Gamma} \tilde{\Phi}^0(P))^2} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2 \end{aligned} \tag{D.3}$$

and

$$\max_{P \in \Gamma} \left| \frac{\tilde{\Phi}^0(P) - \Phi(P)}{\tilde{\Phi}^0(P)} \right|^2 \leq \frac{1}{(\min_{P \in \Gamma} \tilde{\Phi}^0(P))^2} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2. \tag{D.4}$$

From (D.2), (D.3) and (D.4), we obtain

$$\mathcal{P} \left\{ V^1(P) \leq \frac{\tilde{c}_1(1+\kappa)^2 + \tilde{c}_2}{W (\min_{P \in \Gamma} \tilde{\Phi}^0(P))^2} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2 \right\} \geq 1 - \frac{3\varepsilon}{5}.$$

Applying the first inequality (D.1) and Proposition A.2, we obtain

$$\mathcal{P} \left\{ V^1(P) \leq \frac{C_1}{W} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2 \right\} \geq 1 - \frac{4\varepsilon}{5}, \tag{D.5}$$

where $C_1 = \frac{\tilde{c}_1(1+\kappa)^2 + \tilde{c}_2}{\delta_{\tilde{\Phi}}^2}$. Next, according to Chebyshev's inequality,

$$\mathcal{P} \left\{ |\Phi(P) - \tau_w^1(P)| < \frac{\sqrt{V^1(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5},$$

or after we replace $\tau_w^1(P)$ by $\tilde{\Phi}^1(P)$,

$$\mathcal{P} \left\{ |\Phi(P) - \tilde{\Phi}^1(P)| < \frac{\sqrt{V^1(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5}. \tag{D.6}$$

Combining (D.5) with (D.6) and noticing Proposition A.2, we obtain

$$\mathcal{P} \left\{ |\Phi(P) - \tilde{\Phi}^1(P)| < \frac{\sqrt{C_1}}{\sqrt{\frac{W\varepsilon}{5}}} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)| \right\} \geq 1 - \varepsilon.$$

Thus (D.1) is proved for $m = 1$ once we pick a W_4 such that when $W > W_4$,

$$\frac{\sqrt{C_1}}{\sqrt{\frac{W\varepsilon}{5}}} \leq \lambda. \tag{D.7}$$

We now prove (D.1) for the general case. Again, according to Chebyshev's inequality,

$$\mathcal{P} \left\{ |\Phi(P) - \tau_w^{m-1}(P)| < \frac{\sqrt{V^{m-1}(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5},$$

or after replacing $\tau_w^{m-1}(P)$ by $\tilde{\Phi}^{m-1}(P)$,

$$\mathcal{P} \left\{ |\Phi(P) - \tilde{\Phi}^{m-1}(P)| < \frac{\sqrt{V^{m-1}(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5}, \tag{D.8}$$

or

$$\mathcal{P} \left\{ \Phi(P) - \frac{\sqrt{V^{m-1}(P)}}{\sqrt{\frac{\varepsilon}{5}}} < \tilde{\Phi}^{m-1}(P) < \Phi(P) + \frac{\sqrt{V^{m-1}(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5}. \tag{D.9}$$

According to Theorem 7.1,

$$V^{m-1}(P) \leq \frac{C_2}{W}, \quad (\text{D.10})$$

where C_2 does not depend on any specific stages. Substituting (D.10) into (D.9) produces

$$\mathcal{P} \left\{ \min_{P \in \Gamma} \Phi(P) - \frac{\sqrt{C_2}}{\sqrt{\frac{W\varepsilon}{5}}} < \tilde{\Phi}^{m-1}(P) < \max_{P \in \Gamma} \Phi(P) + \frac{\sqrt{C_2}}{\sqrt{\frac{W\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5},$$

or by (2.6),

$$\mathcal{P} \left\{ \delta_\Phi - \frac{\sqrt{C_2}}{\sqrt{\frac{W\varepsilon}{5}}} < \tilde{\Phi}^{m-1}(P) < M_\Phi + \frac{\sqrt{C_2}}{\sqrt{\frac{W\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5}.$$

Noticing (4.2), we can find a W_2 such that

$$\begin{aligned} \delta_\Phi - \frac{\sqrt{C_2}}{\sqrt{\frac{W_2\varepsilon}{5}}} &\geq \delta_{\tilde{\Phi}}, \\ M_\Phi + \frac{\sqrt{C_2}}{\sqrt{\frac{W_2\varepsilon}{5}}} &\leq M_{\tilde{\Phi}}. \end{aligned}$$

We then have

$$\mathcal{P} \left\{ \delta_{\tilde{\Phi}} < \tilde{\Phi}^{m-1}(P) < M_{\tilde{\Phi}} \right\} \geq 1 - \frac{\varepsilon}{5}. \quad (\text{D.11})$$

In order to prove the second inequality of (D.1), we notice

$$\begin{aligned} \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \\ \leq \frac{1+\kappa}{\min_{P \in \Gamma} \tilde{\Phi}^{m-1}(P)} \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^{m-1}(P)|. \end{aligned} \quad (\text{D.12})$$

Combining (D.11) with (D.12) and applying Proposition A.2 (about the lower bound),

$$\begin{aligned} \mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \right. \\ \left. \leq \frac{1+\kappa}{\delta_{\tilde{\Phi}}} \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^{m-1}(P)| \right\} > 1 - \frac{\varepsilon}{5}. \end{aligned} \quad (\text{D.13})$$

Combining (D.8) with (D.13) and applying Proposition A.2, we obtain

$$\mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \leq \frac{1+\kappa}{\delta_{\tilde{\Phi}}} \frac{\sqrt{V^{m-1}(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} > 1 - \frac{2\varepsilon}{5}.$$

Applying (D.10), we obtain

$$\mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \leq \frac{1+\kappa}{\delta_{\tilde{\Phi}}} \frac{\sqrt{C_2}}{\sqrt{\frac{W\varepsilon}{5}}} \right\} > 1 - \frac{2\varepsilon}{5}.$$

(D.14)

We can then pick a W_3 such that

$$\frac{1+\kappa}{\delta_{\tilde{\Phi}}} \frac{\sqrt{C_2}}{\sqrt{W_3 \frac{\varepsilon}{5}}} \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}}. \tag{D.15}$$

Combining (D.14) with (D.15), when $W \geq W_3$,

$$\mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}} \right\} \geq 1 - \frac{2\varepsilon}{5},$$

(D.16)

which is the second inequality of (D.1).

The third inequality of (D.1) can be easily proved. All we have to do is to go over the steps from (D.2) through (D.7) with the superscript 1 replaced by m and 0 replaced by $m - 1$. That is, we can pick the same W_4 as determined by (D.7) such that when $W \geq W_4$,

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^m(P)| < \lambda \max_{P \in \Gamma} |\tilde{\Phi}^{m-1}(P) - \Phi(P)| \right\} \geq 1 - \varepsilon. \tag{D.17}$$

Thus, when $W \geq W_0 = \max\{W_1, W_2, W_3, W_4\}$, (D.11), (D.16) and (D.17) all hold.

The proof of Theorem 8.1 is completed. □

E Geometric convergence for expansion in basis functions

Proof of Theorem 9.1

Using the expressions of the true solution $\Phi(P)$ by (9.1) and the m -th stage approximation $\tilde{\Phi}^m(P)$ by (9.3), we have

$$\begin{aligned} |\Phi(P) - \tilde{\Phi}^m(P)| &= \left| \sum_{i=0}^{\infty} a_i f_i(P) - \sum_{i=0}^N \tilde{a}_i^m f_i(P) \right| \\ &= \left| \sum_{i=0}^N (a_i - \tilde{a}_i^m) f_i(P) + \sum_{i=N+1}^{\infty} a_i f_i(P) \right| \\ &\leq \sum_{i=0}^N |a_i - \tilde{a}_i^m| |f_i(P)| + r_N \\ &\leq M_f \sum_{i=0}^N |a_i - \tilde{a}_i^m| + r_N. \end{aligned} \tag{E.1}$$

Therefore, estimating $|\Phi(P) - \tilde{\Phi}^m(P)|$ becomes estimating $|a_i - \tilde{a}_i^m|$.

The proof is similar to Theorem 8.1. We will prove by induction that, for any $m \geq 1$,

$$\left\{ \begin{array}{l} \mathcal{P} \left\{ \delta_{\tilde{\Phi}} \leq \tilde{\Phi}^{m-1}(P) \leq M_{\tilde{\Phi}} \right\} \geq 1 - \frac{\varepsilon}{5}, \\ \mathcal{P} \left\{ \max_{P \in \Gamma} |D^{m-1}(P)| \leq 1 - \sqrt{\frac{\kappa_P}{\gamma}} \right\} \geq 1 - \frac{2\varepsilon}{5}, \\ \mathcal{P} \left\{ \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^m(P)| < \lambda \max_{P \in \Gamma} |\tilde{\Phi}^{m-1}(P) - \Phi(P)| + r_N \right\} \geq 1 - \varepsilon. \end{array} \right. \tag{E.2}$$

We have added two additional inequalities to the list (first two inequalities of (E.2)). The reason for doing so is that, to prove the third inequality, we need the first two inequalities.

For $m = 1$, the first two inequalities are trivial, as conditions (9.4) and (9.5) imply them. To prove the third one, we appeal to Theorem 7.1, the first inequality of (7.4),

$$\mathcal{P} \left\{ V^1 \leq \frac{c_1}{W} \max_{P \in \Gamma} |D^0(P)|^2 + \frac{c_2}{W} \left(\frac{\int_{\Gamma} (\tilde{\Phi}^0(Q) - \Phi(Q)) S * (Q) dQ}{\int_{\Gamma} \tilde{\Phi}^0(Q) S * (Q) dQ} \right)^2 \right\} \geq 1 - \frac{3\varepsilon}{5}, \tag{E.3}$$

where we have taken $\varepsilon_1 = \frac{\varepsilon}{5}$ and $\varepsilon_2 = \frac{2\varepsilon}{5}$.

We need to estimate $\max_{P \in \Gamma} |D^0(P)|^2$ and $\max_{P \in \Gamma} \left| \frac{\tilde{\Phi}^0(P) - \Phi(P)}{\tilde{\Phi}^0(P)} \right|^2$ by $\max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2$. We have

$$\begin{aligned} \max_{P \in \Gamma} |D^0(P)|^2 &= \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q) (\tilde{\Phi}^0(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^0(P))}{\tilde{\Phi}^0(P)} \right|^2 \\ &\leq \frac{(1+\kappa)^2}{(\min_{P \in \Gamma} \tilde{\Phi}^0(P))^2} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2 \end{aligned} \tag{E.4}$$

and

$$\max_{P \in \Gamma} \left| \frac{\tilde{\Phi}^0(P) - \Phi(P)}{\tilde{\Phi}^0(P)} \right|^2 \leq \frac{1}{(\min_{P \in \Gamma} \tilde{\Phi}^0(P))^2} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2. \tag{E.5}$$

From (E.3), (E.4) and (E.5), we obtain

$$\mathcal{P} \left\{ V^1 \leq \frac{\tilde{c}_1(1+\kappa)^2 + \tilde{c}_2}{W(\min_{P \in \Gamma} \tilde{\Phi}^0(P))^2} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2 \right\} \geq 1 - \frac{3\varepsilon}{5}. \tag{E.6}$$

Applying the first inequality (E.2) and Proposition A.2, we obtain

$$\mathcal{P} \left\{ V^1 \leq \frac{C_1}{W} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)|^2 \right\} \geq 1 - \frac{4\varepsilon}{5}, \tag{E.7}$$

where $C_1 = \frac{\tilde{c}_1(1+\kappa)^2 + \tilde{c}_2}{\delta_{\tilde{\Phi}}^2}$. Next, according to Chebyshev's inequality,

$$\mathcal{P} \left\{ |a_i^1 - \tau_w^1(P)| < \frac{\sqrt{V^1(P)}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5(N+1)}.$$

or after we replace τ_w^1 by \tilde{a}_i^1 ,

$$\mathcal{P} \left\{ |a_i^1 - \tilde{a}_i^1| < \frac{\sqrt{V^1}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5(N+1)}. \tag{E.8}$$

Combining (E.1) with (E.7) and (E.8) and noticing Proposition A.2, we obtain

$$\mathcal{P} \left\{ |\Phi(P) - \tilde{\Phi}^1(P)| < \frac{M_f \sqrt{C_1(N+1)}}{\sqrt{\frac{W\varepsilon}{5}}} \max_{P \in \Gamma} |\tilde{\Phi}^0(P) - \Phi(P)| + r_N \right\} \geq 1 - \varepsilon.$$

Thus (E.2) is proved for $m = 1$ once we pick a W_4 such that when $W = W_4$,

$$\frac{M_f \sqrt{C_1(N+1)}}{\sqrt{\frac{W\varepsilon}{5}}} \leq \lambda. \tag{E.9}$$

We now prove (E.2) for the general case. Again, according to Chebyshev's inequality,

$$\mathcal{P} \left\{ |a_i^{m-1} - \tau_w^{m-1}| < \frac{\sqrt{V^{m-1}}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5(N+1)},$$

or after replacing τ_w^{m-1} by \tilde{a}_i^{m-1} ,

$$\mathcal{P} \left\{ |a_i^{m-1} - \tilde{a}_i^{m-1}| < \frac{\sqrt{V^{m-1}}}{\sqrt{\frac{\varepsilon}{5}}} \right\} \geq 1 - \frac{\varepsilon}{5(N+1)},$$

which leads to

$$\mathcal{P} \left\{ |\Phi(P) - \tilde{\Phi}^{m-1}(P)| < \frac{M_f \sqrt{V^{m-1}(N+1)}}{\sqrt{\frac{\varepsilon}{5}}} + r_N \right\} \geq 1 - \frac{\varepsilon}{5}, \tag{E.10}$$

or using Proposition A.2,

$$\mathcal{P} \left\{ \Phi(P) - \frac{M_f \sqrt{V^{m-1}(N+1)}}{\sqrt{\frac{\varepsilon}{5}}} - r_N < \tilde{\Phi}^{m-1}(P) < \Phi(P) + \frac{M_f \sqrt{V^{m-1}(N+1)}}{\sqrt{\frac{\varepsilon}{5}}} + r_N \right\} \geq 1 - \frac{\varepsilon}{5}.$$

$$\tag{E.11}$$

According to Theorem 7.1,

$$V^{m-1} \leq \frac{C_2}{W}, \tag{E.12}$$

where C_2 does not depend on any specific stages. Substituting (E.12) into (E.11) produces

$$\mathcal{P} \left\{ \min_{P \in \Gamma} \Phi(P) - \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{\frac{W\varepsilon}{5}}} - r_N < \tilde{\Phi}^{m-1}(P) < \max_{P \in \Gamma} \Phi(P) + \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{\frac{W\varepsilon}{5}}} + r_N \right\} \geq 1 - \frac{\varepsilon}{5},$$

or by (2.6),

$$\mathcal{P} \left\{ \delta_\Phi - \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{\frac{W_\varepsilon}{5}}} - r_N < \tilde{\Phi}^{m-1}(P) < M_\Phi + \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{\frac{W_\varepsilon}{5}}} + r_N \right\} \geq 1 - \frac{\varepsilon}{5}.$$

Noticing (4.2), we can find a W_2 such that

$$\begin{aligned} \delta_\Phi - \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{W_2 \frac{\varepsilon}{5}}} - r_N &\geq \delta_{\tilde{\Phi}}, \\ M_\Phi + \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{W_2 \frac{\varepsilon}{5}}} + r_N &\geq \delta_{\tilde{\Phi}}. \end{aligned}$$

We then have

$$\mathcal{P} \left\{ \delta_{\tilde{\Phi}} < \tilde{\Phi}^{m-1}(P) < M_{\tilde{\Phi}} \right\} \geq 1 - \frac{\varepsilon}{5}. \quad (\text{E.13})$$

In order to prove the second inequality of (E.2), we notice

$$\begin{aligned} \max_{P \in \Gamma} \left| \frac{\int_\Gamma K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \\ \leq \frac{1+\kappa}{\min_{P \in \Gamma} \tilde{\Phi}^{m-1}(P)} \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^{m-1}(P)|. \end{aligned} \quad (\text{E.14})$$

Combining (E.13) with (E.14) and applying Proposition A.2 (about the lower bound),

$$\begin{aligned} \mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_\Gamma K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \right. \\ \left. \leq \frac{1+\kappa}{\delta_{\tilde{\Phi}}} \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^{m-1}(P)| \right\} > 1 - \frac{\varepsilon}{5}. \end{aligned} \quad (\text{E.15})$$

Combining (E.10) with (E.15) and applying Proposition A.2, we obtain

$$\begin{aligned} \mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_\Gamma K(P, Q) (\tilde{\Phi}^{m-1}(Q) - \Phi(Q)) dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \right. \\ \left. \leq \frac{1+\kappa}{\delta_{\tilde{\Phi}}} \frac{M_f \sqrt{V^{m-1}(N+1)}}{\sqrt{\frac{\varepsilon}{5}}} + \frac{1+\kappa}{\delta_{\tilde{\Phi}}} r_N \right\} > 1 - \frac{2\varepsilon}{5}. \end{aligned}$$

Applying (E.12), we obtain

$$\begin{aligned} & \left| \frac{\int_{\Gamma} K(P, Q)(\tilde{\Phi}^{m-1}(Q) - \Phi(Q))dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \\ & \leq \frac{1+\kappa}{\delta_{\tilde{\Phi}}} \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{W_3 \frac{\varepsilon}{5}}} + \frac{1+\kappa}{\delta_{\tilde{\Phi}}} r_N \left\{ \max_{P \in \Gamma} \right\} > 1 - \frac{2\varepsilon}{5}. \end{aligned} \quad (\text{E.16})$$

We can pick a sufficiently large N to make r_N sufficiently small and then pick a W_3 such that

$$\frac{1+\kappa}{\delta_{\tilde{\Phi}}} \frac{M_f \sqrt{C_2(N+1)}}{\sqrt{W_3 \frac{\varepsilon}{5}}} + \frac{1+\kappa}{\delta_{\tilde{\Phi}}} r_N \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}}. \quad (\text{E.17})$$

Combining (E.16) with (E.17), when $W = W_3$,

$$\begin{aligned} & \mathcal{P} \left\{ \max_{P \in \Gamma} \left| \frac{\int_{\Gamma} K(P, Q)(\tilde{\Phi}^{m-1}(Q) - \Phi(Q))dQ + (\Phi(P) - \tilde{\Phi}^{m-1}(P))}{\tilde{\Phi}^{m-1}(P)} \right| \right. \\ & \left. \leq 1 - \sqrt{\frac{\kappa_p}{\gamma}} \right\} \geq 1 - \frac{2\varepsilon}{5}, \end{aligned} \quad (\text{E.18})$$

which is the second inequality of (E.2).

The third inequality of (E.2) can be easily proved. All we have to do is to go over the steps from (E.6) through (E.9) with the superscript 1 replaced by m and 0 replaced by $m-1$. That is, we can pick the same W_4 as determined by (E.9) such that when $W = W_4$,

$$\mathcal{P} \left\{ \max_{P \in \Gamma} |\Phi(P) - \tilde{\Phi}^m(P)| < \lambda \max_{P \in \Gamma} |\tilde{\Phi}^{m-1}(P) - \Phi(P)| + r_N \right\} \geq 1 - \varepsilon. \quad (\text{E.19})$$

Thus, when $W = W_0 = \max\{W_1, W_2, W_3, W_4\}$, (E.13), (E.18) and (E.19) all hold.

The proof of Theorem 9.1 is completed. \square

References

1. Booth T. Exponential convergence for Monte Carlo particle transport. *Trans Amer Nuclear Soc.* 1985; 50:267–268.
2. Booth T. Zero-variance solutions for linear Monte Carlo. *Nuclear Sci Eng.* 1989; 102:332–340.
3. Booth T. Exponential convergence on a continuous Monte Carlo transport problem. *Nuclear Sci Eng.* 1997; 127:338–345.
4. Case, KM., Zweifel, PW. *Linear Transport Theory.* Addison-Wesley, Reading; 1967.

5. Kong, R., Spanier, J. Error analysis of sequential monte carlo methods for transport problems. Springer; Berlin: 2000. p. 252-272. Monte Carlo and Quasi-Monte Carlo Methods 1998 (Claremont 1998)
6. Kong, R., Spanier, J. Sequential correlated sampling methods for some transport problems. Springer; Berlin: 2000. p. 238-251. Monte Carlo and Quasi-Monte Carlo Methods 1998 (Claremont 1998)
7. Kong, R., Spanier, J. Residual versus error in transport problems. Springer; Berlin: 2002. p. 306-317. Monte Carlo and Quasi-Monte Carlo Methods 2000 (Hong Kong 2000)
8. Kong R, Spanier J. A new proof of geometric convergence for general transport problems based on sequential correlated sampling methods. J Comput Physics. 2008; 227:9762–9777.
9. Kong R, Spanier J. Geometric convergence of adaptive monte carlo algorithms for radiative transport problems based on importance sampling methods. Nuclear Sci Eng. 2011; 168:197–225.
10. Lai, Y., Spanier, J. Adaptive importance sampling algorithms for transport problems. Springer; Berlin: 2000. p. 273-283. Monte Carlo and Quasi-Monte Carlo Methods 1998 (Claremont 1998)
11. Powell MJD, Swann J. Weighted uniform sampling—a Monte Carlo technique for reducing variance. J Inst Math Appl. 1966; 2:228–236.
12. Ross, S. A First Course in Probability. 5th. Prentice-Hall; Upper Saddle River: 1998.
13. Spanier, J., Gelbard, EM. Monte Carlo Principles and Neutron Transport Problems. Addison-Wesley, Reading; 1969.
14. Spanier, J., Kong, R. A new adaptive method for geometric convergence. Springer; Berlin: 2004. p. 439-449. Monte Carlo and Quasi-Monte Carlo Methods 2002 (Singapore 2002)
15. Silverman, BW. Density Estimation for Statistics and Data Analysis. Chapman and Hall; London: 1986.
16. Advanced Monte Carlo Methods. Feb. 2003 CRIAMS report LANL-03-001 to Los Alamos National Laboratory

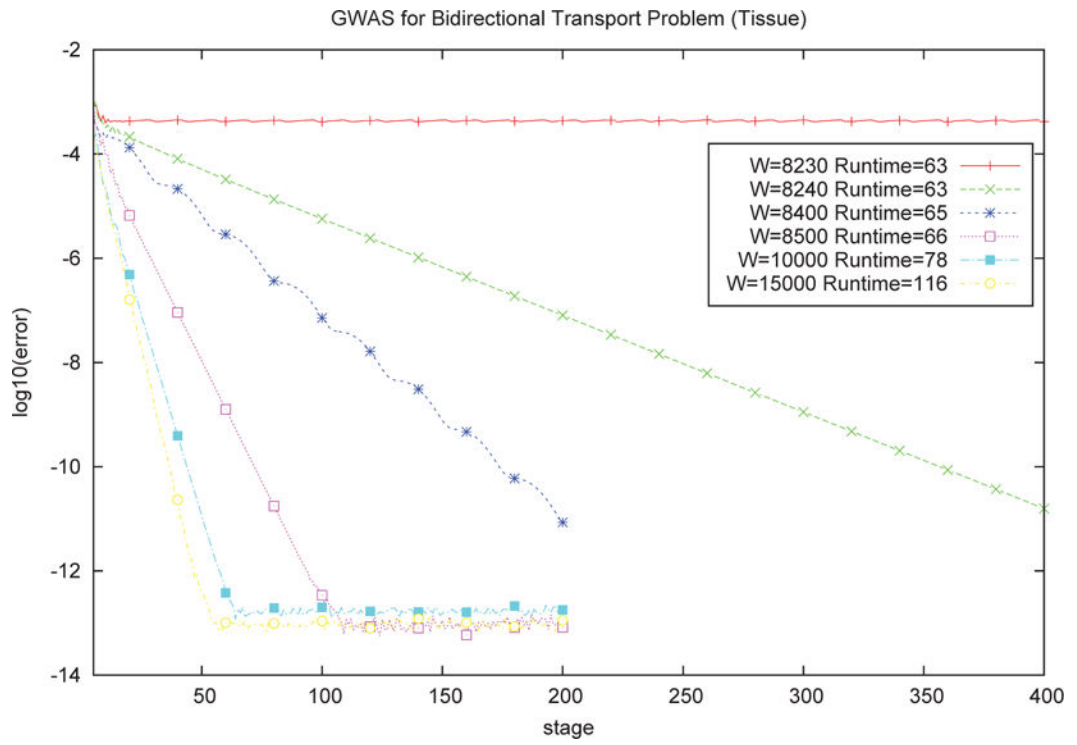


Figure 1.
GWAS for Bidirectional Transport Problem (tissue).

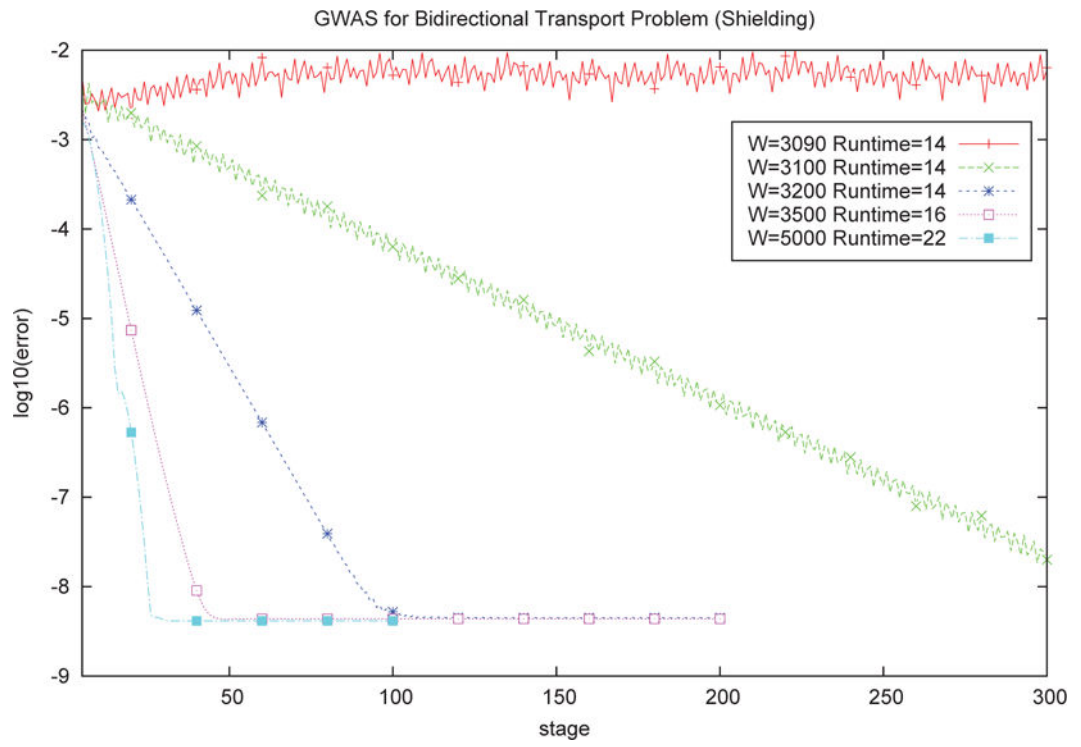


Figure 2.
GWAS for Bidirectional Transport Problem (shielding).

Table 1

Comparison of GWAS and AIS.

Method	S	Est.	R	σ^2	t	RelEff
Exact	-	0.8964537768861041	-	-	-	∞
GWAS	60	0.8964537768857454	5.95×10^{-13}	4.29×10^{-21}	203,940	1.142×10^{15}
AIS	20	0.8964537768868207	5.36×10^{-12}	6.08×10^{-19}	743,100	2.212×10^{12}

Table 2

Comparison of GWAS and AIS.

Method	S	Est.	R	σ^2	t	RelEff
Exact	-	0.0409019938626775	-	-	-	∞
GWAS	500	0.0409019938625908	3.71×10^{-13}	5.03×10^{-26}	34,500	5.763×10^{20}
AIS	174	0.0409019938626479	2.79×10^{-12}	4.25×10^{-24}	148,770	1.582×10^{18}