Combination of general antithetic transformations and control variables

Hatem Ben-Ameur
GERAD and HEC Montréal
hatem.ben-ameur@hec.ca

Pierre L'Ecuyer
GERAD and Département D'informatique et de Recherche Opérationnelle
Université de Montréal
lecuyer@iro.umontreal.ca

Christiane Lemieux
Department of Mathematics and Statistics
University of Calgary
lemieux@math.ucalgary.ca

October 14, 2003

Abstract

Several methods for reducing the variance in the context of Monte Carlo simulation are based on correlation induction. This includes antithetic variates, Latin hypercube sampling, and randomized version of quasi-Monte Carlo methods such as lattice rules and digital nets, where the resulting estimators are usually weighted averages of several dependent random variables that can be seen as function evaluations at a finite set of random points in the unit hypercube. In this paper, we consider a setting where these methods can be combined with the use of control variates and we provide conditions under which we can formally prove that the variance is minimized by choosing equal weights and equal control variate coefficients across the different points of evaluation, regardless of the function (integrand) that is evaluated.

1 Introduction

Suppose we want to compute

$$\mu = E[f(\mathbf{U})] = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u}$$

for some square-integrable function f, where $\mathbf{U}$ denotes a uniform random variable over $[0,1)^s$. The aim of most stochastic (Monte Carlo) simulations is to estimate such integrals, in which $\mathbf{u}$ can be interpreted as the sequence of independent "random numbers" that drive the simulation. Sometimes, f depends on a random and unbounded number of uniforms; in that case s can be taken as infinite.

The *crude Monte Carlo* method estimates μ by the average of $f(\mathbf{u}_0), \ldots, f(\mathbf{u}_{n-1})$, where the $\mathbf{u}_i$'s are independent and uniformly distributed over $[0,1)^s$. Here, we consider the use of a random point set $\tilde{P}_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ such that each $\mathbf{u}_i$ is uniformly distributed over $[0,1)^s$ but where the $\mathbf{u}_i$'s are not necessarily independent. We assume that these random variables $\mathbf{u}_i$ are defined over a common probability space $(\Omega, \mathcal{F}, P)$. The transformations $\mathbf{u}_i : \Omega \to [0,1)^s$ are designed to induce a *dependence structure* between the $\mathbf{u}_i$'s and we refer to them as *general antithetic* (GA) transformations (see also [13, 27]). They are sometimes called *correlation induction* methods [3]. In many cases, P is the uniform distribution over $[0,1)^s$, so ω can be interpreted as a uniform random vector. Actually, it has been shown in [13, 27] that for finite s, the search for transformations minimizing the variance can be restricted to the case where the $\mathbf{u}_i$'s have a common input ω uniformly distributed over $[0,1)$. However, here we do not make this assumption, because alternative interpretations of ω are more natural and convenient in some of our examples. Many well-known variance reduction techniques, including antithetic variates, rotation sampling, Latin hypercube sampling, randomly shifted lattice rules, and other types of randomized quasi-Monte Carlo point sets [3, 4, 12, 16, 18, 23] can be seen as special cases of GA transformations.

To improve the quality of our estimator of μ, we also want to use m measurable functions $C_l : [0,1)^s \to \mathbb{R}$ as control variables (CV), where we assume that $E[C_l(\mathbf{U})] = 0$ and $E[C_l^2(\mathbf{U})] < \infty$ for $l = 1, \ldots, m$, and that the covariance matrix $\Sigma_{C,C}$ with entries $\sigma_{ij} = \mathrm{Cov}(C_i, C_j))$ is positive definite, and therefore nonsingular. We are thus interested in approximating μ by estimators of the form

$$\hat{\mu}_{\mathrm{ga+cv}} = \sum_{i=0}^{n-1} \alpha_i X_i - \sum_{l=1}^{m} \sum_{i=0}^{n-1} \beta_{l,i} C_{l,i}, \tag{1}$$

where $X_i = f(\mathbf{u}_i)$ and $C_{l,i} = C_l(\mathbf{u}_i)$, for $i = 0, \ldots, n-1$, and

$$\sum_{i=0}^{n-1} \alpha_i = 1.$$

The goal is to choose the $n - 1 + nm$ free coefficients $\alpha_1, \ldots, \alpha_{n-1}, \beta_{1,0}, \ldots, \beta_{1,n-1}, \ldots, \beta_{m,0}, \ldots, \beta_{m,n-1}$ so as to minimize the variance of $\hat{\mu}_{\mathrm{ga+cv}}$.

We are interested in conditions on $\tilde{P}_n$ under which the optimal values of these coefficients satisfy

$$\alpha_0 = \ldots = \alpha_{n-1} = 1/n \text{ and } \beta_{l,0} = \ldots = \beta_{l,n-1}, \text{ for } l = 1, \ldots, m. \tag{2}$$

In other words, we want to know under what conditions on $\tilde{P}_n$ should each point $\mathbf{u}_i$ in $\tilde{P}_n$ be given the same weight α_i and the same CV coefficients in the construction of $\hat{\mu}_{\text{ga+cv}}$. Note that when $\tilde{P}_n$ is a set of independent uniformly distributed points as in the crude Monte Carlo method, it is easy to show that (2) holds. The problem is more challenging when $\tilde{P}_n$ has a nontrivial dependence structure, which is the setup considered in this paper. Interestingly, our conditions will turn out to be independent of the function f, so our results will hold *for any* f, as long as it is square-integrable.

At first sight, one might be tempted to believe that equal weights always prevail: Why give more importance to some $\mathbf{u}_i$'s than to others if all are uniformly distributed? Here is a simple counterexample.

Example 1 Let $s = 1$, $n = 3$, and $\mathbf{u}_i = u_i = (U + i/4) \bmod 1$ for $i = 0, 1, 2$, where U is uniformly distributed over $[0, 1)$. Clearly, each $\mathbf{u}_i$ is also uniformly distributed over $[0, 1)$. Suppose now that $f(u) = u$. Then the second moment of $\hat{\mu}_{\text{ga}} = \sum_{i=0}^{2} \alpha_i f(u_i)$ can be written as $V(\alpha_1, \alpha_2) = \int_0^{1/2} (u + \alpha_1/4 + \alpha_2/2)^2 du + \int_{1/2}^{3/4} (u + \alpha_1/4 - \alpha_2/2))^2 du + \int_{3/4}^{1} (u - (3/4)\alpha_1 - \alpha_2/2)^2 du$. The gradient of this expression with respect to (α_1, α_2) is zero when $\alpha_1 = 1/4$ and $\alpha_2 = 3/8$, and the Hessian matrix

$$H_V(\alpha_1, \alpha_2) = \begin{bmatrix} 3/8 & 1/4 \\ 1/4 & 1/2 \end{bmatrix}$$

is positive definite for any $(\alpha_1, \alpha_2) \in \mathbb{R}^2$. Therefore the variance is minimized by taking $\alpha_1 = 1/4$ and $\alpha_0 = \alpha_2 = 3/8$. For a different f, the optimal weights may change. They do depend on f. On the other hand if we take $u_i = (U + i/3) \bmod 1$ with $n = 3$, as a consequence of our results, the optimal weights are $\alpha_i = 1/3$ for all i *whatever be the (square-integrable) function* f.

For the function $f(u) = u$, this second estimator also turns out to have a smaller variance than the first one with the optimal weights $\alpha_1 = 1/4$ and $\alpha_0 = \alpha_2 = 3/8$. The variance is $1/108$ for the second estimator and $5/384$ for the first one. We point out that none of these two estimators minimizes the variance, among all possible GA schemes, for this particular f. For instance, taking $u_0 = U$ and $u_1 = 1 - U$ gives zero variance. However, this paper is not about finding the GA scheme that minimizes the variance for a particular f. This is a different problem. The goal of this paper is simply to provide conditions under which equal weights are optimal for a given GA scheme.

The estimator $\hat{\mu}_{\text{ga+cv}}$ can be rewritten as a Monte Carlo (MC) estimator that uses $n - 1 + nm$ control variates, as follows:

$$\hat{\mu}_{\text{ga+cv}} = X_0 - \alpha_1(X_0 - X_1) - \ldots - \alpha_{n-1}(X_0 - X_{n-1}) - \sum_{l=1}^{m} \sum_{i=0}^{n-1} \beta_{l,i} C_{l,i}. \tag{3}$$

This interpretation of antithetic variates as regression variables was already pointed out long ago by Tukey [26]. We denote by D the vector

$$(X_0 - X_1, \ldots, X_0 - X_{n-1}, C_{1,0}, \ldots, C_{1,n-1}, \ldots, C_{m,0}, \ldots, C_{m,n-1})^T$$

of control variates, and by X the plain MC estimator X_0. From the theory of control variates, the vector of coefficients

$$\beta^* = (\alpha_1, \ldots, \alpha_{n-1}, \beta_{1,0}, \ldots, \beta_{1,n-1}, \ldots, \beta_{m,0}, \ldots, \beta_{m,n-1})^T$$

that minimizes the variance is obtained as a solution of the linear system

$$\Sigma_{D,D}\ \beta = \Sigma_{X,D}, \tag{4}$$

where $\Sigma_{D,D}$ is the $(n-1+nm) \times (n-1+nm)$ covariance matrix of the vector D, and $\Sigma_{X,D}$ is the covariance vector of X_0 with each of the control variates in D, i.e.,

$$\begin{aligned}\Sigma_{X,D} \quad = \quad & (\mathrm{Cov}(X_0, X_0 - X_1), \ldots, \mathrm{Cov}(X_0, X_0 - X_{n-1}), \\ & \mathrm{Cov}(X_0, C_{1,0}), \ldots, \mathrm{Cov}(X_0, C_{m,n-1}))^T.\end{aligned}$$

Our main result, stated in Proposition 1, gives sufficient conditions for (2) to hold. We then consider different settings for $\tilde{P}_n$ under which these conditions are satisfied. Informally speaking, these conditions are that there must be a set of permutations of $[0, 1, \ldots, n-1]$ under which the joint distribution of any pair of points $(\mathbf{u}_i, \mathbf{u}_j)$ in $\tilde{P}_n$ is invariant and this set must be rich enough to sufficiently "shuffle" $[0, 1, \ldots, n-1]$, as we explain in Section 2. This proposition covers results that can be found in [2, 12], and allows us to prove that (2) holds for different types of GA transformations.

Note that property (2) does not imply that for any given l, the optimal coefficients $\beta_{l,0}, \ldots, \beta_{l,n-1}$ are equal to the optimal coefficient β_l that would be used for $n = 1$, i.e., if CV was not combined with GA transformations. It is well known that these optimal coefficients generally differ, because GA changes the covariance structure (see, e.g., [14]).

It is important to point out that in this paper, we assume that the weights have to be chosen *before* the random points in $\tilde{P}_n$ are observed. The case where the weights can be chosen *after* these points are observed is quite different and has been studied, e.g., in [28, 6]. For example, it is shown in [6] that if f is a twice continuously differentiable function over $[0, 1]$ and the u_i's are n i.i.d. $U(0, 1)$ random variables sorted by increasing order, then the variance of the estimator $\hat{\mu}_{\mathrm{ga}} = \sum_{i=0}^{n-1} \alpha_i f(u_i)$ has order $O(n^{-1})$ if $\alpha_i = 1/n$ for all i, whereas it has order $O(n^{-4})$ if $\alpha_0 = (u_0 + u_1)/2$, $\alpha_{n-1} = (2 - u_{n-1} - u_{n-2})/2$, and $\alpha_i = (u_{i+1} - u_{i-1})/2$ for $1 \le i \le n-2$. Similar results in higher dimensions can be found in [28]. However, significant gains by these techniques that assign weights *a posteriori* are difficult to achieve in practice when the dimension exceeds a few units.

Our results assume that all X_i and $C_{l,i}$ are available. They do not take into account the potential savings that can be made by not computing some of them if their weights or coefficients are zero, and the resulting *efficiency* tradeoff. For example, if $n = 2$, (2) is satisfied, and if it takes

twice the amount of time for computing $(X_0, C_{1,0}, \ldots, C_{m,0})$ and $(X_1, C_{1,1}, \ldots, C_{m,1})$ than for computing $(X_0, C_{1,0}, \ldots, C_{m,0})$ alone, then using the GA scheme is more efficient than using independent replications only if it reduces the variance by a factor larger than 2 compared with the case $n = 1$. Several other articles and books concentrate on this efficiency issue (e.g., [3, 4, 9, 11]), usually assuming equal weights *a priori* in the case of GA methods.

The remainder of this paper is organized as follows. In Section 2, we introduce some notation, state our basic result, and prove it. Section 3 considers settings where the $\mathbf{u}_i$'s form an abelian group of random variables and our main result applies. Many examples of sets $\tilde{P}_n$ obtained from different well-known GA techniques are included in these settings and some of them are discussed in Section 4. Section 5 deals with sets $\tilde{P}_n$ obtained by combining two of these GA techniques. The special case where conditional Monte Carlo is combined with GA transformations is examined in Section 6. Concluding comments are given in Section 7.

2 Families of Permutations Preserving Distribution of pairs

We first introduce some notation. If π is a permutation of the indices $[0, 1, \ldots, n-1]$, we denote by D^π the vector of control variates obtained by permuting the order of $\tilde{P}_n$ according to π, i.e., we have

$$\begin{aligned} D^\pi \;=\; & (X_{\pi(0)} - X_{\pi(1)}, \ \ldots, \ X_{\pi(0)} - X_{\pi(n-1)}, \\ & C_{1,\pi(0)}, \ldots, C_{1,\pi(n-1)}, \ldots, C_{m,\pi(0)}, \ldots, C_{m,\pi(n-1)}). \end{aligned}$$

We denote by $\Sigma^\pi_{D,D}$ the covariance matrix of D^π, by $\Sigma^\pi_{X,D}$ the vector of covariances between the plain MC estimator $X_{\pi(0)}$ and D^π, and by $\beta^{\pi,*}$ the vector of optimal coefficients that solves $\Sigma^\pi_{D,D}\beta = \Sigma^\pi_{X,D}$.

Our arguments will be based on the following string of ideas. We first note that for a given permutation π that preserves the joint distribution of pairs of points in $\tilde{P}_n$, it is easy to prove that $\beta^{\pi,*} = \beta^*$. If such a permutation π exchanges two different indices $i, j \in \{0, \ldots, n-1\}$, this implies that $\alpha_i = \alpha_j$ and $\beta_{l,i} = \beta_{l,j}$ for these two specific indices. This can be used to prove (2) if we can identify a family of permutations preserving the joint distribution and such that any given index in $[0, 1, \ldots, n-1]$ can be moved to any position by successively applying an appropriate sequence of permutations from that family. This is what we meant in the introduction by "sufficiently shuffling" the set $[0, 1, \ldots, n-1]$. More precisely, our first result is:

Proposition 1 *Assume that $\Sigma_{D,D}$ is non-singular, that $\Sigma_{X,D} \neq 0$, and that $\tilde{P}_n$ is such that there exists a set of permutations $\Pi = \{\pi^1, \ldots, \pi^d\}$ of $[0, 1, \ldots, n-1]$ satisfying the two following properties:*

(a) For any $\pi \in \Pi$, the joint distribution of $(\mathbf{u}_{\pi(i)}, \mathbf{u}_{\pi(j)})$ is the same as that of $(\mathbf{u}_i, \mathbf{u}_j)$ for all $0 \leq i, j \leq n-1$.

(b) For any $i \in \{0, 1, \ldots, n-1\}$, there is a sequence $\{\pi^{1(i)}, \ldots, \pi^{k(i)}\}$ in Π such that

$$\pi^{k(i)} \circ \cdots \circ \pi^{1(i)}(i) = 0;$$

Then the (unique) vector β^* of optimal coefficients satisfies (2).

Proof: Let Π be a set of permutations that satisfy (a) and (b). For any permutation $\pi \in \Pi$, the vector $\beta^{*,\pi}$ of optimal coefficients is a solution to

$$\Sigma^{\pi}_{D,D}\beta = \Sigma^{\pi}_{X,D}. \tag{5}$$

Note that property (a) implies that $\Sigma^{\pi}_{D,D} = \Sigma_{D,D}$ and $\Sigma^{\pi}_{X,D} = \Sigma_{X,D}$. Combining this with our assumption on $\Sigma_{D,D}$ and $\Sigma_{X,D}$, we get that the solution to (5) is unique, non-zero, and equal to β^*. This implies that for any $i \in \{0, \ldots, n-1\}$, we have

$$\alpha_i = \alpha_{\pi(i)} \text{ and } \beta_{l,i} = \beta_{l,\pi(i)}, \qquad \text{for } l = 1, \ldots, m. \tag{6}$$

Applying (6) to the sequence of permutations $\pi^{1(i)}, \ldots, \pi^{k(i)}$ for which

$$\pi^{k(i)} \circ \cdots \circ \pi^{1(i)}(i) = 0,$$

we obtain that

$$\begin{aligned} \alpha_i &= \alpha_{\pi^{1(i)}(i)} = \alpha_{\pi^{2(i)} \circ \pi^{1(i)}(i)} = \ldots = \alpha_{\pi^{k(i)} \circ \ldots \circ \pi^{1(i)}(i)} = \alpha_0 \qquad \text{and} \\ \beta_{l,i} &= \beta_{l,\pi^{1(i)}(i)} = \beta_{l,\pi^{2(i)} \circ \pi^{1(i)}(i)} = \ldots = \beta_{l,\pi^{k(i)} \circ \ldots \circ \pi^{1(i)}(i)} = \beta_{l,0}, \quad \text{for } l = 1, \ldots, m. \end{aligned}$$

Since this can be done for each i, β^* satisfies (2). □

3 Abelian Groups of Dependent Random Variables

We now study a situation where $\tilde{P}_n$ forms a *abelian group* of random variables with some special properties. This covers many practical settings. Examples of group operators over $\tilde{P}_n$ are given in the proof of Lemma 2 and in Section 4. We introduce three sets of sufficient conditions on $\tilde{P}_n$ under which Proposition 1 can be used to prove the optimality of the uniform weights given in (2). Condition 1 is the most general and will be used directly in Section 4 to prove that (2) holds when $\tilde{P}_n$ is defined by certain types of randomized quasi-Monte Carlo methods. The other conditions are special cases, in the sense that they imply Condition 1, and they turn out to be convenient to verify in a number of practical settings.

Condition 1 The set $\tilde{P}_n$ is an abelian group of random variables uniformly distributed over $[0,1)^s$, and such that for any $\mathbf{u}_i, \mathbf{u}_j, \mathbf{u}_k \in \tilde{P}_n$, the joint distribution of $(\mathbf{u}_i, \mathbf{u}_j)$ is the same as that of $(\mathbf{u}_i \cdot \mathbf{u}_k,\ \mathbf{u}_j \cdot \mathbf{u}_k)$, where "$\cdot$" denotes the group operator. □

Condition 2 The set $\tilde{P}_n$ is an abelian group of random variables uniformly distributed over $[0,1)^s$, the probability measure P corresponds to the uniform distribution over $\Omega = [0,1)^s$, and $(\mathbf{u}_i \cdot \mathbf{u}_j)(\omega) = \mathbf{u}_i(\mathbf{u}_j(\omega))$ for all i, j, and $\omega \in \Omega$. □

Condition 3 The random point set $\tilde{P}_n$ can be ordered so that the infinite sequence $\mathbf{w}_0, \mathbf{w}_1, \ldots$ defined by

$$\mathbf{w}_i = \mathbf{u}_{i \bmod n} \tag{7}$$

for $i \geq 0$ is *pairwise strongly stationary*, i.e., the joint distribution of $\mathbf{w}_i$ and $\mathbf{w}_{i+j}$ only depends on j, for $i, j \geq 0$. □

Lemma 2 *Each of Condition 2 or Condition 3 implies Condition 1.*

Proof. Suppose Condition 2 holds and let $\mathbf{u}_i, \mathbf{u}_j, \mathbf{u}_k \in \tilde{P}_n$. Then, both ω and $\mathbf{u}_k(\omega)$ are uniformly distributed over $[0,1)^s$, and therefore the joint distribution of $(\mathbf{u}_i \cdot \mathbf{u}_k,\ \mathbf{u}_j \cdot \mathbf{u}_k) = ((\mathbf{u}_i \cdot \mathbf{u}_k)(\omega),\ (\mathbf{u}_j \cdot \mathbf{u}_k)(\omega)) = (\mathbf{u}_i(\mathbf{u}_k(\omega)),\ \mathbf{u}_j(\mathbf{u}_k(\omega)))$ is the same as that of $(\mathbf{u}_i(\omega), \mathbf{u}_j(\omega)) = (\mathbf{u}_i, \mathbf{u}_j)$.

Now suppose Condition 3 holds. The set $\tilde{P}_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ equipped with the operator "$\cdot$" defined by $\mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{u}_{(i+j) \bmod n}$ forms an abelian group. Moreover, for every i, j, k, the joint distribution of $(\mathbf{u}_i \cdot \mathbf{u}_k,\ \mathbf{u}_j \cdot \mathbf{u}_k) = (\mathbf{u}_{(i+k) \bmod n},\ \mathbf{u}_{(j+k) \bmod n})$ is the same as that of $(\mathbf{u}_i, \mathbf{u}_j)$ because of the pairwise strong stationarity. □

In [12], the authors show that when $\tilde{P}_n$ satisfies Condition 3, choosing uniform weights α_i minimizes the variance when $\tilde{P}_n$ is used to estimate μ. A similar result for the case where $\tilde{P}_n$ satisfies Condition 2 is discussed in [2] and proved in [1]. Our Condition 1 allows to treat cases not covered by those previous results. Also, we consider in Proposition 6 constructions $\tilde{P}_n$ resulting from the combination of two sets and directly prove that Proposition 1 holds for these $\tilde{P}_n$. Another novelty of this paper is that we look at the combination with control variates $C_1, \ldots, C_m$.

To show that the assumptions of Proposition 1 are satisfied under Condition 1, we need certain properties of abelian groups, which we now recall. See, e.g., [7] for an account of group theory. The fundamental theorem of abelian groups says that $\tilde{P}_n$ can be written as a direct sum

$$\tilde{P}_n = Q_1 \oplus \ldots \oplus Q_r, \tag{8}$$

where for $i = 1, \ldots, r$, Q_i is a cyclic group of order n_i with generator $\mathbf{g}_i$. By asking for the integer r to be as small possible, this decomposition is unique (up to an isomorphism). In this context, r is called the *rank* of $\tilde{P}_n$, and the integers n_i are called the *invariants* and they satisfy $n_1 > 1$, and n_{k+1} divides n_k for $k = 1, \ldots, r-1$. Define $m_l = n_0 \ldots n_l$ for $l = 0, \ldots, r$, where $n_0 = 1$. We will show that the assumptions of Proposition 1 hold for the class of permutations $\Pi = \{\pi_d^l,\ l = 1, \ldots, r,\ d = 0, \ldots, n_l - 1\}$, where

$$\pi_d^l(i) = (i + d m_{l-1}) \bmod m_l + \lfloor i/m_l \rfloor m_l, \qquad i = 0, \ldots, n-1. \tag{9}$$

Let us explain what these permutations $\pi_d^1, \ldots, \pi_d^r$ do. For π_d^1, we partition $[0, 1, \ldots, n-1]$ into n/n_1 disjoint sequences of $n_1 = m_1$ contiguous values, and shift by d (modulo m_1) the position

of each element inside each sequence. For π_d^2, we partition $[0, 1, \ldots, n-1]$ into n/m_2 disjoint sequences of length $n_1 n_2 = m_2$; inside each sequence, the position of the elements is shifted by dm_1 (modulo m_2). More generally, for π_d^l, we partition $[0, 1, \ldots, n-1]$ into n/m_l disjoint sequences of length m_l, and inside each sequence, the position of the elements is shifted to the right by dm_{l-1} positions (modulo m_l). Figure 1 illustrates the effect of these permutations.

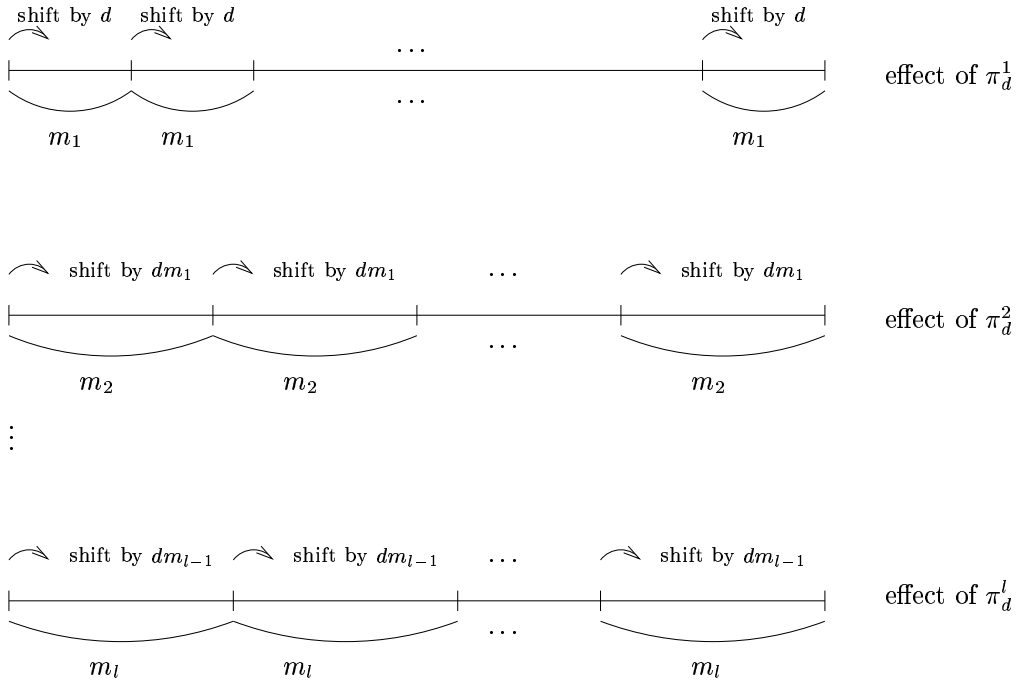

Figure 1: Effect of the permutations on $[0, 1, \ldots, n-1]$

The proof is built on two preliminary results. Lemma 3 says that when an abelian group P_n is initially ordered in a certain way and this order is permuted according to π_d^l, then in the representation of its elements according to the decomposition (8), only the element from the lth cyclic subgroup Q_l changes. Lemma 4 states that the permutations in Π satisfy property (a) of Proposition 1. To complete the proof that (2) holds, it then suffices to show that the family Π is rich enough to ensure property (b) of Proposition 1.

Lemma 3 *Let $\tilde{P}_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ be an abelian group and let r, $n_1, \ldots, n_r$ be its rank and invariants, respectively. Assume that $\tilde{P}_n$ has been ordered so that*

$$\mathbf{u}_i = \mathbf{g}_1^{i_1} \cdot \ldots \cdot \mathbf{g}_r^{i_r}, \qquad i = 0, \ldots, n-1,$$

where $\mathbf{g}^\nu$ is defined by $\mathbf{g}^1 = \mathbf{g}$ and $\mathbf{g}^\nu = \mathbf{g} \cdot \mathbf{g}^{\nu-1}$, and where $i_l = \lfloor i/m_{l-1} \rfloor \bmod n_l$ for $l = 1, \ldots, r$. Then for any $1 \le l \le r$, $0 \le d < n_l$, and $0 \le i < n$, we have that

$$\mathbf{u}_{\pi_d^l(i)} = \mathbf{g}_1^{i_1} \cdot \ldots \cdot \mathbf{g}_{l-1}^{i_{l-1}} \cdot \mathbf{g}_l^{(i_l+d) \bmod n_l} \cdot \mathbf{g}_{l+1}^{i_{l+1}} \cdot \ldots \cdot \mathbf{g}_r^{i_r}.$$

Proof: Let $l \in \{1\ldots,r\}$, $d \in \{0,\ldots,n_l-1\}$, and $i \in \{0,\ldots,n-1\}$. By definition, we have that

$$\mathbf{u}_{\pi_d^l(i)} = \mathbf{g}_1^{\nu_1} \cdot \ldots \cdot \mathbf{g}_r^{\nu_r}$$

where for $u = 1,\ldots,r$,

$$\nu_u = \lfloor \pi_d^l(i)/m_{u-1} \rfloor \bmod n_u.$$

We thus want to show that $\nu_u = i_u$ for $u \neq l$, and that $\nu_l = (i_l + d) \mod n_l$.

If $u < l$, then

$$\begin{aligned}
\nu_u &= \left\lfloor \frac{(i + dm_{l-1}) \bmod m_l + \lfloor i/m_l \rfloor m_l}{m_{u-1}} \right\rfloor \bmod n_u \\
&= \left(\left\lfloor \frac{i + dm_{l-1} - km_l}{m_{u-1}} \right\rfloor + \left\lfloor \frac{i}{m_l} \right\rfloor \frac{m_l}{m_{u-1}} \right) \bmod n_u, \qquad \text{for some integer } k \geq 0, \\
&= \left(\left\lfloor \frac{i}{m_{u-1}} \right\rfloor + \frac{dm_{l-1}}{m_{u-1}} - \frac{km_l}{m_{u-1}} + \left\lfloor \frac{i}{m_l} \right\rfloor \frac{m_l}{m_{u-1}} \right) \bmod n_u \\
&= \lfloor i/m_{u-1} \rfloor \bmod n_u \qquad (10) \\
&= i_u,
\end{aligned}$$

where (10) follows from the fact that m_l/m_{u-1} and m_{l-1}/m_{u-1} are both multiples of n_u.

If $u > l$, then $m_{u-1} = qm_l$ for some positive integer q. Hence

$$\frac{\pi_d^l(i)}{m_{u-1}} = \frac{\pi_d^l(i)}{qm_l} = \frac{i + dm_{l-1} \bmod m_l}{qm_l} + \left\lfloor \frac{i}{m_l} \right\rfloor \frac{1}{q} = \frac{\alpha}{qm_l} + \left\lfloor \frac{i}{m_l} \right\rfloor \frac{1}{q} \qquad (11)$$

$$\text{and} \qquad \frac{i}{m_{u-1}} = \frac{i}{qm_l} = \frac{i \bmod m_l}{qm_l} + \left\lfloor \frac{i}{m_l} \right\rfloor \frac{1}{q} = \frac{\beta}{qm_l} + \left\lfloor \frac{i}{m_l} \right\rfloor \frac{1}{q}, \qquad (12)$$

where $\alpha = (i + dm_{l-1}) \bmod m_l < m_l$ and $\beta = i \bmod m_l < m_l$. Now if we use the (unique) decomposition

$$\left\lfloor i/m_l \right\rfloor = aq + b,$$

where a and b are integers and $0 \leq b < q$ (i.e., take $a = \lfloor \lfloor i/m_l \rfloor / q \rfloor$ and $b = \lfloor i/m_l \rfloor \bmod q$), then from (11) and (12) we get that

$$\frac{\pi_d^l(i)}{m_{u-1}} = \frac{\alpha}{qm_l} + a + \frac{b}{q} \quad \text{and} \quad \frac{i}{m_{u-1}} = \frac{\beta}{qm_l} + a + \frac{b}{q}.$$

Since α/qm_l and β/qm_l are both strictly smaller than $1/q$ and $b \leq q-1$, we have that both $\alpha/qm_l + b/q$ and $\beta/qm_l + b/q$ are strictly smaller than one, and thus

$$\left\lfloor \pi_d^l(i)/m_{u-1} \right\rfloor = \left\lfloor i/m_{u-1} \right\rfloor = a,$$

which means $\nu_u = i_u$.

If $u = l$, then

$$\begin{aligned}\nu_u &= \left\lfloor \frac{(i + dm_{l-1}) \bmod m_l}{m_{l-1}} + \lfloor i/m_l \rfloor \, n_l \right\rfloor \bmod n_l \\ &= (\lfloor (i + dm_{l-1})/m_{l-1} \rfloor) \bmod n_l \\ &= (\lfloor i/m_{l-1} \rfloor + d) \bmod n_l \\ &= (i_u + d) \bmod n_l.\end{aligned}$$

□

Lemma 4 *Under Condition 1, the permutations π_d^l in Π satisfy property (a) of Proposition 1.*

Proof. We suppose that the order of the elements in $\tilde{P}_n$ has been fixed as in Lemma 3. For $\mathbf{u}_i = \mathbf{g}_1^{i_1} \cdot \ldots \cdot \mathbf{g}_r^{i_r}$ and $\mathbf{u}_j = \mathbf{g}_1^{j_1} \cdot \ldots \cdot \mathbf{g}_r^{j_r}$, observe that

$$\mathbf{u}_j \cdot \mathbf{u}_i^{-1} = \mathbf{g}_1^{(j_1 - i_1) \bmod n_1} \cdot \ldots \cdot \mathbf{g}_r^{(j_r - i_r) \bmod n_r}.$$

Using Lemma 3, we have that

$$\begin{aligned}&\mathbf{u}_{\pi_d^l(i)} \cdot \mathbf{u}_{\pi_d^l(j)}^{-1} \\ &= \mathbf{g}_1^{(i_1 - j_1) \bmod n_1} \cdot \ldots \cdot \mathbf{g}_{l-1}^{(i_{l-1} - j_{l-1}) \bmod n_{l-1}} \cdot \mathbf{g}_l^{(i_l + d - j_l - d) \bmod n_l} \cdot \mathbf{g}_{l+1}^{(i_{l+1} - j_{l+1}) \bmod n_{l+1}} \cdot \ldots \cdot \mathbf{g}_r^{(i_r - j_r) \bmod n_r} \\ &= \mathbf{u}_i \cdot \mathbf{u}_j^{-1},\end{aligned}$$

and therefore $\mathbf{u}_j^{-1} \cdot \mathbf{u}_{\pi_d^l(j)} = \mathbf{u}_i^{-1} \cdot \mathbf{u}_{\pi_d^l(i)}$. Let $\mathbf{u}_k$ be this group element. Then,

$$\begin{aligned}\mathbf{u}_{\pi_d^l(i)} &= \mathbf{u}_i \cdot \mathbf{u}_j^{-1} \cdot \mathbf{u}_{\pi_d^l(j)} = \mathbf{u}_i \cdot \mathbf{u}_k \qquad \text{and} \\ \mathbf{u}_{\pi_d^l(j)} &= \mathbf{u}_j \cdot \mathbf{u}_i^{-1} \cdot \mathbf{u}_{\pi_d^l(i)} = \mathbf{u}_j \cdot \mathbf{u}_k.\end{aligned}$$

By assumption, this means that $(\mathbf{u}_{\pi_d^l(i)}, \mathbf{u}_{\pi_d^l(j)})$ have the same joint distribution as $(\mathbf{u}_i, \mathbf{u}_j)$. □

Proposition 5 *Assume that $\Sigma_{D,D}$ is non-singular and that $\Sigma_{X,D} \neq 0$. If $\tilde{P}_n$ satisfies Condition 1,2 or 3, then (2) holds.*

Proof. It suffices to prove that the family Π of permutations introduced in (9) satisfy assumption (b) of Proposition 1.

For an arbitrary $i \in \{0, \ldots, n-1\}$, we define the following sequence of permutations. We first apply $\pi_{d(1)}^1$ with $d(1) = (n_1 - i) \bmod n_1$, which yields

$$\pi_{d(1)}^1(i) = k_1 m_1$$

for some integer $k_1 = \lfloor i/m_1 \rfloor \geq 0$, because $i + d(1) = 0 \pmod{m_1}$. More generally, the choice of permutations $\pi_{d(1)}^1, \ldots, \pi_{d(r)}^r$ is defined recursively as follows: using $k_0 = i$ as an initial value, we successively define for $l = 1, \ldots, r$,

$$d(l) = (n_l - k_{l-1}) \bmod n_l,$$

where $k_{l-1} = \lfloor k_{l-2}m_{l-2}/m_{l-1} \rfloor$ for $l > 1$. Since $k_{l-2}m_{l-2} + ((n_{l-1} - k_{l-2}) \bmod n_{l-1})m_{l-2} = 0 \pmod{m_{l-1}}$, we have that

$$\pi_{d(l-1)}^{l-1}(k_{l-2}m_{l-2}) = k_{l-1}m_{l-1}. \tag{13}$$

This sequence of permutations satisfies

$$\pi_{d(r)}^{r} \circ \cdots \circ \pi_{d(1)}^{1}(i) = k_r m_r \bmod n = 0,$$

because $m_r = n$, and this proves the result. □

4 Application to GA Techniques

We now give several examples of randomized point sets that are frequently used as GA (or quasi-Monte Carlo) methods in simulation and that satisfy our conditions.

Example 2 *Antithetic Variates.*

In this case, $\omega \equiv \mathbf{u}$ is uniformly distributed over $[0,1)^s$ and $\tilde{P}_n = \{\mathbf{u}_0, \mathbf{u}_1\} = \{\mathbf{u}, 1-\mathbf{u}\}$. It is easily seen that both Conditions 2 and 3 are satisfied. For Condition 2, we define $\mathbf{u}_0 \cdot \mathbf{u}_0 = \mathbf{u}_1 \cdot \mathbf{u}_1 = \mathbf{u}_0$, and $\mathbf{u}_0 \cdot \mathbf{u}_1 = \mathbf{u}_1 \cdot \mathbf{u}_0 = \mathbf{u}_1$. □

Example 3 *Randomly Shifted Lattice Rules.*

A *lattice rule* (see, e.g., [5, 17, 24]) estimates μ by averaging the values of f over the point set $P_n = L_s \cap [0,1)^s$, where

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^{s} z_j \mathbf{x}_j \text{ such that each } z_j \in \mathbb{Z} \right\}, \tag{14}$$

$\mathbf{x}_1, \ldots, \mathbf{x}_s$ are linearly independent vectors in $\mathbb{R}^s$, n denotes the cardinality of P_n, and $\mathbb{Z}^s \subseteq L_s$. Under the latter condition, L_s is called an *integration* lattice. The set $P_n = \{\mathbf{v}_0, \ldots, \mathbf{v}_{n-1}\}$, together with the operation $+$ defined over $\mathbb{R}^s$ by $\mathbf{v} + \mathbf{u} = (\mathbf{v} + \mathbf{u}) \mod 1$, where the addition and reduction modulo 1 on the right side are coordinate by coordinate, is an abelian group.

A *randomly-shifted lattice rule* replaces the deterministic point set P_n by $\tilde{P}_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} \subset [0,1)^s$, where $\mathbf{u}_i = \mathbf{v}_i + \mathbf{u}$ and $\mathbf{u} \equiv \omega$ is uniformly distributed over $[0,1)^s$. This $\tilde{P}_n$ is an abelian group under the operation defined by $\mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{u}_i + \mathbf{u}_j - \mathbf{u} = \mathbf{v}_i + \mathbf{v}_j + \mathbf{u}$. In this case, $\mathbf{u}_i \cdot \mathbf{u}_j(\omega) = \mathbf{u}_i(\mathbf{v}_j + \omega) = \mathbf{u}_i(\mathbf{u}_j(\omega))$, so Condition 2 is satisfied.

A lattice rule has *rank 1* if one can take $\mathbf{x}_2 = \mathbf{e}_2$, ..., $\mathbf{x}_s = \mathbf{e}_s$ in (14), where $\mathbf{e}_j$ is the jth unit vector in $\mathbb{R}^s$ or, equivalently, if P_n can be written as $P_n = \{i\mathbf{x}_1 \bmod 1,\ i = 0, \ldots, n-1\}$ for some vector $\mathbf{x}_1 \in [0,1)^s$. In this case, if we define $\mathbf{w}_i$ as in (7), the joint distribution of $(\mathbf{w}_i, \mathbf{w}_{i+j})$ is the same as that of $(\mathbf{u}, (j-i)\mathbf{x}_1 + \mathbf{u})$ for all i and j, so Condition 3 holds.

However, Condition 3 does not hold in general for rules of higher rank. As an example of this, consider a two-dimensional copy-rule with $n = 4$ and $P_n = \{(0,0), (0,1/2), (1/2,0), (1/2,1/2)\}$. Here, there is no way of ordering the points so that $(\mathbf{u}_0, \mathbf{u}_1)$ has the same joint distribution as $(\mathbf{u}_1, \mathbf{u}_2)$. □

Example 4 *Rotation sampling.*

Here, $\omega \equiv \mathbf{u}$ is uniformly distributed over $[0,1)^s$ and $\tilde{P}_n = \{\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\} = \{\mathbf{u}, (\mathbf{x} + \mathbf{u}) \bmod 1, \ldots, ((n-1)\mathbf{x} + \mathbf{u}) \bmod 1\}$, where $\mathbf{x} = (1/n, \ldots, 1/n)$; see [10]. This turns out to be a special case of a randomly shifted rank-1 lattice rule, with $\mathbf{x}_1 = \mathbf{x}$ (see the previous example) and therefore Condition 2 holds. Glynn and Szechtman [12], page 40, did verify Condition 3 for this example. □

Example 5 *Latin Hypercube Sampling.*

This method uses the randomized point set $\tilde{P}_n = \{(\pi^1(i)/n, \ldots, \pi^s(i)/n) + \mathbf{y}_i,\ i = 0, \ldots, n-1\}$, where the π^j's are i.i.d. uniform permutations of $[0, 1, \ldots, n-1]$, and the $\mathbf{y}_i$'s are i.i.d. uniform over $[0, 1/n)^s$ [3, 20, 23]. We can interpret ω as the randomness needed to generate all the π_j's and $\mathbf{y}_i$'s. Here, all pairs $(\mathbf{u}_i, \mathbf{u}_j)$ for $i \neq j$ have the same joint distribution and all pairs $(\mathbf{u}_i, \mathbf{u}_i)$ have the same joint distribution. Thus, Condition 3 holds for any ordering of the $\mathbf{u}_i$'s. □

Example 6 *Digitally Shifted Nets.*

We consider a special case of a *digital net in base b* [8, 21, 25] where the underlying commutative ring is $\mathbb{Z}_b$ and all the bijections are the identity (which is often the case in practice). Such a net corresponds to a deterministic point set $P_n = \{\mathbf{v}_i = (v_{i,1}, \ldots, v_{i,s}),\ i = 0, \ldots, n-1\}$, where $n = b^k$ for some positive integer k, the coordinates $v_{i,j}$ are defined by

$$v_{i,j} = \begin{pmatrix} b^{-1} & b^{-2} & \cdots \end{pmatrix} \mathbf{C}_j \begin{pmatrix} a_{i,0} \\ a_{i,1} \\ \vdots \\ a_{i,k-1} \end{pmatrix} \stackrel{\text{def}}{=} v_{i,j,1} b^{-1} + v_{i,j,2} b^{-2} + \cdots$$

for some carefully selected $\infty \times k$-dimensional matrices $\mathbf{C}_1, \ldots, \mathbf{C}_s$ with elements in $\mathbb{Z}_b$, and the $a_{i,j}$'s are the digits of the expansion of i in base b; i.e., $i = a_{i,0} + a_{i,1} b + \cdots + a_{i,k-1} b^{k-1}$. (We assume that infinitely many coefficients $v_{i,j,l}$ differ from $b-1$ for each (i,j), so the expansion is unique.) The set P_n forms an abelian group under the operation "+" defined by

$$\mathbf{v}_i + \mathbf{v}_{i'} = \left(\sum_{l=1}^{\infty} ((v_{i,1,l} + v_{i',1,l}) \bmod b) b^{-l},\ \ldots,\ \sum_{l=1}^{\infty} ((v_{i,s,l} + v_{i',s,l}) \bmod b) b^{-l} \right). \tag{15}$$

A *digitally shifted net* is a random point set defined as $\tilde{P}_n = P_n + \mathbf{u}$ where P_n is a digital net and $\mathbf{u}$ is uniformly distributed over $[0,1)^s$. For $\mathbf{u}_i = \mathbf{v}_i + \mathbf{u}$ and $\mathbf{u}_j = \mathbf{v}_j + \mathbf{u}$ in $\tilde{P}_n$, define $\mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{v}_i + \mathbf{v}_j + \mathbf{u}$. Under this operation "$\cdot$", it is easily seen that $\tilde{P}_n$ is an abelian group of random variables satisfying Condition 1. Indeed, $(\mathbf{u}_i \cdot \mathbf{u}_m,\ \mathbf{u}_j \cdot \mathbf{u}_m) = (\mathbf{v}_i + \mathbf{v}_m + \mathbf{u},\ \mathbf{v}_j + \mathbf{v}_m + \mathbf{u})$ has the same distribution as $(\mathbf{u}_i,\ \mathbf{u}_j) = (\mathbf{v}_i + \mathbf{u},\ \mathbf{v}_j + \mathbf{u})$, because both $\mathbf{u}$ and $\mathbf{v}_m + \mathbf{u}$ are uniformly distributed. □

Example 7 *Linearly Scrambled Digital Nets.*

This method is very similar to the previous one, except that the deterministic digital net P_n is replaced by a random one, call it $\hat{P}_n$, in which the generating matrices have been randomly "scrambled" [19, 15]. Hence ω in this case is the randomness required to scramble these matrices and to generate $\mathbf{u}$. To prove that $\tilde{P}_n$ satisfies Condition 1, we use the same operator $\cdot$ as in the previous example, but the $\mathbf{v}_i$'s now come from the random digital net $\hat{P}_n$.

Example 8 *Scrambled Digital Nets.*

In this case, the point set $\tilde{P}_n$ is obtained by applying certain random permutations to each digit in the expansion of each coordinate of the points coming from a digital net in base b, P_n, defined as in Example 6. We refer to [22] for the details. Here we only state the facts needed to prove that this case is covered by Condition 1: (i) each $\mathbf{u}_i \in \tilde{P}_n$ is uniformly distributed over $[0,1)^s$; (ii) if $\mathbf{v}_i, \mathbf{v}_j \in P_n$, the joint distribution of $(\mathbf{u}_i, \mathbf{u}_j)$ is completely determined by the vector $(q_1, \ldots, q_s)$, where q_l is such that in dimension l, the first q_l digits of $\mathbf{v}_i$ and $\mathbf{v}_j$ are the same, but they differ on the (q_l+1)th digit. Let φ denote the (random) transformation from P_n to $\tilde{P}_n$, so $\mathbf{u}_i = \varphi(\mathbf{v}_i)$ for each i, and define the operator "$\cdot$" by $\mathbf{u}_i \cdot \mathbf{u}_j = \varphi(\mathbf{v}_i + \mathbf{v}_j)$, where operation $+$ is defined as in (15). It can be verified that $\tilde{P}_n$ is an abelian group under this operation. Moreover, Condition 1 holds because for any $\mathbf{v}_m \in P_n$, the vector $(q_1, \ldots, q_s)$ for $(\mathbf{v}_i + \mathbf{v}_m,\ \mathbf{v}_j + \mathbf{v}_m)$ is the same as that for $(\mathbf{v}_i, \mathbf{v}_j)$. □

Our conditions may apply to cases where different GA methods are combined, as illustrated by the following example.

Example 9 *Modified Latin Hypercube Sampling Combined with a Randomly Shifted Lattice Rule.*

We consider a lattice rule with point set $P_{n_1} = \{\mathbf{v}_0, \ldots, \mathbf{v}_{n_1-1}\}$ and a modified Latin Hypercube sampling scheme with (random) point set $Q_{n_2} = \{\mathbf{w}_j = \boldsymbol{\pi}(j)/n_2 + \mathbf{u},\ j = 0, \ldots, n_2 - 1\}$, where $\boldsymbol{\pi}(j) = (\pi^1(j), \ldots, \pi^s(j))$, the π^l's are i.i.d. uniform permutations of $[0, 1, \ldots, n_2 - 1]$, $\mathbf{u}$ is uniformly distributed over $[0,1)^s$ (instead of $[0, 1/n_2)^s$, and the same $\mathbf{u}$ is used for all j; this is why we say it is a "modified" scheme), and the operation $+$ corresponds to addition modulo 1. Let $n = n_1 n_2$ and $\tilde{P}_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ where $\mathbf{u}_i$ is defined as $\mathbf{u}_i = \boldsymbol{\pi}(j)/n_2 + \mathbf{u} + \mathbf{v}_l$ if $i = n_2 l + j$, where the $+$ operation in the definition of $\mathbf{u}_i$ is again defined as addition modulo 1. In words, this point set corresponds to n_2 randomly shifted copies of P_{n_1}, using the n_2 points of Q_{n_2} for the shifts.

For $i = n_2 l + j$ and $i' = n_2 l' + j'$, $0 \le i, i' < n$, define $\mathbf{u}_i \cdot \mathbf{u}_{i'} = \boldsymbol{\pi}((j + j') \bmod n_2)/n_2 + \mathbf{u} + \mathbf{v}_l + \mathbf{v}_{l'}$. With this operation, $\tilde{P}_n$ is an abelian group that satisfies Condition 1, because for $i'' = n_2 l'' + j''$, $\mathbf{u}_i \cdot \mathbf{u}_{i''} = \boldsymbol{\pi}((j + j'') \bmod n_2)/n_2 + \mathbf{u} + \mathbf{v}_l + \mathbf{v}_{l''}$ and $\mathbf{u}_{i'} \cdot \mathbf{u}_{i''} = \boldsymbol{\pi}((j' + j'') \bmod n_2)/n_2 + \mathbf{u} + \mathbf{v}_{l'} + \mathbf{v}_{l''}$ have the same joint distribution as $\mathbf{u}_i = \boldsymbol{\pi}(j \bmod n_2)/n_2 + \mathbf{u} + \mathbf{v}_l$ and $\mathbf{u}_{i'} = \boldsymbol{\pi}(j' \bmod n_2)/n_2 + \mathbf{u} + \mathbf{v}_{l'}$. □

For certain GA combinations, Condition 1 can be difficult or impossible to verify, but one may still be able to verify the conditions of Proposition 1 via a different path. In the next section we give a result that provides a different set of sufficient conditions for Proposition 1 that are convenient to verify for certain types of combined methods.

5 Combining GA techniques

The next proposition shows that for certain sets $\tilde{P}_n$ obtained by combining two smaller sets, Proposition 1 can be used with a different set of permutations than the one defined in (9) to prove that (2) holds. This bypasses Proposition 5 and the verification of Condition 1.

Proposition 6 *Assume that* $\tilde{V}_q = \{\mathbf{v}_0, \ldots, \mathbf{v}_{q-1}\}$ *satisfies Condition 2 and* $\tilde{W}_t = \{\mathbf{w}_0, \ldots, \mathbf{w}_{t-1}\}$ *satisfies Condition 1. Let* $\tilde{P}_n = \{\mathbf{v}_i(\mathbf{w}_j),\ i = 0, \ldots, q-1,\ j = 0, \ldots, t-1\}$. *(So* $\tilde{P}_n$ *has the same group operation as* $\tilde{V}_q$*). If for any* $\mathbf{v}_k \in \tilde{V}_q, \mathbf{w}_i, \mathbf{w}_j \in \tilde{W}_t$, $(\mathbf{v}_k(\mathbf{w}_i), \mathbf{v}_k(\mathbf{w}_j))$ *has the same distribution as* $(\mathbf{w}_i, \mathbf{w}_j)$ *or* $(\mathbf{w}_j, \mathbf{w}_i)$*, then* $\tilde{P}_n$ *satisfies the conditions of Proposition 1.*

Note that if we were asking for $(\mathbf{v}_k(\mathbf{w}_i), \mathbf{v}_k(\mathbf{w}_j))$ to have the same joint distribution as $(\mathbf{w}_i, \mathbf{w}_j)$ for all k, then the resulting point set $\tilde{P}_n$ would satisfy Condition 1 and there would be nothing more to prove. Our condition is weaker because we allow this joint distribution to be equal to that of $(\mathbf{w}_j, \mathbf{w}_i)$ instead. This weaker condition would be easy to handle if no control variables were used, since $\mathrm{Cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \mathrm{Cov}(f(\mathbf{u}_j), f(\mathbf{u}_i))$. The problem here is that with control variables, we also need to verify that $\mathrm{Cov}(f(\mathbf{u}_i), C_k(\mathbf{u}_j)) = \mathrm{Cov}(f(\mathbf{u}_j), C_k(\mathbf{u}_i))$, and this is not necessarily true under our weaker condition.

Proof. We need to define a set of permutations Π such that part (a) and (b) of Proposition 1 hold. Since $\tilde{P}_n$ is obtained by composing two sets, it seems natural to define the permutations of Π in terms of two permutations that respectively act on $l = \lfloor i/t \rfloor$ and $r = i \bmod t$, for a given $i \in \{0, \ldots, n-1\}$ (i.e., l and $r < t$ are the unique non-negative integers such that $i = lt + r$). Also, since each of $\tilde{V}_q$ and $\tilde{W}_t$ satisfies Condition 1, it seems reasonable to use the sets $\Pi_1 = \{\pi_d^l, 1 \le l \le r_1, 0 \le d < n_{l,1}\}$ and $\Pi_2 = \{\tilde{\pi}_d^l, 1 \le l \le r_2, 0 \le d < n_{l,2}\}$ given in (9) to define these two permutations. Here r_1, $n_{1,1}, \ldots, n_{r_1,1}$ and r_2, $n_{1,2}, \ldots, n_{r_2,2}$ are the rank and invariants associated with $\tilde{V}_q$ and $\tilde{W}_t$, respectively. Before describing Π, recall from the proof of Lemma 4 that for any $i, j \in \{0, \ldots, q-1\}$, $\mathbf{v}_i^{-1} \cdot \mathbf{v}_{\pi_d^l(i)} = \mathbf{v}_j^{-1} \cdot \mathbf{v}_{\pi_d^l(j)}$. Let us call this common element $\mathbf{v}_k$ (where k depends on d and l, for the remainder of this proof) and define the set

$$\begin{aligned} A \;=\; & \{(d,l), 0 \le d < n_{l,1}, 1 \le l \le r_1 : (\mathbf{v}_k(\mathbf{w}_i), \mathbf{v}_k(\mathbf{w}_j)) \text{ has the same} \\ & \text{joint distribution as } (\mathbf{w}_i, \mathbf{w}_j)\}. \end{aligned}$$

Hence $(d,l) \notin A$ means that $(\mathbf{v}_k(\mathbf{w}_i), \mathbf{v}_k(\mathbf{w}_j))$ has the same joint distribution as $(\mathbf{w}_j, \mathbf{w}_i)$. Also, let $(\tilde{\pi}_d^l(r))^{-1}$ be such that $\mathbf{w}_{(\tilde{\pi}_d^l(r))^{-1}} = \mathbf{w}_{\tilde{\pi}_d^l(r)}^{-1}$. We define

$$\Pi = \{\pi_{\delta,\bar{\delta}}^{\ell,\bar{\ell}}, 1 \le \ell \le r_1, 1 \le \bar{\ell} \le r_2, 0 \le \delta < n_{\ell,1}, 0 \le \bar{\delta} < n_{\bar{\ell},2}\},$$

where for $i = tl + r$, $0 \le r < t$, we have

$$\pi_{\delta,\bar{\delta}}^{\ell,\bar{\ell}}(i) = \begin{cases} t\pi_\delta^\ell(l) + \tilde{\pi}_{\bar{\delta}}^{\bar{\ell}}(r) & \text{if } (\delta,\ell) \in A \\ t\pi_\delta^\ell(l) + (\tilde{\pi}_{\bar{\delta}}^{\bar{\ell}}(r))^{-1} & \text{otherwise.} \end{cases}$$

To prove part (a) of Proposition 1, we need to show that for any $\pi_{\delta,\bar{\delta}}^{\ell,\bar{\ell}} \in \Pi$ and any $\mathbf{u}_i, \mathbf{u}_j \in \tilde{P}_n$, if $i' = \pi_{\delta,\bar{\delta}}^{\ell,\bar{\ell}}(i)$ and $j' = \pi_{\delta,\bar{\delta}}^{\ell,\bar{\ell}}(j)$, then $(\mathbf{u}_{i'}, \mathbf{u}_{j'})$ has the same joint distribution as $(\mathbf{u}_i, \mathbf{u}_j)$. Writing

$i = lt + r$, and $j = mt + s$ where $0 \le r, s < t$, assume first that $(\delta, \ell) \in A$. Using the same arguments as in the proof of Lemma 4 and the fact that $\tilde{V}_q$ satisfies Condition 2, we get

$$\mathbf{u}_{i'} = \mathbf{v}_{\pi^{\ell}_{\delta}(l)}(\mathbf{w}_{\tilde{\pi}^{\bar{\ell}}_{\bar{\delta}}(r)}) = \mathbf{v}_l \cdot \mathbf{v}_k(\mathbf{w}_r \cdot \mathbf{w}_\kappa) = \mathbf{v}_l(\mathbf{v}_k(\mathbf{w}_r \cdot \mathbf{w}_\kappa))$$

and

$$\mathbf{u}_{j'} = \mathbf{v}_{\pi^{\ell}_{\delta}(m)}(\mathbf{w}_{\tilde{\pi}^{\bar{\ell}}_{\bar{\delta}}(s)}) = \mathbf{v}_m \cdot \mathbf{v}_k(\mathbf{w}_s \cdot \mathbf{w}_\kappa) = \mathbf{v}_m(\mathbf{v}_k(\mathbf{w}_s \cdot \mathbf{w}_\kappa)),$$

where κ is such that $\mathbf{w}_\kappa = \mathbf{w}_i^{-1} \cdot \mathbf{w}_{\tilde{\pi}^{\bar{\ell}}_{\bar{\delta}}(i)}$ for any $0 \le i < t$. By assumption that $\tilde{W}_t$ satisfies Condition 1, we know that $(\mathbf{w}_r \cdot \mathbf{w}_\kappa, \mathbf{w}_s \cdot \mathbf{w}_\kappa)$ has the same joint distribution as $(\mathbf{w}_r, \mathbf{w}_s)$. Combining this with the fact that $(\delta, \ell) \in A$, we have that $(\mathbf{v}_k(\mathbf{w}_r \cdot \mathbf{w}_\kappa), \mathbf{v}_k(\mathbf{w}_s \cdot \mathbf{w}_\kappa))$ has the same joint distribution as $(\mathbf{w}_r, \mathbf{w}_s)$, as required. If $(\delta, \ell) \notin A$ then,

$$\mathbf{u}_{\pi^{\ell,\bar{\ell}}_{\delta,\bar{\delta}}(i)} = \mathbf{v}_{\pi^{\ell}_{\delta}(l)}(\mathbf{w}^{-1}_{\tilde{\pi}^{\bar{\ell}}_{\bar{\delta}}(r)}) = \mathbf{v}_l \cdot \mathbf{v}_k(\mathbf{w}_r^{-1} \cdot \mathbf{w}_\kappa^{-1}) = \mathbf{v}_l(\mathbf{v}_k(\mathbf{w}_r^{-1} \cdot \mathbf{w}_\kappa^{-1})) \quad (16)$$

$$\mathbf{u}_{\pi^{\ell,\bar{\ell}}_{\delta,\bar{\delta}}(j)} = \mathbf{v}_{\pi^{\ell}_{\delta}(m)}(\mathbf{w}^{-1}_{\tilde{\pi}^{\bar{\ell}}_{\bar{\delta}}(s)}) = \mathbf{v}_m \cdot \mathbf{v}_k(\mathbf{w}_s^{-1} \cdot \mathbf{w}_\kappa^{-1}) = \mathbf{v}_m(\mathbf{v}_k(\mathbf{w}_s^{-1} \cdot \mathbf{w}_\kappa^{-1})), \quad (17)$$

where the second equality in both (16) and (17) comes from the abelian property of $\tilde{W}_t$. By assumption that $\tilde{W}_t$ satisfies Condition 1, we know that $(\mathbf{w}_r^{-1} \cdot \mathbf{w}_\kappa^{-1}, \mathbf{w}_s^{-1} \cdot \mathbf{w}_\kappa^{-1})$ has the same joint distribution as $(\mathbf{w}_r^{-1}, \mathbf{w}_s^{-1})$ and since $(\delta, \ell) \notin A$, $(\mathbf{v}_k(\mathbf{w}_r^{-1}), \mathbf{v}_k(\mathbf{w}_s^{-1}))$ has the same joint distribution as $(\mathbf{w}_s^{-1}, \mathbf{w}_r^{-1})$, which has the same joint distribution as $(\mathbf{w}_r, \mathbf{w}_s)$, again by Condition 1 (multiplying both arguments by $\mathbf{w}_s \cdot \mathbf{w}_r$, and using the abelian property of $\tilde{W}_t$).

We now prove that Π satisfies part (b) of Proposition 1. Let $i = tl + r$ and $0 \le r < t$. Following the proof of Proposition 5, we know that there exists a set of pairs $\{(d(j), j), j = 1, \ldots, r_1\}$ such that $\pi^{r_1}_{d(r_1)} \circ \cdots \circ \pi^{1}_{d(1)}(l) = 0$. Similarly, there exists a set of pairs $\{(e(j), j),\ j = 1, \ldots, r_2\}$ such that $\tilde{\pi}^{r_2}_{e(r_2)} \circ \cdots \circ \tilde{\pi}^{1}_{e(1)}(r) = 0$.

Observe that since π^1_0 is the identity, $(0, 1) \in A$ and it is easy to see that

$$\pi^{r_1,1}_{d(r_1),0} \circ \cdots \circ \pi^{1,1}_{d(1),0} \circ \pi^{1,r_2}_{0,e(r_2)} \circ \cdots \circ \pi^{1,1}_{0,e(1)}(i) = \pi^{r_1,1}_{d(r_1),0} \circ \cdots \circ \pi^{1,1}_{d(1),0}(tl) = 0.$$

□

Example 10 *Antithetic Variates and Randomly Shifted QMC Point Set.*

Let $\omega \equiv \mathbf{u}$, $\tilde{V}_2 = \{\mathbf{v}_0 = \mathbf{u}, \mathbf{v}_1 = 1 - \mathbf{u}\}$, and $\tilde{W}_{n/2}$ be a randomly shifted lattice rule (which satisfies Condition 1). Then for any $\mathbf{w}_i, \mathbf{w}_j \in \tilde{W}_{n/2}$, $(\mathbf{v}_0(\mathbf{w}_i), \mathbf{v}_0(\mathbf{w}_j))$ obviously has the same joint distribution as $(\mathbf{w}_i, \mathbf{w}_j)$. To show that $(\mathbf{v}_1(\mathbf{w}_i), \mathbf{v}_1(\mathbf{w}_j))$ has the same joint distribution as $(\mathbf{w}_j, \mathbf{w}_i)$, let $\mathbf{x}_i$ be such that $\mathbf{w}_i = \mathbf{x}_i + \mathbf{u}$ for $i = 0, \ldots, n/2 - 1$ (as seen in Example 3). We can write $(\mathbf{v}_1(\mathbf{w}_i), \mathbf{v}_1(\mathbf{w}_j)) = (1 - \mathbf{w}_i, 1 - \mathbf{w}_j) = (\mathbf{x}_j + \mathbf{u}', \mathbf{x}_i + \mathbf{u}')$, where $\mathbf{u}' = 1 - \mathbf{x}_i - \mathbf{x}_j - \mathbf{u}$. This pair has the same joint distribution as $(\mathbf{w}_j, \mathbf{w}_i)$ because $\mathbf{u}'$ has the same distribution as $\mathbf{u}$. Since $\tilde{V}_2$ satisfies Condition 2, Proposition 6 applies and thus $\tilde{P}_n$ satisfies the conditions of Proposition 1. This also works if the randomly shifted lattice rule $\tilde{W}_{n/2}$ is replaced by a digitally shifted net, with the $+$ operator defined in (15), and the corresponding $-$ operator in the definition of $\mathbf{v}_1$. □

6 Integrating GA with CMC

For each i, let $\mathcal{F}_i$ be the Borel σ-field generated by $\mathbf{u}_i$, let $\mathcal{G}_i$ be a sub-σ-field of $\mathcal{F}_i$, and define $Y_i = E[X_i \mid \mathcal{G}_i]$, which can also be written as $Y_i = g(\mathbf{u}_i)$ for some measurable function g. One has $E[Y_i] = E[X_i] = \mu$ and Y_i is a *conditional Monte Carlo* (CMC) estimator of μ. Define $C_i = Y_i - X_i$ and consider the estimator

$$\hat{\mu}_n = \sum_{i=0}^{n-1} \alpha_i Y_i - \sum_{i=0}^{n-1} \beta_i C_i, \tag{18}$$

which uses the C_i as control variates, and where $\alpha_0 + \cdots + \alpha_{n-1} = 1$.

If the $\mathbf{u}_i$'s satisfy the assumptions of Proposition 1, we know that the variance of $\hat{\mu}_n$ is minimized by taking $\alpha_i = 1/n$ and $\beta_i = b^*/n$ for all i, for some constant b^* that remains to be determined. With these values, (18) becomes $\hat{\mu}_n = \bar{Y}_n - b^*\bar{C}_n$ where $\bar{Y}_n = \sum_{i=0}^{n-1} Y_i/n$ and $\bar{C}_n = \sum_{i=0}^{n-1} C_i/n$, and it is easily seen that

$$b^* = \frac{\mathrm{Cov}[\bar{Y}_n, \bar{C}_n]}{\mathrm{Var}[\bar{C}_n]}. \tag{19}$$

It is also known that for the case where $n = 1$, one has $b^* = 0$ [12, Theorem 2], which means that the best strategy in this case is to use only Y_i as an estimator and forget about the control variate C_i. The idea is that Y_i has smaller variance than X_i and turns out to be uncorrelated with C_i, so it becomes useless to introduce C_i as a control variate. But for $n > 1$, the C_j are not always independent of the Y_i for $j \neq i$, so it might be worthwhile to have them as control variates. The next example illustrates this.

Example 11 Let $s = 2$, $X_i = f(\mathbf{u}_i) = f(u_{i,1}, u_{i,2}) = u_{i,1} + u_{i,2} + u_{i,1}u_{i,2}$, and $Y_i = E[X_i \mid u_{i,1}u_{i,2}]$. One has $Y_i = u_{i,1}u_{i,2} - 2(1 - u_{i,1}u_{i,2})/\ln(u_{i,1}u_{i,2})$, because $u_{i,1}|(u_{i,1}u_{i,2} = v)$ has density function $h(u) = -1/(u \ln v)$ for $u \geq v$ and $h(u) = 0$ elsewhere. Therefore, $C_i = Y_i - X_i = -2(1 - u_{i,1}u_{i,2})/\ln(u_{i,1}u_{i,2}) - u_{i,1} - u_{i,2}$. Suppose we use antithetic variates (AV), so $n = 2$ and $\mathbf{u}_1 = 1 - \mathbf{u}_0$. Denoting $\mathbf{u}_0 = (u_1, u_2)$, using AV alone gives the estimator

$$\begin{aligned} \bar{X}_2 &= (X_0 + X_1)/2 \\ &= 0.5[u_1 + u_2 + u_1u_2 + (1 - u_1) + (1 - u_2) + (1 - u_1)(1 - u_2)] \\ &= 1 + 0.5(u_1u_2 + (1 - u_1)(1 - u_2)), \end{aligned}$$

while using AV with CMC yields

$$\begin{aligned} \bar{Y}_2 &= (Y_0 + Y_1)/2 \\ &= 0.5\left[u_1u_2 - \frac{2(1 - u_1u_2)}{\ln(u_1u_2)} + (1 - u_1)(1 - u_2) - \frac{2(1 - (1 - u_1)(1 - u_2))}{\ln(1 - u_1)(1 - u_2)}\right] \\ &= 0.5[u_1u_2 + (1 - u_1)(1 - u_2)] - g(u_1, u_2), \end{aligned}$$

where

$$g(u_1, u_2) = \frac{(1 - u_1u_2)}{\ln(u_1u_2)} + \frac{(1 - (1 - u_1)(1 - u_2))}{\ln(1 - u_1)(1 - u_2)}.$$

Their difference is $\bar{C}_2 = (C_0 + C_1)/2 = -1 - g(u_1, u_2)$, and $\mathrm{Cov}[\bar{Y}_2, \bar{C}_2] = \mathrm{Var}[g(u_1, u_2)] - \mathrm{Cov}[g(u_1, u_2), u_1u_2 + (1-u_1)(1-u_2)]/2 = 0.021229 > 0$, which in turn implies that $b^* > 0$. With $\beta_i = b^*/2$, the variance of $\hat{\mu}_n$ turns out to be approximately 0.00215, whereas $\mathrm{Var}[\bar{Y}_2] \approx 0.03614$. Thus, using the control variate $\bar{C}_2$ with its optimal coefficient reduces the variance by a factor of more than 16 compared with using $\beta_1 = \beta_2 = 0$.

Suppose now that we condition on $u_{i,1}$ instead of $u_{i,1}u_{i,2}$. Then, $Y_i = E[X_i \mid u_{i,1}] = 1.5u_{i,1} + 0.5$, $C_i = Y_i - X_i = 0.5 + 0.5u_{i,1} - u_{i,2} - u_{i,1}u_{i,2}$, and since $\bar{Y}_2 = 1.25$, we get that $\mathrm{Cov}[\bar{Y}_n, \bar{C}_n] = 0$, so $b^* = 0$. □

For many of the schemes we have seen in Section 4, such as antithetic variates, randomly shifted lattice rules, and digitally shifted nets, for example, the randomness affects the points in a way that knowing *a single* $\mathbf{u}_i$ reveals enough information to determine the entire point set $\tilde{P}_n$. For instance, knowing $\mathbf{u}_i$ is enough to determine the shift for a randomly-shifted lattice rule, and then to determine all other points $\mathbf{u}_j$. For antithetic variates, knowing $\mathbf{u}$ tells us $1 - \mathbf{u}$.

In such a situation, the σ-fields $\mathcal{F}_i$ are all the same. Frequently, in this context, the $\mathcal{G}_i$ will also be all identical. For example, if Y_i can be written as $Y_i = E[X_i \mid Z_i]$ for some random variable (or vector) Z_i, and if $Z_i = \breve{g}(\breve{\mathbf{u}}_i)$ where $\breve{\mathbf{u}}_i$ represents a subset of the coordinates of $\mathbf{u}_i$ and $\breve{g}$ is a one-to-one transformation, then the $\mathcal{G}_i$'s turn out to be all the same. This implies that $E[X_i \mid \mathcal{G}_j] = E[X_i \mid \mathcal{G}_i] = Y_i$ almost surely (a.s.) for all i and j. The next proposition says that under the latter condition, the optimal strategy is to use the CMC estimator $\bar{Y}_n$ alone, without the control variates C_i.

Proposition 7 *If $E[X_i \mid \mathcal{G}_j] = E[X_i \mid \mathcal{G}_i]$ a.s. for all i and j, then $b^* = 0$.*

Proof. Under the assumption of the proposition, for any j, $\bar{Y}_n$ is $\mathcal{G}_j$-measurable and $E[\bar{X}_n \mid \mathcal{G}_j] = \bar{Y}_n$ a.s., so

$$\begin{aligned}
\mathrm{Cov}[\bar{Y}_n, \bar{C}_n] &= E[\bar{Y}_n\bar{C}_n] \\
&= E[\bar{Y}_n(\bar{Y}_n - \bar{X}_n)] \\
&= E[E[\bar{Y}_n(\bar{Y}_n - \bar{X}_n) \mid \mathcal{G}_j]] \\
&= E[E[\bar{Y}_n(\bar{Y}_n - \bar{Y}_n) \mid \mathcal{G}_j]] \\
&= 0.
\end{aligned}$$

Then, $b^* = 0$. □

Example 12 Let T be the length of the longest path between two given nodes (the origin and the destination) in an acyclic network with s arcs of random length. The aim is to estimate $\mu = P[T > c]$ for some constant c. Here, $X = I[T > c]$ where I is the indicator function. Suppose we generate the vector $\mathbf{V}_i = (V_{i,1}, \ldots, V_{i,s})$ of random arc lengths by using coordinate k of $\mathbf{u}_i$, $u_{i,k}$, to generate $V_{i,k}$ by inversion, for each i and k, and then compute the corresponding values of T_i and X_i. We assume that each arc length distribution has a density, which ensures that the transformation from $u_{i,k}$ to $V_{i,k}$ is invertible.

A CMC estimator for this example can be defined as follows (see, e.g., [3, 17]). Select a set of arcs $\mathcal{L} \subset \{1, \ldots, s\}$ and let $\mathcal{G}_i$ be the σ-field generated by $\{V_{i,k},\, k \notin \mathcal{L}\}$. That is, $Y_i = E[X \mid \mathcal{G}_i] = P[T > c \mid \mathcal{G}_i]$ is a probability conditional on the lengths of all arcs not in $\mathcal{L}$. If $\mathcal{L}$ is chosen so that each path from the origin to the destination contains exactly one arc from $\mathcal{L}$, then this conditional probability is easy to compute [3]. Here, if all $\mathbf{u}_j$ can be recovered from any single $\mathbf{u}_i$, then $\{V_{j,k},\, k \notin \mathcal{L}\}$ can also be recovered from $\{V_{i,k},\, k \notin \mathcal{L}\}$ from each i, j, and therefore the conditions of Proposition 7 are satisfied. This means that if the $\mathbf{u}_i$'s are obtained by AV, or randomly shifted lattice rules, or digitally shifted nets, for example, then $b^* = 0$. □

7 Conclusion

In this work, we have studied estimators based on GA transformations combined with control variates. Our goal was to determine under what conditions using equal weights across all evaluation points minimizes the variance, for a given GA scheme, when the weights must be chosen *a priori*. In Proposition 1, we have obtained general sufficient conditions for the optimality of equal weights. We then provided three different easily verifiable conditions under which Proposition 1 can be applied. For several correlation induction techniques, we have been able to verify these conditions and thus show that equal weights are optimal.

Other questions of interest include: By how much can we reduce the variance if we are allowed to choose the weights *a posteriori*, i.e., after the evaluation points have been observed, and how does the improvement behave as a function of the number of points n and the dimension s? What are the interesting families of GA transformations for which choosing unequal weights is optimal? How does the variance of estimators obtained from these families, with optimal weights, compare with that of estimators for which equal weights prevail?

A different issue, perhaps more important but certainly more difficult to address, is to determine what types of GA transformations provide the largest variance reduction, or the largest efficiency improvement factor if we take the computing cost of the estimator into account, for given classes of functions. The answer obviously depends very much on the functions f we would consider. Further work in that direction is certainly needed.

Acknowledgments

This work has been supported by NSERC-Canada grants to each of the three authors, and by a NATEQ-Québec grant and a Killam Research Fellowship to the second author. We wish to thank Prof. Germund Dahlquist for providing a copy of [2]. We also thank the anonymous referees and the associate editor for their helpful comments.

References

[1] I. J. Andréasson. Combinations of antithetic methods in simulation. Technical Report NA 72.49, Royal Institute of Technology, 1972.

[2] I. J. Andréasson and G. Dahlquist. Groups of antithetic transformations in simulation. Technical Report NA 72.57, Royal Institute of Technology, 1972.

[3] A. N. Avramidis and J. R. Wilson. Integrated variance reduction strategies for simulation. *Operations Research*, 44:327–346, 1996.

[4] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation.* Springer-Verlag, New York, second edition, 1987.

[5] R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914, 1976.

[6] T. J. DiCiccio and P. W. Glynn. On the value of function evaluation location information in Monte Carlo simulation. *Management Science*, 41(4):733–737, 1995.

[7] D. S. Dummit and R. M. Foote. *Abstract Algebra.* John Wiley and Sons, second edition, 1999.

[8] H. Faure. Discrépance des suites associées à un système de numération. *Acta Arithmetica*, 61:337–351, 1982.

[9] G. S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications.* Springer Series in Operations Research. Springer-Verlag, New York, 1996.

[10] G. S. Fishman and B. D. Wang. Antithetic variates revisited. *Communications of the ACM*, 26:964–971, 1983.

[11] P. W. Glynn. Efficiency improvement techniques. *Annals of Operations Research*, 53:175–197, 1994.

[12] P. W. Glynn and R. Szechtman. Some new perspectives on the method of control variates. In K.-T. Fang, F. J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pages 27–49, Berlin, 2002. Springer-Verlag.

[13] B. L. Granovsky. Optimal variance reduction theorem in simulation by the Monte Carlo method. *Österreichische Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse, Sitzungsberichte, Abt. II, Mathematische, Physikalische und Technische Wissenchaften*, 192(8–10):329–335, 1983.

[14] F. J. Hickernell, C. Lemieux, and A. B. Owen. Control variates for quasi-Monte Carlo. Submitted for publication, 2002.

[15] H. S. Hong and F. J. Hickernell. Algorithm 823: Implementing scrambled digital sequences. *ACM Transactions on Mathematical Software*, 29:95–109, 2003.

[16] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, third edition, 2000.

[17] P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. *Management Science*, 46(9):1214–1235, 2000.

[18] P. L'Ecuyer and C. Lemieux. Recent advances in randomized quasi-Monte Carlo methods. In M. Dror, P. L'Ecuyer, and F. Szidarovszki, editors, *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, pages 419–474. Kluwer Academic Publishers, Boston, 2002.

[19] J. Matoušek. On the L_2-discrepancy for anchored boxes. *Journal of Complexity*, 14:527–556, 1998.

[20] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[21] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1992.

[22] A. B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, number 106 in Lecture Notes in Statistics, pages 299–317. Springer-Verlag, 1995.

[23] A. B. Owen. Latin supercube sampling for very high-dimensional simulations. *ACM Transactions of Modeling and Computer Simulation*, 8(1):71–102, 1998.

[24] I. H. Sloan and S. Joe. *Lattice Methods for Multiple Integration*. Clarendon Press, Oxford, 1994.

[25] S. Tezuka. *Uniform Random Numbers: Theory and Practice*. Kluwer Academic Publishers, Norwell, Mass., 1995.

[26] J. W. Tukey. Antithesis or regression? *Proceedings of the Cambridge Philosophical Society*, 54:300–301, 1957.

[27] J. R. Wilson. Antithetic sampling with multivariate inputs. *American Journal of Mathematical and Management Sciences*, 3:121–144, 1983.

[28] S. Yakowitz, J. E. Krimmel, and F. Szidarovszky. Weighted Monte Carlo integration. *SIAM Journal on Numerical Analysis*, 15:1289–1300, 1978.