Methodology and Computing in Applied Probability (2022) 24:1143–1168
https://doi.org/10.1007/s11009-021-09893-3

# On Accelerating Monte Carlo Integration Using Orthogonal Projections

**Huei-Wen Teng**[1] **· Ming-Hsuan Kang**[2]

Received: 31 May 2021 / Revised: 4 August 2021 / Accepted: 19 August 2021 /
Published online: 4 October 2021


## Abstract

Monte Carlo simulation is an indispensable tool in calculating high-dimensional integrals. Although Monte Carlo integration is notoriously known for its slow convergence, it could be improved by various variance reduction techniques. This paper applies orthogonal projections to study the amount of variance reduction, and also proposes a novel *projection estimator* that is associated with a group of symmetries of the probability measure. For a given space of functions, the average variance reduction can be derived. For a specific function, its variance reduction is also analyzed. The well-known antithetic estimator is a special case of the projection estimator, and new results of its variance reduction and efficiency are provided. Various illustrations including pricing financial Asian options are provided to confirm our claims.

**Keywords** Monte Carlo integration · Group · Orthogonal projection · Symmetry · Variance reduction · Financial option pricing

**Mathematics Subject Classification (2010)** 65C05 · 91G60 · 91-08

We are benefited from the very helpful comments of the Editor and two anonymous referees. The first author was supported by the Ministry of Science and Technology of Taiwan, ROC, under Grant 108-2118-M-009-001-MY2, and the second author was supported under Grant 108-2115-M-009-007-MY2.

✉ Ming-Hsuan Kang
kmsming@gmail.com

Huei-Wen Teng
venteng@gmail.com

[1] Department of Information Management and Finance, National Yang Ming Chiao Tung University, University Road 1001, Hsinchu City, Taiwan, Republic of China

[2] Department of Applied Mathematics, National Yang Ming Chiao Tung University, University Road 1001, Hsinchu City, Taiwan, Republic of China

## 1 Introduction

Accurate and efficient calculation of high-dimensional integrals is of considerable importance in various scientific disciplines, such as statistics, engineering, and finance (Asmussen and Glynn 2007; Glasserman 2004; Ross 2013). Monte Carlo simulation is an indispensable tool for calculating them. Although Monte Carlo simulation can avoid the curse of dimensionality, it is notoriously known for its slow convergence. To tackle this problem, various variance reduction techniques have been proposed to improve the efficiency of the standard Monte Carlo estimator.

Variance reduction techniques, such as control variables, importance sampling, conditional Monte Carlo, need additional mathematical treatments for a specific problem, and substantial variance reduction could be obtained. Other techniques, such as the antithetic variates estimator, Latin hypercube sampling, moment matching, stratified sampling, systematic sampling, spherical Monte Carlo, and quasi Monte Carlo, are generally applicable and do not require additional mathematical analysis. They seek procedures to produce samples appearing to be more evenly distributed in the desired space. Recently, it becomes popular to combine two or more variance reduction techniques in a specific application. For example, in pricing financial options, Glasserman et al. combine importance sampling and stratification (Glasserman et al. 1999), and Neddermeyer combines importance sampling and quasi Monte Carlo method (Neddermeyer 2011).

For variance reduction techniques, the amount of variance reduction could be dissatisfied and difficult to analyze in advance. To overcome these problems, we provide a general framework using orthogonal projections. To start, we review orthogonal projections in linear algebra. On the space of square integrable functions, we regard the expectation as an orthogonal projection. Consequently, the variance and squared expectation sum up to the squared norm of the function. Then, we show that an orthogonal projection induces an unbiased estimator which has a smaller variance than the standard estimator, if its image consists of constant functions. However, how to construct an explicit orthogonal projection of a function poses a challenge to this idea.

To conquer this challenge, we provide a feasible solution: We propose the *projection estimator* using a group of symmetries of the probability space. The sophisticated properties of a group enable us to calculate the average variance ratio of the projection estimator to the standard estimator on a given finite-dimensional vector space. This result helps us to determine if a projection estimaftor has a smaller variance on average than the standard estimator. A good introduction of group theory in abstract algebra can be found, for example, in Fraleigh (2019).

Furthermore, to analyze the variance reduction of a projection estimator for a specific function, we decompose the function as the sum of its average function and difference function using the concept of fundamental domains. We show that the variance reduction is not less than the variance of the average function. Doing so guides us in selecting the group of symmetries for a specific problem.

Indeed, through the idea of a group of symmetries, the projection estimator is a generalization of the well-known antithetic variates estimator. Our framework allows to provide new and finer analysis about the variance reduction and efficiency of the antithetic variate estimator. As a note, extensions to the antithetic variate estimator from different angels can be found in Park and Choe (2016) and Ren et al. (2019), for example.

We illustrate our method in a toy example using an indicator function of two different region to contrast the antithetic variates estimator and the projection estimator. Also, we apply our method to a classical problem in financial engineering: pricing Asian options under the GARCH model.

The rest of this paper is organized as follows. Section 2 reviews the orthogonal projection and provides a variance reduction technique using orthogonal projection. Section 3 proposes a projection estimator using the group of symmetries on the probability measure. With the projection estimator, Section 4 calculates the average variance ratio for a given space of functions, and Section 5 analyzes the variance reduction for a specific function. The last section concludes. All proofs are deferred to the Appendix.

## 2 Preliminaries

Let $V$ be a vector space endowed with an inner product $\langle \cdot, \cdot \rangle$. Let $\| \cdot \|$ denote the norm on $V$ given by $\|v\| = \sqrt{\langle v, v \rangle}$. A linear transform $P$ on $V$ is an orthogonal projection if it satisfies the following two conditions:

C1. $P^2 = P$.
C2. For any $v, w \in V$, $\langle P(v), w - P(w) \rangle = 0$.

Figure 1 illustrates the geometry for $v$, $P(v)$, and $v - P(v)$ to illustrate the idea of orthogonality. Consequently, we have the Pythagorean Theorem for orthogonal projections:

$$\|v\|^2 = \|P(v)\|^2 + \|v - P(v)\|^2.$$

Denote the $d$-dimensional real vector space by $\mathbb{R}^d$. Consider a $d$-dimensional real-valued random vector $X = (X_1, \ldots, X_d)$ defined on the probability space $(\Omega, \sigma(\Omega), \mu)$. Then, $X$ induces a measure $\mu_X$ on $\mathbb{R}^d$, such that for any integrable function $f$ on $\mathbb{R}^d$, we have

$$\int_\Omega f(X(\omega))d\mu(\omega) = \int_{\mathbb{R}^d} f(x)d\mu_X(x).$$

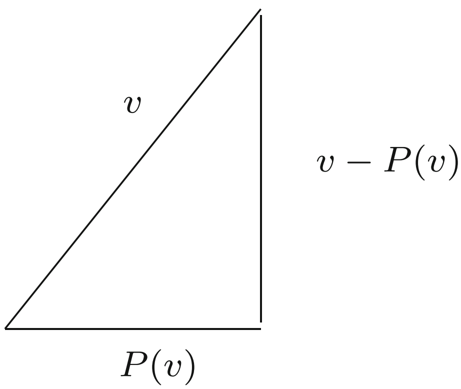
Let $a := b$ abbreviates that $a$ is defined as $b$. Recall that the expectation of $f(X)$ is defined as

$$\mathbb{E}\big(f(X)\big) := \int_{\mathbb{R}^d} f(x)d\mu_X(x),$$

and the variance of $f(X)$ is defined as

$$\mathrm{var}\big(f(X)\big) := \int_{\mathbb{R}^d} \big(f(x) - \mathbb{E}\big(f(X)\big)\big)^2 d\mu_X(x).$$

**Fig. 1** The relation of $v$, $P(v)$, and $v - P(v)$



Let $\mathcal{F}$ be the set of real valued functions on $\mathbb{R}^d$ so that, for $f(x) \in \mathcal{F}$, both $\mathbb{E}\big(f(X)\big)$ and $\mathrm{var}\big(f(X)\big)$ are well-defined. Consider the standard inner product on $\mathcal{F}$ given by

$$\langle f_1, f_2 \rangle = \int_{\mathbb{R}^d} f_1(x) f_2(x) d\mu_X(x),$$

which induces a norm on $\mathcal{F}$ by

$$\|f\|^2 = \langle f, f \rangle.$$

Then, $\mathbb{E}\big(f(X)\big)$ can be represented as the inner product of $f$ and the constant function **1**,

$$\mathbb{E}\big(f(X)\big) = \langle f, \mathbf{1} \rangle.$$

In addition, the expectation defines a linear transform from $\mathcal{F}$ to $\mathbb{R}$. Then, we have the following lemma.

**Lemma 1** *The expectation defines an orthogonal projection $P_{\mathbb{E}}$ given by*

$$P_{\mathbb{E}}(f) := \mathbb{E}\big(f(X)\big).$$

*Proof* See Appendix A. □

Similarly, $\mathrm{var}\big(f(X)\big)$ can be represented as the squared norm of $f(x) - P_{\mathbb{E}}\big(f(X)\big)$,

$$\mathrm{var}\big(f(X)\big) = \langle f - P_{\mathbb{E}}\big(f(X)\big), f - P_{\mathbb{E}}\big(f(X)\big) \rangle = \|f - P_{\mathbb{E}}\big(f(X)\big)\|^2.$$

Recall that Lemma 1 ascertains that the expectation is an orthogonal projection. Applying the Pythagorean theorem, we obtain

$$\|f\|^2 = \|P_{\mathbb{E}}\big(f(X)\big)\|^2 + \|f - P_{\mathbb{E}}\big(f(X)\big)\|^2 = \mathbb{E}\big(f(X)\big)^2 + \mathrm{var}\big(f(X)\big). \tag{1}$$

Equation 1 shows that the squared norm of $f$ equals the sum of the squared expectation of $f(X)$ and the variance of $f(X)$. It also provides a useful formula to calculate the variance of $f(X)$,

$$\mathrm{var}\big(f(X)\big) = \|f\|^2 - \mathbb{E}\big(f(X)\big)^2. \tag{2}$$

The relation of $\|f\|$, $\mathbb{E}\big(f(X)\big)$, and $\mathrm{var}\big(f(X)\big)$ is demonstrated in Fig. 2.

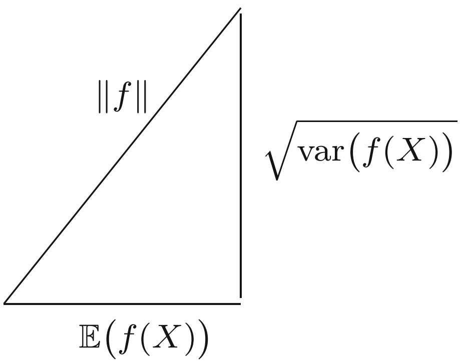
For notational simplicity, we denote the standard estimator by

$$f := f(X).$$

Consider an orthogonal projection $P$. We denote the orthogonal projection of a function $f(x)$ by

$$P(f)(x),$$

**Fig. 2** The relation of $\|f\|$, $\mathbb{E}\big(f(X)\big)$, and $\mathrm{var}\big(f(X)\big)$



which is also a function. We now define the projection estimator associated with the orthogonal projection $P$ as

$$P(f) := P(f)(X).$$

**Theorem 1** *Let $P$ be an orthogonal projection on $\mathcal{F}$ such that $P(\mathcal{F})$ contains constant functions. Then for $f(x) \in \mathcal{F}$, the projection estimator $P(f)$ is an unbiased estimator of $\mathbb{E}\big(f(X)\big)$ and has a smaller variance than the standard estimator $f$. Specifically, we have*

$$\mathbb{E}\big(P(f)(X)\big) = \mathbb{E}\big(f(X)\big),$$

*and*

$$\operatorname{var}\big(P(f)(X)\big) \leq \operatorname{var}\big(f(X)\big).$$

*Proof* See Appendix B. □

Theorem 1 allows us to propose the projection estimator as a new category of variance reduction techniques. However, it remains unclear how to explicitly construct a projection estimator through orthogonal projection. In the next section, we introduce a systematic approach to construct a projection estimator using a group of symmetries of the probability space.

## 3 The Projection Estimator

We introduce symmetries of the probability space in Section 3.1, a group of symmetries $G$ in Section 3.2, and the projection estimator $P_G(f)$ in Section 3.3.

### 3.1 A Symmetry of the Probability Space

A symmetry $g$ of the probability space $(\mathbb{R}^d, \mu_X)$ is a bijection from $\mathbb{R}^d$ to itself, which preserves its structure of probability space. Abbreviate

$$gx := g(x) \text{ and } gD := \{gx : x \in D\}$$

for a subset $D$ of $\mathbb{R}^d$. More precisely, a symmetry $g$ needs to satisfy

$$\mu_X(gD) = \mu_X(D),$$

for any measurable subset $D$ of $\mathbb{R}^d$. This condition is equivalent to

$$d\mu_X(gx) = d\mu_X(x),$$

for any $x \in \mathbb{R}^d$.

For illustration, let $N_d(\mu, \Sigma)$ denote the $d$-dimensional normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$. Let $\mathbf{0}_d$ and $I_d$ denote the zero vector and identity matrix of size $d$, respectively. Then, $N_d(\mathbf{0}_d, I_d)$ denotes the $d$-dimensional standard normal distribution. Let $\sim$ to indicate "has the distribution of", and the superscript $t$ to denote vector or matrix transpose. Let $\|x\| = \sqrt{x^t x}$ denote the Euclidean norm of $x$.

Consider $X \sim N_d(\mathbf{0}_d, I_d)$. Then, we have

$$d\mu_X(x) = \frac{1}{(2\pi)^{d/2}} e^{-\|x\|^2/2}.$$

To describe the symmetries of $(\mathbb{R}^d, \mu_X)$, we recall that a $d \times d$ real matrix $A$ is an orthogonal matrix if $AA^t = I_d$. In this case, we have $\|Ax\|^2 = \|x\|^2$ for all $x \in \mathbb{R}^d$. As a result, each orthogonal matrix $A$ defines a symmetry of $(\mathbb{R}^d, \mu_X)$, because it satisfies the equality,

$$d\mu_X(Ax) = \frac{1}{(2\pi)^{d/2}} e^{-\|Ax\|^2/2} = \frac{1}{(2\pi)^{d/2}} e^{-\|x\|^2/2} = d\mu_X(x).$$

When $d = 2$, there are two types of orthogonal matrices, namely rotations and reflections. When $d > 2$, an orthogonal matrix $A$ of size $d$ is called a rotation (respectively, reflection) if there exists a 2-dimensional subspace $W$ of $\mathbb{R}^d$ such that the restriction of $A$ on $W$ is a rotation (respectively, reflection), and the restriction of $A$ on the orthogonal complement of $W$ is the identity matrix of size $(d - 2)$. Moreover, a general orthogonal matrix is a product of rotations and reflections.

An explicit example of symmetries for $d = 2$ is given as follows. Let $g(x_1, x_2)$ denote the clockwise rotation of 90° around the origin given by

$$g(x_1, x_2) = (x_2, -x_1),$$

which is indeed corresponding to the orthogonal matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then, $g(x_1, x_2)$ is a symmetry of $(\mathbb{R}^d, \mu_X)$. In addition, the powers of $g$:

$$\begin{aligned} g^0(x_1, x_2) &= (x_1, x_2), \\ g^2(x_1, x_2) &= (-x_1, -x_2), \\ g^3(x_1, x_2) &= (-x_2, x_1). \end{aligned}$$

are also symmetries of $(\mathbb{R}^2, \mu_X)$, because they are all rotations. Figure 3 depicts how these four symmetries act on a point $(x_1, x_2)$ in $\mathbb{R}^2$.

### 3.2 A Group of Symmetries

An abstract group is a set equipped with an operation that combines any two elements to form a third element while being associative as well as having an identity element and inverse elements. We denote the cardinality of $G$ by $|G|$.

One important example of groups is the group of symmetries. For instance, consider $\mu_X \sim N_d(\mathbf{0}_d, I_d)$. Let $O_d$ be the set of all orthogonal matrices of size $d$. It is known that $O_d$ is a group (called the orthogonal group) under matrix multiplication and it contains all symmetries of $(\mathbb{R}^d, \mu_X)$. In this case, a group of symmetries $G$ of $(\mathbb{R}^d, \mu_X)$ is a subset of $O_d$ such that $G$ is a group under matrix multiplication. It is easy to check if a finite subset of $O_d$ forms a group by the following criterion.

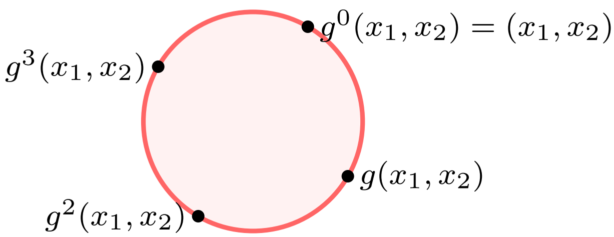



**Fig. 3** An illustration of how symmetries $g^0(x_1, x_2) = (x_1, x_2)$, $g(x_1, x_2) = (x_2, -x_1)$, $g^2(x_1, x_2) = (-x_1, -x_2)$, and $g^3(x_1, x_2) = (-x_2, x_1)$, act on a point $(x_1, x_2)$ in $\mathbb{R}^2$ with $\mu_X \sim N_2(\mathbf{0}_d, I_d)$

*A finite subset $G$ of $O_d$ forms a group if it is closed under multiplication, i.e. for all $g_1, g_2 \in G$, we have $g_1 g_2 \in G$.*

A simple way to construct a such group $G$ is to choose an orthogonal matrix $A$ and consider the subgroup generated by $A$, namely

$$\langle A \rangle := \{A^k | k \in \mathbb{Z}\}.$$

Note that the above set is finite when $A^k = I_d$ for some positive integer $k$. In this case, we have

$$\langle A \rangle = \{I_d, A, A^2, \cdots, A^{k-1}\}.$$

In general, one can select several orthogonal matrices and consider the subgroup generated by them. We provide some examples of the group $G$ as follows.

1. The trivial group is
$$G_0 = \{g_1(x) := x\} = \langle g_1 \rangle,$$
which consists of only the identity element.
2. The group of antithetic variate is
$$G = \{g_1(x) := x, g_2(x) := -x\} = \langle g_2 \rangle.$$
We will connect this group to the antithetic variate estimator soon in Example 2 of Section 3.3.
3. The group of coordinate cyclic permuting is
$$G = \{g^k | k = 0, 1, \cdots d\} = \langle g \rangle,$$
where
$$g(x_1, \cdots, x_d) := (x_d, x_1, \cdots, x_{d-1}).$$
There are a total of $d$ elements in $G$.
4. The group of full coordinate permuting consists of all permutation on coordinates. There are a total of $d!$ such permutations.
5. The group of coordinate sign changing consists all elements of the form:
$$g(x_1, \cdots, x_d) := (\pm x_1, \cdots, \pm x_d).$$
There are a total of $2^d$ such elements.

### 3.3 The Projection Estimator $P_G(f)$

For the rest of the paper, we denote $G$ as a finite group of symmetries of $(\mathbb{R}^d, \mu_X)$. For any $f \in \mathcal{F}$, we define the function $f_g$ associated with a given $g \in G$ by

$$f_g(x) := f(gx) = f(g(x)).$$

Then, for $g, g' \in G$, we have

$$(f_g)_{g'}(x) = (f_g)(g'x) = f(g'gx) = f_{g'g}(x).$$

Recall that $\mathcal{F}$ is the set of real valued functions on $\mathbb{R}^d$ so that for $f(x) \in \mathcal{F}$, both $\mathbb{E}\big(f(X)\big)$ and $\mathrm{var}\big(f(X)\big)$ are well-defined.

**Lemma 2** *For any $f \in \mathcal{F}$ and $g \in G$, $f_g(X)$ is an unbiased estimator, and $f_g$ remains in $\mathcal{F}$.*

*Proof* See Appendix C.

Since $g$ is a symmetry of $\mu_X$, we have

$$\mathbb{E}\big(f_g(X)\big) = \int_{\mathbb{R}^n} f(gx)d\mu_X(x) = \int_{\mathbb{R}^n} f(y)d\mu_X(g^{-1}y) = \int_{\mathbb{R}^n} f(y)d\mu_X(y) = \mathbb{E}\big(f(X)\big).$$

Hence, $f_g(X)$ is an unbiased estimator. By the same token, we also have $\mathbb{E}\big(f_g(X)^2\big) = \mathbb{E}\big(f(X)^2\big)$. Now we have

$$\mathrm{var}\big(f_g(X)\big) = \mathbb{E}\big((f_g(X))^2\big) - \mathbb{E}\big(f_g(X)\big)^2 = \mathbb{E}\big(f(X)^2\big) - \mathbb{E}\big(f(X)\big)^2 = \mathrm{var}\big(f(X)\big).$$

Because both $\mathbb{E}\big(f_g(X)\big)$ and $\mathrm{var}\big(f_g(X)\big)$ are well-defined, $f_g$ remains in $\mathcal{F}$. □

Now, we are ready to define the orthogonal projection of a function $f(x)$ associated with $G$ as

$$P_G(f)(x) := \frac{1}{|G|}\sum_{g\in G} f_g(x).$$

Because $P_G(f)(x)$ remains a function, we define the *projection estimator* associated with $G$ as

$$P_G(f) := \frac{1}{|G|}\sum_{g\in G} f_g(X).$$

*Example 1* With $G = \{g_1(x) := x\}$, $P_G(f) = f(X)$ is the standard estimator.

*Example 2* Suppose $X \sim N_d(\mathbf{0}_d, I_d)$. With $G = \{g_1(x) := x, g_2(x) := -x\}$,

$$P_G(f) = \frac{1}{2}\left(f(X) + f(-X)\right)$$

is exactly the well-known antithetic variable estimator. For notational ease, we define the antithetic variates estimator by

$$f_{AT} := \frac{1}{2}\left(f(X) + f(-X)\right).$$

Lemma 2 ensures that $f_g$ is an unbiased estimator for all $g \in G$, and hence the projection estimator $P_G(f)$ is also an unbiased estimator. Moreover, we have the following theorem.

**Theorem 2** *The projection $P_G$ is an orthogonal projection and $P_G(\mathcal{F})$ contains constant functions.*

*Proof* See Appendix D. □

Applying Theorems 1 and 2, we immediately obtain the following result.

**Corollary 1** *The projection estimator $P_G(f)$ is an unbiased estimator and has a smaller variance than the standard estimator $f$.*

## 4 The Average Variance Reduction

To understand the projection estimator, we first review the variance and efficiency of an estimator in Section 4.1. Then, we study the average variance and efficiency of a given finite-dimensional subspaces of $\mathcal{F}$ in Section 4.2. We provide an example to illuminate the idea of the average variance and efficiency in Section 4.3.

### 4.1 The Variance and Efficiency

Following (L'Ecuyer 1994), we define the efficiency to take into account both the variance and computing cost to implement an estimator. The efficiency of an estimator $h(X)$ is

$$\text{eff}(h(X)) = \frac{1}{\text{var}(h(X))\, c(h(X))}, \tag{3}$$

where $c(h(X))$ is the computing cost of evaluating $h(x)$. An estimator having a smaller variance or computing cost yields a larger value of efficiency. In practice, an estimator with a higher efficiency is regarded as more efficient and is preferred.

To compare different projection estimators, a simple proxy for the computing cost is $|G|$, because if we set the same sample size for each estimator, the computing time mainly differs in $|G|$. We hence define the efficiency by cardinality for a projection estimator as

$$\text{eff}_c(P_G(f)) = \frac{1}{\text{var}(P_G(f))|G|}. \tag{4}$$

On the other hand, recall Example 1 of Section 3.3 that the standard estimator $f$ is a special case of the projection estimator. As a result, the efficiency by cardinality of the standard estimator is

$$\text{eff}_c(f) = \frac{1}{\text{var}(f)}.$$

However, a more realistic alternative for the computing cost is the computing time of implementing an estimator. Thus, we define the efficiency by computing time as

$$\text{eff}_t(h(X)) = \frac{1}{\text{var}(h(X))\text{time}(h(X))}, \tag{5}$$

where $\text{time}(h(X))$ is the computing time of implementing an estimator. The computing time is usually involved with the complexity of the functions, the architecture of the algorithm, and the capability of the computer. If we split the computing time into a sum of the sampling time and estimating time: The former refers to the computing time in generating independent realizations, whereas the later refers to the computing time in calculating multiple function values and taking an average of them. We note that we will provide the efficiency by computing time in a real application of pricing financial options under GARCH model in Section 5.3. But, for tractable mathematical analysis, we only focus on the efficiency by cardinality for the rest of this section.

A projection estimator $P_G(f)$ is said to be more efficient by cardinality than the standard estimator $f$, if and only if,

$$\text{eff}_c(P_G(f)(X)) > \text{eff}_c(f(X)).$$

Plugging the definitions of efficiency by cardinality, the above inequality equals

$$\frac{1}{\text{var}\big(P_G(f)(X)\big)|G|} > \frac{1}{\text{var}\big(f(X)\big)},$$

or equivalently,

$$\frac{\text{var}\big(f(X)\big)}{\text{var}\big(P_G(f)(X)\big)} > |G|.$$

The left-hand-side of the above inequality implies that a projection estimator is more efficient by cardinality than the standard estimator only when its variance is $|G|$ times less than that of the standard estimator.

Let $\mathcal{G}$ be an $n$-dimensional subspace of $\mathcal{F}$ containing constant functions. The space $\mathcal{G}$ is called $G$-invariant, if for any $f \in \mathcal{G}$, we have $f(gx) \in \mathcal{G}$. In this case, $P_G(\mathcal{G})$ is a subspace of $\mathcal{G}$. Suppose that $P_G(\mathcal{G})$ is of $m$ dimension. Fix an orthonormal basis $\alpha = \{f_1, \cdots, f_n\}$ of $\mathcal{G}$, such that $f_1$ is equal to the constant function $\mathbf{1}$, and the first $m$ functions span $P_G(\mathcal{G})$. Denote the coordinate vector of $f$ under the basis $\alpha$ by $[f]_\alpha = (x_1, \cdots, x_n)$. Then, we can calculate $\mathbb{E}\big(f(X)\big)$, $\mathrm{var}\big(f(X)\big)$, and $\mathrm{var}\big(P_G(f)(X)\big)$ easily using $[f]_\alpha$.

First, $\mathbb{E}\big(f(X)\big)$ can be written as

$$\mathbb{E}\big(f(X)\big) = \langle f, 1\rangle = x_1.$$

Because $f \in \mathcal{F}$, we obtain $\mathrm{var}\big(f(X)\big)$ as

$$\mathrm{var}\big(f(X)\big) = \langle f^2, 1\rangle - \big(\mathbb{E}\big(f(X)\big)\big)^2 = \sum_{i=1}^{n} x_i^2 - x_1^2 = \sum_{i=2}^{n} x_i^2, \tag{6}$$

where the summation is from $i = 2$ (rather than from $i = 1$). Similarly, $\mathrm{var}\big(P_G(f)(X)\big)$ equals

$$\mathrm{var}\big(P_G(f)(X)\big) = \sum_{i=2}^{m} x_i^2. \tag{7}$$

We recall that $n$ and $m$ denote the dimensions of $\mathcal{G}$ and $P_G(\mathcal{G})$, respectively.

## 4.2 The Average Variance and Efficiency Ratio

Define the set of $\mathcal{G}$ where $P_G(f)$ is more efficient by cardinality than $f$ by

$$\mathcal{A} = \{f \in \mathcal{G} : \mathrm{eff}_c(P_G(f)(X)) > \mathrm{eff}_c(f(X))\}.$$

With Eqs. 6 and 7, we simplify $\mathcal{A}$ as

$$\begin{aligned}\mathcal{A} &= \{f \in \mathcal{G} : \frac{\mathrm{var}\big(f(X)\big)}{\mathrm{var}(P_G(f)(X))} > |G|.\} \\ &= \left\{(x_1, \cdots, x_n) \in \mathbb{R}^n : \frac{1}{|G|}\left(\sum_{i=2}^{n} x_i^2\right) > \left(\sum_{i=2}^{m} x_i^2\right)\right\}.\end{aligned}$$

This simplification indicates that whether a projection estimator is more efficient than the standard estimator depends on the function $f$ through the coefficients $x_1, \ldots, x_n$.

Now let us define the average variance. First, let $B(r)$ denote the ball of radius $r > 0$ centred at the origin in $\mathcal{G}$. Consider the standard Lebesgue measure on $\mathcal{G} \cong \mathbb{R}^n$. Define the average variance of functions in $\mathcal{F}$ on $B(r)$ as

$$\overline{\mathrm{var}}_r(\mathcal{G}) := \frac{1}{\mathrm{Vol}(B(r))} \int_{B(r)} \mathrm{var}\big(f\big) df.$$

Here, $\mathrm{Vol}(B(r))$ is the volume of the ball $B(r)$. Likewise, define the average variance of the projection estimator as

$$\overline{\mathrm{var}}_r(P_G(\mathcal{G})) := \frac{1}{\mathrm{Vol}(B(r))} \int_{B(r)} \mathrm{var}\big(P_G(f)\big) df.$$

Given a fixed basis $\alpha$, we obtain

$$\overline{\mathrm{var}}_r(\mathcal{G}) = \frac{1}{\mathrm{Vol}(B(r))}\int_{B(r)}\sum_{i=2}^{n} x_i^2 dx = \frac{n-1}{\mathrm{Vol}(B(r))}\int_{B(r)} x_1^2 dx, \tag{8}$$

where the last identity follows by the symmetry of the ball. Similarly, we have

$$\overline{\mathrm{var}}_r(P_G(\mathcal{G})) = \frac{m-1}{\mathrm{Vol}(B(r))}\int_{B(r)} x_1^2 dx. \tag{9}$$

With Eqs. 8 and 9, the average variance ratio of $P_G(f)$ to $f$ on $B(r)$ is independent of $r$:

$$\frac{\overline{\mathrm{var}}_r(P_G(\mathcal{G}))}{\overline{\mathrm{var}}_r(\mathcal{G})} = \frac{m-1}{n-1}.$$

Let the radius approach to infinity to define the average variance ratio of $P_G(f)$ to $f$,

$$\overline{\mathcal{V}}_G := \frac{\overline{\mathrm{var}}(P_G(\mathcal{G}))}{\overline{\mathrm{var}}(\mathcal{G})} = \lim_{r\to\infty}\frac{\overline{\mathrm{var}}_r(P_G(\mathcal{G}))}{\overline{\mathrm{var}}_r(\mathcal{G})} = \frac{m-1}{n-1}.$$

If the average variance ratio $\overline{\mathcal{V}}_G < 1$, we say that the projection estimator has an on average smaller variance than the standard estimator. Note that the average variance ratio only depends on $m$ and $n$. While $n$ is predetermined in the original problem, the selection of $G$ is critical because it directly affects $m$. In other words, if we can construct a group $G$ so that the dimension of $P_G(\mathcal{G})$ is small, the associated projection estimator would has an on average smaller variance than the standard estimator.

Now, we are ready to define the average efficiency by cardinality of an estimator $f(X)$, denoted by $\overline{\mathrm{eff}}_c(f(X))$, by simply replacing the variance with the average variance. We say that a projection estimator is on average more efficient than the standard estimator, if and only if,

$$\overline{\mathrm{eff}}_c\big(P_G(f)(X)\big) > \overline{\mathrm{eff}}_c(f(X)).$$

**Theorem 3** *Let $\mathcal{G}$ be a $G$-invariant subspace $\mathcal{G}$ of $\mathcal{F}$ containing constant functions. The projection estimator $P_G(f)$ is on average more efficient than the standard estimator, if and only if,*

$$\frac{(n-1)}{(m-1)|G|} > 1.$$

*Here, $n = \dim\mathcal{G}$ and $m = \dim P_G(\mathcal{G})$.*

Theorem 3 implies that to have an on average more efficient project estimator $P_G(f)$, both $|G|$ and the dimension of $P_G(\mathcal{G})$ matter: A group $G$ with a larger $|G|$ and a smaller $m$ is preferred.

Let $G$ and $G'$ be two finite groups of symmetries of $\mu_X$. Suppose $\mathcal{G}$ are both $G$-invariant and $G'$-invariant. By Theorem 3, $G'$ induces an on average more efficient projection estimator than $G$ if $\overline{\mathrm{eff}}(f_{P_{G'}}(X)) > \overline{\mathrm{eff}}\big(P_G(f)(X)\big)$. Moreover, to compare multiple projection estimators associated with $G$ with the standard estimator, we define the average efficiency ratio of $\overline{\mathcal{E}}_G$ over $f$ by

$$\overline{\mathcal{E}}_G := \frac{\overline{\mathrm{eff}}\big(P_G(f)(X)\big)}{\overline{\mathrm{eff}}(f(X))} = \frac{(n-1)}{(m-1)|G|}.$$

The one with higher average efficiency ratio is regarded as an on average more efficient estimator.

### 4.3 An Application of Theorem 3: The Antithetic Variates Estimator

Consider $X \sim N_d(\mathbf{0}_d, I_d)$. Let

$$G = \{g_1(x) := x, g_2(x) := -x\}.$$

Recall Example 2 of Section 3.3 that $P_G(f)$ is the antithetic variate estimator. Suppose $d = 2$ and consider the following two subspaces of functions.

(a) $\mathcal{G}_1 = \text{Span}\{1, x_1, x_2, x_1^2, x_1x_2, x_2^2\}$, the subspace of polynomials with degrees less than or equal to two.
(b) $\mathcal{G}_2 = \text{Span}\{1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3\}$, the subspace of polynomials with degrees less than or equal to three.

We summarize the dimension of the subspace of functions, dimension of the projected subspace of functions, the cardinality of $G$, the average variance ratio, and the average efficiency ratio in Table 1. In the following, we compute $m$ for $\mathcal{G}_1$ in details. We can obtain $m$ for $\mathcal{G}_2$ in a similar fashion, so we omit the computation for brevity. Note that $P_G(\mathcal{G}_1)$ is spanned by

$$\{P_G(1), P_G(x_1), P_G(x_2), P_G(x_1^2), P_G(x_1x_2), P_G(x_2^2)\},$$

which is equal to $\{1, 0, 0, x_1^2, x_1x_2, x_2^2\}$. Hence, the dimension of $P_G(\mathcal{G}_1)$ is 4.

Next, let us consider the general case. Suppose $d \geq 1$ and $\mathcal{G}$ is a $G$-invariant space of functions. For all $f(x) \in \mathcal{G}$,

$$P_G(f)(x) = \frac{1}{2}(f(x) + f(-x))$$

is also known as the even part of $f(x)$. Therefore, $P_G(\mathcal{G})$ is the subspace of $\mathcal{G}$ consisting of all even functions. In this case, the efficiency ratio of the antithetic variates estimator $P_G(f)$ equals to $\frac{(n-1)}{2(m-1)}$, which is larger than one if $m < \frac{(n+1)}{2}$. In other words, the antithetic variate estimator on $\mathcal{G}$ is on average more efficient than the standard estimator if the dimension of the subspace of even functions in $\mathcal{G}$ is small.

## 5 The Variance Reduction

We analyze the variance reduction of a specific function using its average and difference function based on the concept of fundamental domain in Section 5.1, provide a simple example to explain how to select the group $G$ for a specific problem in Section 5.2. We then illustrate the superiority of the projection estimator in pricing Asian options under the GARCH model in Section 5.3

In Sections 5.2 and 5.3, we conduct numerical experiments to calculate $\mathbb{E}(f(X))$ using Monte Carlo simulation. Let $N$ be the sample size and $X^{(i)}$ be the $i$-th realization sampled from the given distribution for $i = 1, \ldots, N$. Then, we estimate $\mathbb{E}\big(f(X)\big)$ with the standard

**Table 1** The dimension of $\mathcal{G}$ ($n$), dimension of $P(\mathcal{G})$ ($m$), the cardinality of $G$ ($|G|$), the average variance ratio ($\overline{\mathcal{V}}_G$), and the average efficiency ratio ($\overline{\mathcal{E}}_G$)

| $\mathcal{G}$ | $n$ | $m$ | $\|G\|$ | $\overline{\mathcal{V}}_G$ | $\overline{\mathcal{E}}_G$ |
|---|---|---|---|---|---|
| $\mathcal{G}_1$ | 6 | 4 | 2 | 3/5 | 5/6 |
| $\mathcal{G}_2$ | 10 | 4 | 2 | 1/3 | 3/2 |

estimator $f$ using its sample mean,

$$\frac{1}{N}\sum_{i=1}^{N} f\left(X^{(i)}\right).$$

We estimate $\mathbb{E}\big(f(X)\big)$ with the antithetic variates estimator $f_{AT}$ using its sample mean,

$$\frac{1}{N}\sum_{i=1}^{N} f_{AT}\left(X^{(i)}\right)=\frac{1}{N}\sum_{i=1}^{N}\frac{f\left(X^{(i)}\right)+f\left(-X^{(i)}\right)}{2}.$$

Similarly, we estimate $\mathbb{E}\big(f(X)\big)$ with the projection estimator $P_G(f)$ using its sample mean,

$$\frac{1}{N}\sum_{i=1}^{N} P_G(f)\left(X^{(i)}\right)=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{|G|}\sum_{g\in G} f_g\left(X^{(i)}\right)\right).$$

To obtain accurate results, for the rest of this paper, we set a large number of sample size $N = 100,000$, i.e., we generate independent $= 100,000$ realizations for each of these three estimators $f$, $f_{AT}$, and $P_G(f)$. We remark all simulation results through out this paper are carried out using Matlab (R2019a) in a desktop (Intel Core i9 CPU with 64 GB RAM).

### 5.1 The Average and Difference Function Based on the Fundamental Domain

Let $G$ be a group of symmetries of a probability measure $\mu_X$ and let $D$ be the support of $\mu_X$, which is a subset of $\mathbb{R}^n$. For a point $x$ in $D$, the $G$-orbit of $x$ is the set $\{gx : g \in G\}$. A subset of $D$, denoted as $D_0$, is called a fundamental domain under the action of $G$, if it contains exactly one point of each orbit of $G$. (See Beardon 1983 for more details about fundamental domains.) As a result, $D$ can be written as a union of $\{gD_0 : g \in G\}$, namely

$$D=\bigcup_{g\in G} gD_0.$$

In addition, a point $x$ in $D$ is contained in more than one of $\{gD_0 : g \in G\}$ only when $x$ is fixed by some non-identity element of $G$ (i.e., $gx = x$, for some non-identity $g \in G$). For the rest of this paper, we only consider the case where the subset consisting of all points fixed by some $g$ is of measure zero and $D_0$ is assumed to be a measurable set.

In this case, we have $\mu_X(gD_0) = \frac{1}{|G|}$ and for any integrable function $f(x)$,

$$\mathbb{E}\big(f(X)\big)=\int_{\mathbb{R}^d} f(x)d\mu_X(x)=\int_D f(x)d\mu_X(x)=\sum_{g\in G}\int_{gD_0} f(x)d\mu_X(x).$$

For illustration, Let $D = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$ with the standard Lebesque measure $\mu_X$. Consider three groups of symmetries on $D$ and their associated fundamental domains:

1. For $G = \{g_1(x, y) = (x, y), g_2(x, y) = (-x, -y)\}$, one can choose $D_0$ to be the set $\{(x, y) \in D : x \geq 0\}$.
2. For $G = \{g_1, g_2, g_3, g_4\}$ where $\{g_i(x, y), i = 1, 2, 3, 4\} = \{(\pm x, \pm y)\}$, one can choose $D_0$ to be the set $\{(x, y) \in D : x \geq 0, y \geq 0\}$.
3. For $G = \{g^0, g, g^2, g^3\}$, where $g(x, y) = (y, -x)$ is the clockwise rotation of 90°, one can choose $D_0$ to be the set $\{(x, y) \in D : x \geq 0, y \geq 0\}$.

Using the concept of fundamental domains, we decompose a function as a sum of its average function and difference function as follows. For all $g \in G$, let $I_{gD_0}(x)$ be the indicator function of the set $gD_0$.

Then for all $x \in D$ and $g, h \in G$, it is easy to see that

1. $\langle I_{gD_0}, \mathbf{1}\rangle = \frac{1}{|G|}$;
2. $\langle I_{gD_0}, I_{hD_0}\rangle$ equals to zero if $g \neq h$; it equals to $1/|G|$ if $g = h$;
3. $\sum_{g\in G}\langle I_{gD_0}, f\rangle = \mathbb{E}\big(f(X)\big)$.

Define $f_a(x)$ as the average function of $f(x)$ with respect to $\{gD_0 : g \in G\}$:

$$f_a(x) := |G| \sum_{g\in G} \langle I_{gD_0}, f\rangle I_{gD_0}(x).$$

Moreover, define $f_d(x)$ be the difference between $f(x)$ and its average function:

$$f_d(x) := f(x) - f_a(x).$$

With these definitions, a function $f(x)$ can be written as a sum of its average and difference functions.

Consider a simple example as follows. Suppose $D = \left[-\frac{1}{2}, \frac{1}{2}\right]$ with the standard Lebesque measure $\mu_X$ and $G = \{g_1(x) := x, g_2(x) := -x\}$. Then we can choose $D_0 = \left[0, \frac{1}{2}\right]$. Figure 4 shows the decomposition of a function $f(x)$ as a sum of $f_a(x)$ and $f_d(x)$. The main properties of such decomposition are described in the following proposition.

**Proposition 1** *For any functon $f \in \mathcal{F}$, the following identities hold.*

*1.* $\mathbb{E}\big(f(X)\big) = \mathbb{E}(f_a(X))$.
*2.* $\mathrm{var}\big(f(X)\big) = \mathrm{var}(f_a(X)) + \mathrm{var}(f_d(X))$.
*3.* $\mathrm{var}\big(P_G(f)_a(X)\big) = \mathrm{var}(P_G(f_a)(X)) = 0$.
*4.* $\mathrm{var}\big(P_G(f)_d(X)\big) = \mathrm{var}\big(P_G(f_d)(X)\big) \leq \mathrm{var}(f_d(X))$.

*Proof* See Appendix E. □

Applying Proposition 1, we obtain Theorem 4 immediately which yields the amount of variance reduction of the projection estimator $P_G(f)$ to the standard estimator $f$.

**Theorem 4** *The amount of variance reduction of the projection estimator $P_G(f)$ to the standard estimator $f$ satisfies*

$$\begin{aligned}\mathrm{var}\big(f(X)\big) - \mathrm{var}\big(P_G(f)(X)\big) &= \mathrm{var}(f_a(X)) + \big(\mathrm{var}(f_d(X)) - \mathrm{var}\big(P_G(f_d)(X)\big)\big) \\ &\geq \mathrm{var}(f_a(X)).\end{aligned}$$

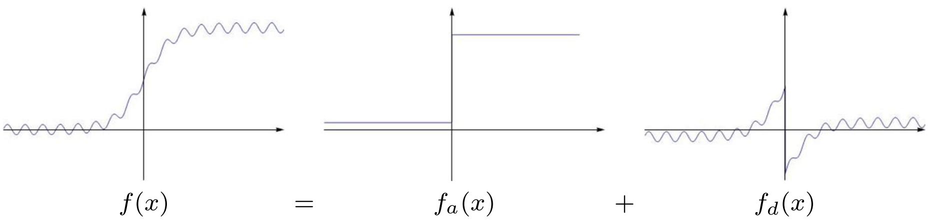



**Fig. 4** The decomposition of the function $f(x)$ as $f_a(x) + f_d(x)$

*Here, both $f_a(X)$ and $f_d(X)$ depend on the choice of the fundamental domain.*

Theorem 4 implicates that, if $\mathrm{var}(f_a(X))$ is large, the amount of variance reduction of $P_G(f)$ to $f$ would be large. On the other hand, by Proposition 1, we have

$$\mathrm{var}\big(f(X)\big) = \mathrm{var}(f_a(X)) + \mathrm{var}(f_d(X)) = \mathrm{var}(f_a(X)) + \mathrm{var}(f(X) - f_a(X)).$$

Therefore, $\mathrm{var}(f_a(X))$ is large if and only if $f_a$ is close to $f$. As a result, for a specific function $f$, the above arguments suggest to find a group $G$ with a suitable fundamental domain $D_0$ so that $f_a$ is close to $f$.

Now, we apply Theorem 4 to give a finer result for the antithetic variates estimator. Let $G = \{g_1(x) := x, g_2(x) := -x\}$. Then the $G$-orbit of $x$ is $\{x, -x\}$. Therefore, one can choose the fundamental domain $D_0$ to satisfy the condition that for all $x \in D_0$, $f(x) \geq f(-x)$. In this case,

$$\mathbb{R}^d = g_1(D_0) \bigcup g_2(D_0) = D_0 \bigcup (-D_0).$$

Consider the following two functions

$$f_1(x) = \max\{f(x), f(-x)\} \quad \text{and} \quad f_2(x) = \min\{f(x), f(-x)\}.$$

Then, it is easy to see that

(a) $f(x) = f_1(x)$ if $x \in D_0$;
(b) $f(x) = f_2(x)$ if $x \in -D_0$;
(c) $f_1(x) - f_2(x) = |f(x) - f(-x)|$ for all $x \in \mathbb{R}^d$;
(d) $f_1(x)$ and $f_2(x)$ are both even functions, i.e. $f_1(x) = f_1(-x)$ and $f_2(x) = f_2(-x)$ for all $x \in \mathbb{R}^d$

Now for any $x \in D_0$, by the definition of $f_a$ we have

$$\begin{aligned} f_a(x) &= |G| \sum_{g \in G} \langle I_{gD_0}, f \rangle I_{gD_0}(x) = 2 \int_{D_0} f(y) d\mu_X(y) \\ &= 2 \int_{D_0} f_1(y) d\mu_X(y) = \int_{\mathbb{R}^d} f_1(y) d\mu_X(y) = \mathbb{E}(f_1(X)). \end{aligned}$$

Similarly, for any $x \in (-D_0)$, we have

$$f_a(x) = \mathbb{E}(f_2(X)).$$

Now, we are ready to compute the variance of $f_a$. Let $e_1 = \mathbb{E}(f_1(X))$ and $e_2 = \mathbb{E}(f_2(X))$ for short. Since $f_a(x)$ is a constant function on both $D_0$ and $(-D_0)$, we have

$$\begin{aligned} \mathrm{var}(f_a) &= \frac{1}{2}\left(e_1 - \frac{(e_1+e_2)}{2}\right)^2 + \frac{1}{2}\left(e_2 - \frac{(e_1+e_2)}{2}\right)^2 = \frac{1}{4}(e_1 - e_2)^2 \\ &= \frac{1}{4}\Big(\mathbb{E}\big(f_1(X)\big) - \mathbb{E}\big(f_2(X)\big)\Big)^2 = \frac{1}{4}\mathbb{E}\Big(f_1(X) - f_2(X)\Big)^2 \\ &= \frac{1}{4}\mathbb{E}\Big(|f(X) - f(-X)|\Big)^2. \end{aligned}$$

Together with Theorem 4, we obtain the following theorem, which implicates that the more $f(X)$ deviates from $f(-X)$, the more variance reduction with the antithetic variates estimator could be obtained.

**Theorem 5** *The amount of variance reduction of the antithetic variate estimator $f_{AT}$ to the standard estimator $f$ satisfies*

$$\operatorname{var}\big(f(X)\big) - \operatorname{var}\big(f_{AT}(X)\big) \geq \frac{1}{4}\mathbb{E}\Big(|f(X) - f(-X)|\Big)^2.$$

### 5.2 An Illustration: Indicator Functions of Two Different Regions

In the following, we provide two examples to numerically verify the usefulness of the proposed projection estimator. Let $X = (X_1, X_2) \sim N_2(\mathbf{0}_2, I_2)$ and let $f(X)$ be the indicator function on the region $R_1$ determined by the conditions

$$3X_1 - X_2 > 0.1 \text{ and } -X_1 + 2X_2 > 0,$$

as depicted in Fig. 5. We are interested in calculating

$$\mathbb{E}\left(f(X)\right) = \int_{\mathbb{R}^2} f(x)d\mu_X(x) = \int_{R_1} d\mu_X(x) = \text{Prob}(X \in R_1).$$

In addition, we choose $G$ to be the group of 8 elements generated by a rotation of degree $\pi/4$ around the origin with a fundamental domain $D_0$ as shown in Fig. 6. As a result, $f$ is very close to $f_a$ and $P_G(f)$ shall has a small variance.

Table 2 compares the estimated variance of the standard estimator $f$, the antithetic variates $f_{AT}$, the projection estimator $P_G(f)$, and also the variance ratios to the standard estimator of $f_{AT}$ and $P_G(f)$. It is shown that the projection estimator produces the least

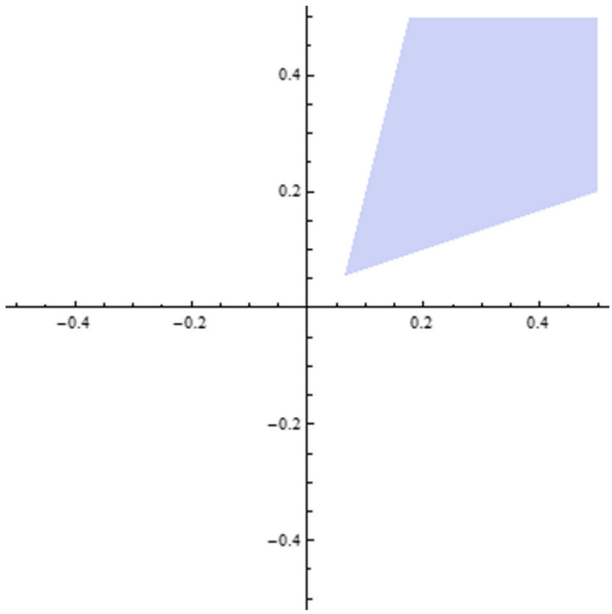



**Fig. 5** The region $R_1$

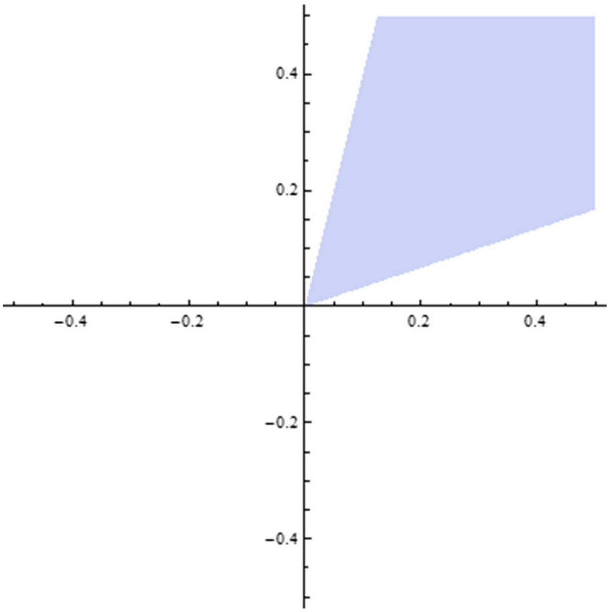



**Fig. 6** The fundamental domain $D_0$

variance, followed by the antithetic variates estimator, and the standard estimator. In addition, the projection estimator has a substantially smaller variance (a variance ratio of 141.3) than the antithetic variates estimator (a variance ratio of simply 2.3).

Now, we consider a different region $R_2$ determined by the conditions

$$(3X_1 - X_2 - 0.1)(-X_1 + 2X_2) > 0,$$

as depicted in Fig. 7. It is shown in Table 3 that the antithetic variates estimator has unfortunately almost the same variance as the standard estimator (a variance ratio of 1.1), but the projection estimator still has a substantially smaller variance (a variance ratio of 79.6).

### 5.3 An Illustration: GARCH Option Pricing

To demonstrate the superiority of the proposed projection estimator, we follow the framework of GARCH option pricing in Duan and Simonato (1998). We use the GARCH(1,1)-in mean to describe the daily asset return dynamics. Let $S_t$ be the asset price at date $t$, and $\sigma_t^2$ be the conditional variance of the logarithmic return over the period $[t, t+1]$, which is a

**Table 2** The estimated mean and variance using the standard estimator $f$, antithetic variates estimator $f_{AT}$, and the projection estimator $P_G(f)$

| Variance | | | $\frac{\mathrm{var}(f(X))}{\mathrm{var}(f_{AT}(X))}$ | $\frac{\mathrm{var}(f(X))}{\mathrm{var}(P_G(f)(X))}$ |
|---|---|---|---|---|
| $f$ | $f_{AT}$ | $P_G(f)$ | | |
| 0.104125 | 0.045305 | 0.000736 | 2.3 | 141.3 |

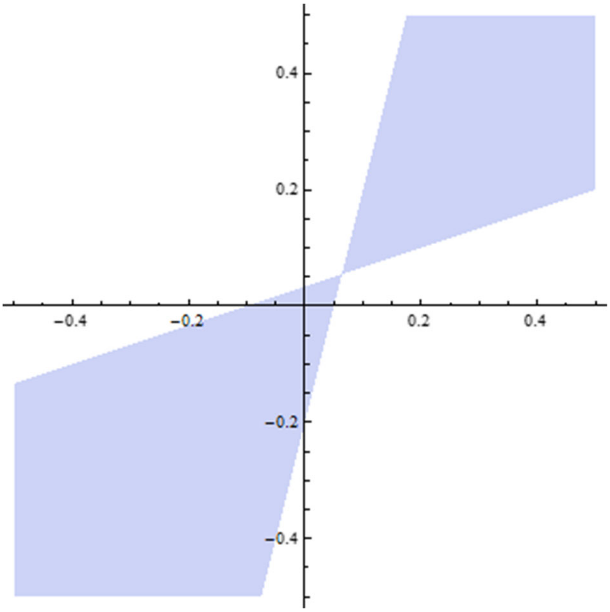



**Fig. 7** The region $R_2$

day. The dynamics under the locally risk-neutralized probability measure $Q$ is

$$\begin{aligned}\log\frac{S_{t+1}}{S_t} &= r-\frac{1}{2}\sigma_{t+1}^2+\sigma_{t+1}\varepsilon_{t+1},\\ \sigma_{t+1}^2 &= \beta_0+\beta_1\sigma_t^2+\beta_2\sigma_t^2(\varepsilon_t-\lambda)^2,\\ \varepsilon_{t+1}|\mathcal{F}_t &\overset{i.i.d.}{\sim} N(0,1),\end{aligned}$$

where $\beta_0$, $\beta_1$, $\beta_2$ are the GARCH(1,1) parameters, $r$ is the daily risk-free rate, and $\lambda$ is the unit risk premium (per unit of conditional standard deviation) parameter that defines the conditional mean equation of the GARCH(1,1)-(in mean) dynamics under the physical probability measure. Viewed from date $t$, $\varepsilon_{t+1}$ is a standard normal random variable.

Let $K$ be the strike price and $T$ be the maturity. We denote the positive function by $(\cdot)^+$. The price of the Asian option is the expectation of the discounted payoff function:

$$\mathbb{E}\left(\mathrm{e}^{-rT}\left(\frac{1}{T}\sum_{t=1}^{T}S_t-K\right)^+\right), \tag{10}$$

**Table 3** The estimated mean and variance using the standard estimator $f$, antithetic variates estimator $f_{AT}$, and the projection estimator $P_G(f)$

| Variance | | | $\frac{\mathrm{var}(f(X))}{\mathrm{var}(f_{AT}(X))}$ | $\frac{\mathrm{var}(f(X))}{\mathrm{var}(P_G(f)(X))}$ |
|---|---|---|---|---|
| $f$ | $f_{AT}$ | $P_G(f)$ | | |
| 0.21824 | 0.20175 | 0.00274312 | 1.1 | 79.6 |

**Table 4** The estimated mean and variance in pricing Asian options under the GARCH model using the standard estimator $f$, antithetic variates estimator $f_{AT}$, and the projection estimator $P_G(f)$ with strike price $K = 90, 100, 110$ and maturity $T = 20, 60, 120, 252, 504$

| $K$ | Mean | | | Variance | | | $\frac{\mathrm{var}(f(X))}{\mathrm{var}(f_{AT}(X))}$ | $\frac{\mathrm{var}(f(X))}{\mathrm{var}(P_G(f)(X))}$ |
|---|---|---|---|---|---|---|---|---|
| | $f$ | $f_{AT}$ | $P_G(f)$ | $f$ | $f_{AT}$ | $P_G(f)$ | | |
| $T = 20$ | | | | | | | | |
| 90 | 10.3336 | 10.3362 | 10.3364 | 7.1370 | 0.0072 | 0.0015 | 990.8 | 4827.1 |
| 100 | 1.2443 | 1.2467 | 1.2461 | 3.0577 | 0.7629 | 0.1767 | 4.0 | 17.3 |
| 110 | 0.0042 | 0.0039 | 0.0035 | 0.0187 | 0.0081 | 0.0020 | 2.3 | 9.3 |
| $T = 60$ | | | | | | | | |
| 90 | 10.9508 | 10.9738 | 10.9733 | 19.6675 | 0.1288 | 0.0261 | 152.7 | 754.1 |
| 100 | 2.3852 | 2.3943 | 2.3889 | 9.7174 | 2.1376 | 0.4537 | 4.5 | 21.4 |
| 110 | 0.0841 | 0.0848 | 0.0846 | 0.4864 | 0.2321 | 0.0662 | 2.1 | 7.3 |
| $T = 120$ | | | | | | | | |
| 90 | 11.9134 | 11.9233 | 11.9228 | 37.6455 | 0.5453 | 0.1072 | 69.0 | 351.1 |
| 100 | 3.7488 | 3.7464 | 3.7512 | 21.1628 | 3.9164 | 0.7945 | 5.4 | 26.6 |
| 110 | 0.4354 | 0.4431 | 0.4433 | 2.9634 | 1.4583 | 0.3643 | 2.0 | 8.1 |
| $T = 252$ | | | | | | | | |
| 90 | 13.8948 | 13.9132 | 13.9073 | 73.5481 | 1.9287 | 0.3326 | 38.1 | 221.1 |
| 100 | 6.2301 | 6.2461 | 6.2473 | 48.7523 | 7.4355 | 1.3394 | 6.6 | 36.4 |
| 110 | 1.8253 | 1.8168 | 1.8288 | 17.4959 | 6.9808 | 1.4084 | 2.5 | 12.4 |
| $T = 504$ | | | | | | | | |
| 90 | 17.2790 | 17.2326 | 17.2369 | 137.2150 | 5 .2350 | 0.8706 | 26.2 | 157.6 |
| 100 | 10.1828 | 10.1742 | 10.1986 | 106.7990 | 12.6402 | 2.1362 | 8.4 | 50.0 |
| 110 | 5.0299 | 5.0652 | 5.0428 | 62.9584 | 19.0828 | 3.0569 | 3.3 | 20.6 |

To formulate Eq. 10 as $\mathbb{E}\big(f(X)\big)$, let $S_0$ denote the initial stock price and assume $X = (X_1, \ldots, X_T)' \sim N_T(\mathbf{0}_T, I_T)$. Then, we have

$$f(X) = e^{-rT}\left(\frac{1}{T}\sum_{t=1}^{T} S_{t-1}e^{r-\sigma_t^2/2+\sigma_t X_t} - K\right)^+,$$

where

$$\sigma_{t+1}^2 = \beta_0 + \beta_1\sigma_t^2 + \beta_2\sigma_t^2(X_t - \lambda)^2.$$

Because $f(X)$ is complicated, there is no simple way to choose a suitable group.

Next, we consider a group of coordinate sign changes as follows. Let

$$G = \{g_1(x), \cdots, g_8(x)\},$$

with

$$g_i(x) = (A_{i1}x_1, A_{i2}x_2, \cdots, A_{ik}x_k, \cdots) \tag{11}$$

**Table 5** The estimated efficiency by cardinality in pricing Asian options under the GARCH model using the standard estimator $f(X)$, antithetic variates estimator $f_{AT}(X)$, and the projection estimator $P_G(f)(X)$ with strike price $K = 90, 100, 110$ and maturity $T = 20, 60, 120, 252, 504$

| $K$ | Efficiency by Cardinality | | | $\frac{\text{eff}_c(f_{AT}(X))}{\text{eff}_c(f(X))}$ | $\frac{\text{eff}_c(P_G(f)(X))}{\text{eff}_c(f(X))}$ |
|---|---|---|---|---|---|
| | $f$ | $f_{AT}$ | $P_G(f)$ | | |
| $T = 20$ | | | | | |
| 90 | 0.14 | 69.42 | 84.54 | 495.4 | 603.4 |
| 100 | 0.33 | 0.66 | 0.71 | 2.0 | 2.2 |
| 110 | 53.38 | 62.11 | 61.75 | 1.2 | 1.2 |
| $T = 60$ | | | | | |
| 90 | 0.05 | 3.88 | 4.79 | 76.4 | 94.3 |
| 100 | 0.10 | 0.23 | 0.28 | 2.3 | 2.7 |
| 110 | 2.06 | 2.15 | 1.89 | 1.0 | 0.9 |
| $T = 120$ | | | | | |
| 90 | 0.03 | 0.92 | 1.17 | 34.5 | 43.9 |
| 100 | 0.05 | 0.13 | 0.16 | 2.7 | 3.3 |
| 110 | 0.34 | 0.34 | 0.34 | 1.0 | 1.0 |
| $T = 252$ | | | | | |
| 90 | 0.01 | 0.26 | 0.38 | 19.1 | 27.6 |
| 100 | 0.02 | 0.07 | 0.09 | 3.3 | 4.5 |
| 110 | 0.06 | 0.07 | 0.09 | 1.3 | 1.6 |
| $T = 504$ | | | | | |
| 90 | 0.01 | 0.10 | 0.14 | 13.1 | 19.7 |
| 100 | 0.01 | 0.04 | 0.06 | 4.2 | 6.2 |
| 110 | 0.02 | 0.03 | 0.04 | 1.6 | 2.6 |

and

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}. \tag{12}$$

Here, the subscript $k$ of $A_{ik}$ is regarded modulo 7.

We compare three estimators in pricing Asian call options under the GARCH model: (1) the standard estimator $f$ (2) the antithetic estimator $f_{AT}$ and (3) the projection estimator,

$$P_G(f) := \frac{1}{8}\left(f(g_1(X)) + f(g_2(X)) + \cdots + f(g_8(X))\right),$$

where $g_i(x)$ is defined through Eqs. 11 and 12.

Parameter values are set as follows: $S_0 = 100$, risk-free rate $r = 0.10/252 = 3.9683 \times 10^{-4}$ (i.e., an annualized risk-free rate of 10%), $\beta_0 = 0.00001$, $\beta_1 = 0.70$, $\beta_2 = 0.20$, $\lambda = 0.01$, and $\sigma_1^2 = \beta_0/(1 - \beta_1 - \beta_2) = 0.0001$. In addition, we vary the strike price

**Table 6** The computing times (in seconds) and estimated efficiency by computing time in pricing Asian options under the GARCH model using the standard estimator $f$, antithetic variates estimator $f_{AT}$, and the projection estimator $P_G(f)$ with strike price $K = 90, 100, 110$ and maturity $T = 20, 60, 120, 252, 504$

| $K$ | Computing Time | | | Efficiency by Time | | | $\frac{\mathrm{eff}_t(f_{AT})}{\mathrm{eff}_t(f)}$ | $\frac{\mathrm{eff}_t\left(P_G(f)(X)\right)}{\mathrm{eff}_t(f)}$ |
|---|---|---|---|---|---|---|---|---|
| | $f$ | $f_{AT}$ | $P_G(f)$ | $f$ | $f_{AT}$ | $P_G(f)$ | | |
| $T = 20$ | | | | | | | | |
| 90 | 4.1 | 5.2 | 5.7 | 0.0344 | 26.8033 | 119.6820 | 779.6 | 3480.9 |
| 100 | 3.9 | 5.1 | 5.7 | 0.0829 | 0.2587 | 0.9968 | 3.1 | 12.0 |
| 110 | 3.9 | 5.1 | 5.6 | 13.5285 | 24.3237 | 87.7911 | 1.8 | 6.5 |
| $T = 60$ | | | | | | | | |
| 90 | 6.1 | 8.8 | 9.7 | 0.0083 | 0.8785 | 3.9558 | 105.6 | 475.5 |
| 100 | 6.2 | 8.9 | 9.8 | 0.0166 | 0.0527 | 0.2250 | 3.2 | 13.5 |
| 110 | 6.1 | 9.0 | 9.8 | 0.3369 | 0.4802 | 1.5362 | 1.4 | 4.6 |
| $T = 120$ | | | | | | | | |
| 90 | 9.3 | 15.3 | 14.5 | 0.0029 | 0.1202 | 0.6414 | 42.1 | 224.8 |
| 100 | 9.1 | 15.3 | 14.7 | 0.0052 | 0.0167 | 0.0857 | 3.2 | 16.6 |
| 110 | 9.2 | 15.2 | 14.6 | 0.0367 | 0.0450 | 0.1876 | 1.2 | 5.1 |
| $T = 252$ | | | | | | | | |
| 90 | 15.7 | 25.9 | 24.8 | 0.0009 | 0.0200 | 0.1211 | 23.1 | 139.6 |
| 100 | 15.7 | 25.5 | 24.8 | 0.0013 | 0.0053 | 0.0301 | 4.0 | 23.0 |
| 110 | 16.0 | 25.9 | 24.9 | 0.0036 | 0.0055 | 0.0285 | 1.5 | 8.0 |
| $T = 504$ | | | | | | | | |
| 90 | 26.9 | 43.0 | 45.2 | 0.0003 | 0.0044 | 0.0254 | 16.4 | 94.0 |
| 100 | 27.0 | 43.3 | 45.2 | 0.0003 | 0.0018 | 0.0104 | 5.3 | 29.8 |
| 110 | 26.7 | 42.9 | 45.1 | 0.0006 | 0.0012 | 0.0072 | 2.1 | 12.2 |

$K = 90, 95, 100, 105, 110$ and maturity $T = 20$ (a month), 60 (a quarter), 120 (half year), 252 (a year), and 504 (two years).

To start, it is shown in Table 4 that these three estimators have very close prices given the same parameters. In addition, the projection estimator has the least variance, followed by the antithetic variates estimator, and the standard estimator. It is interesting that the projection estimator has a substantially smaller variance than the antithetic variates estimator. In addition, these two estimators are particularly useful for in-the-money options.

Table 5 compares the efficiency by cardinality defined in Eq. 4 of each estimator. Recall the number of sample points used in the standard estimator, antithetic variates estimator, and projection estimator are 1, 2, and 8, respectively. The projection estimator produces the highest efficiency by cardinality, followed by the antithetic variates estimator, and the standard estimator. This indicates that even though the projection estimator uses more sample points, it still marginally dominates the other estimators in terms of efficiency by cardinality.

Lastly, we compare the computing time and efficiency by computing time defined in Eq. 5 for each estimator in Table 6. Each estimator takes about the same computing time across different strike prices, but takes longer for longer maturities. The standard estimator takes the least computing time, but the projection estimator and the antithetic variates estimator take about the same computing time. This is likely because generating random variables is more time consuming than calculating multiple function values. As a result, the projection estimator still produces the highest efficiency by time, followed by the antithetic variates estimator, and the standard estimator. In terms of efficiency by computing time, the projection estimator dominates the antithetic variates estimator to a greater degree.

As a remark, this group remains applicable when $T$ is not a multiple of 7. For example, when $T = 7$, we set

$$g_2(x) = (-x_1, x_2, -x_3, x_4, -x_5, x_6, -x_7).$$

When $T = 9$, we set

$$g_2(x) = (-x_1, x_2, -x_3, x_4, -x_5, x_6, -x_7, -x_8, x_9).$$

Here, signs of $x_8$ and $x_9$ are equal to signs of $x_1$ and $x_2$, respectively.

## 6 Conclusion

This paper uses orthogonal projections to analyze the variance of an estimator, and provides a novel projection estimator associated with a group of symmetries of the probability measure as a variance reduction technique. For a finite-dimensional space of functions, the average variance and the efficiency ratio of the projection estimator could be explicitly derived. For a specific function, a guidance on how to select the group of symmetries is given using the average function and the difference function. For further research, the construction for the optimal projection estimator and applications to other examples will be studied.

## Appendix A: Proof of Lemma 1

*Proof* To prove that $P_{\mathbb{E}}$ is a projection, we need to show that it satisfies the conditions C1 and C2. For the condition C1, note that for any $f \in \mathcal{F}$,

$$P_{\mathbb{E}}^2\big(f(X)\big) = P_{\mathbb{E}}(P_{\mathbb{E}}\big(f(X)\big)) = \langle P_{\mathbb{E}}\big(f(X)\big), \mathbf{1}\rangle = P_{\mathbb{E}}\big(f(X)\big).$$

Hence, we have $P_{\mathbb{E}}^2 = P_{\mathbb{E}}$. For the condition C2, for any $f, g \in \mathcal{F}$, we have

$$\begin{aligned}\langle P_{\mathbb{E}}\big(f(X)\big), g - P_{\mathbb{E}}[g]\rangle &= \langle P_{\mathbb{E}}\big(f(X)\big), g\rangle - \langle P_{\mathbb{E}}\big(f(X)\big), P_{\mathbb{E}}[g]\rangle \\ &= P_{\mathbb{E}}\big(f(X)\big)\langle \mathbf{1}, g\rangle - P_{\mathbb{E}}\big(f(X)\big)P_{\mathbb{E}}[g] \\ &= P_{\mathbb{E}}\big(f(X)\big)P_{\mathbb{E}}[g] - P_{\mathbb{E}}\big(f(X)\big)P_{\mathbb{E}}[g] = 0.\end{aligned}$$

As a result, $P_{\mathbb{E}}$ is an orthogonal projection. □

## Appendix B: Proof of Theorem 1

*Proof* Let $V_0$ be the space of constant functions. Since $P(\mathcal{F}) \supset V_0$, we have $P(\mathcal{F})^{\perp} \subset V_0^{\perp}$. For all $f \in \mathcal{F}$, because $(f - P(f)) \in P(\mathcal{F})^{\perp}$, it is clear that

$$\mathbb{E}[f(X) - P(f)(X)] = \langle f - P(f), \mathbf{1}\rangle = 0.$$

As a result, the expectation of $f(X)$ equals

$$\begin{aligned}\mathbb{E}\big(f(X)\big) &= \mathbb{E}\Big[P(f)(X) + f(X) - P(f)(X)\Big) = \mathbb{E}[P(f)(X)] \\ &+\mathbb{E}[f(X) - P(f)(X)] = \mathbb{E}[P(f)(X)].\end{aligned}$$

In addition, we obtain

$$\begin{aligned}\text{var}[P(f)] + \text{var}[f - P(f)] &= \|P(f)\|^2 - \mathbb{E}[P(f)(X)]^2 + \|f - P(f)\|^2 \\ &\quad -\mathbb{E}[f(X) - P(f)(X)]^2 \\ &= (\|P(f)\|^2 + \|f - P(f)\|^2) - \mathbb{E}[P(f)(X)]^2 \\ &\quad -\mathbb{E}[f(X) - P(f)(X)]^2) \\ &= \|f\|^2 - \mathbb{E}\big(f(X)\big)^2 - 0 \\ &= \text{var}\big(f(X)\big),\end{aligned}$$

where the last equality holds by Eq. 2. Therefore, we obtain

$$\text{var}\big(f(X)\big) = \text{var}[P(f)] + \text{var}[f - P(f)] \geq \text{var}[P(f)].$$

□

## Appendix C: Proof of Lemma 2

*Proof* Since $g$ is a symmetry of $\mu_X$, we have

$$\mathbb{E}\big(f_g(X)\big) = \int_{\mathbb{R}^n} f(gx)d\mu_X(x) = \int_{\mathbb{R}^n} f(y)d\mu_X(g^{-1}y) = \int_{\mathbb{R}^n} f(y)d\mu_X(y) = \mathbb{E}\big(f(X)\big).$$

Hence, $f_g(X)$ is an unbiased estimator. By the same token, we also have $\mathbb{E}\big(f_g(X)^2\big) = \mathbb{E}\big(f(X)^2\big)$. Now we have

$$\text{var}\big(f_g(X)\big) = \mathbb{E}\big((f_g(X))^2\big) - \mathbb{E}\big(f_g(X)\big)^2 = \mathbb{E}\big(f(X)^2\big) - \mathbb{E}\big(f(X)\big)^2 = \text{var}\big(f(X)\big).$$

Because both $\mathbb{E}\big(f_g(X)\big)$ and $\text{var}\big(f_g(X)\big)$ are well-defined, $f_g$ remains in $\mathcal{F}$. □

## Appendix D: Proof of Theorem 2

*Proof* First, we show that $P_G$ is a linear transformation. For $f_1, f_2 \in \mathcal{F}$, and $\alpha \in \mathbb{R}$, it is clear that

$$\begin{aligned}P_G(f_1 + \alpha f_2)(x) &= \frac{1}{|G|}\sum_{g\in G}(f_1 + \alpha f_2)(gx) \\ &= \frac{1}{|G|}\sum_{g\in G}(f_1(gx) + \alpha f_2(gx)) \\ &= \frac{1}{|G|}\sum_{g\in G} f_1(gx) + \frac{1}{|G|}\sum_{g\in G}\alpha f_2(gx) \\ &= P_G(f_1) + \alpha P_G(f_2).\end{aligned}$$

Therefore, we conclude that $P_G$ is a linear transformation on $\mathcal{F}$.

Next, let us show that $P_G = P_G^2$. For all $f \in \mathcal{F}$ and all $g \in G$,

$$P_G(f_g) = \frac{1}{|G|}\sum_{g' \in G} f(gg'x) = \frac{1}{|G|}\sum_{g'' \in G} f(g''x) = P_G(f(x)).$$

Here we use the property that the multiplication by $g$ from left just permutes elements of $G$. Therefore, it does not change the summation. Now we have

$$P_G(P_G(f(x))) = \frac{1}{|G|}\sum_{g \in G} P_G(f(gx)) = \frac{1}{|G|}\sum_{g \in G} P_G(f(x)) = P_G(f(x)).$$

Second, let us show that for $f_1, f_2 \in \mathcal{F}$, $\langle P_G(f_1), f_2 - P_G(f_2)\rangle = 0$, or equivalently $\langle P_G(f_1), P_G(f_2)\rangle = \langle P_G(f_1), f_2\rangle$. Now we have

$$\langle P_G(f_1), P_G(f_2)\rangle = \frac{1}{|G|^2}\sum_{g \in G}\left(\sum_{g' \in G}\int_\Omega f_1(gx) f_2(g'x) d\mu(x)\right)$$

Now let us change the variable $x$ by $y = g'x$. Together with the property that $d\mu(gx) = d\mu(x)$ and multiplying $g'^{-1}$ from right is a permutation on $G$, we can rewrite the above equation as

$$\begin{aligned}
\langle P_G(f_1), P_G(f_2)\rangle &= \frac{1}{|G|^2}\sum_{g' \in G}\left(\sum_{g \in G}\int_\Omega f_1(gg'^{-1}y) f_2(y) d\mu(g^{-1}y)\right) \\
&= \frac{1}{|G|^2}\sum_{g' \in G}\left(\sum_{g \in G}\int_\Omega f_1(gg'^{-1}y) f_2(y) d\mu(y)\right) \\
&= \frac{1}{|G|^2}\sum_{g' \in G}\left(\sum_{g \in G}\int_\Omega f_1(gy) f_2(y) d\mu(y)\right) \\
&= \frac{1}{|G|}\left(\sum_{g \in G}\int_\Omega f_1(gy) f_2(y) d\mu(y)\right) = \langle P_G(f_1), f_2\rangle.
\end{aligned}$$

We conclude that $P_G$ is an orthogonal projection. For the last part of the theorem, it is clear that $P_G(\mathbf{1}) = \mathbf{1}$ which implies that $P(\mathcal{F}) \supset P(\mathbb{R}) = \mathbb{R}$. □

## Appendix E: Proof of Proposition 1

*Proof* Let $a_g = \langle I_{gD_0}, f\rangle$ for short.
(1) Consider the following equations.

$$\mathbb{E}(f_a(X)) = \langle f_a, \mathbf{1}\rangle = |G|\sum_{g \in G} a_g \langle I_{gD_0}, \mathbf{1}\rangle = \sum_{g \in G} a_g = \mathbb{E}\big(f(X)\big).$$

(2) Consider the following equations.

$$\langle f_a, f\rangle = |G|\sum_{g \in G} a_g \langle I_{gD_0}, f\rangle = |G|\sum_{g \in G}(a_g)^2$$

On the other hand, we have

$$\langle f_a, f_a \rangle = |G|^2 \sum_{g\in G}\sum_{h\in G} a_g a_h \langle I_{gD_0}, I_{hD_0}\rangle = |G| \sum_{g\in G}(a_g)^2 = \langle f_a, f\rangle$$

From the above result, we have

$$\langle f_a, f_d\rangle = \langle f_a, f - f_a\rangle = \langle f_a, f\rangle - \langle f_a, f_a\rangle = 0,$$

which implies the following

$$\mathrm{var}\big(f(X)\big) = \langle f, f\rangle = \langle f_a + f_b, f_a + f_b\rangle = \langle f_a, f_a\rangle + \langle f_b, f_b\rangle = \mathrm{var}(f_a(X)) + \mathrm{var}[f_b].$$

(3) By definition, we have

$$P_G(f_a)(x) = \sum_{h\in G}\sum_{g\in G} a_g I_{gD_0}(hx).$$

Note that $hx \in gD_0$ if and only if $x \in h^{-1}gD_0$. We can rewrite the above equation as

$$\begin{aligned} P_G(f_a)(x) &= \sum_{g\in G} a_g \left(\sum_{h\in G} I_{h^{-1}gD_0}(x)\right) \\ &= \left(\sum_{g\in G} a_g\right)\left(\sum_{h'\in G} I_{hD_0}(x)\right) \\ &= \mathbb{E}\big(f(X)\big)\left(\sum_{h\in G} I_{hD_0}(x)\right). \end{aligned}$$

Here we use the fact that $\{h^{-1}g : h \in G\}$ equals to $G$ as a set. Next, consider

$$\begin{aligned} P_G(f)_a &= |G| \sum_{g\in G} \langle I_{gD_0}, P_G(f)\rangle I_{gD_0}(x) \\ &= \sum_{g\in G}\left\langle I_{gD_0}(x), \sum_{h\in G} f(hx)\right\rangle I_{gD_0}(x) \\ &= \sum_{g\in G}\sum_{h\in G}\big\langle I_{hgD_0}(x), f(x)\big\rangle I_{gD_0}(x) \\ &= \sum_{g\in G}\left(\sum_{h\in G} a_{hg}\right) I_{gD_0}(x) = \sum_{g\in G}\left(\sum_{h\in G} a_h\right) I_{gD_0}(x) \\ &= \left(\sum_{h\in G} a_h\right)\left(\sum_{g\in G} I_{gD_0}(x)\right) = P_G(f_a). \end{aligned}$$

Combing the above two results, We have shown that

$$P_G(f)_a = P_G(f_a) = \mathbb{E}\big(f(X)\big)\left(\sum_{h\in G} I_{hD_0}(x)\right)$$

which is a constant function except on a measure zero set. implies that $\mathrm{var}\big(P_G(f)_a(X)\big) = \mathrm{var}\big(P_G(f)_a(X)\big) = 0$.

(4) Applying (3), we have

$$P_G(f_d) = P_G(f - f_a) = P_G(f) - P_G(f_a) = P_G(f) - P_G(f)_a = P_G(f)_d.$$

Together with Corollary 1, we have

$$\mathrm{var}\big(P_G(f)_d(X)\big) = \mathrm{var}\big(P_G(f_d)(X)\big) \geq \mathrm{var}(f_d(X)).$$

□

**Data Availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

**Declarations**

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

Asmussen S, Glynn P (2007) Stochastic simulation: algorithms and analysis. Springer-Verlag, New York

Beardon AF (1983) The geometry of discrete groups. Springer-Verlag, New York

Duan JC, Simonato JG (1998) Empirical martingale simulation for asset prices. Manag Sci 44(9):1218–1233

Fraleigh JB (2019) A first course in abstract algebra, 7edn. Pearson Education, India

Glasserman P (2004) Monte carlo methods in financial engineering. Springer, New York

Glasserman P, Heidelberger P, Shahabuddin P (1999) Asymptotically optimal importance sampling and stratification for pricing path-dependent options. Math Financ 9:117–152

L'Ecuyer P (1994) Efficiency improvement and variance reduction. In: Proceedings of the 1994 winter simulation conference, Orlando, pp 122–132

Neddermeyer JC (2011) Non-parametric partial importance sampling for financial derivative pricing. Quant Finance 11:1193–1206

Park JJ, Choe GH (2016) A new variance reduction method for option pricing based on sampling the vertices of a simplex. Quant Finance 16(8):1165–1173

Ren H, Zhao S, Ermon S (2019) Adaptive antithetic sampling for variance reduction. In: International conference on machine learning. PMLR, pp 5420–5428

Ross SM (2013) Simulation. Academic Press, New York