Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Pricing and Hedging with Discontinuous Functions: Quasi-Monte Carlo Methods and Dimension Reduction

Xiaoqun Wang, Ken Seng Tan,

MANAGEMENT SCIENCE
Vol. 59, No. 2, February 2013, pp. 376–389
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1120.1568

Pricing and Hedging with Discontinuous Functions: Quasi–Monte Carlo Methods and Dimension Reduction

Xiaoqun Wang
Department of Mathematical Sciences, Tsinghua University, 100084 Beijing, China, xwang@math.tsinghua.edu.cn

Ken Seng Tan
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; and China Institute for Actuarial Science, Central University of Finance and Economics, 100081 Beijing, China, kstan@uwaterloo.ca

Quasi–Monte Carlo (QMC) methods are important numerical tools in the pricing and hedging of complex financial instruments. The effectiveness of QMC methods crucially depends on the discontinuity and the dimension of the problem. This paper shows how the two fundamental limitations can be overcome in some cases. We first study how path-generation methods (PGMs) affect the structure of the discontinuities and what the effect of discontinuities is on the accuracy of QMC methods. The insight is that the discontinuities can be QMC friendly (i.e., aligned with the coordinate axes) or not, depending on the PGM. The PGMs that offer the best performance in QMC methods are those that make the discontinuities QMC friendly. The structure of discontinuities can affect the accuracy of QMC methods more significantly than the effective dimension. This insight motivates us to propose a novel way of handling the discontinuities. The basic idea is to align the discontinuities with the coordinate axes by a judicious design of a method for simulating the underlying processes. Numerical experiments demonstrate that the proposed method leads to dramatic variance reduction in QMC methods for pricing options and for estimating Greeks. It also reduces the effective dimension of the problem.

Key words: option pricing; Greeks estimation; quasi–Monte Carlo methods; dimension reduction; effective dimension; Brownian bridge; principal component analysis; discontinuity
History: Received July 21, 2010; accepted March 10, 2012, by Assaf Zeevi, stochastic models and simulation. Published online in *Articles in Advance* September 21, 2012.

1. Introduction

The purpose of many stochastic simulations is to estimate the mathematical expectations of some cost functions. In finance, the prices of options and the Greeks can often be expressed as mathematical expectations under the risk-neutral measure (see Glasserman 2004). In most cases the expectations cannot be computed analytically, and one has to resort to Monte Carlo (MC) methods. After suitable transformations, the expectations can be written as integrals over the d-dimensional unit cube $(0,1)^d$, which can be approximated by an equal-weight rule

$$I(f) = \int_{(0,1)^d} f(\mathbf{u})\,d\mathbf{u} \approx \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{u}_k), \tag{1}$$

where the points $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are independent random samples from the uniform distribution on $(0,1)^d$. The MC methods attain a convergence rate $O(n^{-1/2})$ for square integrable functions. Because of their flexibility and simplicity, MC methods are widely used in finance (see Boyle et al. 1997, Glasserman 2004, Staum 2009). However, the convergence of MC methods is very slow.

Quasi–Monte Carlo (QMC) methods are deterministic versions of MC methods, which are also equal-weight rules of the form (1), but they use deterministic *low discrepancy points*. There are two main classes of low discrepancy points: the digital nets and good lattice points. The QMC error can be bounded using the Koksma-Hlawka inequality (see Niederreiter 1992)

$$\left| \int_{(0,1)^d} f(\mathbf{u})\,d\mathbf{u} - \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{u}_k) \right| \le V_{\mathrm{HK}}(f) D^*(\mathbf{u}_1, \ldots, \mathbf{u}_n),$$

where $V_{\mathrm{HK}}(f)$ is the *variation* of f in the sense of Hardy and Krause (which measures the variability of the integrand) and $D^*(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is the *star discrepancy* (which measures the uniformity of the point set). Several constructions of low discrepancy point sets are available, for which $D^*(\mathbf{u}_1, \ldots, \mathbf{u}_n) = O(n^{-1}(\log n)^d)$ (see Niederreiter 1992). Theoretically, the convergence for a QMC algorithm is $O(n^{-1}(\log n)^d)$ for functions with $V_{\mathrm{HK}}(f) < \infty$. This is *asymptotically* much better than that of MC. Note that the factor $(\log n)^d$ grows exponentially with increasing dimension; thus, some experts in number theory believed that QMC methods should not be used for high-dimensional integration. However, many experiments found that QMC substantially outperforms MC for a variety of high-dimensional problems in finance (see Joy et al. 1996, Ninomiya and Tezuka 1996, Paskov and Traub 1995). The theory of weighted function spaces and the concepts of effective dimension (Sloan and Woźniakowski 1998, Caflisch et al. 1997) offer partial explanations. Over the last two decades QMC methods have become an increasingly popular alternative to MC methods for the pricing and hedging of complex financial instruments. We refer to L'Ecuyer (2009) and Lemieux (2009) for recent reviews on QMC methods with applications in finance.

QMC methods are not a panacea for all high-dimensional integrals, and care must be taken in their applications. Two key factors could significantly affect the accuracy of QMC methods: the *discontinuity* and the *dimension* of the integrand. Both factors are especially pronounced in finance. The purpose of this paper is to show how the two fundamental limitations of QMC methods can be overcome for some problems in finance.

1.1. The Impact of Discontinuities in QMC Methods

Empirical studies in the nonfinancial literature indicate that discontinuities in the integrand may significantly deteriorate the performance of QMC methods (see Berblinger et al. 1997, Morokoff and Caflisch 1995, Moskowitz and Caflisch 1996). In particular, the shape of the boundary of jumps has a tremendous influence. For example, it has been observed that the performance of QMC methods for indicator functions of rectangles with sides parallel to the axes is much better than for nonrectangular indicator functions. For the latter, the advantage of QMC over MC is slight.

What can we learn from these empirical observations? The most important lesson is that not all kinds of discontinuities are adverse to QMC methods. Indeed, the discontinuities that are aligned with the coordinate axes can be *QMC friendly*, in the sense that high accuracy can still be expected (see §3 for more examples). This can be understood from the point of view of discrepancy, because discrepancy can be seen as the worst case integration error of a certain class of indicator functions and it is a key measure of the performance of QMC methods. Low discrepancy points (e.g., digital nets and good lattice points) have star discrepancy in the order $O(n^{-1}(\log n)^d)$. Note that star discrepancy is defined in terms of axis-parallel rectangles with one vertex at the origin. In contrast, the so-called isotropic discrepancy, which is defined in terms of convex subsets, has a much worse upper bound $O(n^{-1/d}\log n)$ (see Niederreiter 1992). Results on discrepancies vary, depending on the shape of the regions.

The difficulty with discontinuities severely hinders the usefulness of QMC methods to finance problems because discontinuities are extremely common in pricing options and estimating Greeks. Digital options and barrier options are examples of exotic derivatives with discontinuous payoff functions. Moreover, estimating Greeks by the pathwise method may induce discontinuities even though the underlying payoff functions are continuous (see an example in §4).

There have been some attempts to recover the superiority of QMC methods by smoothing the integrand. Moskowitz and Caflisch (1996) show a method of "smoothing" the integrand by enforcing continuity. Smoothing the integrand by taking a conditional expectation is popular for gradient estimation and for Greeks estimation (L'Ecuyer and Perron 1994, L'Ecuyer and Lemieux 2000, Glasserman 2004). For Value-at-Risk calculation, Fourier transform is used to smooth the indicator function (see Jin and Zhang 2006). In this paper, we offer a novel way to handle the discontinuities in pricing options and in estimating Greeks. This is related to the subject of the next subsection.

1.2. The Impact of Dimension and Path-Generation Methods

It is known that typical high-dimensional low discrepancy points are better distributed in their initial dimensions but exhibit poor uniformity in projections onto later dimensions (Morokoff and Caflisch 1994, Wang and Sloan 2008). This motivates us to improve the performance of QMC methods by reducing the effective dimension such that the variance of the function is concentrated on the first few variables and on the subsets of variables with small cardinality. In the context of financial applications, several methods have been proposed, such as the Brownian bridge (BB) (Caflisch et al. 1997, Moskowitz and Caflisch 1996), the principal component analysis (PCA) (Acworth et al. 1998), and the linear transformation (Imai and Tan 2006). We collectively refer these methods as *path-generation methods* (PGMs), because they are related to the generation of the underlying processes of the financial models (say, Brownian motions). Generalizations to some other processes are available (see Avramidis and L'Ecuyer 2006, Imai and Tan 2009).

The use of PGMs in QMC methods is delicate. Although the literature has documented the great success of good PGMs in enhancing QMC methods, it has also been pointed out that they need not always offer a consistent advantage over the standard construction (STD). For example, Papageorgiou (2002) showed that BB gives consistently worse results than STD for certain types of digital options. No formal explanation was given. This calls for a better understanding of the relationship between a PGM and the accuracy of QMC methods. The impact of a PGM on the effective dimension (an important indicator for characterizing the difficulty of integration) has been studied extensively (Wang and Fang 2003, Liu and Owen 2006, Wang 2006, Wang and Tan 2012). However, the impact of a PGM on the structure of the discontinuities is almost neglected in the literature.

We are interested in how PGMs could change the structure of the discontinuities and what the effect of such discontinuities is on the accuracy of QMC methods. It turns out that for pricing financial derivatives with discontinuous payoff functions, PGMs have a strong impact on the structure of discontinuities. The discontinuities can be QMC friendly or not, depending on the PGM and the problem at hand. The nature of discontinuities is crucial for the success of QMC methods. This critical observation, although seeming surprisingly simple, explains the possible failure of BB or PCA. This further motivates us to develop new PGMs that overcome these deficiencies by taking into account the structure of the discontinuities. Note that the existing PGMs do nothing to deal with discontinuities. Thus their usefulness for discontinuous functions needs further clarification.

1.3. Outline of the Paper

The remainder of this paper is organized as follows. In §2, we present some background on derivative pricing and introduce the concepts of effective dimension and some related quantities. In §3, we perform some numerical experiments to demonstrate that the PGMs commonly used in practice may have good or bad performance in QMC methods, depending on the nature of the payoff functions. We then offer an explanation for this phenomenon by analyzing the impact of each PGM on the structure of the discontinuities. In §4, we develop a new method of handling discontinuities by designing a PGM for simulating the underlying process. Extensive numerical experiments are performed to demonstrate the effectiveness of the proposed method for pricing exotic options with discontinuous payoffs and for estimating Greeks. Conclusions are presented in the last section. Appendices present all the simulated results.

2. Pricing, Hedging, Integration, and Effective Dimension

2.1. Pricing, Hedging, and High-Dimensional Integration

Consider the problem of pricing a financial derivative. Assume the payoff function $g(\mathbf{S})$ at the maturity T, with $\mathbf{S} := (S_1, \ldots, S_d)$, depends on the asset prices $S_j := S_{t_j}$ at equally spaced times $t_j = j\Delta t$ for $j = 1, \ldots, d$, where $\Delta t = T/d$. Assume that under the risk-neutral measure the underlying asset follows the geometric Brownian motion

$$dS_t = rS_t\,dt + \sigma S_t\,dB_t, \tag{2}$$

where r is the risk-free interest rate, σ is the volatility, and B_t is the standard Brownian motion. The analytical solution to (2) is given by

$$S_t = S_0 \exp\big((r - \sigma^2/2)t + \sigma B_t\big), \tag{3}$$

where S_0 is the asset price at time zero. Based on the risk-neutral valuation principle, the value of the financial derivative at time zero is $\mathbb{E}[e^{-rT} g(\mathbf{S})]$, where $\mathbb{E}[\,\cdot\,]$ is the expectation under the risk-neutral measure. For example, the price of a European arithmetic Asian option is $\mathbb{E}[e^{-rT}\max(S_A - K, 0)]$, where S_A is the arithmetic average of the underlying asset prices at times $t_1, \ldots, t_d$.

Moreover, under some regularity conditions, the pathwise or likelihood estimates of the Greeks can be written as expectations. For example, the pathwise estimate for the *delta* of the arithmetic Asian option mentioned above can also be written as (see Broadie and Glasserman 1996, Glasserman 2004)

$$\mathbb{E}\left[e^{-rT}\mathbf{I}_{\{S_A > K\}}(\mathbf{S})\frac{S_A}{S_0}\right], \tag{4}$$

where $\mathbf{I}_{\{\cdot\}}(\mathbf{S})$ is an indicator function, that is, $\mathbf{I}_{\{S_A > K\}}(\mathbf{S}) = 1$ if $S_A > K$, otherwise 0. The computations of the Greeks are important in financial risk management.

We now focus on pricing derivatives. Let $\mathbf{x} := \mathbf{B} := (B_1, \ldots, B_d)^T$ with $B_j = B_{t_j}$ for $j = 1, \ldots, d$. Then $\mathbf{x}$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{C}$; i.e., $\mathbf{x} \sim N(\mathbf{0}, \mathbf{C})$, where the entries of the covariance matrix $\mathbf{C}$ are given by

$$C_{ij} = \min(t_i, t_j).$$

From (3), we may express the payoff function $g(\mathbf{S})$ in terms of $\mathbf{x}$ as

$$\begin{aligned} g(\mathbf{S}) &= g\big(\exp(\mu_1 + \sigma x_1), \ldots, \exp(\mu_d + \sigma x_d)\big) \\ &=: G(\mathbf{x}), \quad \mathbf{x} \sim N(\mathbf{0}, \mathbf{C}), \end{aligned}$$

where $\mu_j = \log S_0 + (r - \sigma^2/2)\,t_j$. The value of the financial derivative at time $t = 0$ can then be expressed as a Gaussian integral

$$V(G) := \mathbb{E}[G(\mathbf{x})] = \frac{e^{-rT}}{(2\pi)^{d/2}\sqrt{\det \mathbf{C}}} \cdot \int_{\mathbb{R}^d} G(\mathbf{x}) \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right) d\mathbf{x}. \tag{5}$$

From the point of view of integration, by setting $\mathbf{x} = A\mathbf{z}$ with $AA^T = \mathbf{C}$ and then imposing the transformation $\mathbf{z} = \Phi^{-1}(\mathbf{u})$, the Gaussian integral (5) is transformed to

$$V(G) = \frac{e^{-rT}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} G(A\mathbf{z}) \exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right) d\mathbf{z}$$
$$= e^{-rT} \int_{(0,1)^d} G(A\Phi^{-1}(\mathbf{u}))\, d\mathbf{u},$$

where $\Phi^{-1}(\mathbf{u}) := (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))^T$ is the componentwise inverse of the standard normal cumulative distribution function. In a QMC setting, $V(G)$ is approximated by

$$V(G) \approx \frac{e^{-rT}}{n} \sum_{k=1}^{n} G(A\Phi^{-1}(\mathbf{u}_k)),$$

where $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a low discrepancy point set over the unit cube $(0,1)^d$.

The change of variables $\mathbf{x} = A\mathbf{z}$ with $AA^T = \mathbf{C}$ is equivalent to a PGM of the Brownian motion

$$(B_1, \ldots, B_d)^T = A(z_1, \ldots, z_d)^T, \quad (z_1, \ldots, z_d)^T \sim N(\mathbf{0}, I_d), \tag{6}$$

where I_d is the $d \times d$ identity matrix. Hence the matrix A is called a *generating matrix* of the Brownian motion. A key insight is that the generating matrix A can be arbitrary as long as it satisfies the decomposition $AA^T = \mathbf{C}$. Thus a PGM is essentially a change-of-variable technique or a decomposition of the covariance matrix $\mathbf{C}$.

We have a great deal of flexibility in choosing the generating matrix A. It is known that MC algorithms based on different PGMs are equivalent in the probabilistic sense, because the mean square error of MC is determined by the variance of the integrand, which is unchanged. However, PGMs have significant impact on QMC methods, because the resulting functions may have not only different effective dimension (to be defined below), but also different smoothness property, in particular, the structure of discontinuity (see §3.2). Both factors could significantly affect the accuracy of QMC methods because of the special distribution property of low discrepancy points.

Several constructions for the Brownian motion are widely used in practice. The simplest one is the STD (or called the Cholesky construction), which generates the Brownian motion sequentially: given $B_0 = 0$,

$$B_j = B_{j-1} + \sqrt{\Delta t} z_j, \quad z_j \sim N(0,1),\ j = 1, \ldots, d. \tag{7}$$

Other popular constructions are BB and PCA constructions. In BB construction, we first generate the value at the end, then the value in the middle conditional on that at the end, and so on (we refer to Glasserman 2004 for details). In a number of cases BB and PCA have been applied successfully. However, they do not guarantee good performance for all problems, especially for discontinuous payoff functions, as we will demonstrate. So they should not be applied blindly. The reason when and why a PGM may succeed or fail is one of the key topics we address (it sheds light on what a "good" financial derivative is for a given PGM). How to find a good PGM for discontinuous payoff functions is the other important topic we consider.

2.2. Effective Dimension

Consider a square integrable function $f(\mathbf{y})$ defined on $(0,1)^d$. We say that the expansion

$$f(\mathbf{y}) = \sum_{u \subseteq \{1,\ldots,d\}} f_u(\mathbf{y}), \quad \mathbf{y} = (y_1, \ldots, y_d)$$

is an ANOVA decomposition of f if the terms f_u satisfy $\int_0^1 f_u(\mathbf{y})\, dy_j = 0$ whenever $j \in u$ for $\varnothing \neq u \subseteq \{1, \ldots, d\}$. The ANOVA decomposition is orthogonal: $\int_{(0,1)^d} f_u(\mathbf{y}) f_v(\mathbf{y})\, d\mathbf{y} = 0$ for $u \neq v$. Based on the orthogonal property, we have the decomposition for the total variance

$$\sigma^2(f) = \sum_{\varnothing \neq u \subseteq \{1,\ldots,d\}} \sigma_u^2(f),$$

where $\sigma^2(f)$ and $\sigma_u^2(f)$ are the variances of $f(\mathbf{y})$ and $f_u(\mathbf{y})$, respectively.

We introduce the concepts of effective dimension following Caflisch et al. (1997). The *effective dimension in the truncation sense* of f is the smallest integer d_t such that $\sum_{u \subseteq \{1,\ldots,d_t\}} \sigma_u^2(f) \geq p\sigma^2(f)$. The *effective dimension in the superposition sense* of f is the smallest integer d_s such that $\sum_{|u| \leq d_s} \sigma_u^2(f) \geq p\sigma^2(f)$. Here p is a parameter close to one and $|u|$ is the cardinality of the set u. In this paper, we use the following effective dimension-related quantities. The *variance ratio* of order l is defined as

$$R_l = \frac{1}{\sigma^2(f)} \sum_{|u|=l} \sigma_u^2(f), \quad l = 1, \ldots, d. \tag{8}$$

The variance ratio of order 1 is called the *degree of additivity*; it measures the inherent additive structure. The *mean dimension* (in the superposition sense) is defined as (see Owen 2003)

$$d_{ms} = \sum_{l=1}^{d} l R_l. \tag{9}$$

If $R_1 \approx 1$ or $d_{ms} \approx 1$, then the function is nearly additive, for which QMC methods are especially suitable due to the perfect one-dimensional projections of low discrepancy points. If $R_1 \ll 1$ or $d_{ms} \gg 1$, the function is dominated by higher order ANOVA terms, for which QMC methods usually behave no better than MC methods. The degree of additivity and the mean dimension can be computed numerically (Wang and Fang 2003, Liu and Owen 2006).

Many finance problems are typically of small effective dimension in either the truncation or superposition sense (Caflisch et al. 1997, Wang and Fang 2003).

Such functions are believed to be easier to integrate numerically by QMC methods. Thus, the effective dimension and the related quantities are useful in measuring the difficulty of high-dimensional integration and in understanding the success of QMC methods in high dimensions. Despite this, we should not overestimate their role, because they may not characterize the QMC error precisely in some cases. The exact theoretical relationship between the accuracy of QMC methods and the effective dimension (or mean dimension) has not yet been established. Moreover, we will demonstrate that the impact of the discontinuities can be even more pronounced than that of the effective dimension.

3. The Impact of Path-Generation Methods

In §3.1, we perform several warm-up numerical experiments on pricing options with discontinuous payoff functions under the Black-Scholes model (2). Our purpose is to get some intuitions about the different performance of various PGMs in QMC methods and to demonstrate that each PGM can easily lose its effectiveness for discontinuous functions. An insightful analysis will be presented in §3.2 for understanding why a PGM may have good or bad performance.

3.1. No PGM Offers a Consistent Advantage

We consider three examples. The first example is the digital option studied in Papageorgiou (2002). The second example is an Asian option with a knock-out feature at the maturity, studied in Glasserman et al. (1999). The last example is a toy option example similar to the second one, except with a modified knock-out feature.

Example A1 (Digital Option). The payoff function at the maturity T of a digital option is defined as

$$g_1(\mathbf{S}) = \frac{1}{d}\sum_{j=1}^{d} S_j \mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}). \tag{10}$$

This option is of interest as it was used in Papageorgiou (2002) to dispel the general belief that BB always outperforms STD.

Example A2 (Asian Options with Knock-Out Feature at the Maturity). The payoff function of the option is

$$g_2(\mathbf{S}) = \max(S_A - K, 0)\mathbf{I}_{\{S_d \le H\}}(\mathbf{S}), \tag{11}$$

where

$$S_A = \sum_{j=1}^{d} w_j S_j \quad \text{with } \sum_{j=1}^{d} w_j = 1, \tag{12}$$

K is the strike price, and H is the barrier. If the price of the underlying asset at the maturity T is below the barrier H, then the option pays $\max(S_A - K, 0)$, as with an ordinary Asian call option; if the terminal price is above the barrier H, then the option pays nothing.

Example A3 (Asian Options with a Modified Knock-Out Feature). Let $\mathbf{V}_1$ be the first column of the generating matrix corresponding to PCA (see Glasserman 2004). Let $\mathbf{V} := b\,\mathbf{V}_1 = (v_1, \ldots, v_d)^T$, satisfying $\sum_{j=1}^{d} v_j = 1$. The payoff function of the option is defined as

$$g_3(\mathbf{S}) = \max(S_A - K, 0)\mathbf{I}_{\{\bar{S} \le H\}}(\mathbf{S}) \quad \text{with } \bar{S} = \prod_{j=1}^{d} S_j^{v_j}. \tag{13}$$

Here $\bar{S}$ is a specific weighted geometric average of asset prices with respect to the weights v_j's.

The parameter values we assume are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, and $T = 1$. For Examples A2 and A3, the strike price is $K = 100$ and the knock-out level is $H \in \{120, 140, 160\}$. We choose equal weights $w_j = 1/d$ in the average S_A. We consider dimensions $d = 16$, 64, and 128 (and also dimension $d = 2$ for Example A1).

Our numerical experiments are conducted as follows. To estimate the price of an option by QMC simulation, we need to simulate a number of paths of the underlying asset, compute the discounted payoff corresponding to each path, and then average the results. We simulate the asset prices by STD, BB, and PCA using the Sobol′ points (Sobol′ 1967) or the Korobov lattice points with good two-dimensional projections (Wang 2007). To assess the accuracy of a QMC estimate, we use *digit shift in base 2* for Sobol′ points and random shift modulo 1 for good lattice points (see L'Ecuyer and Lemieux 2002 for a review of randomized QMC). The variance of a QMC estimate is estimated based on $m = 100$ replications, with each application consisting of $n = 4{,}096$ paths for the Sobol′ points and $n = 4{,}001$ paths for the Korobov points. The variance of MC estimate is based on mn samples (with $n = 4{,}096$). The variance reduction factor (VRF) of a QMC estimate is calculated as the ratio of the variance of MC estimate (with STD) to the variance of the corresponding QMC estimate. The comparisons on VRFs are summarized in Tables A.1–A.3 in Appendix A. We observe the following:

- The relative performances of STD, BB, and PCA in QMC methods for discontinuous functions are quite different for different examples. No method offers a consistent advantage.
- For Example A1, STD is remarkably effective in QMC methods. This is supported by the highest VRFs, and this holds irrespective of the nominal dimension and the low discrepancy point set (the Korobov lattice points seem to give better results than the Sobol′ points). BB and PCA in QMC methods behave consistently worse than STD, not only when the dimension is large, but also when the dimension is as small as two. This highlights that the key factor that determines the different performance of various

PGMs in QMC methods seems something beyond the nominal dimension.

• For Example A2, QMC-based BB is the most effective, irrespective of the nominal dimension and the type of low discrepancy point set, and it remarkably outperforms STD and PCA. STD or PCA in combination with the Sobol′ points or the Korobov points behaves no better or only marginally better than MC.

• For Example A3, PCA with both the Sobol′ points and the Korobov points has the most remarkable success. Most of the VRFs for PCA are in the order of thousands. STD and BB in QMC only lead to moderate VRFs (fewer than 100 in all cases).

The experiments above demonstrate the remarkable differences of the performance of STD, BB, and PCA in QMC methods for different problems with different nature of discontinuities. Any one of the three PGMs can outperform the other two, depending on the problems at hand. In the next subsection, we will try to understand the reason for their different performance in QMC methods. The analysis will shed light on what a "good" financial derivative or a class of financial derivatives is for which a given PGM performs well in QMC methods.

3.2. The Impact of PGMs on the Structure of the Discontinuities

We have demonstrated that there is no uniformly superior PGM in QMC methods. A PGM that works very well in one example may perform badly in another. It is therefore of significant interest to have a better understanding on the mechanism of how a PGM interacts with the underlying problem and what the determinants for the accuracy of QMC methods are. It turns out that PGMs have significant impact on the structure of the discontinuities, on which the accuracy of QMC methods depends crucially. To demonstrate this explicitly, we revisit the examples above and analyze what is the structure of the discontinuities associated with each PGM.

Example A1 (Revisit). It is instructive to first analyze the discontinuities of the following indicator function that is involved in the payoff function (10) of a digital option

$$h_1(\mathbf{S}) = \mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}), \quad \text{for some } j = 1, \dots, d, \tag{14}$$

where

$$S_j = S_j(\mathbf{B}) = S_0 \exp\big((r - \sigma^2/2) j\Delta t + \sigma B_j\big), \tag{15}$$

and B_j is generated by (6); i.e.,

$$B_j = a_{j,1} z_1 + \cdots + a_{j,d} z_d, \quad \mathbf{z} = (z_1, \dots, z_d)^T \sim N(\mathbf{0}, I_d).$$

Here $a_{j,k}$ is the (j,k) entry of the generating matrix A, which satisfies $AA^T = \mathbf{C}$. We may view the indicator function $\mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S})$ as a function of $\mathbf{z}$; i.e.,

$$\mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}) = \mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}(\mathbf{B})) = \mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}(A\mathbf{z})).$$

The jumps of this indicator function occur when

$$S_{j-1}(\mathbf{B}) = S_j(\mathbf{B}).$$

Based on (15), this condition is equivalent to

$$B_{j-1} - B_j = \frac{\Delta t}{\sigma}(r - \sigma^2/2) =: C_1. \tag{16}$$

When the paths are generated by STD, we have (see (7))

$$B_j = B_{j-1} + \sqrt{\Delta t} z_j, \quad z_j \sim N(0,1).$$

For this particular case, the condition (16) is equivalent to

$$z_j = -\frac{1}{\sqrt{\Delta t}} C_1 =: C_2. \tag{17}$$

An important observation is that when the paths are generated by STD, the discontinuities of the indicator function $\mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}(A\mathbf{z}))$ occur only on the axis-parallel hyperplane (17). By componentwise inverse normal transformation, we have

$$\mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}(A\mathbf{z})) = \mathbf{I}_{\{S_j > S_{j-1}\}}\big(\mathbf{S}(A\Phi^{-1}(\mathbf{u}))\big) =: H_1(\mathbf{u}),$$
$$\mathbf{u} = (u_1, \dots, u_d)^T.$$

From (17), the resulting function $H_1(\mathbf{u})$ has jumps on the axis-parallel hyperplane

$$u_j = \Phi(C_2).$$

This kind of discontinuities is QMC friendly from the point of view of QMC integration.

However, when the paths are generated by BB or PCA or by any other method with a generating matrix A, the condition (16) is equivalent to

$$\beta_{j,1} z_1 + \cdots + \beta_{j,d} z_d = C_1, \tag{18}$$

where $\beta_{j,k} = a_{j-1,k} - a_{j,k}$, $k = 1, \dots, d$ (for convention, if $j = 1$, we understand that $a_{j-1,k} = 0$). In such cases, the discontinuities of the indicator function $\mathbf{I}_{\{S_j > S_{j-1}\}}(\mathbf{S}(A\mathbf{z}))$ occur on the hyperplane (18). When transformed to the unit cube by the componentwise inverse normal transformation, the hyperplane (18) is transformed to the manifold with a nonlinear equation

$$\beta_{j,1}\Phi^{-1}(u_1) + \cdots + \beta_{j,d}\Phi^{-1}(u_d) = C_1. \tag{19}$$

Consequently, the discontinuities are in general no longer QMC friendly.

The payoff function (10) of the digital option in Example A1 is essentially the sum of the indicator functions of the form (14). If the paths are generated by STD, the discontinuities of the resulting function are aligned with the coordinate axes, which are QMC friendly, and hence good performance can be expected in QMC methods. In contrast, if the paths are generated by BB or PCA, the resulting integrand has jumps on the manifolds (19) for $j = 1, \dots, d$, which are no longer QMC friendly. We believe this is the reason why in QMC setting STD is remarkably successful for the digital option, whereas both BB and PCA are not so successful. Thus the digital option is STD friendly (in the sense that QMC-based STD has good performance for the digital option).

Example A2 (Revisit). To gain insight to the discontinuities inherent in the knock-out Asian option (11) in Example A2, it suffices to analyze the following indicator function:

$$h_2(\mathbf{S}) = \mathbf{I}_{\{S_d \le H\}}(\mathbf{S}) = \mathbf{I}_{\{S_d \le H\}}(\mathbf{S}(A\mathbf{z})) \\ = \mathbf{I}_{\{S_d \le H\}}\big(\mathbf{S}(A\Phi^{-1}(\mathbf{u}))\big) =: H_2(\mathbf{u}).$$

Its discontinuities occur when

$$S_d = H.$$

This condition is equivalent to

$$B_d = \big[\log(H/S_0) - (r - \sigma^2/2)T\big]\big/\sigma =: C_3. \qquad (20)$$

When the paths are generated by BB, we first generate the value B_d at the end; i.e.,

$$B_d = \sqrt{T} z_1, \quad z_1 \sim N(0, 1).$$

This implies that the condition (20) is equivalent to

$$z_1 = C_3/\sqrt{T}.$$

Transforming by the inverse normal transformation, this is equivalent to

$$u_1 = \Phi\big(C_3/\sqrt{T}\big).$$

Thus, if the paths are generated by BB, the discontinuities of the indicator function $H_2(\mathbf{u})$ are aligned with the coordinate axes, which are QMC friendly.

On the other hand, if the paths are generated by STD or PCA, or by any other method with a generating matrix A, then

$$B_d = a_{d,1} z_1 + \cdots + a_{d,d} z_d.$$

In this case, the condition (20) is equivalent to

$$a_{d,1} z_1 + \cdots + a_{d,d} z_d = C_3,$$

and this is transformed to a nonlinear equation (if more than one coefficient is nonzero)

$$a_{d,1}\Phi^{-1}(u_1) + \cdots + a_{d,d}\Phi^{-1}(u_d) = C_3.$$

This is the manifold on which the jumps of the indicator function $H_2(\mathbf{u})$ occur.

The jumps of the payoff function (11) of an Asian option with a knock-out feature at the maturity in Example A2 occur on the same manifold as the indicator function $H_2(\mathbf{u})$. This explains the superiority of BB for this option in QMC methods. This option is BB friendly.

Finally, we can similarly analyze how the choices of PGM affect the structure of the discontinuities for Example A3. It turns out that if the paths are generated by PCA, then the discontinuities are aligned with the coordinate axes, which are QMC friendly. If the paths are generated by STD or BB, then the discontinuities are unfriendly for QMC. This explains the superiority of PCA for this option in QMC methods. This option is PCA friendly.

The structure of the discontinuities induced by a PGM is usually implicit and requires us to analyze each problem separately. The analysis above clearly shows that the structure of the discontinuities can be quite different when different PGMs are used for the same problem or when the same PGM is used for different problems. The most significant difference is that the discontinuities can be QMC friendly or not. We find that the PGMs that offer the best performance in QMC are those that make the discontinuities QMC friendly. This is consistent with the general belief in QMC community that QMC methods usually enjoy superiority over MC for functions with QMC-friendly discontinuities. This can be understood (at least partially) from the point of view of discrepancy, which is a key factor in gauging the performance of QMC methods. For a low-discrepancy point set, the star discrepancy or extreme discrepancy, which is based on axis-parallel rectangles, behaves much better than the isotropic discrepancy that is based on convex subsets (see Niederreiter 1992). Discontinuity property turns out to be a dominant factor that determines the accuracy of QMC methods. This explains why different PGMs can have different performance for the same problem in QMC methods. In particular, this explains why BB or PCA does not offer a consistent advantage in QMC methods.

4. Handling Discontinuities with a Good PGM

We have analyzed how PGMs affect the structure of the discontinuities and how the discontinuities affect the accuracy of QMC methods. The analysis and observations indicate that not all discontinuities are detrimental to QMC methods. Discontinuities that are aligned with the coordinate axes are QMC friendly, because good performance of QMC can be expected. This motivates us to propose a novel way to recover the superiority of QMC methods for discontinuous functions, namely, to align the discontinuities with the coordinate axes, as much as possible. Can we design a PGM that exploits the QMC friendly discontinuities? This problem can be hard, if not impossible. However, in some special cases a simple solution to the problem is available.

4.1. Discontinuous Functions Involving an Indicator Function

Many functions in pricing and hedging involve indicator functions. Usually, the payoff function of a financial option may be written as

$$g(\mathbf{S}) = f(\mathbf{S})\mathbf{I}_{\{h(\mathbf{S})>0\}}(\mathbf{S}), \tag{21}$$

where $f(\mathbf{S})$ is the payoff of the option if it is exercised; $h(\mathbf{S}) > 0$ defines the exercising condition. For example, the payoff of a binary (or digital) Asian option is $g(\mathbf{S}) = \mathbf{I}_{\{S_A<K\}}(\mathbf{S})$, where S_A is the weighted arithmetic average of the asset prices defined in (12). Moreover, indicator functions naturally appear in the problems of estimating Greeks, even when the original payoff functions are continuous. For example, the pathwise estimate for the delta of an arithmetic Asian option with payoff $\max(S_A - K, 0)$ is given in (4).

Recall that under the Black-Scholes model (2), $\mathbf{S}$ can be expressed in terms of $\mathbf{x} \sim N(\mathbf{0}, \mathbf{C})$; thus, $h(\mathbf{S})$ is a function of $\mathbf{x}$. Below we consider an indicator function of the form

$$\Lambda(\mathbf{x}) = \mathbf{I}_{\{h(\mathbf{q}^T\mathbf{x})<H\}}(\mathbf{x}), \quad \mathbf{x} = (x_1, \ldots, x_d)^T \sim N(\mathbf{0}, \mathbf{C}), \tag{22}$$

where $\mathbf{q} = (q_1, \ldots, q_d)^T$ is a vector of constants and $h(\cdot)$ is a function defined on $\mathbb{R}$. This kind of indicator function is of interest, because the payoff functions of many derivatives can be expressed in terms of (22). In particular, all the indicator functions considered in the previous section have this form. For example, for the indicator function $\mathbf{I}_{\{S_j>S_{j-1}\}}(\mathbf{S})$ involved in the digital option (10), because

$$S_j > S_{j-1} \iff \sigma B_{j-1} - \sigma B_j < (r - \sigma^2/2)\Delta t, \quad \mathbf{B} = (B_1, \ldots, B_d)^T \sim N(\mathbf{0}, \mathbf{C}), \tag{23}$$

the indicator function $\mathbf{I}_{\{S_j>S_{j-1}\}}(\mathbf{S})$, when viewed as a function of $\mathbf{x} := \mathbf{B}$, has the form (22).

The key to designing a good PGM for a function involving an indicator function (22) is to recognize that the discontinuities associated with the indicator function $\Lambda(\mathbf{x})$ depend only on a linear combination of the variables $x_1, \ldots, x_d$ and that $\mathbf{q}^T\mathbf{x}$ is a normally distributed random variable. This implies that we can exploit its intrinsic one-dimensionality.

Theorem 1. *Let $\mathbf{C}$ be a $d \times d$ positive definite matrix and let A_0 be a fixed decomposition matrix such that $A_0A_0^T = \mathbf{C}$. Suppose that the indicator function $\Lambda(\mathbf{x})$ has the form (22), where $\mathbf{q} \in \mathbb{R}^d$ is a nonzero vector of constants and $h(\cdot)$ is a function defined on $\mathbb{R}$. If U is a $d \times d$ orthogonal matrix, whose first column $\mathbf{U}_1$ is given by*

$$\mathbf{U}_1 = \frac{1}{D}A_0^T\mathbf{q}, \tag{24}$$

where $D := \sqrt{\mathbf{q}^T\mathbf{C}\mathbf{q}}$ and the remaining columns of U are arbitrary as long as they satisfy the orthogonality conditions, then by the transformation $\mathbf{x} = A_0U\mathbf{z}$, the function $h(\mathbf{q}^T\mathbf{x})$ involved in the indicator function $\Lambda(\mathbf{x})$ is transformed to a function depending only on the first component of $\mathbf{z}$:

$$h(\mathbf{q}^T\mathbf{x}) = h(Dz_1).$$

Consequently, the indicator function $\Lambda(\mathbf{x})$ is transformed to

$$\Lambda(\mathbf{x}) = \mathbf{I}_{\{h(Dz_1)<H\}}(\mathbf{z}), \quad \mathbf{z} = (z_1, \ldots, z_d)^T \sim N(\mathbf{0}, I_d).$$

If $h(\cdot)$ is strictly increasing on $\mathbb{R}$ and if $\mathbf{I}_{\{h(Dz_1)<H\}}(\mathbf{z})$ is further transformed by the inverse normal transformation $\mathbf{z} = \Phi^{-1}(\mathbf{u})$, then the indicator function $\Lambda(\mathbf{x})$ is transformed to a one-dimensional function:

$$\Lambda(\mathbf{x}) = \mathbf{I}_{\{u_1<c\}}(\mathbf{u}), \quad \mathbf{u} = (u_1, \ldots, u_d)^T \sim U(0,1)^d,$$

where $c = \Phi(D^{-1}h^{-1}(H))$ is a constant. The discontinuities of the indicator function $\mathbf{I}_{\{u_1<c\}}(\mathbf{u})$ are aligned with the coordinate axes, which are QMC friendly.

Proof. From (24) we have

$$\mathbf{q}^T A_0 = D\mathbf{U}_1^T.$$

Denote the columns of U by $\mathbf{U}_1, \ldots, \mathbf{U}_d$. Under the transformation $\mathbf{x} = A_0U\mathbf{z}$, we have

$$\mathbf{q}^T\mathbf{x} = \mathbf{q}^T A_0U\mathbf{z} = D\mathbf{U}_1^T(\mathbf{U}_1z_1 + \mathbf{U}_2z_2 + \cdots + \mathbf{U}_dz_d) = Dz_1,$$

by the orthogonality of the columns $\mathbf{U}_1, \ldots, \mathbf{U}_d$ of U. Thus,

$$h(\mathbf{q}^T\mathbf{x}) = h(Dz_1).$$

This proves the first part of the theorem.

The remaining results follow from the following equivalences (under the transformations $\mathbf{x} = A_0U\mathbf{z}$ and $\mathbf{z} = \Phi^{-1}(\mathbf{u})$) that

$$\{h(\mathbf{q}^T\mathbf{x}) < H\} \iff \{h(Dz_1) < H\} \iff \{u_1 < c\}. \quad \square$$

Theorem 1 indicates the possibility of aligning the discontinuities with the coordinate axes by suitably simulating the paths of the Brownian motion. We refer to the new PGM in Theorem 1 as the *orthogonal transformation* (OT). The OT method is simple to implement. We just need to construct an orthogonal matrix U according to Theorem 1 and take the generating matrix in (6) to be $A = A_0U$ for some fixed A_0 satisfying $A_0A_0^T = \mathbf{C}$. Whenever a function involves an indicator function of the form (22), Theorem 1 guarantees that the discontinuities are aligned with the coordinate axes. When a function involves an indicator function that is not exactly the form (22) but is "close" to this form in some sense, Theorem 1 is still useful in finding a good PGM, as we will demonstrate in Examples C1 and C2. Once the difficulty with the discontinuities has been overcome (at least partially), good performance of QMC methods can be expected. The value of the simple case in Theorem 1 is that it can give insight.

From a practical point of view, two issues are worth mentioning. Theorem 1 gives us only the first column of U. Other columns can be found by the Gram-Schmidt method or modified Gram-Schmidt algorithm (see Wang and Sloan 2011). The determination of other columns also leaves room for further optimization of the generating matrix by taking into account the knowledge of the function. Another issue is the choice of the initial decomposition matrix A_0. If the underlying integrand is, say, a product of an indicator function of the form (22) with another function $G_0(\mathbf{x})$—i.e., $G(\mathbf{x}) = G_0(\mathbf{x})\Lambda(\mathbf{x})$—then the choice of A_0 could impact the practical performance of QMC methods, because Theorem 1 focuses only on the indicator function $\Lambda(\mathbf{x})$. If there is an indication that the function $G_0(\mathbf{x})$ is PCA friendly (in the sense that PCA has good QMC performance for $G_0(\mathbf{x})$), then we may choose the initial decomposition matrix A_0 to be A^{PCA} (the generating matrix corresponding to PCA). If no prior information is available, then A_0 is taken to be the Cholesky decomposition of $\mathbf{C}$. In our examples, we often face an arithmetic average of asset prices S_A with equal weights or increasing weights defined in (12), which is PCA friendly according to Wang and Sloan (2011); thus we choose A_0 to be A^{PCA} (if not indicated otherwise). For equal time intervals, A^{PCA} can be obtained analytically (see Glasserman 2004).

To demonstrate the practical efficiency of the proposed OT method, we perform numerical experiments in two parts using two sets of examples (they differ in whether the involved indicator functions have the form (22) studied in Theorem 1). The efficiencies of various PGMs in QMC methods are assessed by comparing their VRFs (with respect to MC with STD). The VRFs are computed in the same way as in §3 using the same number of points. The degree of additivity and the mean dimension (see (8) and (9)) are computed numerically using the algorithms in Wang and Fang (2003) and Liu and Owen (2006) (but the results are presented only for the second set of examples because of space limitations). The dynamics of the underlying asset are assumed to follow the Black-Scholes model (2).

4.2. Numerical Experiments—Part I

This part of the experiments concerns the pricing of Asian options with different knock-out features.

Example B1. Consider an option with a payoff function defined by

$$g(\mathbf{S}) = \max(S_A - K, 0)\mathbf{I}_{\{S_{j_0} > S_{i_0}\}}(\mathbf{S}), \quad \text{for some } 0 \le i_0 < j_0 \le d.$$

Based on an equivalence relationship similar to (23), the involved indicator function $\mathbf{I}_{\{S_{j_0} > S_{i_0}\}}(\mathbf{S})$ has the form (22), corresponding to the vector $\mathbf{q}$ with $q_{i_0} = 1$, $q_{j_0} = -1$, and $q_j = 0$ for $j \neq i_0, j_0$.

Example B2. Consider a modification of option (13) with payoff at maturity given by

$$g(\mathbf{S}) = \max(S_A - K, 0)\mathbf{I}_{\{S_G < H\}}(\mathbf{S}),$$

where $S_G = \prod_{j=1}^{d} S_j^{v_j}$ (with $\sum_{j=1}^{d} v_j = 1$) is the weighted geometric average of the underlying asset prices with respect to the weights $v_1, \ldots, v_d$. Because the average S_G can be written as

$$S_G = \exp(a + \sigma \mathbf{v}^T \mathbf{x}), \quad \mathbf{v} = (v_1, \ldots, v_d)^T, \mathbf{x} \sim N(\mathbf{0}, \mathbf{C}),$$

with $a := \log S_0 + (r - \sigma^2/2)\sum_{j=1}^{d} v_j t_j$, the indicator function $\mathbf{I}_{\{S_G < H\}}(\mathbf{S})$ has the form (22), corresponding to the vector $\mathbf{q} = \mathbf{v}$.

In both of these examples, the involved indicator functions have the form (22); thus Theorem 1 can be used directly to obtain the generating matrix of the OT method. The resulting discontinuities of using OT are aligned with the coordinate axes, which are QMC friendly. On the contrary, STD, BB, and PCA do not necessarily have this nice property (except in some special cases). In our numerical illustrations, the model parameters are chosen to be

$$S_0 = 100, \quad \sigma = 0.2, \quad r = 0.1,$$
$$T = 1, \quad K \in \{90, 100, 110\}, \quad H \in \{120, 140, 160\}.$$

In Example B1, we set $i_0 = d/4 - 1$ and $j_0 = 3d/4 - 1$. In Example B2, only $K = 100$ is considered. The dimension is $d \in \{16, 64, 128\}$. We choose equal weights $w_j = v_j = 1/d$ $(j = 1, \ldots, d)$ in both averages S_A and S_G. The number of runs is large enough to provide accurate estimates for both the price and its variance. For example, the 99% confidence intervals for the price with crude MC are typically within ± 0.05, and they are better or much better with QMC. The comparisons on VRFs are presented in Tables B.1 and B.2 in Appendix B. We observe the following:

- For both examples, the proposed OT method is the most superior in terms of VRFs, irrespective of the nominal dimension, the strike price, the barrier level, and the type of low discrepancy point set. The VRFs of using OT in QMC can be as large as hundreds or even thousands. On the other hand, the other three PGMs with the Sobol′ points or the Korobov points provide only a moderate or small variance reduction.
- For Example B2, the knock-out level H has a large impact on the performance of PGMs. For the OT method, larger VRF is achieved for smaller knock-out level H, whereas for the other three PGMs, the inverse is true. By the formulation of the indicator function $\mathbf{I}_{\{S_G < H\}}(\mathbf{S})$, when the knock-out level H decreases, the severity of discontinuity increases. This increases the

difficulty of the problem for QMC with the three common PGMs (as expected), but this reduces the difficulty for QMC with the OT method. Thus the OT method is even more attractive when the severity of discontinuity increases.

The numerical results highlight the significance of aligning the discontinuities with the coordinate axes for the success of QMC methods. Numerical results on the degree of additivity and the mean dimension (not presented here) show that the OT method, in addition to making the discontinuities QMC friendly, has an additional effect, namely, it increases the degree of additivity and reduces the mean dimension. Thus the OT method overcomes the two fundamental limitations of QMC simultaneously—the discontinuity and the dimension—at least in these examples. By aligning the discontinuities with the coordinate axes, the OT method changes the function to be well suited to QMC methods, and thus it enjoys a superiority over other PGMs. However, the relative inferior performances of STD, BB, and PCA are attributed to the unfriendly discontinuities. Note that rigorous theoretical QMC error bounds require sufficient smoothness of the functions, and the QMC theory cannot be applied to our applications because of the singularities of the integrands.

4.3. Numerical Experiments—Part II

This part of the experiment concerns the pricing of binary (or digital) Asian options and the estimation of Greeks. Estimating Greeks is among one of the greatest challenges in computational finance. We focus on delta (the conclusions for *vega*, *rho*, and *theta* are similar).

Example C1. Consider the following binary (or digital) Asian option with the payoff

$$g(\mathbf{S}) = \mathbf{I}_{\{S_A<K\}}(\mathbf{S}). \tag{25}$$

Example C2. Consider the problem of estimating the delta of an arithmetic Asian option with payoff $\max(S_A - K, 0)$. The pathwise estimate for its delta is (see Broadie and Glasserman 1996, Glasserman 2004)

$$g(\mathbf{S}) = e^{-rT}\mathbf{I}_{\{S_A>K\}}(\mathbf{S})\frac{S_A}{S_0}. \tag{26}$$

Similar indicator functions appear in pathwise estimates of other Greeks, such as *vega*, *rho*, and *theta* (see Glasserman 2004). We emphasize that the payoff function of the arithmetic Asian option is continuous, but using pathwise approach to estimate Greeks induces discontinuities.

The indicator functions involved in Examples C1 and C2 are of greater interest as they do not conform to the form (22) and hence Theorem 1 cannot be applied directly. We circumvent this problem using a suboptimal strategy, namely, seeking a good PGM for another "auxiliary" indicator function whereby the OT method can be implemented easily. The resulting generating matrix, which is QMC friendly to the auxiliary problem, is applied to the original problem of interest. If a PGM works well for an auxiliary problem, then one hopes it also works well for a similar problem. Naturally, the success of this strategy crucially depends on how "similar" the auxiliary problem is to the original one.

For the indicator functions $\mathbf{I}_{\{S_A<K\}}(\mathbf{S})$ involved in (25), because its discontinuities depend on the weighted arithmetic average S_A, a natural auxiliary indicator function is $\mathbf{I}_{\{S_G<K\}}(\mathbf{S})$, whose discontinuities depend on the weighted geometric average S_G. Recall that we have already shown in Example B2 that Theorem 1 can be applied to the auxiliary indicator function $\mathbf{I}_{\{S_G<K\}}(\mathbf{S})$, transforming the function S_G to univariate. Under the same transformation we can expect that $S_A = f_1(z_1) + f_{\text{rest}}(\mathbf{z})$, with $f_1(z_1)$ being a univariate function and $f_{\text{rest}}(\mathbf{z})$ being "small" in variance (Wang 2006). Therefore, $\mathbf{I}_{\{S_A<K\}}(\mathbf{S})$ is "similar" to $\mathbf{I}_{\{f_1(z_1)<K\}}(\mathbf{S})$, whose discontinuities are QMC friendly. In other words, using OT we can somehow "localize" the discontinuities. The same strategy can be used to the indicator function $\mathbf{I}_{\{S_A>K\}}(\mathbf{S})$ involved in the delta estimate (26) in Example C2.

In these experiments, we use the same set of parameter values as in Example B1. The weights involved in S_A and S_G are assumed to be one of the following:

(A) $w_j = v_j = 1/d, \quad j = 1, \ldots, d;$

(B) $w_j = v_j = b2^j, \quad j = 1, \ldots, d, \text{ with } \sum_{j=1}^{d} w_j = 1.$

The 99% confidence intervals for the price or the delta value with crude MC are typically within ± 0.001, and they are better or much better with QMC. The comparisons on VRFs are presented in Tables C.1 and C.2, and the comparisons on the degree of additivity and the mean dimension are presented in Tables C.3 and C.4 in Appendix C. We observed the following:

- For the pricing of the binary Asian option and the estimation of delta, the OT method has the best performance with both the Sobol' points and the Korobov points, leading to large VRFs (up to thousands), especially for the increasing weights (B). The other three PGMs in QMC methods provide only a moderate or small variance reduction.
- The impact of weights on the effectiveness of OT is significant. The efficiency gain of QMC-based OT for weights (A) is significantly larger than for weights (B). STD seems to be invariant to the weights, BB seems to have greater success with weights (B), and PCA seems to work better with weights (A). In any case, all three common PGMs (i.e., STD, BB, and PCA) in QMC methods are only marginally more efficient than MC.

• In addition to yielding the highest precision for both examples, the OT method leads to the largest degree of additivity (larger than 0.98) and the smallest mean dimension (smaller than 1.03), implying that the obtained functions are nearly additive and hence are well suited to QMC methods. In contrast, other PGMs lead to a smaller (or much smaller) degree of additivity and a larger (or much larger) mean dimension, implying that high-order ANOVA terms are important in the resulting functions.

These examples show that in some cases we cannot make the discontinuities exactly aligned with the coordinate axes; however, by using Theorem 1 we can meticulously "localize" the discontinuities, nearly aligning the discontinuities with the axes. Furthermore, an additional benefit of using the OT method is that it increases the degree of additivity and significantly reduces the mean dimension. The joint effect of localizing the discontinuities and reducing the mean dimension makes the OT method very attractive: a significant gain in efficiency could still be achieved in QMC methods (it is difficult to determine how much there is to gain by making discontinuities QMC friendly, and by reducing the mean dimension, respectively; our experience indicates that the structure of the discontinuities seems to be more important, at least in our examples). The success of the OT method in these examples suggests that it may have a much wider range of applications than stated in Theorem 1. Its success for the estimation of delta is even more reassuring, as the estimation of Greeks is a challenging numerical problem.

The idea of localizing the discontinuities used in Examples C1 and C2 allows us to get a sense of how to apply the same principle to other problems. In particular, it motivates a way of handling the discontinuities involved in a more complicated indicator function of the form $\mathbf{I}_{\{h(\mathbf{S})>0\}}(\mathbf{S})$ in (21), where the function $h(\mathbf{S})$ can have a much more general structure than that considered in (22). We may try to reduce the effective dimension of the function $h(\mathbf{S})$, such that the discontinuities are, as much as possible, aligned with coordinate axes. For example, if $h(\mathbf{S})$ is smooth enough, we may use the linear transformation method of Imai and Tan (2006) on $h(\mathbf{S})$ to reduce its truncation dimension and to find the required generating matrix. This general principle enables us to deal with discontinuities in a much wider range of problems. The efficiency of this general approach remains to be investigated in future work.

5. Conclusions

Many financial derivatives have discontinuities in their payoff functions or Greeks. Discontinuities could severely degrade the performance of QMC methods for pricing or hedging such derivatives. This paper studied the effect of PGMs on the structure of the discontinuities and how to recover the superiority of QMC methods on problems with discontinuities.

The contribution of this paper is twofold. First, it is found that for discontinuous payoff functions, PGMs have a great impact on the structure of the discontinuities. The discontinuities can be QMC friendly or not, depending on the PGM and the problem at hand, and they can significantly affect the accuracy of QMC methods. It is demonstrated that the PGMs that offer the best performance in QMC methods are those that make the discontinuities QMC friendly. For the important example of a digital option, a detailed analysis showed that BB and PCA implicitly induce irregular discontinuities. This provides insight on why BB and PCA do not offer consistent advantages in QMC methods. This critical observation motivated us to exploit the freedom in designing PGMs to handle the discontinuities.

The second contribution of the paper is to propose a new PGM for a special kind of discontinuous functions. The core of the method is to make the discontinuities aligned with the coordinate axes. Aligning the discontinuities with the coordinate axes as much as possible is crucial for the success of QMC methods. Numerical experiments on pricing exotic options with discontinuous payoff functions and on estimating Greeks demonstrated that the proposed method leads to dramatic variance reduction in QMC methods and significantly outperforms STD, BB, and PCA. On the other hand, STD, BB, and PCA can easily lose their effectiveness for discontinuous functions. The traditional idea of designing more advanced PGMs is to reduce the effective dimension. Our idea is to make the discontinuities QMC friendly, opening a new way to improve QMC methods.

The investigation revealed that the OT method, in addition to making the discontinuities QMC friendly, has an additional benefit, namely, it reduces the mean dimension significantly. Thus, the OT method has the advantage of handling discontinuities and high dimensionality simultaneously. It is also found that the structure of the discontinuities has a strong connection with the mean dimension. When the discontinuities are QMC friendly, the mean dimension is usually smaller (this is true for all examples in this paper, but it remains to identify the class of problems for which this statement is true). This observation calls for a theoretical study on the interaction between the structure of the discontinuities and the mean dimension. The QMC friendly discontinuities and the reduced mean dimension may explain the superiority of the OT method to some extent. Rigorous theoretical explanation is an important subject for future study.

Our discussion has been in the context of geometric Brownian motion. The problems studied here share many features with other more complicated problems. It is interesting to generalize the idea to other contexts, such as the Lévy process. Moreover, transforming an arbitrary integrand to make its discontinuities QMC friendly (as much as possible) is also an interesting problem. Other important topics for further research include the development of new PGMs or smoothing-like methods to handle the discontinuities for more complicated structure of the discontinuities (say, for functions involving multiple indicator functions), to better understand the relationship between the PGMs and the accuracy of QMC methods, to develop more efficient QMC algorithms for various financial applications, and to compare their performance with other classes of numerical schemes.

Acknowledgments

The authors thank the department editor, the associate editor, and the anonymous referees for their constructive comments and suggestions. Xiaoqun Wang is the corresponding author. Xiaoqun Wang acknowledges the support from the National Science Foundation of China, and Ken Seng Tan acknowledges the support from the Natural Sciences and Engineering Research Council of Canada and the Ministry of Education Project of Key Research Institute of Humanities and Social Sciences at Universities [No. 11JJD790004].

Appendix A. Comparisons for Examples in §3

Table A.1 Comparison on Variance Reduction Factors for Digital Options (Example A1)

	MC	Sobol'			Korobov		
d	STD	STD	BB	PCA	STD	BB	PCA
2	1	**2,262**	50	110	**2,857**	176	188
16	1	**46**	1	3	**971**	17	8
64	1	**66**	5	11	**545**	17	9
128	1	**60**	13	13	**296**	24	11

Notes. The numbers are the VRFs with respect to MC (using STD) with $m = 100$ replications ($n = 4{,}096$ for MC and the Sobol' points, and $n = 4{,}001$ for the Korobov points). The model parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, and $T = 1$.

Table A.2 Variance Reduction Factors for Asian Options with Knock-Out Feature at Maturity (Example A2)

		MC	Sobol'			Korobov		
d	H	STD	STD	BB	PCA	STD	BB	PCA
16	120	1	1	**105**	6	1	**76**	5
	140	1	1	**179**	7	2	**184**	9
	160	1	2	**279**	24	4	**312**	18
64	120	1	1	**109**	5	1	**39**	5
	140	1	1	**194**	7	2	**80**	7
	160	1	2	**267**	18	5	**133**	12
128	120	1	1	**105**	5	1	**44**	4
	140	1	2	**213**	7	2	**115**	7
	160	1	2	**257**	16	3	**89**	15

Notes. The numbers are the VRFs for Asian options with knock-out feature at the maturity (Example A2). The model parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, $T = 1$, and $K = 100$.

Table A.3 Variance Reduction Factors for Asian Options with a Modified Knock-Out Feature (Example A3)

		MC	Sobol'			Korobov		
d	H	STD	STD	BB	PCA	STD	BB	PCA
16	120	1	1	6	**2,557**	1	7	**1,515**
	140	1	2	20	**2,939**	3	18	**2,096**
	160	1	2	45	**1,773**	11	86	**1,407**
64	120	1	1	5	**3,600**	2	5	**933**
	140	1	2	20	**2,566**	4	11	**2,067**
	160	1	3	62	**1,955**	11	41	**1,393**
128	120	1	1	6	**3,101**	1	4	**815**
	140	1	2	17	**2,525**	3	16	**1,727**
	160	1	4	52	**1,474**	8	38	**1,249**

Notes. The numbers are the VRFs for Example A3. The model parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, $T = 1$, and $K = 100$.

Appendix B. Comparisons for Examples in §4.2

Table B.1 Variance Reduction Factors for Example B1

		MC	Sobol'				Korobov			
d	K	STD	STD	BB	PCA	OT	STD	BB	PCA	OT
16	90	1	1	7	21	**1,303**	4	14	24	**270**
	100	1	1	9	33	**929**	9	12	38	**171**
	110	1	1	24	70	**541**	8	23	80	**108**
64	90	1	5	13	9	**936**	9	15	18	**207**
	100	1	5	23	15	**664**	10	18	29	**148**
	110	1	2	62	37	**441**	7	21	50	**87**
128	90	1	4	25	11	**937**	5	5	4	**92**
	100	1	3	41	18	**694**	7	9	7	**64**
	110	1	2	59	39	**456**	5	10	23	**40**

Notes. The numbers are the VRFs for Example B1. The parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, and $T = 1.0$.

Table B.2 Variance Reduction Factors for Example B2

		MC	Sobol'				Korobov			
d	H	STD	STD	BB	PCA	OT	STD	BB	PCA	OT
16	120	1	1	9	26	**4,233**	1	8	24	**5,201**
	140	1	3	29	70	**2,480**	5	26	82	**3,007**
	160	1	2	117	455	**1,107**	23	170	521	**1,402**
64	120	1	1	7	30	**3,503**	2	5	3	**3,144**
	140	1	2	31	101	**1,969**	6	19	116	**2,195**
	160	1	4	143	870	**1,255**	20	101	631	**1,082**
128	120	1	1	7	29	**4,207**	1	5	31	**3,543**
	140	1	4	31	92	**2,109**	5	22	115	**1,927**
	160	1	5	146	718	**1,101**	14	79	644	**899**

Notes. The numbers are the VRFs for Example B2. The parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, $T = 1.0$, and $K = 100$.

Appendix C. Comparisons for Examples in §4.3

Table C.1 Variance Reduction Factors for Example C1

			MC	Sobol'				Korobov			
d	Weights	K	STD	STD	BB	PCA	OT	STD	BB	PCA	OT
64	(A)	90	1	1	8	29	**116**	3	5	3	**32**
		100	1	2	13	40	**165**	3	11	24	**70**
		110	1	2	15	46	**169**	3	11	13	**45**
64	(B)	90	1	2	35	11	**907**	2	34	8	**1,158**
		100	1	2	46	11	**1,088**	3	25	10	**1,450**
		110	1	3	45	9	**1,206**	3	24	9	**1,065**
128	(A)	90	1	1	9	30	**125**	2	4	25	**132**
		100	1	2	15	41	**261**	3	4	57	**311**
		110	1	2	12	37	**176**	2	4	59	**235**
128	(B)	90	1	2	58	8	**1,179**	3	39	6	**2,525**
		100	1	3	50	12	**2,332**	3	36	8	**2,750**
		110	1	4	40	11	**2,407**	2	47	7	**2,818**

Notes. The parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, and $T = 1.0$. The weights are (A) and (B).

Table C.2 Variance Reduction Factors for Example C2

			MC	Sobol'				Korobov			
d	Weights	K	STD	STD	BB	PCA	OT	STD	BB	PCA	OT
64	(A)	90	1	3	13	45	**172**	5	7	5	**51**
		100	1	4	16	51	**203**	4	14	31	**89**
		110	1	2	18	55	**193**	4	13	15	**52**
64	(B)	90	1	4	63	21	**934**	5	68	15	**1,073**
		100	1	4	71	17	**1,096**	5	39	16	**1,438**
		110	1	4	64	13	**1,192**	4	33	13	**1,117**
128	(A)	90	1	2	14	45	**186**	3	7	39	**188**
		100	1	3	19	51	**327**	4	6	73	**383**
		110	1	2	14	43	**206**	3	5	70	**276**
128	(B)	90	1	5	102	17	**927**	5	74	12	**1,746**
		100	1	5	75	19	**1,806**	5	56	13	**1,907**
		110	1	5	55	15	**1,924**	3	66	10	**2,362**

Notes. The parameters are $S_0 = 100$, $\sigma = 0.2$, $r = 0.1$, and $T = 1.0$. The weights are (A) or (B).

Table C.3 Effective Dimension-Related Characteristics for Example C1

			Degree of additivity R_1				Mean dimension d_{ms}			
d	Weights	K	STD	BB	PCA	OT	STD	BB	PCA	OT
64	(A)	90	0.342	0.551	0.856	**0.983**	7.663	2.319	1.272	**1.030**
		100	0.622	0.703	0.899	**0.991**	6.358	2.024	1.198	**1.022**
		110	0.596	0.687	0.909	**0.983**	6.465	2.025	1.181	**1.028**
64	(B)	90	0.450	0.932	0.674	**0.998**	8.073	1.176	2.119	**1.005**
		100	0.520	0.940	0.709	**0.998**	7.385	1.158	1.986	**1.006**
		110	0.654	0.938	0.727	**0.999**	7.151	1.154	1.967	**1.004**
128	(A)	90	0.244	0.562	0.856	**0.986**	10.834	2.421	1.283	**1.029**
		100	0.597	0.693	0.901	**0.990**	8.940	2.094	1.201	**1.022**
		110	0.604	0.695	0.906	**0.985**	9.123	2.079	1.181	**1.027**
128	(B)	90	0.412	0.953	0.663	**1.000**	11.280	1.132	2.417	**1.005**
		100	0.666	0.955	0.701	**0.999**	10.320	1.121	2.282	**1.004**
		110	0.684	0.957	0.726	**0.999**	10.207	1.118	2.240	**1.004**

Notes. The numbers are the degree of additivity and the mean dimension for Example C1. The model parameters are the same as in Table C.1.

Table C.4 Effective Dimension-Related Characteristics for Example C2

			Degree of additivity R_1				Mean dimension d_{ms}			
d	Weights	K	STD	BB	PCA	OT	STD	BB	PCA	OT
64	(A)	90	0.640	0.740	0.908	**0.988**	5.142	1.808	1.170	**1.019**
		100	0.743	0.773	0.921	**0.993**	5.135	1.788	1.154	**1.016**
		110	0.627	0.722	0.923	**0.985**	5.641	1.876	1.154	**1.024**
64	(B)	90	0.771	0.967	0.862	**0.999**	4.574	1.091	1.557	**1.003**
		100	0.730	0.962	0.827	**0.999**	5.000	1.100	1.614	**1.004**
		110	0.734	0.955	0.806	**0.998**	5.347	1.109	1.687	**1.003**
128	(A)	90	0.555	0.747	0.913	**0.991**	7.158	1.875	1.179	**1.019**
		100	0.685	0.768	0.923	**0.993**	7.150	1.843	1.156	**1.017**
		110	0.631	0.725	0.921	**0.988**	7.892	1.923	1.153	**1.023**
128	(B)	90	0.755	0.976	0.844	**1.000**	6.212	1.069	1.709	**1.003**
		100	0.804	0.973	0.824	**1.000**	6.836	1.076	1.799	**1.002**
		110	0.733	0.970	0.800	**0.999**	7.498	1.083	1.878	**1.003**

Notes. The numbers are the degree of additivity and the mean dimension for Example C2. The model parameters are the same as in Table C.2.

References

Acworth P, Broadie M, Glasserman P (1998) A comparison of some Monte Carlo and quasi–Monte Carlo techniques for option pricing. Niederreiter H, Hellekalek P, Larcher G, Zinterhof P, eds. *Monte Carlo and Quasi-Monte Carlo Methods 1996* (Springer-Verlag, New York), 1–18.

Avramidis AN, L'Ecuyer P (2006) Efficient Monte Carlo and quasi–Monte Carlo option pricing under the variance gamma model. *Management Sci.* 52:1930–1944.

Berblinger M, Schlier Ch, Weiss T (1997) Monte Carlo integration with quasi-random numbers: Experience with discontinuos integrands. *Comput. Phys. Comm.* 99:151–162.

Boyle P, Broadie M, Glasserman P (1997) Monte Carlo methods for security pricing. *J. Econom. Dynam. Control* 21:1267–1321.

Broadie M, Glasserman P (1996) Estimating security price derivatives using simulation. *Management Sci.* 42:269–285.

Caflisch RE, Morokoff W, Owen A (1997) Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Finance* 1:27–46.

Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (Springer-Verlag, New York).

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

Imai J, Tan KS (2006) A general dimension reduction method for derivative pricing. *J. Comput. Finance* 10:129–155.

Imai J, Tan KS (2009) An accelerating quasi-Monte Carlo method for option pricing under the generalized hyperbolic Lévy process. *SIAM J. Sci. Comput.* 31:2282–2302.

Jin X, Zhang AX (2006) Reclaiming quasi-Monte Carlo efficiency in portfolio value-at-risk simulation through Fourier transform. *Management Sci.* 52:925–938.

Joy C, Boyle PP, Tan KS (1996) Quasi-Monte Carlo methods in numerical finance. *Management Sci.* 42:926–938.

L'Ecuyer P (2009) Quasi-Monte Carlo methods with applications in finance. *Finance Stochastics* 13:307–349.

L'Ecuyer P, Lemieux C (2000) Variance reduction via lattice rules. *Management Sci.* 46:1214–1235.

L'Ecuyer P, Lemieux C (2002) Recent advances in randomized quasi-Monte Carlo methods. Dror M, L'Ecuyer P, Szidarovszki F, eds. *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications* (Kluwer Academic, Norwell, MA), 419–474.

L'Ecuyer P, Perron G (1994) On the convergence rates of IPA and FDC derivative estimators. *Oper. Res.* 42:643–656.

Lemieux C (2009) *Monte Carlo and Quasi-Monte Carlo Sampling* (Springer-Verlag, New York).

Liu R, Owen AB (2006) Estimating mean dimensionality of analysis of variance decomposition. *J. Am. Stat. Assoc.* 101: 712–721.

Morokoff WJ, Caflisch RE (1994) Quasi-random sequences and their discrepancies. *SIAM J. Sci. Comput.* 15:1251–1279.

Morokoff WJ, Caflisch RE (1995) Quasi-Monte Carlo integration. *J. Comput. Phys.* 122:218–230.

Moskowitz B, Caflisch RE (1996) Smoothness and dimension reduction in quasi–Monte Carlo methods. *Math. Comput. Modelling* 23:37–54.

Niederreiter H (1992) *Random Number Generation and Quasi-Monte Carlo Methods* (SIAM, Philadelphia).

Ninomiya S, Tezuka S (1996) Toward real-time pricing of complex financial derivatives. *Appl. Math. Finance* 3:1–20.

Owen AB (2003) The dimension distribution and quadrature test functions. *Statist. Sinica* 13:1–17.

Papageorgiou A (2002) The Brownian bridge does not offer a consistent advantage in quasi-Monte Carlo integration. *J. Complexity* 18:171–186.

Paskov SH, Traub JF (1995) Faster valuation of financial derivatives. *J. Portfolio Management* 22:113–120.

Sloan IH, Woźniakowski H (1998) When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. Complexity* 14:1–33.

Sobol' IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *Zh. Vychisli. Mat. i Mat. Fiz.* 7:784–802.

Staum J (2009) Monte Carlo computation in finance. L'Ecuyer P, Owen AB, eds. *Monte Carlo and Quasi-Monte Carlo 2008* (Springer-Verlag, New York), 19–42.

Wang X (2006) On the effects of dimension reduction techniques on high-dimensional problems in finance. *Oper. Res.* 54:1063–1078.

Wang X (2007) Constructing robust good lattice rules for computational finance. *SIAM J. Sci. Comput.* 29:598–621.

Wang X, Fang K-T (2003) The effective dimensions and quasi-Monte Carlo integration. *J. Complexity* 19:101–124.

Wang X, Sloan IH (2008) Projections of low discrepancy sequences: How well are they distributed? *J. Comput. Appl. Math.* 213:366–386.

Wang X, Sloan IH (2011) Quasi-Monte Carlo methods in financial engineering: An equivalent principle and dimension reduction. *Oper. Res.* 59:80–95.

Wang X, Tan KS (2012) How do path generation methods affect the accuracy of quasi-Monte Carlo methods for problems in finance? *J. Complexity* 28:250–277.