Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

ENHANCED QUASI-MONTE CARLO METHODS WITH DIMENSION REDUCTION

Junichi Imai

Iwate Prefectural University,
Faculty of Policy Studies,
152-52, Takizawa-aza-sugo,
Takizawa, Iwate, JAPAN.

Ken Seng Tan

University of Waterloo,
Department of Statistics and Actuarial Science,
University Avenue West,
Waterloo, Ontario, CANADA.

ABSTRACT

In recent years, the quasi-Monte Carlo approach for pricing high-dimensional derivative securities has been used widely relative to other competitive approaches such as the Monte Carlo methods. Such success can be, in part, attributed to the notion of effective dimension of the finance problems. In this paper, we provide additional insight on the connection between the effective dimension and the quasi-Monte Carlo method. We also propose a dimension reduction technique which further enhances the quasi-Monte Carlo method for derivative pricing. The efficiency of the proposed method is illustrated by applying it to high-dimensional multi-factor path-dependent derivative securities.

1 INTRODUCTION AND MOTIVATION

Since the introduction of the Monte Carlo method to derivative pricing by Boyle in 1977, this method has been widely used. The prices of many complex derivative securities can be written as very high dimensional integrals and only in rare cases explicit solutions exist. Hence the Monte Carlo method (MC) becomes the only viable numerical tool. A typical example is mortgage-backed securities where the value can depend on the monthly interest rates over the next 30 years. The dimensions of such a problem can be of several hundreds. Other exotic instruments include multi-factor path-dependent options where the dimensions can be several thousands since these derivatives depend on the trajectories of multiple factors.

Despite its wide applicability and the fact that its convergence rate $O(N^{-1/2})$ is independent of the dimension, the Monte Carlo method is often criticized for its slow rate of convergence. Different techniques for increasing its efficiency have been proposed. These approaches are known as variance reduction techniques which include control variates and antithetic variates. More recently, so-called quasi-Monte Carlo (QMC) or low discrepancy (LD) methods have been introduced to finance applications. Early applications of these methods are discussed in Joy, Boyle, and Tan (1996) and Paskov and Traub (1995). These results have presented something of a puzzle in the fields of computational finance and numerical analysis due to the apparent conflicting conclusion. QMC attains a convergence rate of $O(N^{-1}\log^s N)$ in dimension s, which is better than MC only if N grows exponentially with dimension s. Hence the theoretical higher asymptotic convergence rate of QMC is not achievable for practical applications, particularly for large s. This is supported by the empirical evidence in non-finance applications that QMC offers no practical advantage over MC even for problems with dimensions as low as 12 (see e.g. Bratley, Fox, and Niederreiter (1992)). Nevertheless when the same method is used to price derivative securities, it is superior to MC, even for dimensions that are of several hundreds!

The goal of this paper is to provide further insight on the superior rate of convergence of QMC exhibited by the finance problems. One plausible explanation lies in the distinction between the *nominal* and *effective dimension* of an integrand. Due to the inherence feature of QMC, this method is particularly well suited for problems with low effective dimensions. We explore these relationships in Section 2. Section 3 discusses the simulation techniques used to price the type of options considered in this paper. Section 4 describes the proposed method and Section 5 provides numerical examples. Section 6 concludes the paper.

2 EFFECTIVE DIMENSIONS

In this section we describe the concept related to effective dimension. We discuss its implication to QMC as well as the motivation for our proposed method. Caflisch, Morokoff, and Owen (1997) give two formal definitions of effective dimensions through the “analysis of variance" (ANOVA) decomposition of a function. The relevant definition for our

analysis is the effective dimension in the truncation sense and is stated as follows:

Definition 1 *The effective dimension of f, in the truncation sense, is the smallest integer d_T such that $\sum_{u\in\{1,2,\ldots,d_T\}} \sigma^2(f_u) \geq 0.99\sigma^2(f)$.*

Here f_u is a function which depends on the components of $\boldsymbol{x}$ in the set u and $\sigma^2(f)$ denotes the variance of the function f . An integrand is said to have low effective dimension if it can be captured predominantly by the sum of lower-dimensional integrands.

Low effective dimension occurs naturally in finance problems. For example, mortgage-backed securities depend on the contingent cash flows each month over the next 30 years, leading to a 360 nominal dimensional problem. Its effective dimension, however, is much smaller due to two factors: (i) the time value of money, a dollar in 30 years is worth a lot less than the same dollar in 1 year; (ii) empirical evidence indicates that majority of the cash flows occur in the initial few years. Thus the cash flows in the first few years are most important for pricing of mortgage-backed securities, implying a relatively small effective dimension.

The connection between the effective dimension and the efficiency of QMC can be explained as follows: QMC relies on specially constructed sequences known as low discrepancy sequences. These sequences are deterministic and are designed to have greater uniformity than random sequences. Sobol′ (1967) is an example of a low discrepancy sequence. The left and right panels in Figure 1 give the orthogonal projection (second and third coordinates) of 2048 Sobol′ and random points, respectively. Clearly, the Sobol′ points are more regularly and uniformly distributed over the unit square than the random points which exhibit both clustering and relative sparsity. It is this enhanced uniformity of the points that leads to higher rate of convergence. However as we increase the dimension, the orthogonal projection of the Sobol′ points can yield extremely poor uniform regularity as demonstrated in Figure 2. The left panel of Figure 2 depicts the orthogonal projection of the 27-th and 28-th dimensions. Although it has the same number of points as Figure 1, the points are far from uniformly distributed, exhibiting regular and rigid structures. When we subsequently increase the points to 4096 (as shown in the right panel of Figure 2), the additional points gradually fill up the gaps. However for these higher dimensional set of points, the uniformity is achieved at a much slower rate. In fact, for these points to have the same degree of uniformity as the Sobol′ points in Figure 1, the number of points need to be increased to a very large number!

The deterioration of the uniformity of Sobol′ points as we increase the dimension is unfortunately typical in most low discrepancy sequences. This implies that for practical sample sizes, the uniformity of high dimensional low discrepancy sequences is no better than the random sequences. Hence if the problem of interest is truly s-dimensional, then for large s we should not expect QMC to significantly outperform MC in practical application. On the other hand, if the problem of interest has low effective dimension (in the truncation sense) relative to the nominal dimension, then the lower but more important dimensional structures are evaluated at the much higher precision rate of QMC. Consequently, problem with low effective dimension can recover the QMC rate. This in part accounts for the success of the finance applications of QMC, even for very high dimensional problems.

The above analysis implies that if integrands f and g are such that $\int_{[0,1)^s} f(\boldsymbol{x})d\boldsymbol{x} = \int_{[0,1)^s} g(\boldsymbol{x})d\boldsymbol{x}$ and $d_T(f) > d_T(g)$ where $d_T(\cdot)$ denotes the effective dimension in the truncation sense, then QMC achieves a better rate of convergence when applied to g instead of f. For example, consider the functions

$$f(\boldsymbol{x}) = \sum_{i=1}^{s} 2^{i-1}x_i \tag{1}$$

and

$$g(\boldsymbol{x}) = \sum_{i=1}^{s} 2^{s-i}x_i \tag{2}$$

where $\boldsymbol{x} \in [0,1)^s$. Since the coefficients in these two functions are permutations, their values of the integral must be identical. However, we have $d_T(f) = s$, but $d_T(g) = s$ for $s = 1, 2, 3$ and $d_T(g) = 4$ for $s \geq 4$. Hence for large s, a greater precision can be expected when applying QMC to g than to f since the earlier dimensions of g contribute most to the value of the integral.

The above example also suggests one way of enhancing QMC for estimating an integral value. Suppose the integrand of interest, say f in (1). Rather than applying QMC directly to f, the argument above suggests that we should perform an appropriate transformation on f so as to obtain an integrand such as g in (2) while preserving the value of the integral. Through the transformation, if the effective dimension (in the truncation sense) of the resulting integrand g is reduced substantially, then QMC will have greater success when applied to g than to f. Formally, an optimum application of QMC can be formulated as an optimization algorithm which seeks a transformation $\psi(\cdot)$ within the set of "all possible transformations", say Ψ, that minimizes the effective dimension (in the truncation sense) of the transformed integrand f_ψ while also preserving the integral values. The constraint of preserving the value of the integral is to ensure that the problem of interest is not distorted after the transformation.

The function g in (2) demonstrates one possible transformation which merely re-orders the coefficients of x_i. While it is very difficult to obtain the optimum transformation in general, two types of "transformations" have been

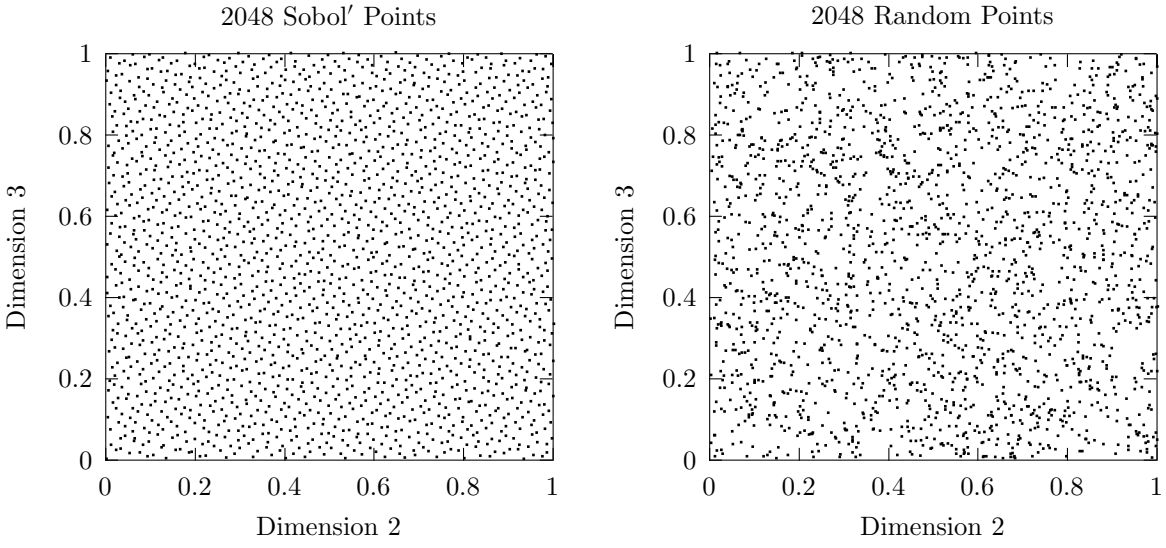

Figure 1: Two-Dimensional Projection of Sobol and Random Sequences

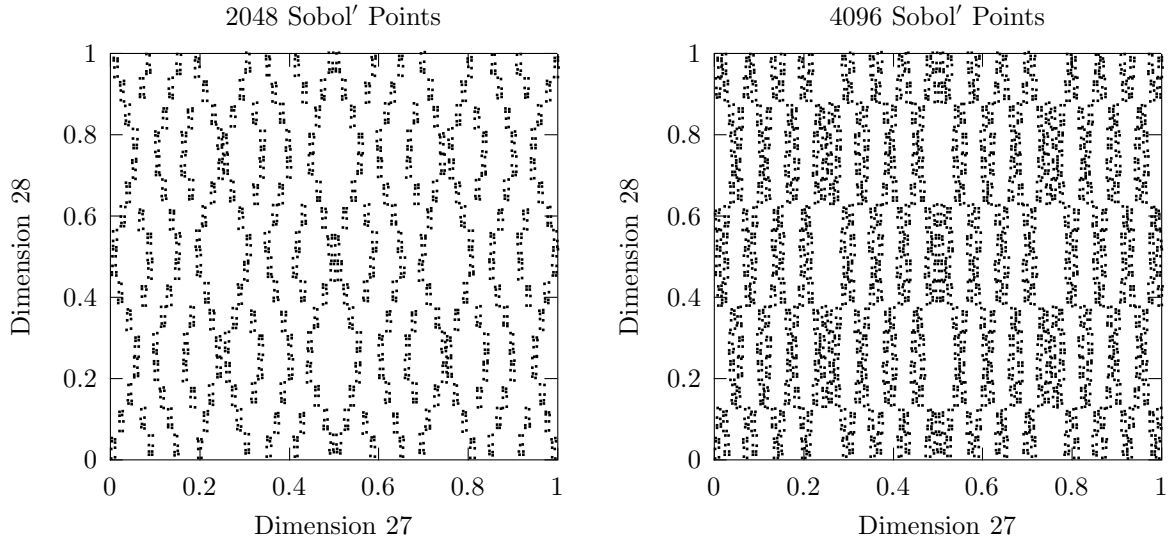

Figure 2: Two-Dimensional Projection of Sobol Sequence

considered in connection with Brownian path generation. These are based on the Brownian bridge (see Moskowitz and Caflisch (1996) and Caflisch, Morokoff, and Owen (1997)) and principal component analysis constructions (see Acworth, Broadie, and Glasserman (1998)). These transformations demonstrate that the efficiency of QMC can be enhanced further. Motivated by these results, this paper confines Ψ to a class of transformation which we denote as "linear transformation". Our objective is to find an optimum transformation over this class so that the effective dimension (in the truncation sense) of the problem of interest is minimized. We describe the proposed transformation in greater details in the following two sections.

3 PATH GENERATION METHODS

It will be of interest to first describe the type of exotic options that will be considered in this paper. We assume the risky assets follow multivariate geometric Brownian motion and that their dynamics under the risk-neutral world (i.e., Q measure) are given by the following stochastic differential equations:

$$dS_i(t) = rS_i(t)dt + \sigma_i S_i(t)dW_i(t), \;\; i = 1, 2, \ldots, m, \quad (3)$$

where $S_i(t)$ denotes i-th asset price at time t, r is the risk-free interest rate, σ_i is the volatility for i-th asset, and

$\boldsymbol{W}(t) = (W_1(t), \ldots, W_m(t))$ is a m-dimensional Brownian motion such that $W_i(t)$ satisfies

$$E[W_i(t)] = 0,$$

$$Var[W_i(t)] = t,$$

and

$$Cov[W_i(t), W_k(t)] = \rho_{ik} t,$$

where ρ_{ik} is the correlation between the i-th and the k-th Brownian motions.

The numerical examples illustrated in Section 5 are based on the European multi-asset path-dependent call options with payoff at maturity $T = t_n$ given by

$$\max\left[\prod_{i=1}^{m}\prod_{j=1}^{n} S_i(t_j)^{w_{ij}} - X,\ 0\right] \tag{4}$$

and

$$\max\left[\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} S_i(t_j) - X,\ 0\right] \tag{5}$$

where $\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} = 1$ and $\{t_1, t_2, \ldots, t_n\}$ are the time points at which the asset prices are sampled and X is the strike price. These options are known as the geometric weighted average options and arithmetic weighted average options, respectively. For an equally weighted case with $m = 1$ and $n > 1$, the contingent claim (5) is commonly referred as an Asian option. When $m > 1$ and $n = 1$ so that the payoff depends only on the terminal prices of a basket of m assets, the resulting option is known as a basket option.

For the assumed model, there exists an explicit closed-form solution for the geometric average case but not the arithmetic average options. In fact, due to the multi-factor and path-dependency, the dimensionality of the arithmetic average option can be very large. For instance with 10 underlying assets and 250 sampling time points, the nominal dimensions of the problem is 2500! Hence the Monte Carlo method becomes an important numerical tool for pricing these derivative securities.

To price the above options using simulation, it is necessary to simulate the multi-asset trajectories at each sampling time point; i.e., $\{S_i(t_j); i = 1, \ldots, m, j = 1, \ldots, n\}$. For simplicity, assume the asset prices are sampled at equal fixed time interval of length Δt so that $\Delta t = T/n$ and $t_j = j\Delta t$. Let $\boldsymbol{\Sigma}$ be an $m \times m$ covariance matrix given by $(\boldsymbol{\Sigma})_{ik} = \rho_{ik}\sigma_i\sigma_k\Delta t$, $i, k = 1, 2, \ldots m$ and define $\boldsymbol{\Sigma}_{mn}$ as an $mn \times mn$ matrix generated from $\boldsymbol{\Sigma}$ via

$$\boldsymbol{\Sigma}_{mn} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} & \cdots & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma} & 2\boldsymbol{\Sigma} & \cdots & 2\boldsymbol{\Sigma} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma} & 2\boldsymbol{\Sigma} & \cdots & n\boldsymbol{\Sigma} \end{bmatrix}.$$

It follows from (3) that the asset prices evolve as

$$S_i(t_j) = S_i(0) e^{(r-\sigma_i^2/2)t_j + Z_i(t_j)}, \tag{6}$$

where

$$\begin{pmatrix} Z_1(t_1) \\ \vdots \\ Z_m(t_1) \\ Z_1(t_2) \\ \vdots \\ Z_m(t_n) \end{pmatrix} = \tilde{\boldsymbol{C}} \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \\ \epsilon_{m+1} \\ \vdots \\ \epsilon_{mn} \end{pmatrix} = \tilde{\boldsymbol{C}}\boldsymbol{\epsilon} \tag{7}$$

and $\tilde{\boldsymbol{C}}$ is a decomposed matrix of $\boldsymbol{\Sigma}_{mn}$ satisfying

$$\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}' = \boldsymbol{\Sigma}_{mn} \tag{8}$$

and $\{\epsilon_k; k = 1, \ldots, mn\}$ are independent standard normal variates. For MC, the samples $\{\epsilon_k; i = 1, \ldots, mn\}$ are typically generated using standard routine such as the Box-Muller or polar transformation. With QMC, $\boldsymbol{\epsilon}$ is obtained via the inverse transformation; i.e., $\{\epsilon_k = F^{-1}(x_k); k = 1, \ldots, mn\}$, where $\boldsymbol{x} = (x_1, \ldots, x_{mn}) \in [0, 1)^{mn}$ corresponds to a point from a mn-dimensional low discrepancy sequence and $F(\cdot)$ is the cumulative standard normal distribution. The inverse transformation is crucial when use with QMC in order to preserve the uniformity of the input low discrepancy points. Since each trajectory depends on mn input components, the problem is said to have a nominal dimension of mn.

The path construction (6) (together with (7)) corresponds to multiplying a vector of mn independent normal variates using the matrix $\tilde{\boldsymbol{C}}$ as long as condition (8) is satisfied. This is a consequence of a property that a Brownian path is completely determined by its covariance structure. The conventional discretization approach uses a lower triangular matrix $\tilde{\boldsymbol{C}} \equiv \boldsymbol{C}^{\mathrm{Ch}}$ based on Cholesky decomposition of $\boldsymbol{\Sigma}_{mn}$.

It was first recognized by Moskowitz and Caflisch (1996) that generating the trajectory using Brownian bridge reduces the effective dimension of a Brownian path and hence increases the efficiency of QMC. This method is simple to use but is restrictive in that it is only applicable to one-dimensional Brownian paths. A more versatile approach is that based on principal component analysis (PCA)

Table 1: Comparison of Explained Variability for each Component under Standard, Brownian Bridge, PCA and LT Constructions

	Explained Variability for Each Component			
	Standard	Brown. Bridge	PCA	LT
ε_1	40.0%	75.0%	82.9%	82.9%
ε_2	30.0%	15.0%	10.0%	10.0%
ε_3	20.0%	5.00%	4.26%	4.26%
ε_4	10.0%	5.00%	2.83%	2.83%

proposed by Acworth, Broadie, and Glasserman (1998). For this method, the generation matrix $\tilde{\boldsymbol{C}} \equiv \boldsymbol{C}^{\mathrm{PCA}}$ is defined as $\boldsymbol{P}\boldsymbol{D}^{1/2}$ where $\boldsymbol{D}$ is the diagonal matrix such that $D_{ii} \geq D_{jj}, i < j$ and $D_{kk}; k = 1, \ldots, mn$, are the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}_{mn}$, and the k-th column of $\boldsymbol{P}$ is the unit-length eigenvector corresponds to eigenvalue D_{kk}.

A standard result in multivariate statistics is that the variability explained by the first k standard normals is equal to the sum of the squared norms of the first k columns of $\tilde{\boldsymbol{C}}$. This quantity therefore provides a way of comparing the relative merits of the generation matrices. For simplicity, let assume the covariance matrix $(\boldsymbol{\Sigma}_{mn})_{ij} = \min(i, j)$, for $i, j = 1, 2, 3$ and 4. Table 1 depicts the explained variability due to each standard normal component. These results indicate that the standard construction is the least favorable. The first component captures only 40% of the variation while the first two 70%. The Brownian bridge construction is an improvement over the standard approach but the PCA approach is even more effective, explaining up to 82.9% of the variation for just one component.

The column labelled "LT" is our proposed dimension reduction transformation which we now explain. We consider a class of transformation by defining

$$\tilde{\boldsymbol{C}} \equiv \boldsymbol{C}^{\mathrm{LT}} = \boldsymbol{C}^{\mathrm{Ch}}\boldsymbol{A}, \tag{9}$$

where $\boldsymbol{A}$ is an orthogonal matrix; i.e., $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}$ and $\boldsymbol{I}$ is the identity matrix. By construction, we have $\boldsymbol{C}^{\mathrm{LT}}(\boldsymbol{C}^{\mathrm{LT}})' = \boldsymbol{\Sigma}_{mn}$ so that condition (8) is satisfied. We refer to this method as linear transformation or simply the LT construction. The optimum $\boldsymbol{C}^{\mathrm{LT}}$ is obtained by optimally choosing $\boldsymbol{A}$ so as the effective dimension (in the truncation sense) of the problem of interest is minimized. For instance in the above example where we wish to maximize the explanatory variability of the covariance structure, the optimum orthogonal matrix $\boldsymbol{A}^*$ can be found by solving the following optimization problem:

$$\max \ ||\boldsymbol{C}_{\cdot k}^{\mathrm{LT}}||^2 = \max_{\boldsymbol{A}_{\cdot k} \in \Re^{mn}} \sum_{i=1}^{mn} (\boldsymbol{C}_{i\cdot}^{\mathrm{Ch}}\boldsymbol{A}_{\cdot k})^2 \tag{10}$$

subject to $||\boldsymbol{A}_{\cdot k}|| = 1$ and $< \boldsymbol{A}_{\cdot i}^*, \boldsymbol{A}_{\cdot k} > = 0, i = 1, 2, \ldots, k-1$. Here we adopt the following matrix notation: if $\boldsymbol{A}$ is a matrix of size m by n. Then $\boldsymbol{A} = (\boldsymbol{A}_{\cdot 1}, \ \ldots, \boldsymbol{A}_{\cdot n}) = \begin{pmatrix} \boldsymbol{A}_{1\cdot} \\ \vdots \\ \boldsymbol{A}_{m\cdot} \end{pmatrix}$, where $\boldsymbol{A}_{i\cdot} \in \Re^n$ is the row vector corresponding to the i-th row of $\boldsymbol{A}$ and $\boldsymbol{A}_{\cdot j} \in \Re^m$ is the column vector corresponding to the j-th column of $\boldsymbol{A}$. Also, the notation $<\boldsymbol{a}, \boldsymbol{b}>$ denotes the dot product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are either both row vectors or column vectors, and $||a||$ denotes the norm of a vector $\boldsymbol{a}$.

The maximization algorithm (10) is carried out iteratively for $k = 1, 2, \ldots, mn$ so that the columns of $\boldsymbol{A}^*$ are obtained sequentially. The objective function maximizes the variance contribution due to the k-th column while the second constraint ensures that the k-th optimum column is orthogonal to the solutions $\boldsymbol{A}_{\cdot i}^*, i = 1, 2, \ldots, k-1$ obtained in the previous step. It can be shown that the solution to the above optimization problem is $\boldsymbol{A}^* = \boldsymbol{P}\boldsymbol{D}^{1/2}$ where $\boldsymbol{D}$ is the diagonal matrix with $D_{ii} \geq D_{jj}, i < j$, and $D_{kk}; k = 1, \ldots, mn$ are the eigenvalues of the mn by mn matrix $\sum_{k=1}^{mn} (\boldsymbol{C}_{k\cdot}^{\mathrm{Ch}})'\boldsymbol{C}_{k\cdot}^{\mathrm{Ch}}$, and the k-th column of $\boldsymbol{P}$ is the unit-length eigenvector corresponds to eigenvalue D_{kk}.

The last column of Table 1 indicates that the proposed LT method is as efficient as PCA. This is in fact not surprising since the matrix $\sum_{k=1}^{mn} (\boldsymbol{C}_{k\cdot}^{\mathrm{Ch}})'\boldsymbol{C}_{k\cdot}^{\mathrm{Ch}}$ is merely a permutation, a rotation and a reflection of $\boldsymbol{\Sigma}_{mn}$, a consequence of the orthogonal matrix $\boldsymbol{A}$. Hence if the objective is to extract key components of a covariance matrix, then both PCA and LT constructions are equivalent. The strength of the proposed method becomes apparent in the following section when we consider more complicated applications.

4 LINEAR TRANSFORMATION

In this section, we consider an European option with payoff at maturity T given by

$$\max[g(\boldsymbol{\epsilon}) - X, 0],$$

where

$$g(\boldsymbol{\epsilon}) = \begin{cases} \prod_{i=1}^{m} \prod_{j=1}^{n} S_i(t_j)^{w_{ij}} \\ \\ \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} S_i(t_j) \end{cases}.$$

It follows from (6) and (7) that g is a function in terms of a vector of mn standard normal variates. Using these two option structures, we now illustrate how to obtain an optimum $\boldsymbol{C}^{\mathrm{LT}}$ under our proposed LT constructions.

We first consider the geometric average option. In this case g becomes

$$g(\boldsymbol{\epsilon}) = e^{\mu + \sum_{k=1}^{mn} \alpha_k \epsilon_k}, \tag{11}$$

where

$$\mu = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\left[\log S_i(0) + \left(r - \frac{\sigma_i^2}{2}\right)t_j,\right] \qquad (12)$$

and

$$\alpha_k = <\tilde{\boldsymbol{C}}_{\cdot k}, \boldsymbol{w}>, \qquad (13)$$

and $\boldsymbol{w} = (w_{11}, w_{21}, \ldots, w_{m1}, w_{12}, \ldots, w_{mn})' \in \Re^{mn}$. Let define $f(\boldsymbol{\epsilon}) = \log g(\boldsymbol{\epsilon})$; i.e.,

$$f(\boldsymbol{\epsilon}) = \mu + \sum_{k=1}^{mn} \alpha_k \epsilon_k.$$

Then the variance of the function f is given by

$$\sigma^2(f(\boldsymbol{\epsilon})) = \sum_{k=1}^{s} \alpha_k^2 \sigma^2(\epsilon_k) = \sum_{k=1}^{s} \alpha_k^2.$$

From the definition of effective dimension, $d_T(f(\boldsymbol{\epsilon}))$ is the smallest integer satisfying

$$\sum_{k=1}^{d_T(f(\boldsymbol{\epsilon}))} \alpha_k^2 \geq 0.99\sigma^2(f(\boldsymbol{\epsilon})).$$

Since $\alpha_k^2 = <\tilde{\boldsymbol{C}}_{\cdot k}, \boldsymbol{w}>^2$, which in turn depends on how the decomposed matrix $\tilde{\boldsymbol{C}}$ is chosen, the above analysis provides a way of reducing the effective dimension through careful construction of the generation matrix $\tilde{\boldsymbol{C}}$. Under our proposed class of transformation where $\tilde{\boldsymbol{C}} \equiv \boldsymbol{C}^{\mathrm{LT}} = \boldsymbol{C}^{\ \mathrm{Ch}}\boldsymbol{A}$, equation (13) reduces to

$$\alpha_k = \ <\boldsymbol{A}_{\cdot k}, \boldsymbol{B}>, \qquad k = 1, \ldots, mn, \qquad (14)$$

where $\boldsymbol{B} = (B_1, \ldots, B_{mn})'$ and $B_k = <\boldsymbol{C}_{\cdot k}^{\mathrm{Ch}}, \boldsymbol{w}>$. The optimum columns of $\boldsymbol{A}$ are then found by modifying the optimization algorithm (10) to

$$\max_{\boldsymbol{A}_{\cdot k}} <\boldsymbol{A}_{\cdot k}, \boldsymbol{B}>^2 \qquad (15)$$

subject to $||\boldsymbol{A}_{\cdot k}|| = 1$ and $<\boldsymbol{A}_{\cdot j}^*, \boldsymbol{A}_{\cdot k}> = 0, \quad j = 1, 2, \ldots, k-1$.

It can be verified that when $k = 1$, the optimum solution to the above maximization problem is $\boldsymbol{A}_{\cdot 1}^* = \pm\frac{\boldsymbol{B}}{||\boldsymbol{B}||}$, which leads to $\alpha_1 = <\boldsymbol{A}_{\cdot 1}^*, \boldsymbol{B}> = \pm||\boldsymbol{B}||$. Furthermore, the orthogonality condition ensures that $<\boldsymbol{A}_{\cdot k}^*, \boldsymbol{B}> = 0$ for $k \geq 2$. Thus the optimum solutions $\boldsymbol{A}_{\cdot k}^*, k = 2, \ldots, mn$ can be arbitrary as long as they satisfy the orthogonal condition. This implies that $\alpha_k = 0$ for $k = 2, \ldots, mn$. Consequently, equation (11) simplifies to

$$g(\boldsymbol{\epsilon}) = e^{\mu \pm ||\boldsymbol{B}||\epsilon_1}, \qquad (16)$$

which depends only on the first component. This indicates that the LT construction effectively collapses an mn-dimensional geometric average option to a 1-dimensional problem. This distinguishes the efficiency of the LT method over the PCA approach. Although the PCA construction is effective at isolating the key components of the covariance matrix $\boldsymbol{\Sigma}_{mn}$, the LT method exploits the linear structure of the function to the fullest. This result is not surprising since the product of lognormals is still a lognormal. Hence simulating a product of lognormal random variables is equivalent to simulating one-dimensional lognormal random variables (with appropriate parameters adjustment). The LT construction is able to exploit the linearity structure.

Note that in our formulation of the LT method, we have $\boldsymbol{C}^{\mathrm{LT}} = \boldsymbol{C}^{\mathrm{Ch}}\boldsymbol{A}$, where $\boldsymbol{C}^{\mathrm{Ch}}$ is a Cholesky decomposition of $\boldsymbol{\Sigma}_{mn}$. We could define other class of transformation such as $\boldsymbol{C}^{\mathrm{LT}} = \boldsymbol{C}^{\mathrm{PCA}}\boldsymbol{A}$. This choice is not explored in this paper since this introduces additional computational effort for the eigenvectors and eigenvalues decomposition of $\boldsymbol{\Sigma}_{mn}$.

We now consider arithmetic average options. In this case, the function g becomes

$$g(\boldsymbol{\epsilon}) = \sum_{i=1}^{mn} e^{\mu_i + \sum_{l=1}^{mn}\alpha_{il}\epsilon_l}, \qquad (17)$$

where

$$\mu_i = \log(w_{i_1,i_2}S_i(0)) + \left(r - \frac{\sigma_{i_1}^2}{2}\right)t_{i_2}, \qquad (18)$$

and

$$\alpha_{ik} = \tilde{C}_{ik}, \qquad (19)$$

with $i_2 = \lfloor (i-1)/m \rfloor + 1$, $i_1 = i - (i_2 - 1)m$, and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

Unlike the geometric average case, the variance of the above function is now quite complex. This complicates the procedure for deriving the optimized matrix $\boldsymbol{A}$ under the LT method. However note that if the function is linear in $\boldsymbol{\epsilon}$ (as in the preceding example), the optimum vector is immediately obtainable. To exploit this property, we obtain the desired column vector $\boldsymbol{A}_{\cdot k}$ iteratively for $k = 1, 2, \ldots, mn$ by linearizing the function through a Taylor expansion. We carry the optimization as follows: Applying a first order

Taylor series around an arbitrary point $\boldsymbol{\epsilon} = \hat{\boldsymbol{\epsilon}} + \Delta\boldsymbol{\epsilon}$ to the function g in (17), we have

$$g(\boldsymbol{\epsilon}) \approx g(\hat{\boldsymbol{\epsilon}}) + \sum_{l=1}^{mn} \left.\frac{\partial g}{\partial \epsilon_l}\right|_{\boldsymbol{\epsilon}=\hat{\boldsymbol{\epsilon}}} \Delta\epsilon_l$$
$$= g(\hat{\boldsymbol{\epsilon}}) + \sum_{l=1}^{mn} \left(\sum_{i=1}^{mn} e^{\mu_i + \sum_{k=1}^{mn} \alpha_{ik}\hat{\epsilon}_k} \alpha_{il} \right) \Delta\epsilon_l. \quad (20)$$

Substituting $\hat{\boldsymbol{\epsilon}} = \mathbf{0}$ in the above expansion yields

$$g(\boldsymbol{\epsilon}) \approx g(\mathbf{0}) + \sum_{l=1}^{mn} \left(\sum_{i=1}^{mn} e^{\mu_i} \alpha_{il} \right) \Delta\epsilon_l, \quad (21)$$

which is linear in the normal random variables $\Delta\boldsymbol{\epsilon}$. The variance of the function due to the k-th component is simply given by $\left(\sum_{i=1}^{mn} e^{\mu_i}\alpha_{ik}\right)^2$. This facilitates the optimization of $\boldsymbol{\alpha}_{\cdot k}; k = 1, \ldots, mn$ in that for the LT construction, the optimum $\boldsymbol{A}^*_{\cdot 1}$ can be shown to be $\frac{\boldsymbol{B}_1}{||\boldsymbol{B}_1||}$ where $\boldsymbol{B}_1 = \sum_{i=1}^{mn} \bar{\boldsymbol{C}}_{i\cdot}$ and $\bar{\boldsymbol{C}}_{i\cdot} = e^{\mu_i}\boldsymbol{C}^{\mathrm{Ch}}_{i\cdot}$. The orthogonality condition implies that $< \boldsymbol{B}_1, \boldsymbol{A}^*_{\cdot l} >= 0; l = 2, \ldots, k$ so that the right hand side of (21) reduces to a function involves only the first component:

$$g(\mathbf{0}) + \sum_{l=1}^{mn} \left(\sum_{i=1}^{mn} e^{\mu_i} \alpha_{il} \right) \Delta\epsilon_l \;=\; g(\mathbf{0}) + ||\boldsymbol{B}_1||\Delta\epsilon_1.$$

It should be emphasized that the underlying g is not linear. The above result is true under the approximation. To find the subsequent optimum $\boldsymbol{A}^*_{\cdot k}, k > 1$, one approach is to consider the Taylor expansion with higher order terms. This implies that optimizing $\boldsymbol{A}_{\cdot k}$ requires the k-th order Taylor expansion, which can be very complex and time-consuming. To mitigate this problem, we optimize $\boldsymbol{A}_{\cdot 2}$ given $\boldsymbol{A}^*_{\cdot 1}$ by considering the Taylor approximation (20) with expansion at a different point, say $\hat{\boldsymbol{\epsilon}} = (1, 0, \ldots, 0)$. This results in

$$g(\boldsymbol{\epsilon}) \approx g(\hat{\boldsymbol{\epsilon}}) + \sum_{l=1}^{mn} \left(\sum_{i=1}^{mn} e^{\mu_i + \alpha_{i1}} \alpha_{il} \right) \Delta\epsilon_l$$

so that the variance of the function becomes

$$\sigma^2(g(\boldsymbol{\epsilon})) \approx \sum_{l=1}^{mn} \left(\sum_{i=1}^{mn} e^{\mu_i + \alpha_{i1}} \alpha_{il} \right)^2.$$

Consequently, to optimize $\boldsymbol{A}_{\cdot 2}$ we maximize the variance contribution due to the second dimension as

$$\max_{\boldsymbol{A}_{\cdot 2}} \left(\sum_{i=1}^{mn} e^{\mu_i + \alpha_{i1}} \alpha_{i2} \right)^2$$
$$= \max_{\boldsymbol{A}_{\cdot 2}} \left(\sum_{i=1}^{mn} e^{\mu_i + <\boldsymbol{C}^{\mathrm{Ch}}_{i\cdot}, \boldsymbol{A}^*_{\cdot 1}>} < \boldsymbol{C}^{\mathrm{Ch}}_{i\cdot}, \boldsymbol{A}_{\cdot 2} > \right)^2$$

subject to $||\boldsymbol{A}_{\cdot 2}|| = 1$ and $< \boldsymbol{A}^*_{\cdot 1}, \boldsymbol{A}_{\cdot 2} > = 0$. Note that the objective function depends explicitly on $\boldsymbol{A}^*_{\cdot 1}$. Without the orthogonal condition, the solution to the above maximization problem is easily found. Gram-Schmidt method is then applied to the resulting vector in order to satisfy the orthogonality condition.

The above procedure is readily extended for optimizing $\boldsymbol{A}_{\cdot k}; k > 2$ by expanding the Taylor series around the point $\hat{\boldsymbol{\epsilon}} = (1, \ldots, 1, 0 \ldots, 0)$, where the first $k-1$ components are ones and the remaining components are zeroes. The resulting function becomes a linear function and the usual optimization problem can be solved easily to obtain the desired vector $\boldsymbol{A}^*_{\cdot k}$.

5 NUMERICAL ILLUSTRATIONS

In this section, we demonstrate the efficiency of the proposed LT constructions by considering the arithmetic average call options on a basket of 10 assets with 250 sampling time points using the following parameter values:

$$\begin{array}{lcll} S_i(0) & = & 100 & \\ r & = & 4\% \text{ p.a.} & \\ T & = & 1 \text{ year} & \\ \sigma_i & = & 10\% + \frac{i-1}{9}40\% & \\ X & = & 100 & \\ \rho_{ij} & = & 0\% \text{ and } 40\% & \text{for } i, j = 1, 2, \ldots, 10. \end{array}$$

Since $m = 10$ and $n = 250$, the nominal dimensions for these examples are 2500, illustrating that the dimensionality can be very high for pricing complex derivatives.

We consider three simulation techniques – Cholesky, PCA and LT. For each method, we simulate the option prices using both MC and QMC. The results are reported, respectively, in Tables 2 and 3. The values in parentheses denote the standard errors based on 10 independent replications with each batch consists of either $N = 4096$ or 8192. Theoretically, the LT construction requires optimizing 2500 columns of $\boldsymbol{A}$ since the problems have nominal dimension of 2500. The reported results are only based on optimal vectors up to either 50 or 100 columns, with the remaining entries randomly generated while satisfying the orthogonality constraints. This explains the entries [50] and [100] under the heading "LT". The loss of accuracy from using the suboptimal matrix $\boldsymbol{A}$ is likely to be negligible in view of the effectiveness of the dimension reduction of the LT methods. However, using such suboptimal $\boldsymbol{A}$ leads to a significant reduction in the computational effort in solving the optimization problem.

For the benchmark, using the standard construction with 1 million randomly generating trajectories, the estimated

prices for the zero and positive correlation cases are 3.4409 and 5.6490 with standard errors of 0.0044 and 0.0082, respectively. One immediate conclusion from Table 2 is that there is no significant difference among the various construction approaches. This is not surprising since the rate of convergence of Monte Carlo methods does not depend on dimensions as well as the decomposed matrix $\tilde{\boldsymbol{C}}$.

Table 2: MC Results: Simulated prices of the Asian basket options using random sequences. The values in parentheses denote the standard errors based on 10 independent replications.

N	Standard	PCA	LT
Zero correlation example			
4096	3.429(0.027)	3.451(0.022)	3.425(0.027) [100]
			3.425(0.028) [50]
8192	3.457(0.022)	3.438(0.017)	3.423(0.018) [100]
			3.422(0.019) [50]
Positive correlation example			
4096	5.673(0.049)	5.609(0.048)	5.618(0.049) [100]
			5.619(0.049) [50]
8192	5.672(0.036)	5.616(0.037)	5.618(0.035) [100]
			5.619(0.035) [50]

In Table 3, the same set of examples and the same techniques are compared. The only difference is that the input $\boldsymbol{\epsilon}$ is drawn from the randomized Sobol′ low discrepancy sequences proposed by Owen (1995), instead of a random sequence. We also avoid generating 2500-dimensional Sobol′ sequences by using the Latin supercube sampling (LSS) method (see Owen (1998)). Briefly, this sampling mechanism is a scheme for creating a high-dimensional sequence from sets of lower dimensional sequences. For instance, a 2500-dimensional low discrepancy sequences can be concatenated from 100 sets of 25-dimensional low discrepancy sequences by appropriately randomizing the run order of the points. For theoretical justification of the LSS method, see Owen (1998). In our examples, we use LSS with 50 and 25 dimensions respectively. Based on these results, we can make the following conclusions:

- Unlike the Monte Carlo method, the performance of QMC depends on the choice of decomposed matrix $\tilde{\boldsymbol{C}}$. In particular, a greater efficiency is achieved with PCA and LT as confirmed by their smaller standard errors.
- The efficiency of the PCA method critically depends on the structure of the covariance matrix. As the correlation increases, so is the effectiveness of the underlying method.
- The LT construction is competitively efficient. There is also no significant loss of accuracy by reducing the optimal columns from 100 to 50, confirming that the effective dimensions using the LT methods are substantially lower than 50. The LT methods are more efficient than the PCA method for the zero-correlated case and comparable for the positively correlated case.
- Relative to the Monte Carlo methods, the use of low discrepancy sequences leads to dramatic improvement, particularly with PCA-based and LT-based methods.

Table 3: QMC results: Simulated prices of the Asian basket options using randomized low discrepancy sequences. The values in parentheses denote the standard errors based on 10 independent replications.

LSS	Standard	PCA	LT
$N = 4096$, *zero correlation*			
50	3.438(0.007)	3.443(0.003)	3.447(0.001) [100]
50			3.448(0.001) [50]
25	3.434(0.007)	3.444(0.003)	3.445(0.001) [100]
25			3.446(0.001) [50]
$N = 8192$, *zero correlation*			
50	3.411(0.006)	3.447(0.002)	3.445(0.001) [100]
50			3.444(0.001) [50]
25	3.414(0.009)	3.448(0.002)	3.445(0.001) [100]
25			3.444(0.001) [50]
$N = 4096$, *positive correlation*			
50	5.657(0.015)	5.658(0.001)	5.657(0.001) [100]
50			5.656(0.001) [50]
25	5.633(0.016)	5.658(0.001)	5.657(0.001) [100]
25			5.657(0.001) [50]
$N = 8192$, *positive correlation*			
50	5.613(0.015)	5.658(0.001)	5.657(0.001) [100]
50			5.658(0.001) [50]
25	5.603(0.021)	5.659(0.001)	5.657(0.001) [100]
25			5.658(0.001) [50]

6 CONCLUSION

In this paper, we propose a new method that enhances the efficiency of QMC for finance applications. The proposed class of transformation reduces the effective dimension of the problem, thus enables us to exploit a particular feature of QMC; namely the earlier dimensions of low discrepancy sequences are more uniformly distributed. By optimally obtaining a decomposed matrix $\boldsymbol{C}^{\mathrm{LT}}$, a significant increase in efficiency of QMC is attained. This is demonstrated in our high-dimensional derivative examples. While we have only illustrated the applicability of our proposed construction in the context of pricing multi-factor path-dependent options, the underlying method in fact can be extended to pricing other more exotic options as well as in optimum asset allocation problems. We will report these applications in future studies.

REFERENCES

Acworth, P., M. Broadie, and P. Glasserman. 1998. A comparison of some Monte Carlo and quasi-Monte Carlo methods for option pricing. In *Monte and quasi-Monte-Carlo Methods 1996: Proceedings of a conference at the University of Salzburg, Austria, July 9-12, 1996*, ed. H. Niederreiter, P. Hellekalek, G. Larcher and P. Zinterhof. *Lecture Notes in Statistics* 127. Springer-Verlag, New York.

Boyle, P. 1977. Options: A Monte Carlo Approach. *Journal of Financial Economics*, 4:323–338.

Bratley, P., B. Fox, and H. Niederreiter. 1992. Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation* 2 (3):195–213.

Caflisch, R., W. Morokoff, and A. Owen. 1997. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance* 1 (1):27–46.

Joy, C., P. Boyle, and K.S. Tan. 1996. Quasi-Monte Carlo methods in numerical finance. *Management Science* 42 (6):926–938.

Moskowitz, B. and R. Caflisch. 1996. Smoothness and dimension reduction in quasi-Monte Carlo methods. *Mathematical and Computer Modelling* 23 (8-9):37–54.

Owen, A. 1995. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte and quasi-Monte-Carlo Methods in Scientific Computing: Proceedings of a conference at the University of Nevada, Las Vegas, Nevada, USA, June 23-25, 1994*, ed. H. Niederreiter, and P.-S. Shiue. *Lecture Notes in Statistics* 106, pages 299–317. Springer-Verlag, New York.

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

Paskov, S. and J. Traub. 1995. Faster valuation of financial derivatives. *Journal of Portfolio Management* 22 (1):113–120.

Sobol′, I. 1967. The distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R Computational Mathematics and Mathematical Physics* 7 (4):86–112.

AUTHOR BIOGRAPHIES

JUNICHI IMAI is a Lecturer in Faculty of Policy Studies at Iwate Prefectural University. His e-mail address is <imai@iwate-pu.ac.jp>.

KEN SENG TAN is an Assistant Professor at the Department of Statistics and Actuarial Science, University of Waterloo, Canada. His e-mail address is <kstan@uwaterloo.ca>.