Quasi-Monte Carlo approximations in stochastic optimization

H. Leövey and W. Römisch

Humboldt-University Berlin
Institute of Mathematics

www.math.hu-berlin.de/~romisch

DFG Research Center MATHEON
Mathematics for key technologies

Stevens Institute of Technology, Department of Mathematical Sciences,
Hoboken, New Jersey, May 9, 2013

Introduction

- Computational methods for solving stochastic programs require (first) a discretization of the underlying probability distribution induced by a numerical integration scheme and (second) an efficient solver for the finite-dimensional program.
- Discretization means scenario or sample generation.
- Standard approach: Variants of Monte Carlo (MC) methods.
- Recent alternative approaches to scenario generation:

  (a) Optimal quantization of probability distributions

  (Pflug-Pichler 2011).

  (b) Quasi-Monte Carlo (QMC) methods

  (Koivu-Pennanen 05, Homem-de-Mello 08).

  (c) Sparse grid quadrature rules

  (Chen-Mehrotra 08).

  (d) Moment matching methods

  (Høyland-Wallace 01, Gülpinar-Rustem-Settergren 04)

- Known convergence rates in terms of scenario or sample size n:
  MC: $\hat{e}_n(f) = O(n^{-\frac{1}{2}})$ if $f \in L_2$,
  (a): $e_n(f) = O(n^{-\frac{1}{d}})$ if $f \in \mathrm{Lip}$,
  (b): classical: $e_n(f) = O(n^{-1}(\log n)^d)$ if $f \in \mathrm{BV}$,
  recently: $\hat{e}_n(f) \leq C(\delta) n^{-1+\delta}$ ($\delta \in (0, \frac{1}{2}]$) if $f \in W^{(1,\ldots,1)}$,
  where $C(\delta)$ does not depend on d,
  (c): $e_n(f) = O(n^{-r}(\log n)^{(d-1)(r+1)})$ if $f \in W^{(r,\ldots,r)}$,
  where d is the dimension of the random vector and $e_n(f)$ the quadrature error for integrand f and sample size n, i.e.,

$$e_n(f) = \Big| \int_{[0,1]^d} f(\xi) d\xi - \frac{1}{n} \sum_{i=1}^{n} f(\xi^i) \Big|$$

  and $\hat{e}_n(f)$ denotes mean (square) quadrature error.

- Monte Carlo methods and (a) may be justified by available stability results for stochastic programs, but there is almost no reasonable justification in cases (b), (c) and (d).

- In applications of stochastic programming d is often large.

Contents

Quasi-Monte Carlo methods

We consider the approximate computation of

$$I_d(f) = \int_{[0,1]^d} f(\xi) d\xi$$

by a QMC algorithm

$$Q_{n,d}(f) = \frac{1}{n} \sum_{i=1}^{n} f(\xi^i)$$

with (non-random) points ξ^i, $i = 1, \ldots, n$, from $[0, 1]^d$.

We assume that f belongs to a linear normed space $\mathbb{F}_d$ of functions on $[0, 1]^d$ with norm $\| \cdot \|_d$ and unit ball $\mathbb{B}_d$.

Worst-case error of $Q_{n,d}$ over $\mathbb{B}_d$:

$$e(Q_{n,d}) = \sup_{f \in \mathbb{B}_d} \left| I_d(f) - Q_{n,d}(f) \right|$$

Classical convergence results:

Theorem: (Proinov 88)
If the real function f is continuous on $[0,1]^d$, then there exists $C>0$ such that

$$|Q_{n,d}(f) - I_d(f)| \leq C\omega_f\big(D_n^*(\xi^1,\dots,\xi^n)^{\frac{1}{d}}\big),$$

where $\omega_f(\delta) = \sup\{|f(\xi) - f(\tilde{\xi})| : \|\xi - \tilde{\xi})\| \leq \delta,\ \xi, \tilde{\xi} \in [0,1]^d\}$ is the modulus of continuity of f and

$$D_n^*(\xi^1,\dots,\xi^n) := \sup_{x\in[0,1]^d} |\mathrm{disc}(x)|, \quad \mathrm{disc}(x) = \lambda^d([0,x)) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,x)}(\xi^i),$$

is the star-discrepancy of $\xi^1,\dots,\xi^n$ (λ^d denotes Lebesgue's measure on $\mathbb{R}^d$).

Theorem: (Koksma-Hlawka 61)
If $V_{\mathrm{HK}}(f)$ is the variation of f in the sense of Hardy and Krause, it holds

$$|I_d(f) - Q_{n,d}(f)| \leq V_{\mathrm{HK}}(f) D_n^*(\xi^1,\dots,\xi^n)$$

for any $n \in \mathbb{N}$ and any $\xi^1,\dots,\xi^n \in [0,1]^d$.

Extended Koksma-Hlawka inequality:

$$|I_d(f) - Q_{n,d}(f)| \leq \|\mathrm{disc}(\cdot)\|_{p,p'} \|f\|_{q,q'},$$

where $1 \leq p, p', q, q' \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, $\frac{1}{p'} + \frac{1}{q'} = 1$, and

$$\|\mathrm{disc}(\cdot)\|_{p,p'} = \left(\sum_{u \subseteq D} \left(\int_{[0,1]^{|u|}} |\mathrm{disc}(x_u, 1)|^{p'} \, dx_u \right)^{\frac{p}{p'}} \right)^{\frac{1}{p}}$$

and

$$\|f\|_{q,q'} = \left(\sum_{u \subseteq D} \left(\int_{[0,1]^{|u|}} \left| \frac{\partial^{|u|} f}{\partial x_u}(x_u, 1) \right|^{q'} dx_u \right)^{\frac{q}{q'}} \right)^{\frac{1}{q}}$$

with the obvious modifications if one or more of p, p', q, q' are infinite.

In particular, the classical Koksma-Hlawka inequality essentially corresponds to $p = p' = \infty$ if f belongs to the tensor product Sobolev space $\mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d)$ which is defined next.

By $(x_u, 1)$ we mean the d-dimensional vector with the same components as x for indices in u and the rest of the components replaced by 1.

The case of kernel reproducing Hilbert spaces

We assume that $\mathbb{F}_d$ is a kernel reproducing Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and kernel $K : [0,1]^d \times [0,1]^d \to \mathbb{R}$, i.e.,

$$K(\cdot, y) \in \mathbb{F}_d \text{ and } \langle f(\cdot), K(\cdot, y) \rangle = f(y) \quad (\forall y \in [0,1]^d, f \in \mathbb{F}_d).$$

If I_d is a linear bounded functional on $\mathbb{F}_d$, the quadrature error $e_n(Q_{n,d})$ allows the representation

$$e_n(Q_{n,d}) = \sup_{f \in \mathbb{B}_d} \left| I_d(f) - Q_{n,d}(f) \right| = \sup_{f \in \mathbb{B}_d} |\langle f, h_n \rangle| = \|h_n\|_d$$

according to Riesz' theorem for linear bounded functionals.

The representer $h_n \in \mathbb{F}_d$ of the quadrature error is of the form

$$h_n(x) = \int_{[0,1]^d} K(x,y)dy - \tfrac{1}{n} \sum_{i=1}^{n} K(x, \xi^i) \quad (\forall x \in [0,1]^d),$$

and it holds

$$e_n^2(Q_{n,d}) = \int_{[0,1]^{2d}} K(x,y)dx\,dy - \tfrac{2}{n} \sum_{i=1}^{n} \int_{[0,1]^d} K(\xi^i, y)dy + \tfrac{1}{n^2} \sum_{i,j=1}^{n} K(\xi^i, \xi^j)$$

(Hickernell 96,98)

Example: Weighted tensor product Sobolev space

$$\mathbb{F}_d = \mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d) = \bigotimes_{i=1}^{d} W_2^1([0,1])$$

equipped with the weighted norm $\|f\|_\gamma^2 = \langle f, f\rangle_\gamma$ and inner product

$$\langle f, g\rangle_\gamma = \sum_{u\subseteq\{1,\ldots,d\}} \gamma_u^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|} f}{\partial x_u}(x_u, 1) \frac{\partial^{|u|} g}{\partial x_u}(x_u, 1) dx_u \,,$$

where $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_d > 0$, $\gamma_u = \prod_{j\in u} \gamma_j$, is a kernel reproducing Hilbert space with the kernel

$$K_{d,\gamma}(x,y) = \prod_{j=1}^{d} (1 + \gamma_j \mu(x_j, y_j)) \quad (x, y \in [0,1]^d),$$

where

$$\mu(t,s) = \begin{cases} \min\{|t-1|, |s-1|\} & , (t-1)(s-1) > 0, \\ 0 & , \text{else.} \end{cases}$$

Note that $f \in \mathbb{F}_d$ iff $\frac{\partial^{|u|} f}{\partial x_u}(\cdot, 1) \in L_2([0,1]^{|u|})$ for all $u \subseteq D$.

Theorem: (Sloan-Woźniakowski 98)
Let $\mathbb{F}_d = \mathcal{W}_{2,\text{mix}}^{(1,\ldots,1)}([0,1]^d)$. Then the worst-case error

$$e^2(Q_{n,d}) = \sup_{\|f\|_\gamma \le 1} |I_d(f) - Q_{n,d}(f)| = \sum_{\emptyset \neq u \subseteq D} \prod_{j \in u} \gamma_j \int_{[0,1]^{|u|}} \text{disc}^2(x_u, 1) dx_u$$

is called weighted L_2-discrepancy of $\xi^1, \ldots, \xi^n$.

Note that any $f \in \mathbb{F}_d$ is of bounded variation $V_{\text{HK}}(f)$ in the sense of Hardy and Krause and it holds

$$V(f) = \sum_{\emptyset \neq u \subseteq D} \int_{[0,1]^{|u|}} \left| \frac{\partial^{|u|} f}{\partial x_u}(x_u, 1) \right| dx_u \,.$$

Problem: Integrands in two-stage stochastic programming do not belong to F_d
(piecewise linear functions are not of bounded variation (Owen 05)).

First general QMC construction: Digital nets (Sobol 69, Niederreiter 87)

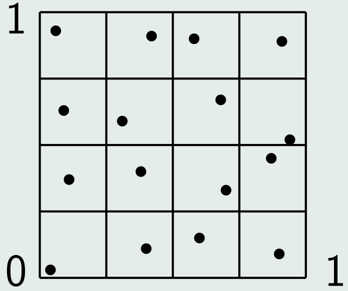
Elementary subintervals E in base b:

$$E = \prod_{j=1}^{d} \left[\frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right),$$

where $a_i, d_i \in \mathbb{Z}_+, 0 \leq a_i < b^{d_i}, i = 1, \dots, d$.

Let $m, t \in \mathbb{Z}_+$, $m > t$. A set of b^m points in $[0,1)^d$ is a (t,m,d)-net in base b if every elementary subinterval E in base b with $\lambda^d(E) = b^{t-m}$ contains b^t points.

Illustration of a $(0,4,2)$-net with $b = 2$



A sequence (ξ^i) in $[0,1)^d$ is a (t,d)-sequence in base b if, for all integers $k \in \mathbb{Z}_+$ and $m > t$, the set

$$\{\xi^i : kb^m \leq i < (k+1)b^m\}$$

is a (t,m,d)-net in base b.

There exist (t,d)-sequences (ξ^i) in $[0,1]^d$ such that

$$D_n^*(\xi^1,\ldots,\xi^n) = O(n^{-1}(\log n)^{d-1}) \le C(\delta,d)\, n^{-1+\delta} \quad (\forall \delta > 0).$$

Specific sequences: Faure, Sobol', Niederreiter and Niederreiter-Xing sequences (Lemieux 09, Dick-Pillichshammer 10).

Recent development: Scrambled (t,m,d)-nets, where the digits are randomly permuted (Owen 95).

Second general QMC construction: Lattices (Korobov 59, Sloan-Joe 94)

Lattice rules: Let $g \in \mathbb{Z}^d$ and consider the lattice points

$$\left\{\xi^i = \left\{\tfrac{i}{n} g\right\} : i = 1,\ldots,n\right\},$$

where $\{z\}$ is defined as *componentwise fractional part* of $z \in \mathbb{R}_+$, i.e., $\{z\} = z - \lfloor z \rfloor \in [0,1)$.

The generator g is chosen such that the lattice rule has good convergence properties. Such lattice rules may achieve better convergence rates $O(n^{-k+\delta})$, $k \in \mathbb{N}$, for integrands in C^k.

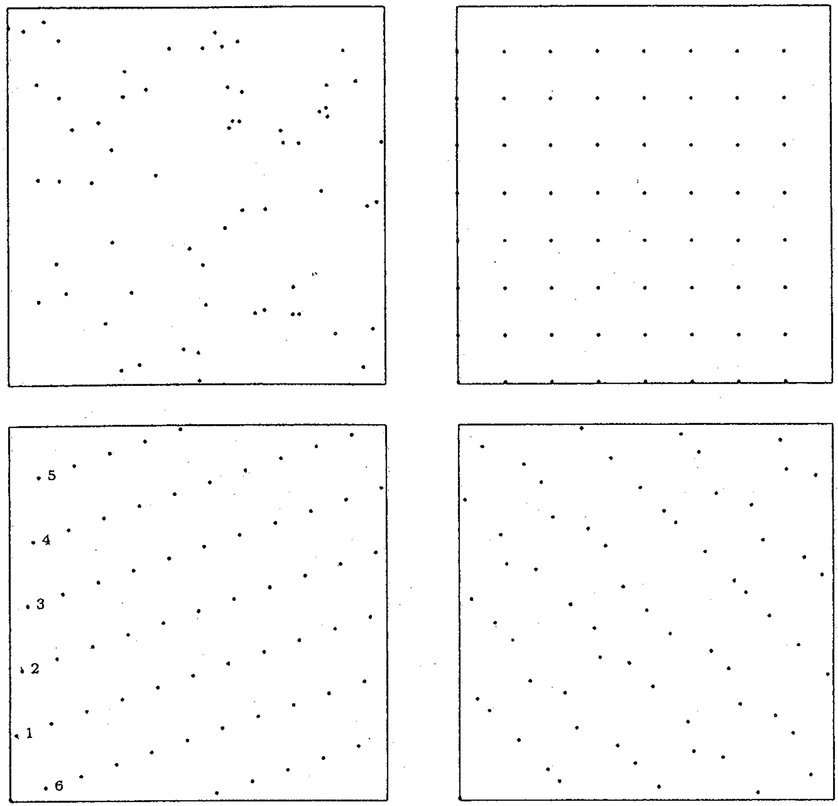


Fig. 5.3 Four different point sets with $n = 64$: random (top left), rectangular grid (top right), Korobov lattice (bottom left), and Sobol' (bottom right).

Recent development: Randomized lattice rules.

Randomly shifted lattice points:
If $\triangle$ is a sample from uniform distribution in $[0,1]^d$. put

$$Q_{n,d}(f) = \frac{1}{n}\sum_{i=1}^{n} f\left(\frac{i}{n}g + \triangle\right).$$

Theorem:
Let n be prime, $\mathbb{F}_d = \mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d)$ and $g \in \mathbb{Z}^d$ be constructed componentwise. Then there exists for any $\delta \in (0, \frac{1}{2}]$ a constant $C(\delta) > 0$ such that the mean quadrature error attains the optimal convergence rate

$$\hat{e}(Q_{n,d}) \leq C(\delta) n^{-1+\delta},$$

where the constant $C(\delta)$ grows when δ decreases, but does not depend on the dimension d if the sequence (γ_j) satisfies the condition

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty \qquad \text{(e.g. } \gamma_j = \frac{1}{j^2}\text{)}.$$

(Sloan/Wožniakowski 98, Sloan/Kuo/Joe 02, Kuo 03)

ANOVA decomposition of multivariate functions

Idea: Decompositions of f may be used, where most of the terms are smooth, but hopefully only some of them relevant.

Let $D = \{1, \ldots, d\}$ and $f \in L_{1,\rho}(\mathbb{R}^d)$ with $\rho(\xi) = \prod_{j=1}^{d} \rho_j(\xi_j)$, where

$$f \in L_{p,\rho}(\mathbb{R}^d) \quad \text{iff} \quad \int_{\mathbb{R}^d} |f(\xi)|^p \rho(\xi) d\xi < \infty \quad (p \geq 1).$$

Let the projection P_k, $k \in D$, be defined by

$$(P_k f)(\xi) := \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \rho_k(s) ds \quad (\xi \in \mathbb{R}^d).$$

Clearly, $P_k f$ is constant with respect to ξ_k. For $u \subseteq D$ we write

$$P_u f = \Big(\prod_{k \in u} P_k \Big)(f),$$

where the product means composition, and note that the ordering within the product is not important because of Fubini's theorem. The function $P_u f$ is constant with respect to all x_k, $k \in u$.

ANOVA-decomposition of f:

$$f = \sum_{u \subseteq D} f_u \,,$$

where $f_\emptyset = I_d(f) = P_D(f)$ and recursively

$$f_u = P_{-u}(f) - \sum_{v \subset u} f_v$$

or (due to Kuo-Sloan-Wasilkowski-Woźniakowski 10)

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u}(f) + \sum_{v \subset u} (-1)^{|u|-|v|} P_{u-v}(P_{-u}(f)),$$

where P_{-u} and P_{u-v} mean integration with respect to ξ_j, $j \in D \backslash u$ and $j \in u \backslash v$, respectively. The second representation motivates that f_u is essentially as smooth as $P_{-u}(f)$.

If f belongs to $L_{2,\rho}(\mathbb{R}^d)$, its ANOVA terms $\{f_u\}_{u \subseteq D}$ are orthogonal in $L_{2,\rho}(\mathbb{R}^d)$.

We set $\sigma^2(f) = \|f - I_d(f)\|_{L_2}^2$ and $\sigma_u^2(f) = \|f_u\|_{L_2}^2$, and have

$$\sigma^2(f) = \|f\|_{L_2}^2 - (I_d(f))^2 = \sum_{\emptyset \neq u \subseteq D} \sigma_u^2(f) \,.$$

Owen's superposition (truncation) dimension distribution of f: Probability measure ν_S (ν_T) defined on the power set of D

$$\nu_S(s) := \sum_{|u|=s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \qquad \left(\nu_T(s) = \sum_{\max\{j:j\in u\}=s} \frac{\sigma_u^2(f)}{\sigma^2(f)}\right) \ (s \in D).$$

Effective superposition (truncation) dimension $d_S(\varepsilon)$ ($d_T(\varepsilon)$) of f is the $(1-\varepsilon)$-quantile of ν_S (ν_T):

$$d_S(\varepsilon) = \min\left\{s \in D : \sum_{|u|\le s} \sigma_u^2(f) \ge (1-\varepsilon)\sigma^2(f)\right\} \le d_T(\varepsilon)$$

$$d_T(\varepsilon) = \min\left\{s \in D : \sum_{u\subseteq\{1,\dots,s\}} \sigma_u^2(f) \ge (1-\varepsilon)\sigma^2(f)\right\}$$

It holds

$$\max\left\{\Big\|f - \sum_{|u|\le d_S(\varepsilon)} f_u\Big\|_{2,\rho}, \Big\|f - \sum_{u\subseteq\{1,\dots,d_T(\varepsilon)\}} f_u\Big\|_{2,\rho}\right\} \le \sqrt{\varepsilon}\sigma(f).$$

(Caflisch-Morokoff-Owen 97, Owen 03, Wang-Fang 03)

Two-stage linear stochastic programs

We consider the linear two-stage stochastic program

$$\min\Big\{\int_{\Xi} f(x,\xi)P(d\xi) : x \in X\Big\},$$

where f is extended real-valued defined on $\mathbb{R}^m \times \mathbb{R}^d$ given by

$$f(x,\xi) = \langle c, x\rangle + \Phi(q(\xi), h(\xi) - T(\xi)x),\ (x,\xi) \in X \times \Xi,$$

$c \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^m$ and $\Xi \subseteq \mathbb{R}^d$ are convex polyhedral, W is an $(r, \overline{m})$-matrix, P is a Borel probability measure on Ξ, and the vectors $q(\xi) \in \mathbb{R}^{\overline{m}}$, $h(\xi) \in \mathbb{R}^r$ and the (r, m)-matrix $T(\xi)$ are affine functions of ξ, Φ is the second-stage optimal value function

$$\Phi(u,t) = \inf\{\langle u, y\rangle : Wy = t, y \geq 0\} \quad ((u,t) \in \mathbb{R}^{\overline{m}} \times \mathbb{R}^r),$$

Let $\mathrm{pos}\, W = W(\mathbb{R}^{\overline{m}}_+)$, $\mathcal{D} = \{u \in \mathbb{R}^{\overline{m}} : \{z \in \mathbb{R}^r : W^\top z \leq u\} \neq \emptyset\}$.

Assumptions:

(A1) $h(\xi) - T(\xi)x \in \mathrm{pos}\, W$ and $q(\xi) \in \mathcal{D}$ for all $(x,\xi) \in X \times \Xi$.

(A2) $\int_{\Xi} \|\xi\|^2 P(d\xi) < \infty$.

Proposition:
(A1) and (A2) imply that the two-stage stochastic program represents a convex minimization problem with respect to the first stage decision x with polyhedral constraints.

Lemma: (Walkup-Wets 69, Nožička-Guddat-Hollatz-Bank 74)
Φ is finite, polyhedral and continuous on the $(\overline{m}+r)$-dimensional polyhedral cone $\mathcal{D} \times \operatorname{pos} W$ and there exist $(r, \overline{m})$-matrices C_j and $(\overline{m}+r)$-dimensional polyhedral cones $\mathcal{K}_j$, $j = 1, ..., \ell$, such that

$$\bigcup_{j=1}^{\ell} \mathcal{K}_j \; = \; \mathcal{D} \times \operatorname{pos} W \quad \text{and} \quad \operatorname{int} \mathcal{K}_i \cap \operatorname{int} \mathcal{K}_j = \emptyset \,,\; i \neq j,$$
$$\Phi(u,t) \; = \; \langle C_j u, t \rangle, \quad \text{for each} \quad (u,t) \in \mathcal{K}_j,\; j = 1, ..., \ell.$$

The function $\Phi(u, \cdot)$ is convex on $\operatorname{pos} W$ for each $u \in \mathcal{D}$, and $\Phi(\cdot, t)$ is concave on $\mathcal{D}$ for each $t \in \operatorname{pos} W$. The intersection $\mathcal{K}_i \cap \mathcal{K}_j$, $i \neq j$, is either equal to $\{0\}$ or contained in a $(\overline{m}+r-1)$-dimensional subspace of $\mathbb{R}^{\overline{m}+r}$ if the two cones are adjacent.

Error estimates for optimal values and solution sets

With $v(P)$ and $S(P)$ denoting the optimal value and solution set of

$$\min \Big\{ \int_\Xi f(x,\xi)P(d\xi) : x \in X \Big\},$$

it holds

$$\begin{aligned} |v(P) - v(Q)| &\leq L \sup_{x\in X} \Big| \int_\Xi f(x,\xi)P(d\xi) - \int_\Xi f(x,\xi)Q(d\xi) \Big| \\ \emptyset \neq S(Q) &\subseteq S(P) + \Psi_P\Big(L \sup_{x\in X} \Big| \int_\Xi f(x,\xi)(P-Q)(d\xi) \Big| \Big), \end{aligned}$$

where $L > 0$ is some constant, P the original probability distribution and Q its perburbation, and Ψ_P the conditioning function given by

$$\Psi_P(\eta) := \eta + \psi_P^{-1}(2\eta) \quad (\eta \in \mathbb{R}_+),$$

where the growth function ψ_P is

$$\psi_P(\tau) := \min \Big\{ \int_\Xi f_0(x,\xi)P(d\xi) - v(P) : d(x, S(P)) \geq \tau, x \in X \Big\}$$

with inverse $\psi_P^{-1}(t) := \sup\{\tau \in \mathbb{R}_+ : \psi_P(\tau) \leq t\}$. (Römisch 03)

ANOVA decomposition of two-stage integrands

Assumptions:

(A1), **(A2)** and

(A3) P has a density of the form $\rho(\xi) = \prod_{j=1}^{d} \rho_j(\xi_j)$ ($\xi \in \mathbb{R}^d$) with continuous marginal densities ρ_j, $j \in D$.

Proposition:

(A1) implies that the function $f(x, \cdot)$, where

$$f_x(\xi) := f(x, \xi) = \langle c, x \rangle + \Phi(q(\xi), h(\xi) - T(\xi)x) \quad (x \in X, \xi \in \Xi)$$

is the two-stage integrand, is continuous and piecewise linear-quadratic.
For each $x \in X$, $f(x, \cdot)$ is linear-quadratic on each polyhedral set

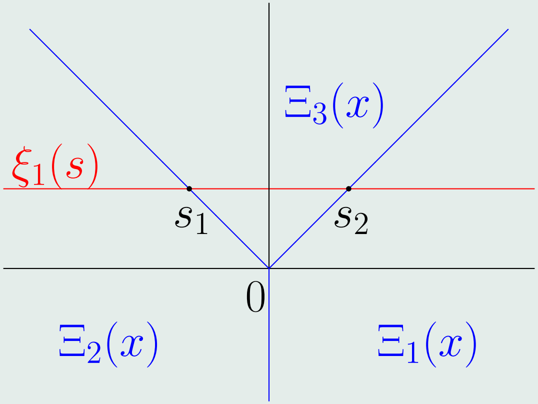
$$\Xi_j(x) = \{\xi \in \Xi : (q(\xi), h(\xi) - T(\xi)x) \in \mathcal{K}_j\} \quad (j = 1, \ldots, \ell).$$

It holds $\operatorname{int} \Xi_j(x) \neq \emptyset$, $\operatorname{int} \Xi_j(x) \cap \operatorname{int} \Xi_i(x) = \emptyset$, $i \neq j$, and the sets $\Xi_j(x)$, $j = 1, \ldots, \ell$, decompose Ξ. Furthermore, the intersection of two adjacent sets $\Xi_i(x)$ and $\Xi_j(x)$, $i \neq j$, is contained in some $(d-1)$-dimensional affine subspace.

To compute projections $P_k f$ for $k \in D$, let $\xi_i \in \mathbb{R}$, $i = 1, \ldots, d$, $i \neq k$, be given. We set $\xi^k = (\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_d)$ and

$$\xi_k(s) = (\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \in \mathbb{R}^d \quad (s \in \mathbb{R}).$$

We fix $x \in X$ and consider the one-dimensional affine subspace $\{\xi_k(s) : s \in \mathbb{R}\}$:



Example with $d = 2 = p$, where the polyhedral sets are cones

It meets the nontrivial intersections of two adjacent polyhedral sets $\Xi_i(x)$ and $\Xi_j(x)$, $i \neq j$, at finitely many points s_i, $i = 1, \ldots, p$ if all $(d-1)$-dimensional subspaces containing the intersections do not parallel the kth coordinate axis.

The $s_i = s_i(\xi^k)$, $i = 1, \ldots, p$, are affine functions of ξ^k. It holds

$$s_i = -\sum_{l=1, l\neq k}^{p} \frac{g_{il}}{g_{ik}} \xi_l + a_i \quad (i = 1, \ldots, p)$$

for some $a_i \in \mathbb{R}$ and $g_i \in \mathbb{R}^d$ belonging to an intersection of polyhedral sets.

Proposition:
Let $k \in D$, $x \in X$. Assume (A1)–(A3) and that all $(d-1)$-dimensional affine subspaces containing nontrivial intersections of adjacent sets $\Xi_i(x)$ and $\Xi_j(x)$ do not parallel the kth coordinate axis.
Then the kth projection $P_k f$ has the explicit representation

$$P_k f(\xi^k) = \sum_{i=1}^{p+1} \sum_{j=0}^{2} p_{ij}(\xi^k; x) \int_{s_{i-1}}^{s_i} s^j \rho_k(s) ds,$$

where $s_0 = -\infty$, $s_{p+1} = +\infty$ and $p_{ij}(\cdot; x)$ are polynomials in ξ^k of degree $2-j$, $j = 0, 1, 2$, with coefficients depending on x, and is continuously differentiable. $P_k f$ is infinitely differentiable if the marginal density ρ_k belongs to $C^\infty(\mathbb{R})$.

Theorem:

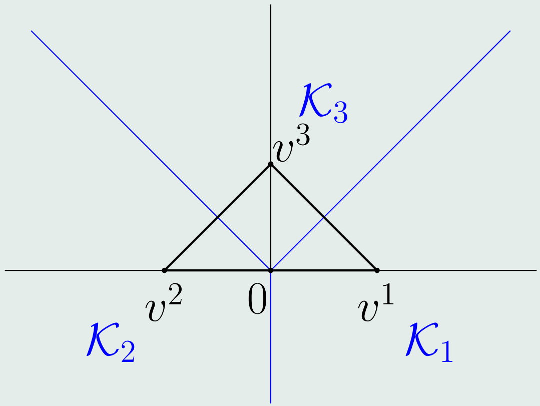
Let $x \in X$, assume (A1)–(A3) and that the following geometric condition (GC) be satisfied: All $(d-1)$-dimensional affine subspaces containing nontrivial intersections of adjacent sets $\Xi_i(x)$ and $\Xi_j(x)$ do not parallel any coordinate axis. Then the ANOVA approximation

$$f_{d-1} := \sum_{|u| \leq d-1} f_u \quad \text{i.e.} \quad f = f_{d-1} + f_D$$

of f is infinitely differentiable if all densities ρ_k, $k \in D$, belong to $C_b^\infty(\mathbb{R})$. Here, the subscript b means that all derivatives of functions belonging to that space are bounded on $\mathbb{R}$.

Example: Let $\bar{m} = 3$, $d = 2$, P denote the two-dimensional standard normal distribution, $h(\xi) = \xi$, q and W be given such that (A1) is satisfied and the dual feasible set is

$$\{z \in \mathbb{R}^2 : -z_1 + z_2 \leq 1, z_1 + z_2 \leq 1, -z_2 \leq 0\}.$$



Dual feasible set, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

The function Φ and the integrand are of the form

$$\Phi(t) = \max_{i=1,2,3} \langle v^i, t\rangle = \max\{t_1, -t_1, t_2\} = \max\{|t_1|, t_2\}$$

$$f(\xi) = \langle c, x\rangle + \Phi(\xi - Tx) = \langle c, x\rangle + \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}$$

and the convex polyhedral sets are $\Xi_j(x) = Tx + \mathcal{K}_j$, $j = 1, 2, 3$.

The ANOVA projection $P_1 f$ is in C^∞, but $P_2 f$ is not differentiable.

QMC quadrature error estimates

If the assumptions of the theorem are satisfied, the two-stage integrand $f = f_x$ (for fixed $x \in X$) allows the representation $f = f_{d-1} + f_D$ with f_{d-1} belonging to $\mathbb{F}_d$. This implies

$$\Big|\int_{[0,1]^d} f(\xi)d\xi - \frac{1}{n}\sum_{j=1}^{n} f(\xi^j)\Big| \leq e(Q_{n,d})\|f_{d-1}\|_\gamma + \Big|\int_{[0,1]^d} f_D(\xi)d\xi - \frac{1}{n}\sum_{j=1}^{n} f_D(\xi^j)\Big|$$
$$\leq e(Q_{n,d})\|f_{d-1}\|_\gamma + \|f_D\|_{L_2} + \Big(\frac{1}{n}\sum_{j=1}^{n}|f_D(\xi^j)|^2\Big)^{\frac{1}{2}}$$

where $\|\cdot\|_\gamma$ is the weighted tensor product Sobolev space norm.

As f_D is (Lipschitz) continuous and if the ξ^j, $j = 1, \ldots, n$ are properly selected, the last term in the above estimate may be assumed to be bounded by $2\|f_D\|_{L_2}$.

Hence, if the effective superposition dimension satisfies $d_S(\varepsilon) \leq d - 1$, i.e., $\|f_D\|_{L_2} \leq \sqrt{\varepsilon}\sigma(f)$ holds for some small $\varepsilon > 0$, the first term $e(Q_{n,d})\|f_{d-1}\|_\gamma$ dominates and the convergence rate of $e(Q_{n,d})$ becomes most important.

Question: How important is the geometric condition (GC) ?

Partial answer: If P is normal with nonsingular covariance matrix, (GC) is satisfied for almost all covariance matrices. Namely, it holds

Proposition: Let $x \in X$, (A1), (A2) be satisfied, $\operatorname{dom} \Phi = \mathbb{R}^r$ and P be a normal distribution with nonsingular covariance matrix Σ. Then the infinite differentiability of the ANOVA approximation f_{d-1} of f is a generic property, i.e., it holds in a residual set (countable intersection of open dense subsets) in the metric space of orthogonal (d, d)-matrices Q (endowed with the norm topology) appearing in the spectral decomposition $\Sigma = Q^\top D\, Q$ of Σ (with a diagonal matrix D containing the eigenvalues of Σ).

Question: For which two-stage stochastic programs is $\|f_D\|_{L_{2,\rho}}$ small, i.e., the effective superposition dimension $d_S(\varepsilon)$ of f is less than $d-1$ or even much less?

Partial answer: In case of a (log)normal probability distribution P the effective dimension depends on the mode of decomposition of the covariance matrix into a diagonal one.

Dimension reduction in case of (log)normal distributions

Let P be the normal distribution with mean μ and nonsingular covariance matrix Σ. Let A be a matrix satisfying $\Sigma = A\,A^\top$. Then η defined by $\xi = A\eta + \mu$ is standard normal.

A universal principle is principal component analysis (PCA). Here, one uses $A = (\sqrt{\lambda_1}u_1, \ldots, \sqrt{\lambda_d}u_d)$, where $\lambda_1 \geq \cdots \geq \lambda_d > 0$ are the eigenvalues of Σ in decreasing order and the corresponding orthonormal eigenvectors u_i, $i = 1, \ldots, d$. Wang-Fang 03, Wang-Sloan 05 report an enormous reduction of the effective truncation dimension in financial models if PCA is used.

A problem-dependent principle may be based on the following equivalence principle (Papageorgiou 02, Wang-Sloan 11).

Proposition: Let A be a fixed $d \times d$ matrix such that $A\,A^\top = \Sigma$. Then it holds $\Sigma = B\,B^\top$ if and only if B is of the form $B = A\,Q$ with some orthogonal $d \times d$ matrix Q.

Idea: Determine Q for given A such that the effective truncation dimension is minimized (Wang-Sloan 11).

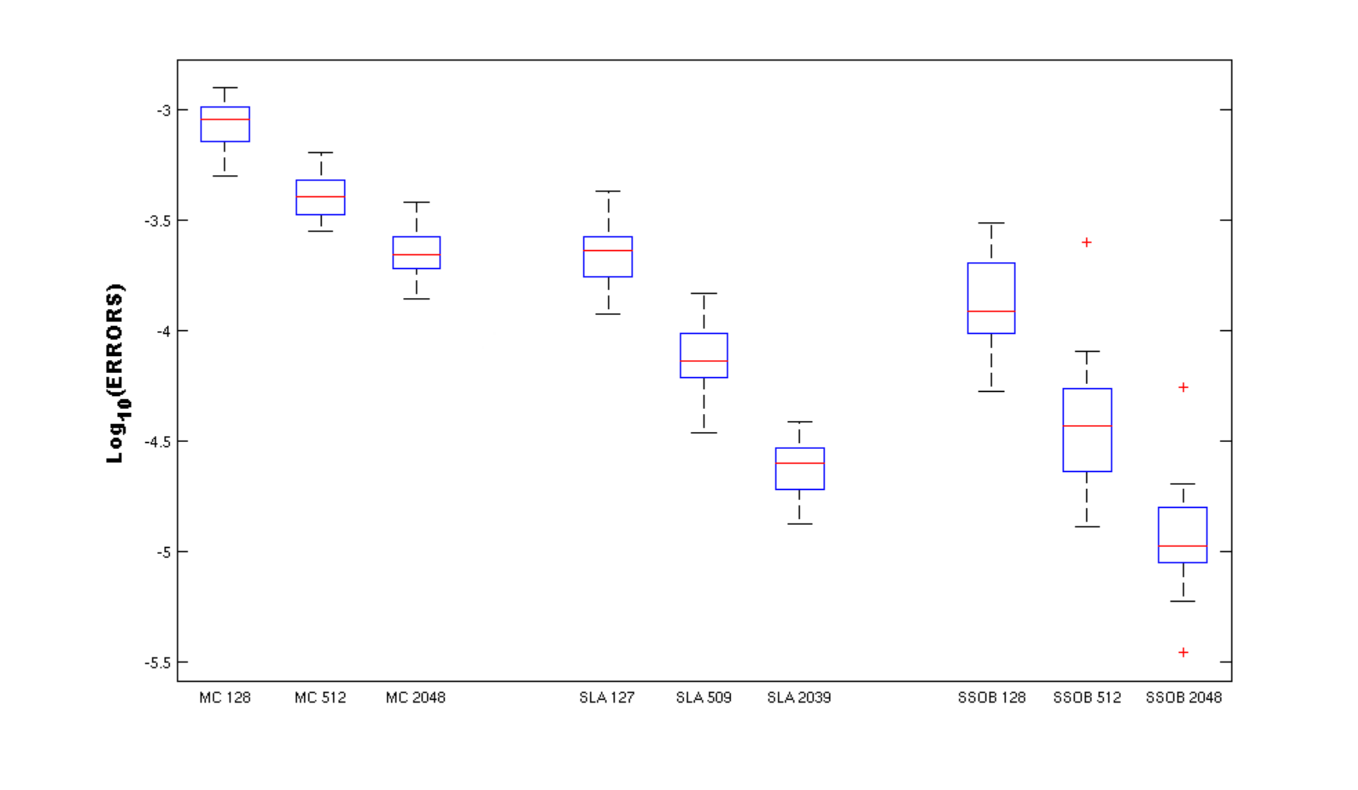
Some computational experience

We considered a two-stage production planning problem for maximizing the expected revenue while satisfying a fixed demand in a time horizon with $d = T = 100$ time periods and stochastic prices for the second-stage decisions. It is assumed that the probability distribution of the prices ξ is log-normal. The model is of the form

$$\max\Big\{\sum_{t=1}^{T}\Big(c_t^\top x_t + \int_{\mathbb{R}^T} q_t(\xi)^\top y_t P(d\xi)\Big) : Wy + Vx = h, y \geq 0, x \in X\Big\}$$

The use of PCA for decomposing the covariance matrix has led to effective truncation dimension $d_T(0.01) = 2$. As QMC methods we used a randomly scrambled Sobol sequence (SSobol)(Owen, Hickernell) with $n = 2^7,\ 2^9,\ 2^{11}$ and a randomly shifted lattice rule (Sloan-Kuo-Joe) with $n = 127, 509, 2039$, weights $\gamma_j = \frac{1}{j^2}$ and for MC the Mersenne-Twister. 10 runs were performed for the error estimates and 30 runs for plotting relative errors.

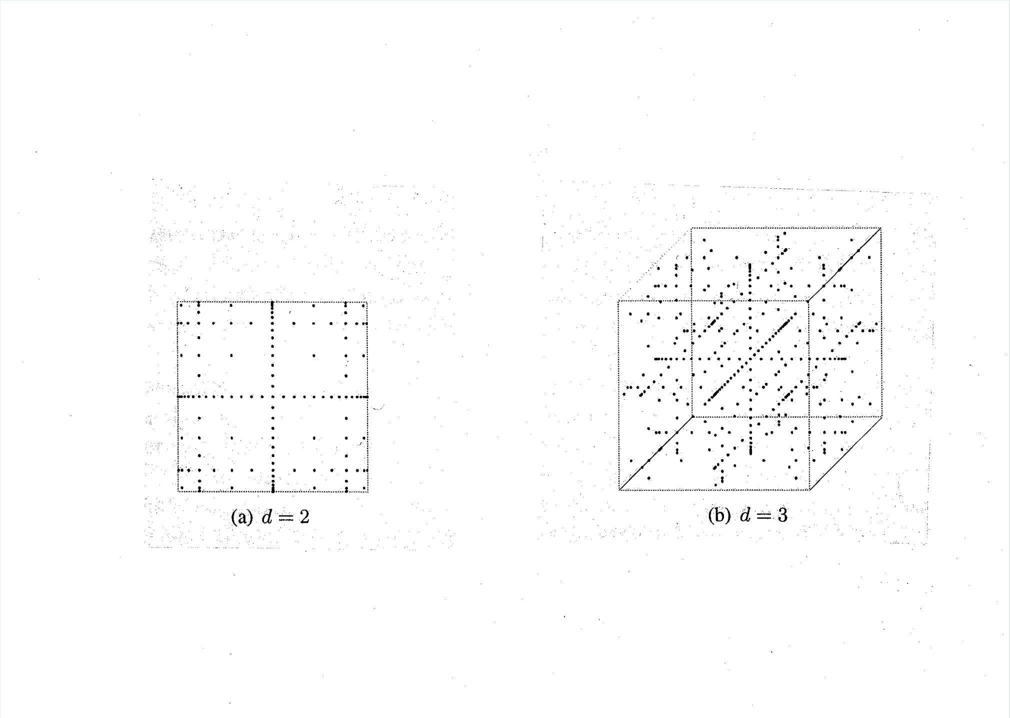
Average rate of convergence for QMC: $O(n^{-0.9})$ and $O(n^{-0.8})$.
Instead of $n = 2^7$ SSobol samples one would need $n = 10^4$ MC samples to achieve a similar accuracy as SSobol.



$\log_{10}$ of the relative errors of MC, SLA (randomly shifted lattice rule) and SSOB (scrambled Sobol' points)

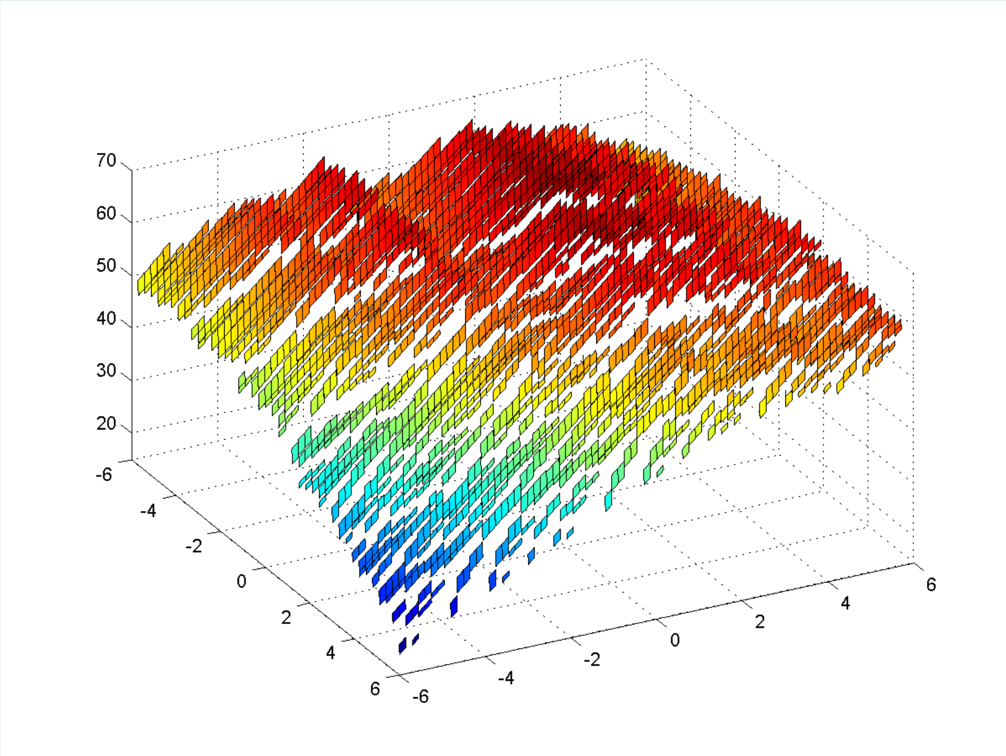
Conclusions

- Our analysis provides a theoretical basis for applying QMC methods accompanied by dimension reduction techniques to two-stage stochastic programs.
- The analysis also applies to sparse grid quadrature techniques.



Sparse grids in the unit cube $[0, 1]^d$

- The results are extendable and will be extended to mixed-integer two-stage models, to multi-stage situations, and to models with stochastic dominance constraints.



Second-stage optimal value function of an integer program (van der Vlerk)

References

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

M. Chen and S. Mehrotra: Epi-convergent scenario generation method for stochastic problems via sparse grid, *Stochastic Programming E-Print Series* 7-2008 (`www.speps.org`).

J. Dick, F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, Cambridge 2010.

M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* 82 (2013), 383-400.

H. Heitsch, H. Leövey and W. Römisch, Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs?, *Stochastic Programming E-Print Series* 5-2012 (`www.speps.org`) and submitted.

F. J. Hickernell: A generalized discrepancy and quadrature error bound, *Mathematics of Computation* 67 (1998), 299-322.

T. Homem-de-Mello: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling, *SIAM Journal on Optimization* 19 (2008), 524-551.

F. Y. Kuo: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces, *Journal of Complexity* 19 (2003), 301-320.

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, B. J. Waterhouse: Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands, *Journal of Complexity* 26 (2010), 135–160.

A. B. Owen: Randomly permuted (t, m, s)-nets and (t, s)-sequences, in: *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Lecture Notes in Statistics, Vol. 106, Springer, New York, 1995, 299–317.

A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.

A. B. Owen: Multidimensional variation for Quasi-Monte Carlo, in J. Fan, G. Li (Eds.), International Conference on Statistics, World Scientific Publ., 2005, 49–74.

T. Pennanen, M. Koivu: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.

G. Ch. Pflug, A. Pichler: Approximations of probability distributions and stochastic optimization problems, in: Stochastic Optimization Methods in Finance and Energy (M.I. Bertocchi, G. Consigli, M.A.H. Dempster, eds.), Springer, 2011.

W. Römisch: Stability of stochastic programming problems, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.

I. H. Sloan and H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.

I. H. Sloan, F. Y. Kuo and S. Joe: Constructing randomly shifted lattice rules in weighted Sobolev spaces, *SIAM Journal Numerical Analysis* 40 (2002), 1650–1665.

X. Wang and K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.

X. Wang and I. H. Sloan: Low discrepancy sequences in high dimensions: How well are their projections distributed ? *Journal of Computational and Applied Mathematics* 213 (2008), 366–386.

X. Wang and I. H. Sloan, Quasi-Monte Carlo methods in financial engineering: An equivalence principle and dimension reduction. *Operations Research* 59 (2011), 80–95.