Are Quasi-Monte Carlo methods efficient for two-stage stochastic programs?

H. Heitsch, 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

ISMP, TU Berlin, August 19–24, 2012

Introduction

- Standard approach for solving stochastic programs are variants of Monte Carlo (MC) for generating scenarios (i.e., samples).
- Recent alternative approaches to scenario generation:
  (a) Optimal quantization of probability distributions (Pflug-Pichler 2010).
  (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, Kaut-Wallace 07, Gülpinar-Rustem-Settergren 04)
- MC and (a) may be justified by available stability results, but there is almost no reasonable justification for (b), (c) and (d).
- Known convergence rates: MC $O(n^{-\frac{1}{2}})$, (a) $O(n^{-\frac{1}{d}})$ (b) $O(n^{-1}(\log n)^d)$, recently: $O(n^{-1+\delta})$ (δ small) (d dimension of random vector, n number of scenarios).

Two-stage linear stochastic programs

Two-stage stochastic programs are of the form

$$\min\left\{\langle c,x\rangle+\int_{\mathbb{R}^d}\Phi(h(\xi)-T(\xi)x)P(d\xi):x\in X\right\},$$

where X is convex polyhedral in $\mathbb{R}^m$, $c\in\mathbb{R}^m$, $h(\xi)\in\mathbb{R}^r$ and the (r,m)-matrix $T(\xi)$ are affine functions of ξ, $q\in\mathbb{R}^{\bar{m}}$, W a $(r,\bar{m})$-matrix, P a probability distribution on $\mathbb{R}^d$, and

$$\Phi(t)=\inf\{\langle q,y\rangle : y\in\mathbb{R}^{\bar{m}}, Wy=t, y\geq 0\}.$$

Then $\operatorname{dom}\Phi=W(\mathbb{R}^{\bar{m}}_+)$ is a polyhedral cone and it holds

$$\Phi(t)=\max_{j=1,\ldots,\ell} t^\top v^j \quad (t\in\operatorname{dom}\Phi),$$

where v^j, $j=1,\ldots,\ell$, are the vertices of $\mathcal{D}=\{z:W^\top z\leq q\}$. Hence, the integrand is the convex piecewise linear function

$$f(\xi)=f_x(\xi)=c^\top x+\max_{j=1,\ldots,\ell}(h(\xi)-T(\xi)x)^\top v^j \quad (x\in X)$$

if $h(\xi)-T(\xi)x\in W(\mathbb{R}^{\bar{m}}_+)$ for every $\xi\in\Xi=\operatorname{supp}P$.

Quasi-Monte Carlo methods

We consider the approximate computation of

$$I_d(f) = \int_{[0,1]^d} f(\xi)d\xi \quad \text{or} \quad I_d(f) = \int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi$$

by a QMC algorithm

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

with (non-random) points ξ^i, $i = 1, \ldots, n$, from $[0,1]^d$ or $\mathbb{R}^d$.
We assume that f belongs to a linear normed space $\mathbb{F}_d$ with norm $\|\cdot\|_d$ and unit ball $\mathbb{B}_d$. Worst-case error of $Q_{n,d}$ over $\mathbb{B}_d$:

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

Example: F_d is a weighted tensor product Sobolev space $\bigotimes_{i=1}^{d} W_2^1([0,1])$, a particular kernel reproducing Hilbert space.

Problem: Integrands in stochastic programming are not in F_d (even not of bounded variation (Owen 05)).

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

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

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, $j \in D \setminus u$ and $j \in u \setminus v$, respectively. The second representation motivates that f_u is essentially as smooth as $P_{-u}(f)$.

If f belongs to $L_{2,\rho}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u \subseteq D}$ 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

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

Sobol's global sensitivity indices of f w.r.t. ξ_j, $j \in u$:

$$\bar{S}_u = \frac{1}{\sigma^2(f)} \sum_{v \cap u \neq \emptyset} \sigma_v^2(f).$$

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

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

Mean superposition dimension of f:

$$\bar{d}_S = \sum_{\emptyset \neq u \subseteq D} |u| \frac{\sigma_u^2(f)}{\sigma^2(f)} = \sum_{i=1}^{d} \bar{S}_{\{i\}}.$$

Efficient superposition (truncation) dimension $d_T(\varepsilon)$ of f is the $(1-\varepsilon)$-quantile of ν_S (ν_T).

ANOVA decomposition of two-stage integrands

Assumption:

(A1) $h(\xi) - Tx \in W(\mathbb{R}^{\bar{m}}_{+})$ for all $x \in X$ and $\xi \in \Xi = \operatorname{supp} P$ (*relatively complete recourse*).

(A2) $\mathcal{D} \neq \emptyset$ (*dual feasibility*).

(A3) $\int_{\mathbb{R}^d} \|\xi\| P(d\xi) < \infty$.

(A4) P has a density of the form $\rho(\xi) = \prod_{j=1}^{d} \rho_j(\xi_j)$ $(\xi \in \mathbb{R}^d)$ with continuous density ρ_j, $j = 1, \ldots, d$.

The integrand $f = f_x$ is convex piecewise linear, i.e.,

$$f(\xi) = f_x(\xi) = \max_{j=1,\ldots,\ell} a_j(x)^\top \xi + \alpha_j(x),$$

where $a_j(x) \in \mathbb{R}^d$ and $\alpha_j(x)$ are affine functions of x. It holds that

$$f_x(\xi) = a_j(x)^\top \xi + \alpha_j(x), \quad \forall \xi \in K_j \quad (j = 1, \ldots, \ell),$$

where $K_j = K_j(x) = \{\xi \in \mathbb{R}^d : h(\xi) - T(\xi)x \in \mathcal{K}_j\}$ is convex polyhedral and $\mathcal{K}_j$ the normal cone to $\mathcal{D}$ at the vertex v^j $(j = 1, \ldots, \ell)$. The intersection $K_j \cap K_{j'}$ of two adjacent polyhedral sets is contained in a $(d-1)$-dimensional affine subspace of $\mathbb{R}^d$.

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

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

Assuming (A1)–(A4) it is possible to derive an explicit representation of $P_k(f)$ depending on ξ^k and on the finitely many points at which the one-dimensional affine subspace $\{\xi_s : s \in \mathbb{R}\}$ meets the intersections of two adjacent polyhedral sets K_j. This leads to

Proposition:

Let $k \in D$, $x \in X$. Assume (A1)–(A4) and that vectors a_j belonging to adjacent polyhedral sets K_j have different kth components. Then the kth projection $P_k f$ is twice continuously differentiable. $P_k f$ is infinitely differentiable if the density ρ_k is in $C^\infty(\mathbb{R})$.

Proof:

$\frac{\partial^2 P_k f}{\partial \xi_l \partial \xi_r}(\xi^k) = \sum_{i=1}^{p} \frac{-w_{il} w_{ir}}{w_{ik}} \rho_k(s_i(\xi^k))$, where $w_i = a_{j_i} - a_{j_{i+1}}$ and s_i is an affine function.

Theorem:
Let $x \in X$, assume (A1)–(A4) and that the following geometric condition (GC) be satisfied: All $(d-1)$-dimensional affine subspaces containing $(d-1)$-dimensional intersections of adjacent polyhedral sets K_j are not parallel to any coordinate axis. Then the ANOVA approximation

$$f_{d-1} := \sum_{u \subset D} f_u \quad \text{with} \quad f = f_{d-1} + f_D$$

of f is infinitely differentiable if all densities ρ_k belong to $C_b^\infty(\mathbb{R})$.

Example: Let $\bar{m} = 3$, $d = 2$, P denote the two-dimensional standard normal distribution, $h(\xi) = \xi$, q and W be given by

$$W = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix} \qquad q = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

Then (A1) and (A2) are satisfied and the dual feasible set $\mathcal{D}$ is

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

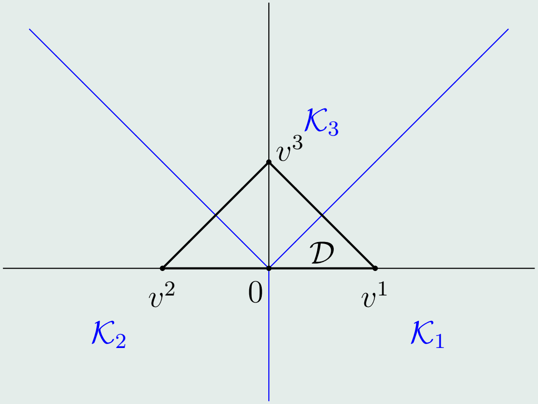


Figure 1: Illustration of $\mathcal{D}$, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

Hence, the second component of the two adjacent vertices v^1 and v^2 coincides. The function Φ is of the form

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

and the integrand is

$$f(\xi) = \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}$$

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

Proposition: Let $x \in X$, (A1), (A2) be satisfied, $\mathrm{dom}\,\Phi = \mathbb{R}^r$ 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 space of orthogonal (d, d)-matrices for the spectral decomposition of Σ.

Question: For which two-stage stochastic programs is $\|f_D\|_{L_{2,\rho}}$ small, i.e., the efficient truncation dimension is less than $d-1$ or even much less?

Dimension reduction in case of 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 efficient truncation dimension in financial models if PCA is used.

A problem-dependent principle may be based on the following equivalence principle (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 efficient truncation dimension is minimized (Wang-Sloan 11).

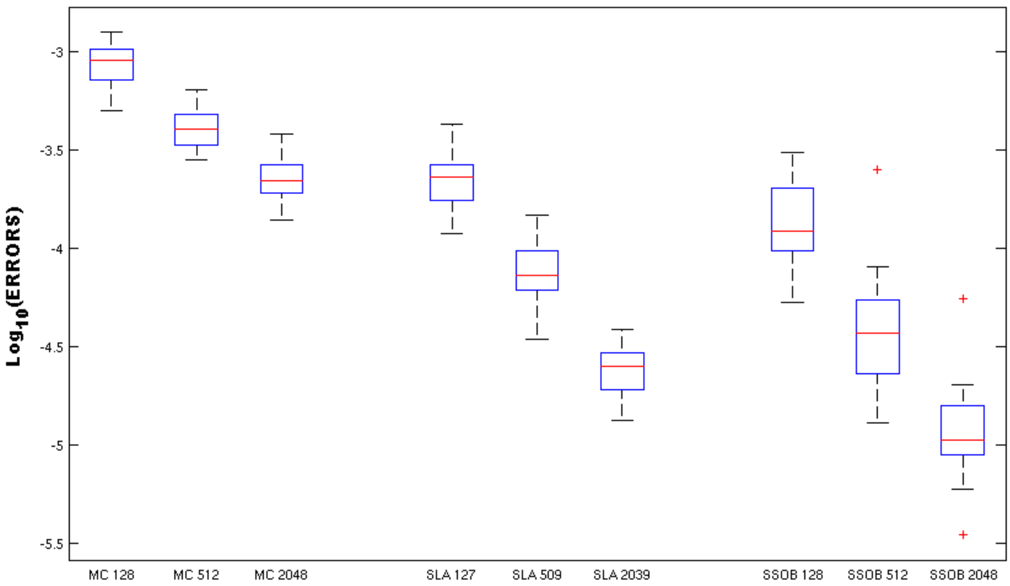
Some computational experience

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

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

The use of PCA for decomposing the covariance matrix has led to efficient 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 used 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.


Log10(ERRORS)
-3
-3.5
-4
-4.5
-5
-5.5
MC 128
MC 512
MC 2048
SLA 127
SLA 509
SLA 2039
SSOB 128
SSOB 512
SSOB 2048

Conclusions

- Our analysis provides a theoretical basis for applying QMC accompanied by dimension reduction techniques to stochastic programs with low efficient dimension.
- The results are extendable and will be extended to more general two-stage and to multi-stage situations.
- The analysis also applies to sparse grid quadrature techniques.

Thank you !

Appendix: QMC quadrature error estimates

The QMC quadrature error allows to derive the following bound (by using the ANOVA decomposition and Hickernell 98)

$$\left|\int_{[0,1]^d} f(\xi)d\xi - \frac{1}{n}\sum_{j=1}^n f(\eta_j)\right| \leq \sum_{0<|u|}\left|\int_{[0,1]^d} f_u(\xi^u)d\xi^u - \frac{1}{n}\sum_{j=1}^n f_u(\eta_j^u)\right|$$

$$\leq \sum_{0<|u|<d} \mathrm{Disc}_{n,u}(\eta_1^u,\ldots,\eta_n^u)\|f_u\|$$

$$+\left|\int_{[0,1]^d} f_D(\xi)d\xi - \frac{1}{n}\sum_{j=1}^n f_D(\eta_j)\right|,$$

where $\mathrm{Disc}_{n,u}$ is a discrepancy for n points in $[0,1]^{|u|}$ and $\|f_u\|$ a compatible norm, e.g. the norm in the weighted tensor product Sobolev space and the corresponding weighted L_2-discrepancy

$$\mathrm{Disc}^2_{n,u}(\eta_1^u,\ldots,\eta_n^u) = \prod_{j\in u}\gamma_j\int_{[0,1]^{|u|}} \mathrm{disc}_u^2(\xi^u)d\xi^u,$$

$$\mathrm{disc}_u(\xi^u) = \prod_{j\in u}\xi_j - \frac{1}{n}|\{j\in\{1,\ldots,n\} : \eta_j^u \in [0,\xi^u)\}|.$$

References

M. Chen and S. Mehrotra: Epi-convergent scenario generation method for stochastic problems via sparse grid, *SPEPS* 7-2008.

J. Dick, F. Pillichshammer: *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, 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* (to appear).

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

Hee Sun Hong and F. J. Hickernell: Algorithm 823: Implementing scrambled digital sequences, *ACM Transactions of Mathematical Software* 29 (2003), 95–109.

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.

J. Imai and K. S. Tan: Minimizing effective dimension using linear transformation, in *Monte Carlo and Quasi-Monte Carlo Methods 2002* (H. Niederreiter Ed.), Springer, Berlin, 2004, 275–292.

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.

D. Nuyens and R. Cools: Fast algorithms for component-by-component constructions of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces, *Mathematics of Computation* 75 (2006), 903-922.

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: Scenario generation for 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: ANOVA decomposition of convex piecewise linear functions,, Preprint Nr. 979, DFG Research Center Matheon Berlin, 2012 and submitted.

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.

I. M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* 55 (2001), 271–280.

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, Why are high-dimensional finance problems often of low effective dimension. *SIAM Journal Scientific Computing* 27 (2005), 159–183.

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.