Convergence of randomized Quasi-Monte Carlo methods for mixed-integer two-stage stochastic programs

H. Leövey und W. Römisch

Axpo, Baden, Switzerland

Humboldt-University Berlin
http://www.math.hu-berlin.de/~romisch

New Directions in Stochastic Optimization,
Mathematisches Forschungsinstitut Oberwolfach (Germany),
August 19–25, 2018

Introduction

- A breakthrough was obtained for computing high-dimensional integrals by means of randomized Quasi-Monte Carlo (RQMC) methods for integrands with mixed first derivatives (Kuo-Sloan 05).
- RQMC methods for the numerical quadrature in stochastic optimization models replace the stochastic parameter by a finite number of (random) scenarios having equal probabilities and lead to large scale approximating programs.
- Stochastic two-stage mixed-integer optimization models lead to discontinuous integrands, where the specific structure of polyhedral continuity regions and discontinuity facets is hidden.
- For RQMC methods on discontinuous integrands of the form $g(x)\mathbf{1}_B(x)$ on $[0,1]^d$, where g has finite HK variation and B is convex polyhedral, He-Wang 15 obtained the convergence rate

  $$O(n^{-\frac{1}{2}-\frac{1}{4d_*-2}+\delta})$$

  where d_* is the number of coordinate axes which are not parallel to the discontinuity faces of B and $\delta > 0$ is small.

Mixed-integer two-stage stochastic programs

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

$$\Phi(u,t):=\inf\left\{\langle u_1,y_1\rangle+\langle u_2,y_2\rangle : W_1y_1+W_2y_2\leq t, y_1\in\mathbb{R}^{m_1}, y_2\in\mathbb{Z}^{m_2}\right\}$$

for all $(u,t)\in\mathbb{R}^{m_1+m_2}\times\mathbb{R}^r$, with $c\in\mathbb{R}^m$, a closed subset X of $\mathbb{R}^m$, (r,m_1) and (r,m_2)-matrices W_1 and W_2, (r,m)-matrix V, affine functions $q(\xi)\in\mathbb{R}^{m_1+m_2}$, $h(\xi)\in\mathbb{R}^r$, and a probability density ρ on $\mathbb{R}^d$.

Assumptions:

(B1) The matrices W_1 and W_2 have only rational elements.

(B2) For each pair $(x,\xi)\in X\times\mathbb{R}^d$ it holds that $h(\xi)-Vx\in\mathcal{T}$, where

$$\mathcal{T}:=\{t\in\mathbb{R}^r : \exists(y_1,y_2)\in\mathbb{R}^{m_1}\times\mathbb{Z}^{m_2}\text{ such that } W_1y_1+W_2y_2\leq t\}.$$

(B3) For each $\xi\in\mathbb{R}^d$ the recourse cost $q(\xi)$ belongs to the dual feasible set

$$\mathcal{U}:=\left\{u=(u_1,u_2)\in\mathbb{R}^{m_1+m_2} : \exists v\in\mathbb{R}^r_-\text{ such that } W_1^\top v=u_1,\, W_2^\top v=u_2\right\}.$$

(B4) The number of integer decisions is finite.

Proposition:

Assume (B1)–(B4). The function Φ is finite and lower semicontinuous on $\mathcal{U}\times\mathcal{T}$ and there exists finitely many Borel sets $U_\nu\times B_\nu$, $\nu\in\mathcal{N}$, covering $\mathcal{U}\times\mathcal{T}$ such that their closure is convex polyhedral and Φ is bilinear in (u,t) on each $U_\nu\times B_\nu$. Φ may have points of discontinuity at the boundaries of $U_\nu\times B_\nu$.

Example: (Schultz-Stougie-van der Vlerk 98)

Stochastic multi-knapsack problem:

$\min \to \max$, $m = 2$, $m_1 = 0$, $m_2 = 4$, $d = s = 2$, $X = [-5, 5]^2$, $c = (1.5, 4)$, $h(\xi) = \xi$, $q(\xi) \equiv q = (16, 19, 23, 28)$, $y_i \in \{0, 1\}$, $i = 1, 2, 3, 4$, $P \sim U(\{5, 10, 15\}^2)$ (discrete)

$$V(\xi) \equiv V = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \qquad W = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 6 & 1 & 3 & 2 \end{pmatrix}$$

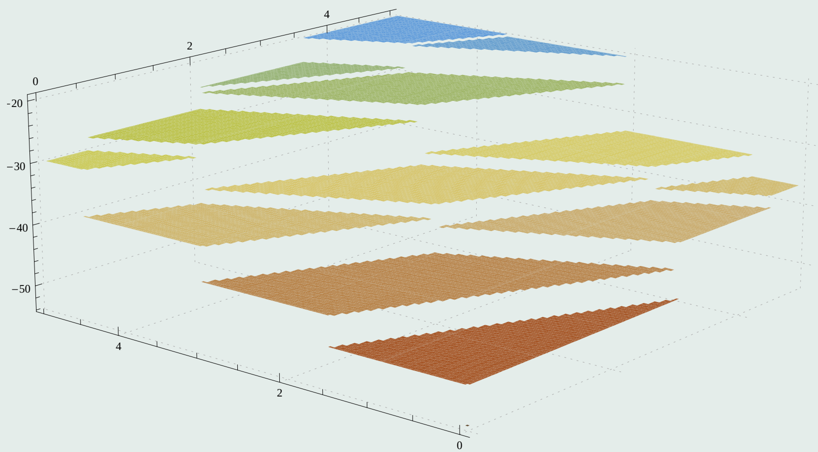


Illustration of the expected recourse function with discrete uniform probability distribution

Quasi-Monte Carlo methods

We consider the approximate computation of

$$I_d(g) = \int_{[0,1]^d} g(x)dx$$

by a Quasi-Monte Carlo (QMC) method

$$Q_n(g) = \frac{1}{n}\sum_{j=1}^{n} g(x^j)$$

with (deterministic) points x^j, $j = 1, \ldots, n$, from $[0,1]^d$.

Worst-case quadrature error in a normed space $(\mathbb{G}_d, \|\cdot\|)$:

$$e_n(\mathbb{G}_d) = \sup_{\|g\|\leq 1} |Q_n(g) - I_d(g)|$$

The convergence rate $e_n(\mathbb{G}_d) = O(n^{-1}(\log n)^{d-1})$ can be achieved in suitable normed spaces $\mathbb{G}_d$ like $\mathrm{BV}_{\mathrm{HK}}([0,1]^d)$.

Quasi-Monte Carlo methods often have good convergence properties if the integrands have low effective dimension.

Randomized QMC methods

Randomized versions of QMC point sets have the properties that

(Owen 95, L'Ecuyer-Lemieux 02, Dick-Pillichshammer 10)

(i) each point of the randomized point set has a uniform distribution over $[0,1)^d$ (uniformity), and

(ii) the QMC properties are preserved under the randomization with probability one (equidistribution).

Examples of randomization techniques are

(a) random shifts of lattice rules,

(b) scrambling, i.e., random permutations of the integers in $\mathbb{Z}_b = \{0, 1, \ldots, b-1\}$ applied to the digits in b-adic representation,

(c) affine scrambling which generates random digits by random linear transformations of the original digits, where the elements of all matrices and vectors are chosen randomly, independently and uniformly over $\mathbb{Z}_b$.

Properties (i) and (ii) enable error estimates and lead to improved convergence results compared to the original QMC method.

Weighted tensor product Sobolev spaces

$$\mathbb{G}_d = \mathcal{W}^{(1,\dots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d) = \bigotimes_{i=1}^{d} W^1_{2,\gamma_i}([0,1]),$$

where $W^1_{2,\gamma_i}([0,1])$ is the Sobolev space of absolutely continuous functions h on $[0,1]$ with derivative $h' \in L_2([0,1])$. Its inner product is

$$\langle h, \tilde{h} \rangle = \Big(\int_0^1 h(t)dt \Big) \Big(\int_0^1 \tilde{h}(t)dt \Big) + \gamma_i^{-1} \int_0^1 h'(t)\tilde{h}'(t)dt \,.$$

The weighted norm $\|g\|_\gamma = \sqrt{\langle g, g \rangle_\gamma}$ and inner product of $\mathbb{G}_d$ are given by

$$\langle g, \tilde{g} \rangle_\gamma = \sum_{u \subseteq \mathfrak{D}} \gamma_u^{-1} \int_{[0,1]^{|u|}} I_u g(t^u) I_u \tilde{g}(t^u) dt^u,$$

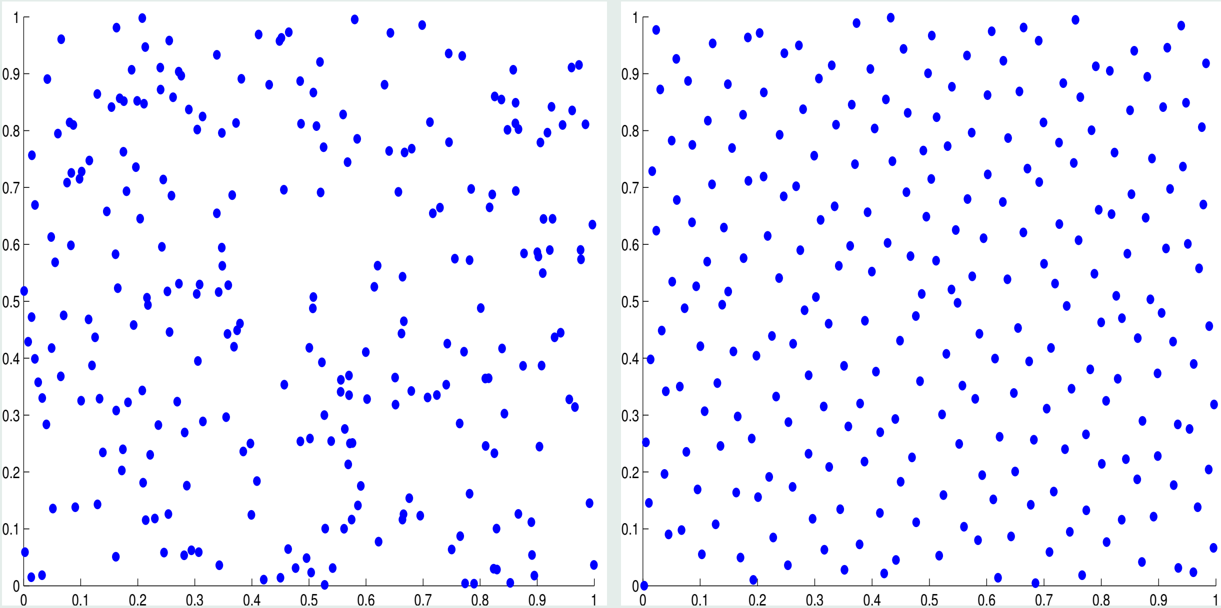
where $\mathfrak{D} = \{1, \dots, d\}$, the weights γ_i are positive nonincreasing, and

$$I_u g(t^u) = \int_{[0,1]^{d-|u|}} \frac{\partial^{|u|}}{\partial t^u} g(t) dt^{-u} \quad \text{and} \quad \gamma_u = \prod_{i \in u} \gamma_i$$

for $u \subseteq \mathfrak{D}$, where $\gamma_\emptyset = 1$. For $u \subseteq \mathfrak{D}$ we use the notation $|u|$ for its cardinality, $-u$ for $\mathfrak{D} \setminus u$ and t^u for the $|u|$-dimensional vector with components t_j, $j \in u$. Moreover, $\mathbb{G}_d$ is a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(t,s) = \prod_{i=1}^{d} \big(1 + \gamma_i (0.5 B_2(|t_i - s_i|) + B_1(t_i) B_1(s_i)) \big) \ (t, s \in [0,1]^d),$$

where $B_1(t) = t - \frac{1}{2}$ and $B_2(t) = t^2 - t + \frac{1}{6}$.



Comparison of $n = 2^7$ MC Mersenne Twister points and randomly binary shifted Sobol' points in dimension $d = 500$, projection onto the 8. and 9. components

Randomly scrambled Sobol' sequences admits the following convergence rate of the root mean square error on $\mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$

$$\sup_{\|g\|_\gamma \leq 1} \sqrt{\mathbb{E}|Q_n(\omega)(g) - I_d(g)|^2} \leq C_d\, n^{-\frac{3}{2}} (\log n)^{\frac{d-1}{2}} \,.$$

(Dick-Pillichshammer 10)

Usually a rate close to $O(n^{-1})$ is observable unless the sample sizes become huge.

Randomly shifted lattice rules

Is the random vector $\triangle$ uniformly distributed on $[0,1]^d$, we consider

$$Q_n(\omega)(g) = \frac{1}{n}\sum_{j=1}^{n} g\Big(\Big\{\frac{(j-1)}{n}\mathbf{g} + \triangle(\omega)\Big\}\Big).$$

($\{z\}$ means taking componentwise the fractional part of z)

Let $n \in \mathbb{N}$ be prime and $g \in \mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$.
Then $\mathbf{g} \in \mathbb{Z}^d_+$ can be constructed componentwise such that for each $\delta \in (0, \frac{1}{2}]$ there exists a constant $C(\delta) > 0$ with

$$\sup_{\|g\|_\gamma \le 1} \sqrt{\mathbb{E}|Q_n(\omega)(g) - I_d(g)|^2} \le C(\delta)\, n^{-1+\delta}\,,$$

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

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty \qquad (\text{e.g. } \gamma_j = \tfrac{1}{j^3}).$$

(Sloan-Kuo-Joe 02, Kuo 03, Nuyens-Cools 06)

Transformation of integrals for general densities ρ

We consider the function $f : \mathbb{R}^d \to \mathbb{R}$ and the integral

$$\int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi \,.$$

Although there are general distributional transforms for probability distributions from $\mathbb{R}^d$ to $[0,1]^d$ based on using conditional distributions (Rosenblatt 52, Rüschendorf 09) we consider here a more restrictive two-step procedure:

Step 1: Transformation of the multivariate density function ρ on $\mathbb{R}^d$ into a product-density $\rho(\xi) = \prod_{k=1}^{d} \rho_k(\xi_k)$ with d independent one-dimensional marginal densities ρ_k.

Example: If P is normal with mean 0 and nonsingular covariance matrix Σ, then for any matrix A with $\Sigma = A\,A^\top$ the density of $P \circ A$ has product form.

Step 2: Let ρ_k denote the independent marginal densities and ϕ_k the marginal distribution functions of the probability distribution P. With the transformations $x_k = \phi_k(\xi_k)$, $k \in \mathfrak{D}$, one obtains

$$\int_{\mathbb{R}^d} f(\xi) \prod_{k=1}^{d} \rho_k(\xi_k) d\xi = \int_{[0,1]^d} f(\phi_1^{-1}(x_1), \ldots, \phi_d^{-1}(x_d)) dx_1 \cdots dx_d$$

ANOVA decomposition and effective dimension

We consider a multivariate function $f : \mathbb{R}^d \to \mathbb{R}$ and intend to compute the mean of $f(\xi)$, i.e.

$$\mathbb{E}[f(\xi)] = I_{d,\rho}(f) = \int_{\mathbb{R}^d} f(\xi_1, \ldots, \xi_d)\rho(\xi_1, \ldots, \xi_d) d\xi_1 \cdots d\xi_d \,,$$

where ξ is a d-dimensional random vector with density

$$\rho(\xi) = \prod_{k=1}^{d} \rho_k(\xi_k) \quad (\xi \in \mathbb{R}^d).$$

We are interested in a representation of f consisting of 2^d terms

$$f(\xi) = f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{\substack{i,j=1 \\ i<j}}^{d} f_{ij}(\xi_i, \xi_j) + \cdots + f_{12\cdots d}(\xi_1, \ldots, \xi_d).$$

The previous representation can be more compactly written as

$$(*) \qquad f(\xi) = \sum_{u \subseteq \mathfrak{D}} f_u(\xi^u) \,,$$

where $\mathfrak{D} = \{1, \ldots, d\}$ and ξ^u contains only the components ξ_j with $j \in u$ and belongs to $\mathbb{R}^{|u|}$. Here, $|u|$ denotes the cardinality of u.

Next we make use of the space $L_{2,\rho}(\mathbb{R}^d)$ of all square integrable functions with inner product

$$\langle f, \tilde{f} \rangle_\rho = \int_{\mathbb{R}^d} f(\xi)\tilde{f}(\xi)\rho(\xi) d\xi \,.$$

A representation of the form $(*)$ of $f \in L_{2,\rho}(\mathbb{R}^d)$ is called ANOVA decomposition of f if

$$\int_{\mathbb{R}} f_u(\xi^u)\rho_k(\xi_k) d\xi_k = 0 \quad (\text{for all } k \in u \text{ and } u \subseteq \mathfrak{D}).$$

The ANOVA terms f_u, $\emptyset \neq u \subseteq \mathfrak{D}$, are orthogonal in $L_{2,\rho}(\mathbb{R}^d)$, i.e.

$$\langle f_u, f_v \rangle_\rho = \int_{\mathbb{R}^d} f_u(\xi) f_v(\xi)\rho(\xi) d\xi = 0 \quad \text{if and only if} \quad u \neq v,$$

The ANOVA terms f_u allow a representation in terms of (so-called) (ANOVA) projections, i.e. for $\xi^k = \Pi_k \xi = (\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_d)$

$$(P_k f)(\xi^k) = \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d)\rho_k(s) ds \;\; (k \in \mathfrak{D}).$$

and

$$P_u f(\xi^u) = \Big(\prod_{k \in u} P_k \Big) f(\xi^u) \quad (\xi^u = \Pi_u \xi, u \subseteq \mathfrak{D}).$$

Then it holds (Kuo-Sloan-Wasilkowski-Woźniakowski 10):

$$f_u = \Big(\prod_{j\in u}(I-P_j)\Big)P_{-u}(f) = P_{-u}(f) + \sum_{v\subsetneq u}(-1)^{|u|-|v|}P_{-v}(f)\,,$$

($-u$ denotes the complement $\mathfrak{D}\setminus u$).

We consider the variances of f and f_u

$$\sigma^2(f) = \|f - I_{d,\rho}(f)\|_{2,\rho}^2 \quad \text{und} \quad \sigma_u^2(f) = \|f_u\|_{2,\rho}^2$$

and obtain

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

The quotients

$$\frac{\sigma_u^2(f)}{\sigma^2(f)} \qquad (u\subseteq\mathfrak{D})$$

are called global sensitivity indices for the importance of the group ξ_j, $j\in u$, of variables of f. For small $\varepsilon\in(0,1)$ (e.g. $\varepsilon=0.01$)

$$d_S(\varepsilon) = \min\Big\{s\in\mathfrak{D} : \sum_{|u|\leq s}\frac{\sigma_u^2(f)}{\sigma^2(f)} \geq 1-\varepsilon\Big\}$$

is called effective (superposition) dimension of f.

The following estimate is valid

$$(+) \qquad \Big\| f - \sum_{|u| \le d_S(\varepsilon)} f_u \Big\|_{2,\rho} \le \sqrt{\varepsilon}\,\sigma(f)\,,$$

i.e., the function f is approximated by a truncated ANOVA decomposition which contains all ANOVA terms f_u such that $|u| \le d_S(\varepsilon)$.

If f is nonsmooth and the ANOVA terms f_u, $|u| \le d_S(\varepsilon)$, are smoother than f, the estimate $(+)$ means an approximate smoothing of f.

Unfortunately, the effective superposition dimension is hardly computable in general, but an upper bound can be computed by finding the smallest $s \in \mathfrak{D}$ such that

$$\sum_{v \subseteq \{1,\ldots,s\}} \sigma_v^2(f) \ge (1-\varepsilon)\sigma^2(f)\,.$$

This relies on a particular integral representation of the left-hand side, where the occuring integrals are computed approximately by means of Monte Carlo or Quasi-Monte Carlo methods based on large samples.

ANOVA terms of mixed-integer two-stage integrands

Example:

Let $d = 3$, P denote a three-dimensional probability distribution with independent continuous marginal densities ρ_i and marginal distribution functions φ_i, $i = 1, 2, 3$. We consider the convex polyhedral cone

$$\mathcal{K} = \{(t_1, t_2, t_3) \in \mathbb{R}^3 : 0 \leq t_3 \leq t_1,\, 0 \leq t_2 \leq t_3,\, 0 \leq t_1 \leq t_2\}$$

and the infimal function

$$\Phi(t) = \begin{cases} 1 & , \; t \in \operatorname{int} \mathcal{K} \\ 0 & , \; \text{otherwise}, \end{cases}$$

which is piecewise constant and lower semicontinuous. The infimal value function is simple, but typical for pure integer optimization models.

Let the integrand f be defined by

$$f(\xi) = \Phi(\xi - Vx),$$

where we let for simplicity $x = 0$.

Then its kth first order ANOVA projection $P_k f$ is

$$(P_k f)(\xi^k) = \int_{-\infty}^{+\infty} \Phi(\xi_s^k) \rho_k(s) ds,$$

where $\xi^k \in \Pi_k \mathbb{R}^3$, $k \in \{1, 2, 3\}$, and

$$\begin{aligned} \Pi_1(\mathcal{K}) &= \{(t_2, t_3) \in \mathbb{R}^2 : 0 \leq t_2 \leq t_3\}, \quad \Pi_2(\mathcal{K}) = \{(t_1, t_3) \in \mathbb{R}^2 : 0 \leq t_3 \leq t_1\} \\ \Pi_3(\mathcal{K}) &= \{(t_1, t_2) \in \mathbb{R}^2 : 0 \leq t_1 \leq t_2\}. \end{aligned}$$

We obtain

$$P_1 f(\xi^1) = P_1 f(\xi_2, \xi_3) = \begin{cases} \int_{\xi_2}^{\xi_3} \rho_1(s)ds = \varphi_1(\xi_3) - \varphi_1(\xi_2) & , 0 \leq \xi_2 \leq \xi_3, \\ 0 & , \text{ otherwise,} \end{cases}$$

$$P_2 f(\xi^2) = P_2 f(\xi_1, \xi_3) = \begin{cases} \int_{\xi_3}^{\xi_1} \rho_2(s)ds = \varphi_2(\xi_1) - \varphi_2(\xi_3) & , 0 \leq \xi_3 \leq \xi_1, \\ 0 & , \text{ otherwise,} \end{cases}$$

$$P_3 f(\xi^3) = P_3 f(\xi_1, \xi_2) = \begin{cases} \int_{\xi_1}^{\xi_2} \rho_3(s)ds = \varphi_3(\xi_2) - \varphi_3(\xi_1) & , 0 \leq \xi_1 \leq \xi_2, \\ 0 & , \text{ otherwise,} \end{cases}$$

and, hence, all first order projections are continuous and piecewise differentiable. Next we consider the second order projections $P_u f$, $u \subset \{1, 2, 3\}$, $|u| = 2$. For example, one obtains that

$$P_{13} f(\xi_2) = \begin{cases} \int_0^{\xi_2} (\varphi_3(\xi_2) - \varphi_3(t))\rho_1(t)dt & , \xi_2 \geq 0 \\ 0 & , \text{ otherwise,} \end{cases}$$

is at least continuously differentiable on $\Pi_{\{1,3\}}(\mathcal{K}) = \mathbb{R}$. The same is true for the other second order projections. Of course, first and second order ANOVA projections gain further smoothness if the densities get smoother. Hence, the first order ANOVA terms of f belong at least to $C^1(\mathbb{R})$.

Note that all facets of $\mathcal{K}$ do not parallel any coordinate axis in $\mathbb{R}^3$.

We consider the objective function

$$f(\xi) = \langle c, x \rangle + \Phi(q(\xi), h(\xi) - Vx)$$

of our two-stage mixed-integer stochastic program for fixed $x \in X$. We know that f is linear-quadratic in ξ on the Borel sets $\Xi_\nu(x) = \{\xi \in \mathbb{R}^d : q(\xi) \in U_\nu, h(\xi) \in Vx + B_\nu\}$, $\nu \in \mathcal{N}$, covering $\mathbb{R}^d$ such that their closures are convex polyhedral.

Assumptions:

(B5) The density ρ has fourth order absolute moments.
(B6) $\rho(\xi) = \prod_{k=1}^{d} \rho_k(\xi_k)$ with $\rho_k \in C^1(\mathbb{R})$, $k \in \mathfrak{D}$.
(B7) All facets of the convex polyhedral sets $\mathrm{cl}\, \Xi_\nu(x)$ do not parallel any coordinate axis in $\mathbb{R}^d$ for any fixed $x \in X$ (geometric condition).

Theorem:

Assume (B1)–(B7). Then the ANOVA terms f_u, $|u| \leq 2$, $u \subset \mathfrak{D}$, of f have all mixed first Sobolev derivatives.

Proposition: If ρ is multivariate normal with nonsingular covariance matrix Σ, the geometric condition (B7) is satisfied almost everywhere with respect to the Haar measure over the topological group of $d \times d$ real orthogonal matrices needed to transform Σ into diagonal form.

RQMC convergence and error analysis

Assume $f \in \mathbb{F}_d = \bigotimes_{i=1}^d W^1_{2,\gamma_i,\psi_i}(\mathbb{R})$, where $W^1_{2,\gamma_i,\psi_i}(\mathbb{R})$ is the Sobolev space of functions $h \in L_{2,\rho_i}(\mathbb{R})$, which are absolutely continuous with derivative $h' \in L_{2,\psi_i}(\mathbb{R})$, and norm

$$\|h\|^2_{\gamma_i,\psi_i} = \Big(\int_{\mathbb{R}} h(\xi)\rho_i(\xi)d\xi\Big)^2 + \frac{1}{\gamma_i}\int_{\mathbb{R}}(h'(\xi)\psi_i(\xi))^2 d\xi\,.$$

The functions ψ_i, $i = 1,\ldots,d$, are selected such that the function

$$g = f(\varphi_1^{-1}(\cdot),\ldots,\varphi_d^{-1}(\cdot))$$

belongs to $\mathbb{G}_d$ and $\mathbb{F}_d$ is a complete tensor product Sobolev space (Kuo-Sloan-Wasilkowski-Waterhouse 10, Nichols-Kuo 14).

The QMC error may be estimated as follows:

$$\begin{aligned}\Big|\int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi - n^{-1}\sum_{j=1}^n f(\xi^j)\Big| &= \Big|\int_{[0,1]^d} g(x)dx - n^{-1}\sum_{j=1}^n g(x^j)\Big| \\ &\le \sum_{0<|u|\le d}\Big|\int_{[0,1]^{|u|}} g_u(x^u)dx^u - n^{-1}\sum_{j=1}^n g_u(x^j)\Big|,\end{aligned}$$

where $x^j = (x_1^j,\ldots,x_d^j)$, $x_i^j = \varphi_i^{-1}(\xi_i^j) \in (0,1)^d$, $i = 1,\ldots,d$, $j = 1,\ldots,n$, are the QMC points.

If the points x^j, $j = 1, \ldots, n$, are randomly shifted lattice points, n is prime and $\delta \in (0, \frac{1}{2}]$, we may continue

$$\Big(\mathbb{E}\Big|\int_{[0,1]^d} g(x)dx - n^{-1}\sum_{j=1}^{n} g(x^j)\Big|^2\Big)^{\frac{1}{2}} \leq \; C(\delta)n^{-1+\delta} + \sum_{|u|>d_S(\varepsilon)}\Big(\mathbb{E}\Big|\int_{[0,1]^{|u|}} g_u(x^u)dx^u - n^{-1}\sum_{j=1}^{n} g_u(x^j)\Big|^2\Big)^{\frac{1}{2}} \leq \; C(\delta)n^{-1+\delta} + O(\sqrt{\varepsilon})$$

if the ANOVA terms g_u, $|u| \leq d_S(\varepsilon)$, belong to $\mathbb{G}_d$ and the sequence (γ_j) is selected properly.

The condition $g_u \in \mathbb{G}_d$ is satisfied if $f_u \in \mathbb{F}_d$.

Numerical results

The optimization model contains the electrical load ξ_δ and the electricity price ξ_c as stochastic parameters. Both are components of the random vector

$$\xi = (\xi_{\delta,1}, \xi_{c,1}, \ldots, \xi_{\delta,T}, \xi_{c,T})^\top.$$

The time horizon consists of T intervals. At each time period $t \in \{1, \ldots, T\}$ the load has to be covered. During peak load periods load covering requires electricity trading based on bilateral contracts with fixed prices or day-ahead trading with stochastic prices. Peak/offpeak load periods typically require to switch on/off (cycling) units.

A two-stage electricity production and trading model is of the form

$$\min \left\{ \sum_{t=1}^{T} \langle c_t, x_t \rangle + \int_{\mathbb{R}^T} \Phi(q(\xi), h(\xi) - Vx) P(d\xi) : x \in X \right\}$$

$$\Phi(q, h) = \inf \left\{ \sum_{t=1}^{T} \langle q_t, y_t \rangle : Wy + Vx \geq h, \, y \in Y \right\},$$

where x_t denotes the outputs of the base load units with costs c_t. The set X contains capacity limits and eventual ramping constraints at t.

The vector y_t of second-stage decisions contains the 0-1 decisions and outputs of cycling units, and the amounts of trading. The constraint $Wy + Vx \geq h(\xi)$ describes load covering at any t and minimum up/down times of the cycling units. The constraint $y \in Y$ describes capacity limits, ramping constraints and integer requirements. P denotes the probability distribution of ξ on $\mathbb{R}^{2T}$.

We assume that the centered stochastic load-price process $\{\bar{\xi}_t = (\bar{\xi}_{\delta,t}, \bar{\xi}_{c,t})\}_{t=1}^{T}$ may be modeled as linear multivariate time series ARMA(p,q)

$$\bar{\xi}_t + \sum_{i=1}^{p} A_i \bar{\xi}_{t-i} = \sum_{i=0}^{q} B_i \eta_{t-i}\,,\ t = 1, \ldots, T,$$

with independent standard normal innovations η_t, $t = 1, \ldots, T$, and suitable matrices A_i and B_i (Eichhorn-Römisch-Wegner 05).

Let m and Σ denote mean and covariance matrix of ξ, respectively.

To generate RQMC samples for the load/price vector ξ with mean $m = \mathbb{E}[\xi] \in \mathbb{R}^{2T}$ and covariance matrix $\Sigma = \mathbb{E}[(\xi - m)(\xi - m)^\top]$ in our two-stage mixed-integer electricity portfolio optimization model, we first decompose Σ by a suitable matrix A such that $\Sigma = A\,A^\top$. In this way we obtain a standard normal random vector $z = (z_1, \ldots, z_{2T})^T$ such that

$$\xi = Az + m\,.$$

If ϕ denotes the standard normal distribution function, then the vector $\eta = (\eta_1, \ldots, \eta_{2T})^\top$ with $z_i = \phi^{-1}(\eta_i)$, $i = 1, \ldots, 2T$, is uniformly distributed in $[0, 1]^{2T}$. We used the triangular Cholesky matrix $A = L_{\mathrm{Ch}}$ and the matrix $A = U_{\mathrm{PCA}}$ of the principal component analysis (PCA) factorization

$$U_{\mathrm{PCA}} = \left(\sqrt{\lambda_1}u_1 \cdots \sqrt{\lambda_d}u_d\right)$$

with the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d > 0$ and eigenvectors $u_1, \ldots, u_d$ of the covariance matrix Σ.

For our tests we used $T = 100$ and, hence, $d = 2T = 200$.

By computing the upper bounds of the effective dimension using 2^{15} randomly scrambled Sobol' points we obtained

$$d_S(0.01) \leq 2 \quad \text{with PCA and} \quad 2 < d_S(0.01) \leq 200 \quad \text{with CH}.$$

Hence, principal component analysis leads to a strong reduction of the effective dimension.

For the numerical tests n samples $\eta^j \in [0,1]^d$, $j = 1, \ldots, n$, of Mersenne Twister MC and of the two RQMC methods were generated and inserted after the transformations $z_i^j = \phi^{-1}(\eta_i^j)$, $i = 1, \ldots, 2T$, and $\xi^j = A\, z^j + m$, $j = 1, \ldots, n$, into

$$\min \left\{ \sum_{t=1}^{T} \langle c_t, x_t \rangle + \frac{1}{n} \sum_{j=1}^{n} \Phi(q(\xi^j), h(\xi^j) - Vx) : x \in X \right\}.$$

For MC and randomly scrambled Sobol' points we used $n = 128, 256$, 512 and for randomly shifted lattice rules $n = 127, 257, 509$ (since prime numbers n are favorable for the latter). The Mersenne Twister was also used for the random scrambling and the random shifts.

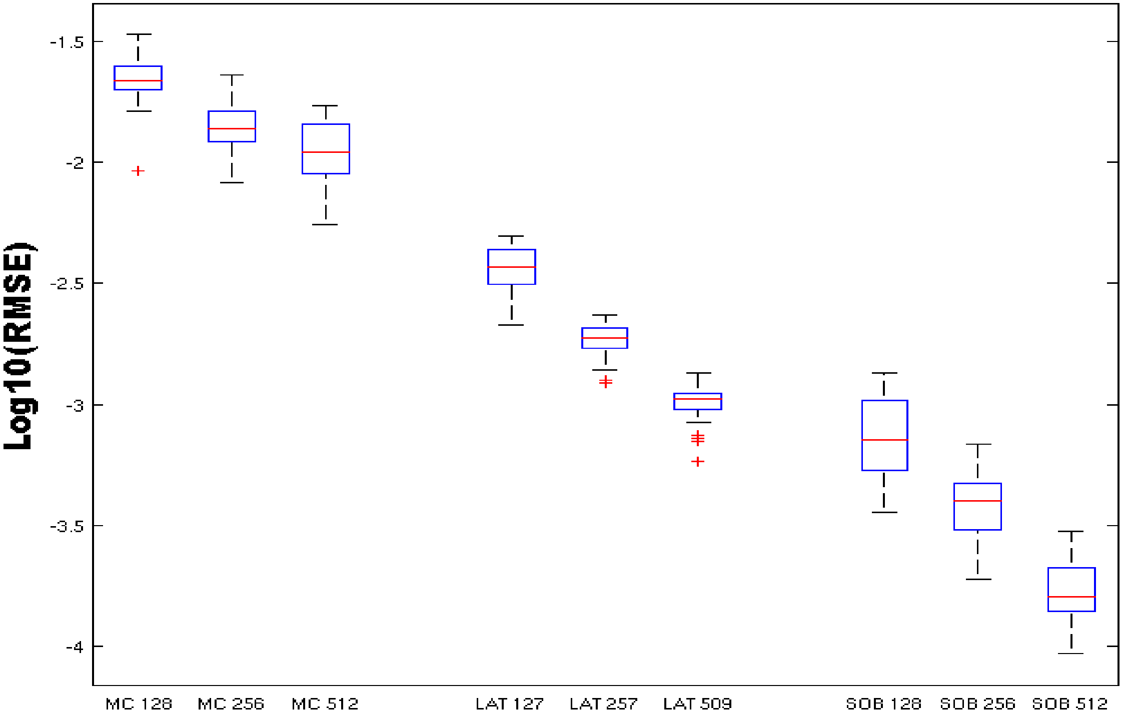
The relative root mean square error (RSME) of the optimal value of the mixed-integer linear two-stage model is estimated by performing 10 runs of every experiment and repeat the process 30 times.

The lower and upper bounds of the boxes correspond to the first and third quartile and the line in between to the median. Outliers not belonging to boxes are marked by plus signs.

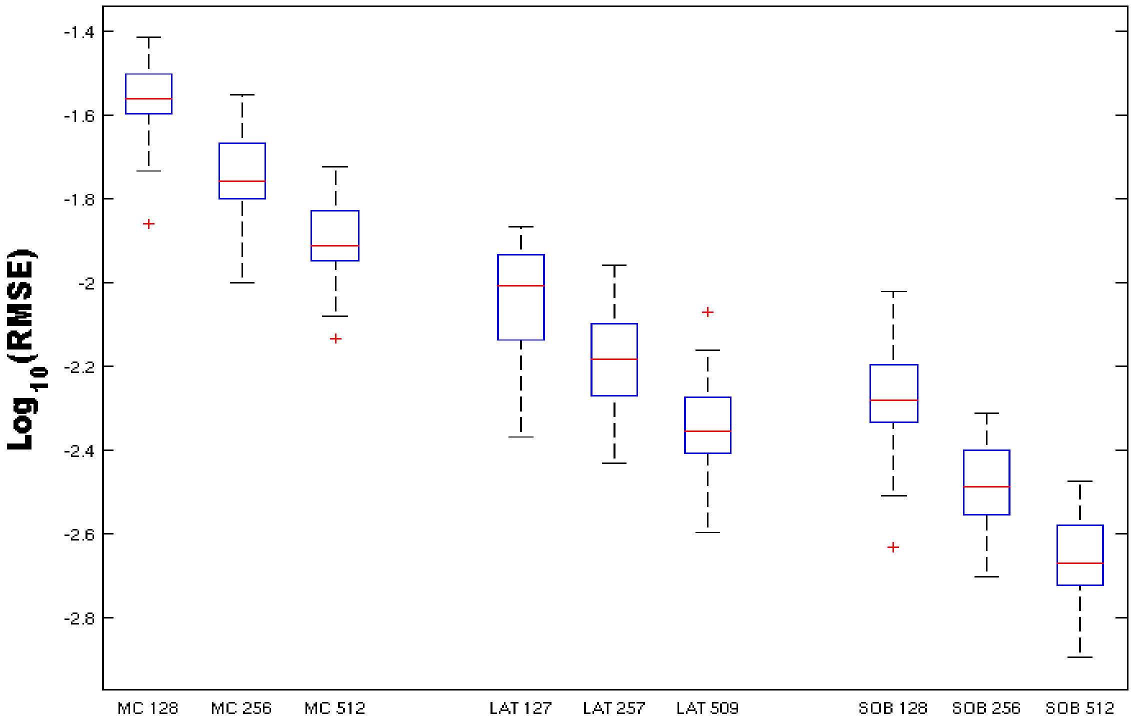
The average convergence rates of the three methods are the theoretical rate -0.5 for MC, about -0.9 for randomly shifted lattice rules und -1.0 for randomly scrambled Sobol' points if PCA factorization is used.

An explanation for the much better behavior of both randomized QMC methods is the approximate smoothing of integrands achieved by the low effective dimension due to the use of PCA.

All three methods showed only average convergence rate $-0,5$ if CH is used. However, it is also visible that both randomized QMC methods lead to smaller errors than MC.



Shown are the Log_{10} of relative RMSE for the optimal values of the two-stage model by using the **PCA** factorization of the covariance matrix. Results for Mersenne Twister MC and randomly scrambled Sobol' QMC with $128, 256$ and 512 points and randomly shifted lattice rules QMC with $127, 257$ and 509 lattice points.



Shown are the Log_{10} of relative RMSE for the optimal values of the two-stage model by using the **CH** factorization of the covariance matrix. Results for Mersenne Twister MC and randomly scrambled Sobol' QMC with $128, 256$ and 512 points and randomly shifted lattice rules QMC with $127, 257$ and 509 lattice points.

Conclusions

- Randomized Quasi-Monte Carlo methods are advantageous compared to MC methods also for integrands having kinks or even discontinuities at least in case of normal distributions and if the effective dimension of the integrand is low.
- Instead of 10^4 MC samples one only needs about 10^2 samples for randomly scrambled Sobol' point sets and randomly shifted lattice rules. The advantages consist in the improved accuracy for given sample size or in shorter running times for smaller sample sizes. The latter becomes crucial for high-dimensional models.

References:

J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 22 (2013), 133–288.

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

A. Eichhorn, W. Römisch and I. Wegner: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. IEEE St. Petersburg PowerTech Proceedings (2005).

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.

A. Griewank, F. Y. Kuo, H. Leövey and I. H. Sloan, High dimensional integration of kinks and jumps – smoothing by preintegration, *Journal of Computational and Applied Mathematics* 344 (2018), 259–274.

Z. He and X. Wang: On the convergence rate of randomized Quasi-Monte Carlo for discontinuous functions, *SIAM Journal on Numerical Analysis* 53 (2015), 2488–2503.

H. Heitsch, H. Leövey and W. Römisch, Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs?, *Computational Optimization and Applications* 65 (2016), 567–603.

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 and H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.

P. L'Ecuyer and Ch. Lemieux: Recent advances in randomized quasi-Monte Carlo methods, in *Modeling Uncertainty* (M. Dror, P. L'Ecuyer, F. Szidarovski eds.), Kluwer, Boston, 2002, 419–474.

H. Leövey and W. Römisch: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems, *Mathematical Programming* 151 (2015), 315–345.

F. Louveaux and R. Schultz: Stochastic integer programming, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 213–266.

M. Matsumoto and T. Nishimura: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* 8 (1998), 3–30.

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* (H. Niederreiter and P. J.-S. Shiue eds.), 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.

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 on Numerical Analysis* 40 (2002), 1650–1665.

I. M. Sobol': The distribution of points in a cube and the approximate evaluation of integrals, *U.S.S.R. Comput. Math. and Math. Phys.* 7 (1967), 86–112.

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