Randomized QMC methods for mixed-integer two-stage stochastic programs with application to electricity optimization

H. Leövey and W. Römisch

Abstract We consider randomized QMC methods for approximating the expected recourse in two-stage stochastic optimization problems containing mixed-integer decisions in the second stage. It is known that the second-stage optimal value function is piecewise linear-quadratic with possible kinks and discontinuities at the boundaries of certain convex polyhedral sets. This structure is exploited to provide conditions implying that first and second order ANOVA terms of the integrand have mixed first order partial derivatives in the sense of Sobolev. This shows that the integrand can be decomposed into a smooth part and a not well-behaved but small part if the effective dimension is low. This leads to good convergence properties of randomized QMC methods. In a case study we consider an optimization model for generating and trading electricity under normal load and price stochasticity. Our numerical experiments where we compare Monte Carlo and two randomized QMC methods indicate that the latter can be superior which confirms our analysis.

1 Introduction

Two-stage stochastic programming models represent a classical approach to deal with optimization problems containing random parameters in the constraints. Its idea is to introduce a two-stage decision process, where the first-stage decision x has to be decided before the randomness occurs, and the second-stage decision y satisfies the constraints that depend on x and the random parameter. Then the sum of the first-stage objective and the expected optimal value of the second-stage problem

H. Leövey
Structured Energy Management Team, Axpo, Baden, Switzerland,
e-mail: hernaneugenio.leoevey@axpo.com

W. Römisch
Humboldt-University Berlin, Institute of Mathematics, Berlin, Germany,
e-mail: romisch@math.hu-berlin.de

is optimized with respect to x. If the second-stage problem contains also integer decisions, we arrive at *mixed-integer two-stage stochastic programs*. We refer to Section 2 for a formal mathematical description and for recalling some structural properties. For further information we refer to [22, 29] and to [30] for a recent monograph on stochastic programming. We also refer to Section 6 for a practical application from electricity management.

Mixed-integer two-stage stochastic programs belong to the most complicated optimization problems. For a long time it was believed that the only way to tackle the solution of such models is by Monte Carlo (MC) methods [13]. In this paper, we study the possibility of applying randomized Quasi-Monte Carlo (QMC) methods and thereby extending our earlier work [11, 20] on two-stage models without integer decisions. In the present paper we review in Sections 2 to 4 theoretical results from [21], but discuss exclusively the error analysis in Section 5 and the numerical experiments on solving a practical optimization problem from electric power industry by using randomized QMC methods.

We consider two specific randomized QMC methods, namely, *randomly scrambled Sobol' point sets* [27, 5] and *randomly shifted lattice rules* [31, 15]. For further reading we refer to a survey of randomized QMC mehods [19] and to the recent survey [4]. It is well known that such methods display their power and fast convergence in weighted tensor product Sobolev spaces of functions on $[0,1]^d$ or $\mathbb{R}^d$ (see [4] and Section 5). However, there exist several attempts to study the convergence behavior also for functions with kinks [8] and discontinuities [9, 10]. The performance of randomized QMC methods may be significantly deteriorated for such functions. In [10] the authors derive convergence rates for functions of the form $g(x)\,\mathbb{1}_B(x)$, $x \in [0,1]^d$, where the function g is smooth and B is a convex polyhedron. They show that the convergence rate can be improved if some of the discontinuity faces of B are parallel to some coordinate axes (best case being all faces parallel to some coordinate axes since then the function exhibits bounded HK variation).

Integrands of mixed-integer two-stage models are piecewise linear-quadratic with kinks and discontinuities at boundaries of convex polyhedral sets. However, the structure of the convex polyhedra is not known, but hidden in the problem data. Therefore, our approach is different and motivated by the work of [8]. We study the smoothness of lower order ANOVA terms of the integrands and show that they are indeed much smoother than the integrand itself under certain conditions (Section 4). Hence, the integrands my be decomposed into a smooth part consisting of lower order ANOVA terms and a nonsmooth part which is small if the effective dimension of the integrand is low (see Section 3). This fact indicates that randomized QMC methods can be applied to mixed-integer two-stage models if the integrand has low effective dimension relative to the underlying probability distribution. Details are discussed in the error analysis for randomly shifted lattice rules (see Section 5) where we derive an error estimate for the root mean square error of true and approximate optimal values. In our numerical experiments we consider a practical electricity optimization model under uncertainty with normal load and price processes (see Section 6). In that case the effective dimension of the integrand can be reduced by factorizing the covariance matrix using principal component analysis.

2 Mixed-integer two-stage stochastic programs

We consider the mixed-integer two-stage stochastic optimization problem

$$\min\Big\{\langle c,x\rangle+\int_{\mathbb{R}^d}\Phi(q(\xi),h(\xi)-Vx)P(d\xi):x\in X\Big\},\tag{1}$$

where Φ denotes the parametric infimal function of the second-stage program

$$\Phi(u,t):=\inf\{\langle u_1,y_1\rangle+\langle u_2,y_2\rangle:W_1y_1+W_2y_2\le t,y_1\in\mathbb{R}^{m_1},y_2\in\mathbb{Z}^{m_2}\}\tag{2}$$

for all $(u,t)\in\mathbb{R}^{m_1+m_2}\times\mathbb{R}^r$, and $c\in\mathbb{R}^m$, 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 Borel probability measure P on $\mathbb{R}^d$. To characterize the domain of Φ we introduce

$$\begin{aligned}\mathscr{T}&=\{t\in\mathbb{R}^r:\exists(y_1,y_2)\in\mathbb{R}^{m_1}\times\mathbb{Z}^{m_2}\text{ such that }W_1y_1+W_2y_2\le t\}\\ \mathscr{U}&=\Big\{u=(u_1,u_2)\in\mathbb{R}^{m_1+m_2}:\exists v\in\mathbb{R}^r_-\text{ such that }W_1^\top v=u_1,W_2^\top v=u_2\Big\}\end{aligned}$$

the primal and dual feasible right-side sets of (2) and assume:
(A1) The matrices W_1 and W_2 have only rational elements.
(A2) The cardinality of the set

$$\bigcup_{t\in\mathscr{T}}\{y_2\in\mathbb{Z}^{m_2}:\exists y_1\in\mathbb{R}^{m_1}\text{ such that }W_1y_1+W_2y_2\le t\}$$

is finite, i.e., the number of integer decisions appearing in (2) is finite.
It is well known that the presence of integer decisions in (2) leads to discontinuities of Φ. By imposing conditions (A1) and (A2) the structure of the function Φ and of its discontinuity and nondifferentiability regions can be further characterized by utilizing results from parametric mixed-integer linear programming [1, Section 5.6].

Proposition 1. *[21] Assume (A1) and (A2). The function Φ is finite and lower semicontinuous on $\mathscr{U}\times\mathscr{T}$ and there exists a finite index set $\mathscr{N}$ and a decomposition of $\mathscr{U}\times\mathscr{T}$ consisting of Borel sets $U_\nu\times B_\nu$, $\nu\in\mathscr{N}$, such that their closure is convex polyhedral and Φ is bilinear in (u,t) on each $U_\nu\times B_\nu$. Φ may have kinks and discontinuities at the boundaries of $U_\nu\times B_\nu$.*

In order to have the integrand in (1) well defined we need the additional assumptions known as *relatively complete recourse* and *dual feasibility*:
(A3) For each pair $(x,\xi)\in X\times\mathbb{R}^d$ it holds that $h(\xi)-Vx\in\mathscr{T}$.
(A4) For each $\xi\in\mathbb{R}^d$ the recourse cost $q(\xi)$ belongs to the dual feasible set $\mathscr{U}$.

Proposition 2. *[21] Assume (A1)–(A4). Then the integrand*

$$f(x,\xi)=\langle c,x\rangle+\Phi(q(\xi),h(\xi)-Vx)\tag{3}$$

in (1) *is finite and lower semicontinuous on $X\times\mathbb{R}^d$.*
For fixed $x\in X$ the function $f(x,\cdot)$ is linear-quadratic in ξ on the Borel sets

$$\Xi_\nu(x) = \{\xi \in \mathbb{R}^d : q(\xi) \in U_\nu, h(\xi) \in Vx + B_\nu\},\ \nu \in \mathscr{N}, \tag{4}$$

that decompose $\mathbb{R}^d$ *and have convex polyhedral closures. Kinks and discontinuities of* $f(x,\cdot)$ *may appear at the boundaries of* $\Xi_\nu(x)$.

If the probability distribution P has at least finite second order moments, the objective function of (1) is finite and lower semicontinuous due to Fatou's lemma. Hence, the minimization problem (1) is well defined and solvable if the objective is inf-compact. Later we assume even a stronger moment condition in order to be able to use properties of the ANOVA decomposition which we recall next.

3 ANOVA decomposition and effective dimension

We consider a nonlinear function $f : \mathbb{R}^d \to \mathbb{R}$ and intend to compute the expectation $\mathbb{E}[f(\xi)]$ with respect to a probability distribution P having a density ρ given in product form

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

In this context, representations of f that are of interest are of the form

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

Such representations can be written more compactly in the form

$$f(\xi) = \sum_{u\subseteq\mathfrak{D}} f_u(\xi^u)\,, \tag{5}$$

where $\mathfrak{D} = \{1,\ldots,d\}$, f_u is defined on $\mathbb{R}^{|u|}$ and ξ^u belongs to $\mathbb{R}^{|u|}$ and contains only the components ξ_j with $j \in u$. Here and in what follows, $|u|$ denotes the cardinality of u and $-u$ the complement $\mathfrak{D} \setminus u$ of u.

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

$$\langle f, \tilde{f}\rangle_{2,\rho} = \int_{\mathbb{R}^d} f(\xi)\tilde{f}(\xi)\rho(\xi)d\xi\,.$$

For each function $f \in L_{2,\rho}(\mathbb{R}^d)$ a representation of the form (5) is called ANOVA decomposition of f and the functions f_u are called ANOVA terms if

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

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

$$\langle f_u, f_v \rangle_{2,\rho} = \int_{\mathbb{R}^d} f_u(\xi) f_v(\xi) \rho(\xi) d\xi = 0 \quad \text{if and only if} \quad u \neq v,$$

and allow a representation by means of (so-called) ANOVA projections. The latter are defined recursively as follows. The first and higher order projections $P_k = P_{-\{k\}}$, $k \in \mathfrak{D}$, and P_u, $u \subseteq \mathfrak{D}$, are given by

$$(P_k f)(\xi^k) = \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \rho_k(s) ds$$
$$P_u f(\xi^u) = \Big(\prod_{k \in u} P_k f\Big)(\xi^u)$$

and it holds (see [17])

$$f_u = \Big(\prod_{j \in u}(I - P_j)\Big) P_{-u}(f) = P_{-u}(f) + \sum_{v \subsetneq u} (-1)^{|u|-|v|} P_{-v}(f). \tag{6}$$

To define the effective dimension we consider the variances of f and f_u

$$\sigma^2(f) = \|f - I_{d,\rho}(f)\|_{2,\rho}^2 \quad \text{and} \quad \sigma_u^2(f) = \|f_u\|_{2,\rho}^2. \tag{7}$$

Due to the orthogonality of the ANOVA terms we obtain

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

Since the quotients $\sigma_u^2(f)/\sigma^2(f)$ indicate for any $u \subseteq \mathfrak{D}$ the importance of the group ξ_j, $j \in u$, of variables of f relative to the underlying distribution P, we define for small $\varepsilon \in (0,1)$ (e.g. $\varepsilon = 0.01$) the *effective (superposition) dimension* $d_S(\varepsilon)$ of f given P [26] as

$$d_S(\varepsilon) = \min\left\{ s \in \mathfrak{D} : \sum_{|u| \leq s} \sigma_u^2(f) \geq (1-\varepsilon)\sigma^2(f) \right\}. \tag{8}$$

An important property of the effective dimension consists in the estimate (see [32])

$$\left\| f - \sum_{|u| \leq d_S(\varepsilon)} f_u \right\|_{2,\rho} \leq \sqrt{\varepsilon}\,\sigma(f) \tag{9}$$

showing that the function f is approximated by a truncated ANOVA decomposition which contains all ANOVA terms f_u such that $|u| \leq d_S(\varepsilon)$.
If the function f is nonsmooth, the ANOVA terms f_u, $|u| \leq d_S(\varepsilon)$, are often smoother than f due to their relation to ANOVA projections and the smoothing effect of integration (see [8, 9]). Hence, the estimate (9) indicates that the main part of f can be smooth and the remaining nonsmooth part be small. 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

$$\sum_{v \subseteq \{1,\ldots,s\}} \sigma_v^2(f) \geq (1-\varepsilon)\sigma^2(f)\,. \tag{10}$$

This relies on a particular integral representation of the left-hand side of (10), where the occuring integrals can be computed approximately by means of Monte Carlo or Quasi-Monte Carlo methods based on large samples. It should be mentioned, however, that the upper bound can be (extremely) conservative.

4 ANOVA terms of mixed-integer two-stage integrands

According to Proposition 2 mixed-integer two-stage integrands (3) are discontinuous and piecewise linear-quadratic and may be written in the form

$$f(x,\xi) = \langle A_\nu(x)\xi,\xi\rangle + \langle b_\nu(x),\xi\rangle + c_\nu(x) \tag{11}$$

for all $\xi \in \Xi_\nu(x)$, $\nu \in \mathcal{N}$, $x \in X$ if (A1)–(A4) are satisfied. Here, $A_\nu(\cdot)$ are (d,d)-matrices, $b_\nu(\cdot) \in \mathbb{R}^d$ and $c_\nu(\cdot) \in \mathbb{R}$, which are all affine functions of x. The sets $\Xi_\nu(x)$, $\nu 1\mathcal{N}$, $x \in X$, are defined in (4). They decompose $\mathbb{R}^d$ and their closures are convex polyhedral.

In this section we need further assumptions to prove our main results:
(A5) The probability distribution P has finite fourth order absolute moments.
(A6) P has a density ρ with respect to the Lebesgue measure on $\mathbb{R}^d$ and ρ admits product form

$$\rho(\xi) = \prod_{i=1}^{d} \rho_i(\xi_i) \quad (\xi = (\xi_1,\ldots,\xi_d) \in \mathbb{R}^d,$$

where the densities ρ_i are positive and continuously differentiable, and ρ_i and its derivative are bounded on $\mathbb{R}$.
(A7) For each face F of dimension greater than zero of the polyhedra $\mathrm{cl}\,\Xi_\nu(x)$, $\nu \in \mathcal{N}$, the affine hull $\mathrm{aff}(F)$ of F does not parallel any coordinate axis in $\mathbb{R}^d$ for each $x \in X$ (*geometric condition*).
Due to (A5) and (A6) we may use the concepts ANOVA decomposition and effective dimension for studying mixed-integer two-stage integrands. Using the representation (11) of f the structure of first and second order ANOVA projections can be computed explicitly. This allows conclusions also on the smoothness of higher order projections and, hence, of lower order ANOVA terms due to (6). Finally this leads to the main result of this section. It is proved in [21] and states that at least lower order ANOVA terms of $f = f(x,\cdot)$ for fixed $x \in X$ have all mixed first order partial derivatives in the sense of Sobolev.

Theorem 1. *Assume (A1)–(A7). For fixed $x \in X$ we consider $f = f(x,\cdot)$. Then the ANOVA terms f_u, $|u| \leq 2$, $u \subset \mathfrak{D}$, of f are continuously differentiable and have partial mixed first Sobolev derivatives which belong to $L_{2,\rho}(\mathbb{R}^d)$.*

We recall that a real-valued function g on $\mathbb{R}^d$ is the partial *weak or Sobolev derivative* $D^\alpha f$ of a given function f if it is measurable on $\mathbb{R}^d$ and satisfies

$$\int_{\mathbb{R}^d} g(\xi)v(\xi)d\xi = (-1)^{|\alpha|}\int_{\mathbb{R}^d} f(\xi)(D^\alpha v)(\xi)d\xi \quad \text{for all} \quad v \in C_0^\infty(\mathbb{R}^d), \tag{12}$$

where $C_0^\infty(\mathbb{R}^d)$ denotes the space of infinitely differentiable functions with compact support in $\mathbb{R}^d$ and

$$D^\alpha v = \frac{\partial^{|\alpha|} v}{\partial \xi_1^{\alpha_1} \cdots \partial \xi_d^{\alpha_d}} \tag{13}$$

is the classical derivative of v of order $|\alpha| = \sum_{i=1}^d \alpha_i$, where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index. The same symbol as in (13) is also used for partial Sobolev derivatives, since classical are also Sobolev derivatives. In the classical case equation (12) is just the classical multivariate integration by parts formula.

Remark 1. Theorem 1 shows that the second order ANOVA approximation

$$f^{(2)} = \sum_{\substack{|u| \leq 2 \\ u \in \mathfrak{D}}} f_u \tag{14}$$

of the mixed-integer two-stage integrand f (see (3)) has all mixed first partial Sobolev derivatives. If the effective dimension $d_S(\varepsilon)$ of f (see (8)) is at most 2, the mean square distance between the integrand f and $f^{(2)}$ satisfies

$$\|f - f^{(2)}\|_{2,\rho}^2 \leq \varepsilon \sigma^2(f)$$

due to (9). For a discussion of techniques for reducing the effective dimension we refer to [32, 33].

While the assumptions (A1)–(A6) are reasonable, assumption (A7) seems somewhat implicit and restrictive at first sight and needs further explanation. For a normal probability distribution P with nonsingular covariance matrix Σ, the orthogonal matrix Q of eigenvectors allows a transformation of Σ into a diagonal matrix D containing the eigenvalues in its main diagonal. This observation enables the following characterization of the geometric condition (A7) using the Haar measure over the topological group of orthogonal matrices. For its proof we refer to [21] and for further information on the Haar measure to [3, Chapter 9].

Theorem 2. *We consider* (1) *and assume (A1)–(A4). If P is multivariate normal on $\mathbb{R}^d$ with nonsingular covariance matrix Σ, the geometric condition (A7) is satisfied almost everywhere with respect to the Haar measure over the topological group of orthogonal (d,d) matrices needed to transform Σ into diagonal form.*

5 Error analysis of randomly shifted lattice rules

In this section we provide an error analysis for randomly shifted lattice rules applied to solving mixed-integer two-stage stochastic programs (1). Since typical integrands in stochastic programming are defined on $\mathbb{R}^d$, we introduce first appropriate Sobolev spaces. Following [17, 25] we start with the weighted Sobolev spaces $W^1_{2,\gamma_i,\rho_i,\psi_i}(\mathbb{R})$ of functions $h \in L_{2,\rho_i}(\mathbb{R})$ that are absolutely continuous with derivatives $h' \in L_{2,\psi_i}(\mathbb{R})$ with positive continuous weight functions ψ_i, $i \in \mathfrak{D}$. They are endowed with the weighted inner product

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

where for each $i \in \mathfrak{D}$ the weight γ_i is positive and we assume that for any $x, \tilde{x} \in \mathbb{R}$

$$\int_x^{\tilde{x}} \psi_i^{-2}(t)dt < \infty\,.$$

The latter condition implies that the weighted Sobolev space is complete [14] and, thus, a Hilbert space. Then the weighted tensor product Sobolev space

$$\mathbb{F}_d = \mathscr{W}^{(1,\dots,1)}_{2,\gamma,\rho,\psi,\mathrm{mix}}(\mathbb{R}^d) = \bigotimes_{i=1}^d W^1_{2,\gamma_i,\rho_i,\psi_i}(\mathbb{R})$$

is equipped with the inner product

$$\langle f, \tilde{f} \rangle_{\gamma,\psi} = \sum_{u \subseteq \mathfrak{D}} \gamma_u^{-1} \int_{\mathbb{R}^{|u|}} I_{u,\rho}(f)(\xi^u) I_{u,\rho}(\tilde{f})(\xi^u) \prod_{i\in u} \psi_i^2(\xi_i)d\xi^u,$$

where the integrands $I_{u,\rho}(f)(\xi^u)$ and the weights γ_u are defined by

$$I_{u,\rho}(f)(\xi^u) = \int_{\mathbb{R}^{|-u|}} \frac{\partial^{|u|} f}{\partial \xi^u}(\xi) \prod_{i\in -u} \rho_i(\xi_i)d\xi^{-u} \quad \text{and} \quad \gamma_u = \prod_{i \in u} \gamma_i\,, \quad \gamma_\emptyset = 1\,.$$

In the QMC literature, this is called the unanchored setting with product weights.
In order to apply QMC methods to the computation of integrals

$$I_\rho(f) = \int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi = \int_{\mathbb{R}^d} f(\xi) \prod_{i=1}^d \rho_i(\xi_i)d\xi$$

with $f \in \mathbb{F}_d$, the Hilbert space $\mathbb{F}_d$ has to be transformed to a Hilbert space $\mathbb{G}_d$ of functions g on $[0,1]^d$ by the isometry

$$f \in \mathbb{F}_d \Longleftrightarrow g(\cdot) = f(\Phi^{-1}(\cdot)) \in \mathbb{G}_d\,,$$

where $\Phi^{-1}(t) = (\phi_1^{-1}(t_1), \ldots, \phi_d^{-1}(t_d))$, $t \in [0,1]^d$, and ϕ_i denotes the one-dimensional distribution function to the density ρ_i, $i \in \mathfrak{D}$. The inner product of $\mathbb{G}_d$ is

$$\langle g, \tilde{g} \rangle_\gamma = \langle f(\Phi^{-1}(\cdot)), \tilde{f}(\Phi^{-1}(\cdot)) \rangle_\gamma = \langle f, \tilde{f} \rangle_{\gamma,\psi}.$$

The choice of the weight functions ψ_i depends on the marginal densities ρ_i, $i \in \mathfrak{D}$. We refer to [16, 25] for a discussion of this aspect and for a list of marginal densities and the corresponding weight functions.

Now we consider randomly shifted lattice rules for numerical integration in $\mathbb{G}_d$ (see [31, 15]). Let $Z_n = \{z \in \mathbb{N} : 1 \leq z \leq n, \gcd(z,n) = 1\}$ denote the set of natural numbers between 1 and n that are relatively prime to n. Given a generating vector $\mathbf{g} \in Z_n^d$ and a random shift vector $\triangle$ which is uniformly distributed in $[0,1]^d$, the shifted lattice rule points are $t^j = \{\frac{j\mathbf{g}}{n} + \triangle\}$, $j = 1, \ldots, n$, where the braces indicate taking componentwise the fractional part. The corresponding randomized QMC method on $\mathbb{G}_d$ is of the form

$$Q_{n,d}(g) = \frac{1}{n} \sum_{j=1}^{n} g(t^j) \quad (g \in \mathbb{G}_d, n \in \mathbb{N}). \tag{15}$$

Let $\varphi(n)$ denote the cardinality of Z_n, thus, $\varphi(n) = n$ if n is prime, and let $\xi^j = \Phi^{-1}(t^j)$ for $j = 1, \ldots, n$. Then we obtain from [25, Theorem 8] that a generating vector $\mathbf{g} \in Z_n^d$ can be constructed by a component-by-component algorithm such that for each $\delta \in (0, \frac{1}{2}]$ there exists $C(\delta) > 0$ with

$$\left(\mathbb{E}\left|I_{d,\rho}(f) - Q_{n,d}(f(\Phi^{-1}(\cdot)))\right|^2\right)^{\frac{1}{2}} \leq C(\delta)\|f\|_{\gamma,\psi}\, \varphi(n)^{-1+\delta} \tag{16}$$

if the following condition

$$\sum_{i=1}^{\infty} \gamma_i^{\frac{1}{2(1-\delta)}} < \infty \tag{17}$$

on the weights is satisfied and f belongs to $\mathbb{F}_d$. To state our next result we denote by $v(P)$ the infimal value of (1) and by $v(Q_{n,d})$ the infimum if the integral in (1) is replaced by the randomly shifted lattice rule (15).

Theorem 3. *Let (A1)–(A7) be satisfied and X be compact. Assume that all integrands $f = f_x$, $x \in X$, of the form (3) have at most effective superposition dimension $d_S(\varepsilon) = 2$ for some $\varepsilon > 0$ and that the second order ANOVA approximation $f^{(2)}$ of f belongs to $\mathbb{F}_d$. Furthermore, we assume that $Q_{n,d}$ is a randomly shifted lattice rule* (15) *satisfying* (16)*. Then, for each $\delta \in (0, \frac{1}{2}]$, there exists $\hat{C}(\delta) > 0$ such that*

$$\left(\mathbb{E}\left|v(P) - v(Q_{n,d})\right|^2\right)^{\frac{1}{2}} \leq \hat{C}(\delta)\varphi(n)^{-1+\delta} + a_n\,, \tag{18}$$

where the sequence (a_n) converges to zero and allows the estimate

$$a_n \leq \sqrt{\varepsilon}\,\sigma(f) \tag{19}$$

with $\sigma(f)$ denoting the variance (7) *of f.*

Proof. Let $x \in X$ be fixed and we consider $f = f_x$. The QMC error may be estimated using the ANOVA approximation $f^{(2)}$ of f of order 2 as follows:

$$\left|I_{d,\rho}(f) - Q_{n,d}(f(\Phi^{-1}(\cdot)))\right| \leq \Big|\int_{\mathbb{R}^d} f^{(2)}(\xi)\rho(\xi)d\xi - \frac{1}{n}\sum_{j=1}^{n} f^{(2)}(\xi^j)\Big| + \Big|\int_{\mathbb{R}^d} f^{-(2)}(\xi)\rho(\xi)d\xi - \frac{1}{n}\sum_{j=1}^{n} f^{-(2)}(\xi^j)\Big|,$$

where $f^{-(2)} = f - f^{(2)}$. For any $\delta \in (0, \frac{1}{2}]$ we continue

$$\left(\mathbb{E}\left|I_{d,\rho}(f) - Q_{n,d}(f(\Phi^{-1}(\cdot)))\right|^2\right)^{\frac{1}{2}} \leq \left(\mathbb{E}\left|I_{d,\rho}(f^{(2)}) - Q_{n,d}(f^{(2)}(\Phi^{-1}(\cdot)))\right|^2\right)^{\frac{1}{2}} + \left(\mathbb{E}\left|I_{d,\rho}(f^{-(2)}) - Q_{n,d}(f^{-(2)}(\Phi^{-1}(\cdot)))\right|^2\right)^{\frac{1}{2}} \quad (20)$$

$$\leq C(\delta)\|f^{(2)}\|_{\gamma,\psi}\,\varphi(n)^{-1+\delta} + a_n\,, \quad (21)$$

where we use (16) with $f = f^{(2)}$ to estimate the first term and denote the second term by a_n. Since the integrand $f^{-(2)}$ is Riemann-integrable, the sequence (a_n) converges to zero. Next we utilize [18, Proposition 4] on expressing the variance of randomly shifted lattice rules in terms of squared Fourier coefficients, Parseval's identity for $\|f - f^{(2)}\|_{2,\rho}^2$ and the estimate (9) to obtain

$$a_n \leq \|f - f^{(2)}\|_{2,\rho} \leq \sqrt{\varepsilon}\,\sigma(f).$$

Our next step is to study how the right-hand side in the estimate (20), (21) depends on $x \in X$. The only term depending on x is the $\mathbb{F}_d$-norm of $f^{(2)} = f_x^{(2)}$. Since $f^{(2)}$ contains only ANOVA terms of order 1 and 2, its norm is given by

$$\|f^{(2)}\|_{\gamma,\psi}^2 = \sum_{|u|\leq 2} \gamma_u^{-1} \int_{\mathbb{R}^{|u|}} \Big|\int_{\mathbb{R}^{|-u|}} \frac{\partial^{|u|} f^{(2)}}{\partial \xi^u}(\xi) \prod_{i\in -u} \rho_i(\xi_i) d\xi^{-u}\Big|^2 \prod_{i\in u} \psi_i^2(\xi_i) d\xi^u.$$

Due to (14) and (6) the second order ANOVA approximation allows a representation in terms of ANOVA projections $P_u f$ with $d-2 \leq |u| \leq d$. The modulus of such ANOVA projections and of their first and second order derivatives can be bounded by some constant times $\max\{1, \|x\|\}\|\xi^{-u}\|^2$ (at least almost everywhere). Since X is compact, those bounds being continuous functions with respect to x are uniformly bounded on X. Using (A5) this implies that $\|f^{(2)}\|_{\gamma,\psi}$ can be bounded by some uniform constant $\bar{C}$. Now, it remains to appeal to a standard stability result for stochastic programs (see [28, Theorem 5]) to obtain

$$\left(\mathbb{E}\left|v(P) - v(Q_{n,d})\right|^2\right)^{\frac{1}{2}} \leq \sup_{x\in X}\left(\mathbb{E}\left|I_{d,\rho}(f_x) - Q_{n,d}(f_x(\Phi^{-1}(\cdot)))\right|^2\right)^{\frac{1}{2}} \leq C(\delta)\bar{C}\varphi(n)^{-1+\delta} + a_n\,,$$

which completes the proof. □

We note that the differentiability properties of $f^{(2)}$ in Theorem 1 motivate the condition for $f^{(2)}$ imposed in Theorem 3.

6 Application to electricity optimization under uncertainty

We consider a model for the optimal operation of an electricity company in the presence of stochasticity of the electrical load ξ_λ and market price ξ_π. The company owns a number of thermal units and bilateral contracts with other power producers. In addition it trades at electricity markets. Load and price are components of the random vector

$$\xi = (\xi_{\lambda,1}, \dots, \xi_{\lambda,T}, \xi_{\pi,1}, \dots, \xi_{\pi,T})^\top.$$

The time horizon consists of T hourly intervals. At each time period $t \in \{1, \dots, T\}$ the load has to be covered. During peak load periods the production capacity based on their own m units does eventually not suffice to cover the load. Hence, it has to buy the necessary extra amounts from other m_1 markets and m_2 producers at prices

$$p_{1,j_1,t}(\xi) = \bar{p}_{1,j_1,t} + \xi_{\pi,t},\ p_{2,j_2,t} = \bar{p}_{2,j_2,t},\ t = 1, \dots, T,\ j_1 = 1, \dots, m_1,\ j_2 = 1, \dots, m_2,$$

where the vector $\xi_{\pi,t}$ represents the stochastic part of the prices $p_{1,j_1,t}$ at the markets, and $\bar{p}_{1,j_1,t}$, $\bar{p}_{2,j_2,t}$, $t = 1, \dots, T$, represent contractual fixed prices. The aim of the company consists in minimizing its expected costs in the presence of uncertain load and prices. The two-stage stochastic electricity optimization model is of the form

$$\min \Big\{ \sum_{t=1}^{T} \sum_{j=1}^{m} c_{j,t} x_{j,t} + \int_{\mathbb{R}^{2T}} \inf\{g(x,y,u,\xi) : (y,u) \in Y(x,\xi)\} P(d\xi) : x \in X \Big\} \quad (22)$$

with the convex polyhedral feasible set

$$X := \left\{ x \in \mathbb{R}^{mT} \,\middle|\, \begin{array}{l} a_{i,t} \le x_{i,t} \le b_{i,t}\ , i = 1, \dots, m\,, t = 1, \dots, T \\ |x_{i,t} - x_{i,t+1}| \le \delta_{i,t}\ , i = 1, \dots, m\,, t = 1, \dots, T-1 \end{array} \right\},$$

where the linear constraints model capacity limits and ramping constraints. The second-stage objective function g is given by

$$g(x,y,u,\xi) = \sum_{t=1}^{T} \left[\sum_{j_1=1}^{m_1} p_{1,j_1,t}(\xi)(y_{1,j_1,t} + \eta_{j_1} u_{j_1,t}) + \sum_{j_2=1}^{m_2} p_{2,j_2,t} y_{2,j_2,t} \right]$$

and the second-stage constraint set $Y(x,\xi)$ as subset of points $(y,u) \in \mathbb{R}^{(m_1+m_2)T} \times \{0,1\}^{m_1 T}$ such that

$$\sum_{i=1}^{m} x_{i,t} + \sum_{j_1=1}^{m_1} y_{1,j_1,t} + \sum_{j_2=1}^{m_2} y_{2,j_2,t} \geq \xi_{\lambda,t}\ , \quad t = 1,\ldots,T,$$

$$\begin{aligned}
& w_{2,j_2,t} \leq y_{2,j_2,t}\ , \quad j_2 = 1,\ldots,m_2\,, t = 1,\ldots,T,\\
& |y_{2,j_2,t} - y_{2,j_2,t+1}| \leq \rho_{j_2,t}\ , \quad j_2 = 1,\ldots,m_2\,, t = 1,\ldots,T-1,\\
& w_{1,j_1,t} u_{j_1,t} \leq y_{j_1,t} \leq z_{j_1,t} u_{j_1,t}\ , \quad j_1 = 1,\ldots,m_1\,, t = 1,\ldots,T,\\
& u_{j_1,\tau} - u_{j_1,\tau-1} \leq u_{j_1,t}\ , \quad \tau = t-\overline{\tau},\ldots,t-1\,, j_1 = 1,\ldots,m_1\,, t = 1,\ldots,T,\\
& u_{j_1,\tau-1} - u_{j_1,\tau} \leq 1 - u_{j_1,t}\ , \quad \tau = t-\underline{\tau},\ldots,t-1\,, j_1 = 1,\ldots,m_1\,, t = 1,\ldots,T,
\end{aligned}$$

with fixed positive costs $c_{i,t}$, up/down price proportion η_{j_1}, bounds $a_{i,t}$, $b_{i,t}$, $\delta_{i,t}$, $w_{1,j_1,t}$,$w_{2,j_2,t}$, $z_{j_1,t}$, $\rho_{j_2,t}$ modeling capacity limits and ramp constraints. The variables $u_{j_1,t} \in \{0,1\}$, $j_1 = 1,\ldots,m_1$, $t = 1,\ldots,T$, model on/off decisions for external units and the bounds $\overline{\tau}$, $\underline{\tau}$ are their minimum up/down times.

We assume that the stochastic loads and prices $\xi_{\lambda,t}, \xi_{\pi,t}$ follow the condition

$$\begin{pmatrix}\xi_{\lambda,t}\\ \xi_{\pi,t}\end{pmatrix} = \begin{pmatrix}\bar{\xi}_{\lambda,t}\\ \bar{\xi}_{\pi,t}\end{pmatrix} + \begin{pmatrix}E_{1,t}\\ E_{2,t}\end{pmatrix},\ t = 1,\ldots,T,$$

$$\begin{pmatrix}\bar{\xi}_{\lambda,1}\\ \bar{\xi}_{\pi,1}\end{pmatrix} = B_1 \begin{pmatrix}\gamma_{1,1}\\ \gamma_{2,1}\end{pmatrix},\ \begin{pmatrix}\bar{\xi}_{\lambda,t}\\ \bar{\xi}_{\pi,t}\end{pmatrix} = A \begin{pmatrix}\bar{\xi}_{\lambda,t-1}\\ \bar{\xi}_{\pi,t-1}\end{pmatrix} + B_1 \begin{pmatrix}\gamma_{1,t}\\ \gamma_{2,t}\end{pmatrix} + B_2 \begin{pmatrix}\gamma_{1,t-1}\\ \gamma_{2,t-1}\end{pmatrix},\ t = 2,\ldots,T,$$

where $(E_{1,1},\ldots,E_{1,T})$ and $(E_{2,1},\ldots,E_{2,T})$ are fixed mean vectors for loads and prices simulating the *trend* or *seasonality*, $A, B_1, B_2 \in \mathbb{R}^{2\times 2}$, and $\gamma_{1,t}, \gamma_{2,t} \sim$ N(0,1) are independent standard normal random variables. The resulting stochastic process $\xi = \{(\xi_{\lambda,t}, \xi_{\pi,t})\}_{t=1}^{T}$ is thus a multivariate ARMA(1,1) process. Similar models have been considered for simulating prices and demands in the energy industry in the literature, see e.g. [6]. Note that since the model contains unbounded demands $\xi_{\lambda,1},\ldots,\xi_{\lambda,T}$, no upper bounds on the variables $y_{2,j_2,t}$, $j_2 = 1,\ldots,m_2\,, t = 1,\ldots,T$ were imposed, allowing the latter to cover arbitrarily large demand values. We select in addition the prices $\bar{\pi}_{2,j_2,t}$ significantly higher than the prices $\bar{\pi}_{1,j_1,t}$, such that the variables $y_{2,j_2,t}$, $j_2 = 1,\ldots,m_2\,, t = 1,\ldots,T$ do not always represent the trivial choice for costs minimization. For our tests, we chose the time horizon $T = 100$, therefore the real dimension of the model is $d = 2T = 200$. Further model constants were set to

$$A = \begin{pmatrix}0.29 & 0.44\\ 0.44 & 0.70\end{pmatrix}, \quad B_1 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad B_2 = \begin{pmatrix}0.75 & 0.053\\ 0.053 & 0.43\end{pmatrix}.$$

We refer to [2, Section 7] for detailed information about modeling with multivariate ARMA processes. The resulting joint probability distribution P of the process is normal with dimension $d = 2T$ and covariance matrix Σ. The expectation integral is transformed by factorizing the covariance matrix $\Sigma = AA^\top$ as usually recommended in normal high-dimensional integration (see [7, Sect. 2.3.3]). We carry out our tests using the standard lower triangular Cholesky matrix for A (CH) and the principal component analysis factorization, in which $A = (\sqrt{\lambda_1}u_1, \ldots, \sqrt{\lambda_d}u_d)$ with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots, \lambda_d > 0$ of Σ in decreasing order and the corresponding orthonormal eigenvectors u_i, $i = 1, \ldots, d$. Another description of PCA is

$$\Sigma = Q\,\mathrm{diag}(\lambda_1, \ldots, \lambda_d)\, Q^\top,$$

where Q denotes the orthogonal matrix $Q = (u_1 \cdots u_d)$. While the Cholesky factorization seems to assign the same importance to every variable and, hence, is not suitable to reduce the effective dimension, several authors report an enormous reduction of the effective dimension in financial models if PCA is used (e.g., [32]).
A simulated demands & prices-path ξ can then be obtained by

$$\xi = A\,(\phi^{-1}(z_1), \ldots, \phi^{-1}(z_{2T}))^\top + (E_{1,1}, \ldots, E_{1,T}, E_{2,1}, \ldots, E_{2,T}),$$

where $Z = (z_1, \ldots, z_{2T}) \sim U([0,1]^{2T})$ (i.e., the probability distribution of Z is the uniform distribution on $[0,1]^{2T}$), and $\phi^{-1}(.)$ represents the inverse cumulative normal distribution function, which can be efficiently and accurately calculated by Moro's algorithm (see [7, Sect. 2.3.2]). The evaluation begins then with MC or randomized QMC points for the samples $Z \sim U([0,1]^{2T})$. For MC points in $[0,1]^{2T}$ we used the Mersenne Twister [24] as pseudo random number generator. For QMC, we use randomly scrambled Sobol' points with direction numbers given in [12] and randomly shifted lattice rules [31, 15]. As scrambling technique we used random linear scrambling described in [23]. For our tests, we considered cubic decaying weights $\gamma_j = \frac{1}{j^3}$ for constructing the lattice rules.
We chose the following parameters for the numerical experiments:

- $m = 8, m_1 = 3, m_2 = 4$.
- For all i, j_1, j_2, t, we select randomly $a_{i,t} \in [0.001, 0.003]$, $b_{i,t} \in [0.3, 0.6]$, $\delta_{i,t} \in [0.3, 0.35]$, $w_{1,j_1,t}, w_{2,j_2,t} \in [0.000001, 0.00002]$, $z_{j_1,t} \in [5,7]$, $\gamma \in [0.1, 0.3]$, $\rho_{j_2,t} \in [1.0, 1.1]$, and $\overline{\tau} = \underline{\tau} = 2$.
- For all i, j_1, j_2, t, we select randomly $c_{i,t} \in [7,9]$, $\bar{c}_{1,j_1,t} \in [8,10]$, and $\bar{c}_{2,j_2,t} \in [11,13]$. We fixed $(E_{1,1}, \ldots, E_{1,d}) = (6, 6, \ldots, 6)$, and $(E_{2,1}, \ldots, E_{2,d}) = (0, 0, \ldots, 0)$.

We performed the following computational experiments. We fixed N sampling points ξ^j and replaced the expected recourse costs by the corresponding equal-weight MC or randomized QMC quadrature rule. Then the resulting approximate stochastic program is of the form

$$\min_{x \in X} \left\{ \sum_{t=1}^{T} \sum_{i=1}^{m} c_{i,t}\, x_{i,t} + \frac{1}{N} \sum_{j=1}^{N} g(x, y^j, u^j, \xi^j) : (y^j, u^j) \in Y(x, \xi^j), j = 1, \ldots, N \right\}. \quad (23)$$

It represents a mixed-integer linear program comprising $(m + (m_1 + m_2)N)T$ continuous and $m_1 NT$ binary variables. Since N ranges between 2^7 and 2^9, the program (23) contains more than 30.000 to 150.000 binary variables. These large scale mixed-integer linear programs are solved by means of the standard solver ILOG CPLEX (2014). The aim of the experiments is to examine the convergence rate with respect to the sample size N of the estimated optimal value from (23) obtained by replacing the expectation with MC or randomized QMC quadrature rules. We

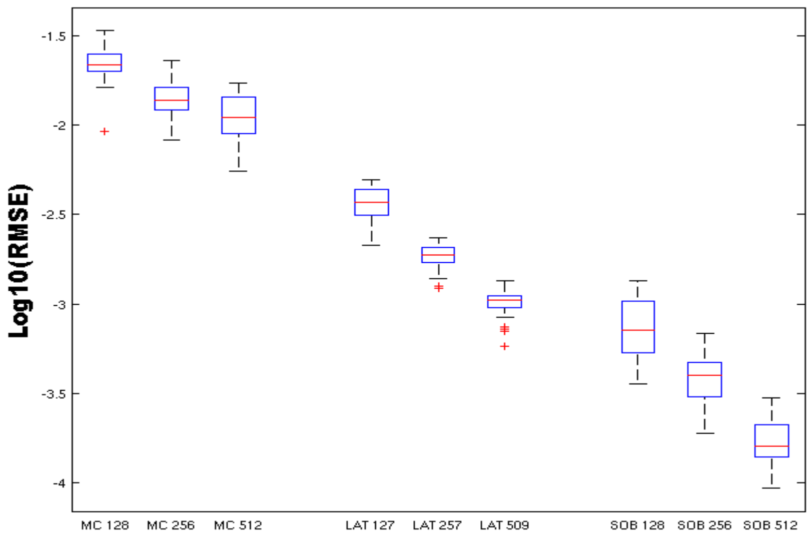

Fig. 1 Shown are the Log_{10} of the relative RMSE with PCA factorization of covariance matrix for computing the optimal value of (23) for parameters as stated above. Results for Mersenne Twister MC and randomly scrambled Sobol' QMC with $N_1 = 128$, $N_2 = 256$ and $N_3 = 512$ points (MC 128,... or SOB 128,...), and randomly shifted lattice rules QMC with $N_1 = 127$, $N_2 = 257$ and $N_3 = 509$ lattice points (LAT 127,...).

performed 5 runs for all experiments by changing the set of randomly selected parameters. But the qualitative results remained very similar, therefore we only expose one of these results in the figures. Figure 1 summarizes the convergence behavior under PCA factorizations and Table 1 shows the mean and standard deviation of the estimated optimal values under PCA for each sampling method and each sample size over the 300 replications. We chose $N_1 = 128$, $N_2 = 256$, $N_3 = 512$ as sample sizes for the Mersenne Twister and for the scrambled Sobol' points. For randomly shifted lattices, we chose $N_1 = 127$, $N_2 = 257$, $N_3 = 509$. The random shifts were generated using the Mersenne Twister. We estimate the relative root mean square errors (RMSE) of the optimal values by taking 10 runs of every experiment, and repeat the process 30 times for the box plots in the figures. The box-plots show the median value (red line), first quartile (lower bound of the box) and third quartile (upper bound of the box). Outliers are marked in red and the rest of the results lie between the brackets. The average of the estimated rates of convergence for the tests under PCA were approximately -0.91 for randomly shifted lattice rules, and

PCA	mean			standard deviation		
	N_1	N_2	N_3	N_1	N_2	N_3
MC	5022.61	5024.13	5026.24	121.86	77.53	62.28
LAT	5026.65	5026.79	5026.99	19.60	9.90	5.41
SOB	5027.14	5027.50	5027.53	4.34	2.16	0.96

Table 1 Mean and standard deviation of the estimated optimal values under PCA for different sampling methods and sample sizes.

-1.05 for the randomly scrambled Sobol' points, for different price- and bound-parameters as listed above. This is clearly superior to the MC convergence rate of -0.5. The upper bound for the effective dimension of the integrand $f(x,\cdot)$ in (22) was computed by means of (10) at 5 different feasible vertices x. We used the algorithm proposed in [32] with 2^{16} randomly scrambled Sobol' points ensuring that all results for the ANOVA total and partial variances were obtained with at least 3 digits accuracy. The upper bound of $d_S(\varepsilon)$ with $\varepsilon = 0.01$ is computed by using (10) and remained always equal to 2. We observed also that the first variable under PCA seems to accumulate always more than 90% of the total variance $\sigma^2(f(x,\cdot))$. Hence, PCA serves as excellent dimension reduction technique in this case. Additionally, we performed the same test runs by using the Cholesky decomposition CH instead of PCA for factorizing the covariance matrix. Using CH the observed results, see Figure 2, were completely different than those under PCA. The average of the estimated rates of convergence of randomized QMC was approximately -0.5, which is the same as the expected MC rate, although the implied error constants seem to be smaller for randomly shifted lattice rules and randomly scrambled Sobol' points than for MC. The upper bound for the effective dimension of the integrand $f(x,\cdot)$ in (22) was estimated by using (10) to be 200 in all tests.

7 Conclusions

The theoretical and numerical results indicate that randomized QMC methods can be superior to MC for solving two-stage stochastic programming problems at least if the recourse cost function has low effective dimension and $\sqrt{\varepsilon}\sigma(f)$ is smaller than the target accuracy for solving the optimization problem. Then using randomized QMC methods instead of MC allows a reduction of sample sizes from N approximately to $\sqrt{N}$. This fact becomes especially important when solving practical mixed-integer stochastic programming models because it reduces the dimension of the large scale mixed-integer linear programs of type (23) and, hence, leads to a considerable reduction of running time. But, Figure 2 shows that the error constants for randomized QMC methods tend to be smaller than for MC even if the effective dimension is not low. Hence, the use of randomized QMC methods for solving stochastic programs instead of MC seems to pay in any case.

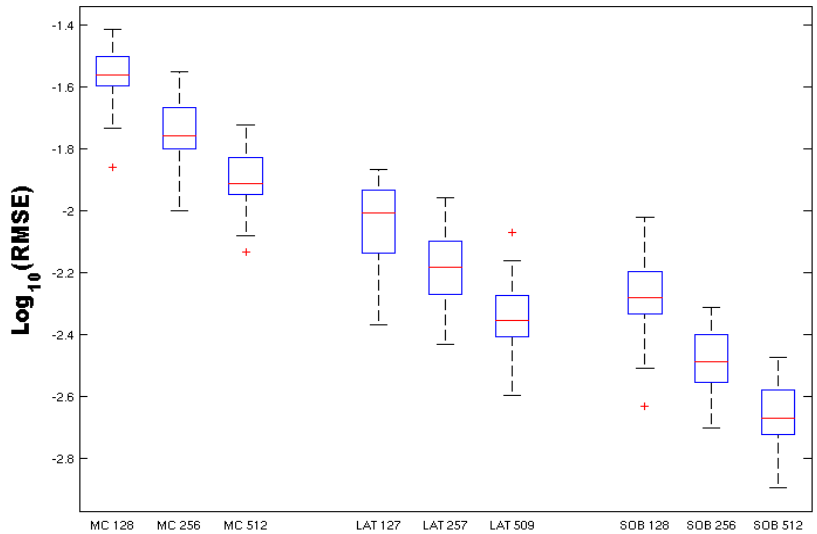

Fig. 2 Shown are the Log_{10} of relative RMSE with Cholesky factorization of covariance matrix for computing the optimal value of (23) for parameters as stated above. Results for Mersenne Twister MC and randomly scrambled Sobol' QMC with $N_1 = 128$, $N_2 = 256$ and $N_3 = 512$ points (MC 128,... or SOB 128,...), and randomly shifted lattice rules QMC with $N_1 = 127$, $N_2 = 257$ and $N_3 = 509$ lattice points (LAT 127,...).

Acknowledgement: The authors wish to thank the Editors and the referee for their detailed and valuable comments on an earlier version of this paper.

References

1. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Non-Linear Parametric Optimization. Akadamie-Verlag, Berlin (1982).
2. Brockwell, P. J., Davis, R. A.: Introduction to Time Series and Forecasting (Second Edition). Springer, New York (2002).
3. Cohn, D. L.: Measure Theory (Second Edition). Springer, New York (2013).
4. Dick, J., Kuo, F. Y., Sloan, I. H.: High-dimensional integration – the Quasi-Monte Carlo way. Acta Numerica **22**, 133–288 (2013).
5. Dick, J., Pillichshammer, F.: Digital Nets and Sequences. Cambridge University Press, Cambridge (2010).
6. Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. IEEE St. Petersburg Power Tech (2005).
7. Glasserman, P.: Monte-Carlo Methods in Financial Engineering. Springer, New York (2003).
8. Griebel, M., Kuo, F. Y., Sloan, I. H.: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition. Mathematics of Computation **82**, 383–400 (2013).
9. Griewank, A., Kuo, F. Y., Leövey, H., Sloan, I. H.: High dimensional integration of kinks and jumps – smoothing by preintegration. Journal of Computational and Applied Mathematics **344**, 259–274 (2018).

10. He, Z., Wang, X.: On the convergence rate of randomized Quasi-Monte Carlo for discontinuous functions. SIAM Journal on Numerical Analysis **53**, 2488–2503 (2015).
11. Heitsch, H., Leövey, H., Römisch, W.: Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs? Computational Optimization and Applications **65**, 567–603 (2016).
12. Joe, S., Kuo, F. Y.: Remark on Algorithm 659: Implementing Sobol's quasirandom sequence generator. ACM Transactions on Mathematical Software **29**, 49–57 (2003).
13. Kleywegt, A. J., Shapiro, A., Homem-de-Mello, T.: The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization **12**, 479–502 (2001).
14. Kufner, A., Opic, B.: How to define reasonably weighted Sobolev spaces, Commentationes Mathematicae Universitatis Carolinae **25**, 537–554 (1984).
15. Kuo, F. Y.: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces. Journal of Complexity **19**, 301-320 (2003).
16. Kuo, F. Y., Sloan, I. H., Wasilkowski, G. W., Waterhouse, B. J.: Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands. Journal of Complexity **26**, 135–160 (2010).
17. Kuo, F. Y., Sloan, I. H., Wasilkowski, G. W., Woźniakowski, H.: On decomposition of multivariate functions. Mathematics of Computation **79**, 953–966 (2010).
18. L'Ecuyer, P., Lemieux, Ch.: Variance reduction via lattice rules. Management Science **46**, 1214–1235 (2000).
19. L'Ecuyer, P., Lemieux, Ch.: Recent advances in randomized quasi-Monte Carlo methods. In: Dror, M., L'Ecuyer, P., Szidarovski, F. (eds.) Modeling Uncertainty, pp. 419–474, Kluwer, Boston (2002).
20. Leövey, H., Römisch, W.: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems. Mathematical Programming **151**, 315–345 (2015).
21. Leövey, H., Römisch, W.: Quasi-Monte Carlo methods for two-stage stochastic programs: Mixed-integer models, Optimization Online 2019–12–7546 and submitted.
22. F. Louveaux and R. Schultz: Stochastic integer programming. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, Volume 10, pp. 213–266, Elsevier, Amsterdam (2003).
23. Matoušek, J.: On the L_2-discrepancy for anchored boxes. Journal of Complexity **14**, 527–556 (1998).
24. Matsumoto, M., Nishimura, T.: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation **8**, 3–30 (1998).
25. Nichols, J. A., Kuo, F. Y.: Fast CBC construction of randomly shifted lattice rules achieving $O(n^{-1+\delta})$ convergence for unbounded integrands over $\mathbb{R}^s$ in weighted spaces with POD weights. Journal of Complexity **30**, 444–468 (2014).
26. Owen, A. B.: The dimension distribution and quadrature test functions. Statistica Sinica **13**, 1–17 (2003).
27. Owen, A. B.: Scrambling Sobol' and Niederreiter-Xing points. Journal of Complexity **14**, 466–489 (1998).
28. Römisch, W.: Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, Volume 10, pp. 483–554, Elsevier, Amsterdam (2003).
29. Schultz, R.: Stochastic programming with integer variables. Mathematical Programming **97**, 285–309 (2003).
30. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming (Second Edition). MPS-SIAM Series on Optimization, Philadelphia (2014).
31. Sloan, I. H., Kuo, F. Y., Joe, S.: Constructing randomly shifted lattice rules in weighted Sobolev spaces. SIAM Journal Numerical Analysis **40**, 1650–1665 (2002).
32. Wang, X., Fang, K.-T.: The effective dimension and Quasi-Monte Carlo integration. Journal of Complexity **19**, 101–124 (2003).
33. Wang, X., Sloan, I.H.: Quasi-Monte Carlo methods in financial engineering: An equivalence principle and dimension reduction. Operations Research **59**, 80–95 (2011).