Math. Program., Ser. B (2015) 151:315–345
DOI 10.1007/s10107-015-0898-x

FULL LENGTH PAPER

Quasi-Monte Carlo methods for linear two-stage stochastic programming problems

H. Leövey[1] · W. Römisch[1]

Received: 29 October 2014 / Accepted: 17 February 2015 / Published online: 28 March 2015

Abstract Quasi-Monte Carlo (QMC) algorithms are studied for generating scenarios to solve two-stage linear stochastic programming problems. Their integrands are piecewise linear-quadratic, but do not belong to the function spaces considered for QMC error analysis. We show that under some weak geometric condition on the two-stage model all terms of their ANOVA decomposition, except the one of highest order, are continuously differentiable and second order mixed derivatives exist almost everywhere and belong to L_2. This implies that randomly shifted lattice rules may achieve the optimal rate of convergence $O(n^{-1+\delta})$ with $\delta \in (0, \frac{1}{2}]$ and a constant not depending on the dimension if the effective superposition dimension is less than or equal to two. The geometric condition is shown to be satisfied for almost all covariance matrices if the underlying probability distribution is normal. We discuss effective dimensions and techniques for dimension reduction. Numerical experiments for a production planning model with normal inputs show that indeed convergence rates close to the optimal rate are achieved when using randomly shifted lattice rules or scrambled Sobol' point sets accompanied with principal component analysis for dimension reduction.

Mathematics Subject Classification 90C15 · 65D30 · 11K16 · 11K45

✉ W. Römisch
romisch@math.hu-berlin.de

H. Leövey
leovey@math.hu-berlin.de

1 Institute of Mathematics, Humboldt-University Berlin, Berlin, Germany

1 Introduction

During the last decade much progress has been achieved in *Quasi-Monte Carlo (QMC) theory* for computing multidimensional integrals. Appropriate function spaces of integrands were discovered that allowed to improve classical convergence rates. We refer to the classical books [31,49] for providing an overview of earlier work, and to the monographs [5,27] and the recent surveys [4,22] for presenting much of the more recent achievements.

Many stochastic programming problems may be formulated in the form

$$\min\left\{\int_{\mathbb{R}^d} f(x,\xi)P(d\xi) : x \in X\right\}, \tag{1}$$

where the integrand f is convex with respect to the first and measurable with respect to the second variable, X is a closed convex subset of $\mathbb{R}^m$ and P is a probability distribution on $\mathbb{R}^d$. We assume that P has a density ρ with respect to the Lebesgue measure λ^d. For linear two-stage stochastic programming problems the integrand f is of the form

$$f(x,\xi) = \langle c, x\rangle + \Phi(q(\xi), h(\xi) - T(\xi)x) \quad ((x,\xi) \in X \times \mathbb{R}^d), \tag{2}$$

where $c \in \mathbb{R}^m$, $q(\cdot)$, $h(\cdot)$ and $T(\cdot)$ are affine mappings defined on $\mathbb{R}^d$, Φ is the optimal value function of the second stage linear program (see Sect. 3) and X is convex polyhedral.

Most solution methods for (1) require an approximation of P by a probability measure based on a finite (possibly random or randomized) sample $\xi^1, \ldots, \xi^n$ with probabilities $p_1, \ldots, p_n$ and on solving the convex program

$$\min\left\{\sum_{j=1}^{n} p_j f(x,\xi^j) : x \in X\right\}$$

by suitable decomposition methods. So far only a few papers applied Quasi-Monte Carlo methods to stochastic programs and established, for example, convergence results (see [7,18,20,41]).

The aim of the present paper is to make use of the enormous progress in *Quasi-Monte Carlo* theory and practice, in particular, of randomly shifted lattice rules (see Sect. 2) and to provide theoretical arguments of their superiority over standard *Monte Carlo* methods with slow convergence rate $O(n^{-\frac{1}{2}})$. Randomly shifted lattice rules are known to lift the curse of dimension in numerical integration [23] if the integrands belong to certain mixed Sobolev spaces. Although typical integrands (as function of ξ) of linear two-stage stochastic programming problems do not belong to such spaces, we provide theoretical arguments that explain why randomly shifted lattice rules may converge with nearly the optimal rate $O(n^{-1})$. In comparison with our earlier work [15] the present paper extends the range of two-stage models considerably.

As a first step of our arguments we introduce ANOVA representations of multivariate functions and discuss the notion of effective dimension of such functions in Sect. 4. Section 5 contains our main theoretical results. We show that integrands $f(x,\cdot)$ given by (2) may be approximated in the L_2 sense by a function belonging to the relevant Sobolev space. More precisely, it is shown that all ANOVA terms of $f(x,\cdot)$ except the one of highest order are continuously differentiable and possess second order partial derivatives almost everywhere under some geometric condition on the dual of the second stage program. Moreover, the first and second order ANOVA terms belong to the Sobolev space and approximate the integrand if the effective superposition dimension is at most 2 (Remark 2). Error estimates show that the QMC convergence rate dominates the error in that case. In addition, we show in Sect. 6 that the geometric condition is satisfied for almost all covariance matrices if the underlying random vector is normal. In Sect. 7 we discuss techniques for reducing the effective (superposition) dimension. In accordance with the theory in Sect. 5 our computational results in Sect. 8 show that scrambled Sobol' sequences and randomly shifted lattice rules applied to a large scale two-stage stochastic programming problem achieve convergence rates close to the optimal if principal component analysis is employed for dimension reduction. Tests show that indeed the effective superposition dimension does not exceed 2.

2 Modern QMC methods: randomly shifted lattice rules and scrambled Sobol' sequences

QMC methods are designed for computing integrals of the form

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

on the domain $[0,1]^d$. QMC algorithms are equal-weight quadrature rules of the form

$$Q_{n,d}(g) = n^{-1}\sum_{j=1}^{n} g(\mathbf{t}^j) \quad (n \in \mathbb{N}),$$

where the points $\mathbf{t}^j \in [0,1]^d$ are chosen to be deterministic. There are two main groups of QMC methods (see [4,5,27,31]):

- *digital nets and sequences*,
- *lattice rules*.

The two methods we are going to describe here are randomized versions of a digital sequence and of a lattice rule, respectively. A randomized version of a QMC point set has the properties that (i) each point in the randomized point set has a uniform distribution over $[0,1)^d$ (*uniformity*), (ii) the QMC properties are preserved under the randomization with probability one (*equidistribution*). Randomization procedures for digital sequences, in particular, for Sobol' sequences, were first considered in [35]. For an overview on randomization techniques we refer to [26, Section 5] and [5, Chapter 13]. Examples of such techniques are (a) *random shifts* of lattice rules,

(b) *scrambling*, i.e., random permutations of the integers $\mathbb{Z}_b = \{0, 1, \ldots, b-1\}$ applied to the digits in b-adic representations, and (c) *affine matrix 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 $\mathbb{Z}_b$. The two properties (i) and (ii) allow to consider randomized QMC methods as variance reduction techniques that preserve the unbiasedness of the Monte Carlo estimator. They allow for error estimates and may lead to improved convergence properties compared to the original QMC method.

The first method we consider here is a *randomly shifted lattice rule* (see [21,24,34,47]) in which the QMC points are

$$\mathbf{t}^j = \left\{ \frac{(j-1)\,\mathbf{g}}{n} + \Delta \right\} \quad (j = 1, \ldots, n), \tag{3}$$

where Δ is a uniformly distributed in $[0,1)^d$ random vector, $\mathbf{g} \in \mathbb{Z}^d$ is the generator of the lattice which is obtained by a component-by-component construction and the braces $\{\cdot\}$ mean taking componentwise the fractional part. While the term $\frac{(j-1)\,\mathbf{g}}{n}$ corresponds to a classical rank-1 lattice rule, the randomization occurs by adding a random shift.

For analyzing the convergence properties of this and many other QMC methods of both groups important observations are due to [16] and [46], namely, the use of reproducing kernel Hilbert spaces of functions in general and of tensor product Sobolev spaces endowed with a weighted inner product and norm, respectively, in particular.

Let us consider a reproducing kernel Hilbert space $\mathbb{G}_d$ of functions $g : [0,1]^d \to \mathbb{R}$ with a kernel $K : [0,1]^d \times [0,1]^d \to \mathbb{R}$ satisfying $K(\cdot, t) \in \mathbb{G}_d$ and $\langle g, K(\cdot, t)\rangle = g(t)$ for each $t \in [0,1]^d$ and $g \in \mathbb{G}_d$. If $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denote the inner product and norm in $\mathbb{G}_d$, and I_d is a continuous functional on $\mathbb{G}_d$, the worst-case quadrature error $e_n(\mathbb{G}_d)$ allows the representation

$$e_n(\mathbb{G}_d) = \sup_{g \in \mathbb{G}_d, \|g\| \le 1} \left| I_d(g) - Q_{n,d}(g) \right| = \sup_{g \in \mathbb{G}_d, \|g\| \le 1} |\langle g, h_n \rangle| = \|h_n\| \tag{4}$$

for some $h_n \in \mathbb{G}_d$ according to Riesz' representation theorem for linear bounded functionals on Hilbert spaces. The *representer* h_n of the quadrature error is of the form

$$h_n(t) = \int_{[0,1]^d} K(t, s) ds - n^{-1} \sum_{j=1}^{n} K(t, \mathbf{t}^j) \quad (\forall t \in [0,1]^d).$$

An important example is the weighted tensor product Sobolev space [4]

$$\mathbb{G}_d = \mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d) = \bigotimes_{i=1}^{d} W^1_{2,\gamma_i}([0,1]), \tag{5}$$

where $W^1_{2,\gamma_i}([0,1])$ is the classical Sobolev space of single-variable absolutely continuous functions h on $[0,1]$ with derivative h' belonging to $L_2([0,1])$. Its scalar product is defined by

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

Then the tensor product $\bigotimes_{i=1}^d W^1_{2,\gamma_i}([0,1])$ is the completion of the span of products $\prod_{i=1}^d h_i(x_i)$ where h_i belongs to $W^1_{2,\gamma_i}([0,1])$ and the completion is understood in the sense of its norm $\|\cdot\|_\gamma$. The weighted norm $\|g\|_\gamma = \sqrt{\langle g, g\rangle_\gamma}$ and inner product of the tensor product space are given by

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

where $\mathfrak{D} = \{1,\ldots,d\}$, the weights γ_i are positive and nonincreasing, and γ_u is given in product form by

$$\gamma_u = \prod_{i\in u}\gamma_i$$

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 for $j \in u$. Consequently, the tensor product space $\mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$ contains functions g of d variables which have square-integrable mixed first partial derivatives $\partial^{|u|}g/\partial t^u$ for each $u \in \mathfrak{D}$. To indicate that this space is a nonclassical Sobolev space we used the sign $\mathcal{W}$ instead of the classical Sobolev space denoted by W.

Moreover, the space $\mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$ is a reproducing kernel Hilbert space with the kernel

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

where $B_1(t) = t - \frac{1}{2}$ and $B_2(t) = t^2 - t + \frac{1}{6}$ are the Bernoulli polynomials of order 1 and 2, respectively, and each factor is the kernel of the Hilbert space $W^1_{2,\gamma_i}([0,1])$.

If the integrand g belongs to the tensor product Sobolev space $\mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$, the root mean square error of randomly shifted lattice rules can be bounded by [6,21,47]

$$\sqrt{\mathbb{E}_\Delta\left|I_d(g) - Q_{n,d}(g)\right|^2} \leq C(\delta)n^{-1+\delta}\|g\|_\gamma, \tag{6}$$

where $n \in \mathbb{N}$ is prime, $\delta \in (0, \frac{1}{2}]$ and the constant $C(\delta) > 0$ does not depend on the dimension d if the sequence of weights (γ_j) satisfies

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty. \tag{7}$$

The condition (7) is satisfied, for example, for $\gamma_j = j^{-3}$, $j \in \mathbb{N}$.

The second method is a *scrambled Sobol' sequence*. Sobol' introduced the first known construction of a digital (t, d)-sequence in base $b = 2$ ([48], see also [31] and [5, Chapter 8]). The construction of Sobol' sequences is described in [5, Section 8.1.3] or [4, Example 2.18]. The quality of low dimensional projections of the points in Sobol' sequences is determined by certain parameters (called direction numbers). In our tests we used the direction numbers suggested in [19]. For practical implementations we refer to [1]. Recent developments of Sobol' sequences and comparison between available implementations can be found in [51].

As randomization technique we used the *affine matrix scrambling* proposed in [29] instead of Owen's scrambling [35] due to reductions in the implementation cost. To obtain estimates on the variance of a scrambled QMC estimator $\hat{Q}_{n,d}(g)$ for functions $g : [0, 1]^d \to \mathbb{R}$ one needs a certain degree of smoothness of g. For example, if g belongs to the tensor product Sobolev space (5), the QMC estimator $\hat{Q}_{n,d}(g)$ based on n scrambled points of a (t, d)-sequence satisfies

$$\sqrt{\mathrm{Var}(\hat{Q}_{n,d}(g))} \leq C\, n^{-\frac{3}{2}} (\log n)^{\frac{d-1}{2}} \tag{8}$$

for some constant $C > 0$ depending on g (see [5, Theorem 13.25]). Usually a rate close to $O(n^{-1})$ is observable for the QMC estimator unless the sample sizes become huge, as reported in [36].

3 Two-stage linear stochastic programming problems

We consider the linear two-stage stochastic programming problem with fixed recourse

$$\min \left\{ \langle c, x\rangle + \int_{\mathbb{R}^d} \langle q(\xi), y(\xi)\rangle P(d\xi) : Wy(\xi) = h(\xi) - T(\xi)x,\ y(\xi) \geq 0,\ x \in X \right\}, \tag{9}$$

where $c \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^m$ is convex polyhedral, W is an $(r, \overline{m})$-matrix, P is a Borel probability measure on Ξ, and the vectors $q(\xi) \in \mathbb{R}^{\overline{m}}$, $h(\xi) \in \mathbb{R}^r$ and the (r, m)-matrix $T(\xi)$ are affine functions of ξ. We define the function $f : \mathbb{R}^m \times \mathbb{R}^d \to \overline{\mathbb{R}}$ by

$$f(x, \xi) = \begin{cases} \langle c, x\rangle + \Phi(q(\xi), h(\xi) - T(\xi)x), & h(\xi) - T(\xi)x \in \operatorname{pos} W,\ q(\xi) \in \mathcal{D} \\ +\infty, & \text{otherwise} \end{cases} \tag{10}$$

where

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

and Φ the optimal value function of the second-stage problem, i.e.,

$$\Phi(u,t) = \inf\{\langle u, y\rangle : Wy = t, y \geq 0\} \quad ((u,t) \in \mathbb{R}^{\overline{m}} \times \mathbb{R}^r). \tag{11}$$

Then problem (9) may be rewritten equivalently in form (1) as a convex minimization problem with respect to the first stage decision x. Next we recall some well-known properties of the function Φ, which were derived in [54] (see also [33]).

Lemma 1 *The function Φ is finite and continuous on the $(\overline{m}+r)$-dimensional polyhedral cone $\mathcal{D} \times \operatorname{pos} W$ and there exist $(r, \overline{m})$-matrices C_j and $(\overline{m}+r)$-dimensional polyhedral cones $\mathcal{K}_j$, $j = 1, \ldots, \ell$, such that*

$$\bigcup_{j=1}^{\ell} \mathcal{K}_j = \mathcal{D} \times \operatorname{pos} W \quad \textit{and} \quad \operatorname{int} \mathcal{K}_i \cap \operatorname{int} \mathcal{K}_j = \emptyset, i \neq j,$$

$$\Phi(u,t) = \max_{j=1,\ldots,\ell} \langle C_j u, t\rangle = \max\{\langle z, t\rangle : W^\top z \leq u\} \quad ((u,t) \in \mathcal{D} \times \operatorname{pos} W),$$

$$\Phi(u,t) = \langle C_j u, t\rangle, \textit{ for each } (u,t) \in \mathcal{K}_j, j = 1, \ldots, \ell.$$

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

Next we introduce conditions on problem (9) that are needed in the next sections.

(A1) For each $(x, \xi) \in X \times \mathbb{R}^d$ it holds that $h(\xi) - T(\xi)x \in \operatorname{pos} W$ and $q(\xi) \in \mathcal{D}$.
(A2) P has finite fourth order absolute moments, i.e., $\int_{\mathbb{R}^d} \|\xi\|^4 P(d\xi) < \infty$.
(A3) P has a density of the form $\rho(\xi) = \prod_{i=1}^d \rho_i(\xi_i)$ $(\xi \in \mathbb{R}^d)$, where ρ_i is a continuous (marginal) density on $\mathbb{R}$, $i = 1, \ldots, d$ (*independent components*).
(A4) All common closed faces of adjacent polyhedral sets

$$\Xi_j(x) = \{\xi \in \mathbb{R}^d : (q(\xi), h(\xi) - T(\xi)x) \in \mathcal{K}_j\}, \; j = 1, \ldots, \ell, \tag{12}$$

do not parallel any coordinate axis for every $x \in X$ (*geometric condition*).

(A1) combines the two usual conditions: *relatively complete recourse* and *dual feasibility* and implies $X \times \mathbb{R}^d \subseteq \operatorname{dom} f$. Condition (A2) is stronger than the usually required condition that P has finite second order moments. We note, however, that later the integrands have to be quadratically integrable with respect to P. Condition (A3) is needed in the next sections to introduce and analyze the ANOVA decomposition of two-stage integrands. Since this condition is not satisfied for the underlying probability distribution P in general, it means practically that the probability distribution P has a Lebesgue density and may be transformed such that (A3) is satisfied. Condition (A4) is (only) needed in the smoothness analysis of the ANOVA terms of the integrand f in Sect. 5. Determining the polyhedral cones $\Xi_j(x)$ and, hence, checking (A4) for a particular optimization model is too costly in general. (A4) is further discussed in Sect. 6.

Proposition 1 *Let (A1) be satisfied and $x \in X$. Then the function $f(x,\cdot)$ is continuous and piecewise linear-quadratic, and of the form*

$$f(x,\xi) = \langle (A_j - B_j(x))\xi,\ \xi\rangle + \langle c_j - G_j(x), \xi\rangle + \alpha_j - a_j^\top x \quad (\xi \in \Xi_j(x)), \quad (13)$$

where A_j and $B_j(x)$ are (d,d)-matrices, $G_j(x) \in \mathbb{R}^d$, $c_j \in \mathbb{R}^d$, $\alpha_j \in \mathbb{R}$, $a_j \in \mathbb{R}^m$ and $\Xi_j(x)$ defined by (12) with $B_j(\cdot)$ and $G_j(\cdot)$, $j = 1,\ldots,\ell$, depending linearly on x.
It holds $\operatorname{int} \Xi_j(x) \neq \emptyset$, $\operatorname{int} \Xi_j(x) \cap \operatorname{int} \Xi_{j'}(x) = \emptyset$ *for $j \neq j'$ and*

$$\bigcup_{j=1}^{\ell} \Xi_j(x) = \mathbb{R}^d. \quad (14)$$

Furthermore, the intersection of two adjacent convex polyhedral sets $\Xi_i(x)$ and $\Xi_j(x)$ is contained in a $(d-1)$-dimensional affine subspace.

Proof Since $q(\cdot)$, $h(\cdot)$ and $T(\cdot)$ are affine functions of ξ, there exist q_0, q_i in $\mathbb{R}^{\overline{m}}$, h_0, h_i in $\mathbb{R}^r$ and (r,m)-matrices T_0, T_i, $i = 1,\ldots,d$, such that

$$q(\xi) = q_0 + \sum_{i=1}^{d} q_i\xi_i \quad \text{and} \quad h(\xi) - T(\xi)x = h_0 - T_0x + \sum_{i=1}^{d}(h_i - T_ix)\xi_i\,.$$

After some calculations one obtains for $\xi \in \Xi_j(x)$

$$\begin{aligned} f(x,\xi) &= \langle C_jq(\xi), h(\xi) - T(\xi)x\rangle \\ &= \sum_{i=1}^{d}\sum_{k=1}^{d} \xi_i\xi_k\langle C_jq_i, h_k - T_kx\rangle + \langle C_jq_0, h_0 - T_0x\rangle \\ &\quad + \sum_{i=1}^{d} \xi_i(\langle C_jq_i, h_0 - T_0x\rangle + \langle C_jq_0, h_i - T_ix\rangle) \\ &= \langle (A_j - B_j(x))\xi, \xi\rangle + \langle c_j - G_j(x), \xi\rangle + \alpha_j - a_j^\top x \end{aligned}$$

with the (d,d)-matrices $A_j = (\langle C_jq_i, h_k\rangle)_{i,k=1,\ldots,d}$, $B_j(x) = (\langle C_jq_i, T_kx\rangle)_{i,k=1,\ldots,d}$, the d-dimensional vectors $c_j = (\langle C_jq_i, h_0\rangle + \langle C_jq_0, h_i\rangle)$ and $G_j(x) = (\langle C_jq_i, T_0x\rangle + \langle C_jq_0, T_ix\rangle)$ with the components $i = 1,\ldots,d$, the real number $\alpha_j = \langle C_jq_0, h_0\rangle$, the m-dimensional vector $a_j = T_0^\top C_jq_0$ and $\Xi_j(x)$ as defined by (12). □

Conditions (A1) and (A2) imply that the two-stage stochastic program (2) is well defined and represents an optimization problem with finite convex objective and polyhedral convex feasible set. If X is compact its optimal value $v(P)$ is finite and its solution set $S(P)$ is nonempty, closed and convex. The quantitative stability results

[43, Theorems 5 and 9] for general stochastic programming problems imply the perturbation estimate

$$|v(P) - v(Q)| \leq \sup_{x \in X} \left| \int_{\mathbb{R}^d} f(x, \xi)(P - Q)(d\xi) \right| \tag{15}$$

$$\sup_{x \in S(Q)} d(x, S(P)) \leq \psi_P^{-1} \left(\sup_{x \in X} \left| \int_{\mathbb{R}^d} f(x, \xi)(P - Q)(d\xi) \right| \right), \tag{16}$$

where ψ_P is the growth function of the objective

$$\psi_P(\tau) = \inf \left\{ \int_{\mathbb{R}^d} f(x, \xi) P(d\xi) - v(P) : d(x, S(P)) \leq \tau, \ x \in X \right\} \quad (\tau \geq 0),$$

its inverse is defined by $\psi_P^{-1}(t) = \sup\{\tau \in \mathbb{R}_+ : \psi_P(\tau) \leq t\}$, and Q is a probability measure satisfying (A2), too.

For further information on linear parametric programming and two-stage stochastic programming we refer to [33,54] and [44,45,58].

4 ANOVA decomposition and effective dimension

The analysis of variance (ANOVA) decomposition of a multivariate function was first proposed as a tool in statistical analysis (see [17] and the survey [52]). Later it was often used for the analysis of quadrature methods mainly on $[0, 1]^d$. Here, we will use it on $\mathbb{R}^d$ equipped with a probability measure P satisfying (A3).

As in [14] we consider the weighted $\mathcal{L}_p$ space over $\mathbb{R}^d$, i.e., $\mathcal{L}_{p,\rho}(\mathbb{R}^d)$, with the norm

$$\|f\|_{p,\rho} = \begin{cases} \left(\int_{\mathbb{R}^d} |f(\xi)|^p \rho(\xi) d\xi \right)^{\frac{1}{p}} & \text{if } 1 \leq p < +\infty, \\ \operatorname{ess\,sup}_{\xi \in \mathbb{R}^d} \rho(\xi) |f(\xi)| & \text{if } p = +\infty. \end{cases}$$

Let $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$ and $\mathfrak{D}$ be as in Sect. 2. The projection P_k, $k \in \mathfrak{D}$, is defined by

$$(P_k f)(\xi) := \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \rho_k(s) ds \quad (\xi \in \mathbb{R}^d). \tag{17}$$

Clearly, the function $P_k f$ is constant with respect to ξ_k. For $u \subseteq \mathfrak{D}$ we write

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

where the product sign means composition. Due to Fubini's theorem the ordering within the product is not important and $P_u f$ is constant with respect to all ξ_k, $k \in u$. The ANOVA decomposition of $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$ is of the form [25,50,55]

$$f = \sum_{u \subseteq \mathfrak{D}} f_u \tag{18}$$

with f_u depending only on ξ^u, i.e., on the variables ξ_j with indices $j \in u$. It satisfies the property $P_j f_u = 0$ for all $j \in u$ and the recurrence relation

$$f_\emptyset = P_{\mathfrak{D}}(f) \quad \text{and} \quad f_u = P_{-u}(f) - \sum_{v \subsetneq u} f_v \quad (u \subseteq \mathfrak{D}).$$

It is known from [25] that the ANOVA terms are given explicitly in terms of the projections by

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u}(f) + \sum_{v \subsetneq u} (-1)^{|u|-|v|} P_{u-v}(P_{-u}(f)), \tag{19}$$

where P_{-u} and P_{u-v} mean integration with respect to ξ_j, $j \in \mathfrak{D} \setminus u$ and $j \in u \setminus v$, respectively. The second representation of f_u implies that the smoothness of f_u is determined by $P_{-u}(f)$ due to the Inheritance Theorem [14, Theorem 2]. The latter result shows that projections do not reduce the smoothness.

If f belongs to $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u \subseteq \mathfrak{D}}$ are orthogonal in the Hilbert space $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$ (see e.g. [55]). Let the variances of f and f_u be defined by $\sigma^2(f) = \|f - P_{\mathfrak{D}}(f)\|_{2,\rho}^2$ and $\sigma_u^2(f) = \|f_u\|_{2,\rho}^2$. Then it holds

$$\sigma^2(f) = \|f\|_{2,\rho}^2 - (P_{\mathfrak{D}}(f))^2 = \sum_{\emptyset \neq u \subseteq \mathfrak{D}} \sigma_u^2(f). \tag{20}$$

To avoid trivial cases we assume $\sigma(f) > 0$ in the following. Due to (20) the normalized ratios $\frac{\sigma_u^2(f)}{\sigma^2(f)}$ serve as indicators for the importance of the variable ξ^u in f. They are used to define sensitivity indices of a set $u \subseteq D$ for f in [50] and the dimension distribution of f in [28,37].

For small $\varepsilon \in (0, 1)$ ($\varepsilon = 0.01$ is suggested in a number of papers), the *effective superposition (truncation) dimension* $d_S(\varepsilon) \in \mathfrak{D}$ ($d_T(\varepsilon) \in \mathfrak{D}$) of f is defined by

$$d_S(\varepsilon) = \min \left\{ s \in \mathfrak{D} : \sum_{0<|u|\le s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \ge 1 - \varepsilon \right\} \tag{21}$$

$$d_T(\varepsilon) = \min \left\{ s \in \mathfrak{D} : \sum_{u \subseteq \{1,\ldots,s\}} \frac{\sigma_u^2(f)}{\sigma^2(f)} \ge 1 - \varepsilon \right\}. \tag{22}$$

Note that $d_S(\varepsilon) \le d_T(\varepsilon)$ and it holds (see [12,55])

$$\max \left\{ \left\| f - \sum_{|u| \le d_S(\varepsilon)} f_u \right\|_{2,\rho}, \left\| f - \sum_{u \subseteq \{1,\ldots,d_T(\varepsilon)\}} f_u \right\|_{2,\rho} \right\} \le \sqrt{\varepsilon}\sigma(f). \tag{23}$$

The effective truncation dimension $d_T(\varepsilon)$ is much easier to estimate than $d_S(\varepsilon)$, namely, by computing the integrals [55]

$$D_u = \int_{\mathbb{R}^d} \int_{\mathbb{R}^{d-|u|}} f(\xi) f(\xi^u, \tilde{\xi}^{-u}) \rho(\xi) \rho_{-u}(\tilde{\xi}^{-u}) d\xi d\tilde{\xi}^{-u} - P_{\mathfrak{D}}^2(f) \qquad (24)$$

for $u = \{1, \ldots, s\}$, $s = 1, 2, \ldots, d$, numerically with MC or QMC methods until $D_u \geq (1 - \varepsilon)\sigma^2(f)$ and by setting $d_T(\varepsilon) = s$.

The importance of the ANOVA decomposition in the context of this paper is due to the fact that the ANOVA terms f_u with $|u| < d$ may be smoother than the original integrand f under certain conditions (see [13,14]).

As in [14] we use the notation $D_i f$ for $i \in \mathfrak{D}$ to denote the classical partial derivative $\frac{\partial f}{\partial x_i}$. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $\alpha_i \in \mathbb{N}_0$ we set

$$D^\alpha f = \left(\prod_{i=1}^d D_i^{\alpha_i} \right) f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}},$$

and call $D^\alpha f$ the partial derivative of order $|\alpha| = \sum_{i=1}^d \alpha_i$. A real-valued function g on $\mathbb{R}^d$ is called *weak or Sobolev derivative* of f of order $|\alpha|$ if it is measurable and satisfies

$$\int_{\mathbb{R}^d} g(\xi) v(\xi) d\xi = (-1)^{|\alpha|} \int_{\mathbb{R}^d} f(\xi) (D^\alpha v)(\xi) d\xi \quad \text{for all} \quad v \in C_0^\infty(\mathbb{R}^d), \qquad (25)$$

where $C_0^\infty(\mathbb{R}^d)$ denotes the space of infinitely differentiable functions with compact support in $\mathbb{R}^d$ and $D^\alpha v$ is a classical derivative. We will use the same symbol for the weak derivative as for the classical one, i.e., we set $D^\alpha f = g$, since classical derivatives are also weak derivatives. The latter follows because classical derivatives satisfy (25) which is just the multivariate integration by parts formula in the classical sense. In accordance with the notation (5) we consider in the next section the *mixed Sobolev space* of functions having mixed first order weak derivatives

$$\mathcal{W}_{2,\rho,\mathrm{mix}}^{(1,\ldots,1)}(\mathbb{R}^d) = \left\{ f \in \mathcal{L}_{2,\rho}(\mathbb{R}^d) : D^\alpha f \in \mathcal{L}_{2,\rho}(\mathbb{R}^d) \ \text{ if } \ \alpha_i \leq 1, i \in \mathfrak{D} \right\}. \qquad (26)$$

In [53] such spaces are called Sobolev spaces with dominating mixed smoothness.

5 ANOVA decomposition of linear two-stage integrands

We assume(A1)–(A4). According to Proposition 1 linear two-stage integrands may be written in the form

$$f_x(\xi) := f(x, \xi) = \langle A^j(x)\xi, \xi \rangle + \langle B^j(x), \xi \rangle + c^j(x) \quad (\xi \in \varXi_j(x),\ x \in X), \ (27)$$

where $A^j(\cdot)$, $B^j(\cdot)$ and $c^j(\cdot)$ are affine mappings to the linear space of (d, d)-matrices, to $\mathbb{R}^d$ and to $\mathbb{R}$, respectively, $\varXi_j(x)$ is a d-dimensional polyhedral subset (12) of $\mathbb{R}^d$

for every $j = 1, \ldots, \ell$ and $x \in X$. When $x \in X$ is given and fixed, we will also write f instead of f_x.

The integrands do not belong to the mixed Sobolev spaces (26) and are not of bounded variation in the sense of Hardy and Krause on any hyperrectangle (see [38, Proposition 17]) in general. For example, the function $f(\xi_1, \xi_2) = \max\{\xi_1, \xi_2\}$ of two variables does not have a mixed derivative $\partial^2 f/\partial\xi_1\partial\xi_2$ in the Sobolev sense.

We intend to show that all but one ANOVA terms of f are smoother than the function f itself. Since the function f is piecewise linear-quadratic in the sense of [42, Section 10.E], it is locally Lipschitz continuous and, hence, differentiable almost everywhere due to Rademacher's theorem (see, for example, [9, Section 3.1.2]). Since the ANOVA terms are given in terms of projections (see (19)), we study the smoothness of first order projections.

Let $k \in \mathfrak{D}$, fix $x \in X$. For $\xi \in \mathbb{R}^d$ we set

$$\begin{aligned} \xi^k &= (\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_d) \\ \xi^k_s &= (\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \quad (s \in \mathbb{R}). \end{aligned}$$

We know that

$$\xi^k_s \in \bigcup_{j=1}^{\ell} \Xi_j(x) \tag{28}$$

holds for every $s \in \mathbb{R}$. By definition the kth projection is of the form

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

Due to (28) the one-dimensional affine subspace $\{\xi^k_s : s \in \mathbb{R}\}$ intersects finitely many polyhedral sets $\Xi_j(x)$. Hence, there exist $p = p(k) \in \mathbb{N} \cup \{0\}$, $s_i = s^k_i \in \mathbb{R}$, $i = 1, \ldots, p$, and $j_i = j^k_i \in \{1, \ldots, \ell\}$, $i = 1, \ldots, p+1$, such that $s_i < s_{i+1}$ and

$$\begin{aligned} &\xi^k_s \in \Xi_{j_1}(x) && \forall s \in (-\infty, s_1] \\ &\xi^k_s \in \Xi_{j_i}(x) && \forall s \in [s_{i-1}, s_i] \quad (i = 2, \ldots, p) \\ &\xi^k_s \in \Xi_{j_{p+1}}(x) && \forall s \in [s_p, +\infty). \end{aligned}$$

Clearly, the real numbers s_i depend on k, x and ξ^k, but later we write only s_i or $s_i(\xi^k)$ to emphasize their dependence on ξ^k which is of particular importance here. By setting $s_0 := -\infty$, $s_{p+1} := \infty$, we obtain the following representation of $P_k f$.

$$\begin{aligned} (P_k f)(\xi^k) &= \sum_{i=1}^{p+1} \int_{s_{i-1}}^{s_i} (\langle A^{j_i}(x)\xi^k_s, \xi^k_s\rangle + \langle b^{j_i}(x), \xi^k_s\rangle + c^{j_i}(x))\rho_k(s)ds \\ &= \sum_{i=1}^{p+1} \Big((\langle A^{j_i}(x)\xi^k_0, \xi^k_0\rangle + \langle b^{j_i}(x), \xi^k_0\rangle + c^{j_i}(x)) \int_{s_{i-1}}^{s_i} \rho_k(s)ds \end{aligned}$$

$$+\sum_{\substack{l=1\\l\neq k}}^{d}(a_{lk}^{j_i}(x)\xi_l + a_{kl}^{j_i}(x)\xi_l + b_k^{j_i}(x))\int_{s_{i-1}}^{s_i} s\rho_k(s)ds$$

$$+ a_{kk}^{j_i}(x)\int_{s_{i-1}}^{s_i} s^2\rho_k(s)ds\Big) \tag{29}$$

$$= \sum_{i=1}^{p+1}\sum_{j=0}^{2} p_{ij}(\xi^k; x)\int_{s_{i-1}}^{s_i} s^j\rho_k(s)ds \tag{30}$$

$$= \sum_{i=1}^{p+1}\sum_{j=0}^{2} p_{ij}(\xi^k; x)[\varphi_{kj}(s_i(\xi^k)) - \varphi_{kj}(s_{i-1}(\xi^k))] \tag{31}$$

where $a_{lk}^j(x)$ and $a_{kl}^j(x), l = 1, \ldots, d$, are the elements of the kth column and kth row of the matrix $A^j(x)$, respectively, $b_k^j(x)$ the kth component of $b^j(x)$ and $p_{ij}(\cdot; x)$ are $(d-1)$-variate polynomials in ξ^k of degree $2-j$ with coefficients depending on the first-stage variable x. The function φ_{k0} is the kth marginal distribution function and φ_{kj}, $j = 1, 2$, are the corresponding first and second order moment functions, i.e.,

$$\varphi_{kj}(t) = \int_{-\infty}^{t} s^j\rho_k(s)ds \quad (j = 0, 1, 2). \tag{32}$$

According to Proposition 1 the points $\xi_{s_i}^k$, $i = 1, \ldots, p$, belong to the boundary of $\Xi_{j_i}(x)$, thus, to a $(d-1)$-dimensional affine subspace H_i of $\mathbb{R}^d$. Hence, there exist $g_i \in \mathbb{R}^d$ and $a_i \in \mathbb{R}$ such that

$$\langle g_i, \xi_{s_i}^k\rangle = \sum_{\substack{l=1\\l\neq k}}^{d} g_{il}\xi_l + g_{ik}s_i = a_i \quad (i = 1, \ldots, p). \tag{33}$$

Note that $g_{ik} \neq 0$, since the condition $g_{ik} = 0$ is equivalent to the orthogonality of g_i to the kth coordinate axis and, thus, to the fact that the affine subspace H_i is parallel to the kth coordinate axis which is excluded according to (A4). Hence, $s_i = s_i(\xi^k)$ is an affine function of ξ^k and the projection $P_k f$ represents a sum of products of functions (depending on ξ^k) that are continuously differentiable if the polyhedra $\Xi_{j_i}(x)$, $i = 1, \ldots, p+1$, do not change in some neighborhood of ξ^k.

In order to study the behavior of $P_k f$ also at points ξ^k where the polyhedra $\Xi_{j_i}(x)$, $i = 1, \ldots, p+1$, do change in any neighborhood of ξ^k, we introduce some additional notation. Let $\bar{\xi}^k \in \mathbb{R}^{d-1}$, $\mathbb{B}_\epsilon(\bar{\xi}^k)$ denote the open ball around $\bar{\xi}^k$ with radius $\epsilon > 0$ and

$$\mathcal{P}_\epsilon(\bar{\xi}^k) := \Big\{\Xi_j(x) : \xi_s^k \in \Xi_j(x) \text{ for some } s \in \mathbb{R},\ \xi^k \in \mathbb{B}_\epsilon(\bar{\xi}^k)\Big\} \tag{34}$$

$$\mathcal{P}(\xi^k) := \Big\{\Xi_j(x) : \xi_s^k \in \Xi_j(x) \text{ for some } s \in \mathbb{R}\Big\} \tag{35}$$

denote sets of polyhedra $\Xi_j(x)$ that are met by the affine one-dimensional space $\{\xi_s^k : s \in \mathbb{R}\}$. Because any affine one-dimensional space $\{\xi_s^k : s \in \mathbb{R}\}$ for some $\xi^k \in \mathbb{B}_\epsilon(\bar{\xi}^k)$ is a parallel translation of $\{\bar{\xi}_s^k : s \in \mathbb{R}\}$, ϵ_0 can be chosen even small enough such that $\mathcal{P}(\xi^k) \subseteq \mathcal{P}(\bar{\xi}^k)$ for every $\xi^k \in \mathbb{B}_{\epsilon_0}(\bar{\xi}^k)$. Therefore we have

$$\mathcal{P}(\bar{\xi}^k) = \mathcal{P}_{\epsilon_0}(\bar{\xi}^k). \tag{36}$$

Since the polyhedra $\Xi_j(x)$ are convex, the sets $\{\xi_s^k : \xi_s^k \in \Xi_j(x),\ s \in \mathbb{R}\}$ are convex, too, and, hence, represent either an interval or a single point if $\Xi_j(x)$ belongs to $\mathcal{P}(\xi^k)$. The latter is only possible if the affine one-dimensional space meets a vertex or an edge (i.e., faces of dimension zero or one) of $\Xi_j(x)$. The subset of $\mathbb{R}^d$ that contains all vertices and edges of all such polyhedra $\Xi_j(x)$ has Lebesgue measure zero in $\mathbb{R}^d$. If the set $\{\xi_s^k : \xi_s^k \in \Xi_j(x),\ s \in \mathbb{R}\}$ is an interval denoted by $I_j(\xi^k)$, the set $\{\xi_s^k : s \in I_j^\circ(\xi^k)\}$, where $I_j^\circ(\xi^k)$ denotes the interior of $I_j(\xi^k)$, belongs to the interior of $\Xi_j(x)$. Otherwise, the interval $I_j(\xi^k)$ belongs to a facet of $\Xi_j(x)$ which in turn is parallel to the canonical basis element e_k contradicting (A4).
Now, we are ready to prove our first result on smoothness properties of $P_k f$.

Theorem 1 *Let (A1)–(A4) be satisfied, $k \in \mathfrak{D}$, $x \in X$ and we consider an integrand f of the form (27). Then the kth projection $P_k f$ is continuously differentiable on $\mathbb{R}^d$. $P_k f$ is second order continuously partially differentiable almost everywhere on $\mathbb{R}^d$ if ρ_k is continuously differentiable.*

Proof There are two possible cases for any point $\bar{\xi}^k \in \mathbb{R}^{d-1}$:

(i) There exists $\epsilon_0 > 0$ such that $\mathcal{P}(\xi^k) = \mathcal{P}(\bar{\xi}^k)$ for all $\xi^k \in \mathbb{B}_{\epsilon_0}(\bar{\xi}^k)$.
(ii) For each $\epsilon > 0$ there exists $\xi^k \in \mathbb{B}_\epsilon(\bar{\xi}^k)$ such that $\mathcal{P}(\xi^k) \subsetneq \mathcal{P}(\bar{\xi}^k)$.

The case (i) corresponds to the consideration before stating the theorem where we arrived at (see also (31))

$$(P_k f)(\bar{\xi}^k) = \sum_{i=1}^{p(\bar{\xi}^k)+1} \sum_{j=0}^{2} S_{ij}(\bar{\xi}^k; x),$$

where $S_{ij}(\bar{\xi}^k; x) := p_{ij}(\bar{\xi}^k; x)[\varphi_{kj}(s_i(\bar{\xi}^k)) - \varphi_{kj}(s_{i-1}(\bar{\xi}^k))]$ for $j = 0, 1, 2$, $i = 1, \ldots, p+1$, and the functions s_i, $i = 1, \ldots, p$, and φ_{kj}, $j = 0, 1, 2$, are defined by (32) and (33), respectively. Furthermore, $s_0 = -\infty$, $s_{p+1} = +\infty$ and the functions $p_{ij}(\cdot; x)$ are $(d-1)$-variate polynomials of degree $2-j$, $j = 0, 1, 2$, $i = 1, \ldots, p+1$.

Now, let $l \in \mathfrak{D}$, $l \neq k$. Then all partial derivatives $\frac{\partial S_{ij}}{\partial \xi_l}$ and, hence, the first partial derivative of $P_k f$ with respect to ξ_l exists at $\bar{\xi}^k$ and it holds

$$\frac{\partial P_k f}{\partial \xi_l}(\bar{\xi}^k) = \sum_{i=1}^{p(\bar{\xi}^k)+1} \sum_{j=0}^{2} \frac{\partial S_{ij}}{\partial \xi_l}(\bar{\xi}^k; x) \tag{37}$$

$$\frac{\partial S_{ij}}{\partial \xi_l}(\bar{\xi}^k; x) = \frac{\partial p_{ij}}{\partial \xi_l}(\bar{\xi}^k; x)\left[\varphi_{kj}(s_i(\bar{\xi}^k)) - \varphi_{kj}(s_{i-1}(\bar{\xi}^k))\right] \tag{38}$$

$$+\frac{p_{ij}(\bar{\xi}^k;x)}{g_{ik}}\left[s_i^j(\bar{\xi}^k)\rho_k(s_i(\bar{\xi}^k))g_{il}-s_{i-1}^j(\bar{\xi}^k)\rho_k(s_{i-1}(\bar{\xi}^k))g_{i-1,l}\right], \tag{39}$$

where the first and second term in (39) disappear for $i=p+1$ and $i=1$, respectively. We note that the partial derivative $\frac{\partial p_{ij}}{\partial \xi_l}(\cdot;x)$ is a $(d-1)$-variate polynomial of degree $1-j$ for $j=0,1$ and vanishes for $j=2$. The term in (39) is equal to the sum of polynomials of degree 2 multiplied by the density ρ_k evaluated at $s_i(\bar{\xi}^k)$ or $s_{i-1}(\bar{\xi}^k)$. Since the kth marginal density ρ_k is continuous, the partial derivative is also continuous. The term in (38) is continuously differentiable once again and the term in (39) is continuously differentiable if ρ_k is continuously differentiable. Hence, $P_k f$ is second order partially differentiable at points $\bar{\xi}^k$ which satisfy (i).

In case (ii) we use the identity (36) and consider all polyhedra belonging to $\mathcal{P}(\bar{\xi}^k)$. Let $\Xi_{j_i}(x), i=1,\ldots,p+1$, be all such polyhedra. Furthermore, let $s_i, i=1,\ldots,p$, be nondecreasing and defined by

$$\bar{\xi}^k_{s_i}\in\Xi_{j_i}(x)\cap\Xi_{j_{i+1}}(x)\quad(i=1,\ldots,p)$$

and we set $s_0=-\infty$ and $s_{p+1}=+\infty$. We allow explicitly that $s_i=s_{i+1}$ holds for some $i\in\{1,\ldots,p-1\}$. Then we obtain

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

where $p=p(\bar{\xi}^k)$ and $s_i=s_i(\bar{\xi}^k)$, $i=1,\ldots,p$, are given by (33). Now, let $\xi^k\in\mathbb{B}_\epsilon(\bar{\xi}^k)$ for some $\epsilon>0$. Due to (36) the kth projection may be represented by a subset of the set $\mathcal{P}(\bar{\xi}^k)$. Of course, $\Xi_{j_1}(x)$ and $\Xi_{j_{p+1}}(x)$ and all polyhedra $\Xi_{j_i}(x)$ such that $s_i(\bar{\xi}^k)<s_{i+1}(\bar{\xi}^k)$ appear also in the representation of $P_k f(\xi^k)$. Those polyhedra $\Xi_{j_i}(x)$ with $s_i(\bar{\xi}^k)=s_{i+1}(\bar{\xi}^k)$ may either disappear or appear with $s_i(\xi^k)<s_{i+1}(\xi^k)$. If they disappear we set $s_i(\xi^k)=s_{i+1}(\xi^k)$ and include them formally into the representation of $P_k f(\xi^k)$ which is of the form

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

In a small ball around ξ^k this representation doesn't change. Hence, $P_k f$ is differentiable also in case (ii) and the partial derivative is of the form

$$\frac{\partial P_k f}{\partial \xi_l}(\xi^k)=\sum_{i=1}^{p+1}\sum_{j=0}^{1}\frac{\partial S_{ij}}{\partial \xi_l}(\xi^k;x), \tag{40}$$

where the partial derivative of S_{ij} at ξ^k is of the same form as in (38) and (39) and, thus, as in case (i). This means that the partial derivative of $P_k f$ is also continuous at ξ^k. The integrals with $s_{i-1}(\xi^k) = s_i(\xi^k)$ are again formally included into (40). The second order partial derivative at ξ^k does not exist in general, since left and right one-sided second order partial derivatives differ in general since different summands from (37) may appear.

Hence, $P_k f$ is continuously differentiable on $\mathbb{R}^d$, but the mixed second order partial derivatives exist only almost everywhere on $\mathbb{R}^d$, where the relevant set of Lebesgue measure zero is just contained in the union of all faces of all polyhedra $\Xi_j(x)$, $j = 1, \ldots, \ell$. □

The following example shows that the geometric condition (A4) imposed in the previous result is necessary for Theorem 1 to hold.

Example 1 We assume that the stochasticity only appears at right-hand sides and that the dual feasible set $\mathcal{D}$ is given as the line segment

$$\mathcal{D} = \mathrm{conv}\{-v, v\} = \{-\lambda v + (1-\lambda)v : \lambda \in [0,1]\} = \{(1-2\lambda)v : \lambda \in [0,1]\}$$

for some $v \in \mathbb{R}^3$, $v \neq 0$. With the two vertices v and $-v$ of $\mathcal{D}$ the optimal value function of the second-stage problem is

$$\Phi(t) = \max\{\langle -v, t\rangle, \langle v, t\rangle\} = |\langle v, t\rangle| \quad (t \in \mathbb{R}^3).$$

With the right-hand side $h(\xi) = \xi$ and the choice $x = 0$ without loss of generality, we obtain the integrand

$$f(\xi) = f(0, \xi) = |\langle v, \xi\rangle| \quad (\xi \in \mathbb{R}^3).$$

Let $k \in \{1, 2, 3\}$. If the kth component of v vanishes, the face of the normal cone to $\mathcal{D}$ at v parallels the kth coordinate axis, i.e., the geometric condition (A4) is not satisfied. The kth projection $P_k f$ of f is then of the form

$$P_k f(\xi^k) = \left|\langle v^k, \xi^k\rangle\right| \quad (\xi^k \in \mathbb{R}^2)$$

and, hence, *not differentiable* on $\mathbb{R}^3$. For $v_k > 0$ we obtain

$$\begin{aligned} P_k f(\xi^k) &= \int_{-\infty}^{+\infty} \left|\langle v^k, \xi^k\rangle + v_k s\right| \rho_k(s)ds \\ &= v_k\big(\varphi_{k1}(+\infty) - 2\varphi_{k1}(s_1(\xi^k))\big) + \langle v^k, \xi^k\rangle\big(1 - 2\varphi_{k0}(s_1(\xi^k))\big) \end{aligned}$$

where $s_1(\xi^k) = -v_k^{-1}\langle v^k, \xi^k\rangle$ and we used the notation (32). Hence, $P_k f$ is twice continuously differentiable with

$$\frac{\partial P_k f(\xi^k)}{\partial \xi_j} = v_j\big(1 - 2\varphi_{k0}(s_1(\xi^k))\big) \quad \text{and} \quad \frac{\partial^2 P_k f(\xi^k)}{\partial \xi_j \partial \xi_l} = 2\frac{v_j v_l}{v_k}\rho_k(s_1(\xi^k))$$

for $j, l \in \{1, 2, 3\} \setminus \{k\}$ and each $\xi^k \in \mathbb{R}^2$. This implies that in this particular case all ANOVA terms f_u with $|u| \leq 2$ possess even continuous classical mixed derivatives.

Example 4 in [15] shows that $P_k f$ is not second order continuously differentiable on the entire $\mathbb{R}^d$ in general. The geometric condition is further discussed in Sect. 6. Theorem 1 extends to more general projections P_u.

Corollary 1 *Let (A1)–(A4) be satisfied, $\emptyset \neq u \subseteq \mathfrak{D}$, $x \in X$ and we consider an integrand f of the form (27). Then the projection $P_u f$ is continuously differentiable on $\mathbb{R}^d$. $P_u f$ is mixed second order continuously partially differentiable almost everywhere on $\mathbb{R}^d$ if ρ_k is continuously differentiable for some $k \in u$.*

Proof If $|u| = 1$ the result follows from Theorem 1. For $u = \{k, r\}$ with $k, r \in \mathfrak{D}$, $k \neq r$, and we obtain from the Leibniz theorem [14, Theorem 1] for $l \notin u$

$$D_l P_u f(\xi^u) = \frac{\partial}{\partial \xi_l} P_u f(\xi^u) = P_r \frac{\partial}{\partial \xi_l} P_k f(\xi^u)$$

and from the proof of Theorem 1

$$D_l P_u f(\xi^u) = \sum_{i=1}^{p+1} \sum_{j=0}^{2} \int_{\mathbb{R}} \frac{\partial S_{ij}}{\partial \xi_l}(\xi^k; x) \rho_r(\xi_r) d\xi_r.$$

A description of the partial derivative of S_{ij} is given by (38) and (39).

If u contains more than two elements, the integral on the right-hand side becomes a multiple integral. In all cases, however, such an integral is a function of the remaining variables ξ_j, $j \in \mathfrak{D} \setminus u$, whose continuity and differentiability properties correspond at least to those of φ_{kj} and ρ_k. This follows using Lebesgue's dominated convergence theorem as φ_{kj} and all densities ρ_j, $j \in u$, and their derivatives are bounded on $\mathbb{R}$. □

The following is the main result of this section.

Theorem 2 *Assume (A1)–(A4) and that all marginal densities ρ_i, $i = 1, \ldots, d$, are continuously differentiable. Then all ANOVA terms of f except the one of highest order are first order continuously differentiable on $\mathbb{R}^d$ and all mixed second order partial derivatives exist and are continuous except on a set of Lebesgue measure zero, and quadratically integrable with respect to the density ρ. In particular, the first and second order ANOVA terms of f belong to the tensor product Sobolev space $\mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}(\mathbb{R}^d)$.*

Proof According to (19) the ANOVA terms of f are defined by

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

for all nonempty subsets u of $\mathfrak{D}$. Hence, all ANOVA terms of f for $u \neq \mathfrak{D}$ are continuously differentiable on $\mathbb{R}^d$. Second order partial derivatives of those ANOVA terms exist and are continuous almost everywhere in $\mathbb{R}^d$. The non-vanishing first order partial derivatives of the second order ANOVA terms are of the form

$$
\begin{aligned}
D_l f_{\{l,r\}}(\xi_l, \xi_r) &= D_l P_{\mathfrak{D}\setminus\{l,r\}} f(\xi_l, \xi_r) - D_l P_{\mathfrak{D}\setminus\{l\}} f(\xi_l) \\
&= \sum_{i=1}^{p+1} \sum_{j=0}^{1} \int_{\mathbb{R}^{d-2}} \frac{\partial S_{ij}}{\partial \xi_l}(\xi^k; x) \prod_{\substack{r\in\mathfrak{D} \\ r\neq k}} \rho_r(\xi_r) d\xi^{-\{r,l\}} - D_l P_{\mathfrak{D}\setminus\{l\}} f(\xi_l)
\end{aligned}
$$

for all $l, r \in \mathfrak{D}$ and some $k \in \mathfrak{D}$. Due to the structure of the partial derivative $\frac{\partial S_{ij}}{\partial \xi_l}(\xi^k; x)$ (see (38) and (39)) and the local Lipschitz continuity of φ_{kj}, $j = 0, 1$, and of ρ_k, it is a locally Lipschitz continuous function of ξ^k. Hence, the functions $D_l f_{\{l,r\}}$ and $D_r f_{\{l,r\}}$ are locally Lipschitz continuous with respect to each of the two variables ξ_l and ξ_r independently when the other variable is fixed almost everywhere. Hence, $D_l f_{\{l,r\}}$ and $D_r f_{\{l,r\}}$ are partially differentiable with respect to ξ_r and ξ_l, respectively, in the sense of Sobolev (see, for example, [9, Section 4.2.3]). Furthermore, the second order mixed partial derivatives are quadratically integrable with respect to ρ due to (A2). $\square$

Remark 1 The second order ANOVA approximation of f, i.e.,

$$
f^{(2)} := \sum_{\substack{|u|\leq 2 \\ u\subseteq\mathfrak{D}}} f_u \tag{41}
$$

belongs to the mixed Sobolev space $\mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}(\mathbb{R}^d)$ due to Theorem 2. Since the estimate (23) implies

$$
\|f - f^{(2)}\|_{2,\rho}^2 = \sum_{\substack{|u|=3 \\ u\subseteq\mathfrak{D}}}^{d} \|f_u\|_{2,\rho}^2 \leq \varepsilon\sigma^2(f)
$$

if the effective superposition dimension of f satisfies $d_S(\varepsilon) \leq 2$, the function f is representable as sum of an element of the mixed Sobolev space and of a function which is small in L_2. Based on this observation we derive in Remark 2 an error estimate for randomly shifted lattice rules showing that essentially the convergence rate (6) is valid for optimal values of two-stage stochastic programming problems, too. We note that, in general, the property $d_S(\varepsilon) \leq 2$ is known as a good sign for the favorable behavior of QMC methods compared to MC.

Remark 2 We assume that all marginal densities ρ_k, $k = 1, \ldots, d$, are continuously differentiable and positive. Then the corresponding marginal distribution functions

$$
\varphi_k(t) = \int_{-\infty}^{t} \rho_k(s) ds \quad (t \in \mathbb{R},\ k = 1, \ldots, d)
$$

are invertible on $(0, 1)$ and the mapping

$$\varphi^{-1}(t) = (\varphi_1^{-1}(t_1), \ldots, \varphi_d^{-1}(t_d))^\top \quad (t \in (0,1)^d)$$

is well defined and twice continuously differentiable. We consider the function $g : [0,1]^d \to \mathbb{R}$ defined by the *transformation*

$$g(t) = f \circ \varphi^{-1}(t) \quad \text{for} \quad t \in (0,1)^d$$

and $g(t) = 0$ else. The well known difficulty of this transformation is that g is either unbounded near the boundary of the cube $[0,1]^d$ or has very large derivatives near the boundary. In [24] the authors developed a theory that overcomes this difficulty and derives the same rate of convergence as (6) also for unbounded integrands and a number of one-dimensional density functions (including the normal density).

Here, we assume for simplicity that the ANOVA terms g_u, $|u| = 1, 2$, of the transformed function g belong to the tensor product Sobolev space (5) if the ANOVA terms f_u, $|u| = 1, 2$, of the function f belong to the mixed Sobolev space (26). This is true, for example, if the support of P is compact. Theorem 2 contains conditions implying that the ANOVA terms f_u, $|u| = 1, 2$, of two-stage integrands f belong to the mixed Sobolev space (26) if (A1)–(A4) are satisfied. Notice the following relations of g_u and f_u

$$f_u(\xi^u) = g_u \circ \varphi_u(\xi^u) \text{ for } \xi^u \in \mathbb{R}^{|u|}, \;\; g_u(t^u) = (f_u \circ \varphi_u^{-1})(t^u) \text{ for } t^u \in (0,1)^{|u|},$$

where

$$\varphi_u := (\varphi_{j_1}, \ldots, \varphi_{j_{|u|}}), \;\; \varphi_u^{-1} := (\varphi_{j_1}^{-1}, \ldots, \varphi_{j_{|u|}}^{-1}),$$
$$(j_k \in u, \; 1 \le k \le |u|, \; j_k < j_l, \; k < l).$$

Then the QMC quadrature error may be estimated as follows:

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

where $\mathbf{t}^j = (t_1^j, \ldots, t_d^j)$, $t_i^j = \varphi_i(\xi_i^j) \in (0,1)$, $i = 1, \ldots, d$, $j = 1, \ldots, n$, are the QMC points. If the $\mathbf{t}^j$, $j = 1, \ldots, n$, are randomly shifted lattice points, n is prime and $\delta \in (0, \frac{1}{2}]$, we may continue

$$\left(\mathbb{E} \left| \int_{[0,1]^d} g(t)dt - n^{-1} \sum_{j=1}^n g(\mathbf{t}^j) \right|^2 \right)^{\frac{1}{2}} \le C(\delta) n^{-1+\delta}$$

$$+\sum_{|u|=3}^{d}\left(\mathbb{E}\Big|\int_{[0,1]^d} g_u(t)dt - n^{-1}\sum_{j=1}^{n} g_u(\mathbf{t}^j)\Big|^2\right)^{\frac{1}{2}}.$$

The latter sum can be further estimated by

$$\sum_{|u|=3}^{d}\left(\|g_u\|_2^2 + n^{-1}\sum_{j=1}^{n}\mathbb{E}(g_u^2(\mathbf{t}^j))\right)^{\frac{1}{2}} = \sum_{|u|=3}^{d}\left(\|f_u\|_{2,\rho}^2 + n^{-1}\sum_{j=1}^{n}\mathbb{E}(f_u^2(\xi^j))\right)^{\frac{1}{2}}. \tag{42}$$

Since (23) implies $\sum_{|u|=3}^{d}\|f_u\|_{L_2}^2 \leq \varepsilon\sigma^2(f)$ if $d_S(\varepsilon) \leq 2$ and the second term on the right-hand side of (42) represents a QMC approximation of the first term, we may conclude that the right-hand side in (42) is of the form $O(\sqrt{\varepsilon})$. Hence, we obtain

$$\left(\mathbb{E}\Big|\int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi - n^{-1}\sum_{j=1}^{n} f(\xi^j)\Big|^2\right)^{\frac{1}{2}} \leq C(\delta)n^{-1+\delta} + O(\sqrt{\varepsilon}) \tag{43}$$

if the condition $d_S(\varepsilon) \leq 2$ is satisfied. The latter may eventually be achieved by applying dimension reduction techniques (see Sect. 7).

Finally, we note that the constants involved in the estimate (43) may be chosen to be uniform with respect to $x \in X$. Hence, using the perturbation estimate (15) for the optimal values in Sect. 3 we obtain

$$(\mathbb{E}|v(P) - v(P_n)|^2)^{\frac{1}{2}} \leq \hat{C}(\delta)n^{-1+\delta} + O(\sqrt{\varepsilon}), \tag{44}$$

if $d_S(\varepsilon) \leq 2$. Hence, the estimate (43) carries over to optimal values. A similar result can also be obtained for solution sets by relying on (16). Here, P_n is the discrete probability measure representing the randomized QMC method, i.e., $P_n = n^{-1}\sum_{j=1}^{n}\delta_{\xi^j}$, where δ_ξ denotes the Dirac measure placing unit mass at ξ.

6 Generic smoothness in the normal case

Let ξ be a d-dimensional normal random vector with mean μ and nonsingular covariance matrix Σ. Then there exists an orthogonal matrix Q such that $Q\,\Sigma\,Q^\top$ is a diagonal matrix. Then the d-dimensional random vector η given by the transformation

$$\xi = Q\eta + \mu \quad \text{or} \quad \eta = Q^\top(\xi - \mu) \tag{45}$$

is normal with zero mean and diagonal covariance matrix, i.e., η has independent components. For fixed $x \in X$, let $\Xi_j(x)$, $j = 1, \ldots, \ell$, denote the polyhedral decomposition (12) of $\mathbb{R}^d$. The transformed function $\hat{f}(x, \eta) = f(x, Q\eta + \mu)$ is defined on the polyhedral sets $Q^\top\Xi_j(x) - Q^\top\mu$, $j = 1, \ldots, \ell$, and still linear-quadratic in η on each such set. The intersections of two adjacent polyhedral sets $\Xi_j(x)$ are subsets of $(d-1)$-dimensional affine subspaces $H_j(x)$. The orthogonal matrix $Q^\top$ causes a rotation of $H_j(x)$. However, there are only finitely many of such subspaces $H_j(x)$ and, thus, only finitely many orthogonal matrices Q causing rotations modulo 2π such that the subspace $Q^\top H_j(x)$ is parallel to some coordinate axis. Hence, altogether, the set of all orthogonal matrices Q such that (A4) for the polyhedral sets $Q^\top \Xi_j(x) - Q^\top \mu$ is *not* satisfied, is *countable*. When equipping the metric space of all orthogonal $d \times d$-matrices with the standard norm topology, the set of all orthogonal matrices Q that satisfy (A4), is a *residual set*, i.e., the countable intersection of open dense subsets. A property for elements of a topological space is called *generic* or is said to hold *almost everywhere* if it is valid in a residual set. By referring to Theorem 2 this proves

Corollary 2 *Let (A1) and (A2) be satisfied, $x \in X$, $f = f(x, \cdot)$ be given by (27) and ξ be normally distributed with nonsingular covariance matrix. The transformation (45) of ξ implies the fact that the second order ANOVA approximation $f^{(2)}$ of f (given by (41)) belongs to $\mathcal{W}^{(1,\ldots,1)}_{2,\rho,\mathrm{mix}}(\mathbb{R}^d)$ is a generic property.*

7 Dimension reduction

It is known (see [37,55]) that Quasi-Monte Carlo methods for high-dimensional numerical integration may be more efficient if the integrands f have low effective superposition dimension. For integrals appearing in two-stage stochastic programming one needs in addition that the effective superposition dimension satisfies $d_S(\varepsilon) \le 2$ for some sufficiently small $\varepsilon > 0$ (see Sect. 5). Hence, one is usually interested in determining and reducing the effective dimension. This topic is discussed in a number of papers, e.g.,in [7,28,37,50,55,57]. Here, we concentrate on the *normal case*.

Several *dimension reduction techniques* exploit the fact that a normal random vector ξ with mean μ and non-singular covariance matrix Σ can be transformed by $\xi = B\eta + \mu$ and *any* matrix B satisfying $\Sigma = B\,B^\top$ into a standard normal random vector η with independent components. The choice of B may change the QMC error and the effective dimension of the integrand f_x (cf. [40]). As observed in [40,57], however, there is no *consistent* dimension reduction effect for any such matrix B. This means that a specific choice of the matrix B may result in a dimension reduction for one integrand, but eventually not for another one. For example, the *standard* (lower triangular) Cholesky matrix L_C performing the factorization $\Sigma = L_C L_C^\top$ seems to assign the same importance to every variable in option pricing models (cf. [55]) and, hence, is not suitable for reducing the effective dimension of such models. The same effect is observed in our numerical experiments for production planning models (see Sect. 8).

A universal principle for dimension reduction in the normal case is *principal component analysis* (PCA). It is universal in the sense that it does not depend on the structure of the underlying integrand f. The basic idea of PCA is to determine the best mean square approximation of the form $\sum_{i=1}^d v_i z_i$ to a d-dimensional normal random vector ξ, where $v_i \in \mathbb{R}^d$, $i = 1, \ldots, d$, and $(z_1, \ldots, z_d)$ is normal with zero mean and the $\mathbb{R}^d$ identity matrix as covariance matrix. The solution is $v_i = \sqrt{\lambda_i}u_i$ and $z_i = (\sqrt{\lambda_i})^{-1} u_i^\top \xi$, where $\lambda_1 \ge \cdots \ge \lambda_d > 0$ are the eigenvalues of Σ in decreasing order and u_i, $i = 1, \ldots, d$, the corresponding orthonormal eigenvectors (see [57]). Hence, PCA consists in using the factorization

$$\Sigma = U_P U_P^\top \quad \text{or} \quad \Sigma = (u_1, \ldots, u_d)\text{diag}(\lambda_1, \ldots, \lambda_d)(u_1, \ldots, u_d)^\top,$$

where $U_P = (\sqrt{\lambda_1}u_1, \ldots, \sqrt{\lambda_d}u_d)$. Several authors report an enormous reduction of the effective truncation dimension in financial models if PCA is used (see, for example, [55,56]). We observed the same effect in our numerical experiments (see Sect. 8). However, the reduction effect certainly depends on the eigenvalues of Σ. If the ratio $\frac{\lambda_1}{\lambda_d}$ is close to 1, the performance of PCA gets worse. Nevertheless we recommend to use first PCA and to resort to other ideas only after its failure.

8 Numerical results

We consider a stochastic production planning problem which consists in minimizing the expected costs of a company during a certain time horizon. The model contains stochastic demands ξ_δ and prices ξ_c as components of

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

The company aims to satisfy stochastic demands $\xi_{\delta,t}$ in a time horizon $\{1, \ldots, T\}$, but its production capacity based on their own N units does eventually not suffice to cover the demand. Hence, it has to buy the necessary extra amounts on m_1 markets or from m_2 other providers at prices $p_{1,j_1,t}(\xi_{c,t}) := \bar{c}_{1,j_1,t} + \xi_{c,t}$ and $p_{2,j_1,t} := \bar{c}_{2,j_2,t}$, $t = 1, \ldots, T$, $1 \le j_1 \le m_1$, $1 \le j_2 \le m_2$, where the vector $(\xi_{c,1}, \ldots, \xi_{c,T})$ represents the stochastic part of the prices $p_{1,j_1,t}$, $1 \le t \le T$, at the markets, and $\bar{c}_{1,j_1,t}, \bar{c}_{2,j_1,t}$ represent contractual fixed prices. At the end, the company aims at minimizing the expected costs. The optimization model is of the form

$$\min_{x \in \mathbb{R}^{NT}} \left\{ \sum_{t=1}^{T} \sum_{i=1}^{N} c_{i,t}\, x_{i,t} + \int_{\mathbb{R}^{2T}} \Phi(x,\xi) P(d\xi) \;:\; x \in X \right\},$$

where the feasible set X is convex polyhedral and given by

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

The constraints in X model capacity limits and ramping constraints, i.e., limits on the rate of capacity changes. The recourse costs Φ are given by

$$\Phi(x,\xi) = \min_{y \in \mathbb{R}^{(m_1+m_2)T}} \left\{ \sum_{t=1}^{T} \left(\sum_{j_1=1}^{m_1} p_{1,j_1,t}(\xi_{c,t})\, y_{j_1,t} + \sum_{j_2=1}^{m_2} p_{2,j_2,t}\, y_{m_1+j_2,t} \right) : y \in Y(x,\xi) \right\}$$

with convex polyhedral feasible set

$$Y(x,\xi) := \left\{ y \in \mathbb{R}^{mT} \;\middle|\; \begin{array}{l} \sum_{i=1}^{N} x_{i,t} + \sum_{j=1}^{m_1+m_2} y_{j,t} \geq \xi_{\delta,t}\,,\ t = 1,\ldots,T \\ w_{1,j_1,t} \leq y_{j_1,t} \leq z_{1,j_1,t}\,,\ j_1 = 1,\ldots,m_1\,,\ t = 1,\ldots,T \\ w_{2,j_2,t} \leq y_{m_1+j_2,t}\,,\ j_2 = 1,\ldots,m_2\,,\ t = 1,\ldots,T \\ |y_{j_1,t} - y_{j_1,t+1}| \leq \rho_{1,j_1,t}\,,\ j_1 = 1,\ldots,m_1\,,\ t = 1,\ldots,T-1 \\ |y_{m_1+j_2,t} - y_{m_1+j_2,t+1}| \leq \rho_{2,j_2,t}\,,\ j_2 = 1,\ldots,m_2\,,\ t = 1,\ldots,T-1 \end{array} \right\},$$

with fixed $c_{i,t}, \bar{c}_{1,j_1,t}, \bar{c}_{2,j_2,t}$ and $a_{i,t}, b_{i,t}, \delta_{i,t}, w_{1,j_1,t}, w_{2,j_2,t}, z_{1,j_1,t}, \rho_{1,j_1,t}$, and $\rho_{2,j_2,t}$. The constraints in $Y(x,\xi)$ describe again capacity limits and ramping constraints. We assume that the stochastic demands and prices $\xi_{\delta,t}, \xi_{c,t}$ follow the condition

$$\begin{pmatrix} \xi_{\delta,t} \\ \xi_{c,t} \end{pmatrix} = \begin{pmatrix} \bar{\xi}_{\delta,t} \\ \bar{\xi}_{c,t} \end{pmatrix} + \begin{pmatrix} E_{1,t} \\ E_{2,t} \end{pmatrix}, \quad \text{for } t = 1,\ldots,T, \text{ and}$$

$$\begin{pmatrix} \bar{\xi}_{\delta,1} \\ \bar{\xi}_{c,1} \end{pmatrix} = B_1 \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \end{pmatrix}, \quad \begin{pmatrix} \bar{\xi}_{\delta,t} \\ \bar{\xi}_{c,t} \end{pmatrix} = A_1 \begin{pmatrix} \bar{\xi}_{\delta,t-1} \\ \bar{\xi}_{c,t-1} \end{pmatrix} + B_1 \begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{pmatrix} + B_2 \begin{pmatrix} \gamma_{1,t-1} \\ \gamma_{2,t-1} \end{pmatrix}$$

for $t = 2,\ldots,T$, where $(E_{1,1},\ldots,E_{1,T})$ and $(E_{2,1},\ldots,E_{2,T})$ are fixed expectation vectors for demands and prices simulating the *trend* or *seasonality*, $A_1, B_1, B_2 \in \mathbb{R}^{2\times 2}$, and stochastic i.i.d. Gaussian noise $\gamma_{1,t}, \gamma_{2,t} \sim$ N(0,1). The resulting stochastic process for demands and prices is therefore a multivariate ARMA (1,1) process. Similar models have been considered for simulating prices and demands in energy industry, see e.g. [8]. Note that since the model includes unbounded demands $\xi_{\delta,1},\ldots,\xi_{\delta,T}$, no upper bounds in the variables $y_{m_1+j_2,t}$, $j_2 = 1,\ldots,m_2$, $t = 1,\ldots,T$ were imposed, allowing the latter to cover arbitrarily large demand values. We select in addition the prices values $\bar{c}_{2,j_2,t}$ significantly higher than the prices values $\bar{c}_{1,j_1,t}$, such that the variables $y_{m_1+j_2,t}$, $j_2 = 1,\ldots,m_2$, $t = 1,\ldots,T$, do not represent always 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

$$A_1 = \begin{pmatrix} 0.29 & 0.44 \\ 0.44 & 0.70 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0.75 & 0.053 \\ 0.053 & 0.43 \end{pmatrix}.$$

For detailed information about modeling with multivariate ARMA processes we refer to [2].

The resulting joint distribution of the process is Gaussian, with dimension $d = 2T$ and covariance matrix Σ. The integration problem is transformed by factorizing the covariance matrix $\Sigma = A\,A^\top$ as usually recommended in Gaussian high-dimensional integration (see [11, Sect. 2.3.3]). We carry out our tests using the Cholesky factorization $A = L_C$ (CH) and the principal component analysis factorization $A = U_P$ (PCA).

A simulated demand-price path $(\xi_{\delta,1},\ldots,\xi_{\delta,T},\xi_{c,1},\ldots,\xi_{c,T})$ can then be obtained by

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

where $Z = (z_1,\ldots,z_{2T}) \sim U([0,1]^{2T})$ (i.e., the probability distribution of Z is uniform distribution on $[0,1]^{2T}$), and $\phi^{-1}(.)$ represents the inverse standard normal distribution function, which can be efficiently and accurately calculated by Moro's algorithm (see [11, 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 [30] as pseudo random number generator. For QMC, we use randomly scrambled Sobol' points with direction numbers given in [19] and randomly shifted lattice rules [22,47]. The used scrambling technique is affine matrix scrambling described in [29] under the name random linear scrambling. 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:

- $N = 8, m_1 = 4, m_2 = 2$.
- For all i, j_1, j_2, t, we select randomly $a_{i,t} \in [0.001, 0.003]$, $b_{i,t} \in [0.3, 0.6]$, $\delta_{i,t} \in [0.3, 0.35]$, $w_{1,j_1,t}, w_{2,j_2,t} \in [0.000001, 0.00002]$, $z_{1,j_1,t} \in [5, 7]$, and $\rho_{1,j_1,t}, \rho_{2,j_2,t} \in [1.0, 1.1]$.
- 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 $\bar{c}_{2,j_2,t} \in [12, 14]$. We fixed $(E_{1,1},\ldots,E_{1,T}) = (6,6,\ldots,6)$, $(E_{2,1},\ldots,E_{2,T}) = (0,0,\ldots,0)$.

The given parameters were chosen as an attempt to avoid trivial solutions of the linear programs.
We performed two different kinds of computational tests. First we studied the convergence behavior and the error of the estimated optimal values of the resulting large linear optimization problem

$$\min_{x\in\mathbb{R}^{NT}} \left\{ \sum_{t=1}^{T}\sum_{i=1}^{N} c_{i,t}\, x_{i,t} + n^{-1}\sum_{j=1}^{n} \Phi(x,\xi^j)P(d\xi) \;:\; x \in X \right\} \tag{46}$$

by increasing the sample sizes n, under Cholesky and PCA factorizations of the covariance matrix Σ. The sample sizes for Monte Carlo (MC, Mersenne Twister) and scrambled Sobol' sequences (SOB) were $n_1 = 128$, $n_2 = 256$ and $n_3 = 512$. For randomly shifted lattice rules (LAT) we have taken $n_1 = 127$, $n_2 = 257$ and $n_3 = 509$. The experiments where repeated 300 times for each sampling method, each sample size and each matrix factorization technique. Figure 1 illustrates the convergence behavior and Table 1 shows the mean and standard deviation of the optimal values for each sampling method, each sample size and each matrix factorization technique over the 300 replications. It is clearly visible that the matrix factorization does not affect significantly the behavior of the Monte Carlo convergence, while the QMC convergence is improved under PCA. More precise estimations of the errors and convergence rates can be found in Figs. 2 and 3.

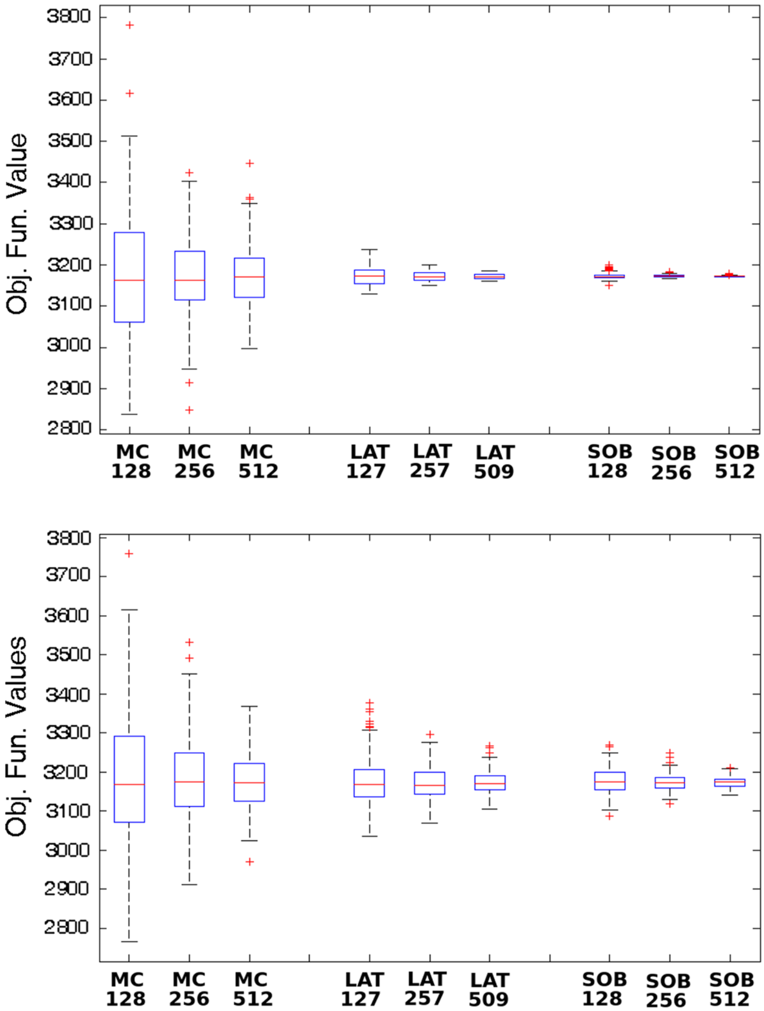

Fig. 1 Shown are the optimal values of (46) obtained with PCA factorization (*top figure*) and Cholesky factorization (*bottom figure*) of the covariance matrix for integration of $\Phi(x, \xi)$ for parameters as stated above. The results for Mersenne Twister MC and scrambled Sobol' QMC (SOB) were obtained with $n_1 = 128$, $n_2 = 256$ and $n_3 = 512$ points, and for randomly shifted lattice rules QMC (LAT) with $n_1 = 127$, $n_2 = 257$ and $n_3 = 509$ lattice points

For the second kind of tests we selected fixed feasible points $x \in X$ and examined the integration errors for the expected recourse

$$\int_{\mathbb{R}^{2T}} \Phi(x, \xi) P(d\xi) \tag{47}$$

by equal weight MC or randomized QMC quadrature rules. For simplicity we chose fixed feasible points $x \in X$ that are also optimal solutions of the tests of the first kind,

Table 1 Mean and standard deviation of the optimal values for different sampling methods, sample sizes and covariance matrix factorizations

	Mean			Standard deviation		
	n_1	n_2	n_3	n_1	n_2	n_3
Cholesky fact.						
MC	3180	3176	3174	151.2	105.2	73.3
LAT	3173	3170	3172	57.4	38.7	26.4
SOB	3176	3173	3173	31.0	19.1	12.9
PCA fact.						
MC	3172	3170	3172	147.0	92.6	73.6
LAT	3172	3172	3172	22.2	11.4	6.2
SOB	3173	3173	3173	6.2	2.9	1.3

which were obtained by solving the resulting large linear program for different costs while keeping the constraint set unchanged.

To determine the errors in the tests of first and second kind we performed 5 runs by changing the set of randomly selected parameters but the qualitative results remained very similar, therefore we only display one of these results in the figures in order to summarize the work done. The results under PCA factorization are summarized in Fig. 2. The sample sizes are chosen as described earlier. The random shifts were generated using the Mersenne Twister. We estimated the relative root mean square errors (RMSE) of the estimated integrals (tests of second kind) and of the optimal values (tests of first kind) by taking 10 runs of every experiment, and repeat the process 30 times for the box plots in the figures. The box-plots of Figs. 1, 2 and 3 show the first quartile as lower bound of the box, the third quartile as upper bound and the median as line between the bounds. Outliers are marked as plus signs and the rest of the results lie between the brackets.

The average of the estimated rates of convergence for the tests of first kind under PCA was approximately -0.9 for randomly shifted lattice rules, and -1.0 for the randomly scrambled Sobol' points, for different price- and bound-parameters as mentioned above. This is clearly superior to the MC convergence rate of -0.5. The effective truncation dimension of $\Phi(x, \xi)$ was tested at 5 different optimal first-stage solutions x obtained as mentioned above. We used the algorithm proposed in [55], namely, computing the integrals (24) with 2^{15} randomly scrambled Sobol' points ensuring that all results for the ANOVA total and partial variances were obtained with at least 3 digits accuracy. The effective dimension $d_T(0.01)$ remained always equal to 2. Hence, Theorem 2 and Remark 1 apply if (A4) is satisfied. But, the latter may be assumed due to Corollary 2. Hence, the theory of Sects. 5 and 6 justifies the application of both randomized QMC methods.

Moreover, further tests showed that the variance accumulated by the sum of the ANOVA terms f_i, $3 \leq i \leq d$, did not exceed 0.6 % of the total variance $\sigma^2(f)$. These results were obtained by using the special estimator *Correlator 2* proposed in [39] for accurate estimation of relatively small partial variances. We observed also that the first variable under PCA seems to accumulate always more than 90 % of the total variance $\sigma^2(\Phi(x, \xi))$. Hence, PCA serves as excellent dimension reduction technique in this case.

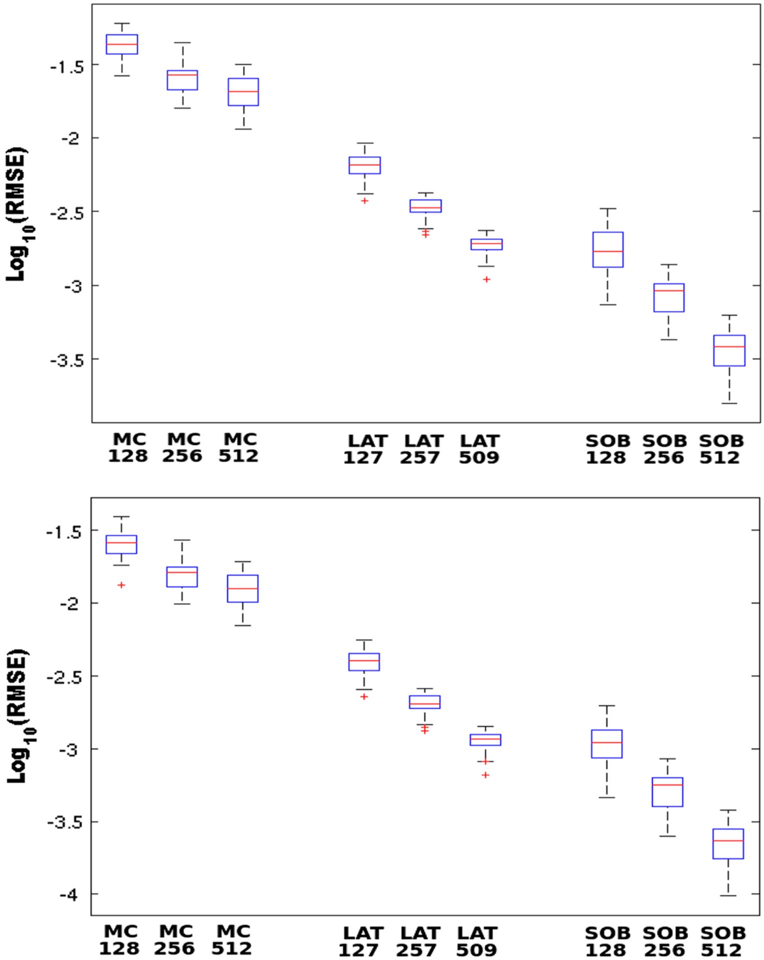

Fig. 2 Shown are the Log_{10} of relative RMSE with PCA factorization of covariance matrix for integration of $\Phi(x,\xi)$ (*top figure*) and for the optimal values of (46) (*bottom figure*) for parameters as stated above. The results for Mersenne Twister MC and scrambled Sobol' QMC (SOB) were obtained with $n_1 = 128$, $n_2 = 256$ and $n_3 = 512$ points, and for randomly shifted lattice rules QMC (LAT) with $n_1 = 127$, $n_2 = 257$ and $n_3 = 509$ lattice points

Using the Cholesky factorization the numerical results were completely different than those under PCA, see Fig. 3. The average of the estimated rates of convergence of both randomized QMC methods were 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 scrambled Sobol' points than for MC. In this case the theory of Sect. 5 does not apply since the effective truncation dimension of $\Phi(x,\xi)$ was estimated to be $d_T = 200$ in all tests. Further tests showed that the variance accumulated by the sum of the first order ANOVA terms f_i, $1 \le i \le d$, did not exceed 30 % of the total variance $\sigma^2(f)$.

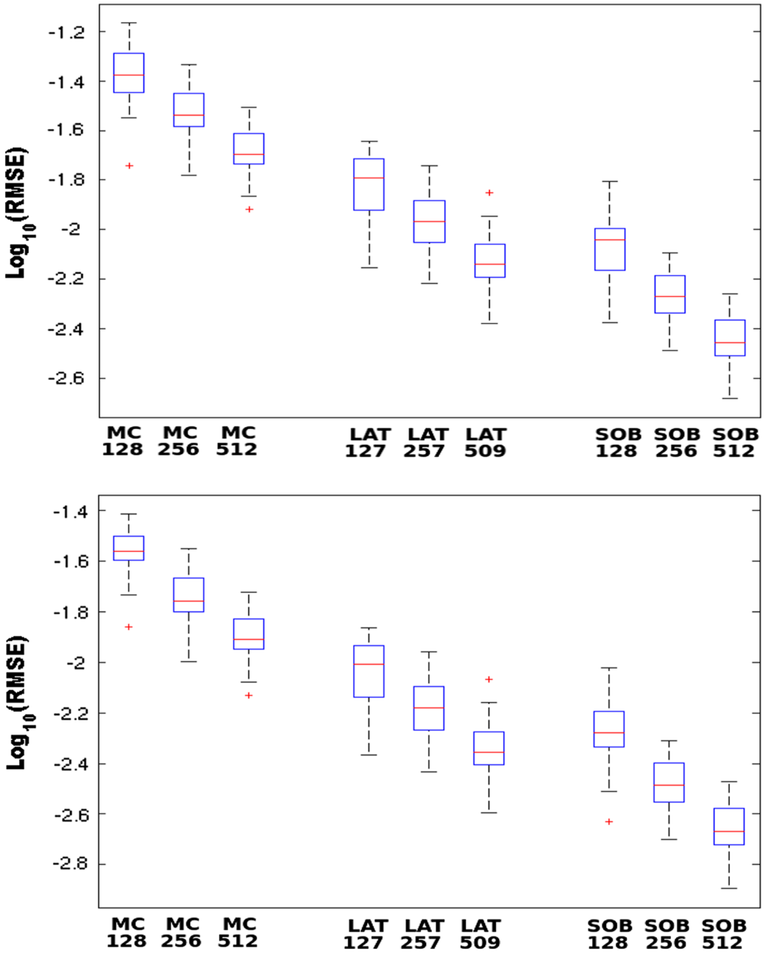

Fig. 3 Shown are the Log_{10} of relative RMSE with Cholesky factorization of covariance matrix for integration of $\Phi(x, \xi)$ (*top figure*) and for the optimal values of (46) (*bottom figure*). The results for Mersenne Twister MC and scrambled Sobol' QMC (SOB) were obtained with $n_1 = 128$, $n_2 = 256$ and $n_3 = 512$ points, and for randomly shifted lattice rules QMC (LAT) with $n_1 = 127$, $n_2 = 257$ and $n_3 = 509$ lattice points

9 Conclusions

Quasi-Monte Carlo methods were developed as alternative to Monte Carlo methods for numerical integration in higher dimensions. Their original convergence rate $O(n^{-1}(\log n)^{d-1})$ is clearly superior to the Monte Carlo rate $O(n^{-\frac{1}{2}})$, but required integrands that are of bounded variation in the sense of Hardy and Krause. Moreover, the term n^{-1} becomes effective only for very large sample sizes at least for higher dimensions. Meanwhile the enormous progress in Quasi-Monte Carlo theory has led to improved rates which may be effective already for smaller sample sizes like for randomly shifted lattice rules. The additional requirement is that the integrands belong to a mixed first order Sobolev space.

Our theoretical results in Sect. 5 show that at least the first and second order ANOVA terms of two-stage integrands satisfy this smoothness property. Hence, randomly shifted lattice rules and scrambled Sobol' sequences applied to two-stage stochastic programs may converge with the rates (6) and (8), respectively, if the first and second order ANOVA terms represent already a good approximation of them. The latter means that the effective superposition dimension of the integrands is at most 2. At first moment this appears as a serious restriction, but such low effective dimensions may be achieved by dimension reduction methods as computational results for option pricing models in the literature indicate. Our computational tests for a production planning model under price and demand uncertainty show that in case of normal distributions for prices and demands principal component analysis may lead to effective superposition dimension 2. Indeed our computational results proved the superior convergence behavior of both randomized QMC methods. Both methods lead to a *substantial improvement* compared to Monte Carlo schemes. We note that the results of Sect. 5 also justify the use of *sparse grid quadrature rules* [10,32] for two-stage stochastic programs.

A number of questions still remain open, for example, the smoothness of higher order ANOVA terms and, thus, the possible validity of the theory in Sect. 5 also for effective superposition dimensions larger than 2 or extensions of the theory regarding the geometric condition (A4) and of dimension reduction techniques beyond the case of (log)normal distributions. In addition, extensions to other stochastic optimization models like multistage ones deserve further efforts.

Acknowledgments The research of the first author was supported by grant GR 705/13 of the German Research Foundation (DFG) and the second by the DFG Research Center MATHEON at Berlin. The authors wish to thank all referees for their helpful comments and valuable suggestions that improved readability and presentation.

References

1. Bratley, P., Fox, B.L.: Algorithm 659: implementing Sobol's quasirandom sequence generator. ACM Trans. Math. Softw. **14**, 88–100 (1988)
2. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting, 2nd edn. Springer, New York (2002)
3. Dick, J.: Walsh spaces containing smooth functions and Quasi-Monte Carlo rules of arbitrary high order. SIAM J. Numer. Anal. **46**, 1519–1553 (2008)
4. Dick, J., Kuo, F.Y., Sloan, I.H.: High-dimensional integration: the Quasi-Monte Carlo way. Acta Numer. **22**, 133–288 (2013)
5. Dick, J., Pillichshammer, F.: Digital Nets and Sequences. Cambridge University Press, Cambridge (2010)
6. Dick, J., Sloan, I.H., Wang, X., Woźniakowski, H.: Liberating the weights. J. Complex. **20**, 593–623 (2004)
7. Drew, S.S., Homem-de-Mello, T.: Quasi-Monte Carlo strategies for stochastic optimization. In: Proceedings of the 2006 Winter Simulation Conference, IEEE pp. 774–782 (2006)
8. Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. In: IEEE St. Petersburg Power Tech (2005)
9. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Boca Raton (1992)

10. Gerstner, T., Griebel, M.: Numerical integration using sparse grids. Numer. Algorithms **18**, 209–232 (1998)
11. Glasserman, P.: Monte-Carlo Methods in Financial Engineering. Springer, New York (2003)
12. Griebel, M., Holtz, M.: Dimension-wise integration of high-dimensional functions with applications to finance. J. Complex. **26**, 455–489 (2010)
13. Griebel, M., Kuo, F.Y., Sloan, I.H.: The smoothing effect of the ANOVA decomposition. J. Complex. **26**, 523–551 (2010)
14. Griebel, M., Kuo, F.Y., Sloan, I.H.: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition. Math. Comput. **82**, 383–400 (2013)
15. Heitsch, H., Leövey, H., Römisch, W.: Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs? Stochastic Programming E-Print Series 5 (2012) (http://www.speps.org) and submitted
16. Hickernell, F.J.: A generalized discrepancy and quadrature error bound. Math. Comput. **67**, 299–322 (1998)
17. Hoeffding, W.: A class of statistics with asymptotically normal distribution. Ann. Math. Stat. **19**, 293–325 (1948)
18. Homem-de-Mello, T.: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling. SIAM J. Optim. **19**, 524–551 (2008)
19. Joe, S., Kuo, F.Y.: Remark on algorithm 659: implementing Sobol's quasirandom sequence generator. ACM Trans. Math. Softw. **29**, 49–57 (2003)
20. Koivu, M.: Variance reduction in sample approximations of stochastic programs. Math. Program. **103**, 463–485 (2005)
21. Kuo, F.Y.: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces. J. Complex. **19**, 301–320 (2003)
22. Kuo, F.Y., Schwab, Ch., Sloan, I.H.: Quasi-Monte Carlo methods for high-dimensional integration: the standard (weighted Hilbert space) setting and beyond. ANZIAM J. **53**, 1–37 (2011)
23. Kuo, F.Y., Sloan, I.H.: Lifting the curse of dimensionality. Notices Am. Math. Soc. **52**, 1320–1328 (2005)
24. 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. J. Complex. **26**, 135–160 (2010)
25. Kuo, F.Y., Sloan, I.H., Wasilkowski, G.W., Woźniakowski, H.: On decomposition of multivariate functions. Math. Comput. **79**, 953–966 (2010)
26. 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)
27. Lemieux, C.: Monte Carlo and Quasi-Monte Carlo Sampling. Springer, New York (2009)
28. Liu, R., Owen, A.B.: Estimating mean dimensionality of analysis of variance decompositions. J. Am. Stat. Assoc. **101**, 712–721 (2006)
29. Matoušek, J.: On the L_2-discrepancy for anchored boxes. J. Complex. **14**, 527–556 (1998)
30. Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. **8**, 3–30 (1998)
31. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia (1992)
32. Novak, E., Ritter, K.: High dimensional integration of smooth functions over cubes. Numer. Math. **75**, 79–97 (1996)
33. Nožička, F., Guddat, J., Hollatz, H., Bank, B.: Theory of Linear Parametric Programming (in German). Akademie-Verlag, Berlin (1974)
34. Nuyens, D., Cools, R.: Fast algorithms for component-by-component constructions of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces. Math. Comput. **75**, 903–922 (2006)
35. Owen, A.B.: Randomly permuted (t, m, s)-nets and (t, s)-sequences. In: Niederreiter, H., Shiue, P.J.S. (eds.) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Lecture Notes in Statistics, vol. 106, pp. 299–317. Springer, New York (1995)
36. Owen, A.B.: Monte Carlo variance of scrambled net quadrature. SIAM J. Numer. Anal. **34**, 1884–1910 (1997)
37. Owen, A.B.: The dimension distribution and quadrature test functions. Stat. Sinica **13**, 1–17 (2003)
38. Owen, A.B.: Multidimensional variation for Quasi-Monte Carlo. In: Fan, J., Li, G. (eds.) International Conference on Statistics, pp. 49–74. World Scientific, Singapore (2005)

39. Owen, A.B.: Better estimation of small Sobol' sensitivity indices. ACM Trans. Model. Comput. Simul. **23**, 11:1–11:17 (2013)
40. Papageorgiou, A.: Brownian bridge does not offer a consistent advantage in Quasi-Monte Carlo integration. J. Complex. **18**, 171–186 (2002)
41. Pennanen, T., Koivu, M.: Epi-convergent discretizations of stochastic programs via integration quadratures. Numer. Math. **100**, 141–163 (2005)
42. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
43. Römisch, W.: Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, vol. 10, pp. 483–554. Elsevier, Amsterdam (2003)
44. Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming, Handbooks in Operations Research and Management Science, vol. 10. Elsevier, Amsterdam (2003)
45. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming, MPS-SIAM Series on Optimization, Philadelphia, 2009 (2014 Second Edition)
46. Sloan, I.H., Woźniakowski, H.: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration. J. Complex. **14**, 1–33 (1998)
47. Sloan, I.H., Kuo, F.Y., Joe, S.: Constructing randomly shifted lattice rules in weighted Sobolev spaces. SIAM J. Numer. Anal. **40**, 1650–1665 (2002)
48. Sobol', I.M.: The distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. **7**, 86–112 (1967)
49. Sobol', I.M.: Multidimensional Quadrature Formulas and Haar Functions. Nauka, Moscow (1969). (in Russian)
50. Sobol', I.M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. **55**, 271–280 (2001)
51. Sobol', I.M., Asotsky, D., Kreinin, A., Kucherenko, S.: Construction and comparison of high-dimensional Sobol' generators. Wilmott **56**, 64–79 (2011)
52. Takemura, A.: Tensor analysis of ANOVA decomposition. J. Am. Stat. Assoc. **78**, 894–900 (1983)
53. Triebel, H.: Bases in Function Spaces, Sampling, Discrepancy, Numerical Integration. European Mathematical Society, Zürich (2010)
54. Walkup, D., Wets, R.J.-B.: Lifting projections of convex polyedra. Pac. J. Math. **28**, 465–475 (1969)
55. Wang, X., Fang, K.-T.: The effective dimension and Quasi-Monte Carlo integration. J. Complex. **19**, 101–124 (2003)
56. Wang, X., Sloan, I.H.: Why are high-dimensional finance problems often of low effective dimension. SIAM J. Sci. Comput. **27**, 159–183 (2005)
57. Wang, X., Sloan, I.H.: Quasi-Monte Carlo methods in financial engineering: an equivalence principle and dimension reduction. Oper. Res. **59**, 80–95 (2011)
58. Wets, R.J.-B.: Stochastic programs with fixed recourse: the equivalent deterministic program. SIAM Rev. **16**, 309–339 (1974)