Noname manuscript No.
(will be inserted by the editor)

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

H. Heitsch · H. Leövey · W. Römisch

the date of receipt and acceptance should be inserted later

Abstract Quasi-Monte Carlo algorithms are studied for designing discrete approximations of two-stage linear stochastic programs with random right-hand side and continuous probability distribution. The latter should allow for a transformation to a distribution with independent marginals. The two-stage integrands are piecewise linear, but neither smooth nor lie in 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 that first and second order ANOVA terms have mixed first order partial derivatives and belong to L_2. Hence, randomly shifted lattice rules (SLR) 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 at most two. We discuss effective dimensions and dimension reduction for two-stage integrands. The geometric condition is shown to be satisfied almost everywhere if the underlying probability distribution is normal and principal component analysis (PCA) is used for transforming the covariance matrix. Numerical experiments for a large scale two-stage stochastic production planning model with normal demand show that indeed convergence rates close to the optimal are achieved when using SLR and randomly scrambled Sobol' point sets accompanied with PCA for dimension reduction.

1 Introduction

Two-stage stochastic programs arise as deterministic equivalents of improperly posed random linear programs

$$\min\{\langle c, x\rangle : x \in X,\, Tx = h(\xi)\}, \tag{1}$$

where X is a convex polyhedral subset of $\mathbb{R}^m$, T a matrix, ξ is a d-dimensional random vector, $h(\xi) = (\xi, \bar{h})^\top$ for some $\bar{h} \in \mathbb{R}^{r-d}$, $d \le r$, and $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^m$. The modeling idea consists in the compensation of a possible deviation $h(\xi(\omega)) - Tx$ for a given realization $\xi(\omega)$ of ξ, by introducing additional costs $\Phi(x, \xi(\omega))$ whose

Weierstrass Institute, Berlin, Germany, `heitsch@wias-berlin.de`, Humboldt-University Berlin, Institute of Mathematics, Berlin, Germany `(leovey,romisch)@math.hu-berlin.de`

mean with respect to the probability distribution P of ξ is added to the objective of (1). In two-stage stochastic programming it is assumed that the additional costs represent the optimal value of a second-stage linear program, i.e.,

$$\Phi(x,\xi) = \inf\{\langle q,y\rangle : y \in \mathbb{R}^{\bar{m}},\, Wy = h(\xi) - Tx, y \geq 0\}, \tag{2}$$

where W is a $(r,\bar{m})$-matrix called recourse matrix, $q \in \mathbb{R}^{\bar{m}}$ the recourse costs and y the recourse decision. The deterministic equivalent program then is of the form

$$\min\Big\{\langle c,x\rangle + \int_{\mathbb{R}^d} \Phi(x,\xi)P(d\xi) : x \in X\Big\}. \tag{3}$$

In practical applications of stochastic programming the dimension d is often large, e.g., in economics, energy, finance or transportation (see [62] for a survey of applied models). It is worth noting that the option pricing models that served as motivating examples for the further development of Quasi-Monte Carlo algorithms (e.g. in [64,65,68]) may be reformulated as linear two-stage stochastic programs whose stochastic inputs are means of geometric Brownian motions paths. So, in a sense, the models considered here may be regarded as extensions of such financial models (see Example 1).
The standard approach to solving the optimization model (3) consists in approximating the underlying probability distribution by discrete distributions P_n based on a finite number n of *samples* or *scenarios* $\xi^j \in \mathbb{R}^d$ with probabilities p_j, $j = 1,\dots,n$, and to consider the approximate stochastic program

$$\min\Big\{\langle c,x\rangle + \sum_{j=1}^{n} p_j \Phi(x,\xi^j) : x \in X\Big\}.$$

The case of random samples is studied in detail at least for independent and identically distributed (iid) samples (see e.g. Chapters 6 and 7 in [51], [49, Sect. 4]), where the convergence rate (in probability or quadratic mean) is $O(n^{-\frac{1}{2}})$. Only a few papers related to stochastic programming deal with the situation of deterministic samples with identical weights $p_j = n^{-1}$ and proved (general) convergence results (see [7,45, 19,46], [23] for randomized samples or [50] for an overview).
There exist two main approaches for the generation of discrete approximations to P based on deterministic samples with identical weights. The first one is called *optimal quantization of probability distributions* (see [12], [42]) and determines such quantizations by (approximately) solving best approximation problems for P in terms of the L_p-minimal (or L_p-Wasserstein) metric ℓ_p, $p \geq 1$ (see Section 2.5 in [48]). The primal and dual representations of ℓ_1 together with a classical result (see [8, Proposition 2.1]) imply that

$$c\,n^{-\frac{1}{d}} \leq \ell_1(P,P_n) = \sup_{f\in\mathbb{F}_d,\,\|f\|_L\leq 1} \Big|\int_{\mathbb{R}^d} f(\xi)(P-P_n)(d\xi)\Big| \leq \ell_p(P,P_n)$$

holds for sufficiently large n and some constant $c > 0$ if P has a density on $\mathbb{R}^d$ and $\mathbb{F}_d$ denotes the Banach space of Lipschitz functions on $\mathbb{R}^d$ equipped with the Lipschitz norm $\|\cdot\|_L$. This shows that the convergence rate of $\ell_p(P,P_n)$ is at best $O(n^{-\frac{1}{d}})$. This rate is indeed established in [12, Theorem 6.2] under certain additonal conditions on P. It is well known that the unit ball $\{f \in \mathbb{F}_d : \|f\|_L \leq 1\}$ is too large for obtaining better rates.

The second approach utilizes *Quasi-Monte Carlo algorithms* that are of the form

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

and relies on the concept of equidistributed or low discrepancy point sets $\{x^j\}_{j=1}^n$ or sequences $(x^j)_{j\in\mathbb{N}}$ in $[0,1)^d$ (see [56,33,29,5]). As observed in [16] certain reproducing kernel Hilbert spaces $\mathbb{F}_d$ of functions $f : [0,1]^d \to \mathbb{R}$ are particularly useful for estimating the quadrature error. Let $K : [0,1]^d \times [0,1]^d \to \mathbb{R}$ be a kernel satisfying $K(\cdot, y) \in \mathbb{F}_d$ and $\langle f, K(\cdot, y)\rangle = f(y)$ for each $y \in [0,1]^d$ and $f \in \mathbb{F}_d$. If $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denote the inner product and norm in $\mathbb{F}_d$, and the integral

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

is a continuous functional on $\mathbb{F}_d$, the worst-case quadrature error $e_n(\mathbb{F}_d)$ allows the representation

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

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

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

Here, we consider the weighted tensor product Sobolev space [54,3]

$$\mathbb{F}_d = \mathcal{W}_{2,\text{mix}}^{(1,\dots,1)}([0,1]^d) = \bigotimes_{i=1}^{d} W_2^1([0,1]) \tag{5}$$

equipped with a weighted norm $\|f\|_\gamma$ where the index γ refers to a positive and non-increasing sequence (γ_i). The norm is given by $\|f\|_\gamma^2 = \langle f, f\rangle_\gamma$ and the inner product (see Section 4 for the notation)

$$\langle f, g\rangle_\gamma = \sum_{u\subseteq\{1,\dots,d\}} \gamma_u^{-1} \int_{[0,1]^{|u|}} I_u(f)(x^u) I_u(g)(x^u) dx^u \,, \tag{6}$$

where $I_u(f)$ and γ_u are defined for any subset u of $\{1,\dots,d\}$ by

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

The space $\mathbb{F}_d$ is also a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(x,y) = \prod_{i=1}^{d} \left(1 + \gamma_i (0.5 B_2(|x_i - y_i|) + B_1(x_i)B_1(y_i))\right) \quad (x, y \in [0,1]^d),$$

where $B_1(x) = x - \frac{1}{2}$ and $B_2(x) = x^2 - x + \frac{1}{6}$ are the Bernoulli polynomials of order 1 and 2, respectively [3,25].

Another example is a weighted tensor product Walsh space consisting of Walsh series (see [5, Example 2.8] and [4]). These spaces became important for analyzing the recently developed randomized lattice rules, namely, randomly shifted lattice rules [55, 24, 26, 35]) and random digitally shifted polynomial lattice rules (see [4, 5]). Both are special cases of *randomized Quasi-Monte Carlo algorithms (RQMC)* which will be discussed in Section 2.
Here, we just mention that randomly shifted lattice rules

$$Q_{n,d}(\Delta, f) = n^{-1} \sum_{j=0}^{n-1} f\left(\left\{\frac{jg}{n} + \Delta\right\}\right) \tag{7}$$

can be constructed, where Δ is uniformly distributed in $[0,1)^d$, $g \in \mathbb{Z}^d$ is the generator of the lattice which is obtained by a component-by-component algorithm [24] and $\{\cdot\}$ means taking componentwise the fractional part. For f belonging to the weighted (un)anchored tensor product Sobolev space $\mathbb{F}_d$ the root mean square error of such randomly shifted lattice rules can be bounded by [55, 24, 3]

$$\sqrt{\mathbb{E}_\Delta \left|I_d(f) - Q_{n,d}(\Delta, f)\right|^2} \leq C(\delta) n^{-1+\delta} \|f\|_\gamma \quad (\delta \in (0, 0.5]), \tag{8}$$

where the constant $C(\delta)$ does not depend on the dimension d if the sequence of nonnegative weights (γ_j) satisfies

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty \,. \tag{9}$$

Unfortunately, typical integrands in linear two-stage stochastic programming (see Section 3) do not belong to such tensor product Sobolev or Walsh spaces and are even not of bounded Hardy and Krause variation (on $[0,1]^d$). The latter condition represents the standard requirement on the integrand f to justify Quasi-Monte Carlo algorithms via the Koksma-Hlawka theorem [33, Theorem 2.11].
Alternatively, it is suggested in the literature to study the so-called ANOVA decomposition (see Section 4) of such integrands, the smoothness of the ANOVA terms, effective dimensions and/or sensitivity indices of the integrands.
The aim of the present paper is to follow the suggestions and to derive theoretical arguments that explain why modern RQMC methods, with focus on randomly shifted lattice rules (7), converge with nearly the optimal rate (8) for the considered class of stochastic programs although the integrands do not satisfy standard requirements in QMC analysis, e.g., do not belong to the weighted tensor product Sobolev space (5).
As a first step in this direction we show in Section 5 that all ANOVA terms except the one of highest order are continuously differentiable and possess second order partial derivatives almost everywhere under some geometric condition on the second stage program. In particular, the first and second order ANOVA terms belong to a mixed Sobolev space which is defined in Section 4. Error estimates show that the QMC convergence rate dominates the error if the effective superposition dimension is equal to 2 (Remark 1 in Section 5). In addition, we show in Section 6 that the geometric condition is satisfied for almost all covariance matrices if the underlying random vector is Gaussian. The meaning of "almost all" is also explained there. We also provide estimates of sensitivity indices and mean dimension in Section 7 and discuss techniques for dimension reduction. In accordance with the theoretical results in Section 5 our preliminary computational results in Section 8 show that scrambled Sobol' sequences

and randomly shifted lattice rules applied to a large scale two-stage stochastic program achieve convergence rates close to the optimal rate (8) if principal component analysis (PCA) is employed for dimension reduction. Both randomized QMC algorithms clearly outperform Monte Carlo methods.

2 Randomized Quasi-Monte Carlo methods

Randomized Quasi-Monte Carlo algorithms (RQMC) permit us to combine the good features of Monte Carlo within Quasi-Monte Carlo methods for practical error estimation.
If f has mixed partial derivatives of second order in each variable in $L_2([0,1]^d)$, then the convergence rate (8) can be improved to nearly $O(n^{-2})$ by embedding the function into an appropriate Korobov space through the so called *tent* or *baker's* transformation (see [6, Section 5]). Although this is theoretically true, this "extra" improved rate of convergence (over the already good $O(n^{-1+\delta})$) for smoother integrands is rarely observed for RQMC in practical applications of high-dimensional integration where only moderate or small sample sizes n are affordable for computations [17].
A large class of QMC rules that can be randomized are the well known (t, m, d)-nets and (t, d)-sequences [33]. The randomization techniques for these constructions follow mainly two schemes: *random digital shifts* and *random scramblings*. Random digital shifting of (t, m, d)-nets and (t, d)-sequences can be performed in a similar way as mentioned for randomly shifting lattice rules, but the operations to add the shift must be carried out in the basis b used to define the (t, m, d)-nets (see [6, Section 6]). The resulting RQMC point set preserves the original net structure. Similar bounds for the root mean square error as in (8) can be obtained for integrands belonging to the weighted (anchored and unanchored) tensor product Sobolev space $\mathbb{F}_d$ by using a special class of (t, m, d)-nets called *polynomial lattice rules*, see again [6, Section 6].
The random scrambling method was first introduced by Owen in [36]. The basic properties of Owen's scrambling are the following:

Proposition 1 ***(Equidistribution)***
A randomized (t, m, d)-net in base b using Owen's scrambling is again a (t, m, d)-net in base b with probability 1. A randomized (t, d)-sequence in base b using Owen's scrambling is again a (t, d)-sequence in base b with probability 1.

Proposition 2 ***(Uniformity)***
Let $\tilde{z}_i$ be the randomized version of a point z_i originally belonging to a (t, m, d)-net in base b or a (t, d)-sequence in base b, using Owen's scrambling. Then $\tilde{z}_i$ has a uniform distribution in $[0,1)^d$, that is, for any Lebesgue measurable set $G \subseteq [0,1)^d$, $P(\tilde{z}_i \in G) = \lambda_d(G)$, with λ_d the d-dimensional Lebesgue measure.

Note that the uniformity property stated above ensures that the resulting RQMC estimator $\hat{Q}_{n,d}(.)$ is unbiased. We mention here the general results about the variance of a RQMC estimator $\hat{Q}_{n,d}(.)$ after Owen's random scrambling technique to (t, m, d)-nets in base b for functions $f \in L_2([0,1]^d)$ (see [37]).

Theorem 1 *Let $\tilde{z}_i$, $1 \leq i \leq n$, be the points of a scrambled (t, m, d)-net in base b, and let f be a function on $[0,1)^d$ with integral I and variance $\sigma^2 := \int (f(z) - I_d(f))^2 dz < \infty$.*

Let $\hat{Q}_{n,d}(f) = n^{-1}\sum_{i=1}^{n} f(\tilde{z}_i)$ *with* $n = b^m$ *be the RQMC estimator. Then its variance* $\mathrm{Var}(\hat{Q}_{n,d}(f))$ *has the properties*

$$\mathrm{Var}(\hat{Q}_{n,d}(f)) = o(n^{-1}) \quad \text{as } n \to \infty \quad \text{and} \quad \mathrm{Var}(\hat{Q}_{n,d}(f)) \leq \frac{b^t}{n}\left(\frac{b+1}{b-1}\right)^d \sigma^2.$$

For $t = 0$ *we have*

$$\mathrm{Var}(\hat{Q}_{n,d}(f)) \leq \frac{1}{n}\left(\frac{b}{b-1}\right)^{d-1} \sigma^2 \leq \frac{1}{n} e\sigma^2.$$

Note that the last inequality for $t = 0$ above holds since in this case one must have $b \geq d$. If the function f has bounded variation in the sense of Hardy and Krause $V_{\mathrm{HK}}(f) < \infty$, then by the equidistribution property stated above the classical Koksma-Hlawka inequality holds with probability 1 for random scrambled (t, m, d)-nets, therefore the classical discrepancy bounds for (t, m, d)-nets [33] lead to

$$\mathrm{Var}(\hat{Q}_{n,d}(f)) = O\left(n^{-2}(\log n)^{2(d-1)}\right).$$

If the integrand f has a mixed partial derivative of order d which satisfies a Hölder condition, the above rate of convergence can be improved to [37,38]

$$\mathrm{Var}(\hat{Q}_{n,d}(f)) = O\left(n^{-3}(\log n)^{d-1}\right).$$

Further improved results for functions having *finite generalized Hardy and Krause variation* can be found in [5, Theorem 13.25]. Note, however, that distinct from (8) sequences of the form $(n^{-\alpha}(\log n)^{d-1})$ increase as long as $n < \exp\frac{d-1}{\alpha}$ and, hence, require extremely large sample sizes n for higher dimensions d to get small.

The piecewise linear convex functions arising in stochastic programming (see Section 3) do not even have mixed partial derivatives (in the sense of Sobolev) in general. They do not have finite (generalized) Hardy and Krause variation either. The latter is shown for the classical Hardy and Krause variation of the special function $f_d(x) = \max\{x_1 + x_2 + \cdots + x_d - \frac{1}{2}, 0\}$, $d \geq 3$, in [40, Proposition 17], but its proof carries over to the generalized variation. Thus, none of the results stated or mentioned above for RQMC can be used to formally justify an observed root mean square error convergence near to $O(n^{-1})$ (see Section 8) for integrands appearing in linear two-stage stochastic programming.

Several modifications of the original scrambling method proposed by Owen have been investigated in order to provide efficient implementations of scramblings for practical applications, see the survey [28] and [31,20,60,41] for example.

Recent QMC constructions that aim to advantage from a setting with even higher smoothness of the integrands are the so called *higher order digital nets* in combination with *higher order scramblings*. For further information on this topic we refer the reader to [1,5].

3 Integrands of linear two-stage stochastic programs

As described in the introduction, the integrands of two-stage linear stochastic programs with random right-hand sides are

$$\Phi(x, \xi) = \phi(h(\xi) - Tx), \tag{10}$$

where ϕ denotes the optimal value function assigning to each $t \in \mathbb{R}^r$ the infimum $\phi(t) = \inf\{\langle q, y\rangle : Wy = t, y \geq 0\}$ in $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Due to duality in linear programming, the function ϕ is finite and

$$\phi(t) = \sup\{\langle t, z\rangle : W^\top z \leq q\}, \tag{11}$$

if $t \in \operatorname{dom}\phi = \{t \in \mathbb{R}^r : \phi(t) < \infty\}$ and the dual feasible set $\mathcal{D} = \{z \in \mathbb{R}^r : W^\top z \leq q\}$ is nonempty. Here, $q \in \mathbb{R}^{\bar{m}}$, W is a $(r, \bar{m})$-matrix and t varies in the polyhedral cone $\operatorname{dom}\phi = W(\mathbb{R}^{\bar{m}}_+)$. If $\mathcal{D}$ is nonempty, it is of the form

$$\mathcal{D} = \operatorname{conv}\{v^1, \ldots, v^\ell\} + (\operatorname{dom}\phi)^*,$$

where $v^1, \ldots, v^\ell$ are the vertices of $\mathcal{D}$, conv means convex hull and $(\operatorname{dom}\phi)^*$ is the polar cone to the cone $\operatorname{dom}\phi = W(\mathbb{R}^{\bar{m}}_+)$, i.e.,

$$(\operatorname{dom}\phi)^* = \{d \in \mathbb{R}^r : \langle d, t\rangle \leq 0, \forall t \in W(\mathbb{R}^{\bar{m}}_+)\} = \{d \in \mathbb{R}^r : W^\top d \leq 0\}.$$

Furthermore, there exist polyhedral cones $\mathcal{K}_j$, $j = 1, \ldots, \ell$, decomposing $\operatorname{dom}\phi$. The cone $\mathcal{K}_j$ is the normal cone to the vertex v^j, i.e.,

$$\begin{aligned} \mathcal{K}_j &= \{t \in \operatorname{dom}\phi : \langle t, z - v^j\rangle \leq 0, \ \forall z \in \mathcal{D}\} \quad (j = 1, \ldots, \ell) && (12)\\ &= \{t \in \operatorname{dom}\phi : \langle t, v^i - v^j\rangle \leq 0, \ \forall i = 1, \ldots, \ell, \ i \neq j\}. && (13) \end{aligned}$$

Moreover,

$$\phi(t) = \langle v^j, t\rangle \quad (\forall t \in \mathcal{K}_j) \quad \text{and} \quad \phi(t) = \max_{j=1,\ldots,\ell} \langle v^j, t\rangle \quad (\forall t \in \operatorname{dom}\phi) \tag{14}$$

and $\cup_{j=1,\ldots,\ell} \mathcal{K}_j = \operatorname{dom}\phi$. The intersection $\mathcal{K}_j \cap \mathcal{K}_{j'}$ for $j \neq j'$ coincides with a common closed face of dimension less than r. It is a common closed face of dimension $r - 1$ iff the two cones are adjacent. In the latter case, the intersection is contained in

$$\{t \in \operatorname{dom}\phi : \langle t, v^{j'} - v^j\rangle = 0\}. \tag{15}$$

If there exists $k \in \{1, \ldots, d\}$ such that the kth components of v^j and $v^{j'}$ coincide, the common closed face of $\mathcal{K}_j$ and $\mathcal{K}_{j'}$ contains at least one of the two one-dimensional cones

$$\{(0, \ldots, 0, t_k, 0, \ldots, 0) : t_k \geq 0\} \quad \text{and} \quad \{(0, \ldots, 0, t_k, 0, \ldots, 0) : t_k \leq 0\}.$$

The cones $\mathcal{K}_j$ may also be represented by

$$\mathcal{K}_j = \Big\{ \sum_{i \in I_j} \lambda_i w^i : \lambda_i \geq 0, \, i \in I_j \Big\},$$

where $w^i \in \mathbb{R}^r$ are the columns of W and $I_j = \{i \in \{1, \ldots, \bar{m}\} : \langle w^i, v^j\rangle = q_i\}$. Each vertex v^j is determined by r linear independent equations out of the $\bar{m}$ equations $\langle w^i, v\rangle = q_i$, $i = 1, \ldots, \bar{m}$.

In the following sections we need the conditions

(A1) $h(\xi) - Tx \in W(\mathbb{R}^{\bar{m}}_+)$ for all $\xi \in \mathbb{R}^d$, $x \in X$ (*relatively complete recourse*).

(A2) The dual feasible set $\mathcal{D}$ is nonempty (*dual feasibility*).

(A3) $\int_{\mathbb{R}^d} \|\xi\|^2 P(d\xi) < \infty$ (*finite second order moment*).

(A4) 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 $\mathbb{R}$, $i = 1, \dots, d$ (*independent components*).

(A5) The first d components of the adjacent vertices of $\mathcal{D}$ are distinct, i.e., all common closed faces of the normal cones to two adjacent vertices of $\mathcal{D}$ do not parallel the first d coordinate axes in $\mathbb{R}^r$ (*geometric condition*).

Conditions (A3) and (A4) are needed for defining and using the ANOVA decomposition in Sections 4 and 5. Our *present analysis applies to random vectors ξ of the form* $\xi = B\zeta$, where B is a nonsingular (d, d)-matrix and the d-dimensional random vector ζ has independent components. Then the equality constraint $Wy = h(\xi) - Tx$ in (2) has to be replaced by $\hat{B}^{-1}Wy = h(\zeta) - \hat{B}^{-1}Tx$, where

$$\hat{B} = \begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix}. \tag{16}$$

and I denotes the $(r-d, r-d)$ identity matrix. The structure $\xi = B\zeta$ applies, in particular, to multivariate normal probability distributions where the matrix B may be chosen to be orthogonal (see Section 6). The geometric condition (A5) has technical character and will be further discussed in Section 6.

Conditions (A1), (A2), (A3) imply that the two-stage stochastic program (3) 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 [49, Theorems 5 and 9] for general stochastic programs imply the perturbation estimate

$$|v(P) - v(Q)| \le L \sup_{x \in X} \Big| \int_{\mathbb{R}^d} \Phi(x,\xi)(P-Q)(d\xi) \Big| \tag{17}$$

$$\emptyset \ne S(Q) \subseteq S(P) + \psi_P^{-1}\Big(\sup_{x \in X} \Big| \int_{\mathbb{R}^d} \Phi(x,\xi)(P-Q)(d\xi) \Big| \Big) \mathbb{B}, \tag{18}$$

where $\mathbb{B}$ is the unit ball in $\mathbb{R}^m$, ψ_P is the growth function of the objective

$$\psi_P(\tau) = \inf \Big\{ \langle c, x \rangle + \int_{\mathbb{R}^d} \Phi(x,\xi) P(d\xi) - v(P) : d(x, S(P)) \le \tau,\, x \in X \Big\} \quad (\tau \ge 0),$$

its inverse is defined by $\psi_P^{-1}(t) = \sup\{\tau \in \mathbb{R}_+ : \psi_P(\tau) \le t\}$, and Q is a probability measure satisfying (A3), too.

For further information on linear parametric programming and two-stage stochastic programming we refer to [61,34] and [51,52,69].

To give an example for (3) we show that option pricing models considered as stimulating examples for the recent developments in QMC theory (see e.g. [65,66]) may be reformulated as linear two-stage stochastic programs.

Example 1 Let the first stage variable x represent the strike price at the expiration date T_e. The dimensions are set to $m = 1$, $\bar{m} = 2$ and the matrix W is set to $W = (w, -w)$ with $w = \exp(rT_e)$ and r denoting the risk-free interest rate. The second stage program and its dual are

$$\begin{aligned} \min\{y_1 : Wy = \xi - x, y \in \mathbb{R}^2, y \ge 0\} &= \max\{(\xi - x)z : z \in \mathbb{R}, W^\top z \le (1,0)^\top\} \\ &= \max\{(\xi - x)z : 0 \le wz \le 1\}. \end{aligned}$$

The terminal payoff is $\exp(-rT_e)\max\{0, \xi - x\}$ and $v = 0$ and $v = \frac{1}{w}$ are the only vertices. Taking the expectation then leads to the optimization model

$$\min \Big\{ -x + \int_{\mathbb{R}} \exp(-rT_e) \max\{0, \xi - x\} \rho(\xi) d\xi : x \ge 0 \Big\}$$

for maximizing the strike price. Now, it depends on the kind of option how the random variable ξ depends on the geometric Brownian motion S given by

$$S_t = S_0 \exp\left((r - \tfrac{1}{2}\sigma^2)t + \sigma B_t\right)$$

with volatility σ and standard Brownian motion $(B_t)_{t\geq 0}$. For example, for arithmetic Asian options one has [64]

$$\xi = \frac{1}{d}\sum_{i=1}^{d} S_{t_i} \quad \text{with} \quad t_i = \frac{iT_e}{d},\ i = 1, \dots, d.$$

Hence, in a sense, the integrand (10) extends the situations encountered in such option pricing models. It is, however, much more involved.

4 ANOVA decomposition of integrands and effective dimension

The analysis of variance (ANOVA) decomposition of a function was first proposed as a tool in statistical analysis (see [18] and the survey [59]). In [56] it was first used for the analysis of quadrature methods.

We consider a density function ρ on $\mathbb{R}^d$ and assume (A4) from Section 3. As in [15] we consider the weighted $\mathcal{L}_p$ space over $\mathbb{R}^d$, i.e., $\mathcal{L}_{p,\rho}(\mathbb{R}^d)$, with the norm

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

Let $\mathfrak{D} = \{1, \dots, d\}$ and $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$. The projection P_k, $k \in \mathfrak{D}$, is given by

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

Clearly, the function $P_k f$ is constant with respect to ξ_k. For $u \subseteq \mathfrak{D}$ we use $|u|$ for its cardinality, $-u$ for $\mathfrak{D} \setminus u$ and write

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

where the product means composition. We note that the ordering within the product is not important because of Fubini's theorem. The function $P_u f$ is constant with respect to all ξ_k, $k \in u$. Note that P_u satisfies the properties of a projection, namely, P_u is linear and $P_u^2 = P_u$.

The ANOVA decomposition of $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$ is of the form [64, 27]

$$f = \sum_{u\subseteq\mathfrak{D}} f_u \tag{19}$$

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

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

It is known from [27] that the ANOVA terms are given explicitly by

$$f_u = \sum_{v\subseteq u}(-1)^{|u|-|v|}P_{-v}f = P_{-u}(f) + \sum_{v\subsetneq u}(-1)^{|u|-|v|}P_{u-v}(P_{-u}(f)), \qquad (20)$$

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 motivates that f_u is essentially as smooth as $P_{-u}(f)$ due to the Inheritance Theorem [15, Theorem 2]. The following result is well known (e.g. [64]).

Proposition 3 *If f belongs to $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u\subseteq\mathfrak{D}}$ are orthogonal in $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$.*

We define the variance of f and f_u by $\sigma^2(f) = \|f - I_{d,\rho}(f)\|_{2,\rho}^2$, $\sigma_u^2(f) = \|f_u\|_{2,\rho}^2$, and have

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

In the literature, the ANOVA decomposition is often considered for functions $g \in \mathcal{L}_1([0,1]^d)$. Then the projections are defined by

$$(P_k^\star g)(\upsilon) := \int_0^1 g(\upsilon_1,\dots,\upsilon_{k-1},s,\upsilon_{k+1},\dots,\upsilon_d)ds \quad (\upsilon\in[0,1]^d)$$

and

$$P_u^\star g := \Big(\prod_{k\in u} P_k^\star\Big)(g) \quad (u\subseteq\mathfrak{D}).$$

Similarly to the case in $\mathbb{R}^d$ the ANOVA decomposition of $g \in \mathcal{L}_1([0,1]^d)$ is of the form

$$g = \sum_{u\subseteq\mathfrak{D}} g_u, \quad g_\emptyset := I_d(g) := P_{\mathfrak{D}}^\star(g) \quad \text{and} \quad g_u := P_{-u}^\star(g) - \sum_{v\subsetneq u} g_v$$

with g_u depending only on υ^u, i.e., on the variables υ_j with indices $j\in u$. Note that $P_u^\star$ is indeed again a projection and, assuming that $g \in \mathcal{L}_2([0,1]^d)$, the same orthogonality property (now over $\mathcal{L}_2([0,1]^d)$) as in Proposition 3 follows.
Assuming now for simplicity that $\rho_j(t) > 0$ for all $t\in\mathbb{R}$, $j = 1,\dots,d$, an integrand $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$ can be transformed into a function g defined on $[0,1]^d$ by inverting the function

$$\varphi := (\varphi_1,\dots,\varphi_d), \quad \varphi_i(t) := \int_{-\infty}^t \rho_i(s)ds \quad (i\in\mathfrak{D}) \qquad (21)$$

and by defining

$$g(\upsilon) := \begin{cases} (f\circ\varphi^{-1})(\upsilon) & \text{if } \upsilon\in(0,1)^d, \\ 0 & \text{if } \upsilon\in[0,1]^d\setminus(0,1)^d. \end{cases} \qquad (22)$$

Then the ANOVA terms g_u of g satisfy

$$f_u(\xi^u) = g_u\circ\varphi_u(\xi^u) \text{ for } \xi^u\in\mathbb{R}^{|u|}, \;\; g_u(\upsilon^u) = (f_u\circ\varphi_u^{-1})(\upsilon^u) \text{ for } \upsilon^u\in(0,1)^{|u|}, \;\; (23)$$

where

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

When setting $\sigma_u^2(g) := \int_{[0,1]^{|u|}} g_u^2(\upsilon^u)d\upsilon^u$ for $\emptyset \neq u \subseteq \mathfrak{D}$ and $\sigma_\emptyset^2(g) := 0$ one obtains $\sigma_u^2(g) = \sigma_u^2(f)$ for $u \subseteq \mathfrak{D}$.
We return to the $\mathbb{R}^d$ case and assume $\sigma(f) > 0$ in the following to avoid trivial cases. 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 in [57] to define *global sensitivity indices* of a set $u \subseteq \mathfrak{D}$ by

$$S_u = \frac{1}{\sigma^2(f)} \sum_{v \subseteq u} \sigma_v^2(f) \quad \text{and} \quad \bar{S}_u = 1 - S_{-u} = \frac{1}{\sigma^2(f)} \sum_{v \cap u \neq \emptyset} \sigma_v^2(f).$$

If $\bar{S}_u$ is small, then the variable ξ^u is considered inessential for f in [57].
The normalized ratios are also used in [39, 30] to define and study the *dimension distribution* of a function f in two ways. The dimension distribution of f in the *superposition (truncation) sense* is a probability measure ν_S (ν_T) defined on the power set of $\mathfrak{D}$ by

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

Hence, the *mean dimension* in the superposition (truncation) sense is

$$\bar{d}_S = \sum_{\emptyset \neq u \subseteq \mathfrak{D}} |u| \frac{\sigma_u^2(f)}{\sigma^2(f)} \qquad \Big(\bar{d}_T = \sum_{\emptyset \neq u \subseteq \mathfrak{D}} \max\{j : j \in u\} \frac{\sigma_u^2(f)}{\sigma^2(f)}\Big). \tag{24}$$

It is proved in [30, Theorem 2] that the mean dimension $\bar{d}_S$ in the superposition sense is closely related to the global sensitivity indices of subsets of $\mathfrak{D}$ containing a single element. Namely,

$$\bar{d}_S = \sum_{j=1}^{d} \bar{S}_{\{j\}}. \tag{25}$$

The paper [30] also provides a formula for the dimension variance based on $\bar{S}_u$ for all subsets u of $\mathfrak{D}$ containing two indices.
For small $\varepsilon \in (0, 1)$ ($\varepsilon = 0.01$ is suggested in a number of papers), the *effective superposition (truncation) dimension* $d_S(\varepsilon) \in \mathfrak{D}$ ($d_T(\varepsilon) \in \mathfrak{D}$) is the $(1 - \varepsilon)$-quantile of ν_S (ν_T), i.e.,

$$d_S(\varepsilon) = \min\{s \in \mathfrak{D} : \nu_S(u) \geq 1 - \varepsilon, |u| \leq s\}$$
$$d_T(\varepsilon) = \min\{s \in \mathfrak{D} : \nu_T(\{1, \ldots, s\}) \geq 1 - \varepsilon\}.$$

Note that $d_S(\varepsilon) \leq d_T(\varepsilon)$ and (see [64, 13])

$$\max\Big\{\Big\|f - \sum_{|u| \leq d_S(\varepsilon)} f_u\Big\|_{2,\rho}, \Big\|f - \sum_{u \subseteq \{1,\ldots,d_T(\varepsilon)\}} f_u\Big\|_{2,\rho}\Big\} \leq \sqrt{\varepsilon}\sigma(f). \tag{26}$$

Small effective superposition dimension $d_S(\varepsilon)$, even if $d_T(\varepsilon)$ is large, suggests that we may expect superiority of QMC over MC. We note that there exist algorithms based on MC or QMC to compute global sensitivity indices and effective dimensions approximately (see [57, 64, 58, 65] for example). Since the algorithms are often described for functions on $[0, 1]^d$, we mention that

– the dimension distribution and, hence, any effective dimension of f is the same as for g given by (22).

– The algorithm of [64] for estimating the effective truncation dimension can be carried out equivalently for f, with its obvious adaption to the $\mathbb{R}^d$ setting.

All these notions are discussed in [39] for different classes of functions, including additive and multiplicative functions. We record here the results for additive functions for later reference.

Example 2 For functions f having separability structure, i.e., f is of the form

$$f(\xi) = \sum_{j=1}^{d} g_j(\xi_j) \quad (\xi \in \mathbb{R}^d)$$

with $g_j \in \mathcal{L}_{2,\rho_j}(\mathbb{R})$, $j = 1, \ldots, d$, the second and higher order ANOVA terms vanish (see [39]). Hence, the effective superposition dimension $d_S(\varepsilon)$ is equal to 1 for every $\varepsilon \in (0, 1)$ while the effective truncation dimension $d_T(\varepsilon) = s$ if

$$\sum_{j=s+1}^{d} \sigma_j^2 \leq \varepsilon \Big(\sum_{j=1}^{d} \sigma_j^2 \Big),$$

where σ_j is the variance of g_j, $j = 1, \ldots, d$.

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 (much) smoother than the original integrand f under certain conditions (see [14,15]). As in [15] 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

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

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

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

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. Then classical derivatives are also weak derivatives. In accordance with the notation (5) we consider in the next section the *mixed Sobolev space*

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

5 ANOVA decomposition of linear two-stage integrands

According to Section 3 the integrands in linear two-stage stochastic programming map from $\mathbb{R}^d$ to $\mathbb{R}$ and are given by

$$f(\xi) = f_x(\xi) = \max_{j=1,\ldots,\ell} \langle v^j, (\xi, \bar{h}) - Tx \rangle \quad (x \in X), \tag{27}$$

where the v^j, $j = 1, \dots, \ell$, are the vertices of the dual feasible set $\mathcal{D} = \{z \in \mathbb{R}^r : W^\top z \le q\}$ and $\mathcal{K}_j$ are the normal cones to v^j, $j = 1, \dots, \ell$.

The integrands are parametrized by the first-stage decision x varying in X. Such functions do not belong to the tensor product Sobolev spaces described in Section 1 and, in general, are not of bounded variation in the sense of Hardy and Krause (see [40, Proposition 17]).

Next we intend to compute projections $P_k(f)$ for $k \in \mathfrak{D}$. Let $x \in X$ be fixed, $\xi_i \in \mathbb{R}$, $i = 1, \dots, d$, $i \neq k$, be given. We set $\xi^k = (\xi_1, \dots, \xi_{k-1}, \xi_{k+1}, \dots, \xi_d)$ and $\xi^k_s = (\xi_1, \dots, \xi_{k-1}, s, \xi_{k+1}, \dots, \xi_d)$. We assume (A1)–(A5) and have according to Section 3

$$(\xi^k_s, \bar{h}) - Tx \in \operatorname{dom} \phi = \bigcup_{j=1}^{\ell} \mathcal{K}_j$$

for every $s \in \mathbb{R}$ and by definition of the projection

$$(P_k f)(\xi^k) = \int_{-\infty}^{\infty} f(\xi^k_s)\rho_k(s)ds = \int_{-\infty}^{\infty} f(\xi_1, \dots, \xi_{k-1}, s, \xi_{k+1}, \dots, \xi_d)\rho_k(s)ds. \quad (28)$$

The one-dimensional affine subspace $\{(\xi^k_s, \bar{h}) - Tx : s \in \mathbb{R}\}$ intersects a finite number of the polyhedral cones $\mathcal{K}_j$. Hence, there exist $p = p(k) \in \mathbb{N} \cup \{0\}$, $s_i = s^k_i \in \mathbb{R}$, $i = 1, \dots, p$, and $j_i = j^k_i \in \{1, \dots, \ell\}$, $i = 1, \dots, p+1$, such that $s_i < s_{i+1}$ and

$$\begin{aligned}
&(\xi^k_s, \bar{h}) - Tx \in \mathcal{K}_{j_1} && \forall s \in (-\infty, s_1] \\
&(\xi^k_s, \bar{h}) - Tx \in \mathcal{K}_{j_i} && \forall s \in [s_{i-1}, s_i] \quad (i = 2, \dots, p) \\
&(\xi^k_s, \bar{h}) - Tx \in \mathcal{K}_{j_{p+1}} && \forall s \in [s_p, +\infty).
\end{aligned}$$

By setting $s_0 := -\infty$, $s_{p+1} := \infty$, we obtain the following explicit representation of $P_k f$

$$(P_k f)(\xi^k) = \sum_{i=1}^{p+1} \int_{s_{i-1}}^{s_i} \langle v^{j_i}, (\xi^k_s, \bar{h}) - Tx\rangle \rho_k(s)ds, \quad (29)$$

where the points s_i, $i = 1, \dots, p$, satisfy the equations

$$\begin{aligned}
0 &= \langle (\xi^k_{s_i}, \bar{h}) - Tx, v^{j_{i+1}} - v^{j_i}\rangle = \langle (\xi^k_{s_i}, 0) + (0, \bar{h}) - Tx, v^{j_{i+1}} - v^{j_i}\rangle \\
&= \sum_{\substack{j=1 \\ j \neq k}}^{d} \xi_j (v^{j_{i+1}}_j - v^{j_i}_j) + s_i (v^{j_{i+1}}_k - v^{j_i}_k) + \langle (0, \bar{h}) - Tx, v^{j_{i+1}} - v^{j_i}\rangle
\end{aligned}$$

according to (15). By setting $w^i = v^{j_{i+1}} - v^{j_i}$ for $i = 1, \dots, p$ and $z(x) = (0, \bar{h}) - Tx$ this leads to the explicit formula

$$s_i = s_i(\xi^k, x) = \frac{1}{w^i_k}\Big[- \sum_{\substack{j=1 \\ j \neq k}}^{d} w^i_j \xi_j - \langle z(x), w^i\rangle \Big] \quad (i = 1, \dots, p). \quad (30)$$

Hence, all s_i, $i = 1, \dots, p$, are affine functions of the remaining components ξ_j, $j \neq k$. The first step in our analysis consists in studying smoothness properties of the projection $P_k f$ on $\mathbb{R}^d$. We note that f and $P_k f$ are finite convex functions on $\mathbb{R}^d$ and $\mathbb{R}^{d-1}$, respectively and, hence, twice differentiable almost everywhere on their domains

due to Alexandroff's theorem (see, for example, [10, Section 6.4]). Our analysis shows that the integration in (28) even improves the smoothness properties.
In the following, we consider a point $\xi_0^k \in \mathbb{R}^{d-1}$ and an open ball $\mathbb{B}_{\epsilon_0}(\xi_0^k)$ around ξ_0^k with radius ϵ_0. Assume that the ball $\mathbb{B}_{\epsilon_0}(\xi_0^k)$ is small enough such that the set of cones

$$\mathcal{K}_\epsilon(\xi_0^k) := \Big\{\mathcal{K}_j : \ \mathcal{K}_j \cap \{(\xi_s^k, \bar{h}) - Tx : s \in \mathbb{R}\} \neq \emptyset \text{ for some } \xi^k \in \mathbb{B}_\epsilon(\xi_0^k)\Big\} \tag{31}$$

satisfies $\mathcal{K}_\epsilon(\xi_0^k) = \mathcal{K}_{\epsilon_0}(\xi_0^k)$ for $0 < \epsilon < \epsilon_0$. Thus, the relevant intersecting cones are fixed in $\mathbb{B}_\epsilon(\xi_0^k)$. We consider also the sets of intersecting cones at an arbitrary point $\xi^k \in B_\epsilon(\xi_0^k)$

$$\mathcal{K}(\xi^k) := \Big\{\mathcal{K}_j : \mathcal{K}_j \cap \{(\xi_s^k, \bar{h}) - Tx : s \in \mathbb{R}\} \neq \emptyset\Big\}. \tag{32}$$

Because any affine one-dimensional space $\{(\xi_s^k, \bar{h}) - Tx : s \in \mathbb{R}\}$ for some $\xi^k \in \mathbb{B}_\epsilon(\xi_0^k)$ is in fact a parallel translation of $\{(\xi_{0,s}^k, \bar{h}) - Tx : s \in \mathbb{R}\}$, ϵ_0 can be chosen even small enough such that $\mathcal{K}(\xi^k) \subseteq \mathcal{K}_{\epsilon_0}(\xi_0^k)$ for every $\xi^k \in \mathbb{B}_{\epsilon_0}(\xi_0^k)$. Therefore we have

$$\mathcal{K}(\xi_0^k) = \mathcal{K}_{\epsilon_0}(\xi_0^k). \tag{33}$$

Moreover, since the cones $\mathcal{K}_j \in \mathcal{K}_{\epsilon_0}(\xi_0^k)$ are convex, their intersection with the affine one-dimensional space $\{(\xi_{0,s}^k, \bar{h}) - Tx : s \in \mathbb{R}\}$ is given either by an interval or by a single point. The latter can only appear if the intersection meets just the vertex or an edge (i.e., faces of dimension zero or one) of a polyhedral cone belonging to $\mathcal{K}(\xi_0^k)$. Hence, the subset of $\mathbb{R}^d$ that corresponds to such single points has Lebesgue measure 0 in $\mathbb{R}^d$. In case that the intersection is given by an interval $I_{\mathcal{K}_j}(\xi_0^k)$, we have due to (A5) that the interior of $I_{\mathcal{K}_j}(\xi_0^k)$, denoted $I^\circ_{\mathcal{K}_j}(\xi_0^k)$, contains only interior points of $\mathcal{K}_j$. This is true because otherwise the interval $I_{\mathcal{K}_j}(\xi_0^k)$ must lie in a facet of $\mathcal{K}_j$, and this would imply that there is a facet that is parallel to one of the canonical basis elements e_k, $1 \leq k \leq d$, in contradiction to (A5). This implies that we can partition the affine one-dimensional space $\{(\xi_{0,s}^k, \bar{h}) - Tx : s \in \mathbb{R}\}$ by considering the intervals (s_i, s_{i+1}), $0 \leq i \leq p(\xi_0^k)$, $s_0 = -\infty$ and $s_{p(\xi_0^k)+1} = +\infty$, such that

$$\{(\xi_{0,s}^k, \bar{h}) - Tx : s \in (s_i, s_{i+1})\} \subset \mathcal{K}^\circ_{j_i}.$$

Now, we are ready to state our first result on smoothness properties of $P_k f$.

Theorem 2 *Let $k \in D$ and $x \in X$. Assume (A1)–(A5) and let $f = f_x$ be the integrand (27) of the linear two-stage stochastic program (3). Then the kth projection $P_k f$ of f is continuously differentiable on $\mathbb{R}^{d-1}$. $P_k f$ is s-times continuously differentiable almost everywhere if the density ρ_k belongs to $C^{s-2}(\mathbb{R})$ for some $s \in \mathbb{N}$, $s \geq 2$.*

Proof We consider the two possible cases for open balls around ξ_0^k:
$P1$.) There exists $\epsilon_0 > 0$ such that $\mathcal{K}(\xi^k) = \mathcal{K}(\xi_0^k)$ for all $\xi^k \in \mathbb{B}_{\epsilon_0}(\xi_0^k)$.
$P2$.) For each $\epsilon > 0$ there exists $\xi^k \in \mathbb{B}_\epsilon(\xi_0^k)$ such that $\mathcal{K}(\xi^k) \subsetneq \mathcal{K}(\xi_0^k)$.

In case $P1$.), the functions s_i are differentiable in the entire neighborhood $\mathbb{B}_{\epsilon_0}(\xi_0^k)$, because they admit a representation as an affine function. Thus, we obtain from (29) for any $l \in \mathfrak{D}$, $l \neq k$, that $P_k f$ is partially differentiable with respect to ξ_l at ξ^k and

$$\begin{aligned}\frac{\partial P_k f}{\partial \xi_l}(\xi^k) &= \sum_{i=1}^{p+1} \frac{\partial}{\partial \xi_l} \int_{s_{i-1}}^{s_i} \langle v^{j_i}, (\xi^k_s, \bar{h}) - Tx\rangle \rho_k(s) ds \\ &= \sum_{i=1}^{p+1} \int_{s_{i-1}}^{s_i} v_l^{j_i} \rho_k(s) ds + \sum_{i=1}^{p} \langle v^{j_i}, (\xi^k_{s_i}, \bar{h}) - Tx\rangle \rho_k(s_i) \frac{\partial s_i}{\partial \xi_l} \\ &\quad - \sum_{i=2}^{p+1} \langle v^{j_i}, (\xi^k_{s_{i-1}}, \bar{h}) - Tx\rangle \rho_k(s_{i-1}) \frac{\partial s_{i-1}}{\partial \xi_l} \\ &= \sum_{i=1}^{p+1} v_l^{j_i} \int_{s_{i-1}}^{s_i} \rho_k(s) ds = \sum_{i=1}^{p+1} v_l^{j_i} (\varphi_k(s_i) - \varphi_k(s_{i-1})),\end{aligned}$$

where we used the identity $\langle v^{j_i}, (\xi^k_{s_i}, \bar{h}) - Tx\rangle = \langle v^{j_{i+1}}, (\xi^k_{s_i}, \bar{h}) - Tx\rangle$ for each $i = 1, \dots, p$ and φ_k denotes the marginal distribution function with density ρ_k. By reordering the latter sum we have

$$\frac{\partial P_k f}{\partial \xi_l}(\xi^k) = -\sum_{i=1}^{p} w_l^i \varphi_k(s_i) + v_l^{j_{p+1}} \tag{34}$$

Hence, the behavior of all first order partial derivatives of $P_k f$ only depends on the kth marginal distribution function φ_k. The latter are again differentiable and it follows for $r \in \mathfrak{D}$, $r \neq k$,

$$\frac{\partial^2 P_k f}{\partial \xi_r \partial \xi_l}(\xi^k) = \sum_{i=1}^{p} \frac{w_l^i w_r^i}{w_k^i} \rho_k(s_i). \tag{35}$$

Hence, $P_k f$ is second order continuously differentiable on the neighborhood $\mathbb{B}_{\epsilon_0}(\xi_0^k)$. More generally, if $\rho_k \in C^{s-2}(\mathbb{R})$ for some $s \in \mathbb{N}$, $s \geq 2$, $P_k f$ is s-times continuously differentiable on the neighborhood $\mathbb{B}_{\epsilon_0}(\xi_0^k)$.
In case $P2.$), we use the identity (33) and consider all cones belonging to $\mathcal{K}(\xi_0^k)$. Let $\mathcal{K}_i$, $i = 1, \dots, p+1$, denote all such cones which are normal cones to the vertices v^{j_i} of $\mathcal{D}$. Furthermore, let s_i, $i = 1, \dots, p$, be nondecreasing and defined by

$$(\xi^k_{s_i}, \bar{h}) - Tx \in \mathcal{K}_i \cap \mathcal{K}_{i+1}$$

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

$$P_k f(\xi_0^k) = \sum_{i=1}^{p+1} \int_{s_{i-1}}^{s_i} \langle v^{j_i}, (\xi^k_{0,s}, \bar{h}) - Tx\rangle \rho_k(s) ds,$$

where $p = p(\xi_0^k)$ and $s_i = s_i(\xi_0^k, x)$, $i = 1, \dots, p$, is given by (30).
Now, let $\xi^k \in \mathbb{B}_{\epsilon_0}(\xi_0^k)$. Due to (33) $P_k f(\xi^k)$ may be represented by a subset of the set $\mathcal{K}(\xi_0^k)$. Of course, $\mathcal{K}_1$ and $\mathcal{K}_{p+1}$ and all cones $\mathcal{K}_i$ such that $s_{i-1}(\xi_0^k, x) < s_i(\xi_0^k, x)$ appear also in the representation of $P_k f(\xi^k)$. Those cones $\mathcal{K}_i$ with $s_{i-1}(\xi_0^k, x) = s_i(\xi_0^k, x)$ may either disappear or appear with $s_{i-1}(\xi^k, x) < s_i(\xi^k, x)$. If they disappear we set $s_{i-1}(\xi^k, x) = s_i(\xi^k, x)$ and include them formally into the representation which is of the form

$$P_k f(\xi^k) = \sum_{i=1}^{p+1} \int_{s_{i-1}(\xi^k, x)}^{s_i(\xi^k, x)} \langle v^{j_i}, (\xi^k_s, \bar{h}) - Tx\rangle \rho_k(s) ds.$$

In a small ball around ξ^k this representation doesn't change. Hence, $P_k f$ is differentiable also in case $P2.)$ and the partial derivative is of the form

$$\frac{\partial P_k f}{\partial \xi_l}(\xi^k) = \sum_{i=1}^{p+1} v_l^{j_i} \int_{s_{i-1}(\xi^k,x)}^{s_i(\xi^k,x)} \rho_k(s)ds \tag{36}$$

as in case $P1.)$ and, hence, continuous. The integrals with $s_{i-1}(\xi^k, x) = s_i(\xi^k, x)$ are again formally included into (36). The second order partial derivative at ξ^k

$$\frac{\partial^2 P_k f}{\partial \xi_r \partial \xi_l}(\xi^k)$$

for $r \in \mathfrak{D}$, $r \neq k$, is, however, not of the form (35) in general since summands are missing that correspond to integrals with $s_{i-1}(\xi^k, x) = s_i(\xi^k, x)$. Moreover, note that the points $\xi_0^k \in \mathbb{R}^{d-1}$ with the property $s_{i-1}(\xi_0^k, x) = s_i(\xi_0^k, x)$ and $s_{i-1}(\xi^k, x) < s_i(\xi^k, x)$ for some ξ^k in $\mathbb{B}_{\epsilon_0}(\xi_0^k)$, $\xi^k \neq \xi_0^k$, and $i \in \{1, \ldots, p\}$, belong to a solution set of linear (see (30)) equations in $d-1$ variables of the form $s_{i-1}(\omega^k, x) = s_i(\omega^k, x)$ with s_{i-1} and s_i being not parallel each other, $i \in \{1, \ldots, p\}$. Thus, those singular points $\xi_0^k \in \mathbb{R}^{d-1}$, where the second order partial derivatives may not exist, belong to a subset of a finite union of hyperplanes in $\mathbb{R}^{d-1}$, each one of dimension less than or equal to $d-2$, and thus they are a measure-zero subset of $\mathbb{R}^{d-1}$. Hence, $P_k f$ is continuously differentiable on $\mathbb{R}^{d-1}$, but the mixed second order partial derivatives may exist only almost everywhere on $\mathbb{R}^{d-1}$. □

Corollary 1 *Let $\emptyset \neq u \subseteq \mathfrak{D}$ and $x \in X$. Assume (A1)–(A5). Then the projection $P_u f$ is continuously differentiable on $\mathbb{R}^{d-|u|}$ and second order continuously differentiable almost everywhere in $\mathbb{R}^{d-|u|}$.*

Proof If $|u| = 1$ the result follows from Theorem 2. For $u = \{k, j\}$ with $k, j \in \mathfrak{D}$, $k \neq j$, we obtain from the Leibniz theorem [15, Theorem 1] for $l \notin u$ and $r \notin u$

$$D_l P_u f(\xi^u) := \frac{\partial}{\partial \xi_l} P_u f(\xi^u) = P_j \frac{\partial}{\partial \xi_l} P_k f(\xi^u)$$
$$D_r D_l P_u f(\xi^u) := \frac{\partial^2}{\partial \xi_l \partial \xi_r} P_u f(\xi^u) = P_j \frac{\partial^2}{\partial \xi_l \partial \xi_r} P_k f(\xi^u)$$

if the derivatives exist, and since integration by P_j preserves the smoothness (see [15, Theorem 2]), the claim is proved. Moreover, from the proof of Theorem 2 we obtain the explicit form

$$D_l P_u f(\xi^u) = -\sum_{i=1}^{p} w_l^i \int_{\mathbb{R}} \varphi_k(s_i(\xi^k))\rho_j(\xi_j)d\xi_j + v_l^{j_{p+1}}, \text{ for all } \xi^u \text{ in } \mathbb{R}^{d-|u|} \tag{37}$$

$$D_r D_l P_u f(\xi^u) = \sum_{i=1}^{p} \frac{w_l^i w_r^i}{w_k^i} \int_{\mathbb{R}} \rho_k(s_i(\xi^k))\rho_j(\xi_j)d\xi_j, \ \ a.e. \text{ in } \mathbb{R}^{d-|u|}. \tag{38}$$

If u contains more than two elements, the integrals on the right-hand side become multiple integrals. In all cases, however, such an integral is a continuous function of the remaining variables ξ_i, $i \in \mathfrak{D} \setminus u$. This can be shown using Lebesgue's theorem as φ_k and ρ_k are continuous and bounded on $\mathbb{R}$. □

The following is the main result of this section.

Theorem 3 *Assume (A1)–(A5). Then all ANOVA terms of f except the one of highest order are first order continuously differentiable on $\mathbb{R}^d$ and all second order partial derivatives exist and are continuous except in 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 mixed Sobolev space $\mathcal{W}^{(1,\ldots,1)}_{2,\rho,\mathrm{mix}}(\mathbb{R}^d)$.*

Proof According to (20) 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

$$\begin{aligned} D_l f_{\{l,r\}}(\xi_l, \xi_r) &= D_l P_{D\setminus\{l,r\}} f(\xi_l, \xi_r) - D_l P_{D\setminus\{l\}} f(\xi_l) \\ &= -\sum_{i=1}^{p} w_l^i \int_{\mathbb{R}} \varphi_k(s_i(\xi^k)) \prod_{\substack{i \in D\setminus\{l,r\} \\ i \neq k}} \rho_i(\xi_i) d\xi^{-\{l,r\}} - D_l P_{D\setminus\{l\}} f(\xi_l) \end{aligned}$$

for all $l, r \in \mathfrak{D}$ and some $k \in \mathfrak{D}$. Since φ_k is Lipschitz continuous, the functions $D_l f_{\{l,r\}}$ and $D_r f_{\{l,r\}}$ are 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, [10, Section 4.2.3]). Furthermore, the second order partial derivative is almost everywhere bounded (see also (35)) and due to (A3) quadratically integrable with respect to ρ. □

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

$$f^{(2)} := \sum_{\substack{|u|=1 \\ u \subseteq D}}^{2} f_u \qquad (39)$$

belongs to the tensor product Sobolev space $\mathcal{W}^{(1,\ldots,1)}_{2,\mathrm{mix}}(\mathbb{R}^d)$. Hence, if the effective superposition dimension is at most 2, $f^{(2)}$ is a good approximation of f due to (26) and a favorable behavior of randomly shifted lattice rules may be expected.

The following two examples provide some more insight into the smoothness of the first order projections. The first example verifies that some projection is not continuously differentiable if (A5) is violated while the second example shows that conditions (A1)–(A5) do not imply second order partial differentiability of all projections.

Example 3 Let $\bar{m} = 3$, $d = 2$, $\Xi = \mathbb{R}^2$, P denote a probability distribution with independent marginal densities ρ_i, $i = 1, 2$, whose means are w.l.o.g. equal to 0. We assume that (A3) is satisfied for P. Let the vector q and matrix W

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

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

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

i.e., $\mathcal{D}$ is a triangle and has the three vertices

$$v^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v^2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad v^3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence, the second component of the two adjacent vertices v^1 and v^2 coincides. Ac-

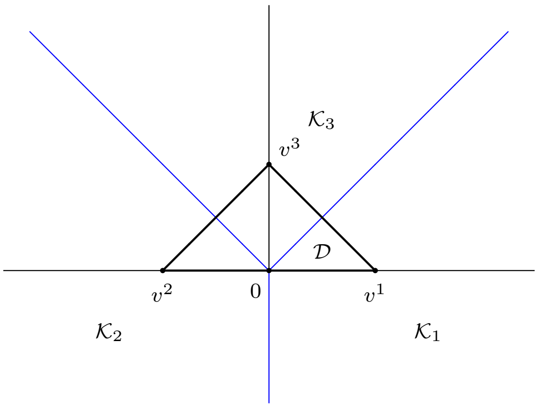

Fig. 1 Illustration of $\mathcal{D}$, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

cording to (13) the normal cones $\mathcal{K}_j$ to $\mathcal{D}$ at v^j, $j = 1, 2, 3$, are

$$\mathcal{K}_1 = \{z \in \mathbb{R}^2 : z_1 \ge 0, z_2 \le z_1\}, \quad \mathcal{K}_2 = \{z \in \mathbb{R}^2 : z_1 \le 0, z_2 \le -z_1\},$$
$$\mathcal{K}_3 = \{z \in \mathbb{R}^2 : z_2 \ge z_1, z_2 \ge -z_1\}.$$

The function ϕ (see (11)) and the integrand are of the form

$$\phi(t) = \max_{i=1,2,3} \langle v^i, t \rangle = \max\{t_1, -t_1, t_2\} = \max\{|t_1|, t_2\}$$
$$f(x, \xi) = \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}.$$

The ANOVA projection $P_1 f$ is defined by

$$(P_1 f)(\xi_2) = \int_{-\infty}^{+\infty} \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\} \rho_1(\xi_1) d\xi_1 \quad (\xi_2 \in \mathbb{R}).$$

For $\xi_2 - [Tx]_2 \le 0$ one obtains

$$\begin{aligned}(P_1 f)(\xi_2) &= \int_{-\infty}^{+\infty} |\xi_1 - [Tx]_1| \rho_1(\xi_1) d\xi_1 \\ &= \int_{-\infty}^{+\infty} (\xi_1 - [Tx]_1) \rho_1(\xi_1) d\xi_1 - 2 \int_{-\infty}^{[Tx]_1} (\xi_1 - [Tx]_1) \rho_1(\xi_1) d\xi_1\end{aligned}$$

and in case $\xi_2 - [Tx]_2 \geq 0$

$$(P_1 f)(\xi_2) = \int_{-\infty}^{+\infty} |\xi_1 - [Tx]_1| \rho_1(\xi_1) d\xi_1 - \int_0^{\xi_2 - [Tx]_2} (\xi_1 + \xi_2 - [Tx]_1 - [Tx]_2) \rho_1(\xi_1) d\xi_1.$$

Hence, $P_1 f$ belongs to $C^1(\mathbb{R})$ for all $x \in X$ if ρ is continuous.
When calculating the ANOVA projection $P_2 f$, notice that assumption (A5) is violated. We obtain

$$(P_2 f)(\xi_1) = |\xi_1 - [Tx]_1| \int_{-\infty}^{|\xi_1 - [Tx]_1|} \rho_2(\xi_2) d\xi_2 + \int_{|\xi_1 - [Tx]_1|}^{+\infty} (\xi_2 - [Tx]_2) \rho_2(\xi_2) d\xi_2$$

and $P_2 f$ *does not belong to* $C^1(\mathbb{R})$ for all $x \in X$.

Example 4 Let $\bar{m} = 3$, $d = 2$, P denote a two-dimensional probability distribution with independent continuous marginal densities ρ_i, $i = 1, 2$, whose means are w.l.o.g. equal to 0. Again we assume that (A3) is satisfied for P. Let the vector q and matrix W

$$W = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix} \qquad q = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

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

$$\mathcal{D} = \{z \in \mathbb{R}^2 : W^\top z \leq q\} = \{z \in \mathbb{R}^2 : -z_1 + z_2 \leq 1, z_1 + z_2 \leq 1, -z_1 - 3z_2 \leq 1\},$$

i.e., $\mathcal{D}$ is also a triangle and has the three vertices

$$v^1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \quad v^2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad v^3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence, both components of the vertices v^j, $j = 1, 2, 3$, are distinct. This means that (A4) and (A5) are satisfied. The normal cones $\mathcal{K}_j$ to $\mathcal{D}$ at v^j, $j = 1, 2, 3$, are

$$\mathcal{K}_1 = \{z \in \mathbb{R}^2 : z_2 \leq z_1, z_2 \leq 3z_1\}, \quad \mathcal{K}_2 = \{z \in \mathbb{R}^2 : z_2 \geq 3z_1, z_2 \leq -z_1\},$$
$$\mathcal{K}_3 = \{z \in \mathbb{R}^2 : z_2 \geq z_1, z_2 \geq -z_1\}.$$

The function ϕ in (11) and the integrand are of the form

$$\phi(t) = \max_{i=1,2,3} \langle v^i, t \rangle = \max\{2t_1 - t_2, -t_1, t_2\}$$
$$f(x, \xi) = \phi(\xi_1 - [Tx]_1, \xi_2 - [Tx]_2)$$

The ANOVA projection $P_1 f$ is given by

$$(P_1 f)(\xi_2) = \int_{-\infty}^{+\infty} \max\{2(s - [Tx]_1) - \xi_2 + [Tx]_2, -s + [Tx]_1, \xi_2 - [Tx]_2\} \rho_1(s) ds$$

for every $\xi_2 \in \mathbb{R}$. For simplicity let $x = 0$. First let $\xi_2 > 0$.

$$(P_1 f)(\xi_2) = \int_{-\infty}^{+\infty} \max\{2s - \xi_2, -s, \xi_2\} \rho_1(s) ds$$
$$= \int_{-\infty}^{s_1} -s \rho_1(s) ds + \int_{s_1}^{s_2} \xi_2 \rho_1(s) ds + \int_{s_2}^{+\infty} (2s - \xi_2) \rho_1(s) ds,$$

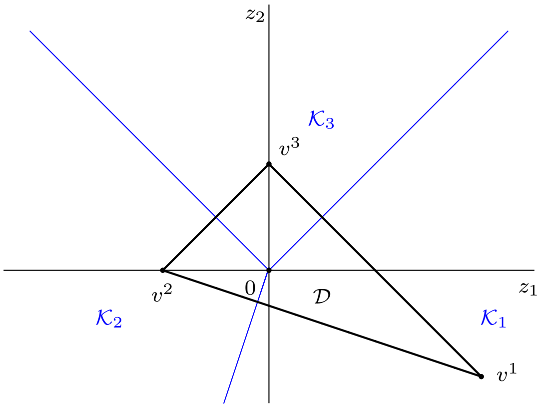

Fig. 2 Illustration of $\mathcal{D}$, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

where $s_1 = s_1(\xi_2) = -\xi_2$ and $s_2 = s_2(\xi_2) = \xi_2$. Hence,

$$(P_1 f)(\xi_2) = 3\int_{\xi_2}^{+\infty} s\rho_1(s)ds + \xi_2\Big(\int_{-\xi_2}^{\xi_2} \rho_1(s)ds - \int_{\xi_2}^{+\infty} \rho_1(s)ds\Big).$$

Now, we compute the partial derivatives for $\xi_2 > 0$ and obtain

$$\frac{\partial P_1 f}{\partial \xi_2}(\xi_2) = -\xi_2(\rho_1(\xi_2) - \rho_1(-\xi_2)) + (2\varphi_1(\xi_2) - \varphi_1(-\xi_2) - 1)$$
$$\frac{\partial P_1 f}{\partial \xi_2}(0+) = \varphi_1(0) - 1$$
$$\frac{\partial^2 P_1 f}{\partial \xi_2^2}(\xi_2) = 2\rho_1(\xi_2) + \rho_1(-\xi_2) \;=\; 3\rho_1(\xi_2)$$

$P_1 f$ is for $\xi_2 > 0$ s-times continuously differentiable if $\rho_1 \in C^{s-2}(\mathbb{R})$ for any $s \in \mathbb{N}$. Now, let $\xi_2 < 0$. Then we obtain with $s_1(\xi_2) = \frac{\xi_2}{3}$

$$(P_1 f)(\xi_2) = \int_{-\infty}^{+\infty} \max\{2s - \xi_2, -s, \xi_2\}\rho_1(s)ds$$
$$= \int_{-\infty}^{s_1} -s\rho_1(s)ds + \int_{s_1}^{+\infty} (2s - \xi_2)\rho_1(s)ds$$
$$\frac{\partial P_1 f}{\partial \xi_2}(\xi_2) = -\tfrac{\xi_2}{3}\rho_1(\tfrac{\xi_2}{3}) + (\varphi_1(\tfrac{\xi_2}{3}) - 1) + \tfrac{\xi_2}{3}\rho_1(\tfrac{\xi_2}{3}) = \varphi_1(\tfrac{\xi_2}{3}) - 1$$
$$\frac{\partial P_1 f}{\partial \xi_2}(0-) = \varphi_1(0) - 1$$
$$\frac{\partial^2 P_1 f}{\partial \xi_2^2}(\xi_2) = \tfrac{1}{3}\rho_1(\tfrac{\xi_2}{3})$$

$P_1 f$ is for $\xi_2 < 0$ s-times continuously differentiable if $\rho_1 \in C^{s-2}(\mathbb{R})$ for any $s \in \mathbb{N}$. Hence, $P_1 f$ *belongs to* $C^1(\mathbb{R})$, but its second derivative is discontinuous at $\xi_2 = 0$. The same holds for $P_2 f$.

Remark 2 (error estimate)
If the assumptions of Theorem 3 are satisfied and all marginal densities ρ_j, $j \in \mathfrak{D}$, are positive and continuously differentiable, all ANOVA terms g_u, $|u| = 1$, 2, of g given by (23) are continuously differentiable on $\mathbb{R}^d$ and the second order partial derivatives exist in the sense of Sobolev. However, as observed already in [53], the partial derivatives are not quadratically integrable and, hence, g_u does not belong to the tensor product Sobolev space (5). Meanwhile, a theory was developed in [26] in which a suitable weight function is incorporated into the integrals defining the inner product (6) and into the kernel to remedy this issue. In this way, the same convergence rate (8) is obtained in [26] for a number of probability distributions (including the normal distribution).
In order not to complicate the argument with technical details we assume here for simplicity that the ANOVA terms g_u, $|u| = 1$, 2, belong to (5). Then the QMC quadrature error may be estimated as follows:

$$\begin{aligned}\Big|\int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi - n^{-1}\sum_{j=1}^n f(\xi^j)\Big| &= \Big|\int_{[0,1]^d} g(x)dx - n^{-1}\sum_{j=1}^n g(x^j)\Big| \\ &\le \sum_{0<|u|\le d}\Big|\int_{[0,1]^d} g_u(x^u)dx^u - n^{-1}\sum_{j=1}^n g_u(x^j)\Big| \\ &\le \sum_{|u|=1}^{2} \mathrm{Disc}_{n,u}(x^1,\dots,x^n)\|g_u\|_\gamma + \qquad (40) \\ &\quad \sum_{|u|=3}^{d}\Big|\int_{[0,1]^d} g_u(x)dx - n^{-1}\sum_{j=1}^n g_u(x^j)\Big|, \qquad (41)\end{aligned}$$

where $x_i^j = \varphi_i(\xi_i^j) \in (0,1)^d$, $j = 1,\dots,n$, $i = 1,\dots,d$, are the QMC points and $\mathrm{Disc}_{n,u}$ is the weighted L_2- discrepancy

$$\mathrm{Disc}^2_{n,u}(x^1,\dots,x^n) = \gamma_u \int_{[0,1]^{|u|}} \mathrm{disc}_u^2(x^u)dx^u,$$

where the discrepancy disc is given by

$$\mathrm{disc}_u(x^u) = \prod_{i\in u} x_i - n^{-1}\big|\{j \in \{1,\dots,n\} : x^j \in [0,x^u)\}\big|,$$

and $\|g_u\|_\gamma$ the weighted norm of g_u given by (6) in the weighted tensor product Sobolev space (5). Recalling the arguments in the introduction one may conclude that all terms in (40) converge with the optimal rate (8) while all terms in (41) also converge to 0 due to Proinov's convergence result [47] (as the g_u are continuous). In addition, the sum (41) can be further estimated by

$$\sum_{|u|=3}^{d}\Big(\int_{[0,1]^d} g_u^2(x)dx + n^{-1}\sum_{j=1}^n g_u^2(x^j)\Big)^{\frac{1}{2}} = \sum_{|u|=3}^{d}\Big(\|f_u\|^2_{L_2} + n^{-1}\sum_{j=1}^n f_u^2(\xi^j)\Big)^{\frac{1}{2}}. \quad (42)$$

Since (26) implies $\sum_{|u|=3}^d \|f_u\|^2_{L_2} \le \varepsilon\sigma^2(f)$ if $d_S(\varepsilon) \le 2$ and the second term on the right-hand side of (42) represents a QMC approximation of the first term, we may conclude that the term in (41) is of the form $O(\sqrt{\varepsilon})$. Hence, we obtain the estimate

$$\Big|\int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi - n^{-1}\sum_{j=1}^n f(\xi^j)\Big| \le C(\delta)n^{-1+\delta} + O(\sqrt{\varepsilon}) \quad (43)$$

if the condition $d_S(\varepsilon) \leq 2$ is satisfied. The latter may eventually be achieved by applying dimension reduction techniques (see Section 7).
Moreover, when recalling the results in [67], one may hope that the convergence rate for the terms in (40) is even better.
Finally, we note that the constants involved in the estimate (43) may be chosen to be uniform with respect to $x \in X$. Together with the perturbation estimates (17) and (18) in Section 3 one, hence, obtains

$$|v(P) - v(P_n)| \leq \hat{C}(\delta)n^{-1+\delta} + O(\sqrt{\varepsilon}),$$
$$S(P_n) \subseteq S(P) + \psi_P^{-1}(\hat{C}(\delta)n^{-1+\delta} + O(\sqrt{\varepsilon}))$$

if $d_S(\varepsilon) \leq 2$. Here, P_n is the discrete probability measure representing the 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 Orthogonal transformations and the Gaussian case

We consider the stochastic program (3) with

$$\Phi(x, \xi) = \phi(h(\xi) - Tx)$$

and the infimal function (11) of the second stage problem given by (14). Assume that (A1)–(A3) is satisfied. Further we assume that ξ is of the form $\xi = Q\zeta$ with some orthogonal (d, d) matrix Q and with ζ satisfying (A4). Then the relevant integrand is of the form

$$f(\xi) = \max_{j=1,\ldots,\ell} \langle v^j, (Q\zeta, \bar{h}) - Tx\rangle = \max_{j=1,\ldots,\ell} \langle \hat{Q}^\top v^j, (\zeta, \bar{h}) - \hat{Q}^\top Tx\rangle,$$

where the (r, r) matrix $\hat{Q}$ is defined as in (16) with B replaced by Q. Hence, the results of Section 5 apply if the vertices $\hat{Q}^\top v^j$, $j = 1, \ldots, \ell$, of the linearly transformed dual feasible set $\hat{Q}^\top \mathcal{D}$ satisfy the corresponding assumption (A5). The set $\hat{Q}^\top \mathcal{D}$ may be represented in the form

$$\hat{Q}^\top \mathcal{D} = \{\hat{Q}^\top z : W^\top z \leq q\} = \{z \in \mathbb{R}^r : (\hat{Q}^\top W)^\top z \leq q\}.$$

The geometric condition (A5) is violated only if some facet of $\hat{Q}^\top \mathcal{D}$ is parallel to some of the first d coordinate axis. To study the set of orthogonal (d, d) matrices Q such that the latter property is true, we consider the set $\mathcal{M}$ of all nonsingular (d, d) matrices as elements of $\mathbb{R}^{d^2}$ and the subset $\mathcal{M}_o$ of all orthogonal (d, d) matrices. We recall that $\mathcal{M}_o$ is a compact subset of $\mathcal{M}$ and is of dimension $l(d) = \frac{1}{2}d(d-1)$. Hence, $\mathcal{M}_o$ can be considered as a subset of $\mathbb{R}^{l(d)}$ which is equipped with the Lebesue measure $\lambda^{l(d)}$. The (d, d) matrices Q cause rotations of $\mathbb{R}^d$ and, thus, changes of the first d variables of the elements of $\mathcal{D}$. If all d variables change, i.e., if Q causes a full-dimensional rotation, then there exist only countably many matrices Q such that (A5) is violated for $\hat{Q}\mathcal{D}$. All orthogonal matrices Q such that less than d variables change, form a subset of $\mathcal{M}_o$ of dimension smaller than $l(d)$. In both situations the set of all orthogonal matrices $Q \in \mathcal{M}_o$ such that (A5) is violated for $\hat{Q}\mathcal{D}$ has Lebesgue measure $\lambda^{l(d)}$ zero. Hence, we may say that (A5) is violated for almost every orthogonal matrix Q if ξ is given in the form $\xi = Q\zeta$ with ζ satisfying (A4). We note that this argument applies to the principal component analysis (PCA) decomposition of a normal random vector with nonsingular covariance matrix (see Section 7).

Corollary 2 *Let $x \in X$ be fixed and assume (A1), (A2) and that ξ is normal with nonsingular covariance matrix. For almost every covariance matrix the ANOVA approximation $f^{(2)}$ of f given by (39) belongs to the mixed Sobolev space $\mathcal{W}_{2,\rho,\mathrm{mix}}^{(1,\ldots,1)}(\mathbb{R}^d)$.*

Proof First we note that normal random vectors satisfy (A3). Now, let λ_j denote the positive eigenvalues and u_j, $j = 1, \ldots, d$, the corresponding orthogonal eigenvectors of Σ. Then it holds $\Sigma = Q\,D\,Q^\top$, where $Q = (u_1 \cdots u_d)$ and D is the diagonal matrix with $\lambda_1, \ldots, \lambda_d$ in the main diagonal. Hence, $\zeta = Q^\top \xi$ is normal with covariance matrix D, thus, satifies (A4) and we may apply the result derived before stating the corollary. The result then follows from Theorem 3. □

7 Sensitivity and dimension reduction of two-stage stochastic programs

In this section we discuss sensitivity and possibilities for reducing the effective dimension of two-stage models. First, we derive an upper bound for the global sensitivity indices $\bar{S}_{\{i\}}$, $i = 1, \ldots, d$, and the mean dimension $\bar{d}_S$ in the superposition sense, respectively.

Proposition 4 *Let (A1)–(A4) with $h(\xi) = (\xi, \bar{h})$ with fixed $\bar{h} \in \mathbb{R}^{r-d}$ be satisfied and σ_i^2 denote the variance of ξ_i, $i = 1, \ldots, d$. Then*

$$\bar{S}_{\{i\}} \leq \frac{\sigma_i^2}{\sigma^2(f)} \max_{j=1,\ldots,\ell} |v_i^j|^2 \quad (i = 1, \ldots, d)$$

$$\bar{d}_S \leq \frac{1}{\sigma^2(f)} \max_{j=1,\ldots,\ell} \|v^j\|_\infty^2 \sum_{i=1}^d \sigma_i^2,$$

where v^j, $j = 1, \ldots, \ell$, are the vertices of the dual polyhedron.

Proof We use [58, Theorem 3] and compute the partial derivatives of f with respect to ξ_i, $i = 1, \ldots, d$, which exist almost everywhere on $\mathbb{R}^d$. If $h(\xi) - Tx$ belongs to the cone $\mathcal{K}_j$, then

$$f(\xi) = \sum_{i=1}^d v_i^j (\xi_i - [Tx]_i) + \sum_{i=d+1}^r v_i^j (\bar{h}_i - [Tx]_i),$$

where $x \in X$ is fixed. We obtain for $\xi \in \mathbb{R}^d$ such that $h(\xi) - Tx$ belongs to the interior of $\mathcal{K}_j$ that

$$\frac{\partial f}{\partial \xi_i} = v_i^j.$$

Hence, the partial derivative is piecewise constant and may be bounded from above by $\max_{j=1,\ldots,\ell} |v_i^j|$. Using [58, Theorem 3] this proves our estimate for the global sensitivity index $\bar{S}_{\{i\}}$. The second estimate is a consequence of formula (25). □

Proposition 4 provides an estimate for the importance of variable i on f and sheds light on the role of σ_i.

If ξ is normal with nonsingular covariance matrix Σ, 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 and, hence, is not suitable to reduce the effective

dimension (at least in the truncation sense). This fact is confirmed in our numerical experiments (see Section 8).
Several *dimension reduction techniques* exploit the fact that a normal random vector ξ with mean μ and 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. As observed in [43,68], 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.
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 mean 0 and covariance matrix I. The solution is $v_i = \sqrt{\lambda_i}u_i$ and $z_i = (\sqrt{\lambda_i})^{-1}u_i^\top\xi$, where $\lambda_1 \geq \cdots \geq \lambda_d > 0$ are the eigenvalues of Σ in decreasing order and u_i, $i = 1,\ldots,d$, the corresponding orthonormal eigenvectors (see [68]). Hence, PCA consists in using the factorization

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

where $U_P = (\sqrt{\lambda_1}u_1,\ldots,\sqrt{\lambda_d}u_d)$ and $Q_P = (u_1,\ldots,u_d)$. Several authors report an enormous reduction of the effective truncation dimension in financial models if PCA is used (see, for example, [64–66]). We observed the same effect in our numerical experiments (see Section 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.
One such idea is based on the following observation which is seemingly due to [43], too (see also [68, Lemma 1]).

Proposition 5 *Let Σ be a $d \times d$ nonsingular covariance matrix and A be a fixed $d \times d$ matrix such that $A\,A^\top = \Sigma$. Then $\Sigma = B\,B^\top$ if and only if B is of the form $B = A\,Q$ for some orthogonal $d \times d$ matrix Q.*

This observation is used in the dimension reduction approach for (log)normal random vectors ξ proposed in [21] and extended in [68]. For general (non-normal) random vectors ξ the influence (of groups) of variables and the computation of effective dimensions are studied, e.g., in [7,57,58,64].

8 Numerical experiments

For our tests we consider a two-stage stochastic production planning problem which consists in minimizing costs of a company. The company aims to satisfy stochastic demands ξ_t in a time horizon $\{1,\ldots,T\}$ with multivariate probability distribution P (on $\mathbb{R}^T$), but its production capacity based on I company owned units does eventually not suffice to cover the demand. Hence, it has to buy the necessary amounts from other $m = m_1 + m_2$ providers or markets at fixed prices $\bar{c}_{1,j_1,t}$ and $\bar{c}_{2,j_2,t}$, $t = 1,\ldots,T, 1 \leq j_1 \leq m_1, 1 \leq j_2 \leq m_2$, and aims at minimizing the expected costs.

The optimization model is of the form

$$\min_{x\in\mathbb{R}^{IT}} \Big\{ \sum_{t=1}^{T}\sum_{i=1}^{I} c_{i,t}\, x_{i,t} + \int_{\mathbb{R}^T} \Phi(x,\xi) P(d\xi) \,:\, x \in X \Big\},$$

where the recourse costs Φ are given by

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

with the polyhedral constraint sets

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

and

$$Y(x,\xi) := \left\{ y \in \mathbb{R}^{mT} \,\middle|\, \begin{array}{l} \displaystyle\sum_{i=1}^{I} x_{i,t} + \sum_{j=1}^{m_1+m_2} y_{j,t} \ge \xi_t\,, \\ w_{1,j_1,t} \le y_{j_1,t} \le z_{1,j_1,t}\,, j_1 = 1,\dots,m_1 \\ w_{2,j_2,t} \le y_{m_1+j_2,t}\,, j_2 = 1,\dots,m_2 \\ (t = 1,\dots,T) \\ |y_{j_1,t} - y_{j_1,t+1}| \le \rho_{1,j_1,t}\,, j_1 = 1,\dots,m_1, \\ |y_{m_1+j_2,t} - y_{m_1+j_2,t+1}| \le \rho_{2,j_2,t}\,, j_2 = 1,\dots,m_2 \\ (t = 1,\dots,T-1) \end{array} \right\}$$

with fixed positive prices $c_{i,t}, \bar{c}_{1,j_1,t}, \bar{c}_{2,j_2,t}$ and bounds $a_{i,t},\ b_{i,t},\ \delta_{i,t},\ w_{1,j_1,t},\ w_{2,j_2,t},$ $z_{1,j_1,t},\ \rho_{1,j_1,t},\ \rho_{2,j_2,t}$. We assume that the demands ξ_t follow the condition

$$\xi_t = m_t + \eta_t, \quad \text{for } 1 \le t \le T, \tag{44}$$

where $m = (m_1,\dots,m_T)$ is a vector of expected values simulating the *trend* or *seasonality* of the demands, and η is an *ARMA(p,q) process* given by the recurrence equation

$$\eta_t = \sum_{i=1}^{p} \alpha_i \eta_{t-i} + \sum_{j=1}^{q} \beta_j \gamma_{t-j} + \gamma_t \quad (t \in \mathbb{Z}) \tag{45}$$

with i.i.d. Gaussian noise $\gamma_t \sim$ N(0,1) and characteristic polynomials $P(z) = 1 - \sum_{i=1}^{p} \alpha_i z^i$ and $Q(z) = 1 + \sum_{i=1}^{q} \beta_i z^i$. An ARMA(p,q) process is stationary (i.e., the covariance function $R(t,s) = \mathbb{E}(\eta_t\eta_s)$ is of the form $R(t,s) = \lambda(|t-s|+1)$, $1 \le t,s \le T$) iff the polynomials P and Q do not have common zeros and $P(z) \ne 0$ for all $z \in \mathbb{C}$ with $|z| \le 1$ (see [2, Chapter 3]).

The vector of demands $\xi_1,\dots,\xi_T$ is then normally distributed with mean vector m and covariance matrix dependending on the constants $\alpha_i,\ \beta_j,\ 1 \le i \le p,\ 1 \le j \le q,$ $p,q \in \mathbb{N}$. Such models have been considered for simulating electricity load demands in energy industry, see e.g. [44] and [9]. Note that since the model includes unbounded demands ξ, no upper bounds in the variables $y_{m_1+j_2,t},\ j_2 = 1,\dots,m_2,\ t = 1,\dots,T,$ were imposed, allowing to cover arbitrarily large demand values. We select in addition the prices $\bar{c}_{2,j_2,t}$ significantly higher than the prices $\bar{c}_{1,j_1,t}$, 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 choose the real dimension $d = T = 100$, and the model constants $p = 2$, $q = 6$, $\alpha_1 = -0.52$, $\alpha_2 = 0.45$, $\beta_1 = -0.17$, $\beta_2 = 0.12$, $\beta_3 = 0.05$, $\beta_4 = -0.07$, $\beta_5 = 0.06$, $\beta_6 = 0.04$. The resulting ARMA process η is stationary and, hence, $\mathbb{E}(\eta_t\eta_s) = \lambda(|t-s|+1)$, $1 \le t, s \le T$. The values $\lambda(t), 1 \le t \le T$, can be obtained by solving a system of linear equations with coefficients depending on the constants α_i, β_j, $1 \le i \le 2$, $1 \le j \le 6$ (see [2] for detailed information about modeling with ARMA processes). The resulting covariance matrix Σ is Toeplitz symmetric, with entry values $\Sigma(i,j) = \lambda(|i-j|+1)$. 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) (see Section 7). After the factorization of Σ assumptions (A1)–(A4) (see Section 3) are satisfied. Hence, Theorem 3 applies if (A5) is satisfied. A simulated demands-path $\xi_1, \ldots, \xi_d$ can then be obtained by

$$(\xi_1, \ldots, \xi_d)^\top = A\,(\phi^{-1}(z_1), \ldots, \phi^{-1}(z_d))^\top + (m_1, \ldots, m_d),$$

where $Z = (z_1, \ldots, z_d) \sim U([0,1]^d)$ (i.e., the probability distribution of Z is uniform on $[0,1]^d$), and $\phi^{-1}(.)$ represents the inverse cumulative 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 $Z \sim U([0,1]^d)$. For MC points in $[0,1]^d$ we used the Mersenne Twister [32] as pseudo random number generator. For QMC, we use randomly scrambled Sobol' points with direction numbers given in [22] and randomly shifted lattice rules [55, 25]. The implemented scrambling technique is random linear scrambling described in [31]. 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:

- $I = 10$, $m_1 = 6$, $m_2 = 2$.
- 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 $\bar{c}_{2,j_2,t} \in [12, 14]$.

The given parameters were chosen to attempt avoiding trivial solutions of the linear programs.

We perform two different kinds of tests in our experiments. For the first kind of tests we fix n sampling points ξ^j and replace the integral of the second stage function $\Phi(x, \cdot)$ by the equal weight MC and randomized QMC quadrature rule, respectively. Then we solve the resulting large linear program

$$\min_{x \in \mathbb{R}^{IT}} \Big\{ \sum_{t=1}^{T} \sum_{i=1}^{I} c_{i,t}\, x_{i,t} + \frac{1}{n} \sum_{j=1}^{n} \Phi(x, \xi^j) P(d\xi) \; : \; x \in X \Big\}. \tag{46}$$

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

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

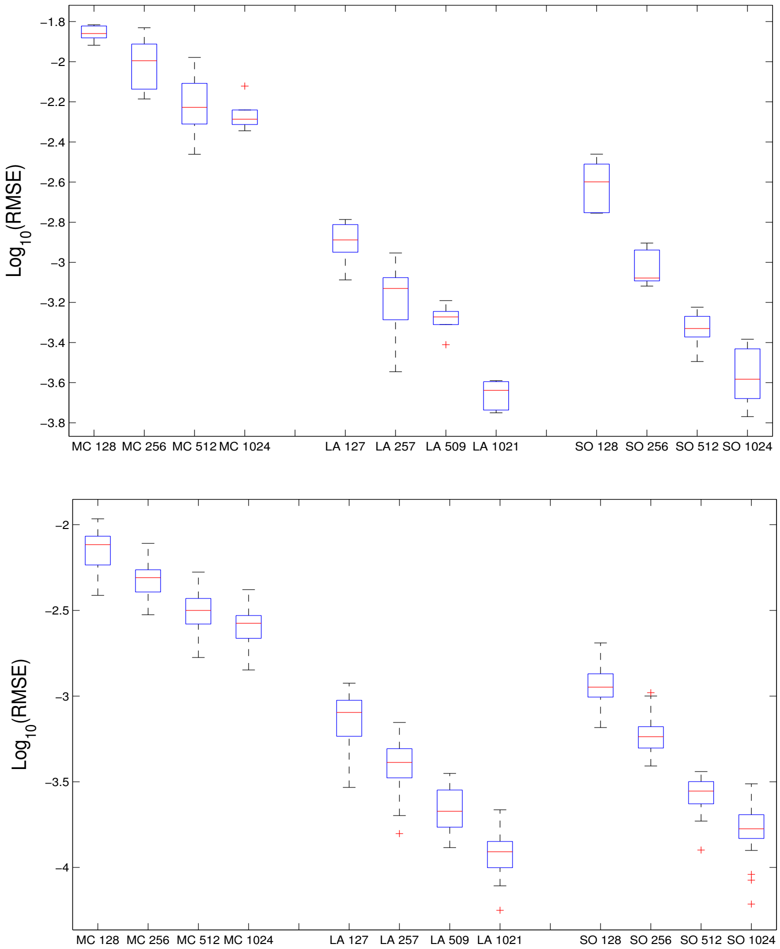

Fig. 3 Shown are the Log_{10} of relative RMSE with PCA factorization of covariance matrix for integrating $\Phi(x, \cdot)$ (upper figure) and for the minimum in (46) (lower figure). Results for Mersenne Twister MC and randomly scrambled Sobol' QMC with $128, 256, 512$ and 1024 points (MC 128,... or SO 128,...), and randomly shifted lattice rules QMC with $127, 257, 509$ and 1021 lattice points (LA 127,...).

by equal weight MC or randomized QMC quadrature rules. For simplicity we choose the fixed feasible points $x \in X$ to be the optimal solutions of the tests of the first kind, which were obtained by solving the resulting linear program for different costs while keeping the constraint set unchanged. The aim of these experiments is twofold. First we examine the convergence rate of the MC or randomized QMC quadrature rules with some fixed feasible points $x \in X$ for the expected recourse in the tests of second kind. Secondly we examine if these convergence rates in terms of sample sizes n are translated to the resulting large linear programs for the tests of first kind. The results for the tests of first and second kind under PCA factorization are summarized in Figure

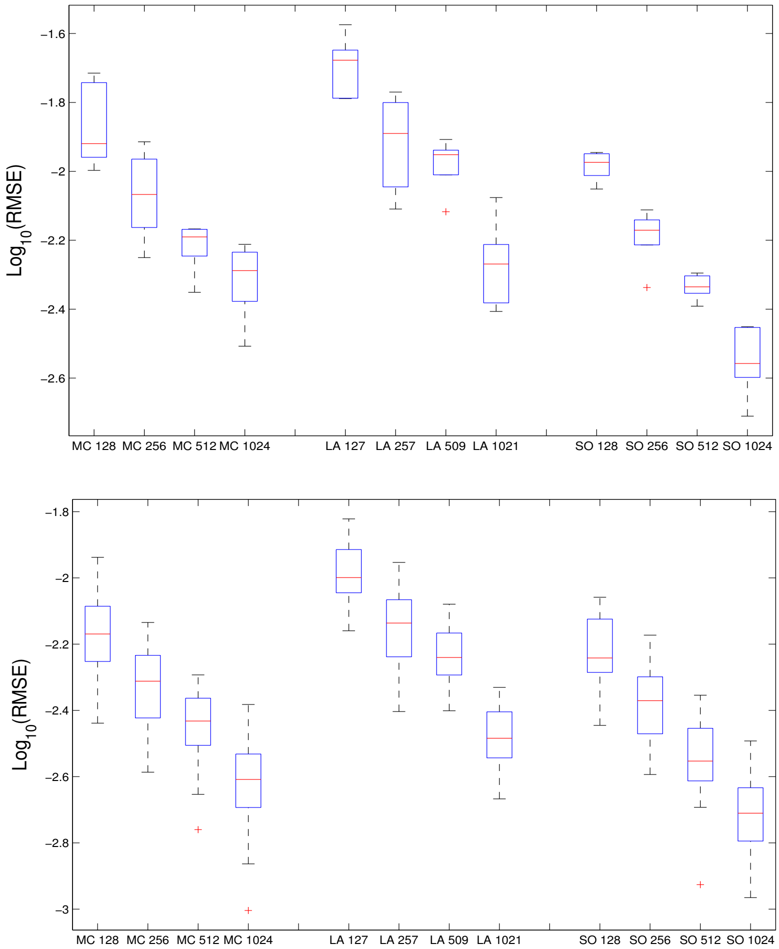

Fig. 4 Shown are the Log_{10} of relative RMSE with Cholesky factorization of covariance matrix for integration of $\Phi(x, \xi)$ (upper figure) and for the minimum in (46) (lower figure). Results for Mersenne Twister MC and randomly scrambled Sobol' QMC with $128, 256, 512$ and 1024 points (MC 128,... or SO 128,...), and randomly shifted lattice rules QMC with $127, 257, 509$ and 1021 lattice points (LA 127,...).

3. We chose $n = 128, 256, 512, 1024$ for the Mersenne Twister and for Sobol' points. For randomly shifted lattices, we chose the primes $n = 127, 257, 509, 1021$. The random shifts were generated using the Mersenne Twister. We estimate the relative root mean square errors (RMSE) of the estimated integrals (for the tests of the first kind) and of the optimal objective values (for the tests of the second kind) 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 first (lower bound of the box) and third quartiles (upper bound of the box), and the median (line between lower and upper bound). Outliers are marked by plus signs and the remaining results lie between the bounds.

The average of the estimated rates of convergence for both kind of tests under PCA ranged in $[-0.95, -0.85]$ for randomly shifted lattice rules, and in $[-1, -0.9]$ for 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 effective truncation dimension of $\Phi(x, \cdot)$ was tested at 20 different feasible vertices x (obtained from the tests of first kind with different price parameters and fixed bounds). We used the algorithm proposed in [64] 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 effective dimension d_T remained close to 2 in most cases and always ≤ 6. Further tests for the case $d_T = 6$ showed that the variance accumulated by the first order ANOVA terms $f_{\{i\}}$, $1 \leq i \leq 6$ was approximately 95% of the total variance. The first order ANOVA terms $f_{\{i\}}$, $7 \leq i \leq d$ accumulated in total approximately 0,5% of the total variance. Moreover, adding the variance of the ANOVA terms $f_{\{1,2\}}$ and $f_{\{1,3\}}$ to the variance of the terms $f_{\{i\}}$, $1 \leq i \leq 6$ resulted in a variance accumulation higher than 99%. Therefore we can conclude that the effective superposition dimension in case of using PCA is $d_S(0.01) = 2$. Intensive computations seem to show that we may have $d_S(\varepsilon) = 2$ for even smaller values of ε than 0.01. Hence, PCA serves as excellent dimension reduction technique in this case.
Although the geometric condition (A5) seems difficult to prove in this case (and maybe in many high-dimensional realistic examples encountered in energy industry), we may rely on Corollary 2 which states that the condition is satisfied for almost all covariance matrices except for countably many. Indeed, it seems that the recourse function $\Phi(x, \cdot)$ is well approximated by a low dimensional smooth function as is the case in many practical examples considered in finance (see [15]), for different feasible vertices $x \in X$. Further tests were carried out by combining randomly shifted lattice rules with the tent transformation as described in [17], but no improvements in the convergence rates beyond $O(n^{-1})$ were observed for our feasible range of sample sizes. Similarly no improvement beyond the rate $O(n^{-1})$ was observed for scrambled Sobol' sequences as might be expected for smooth integrands (see Section 2)). This may be explained by the lack of the required smoothness properties of the second order ANOVA approximation.
Using the Cholesky factorization, the results for both kind of tests were completely different than those under PCA. The average of the estimated rates of convergence of randomized QMC ranged in $[-0.6, -0.5]$, which is very close to the expected MC rate of -0.5. The results for the Cholesky factorization are presented in Figure 4. The effective truncation dimension of $\Phi(x, \cdot)$ was estimated to be equal to $d_T = 100$, which is just the real dimension d of the problem. Tests showed that the variance accumulated by the first order ANOVA terms $f_{\{i\}}$, $i \in \mathfrak{D}$ was approximately 20% of the total variance. It seems very likely that the effective superposition dimension for the case of using the Cholesky factorization is really high-dimensional.

9 Conclusions

Our theoretical results in Section 5 imply that all ANOVA terms except the one of highest order of integrands f appearing in linear two-stage stochastic programs are smoother than f. More precisely, the ANOVA terms of first and second order belong to a mixed Sobolev space which is important for optimal convergence rates of randomly shifted lattice rules. Error estimates as in Remark 2 then indicate that we may expect that Quasi-Monte Carlo approximations of two-stage stochastic programs converge with the optimal rate (8) even for high dimensions d if the effective superposition dimension satisfies $d_S \leq 2$. Since we estimate the effective truncation dimension d_T and it holds $d_S \leq d_T$, it is important that d_T is equal to or at least close to 2. This requires the use of dimension reduction techniques, for example, principal component analysis for (log)normal distributions P.

Our preliminary computational experience on applying Quasi-Monte Carlo methods to a two-stage stochastic production planning problem confirms the theoretical results. They show that using appropriate Quasi-Monte Carlo methods instead of Monte Carlo may lead to a *substantial improvement*, because one may work with a much smaller number n of scenarios if suitable dimension reduction techniques allow for an essential reduction from $d_T = d$ to d_T close to 2.

Altogether, there are good reasons to conclude that recent Quasi-Monte Carlo methods (like (scrambled) Sobol' sequences and randomly shifted lattice rules) *may be efficient* for two-stage linear stochastic programs (even if the programs are large scale) if they allow for a clear dimension reduction. However, our present theoretical results do not support the use of higher order QMC methods (see [4,5]) since the first and second order ANOVA terms do not seem to satisfy the required smoothness conditions.

Acknowledgements The authors wish to express their gratitude to Prof. Ian Sloan (University of New South Wales, Sydney) for inspiring conversations during his visit of the Humboldt-University Berlin in 2011. Much of the work on this paper was done during the first author held a position at the Humboldt-University Berlin. The research of the first author is partially supported by a grant of Kisters AG and by the Deutsche Forschungsgemeinschaft within SFB Transregio 154: Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks. The research of the second author is supported by a grant of the German Bundesministerium für Wirtschaft und Technologie (BMWi) and the third by the DFG Research Center MATHEON at Berlin. The authors extend their gratitude to two anonymus referees and to the Associate Editor for their constructive and stimulating criticism.

References

1. J. Baldeaux: *Higher order nets and sequences*, PhD Thesis, The University of New South Wales, 2010.
2. P. J. Brockwell and R. A. Davis: *Introduction to Time Series and Forecasting* (Second Edition), Springer, New York, 2002.
3. J. Dick, I. H. Sloan, X. Wang, and H. Woźniakowski: Liberating the weights, *Journal of Complexity* 20 (2004), 593–623.
4. J. Dick: Walsh spaces containing smooth functions and Quasi-Monte Carlo rules of arbitrary high order, *SIAM Journal Numerical Analysis* 46 (2008), 1519–1553.
5. J. Dick and F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, 2010.
6. J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 22 (2013), 133–288.
7. S. S. Drew and T. Homem-de-Mello: Quasi-Monte Carlo strategies for stochastic optimization, *Proceedings of the 2006 Winter Simulation Conference*, IEEE, 2006, 774–782.
8. R. M. Dudley: The speed of mean Glivenko-Cantelli convergence, *The Annals of Mathematical Statistics* 40 (1969), 40–50.
9. A. Eichhorn, W. Römisch and I. Wegner: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, IEEE St. Petersburg Power Tech 2005.
10. L. C. Evans and R. F. Gariepy: *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, 1992.
11. P. Glasserman: *Monte-Carlo Methods in Financial Engineering*, Springer, New York, 2003.
12. S. Graf and H. Luschgy: *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.

13. M. Griebel and M. Holtz: Dimension-wise integration of high-dimensional functions with applications to finance, *Journal of Complexity* 26 (2010), 455–489.
14. M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of the ANOVA decomposition, *Journal of Complexity* 26 (2010), 523–551.
15. M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* 82 (2013), 383–400.
16. F. J. Hickernell: A generalized discrepancy and quadrature error bound, *Mathematics of Computation* 67 (1998), 299-322.
17. F. J. Hickernell: Obtaining $O(N^{-2+\epsilon})$ convergence for lattice quadrature rules, in *Monte Carlo and Quasi-Monte Carlo Methods 2000* (K.-T. Fang, F. J. Hickernell, H. Niederreiter eds.), Springer, Berlin, 2002, 274–289.
18. W. Hoeffding: A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics* 19 (1948), 293–325.
19. T. Homem-de-Mello: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling, *SIAM Journal on Optimization* 19 (2008), 524-551.
20. H. S. Hong and F. J. Hickernell: Algorithm 823: Implementing scrambled digital sequences, *ACM Trans. Math. Softw.* 29 (2003), 95–109.
21. J. Imai and K. S. Tan: Minimizing effective dimension using linear transformation, in *Monte Carlo and Quasi-Monte Carlo Methods* (H. Niederreiter Ed.), Springer, Berlin, 2004, 275–292.
22. S. Joe and F. Y. Kuo: Remark on Algorithm 659: Implementing Sobol's quasirandom sequence generator, *ACM Transactions on Mathematical Software* 29 (2003), 49–57.
23. M. Koivu: Variance reduction in sample approximations of stochastic programs, *Mathematical Programming* 103 (2005), 463–485.
24. F. Y. Kuo: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces, *Journal of Complexity* 19 (2003), 301-320.
25. F. Y. Kuo, Ch. Schwab and I. H. Sloan: Quasi-Monte Carlo methods for high-dimensional integration: the standard (weighted Hilbert space) setting and beyond, ANZIAM Journal 53 (2011), 1–37.
26. F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and B. J. Waterhouse: Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands, *Journal of Complexity* 26 (2010), 135–160.
27. F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.
28. P. L'Ecuyer and Ch. Lemieux: Recent advances in randomized quasi-Monte Carlo methods, in *Modeling Uncertainty* (M. Dror, P. L'Ecuyer, F. Szidarovski eds.), Kluwer, Boston, 2002, 419–474.
29. Ch. Lemieux: *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, New York, 2009.
30. R. Liu and A. B. Owen: Estimating mean dimensionality of analysis of variance decompositions, *Journal of the American Statistical Association* 101 (2006), 712–721.
31. J. Matoušek: On the L_2-discrepancy for anchored boxes, *Journal of Complexity* 14 (1998), 527–556.
32. M. Matsumoto, T. Nishimura: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* 8 (1998), 3–30.
33. H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
34. F. Nožička, J. Guddat, H. Hollatz and B. Bank: *Theory of Linear Parametric Programming* (in German), Akademie-Verlag, Berlin 1974.
35. D. Nuyens and R. Cools: Fast algorithms for component-by-component constructions of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces, *Mathematics of Computation* 75 (2006), 903-922.
36. A. B. Owen: Randomly Permuted (t, m, s)-Nets and (t, s)-Sequences. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.-S. Shiue eds.), Lecture Notes in Statistics, Vol. 106, Springer, New York, 1995, 299–317.
37. A. B. Owen: Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.* 34 (1997), 1884–1910.
38. A. B. Owen: Scrambled net variance for integrals of smooth functions, *Annals of Statistics* 25 (1997), 1541–1562.
39. A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.

40. A. B. Owen: Multidimensional variation for Quasi-Monte Carlo, in J. Fan, G. Li (Eds.), International Conference on Statistics, World Scientific Publ., 2005, 49–74.
41. A. B. Owen: Local antithetic sampling with scrambled nets, *Annals of Statistics* 36 (2008), 2319–2343.
42. G. Pagès: A space quantization method for numerical integration, *Journal Computational and Applied Mathematics* 89 (1997), 1–38.
43. A. Papageorgiou: Brownian bridge does not offer a consistent advantage in Quasi-Monte Carlo integration, *Journal of Complexity* 18 (2002), 171–186.
44. S. Sp. Pappas, L. Ekonomou, P. Karampelas, D. C. Karamousantas, S.K. Katsikas, G.E. Chatzarakis and P.D. Skafidas: Electricity demand load forecasting of the Hellenic power system using an ARMA model, *Electric Power Systems Research* 80 (2010), 256–264.
45. T. Pennanen and M. Koivu: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.
46. G. Ch. Pflug and A. Pichler: Approximations for probability distributions and stochastic optimization problems, in: *Stochastic Optimization Methods in Finance and Energy* (M.I. Bertocchi, G. Consigli, M.A.H. Dempster eds.), Springer, 2011, 343–387.
47. P. D. Proinov: Discrepancy and integration of continuous functions, *Journal Approximation Theory* 52 (1998), 121–131.
48. S. T. Rachev and L. Rüschendorf: *Mass Transportation Problems*, Vol. I, Springer, New York, 1998.
49. W. Römisch: Stability of stochastic programming problems, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.
50. W. Römisch: Scenario generation, in: Wiley Encyclopedia of Operations Research and Management Science (J.J. Cochran ed.), Wiley, 2010.
51. A. Ruszczyński and A. Shapiro (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.
52. A. Shapiro, D. Dentcheva and A. Ruszczyński: *Lectures on Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2009.
53. I. H. Sloan: QMC integration – Beating intractability by weighting the coordinate directions, in *Monte Carlo and Quasi-Monte Carlo Methods 2000* (K.-T. Fang, F. J. Hickernell and H. Niederreiter eds.), Springer, Berlin, 2002, 103–123.
54. I. H. Sloan and H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.
55. I. H. Sloan, F. Y. Kuo and S. Joe: Constructing randomly shifted lattice rules in weighted Sobolev spaces, *SIAM Journal Numerical Analysis* 40 (2002), 1650–1665.
56. I. M. Sobol': *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).
57. I. M. Sobol': Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* 55 (2001), 271–280.
58. I. M. Sobol' and S. Kucherenko: Derivative based global sensitivity measures and their link with global sensitivity indices, *Mathematics and Computers in Simulation* 79 (2009), 3009-3017.
59. A. Takemura: Tensor analysis of ANOVA decomposition, *Journal of the American Statistical Association* 78 (1983), 894-900.
60. Shu Tezuka and Henri Faure: I-binomial scrambling of digital nets and sequences, *Journal of Complexity* 19 (2003), 744–757.
61. D. Walkup and R. J-B Wets: Lifting projections of convex polyedra, *Pacific Journal of Mathematics* 28 (1969), 465–475.
62. S. W. Wallace and W. T. Ziemba (Eds.): *Applications of Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2005.
63. X. Wang: Tractability of multivariate integration using Quasi-Monte Carlo algorithms, *Mathematics of Computation* 72 (2003), 823–838.
64. X. Wang and K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.
65. X. Wang and I. H. Sloan: Why are high-dimensional finance problems often of low effective dimension, *SIAM Journal Scientific Computing* 27 (2005), 159–183.
66. X. Wang and I. H. Sloan: Brownian bridge and principal component analysis: towards removing the curse of dimensionality, *IMA Journal of Numerical Analysis* 27 (2007), 631–654.

67. X. Wang and I. H. Sloan: Low discrepancy sequences in high dimensions: How well are their projections distributed? *Journal of Computational and Applied Mathematics* 213 (2008), 366–386.
68. X. Wang and I. H. Sloan: Quasi-Monte Carlo methods in financial engineering: An equivalence principle and dimension reduction, *Operations Research* 59 (2011), 80–95.
69. R. J-B Wets: Stochastic programs with fixed recourse: The equivalent deterministic program, *SIAM Review* 16, 309–339.