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

H. Heitsch and W. Römisch

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

**Summary.** Quasi-Monte Carlo algorithms are studied for designing discrete approximations of two-stage linear stochastic programs. Their integrands are piecewise linear, but neither smooth nor of bounded variation in the sense of Hardy and Krause. 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 smooth and, hence, certain Quasi-Monte Carlo algorithms 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 integrands belong to weighted tensor product Sobolev spaces with properly selected weights. The geometric condition is generically (i.e., almost everywhere) satisfied if the underlying distribution is normal. We also discuss sensitivity indices and efficient dimensions of two-stage integrands, and suggest a dimension reduction heuristic for such integrands.

## 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, $\xi$ is a $d$-dimensional random vector, $h$ represents an affine function from $\mathbb{R}^d$ to $\mathbb{R}^r$ 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 $\xi$, by introducing additional costs $\Phi(x, \xi(\omega))$ whose mean with respect to the probability distribution $P$ of $\xi$ 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 program, i.e.,

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

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 [40] 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 [42, 43, 46]) 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 based on a finite number $n$ of *samples* or *scenarios* $\xi^j\in\mathbb{R}^d$ with probabilities $p_j$, $j=1,\dots,n$. While 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 [31], [29, 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 dealt with the situation of deterministic samples with identical weights $p_j=\frac{1}{n}$ (see [3, 25, 10, 26] and [30] 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 [4], [24]) and determines such quantizations by (approximately) solving best approximation problems for $P$ in terms of the $L_p$-minimal (or $L_p$-Wasserstein) metric $\ell_p$, $p\geq 1$ (see Section 2.5 in [28]). It is shown in [4] that, under certain conditions, it holds

$$\ell_p(P,P_n)=O(n^{-\frac{1}{d}}) \tag{4}$$

for the optimal quantization $P_n$ of $P$ having $n$ samples. The dual representation of $\ell_1$ reveals that the dimension dependent convergence rate is due to the fact that the identity

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

holds, where the supremum is taken with respect to the unit ball in the normed space $\mathbb{F}_d=\mathrm{Lip}\,(\mathbb{R}^d)$ of Lipschitz functions with the Lipschitz norm $\|\cdot\|_L$. This unit ball is known to be too large for obtaining better rates.

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

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

and relies on the concept of equidistributed point sets or low discrepancy point sets $\{\eta_j\}_{j=1}^n$ or sequences $(\eta_j)_{j\in\mathbb{N}}$ in $[0,1)^d$ (see [35, 19, 17, 18, 2]). It turned out that 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 (see [8]). Let $K : [0,1]^d \times [0,1]^d \to \mathbb{R}$ be a kernel satisfying $K(\cdot, y) \in \mathbb{F}_d$ for each $y \in [0,1]^d$ and $\langle f, K(\cdot, y)\rangle = f(y)$ for each $y \in [0,1]^d$ and $f \in \mathbb{F}_d$. With $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denoting the inner product and norm in $\mathbb{F}_d$, the quadrature error $e_n(\mathbb{F}_d)$ allows the representation

$$e_n(\mathbb{F}_d) = \sup_{f \in \mathbb{F}_d, \|f\| \le 1} \left| \int_{[0,1]^d} f(x)dx - Q_{n,d}(f) \right| = \sup_{\|f\| \le 1} |\langle f, h_n \rangle| = \|h_n\| \quad (5)$$

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

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

In particular, the weighted tensor product Sobolev space [33]

$$\mathbb{F}_d = \mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d) = \bigotimes_{i=1}^{d} W_2^1([0,1]) \quad (6)$$

equipped with the weighted norm $\|f\|_\gamma^2 = \langle f, f \rangle_\gamma$ with the inner product (see Section 3 for the notation)

$$\langle f, g \rangle_\gamma = \sum_{u \subseteq \{1,\ldots,d\}} \gamma_u^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial x^u} f(x^u, 1^{-u}) \frac{\partial^{|u|}}{\partial x^u} g(x^u, 1^{-u}) dx^u \quad (7)$$

and a weighted Walsh space consisting of Walsh series (see [2, Example 2.8] and [1]) are reproducing kernel Hilbert spaces and became important for analyzing the recently developed randomized lattice rules (see [34, 13, 14, 21] and [1, 2]). They allow for deriving optimal error estimates of the form

$$e_n(\mathbb{F}_d) \le C(\delta) n^{-1+\delta} \quad (n \in \mathbb{N}, \delta \in (0, \tfrac{1}{2})), \quad (8)$$

where the constant $C(\delta)$ does not depend on the dimension $d$, if the weights $\gamma_u$ are multiplicatively defined from a decreasing sequence of nonnegative weights $(\gamma_j)$ satisfying

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

The rate (8) is also obtained for Niederreiter and Sobol' sequences in [41] if the integrands belong to weighted tensor product Sobolev spaces and the weights $\gamma_j$, $j \in \mathbb{N}$, satisfy specific conditions.
Unfortunately, typical integrands in linear two-stage stochastic programming (see Section 2) do not belong to such tensor product Sobolev or Walsh spaces and are even not of bounded variation in the sense of Hardy and Krause. The

latter condition represents the standard requirement on the integrand $f$ to justify Quasi-Monte Carlo algorithms via the Koksma-Hlawka theorem [19, Theorem 2.11].
Alternatively, it is suggested in the literature to study the so-called ANOVA decomposition (see Section 3) of such integrands, the smoothness of the ANOVA terms, the dimension distribution and sensitivity indices of the integrands. As a first step in this direction we show in Section 4 that all ANOVA terms except the one of highest order are smooth under some algebraic condition on the second stage program and smoothness conditions on the density of the underlying probability distribution. In addition, we show in Section 5 that the algebraic condition is generically satisfied if the underlying random vector is Gaussian. Finally, we provide estimates of sensitivity indices and of the mean dimension in Section 6. We also discuss known techniques for dimension reduction and suggest a dimension reduction heuristic for two-stage models. In the conclusions we argue that our results indicate the efficiency of randomly shifted and digitally shifted polynomial lattice rules for large scale two-stage stochastic programs.

## 2 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

$$\Phi(x,\xi) = \phi(h(\xi) - Tx), \tag{9}$$

where $\phi$ 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 it holds for $t \in \operatorname{dom}\phi = \{t \in \mathbb{R}^r : |\phi(t)| < \infty\}$

$$\phi(t) = \sup\{\langle t, z\rangle : W^\top z \leq q\}, \tag{10}$$

if 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, the dual feasible set $\mathcal{D}$ 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.,

$$\mathcal{K}_j = \{t \in \operatorname{dom}\phi : \langle t, z - v^j\rangle \leq 0, \forall z \in \mathcal{D}\} \quad (j = 1, \ldots, \ell) \tag{11}$$

$$= \{t \in \operatorname{dom}\phi : \langle t, v^i - v^j\rangle \leq 0, \forall i = 1, \ldots, \ell, i \neq j\}. \tag{12}$$

Moreover, it holds

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

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

$$\{t \in \operatorname{dom}\phi : \langle t, v^{j'} - v^j\rangle = 0\}. \tag{13}$$

If there exists $k \in \{1,\dots,d\}$ such that the $k$th 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

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

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 $w^i \in \mathbb{R}^d$ are the columns of $W$ and $I_j = \{i \in \{1,\dots,\bar{m}\} : \langle w^i, v^j\rangle = q_i\}$. Each vertex $v^j$ is determined by $d$ linear independent equations out of the $\bar{m}$ equations $\langle w^i, v\rangle = q_i$, $i = 1,\dots,\bar{m}$.
In the following we assume
**(A1)** $h(\mathbb{R}^d) \subseteq W(\mathbb{R}^{\bar{m}}_+)$ (*relatively complete recourse*).
**(A2)** The dual feasible set $\mathcal{D}$ is nonempty (*dual feasibility*).
**(A3)** $\int_{\mathbb{R}^d} \|\xi\| P(d\xi) < \infty$ (*finite first moment*).
**(A4)** $P$ has a density of the form $\rho_D(\xi) = \prod_{i=1}^d \rho_i(\xi_i)$ ($\xi \in \mathbb{R}^d$), where $\rho_i$ is a continuous density on $\mathbb{R}$, $i = 1,\dots,d$.
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. For further information on linear parametric programming and two-stage stochastic programming we refer to [39, 20] and [31, 32, 47].
To give an example we show that option pricing models considered as stimulating examples for the recent developments in QMC theory (see e.g. [43, 44]) 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

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

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

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

for minimizing the strike price. Now, it depends on the kind of option how the random variable $\xi$ depends on the geometric Brownian motion $S$ given by

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

with volatility $\sigma$ and standard Brownian motion $(B_t)_{t\geq 0}$. For example, for arithmetic Asian options one has [42]

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

Hence, in a sense, the integrands considered in this paper extend the situations encountered in such option pricing models.

## 3 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 [9] and the survey [38]). In [35] it was first used for the analysis of quadrature methods.
We consider a density function $\rho_D$ on $\mathbb{R}^d$ and assume (A4) from Section 2. As in [7] we consider the weighted $\mathcal{L}_p$ space over $\mathbb{R}^d$, i.e., $\mathcal{L}_{p,\rho_D}(\mathbb{R}^d)$, with the norm

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

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

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

Clearly, the function $P_k f$ is constant with respect to $\xi_k$. For $u \subseteq D$ we use $|u|$ for its cardinality, $-u$ for $D \setminus u$ 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 $\xi_k$, $k \in u$. Note that $P_u$ satisfies the properties of a projection, namely, $P_u$ is linear and it holds $P_u^2 = P_u$.
The ANOVA decomposition of $f \in \mathcal{L}_{1,\rho_D}(\mathbb{R}^d)$ is of the form [42, 15]

$$f = \sum_{u \subseteq D} f_u \tag{14}$$

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

$$f_\emptyset = I_{d,\rho_d}(f) = P_D(f) \quad \text{and} \quad f_u = P_{-u}(f) - \sum_{v \subseteq u} f_v \,.$$

It is known from [15] that the ANOVA terms are given explicitly by

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u}(f) + \sum_{v \subset u} (-1)^{|u|-|v|} P_{u-v}(P_{-u}(f)), \tag{15}$$

where $P_{-u}$ and $P_{u-v}$ mean integration with respect to $\xi_j$, $j \in D \setminus u$ and $j \in u \setminus v$, respectively. The second representation motivates that $f_u$ is essentially as smooth as $P_{-u}(f)$ due to the Inheritance Theorem [7, Theorem 2]. The following result is well known (e.g. [42]).

**Proposition 1.** *If $f$ belongs to $\mathcal{L}_{2,\rho_D}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u \subseteq D}$ are orthogonal in $\mathcal{L}_{2,\rho_D}(\mathbb{R}^d)$.*

If the variance of $f$ is defined by $\sigma^2(f) = \|f - I_{d,\rho_D}(f)\|_{L_2}^2$, it holds that

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

To avoid trivial cases we assume $\sigma(f) > 0$ in the following. The normalized ratios $\frac{\sigma_u^2(f)}{\sigma^2(f)}$ serve as indicators for the importance of the variable $\xi^u$ in $f$. They are used in [36] to define *global sensitivity indices* of a set $u \subseteq D$ 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 $\xi^u$ is considered inessential for $f$ in [36].
The normalized ratios are also used in [22, 16] 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 $\nu_S$ ($\nu_T$) defined on the power set of $D$ by

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

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

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

It is proved in [16, Theorem 2] that the mean dimension $\bar{d}_S$ in the superposition sense is closely related to the global sensitivity indices of subsets of $D$ containing a single element. Namely, it holds

$$\bar{d}_S = \sum_{j=1}^{d} \bar{S}_{\{j\}}. \tag{16}$$

The paper [16] also provides a formula for the dimension variance based on $\bar{S}_u$ for all subsets $u$ of $D$ containing two indices.
For small $\varepsilon \in (0,1)$ ($\varepsilon = 0.01$ is suggested in a number of papers), the *effective superposition (truncation) dimension* $d_S(\varepsilon) \in D$ ($d_T(\varepsilon) \in D$) is the $(1-\varepsilon)$-quantile of $\nu_S$ ($\nu_T$), i.e.,

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

Note that $d_S(\varepsilon) \leq d_T(\varepsilon)$ and it holds (see [42, 5])

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

Small effective superposition dimension $d_S(\varepsilon)$, even if $d_T(\varepsilon)$ is large, is considered as good hint to 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 [36, 37, 43] for example). For large $d$, however, this becomes computationally expensive.
All these notions are discussed in [22] for different classes of functions, including additive and multiplicative functions. We record here the results in case of additive functions for later reference.

*Example 2.* For functions $f$ having separability structure, i.e.,

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

with $g_j \in \mathcal{L}_{2,\rho_j}(\mathbb{R})$, $j = 1, \ldots, d$, the ANOVA terms are (see [22])

$$f_\emptyset(\xi) = \sum_{j=1}^{d} \mu_j, \quad f_{\{j\}}(\xi) = g_j(\xi_j) - \mu_j, \quad f_u(\xi) = 0 \text{ if } |u| > 1,$$

where $\mu_j = \int_{\mathbb{R}} g_j(t)\rho_j(t)dt$, $\sigma_j^2 = \int_{\mathbb{R}}(g_j(t)-\mu_j)^2\rho_j(t)dt$, $j = 1,\dots,d$. Hence, one obtains for the global sensitivity indices, and the mean dimension in the superposition and truncation sense, respectively,

$$S_{\{j\}} = \frac{\sigma_j^2}{\sigma^2(f)}, \quad \bar{d}_S = 1 \quad \text{and} \quad \bar{d}_T = \sum_{j=1}^{d} j\Big(\frac{\sigma_j}{\sigma(f)}\Big)^2,$$

while the superposition and truncation dimensions are

$$d_S(\varepsilon) = 1 \quad (\forall \varepsilon \in (0,1)) \quad \text{and} \quad d_T(\varepsilon) = s \quad \text{if} \sum_{j=s+1}^{d} \Big(\frac{\sigma_j}{\sigma(f)}\Big)^2 \le \varepsilon.$$

with $\sigma^2(f) = \sum_{j=1}^{d} \sigma_j^2$.

The importance of the ANOVA decomposition in the context of this paper is also due to the fact that the functions $f_u$ can be (much) smoother than the original integrand $f$ under some conditions (see [6, 7] and the next section).

## 4 ANOVA decomposition of linear two-stage integrands

According to Section 2 the integrands in linear two-stage stochastic programming map from $\mathbb{R}^d$ to $\mathbb{R}$ and are given by

$$f(\xi) = f_x(\xi) = \max_{j=1,\dots,\ell} \langle v^j, h(\xi) - Tx\rangle \quad (x \in X), \tag{17}$$

where the affine function $h$ is assumed to be of the form $h(\xi) = (\xi, \bar{h}) = (\xi, 0) + (0, \bar{h})$ with some fixed element $\bar{h} \in \mathbb{R}^{r-d}$. The integrands are parametrized by the first-stage decision $x$. Such functions do not belong to tensor product Sobolev spaces and, in general, are not of bounded variation in the sense of Hardy and Krause (see [23, Proposition 17]).
Next we intend to compute projections $P_k(f)$ for $k \in D$. Let $x \in X$ be fixed, $\xi_i \in \mathbb{R}$, $i = 1,\dots,d$, $i \ne k$, be given. We set $\bar{\xi}^k = (\xi_1,\dots,\xi_{k-1},\xi_{k+1},\dots,\xi_d)$ and $\bar{\xi}^k_s = (\xi_1,\dots,\xi_{k-1},s,\xi_{k+1},\dots,\xi_d)$. We assume (A1)–(A4) and have

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

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

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

The one-dimensional affine subspace $\{(\bar{\xi}^k_s, \bar{h}) - Tx : s \in \mathbb{R}\}$ intersects a finite number of the polyhedral cones $\mathcal{K}_j$. Hence, there exist $p = p(k) \in \mathbb{N} \cup \{0\}$,

$s_i = s_i^k \in \mathbb{R}$, $i = 1, \ldots, p$, and $j_i = j_i^k \in D$, $i = 1, \ldots, p+1$, such that $s_i < s_{i+1}$ and

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

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

$$(P_k f)(\bar{\xi}^k) = \sum_{i=1}^{p+1} \int_{s_{i-1}}^{s_i} \langle v^{j_i}, (\bar{\xi}_s^k, \bar{h}) - Tx \rangle \rho_k(s) ds \tag{18}$$

$$= \sum_{i=1}^{p+1} \Big( \Big( \sum_{\substack{j=1 \\ j \neq k}}^{d} v_j^{j_i} \xi_j + \langle v^{j_i}, (0, \bar{h}) - Tx \rangle \Big) \int_{s_{i-1}}^{s_i} \rho_k(s) ds + v_k^{j_i} \int_{s_{i-1}}^{s_i} s \rho_k(s) ds \Big)$$

$$= \sum_{i=1}^{p+1} \Big( \Big( \sum_{\substack{j=1 \\ j \neq k}}^{d} v_j^{j_i} \xi_j + \langle v^{j_i}, (0, \bar{h}) - Tx \rangle \Big) (\varphi_k(s_i) - \varphi_k(s_{i-1})) \tag{19}$$

$$+ v_k^{j_i} (\psi_k(s_i) - \psi_k(s_{i-1})) \Big)$$

Here, $\varphi_k$ is the one-dimensional distribution function with density $\rho_k$, $\psi_k$ the corresponding mean value function and $\mu_k$ the mean value, i.e.,

$$\varphi_k(u) = \int_{-\infty}^{u} \rho_k(s) ds, \quad \psi_k(u) = \int_{-\infty}^{u} s \rho_k(s) ds, \quad \mu_k = \int_{-\infty}^{+\infty} s \rho_k(s) ds.$$

Next we reorder the outer sum to collect the factors of $\varphi_k(s_i)$ and $\psi_k(s_i)$, and a remainder.

$$(P_k f)(\bar{\xi}^k) = \sum_{i=1}^{p} \Big( \Big( \sum_{\substack{j=1 \\ j \neq k}}^{d} (v_j^{j_i} - v_j^{j_{i+1}}) \xi_j + \langle v^{j_i} - v^{j_{i+1}}, (0, \bar{h}) - Tx \rangle \Big) \varphi_k(s_i) + \tag{20}$$

$$(v_k^{j_i} - v_k^{j_{i+1}}) \psi_k(s_i) \Big) + \sum_{\substack{j=1 \\ j \neq k}}^{d} v_j^{j_{p+1}} \xi_j + \langle v^{j_{p+1}}, (0, \bar{h}) - Tx \rangle + v_k^{j_{p+1}} \mu_k.$$

Moreover, the points $s_i$, $i = 1, \ldots, p$, satisfy the equations

$$\langle (\xi_{s_i}, 0), v^{j_{i+1}} - v^{j_i} \rangle = \sum_{\substack{j=1 \\ j \neq k}}^{d} \xi_j (v_j^{j_{i+1}} - v_j^{j_i}) + s_i (v_k^{j_{i+1}} - v_k^{j_i}) = 0 \ \ (i = 1, \ldots, p),$$

according to (13). This leads to the explicit formula

$$s_i = \frac{1}{v_k^{j_i} - v_k^{j_{i+1}}} \sum_{\substack{j=1 \\ j \neq k}}^{d} \xi_j (v_j^{j_{i+1}} - v_j^{j_i}) \text{ if } v_k^{j_i} \neq v_k^{j_{i+1}} \quad (i = 1, \ldots, p). \tag{21}$$

Hence, all $s_i$, $i = 1, \ldots, p$, are linear combinations of the remaining components $\xi_j$, $j \neq k$, of $\xi$ if the following *geometric condition* is satisfied: All $k$th components of adjacent vertices of $\mathcal{D}$ are different from each other, i.e., all facets of $\mathcal{D}$ are not parallel to the $k$th coordinate axis in $\mathbb{R}^r$ or, with other words, not parallel to the canonical basis element $e_k$ (whose components are equal to $\delta_{ik}$, $i = 1, \ldots, r$).
To simplify notation we set $w^i = v^{j_i} - v^{j_{i+1}}$ and $z(x) = (0, \bar{h}) - Tx$. If the above geometric condition is satisfied, we obtain the following form of the projection $P_k f$:

$$(P_k f)(\bar{\xi}^k) = \sum_{i=1}^{p} \Big( \Big( \sum_{\substack{j=1 \\ j \neq k}}^{d} w_j^i \xi_j + \langle w^i, z(x) \rangle \Big) \varphi_k(s_i(\bar{\xi}^k)) + w_k^i \psi_k(s_i(\bar{\xi}^k)) \Big) \tag{22}$$

$$+ \sum_{\substack{j=1 \\ j \neq k}}^{d} v_j^{j_{p+1}} \xi_j + \langle v^{j_{p+1}}, z(x) \rangle + v_k^{j_{p+1}} \mu_k \tag{23}$$

$$s_i = s_i(\bar{\xi}^k) = -\frac{1}{w_k^i} \sum_{\substack{j=1 \\ j \neq k}}^{d} w_j^i \xi_j. \tag{24}$$

Hence, the projection represents a sum of products of differentiable functions and of affine functions of $\xi^k$.

**Proposition 2.** *Let $k \in D$ and $x \in X$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different $k$th components. Let $f_x$ be the integrand (17) of the linear two-stage stochastic program (3).*
*Then the $k$th projection $P_k f_x$ is continuously differentiable.*
*The projection $P_k f_x$ belongs to $C^s(\mathbb{R}^d)$ if the density $\rho_k$ is in $C^{s-1}(\mathbb{R})$ ($s \in \mathbb{N}$).*
*$P_k f_x$ is infinitely differentiable if the density $\rho_k$ is in $C^\infty(\mathbb{R})$.*

*Proof.* Let $l \in D$, $l \neq k$. The projection $P_k f$ is partially differentiable with respect to $\xi_l$ and it holds according to (22)–(24)

$$\begin{aligned} \frac{\partial P_k f}{\partial \xi_l}(\bar{\xi}^k) &= \sum_{i=1}^{p} \Big( w_l^i \varphi_k(s_i(\bar{\xi}^k)) - \Big( \sum_{\substack{j=1 \\ j \neq k}}^{d} w_j^i \xi_j + \langle w^i, z(x) \rangle \Big) \rho_k(s_i(\bar{\xi}^k)) \frac{w_l^i}{w_k^i} \\ &\qquad - w_l^i s_i(\bar{\xi}^k) \rho_k(s_i(\bar{\xi}^k)) \Big) + v_l^{j_{p+1}} \\ &= \sum_{i=1}^{p} \Big( w_l^i \varphi_k(s_i(\bar{\xi}^k)) - \langle w^i, z(x) \rangle \rho_k(s_i(\bar{\xi}^k)) \frac{w_l^i}{w_k^i} \Big) + v_l^{j_{p+1}} \end{aligned}$$

Hence, the behavior of all partial derivatives of $P_k f$ only depends on the $k$th marginal densities and distribution functions. They are again continuous with

respect to $\bar{\xi}^k$ if $\rho_k$ is continuous. If $\rho_k \in C^{s-1}(\mathbb{R})$ for some $s \in \mathbb{N}$, $\varphi_k$ and, thus, $P_k f$ belong to $C^s(\mathbb{R}^d)$. If $\rho_k \in C^\infty(\mathbb{R})$, $P_k f$ is in $C^\infty(\mathbb{R}^d)$. □

*Remark 1.* If $\rho_k$ is not continuous, but belongs to $L_p(\mathbb{R})$, the projection $P_k f$ is in the Sobolev space $\mathcal{W}_p^1(\mathbb{R}^d)$. For the definition of the Sobolev spaces, we refer to [7].

**Proposition 3.** *Let $\emptyset \neq u \subseteq D$ and $x \in X$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different $k$th components for some $k \in u$. Then the projection $P_u f_x$ is continuously differentiable. The projection $P_u f_x$ is infinitely differentiable if $\rho_k \in C_b^\infty(\mathbb{R})$. Here, the subscript $b$ at $C_b^\infty(\mathbb{R})$ indicates that all derivatives of functions in that space are bounded on $\mathbb{R}$.*

*Proof.* If $|u| = 1$ the result follows from Proposition 2. For $u = \{k, r\}$ with $k, r \in D$, $k \neq r$, we obtain from the Leibniz theorem [7, Theorem 1] for $l \notin u$

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

and from the proof of Proposition 2

$$D_l P_u f = \sum_{i=1}^{p} \Big( w_l^i \int_{\mathbb{R}} \varphi_k(s_i)\rho_r(\xi_r) d\xi_r - \langle w^i, z(x)\rangle \frac{w_l^i}{w_k^i} \int_{\mathbb{R}} \rho_k(s_i)\rho_r(\xi_r) d\xi_r \Big) + v_l^{j_{p+1}}$$

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 function of the remaining variables $\xi_j$, $j \in D \setminus u$, whose continuity and differentiability properties correspond to those of $\varphi_k$ and $\rho_k$. This can be shown using Lebesgue's theorem as $\varphi_k$ and all densities $\rho_j$, $j \in u$, and their derivatives are bounded on $\mathbb{R}$. □

The following is the main result of this section.

**Theorem 1.** *Let $u \subset D$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different $k$th components for some $k \in -u = D \setminus u$. Then the ANOVA term $f_u$ is infinitely differentiable if $\rho_k \in C_b^\infty(\mathbb{R})$.*

*Proof.* The result follows from Proposition 3 applied to $P_{-u}(f_x)$, the Inheritance Theorem [7, Theorem 2] and the second part of formula (15). □

**Corollary 1.** *Let $x \in X$. Assume (A1)–(A4) and that all components of all adjacent vertices of $\mathcal{D}$ are different. Then the ANOVA approximation*

$$f_{d-1} = \sum_{u \subset D} f_u \tag{25}$$

*of $f_x$ is infinitely differentiable if all densities $\rho_k$, $k \in D$, belong to $C_b^\infty(\mathbb{R})$.*

*Proof.* The result follows immediately from Theorem 1 when applying it to all nonempty strict subsets of $D$. □

*Remark 2.* Under the assumptions of Corollary 1 all ANOVA terms $f_u$ are at least continuously differentiable if $\rho$ is continuous and $|u| \leq d-1$. Hence, the function $f_{d-1}$ is in $C^1(\mathbb{R}^d)$ ($C^\infty(\mathbb{R}^d)$) if each $\rho_k$, $k \in D$, belongs to $C(\mathbb{R})$ ($C_b^\infty(\mathbb{R})$). On the other hand, it holds

$$f = f_{d-1} + f_D \quad \text{and} \quad \|f - f_{d-1}\|_{L_2}^2 = \|f_D\|_{L_2}^2$$

according to (14) and Proposition 1. Hence, the question arises: For which two-stage linear stochastic programs is the $L_2$-norm of $f_D$ small or, equivalently, is $f_{d-1}$ a good approximation of $f$ in $L_{2,\rho_d}$? This means in terms of the truncation dimension that $d_t$ should at least be smaller than $d$. The latter condition appears to be realistic in models with medium- or long-term time horizon.

The following example shows that the condition on adjacent vertices in all results cannot be removed in general.

*Example 3.* Let $\bar{m} = 3$, $d = 2$, $\Xi = \mathbb{R}^2$, $P$ denote the two-dimensional standard normal distribution and let the following 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

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

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. According to (12) the normal cones $\mathcal{K}_j$ to $\mathcal{D}$ at $v^j$, $j = 1, 2, 3$, are

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

The function $\Phi$ is of the form

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

and the two-stage stochastic program is

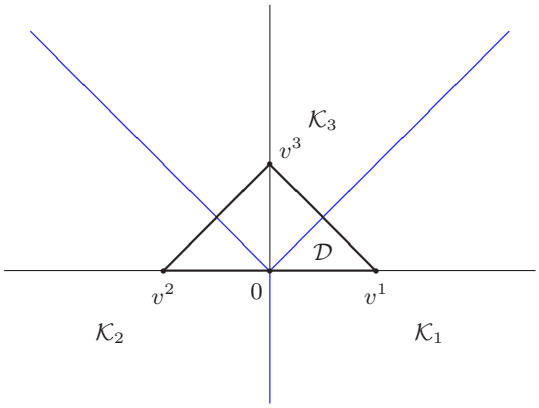


**Fig. 1.** Illustration of $\mathcal{D}$, its vertices $v^j$ and the normal cones $\mathcal{K}_j$ to its vertices

$$\min\Big\{\langle c,x\rangle + \int_{\mathbb{R}^2} \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}\rho_2(\xi)d\xi : x \in X\Big\} \qquad (26)$$

with the standard normal density

$$\rho_2(\xi_1,\xi_2) = \rho(\xi_1)\rho(\xi_2) = \frac{1}{2\pi}\exp\Big(-\frac{\xi_1^2}{2}\Big)\exp\Big(-\frac{\xi_2^2}{2}\Big) \quad ((\xi_1,\xi_2)\in\mathbb{R}^2).$$

The integral $I_2$ in (26) may be rewritten as

$$I_2(f) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}\exp\Big(-\frac{\xi_1^2+\xi_2^2}{2}\Big)d\xi_1 d\xi_2$$

The ANOVA projection $P_1 f$ is

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

Case $\xi_2 - [Tx]_2 \le 0$:

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

Case $\xi_2 - [Tx]_2 \ge 0$:

$$(P_1 f)(\xi_2) = \int_{-\infty}^{0} (-\xi_1 + [Tx]_1)\rho(\xi_1)d\xi_1 + \int_{0}^{\xi_2 - [Tx]_2} (\xi_2 - [Tx]_2)\rho(\xi_1)d\xi_1$$
$$+ \int_{\xi_2 - [Tx]_2}^{+\infty} (\xi_1 - [Tx]_1)\rho(\xi_1)d\xi_1$$
$$= \int_{-\infty}^{+\infty} |\xi_1 - [Tx]_1|\rho(\xi_1)d\xi_1 - \int_{0}^{\xi_2 - [Tx]_2} (\xi_1 + \xi_2 - [Tx]_1 - [Tx]_2)\rho(\xi_1)d\xi_1$$

Hence, $P_1 f$ belongs to $C^1(\mathbb{R})$ for all $x \in X$ if $\rho$ is continuous.
Next we calculate the ANOVA projection $P_2 f$. Notice that the assumption of Proposition 1 is violated for $k = 2$.
Case $\xi_1 - [Tx]_1 \geq 0$:

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

Case $\xi_1 - [Tx]_1 \leq 0$:

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

Hence, we obtain

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

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

## 5 Orthogonal transformations and the Gaussian case

We consider the stochastic program (3) with

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

as in Section 4 and assume that (A1)–(A3) is satisfied. Further we assume that $h(\xi)$ is of the form $h(\xi) = (Q\xi, \bar{h})$ with some orthogonal $d \times d$ matrix $Q$ and with $\xi$ satisfying (A4). Then the relevant integrand is of the form

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

where the $r \times r$ matrix $\hat{Q}$ is given by

$$\hat{Q} = \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix} \tag{27}$$

with $I$ denoting the $(r-d)\times(r-d)$ identity matrix. Hence, the results of Section 4 apply if the vertices $\hat{Q}^\top v^j$, $j=1,\dots,\ell$, of the linearly transformed dual feasible set $\hat{Q}^\top\mathcal{D}$ satisfy the corresponding assumptions. The set $\hat{Q}^\top\mathcal{D}$ may be represented in the form

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

The geometric condition on the vertices is violated only if some face of $\hat{Q}^\top\mathcal{D}$ is parallel to some coordinate axis. Clearly, there are only countably many orthogonal matrices $Q$ for which this is the case.

Assume now that $\xi$ is normally distributed with zero mean and nonsingular covariance matrix $\Sigma$. Let the nonsingular diagonal matrix $D$ be the result of a unitary decomposition of $\Sigma$, i.e., $D=Q\,\Sigma\,Q^\top$ with an orthogonal matrix $Q$. If $h(\xi)=(\xi,\bar{h})$ enters the integrand (17) with given dual feasible polyhedron $\mathcal{D}$ and vertices $v^j$, $j=1,\dots,\ell$, and $\hat{Q}$ is defined as in (27), the integrand may be rewritten as

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

As $Q\xi$ is normal with covariance matrix $D$ and, thus, satisfies (A4), the results of the preceding section apply when using the transformed dual feasible set $\hat{Q}\mathcal{D}$ and normal cones $\hat{Q}^\top\mathcal{K}_j$, $j=1,\dots,\ell$, respectively. However, given $\mathcal{D}$, there are only countably many orthogonal matrices $Q$ such that the algebraic condition on the vertices of $\hat{Q}\mathcal{D}$ is not satisfied. When equipping the linear space of all orthogonal $d\times d$ matrices with the standard norm topology, the set of all orthogonal matrices $Q$ such that $\hat{Q}\mathcal{D}$ satisfies the algebraic condition on the vertices is *residual*, i.e., the countable intersection of open dense subsets. A property for elements of a topological space is called *generic* if it holds in a residual set. This proves part (a) of

**Corollary 2.** *Let $x\in X$ and assume (A1)–(A3) with $h(\xi)=(\xi,\bar{h})$ with fixed $\bar{h}\in\mathbb{R}^{r-d}$ to be satisfied.*

*(a) The algebraic condition that all components of all adjacent vertices of $\hat{Q}\mathcal{D}$ are different is a generic property in the space of all $d\times d$ orthogonal matrices $Q$ where $\hat{Q}$ is defined by (27).*

*(b) Let $\xi$ be normally distributed with mean $m\in\mathbb{R}^d$ and nonsingular covariance matrix $\Sigma$, and let the orthogonal matrix $Q$ be chosen such that $Q\,\Sigma\,Q^\top=\operatorname{diag}(\sigma_1^2,\dots,\sigma_d^2)$. Let $\rho_d$ be the normal density with mean $m$ and covariance matrix $\operatorname{diag}(\sigma_1^2,\dots,\sigma_d^2)$. If $Q$ belongs to the residual set of orthogonal matrices satisfying the generic property, the ANOVA approximation $f_{d-1}$ of $f$ given by (25) is infinitely differentiable.*

*Proof.* For part (b) it remains to note that the marginal normal densities

$$\rho_k(t)=\frac{1}{\sqrt{2\pi}\sigma_k}\exp\Big(-\frac{(t-m_k)^2}{2\sigma_k^2}\Big)\quad(k=1,\dots,d)$$

belong to $C_b^\infty(\mathbb{R})$ and, hence, the result follows from Corollary 1. □

## 6 Sensitivity and dimension reduction of two-stage stochastic programs

In this section we discuss sensitivity and possibilities for reducing the efficient dimension of two-stage models. First, we derive an upper bound for the global sensitivity indices $\bar{S}_{\{i\}}$, $i = 1, \dots, d$, 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 $\sigma_i^2$ denote the variance of $\xi_i$, $i = 1, \dots, d$. Then it holds*

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

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

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

*Proof.* We use [37, Theorem 3] and compute the partial derivatives of $f$ with respect to $\xi_i$, $i = 1, \dots, d$, which exist almost everywhere on $\mathbb{R}^d$. If $h(\xi) - Tx$ belongs to the cone $\mathcal{K}_j$, it holds

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

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 $\max_{j=1,\dots,\ell} |v_i^j|$. Using [37, Theorem 3] this proves our estimate for the global sensitivity index $\bar{S}_{\{i\}}$. The second estimate is a consequence of formula (16). □

The variance $\sigma^2(f)$ is greater and grows faster than $\sum_{i=1}^{d} \sigma_i^2$ with increasing dimension $d$ except in case that $f$ is affine (i.e. $\ell = 1$). Hence, Proposition 4 implies that the mean dimension in the superposition sense gets smaller with increasing $d$. Much better results may be obtained under further assumptions on the underlying stochastic program as indicated in the following example.

*Example 4.* We consider two-stage stochastic programs with generalized simple recourse (and $r = d$ for simplicity), i.e., the recourse matrix $W$ is of the form $W = (V, -V)$ with an invertible $d \times d$ matrix $V$. Let the second-stage costs $q \in \mathbb{R}^{2r}$ be decomposed into $q = (q^+, q^-)$, where $q+$, $q^-$ belong to $\mathbb{R}^r$, and we assume $q^+ + q^- \ge 0$. Then the dual feasible polyhedron is of the form

$\mathcal{D} = (V^\top)^{-1}([-q^-, q^+])$, where $[-q^-, q^+] = \times_{i=1}^r [-q_i^-, q_i^+]$ represents a rectangle in $\mathbb{R}^r$. Then (A1) and (A2) are satisfied and it holds for the integrand

$$f(\xi) = \max_{z\in\mathcal{D}} \langle \xi - Tx, (V^\top)^{-1} V^\top z\rangle = \max_{x\in[-q^-,q^+]} \langle V^{-1}(\xi - Tx), z\rangle = \sum_{i=1}^d f_i(\xi),$$

where

$$f_i(\xi) = \begin{cases} q_i^+ [V^{-1}(\xi - Tx)]_i, \ [V^{-1}(\xi - Tx)]_i > 0 \\ -q_i^- [V^{-1}(\xi - Tx)]_i, \ [V^{-1}(\xi - Tx)]_i \le 0. \end{cases}$$

If the matrix $V^{-1}$ is banded (say, containing at most $k$ nonzero elements in each row), $f_i$ depends on at most $k$ components of $\xi$, hence, the efficient superposition dimension is at most $k$. If $V$ is diagonal, Example 2 applies.

Another conclusion from Proposition 4 is that a low truncation dimension can only be achieved if the variances $\sigma_i$ are decreasing with increasing $i$ and if the first few variances are dominating.
If $\xi$ is normal with nonsingular covariance matrix $\Sigma$, the *standard* (lower triangular) Cholesky matrix $L_C$ performing the decomposition $L_C L_C^\top = \Sigma$ leads to $\sigma_i \equiv 1$ and is, hence, not suitable to reduce the efficient dimension.
A universal principle for dimension reduction is *principal component analysis* (PCA). It is universal in the sense that it does not depend on the structure of the underlying integrand $f$. In PCA one uses the decomposition $U_P\, U_P^\top = \Sigma$, where $U_P = (\sqrt{\lambda_1}u_1, \ldots, \sqrt{\lambda_d}u_d)$, $\lambda_1 \ge \cdots \ge \lambda_d \ge 0$ are the eigenvalues of $\Sigma$ in decreasing order and $u_i$, $i = 1, \ldots, d$, are the corresponding orthonormal eigenvectors of $\Sigma$. Several authors report an enormous reduction of the efficient truncation dimension in financial models if PCA is used (see, for example, [42], [43, 44]). However, the reduction effect depends on the eigenvalues of $\Sigma$. If the ratio $\frac{\lambda_1}{\lambda_d}$ is even close to 1, the performance of PCA gets worse. Furthermore, PCA may become expensive for large $d$.
Several *dimension reduction techniques* exploit the fact that a normal random vector $\xi$ with mean $\mu$ and covariance matrix $\Sigma$ can be transformed by $\xi = B\eta + \mu$ and any matrix $B$ satisfying $\Sigma = B\,B^\top$ into a standard normal random vector $\eta$ with independent components. The following fact is proved as Lemma 1 in [46].

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

To apply the proposition, one may choose $A = L_C$ since computing the standard Cholesky matrix $L_C$ requires only $\frac{1}{6}d^3$ operations. Then any other decomposition matrix $B$ with $\Sigma = B\,B^\top$ is of the form $B = L_C\,Q$ with some orthogonal matrix $Q$. The approach now consists in determining a *good* orthogonal matrix $Q$ such that the (mean) truncation dimension is minimized by exploiting the structure of the underlying integrand $f$. Such an approach is proposed in [11] for linear functions and refined and extended in [46].

Linear two-stage integrands are of the form (see proof of Proposition 4)

$$f(\xi) = G(\langle w^1, \xi\rangle + a_1(x), \ldots, \langle w^\ell, \xi\rangle + a_\ell(x)), \tag{28}$$

where $G(t_1, \ldots, t_\ell) = \max\{t_1, \ldots, t_\ell\}$, $w^j \in \mathbb{R}^d$, $w_i^j = v_i^j$, $i = 1, \ldots, d$, and

$$a_j(x) = -\langle v^j, Tx\rangle + \sum_{i=d+1}^{r} v_i^j \bar{h}_i,$$

the vertices $v^j \in \mathbb{R}^r$, $j = 1, \ldots, \ell$, of the dual polyhedron and $x \in X$, $\xi \in \mathbb{R}^d$. Hence, $f$ is of the form considered in [46] shortly after Theorem 3.
If $\xi$ is normal with mean $\mu$ and nonsingular covariance matrix $\Sigma$, we consider the decomposition $\Sigma = B\,B^\top$, with $B = L_C Q$ and some orthogonal matrix $Q$, and know that $\eta$ such that $\xi = L_C Q\eta + \mu$ has independent standard normally distributed components. The transformed function is

$$\hat{f}(\eta) = G(\langle B^\top w^1, \eta\rangle + \langle w^1, \mu\rangle + a_1(x), \ldots, \langle B^\top w^\ell, \eta\rangle + \langle w^\ell, \mu\rangle + a_\ell(x)). \tag{29}$$

It is suggested in [11, 46] to determine the orthogonal matrix $Q = (q_1, \ldots, q_d)$ with columns $q_j \in \mathbb{R}^d$, $j = 1, \ldots, d$, such that the (mean) truncation dimension of $\hat{f}$ in (29) is minimized. In the following we record a result proved as Theorem 2 in [46] (see also Proposition 1 in [11]).

**Proposition 6.** *Let $\ell = 1$ and set $w = w^1$. If the matrix $Q = (q_1, \ldots, q_d)$ is determined such that*

$$q_1 = \pm\frac{L_C^\top w}{\|L_C^\top w\|} \quad \textit{and} \quad Q \textit{ is orthogonal,} \tag{30}$$

*the decomposition $\Sigma = L_C Q\,(L_C Q)^\top$ leads to the transformed function*

$$\hat{f}(\eta) = G(\|L_C^\top w\|\eta_1 + \langle w, \mu\rangle + a_1(x)).$$

*Hence, the mean efficient dimension of $\hat{f}$ is $\bar{d}_T = 1$.*

The orthogonal columns $q_2, \ldots, q_d$ may be computed with $\frac{4}{3}d^3$ operations by the Householder transformation or fast Givens rotation starting from the canonical basis vectors $e_2, \ldots, e_d$ (see e.g. [12, Chapter 3]).

In case $1 < \ell \le d$ it is proposed in [46] to determine the orthogonal matrix $Q = (q_1, \ldots, q_d)$ by applying an orthogonalization technique to the matrix

$$M = (L_C^\top w^1, \ldots, L_C^\top w^\ell, b_{\ell+1}, \ldots, b_d), \tag{31}$$

where we assume that the $w^1, \ldots, w^\ell$ are linearly independent and $b_{\ell+1}, \ldots, b_d$ are selected such that $M$ has rank $d$. It is shown in [46, Theorem 3] that then the function $\hat{f}$ depends only on $\eta_1, \ldots, \eta_\ell$. The practical computation may

again be done by the Householder transformation or fast Givens rotation applied to the matrix $M$ in (31).
Applying the orthogonalization technique to the two-stage integrand (28) is not straightforward since the vertices $v^j$ of the dual polyhedron are not known in general and the computation of all of them is too expensive. Next we sketch a possible heuristic for dimension reduction of two-stage models.

*Remark 3.* (Dimension reduction heuristic)
Let us recall the structure of two-stage integrands from (28)

$$f(\xi) = \max\{\langle w^1, \xi\rangle + a_1(x), \ldots, \langle w^\ell, \xi\rangle + a_\ell(x)\},$$

where $x$ varies in the polyhedral set $X$, $w^j \in \mathbb{R}^d$, $w_i^j = v_i^j$, $i = 1, \ldots, d$, and $a_j(x) = -\langle v^j, Tx\rangle + \sum_{i=d+1}^r v_i^j \bar{h}_i$. The unknown vertices are again denoted by $v^j$, $j = 1, \ldots, \ell$.
An optimal vertex, i.e., a vertex $v^{j_*}$ such that

$$f(\xi) = \langle w^{j_*}, \xi\rangle + a_{j_*}(x)$$

depends on $\xi$ and $x$, i.e., $v^{j_*} = v^{j_*}(\xi, x)$.
A heuristic may be based on a small sample $\xi^j \in \mathbb{R}^d$, $j = 1, \ldots, n < d$, of the underlying probability distribution with uniform weights $\frac{1}{n}$. The two-stage linear stochastic program based on this sample is solved. Let $x^*$ be a first-stage solution and $v^j = v^j(\xi^j, x^*)$ a dual optimal vertex that corresponds to $\xi^j$, $j = 1, \ldots, n$. We compute the $w^j$, $j = 1, \ldots, n$ and select a linearly independent subset $w^j$, $j \in J \subseteq \{1, \ldots, n\}$. Next the orthogonal matrix $Q$ and the decomposition $\Sigma = L_C Q \left(L_C Q\right)^\top$ are computed based on $w^j$, $j \in J$. The transformed function $\hat{f}$ then hopefully has a lower dimension than $f$.

## 7 Conclusions

We have shown in Section 4 that all but one ANOVA terms of two-stage integrands are infinitely differentiable if all marginal densities belong to $C_b^\infty(\mathbb{R})$ and a geometric condition on the vertices is satisfied. The QMC quadrature error (5) allows to derive the following bound (by using the ANOVA decomposition and then the techniques in [8])

$$\Big|\int_{[0,1]^d} f(\xi)d\xi - \frac{1}{n}\sum_{j=1}^n f(\eta_j)\Big| \le \sum_{0<|u|}\Big|\int_{[0,1]^d} f_u(\xi^u)d\xi^u - \frac{1}{n}\sum_{j=1}^n f_u(\eta_j^u)\Big| \quad (32)$$

$$\le \sum_{0<|u|<d} \mathrm{Disc}_{n,u}(\eta_1^u, \ldots, \eta_n^u)\|f_u\| \quad (33)$$

$$+\Big|\int_{[0,1]^d} f_D(\xi)d\xi - \frac{1}{n}\sum_{j=1}^n f_D(\eta_j)\Big|, \quad (34)$$

where $\mathrm{Disc}_{n,u}$ is a discrepancy for $n$ points in $[0,1]^{|u|}$ and $\|f_u\|$ a compatible norm. In particular, this holds for the weighted norm $\|\cdot\|_\gamma$ given by (7) in the weighted tensor product Sobolev space (6) and the corresponding weighted $L_2$-discrepancy

$$\mathrm{Disc}^2_{n,u}(\eta^u_1,\ldots,\eta^u_n) = \gamma_u \int_{[0,1]^{|u|}} \mathrm{disc}^2_u(\xi^u) d\xi^u,$$

where the discrepancy disc is given by

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

Recalling the arguments in the introduction one may conclude that all terms in (33) converge with the optimal rate while the term in (34) also converges to 0 due to Proinov's convergence result [27] (as $f_D$ is continuous). In addition, one may hope, that this term is small as $f_D$ is small by assuming that the truncation dimension $d_T$ has been reduced (see Section 6) and is, thus, essentially smaller than $d$. Moreover, by recalling the results in [45], one may hope that the convergence rate for terms with $|u| \le d_T$ is even better (at least if $d_T$ is small) than for $|u| > d_T$. In the latter case, however, one can hope that $\|f_u\|$ is small.
Altogether, there are good reasons to conclude that recent Quasi-Monte Carlo methods (like randomly shifted and digitally shifted polynomial lattice rules) are efficient for two-stage linear stochastic programs (even if the programs are large scale) if they are accompanied by dimension reduction techniques.
The extension of the results to two-stage stochastic programs with stochastic costs and right-hand sides and to multi-stage models as well as the presentation of numerical experience is left to a projected paper.

*Acknowledgement.* 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. They extend their gratitude to Hernan Leövey (Humboldt-University Berlin) for a number of very useful discussions. The research of the first author is partially supported by a grant of Kisters AG and the second by the DFG Research Center Matheon at Berlin.

## References

1. J. Dick: Walsh spaces containing smooth functions and Quasi-Monte Carlo rules of arbitrary high order, *SIAM Journal Numerical Analysis* 46 (2008), 1519–1553.
2. J. Dick and F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, 2010.
3. 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.

4. S. Graf and H. Luschgy: *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.
5. M. Griebel and M. Holtz: Dimension-wise integration of high-dimensional functions with applications to finance, *Journal of Complexity* 26 (2010), 455–489.
6. M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of the ANOVA decomposition, *Journal of Complexity* 26 (2010), 523–551.
7. M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* (to appear).
8. F. J. Hickernell: A generalized discrepancy and quadrature error bound, *Mathematics of Computation* 67 (1998), 299-322.
9. W. Hoeffding: A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics* 19 (1948), 293–325.
10. 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.
11. 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.
12. C. Kanzow: *Numerical Methods for systems of linear equations* (in German), Springer, Berlin 2005.
13. 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.
14. 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.
15. F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.
16. R. Liu and A. B. Owen: Estimating mean dimensionality of analysis of variance decompositions, *Journal of the American Statistical Association* 101 (2006), 712–721.
17. P. L'Ecuyer: Quasi-Monte Carlo methods with applications in finance, *Finance and Stochastics* 13 (2009), 307–349.
18. C. Lemieux: *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, 2009.
19. H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
20. F. Nožička, J. Guddat, H. Hollatz and B. Bank: *Theory of Linear Parametric Programming* (in German), Akademie-Verlag, Berlin 1974.
21. 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.
22. A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.
23. 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.
24. G. Pagès: A space quantization method for numerical integration, *Journal Computational and Applied Mathematics* 89 (1997), 1–38.

25. T. Pennanen and M. Koivu: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.
26. G. Ch. Pflug and A. Pichler: Scenario generation for stochastic optimization problems, in: Stochastic Optimization Methods in Finance and Energy (M.I. Bertocchi, G. Consigli, M.A.H. Dempster eds.), Springer, 2011.
27. P. D. Proinov: Discrepancy and integration of continuous functions, *Journal Approximation Theory* 52 (1998), 121–131.
28. S. T. Rachev and L. Rüschendorf: *Mass Transportation Problems*, Vol. I, Springer, New York, 1998.
29. 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.
30. W. Römisch: Scenario generation, in: Wiley Encyclopedia of Operations Research and Management Science (J.J. Cochran ed.), Wiley, 2010.
31. A. Ruszczyński and A. Shapiro (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.
32. A. Shapiro, D. Dentcheva and A. Ruszczyński: *Lectures on Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2009.
33. I. H. Sloan and H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.
34. I. H. Sloan, F. Y. Kuo and S. Joe: On the step-by-step construction of Quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces, *Mathematics of Computation* 71 (2002), 1609-1640.
35. I. M. Sobol': *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).
36. I. M. Sobol': Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* 55 (2001), 271–280.
37. 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.
38. A. Takemura: Tensor analysis of ANOVA decomposition, *Journal of the American Statistical Association* 78 (1983), 894-900.
39. D. Walkup and R. J-B Wets: Lifting projections of convex polyedra, *Pacific Journal of Mathematics* 28 (1969), 465–475.
40. S. W. Wallace and W. T. Ziemba (Eds.): *Applications of Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2005.
41. X. Wang: Tractability of multivariate integration using Quasi-Monte Carlo algorithms, *Mathematics of Computation* 72 (2003), 823–838.
42. X. Wang and K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.
43. 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.
44. 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.

45. 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.
46. 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.
47. R. J-B Wets: Stochastic programs with fixed recourse: The equivalent deterministic program, *SIAM Review* 16, 309–339.