ANOVA decomposition of convex piecewise linear functions

W. Römisch

Abstract Piecewise linear convex functions arise as integrands in stochastic programs. They are Lipschitz continuous on their domain, but do not belong to tensor product Sobolev spaces. Motivated by applying Quasi-Monte Carlo methods we show that all terms of their ANOVA decomposition, except the one of highest order, are smooth if the underlying densities are smooth and a certain geometric condition is satisfied. The latter condition is generically satisfied in the normal case.

1 Introduction

During the last decade much progress has been achieved in *Quasi-Monte Carlo (QMC) theory* for computing multidimensional integrals. Appropriate function spaces of integrands were discovered that allowed to improve the classical convergence rates. It is referred to the monographs [27, 17] for providing an overview of the earlier work and to [16, 2, 12] for presenting much of the more recent achievements.
In particular, certain reproducing kernel Hilbert spaces $\mathbb{F}_d$ of functions $f : [0,1]^d \to \mathbb{R}$ became important for estimating the quadrature error (see [7]). If the integral $I_d(f) = \int_{[0,1]^d} f(x)dx$ defines a linear continuous functional on $\mathbb{F}_d$ and $Q_{n,d}(f)$ denotes a Quasi-Monte Carlo method for computing $I_d(f)$, i.e.,

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

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

for some sequence $x_i \in [0,1)^d$, $i \in \mathbb{N}$, the quadrature error $e_n(\mathbb{F}_d)$ allows the representation

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

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

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

where $K : [0,1]^d \times [0,1]^d \to \mathbb{R}$ denotes the kernel of $\mathbb{F}_d$. It satisfies the conditions $K(\cdot,y) \in \mathbb{F}_d$ and $\langle f, K(\cdot,y)\rangle = f(y)$ for each $y \in [0,1]^d$ and $f \in \mathbb{F}_d$, where $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote inner product and norm in $\mathbb{F}_d$.
In particular, the weighted tensor product Sobolev space [25]

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

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

$$\langle f, g\rangle_\gamma = \sum_{u\subseteq\{1,\ldots,d\}} \prod_{j\in u} \gamma_j^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial x^u} f(x^u, 1^{-u}) \frac{\partial^{|u|}}{\partial x^u} g(x^u, 1^{-u}) dx^u, \qquad (3)$$

and a weighted Walsh space consisting of Walsh series (see [2, Example 2.8] and [1]) are reproducing kernel Hilbert spaces.
They became important for analyzing the recently developed randomized lattice rules (see [26, 11, 13] and [1, 2]) and allowed for deriving optimal error estimates of the form

$$e_n(\mathbb{F}_d) \leq C(\delta) n^{-1+\delta} \quad (n \in \mathbb{N}, \delta \in (0, \tfrac{1}{2}]), \qquad (4)$$

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

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

Unfortunately, a number of integrands 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 an integrand f to justify Quasi-Monte Carlo algorithms via the Koksma-Hlawka theorem [17, Theorem 2.11].

Often integrands are non-differentiable like those in option pricing models [31] or max-type functions in general. It has been discovered in [4, 5] that the so-called ANOVA decomposition (see Section 2) of such integrands may have a smoothing effect in the sense that many ANOVA terms are smooth if the underlying densities are sufficiently smooth.
In this paper we show that such a smoothing effect occurs also in case of piecewise linear convex functions f. More precisely, we show that all ANOVA terms except the one of highest order of such functions are infinitely differentiable if the densities are sufficiently smooth and a geometric property is satisfied. This geometric property is generic if the underlying densities are normal. The results pave the way to extensions for composite functions $f(g(\cdot))$ with a smooth mapping g. Since piecewise linear convex functions appear as the result of linear optimization processes, our results apply to linear two-stage stochastic programs and (slightly) extend the main result of [6]. Hence, the results justify earlier studies of QMC methods in stochastic programming [3, 9, 21] and motivate that the recently developed randomized lattice rules [26, 2] may be efficient for stochastic programming models if their superposition dimension is small. The computational experience reported in [6] confirms the efficiency of randomly shifted lattice rules.
The paper starts by recalling the ANOVA decomposition in Section 2 and convex piecewise linear functions in Section 3. Section 4 contains the main results on the smoothing effect of the ANOVA decomposition of convex piecewise linear functions, followed by discussing the generic character of the geometric property (Section 5) and dimension reduction (Section 6) both in the normal case.

2 ANOVA decomposition and effective dimension

The analysis of variance (ANOVA) decomposition of a function was first proposed as a tool in statistical analysis (see [8] and the survey [29]). Later it was often used for the analysis of quadrature methods mainly on $[0,1]^d$. Here, we will use it on $\mathbb{R}^d$ equipped with a probability measure given by a density function ρ of the form

$$\rho(\xi) = \prod_{j=1}^{d} \rho_j(\xi_j) \quad (\forall \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d) \tag{5}$$

with continuous one-dimensional marginal densities ρ_j on $\mathbb{R}$. As in [5] we consider the weighted $\mathcal{L}_p$ space over $\mathbb{R}^d$, i.e., $\mathcal{L}_{p,\rho}(\mathbb{R}^d)$, with the norm

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

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

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

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

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

where the product sign means composition. Due to Fubini's theorem the ordering within the product is not important and $P_u f$ is constant with respect to all ξ_k, $k \in u$.
The ANOVA decomposition of $f \in \mathcal{L}_{1,\rho}(\mathbb{R}^d)$ is of the form [30, 14]

$$f = \sum_{u \subseteq I} f_u \tag{6}$$

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

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

It is known from [14] that the ANOVA terms are given explicitly in terms of the projections by

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u}(f) + \sum_{v \subset u} (-1)^{|u|-|v|} P_{u-v}(P_{-u}(f)), \tag{7}$$

where P_{-u} and P_{u-v} mean integration with respect to ξ_j, $j \in I \setminus u$ 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 [5, Theorem 2].
If f belongs to $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u \subseteq I}$ are orthogonal in the Hilbert space $\mathcal{L}_{2,\rho}(\mathbb{R}^d)$ (see e.g. [30]).
Let the variance of f be defined by $\sigma^2(f) = \|f - P_I(f)\|_{L_2}^2$. Then it holds

$$\sigma^2(f) = \|f\|_{2,\rho}^2 - (P_I(f))^2 = \sum_{\emptyset \neq u \subseteq I} \|f_u\|_{2,\rho}^2 =: \sum_{\emptyset \neq u \subseteq I} \sigma_u^2(f).$$

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 ξ^u in f. They are used to define sensitivity indices of a set $u \subseteq I$ for f in [28] and the dimension distribution of f in [18, 15].
For small $\varepsilon \in (0,1)$ ($\varepsilon = 0.01$ is suggested in a number of papers), the *effective superposition (truncation) dimension* $d_S(\varepsilon)$ ($d_T(\varepsilon)$) is defined by

$$d_S(\varepsilon) = \min\Big\{ s \in I : \sum_{|u| \le s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \ge 1 - \varepsilon \Big\}$$

$$d_T(\varepsilon) = \min\Big\{ s \in I : \sum_{u \subseteq \{1,\ldots,s\}} \frac{\sigma_u^2(f)}{\sigma^2(f)} \ge 1 - \varepsilon \Big\}$$

and it holds $d_S(\varepsilon) \le d_T(\varepsilon)$ and (see [30])

$$\Big\| f - \sum_{|u| \le d_S(\varepsilon)} f_u \Big\|_{2,\rho} \le \sqrt{\varepsilon}\sigma(f). \tag{8}$$

For linear functions f one has $\sigma_u(f) = 0$ for $|u| > 1$, $d_S(\varepsilon) = 1$, but $d_T(\varepsilon)$ may be close to d [18, 30]. For the simple convex piecewise linear function $f(\xi_1, \xi_2) = \max\{\xi_1, \xi_2\}$ on $[0,1]^2$ with the uniform distribution it holds $f_\emptyset = \frac{1}{3}$, $\sigma^2(f) = \frac{1}{18}$,

$$f_{\{i\}}(\xi_i) = -\frac{1}{2}\xi_i^2 + \xi_i - \frac{1}{3}, \;\; \sigma_{\{i\}}^2(f) = \frac{2}{45}, \;\; (i = 1,2), \;\; \sigma_{\{1,2\}}^2(f) = \frac{1}{90}.$$

Hence, we obtain $d_S(\varepsilon) = 2$ for $\varepsilon \in (0, \frac{1}{5})$ and the situation is entirely different for convex piecewise linear functions.

3 Convex piecewise linear functions

Convex piecewise linear functions appear as optimal value functions of linear programs depending on parameters in right-hand sides of linear constraints or in the objective function. In general, they are nondifferentiable and not of bounded variation in the sense of Hardy and Krause (for the latter see [19]). On the other hand, such functions enjoy structural properties which make them attractive for variational problems.
As in [22, Section 2.I] a function f from $\mathbb{R}^d$ to the extended reals $\bar{\mathbb{R}}$ is called *piecewise linear* on $D = \operatorname{dom} f = \{\xi \in \mathbb{R}^d : f(\xi) < \infty\}$ if D can be represented as the union of finitely many polyhedral sets relative to each of which $f(\xi)$ is given by $f(\xi) = a^\top \xi + \alpha$ for some $a \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$.

Proposition 1. *Let $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ be proper, i.e., $f(\xi) > -\infty$ and $D = \operatorname{dom} f$ be nonempty. Then the function f is convex and piecewise linear if and only* *if it has a representation of the form*

$$f(\xi) = \begin{cases} \max\{a_1^\top \xi + \alpha_1, \dots, a_\ell^\top \xi + \alpha_\ell\} & , \ \xi \in D, \\ \infty & , \ \xi \notin D, \end{cases} \tag{9}$$

for some $\ell \in \mathbb{N}$, $a_j \in \mathbb{R}^d$ *and* $\alpha_j \in \mathbb{R}$, $j = 1, \dots, \ell$. *Moreover,* D *is polyhedral and, if* $\operatorname{int} D$ *is nonempty,* D *may represented as the union of a finite collection of polyhedral sets* D_j, $j = 1, \dots, \ell$, *such that* $\operatorname{int} D_j \neq \emptyset$ *and* $\operatorname{int} D_j \cap \operatorname{int} D_{j'} = \emptyset$ *when* $j \neq j'$.

Proof. The two parts of the results are proved as Theorem 2.49 and Lemma 2.50 in [22, Section 2.I]. □

Example 1. (Linear two-stage stochastic programs)
We consider the linear optimization problem

$$\min\Big\{c^\top x + \mathbb{E}_P[q^\top y(\xi)] : Wy(\xi) + T(\xi)x = h(\xi),\ x \in X,\ y(\xi) \geq 0, \forall \xi \in \mathbb{R}^d\Big\},$$

where $c \in \mathbb{R}^m$, $q \in \mathbb{R}^{\bar{m}}$, W is a $r \times \bar{m}$ matrix, $T(\xi)$ a $r \times m$ matrix, $h(\xi) \in \mathbb{R}^r$ for each $\xi \in \mathbb{R}^d$, X is convex and polyhedral in $\mathbb{R}^m$, P is a probability measure on $\mathbb{R}^d$ and $\mathbb{E}_P$ denotes expectation with respect to P. We assume that $T(\cdot)$ and $h(\cdot)$ are affine functions of ξ. The above problem may be reformulated as minimizing a convex integral functional with respect to x, namely,

$$\min\Big\{c^\top x + \int_{\mathbb{R}^d} \Phi(h(\xi) - T(\xi)x) P(d\xi) : x \in X\Big\}, \tag{10}$$

where Φ is the optimal value function assigning to each parameter $t \in \mathbb{R}^r$ an extended real number by $\Phi(t) = \inf\{q^\top y : Wy = t, y \geq 0\}$. The value $\Phi(t) = -\infty$ appears if there exists $y \in \mathbb{R}^{\bar{m}}_+$, $y \neq 0$ such that $Wy = 0$ and $\Phi(t) = +\infty$ means infeasibility, i.e., $\{y \in \mathbb{R}^{\bar{m}} : Wy = t, y \geq 0\}$ is empty. The integrand in (10) is $f(\xi) = c^\top x + \Phi(h(\xi) - T(\xi)x)$ for every $x \in X$.
Now, we assume that both $\operatorname{dom} \Phi = \{t \in \mathbb{R}^r : \Phi(t) < +\infty\}$ and the dual polyhedron $\mathcal{D} = \{z \in \mathbb{R}^r : W^\top z \leq q\}$ are nonempty. Then $\Phi(t) > -\infty$ holds for all $t \in \mathbb{R}^r$ and the original primal as well as the dual linear program $\max\{t^\top z : z \in \mathcal{D}\}$ are solvable due to the duality theorem. If v^j, $j = 1, \dots, l$, denote the vertices of $\mathcal{D}$, it holds

$$\Phi(t) = \max_{j=1,\dots,l} t^\top v^j \quad (t \in \operatorname{dom} \Phi = \mathbb{R}^r),$$

i.e., the integrand $f(\cdot)$ is convex and piecewise linear on $D = \mathbb{R}^d$ for every $x \in X$. For more information on stochastic programming see [23, 24].

4 ANOVA decomposition of convex piecewise linear functions

We consider a piecewise linear convex function f and assume that its polyhedral domain $D = \operatorname{dom} f$ has nonempty interior. Let D_j, $j = 1, \ldots, \ell$, be the polyhedral subsets of D according to Proposition 1 such that

$$f(\xi) = a_j^\top \xi + \alpha_j \quad (\forall \xi \in D_j)$$

holds for some $a_j \in \mathbb{R}^d$, $\alpha_j \in \mathbb{R}$, $j = 1, \ldots, \ell$. For each $i \in I = \{1, \ldots, d\}$ there exist finitely many $(d-1)$-dimensional intersections H_{ij}, $j = 1, \ldots, J(i)$, of D_i with adjacent polyhedral sets D_j, $j \in \{1, \ldots, d\} \setminus \{i\}$. These polyhedral sets are subsets of finitely many $(d-1)$-dimensional affine subspaces of $\mathbb{R}^d$ which are renumbered by H_i, $i = 1, \ldots, \theta(f)$.
Furthermore, we assume that the support Ξ of the probability measure is contained in D and its density ρ is of the form (5). For any $k \in I$ we denote the kth coordinate projection of D by $\pi_k(D)$, i.e.,

$$\pi_k(D) = \{\xi_k \in \mathbb{R} : \exists \xi_j,\ j \in I, j \neq k, \text{ such that } \xi = (\xi_1, \ldots, \xi_d) \in D\}.$$

Next we intend to compute projections $P_k(f)$ for $k \in I$. For $\xi \in D$ we set $\bar{\xi}^k = (\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_d)$, and $\bar{\xi}^k_s = (\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d)$ for $s \in \pi_k(D)$. We know that

$$\bar{\xi}^k_s \in \bigcup_{j=1}^{\ell} D_j = D \tag{11}$$

for every $s \in \pi_k(D)$ and assume $\rho_k(s) = 0$ for every $s \in \mathbb{R} \setminus \pi_k(D)$. Hence, we obtain by definition of the projection

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

Due to (11) the one-dimensional affine subspace $\{\bar{\xi}^k_s : s \in \mathbb{R}\}$ intersects a finite number of the polyhedral sets D_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 $j_i = j_i^k \in \{1, \ldots, \ell\}$, $i = 1, \ldots, p+1$, such that $s_i < s_{i+1}$ and

$$\begin{array}{ll} \bar{\xi}^k_s \in D_{j_1} & \forall s \in (-\infty, s_1] \cap \pi_k(D) \\ \bar{\xi}^k_s \in D_{j_i} & \forall s \in [s_{i-1}, s_i] \quad (i = 2, \ldots, p) \\ \bar{\xi}^k_s \in D_{j_{p+1}} & \forall s \in [s_p, +\infty) \cap \pi_k(D). \end{array}$$

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

$$
\begin{aligned}
(P_k f)(\bar{\xi}^k) &= \sum_{i=1}^{p+1} \int_{s_{i-1}}^{s_i} (a_{j_i}^\top \bar{\xi}_s^k + \alpha_{j_i}) \rho_k(s) ds & (12)\\
&= \sum_{i=1}^{p+1} \Big(\Big(\sum_{\substack{l=1\\ l\neq k}}^{d} a_{j_i l}\xi_l + \alpha_{j_i} \Big) \int_{s_{i-1}}^{s_i} \rho_k(s) ds + a_{j_i k} \int_{s_{i-1}}^{s_i} s\rho_k(s) ds \Big) \\
&= \sum_{i=1}^{p+1} \Big(\Big(\sum_{\substack{l=1\\ l\neq k}}^{d} a_{j_i l}\xi_l + \alpha_{j_i} \Big) (\varphi_k(s_i) - \varphi_k(s_{i-1})) \\
&\qquad + a_{j_i k}(\psi_k(s_i) - \psi_k(s_{i-1})) \Big) & (13)
\end{aligned}
$$

Here, φ_k is the one-dimensional distribution function with density ρ_k, ψ_k the corresponding mean value function and μ_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.

$$
\begin{aligned}
(P_k f)(\bar{\xi}^k) = \sum_{i=1}^{p} \Big(\Big(\sum_{\substack{l=1\\ l\neq k}}^{d} (a_{j_i l} - a_{j_{i+1} l})\xi_l + (\alpha_{j_i} - \alpha_{j_{i+1}}) \Big) \varphi_k(s_i) + \\
(a_{j_i k} - a_{j_{i+1} k}) \psi_k(s_i) \Big) + \sum_{\substack{l=1\\ l\neq k}}^{d} a_{j_{p+1} l}\xi_l + \alpha_{j_{p+1}} + a_{j_{p+1} k}\mu_k. \quad (14)
\end{aligned}
$$

As the convex function f is continuous on $\operatorname{int} D$, it holds

$$
a_{j_i}^\top \bar{\xi}_s^k + \alpha_{j_i} = a_{j_{i+1}}^\top \bar{\xi}_s^k + \alpha_{j_{i+1}}
$$

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

$$
\sum_{\substack{l=1\\ l\neq k}}^{d} \xi_l (a_{j_{i+1} l} - a_{j_i l}) + s_i (a_{j_{i+1} k} - a_{j_i k}) + \alpha_{j_{i+1}} - \alpha_{j_i} = 0 \;\; (i = 1, \ldots, p).
$$

This leads to the explicit formula

$$
s_i = \frac{1}{a_{j_i k} - a_{j_{i+1} k}} \Big(\sum_{\substack{l=1\\ l\neq k}}^{d} \xi_l (a_{j_{i+1} l} - a_{j_i l}) + \alpha_{j_{i+1}} - \alpha_{j_i} \Big) \;\text{ if }\; a_{j_i k} \neq a_{j_{i+1} k}. \quad (15)
$$

for $i = 1, \ldots, p$. Hence, all s_i, $i = 1, \ldots, p$, are linear combinations of the remaining components ξ_j, $j \neq k$, of ξ if the following *geometric condition* is satisfied: All kth components of adjacent vectors a_j are different from each

other, i.e., all polyhedral sets H_j are subsets of $(d-1)$-dimensional subspaces that are not parallel to the kth coordinate axis in $\mathbb{R}^r$ or, with other words, not parallel to the canonical basis element e_k (whose components are equal to δ_{ik}, $i = 1, \ldots, d$).
To simplify notation we set $w_i = a_{j_i} - a_{j_{i+1}}$ and $v_i = \alpha_{j_i} - \alpha_{j_{i+1}}$. If the above geometric condition is satisfied, we obtain the following representation of $P_k f$:

$$(P_k f)(\bar{\xi}^k) = \sum_{i=1}^{p} w_{ik} \Big(- s_i(\bar{\xi}^k)\varphi_k(s_i(\bar{\xi}^k)) + \psi_k(s_i(\bar{\xi}^k)) \Big) + \sum_{\substack{l=1 \\ l \neq k}}^{d} a_{j_{p+1}l}\xi_l + \alpha_{j_{p+1}} + a_{j_{p+1}k}\mu_k \tag{16}$$

$$s_i = s_i(\bar{\xi}^k) = -\frac{1}{w_{ik}} \Big(\sum_{\substack{l=1 \\ l \neq k}}^{d} w_{il}\xi_l + v_i \Big). \tag{17}$$

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

Proposition 2. *Let f be piecewise linear convex having the form*

$$f(\xi) = a_j^\top \xi + \alpha_j \quad (\forall \xi \in D_j). \tag{18}$$

Let $k \in I$ and assume that vectors a_j belonging to adjacent polyhedral sets D_j have different kth components. Then the kth projection $P_k f$ is twice continuously differentiable. The projection $P_k f$ belongs to $C^{s+1}(\mathbb{R}^d)$ if the density ρ_k is in $C^{s-1}(\mathbb{R})$ $(s \in \mathbb{N})$. $P_k f$ is infinitely differentiable if the density ρ_k is in $C^\infty(\mathbb{R})$.

Proof. Let $l \in I$, $l \neq k$. The projection $P_k f$ is partially differentiable with respect to ξ_l and it holds

$$\begin{aligned}
\frac{\partial P_k f}{\partial \xi_l}(\bar{\xi}^k) &= \sum_{i=1}^{p} w_{ik} \frac{\partial}{\partial \xi_l} \Big(- s_i(\bar{\xi}^k)\varphi_k(s_i(\bar{\xi}^k)) + \psi_k(s_i(\bar{\xi}^k)) \Big) + a_{j_{p+1}l} \\
&= \sum_{i=1}^{p} w_{il} \big(\varphi_k(s_i(\bar{\xi}^k)) + s_i(\bar{\xi}^k)\varphi_k'(s_i(\bar{\xi}^k)) - \psi_k'(s_i(\bar{\xi}^k)) \big) + a_{j_{p+1}l} \\
&= \sum_{i=1}^{p} w_{il}\varphi_k(s_i(\bar{\xi}^k)) + a_{j_{p+1}l}
\end{aligned}$$

due to (16)–(17) and $\varphi_k'(s) = \rho_k(s)$ and $\psi_k'(s) = s\rho_k(s)$. Hence, the behavior of all first order partial derivatives of $P_k f$ only depends on the kth marginal distribution functions. The first order partial derivatives are continuous and again partially differentiable. The second order partial derivatives are of the form

$$\frac{\partial^2 P_k f}{\partial \xi_l \partial \xi_r}(\bar{\xi}^k) = \sum_{i=1}^{p} \frac{-w_{il} w_{ir}}{w_{ik}} \rho_k(s_i(\bar{\xi}^k))$$

and, thus, only depend on the marginal density ρ_k. Hence, $P_k f$ is twice continuously differentiable as ρ_k is continuous. If $\rho_k \in C^{s-1}(\mathbb{R})$ for some $s \in \mathbb{N}$, $P_k f$ belongs to $C^{s+1}(\mathbb{R}^d)$. If $\rho_k \in C^\infty(\mathbb{R})$, $P_k f$ is in $C^\infty(\mathbb{R}^d)$. □

Our next example shows that the geometric condition imposed in Proposition 2 is not superfluous.

Example 2. Let us consider the function

$$f(\xi) = \max\{\xi_1, -\xi_1, \xi_2\} \quad (\forall \xi = (\xi_1, \xi_2) \in \mathbb{R}^2)$$

on $D = \mathbb{R}^2$, i.e., we have $\alpha_1 = \alpha_2 = \alpha_3 = 0$ and $a_1 = (1, 0)^\top$, $a_2 = (-1, 0)^\top$ and $a_3 = (0, 1)^\top$. The decomposition of D according to Proposition 1 consists of

$$D_1 = \{\xi \in \mathbb{R}^2 : \xi_1 \geq 0, \xi_2 \leq \xi_1\}, \quad D_2 = \{\xi \in \mathbb{R}^2 : \xi_1 \leq 0, \xi_2 \leq -\xi_1\},$$
$$D_3 = \{\xi \in \mathbb{R}^2 : \xi_2 \geq \xi_1, \xi_2 \geq -\xi_1\}.$$

All polyhedral sets are adjacent and the second component of two of the vectors a_j, $j = 1, 2, 3$, coincides. Hence, the geometric condition in Proposition 2 is violated. Indeed, the projection $P_2 f$ is of the form

$$(P_2 f)(\xi_1) = |\xi_1| \int_{-\infty}^{|\xi_1|} \rho(\xi_2) d\xi_2 + \int_{|\xi_1|}^{+\infty} \xi_2 \rho(\xi_2) d\xi_2$$

and, thus, nondifferentiable on $\mathbb{R}$ (see also [6, Example 3]).

The previous result extends to more general projections P_u.

Proposition 3. *Let $\emptyset \neq u \subseteq I$, f be given by (18) and the vectors a_j belonging to adjacent polyhedral sets D_j have kth components which are all different for some $k \in u$. Then the projection $P_u f$ is continuously differentiable. The projection $P_u f$ 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 $k, r \in I$, $k \neq r$, we obtain from the Leibniz theorem [5, Theorem 1] for $l \notin u$

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

and from the proof of Proposition 2

$$D_l P_u f(\xi^u) = \sum_{i=1}^{p} w_{il} \int_{\mathbb{R}} \varphi_k(s_i(\bar{\xi}^k)) \rho_r(\xi_r) d\xi_r + a_{j_{p+1} l} \, .$$

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

The following is the main result of this section.

Theorem 1. *Let $u \subset I$, f given by (18) and the vectors a_j belonging to adjacent polyhedral sets D_j have kth components which are all different for some $k \in -u = I \setminus u$. Then the ANOVA term f_u is infinitely differentiable if $\rho_k \in C_b^\infty(\mathbb{R})$.*

Proof. According to formula (7) it holds

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

and Proposition 3 implies that $P_{-u}f$ is infinitely differentiable. The result follows from the Inheritance Theorem [5, Theorem 2] applied to $P_{u-v}(P_{-u}(f))$ for each $v \subset u$. □

Corollary 1. *Let f be given by (18) and the following geometric condition (GC) be satisfied: All $(d-1)$-dimensional subspaces containing $(d-1)$-dimensional intersections of adjacent polyhedral sets D_j are not parallel to any coordinate axis. Then the ANOVA approximation*

$$f_{d-1} := \sum_{u \subset I} f_u \tag{19}$$

of f is infinitely differentiable if all densities ρ_k, $k \in I$, belong to $C_b^\infty(\mathbb{R})$.

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

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

$$f = f_{d-1} + f_I \quad \text{and} \quad \|f - f_{d-1}\|_{L_2}^2 = \|f_I\|_{L_2}^2 = \sigma_I^2(f)$$

according to (6). Hence, the question arises: For which convex piecewise linear functions f is $\sigma_I^2(f)$ small or, in terms of the effective superposition dimension $d_S(\varepsilon)$ of f, is $d_S(\varepsilon)$ smaller than d (see also (8))?

5 Generic smoothness in the normal case

We consider the convex, piecewise linear function

$$f(\xi) = \max\{a_1^\top \xi + \alpha_1, \ldots, a_\ell^\top \xi + \alpha_\ell\} \quad (\forall \xi \in \mathbb{R}^d)$$

on $\operatorname{dom} f = \mathbb{R}^d$ and assume that ξ is normal with mean μ and nonsingular covariance matrix Σ. Then there exists an orthogonal matrix Q such that $\Delta = Q\,\Sigma\,Q^\top$ is a diagonal matrix. Then the d-dimensional random vector η given by

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

is normal with zero mean and covariance matrix Δ, i.e., has independent components. The transformed function $\hat{f}$

$$\hat{f}(\eta) = f(Q\eta+\mu) = \max_{j=1,\ldots,\ell}\{a_j^\top(Q\eta+\mu)+\alpha_j\} = \max_{j=1,\ldots,\ell}\{(Q^\top a_j)^\top \eta + a_j^\top \mu + \alpha_j\}$$

is defined on the polyhedral set $Q^\top D - Q^\top \mu$ and it holds

$$\hat{f}(\eta) = (Q^\top a_j)^\top \eta + a_j^\top \mu + \alpha_j \quad \text{for each } \eta \in Q^\top(D_j - \mu).$$

We consider now $(d-1)$-dimensional intersections H_{ij} of two adjacent polyhedral sets D_i and D_j, $i, j = 1, \ldots, \ell$. They are polyhedral subsets of $(d-1)$-dimensional affine subspaces H_i. The orthogonal matrix $Q^\top$ causes a rotation of the sets H_{ij} and the corresponding affine subspaces H_i. However, there are only countably many orthogonal matrices Q such that the geometric condition (GC) (see Corollary 1) on the subspaces 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 that satisfy the geometric condition, is a *residual set*, 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

Corollary 2. *Let f be a piecewise linear convex function on $\operatorname{dom} f = \mathbb{R}^d$ and let ξ be normally distributed with nonsingular covariance matrix. Then the infinite differentiability of the ANOVA approximation f_{d-1} of f (given by (19)) is a generic property.*

Proof. Let μ be the mean vector and Σ be the nonsingular covariance matrix of ξ. Let Q be the orthogonal matrix satisfying $Q\,\Sigma\,Q^\top = \Delta = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_d^2)$ and ρ be the normal density with mean μ and covariance matrix Δ. Then $\sigma_j > 0$, $j = 1, \ldots, d$, and ρ is equal to the product of all one-dimensional marginal densities ρ_k, where

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

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

6 Dimension reduction of piecewise linear convex functions

In order to replace a piecewise linear convex function f by the sum f_{d-1} of ANOVA terms until order $d-1$ (see Corollary 1), we need that the effective superposition dimension d_S of f is is smaller than d. Hence, one is usually interested in determining and reducing the effective dimension. This topic is discussed in a number of papers, e.g., [3, 15, 18, 28, 30, 32].
In the *normal* or *lognormal case* there exist universal (i.e., independent on the structure of f) and problem dependent principles for dimension reduction.
A universal principle for dimension reduction is *principal component analysis* (PCA). In PCA one uses the decomposition $\Sigma = U_P U_P^\top$ of Σ with the matrix $U_P = (\sqrt{\lambda_1}u_1, \ldots, \sqrt{\lambda_d}u_d)$, the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d > 0$ of Σ in decreasing order and the corresponding orthonormal eigenvectors u_i, $i = 1, \ldots, d$, of Σ. Several authors report an enormous reduction of the effective truncation dimension in financial models if PCA is used (see, for example, [30, 31]). However, PCA may become expensive for large d and the reduction effect depends on the eigenvalue distribution.
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 following equivalence principle is [32, Lemma 1] and already mentioned in [20, p. 182].

Proposition 4. *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 it holds $\Sigma = B B^\top$ if and only if B is of the form $B = A Q$ with some orthogonal $d \times d$ matrix Q.*

To apply the proposition, one may choose $A = L_C$, where L_C is the standard Cholesky matrix, or $A = U_P$. Then any other decomposition matrix B with $\Sigma = B B^\top$ is of the form $B = A Q$ with some orthogonal matrix Q.
A dimension reduction approach now consists in determining a *good* orthogonal matrix Q such that the truncation dimension is minimized by exploiting the structure of the underlying integrand f. Such an approach is proposed in [10] for linear functions and refined and extended in [32].
Piecewise linear convex functions are of the form

$$f(\xi) = G(a_1^\top \xi + \alpha_1, \ldots, a_\ell^\top \xi + \alpha_\ell), \tag{20}$$

where $G(t_1, \ldots, t_\ell) = \max\{t_1, \ldots, t_\ell\}$. Hence, f is of the form as considered in [32] shortly after Theorem 3. The transformed function is

$$\hat{f}(\eta) = f(B\eta + \mu) = G((B^\top a_1)^\top \eta_1 + a_1^\top \mu + \alpha_1, \ldots, (B^\top a_\ell)^\top \eta_d + a_\ell^\top \mu + \alpha_\ell). \tag{21}$$

In order to minimize the truncation dimension of $\hat{f}$ in (21), the following result is recorded from [32, Theorem 2] (see also Proposition 1 in [10]).

Proposition 5. *Let $\ell = 1$. If the matrix $Q = (q_1, \ldots, q_d)$ is determined such that*

$$q_1 = \pm \frac{A^\top a_1}{\|A^\top a_1\|} \quad \text{and} \quad Q \text{ is orthogonal,} \tag{22}$$

the transformed function is

$$\hat{f}(\eta) = G(\|A^\top a_1\| \eta_1 + a_1^\top \mu + \alpha_1)$$

and has effective truncation dimension $d_T = 1$.

The orthogonal columns $q_2, \ldots, q_d$ may be computed by the Householder transformation. In case $1 < \ell \leq d$ it is proposed in [32] to determine the orthogonal matrix $Q = (q_1, \ldots, q_d)$ by applying an orthogonalization technique to the matrix

$$M = (A^\top a_1, \ldots, A^\top a_\ell, b_{\ell+1}, \ldots, b_d), \tag{23}$$

where we assume that the $a_1, \ldots, a_\ell$ are linearly independent and $b_{\ell+1}, \ldots, b_d$ are selected such that M has rank d. It is shown in [32, 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 applied to M in (23).

Acknowledgements The author wishes to express his gratitude to Prof. Ian Sloan (University of New South Wales, Sydney) for inspiring conversations during his visit to the Humboldt-University Berlin in 2011 and the referee for his/her constructive criticism. The research of the author is supported 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 on 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. M. Griebel, F. Y. Kuo and I. H. Sloan, The smoothing effect of the ANOVA decomposition. *Journal of Complexity* 26 (2010), 523–551.
5. 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.
6. H. Heitsch, H. Leövey and W. Römisch, Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs? *Stochastic Programming E-Print Series* 5-2012 (`www.speps.org`).

7. F. J. Hickernell, A generalized discrepancy and quadrature error bound. *Mathematics of Computation* 67 (1998), 299–322.
8. W. Hoeffding, A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* 19 (1948), 293–325.
9. 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.
10. 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.
11. 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.
12. 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. The ANZIAM Journal 53 (2011), 1–37.
13. 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.
14. F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski, On decomposition of multivariate functions. *Mathematics of Computation* 79 (2010), 953–966.
15. R. Liu and A. B. Owen, Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association* 101 (2006), 712–721.
16. C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling.* Springer, 2009.
17. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia, 1992.
18. A. B. Owen, The dimension distribution and quadrature test functions. *Statistica Sinica* 13 (2003), 1–17.
19. 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.
20. A. Papageorgiou, The Brownian bridge does not offer a consistent advantage in Quasi-Monte Carlo integration, *Journal of Complexity* 18 (2002), 171–186.
21. T. Pennanen and M. Koivu, Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerische Mathematik* 100 (2005), 141–163.
22. R. T. Rockafellar and R. J-B. Wets, *Variational Analysis.* Springer, Berlin 1998.
23. A. Ruszczyński and A. Shapiro (Eds.), *Stochastic Programming.* Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.
24. A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming.* MPS-SIAM Series on Optimization, Philadelphia, 2009.
25. I. H. Sloan and H. Woźniakowski, When are Quasi Monte Carlo algorithms efficient for high-dimensional integration. *Journal of Complexity* 14 (1998), 1–33.
26. 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.
27. I. M. Sobol', *Multidimensional Quadrature Formulas and Haar Functions.* Nauka, Moscow, 1969 (in Russian).
28. I. M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* 55 (2001), 271–280.
29. A. Takemura, Tensor analysis of ANOVA decomposition. *Journal of the American Statistical Association* 78 (1983), 894–900.
30. X. Wang and K.-T. Fang, The effective dimension and Quasi-Monte Carlo integration. *Journal of Complexity* 19 (2003), 101–124.
31. X. Wang and I. H. Sloan, Why are high-dimensional finance problems often of low effective dimension. *SIAM Journal on Scientific Computing* 27 (2005), 159–183.

32. 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.