Quasi-Monte Carlo methods for two-stage stochastic programs: Mixed-integer models

H. Leövey · W. Römisch

Abstract We consider randomized QMC methods for approximating the expected recourse in two-stage stochastic optimization problems containing mixed-integer decisions in the second stage. It is known that the second-stage optimal value function is piecewise linear-quadratic with possible kinks and discontinuities at the boundaries of certain convex polyhedral sets. This structure is exploited to provide conditions implying that first and higher order ANOVA terms of the integrand have mixed first order partial derivatives in the sense of Sobolev. This leads to an approximate smoothing of the integrand and to good convergence rates of randomized QMC methods if the effective superposition dimension is low.

1 Introduction

Optimizing a mathematical model containing integer variables and uncertainty in constraints often leads to mixed-integer two-stage stochastic programs. Such stochastic programs belong to the most complicated optimization problems due to multivariate integrals and discontinuous integrands (see [28, 42]). For some time Monte Carlo methods appeared as the only numerical integration technique that applies to such optimization models [18]. Motivated by recent developments in Quasi-Monte Carlo (QMC) theory and practice, we study here the applicability of randomized QMC methods (see [35, 36] and the survey [24]) and extend our earlier work [25] for two-stage models without integer decisions considerably.

In recent years much progress has been achieved in the construction and analysis of QMC methods for computing high-dimensional integrals. For example, it is known that certain randomized QMC methods can achieve almost the optimal convergence rate $O(n^{-1/2})$ if the integrands belong to certain weighted tensor product Sobolev spaces on the unit cube $[0,1]^d$ or on $\mathbb{R}^d$. We refer to the monograph [7], the survey [6] and the state-of-the-art [21] for presenting recent developments.

Structured Energy Management Team, Axpo, Baden, Switzerland, E-mail: hernaneugenio.leoevey@axpo.com
Humboldt-University Berlin, Institute of Mathematics, Berlin, Germany, E-mail: romisch@math.hu-berlin.de
Although randomized QMC methods display their fast convergence for integrands belonging to weighted tensor product spaces of functions containing mixed first Sobolev derivatives, there exist convergence studies also for functions with kinks [12] and discontinuities [13,14]. The performance of randomized QMC methods may be significantly deteriorated for such functions. In [14], for example, the authors derive convergence rates for functions of the form \( g(x) \mathbf{1}_B(x), \ x \in [0,1]^d \), where \( g \) is smooth and \( B \) is convex polyhedral. They show that the convergence rate is much lower than optimal, but can be improved if some of the discontinuity faces of \( B \) are parallel to some coordinate axes (best case being all faces parallel to some coordinate axes).

The functions appearing as integrands in mixed-integer two-stage models are piecewise linear-quadratic with kinks and discontinuities at boundaries of convex polyhedral sets. However, the structure of the convex polyhedra is hidden in the problem data. Therefore, our approach is different and motivated by the papers [11,12]. We study the smoothness of lower order ANOVA terms of the integrands and show that they are indeed much smoother than the integrand itself under certain conditions. Hence, truncated ANOVA decompositions can lead to an approximate smoothing of the integrand if its effective superposition dimension (Section 3) is low. The mentioned conditions on the problem data are partly canonical, but contain also a geometric condition on the position of the implicitly given convex polyhedra. To prove the results in the present paper requires tools from polyhedral theory and Haar measure theory on the topological group of real orthogonal matrices. The latter is needed, for example, to show that for multivariate normal distributions the geometric condition is satisfied almost everywhere with respect to the Haar measure defined on the group of orthogonal matrices needed for transforming the covariance matrix to diagonal form. Altogether, we obtain that randomized QMC methods can successfully be applied to mixed-integer two-stage models if the densities are sufficiently smooth and the integrand has low effective dimension relative to the underlying probability distribution.

The paper starts by recalling the structure of integrands in mixed-integer two-stage stochastic programming models in Section 2 and the ANOVA decomposition of multivariate functions in Section 3. Section 4 contains the main results on the smoothing effect of truncated ANOVA decompositions of mixed-integer two-stage integrands, followed by an error analysis of randomized QMC methods for such integrands (Section 5). Section 6 contains the study of the geometric condition in the multivariate normal case and Section 7 serves for a discussion of our numerical experience reported in the accompanying paper [26].

2 Mixed-integer two-stage stochastic programs

Let us consider the two-stage mixed-integer stochastic program

\[
\min \left\{ \langle c, x \rangle + \int_{\mathbb{R}^d} \Phi(q(\xi), h(\xi) - T(\xi)x) \rho(\xi) d\xi : x \in X \right\},
\]

where \( \Phi \) is the infimum function of the second-stage mixed-integer linear program

\[
\Phi(u,t) := \inf \left\{ \langle u_1, y_1 \rangle + \langle u_2, y_2 \rangle : W_1 y_1 + W_2 y_2 \leq t, y_1 \in \mathbb{R}^{m_1}, y_2 \in \mathbb{Z}^{m_2} \right\}
\]
for all pairs \((u, t) \in \mathbb{R}^{m_1 + m_2} \times \mathbb{R}^r\), \(c \in \mathbb{R}^m\), \(X\) is a closed subset of \(\mathbb{R}^m\), \(W_1\) and \(W_2\) are \((r, m_1)\) and \((r, m_2)\)-matrices, respectively, \(q(\xi) \in \mathbb{R}^{m_1 + m_2}\), \(h(\xi) \in \mathbb{R}^r\), and the \((r, m)\)-matrix \(T(\xi)\) are affine functions of \(\xi \in \mathbb{R}^d\), and \(\rho\) is the probability density of a Borel probability measure \(\mathbb{P}\) on \(\mathbb{R}^d\).

The primal and dual feasible right-hand side sets for the second-stage program are

\[
\mathcal{T} = \{ t \in \mathbb{R}^r : \exists (y_1, y_2) \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2} \text{ such that } W_1 y_1 + W_2 y_2 \leq t \}, \quad \text{and} \\
\mathcal{U} = \{ u = (u_1, u_2) \in \mathbb{R}^{m_1 + m_2} : \exists v \in \mathbb{R}^r \text{ such that } W_1^T v = u_1, W_2^T v = u_2 \}.
\]

Clearly, \(\Phi(u, t)\) is finite for all \((u, t) \in \mathcal{U} \times \mathcal{T}\), it holds \((0, 0) \in \mathcal{U} \times \mathcal{T}\) and \(\Phi(0, t) = 0\) for any \(t \in \mathcal{T}\). While \(\mathcal{U}\) is a convex polyhedral cone in \(\mathbb{R}^{m_1 + m_2}\), the structure of \(\mathcal{T}\) is more complicated. The latter has the representation

\[
\mathcal{T} = \bigcup_{z \in \mathbb{Z}^{m_2}} (W_2 z + \mathcal{K}), \tag{3}
\]

where \(\mathcal{K}\) is the convex polyhedral cone

\[
\mathcal{K} = \{ t \in \mathbb{R}^r : \exists y_1 \in \mathbb{R}^{m_1} \text{ such that } W_1 y_1 \leq t \} = W_1 (\mathbb{R}^{m_1}) + \mathbb{R}_-^r. \tag{4}
\]

Specific cases are (i) \(W_2 = 0\) (pure continuous recourse) implying \(\mathcal{T} = \mathcal{K}\) and (ii) \(W_1 = 0\) (pure integer recourse) leading to \(\mathcal{K} = \mathbb{R}^r_+\).

Next we introduce the assumptions

(A1) The matrices \(W_1\) and \(W_2\) have only rational elements.

(A2) The cardinality of the set

\[
Z = \bigcup_{t \in \mathcal{T}} Z(t), \quad \text{where} \quad Z(t) = \{ y_2 \in \mathbb{Z}^{m_2} : \exists y_1 \in \mathbb{R}^{m_1} \text{ such that } W_1 y_1 + W_2 y_2 \leq t \},
\]

is finite, i.e., the number of integer decisions in (1) is finite.

It is known that the set \(\mathcal{T}\) is always connected (i.e., there exists a polygon connecting two arbitrary points of \(\mathcal{T}\)) and closed if (A1) is satisfied (see [3, Theorems 5.6.1 and 5.6.2]). The representation (3) implies that \(\mathcal{T}\) can be decomposed into subsets of the form

\[
\mathcal{T}(t_0) := \{ t \in \mathcal{T} : Z(t) = Z(t_0) \} = \bigcap_{z \in Z(t_0)} (W_2 z + \mathcal{K}) \setminus \bigcup_{z \in Z \setminus Z(t_0)} (W_2 z + \mathcal{K}) \tag{5}
\]

for each fixed \(t_0 \in \mathcal{T}\). Condition (A1) implies that the intersection in (5) may be replaced by \(t + \mathcal{K}\) for some \(t \in \mathcal{T}\) (see [3, Lemma 5.6.1]).

Hence, if (A1) is satisfied, there exist a finite subset \(N\) of \(N\) and elements \(t_i \in \mathcal{T}\) and \(z_{ij} \in \mathbb{Z}^{m_2}\) for \(i \in N\) and \(j\) belonging to a finite subset \(N_i\) of \(N_i\), such that \(\mathcal{T}\) admits the representation

\[
\mathcal{T} = \bigcup_{i \in N} \mathcal{T}(t_i) \quad \text{with} \quad \mathcal{T}(t_i) = (t_i + \mathcal{K}) \setminus \bigcup_{j \in N_i} (W_2 z_{ij} + \mathcal{K}). \tag{6}
\]

The sets \(\mathcal{T}(t_i), \ i \in N,\) are nonempty and connected (even star-shaped cf. [3, Theorem 5.6.3]), but nonconvex in general. If for some \(i \in N\) the set \(\mathcal{T}(t_i)\) is nonconvex, it can be decomposed into a finite number of disjoint subsets whose closures are convex polyhedra with facets parallel to suitable facets of \(\mathcal{K}\). By renumbering all
such subsets (for every \(i \in N\)) one obtains a finite index set which is again denoted by \(N\) and subsets \(B_i, i \in N\), forming a partition of \(\mathcal{T}\).

We will need the following result on optimal value functions of linear programs. For a given \((r, \mathbb{M})\)-matrix \(W\) we consider the function

\[
\Phi_L(u, t) = \inf \{ \langle u, y \rangle : Wy \leq t \} = \sup \{ \langle t, z \rangle : W^T z = u, z \leq 0 \}
\]

(7)

from \(\mathbb{R}^r \times \mathbb{R}^r\) to \(\mathbb{R}\). We define the primal and dual feasibility sets

\[
\mathcal{P} = W(\mathbb{R}^r) + \mathbb{R}^r_+ \quad \text{and} \quad \mathcal{D} = W^T(\mathbb{R}^r_+)
\]

and recall some well-known properties of \(\Phi_L\) (see [50] and [33]).

**Lemma 1** The function \(\Phi_L\) is finite and continuous on the \((\mathbb{M} + r)\)-dimensional convex polyhedral cone \(\mathcal{D} \times \mathcal{P}\) and there exist \((\mathbb{M}, r)\)-matrices \(C_j\) and \((\mathbb{M} + r)\)-dimensional convex polyhedral cones \(K_j, j = 1, \ldots, \ell\), such that

\[
\bigcup_{j=1}^\ell K_j = \mathcal{D} \times \mathcal{P} \quad \text{and} \quad \text{int } K_j \cap \text{int } K_{j'} = \emptyset, \quad j \neq j',
\]

\[
\Phi_L(u, t) = \max_{j=1, \ldots, \ell} \langle C_j u, t \rangle \quad ((u, t) \in \mathcal{D} \times \mathcal{P}),
\]

\[
\Phi_L(u, t) = \langle C_j u, t \rangle, \quad \text{for each } (u, t) \in K_j, \quad j = 1, \ldots, \ell.
\]

The function \(\Phi_L(u, \cdot)\) is convex on \(\mathcal{P}\) for each \(u \in \mathcal{D}\), and \(\Phi_L(\cdot, t)\) is concave on \(\mathcal{D}\) for each \(t \in \mathcal{P}\). Furthermore, the intersection \(K_j \cap K_{j'}, j \neq j'\), is either equal to \(\{0\}\) or contained in a \((\mathbb{M} + r - 1)\)-dimensional subspace of \(\mathbb{R}^{\mathbb{M} + r}\) if the two cones are adjacent.

Now we are in the position to prove the following result on the representation and properties of the infimum function \(\Phi\) (see also [28, (2.10)] for the case of fixed \(u\)).

**Lemma 2** Assume (A1) and (A2). Then there exists a finite set \(N\) and Borel sets \(B_i, i \in N\), such that \(\mathcal{T} = \bigcup_{i \in N} B_i\), and the closures of \(B_i\) are convex polyhedral with facets parallel to suitable facets of \(\mathcal{K} = W_1(\mathbb{R}^{m_1}) + \mathbb{R}^r_+\).

The function \(\Phi\) is lower semicontinuous on \(\mathcal{U} \times \mathcal{T}\) and there exist \((r, m_1)\) matrices \(C_j, j = 1, \ldots, \ell, \ell \in \mathbb{N}\), such that

\[
\Phi(u, t) = \min_{y_2 \in Z_i(t)} \left( \langle u_2, y_2 \rangle + \max_{j=1, \ldots, \ell} \langle C_j u_1, t - W_2 y_2 \rangle \right) \quad ((u, t) \in \mathcal{U} \times B_i),
\]

(8)

where \(Z_i(t) = Z(t)\) is fixed for \(t \in B_i, i \in N\). \(\Phi\) is continuous on \(\mathcal{U} \times B_i\) for each \(i \in N\) and there exists a constant \(C > 0\) and such that

\[
|\Phi(u, t)| \leq C \max \{1, \|t\|\} \max \{1, \|u\|\}
\]

(9)

holds for all pairs \((u, t) \in \mathcal{U} \times \mathcal{T}\).

**Proof** The existence of the sets \(B_i\) and their properties are discussed after equation (6). The lower semicontinuity of \(\Phi\) follows from general results in parametric optimization, for example, [3, Theorem 4.2.1]. Next we prove the representation (8) of \(\Phi\). Due to the above construction the set \(Z(t)\) remains constant for all \(t \in B_i\). Hence, \(Z_i(t)\) is well defined and

\[
\Phi(u, t) = \inf_{y_2 \in Z_i(t)} \langle u_2, y_2 \rangle + \inf_{y_1 \in \mathbb{R}^{m_1}} \left\{ \langle u_1, y_1 \rangle : W_1 y_1 \leq t - W_2 y_2 \right\}
\]

(10)
holds for every \((u,t) \in U \times B_i\) and \(i \in N\). Due to Lemma 1 there exist \((r,m_1)\) matrices \(C_j\), \(j = 1, \ldots, \ell\), such that

\[
\inf_{y_1 \in \mathbb{R}^{m_1}} \{(u_1, y_1) : W_1 y_1 \leq t - W_2 y_2\} = \max_{j=1, \ldots, \ell} \langle C_j u_1, t - W_2 y_2 \rangle \quad ((u, t) \in U \times B_i).
\]

The first infimum in (10) is lower bounded and, hence, attained. Hence, one obtains

\[
\Phi(u,t) = \min_{y_2 \in \mathbb{Z}(t)} \left\{ \langle u_2, y_2 \rangle + \max_{j=1, \ldots, \ell} \langle C_j u_1, t - W_2 y_2 \rangle \right\}
\]

for every pair \((u,t) \in U \times B_i\). For the remaining statements we refer to [39]. □

For more information on the continuity properties of \(\Phi\) on \(U \times B_i\) for any \(i \in N\), we refer to [39]. Next we state our main representation result of the function \(\Phi\).

**Proposition 1** Assume (A1) and (A2). The function \(\Phi\) is finite and lower semicontinuous on \(U \times T\). There exists a finite decomposition of \(U \times T\) consisting of Borel sets \(U_\nu \times B_\nu\), \(\nu \in \mathcal{N}\), such that their closures are convex polyhedral and \(\Phi\) is bilinear on each \(U_\nu \times B_\nu\). More precisely, there exist \((r, m_1)\) matrices \(C_\nu\) and elements \(z_\nu \in \mathbb{Z}^{m_2}\) such that \(\Phi\) is of the form

\[
\Phi(u,t) = \langle u_2 - W_2^T C_\nu u_1, z_\nu \rangle + \langle C_\nu u_1, t \rangle
\]

(11)

for each \((u,t) \in U_\nu \times B_\nu\). The function \(\Phi\) may have kinks or discontinuities at the boundaries of \(U_\nu \times B_\nu\), \(\nu \in \mathcal{N}\).

**Proof** We start from the representation (8) of \(\Phi\) on \(U \times B_i\) for some \(i \in N\) and derive a further partition of \(U \times B_i\). To this end we consider the sets \(Z_i(t) = \{z_k : k \in N_i(t)\}\) and \(V_{il}(t) = \{v \in \mathbb{R}^{m_2} : \langle v, z_l \rangle \leq \langle v, z_k \rangle, k \in N_i(t)\}\), for \(t \in B_i\), \(l \in N_i(t)\). In addition, we consider the \((r, m_1)\) matrices \(C_j\) and the polyhedral cones \(K_j, j = 1, \ldots, \ell\), needed in Lemma 2. More precisely, we need the projections \(pr_1\) and \(pr_2\) from \(\mathbb{R}^{m_1+r}\) to \(\mathbb{R}^r\) and \(\mathbb{R}^m\), respectively, and the fact that \(pr_1(K_j)\) and \(pr_2(K_j)\) are also polyhedral cones for each \(j = 1, \ldots, \ell\). For each \(i \in N\) we define the following subsets of \(U\) and of \(B_i\):

\[
U_{ijl} = \{u = (u_1, u_2) \in U : u_1 \in pr_1(K_j), u_2 = W_2^T C_j u_1 \in V_{il}\}
\]

\[
B_{ijl} = \{t \in B_i : t \in W_2 z_l + pr_2(K_j)\}
\]

for all \(i \in N, j = 1, \ldots, \ell\) and \(l \in N_i\). For any \((u,t) \in U_{ijl} \times B_{ijl}\) we obtain

\[
\Phi(u,t) = \min_{k \in N_i} \left\{ \langle u_2, z_k \rangle + \langle C_j u_1, t - W_2 z_k \rangle \right\} = \min_{k \in N_i} \langle u_2 - W_2^T C_j u_1, z_k \rangle + \langle C_j u_1, t \rangle
\]

\[
= \langle u_2 - W_2^T C_j u_1, z_l \rangle + \langle C_j u_1, t \rangle
\]

starting from (8) in Lemma 2, using Lemma 1 and the definition of \(V_{ij}\). Finally, we introduce a new index \(\nu\) varying in a new (finite) index set \(\mathcal{N}\) and a bijective mapping \(\nu \leftrightarrow (i,j,l)\). By writing \(U_\nu\) instead of \(U_{ijl}\) and \(B_\nu\) instead of \(B_{ijl}\) we arrive at (11) by noting that \(C_\nu = C_j\) and \(z_\nu = z_l\) if \(\nu \leftrightarrow (i,j,l)\). We also note that the sets \(U_\nu\) and the closures of \(B_\nu\) are convex polyhedral. □
When defining the mixed-integer two-stage integrand \( f : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R} \) by
\[
f(x, \xi) = \begin{cases} 
(c, x) + \Phi(q(\xi), h(\xi) - T(\xi)x) + \delta(\xi) & , \text{if } (q(\xi), h(\xi) - T(\xi)x) \in U, \xi \in \mathcal{T}, \xi \in \mathcal{U}, \\
+ \infty & , \text{otherwise.}
\end{cases}
\] (12)
problem (1) may be rewritten as
\[
\min \left\{ \int_{\mathbb{R}^d} f(x, \xi) P(\xi) : x \in X \right\}.
\] (13)

We introduce the additional assumptions
(A3) For each pair \((x, \xi) \in X \times \mathbb{R}^d\) it holds \((q(\xi), h(\xi) - T(\xi)x) \in U \times \mathcal{T}\).
Condition (A3) refers to the standard requirements relatively complete recourse and
dual feasibility (see [43, Section 2.1]). The structural result for \( \Phi \) in Proposition 1
leads to the following representation of the integrand \( f \).

**Proposition 2** Assume (A1)-(A3) and let \( x \in X \). Then the integrand \( f \) is lower
semicontinuous on \( X \times \mathbb{R}^d \) and \( f(x, \cdot) \) is finite and linear-quadratic on the sets
\[
\Xi(\nu)(x) = \{ \xi \in \mathbb{R}^d : q(\xi) \in U(\nu), h(\xi) - T(\xi)x \in B(\nu) \}
\] (14)
for each \( \nu \in \mathcal{N} \), where \( N, U, U(\nu), B(\nu) \) are defined in Proposition 1. The function \( f(x, \cdot) \) is of the form
\[
f(x, \xi) = (c, x) + q_2(\xi) - W_2^T C_\nu q_1(\xi), z_\nu \) + \langle C_\nu q_1(\xi), h(\xi) - T(\xi)x \rangle
\] (15)
on the sets \( \Xi(\nu)(x) \), where the \((r, m_1)\) matrix \( C_\nu \) and \( \nu \in \mathbb{Z}^{m_2} \) are explained in
Proposition 1. The functions \( f(x, \cdot) \) may have points of discontinuity or nondifferentiability
at the boundaries of \( \Xi(\nu)(x) \). The union of all \( \Xi(\nu)(x) \) equals \( \mathbb{R}^d \) and their closures
are convex polyhedral. Moreover, the estimate
\[
|f(x, \xi)| \leq C \max \{1, \|x\|\} \max \{1, \|\xi\|^2\}
\] (16)
is valid for every pair \((x, \xi) \in X \times \mathbb{R}^d\) and some constant \( C > 0 \).

**Proof** The sets \( \Xi(\nu)(x), \nu \in \mathcal{N} \), form a partition of \( \mathbb{R}^d \) into Borel sets whose closures,
denoted by \( \text{cl} \Xi(\nu)(x) \), are of the form
\[
\text{cl} \Xi(\nu)(x) = \{ \xi \in \mathbb{R}^d : q(\xi) \in U(\nu), h(\xi) - T(\xi)x \in \text{cl} B(\nu) \}
\]
and, thus, are convex polyhedral, since \( h(\cdot), T(\cdot) \) and \( q(\cdot) \) are affine functions,
\( \text{cl} B(\nu) \), the closure of \( B(\nu) \), is convex polyhedral and \( U(\nu) \) is convex polyhedral, too.
The lower semicontinuity of \( f \) follows from Lemma 2.
The representation (15) of \( f(x, \xi) \) for every pair \((x, \xi) \in X \times \Xi(\nu)(x)\) and \( \nu \in \mathcal{N} \)
follows immediately from (11). Since \( q_1(\cdot), q_2(\cdot), h(\cdot) \) and \( T(\cdot) \) are affine functions
of \( \xi \), the second summand of (15) is an affine function of \( \xi \) while the third represents a
quadratic function. The final statement follows from (9) and the estimate
\[
|f(x, \xi)| \leq |(c, x)| + |\Phi(q(\xi), h(\xi) - T(\xi)x)|
\leq \|c\|\|x\| + \max \{1, \|h(\xi) - T(\xi)x\|\} \max \{1, \|q(\xi)\|\}
\leq \hat{C} \max \{1, \|x\|\} \max \{1, \|\xi\|^2\}
\] after a few calculations for all pairs \((x, \xi) \in X \times \mathbb{R}^d\) and some constant \( \hat{C} > 0 \). □
3 ANOVA decomposition and effective dimension

The analysis of variance (ANOVA) decomposition of a function was first proposed as a tool in statistical analysis (see [15] and the survey [48]). Later it was often used for the analysis of quadrature methods mainly on $[0, 1]^d$. Here, we use it on $\mathbb{R}^d$ equipped with a probability density function $\rho$ given in product form

$$\rho(\xi) = \prod_{k=1}^{d} \rho_k(\xi_k) \quad (\forall \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d).$$  \hspace{1cm} (17)

As in [12] we consider the weighted $L_p$ space over $\mathbb{R}^d$, i.e., $L_{p, \rho}(\mathbb{R}^d)$, with the norm

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

Let $f \in L_{1, \rho}(\mathbb{R}^d)$. The ANOVA projection $P_k$, $k \in \mathcal{D}$, is defined by

$$P_k f(\xi) := \int_{-\infty}^{\infty} f(\xi, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \rho_k(s) ds \quad (\xi \in \mathbb{R}^d).$$ \hspace{1cm} (18)

Clearly, the function $P_k f$ is constant with respect to $\xi_k$. For $u \subseteq \mathcal{D}$ we use $|u|$ for its cardinality, $-u$ for $\mathcal{D} \setminus u$ and define the higher order ANOVA projection by

$$P_u f = \left( \prod_{k \in u} P_k \right) (f),$$ \hspace{1cm} (19)

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 $\xi_k$, $k \in u$. The ANOVA decomposition of $f \in L_{1, \rho}(\mathbb{R}^d)$ is of the form [51, 23]

$$f = \sum_{u \subseteq \mathcal{D}} f_u,$$ \hspace{1cm} (20)

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

$$f_{\emptyset} = P_D f, \quad f_{\{k\}} = P_{-\{k\}} f, \quad f_u = P_{-u} f - \sum_{v \subseteq u} f_v, \quad u \subseteq \mathcal{D}.$$  

It is known from [23] that the ANOVA terms are given explicitly in terms of the ANOVA projections by

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u} f + \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{u-v}(P_{-u} f),$$ \hspace{1cm} (21)

where $P_{-v}$ and $P_{u-v}$ mean integration with respect to $\xi_j$, $j \in \mathcal{D} \setminus u$ and $j \in u \setminus v$, respectively. The first representation shows that lower order ANOVA terms $f_u$ with small $|u|$ are given by higher order projections. The second representation reveals
that the ANOVA term \(f_u\) is essentially as smooth as the ANOVA Projection \(P_u(f)\) due to the Inheritance theorem [12, Theorem 2]. If \(f\) belongs to \(L_{2,p}(\mathbb{R}^d)\), the projections \(P_u f\) and the ANOVA terms \(f_u\) also belong to \(L_{2,p}(\mathbb{R}^d)\), and the system \(\{f_u\}_{u \subseteq \mathcal{D}}\) is orthogonal in \(L_{2,p}(\mathbb{R}^d)\) (see e.g. [51]).

Let the variance of \(f\) be given by

\[
\sigma^2(f) = \|f - P_D(f)\|_{2,p}^2 = \|f\|_{2,p}^2 - (P_D(f))^2 = \sum_{\emptyset \neq u \subseteq \mathcal{D}} \|f_u\|_{2,p}^2 .
\]  

(22)

To avoid trivial cases we assume \(\sigma(f) > 0\) in the following. The normalized ratios \(\frac{\sigma^2_u(f)}{\sigma^2(f)}\), where \(\sigma_u(f) = \|f_u\|_{2,p}\), serve as indicators for the importance of the variable \(\xi^u\) in \(f\). They are used to define sensitivity indices of a set \(u \subseteq \mathcal{D}\) for \(f\) in [47] and the dimension distribution of \(f\) in [37, 27]. For small \(\varepsilon \in (0, 1)\) (\(\varepsilon = 0.01\) is suggested in a number of papers), the effective superposition (truncation) dimension \(d_S(\varepsilon) \in \mathcal{D}\) \((d_T(\varepsilon) \in \mathcal{D}\) of \(f\) is defined by

\[
d_S(\varepsilon) = \min \left\{ s \in \mathcal{D} : \sum_{0 < |u| \leq s} \frac{\sigma^2_u(f)}{\sigma^2(f)} \geq 1 - \varepsilon \right\}
\]

(23)

\[
d_T(\varepsilon) = \min \left\{ s \in \mathcal{D} : \sum_{u \subseteq \{1, \ldots, s\}} \frac{\sigma^2_u(f)}{\sigma^2(f)} \geq 1 - \varepsilon \right\}.
\]

(24)

We note that the effective superposition dimension \(d_S(\varepsilon)\) is important for the error analysis of Quasi-Monte Carlo methods, but its computation is complicated. The effective truncation dimension is computationally accessible (see [47, 51]). Note also that \(d_S(\varepsilon) \leq d_T(\varepsilon)\) holds and the estimate

\[
\max \left\{ \left\| f - \sum_{|u| \leq d_S(\varepsilon)} f_u \right\|_{2,p}, \left\| f - \sum_{u \subseteq \{1, \ldots, d_T(\varepsilon)\}} f_u \right\|_{2,p} \right\} \leq \sqrt{\varepsilon} \sigma(f)
\]

(25)

is valid (see [10, 51]). The estimate (25) means that the truncated ANOVA decomposition of \(f\) containing all ANOVA terms \(f_u\) until \(|u| \leq d_S(\varepsilon)\) (or \(|u| \leq d_T(\varepsilon)\)) represents an approximation of \(f\). The importance of (25) is due to the fact that lower order ANOVA terms of \(f\) may have smoothness properties even if \(f\) is known to be nondifferentiable or discontinuous (see [11, 12]). Then (25) may be used in error estimates by exploiting the eventual smoothness of the lower order ANOVA terms.

To formulate smoothness conditions we follow [12] and use the notation \(D_i f, i \in \mathcal{D}\), to denote the classical partial derivative \(\frac{\partial f}{\partial \xi_i}\). For a multi-index \(\alpha = (\alpha_1, \ldots, \alpha_d)\) with \(\alpha_i \in \mathbb{N}_0\) we set

\[
D^\alpha f = \prod_{i=1}^d D_i^{\alpha_i} f = \frac{\partial^{\alpha} f}{\partial \xi_1^{\alpha_1} \cdots \partial \xi_d^{\alpha_d}}.
\]

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

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

(26)
where $C_0^\infty(\mathbb{R}^d)$ denotes the space of infinitely differentiable functions with compact support in $\mathbb{R}^d$ and $D^\alpha v$ is the classical derivative of $v$. We will use the same symbol for the weak derivative as for the classical one, i.e., we set $D^\alpha f = g$ if (26) is satisfied, since classical derivatives are also weak derivatives. The latter follows because classical derivatives satisfy (26) which is just the multivariate integration by parts formula in the classical sense. We consider in the next sections the mixed Sobolev space of functions on $\mathbb{R}^d$

$$\mathcal{W}^{(1,\ldots,1)}_{2,\rho,\text{mix}}(\mathbb{R}^d) = \{ f \in L^2_{\rho}(\mathbb{R}^d) : D^\alpha f \in L^2_{\rho}(\mathbb{R}^d) \text{ if } \alpha_i \leq 1, i \in \mathcal{D} \}. \quad (27)$$

having weak mixed first order derivatives that are quadratically integrable. In [49] such spaces are called Sobolev spaces with dominating mixed smoothness.

4 ANOVA decomposition of mixed-integer two-stage integrands

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

$$f_\nu(\xi) := f(x,\xi) = \langle A_\nu(x)\xi,\xi \rangle + \langle b_\nu(x),\xi \rangle + c_\nu(x) \quad (28)$$

for all $(x,\xi) \in X \times \Xi_\nu(x)$ and some symmetric $(d,d)$-matrices $A_\nu(\cdot)$, $d$-dimensional vectors $b_\nu(\cdot)$ and real numbers $c_\nu(\cdot)$, which are all affine functions of $x$. The Borel sets $\Xi_\nu(x)$, $\nu \in \mathcal{N}$ are defined by (14) and have convex polyhedral closures.  

In addition to the conditions (A1)–(A3) we need to impose:

(A4) The probability distribution $\mathbb{P}$ has finite fourth order absolute moments. Due to (16) the mixed-integer two-stage stochastic program (1) is already well defined if $\mathbb{P}$ has finite second order moments. However, the stronger condition (A5) together with the next one enable the use of the concepts from Section 3.

(A5) $\mathbb{P}$ has a density $\rho$ with respect to the Lebesgue measure on $\mathbb{R}^d$ and $\rho$ admits product form

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

where the densities $\rho_i$ are positive and continuously differentiable, and $\rho_i$ and its derivative are bounded on $\mathbb{R}$. 

To apply the results in this section to general probability distributions $\mathbb{P}$, one has to decompose the dependence structure of $\mathbb{P}$. The latter is always possible using the multivariate distributional transform, which was first established in [40] in case that the conditional distribution functions of $\mathbb{P}$ are absolutely continuous. Later the distributional transform was extended to the general case (see [41]).

(A6) For each face $F$ of dimension greater than zero of the convex polyhedral sets $\text{cl} \Xi_\nu(x)$, $\nu \in \mathcal{N}$, the affine hull $\text{aff}(F)$ of $F$ does not parallel any coordinate axis $\text{in } \mathbb{R}^d$ for each $x \in X$ (geometric condition).

Recall that a face $F$ of a polyhedron $P$ in $\mathbb{R}^d$ is defined by $F = \{ \xi \in P : \langle a,\xi \rangle = b \}$ for some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $P$ is contained in the halfspace $\{ \xi \in \mathbb{R}^d : \langle a,\xi \rangle \leq b \}$. The face $F$ is said to be defined by the inequality $\langle a,\xi \rangle \leq b$. Clearly, each face is itself a polyhedron. The dimension $\text{dim}(P)$ of a polyhedron $P$ is the dimension of its affine hull $\text{aff}(P)$. A facet $F$ of a polyhedron $P$ with $P \neq F$ is a
face of dimension $\dim(P) - 1$. Vertices of polyhedra are faces of dimension zero.

For a short review of basic polyhedral theory we refer the reader to [17]. If $F$ is any face of a polyhedron $\text{cl}\,\Xi(x)$ for some $\nu \in \mathcal{N}$ defined by the inequality $(g, \xi) \leq a$ for some $g \in \mathbb{R}^d$ and $a \in \mathbb{R}$, then (A6) means that all components of $g$ do not vanish. This condition is important for deriving the results in this section. It will be illustrated in Example 1 and discussed in Section 6. Since the polyhedra $\text{cl}\,\Xi(x)$ are not explicitly given, condition (A6) has implicit character.

In the following we consider the ANOVA decomposition of $f = f_x$ for any fixed $x \in X$ and show that lower order ANOVA terms of $f$ are smoother than the function $f$ itself. Since the ANOVA terms are given in terms of (ANOVA) projections (see (21)), we study first properties of projections.

Let $u \subset \mathcal{D} = \{1, \ldots, d\}$ and fix $x \in X$. For $\xi \in \mathbb{R}^d$ we define

$$\Pi_u \xi := \xi^{-u},$$

where $\xi = (\xi^u, \xi^{-u})$ and $\xi^{-u} := (s^u, \xi^{-u})$, where $s \in \mathbb{R}^{|u|}$.

If $u = \{k\}$ for some $k \in \mathcal{D}$ we write $\xi^{-k}$ and $\xi^{-k}$ with $s \in \mathbb{R}$.

First we derive bounds for $P_u f$ where $f$ is given by (28).

**Proposition 3** Let (A1)–(A5) be satisfied, $x \in X$ be fixed, $u \subset \mathcal{D}$ and we consider an integrand $f = f_x$ of the form (28). Then there exists a constant $\tilde{C} > 0$ such that

$$|P_u f(\xi^{-u})| \leq \tilde{C} \max\{1, \|x\|\} \max\{1, \|\xi^{-u}\|^2\}$$

holds for all $\xi^{-u} \in \Pi_u(\mathbb{R}^d)$.

**Proof** Using the representation $u = \{i_1, \ldots, i_{|u|}\}$ and the definition (19) of $P_u f$, we obtain (16)

$$P_u f(\xi^{-u}) = \int_{\mathbb{R}^{|u|}} f(s^u, \xi^{-u}) \prod_{k=1}^{|u|} \rho_{i_k}(s_{i_k}) ds^u$$

$$\leq C \max\{1, \|x\|\} \int_{\mathbb{R}^{|u|}} \max\{1, \|s^u\|^2 + \|\xi^{-u}\|^2\} \prod_{k=1}^{|u|} \rho_{i_k}(s_{i_k}) ds^u$$

$$\leq C \max\{1, \|x\|\} \left(1 + \|\xi^{-u}\|^2 + \sum_{j=1}^{|u|} s_{i_j}^2 \prod_{k=1}^{|u|} \rho_{i_k}(s_{i_k}) ds_{i_1} \cdots ds_{i_{|u|}}\right)$$

$$= C \max\{1, \|x\|\} \left(1 + \|\xi^{-u}\|^2 + \sum_{j=1}^{|u|} s_{i_j}^2 \rho_{i_j}(s_{i_j}) ds_{i_j}\right)$$

$$\leq \tilde{C} \max\{1, \|x\|\} \max\{1, \|\xi^{-u}\|^2\}$$

for some positive constant $\tilde{C}$ and all $\xi^{-u} \in \Pi_u(\mathbb{R}^d)$.

Next we study continuity and differentiability properties of projections and we start with first order projections $P_k f$ of $f = f_x$ for some $k \in \mathcal{D}$. We know that

$$\xi^{-k} \in \bigcup_{\nu \in \mathcal{N}} \Xi_{\nu}(x) = \mathbb{R}^d$$

(30)

holds for fixed $\xi^{-k}$ and every $s \in \mathbb{R}$. According to (18) we have

$$(P_k f)(\xi^{-k}) = \int_{-\infty}^{\infty} f(\xi^{-k}) \rho_k(s) ds = \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \rho_k(s) ds.$$
Due to (30) there exists a finite subset $\hat{\mathcal{N}} = \hat{\mathcal{N}}(\xi^{-k})$ of $\mathcal{N}$ such that the one-dimensional affine subspace $\{\xi s^{-k} : s \in \mathbb{R}\}$ intersects the sets $\mathcal{E}_\nu(x)$ for $\nu \in \mathcal{N}$, where $\text{cl} \mathcal{E}_\nu(x)$ and its adjacent sets have a common facet for every $\nu \in \mathcal{N}$. Hence, there exists a partition of $\mathbb{R}$ into subintervals $I_\nu = I_\nu(\xi^{-k})$, $\nu \in \mathcal{N}$, such that $\xi s^{-k} \in \mathcal{E}_\nu(x)$ for all $s \in I_\nu$ and $\nu \in \mathcal{N}$. We obtain the following representation of $P_k f$ by setting $f^\nu(x, \xi s^{-k}) := (A_\nu(x)\xi s^{-k}, \xi_s^{-k}) + (b_\nu(x), \xi_s^{-k}) + c_\nu(x)$ and using the identity $\xi s^{-k} = \xi 0^{-k} + s e_k$ with $e_k$ denoting the element of $\mathbb{R}^d$ having components $\delta_{ik}$, $i = 1, \ldots, d$:

\[
(P_k f)(\xi^{-k}) = \sum_{\nu \in \mathcal{N}} \int_{I_\nu} f^\nu(x, \xi s^{-k}) \rho_k(s) ds = \sum_{\nu \in \mathcal{N}} \left( \int_{I_\nu} f^\nu(x, \xi 0^{-k}) \rho_k(s) ds + \langle A_\nu(x)e_k, e_k \rangle \int_{I_\nu} s^2 \rho_k(s) ds + \langle 2A_\nu(x)\xi 0^{-k} + b_\nu(x), e_k \rangle \int_{I_\nu} s \rho_k(s) ds \right)
\]

\[
= \sum_{\nu \in \mathcal{N}} \sum_{j=0}^{2} p_{j,\nu}(\xi^{-k}) \int_{I_\nu} s^j \rho_k(s) ds = \sum_{\nu \in \mathcal{N}} \sum_{j=0}^{2} p_{j,\nu}(\xi^{-k}) \int_{s_{\nu}(\xi^{-k})}^{s_{\nu+1}(\xi^{-k})} s^j \rho_k(s) ds
\]

where we define $s_{\nu} = s_{\nu}(\xi^{-k}) = \inf I_\nu(\xi^{-k})$ and $s_{\nu+1} = s_{\nu+1}(\xi^{-k}) = \sup I_\nu(\xi^{-k})$, and the functions $p_{j,\nu}(\cdot)$ are $(d - 1)$-variate polynomials in $\xi^{-k}$ of degree $2 - j$ with coefficients being affine functions of $x$. If $s_{\nu}$ is finite, the point $\xi_s^{-k}$ belongs to the common facet of $\text{cl} \mathcal{E}_\nu(x)$ and $\text{cl} \mathcal{E}_{\nu-1}(x)$. Let $g_\nu = (g_{\nu,1}, \ldots, g_{\nu,d}) \in \mathbb{R}^d$ and $a_\nu \in \mathbb{R}$ be selected such that the facet is defined by the inequality

\[
\{g_\nu, \xi \} = \sum_{i=1}^{d} g_{\nu,i} \xi_i \leq a_\nu.
\]

Hence, for finite $s_{\nu}$ we obtain

\[
\{g_\nu, \xi^{-k}_{s_{\nu}} \} = \sum_{i=1}^{d} g_{\nu,i} \xi_i + g_{\nu,k} s_{\nu} = a_\nu.
\]

Since $g_{\nu,k} \neq 0$ due to condition (A6), we arrive at the representation

\[
s_{\nu} = s_{\nu}(\xi^{-k}) = \frac{1}{g_{\nu,k}} \left( - \sum_{i=1, i \neq k}^{d} g_{\nu,i} \xi_i + a_\nu \right).
\]

Since the points $s_{\nu} = s_{\nu}(\xi^{-k})$ are affine functions of $\xi^{-k}$ and the integrands $f^\nu(\cdot, \xi^{-k})$ are linear-quadratic in $I_\nu(\xi^{-k})$, classical results on integrals depending on parameters may be used to derive continuity and continuous differentiability of the projections $P_k f$ at any $\xi^{-k} \in I_\nu(\mathbb{R}^d)$ as functions of $\xi^{-k}$ if the index set $\hat{\mathcal{N}}(\xi^{-k})$ does not change in some neighborhood of $\xi^{-k}$. In order to study the continuity of
$P_k f$ also at points $\xi^{-k}$ where the index sets $\mathcal{N}(\xi^{-k})$ do change in any neighborhood of $\xi^{-k}$, we introduce some additional notation. Let $\mathbb{B}_s(\xi^{-k})$ denote the open ball around $\xi^{-k}$ with radius $\epsilon > 0$ and let

$$\mathcal{P}(\xi^{-k}) := \{ \text{cl } \Xi_v(x) : \xi_s^{-k} \in \text{cl } \Xi_v(x) \text{ for some } s \in \mathbb{R} \}$$

(34)

$$\mathcal{P}_r(\xi^{-k}) := \{ \text{cl } \Xi_v(x) : \xi_s^{-k} \in \text{cl } \Xi_v(x) \text{ for some } s \in \mathbb{R}, \xi^{-k} \in \mathbb{B}_r(\xi^{-k}) \}$$

(35)

denote sets of convex polyhedra $\text{cl } \Xi_v(x)$ that are met by the one-dimensional affine subspace $\{ \xi_s^{-k} : s \in \mathbb{R} \}$. Because any such subspace $\{ \xi_s^{-k} : s \in \mathbb{R} \}$ for some $\xi^{-k} \in \mathbb{B}_r(\xi^{-k})$ is a parallel translation of $\{ \xi_s^{-k} : s \in \mathbb{R} \}$, $\epsilon_0$ can be chosen small enough such that $\mathcal{P}(\xi^{-k}) \subseteq \mathcal{P}(\xi^{-k})$ holds for every $\xi^{-k} \in \mathbb{B}_r(\xi^{-k})$. Therefore we have

$$\mathcal{P}(\xi^{-k}) = \mathcal{P}_0(\xi^{-k}).$$

(36)

Since the polyhedra $\text{cl } \Xi_v(x)$ are convex, the sets $\{ \xi_s^{-k} \in \text{cl } \Xi_v(x) : s \in \mathbb{R} \}$ are convex, too, and, hence, represent either an interval or a single point if $\text{cl } \Xi_v(x)$ belongs to $\mathcal{P}(\xi^{-k})$. The latter is only possible if the one-dimensional affine space meets a vertex or an edge (i.e., faces of dimension zero or one) of $\text{cl } \Xi_v(x)$. The subset of $\mathbb{R}^d$ that contains all vertices and edges of all such polyhedra $\text{cl } \Xi_v(x)$ has Lebesgue measure zero in $\mathbb{R}^d$. If the set $\{ \xi_s^{-k} \in \text{cl } \Xi_v(x) : s \in \mathbb{R} \}$ is an interval $I_v$, the set $\{ \xi_s^{-k} : s \in \text{int } I_v \}$ belongs to the interior of $\text{cl } \Xi_v(x)$. Otherwise, the interval $I_v$ belongs to a facet of $\text{cl } \Xi_v(x)$ which in turn is parallel to the canonical basis element $e_k$ contradicting the geometric condition (A6).

**Proposition 4** Let $(A1)$–$(A6)$ be satisfied, $x \in X$ be fixed, $k \in \mathfrak{D}$ and we consider an integrand $f = f_x$ of the form (28). Then its $k$th projection $P_k f$ is continuous on $\Pi_k(\mathbb{R}^d)$ and first order continuously differentiable on $\Pi_k(\mathbb{R}^d) \setminus M$, where $M$ is a closed set that is contained in the union of finitely many hyperplanes of dimension at most $d - 2$ and, thus, has Lebesgue measure zero in $\mathbb{R}^{d-1}$. Moreover, the estimate

$$\left| \frac{\partial P_k f}{\partial \xi_r}(\xi^{-k}) \right| \leq C \max\{1, \|x\|\} \max\{1, \|\xi^{-k}\|^2\}$$

(37)

holds for almost every $\xi^{-k} \in \Pi_k(\mathbb{R}^d)$, for all $r \in \mathfrak{D}$, $r \neq k$, and some constant $C > 0$.

**Proof** Let $x \in X$, $k \in \mathfrak{D}$ and $\xi^{-k} \in \Pi_k(\mathbb{R}^d)$. First we prove continuity of $P_k f$ at $\xi^{-k}$ and distinguish the following two cases:

(i) There exists $\epsilon_0 > 0$ such that $\mathcal{P}(\xi^{-k}) = \mathcal{P}(\xi^{-k})$ for all $\xi^{-k} \in \mathbb{B}_{\epsilon_0}(\xi^{-k})$.

(ii) For each $\epsilon > 0$ there exists $\xi^{-k} \in \mathbb{B}_\epsilon(\xi^{-k})$ such that $\mathcal{P}(\xi^{-k}) \subseteq \mathcal{P}(\xi^{-k})$.

In case (i) we know that the function $\xi^{-k} \mapsto f(\xi^{-k}) = f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d)$ is continuous in $\mathbb{B}_{\epsilon_0}(\xi^{-k})$ for all $s \in \mathbb{R}$ except at the points $s_{\nu}$, $\nu \in \mathcal{N}(\xi^{-k})$. Due to Proposition 2 the estimate

$$\left| f(\xi^{-k}) \right| \leq C \max\{1, \|x\|\} \max\{1, s^2 + \|\xi^{-k}\|^2\}$$

holds for all $(s, \xi^{-k}) \in \mathbb{R} \times \mathbb{B}_{\epsilon_0}(\xi^{-k})$ and some constant $C > 0$. The latter right-hand side represents an integrable majorant for $f(\xi^{-k})$ and, hence, Lebesgue’s dominated convergence theorem implies that $P_k f$ is continuous at $\xi^{-k}$.

In case (ii) we choose $\epsilon_0 > 0$ small enough such that the identity (36) is valid. We consider the index set $\mathcal{N}(\xi^{-k})$ of all intervals $I_v(\xi^{-k})$ with left end points $s_{\nu}(\xi^{-k})$
and allow explicitly that \( s_\nu(\xi^{-k}) = s_{\nu+1}(\xi^{-k}) \) holds for some \( \nu \in \mathcal{N}(\xi^{-k}) \). Then the representation

\[
(P_k f)(\xi^{-k}) = \sum_{\nu \in \mathcal{N}(\xi^{-k})} \sum_{j=0}^{2} p_{j,\nu}(\xi^{-k}; x) \int_{s_\nu(\xi^{-k})}^{s_{\nu+1}(\xi^{-k})} s^j \rho_k(s) ds
\]

(38)
is valid, where \( s_\nu(\xi^{-k}) \) is given by (33). Now, let \( \xi^{-k} \in B_{\epsilon_0}(\xi^{-k}) \). Then \( P_k f(\xi^{-k}) \)
may be represented by a subset of the set \( \mathcal{N}(\xi^{-k}) \). Of course, all intervals \( L_\nu(\xi^{-k}) \)
with \( s_\nu(\xi^{-k}) < s_{\nu+1}(\xi^{-k}) \) appear also in the representation of \( P_k f(\xi^{-k}) \) if \( \epsilon_0 \) is
small enough. Here, we used that \( \mathcal{N} \) and, hence, \( \mathcal{N}(\xi^{-k}) \) are finite. Those
intervals with \( s_\nu(\xi^{-k}) = s_{\nu+1}(\xi^{-k}) \) may either disappear or appear with \( s_\nu(\xi^{-k}) < s_{\nu+1}(\xi^{-k}) \). If they disappear we set \( s_\nu(\xi^{-k}) = s_{\nu+1}(\xi^{-k}) \) and include them formally into
the representation of \( P_k f(\xi^{-k}) \) which is of the form

\[
(P_k f)(\xi^{-k}) = \sum_{\nu \in \mathcal{N}(\xi^{-k})} \sum_{j=0}^{2} p_{j,\nu}(\xi^{-k}; x) \int_{s_\nu(\xi^{-k})}^{s_{\nu+1}(\xi^{-k})} s^j \rho_k(s) ds .
\]

(39)

Letting \( \xi^{-k} \) in (39) tend to \( \xi^{-k} \) and using the continuity of \( s_\nu(\cdot) \) and of \( p_{j,\nu}(\cdot; x) \)
for \( \nu \) belonging to the finite index set \( \mathcal{N}(\xi^{-k}) \), a comparison with (38) proves the
continuity of \( P_k f \) at \( \xi^{-k} \) in case (ii), too.

Finally, we return to case (i) and study differentiability properties of \( P_k f \) at such
points \( \xi^{-k} \in \Pi_k(\mathbb{R}^d) \). From (32) we obtain for \( r \in \mathcal{D}, r \neq k \), that

\[
\frac{\partial P_k f(\xi^{-k})}{\partial x} = \sum_{\nu \in \mathcal{N}} \left( \sum_{j=0}^{2} \frac{\partial}{\partial x} p_{j,\nu}(\xi^{-k}; x) \int_{s_\nu(\xi^{-k})}^{s_{\nu+1}(\xi^{-k})} s^j \rho_k(s) ds + \sum_{j=0}^{2} p_{j,\nu}(\xi^{-k}; x) \right)
\]

(40)

\[
\left( \frac{g_{\nu,r}}{g_{\nu,k}} s^j(\xi^{-k}) \rho_k(s_\nu(\xi^{-k})) - \frac{g_{\nu+1,r}}{g_{\nu+1,k}} s^j(\xi^{-k}) \rho_k(s_{\nu+1}(\xi^{-k})) \right)
\]

(41)

holds, where the corresponding term in (41) vanishes if \( s_\nu \) and \( s_{\nu+1} \), respectively,
are not finite. Hence, \( P_k f \) is first order continuously differentiable at points \( \xi^{-k} \)
satisfying (i) and, thus, on \( \Pi_k(\mathbb{R}^d) \) except at all boundary points of the polyhedra
\( \Pi_k \mathcal{Z}_\nu(x), \nu \in \mathcal{N} \). All boundaries are contained in a finite union of hyperplanes
of dimension at most \( d - 2 \) which has Lebesgue measure 0 in \( \mathbb{R}^{d-1} \).

To prove the estimate (37) we fix \( n \in \mathcal{N} \). To bound the first summand in (40) we note that
\( \frac{\partial}{\partial x} p_{j,\nu}(\xi^{-k}; x) \) is linear in \( \xi^{-k} \) for \( j = 0 \) and constant for \( j = 1 \).

The integral \( \int_{s_\nu(\xi^{-k})}^{s_{\nu+1}(\xi^{-k})} s^j \rho_k(s) ds \) is bounded by 1 for \( j = 0 \) and by a constant
for \( j = 1 \). To bound the second summand in (40) and (41) we observe that the
first factor \( p_{j,\nu}(\xi^{-k}; x) \) of each summand is bounded by a constant for \( j = 2 \), by
a constant times \( \|\xi^{-k}\| \) for \( j = 1 \), and by a constant times \( \|\xi^{-k}\|^2 \) for \( j = 0 \). The second
factor is bounded by a constant for \( j = 0 \), by a constant times \( \|\xi^{-k}\| \) for \( j = 1 \), and by a constant times \( \|\xi^{-k}\|^2 \) for \( j = 2 \). Furthermore, the coefficients of the
polynomials \( p_{j,\nu} \) are affine functions of \( x \), thus, can bounded by a constant times
\( \max\{|1,\|x\|\} \). Altogether, both summands can be estimated by a constant times
\( \max\{|1,\|x\|\} \max\{|1,\|\xi^{-k}\|\}^2 \} \), where the constant depends on \( \nu \) and \( r \). Finally, we
note that \( \nu \) and \( r \) vary in finite sets and arrive at the desired estimate (37). \( \square \)
Remark 1 When looking at the formula (40), (41) for the first order partial derivative of $P_k f$ in the proof of Proposition 4, which exists on $\Pi_k(\mathbb{R}^d) \setminus M$, it becomes evident that this result can be extended to twice partial differentiability if the conditions (A1)–(A6) are satisfied. Moreover, we state without recording the elementary proof and analogous arguments as in the last part of the proof that the estimate
\begin{equation}
\left| \frac{\partial^2 P_k f}{\partial \xi_r \partial \xi_q}(\xi^{-k}) \right| \leq C \max\{1, \|x\|\} \max\{1, \|\xi^{-k}\|^2\}
\end{equation}
holds for all $\xi^{-k} \in \Pi_k(\mathbb{R}^d) \setminus M$, all $q, r \in \mathcal{D} \setminus \{k\}$ and some constant $C > 0$.

The following example shows that the geometric condition (A6) is indispensable for Proposition 4 to hold true.

Example 1 Let $d = 2$ and $P$ denote a two-dimensional probability distribution with independent continuous marginal densities $\rho_k$, $k = 1, 2$. We consider the two convex polyhedral cones (see the picture below)
\[ K_1 = \{(t_1, t_2) \in \mathbb{R}^2 : 0 \leq t_2 \leq t_1\}, \quad K_2 = \{(t_1, t_2) \in \mathbb{R}^2 : 0 \leq t_1, t_2 \leq t_1, -2t_1 \leq t_2\} \]
and the infimal functions
\[ \Phi_i(t) = \begin{cases} 1, & t \in \text{int} K_i, \\ 0, & \text{otherwise} \end{cases} \quad (i = 1, 2), \]
which are piecewise constant and lower semicontinuous functions. Both are simple (but typical) infimal value functions for pure integer optimization models.

Let the integrands $f_i$ be defined by
\[ f_i(\xi) = \Phi_i(\xi - Tx), \]
where we let for simplicity $x = 0$.

Then its $k$th first order ANOVA projections $P_k f_i$ are
\[ (P_k f_i)(\xi^{-k}) = \int_{-\infty}^{+\infty} \Phi_i(\xi^{-k}) \rho_k(s) ds, \]
where $\xi^{-k} \in \mathbb{R}$, $k \in \{1, 2\}$. We obtain for $i = 1$
\[ P_1 f_1(\xi^{-1}) = P_1 f_1(\xi_2) = \begin{cases} \int_{\xi_2}^{+\infty} \rho_1(s) ds, & 0 \leq \xi_2, \\ 0, & \text{otherwise}, \end{cases} \]
\[ P_2 f_1(\xi^{-2}) = P_2 f_1(\xi_1) = \begin{cases} \int_{0}^{\xi_1} \rho_2(s) ds, & 0 \leq \xi_1, \\ 0, & \text{otherwise}. \end{cases} \]
Hence, in general $P_1 f_1$ isn’t continuous, but $P_2 f_2$ is continuous on $\mathbb{R}$. The reason is that a facet of $K_1$ is parallel to the $t_1$-axis. For $i = 2$ we have

$$P_1 f_2(\xi^{-1}) = P_1 f_2(\xi_2) = \begin{cases} \int_{\xi_2}^{+\infty} \rho_1(s) \, ds, & \xi_2 \geq 0, \\ \int_{\xi_2}^{-\infty} \rho_1(s) \, ds, & \text{otherwise}, \end{cases}$$

$$P_2 f_2(\xi^{-2}) = P_2 f_2(\xi_1) = \begin{cases} \int_{\xi_1}^{-\xi_1} \rho_2(s) \, ds, & 0 \leq \xi_1, \\ 0, & \text{otherwise}, \end{cases}$$

and, thus, $P_1 f_2$ and $P_2 f_2$ are continuous and piecewise continuously differentiable. For a discussion of the geometric condition (A6) we refer the reader to Section 6.

Using Proposition 4 we show now that second order projections $P_u f$, of $f$ with $u \subseteq \mathcal{D}$, $|u| = 2$, are even continuously differentiable on the entire space $P_u \mathbb{R}^d$.

For $k, l \in \mathcal{D}$, $k \neq l$, we consider $P_k f$ and its projection $P_l P_k f$, i.e., the second order projection $P_u f$ of $f$ with $u = \{k, l\}$. The function $P_k f$ is given on the space $P_k \mathbb{R}^d$ which is subdivided into the sets $P_k(\Xi_{\nu})$, i.e., the $k$th projections of the original sets $\Xi_{\nu}$, $\nu \in \mathcal{N}$, in $\mathbb{R}^d$. The closures $P_k(\text{cl } \Xi_{\nu})$ of the sets $P_k(\Xi_{\nu})$ are convex polyhedral and have dimension $d - 1$ [2, Proposition 2.1]. We obtain

$$P_u f(\xi^{-u}) = P(P_k f)(\xi^{-u}) = \int_{-\infty}^{+\infty} P_k f(\xi^{-u}) \rho_1(s) \, ds,$$

where $\xi^{-u} = P_u \xi$ and $\xi_s^{-u} = P_k \xi_s^{-l}$, $s \in \mathbb{R}$, and know that

$$\xi_s^{-u} \in \bigcup_{\nu \in \mathcal{N}} P_k(\Xi_{\nu}) = P_k \mathbb{R}^d$$

holds for each $s \in \mathbb{R}$. Hence, similar as before Proposition 4 for $\xi^{-k}$, for given $\xi^{-u}$ there exist a finite index set $\mathcal{N}_1 = N_1(\xi^{-u})$ and intervals $I_{1, \nu}$ with $s_{1, \nu} = \inf I_{1, \nu}$ and $s_{1, \nu+1} = \sup I_{1, \nu}$ for $\nu \in \mathcal{N}_1$ such that

$$P_u f(\xi^{-u}) = \sum_{\nu \in \mathcal{N}_1(\xi^{-u})} \int_{I_{1, \nu}(\xi^{-u})} P_k f(\xi^{-u}) \rho_1(s) \, ds$$

$$= \sum_{\nu \in \mathcal{N}_1(\xi^{-u})} \int_{s_{1, \nu}(\xi^{-u})}^{s_{1, \nu+1}(\xi^{-u})} P_k f(\xi^{-u}) \rho_1(s) \, ds,$$

(43)

where the first and the last interval are unbounded and the finite points $s_{1, \nu}$ belong to common facets $G_{\nu}$ of two adjacent convex polyhedra of the form $P_k(\text{cl } \Xi_{\nu})$. All such facets are $k$th projections of certain faces $F_{\nu}$ of the polyhedra $\text{cl } \Xi_{\nu}$, i.e., $P_k(F_{\nu}) = G_{\nu}$ (see [17, Theorem 16] or [54, Lemma 7.10]). If the faces $F_{\nu}$ are defined by the inequalities $\langle g_{1, \nu}, \xi \rangle \leq a_{1, \nu}$, the points $s_{1, \nu}$ may be represented in the form

$$s_{1, \nu} = s_{1, \nu}(\xi^{-u}) = \frac{1}{g_{1, \nu,l}} \left( a_{1, \nu} - \sum_{i=1}^{d} g_{1, \nu,i} \xi_i \right)$$

as in (33). Note that $g_{1, \nu,l} \neq 0$ holds due to condition (A6).

To state our next result, we need the following notion. A real function $g$ on $\mathbb{R}^d$ is called locally Lipschitz continuous on lines if for each $k \in \mathcal{D}$ the function $t \mapsto g(\xi^{-k})$ is Lipschitz continuous in $t$ on compact subsets of $\mathbb{R}$ for almost every $\xi^{-k} \in P_k \mathbb{R}^d$. 


Proposition 5 Let (A1)–(A6) be satisfied, \( x \in X \) be fixed and consider the integrand \( f = f_x \) in (28). For any \( k, l \in \mathcal{D} \), \( k \neq l \), \( u = \{k, l\} \), the (ANOVA) projection \( P_u f \) is continuously differentiable on \( \Pi_u(\mathbb{R}^d) \). In addition, the partial derivatives \( \frac{\partial P_u f}{\partial \xi^r}(\xi^{-u}) \) are locally Lipschitz continuous on lines and there exists \( C > 0 \) such that
\[
\left| \frac{\partial P_u f}{\partial \xi^r}(\xi^{-u}) \right| \leq C \max\{1, \|x\|\} \max\{1, \|\xi^{-u}\|^2\} \tag{44}
\]
holds for every \( \xi^{-u} \in \Pi_u(\mathbb{R}^d) \) and \( r \in -u \).

**Proof** Let \( M \) be the closed set in Proposition 4. We consider \( k, l \in \mathcal{D} \) with \( k \neq l \) and set \( u = \{k, l\} \). From Proposition 4 we know that \( P_k f \) is continuously differentiable at any \( \xi^{-k} \in \Pi_k(\mathbb{R}^d) \setminus M \). Hence, for any such \( \xi^{-k} \) and \( \xi^{-u} = \Pi_k \xi^{-k} \) we know that \( P_k f \) is continuously differentiable at \( \xi^{-u} \in \Pi_k(\mathbb{R}^d) \) if \( \xi^{-u} \not\in M \). Since \( \xi^{-u} \in M \) happens only at the finitely many points \( s = s_{1,\nu}(\xi^{-u}) \) due to (A6) and the bound (37) is valid, we can use Lebesgue’s theorem on dominated convergence. We conclude that \( P_u f \) is continuously differentiable at \( \xi^{-u} \) and the identity
\[
\frac{\partial P_u f}{\partial \xi^r}(\xi^{-u}) = \int_{-\infty}^{\infty} \frac{\partial P_k f}{\partial \xi^r}(\xi^{-u}) \rho(s) ds \tag{45}
\]
holds for any \( r \in -u \). To prove that \( P_u f \) is continuously differentiable at any \( \xi^{-u} \in \Pi_u(\mathbb{R}^d) \) we proceed as in the proof of Proposition 4 and consider the set \( P_k(\xi^{-u}) \) of all convex polyhedra \( \Pi_k(\Xi_{\nu}) \) such that \( \xi^{-u} \in \Pi_k(\Xi_{\nu}) \) for some \( s \in \mathbb{R} \). The first case in the proof of Proposition 4 corresponds to the result (45). In the second case we know that for each \( \epsilon > 0 \) there exists \( \xi^{-u} \in \mathcal{B}_\epsilon(\xi^{-u}) \) such that
\[
P_k(\xi^{-u}) \subseteq P_k(\xi^{-u}) \tag{46}
\]
and the representation
\[
P_u f(\xi^{-u}) = \sum_{\nu \in N_1(\xi^{-u})} \int_{s_{1,\nu}(\xi^{-u})}^{s_{1,\nu+1}(\xi^{-u})} P_k f(\xi^{-u}) \rho(s) ds, \tag{47}
\]
holds according to (43). We choose \( \epsilon > 0 \) small enough such that the relation \( s_{1,\nu}(\xi^{-u}) < s_{1,\nu+1}(\xi^{-u}) \) in the representation (47) leads to \( s_{1,\nu}(\xi^{-u}) < s_{1,\nu+1}(\xi^{-u}) \), too. Those \( \nu \in N_1(\xi^{-u}) \) with \( s_{1,\nu}(\xi^{-u}) = s_{1,\nu+1}(\xi^{-u}) \) may either disappear or appear with \( s_{1,\nu}(\xi^{-u}) < s_{1,\nu+1}(\xi^{-u}) \). If they disappear we set \( s_{1,\nu}(\xi^{-u}) = s_{1,\nu+1}(\xi^{-u}) \) and include them formally into the representation of \( P_u f(\xi^{-u}) \) which is of the form
\[
P_u f(\xi^{-u}) = \sum_{\nu \in N_1(\xi^{-u})} \int_{s_{1,\nu}(\xi^{-u})}^{s_{1,\nu+1}(\xi^{-u})} P_k f(\xi^{-u}) \rho(s) ds.
\]
In a small ball around $\xi^{-u}$ this representation doesn’t change. Hence, $P_u f$ is differentiable at $\xi^{u}$ and the partial derivative for any $r \in -u$ is of the form

$$
\frac{\partial P_u f}{\partial \xi_r}(\xi^{-u}) = \sum_{\nu \in N_1(\xi^{-u})} \left( \int_{s_1,\nu}(\xi^{-u}) \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \rho(s) ds \right)
+ \frac{\partial s_{\nu+1}}{\partial \xi_r} P_k f(s_{\nu+1}) \rho(s_{\nu+1}) - \frac{\partial s_\nu}{\partial \xi_r} P_k f(s_{\nu}) \rho(s_{\nu}) \right) (48)
$$

$$
= \sum_{\nu \in N_1(\xi^{-u})} \int_{s_1,\nu}(\xi^{-u}) \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \rho(s) ds
$$

$$
= \int_{-\infty}^{\infty} \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \rho(s) ds, (49)
$$

where the summands in (48) cancel successively, and the first and the last term in (48) vanish by definition. Letting $\xi^{-u}$ converge to $\xi^{-u}$ in the right-hand side of (49) proves the continuous differentiability of $P_u f$ at $\xi^{u}$, where the partial derivative with respect to $\xi$ is given by (45).

To show that the partial derivatives $\frac{\partial P_u f}{\partial \xi_r}$ are locally Lipschitz continuous on lines, let $p \in -u$. First we consider the partial derivative of the first order projection $\frac{\partial P_u f}{\partial \xi_r}(\xi^{-k})$ given by the equations (40) and (41). We fix all components of $\xi^{-k}$ except the $p$th component $\xi_p$. Then the representation (40) and (41) of $\frac{\partial P_u f}{\partial \xi_r}(\xi^{-k})$ is valid for all $\xi_p \in \mathbb{R}$ except at finitely many points $\xi_{p,\nu}, \nu = 1, \ldots, N_p = N_p(\xi^{-k,p})$.

We assume that the points are ordered with respect to the natural order and observe that in each of the open intervals $I_{p,0} = (-\infty, \xi_{p,0})$, $I_{p,\nu} = (\xi_{p,\nu}, \xi_{p,\nu+1})$ and $I_{p,N_p} = (\xi_{p,N_p}, +\infty)$ the partial derivative $\frac{\partial P_u f}{\partial \xi_r}(\xi^{-k})$ is equal to a sum of products of functions that are locally Lipschitz continuous with respect to $\xi_p$. Hence, $\frac{\partial P_u f}{\partial \xi_r}(\xi^{-k})$ is Lipschitz continuous on each bounded part of $I_{p,0}$ and $I_{p,N_p}$, and on each interval $I_{p,\nu}, \nu = 1, \ldots, N_p - 1$, respectively. Now, let $I_{p,0}$ denote a bounded interval and let $\xi_p \in I_{p,0}, \xi_p \in I_{p,0}, \xi_p < \xi_p$. We choose $\varepsilon > 0$ and $\kappa \leq \mu$ such that $\xi_{p,\mu-1} < \xi_p + \varepsilon < \xi_{p,\mu} - \varepsilon < \xi_{p,\mu} + \varepsilon < \xi_p - \varepsilon < \xi_p \leq \xi_{p,\mu+1}$. Then observe that $\xi_{\nu}^{-u}$ and $\xi_{\nu}^{-u}$ are the elements in $\mathbb{R}^{d-2}$ in which the $p$th components are $\xi_p + \varepsilon$ and $\xi_p - \varepsilon$, respectively, and all other components be fixed. Similarly, we introduce the notations $\xi_{\nu}^{-u}$ and $\xi_{\nu}^{-u}$. Then we obtain

$$
\left| \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u}) - \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u}) \right| \leq \int_{-\infty}^{\infty} \left| \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s - \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \right| ds
$$

$$
+ \sum_{\nu = k}^{\mu - 1} \left| \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s - \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \right| ds
$$

$$
+ \left| \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s - \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \right| ds
$$

$$
+ \sum_{\nu = k}^{\mu} \left| \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s - \frac{\partial P_u f}{\partial \xi_r}(\xi^{-u})_s \right| ds
$$

where the summands in (48) cancel successively, and the first and the last term in (48) vanish by definition. Letting $\xi^{-u}$ converge to $\xi^{-u}$ in the right-hand side of (49) proves the continuous differentiability of $P_u f$ at $\xi^{u}$, where the partial derivative with respect to $\xi$ is given by (45).
Next we let \( \varepsilon \) tend to zero and make use of the continuity of \( \frac{\partial P_uf}{\partial \xi_r}(\xi^{-u}) \). This leads to

\[
\left| \frac{\partial P_uf}{\partial \xi_r}(\xi^{-u}) - \frac{\partial P_uf}{\partial \xi_r}(\xi^0) \right| \leq L \left( (\xi_{p,k} - \xi_p) + \sum_{\nu=k}^{\mu-1} (\xi_{p,k+1} - \xi_{p,\nu}) + (\xi_{p,k} - \xi_{p,\mu}) \right)
\]

and, hence, to the desired Lipschitz continuity property on lines.

For \( r \in -u \) and \( \xi^{-u} \in \Pi_u(\mathbb{R}^d) \) we conclude finally from (49) and (37) that

\[
\left| \frac{\partial P_uf}{\partial \xi_r}(\xi^{-u}) \right| \leq C \max\{1, \|x\|\} \int_{-\infty}^{\infty} \max\{1, s^2 + \|\xi^{-u}\|^2\} \rho_l(s)\,ds
\]

\[
\leq C \max\{1, \|x\|\} \left( \int_{-\infty}^{\infty} \max\{1, s^2\} \rho_l(s)\,ds + \max\{1, \|\xi^{-u}\|^2\} \right)
\]

\[
\leq C \max\{1, \|x\|\} \max\{1, \|\xi^{-u}\|^2\}
\]

and, thus, (44) holds for some sufficiently large constant \( C > 0 \).

The following is our main result in this section. It states that the first and second order ANOVA terms of \( f \) have mixed first weak derivatives.

**Theorem 1** Let (A1)–(A6) be satisfied, \( x \in X \) be fixed and we consider an integrand \( f = f_x \) of the form (28). Then all first and second order ANOVA terms \( f_u, 0 \neq |u| \leq 2, u \subseteq \mathcal{D} \), are first order continuously differentiable and have second order mixed first weak derivatives that belong to \( L_{2,\rho}(\mathbb{R}^d) \). Hence, they belong to the mixed Sobolev space \( W_{2,\rho,\text{mil}}(\mathbb{R}^d) \).

**Proof** Due to Proposition 5 all second order projections \( P_uf \) of \( f \) with \( |u| = 2 \) are continuously differentiable and their partial derivatives are locally Lipschitz continuous on lines on \( \Pi_u(\mathbb{R}^d) \). These properties carry over to higher order projections \( P_vf \) with \( 2 < |v| < d \). While the continuous differentiability follows from the dominated convergence theorem using the bound (44), the local Lipschitz continuity on lines of partial derivatives is a consequence of Proposition 5 and of the following estimate for subsets \( u, v \) of \( \mathcal{D} \) with \( u \subset v \) and \( 2 = |u| < |v| \):

\[
\left| \frac{\partial P_uf}{\partial \xi_r}(\xi^{-u}) - \frac{\partial P_uf}{\partial \xi_r}(\xi^0) \right| = \left| P_{v-u} \frac{\partial P_uf}{\partial \xi_r}(\xi^{-v}) - P_{v-u} \frac{\partial P_uf}{\partial \xi_r}(\xi^v) \right|
\]

\[
\leq \int_{\mathbb{R}^{|v-u|}} \left| \frac{\partial P_uf}{\partial \xi_r}(\xi^{-v}, s^{-u}) - \frac{\partial P_uf}{\partial \xi_r}(\xi^{-v}, s^{-u}) \right| \prod_{j \in v-u} \rho_j(s_j)\,ds_j.
\]
According to (21) the ANOVA terms of $f$ are given by

$$f_u = P_{-u}(f) + \sum_{0 \leq u \leq d} (-1)^{|u|} |v|^{-d} P_{-v}(f)$$

for all nonempty subsets $u$ of $\mathcal{D}$. Hence, all ANOVA terms $f_u$ of $f$ for $|u| < d - 1$ are continuously differentiable on $\Pi_u(\mathbb{R}^d)$. The non-vanishing first order partial derivatives of the first and second order ANOVA terms are of the form

$$D_1 f_{\{l\}}(\xi_l) = \frac{\partial f_{\{l\}}}{\partial \xi_l}(\xi_l) = \frac{\partial P_{-\{l\}} f}{\partial \xi_l}(\xi_l)$$

$$D_1 f_{\{l,k\}}(\xi_l, \xi_k) = \frac{\partial f_{\{l,k\}}}{\partial \xi_l}(\xi_l, \xi_k) - \frac{\partial P_{-\{l\}} f}{\partial \xi_l}(\xi_l, \xi_k) - \frac{\partial P_{-\{l\}} f}{\partial \xi_l}(\xi_l),$$

for any $l, k \in \mathcal{D}$. Hence, the functions $D_1 f_{\{l,k\}}$ and $D_2 f_{\{l,k\}}$ are locally Lipschitz continuous with respect to each of the two variables $\xi_l$ and $\xi_k$, independently when the other variable is fixed almost everywhere. Hence, $D_1 f_{\{l,k\}}$ and $D_2 f_{\{l,k\}}$ are partially differentiable with respect to $\xi_l$ and $\xi_k$, respectively, in the sense of Sobolev (see, for example, [8, Section 4.2.3]). Furthermore, the mixed first weak derivatives coincide with the mixed first classical derivatives at some point if the latter exist at this point. We know from Remark 1 that second order classical mixed first derivatives of $P_k f$ and, thus, of all projections $P_k f$ with $|v| \leq d - 1$ exist almost everywhere due to the dominated convergence theorem. Hence, the classical mixed first derivatives $D_{lk} f_{\{l,k\}} = \frac{\partial^2 f_{\{l,k\}}}{\partial \xi_l \partial \xi_k}$ exist almost everywhere and coincide there with the weak mixed first derivatives. The bound (42) then implies that the estimate

$$|D_{lk} f_{\{l,k\}}(\xi_l, \xi_k)| \leq C \max\{1, \|x\|\} \max\{1, \xi_l^2 + \xi_k^2\}$$

is valid for almost every pair $(\xi_l, \xi_k) \in \Pi_{\{l,k\}} \mathbb{R}^d$, any $l, k \in \mathcal{D}$, any $x \in X$ and some constant $C > 0$. We conclude that $D_{lk} f_{\{l,k\}}$ belongs to $L_{2,\rho}(\mathbb{R}^d)$ for all $l, k \in \mathcal{D}$ due to (A4). □

Remark 2 Let $f^{(k)}$ denote the $k$th order ANOVA approximation

$$f^{(k)} = \sum_{n < |u| \leq k} f_u$$

(50)

of the mixed-integer two-stage integrand $f$ (see (28)) for some $1 \leq k < d$. Theorem 1 furnishes conditions implying that $f^{(2)}$ has all weak mixed first order partial derivatives and our next Remark discusses its extension to $f^{(k)}$ for $k > 2$. According to (20) and to the orthogonality of the ANOVA terms $f_u$ in $L_{2,\rho}$ one has

$$\|f - f^{(k)}\|_{L_{2,\rho}}^2 = \sum_{|u| > k} \|f_u\|_{L_{2,\rho}}^2.$$

If the effective superposition dimension $d_S(\varepsilon)$ of $f$ (see (23)) is at most $k$, the mean square error of the integrands $f$ and $f^{(k)}$ satisfies

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

due to (25). For a discussion of techniques for determining and reducing the effective superposition dimension in case of (log)normal probability distributions we refer to [27, 37, 51–53].
Remark 3 If in addition to (A1)–(A6) the densities \( \rho_i, i \in \mathcal{D} \), have also second and higher order derivatives and all derivatives are bounded on \( \mathbb{R} \), the result in Remark 1 extends to third and higher order mixed first derivatives of \( P_k f \) on \( \Pi_k(\mathbb{R}^d) \setminus M \).

With the techniques used in Proposition 5 this allows to prove that \( P_k f \) has second order mixed first derivatives if \( |u| = 3 \) and, more general, that \( P_k f \) has \( k \)th order mixed first derivatives which are locally Lipschitz continuous on lines if \( |u| = k + 1 \). The corresponding bounds for the mixed derivatives can be proved, too. Finally, Theorem 1 extends to the existence of \( k \)th order weak mixed first derivatives for \( P_k f \) if \( |u| = k, 1 \leq k \leq \frac{d}{2} \). The representation (21) of ANOVA terms then implies that \( f_u \) with \( |u| = k \) belongs to \( W^{(1,\ldots,1)}_{2,p}(\mathbb{R}^d) \) for \( 1 \leq k \leq \frac{d}{2} - 1 \).

An important consequence is that Theorem 2 in Section 5 remains valid for mixed-
integer two-stage integrands with effective superposition dimension \( 2 < d_p(\varepsilon) = k \leq \frac{d}{2} - 1 \) for some \( \varepsilon > 0 \) by arguing with \( k \)th order ANOVA approximations.

Remark 4 For the special case of linear two-stage integrands \( f \) it is shown in [25] that \( P_k f \) is continuously differentiable on \( \mathbb{R}^d \) and has second order weak first derivatives. Under the assumptions imposed in [25] the projection \( P_k f \) with \( |u| = k \) has weak mixed first derivatives of order \( k + 1 \) for \( 1 \leq k \leq \frac{d}{2} \) and the ANOVA term \( f_u \) with \( |u| = k \) belongs to \( W^{(1,\ldots,1)}_{2,p}(\mathbb{R}^d) \) for \( 1 \leq k \leq \frac{d}{2} \).

5 Error analysis for randomly shifted lattice rules

In this section we provide an error analysis for randomly shifted lattice rules. Convergence results for this method are known for integrands from weighted tensor product Sobolev spaces on \([0,1]^d\) (see [6, 20]). Since typical integrands in stochastic programming are defined on \( \mathbb{R}^d \), we introduce first appropriate Sobolev spaces.

Following [23, 32] we start with the weighted Sobolev spaces \( W^{1}_{2,\gamma_i,\rho_i,\psi_i}(\mathbb{R}) \) of functions \( h \in L_2, \rho_i(\mathbb{R}) \) that are absolutely continuous with derivatives \( h' \in L_2, \psi_i(\mathbb{R}) \) with positive continuous weight functions \( \psi_i, i \in \mathcal{D} \). They are endowed with the weighted inner product

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

where for each \( i \in \mathcal{D} \) the weight \( \gamma_i \) is positive and we assume that for any \( x, \tilde{x} \in \mathbb{R} \)

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

The latter condition implies that the weighted Sobolev space is complete [19] and, thus, a Hilbert space. Furthermore, it is known that there exists a reproducing kernel, i.e., a function \( K_i(x, \tilde{x}) = 1 + \gamma_i \eta_i(x, \tilde{x}) \) for \( x, \tilde{x} \in \mathbb{R} \), where

\[
\eta_i(x, \tilde{x}) = \int_{-\infty}^{\min(x, \tilde{x})} \frac{\phi_i(t)}{\psi_i^2(t)}dt + \int_{\max(x, \tilde{x})}^{\infty} \frac{1 - \phi_i(t)}{\psi_i^2(t)}dt - \int_{-\infty}^{\infty} \frac{\phi_i(t)(1 - \phi_i(t))}{\psi_i^2(t)}dt,
\]

and \( \phi_i \) is the distribution function of the density \( \rho_i \) (see [32, Lemma 1]). This means that \( K_i : \mathbb{R} \times \rightarrow \mathbb{R} \) satisfies \( K_i(\cdot, x) \in W^1_{2,\gamma_i,\rho_i,\psi_i}(\mathbb{R}) \) and \( (h, K_i(\cdot, x))_{\gamma_i, \psi_i} = h(x) \) for all \( x \in \mathbb{R} \) and \( h \in W^1_{2,\gamma_i,\rho_i,\psi_i}(\mathbb{R}) \). For more information on kernel reproducing
Hilbert spaces we refer to the seminal paper [1] and to the monograph [4]. It is known from [1] that the weighted tensor product Sobolev space
\[ F_d = \mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\rho,\psi,\text{mix}}(\mathbb{R}^d) = \bigotimes_{i=1}^{d} W^{1}_{2,\gamma_i,\rho_i,\psi_i}(\mathbb{R}) \]
is also a reproducing kernel Hilbert space with the reproducing kernel
\[ K_{d,\gamma,\rho,\psi}(\xi,\tilde{\xi}) = \prod_{i=1}^{d} \left( 1 + \gamma_i \rho_i(\xi_i,\tilde{\xi}_i) \right) = \sum_{\emptyset \neq u \subseteq D}^{} \gamma_u \prod_{i \in u} \rho_i(\xi_i,\tilde{\xi}_i). \]
The inner product of \( F_d \) is given by
\[ \langle f, \tilde{f} \rangle_{\gamma,\psi} = \sum_{u \subseteq D}^{\gamma_u^{-1}} \int_{\mathbb{R}^{|u|}} I_{u,\rho}(f)(\xi^u) I_{u,\rho}(\tilde{f})(\xi^u) \prod_{i \in u} \psi_i^2(\xi_i)d\xi^u, \]
where the integrands \( I_{u,\rho}(f)(\xi^u) \) and the weights \( \gamma_u \) are defined by
\[ I_{u,\rho}(f)(\xi^u) = \int_{\mathbb{R}^{|u|-1}} \frac{\partial^{|u|}}{\partial \xi^u} f(\xi) \prod_{i \in -u} \rho_i(\xi_i)d\xi^{-u} \quad \text{and} \quad \gamma_u = \prod_{i \in u} \gamma_i, \quad \gamma_\emptyset = 1. \]
In the QMC literature, this is called the unanchored setting with product weights. In order to apply QMC methods to the computation of integrals
\[ I_\rho(f) = \int_{\mathbb{R}^d} f(\xi) \rho(\xi)d\xi = \int_{\mathbb{R}^d} f(\xi) \prod_{i=1}^{d} \rho_i(\xi_i)d\xi \]
with \( f \in F_d \) the Hilbert space \( F_d \) has to be transformed to a Hilbert space \( G_d \) of functions \( g \) on \([0,1]^d\) by the isometry
\[ f \in F_d \iff g(\cdot) = f(\Phi^{-1}(\cdot)) \in G_d, \]
where \( \Phi^{-1}(t) = (\phi_1^{-1}(t_1), \ldots, \phi_d^{-1}(t_d)), \quad t \in [0,1]^d \). The reproducing kernel and inner product of \( G_d \) are
\[ K_{d,\gamma}(t,\tilde{t}) = K_{d,\gamma,\rho,\psi}(\Phi^{-1}(t),\Phi^{-1}(\tilde{t})), \quad (t,\tilde{t} \in [0,1]^d) \]
\[ \langle g, \tilde{g} \rangle_\gamma = \langle f(\Phi^{-1}(\cdot)), \tilde{f}(\Phi^{-1}(\cdot)) \rangle_\gamma = \langle f, \tilde{f} \rangle_{\gamma,\psi}. \]
The choice of the weight functions \( \psi_i \) depends on the marginal densities \( \rho_i, \ i \in D \).
We refer to [22,32] for a discussion of this aspect and for a list of marginal densities and the corresponding weight functions.
Now we consider randomly shifted lattice rules for numerical integration in \( G_d \) (see [45,20]). Let \( Z_n = \{ z \in \mathbb{N} : 1 \leq z \leq n, \gcd(z,n)=1 \} \) denote the set of natural numbers between 1 and \( n \) that are relatively prime to \( n \). Given a generating vector \( g \in \mathbb{Z}_n^d \) and a random shift vector \( \triangle \) which is uniformly distributed in \([0,1]^d\), the shifted lattice rule points are \( t^j = \left\{ \frac{j g}{n} + \triangle \right\}, \quad j = 1, \ldots, n, \)
where the braces indicate taking componentwise the fractional part. The corresponding randomized QMC method on \( G_d \) is of the form
\[ Q_{n,d}(g) = \frac{1}{n} \sum_{j=1}^{n} g(t^j) \quad (g \in G_d, n \in \mathbb{N}) \quad (51) \]
and its shift-averaged worst-case error can be computed using the reproducing kernel. Let \( \varphi(n) \) denote the cardinality of \( \mathcal{Z}_n \), thus, \( \varphi(n) = n \) if \( n \) is prime, and \( \xi^j = \Phi^{-1}(v^j) \) for \( j = 1, \ldots, n \). Then we obtain from [32, Theorem 8] that a generating vector \( \mathbf{g} \in \mathbb{Z}_n^d \) can be constructed by a component-by-component algorithm such that for each \( \delta \in (0, \frac{1}{2}] \) there exists \( C(\delta) > 0 \) with
\[
(\mathbb{E}|\nu_\delta(f) - Q_{n,d}(f(\Phi^{-1}(\cdot)))|^2)^{\frac{1}{2}} \leq C(\delta)\|f\|_{V,\psi} \varphi(n)^{-1+\delta}
\]
if the following condition
\[
\sum_{i=1}^{\infty} \delta_{i,\alpha(i)} < \infty
\]
on the weights is satisfied and \( f \) belongs to \( \mathbb{F}_d \). To state our next result we denote by \( \nu(P) \) the infimal value of (1) and by \( \nu(Q_{n,d}) \) the infimum if the integral in (1) is replaced by the randomly shifted lattice rule (51).

**Theorem 2** Let (A1)-(A6) be satisfied, the densities \( \rho_i, i \in \mathcal{D} \), be \( s \) times differentiable and all derivatives be bounded on \( \mathbb{R} \) and \( X \) be compact. Assume that all integrands \( f = f_x, x \in X \), of the form (28) have effective superposition dimension \( d_2(\varepsilon) = k \leq \frac{d}{2} - 1 \) for some \( \varepsilon > 0 \) and that the \( k \)th order ANOVA approximation \( f^{(k)} \) of \( f \) (see (50)) belongs to \( \mathbb{F}_{d} \). Furthermore, we assume that \( Q_{n,d} \) is a randomly shifted lattice rule (51) satisfying (52). For each \( \delta \in (0, \frac{1}{2}] \) there exists \( \hat{C}(\delta) > 0 \) such that
\[
(\mathbb{E}|\nu(P) - \nu(Q_{n,d})|^2)^{\frac{1}{2}} \leq \hat{C}(\delta)\varphi(n)^{-1+\delta} + a_n,
\]
where the sequence \( (a_n) \) converges to zero and allows the estimate
\[
a_n \leq \sqrt{\varepsilon}\sigma(f)
\]
with \( \sigma(f) \) denoting the standard deviation of \( f \) (22).

The result can be proved by following the lines of the proof of [26, Theorem 3] with obvious modifications. We note that the differentiability properties of the \( k \)th order ANOVA approximation \( f^{(k)} \) of \( f \) discussed in Remark 3 motivate the condition for \( f^{(k)} \) imposed in Theorem 2.

### 6 Generic smoothness in the normal case

Let \( \xi \) be a \( d \)-dimensional normal random vector with mean \( \mu \) and nonsingular covariance matrix \( \Sigma \). Then there exists an orthogonal matrix \( Q \) such that \( Q \Sigma Q^\top \) is a diagonal matrix. Then the \( d \)-dimensional random vector \( \eta \) given by the transformation
\[
\xi = Q\eta + \mu \quad \text{or} \quad \eta = Q^\top(\xi - \mu)
\]
is normal with zero mean and diagonal covariance matrix, i.e., \( \eta \) has independent components. For fixed \( x \in X \), let \( \Xi_\nu(x), \nu \in \mathcal{N} \), denote the decomposition (14) of \( \mathbb{R}^d \) into Borel sets whose closures are convex polyhedral. The transformed function \( \tilde{f}(x, \eta) = f(x, Q\eta + \mu) \) is linear-quadratic on the sets \( Q^\top \Xi_\nu(x) - Q^\top \mu, \nu \in \mathcal{N} \), whose closures are again convex polyhedral.

The intersections of two adjacent convex polyhedral sets \( \text{cl} \Xi_\nu(x) \) are facets, which are contained in \( (d - 1) \)-dimensional affine subspaces \( H_\nu(x), \nu \in \mathcal{N} \). The space
$H_{v}(x)$ can be described by an equation $v_{H_{v}(x)}^T \xi = b_{H_{v}(x)}$ with a $d$-dimensional nonzero vector $v_{H_{v}(x)}$ and a constant $b_{H_{v}(x)} \in \mathbb{R}$. Since the number of facets of each polyhedral set $\text{cl} \Xi_{\nu}(x)$ is finite, there are finitely many equations

$$v_{H_{1,\nu}(x)}^T \xi = b_{H_{1,\nu}(x)}, \quad i \in I_{\nu}, \ \nu \in \mathcal{N},$$

that describe all $(d - 1)$-dimensional affine subspaces each containing at least one facet of some polyhedron $\text{cl} \Xi_{\nu}(x)$. A $d - k$ dimensional face of a given polyhedron $\text{cl} \Xi_{\nu}(x)$ is then a subset of an affine subspace described by a system of $k$ linear independent equations (intersection of $k$ hyperplanes)

$$v_{H_{1,\nu}(x)}^T \xi = b_{H_{1,\nu}(x)}, \ldots, v_{H_{k,\nu}(x)}^T \xi = b_{H_{k,\nu}(x)}$$

or shortly $V_{k} \xi = b$, where $V_{k}$ is a $(k,d)$-matrix and $b \in \mathbb{R}^k$. Under the linear transformation (56), the corresponding face of the transformed polyhedron $\text{cl}(Q^T \Xi_{\nu}(x) - Q^T \mu)$ is a subset of an affine space described by the system

$$V_{k} Q \eta = b', \quad b' := b - Q \mu.$$

In order to make sure that the face of the transformed polyhedron does not parallel any coordinate axis, it is sufficient to show that the system $V_{k} Q \eta = b'$ must be solvable for each subset of $k$ variables $\eta_{i_1}, \ldots, \eta_{i_k}$ in terms of the remaining components of $\eta$. The latter condition is satisfied if each square submatrix of the $(k,d)$-matrix $A = V_{k} Q$ is nonsingular or, equivalently, each minor of order $r$ for $1 \leq r \leq k$ of the matrix $A$ is nonzero. Now, let $1 \leq r \leq k$ and $A_r$ be any $(r,r)$-submatrix of $A$. Then $A_r$ is given as product of $r$ rows of the matrix $V_{k} = (v_{il})$ and $r$ columns of the matrix $Q = (q_{ij})$, i.e.,

$$A_r = (a_{i_lj_l}) = \left( \sum_{s,t=1}^{d} v_{i_l}s q_{s,j_l} \right) (s, t = 1, \ldots, r).$$

According to the Cauchy-Binet formula the minor $|A_r| = \det(A_r)$ is of the form

$$|A_r| = \sum_{1 \leq l_1 < l_2 < \cdots < l_r \leq d} \det \begin{pmatrix} v_{i_{1}l_1} & \cdots & v_{i_{1}l_r} \\ \vdots & \ddots & \vdots \\ v_{i_{r}l_1} & \cdots & v_{i_{r}l_r} \end{pmatrix} \det \begin{pmatrix} q_{i_{1}j_1} & \cdots & q_{i_{1}j_r} \\ \vdots & \ddots & \vdots \\ q_{i_{r}j_1} & \cdots & q_{i_{r}j_r} \end{pmatrix}.$$

In particular, the minor $|A_r| = \det(A_r)$ can be interpreted as a multivariate polynomial function where the variables are the entries of the $r$ columns of $Q$, and the coefficients are given in terms of the entries of the $r$ selected rows of $V_{k}$. Hence, zeros of the multivariate polynomial correspond to an orthogonal matrix $Q$ for condition (A6) after the transformation is violated.

Next we argue that the multivariate polynomial $|A_r|$ is non-constant. Assuming the contrary means that all $r$-minors that can be obtained from the selected $r$ rows of the matrix $V_{k}$ must be zero. This implies that those $r$ rows are not linearly independent which contradicts the construction of $V_{k}$. We also note that for any $d-k$ dimensional face with $1 \leq k < d$ which defines a system $V_{k} \xi = b$, a multivariate polynomial $|A_r|$ as considered above cannot contain all entries of a column of $Q$ in its variables. It follows that the equations on the entries of $Q$ defining the matrix $Q$ as orthogonal can not imply that $|A_r|$ is a constant polynomial (as it is for the polynomial $|Q|$ over $O(d, \mathbb{R})$). We refer for the following part to [5] for an
introductory presentation of the Haar measure on topological groups. It is known that $O(d, \mathbb{R})$ is a compact topological group and a smooth manifold of dimension

$$\binom{d}{2} = \frac{d(d-1)}{2} = d^2 - \frac{d(d+1)}{2},$$

where the first term on the right-hand side corresponds to the number of elements of a matrix $Q \in O(d, \mathbb{R})$ and the second term is the number of equations $\langle Q_i, Q_j \rangle = \delta_{i,j}$, $i, j \in \mathcal{D}$, $i \leq j$, describing the orthonormality of the columns of $Q$. One important fact of $O(d, \mathbb{R})$ is that this group has two connected components, one for matrices having determinant equal to 1 and including the identity matrix, and the other one for matrices having determinant equal to $-1$. The set of real orthogonal matrices having determinant equal to 1 build a subgroup, called the special orthogonal group, and denoted by $SO(d, \mathbb{R})$.

If a matrix $Q$ belongs to $SO(d, \mathbb{R})$, then by multiplying $Q$ by the $d \times d$ matrix $I_- = \text{diag}(1, \ldots, 1, -1)$, we obtain $\det(I_-Q) = -1$ and, hence, $I_-Q$ belongs to the connected component of $O(d, \mathbb{R})$ having determinant equal to $-1$. The matrix $I_-$ just creates a mirroring of the last coordinate without affecting the others. Therefore if $Q \in O(d, \mathbb{R})$ and $Q = I_-Q_+$, with $Q_+ \in SO(d, \mathbb{R})$, then $Q$ transforms a polyhedron such that a face of the transformed polyhedron parallels a coordinate axis if and only if $Q_+$ parallels the same axis. Therefore, the set of orthogonal matrices transforming a polyhedron such that a resulting face parallels some axis can be described as a set $S_{Q_+} \subset SO(d, \mathbb{R})$ having this property, or a set $S_{Q_-} \subset (O(d, \mathbb{R}) \setminus SO(d, \mathbb{R}))$ having this property, where $S_{Q_-}$ can be described as $S_{Q_-} = I_+S_{Q_+}$ (that is, every matrix in $S_{Q_-}$ is given as a matrix in $S_{Q_+}$ multiplied by $I_+$). By the invariance property of the Haar measure $\lambda$ over $O(d, \mathbb{R})$, we have that $\lambda(S_{Q_-}) = \lambda(I_+S_{Q_+}) = \lambda(S_{Q_+})$.

The restriction of the Haar measure over $O(d, \mathbb{R})$ to its subgroup $SO(d, \mathbb{R})$ coincides with the Haar measure of $SO(d, \mathbb{R})$ up to a normalization constant. Considering now specially $S_{Q_+}$, our aim now is to show that the zero-set of the multivariate polynomial $|A_r|$ is a set of measure zero with respect to the Haar measure over the group $SO(d, \mathbb{R})$. The special orthogonal group $SO(d, \mathbb{R})$ allows a parametrization via hyperspherical coordinates. We follow the presentation in [9, Chapter 1] and obtain that each $Q \in SO(d, \mathbb{R})$ allows a representation in the form

$$Q^\top = \prod_{i=1}^d \prod_{j=i+1}^d T_{ij} (\varphi_{ij}),$$

where the orthogonal matrices $T_{ij}(\varphi_{ij})$ define a rotation in the coordinate plane $\xi'_i = \cos \varphi_{ij} \xi_i + \sin \varphi_{ij} \xi_j$, $\xi'_j = -\sin \varphi_{ij} \xi_i + \cos \varphi_{ij} \xi_j$, $\xi'_l = \xi_l$ for $l \notin \{i, j\}$, $i < j$, $i, j = 1, \ldots, d$. The representation of $Q$ in this form is unique for almost all values of the angles $\varphi_{ij}$. The angles vary in $0 \leq \varphi_{id} < 2\pi$, $0 \leq \varphi_{ij} < \pi$, $j = i+1, \ldots, d-1$, $i = 1, \ldots, d$. Moreover, the Haar measure on $SO(d, \mathbb{R})$ is absolutely continuous with respect to the Lebesgue measure with the density [9, Theorem 1.2.1]
where \( c_d \) denotes some normalizing constant. By applying this parametrization to the multivariate polynomial \( |A_r| \), one obtains a non-constant analytic function. We recall that the zero-set of a non-constant multivariate analytic function has Lebesgue measure zero [31]. Therefore the restriction of the zero-set of the parametrized multivariate polynomial \( |A_r| \) to the parametrization domain box of the special orthogonal group has zero Lebesgue measure. Hence, the zero-set of the multivariate polynomial \( |A_r| \) has measure zero with respect to the Haar measure over \( SO(d, \mathbb{R}) \). By taking finite unions of the corresponding sets of zero Haar measure over \( SO(d, \mathbb{R}) \) with respect to all \( r \)-minors of \( A \) and all transformed polyhedra \( (Q^\top \Xi)(x) - Q^\top \mu) \) having a face parallel to some coordinate axis, the set of the corresponding special orthogonal matrices has Haar measure zero. Since the latter set transformed by \( I - \) also has zero Haar measure, we arrive at the following statement as a consequence of Theorem 1.

**Theorem 3** Let (A1)–(A5) be satisfied, \( x \in X \) be fixed, \( f = f(x, \cdot) \) be given by \( (28) \) and \( \xi \) be normally distributed with nonsingular covariance matrix \( \Sigma \). After the orthogonal transformation \( (56) \) of \( \xi \) the second order ANOVA approximation \( f^{(2)} \) of \( f \) (see Remark 2) belongs to \( W_{2,\text{mix}}(\mathbb{R}^d) \) for all orthogonal matrices in \( O(d, \mathbb{R}) \) except for those belonging to a subset of \( O(d, \mathbb{R}) \) having Haar measure zero.

According to Remarks 3 and 4 Theorem 3 remains valid for \( k \)th order ANOVA approximations (50) of the integrand \( f \) for \( k \leq \frac{d^2}{2} - 1 \) in the mixed-integer two-stage case and for \( k \leq \frac{d^2}{2} \) in the linear two-stage case.

### 7 Discussion and conclusions

In our numerical experiments reported in the companion paper [26] we compare two randomized QMC methods, namely, randomly shifted lattice rules [45,34] and randomly scrambled Sobol’ point sets (based on [16] and random linear scrambling [29]) with Monte Carlo methods [30] by applying them to a two-stage mixed-integer stochastic electricity portfolio optimization model. Its aim consists in minimizing the expected costs over a time horizon with \( T \) time intervals in the presence of stochastic load and prices. The latter are modelled as multivariate ARMA(1,1) process. The resulting multivariate probability distribution is normal with covariance matrix \( \Sigma \) of dimension \( d = 2T \) which is factorized in the form \( \Sigma = BB^\top \). Two types of factorizations \( B \) are used, namely, (i) standard Cholesky (CH) and (ii) principal component analysis (PCA). Under PCA we obtained in our test runs with \( T = 100 \) that the effective truncation dimension \( d_T(\varepsilon) \) is equal to 2 for \( \varepsilon = 0.01 \) and the mixed-integer two-stage integrands \( f \). We also observed that under PCA the first variable accumulates more than 90% of the total variance \( \sigma^2(f) \). This means \( d_T(0.01) = 2 \) and indeed PCA is an excellent dimension reduction technique in this case. The average of the estimated rates of convergence \( O(n^{-\alpha}) \) for the root mean square error of the optimal values under PCA in our computational tests were approximately \( \alpha = 0.91 \) for randomly shifted lattice rules, and \( \alpha = 1.05 \) for the randomly scrambled Sobol’ points. This is clearly superior to the MC convergence rate \( \alpha = 0.5 \). The same test runs were performed by using CH instead of PCA for factorizing \( \Sigma \). The average of the estimated rates of convergence were \( O(n^{-0.5}) \) for all three methods under CH although the error constants of the randomized QMC methods seemed to be smaller. An explanation for the worse rates is that
the approximate smoothing effect due to the smoothness of lower order ANOVA terms does not occur since the effective truncation dimension under CH always remained $d_T(0.01) = 200$. Compared to our earlier work in [25] we showed for linear two-stage integrands $f$ in the present paper that even ANOVA terms $f_u$ of order $2 \leq |u| < \frac{d}{2}$ have weak mixed first derivatives (Remark 4) and that this property extends to mixed-integer two-stage integrands for $2 \leq |u| \leq \frac{d}{2} - 1$ (Remark 3). However, several questions still remain open. For example, a sufficient condition is desirable implying that lower order ANOVA terms belong to the tensor product Sobolev space $S_d$ (see Theorem 2 in Section 5). Furthermore, a discussion of the geometric condition $(A6)$ beyond the case of normal probability distributions in Section 6 is important.

Acknowledgement: The second author wishes to thank Volker Kaibel (Otto-von-Guericke Universität Magdeburg) for helpful conversations on polyhedral theory.

References


