decomposition. Other choices are also possible, see e.g. [74], but not further investigated here.

2.1 Classical ANOVA Decomposition

For $\Omega = [0,1]$ and the example of the Lebesgue measure $d\mu(\mathbf{x}) = d\mathbf{x}$ in (2.1), the space $V^{(d)}$ is the space of square integrable functions and the projections are given by

$$P_{\mathbf{u}} f(\mathbf{x}_{\mathbf{u}}) = \int_{[0,1]^{d-|\mathbf{u}|}} f(\mathbf{x})\, d\mathbf{x}_{\mathscr{D}\setminus\mathbf{u}}.$$

The decomposition (2.3) then corresponds to the well-known analysis of variance (ANOVA) decomposition which is used in statistics to identify important variables and important interactions between variables in high-dimensional models. It goes back to [73] and has been studied in many different contexts and applications, e.g., [34, 66, 81, 155]. Recently, it has extensively been used for the analysis of quasi-Monte Carlo methods, see, e.g., [17, 97, 98, 102, 148] and the references cited therein.

In the ANOVA the orthogonality (2.6) implies that the variance

$$\sigma^2(f) := \int_{\Omega^d} (f(\mathbf{x}) - If)^2\, d\mu(\mathbf{x})$$

of the function f can be written as

$$\sigma^2(f) = \sum_{\substack{\mathbf{u}\subseteq\mathscr{D}\\ \mathbf{u}\neq\emptyset}} \sigma^2(f_{\mathbf{u}}), \tag{2.7}$$

where $\sigma^2(f_{\mathbf{u}})$ denotes the variance of the term $f_{\mathbf{u}}$.[1] The values $\sigma^2(f_{\mathbf{u}})/\sigma^2(f)$, called global sensitivity indices in [147, 148], can then be used to measure the relative importance of the term $f_{\mathbf{u}}$ with respect to the function f.

Example 2.2. For the class of polynomials in two variables of the form

$$f(x_1, x_2) := a + bx_1 + cx_2 + dx_1x_2$$

with parameters a, b, c and $d \in \mathbb{R}$ one easily calculates that the terms of the ANOVA decomposition are given by

[1] Note that $If_u := \int_{\Omega^d} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) d\mu(\mathbf{x}) = 0$ for $\mathbf{u} \neq \emptyset$.

$$
\begin{aligned}
f_0 &= a+\tfrac{b}{2}+\tfrac{c}{2}+\tfrac{d}{4}\\
f_1(x_1) &= (b+\tfrac{d}{2})(x_1-\tfrac{1}{2})\\
f_2(x_2) &= (c+\tfrac{d}{2})(x_2-\tfrac{1}{2})\\
f_{1,2}(x_1,x_2) &= \tfrac{d}{4}(2x_1-1)(2x_2-1).
\end{aligned}
$$

For the specific case $a=0$, $b=12$, $c=6$ and $d=-6$, i.e. for the polynomial

$$f(x_1,x_2) := 12x_1+6x_2-6x_1x_2$$

one obtains

$$
\begin{aligned}
f_0 &= 15/2\\
f_1(x_1) &= 9x_1-9/2\\
f_2(x_2) &= 3x_2-3/2\\
f_{1,2}(x_1,x_2) &= -6x_1x_2+3x_1+3x_2-3/2.
\end{aligned}
$$

The variance of f is given by $\sigma^2(f) = 31/4$ and we see that $\sigma^2(f_1) = 27/4$, $\sigma^2(f_2) = 3/4$ and $\sigma^2(f_{1,2}) = 1/4$. Hence, the one-dimensional terms f_1 and f_2 explain about 87% and 10% of $\sigma^2(f)$, respectively. The highest-order term $f_{1,2}$ contributes the remaining 3% of the total variance.

Example 2.3. For the slightly modified polynomial

$$f(x_1,x_2) := 12x_1^2+6x_2^2-6x_1x_2$$

we obtain

$$
\begin{aligned}
f_0 &= 9/2\\
f_1(x_1) &= 12x_1^2-3x_1-5/2\\
f_2(x_2) &= 6x_2^2-3x_2-1/2\\
f_{1,2}(x_1,x_2) &= -6x_1x_2+3x_1+3x_2-3/2.
\end{aligned}
$$

It holds $\sigma^2(f)=35/4$, $\sigma^2(f_1)=151/20$, $\sigma^2(f_2)=19/20$ and $\sigma^2(f_{1,2})=1/4$. The first-order terms f_1 and f_2 explain about 86% and 11% of the variance of f and the second-order term $f_{1,2}$ about 3% of the variance.

Example 2.4. For given univariate functions $g_j \in L^2([0,1])$, $j=1,\ldots,d$, let

$$Ig_j := \int_{[0,1]} g_j(x)dx \quad \text{and} \quad \sigma^2(g_j) := \int_{[0,1]} (g_j(x)-Ig_j)^2\,dx.$$

For the classes of purely additive or multiplicative functions

$$f^+(\mathbf{x}) := \sum_{j=1}^d g_j(x_j) \quad \text{and} \quad f^*(\mathbf{x}) := \prod_{j=1}^d g_j(x_j)$$

the ANOVA decomposition can easily be derived analytical, see also [124]. We obtain

$$If^+ = \sum_{j=1}^{d} Ig_j \quad \text{and} \quad \sigma^2(f^+) = \sum_{j=1}^{d} \sigma^2(g_j).$$

Furthermore,

$$If^* = \prod_{j=1}^{d} Ig_j \quad \text{and} \quad \sigma^2(f^*) = \prod_{j=1}^{d} (I^2 g_j + \sigma^2(g_j)) - If^*.$$

The ANOVA terms of f^+ and their variances are given by

$$f_{\mathbf{u}}^+(\mathbf{x}_{\mathbf{u}}) = \begin{cases} If^+ & \text{if } \mathbf{u} = \emptyset \\ g_j(x_j) - Ig_j & \text{if } \mathbf{u} = \{j\} \\ 0 & \text{if } |\mathbf{u}| > 1 \end{cases} \quad \text{and} \quad \sigma^2(f_{\mathbf{u}}^+) = \begin{cases} 0 & \text{if } \mathbf{u} = \emptyset \\ \sigma^2(g_j) & \text{if } \mathbf{u} = \{j\} \\ 0 & \text{if } |\mathbf{u}| > 1. \end{cases}$$

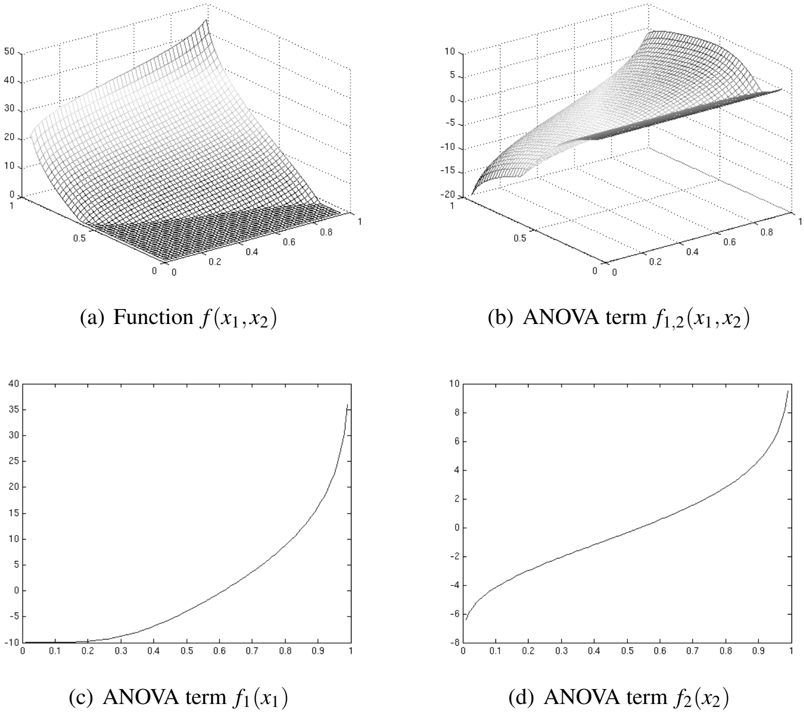
For the function f^*, a simple computation yields

$$f_{\mathbf{u}}^*(\mathbf{x}_{\mathbf{u}}) = \prod_{j \in \mathbf{u}} (g_j(x_j) - Ig_j) \prod_{j \notin \mathbf{u}} Ig_j \quad \text{and} \quad \sigma^2(f_{\mathbf{u}}^*) = \prod_{j \in \mathbf{u}} \sigma^2(g_j) \prod_{j \notin \mathbf{u}} I^2 g_j.$$

Example 2.5. We consider a two-dimensional sample function f which appears in the problem to price Asian options using the Brownian bridge path constructions.[2] The function f and its three ANOVA terms f_1, f_2 and $f_{1,2}$ are displayed in Figure 2.1. Looking at the scales of Figure 2.1(c) and 2.1(d) one observes that the variance of the ANOVA term f_2 is significantly smaller than the variance of f_1. Such a concentration of the variance in the first variables can be exploited for the efficient numerical treatment of high-dimensional integration problems as we will explain in Chapter 3 and Chapter 4. Note further that the first order ANOVA terms are smooth functions despite the fact that f is not differentiable. This smoothing effect of the ANOVA decomposition indicates that also functions of low regularity may be integrated efficiently in high dimensions if the higher-order ANOVA terms are sufficiently small. This smoothing effect was first observed in [102] and is further investigated in [60].

We will come back to these examples in Section 2.2 when we consider the anchored-ANOVA decomposition. The ANOVA decompositions of more complex functions appearing in application problems from finance are investigated in Chapter 6.

[2] This application problem is described in detail in Section 6.2.1. Path constructions are the topic of Section 5.1.



(a) Function $f(x_1,x_2)$

(b) ANOVA term $f_{1,2}(x_1,x_2)$

(c) ANOVA term $f_1(x_1)$

(d) ANOVA term $f_2(x_2)$

Fig. 2.1 Sample function $f(x_1,x_2)$ and its ANOVA terms.

2.1.1 Effective Dimensions

Based on the ANOVA decomposition, different notions of effective dimensions have been introduced in [17]. For the proportion $\alpha \in (0,1]$, the effective dimension in the truncation sense (the *truncation dimension*) of the function f is defined as the smallest integer d_t, such that

$$\sum_{\substack{\mathbf{u}\subseteq\{1,\ldots,d_t\}\\ \mathbf{u}\neq\emptyset}} \sigma^2(f_\mathbf{u}) \geq \alpha\, \sigma^2(f). \tag{2.8}$$

Here, often the proportion $\alpha = 0.99$ is used. The effective dimension in the superposition sense (the *superposition dimension*) is defined as the smallest integer d_s, such that

$$\sum_{\substack{|\mathbf{u}|\leq d_s\\ \mathbf{u}\neq\emptyset}} \sigma^2(f_\mathbf{u}) \geq \alpha\, \sigma^2(f). \tag{2.9}$$

If the variables are ordered according to their importance, the truncation dimension d_t roughly describes the number of important variables of the function f. The super-

Algorithm 2.1: Computing the effective truncation dimension.

```
Compute If and σ²(f)                          /* d-dim. (Q)MC integration */
for j = 1,2,...,d do
    Compute D_{1,2,...,j} using (2.10)       /* (2d − j)-dim. (Q)MC integration */
    if D_{1,2,...,j} > α σ²(f) then return j  /* estimation of d_t */
```

position d_s dimension roughly describes the highest order of important interactions between variables.[3]

Example 2.6. For the function $f(x_1,x_2,x_3) = \exp\{x_1\} + x_2$ with $d = 3$, we obtain (independently of α) that $d_t = 2$ and $d_s = 1$ since the third variable as well as the interaction between two and more variables are unimportant.

For large d, it is no longer possible to compute all 2^d ANOVA terms, but the effective truncation dimension can still be computed in many cases. To this end, let

$$D_{\mathbf{u}} := \sum_{\mathbf{v}\subseteq\mathbf{u}} \sigma^2(f_{\mathbf{v}})$$

denote the variance corresponding to $\mathbf{u}$. Then it holds

$$D_{\mathbf{u}} = \int_{[0,1]^{2d-|\mathbf{u}|}} f(\mathbf{x}) f(\mathbf{x}_{\mathbf{u}}, \mathbf{y}_{\mathscr{D}\setminus\mathbf{u}}) d\mathbf{x}\, d\mathbf{y}_{\mathscr{D}\setminus\mathbf{u}} - (If)^2 \tag{2.10}$$

with $\mathbf{x} = (\mathbf{x}_{\mathbf{u}}, \mathbf{x}_{\mathscr{D}\setminus\mathbf{u}})$ and $\mathbf{y} = (\mathbf{y}_{\mathbf{u}}, \mathbf{y}_{\mathscr{D}\setminus\mathbf{u}})$, see [147, 159]. Using (2.10) the value $D_{\mathbf{u}}$ can be computed by computing a $(2d - |\mathbf{u}|)$-dimensional integral which is possible e.g. by using a Monte Carlo or quasi-Monte Carlo method.[4] Now the effective truncation dimension can be computed as shown in Algorithm 2.1. The more precisely the integrals (2.10) are computed the better is the estimation of the effective truncation dimension d_t which is returned by Algorithm 2.1. Overall the algorithm requires the computation of at most $d+1$ many integrals with up to $2d-1$ dimensions.

Without additional effort, the algorithm also yields the values

$$T_j := \frac{1}{\sigma^2(f)} \sum_{\mathbf{u}\not\subseteq\{1,\ldots,j\}} \sigma^2(f_{\mathbf{u}}) \tag{2.11}$$

for $j = 0,\ldots,d$. The value T_j denotes the percentage of the variance which is not explained by the leading j dimensions. Note that $T_0 = 1$ and $T_d = 0$. The decay of these values often more clearly illustrate the importance of the first dimensions than the truncation dimension since the dependence on the proportion α is avoided.

---

[3] In [102, 124] dimension distributions of f are considered whose 99th percentiles correspond the truncation and superposition dimension, respectively. In addition, the mean dimension $d_m := \sum_{j=1}^{d} \frac{j}{\sigma^2(f)} \sum_{|\mathbf{u}|=j} \sigma^2(f_{\mathbf{u}})$ is defined, which has a similar interpretation as the superposition dimension but is often easier to compute.

[4] (Quasi-) Monte Carlo methods are the topic of Section 3.1.

Algorithm 2.2: Computing the effective superposition dimension.

```
Compute If and σ²(f), D_tot = 0                    /* d-dim. (Q)MC integration */
for i = 1,2,...,d do
    Compute σ_i²(f) = D_i using (2.10)             /* (2d−1)-dim. (Q)MC integ. */
    D_tot = D_tot + σ_i²(f)
if D_tot > α σ²(f) then return d_s = 1             /* superposition dim. of one */
for i = 1,2,...,d do
    for j = i+1, i+2 ..., d do
        Compute D_{i,j} using (2.10)               /* (2d−2)-dim. (Q)MC integ. */
        σ_{i,j}²(f) = D_{i,j} − σ_i²(f) − σ_j²(f)
        D_tot = D_tot + σ_{i,j}²(f)
if D_tot > α σ²(f) then return d_s = 2             /* superposition dim. of two */
else return d_s ≥ 3                                /* superposition dim. of at least three */
```

For the more difficult problem to compute the superposition dimension or the values

$$S_j := \frac{1}{\sigma^2(f)} \sum_{|\mathbf{u}|>j} \sigma^2(f_{\mathbf{u}}). \tag{2.12}$$

the recursive method described in [160] can be used. In Algorithm 2.2 it is shown how functions with low superposition dimension can be identified by computing the variance contributions of the order-1 and order-2 terms of the ANOVA decomposition. The algorithms can be generalized to higher orders but suffers in this case from cancellation problems and costs which grow exponential in d_s. In general the computation of the superposition dimension is only feasible for lower-dimensional functions or for functions with very low superposition dimension.

Example 2.7. As in [159] we consider the multiplicative function

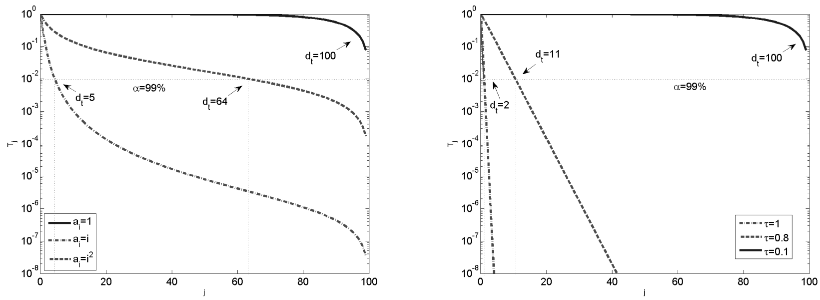
$$f(\mathbf{x}) = \prod_{i=1}^{d} \frac{|4x_i - 2| + a_i}{1 + a_i} \tag{2.13}$$

with $a_i \in \mathbb{R}$ for $i = 1, \ldots, d$. For $\alpha = 99\%$ and $d = 100$ this function has the truncation dimension $d_t = 100$, $d_t = 64$ and $d_t = 5$, respectively, if we choose $a_i = 1$, $a_i = i$ and $a_i = i^2$ for $i = 1, \ldots, d$. The different decay of the importance of the dimensions in these cases is visualized in Figure 2.2(a). There one can see the values T_j from (2.11) for $j = 0, \ldots, 100$ and their relation to d_t.

Example 2.8. We consider the function

$$f(\mathbf{x}) = \prod_{i=1}^{d} \left(1 + a\tau^i (x_i - 1/2)\right) \tag{2.14}$$

from [160]. For $\alpha = 99\%$, $d = 100$ and $a = 1$ one obtains $d_t = 100$, $d_t = 11$ and $d_t = 2$ for the three cases $\tau = 1$, $\tau = 0.8$ and $\tau = 0.1$, respectively. The corresponding values T_j from (2.11) are shown in Figure 2.2(b). A fast decay of T_j means that most of the variance is concentrated in the leading dimensions.



(a) Function (2.13) (b) Function (2.14)

Fig. 2.2 The values T_j, $j = 0, \ldots, 100$, from (2.11) for different sample functions with nominal dimension $d = 100$.

2.1.2 Error Bounds

The following two lemmas relate the effective dimensions to approximation errors. The second lemma is taken from [147].

Lemma 2.1. *Let d_t denote the truncation dimension of the function f with proportion α and let $f_{d_t}(\mathbf{x}) := \sum_{\mathbf{u}\subseteq\{1,\ldots,d_t\}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})$. Then*

$$\|f - f_{d_t}\|_{L_2}^2 \leq (1-\alpha)\sigma^2(f).$$

Proof. Note that $\sigma^2(f_{\mathbf{u}}) = \|f_{\mathbf{u}}\|_2^2$ for $\mathbf{u} \neq \emptyset$ since $\int_{[0,1]^{|\mathbf{u}|}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})\, d\mathbf{x}_{\mathbf{u}} = 0$ for $\mathbf{u} \neq \emptyset$. From (2.3), one obtains

$$\begin{aligned} \|f - f_{d_t}\|_{L_2}^2 &= \|\sum_{\mathbf{u}\not\subseteq\{1,\ldots,d_t\}} f_{\mathbf{u}}\|_{L_2}^2 = \sum_{\mathbf{u}\not\subseteq\{1,\ldots,d_t\}} \|f_{\mathbf{u}}\|_{L_2}^2 \\ &= \sum_{\mathbf{u}\subseteq\mathscr{D}} \sigma^2(f_{\mathbf{u}}) - \sum_{\mathbf{u}\subseteq\{1,\ldots,d_t\}} \sigma^2(f_{\mathbf{u}}) \leq (1-\alpha)\sigma^2(f), \end{aligned}$$

where the second equality holds by orthogonality and where the inequality follows from (2.7) and (2.8).

Lemma 2.2. *Let d_s denote the superposition dimension of the function f with proportion α and let $f_{d_s}(\mathbf{x}) := \sum_{|\mathbf{u}|\leq d_s} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})$. Then*

$$\|f - f_{d_s}\|_{L_2}^2 \leq (1-\alpha)\sigma^2(f).$$

Proof. Similar to Lemma 2.1 we compute

$$\|f - f_{d_s}\|_{L_2}^2 = \| \sum_{|\mathbf{u}|>d_s} f_{\mathbf{u}}\|_{L_2}^2 = \sum_{|\mathbf{u}|>d_s} \|f_{\mathbf{u}}\|_{L_2}^2 = \sum_{|\mathbf{u}|>d_s} \sigma^2(f_{\mathbf{u}}) \leq (1-\alpha)\,\sigma^2(f)$$

using orthogonality, (2.3), (2.7) and (2.9).

For integration, we immediately obtain as corollary the error bound

$$|If - If_{\text{tr}}| \leq \|f - f_{\text{tr}}\|_{L_1} \leq \|f - f_{\text{tr}}\|_{L_2} \leq \sqrt{1-\alpha}\,\sigma(f) \tag{2.15}$$

either if $f_{\text{tr}} := f_{d_t}$ (as in Lemma 2.1) or if $f_{\text{tr}} := f_{d_s}$ (as in Lemma 2.2) and if α is the proportion corresponding to d_t and d_s, respectively. One can see that quadrature methods produce small errors if α is close to one and if the methods can compute If_{tr} efficiently.

Remark 2.1. Quasi-random points[5] are usually more uniformly distributed in smaller dimensions than in higher ones such that we can expect that If_{d_t} is well approximated for small d_t. Moreover, quasi-random points usually have very well distributed low dimensional projections such that we can expect that If_{d_s} is efficiently computed for small d_s. Hence, the bound (2.15) partly explains the success of quasi-Monte Carlo methods for high-dimensional integrals with functions of low truncation dimension or low superposition dimension.[6]

The bound (2.15) also partly explains the success of sparse grid methods[7] for high-dimensional integrals with functions of low effective dimension since these methods can compute If_{tr} very efficiently for small d_s or small d_t with the help of a dimension-adaptive grid refinement.

Remark 2.2. We can also choose $\Omega = \mathbb{R}$ and the Gaussian measure $d\mu(\mathbf{x}) = \varphi_d(\mathbf{x})d\mathbf{x}$ in (2.1) where

$$\varphi_d(\mathbf{x}) := e^{-\mathbf{x}^T\mathbf{x}/2}/(2\pi)^{d/2} \tag{2.16}$$

denotes the standard Gaussian density in d dimensions. This induces projections

$$P_{\mathbf{u}}f(\mathbf{x_u}) = \int_{\mathbf{R}^{d-|\mathbf{u}|}} f(\mathbf{x})\,\varphi_{d-|\mathbf{u}|}(\mathbf{x_u})\,d\mathbf{x}_{\mathscr{D}\setminus\mathbf{u}}.$$

Then, by (2.3), a corresponding decomposition of functions $f : \mathbb{R}^d \to \mathbb{R}$ results, which we refer to as *ANOVA decomposition with Gaussian weight*. Based on this

[5] Quasi-Monte Carlo methods are the topic of Section 3.1.2.

[6] Note that low effective dimension is not necessary for the situation that quasi-Monte Carlo methods are more efficient than Monte Carlo methods for large d and small sample sizes n. This is shown in [153]. Note also that the combination of low effective dimension and of good low dimensional projections is not sufficient to imply that quasi-Monte Carlo is more efficient than Monte Carlo.

[7] We discuss sparse grid methods and their combination with dimension-adaptivity in Chapter 4.

decomposition, effective dimensions for the ANOVA decomposition with Gaussian weight can be defined as in (2.8) and (2.9).

2.2 Anchored-ANOVA Decomposition

For $\Omega = [0,1]$ and the example of the Dirac measure located at a fixed anchor point $\mathbf{a} \in [0,1]^d$, i.e. $d\mu(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{a})d\mathbf{x}$, we obtain from (2.2) the projections

$$P_{\mathbf{u}} f(\mathbf{x}_{\mathbf{u}}) = f(\mathbf{x})|_{\mathbf{x}=\mathbf{a}\backslash\mathbf{x}_{\mathbf{u}}}$$

where we use the notation $f(\mathbf{x})|_{\mathbf{x}=\mathbf{a}\backslash x_i} = f(a_1,\ldots,a_{i-1},x_i,a_{i+1},\ldots,a_d)$ with its obvious generalisation to $\mathbf{a}\backslash\mathbf{x}_{\mathbf{u}}$. We refer to the corresponding decomposition as anchored-ANOVA decomposition. The terms of the anchored-ANOVA decomposition are thus related to the terms of the classical ANOVA decomposition in the sense that all integrals are replaced by point evaluations at a fixed anchor point $\mathbf{a} \in [0,1]^d$. This approach is considered in [135] under the name CUT-HDMR. The decomposition expresses f as superposition of its values on lines, faces, hyperplanes, etc., which intersect the anchor point $\mathbf{a}$ and are parallel to the coordinate axes. It is closely related to the multivariate Taylor expansion [57], and to the anchor spaces considered, e.g., in [29, 70, 143, 165]. There are various generalisations of the anchored-ANOVA decomposition such as the multi-CUT-HDMR, the mp-CUT-HDMR and the lp-RS, see [57] and the references cited therein.

Example 2.9. If we use the anchor point $\mathbf{a} = (1/2, 1/2)$ and again consider the class of polynomials

$$f(x_1,x_2) := a + bx_1 + cx_2 + dx_1x_2 \tag{2.17}$$

with parameters a,b,c and $d \in \mathbb{R}$, then one easily sees that all terms $f_{\mathbf{u}}$ of the anchored-ANOVA decomposition exactly coincides with the terms $f_{\mathbf{u}}$ of the classical ANOVA decomposition shown in Example 2.2.

Example 2.10. For the anchor point $\mathbf{a} = (1/2, 1/2)$ and the polynomial

$$f(x_1,x_2) := 12x_1^2 + 6x_2^2 - 6x_1x_2 \tag{2.18}$$

we obtain the anchored-ANOVA terms

$$\begin{aligned} f_0 &= 3 \\ f_1(x_1) &= 12x_1^2 - 3x_1 - 3/2 \\ f_2(x_2) &= 6x_2^2 - 3x_2 \\ f_{1,2}(x_1,x_2) &= -6x_1x_2 + 3x_1 + 3x_2 - 3/2 \end{aligned}$$

which differ in the constants in comparison with the terms of the classical case from Example 2.3.

http://www.springer.com/978-3-642-16003-5

Sparse Grid Quadrature in High Dimensions with
Applications in Finance and Insurance
Holtz, M.
2011, VIII, 192 p. 32 illus., Hardcover
ISBN: 978-3-642-16003-5