

# Dimension-wise Integration of High-dimensional Functions with Applications to Finance

Michael Griebel, Markus Holtz

*Institute for Numerical Simulation, University of Bonn, Wegelestr. 6, 53115 Bonn, Germany*

---

## Abstract

We present a new general class of methods for the computation of high-dimensional integrals. The quadrature schemes result by truncation and discretization of the anchored-ANOVA decomposition. They are designed to exploit low effective dimensions and include sparse grid methods as special case. To derive bounds for the resulting modelling and discretization errors, we introduce effective dimensions for the anchored-ANOVA decomposition. We show that the new methods can be applied in a locally-adaptive and dimension-adaptive way and demonstrate their efficiency by numerical experiments with high-dimensional integrals from finance.

*Key words:*

ANOVA decomposition, numerical integration, sparse grids, effective dimension

---

## 1. Introduction

High-dimensional integrals appear in various mathematical models from physics, chemistry and finance. Their large number of dimensions arises, e.g., from small time steps in time discretizations and/or a large number of state variables. Examples from finance are option pricing, bond valuation or the pricing of collateral mortgage backed securities. In most cases, the arising integrals cannot be calculated analytically and numerical methods must be applied. Here, one of the key prerequisites is that the *curse of dimension* can be avoided at least to some extent. The curse of dimension states that the cost to compute an approximation with a prescribed accuracy  $\varepsilon$  depends exponentially on the dimension  $d$  of the problem. It is one of the main obstacles for a conventional numerical treatment of high dimensional problems, see, e.g., Griebel (2006). Classical quadrature methods for the computation of multivariate integrals (e.g. based on product rules) which use  $n$  function evaluations achieve an accuracy of

$$\varepsilon(n) = O(n^{-r/d})$$

for functions with bounded derivatives up to order  $r$ , see, e.g., Davis and Rabinowitz (1984). For fixed  $r$ , their convergence rates thus decrease exponentially with increasing dimension. On the positive side, the case  $r = d$  indicates that the problem of a high dimension can sometimes be compensated by, e.g., a high degree of smoothness. Also other aspects such as the concentration

of measure phenomenon<sup>1</sup> or the superposition theorem of Kolmogorov<sup>2</sup> show that there is some chance to treat high-dimensional problems despite the curse of dimension.

Furthermore, the curse of dimension can be approached from the point of numerical complexity theory. There it is shown that for some integration problems even the best algorithm of a specific algorithm class cannot avoid the curse of dimension. Such problems are therefore called intractable, see, e.g. Traub et al. (1988). However, application problems are often in different or smaller problem classes and thus may be tractable. In addition, there may exist algorithms from other settings which are able to break the curse of dimension. Randomised algorithms (e.g. Monte Carlo methods) form one such class of algorithms. For square integrable functions, the expected mean square error of the Monte Carlo method with  $n$  samples is

$$\varepsilon(n) = O(n^{-1/2})$$

and is thus independent of the dimension.<sup>3</sup> Nevertheless, the convergence rate is quite low and a high accuracy is only achievable with tremendously many function evaluations. Deterministic numerical integration schemes, such as quasi-Monte Carlo methods (see, e.g., Glasserman (2003)) and sparse grids (see, e.g., Bungartz and Griebel (2004); Gerstner and Griebel (1998, 2003); Griebel (2006); Novak and Ritter (1996); Smolyak (1963)) are alternatives to the Monte Carlo method, which can attain faster rates of convergence. The error of quasi-Monte Carlo methods can be shown to be

$$\varepsilon(n) = O(n^{-1}(\log n)^d)$$

for integrands of bounded variation and thus decreases with the number of function evaluations  $n$  asymptotically faster than the expected mean square error of the Monte Carlo method. Sparse grids as introduced in Smolyak (1963) achieve

$$\varepsilon(n) = O(n^{-r}(\log n)^{(d-1)(r+1)})$$

for integrands which have bounded mixed partial derivatives of order  $r$  and can thus also make use of higher smoothness of the integrand. The convergence rates of quasi-Monte Carlo and sparse grid methods still exhibit a logarithmic dependence on the dimension, however. Furthermore, the implicit constants depend on  $d$  and often increase exponentially with the dimension. For problems with higher dimensions the asymptotic advantages of the deterministic numerical methods over the Monte Carlo method might thus not pay off for practical sample sizes  $n$ .<sup>4</sup>

However, for many problems of finance, numerical experiments show that quasi-Monte Carlo methods converge nearly independent of the dimension and are faster and more accurate than Monte Carlo, see, e.g., Caffisch et al. (1997); Paskov and Traub (1995). For sufficiently smooth integrands, similar results have been observed for sparse grid methods, see Gerstner and Griebel (1998, 2003). One explanation of this success is based on the analysis of variance (ANOVA) decomposition of the integrands. There it turns out that in many finance applications the importance of each dimension is naturally weighted by certain hidden weights where with the increase

---

<sup>1</sup>The concentration of measure phenomenon says that every Lipschitz function on a sufficiently high dimensional domain is well approximated by a constant function.

<sup>2</sup>The theorem of Kolmogorov shows that every continuous function of several variables can be represented by the superposition of continuous functions with only one variable.

<sup>3</sup>The probabilistic error of a Monte Carlo estimate with  $n$  samples is  $\sigma^2(f)/\sqrt{n}$ . Here, the term  $n^{-1/2}$  describes the convergence rate and the variance  $\sigma^2(f)$  can be considered as the constant of the Monte Carlo method.

<sup>4</sup>Note however that the constants of the Monte Carlo method depend on the variance of the integrand, which can exponentially grow with the dimension. In this case also Monte Carlo integration is intractable.

of dimension the lower-order terms continue to play a significant role and the higher-order terms tend to be negligible, see, e.g., Caffisch et al. (1997); Wang and Sloan (2005). Moreover, often coordinate transformations, such as the Brownian bridge, see Moskowitz and Caffisch (1996), the PCA-construction, see Acworth et al. (1998), or the LT-construction, see Imai and Tan (2006), can be used to enforce the importance of the first few dimensions. The integrands are thus of so called *low effective dimension* and can be well approximated by a sum of low-dimensional functions. In Caffisch et al. (1997), two notions of effective dimension have been introduced, the truncation and the superposition dimension. Quasi-Monte Carlo methods profit from a low superposition dimension by their well-distributed low dimensional projections and from a low truncation dimension by the fact that their points are usually more uniformly distributed in smaller dimensions than in higher ones. Sparse grid methods can exploit low truncation and low superposition dimension in a very general way by a dimension-adaptive grid refinement. If the weights decay sufficiently fast with increasing dimension and if the integrands are of bounded mixed regularity then tractability of the deterministic algorithms can be proved, see Sloan and Woźniakowski (1998). A further argument given in Griebel et al. (2008); Liu and Owen (2006) states that the lower order terms in the ANOVA decomposition are in certain cases smoother than the original function. This may explain the high convergence rates of the deterministic methods despite the fact that application problems do usually not fulfill the smoothness assumptions on bounded mixed regularity.

While the classical ANOVA decomposition is very useful to analyse the importance of different dimensions and their interactions of a high-dimensional function it cannot be used as a tool for the design of new integration schemes since already the first term in the decomposition requires to integrate the function. In this article, we present a new general class of quadrature methods for the computation of high-dimensional integrals, which is based on the anchored-ANOVA decomposition. The anchored-ANOVA decomposition has the advantage that its sub-terms are much cheaper to compute than the terms of the classical ANOVA decomposition since instead of integrals only point evaluations are required. The new quadrature methods, which we refer to as *dimension-wise quadrature methods*, are defined in two steps: First, the anchored-ANOVA decomposition is truncated either a priori based on function space weights or in an *dimension-adaptive* fashion, which automatically detects the important terms of the decomposition. This truncation introduces a *modeling error* which can be shown to be small, however, if the integrand is of low effective dimension. Then, the integrals of the kept terms are computed using appropriate low-dimensional quadrature rules which may be different from term to term and may refine the approximation in a *locally-adaptive* way. This introduces a *discretization error* which, however, only depends on the superposition dimension of the integrand and not on the nominal dimension  $d$ , if we correctly balance the costs of the quadrature methods with the importance of the corresponding anchored-ANOVA terms. We further show that the new method includes the class of generalised sparse grid methods, see Gerstner and Griebel (1998); Hegland (2003); Plaskota (2000); Wasilkowski and Woźniakowski (1999), as special case which results if we use particular tensor product methods for the integration of the sub-terms. This allows us to intertwine the truncation of the anchored-ANOVA series and the subsequent discretization and to balance modelling and discretization error in an optimal way. We demonstrate this approach in more detail for integrands from weighted tensor product Sobolev spaces.

We also present dimension-wise quadrature methods which are not of sparse grid form. Here we use the CUHRE method Bernstein et al. (1991) for the integration of the low-order anchored ANOVA terms and quasi-Monte Carlo methods for the higher order terms. This way, we obtain *mixed CUHRE/QMC methods* which are to our knowledge the first numerical quadrature methods

which can profit from low effective dimension by dimension-adaptivity and can at the same time resolve low regularity by local adaptivity to some extent. Numerical experiments with a test function from finance with discontinuous first derivatives demonstrates the efficiency of this new approach.

We provide several more numerical experiments based on several application problems from finance with up to 512 dimension. The results show that our dimension-wise quadrature methods can profit significantly more than quasi-Monte Carlo methods from coordinate transformation, which reduce the effective dimension, such as the LT-construction. Dimension-adaptive sparse grid methods based on univariate Gauss-Hermite formulas can in addition optimally profit from smoothness by addressing the arising integrals directly on  $\mathbb{R}^d$ . By exploiting both low effective dimension and smoothness, these methods can outperform quasi-Monte Carlo methods by several orders of magnitude even in hundreds of dimensions as our numerical results demonstrate.

The remainder of this article is organised as follows: In Section 2, we recall two different dimension-wise decompositions of multivariate functions, the classical ANOVA and the anchored-ANOVA decomposition, and give corresponding notions of effective dimensions. In Section 3, we then use the anchored-ANOVA to define a new general class of methods for multivariate integration. In Section 4, we discuss its relation to sparse grid methods. In Section 5, we then specify the components of our general approach to balance error and costs for integrands of particular weighted tensor product Sobolev spaces. In Section 6, we show that the methods can be applied in a dimension-adaptive way. In Section 7, we present numerical results with applications from finance. The article finally closes in Section 8 with concluding remarks.

## 2. Dimension-wise Decompositions

In this section, we introduce the classical ANOVA and the anchored-ANOVA of a multivariate function  $f$ . Based on these decomposition, we then define different notions of effective dimensions of  $f$ . To this end, let  $\Omega \subseteq \mathbb{R}$  be a domain and let

$$d\mu(\mathbf{x}) = \prod_{j=1}^d d\mu_j(x_j) \quad (1)$$

denote a  $d$ -dimensional product measure defined on Borel subsets of  $\Omega^d$ . Here,  $\mathbf{x} = (x_1, \dots, x_d)$  and  $\mu_j, j = 1, \dots, d$ , are probability measures on Borel subsets of  $\Omega$ . Let  $V^{(d)}$  denote the Hilbert space of all functions  $f : \Omega^d \rightarrow \mathbb{R}$  with the inner product

$$(f, g) := \int_{\Omega^d} f(\mathbf{x})g(\mathbf{x}) d\mu(\mathbf{x}).$$

For a given set  $\mathbf{u} \subseteq \mathcal{D}$ , where  $\mathcal{D} := \{1, \dots, d\}$  denotes the set of coordinate indices, the measure  $\mu$  induces projections  $P_{\mathbf{u}} : V^{(d)} \rightarrow V^{(|\mathbf{u}|)}$  by

$$P_{\mathbf{u}}f(\mathbf{x}_{\mathbf{u}}) := \int_{\Omega^{d-|\mathbf{u}|}} f(\mathbf{x})d\mu_{\mathcal{D}\setminus\mathbf{u}}(\mathbf{x}). \quad (2)$$

Here,  $\mathbf{x}_{\mathbf{u}}$  denotes the  $|\mathbf{u}|$ -dimensional vector containing those components of  $\mathbf{x}$  whose indices belong to the set  $\mathbf{u}$  and  $d\mu_{\mathcal{D}\setminus\mathbf{u}}(\mathbf{x}) := \prod_{j \notin \mathbf{u}} d\mu_j(x_j)$ . The projections define a decomposition of

$f \in V^{(d)}$  into a finite sum according to

$$f(x_1, \dots, x_d) = f_0 + \sum_{i=1}^d f_i(x_i) + \sum_{\substack{i,j=1 \\ i < j}}^d f_{i,j}(x_i, x_j) + \dots + f_{1,\dots,d}(x_1, \dots, x_d)$$

which is often written in the more compact notation

$$f(\mathbf{x}) = \sum_{\mathbf{u} \subseteq \mathcal{D}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}). \quad (3)$$

The  $2^d$  terms  $f_{\mathbf{u}}$  describe the dependence of the function  $f$  on the dimensions  $j \in \mathbf{u}$  with respect to the measure  $\mu$ . They are recursively defined by

$$f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) := P_{\mathbf{u}}f(\mathbf{x}_{\mathbf{u}}) - \sum_{\mathbf{v} \subset \mathbf{u}} f_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}) \quad (4)$$

and can also be given explicitly by

$$f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) = \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} P_{\mathbf{v}}f(\mathbf{x}_{\mathbf{v}}), \quad (5)$$

see Kuo et al. (2008). The resulting decomposition (3) is unique for a fixed measure  $\mu$  and orthogonal in the sense that

$$(f_{\mathbf{u}}, f_{\mathbf{v}})_{L_2} = 0 \quad (6)$$

for  $\mathbf{u} \neq \mathbf{v}$ , see, e.g., Griebel (2006); Rabitz and Alis (1999).

### 2.1. Classical ANOVA Decomposition

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

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

The decomposition (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. Recently, it has extensively been used for the analysis of QMC methods, see, e.g., Caflisch et al. (1997); Lemieux and Owen (2002); Liu and Owen (2006); Sobol (2001) and the references cited therein.

Here, the orthogonality (6) implies that the variance of the function  $f$  can be written as

$$\sigma^2(f) = \sum_{\substack{\mathbf{u} \subseteq \mathcal{D} \\ \mathbf{u} \neq \emptyset}} \sigma^2(f_{\mathbf{u}}), \quad (7)$$

where  $\sigma^2(f_{\mathbf{u}})$  denotes the variance of the term  $f_{\mathbf{u}}$ . The values  $\sigma^2(f_{\mathbf{u}})/\sigma^2(f)$ , called global sensitivity indices in Sobol (1993, 2001), can then be used to measure the relative importance of the term  $f_{\mathbf{u}}$  with respect to the function  $f$ . Based on the ANOVA decomposition, different notions of effective dimensions have been introduced in Caflisch et al. (1997). 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

$$\sum_{\substack{\mathbf{u} \subseteq \{1, \dots, d_t\} \\ \mathbf{u} \neq \emptyset}} \sigma^2(f_{\mathbf{u}}) \geq \alpha \sigma^2(f). \quad (8)$$

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

$$\sum_{\substack{|\mathbf{u}| \leq d_s \\ \mathbf{u} \neq \emptyset}} \sigma^2(f_{\mathbf{u}}) \geq \alpha \sigma^2(f). \quad (9)$$

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 superposition dimension  $d_s$  roughly describes the highest order of important interactions between variables.

The following two lemmas relate the effective dimensions to approximation errors. The second lemma is taken from Sobol (1993).

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

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

**PROOF.** Note that  $\sigma^2(f_{\mathbf{u}}) = \|f_{\mathbf{u}}\|_{L_2}^2$  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 (3), one obtains

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

where the second equality holds by orthogonality and where the inequality follows from (7) and (8).

**Lemma 2.** *Let  $d_s$  denote the superposition dimension of the function  $f$  with proportion  $\alpha$  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 1 we compute

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

using orthogonality, (3), (7) and (9).

For integration, we immediately obtain as corollary the error bound

$$|If - If_{\text{tr}}| \leq \|f - f_{\text{tr}}\|_{L_1} \leq \|f - f_{\text{tr}}\|_{L_2} \leq \sqrt{1 - \alpha} \sigma(f) \quad (10)$$

either if the function  $f_{\text{tr}} := f_{d_t}$  as in Lemma 1 or if  $f_{\text{tr}} := f_{d_s}$  as in Lemma 2 and if  $\alpha$  is the proportion corresponding to  $d_t$  and  $d_s$ , respectively. One can see that quadrature methods produce small errors if  $\alpha$  is close to one and if the methods can compute  $If_{\text{tr}}$  efficiently.

Quasi-random points 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 (10) partly explains the success of quasi-Monte Carlo methods for high-dimensional integrals with functions of low truncation dimension or low superposition dimension. The bound (10) also partly explains the success of sparse grid methods for high-dimensional integrals with functions of low effective dimension since these methods can compute  $If_{\text{tr}}$  very efficiently for small  $d_s$  or small  $d_t$  with the help of a dimension-adaptive grid refinement.

**Remark 1.** We can also choose  $\Omega = \mathbb{R}$  and the Gaussian measure  $d\mu(\mathbf{x}) = \varphi_d(\mathbf{x})d\mathbf{x}$  in (1) where

$$\varphi_d(\mathbf{x}) := e^{-\mathbf{x}^T \mathbf{x} / 2} / (2\pi)^{d/2} \quad (11)$$

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

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

Then, by (3), a corresponding decomposition of the function  $f$  on  $\mathbb{R}^d$  results, which we refer to as *ANOVA decomposition with Gaussian weight*. Based on this decomposition, effective dimensions for the ANOVA decomposition with Gaussian weight can be defined as in (8) and (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) the projections

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

where we use the notation  $f(\mathbf{x})|_{\mathbf{x}=\mathbf{a}\setminus x_i} = f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_d)$  with its obvious generalisation to  $\mathbf{a}\setminus\mathbf{x}_{\mathbf{u}}$ . 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 Rabitz and Alis (1999) 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 and to anchored Sobolev spaces, see Griebel (2006) and the references cited therein.

While the classical ANOVA decomposition is very useful to analyse the importance of different dimensions and of their interactions it cannot be used as a tool for the design of new integration schemes since already the constant term in the classical ANOVA decomposition requires to compute the integral. The anchored-ANOVA decomposition has the advantage that its sub-terms are much cheaper to compute since instead of integrals only point evaluations at the anchor point  $\mathbf{a} \in [0, 1]^d$  are required. We will use this property in Section 3 to design new quadrature methods for high-dimensional functions.

We next define a new notion of effective dimension which is based on the anchored-ANOVA decomposition. While the effective dimensions in the classical case are based on the  $L_2$ -norm,

we now introduce effective dimensions for the anchored case, which are based on the operator  $|I(\cdot)|$  and, since  $|I(f)| = |\int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x}| \leq \|f\|_{L_1}$ , which are thus related to the  $L_1$ -norm. While the effective dimensions in the classical case directly lead to error bounds for approximation (see Lemma 1 and Lemma 2), we will use the effective dimensions in the anchored case to derive error bounds for integration (see Lemma 3 and Lemma 4).<sup>5</sup> To this end, let

$$\hat{\sigma}(f) := \sum_{\substack{\mathbf{u} \subseteq \mathcal{D} \\ \mathbf{u} \neq \emptyset}} |If_{\mathbf{u}}| \leq \sum_{\substack{\mathbf{u} \subseteq \mathcal{D} \\ \mathbf{u} \neq \emptyset}} \|f_{\mathbf{u}}\|_{L_1} \quad (12)$$

be the sum of the absolute values of the integrals of all anchored-ANOVA terms. Then, analogous to (8), for the proportion  $\alpha \in (0, 1]$ , the *truncation dimension in the anchored case* is defined as the smallest integer  $d_t$ , such that

$$\sum_{\substack{\mathbf{u} \subseteq \{1, \dots, d_t\} \\ \mathbf{u} \neq \emptyset}} |If_{\mathbf{u}}| \geq \alpha \hat{\sigma}(f), \quad (13)$$

whereas, analogous to (9), the *superposition dimension in the anchored case* is defined as the smallest integer  $d_s$ , such that

$$\sum_{\substack{|\mathbf{u}| \leq d_s \\ \mathbf{u} \neq \emptyset}} |If_{\mathbf{u}}| \geq \alpha \hat{\sigma}(f). \quad (14)$$

As in the classical case, these notions describe roughly the number of important dimensions and the order of important interactions, respectively. Compared to the classical case, the effective dimensions in the anchored case have the following advantages: They are directly related to integration errors as we show below and they can easily be determined by dimension-wise integration methods as we will explain in Section 3 in more detail. We also have a direct relation of the effective dimensions in the anchored case to sparse grid methods as we will show in Section 4. As  $|I(\cdot)|$  is not a norm, it may happen, however, that the effective dimensions in the anchored case fail to detect some important dimensions and interactions.<sup>6</sup> This may be the case if  $f$  is a function of varying sign. For instance, let  $\mathbf{a} = (1/2, 1/2)$  and consider the function  $f(x_1, x_2) = e^{x_1} - e^{0.5} + x_2 - \frac{1}{2}$ . Then we obtain  $f_2(x_2) = x_2 - \frac{1}{2}$  such that  $\sigma^2(f_2) > 0$  but  $|If_2| = 0$  which misleadingly indicates independence of  $x_2$ . This effect which we have not yet observed in practical applications from finance, though, is closely related to the early determination problem of dimension-adaptive sparse grid methods which is discussed in Gerstner and Griebel (2003).

To compute the effective dimensions in the anchored case, we use multivariate quadrature methods to compute approximations  $q_{\mathbf{u}} \approx If_{\mathbf{u}}$ . By summation of the computed values  $q_{\mathbf{u}}$ ,  $\mathbf{u} \subseteq \mathcal{D}$ , we estimate  $\hat{\sigma}(f)$  and the effective dimensions  $d_t$  and  $d_s$  according to (12), (13) and (14).

The following two estimates relate effective dimensions in the anchored case and integration errors.

**Lemma 3.** *Let  $d_t$  denote the truncation dimension of the function  $f$  in the anchored case with proportion  $\alpha$  and let  $f_{d_t}(\mathbf{x}) := \sum_{\mathbf{u} \subseteq \{1, \dots, d_t\}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})$ . Then*

$$|If - If_{d_t}| \leq (1 - \alpha) \hat{\sigma}(f).$$

<sup>5</sup>The approximation error of the truncated anchored ANOVA decomposition is studied in Sobol (2003); Wang (2008). There also the impact of the choice of the anchor point is investigated.

<sup>6</sup>It is also possible to base the definition on the  $L_1$ -norm. Then this drawback disappears. Nevertheless we here stick to the operator  $|I(\cdot)|$  to exploit a more direct relation to dimension-adaptive sparse grid methods.



PROOF. We obtain

$$|If - If_{d_s}| = \left| \sum_{\mathbf{u} \notin \{1, \dots, d_s\}} If_{\mathbf{u}} \right| \leq \sum_{\mathbf{u} \notin \{1, \dots, d_s\}} |If_{\mathbf{u}}| = \sum_{\mathbf{u} \subseteq \mathcal{D}} |If_{\mathbf{u}}| - \sum_{\mathbf{u} \subseteq \{1, \dots, d_s\}} |If_{\mathbf{u}}| \leq (1 - \alpha) \hat{\sigma}(f)$$

where the first equality results from (3) and from the definition of the function  $f_{d_s}$ . The last inequality follows from (12) and (13).

**Lemma 4.** *Let  $d_s$  denote the superposition dimension of the function  $f$  in the anchored case with proportion  $\alpha$  and let  $f_{d_s}(\mathbf{x}) := \sum_{|\mathbf{u}| \leq d_s} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})$ . Then*

$$|If - If_{d_s}| \leq (1 - \alpha) \hat{\sigma}(f).$$

PROOF. Similar to Lemma 3 we obtain

$$|If - If_{d_s}| = \left| \sum_{|\mathbf{u}| > d_s} If_{\mathbf{u}} \right| \leq \sum_{|\mathbf{u}| > d_s} |If_{\mathbf{u}}| \leq (1 - \alpha) \hat{\sigma}(f)$$

using (12) and (14) for the last inequality.

**Remark 2.** We can also choose  $\Omega = \mathbb{R}$  and the measure  $d\mu(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{a})\varphi_d(\mathbf{x})d\mathbf{x}$  with a fixed anchor point  $\mathbf{a} \in \mathbb{R}^d$ , where  $\varphi_d$  is the Gaussian density (11). This example induces projections

$$P_{\mathbf{u}}f(\mathbf{x}_{\mathbf{u}}) = (f(\mathbf{x})\varphi_{d-|\mathbf{u}|}(\mathbf{x}_{\mathbf{u}}))|_{\mathbf{x}=\mathbf{a}\setminus\mathbf{x}_{\mathbf{u}}}$$

and, by (3), a corresponding decomposition of functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  which we refer to as *anchored-ANOVA decomposition with Gaussian weight*. Based on this decomposition, effective dimensions for the anchored-ANOVA decomposition with Gaussian weight can be defined as in (13) and (14).

### 3. Dimension-wise Quadrature Methods

Next, we use the anchored-ANOVA decomposition to define a new class of methods for the computation of high-dimensional integrals

$$If := \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} \tag{15}$$

on the unit cube and for integrals

$$I_{\varphi}f := \int_{\mathbb{R}^d} f(\mathbf{z}) \varphi_d(\mathbf{z}) d\mathbf{z} \tag{16}$$

on  $\mathbb{R}^d$  with the Gaussian weight function  $\varphi_d$  from (11). Note that these two domains typically appear in high-dimensional applications.

### 3.1. Truncation and Discretization

In the following, we develop our new class of quadrature methods. We start with  $\Omega = [0, 1]$  and take  $\mu$  as the Dirac measure with anchor point  $\mathbf{a} \in [0, 1]^d$ . Then, (2) and (4) imply

$$f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) = P_{\mathbf{u}}f(\mathbf{x}_{\mathbf{u}}) - \sum_{\mathbf{v} \subset \mathbf{u}} f_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}) \quad \text{where } P_{\mathbf{u}}f(\mathbf{x}_{\mathbf{u}}) = f(\mathbf{x})|_{\mathbf{x}=\mathbf{a} \setminus \mathbf{x}_{\mathbf{u}}}. \quad (17)$$

Applying the integral operator to the anchored-ANOVA decomposition (3), the  $d$ -dimensional integral is decomposed, by linearity, into the finite sum

$$If = \sum_{\mathbf{u} \subseteq \mathcal{D}} If_{\mathbf{u}} = f(\mathbf{a}) + \sum_{i=1}^d If_i + \sum_{\substack{i,j=1 \\ j < i}}^d If_{i,j} + \dots + If_{1,\dots,d} \quad (18)$$

which contains  $\binom{d}{j}$  many  $j$ -dimensional integrals for  $j = 0, \dots, d$ . Starting with the decomposition (18) we now define a general class of quadrature methods for the approximation of  $If$ . We proceed as follows:

1. *Truncation:* We take only a subset  $\mathcal{S}$  of all indices  $\mathbf{u} \subseteq \mathcal{D}$  and thus truncate the sum in (18). Here, we assume that the set  $\mathcal{S}$  satisfies the admissibility condition<sup>7</sup>

$$\mathbf{u} \in \mathcal{S} \text{ and } \mathbf{v} \subset \mathbf{u} \implies \mathbf{v} \in \mathcal{S}. \quad (19)$$

For example, the set  $\mathcal{S}_{d_s} := \{\mathbf{u} \subseteq \mathcal{D} : |\mathbf{u}| \leq d_s\}$  or the set  $\mathcal{S}_{d_t} := \{\mathbf{u} \subseteq \{1, \dots, d_t\}\}$  could be used to take into account the superposition and the truncation dimension of the function  $f$ , respectively. Alternatively, dimension-wise adaptive methods can be applied to build up an appropriate index set  $\mathcal{S}$ . This will be later discussed in more detail.

2. *Discretization:* For each  $\mathbf{u} \in \mathcal{S}$ , we compute approximations to  $If_{\mathbf{u}}$ . To this end, we choose  $|\mathbf{u}|$ -dimensional quadrature rules  $Q_{\mathbf{u}}$ . Starting with  $q_{\emptyset} = f(\mathbf{a})$  we recursively compute

$$q_{\mathbf{u}} := Q_{\mathbf{u}}(P_{\mathbf{u}}f) - \sum_{\mathbf{v} \subset \mathbf{u}} q_{\mathbf{v}}. \quad (20)$$

Then,  $q_{\mathbf{u}}$  is an approximation to  $If_{\mathbf{u}}$  due to the recursive representation (17) of  $f_{\mathbf{u}}$ . Observe that we avoid to compute and integrate the functions  $f_{\mathbf{u}}$  explicitly. Instead we numerically integrate  $P_{\mathbf{u}}f$  and correct the resulting value by the (previously computed) values  $q_{\mathbf{v}}$  using (17). The admissibility condition ensures that we can run over the set  $\mathcal{S}$  by starting with  $\mathbf{u} = \emptyset$  and proceeding with indices  $\mathbf{u}$  for which the values  $q_{\mathbf{v}}$ ,  $\mathbf{v} \subset \mathbf{u}$ , have already been computed in previous steps. Note that we allow for arbitrary quadrature methods  $Q_{\mathbf{u}}$  in (20) which can be different for each  $\mathbf{u}$ . Specific choices for  $Q_{\mathbf{u}}$  will be discussed later.

Altogether, this defines a quadrature formula  $A_S f$  for the approximation of  $If$  which is given by

$$A_S f := \sum_{\mathbf{u} \in \mathcal{S}} q_{\mathbf{u}} \quad (21)$$

---

<sup>7</sup>Note that this condition is closely related to the admissibility condition (30) for sparse grid indices in Section 4.1.

and which we refer to as *dimension-wise quadrature method* in the following. Note that the method  $A_S f$  requires

$$n = \sum_{\mathbf{u} \in \mathcal{S}} n_{\mathbf{u}}$$

evaluations of the function  $f$ , where  $n_{\mathbf{u}}$  denotes the number of function evaluations of  $Q_{\mathbf{u}}$ .

**Remark 3.** Dimension-wise quadrature methods for integrals on  $\mathbb{R}^d$  with Gaussian weight can be constructed analogously to (21). To this end, we set  $\Omega = \mathbb{R}$  and use the measure  $d\mu(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{a})\varphi_d(\mathbf{x})d\mathbf{x}$  such that the anchored-ANOVA decomposition with Gaussian weighted results. Then, we select as above a suitable index set  $\mathcal{S}$  and appropriate quadrature rules  $Q_{\mathbf{u}}$  to integrate the resulting functions  $P_{\mathbf{u}}f$ . Since now  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , either quadrature rules for unbounded domains, e.g. Gauss-Hermite rules, or transformations of the resulting integrals to the unit cube must be used.

### 3.2. Error and Costs

We first consider the case of arbitrary quadrature methods  $Q_{\mathbf{u}}$ . By construction, we then have the error bound

$$|If - A_S f| = \left| \sum_{\mathbf{u} \in \mathcal{D}} If_{\mathbf{u}} - \sum_{\mathbf{u} \in \mathcal{S}} q_{\mathbf{u}} \right| \leq \sum_{\mathbf{u} \in \mathcal{S}} |If_{\mathbf{u}} - q_{\mathbf{u}}| + \sum_{\mathbf{u} \notin \mathcal{S}} |If_{\mathbf{u}}|. \quad (22)$$

This shows how the error of the method (21) depends on the quadrature rules  $Q_{\mathbf{u}}$  (which determine  $q_{\mathbf{u}}$ ) and on the choice of the index set  $\mathcal{S}$ . Here, the second term describes the *modelling error* which is introduced by the truncation of the anchored-ANOVA series whereas the first term describes the *discretization error* which results from the subsequent discretization of the remaining subspaces.

In the following, we aim to balance costs and accuracies by relating the cost of the quadrature method  $Q_{\mathbf{u}}$  to the importance of the anchored-ANOVA term  $f_{\mathbf{u}}$ . We first relate the accuracy of the methods  $Q_{\mathbf{u}}$  to the accuracy of the method  $A_S f$ .

To this end, we fix  $\alpha \in (0, 1]$  and assume that  $d_s$  and  $d_t$ , the corresponding superposition and truncation dimensions in the anchored case, are known. With help of these effective dimensions we define the index set

$$\mathcal{S}_{d_t, d_s} := \{\mathbf{u} \subseteq \{1, \dots, d_t\} : |\mathbf{u}| \leq d_s\}. \quad (23)$$

We now have the following lemma.

**Lemma 5.** *Let  $\mathcal{S} = \mathcal{S}_{d_t, d_s}$ . For  $\varepsilon > 0$ , let furthermore  $Q_{\mathbf{u}}$  be such that  $|I(P_{\mathbf{u}}f) - Q_{\mathbf{u}}(P_{\mathbf{u}}f)| \leq \varepsilon(|\mathbf{u}|)$  with  $\varepsilon(j) := \varepsilon/(e d_t^{d_s} \binom{d_t}{j})$  for all  $\mathbf{u} \in \mathcal{S}$ . Then, it holds*

$$|If - A_S f| \leq \varepsilon + 2(1 - \alpha) \hat{\sigma}(f).$$

**PROOF.** Starting with  $|If - A_S f| \leq |If - If_{d_t, d_s}| + |If_{d_t, d_s} - A_S f|$  with the function  $f_{d_t, d_s} := \sum_{\mathbf{u} \in \mathcal{S}_{d_t, d_s}} f_{\mathbf{u}}$ , we observe that the modelling error is bounded by

$$|If - If_{d_t, d_s}| = \left| \sum_{\mathbf{u} \notin \mathcal{S}_{d_t, d_s}} If_{\mathbf{u}} \right| \leq \sum_{|\mathbf{u}| > d_s} |If_{\mathbf{u}}| + \sum_{\mathbf{u} \notin \{1, \dots, d_t\}} |If_{\mathbf{u}}| \leq 2(1 - \alpha) \hat{\sigma}(f),$$

see the proofs of Lemma 3 and 4. From (5) and (20), we have the explicit representation

$$If_{\mathbf{u}} - q_{\mathbf{u}} = \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} (I(P_{\mathbf{v}}f) - Q_{\mathbf{v}}(P_{\mathbf{v}}f)). \quad (24)$$

Since  $|I(P_{\mathbf{v}}f) - Q_{\mathbf{v}}(P_{\mathbf{v}}f)| \leq \varepsilon(|\mathbf{v}|)$  for all  $\mathbf{v} \subseteq \mathbf{u}$ , we obtain for all  $\mathbf{u}$  with  $|\mathbf{u}| \leq d_t$  that

$$|If_{\mathbf{u}} - q_{\mathbf{u}}| \leq \sum_{\mathbf{v} \subseteq \mathbf{u}} \varepsilon(|\mathbf{v}|) = \sum_{j=1}^{|\mathbf{u}|} \binom{|\mathbf{u}|}{j} \varepsilon(j) \leq \sum_{j=1}^{|\mathbf{u}|} \binom{d_t}{j} \varepsilon(j),$$

where we used that there are  $\binom{|\mathbf{u}|}{j}$  many sets  $\mathbf{v} \subseteq \mathbf{u}$  which satisfy  $|\mathbf{v}| = j$ . Using the definition of  $\varepsilon(j)$ , we can bound the discretization error by

$$\begin{aligned} |If_{d_t, d_s} - A_{S_{d_t, d_s}} f| &\leq \sum_{\mathbf{u} \in S_{d_t, d_s}} |If_{\mathbf{u}} - q_{\mathbf{u}}| \leq \sum_{k=1}^{d_s} \binom{d_t}{k} \sum_{j=1}^k \binom{d_t}{j} \varepsilon(j) \leq \sum_{k=1}^{d_s} \binom{d_t}{k} \sum_{j=1}^k \frac{\varepsilon}{e d_t^{d_s}} \\ &= \frac{\varepsilon}{e d_t^{d_s}} \sum_{k=1}^{d_s} \binom{d_t}{k} k \leq \frac{\varepsilon}{e d_t^{d_s}} \sum_{k=1}^{d_s} \frac{d_t^{d_s}}{k!} k = \frac{\varepsilon}{e} \sum_{k=1}^{d_s} \frac{1}{(k-1)!} \leq \varepsilon, \end{aligned}$$

which concludes the proof.

We next relate the error  $|If - A_S f|$  to the cost  $n = \sum_{\mathbf{u} \in S} n_{\mathbf{u}}$  of the method  $A_S f$ . Furthermore, we aim to balance the cost  $n_{\mathbf{u}}$  of the methods  $Q_{\mathbf{u}}$  with their accuracy. Here, we restrict ourselves to the case that all employed methods  $Q_{\mathbf{u}}$  are based on a univariate quadrature formula  $U_m$  with  $m$  points, which converges for  $f \in C^r([0, 1])$  with rate  $m^{-r}$ . An examples for such a univariate formula with  $r = 1$  is the trapezoidal rule. For arbitrary  $r$ , Gauss rules can be used.

**Theorem 6 (Error versus cost).** *If we choose  $S = S_{d_t, d_s}$  and  $Q_{\mathbf{u}}$  to be the  $|\mathbf{u}|$ -dimensional tensor product of the rule  $U_m$  with  $m := \lfloor n^{1/d_s} \rfloor$  then*

$$|If - A_S f| \leq c(d_t, d_s) n^{-r/d_s} + 2(1 - \alpha) \hat{\sigma}(f)$$

for all  $f \in C^r([0, 1]^d)$ . Here, the constant  $c(d_t, d_s)$  depends on the effective dimensions  $d_t$  and  $d_s$  in the anchored case, but not on the nominal dimension  $d$ .

**PROOF.** We have the same modelling error as in Lemma 5, i.e.,

$$|If - A_S f| \leq |If_{d_t, d_s} - A_S f| + 2(1 - \alpha) \hat{\sigma}(f)$$

with the function  $f_{d_t, d_s} := \sum_{\mathbf{u} \in S_{d_t, d_s}} f_{\mathbf{u}}$ . Since  $f \in C^r([0, 1]^d)$  also  $f_{\mathbf{u}} \in C^r([0, 1]^{|\mathbf{u}|})$  for all  $\mathbf{u} \subseteq \mathcal{D}$ . Consequently,  $Q_{\mathbf{u}}$  converge with rate  $r/|\mathbf{u}|$ . By definition,  $Q_{\mathbf{u}}$  requires  $n_{\mathbf{u}} = \lfloor n^{|\mathbf{u}|/d_s} \rfloor$  function evaluations such that

$$|I(P_{\mathbf{u}}f) - Q_{\mathbf{u}}(P_{\mathbf{u}}f)| \leq c(|\mathbf{u}|) n_{\mathbf{u}}^{-r/|\mathbf{u}|} \leq c(|\mathbf{u}|) n^{-r/d_s}$$

for a constant  $c(|\mathbf{u}|) > 0$  which depends on the order  $|\mathbf{u}|$ .<sup>8</sup> With the help of (24) we estimate

$$\begin{aligned} |If_{d_t, d_s} - A_{S_{d_t, d_s}} f| &\leq \sum_{\mathbf{u} \in S_{d_t, d_s}} |If_{\mathbf{u}} - q_{\mathbf{u}}| \leq \sum_{\mathbf{u} \in S_{d_t, d_s}} \sum_{\mathbf{v} \subseteq \mathbf{u}} |I(P_{\mathbf{v}}f) - Q_{\mathbf{v}}(P_{\mathbf{v}}f)| \\ &\leq \sum_{\mathbf{u} \in S_{d_t, d_s}} \sum_{\mathbf{v} \subseteq \mathbf{u}} c(|\mathbf{v}|) n^{-r/d_s} = \sum_{k=1}^{d_s} \binom{d_t}{k} \sum_{j=1}^k \binom{k}{j} c(j) n^{-r/d_s} = c(d_t, d_s) n^{-r/d_s} \end{aligned}$$

<sup>8</sup>Note that the constant  $c(|\mathbf{u}|)$  depends on the norm of  $P_{\mathbf{u}}f$  and thus also on the smoothness parameter  $r$ .

with the constant

$$c(d_t, d_s) := \sum_{k=1}^{d_s} \binom{d_t}{k} \sum_{j=1}^k \binom{k}{j} c(j) \leq \bar{c}(d_s) \sum_{k=1}^{d_s} \binom{d_t}{k} 2^k \leq \bar{c}(d_s) \sum_{k=1}^{d_s} \frac{d_t^{d_s}}{k!} 2^k \leq \bar{c}(d_s)(e^2 - 1) d_t^{d_s},$$

where  $\bar{c}(d_s) := \max_{j=1, \dots, d_s} c(j)$ . This completes the proof.

Note that the first term in the error bound describes the discretization error which depends on  $n$  whereas the second term corresponds to the modeling error which depends on the proportion  $\alpha$ . Note furthermore that the cost to obtain a prescribed discretization error does not exponentially depend on the nominal dimension  $d$ , but only on the superposition dimension  $d_s$  in the anchored case.

### 3.3. A Priori Construction using Function Space Weights

In applications, the effective dimensions of  $f$  are usually unknown. These dimensions can also not be computed since this would be at least as expensive as the integration of  $f$ . In general it is thus difficult to determine the set  $\mathcal{S}_{d_t, d_s}$  in (23).

To overcome this obstacle, we here assume that the integrand is contained in some function class that is defined by certain function space weights  $\gamma_{\mathbf{u}} \geq 0$ , which describe the importance of the term  $f_{\mathbf{u}}$  of the anchored-ANOVA decomposition. Using this a priori information, we then determine the index set  $\mathcal{S}$  by including those indices  $\mathbf{u}$  which correspond to the largest weights  $\gamma_{\mathbf{u}}$ . In the following, we use the set

$$\mathcal{S}_{\gamma} := \{\mathbf{u} \subseteq \mathcal{D} : \gamma_{\mathbf{u}} > \varepsilon\}$$

which includes all indices  $\mathbf{u}$  for which  $\gamma_{\mathbf{u}}$  is larger than some threshold  $\varepsilon$ . It is known (see, e.g. Wang and Sloan (2005)) that many functions  $f$  in practise are of low effective dimension either in the truncation or in the superposition sense. For these two classes of functions we can hope to determine the index set  $\mathcal{S}_{\gamma}$  of the most important terms by the following approaches to define the weights: *order-dependent weights* for functions  $f$  with low superposition dimension and *product weights* for functions  $f$  with low truncation dimension.

- *Order-dependent Weights:* We define the order-dependent weights  $\gamma_{\mathbf{u}} = 1/|\mathbf{u}|$ . Then, the indices are added according to their order  $|\mathbf{u}|$ . Note that by construction, the admissibility condition (19) is always satisfied. The weights are the larger the lower the order of the anchored-Anova term. If the function  $f$  has a small superposition dimension then we can hope that the resulting index set  $\mathcal{S}_{\gamma}$  includes the most important terms.
- *Product Weights:* As in Sloan and Woźniakowski (1998), we assume that the dimensions are ordered according to their importance which is modulated by a sequence of weights

$$\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_d \geq 0.$$

Using the weights  $\gamma_i$ , we then assign the product weight

$$\gamma_{\mathbf{u}} := \prod_{j \in \mathbf{u}} \gamma_j \tag{25}$$

to the index  $\mathbf{u} \subseteq \mathcal{D}$ . The weights  $\gamma_i$  can here either be input parameters of the algorithm similar to the CBC construction of lattice rules Sloan et al. (2002) or they can be derived

from the first order terms  $f_j$  of the anchored-ANOVA decomposition. In the latter case they are defined by

$$\gamma_j := \frac{|q_j|}{|q_0|} = \frac{|Q_j(P_j f) - f(\mathbf{a})|}{|f(\mathbf{a})|}$$

for  $j = 1, \dots, d$ , see (20). The weights  $\gamma_{\mathbf{u}}$  are the larger the lower the dimensions that are associated with their index set  $\mathbf{u}$ . In this way we can hope that the resulting index set  $\mathcal{S}_\gamma$  includes the most important terms if the function  $f$  has a small truncation dimension.

We will use these weights and the resulting index sets  $\mathcal{S}_\gamma$  in our numerical experiments in Section 7. Note that also more general weights can be used in our construction as long as the admissibility condition (19) is satisfied. Nevertheless, we here only shifted the problem of the choice of  $\mathcal{S}$  to the problem of determining the weights  $\gamma_{\mathbf{u}}$ . In Section 6, we will consider a different approach. There, we will determine the index set  $\mathcal{S}$  a posteriori in a dimension-adaptive way.

#### 4. Sparse Grids Quadrature

In this section, we use tensor product methods  $Q_{\mathbf{u}}$  for the approximation of the integrals  $I_{\mathbf{u}} f$  in (18). This allows us to intertwine the truncation of the anchored-ANOVA series and the subsequent discretization and allows to balance modelling and discretization error in an optimal way. We will demonstrate this in Section 5 for integrands from weighted tensor product Sobolev spaces in more detail.

##### 4.1. Generalised Sparse Grids

For a univariate function  $f : [0, 1] \rightarrow \mathbb{R}$  and a sequence of non-decreasing integers  $m_k, k \in \mathbb{N}$ , let

$$U_{m_k} f := \sum_{i=1}^{m_k} w_{i,k} f(x_{i,k}) \quad (26)$$

denote a sequence of univariate quadrature rules with  $m_k$  points  $x_{i,k}$  and weights  $w_{i,k}$ , which converges pointwise to  $I f$  for  $k \rightarrow \infty$ . We assume  $m_1 = 1$  and  $U_1 f = f(1/2)$  and define the difference quadrature formulae

$$\Delta_k := U_{m_k} - U_{m_{k-1}} \quad \text{with } U_{m_0} := 0 \quad (27)$$

for  $k \geq 1$ .

Now let  $f : [0, 1]^d \rightarrow \mathbb{R}$  be a multivariate function. Then, the  $d$ -dimensional integral  $I f$  can be represented by the infinite telescoping sum

$$I f = \sum_{\mathbf{k} \in \mathbb{N}^d} \Delta_{\mathbf{k}} f \quad (28)$$

which collects the products of each possible combination of the univariate difference formulae. Here,  $\mathbf{k} \in \mathbb{N}^d$  denotes a multi-index with  $k_j > 0$  and

$$\Delta_{\mathbf{k}} f := (\Delta_{k_1} \otimes \dots \otimes \Delta_{k_d}) f. \quad (29)$$

A specific class of quadrature methods for the approximation of  $I f$  is then obtained by a truncation of the sum (28) using an appropriate index set  $\mathcal{I} \subset \mathbb{N}^d$ , which can be regarded as a

refinement of the index set  $\mathcal{S} \subseteq \mathcal{D}$  from Section 3.1. To ensure the validity of the telescoping sum expansion the index set  $\mathcal{I}$  has to satisfy the admissibility condition

$$\mathbf{k} \in \mathcal{I} \text{ and } \mathbf{l} \leq \mathbf{k} \implies \mathbf{l} \in \mathcal{I}, \quad (30)$$

where  $\mathbf{l} \leq \mathbf{k}$  is defined by  $l_j \leq k_j$  for  $j = 1, \dots, d$ . In this way, the generalised sparse grid method

$$\text{SG}_{\mathcal{I}}f := \sum_{\mathbf{k} \in \mathcal{I}} \Delta_{\mathbf{k}}f \quad (31)$$

is obtained, see, e.g., Gerstner and Griebel (1998); Hegland (2003); Plaskota (2000); Wasilkowski and Woźniakowski (1999). Different ways to truncate the sum then correspond to different quadrature methods. Examples are the classical sparse grid construction from Smolyak (1963), often denoted as Smolyak method, which is, on level  $\ell \in \mathbb{N}$ , recovered with the index set

$$\mathcal{I} = \{\mathbf{k} \in \mathbb{N}^d : |\mathbf{k}|_1 \leq \ell + d - 1\} \quad (32)$$

where  $|\mathbf{k}|_1 := \sum_{j=1}^d k_j$ , or product methods, which correspond on level  $\ell$  to index sets of the form

$$\mathcal{I} = \{\mathbf{k} \in \mathbb{N}^d : |\mathbf{k}|_{\infty} \leq \ell\} \quad (33)$$

where  $|\mathbf{k}|_{\infty} := \max\{k_j : j = 1, \dots, d\}$ .

**Remark 4.** Sparse grid methods can directly be applied to the numerical computation of integrals on  $\mathbb{R}^d$  with Gaussian weight. To this end, only the sequence of univariate quadrature rules  $U_{m_k}$  must be replaced by quadrature formulas for functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  on unbounded domains, such as Gauss-Hermite rules, see, e.g., Novak et al. (1998).

#### 4.2. Relation to Dimension-wise Quadrature Methods

There is a close relation of the sparse grid approach and the anchored-ANOVA decomposition. The sparse grid approach (31) can indeed be interpreted as a refinement of the anchored-ANOVA decomposition by first expanding each term of the decomposition into an infinite basis and then truncating this expansion appropriately.<sup>9</sup> It can thus be regarded as special case of the method (21) where the set  $\mathcal{S}$  and the rules  $Q_{\mathbf{u}}$  are chosen in a systematic way to exploit smoothness of the integrand. We now show this in more detail. To this end, we always use the anchor  $\mathbf{a} = (1/2, \dots, 1/2)$ . We start with the following lemma.

**Lemma 7.** *Let  $f_{\mathbf{u}}$  and  $P_{\mathbf{u}}f$  be as in (17) and*

$$\mathcal{N}_{\mathbf{u}} := \{\mathbf{k} \in \mathbb{N}^d : k_j > 1 \text{ if and only if } j \in \mathbf{u}\}. \quad (34)$$

*Then,  $\Delta_{\mathbf{k}}f = \Delta_{\mathbf{k}}(P_{\mathbf{u}}f)$  if  $\mathbf{k} \in \mathcal{N}_{\mathbf{v}}$  with  $\mathbf{v} \subseteq \mathbf{u}$ . Moreover,  $\Delta_{\mathbf{k}}f = \Delta_{\mathbf{k}}f_{\mathbf{u}}$  if  $\mathbf{k} \in \mathcal{N}_{\mathbf{u}}$ .*

**PROOF.** The proof follows from the fact that the projection  $P_{\mathbf{u}}$  and the the operator  $\Delta_{\mathbf{l}}$  are defined in a way such that  $f$  is evaluated at the same anchor point. Indeed, if  $\mathbf{k} \in \mathcal{N}_{\mathbf{v}}$  and  $\mathbf{v} \subseteq \mathbf{u}$  then  $k_j = 1$  for all  $j \notin \mathbf{u}$  and thus  $\Delta_{\mathbf{k}}f = \Delta_{\mathbf{k}}(P_{\mathbf{u}}f)$  since  $\Delta_{\mathbf{k}} = \Delta_{k_1} \otimes \dots \otimes \Delta_{k_d}$  and  $\Delta_{\mathbf{l}}f = P_{\mathbf{l}}f = f(1/2)$

<sup>9</sup>Note the close relation to Lemieux and Owen (2002).

for all univariate functions  $f$ . To show the second assertion, let  $\mathbf{k} \in \mathcal{N}_{\mathbf{u}}$ . Then, we obtain from (4) that

$$\Delta_{\mathbf{k}}(P_{\mathbf{u}}f) = \Delta_{\mathbf{k}}f_{\mathbf{u}} + \sum_{\mathbf{v} \subset \mathbf{u}} \Delta_{\mathbf{k}}f_{\mathbf{v}}.$$

Since  $f_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}})|_{x_j=1/2} = 0$  for all  $j \in \mathbf{v}$ , which is a direct consequence of the orthogonality (6), we conclude  $\Delta_{\mathbf{k}}f_{\mathbf{v}} = 0$  for all  $\mathbf{v} \subset \mathbf{u}$  and  $\mathbf{k} \in \mathcal{N}_{\mathbf{u}}$ . This proves  $\Delta_{\mathbf{k}}f = \Delta_{\mathbf{k}}(P_{\mathbf{u}}f) = \Delta_{\mathbf{k}}f_{\mathbf{u}}$  for all  $\mathbf{k} \in \mathcal{N}_{\mathbf{u}}$ .

By Lemma 7 and (28) we obtain

$$If = \sum_{\mathbf{u} \subseteq \mathcal{D}} \sum_{\mathbf{k} \in \mathcal{N}_{\mathbf{u}}} \Delta_{\mathbf{k}}f$$

since  $\mathbb{N}^d$  is the disjoint union of the sets  $\mathcal{N}_{\mathbf{u}}$ ,  $\mathbf{u} \subseteq \mathcal{D}$ . By (18), we also have  $If = \sum_{\mathbf{u} \subseteq \mathcal{D}} If_{\mathbf{u}}$ , which yield a decomposition

$$If_{\mathbf{u}} = \sum_{\mathbf{k} \in \mathcal{N}_{\mathbf{u}}} \Delta_{\mathbf{k}}f$$

of the integrals of the anchored-ANOVA terms into an infinite sum. Next, we truncate this sum. To this end, we select index sets  $\mathcal{I}_{\mathbf{u}} \subset \mathcal{N}_{\mathbf{u}}$  for all  $\mathbf{u} \subseteq \mathcal{D}$ , which satisfy the condition (30), and use

$$q_{\mathbf{u}} := \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{u}}} \Delta_{\mathbf{k}}f \quad (35)$$

as approximation to  $If_{\mathbf{u}}$ . The corresponding method (21) with  $\mathcal{S} = \mathcal{D}$  can then be represented as

$$A_{\mathcal{S}}f = \sum_{\mathbf{u} \subseteq \mathcal{D}} q_{\mathbf{u}} = \sum_{\mathbf{k} \in \mathcal{I}} \Delta_{\mathbf{k}}f$$

with the index set  $\mathcal{I} = \bigcup_{\mathbf{u} \subseteq \mathcal{D}} \mathcal{I}_{\mathbf{u}}$ . We see that in this way both, the modelling and the discretization error is expressed<sup>10</sup> in terms of the values  $\Delta_{\mathbf{k}}f$ . We further see that the resulting method  $A_{\mathcal{S}}f$  coincides with the generalised sparse grid approach (31). To this end, we define

$$\mathcal{I}_{\mathbf{u}} := \left\{ \mathbf{k} \in \mathcal{I} : k_j > 1 \text{ if and only if } j \in \mathbf{u} \right\} = \mathcal{I} \cap \mathcal{N}_{\mathbf{u}}. \quad (36)$$

**Theorem 8.** *The dimension-wise quadrature method (21) with anchor  $\mathbf{a} = (\frac{1}{2}, \dots, \frac{1}{2})$ , the index set  $\mathcal{S} = \mathcal{D}$  and the quadrature methods*

$$Q_{\mathbf{u}}f := \sum_{\mathbf{v} \subset \mathbf{u}} \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{v}}} \Delta_{\mathbf{k}}f \quad (37)$$

*coincides with the generalised sparse grid method (31).*

**PROOF.** We have to show that (20) holds with  $q_{\mathbf{u}}$  as in (35) and  $Q_{\mathbf{u}}$  as in (37). In fact,

$$\begin{aligned} Q_{\mathbf{u}}(P_{\mathbf{u}}f) - \sum_{\mathbf{v} \subset \mathbf{u}} q_{\mathbf{v}} &= \sum_{\mathbf{v} \subset \mathbf{u}} \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{v}}} \Delta_{\mathbf{k}}(P_{\mathbf{u}}f) - \sum_{\mathbf{v} \subset \mathbf{u}} \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{v}}} \Delta_{\mathbf{k}}f \\ &= \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{u}}} \Delta_{\mathbf{k}}(P_{\mathbf{u}}f) + \sum_{\mathbf{v} \subset \mathbf{u}} \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{v}}} (\Delta_{\mathbf{k}}(P_{\mathbf{u}}f) - \Delta_{\mathbf{k}}f) = \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{u}}} \Delta_{\mathbf{k}}f = q_{\mathbf{u}} \end{aligned}$$

where we twice used Lemma 7.

<sup>10</sup>Modelling errors are here represented by the case  $\mathcal{I}_{\mathbf{u}} = \emptyset$  for any  $\mathbf{u} \subseteq \mathcal{D}$ .



**Remark 5.** Similar to Theorem 8, we see that generalised sparse grid methods for integrals on  $\mathbb{R}^d$  with Gaussian weight (e.g. sparse grids based on Gauss-Hermite rules) are special cases of the dimension-wise quadrature method for integrals with Gaussian weight, see Remark 3. Both methods result from a discretization of the terms of the anchored-ANOVA decomposition with Gaussian weights using the anchor  $\mathbf{a} = (0, \dots, 0)$ , see Remark 2.

## 5. Optimal Sparse Grids in Weighted Spaces

In Section 4, we introduced the index set  $\mathcal{I} \subset \mathbb{N}^d$  as a refinement of the index set  $\mathcal{S} \subseteq \mathcal{D}$  and specified the quadrature rules  $Q_{\mathbf{u}}$  such that the general approach (21) corresponds to the class of generalised sparse grid methods. Now we determine the index set  $\mathcal{I}$  which balances the resulting modelling and discretization errors in an optimal way for integrands from weighted tensor product Sobolev spaces, see, e.g., Sloan and Woźniakowski (1998); Wasilkowski and Woźniakowski (1999). To this end, we partly proceed as in Bungartz and Griebel (2004); Wasilkowski and Woźniakowski (1999).

For reasons of simplicity, we restrict ourselves to the case that the univariate quadrature rules  $U_{m_i}$  in (26) are given by the trapezoidal rule with  $m_1 = 1$ ,  $U_1 f = f(0)$  and  $m_i = 1 + 2^{i-2}$  points for  $i \geq 2$ . Our analysis is based on the univariate function space

$$H_\gamma^1([0, 1]) := \{f : [0, 1] \rightarrow \mathbb{R} : \|f\|_{1,\gamma} < \infty\}$$

with the norm

$$\|f\|_{1,\gamma}^2 := f(0)^2 + \gamma^{-1} \|f'\|_{L_2}^2, \quad (38)$$

where  $\gamma \in (0, 1]$  denotes a weight. In the multivariate case we consider a given sequence of weights

$$1 = \gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_d \geq 0$$

and assign to each set  $\mathbf{u} \subseteq \mathcal{D}$  the product weight  $\gamma_{\mathbf{u}}$  from (25). We then define the tensor product space<sup>11</sup>

$$H_\gamma^{1,\text{mix}}([0, 1]^d) := \bigotimes_{j=1}^d H_{\gamma_j}^1([0, 1])$$

with the norm

$$\|f\|_{1,\gamma}^2 := \sum_{\mathbf{u} \in \mathcal{D}} \gamma_{\mathbf{u}}^{-1} \|f_{\mathbf{u}}\|_{1,\text{mix}}^2$$

with

$$\|f_{\mathbf{u}}\|_{1,\text{mix}}^2 := \int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial^{|\mathbf{u}|}}{\partial \mathbf{x}_{\mathbf{u}}} f(\mathbf{x}_{\mathbf{u}}, \mathbf{0}) \right|^2 d\mathbf{x}_{\mathbf{u}},$$

where  $f_{\mathbf{u}}$  denote the sub-terms of the anchored-ANOVA decomposition anchored at the origin.<sup>12</sup>

<sup>11</sup>Note that  $H_\gamma^{1,\text{mix}}([0, 1]^d)$  is the reproducing kernel Hilbert space to the product kernel  $K(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^d k(x_j, y_j)$ , where  $k(x, y) := 1 + \gamma \min\{x, y\}$  is the reproducing kernel of the space  $H_\gamma^1([0, 1])$ .

<sup>12</sup>It is also possible to anchor the space  $H_\gamma^1([0, 1])$  at the point  $1/2$ .

### 5.1. Cost-Benefit Ratio

For the space  $H_\gamma^{\text{mix}}([0, 1]^d)$  we next determine the index set  $\mathcal{I}$  of the generalised sparse grid method  $\text{SG}_\mathcal{I}$ , which has the best possible cost-benefit ratio. To this end, we first associate each index  $\mathbf{k} \in \mathbb{N}^d$  with a local cost value, namely the number of function evaluations  $n_{\mathbf{k}}$  required by  $\Delta_{\mathbf{k}}f$ . Since the methods  $U_{m_i}$  are nested and since  $m_i \leq 2^{i-1}$ , we have

$$n_{\mathbf{k}} = \prod_{j=1}^d m_{k_j} \leq 2^{|\mathbf{k}-\mathbf{1}|_1} =: c_{\mathbf{k}}.$$

For the global costs of (31) we thus have the bound

$$\sum_{\mathbf{k} \in \mathcal{I}} n_{\mathbf{k}} \leq \sum_{\mathbf{k} \in \mathcal{I}} c_{\mathbf{k}} =: n_{\mathcal{I}}. \quad (39)$$

We now consider the error of the method  $\text{SG}_\mathcal{I}$ . To this end, note that

$$|If - \text{SG}_\mathcal{I}f| = \left| \sum_{\mathbf{k} \in \mathbb{N}^d} \Delta_{\mathbf{k}}f - \sum_{\mathbf{k} \in \mathcal{I}} \Delta_{\mathbf{k}}f \right| \leq \sum_{\mathbf{k} \in \mathbb{N}^d \setminus \mathcal{I}} |\Delta_{\mathbf{k}}f|. \quad (40)$$

To derive bounds for  $\Delta_{\mathbf{k}}f$ , we associate to each index  $\mathbf{k} \in \mathbb{N}^d$  the product weight

$$\gamma_{\mathbf{k}} := \prod_{\substack{j=1, \dots, d \\ k_j > 1}} \gamma_j,$$

where the product is taken over all  $j$  for which  $k_j > 1$  holds.

**Lemma 9.** *It holds*

$$|\Delta_{\mathbf{k}}f| \leq b_{\mathbf{k}} \|f\|_{1,\gamma} \quad (41)$$

where

$$b_{\mathbf{k}} := 2^{-|\mathbf{k}-\mathbf{1}|_1} \gamma_{\mathbf{k}}^{1/2}. \quad (42)$$

**PROOF.** We first consider the univariate case and show that

$$|\Delta_i f| \leq \gamma^{1/2} 2^{-i+1} \|f\|_{1,\gamma} \quad (43)$$

for  $i \geq 2$ . In fact, by (38) we have

$$\|f'\|_{L_2}^2 = \sqrt{\gamma(\|f\|_{1,\gamma}^2 - f(0)^2)} \leq \gamma^{1/2} \|f\|_{1,\gamma}.$$

Therefore,

$$|\Delta_i f| = |U_{m_i}f - U_{m_{i-1}}f| \leq 2^{-i+1} \|f'\|_{L_2}^2 \leq \gamma^{1/2} 2^{-i+1} \|f\|_{1,\gamma}$$

for  $i \geq 2$ , where a proof of the first inequality can be found in Wasilkowski and Woźniakowski (1999). For  $i = 1$ , we have  $|\Delta_i f| = |U_{m_1}f| = |f(0)| \leq \|f\|_{1,\gamma}$ . Using the tensor product structure  $\Delta_{\mathbf{k}} = \bigotimes_{i=1}^d \Delta_{k_i}$  we obtain the assertion.

Motivated by (41), we refer to  $b_{\mathbf{k}}$  in the following as the local benefit associated with the index  $\mathbf{k} \in \mathbb{N}^d$ . The global benefit of the method (31) is then given by

$$B_{\mathcal{I}} := \sum_{\mathbf{k} \in \mathcal{I}} b_{\mathbf{k}}. \quad (44)$$

This leads to the restricted optimization problem

$$\max_{n_{\mathcal{I}}=w} B_{\mathcal{I}}, \quad w \in \mathbb{N}$$

to maximize the global benefit  $B_{\mathcal{I}}$  for fixed global costs  $n_{\mathcal{I}}$ . Using the argument from Bungartz and Griebel (2004), this *global* optimization problem can be reduced to the problem of ordering the *local* cost-benefit ratios

$$cbr_{\mathbf{k}} := b_{\mathbf{k}}/c_{\mathbf{k}} = 2^{-2|\mathbf{k}-\mathbf{1}|_1} \gamma_{\mathbf{k}}^{1/2} \quad (45)$$

associated with the index  $\mathbf{k}$  according to their size. The optimal index set  $\mathcal{I}$  then contains all indices whose local cost-benefit ratios are larger than or equal to some constant value. Here, we use the value

$$cbr_{\bar{\mathbf{k}}} := 2^{-2(\ell-1)} \quad (46)$$

as threshold, which is associated with the index  $\bar{\mathbf{k}} = (\ell, 1, \dots, 1)$ .

**Theorem 10 (Optimal sparse grids in the weighted case).** *The optimal index set in the weighted case is given by*

$$\mathcal{I}_{\ell, \gamma} := \{\mathbf{k} \in \mathbb{N}^d : |\mathbf{k}|_1 + \sigma_{\mathbf{k}} \leq \ell + d - 1\} \quad (47)$$

where

$$\sigma_{\mathbf{k}} := \sum_{\substack{j=1, \dots, d \\ k_j > 1}} \sigma_j \text{ with } \sigma_j := -\log_2(\gamma_j)/4.$$

PROOF. Using

$$\gamma_{\mathbf{k}}^{1/2} = 2^{\sum_{j \in \mathcal{D}_{\mathbf{k}}} \log_2(\gamma_j)/2} = 2^{-2\sigma_{\mathbf{k}}} \text{ with } \mathcal{D}_{\mathbf{k}} := \{j \in \mathcal{D} : k_j > 1\}$$

we obtain from (45) that

$$cbr_{\mathbf{k}} = 2^{-2(|\mathbf{k}-\mathbf{1}|_1 + \sigma_{\mathbf{k}})}.$$

The comparison with (46) shows that  $cbr_{\mathbf{k}} \geq cbr_{\bar{\mathbf{k}}}$  if and only if  $-2(|\mathbf{k}-\mathbf{1}|_1 + \sigma_{\mathbf{k}}) \geq -2(\ell-1)$ , i.e., if  $|\mathbf{k}|_1 - d + \sigma_{\mathbf{k}} \leq \ell - 1$ , which proves the assertion.

The resulting sparse grid method with the index set  $\mathcal{I}_{\ell, \gamma}$  is then given by

$$\text{SG}_{\ell, \gamma} f := \sum_{\mathbf{k} \in \mathcal{I}_{\ell, \gamma}} \Delta_{\mathbf{k}} f. \quad (48)$$

Note that the method  $\text{SG}_{\ell, \gamma}$  is the classical sparse grid approach (32) in the unweighted case  $\gamma_j = 1, j = 1, \dots, d$ .

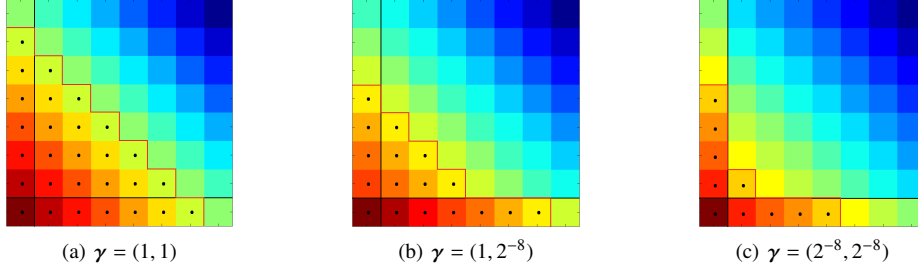


Figure 1: Optimal index sets  $\mathcal{I}_{\ell, \gamma}$  on the level  $\ell = 7$ .

**Example 1.** For illustration, the resulting optimal index sets  $\mathcal{I}_{\ell, \gamma}$  on level  $\ell = 7$  are shown in Figure 1 for  $d = 2$  and different choices of the weights  $\gamma = (\gamma_1, \gamma_2)$ . There, the local cost-benefit ratios of the indices  $\mathbf{k} = (k_1, k_2)$ ,  $k_i \in \{1, \dots, 8\}$ ,  $i = 1, 2$ , are color-coded. In addition the indices  $\mathbf{k}$  which belong to  $\mathcal{I}_{\ell, \gamma}$  with  $\ell = 7$  are marked with a dot.

One can see that the index sets  $\mathcal{I}_{\ell, \gamma}$  can be represented by  $\mathcal{I}_{\ell, \gamma} = \bigcup_{\mathbf{u} \subset \{1, 2\}} \mathcal{I}_{\mathbf{u}}$ , i.e., the disjoint union of the four subsets

$$\begin{aligned}
 \mathcal{I}_0 &= \{(1, 1)\}, \\
 \mathcal{I}_1 &= \{(k_1, 1) : k_1 > 1 \text{ and } k_1 < b_1\}, \\
 \mathcal{I}_2 &= \{(1, k_2) : k_2 > 1 \text{ and } k_2 < b_2\}, \\
 \mathcal{I}_{12} &= \{(k_1, k_2) : k_1, k_2 > 1 \text{ and } |\mathbf{k}|_1 < b_{12}\},
 \end{aligned} \tag{49}$$

with  $b_1, b_2, b_{12} \in \mathbb{N}$  which depend on the weights  $\gamma_1$  and  $\gamma_2$ , the dimension  $d$  and on the level  $\ell$ . Note that all four subsets correspond to index sets of classical sparse grid methods. In general  $2^d$  subsets are required; one for each ANOVA subterm. We will use this decomposition of the index set in the next two sections to derive cost and error bounds for the generalized sparse grid method  $\text{SG}_{\ell, \gamma}$  from the known cost and error bounds in the unweighted case.

## 5.2. Cost Analysis

In the following, we use  $n(d, \ell, \gamma)$  to denote the number of function evaluations of the method  $\text{SG}_{\ell, \gamma}$ . To analyse these costs we first recall the well-known cost bound for classical sparse grids, see, e.g., Bungartz and Griebel (2004); Wasilkowski and Woźniakowski (1995). In this case, we omit the index  $\gamma$  and write  $n(d, \ell) := n(d, \ell, \mathbf{1})$ .

**Lemma 11 (Costs of classical sparse grids).** *Let  $m_i \leq 2^{i-1}$ . Then*

$$n(d, \ell) \leq 2^\ell \binom{\ell + d - 2}{d - 1}.$$

In the following, we will also consider sparse grid methods that start with  $m_1 = 2$  points on their lowest level  $\ell = 1$  instead of  $m_1 = 1$  as in the case of Lemma 11. The index set of such a sparse grid method with level  $\ell$  and dimension  $d$  can be written in the form

$$\tilde{\mathcal{I}} = \{\mathbf{k} \in (\mathbb{N} \setminus \{1\})^d : |\mathbf{k}|_1 \leq \ell + 2d - 1\}. \tag{50}$$

As a corollary of Lemma 11, we see that if  $m_i \leq 2^i$  then the number of points in such a sparse grid satisfies

$$\bar{n}(d, \ell) \leq 2^{d+\ell} \binom{\ell + d - 2}{d - 1}. \quad (51)$$

We now present a generalised cost bound for the weighted case. The cost bound results from the insight that the sparse grid method  $\text{SG}_{\ell, \gamma}$  can be represented by the combination of  $2^d$  many classical sparse grids (one for each anchored-ANOVA term). Here and in the following, we define that  $\binom{n}{d} := 0$  for  $n < d$  and that  $\binom{x}{d} := \binom{\lfloor x \rfloor}{d}$  for  $x \in \mathbb{R}$ .

**Theorem 12 (Costs of weighted sparse grids).** *Let  $m_i \leq 2^{i-1}$ . Then*

$$n(d, \ell, \gamma) \leq 2^\ell \sum_{\mathbf{u} \subseteq \mathcal{D}} \gamma_{\mathbf{u}}^{1/4} \binom{\ell + \log_2(\gamma_{\mathbf{u}})/4 - 2}{|\mathbf{u}| - 1}.$$

**PROOF.** We start with the remark that the index set  $\mathcal{I}_{\ell, \gamma}$  from (47) can be represented by  $\mathcal{I}_{\ell, \gamma} = \bigcup_{\mathbf{u} \subseteq \mathcal{D}} \mathcal{I}_{\mathbf{u}}$  as the disjoint union of the sets

$$\mathcal{I}_{\mathbf{u}} = \{\mathbf{k} \in \mathbb{N}^d : |\mathbf{k}|_1 + \sigma_{\mathbf{k}} \leq \ell + d - 1 \text{ and } k_j > 1 \text{ if and only if } j \in \mathbf{u}\},$$

see also (47). Thus,

$$n(d, \ell, \gamma) \leq \sum_{\mathbf{k} \in \mathcal{I}_{\ell, \gamma}} n_{\mathbf{k}} = \sum_{\mathbf{u} \subseteq \mathcal{D}} \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{u}}} n_{\mathbf{k}}. \quad (52)$$

Let  $\mathbf{l} \in (\mathbb{N} \setminus \{1\})^{|\mathbf{u}|}$  denote the vector that collects all components of  $\mathbf{k} \in \mathcal{I}_{\mathbf{u}}$  that are larger than one. We then can write

$$\begin{aligned} \mathcal{I}_{\mathbf{u}} &= \{\mathbf{l} \in (\mathbb{N} \setminus \{1\})^{|\mathbf{u}|} : |\mathbf{l}|_1 + d - |\mathbf{u}| + \sigma_{\mathbf{u}} \leq \ell + d - 1\} \\ &= \{\mathbf{l} \in (\mathbb{N} \setminus \{1\})^{|\mathbf{u}|} : |\mathbf{l}|_1 \leq \ell - \sigma_{\mathbf{u}} - |\mathbf{u}| + 2|\mathbf{u}| - 1\}, \end{aligned}$$

where  $\sigma_{\mathbf{u}} = \sum_{j \in \mathbf{u}} \sigma_j$ . By comparison with (50), we see that  $\mathcal{I}_{\mathbf{u}}$  is the index set of a  $|\mathbf{u}|$ -dimensional classical sparse grid starting with  $m_1 = 2$  points and with level  $\ell - \sigma_{\mathbf{u}} - |\mathbf{u}|$ . Hence by (51),

$$\sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{u}}} n_{\mathbf{k}} = \bar{n}(|\mathbf{u}|, \ell - \sigma_{\mathbf{u}} - |\mathbf{u}|) \leq 2^{\ell - \sigma_{\mathbf{u}}} \binom{\ell - \sigma_{\mathbf{u}} - 2}{|\mathbf{u}| - 1}. \quad (53)$$

Using  $\sigma_{\mathbf{u}} = -\log_2(\gamma_{\mathbf{u}})/4$  and  $2^{-\sigma_{\mathbf{u}}} = \gamma_{\mathbf{u}}^{1/4}$  we obtain the assertion by combining (52) and (53).

Note that Theorem 12 recovers Lemma 11 in the unweighted case  $\gamma_j = 1, j = 1, \dots, d$ . This can be seen as follows: We have

$$n(d, \ell, \mathbf{1}) \leq 2^\ell \sum_{\mathbf{u} \subseteq \mathcal{D}} \binom{\ell - 2}{|\mathbf{u}| - 1} = 2^\ell \sum_{j=1}^d \binom{d}{j} \binom{\ell - 2}{j - 1}$$

and use

$$\sum_{j=1}^d \binom{d}{j} \binom{\ell - 2}{j - 1} = \sum_{j=1}^d \binom{d}{d-j} \binom{\ell - 2}{j - 1} = \sum_{j=0}^{d-1} \binom{d}{d-1-j} \binom{\ell - 2}{j} = \binom{d + \ell - 2}{d - 1}$$

where we applied the Vandermonde's identity for the last equality.

### 5.3. Error Analysis

We now consider the error of the method  $SG_{\ell,\gamma}$ . To this end, we start with an error bound for classical sparse grids, see, e.g., Bungartz and Griebel (2004). In the unweighted case, we again omit the index  $\gamma$  and write  $\|f\|_1 := \|f\|_{1,1}$ .

**Lemma 13 (Error of classical sparse grids).** *Let  $SG_\ell$  denote the classical sparse grid method with the index set (32) and let  $\ell \geq d - 1$ . Then*

$$\|f - SG_\ell f\| \leq \sum_{\mathbf{k} \in \mathbb{N}^d \setminus \mathcal{I}} 2^{-|\mathbf{k}-1|_1} \|f\|_1 \leq 2^{-\ell} A(d, \ell) \|f\|_1$$

where

$$A(d, \ell) := 2d \binom{\ell + d - 1}{d - 1}.$$

If  $\overline{SG}_\ell$  denotes the classical sparse grid method with the index set  $\overline{\mathcal{I}}$  from (50) that starts with  $m_1 = 2$  points and if  $\ell \geq d - 1$ , then we obtain as a corollary of Lemma 13 that

$$\|f - \overline{SG}_\ell f\| \leq \sum_{\mathbf{k} \in (\mathbb{N} \setminus \{1\})^d \setminus \overline{\mathcal{I}}} 2^{-|\mathbf{k}-1|_1} \|f\|_1 \leq 2^{-d-\ell} A(d, \ell) \|f\|_1. \quad (54)$$

We now present an error bound for the weighted case. To derive this bound, we again use the fact that the error of the sparse grid method  $SG_{\ell,\gamma}$  can be bounded by the sum of the errors of  $2^d$  many classical sparse grids (one for each anchored-ANOVA term).

**Theorem 14 (Error of weighted sparse grids).** *Let  $\ell \geq d - \log_2(\gamma_{(1,\dots,d)})/4 - 1$ . Then*

$$\|f - SG_{\ell,\gamma} f\| \leq 2^{-\ell} \sum_{\mathbf{u} \subseteq \mathcal{D}} \gamma_{\mathbf{u}}^{1/4} 2^{|\mathbf{u}|} \binom{\ell + \log_2(\gamma_{\mathbf{u}})/4 - 1}{|\mathbf{u}| - 1} \|f\|_{1,\gamma}.$$

PROOF. We start with

$$\|f - SG_{\ell,\gamma} f\| \leq \sum_{\mathbf{k} \in \mathbb{N}^d \setminus \mathcal{I}_{\ell,\gamma}} b_{\mathbf{k}} \|f\|_{1,\gamma} \quad (55)$$

which follows from (40) and (41). Note that

$$\mathbb{N}^d \setminus \mathcal{I}_{\ell,\gamma} = \bigcup_{\mathbf{u} \subseteq \mathcal{D}} (\mathcal{N}_{\mathbf{u}} \setminus \mathcal{I}_{\mathbf{u}})$$

with  $\mathcal{N}_{\mathbf{u}}$  as in (34) and  $\mathcal{I}_{\mathbf{u}}$  as in the proof of Theorem 12. By (42), we thus have

$$\sum_{\mathbf{k} \in \mathbb{N}^d \setminus \mathcal{I}_{\ell,\gamma}} b_{\mathbf{k}} \|f\|_{1,\gamma} = \sum_{\mathbf{u} \subseteq \mathcal{D}} \gamma_{\mathbf{u}}^{1/2} \sum_{\mathbf{k} \in \mathcal{N}_{\mathbf{u}} \setminus \mathcal{I}_{\mathbf{u}}} 2^{-|\mathbf{k}-1|_1} \|f\|_{1,\gamma} \quad (56)$$

with  $\gamma_{\mathbf{u}} = \prod_{j \in \mathbf{u}} \gamma_j = 2^{-4\sigma_{\mathbf{u}}}$ . As in the proof of Theorem 12, we see that  $\mathcal{I}_{\mathbf{u}}$  corresponds to the index set of a  $|\mathbf{u}|$ -dimensional classical sparse grid starting with  $m_1 = 2$  points and with level  $\ell - \sigma_{\mathbf{u}} - |\mathbf{u}|$ . By (54) we then obtain

$$\sum_{\mathbf{k} \in \mathcal{N}_{\mathbf{u}} \setminus \mathcal{I}_{\mathbf{u}}} 2^{-|\mathbf{k}-1|_1} \leq 2^{-\ell + \sigma_{\mathbf{u}}} A(|\mathbf{u}|, \ell - \sigma_{\mathbf{u}} - |\mathbf{u}|). \quad (57)$$

If  $\ell \geq d - \log_2(\gamma_{\{1, \dots, d\}})/4 - 1$ , it holds  $\ell - \sigma_{\mathbf{u}} \geq |\mathbf{u}| - 1$  for all  $\mathbf{u} \subseteq \mathcal{D}$  and thus

$$A(|\mathbf{u}|, \ell - \sigma_{\mathbf{u}} - |\mathbf{u}|) = 2^{|\mathbf{u}|} \binom{\ell - \sigma_{\mathbf{u}} - 1}{|\mathbf{u}| - 1} \quad (58)$$

by the definition of  $A(d, \ell)$ , see Lemma 13. Using  $\sigma_{\mathbf{u}} = -\log_2(\gamma_{\mathbf{u}})/4$  and  $\gamma_{\mathbf{u}}^{1/2} 2^{\sigma_{\mathbf{u}}} = \gamma_{\mathbf{u}}^{1/4}$ , we finally obtain the assertion by combining (55) - (58).

Note that Theorem 14 recovers Lemma 13 in the unweighted case  $\gamma_j = 1$ ,  $j = 1, \dots, d$ . We see this from

$$\begin{aligned} |If - SG_{\ell, \mathbf{1}}f| &\leq 2^{-\ell} \sum_{\mathbf{u} \subseteq \mathcal{D}} 2^{|\mathbf{u}|} \binom{\ell - 1}{|\mathbf{u}| - 1} \|f\|_{1, \mathbf{1}} = 2^{-\ell} 2d \sum_{j=1}^d \binom{d}{j} \binom{\ell - 1}{j - 1} \|f\|_{1, \mathbf{1}} \\ &= 2^{-\ell} 2d \binom{\ell + d - 1}{d - 1} \|f\|_{1, \mathbf{1}} = 2^{-\ell} A(d, \ell) \|f\|_{1, \mathbf{1}}. \end{aligned}$$

Here, the second equality follows from the Vandermonde's identity.

#### 5.4. Analysis of Error versus Cost

Using the results of Section 5.2 and Section 5.3, we now represent the error of the method  $SG_{\ell, \gamma}$  as a function of its costs  $n = n(d, \ell, \gamma)$ . We again start with the classical case.

**Lemma 15 (Error versus cost of classical sparse grids).** *For  $f \in H_1^{\text{mix}}([0, 1]^d)$  and  $\ell \geq d - 1$  it holds*

$$|If - SG_{\ell}f| = O(n^{-1} (\log_2 n)^{2(d-1)})$$

where  $n$  denotes the number of points used by the method  $SG_{\ell}$ .

PROOF. Note that  $A(d, \ell) = O(\ell^{d-1})$  in Lemma 13. By Lemma 13 we can thus estimate

$$|If - SG_{\ell}f| = O(2^{-\ell} \ell^{d-1}) = O\left(\frac{\ell^{2(d-1)}}{2^{\ell} \ell^{d-1}}\right) = O\left(\frac{(\log_2 n)^{2(d-1)}}{n}\right).$$

Here, we used  $\ell \leq \log_2(n)$  and  $n = O(2^{\ell} \ell^{d-1})$  where the latter bound can be derived with the help of Lemma 11.

We now generalise this result such that also the weighted case is covered.

**Theorem 16 (Error versus cost of weighted sparse grids).** *Let  $\ell \geq d - \log_2(\gamma_{\{1, \dots, d\}})/4 - 1$ . Then*

$$|If - SG_{\ell, \gamma}f| \leq n^{-1} 2d B(d, \ell, \gamma)^2 \|f\|_{1, \gamma} \quad (59)$$

where

$$B(d, \ell, \gamma) := \sum_{j=1}^d \gamma_{\{1, \dots, j\}}^{1/4} \binom{d}{j} \binom{\ell + \log_2(\gamma_{\{1, \dots, j\}})/4 - 1}{j - 1}.$$

PROOF. We first show that

$$n(d, \ell, \boldsymbol{\gamma}) \leq 2^\ell B(d, \ell, \boldsymbol{\gamma}). \quad (60)$$

To this end, note that  $\gamma_{\{1, \dots, j\}} = \prod_{i=1}^j \gamma_i \geq \gamma_{\mathbf{u}}$  for all  $\mathbf{u}$  with  $|\mathbf{u}| = j$  since the weights are ordered according to their size. Thus, by Lemma 12,

$$n(d, \ell, \boldsymbol{\gamma}) \leq 2^\ell \sum_{j=1}^d \sum_{|\mathbf{u}|=j} \gamma_{\mathbf{u}}^{1/4} \binom{\ell + \log_2(\gamma_{\mathbf{u}})/4 - 2}{|\mathbf{u}| - 1} \leq 2^\ell \sum_{j=1}^d \binom{d}{j} \gamma_{\{1, \dots, j\}}^{1/4} \binom{\ell + \log_2(\gamma_{\{1, \dots, j\}})/4 - 1}{j - 1}$$

where we use the fact that  $\binom{n}{d}$  is monotone increasing in  $n$ . Similarly, we derive

$$|f - \text{SG}_{\ell, \boldsymbol{\gamma}} f| \leq 2^{-\ell} 2d B(d, \ell, \boldsymbol{\gamma}) \|f\|_{1, \boldsymbol{\gamma}} \quad (61)$$

from Lemma 14. From (61) and (60) with  $n = n(d, \ell, \boldsymbol{\gamma})$  we conclude that

$$|f - \text{SG}_{\ell, \boldsymbol{\gamma}} f| \leq \frac{2d B(d, \ell, \boldsymbol{\gamma})^2}{2^\ell B(d, \ell, \boldsymbol{\gamma})} \|f\|_{1, \boldsymbol{\gamma}} \leq 2d B(d, \ell, \boldsymbol{\gamma})^2 n^{-1} \|f\|_{1, \boldsymbol{\gamma}}$$

which proves the theorem.

We now comment on Theorem 16:

- In the unweighted case  $\gamma_j = 1$ ,  $j = 1, \dots, d$ , we obtain  $B(d, \ell, \boldsymbol{\gamma}) = A(d, \ell) = O((\log_2 n)^{d-1})$ . Theorem 16 is thus a generalisation of the classical case in Lemma 15.
- Theorem 16 shows that the method  $\text{SG}_{\ell, \boldsymbol{\gamma}}$  converges with rate  $n^{-1}$  which is independent of the dimension. The error bound still depends on the value  $B(d, \ell, \boldsymbol{\gamma})$ , however. In general, we see that  $B(d, \ell, \boldsymbol{\gamma}) = O(\ell^{d-1})$ . Expressed in  $n$ , the value  $B(d, \ell, \boldsymbol{\gamma})$  thus introduces a logarithmic dependence on  $n$  and an exponential dependence on the dimension  $d$ .
- The value  $B(d, \ell, \boldsymbol{\gamma})$  is decreasing with the size of the weights  $\gamma_j$ ,  $j = 1, \dots, d$ , however. Moreover, the level

$$\ell^* := d - \log_2(\gamma_{\{1, \dots, j\}})/4 \quad (62)$$

grows with decreasing weights, see also Example 2. It describes the level where the asymptotic regime in the error bound of Theorem 16 starts and thus gives the point where the logarithmic factor  $\ell^{d-1}$  appears in the complexity.

- If the weights decay sufficiently fast such that

$$\sup_d \sum_{j=1}^d \gamma_j^{1/2} < \infty \quad (63)$$

then the general results of Wasilkowski and Woźniakowski (1999) indicate that  $B(d, \ell, \boldsymbol{\gamma})$  and hence also the method  $\text{SG}_{\ell, \boldsymbol{\gamma}}$  depends only polynomially on the dimension.

- Note that the error bound (59) also depends on the norm  $\|f\|_{1, \boldsymbol{\gamma}}$  of the integrand. This norm may grow exponentially fast for increasing  $d$  which can cause problems in higher dimensions. Note that this effect is not included in the notion of tractability in Sloan and Woźniakowski (1998); Wasilkowski and Woźniakowski (1999) since only functions with norm  $\|f\|_{1, \boldsymbol{\gamma}} \leq 1$  are addressed there.



**Example 2.** As in Wasilkowski and Woźniakowski (1999), we consider the family of weights

$$\gamma_j = j^{-\alpha}, \quad \text{with } \alpha \geq 0.$$

This example with  $\alpha = 2$  is motivated by the fact that in many application problems from finance the high nominal dimension arises from the discretization of an underlying continuous time process. The corresponding integrals can thus be written as an approximation to some infinite-dimensional integrals with respect to the Wiener measure.<sup>13</sup> In these cases, the integrands are contained in some weighted function spaces whose weights are related to the eigenvalues of the covariance operator of the Wiener measure. These eigenvalues, sorted by their magnitude, are decaying proportionally to  $j^{-2}$ , where  $j$  is the number of the eigenvalue.

For  $\gamma_j = j^{-\alpha}$ , we obtain  $\gamma_{\{1, \dots, j\}} = (j!)^{-\alpha}$  and  $\ell^* = d + \alpha/4 \log_2(d!)$ . We thus can compute the level  $\ell^*$  in (62) for different exponents  $\alpha$  and different dimensions  $d$ . The results are shown in Table 2. For instance, let  $d = 360$  and  $\alpha = 2$ . Then, one can see that the asymptotic log-factor  $\ell^{d-1}$  in the error bound from Theorem 16 does not appear as long as  $\ell < \ell^* = 1633$ . In the unweighted case, the level  $\ell^* = 361$  is significantly smaller.

Table 1: The values  $\ell^* := d + \alpha/4 \log_2(d!)$  for  $\alpha \in \{0, 1, 2, 3\}$  and different values  $d$ .

$\alpha \setminus d$	1	3	5	10	50	100	360	1024
0	1	4	6	11	51	101	361	1025
1	1	4	8	16	105	232	997	3217
2	1	5	9	22	158	363	1633	5409
3	1	6	11	27	212	495	2268	7601

If  $\alpha > 2$  holds, then the condition (63) is satisfied and we can use the general results of Wasilkowski and Woźniakowski (1999) to see that the  $\varepsilon$ -cost of the method  $\text{SG}_{\ell, \gamma}$  is independent of the dimension. In this case the number of function evaluations  $n(\varepsilon)$  to obtain an accuracy of  $\varepsilon$  can be bounded by

$$n(\varepsilon) \leq c \varepsilon^{-\max\{1, \frac{2}{\alpha-1}\}}$$

for integrands from the unit ball  $\|f\|_{\mathcal{Y}} \leq 1$ , where the constant  $c$  is independent of  $d$  and  $\varepsilon$ . It is known, see Plaskota and Wasilkowski (2001), that the  $\varepsilon$ -exponent in this bound cannot be improved using generalised sparse grid methods. It is optimal for  $\alpha \geq 3$  but far from optimal for  $\alpha \approx 1$ .

## 6. Dimension-adaptivity

In Section 3.3, we shifted the problem of the choice of the index set  $\mathcal{S}$  to the problem of determining the weights  $\gamma_{\mathbf{u}}$ . In Section 5.1, we then determined optimal index sets for sparse grid methods in weighted tensor product Sobolev spaces. In practise, however, the weights are usually unknown and can also not be computed as this would be more expensive than computing the integral.

In these cases adaptive algorithms are required which can estimate the weights a posteriori during the actual calculation of the integral. This way appropriate index sets can be constructed

<sup>13</sup>See, e.g., Linetsky (1998) and the references listed there.

automatically for a given function  $f$  without any a priori information on the dimension structure of the integrand being required.

In the following we determine the index set  $\mathcal{S}$  in a dimension-adaptive fashion. To estimate the importance of the term  $I(f_{\mathbf{u}})$  we define  $\gamma_{\mathbf{u}} := |q_{\mathbf{u}}| \approx I(f_{\mathbf{u}})$  with  $q_{\mathbf{u}}$  from (20). Furthermore, we denote by  $\mathcal{A}$  the subset of all indices  $\mathbf{u} \in \mathcal{D} \setminus \mathcal{S}$  which satisfy the admissibility condition (19) with respect to the set  $\mathcal{S}$ . With help of the weights  $\gamma_{\mathbf{u}}$ , we then heuristically build up the index set  $\mathcal{S}$  in a general and automatic way by the following Greedy-approach: We start with the initial set  $\mathcal{S} = \{\emptyset\}$  and add step by step the index  $\mathbf{u} \in \mathcal{A}$  with the largest weight  $\gamma_{\mathbf{u}}$  to  $\mathcal{S}$  until the largest weight is below some threshold  $\varepsilon$ , see Algorithm 6.1.

Note that in the first step of Algorithm 6.1 we still have the flexibility to choose the quadrature rules  $Q_{\mathbf{u}}$  for the computation of  $q_{\mathbf{u}}$ . They can be different for each  $\mathbf{u}$  and can be tailored to the dimension and smoothness of the terms  $f_{\mathbf{u}}$ , e.g., by the use of local adaptivity. Furthermore, note that in the first step of Algorithm 6.1 the values  $q_{\mathbf{u}}$  have to be computed for all  $\mathbf{u} \in \mathcal{A}$  for which  $q_{\mathbf{u}}$  has not yet been computed in previous steps. In high dimensions  $d$ , this may result in a certain overhead since not all of these values significantly contribute to the integral value.

Similarly as in Section 4, we now restrict ourselves to tensor product methods  $Q_{\mathbf{u}}$  for the approximation of the integrals  $I f_{\mathbf{u}}$  in (18). Then, the truncation and discretization of the anchored-ANOVA series can be intertwined if we use the refined index set  $\mathcal{I}$  instead of the finite set  $\mathcal{S}$ . In this case, we see, as a corollary of Theorem 8, that our dimension-adaptive approach corresponds to the *dimension-adaptive sparse grid method* as introduced in Gerstner and Griebel (2003).

This method finds the index set  $\mathcal{I}$  in a dimension-adaptive way with the help of the error indicators  $|\Delta_{\mathbf{k}}f|$ . Starting with the smallest index set  $\mathcal{I} = \{(1, \dots, 1)\}$ , those admissible indices  $\mathbf{k}$  are added step by step to  $\mathcal{I}$  which are expected to provide the largest error reduction. The resulting dimension-adaptive construction of the index set<sup>14</sup> is shown in Algorithm 6.2. There,  $\mathcal{A}$  denotes the subset of all indices  $\mathbf{k} \in \mathbb{N}^d \setminus \mathcal{I}$  which satisfy the admissibility condition (30) with respect to the set  $\mathcal{I}$ . Altogether, the algorithm allows for an adaptive detection of the important dimensions and heuristically constructs optimal index sets  $\mathcal{I}$  in the sense of Bungartz and Griebel (2004). Note that this is closely related to best  $N$ -term approximation DeVore (1998).

Remember that the  $\varepsilon$ -cost analysis in Section 5 was also based on the values  $\Delta_{\mathbf{k}}f$ , which are used here for the error estimation. We can thus expect that dimension-adaptive sparse grid methods correctly identify the optimal index sets  $\mathcal{I}_{\varepsilon, \gamma}$  from (47) provided no early determination problems occur. In this case, the results of Section 5 can also be used to show that dimension-adaptive sparse grid methods can avoid the curse of dimension in weighted function spaces whose weights decay sufficiently fast.

For comparison recall that Algorithm 6.1 can be based on two *separate* types of adaptivity. The important anchored-ANOVA terms  $I f_{\mathbf{u}}$  are found in a dimension-adaptive fashion with help of the weights  $\gamma_{\mathbf{u}}$  and are approximated by  $q_{\mathbf{u}}$  using possibly locally adaptive methods. In Algorithm 6.2 the calculation of the contributions  $\Delta_{\mathbf{k}}$  is more restrictive since the telescoping sum expansion has to hold. The algorithm is already completely determined by the choice of the univariate quadrature rule  $U_{m_{\mathbf{k}}}$ . While Algorithm 6.1 has the advantage that low regularity of low order anchored-ANOVA terms can be resolved by local adaptivity, Algorithm 6.2 has the advantage that modelling and discretization errors are *simultaneously* taken into account and can thus be balanced in an optimal way.

---

<sup>14</sup>Note that in Gerstner and Griebel (2003) more sophisticated stopping criterions are used than  $|\Delta_{\mathbf{k}}f| \leq \varepsilon$ .

---

**Algorithm 6.1:** Dimension-adaptive constructions of the index set  $\mathcal{S}$ .

---

**Initialise:** Let  $\mathcal{S} = \{\emptyset\}$ ,  $q_\emptyset := f(\mathbf{a})$  and  $s = f(\mathbf{a})$ .

**repeat**

- 1) Compute the values  $q_{\mathbf{u}}$  from (20) for all  $\mathbf{u} \in \mathcal{A}$  for which  $q_{\mathbf{u}}$  has not yet been computed and set  $s = s + q_{\mathbf{u}}$ .
- 2) Move the index  $\mathbf{u} \in \mathcal{A}$  with the largest weight  $\gamma_{\mathbf{u}} = |q_{\mathbf{u}}|$  from  $\mathcal{A}$  to  $\mathcal{S}$ .

**until**  $\gamma_{\mathbf{u}} \leq \varepsilon$  ;

Set  $A_{\mathcal{S}}f = s$ .

---



---

**Algorithm 6.2:** Dimension-adaptive sparse grid construction of the index set  $\mathcal{I}$ .

---

**Initialise:** Let  $\mathcal{I} = \{(1, \dots, 1)\}$  and  $s = \Delta_{(1, \dots, 1)}f$ .

**repeat**

- 1) Compute the values  $\Delta_{\mathbf{k}}f$  from (29) for all  $\mathbf{k} \in \mathcal{A}$  for which  $\Delta_{\mathbf{k}}f$  has not yet been computed and set  $s = s + \Delta_{\mathbf{k}}f$ .
- 2) Move the index  $\mathbf{k} \in \mathcal{A}$  with the largest weight  $|\Delta_{\mathbf{k}}f|$  from  $\mathcal{A}$  to  $\mathcal{I}$ .

**until**  $|\Delta_{\mathbf{k}}f| \leq \varepsilon$  ;

Set  $SG_{\mathcal{I}}f = s$ .

---

## 7. Numerical Results

In this section, we use multivariate integrals from finance (corresponding to Asian options, zero coupon bonds and collateralized mortgage obligations) to investigate the performance of sparse grids and other dimension-wise quadrature methods of the form (21). We first describe the setting of our numerical experiments. Then we compare the convergence behaviour of different numerical methods and relate the results to the effective dimensions of our model problems in the classical and in the anchored case.<sup>15</sup>

### 7.1. Setting

The model problems considered in our numerical experiments all lead to high-dimensional integrals of the form

$$I_d := \frac{1}{(2\pi)^{d/2} \sqrt{\det(\mathbf{C})}} \int_{\mathbb{R}^d} f(\mathbf{W}) e^{-\frac{1}{2}\mathbf{W}^T \mathbf{C}^{-1} \mathbf{W}} d\mathbf{W}, \quad (64)$$

---

<sup>15</sup>The effective dimensions in the anchored case can directly be derived from the  $q_{\mathbf{u}}$  values, see Section 2.2. In the worst case  $2^d$  many integrals are needed to compute  $\hat{\sigma}(f)$  with sufficient precision. In practice, however, often many of the terms  $|f_{\mathbf{u}}|$  are very small or zero, in particular if  $|\mathbf{u}|$  is large. Then Algorithm 6.1 can be used to compute  $\hat{\sigma}(f)$  even in high dimensions, which we demonstrate in Section 7.3 and Section 7.4 using examples with  $d = 256$  and  $d = 512$ . In the numerical experiments in Section 7 we apply Algorithm 6.1 with the threshold  $\varepsilon = 10^{-6} f(\mathbf{a})$ . For the derivation of the truncation dimensions in the classical case, we use the algorithm from Wang and Fang (2003). It requires the computation of several integrals with up to  $2d - 1$  dimensions. For their approximation we used  $2^{16}$  Sobol quasi-Monte Carlo points. To compute the superposition dimension the recursive method described in Wang and Sloan (2005) can be used. Because of cancellation problems and costs which are exponential in the superposition dimension, the computation of the superposition dimension is only feasible for moderately high dimensional function.

where  $\mathbf{W} \in \mathbb{R}^d$ ,  $\mathbf{C} \in \mathbb{R}^{d \times d}$  and  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . In our examples, the integral value  $I_d$  describes the price of a financial derivative. The vector  $\mathbf{W}$  relates to the path of an underlying stochastic process,  $\mathbf{C}$  to the covariance matrix of the process and  $f$  to the payoff function of the financial contract. Using the factorization  $\mathbf{C} = \mathbf{A}\mathbf{A}^T$  with  $\mathbf{A} \in \mathbb{R}^{d \times d}$  and the substitution  $\mathbf{W} = \mathbf{A}\mathbf{z}$  the integral is transformed into the integral

$$I_d := \int_{\mathbb{R}^d} f(\mathbf{A}\mathbf{z}) \varphi_d(\mathbf{z}) d\mathbf{z} \quad (65)$$

with standard Gaussian weight  $\varphi_d$ . It is well known that the matrix  $\mathbf{A}$  can be chosen in many different ways. Here we consider the following methods, which are usually interpreted as different generating methods of the path of a Brownian motion Imai and Tan (2006).

- In the random walk (RW) construction the path of the Brownian motion is generated sequentially in time. Here  $\mathbf{A}$  is the Cholesky matrix of  $\mathbf{C}$ .
- In the Brownian bridge (BB) construction, see Moskowitz and Caflisch (1996), the path is constructed in a hierarchical way. This has the effect that more importance is placed on the leading variables. The corresponding matrix  $\mathbf{A}$  is given in Wang (2006) explicitly.
- In the principal component (PCA) construction Acworth et al. (1998) the matrix  $\mathbf{A}$  results from the eigenvalue decomposition of  $\mathbf{C}$ . This construction maximizes the concentration of the total variance in the leading dimensions. The corresponding matrix  $\mathbf{A}$  can be found in Glasserman (2003).
- The linear transformation (LT) construction, see Imai and Tan (2006), aims to identify the matrix  $\mathbf{A}$  which minimizes the effective truncation dimension of the integrand in the classical ANOVA sense. While BB and PCA do not respect the particular structure of the function  $f$ , the LT-construction in addition takes the gradient of  $f$  at a certain anchor point into account. The corresponding matrix  $\mathbf{A}$  is constructed column by column from a Gram-Schmidt like local minimization, which involves  $f$ . For details see Imai and Tan (2006), where this approach is referred to by LT-II.

Below we will study the efficiency of different sparse grid methods to compute the integral (65). To this end, remember that the univariate quadrature rule  $U_{m_k}$  in the sparse grid construction of Section 4 was left open. If a quadrature rule is used which is defined on  $\mathbb{R}$ , then the integral (65) can be treated with a sparse grid method directly on  $\mathbb{R}^d$ . In our tests, we will use the dimension-adaptive sparse grid method based on the Gauss-Hermite rule and refer to this specific method as SGH.

To apply a univariate quadrature rule on  $[0, 1]$ , it is necessary to transform the integral (65) over  $\mathbb{R}^d$  into an integral over the unit cube  $[0, 1]^d$ . To this end, we use the standard component-wise substitution  $\mathbf{z} = \Phi^{-1}(\mathbf{x})$ , where  $\Phi$  denotes the cumulative normal distribution function.<sup>16</sup>

---

<sup>16</sup>The resulting transformed integrand is unbounded on the boundary of the unit cube, which is undesirable from a numerical as well as from a theoretical point of view. Nevertheless, in combination with (quasi-) Monte Carlo methods this transformation turns out to be very effective because it cancels the Gaussian weight. But this singular transformation deteriorates the efficiency of quadrature methods which take advantage of higher smoothness, such as sparse grids. Here, it is often better to avoid the transformation and the corresponding loss of regularity and to address the integral directly on  $\mathbb{R}^d$ .

This yields

$$I_d := \int_{[0,1]^d} g(\mathbf{x}) d\mathbf{x} \quad (66)$$

with the integrand  $g(\mathbf{x}) := f(\mathbf{A}\Phi^{-1}(\mathbf{x}))$ .

This way we can apply the dimension-adaptive sparse grid method based on the Gauss-Patterson rule as univariate rule  $U_{m_k}$ , which we refer to as SGP. This method was first presented in Gerstner and Griebel (2003). It is a special case of the method (21), compare Theorem 8.

In our tests, we will also consider different variants of the dimension-wise quadrature method (21), which are not of sparse grid form, but which can resolve low regularity in the low-order anchored ANOVA terms using local adaptivity. To this end, recall that we still have to specify the quadrature rules  $Q_{\mathbf{u}}$  and the index set  $\mathcal{S}$  to finalize the construction of the method (21) from Section 3. Here we choose the  $Q_{\mathbf{u}}$ 's as follows: If  $|\mathbf{u}| < 4$ , we use the locally-adaptive product method CUHRE, see Bernsten et al. (1991); Hahn (2005), to address low regularity of the terms  $P_{\mathbf{u}}f$ . If  $|\mathbf{u}| \geq 4$ , we use a randomised quasi-Monte Carlo method based on Sobol point sets to lift the dependence on the dimension. We always use the anchor point  $(1/2, \dots, 1/2)$ . For the construction of the index set  $\mathcal{S}$  (i.e. for finding the most important terms of the anchored-ANOVA decomposition) we employ the a priori constructions from Section 3.3 and the a posteriori construction from Section 6. This defines the following three new quadrature methods:

- mixed CUHRE/QMC method with order-dependent weights (COW),
- mixed CUHRE/QMC method with product weights (CPW),
- mixed CUHRE/QMC method with dimension-adaptivity (CAD).

We refer to COW, CPW and CAD as mixed CUHRE/QMC methods. To our knowledge these methods are the first numerical quadrature methods which can profit from low effective dimension (by the selection of appropriate function space weights or by dimension-adaptivity) and which can at the same time resolve low regularity to some extent by local adaptivity.

For comparison we also consider Monte Carlo integration (MC) and quasi-Monte Carlo integration based on Sobol point sets (QMC). These two methods are most commonly used for the computation of high-dimensional integrals. In preliminary numerical experiments, Sobol point sets turned out to be the most efficient representative of several quasi-Monte Carlo variants.<sup>17</sup>

## 7.2. Asian Options

We first consider the commonly used test problem to determine the fair value of an Asian option with geometric average. The arising integrands are explicitly given in Wang and Fang (2003); Wang and Sloan (2005). Here, we distinguish the two cases  $K = 0$  and  $K = 100$  (where  $K$  denotes the strike price of the option) which we refer to as Asian0 and Asian100, respectively. While we obtain a smooth integrand in the first case, the integrand has discontinuous first derivatives in the latter case. We use the same parameters as in Wang and Fang (2003); Wang and Sloan (2005) and consider the case  $d = 16$ .

---

<sup>17</sup>We compared Halton, Faure, Sobol low discrepancy point sets and three different lattice rules based on the CBC construction from Sloan et al. (2002). The lattice rules yield in many cases equal or even more precise results as Sobol points if good function space weights are used for their construction. But the selection of good weights is a priori not always clear. It would be interesting to see if our anchored-ANOVA weights  $\gamma_i$  in (25) can successfully be used in the CBC construction.

Table 2: Truncation dimensions of the Asian option pricing problems.<sup>18</sup>

$1-\alpha$	(a) Asian0 ( $d = 16$ )				(b) Asian100 ( $d = 16$ )			
	anchored-ANOVA		classical ANOVA		anchored-ANOVA		classical ANOVA	
	RW	BB	PCA	LT	RW	BB	PCA	LT
1e-1	9	2	1	1	9	2	1	1
1e-2	13	6	2	1	13	6	2	1
1e-3	15	14	3	1	15	14	3	1
1e-4	16	16	6	1	16	16	6	1

Table 3: Superposition dimensions in the anchored case.<sup>18</sup>

$1-\alpha$	(a) Asian0 ( $d = 16$ )				(b) Asian100 ( $d = 16$ )			
	RW	BB	PCA	LT	RW	BB	PCA	LT
1e-1	1	1	1	1	7	2	1	1
1e-2	1	1	1	1	8	3	1	1
1e-3	2	2	1	1	8	4	2	1
1e-4	2	2	2	1	8	4	3	1

We first show in Table 2 the truncation dimensions of the arising integral in the classical and in the anchored case for different proportions  $\alpha \in [0.9, 0.9999]$ . One can see that the truncation dimensions in the classical case almost coincide with the truncation dimensions in the anchored case.<sup>18</sup> For instance, for the case  $K = 0$  with  $\alpha = 0.999$  we obtain  $d_t = 15, 14, 3, 1$  using RW, BB, PCA and LT, respectively, for the anchored case as well as for the classical one. The LT-construction achieves the optimal result  $d_t = 1$  in both cases. One can show that this holds even for the extreme case  $\alpha = 1$ , i.e. this problem can be reduced by LT to one with only one nominal dimension.

Further numerical calculations, see Table 3(a), show that the Asian0 problem is of very low superposition dimension  $d_s \leq 2$  independent of the employed path construction. For the Asian100 problem,  $d_s$  increases if we switch from LT to PCA, to BB and RW as can be seen in Table 3(b).<sup>18</sup>

Next we compute the integral values using different numerical approaches (MC, QMC, COW, CPW, CAD, SGP and SGH) and different path constructions (RW, BB, PCA, LT) as introduced in Section 7.1. We display the convergence behaviour of these methods in Figure 3 and Figure 4. There, we show the number of function evaluations which is required by each of the different numerical methods to obtain a fixed accuracy.

One can see that the convergence rate of the MC method is always about 0.5 as predicted by the law of large numbers. The rate is not affected by the path construction since the total variance stays unchanged. The convergence rate of the QMC method increases if BB, PCA or LT is used since these path constructions concentrate the total variance in the first few dimensions. This way QMC outperforms MC and achieves higher convergence rates of almost one, smaller relative errors and a less oscillatory convergence behaviour.

<sup>18</sup>Note that our numerical computations for the anchored case with  $K = 100$  and RW or BB might be inaccurate. For these particular problems accurate results are difficult to obtain since the truncation dimension is high (hence many terms have to be integrated) and the integrals are not smooth (hence their computation is expensive).

From Figure 3 and Figure 4 we also observe that the impact of the path construction is considerably bigger in case of the dimension-wise quadrature methods COW, CPW, CAD, SGP and SGH. This is explained by the fact that these methods are tailored to the effective dimension of the problem by the choice of the respective function space weights from Section 3.3, or by the dimension-adaptive grid refinement. The convergence of these methods is thus significantly accelerated by path constructions which reduce the effective dimension of the associated integrand. For instance, in the case  $K = 0$ , compare Figure 3, one can see that the performance of dimension-wise quadrature methods significantly improves if we switch from RW to BB, to PCA and then to LT. While COW, CPW, CAD and SGP provide results which are similar to or even worse than (Q)MC in case of RW, they outperform (Q)MC slightly, clearly and drastically in case of BB, PCA and LT, respectively. Note here that two different regimes have to be distinguished to describe the convergence behaviour of these methods, compare, e.g., Figure 3(d). In the preasymptotic regime, the methods COW, CPW, CAD, SGP and SGH first search for the important dimensions and interactions, whereas, in the asymptotic regime, the important dimensions are identified and the grid is then refined only in these directions.

Since the LT-construction reduces the problem to only one dimension, its combination with dimension-adaptive methods is particularly efficient. We see from Figure 3(d) that the sparse grid and the dimension-adaptive methods correctly identify the important dimension and then only refine in this respect, which leads to an extremely rapid convergence in the asymptotic regime.

A comparison of the convergence rates of the COW, CPW and CAD method shows that the a priori constructions (with order-dependent weights or product weights) and the dimension-adaptive construction of the index set  $\mathcal{S}$  lead to very similar results. The results of COW and CPW even coincide in most cases.

For the Asian0 problem, SGH is by far the most efficient method independent of the employed path construction. It exploits the low effective dimension by its dimension-adaptive grid refinement and can profit from the smoothness of the integrand much better than all other approaches since it avoids the transformation to the unit cube. This way we obtain relative errors smaller than  $10^{-12}$  with only about  $10^5$ ,  $10^4$ ,  $10^3$  and  $10^2$  function evaluations in case of RW, BB, PCA and LT, respectively, which is 7 – 10 orders of magnitude more precise than the results of QMC.

Comparing the two cases  $K = 0$  and  $K = 100$ , we furthermore see that the convergence rates of the QMC method are only slightly affected by the kink in the integrand, whereas the SG methods clearly suffer from the low degree of regularity. This drawback is to some extent overcome by the COW, the CPW and the CAD method, which are in combination with PCA or LT the most efficient approaches for the Asian100 problem. These methods profit from the low effective dimension and can in addition deal with the low regularity of the integrand by local adaptivity in the low order anchored-ANOVA terms due to the CUHRE approach. With LT, these methods obtain relative errors smaller than  $10^{-8}$  with only about 1,000 function evaluations, see Figure 4(d), which is about 100,000 times more accurate than the results of QMC.

### 7.3. Zero Coupon Bonds

We now consider the problem to price zero coupon bonds using the Vasicek model. This model problem is taken from Ninomiya and Tezuka (1996). It is also studied in Wang (2006); Wang and Sloan (2005). We use the same parameters as in Ninomiya and Tezuka (1996) and consider the case  $d = 512$ .

First we again study the effective dimensions of this problem. In Table 4(a) one can see that the effective dimensions in the classical case almost coincide with the effective dimensions in the anchored case. For instance, for  $\alpha = 0.99$  we obtain  $d_t = 420, 7, 1, 1$  using RW, BB, PCA and LT, respectively, for the anchored case and  $d_t = 419, 7, 1, 1$  for the classical one. We conjecture that a more precise numerical computation would yield even exact equal results and we believe that this equality holds for a wider class of functions. Note however that this does not hold in general. We will give a counterexample in Section 7.4.

Observe that the path construction has a significant impact on the truncation dimensions. For  $\alpha$  close to one the dimensions  $d_t$  are almost as large as the nominal dimension  $d = 512$  if we employ the RW approach. The dimensions  $d_t$  are significantly smaller if BB, PCA or LT is used instead. The LT-construction even obtains the optimal result  $d_t = 1$  for this problem. While it is not surprising that such an optimal transformation exists,<sup>19</sup> it is nevertheless interesting that it is correctly identified by the LT-construction, which takes only the gradient of the integrand at a certain anchor point into account. Further computations show, see Table 5(a), that the integral is of very low superposition dimension  $d_s \leq 2$  in the anchored case and that  $d_s$  is almost independent of the path construction.

We display the convergence behaviour of the different numerical methods using the different path generating methods for the Vasicek problem in Figure 5. We observe similar results as for the Asian0 problem despite the fact that the nominal dimension  $d = 512$  is significantly higher here. With RW and BB, the methods COW, CPW and CAD provide similar results as QMC. They partly seem to stop to converge, which happens if important contributions to the integral value are contained in anchored-ANOVA terms of higher order that are not yet identified. For instance, in Figure 5(b) one can see that  $10^6$  function evaluations are not sufficient to find those contributions which would reduce the relative error to be below  $10^{-6}$ . The dimension-wise integration methods again clearly profit from path constructions which lead to low effective dimension. In combination with LT, see Figure 5(d), all these methods again correctly identify the important dimension of the problem. This way the dependence on the dimension is completely avoided in the asymptotic regime. There, the convergence of the methods is as fast as it is known for univariate problems despite the high nominal dimension  $d = 512$ . SGH is again by far the most efficient method. It outperforms (Q)MC by several orders of magnitude independent of the employed path construction. By exploiting the low effective dimension and smoothness of the integrand, SGH achieves in combination with PCA or LT almost machine accuracy with only about 1,000 function evaluations.

To better understand the fast performance of the SGH method we look at the index set  $\mathcal{I}$  that is build up by Algorithm 6.2 in a dimension-adaptive way. For the visualization we consider the two-dimensional slices through the index set  $\mathcal{I}$  that correspond to the set of 512-dimensional indices  $\mathbf{k}$  of the form  $(k_1, k_2, 1, \dots, 1)$  and  $(k_1, 1, \dots, 1, k_2)$  for  $k_1, k_2 \geq 1$ , respectively. The resulting index sets are shown in Figure 2 for the example of the Brownian bridge construction. In Figure 2 all indices  $\mathbf{k}$  are marked with a dot that are included in the index set  $\mathcal{I}$  which is build up by Algorithm 6.2 if the threshold  $\varepsilon = 10^{-3}$  is used, compare also Figure 1. For instance, one can see that the index  $(3, 1, \dots, 1)$  is included in  $\mathcal{I}$  but not the index  $(1, \dots, 1, 3)$ . Moreover, the values  $|\Delta_{\mathbf{k}}f|$  are shown color-coded from  $10^0$  (dark red) to  $10^{-16}$  (dark blue) for  $1 \leq k_1, k_2 \leq 7$ . One can see that these values are decaying rapidly for increasing  $k_1, k_2$ . They are already below  $10^{-10}$  if  $k_1 > 4$  or  $k_2 > 4$ . With respect to the dimension 512 we see in Figure 2(b) that  $|\Delta_{\mathbf{k}}f| < 10^{-10}$

---

<sup>19</sup>This way the closed-form pricing formula is derived.



already if  $k_2 > 1$ , which reflects the low truncation dimension of the Vasicek problem with the Brownian bridge construction. The results shown in Figure 2 indicate for this particular example that already a rather small index set  $\mathcal{I}$  is sufficient to capture all indices  $\mathbf{k}$  that correspond to significant contributions to the integral value. This explains why high precisions can be achieved with only little costs by the dimension-adaptive SGH method.<sup>20</sup>

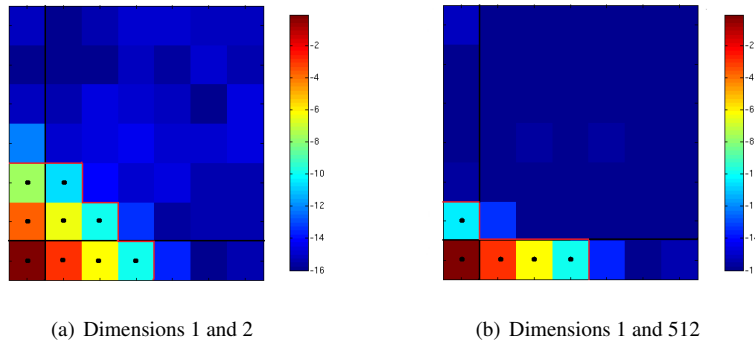


Figure 2: Two-dimensional slices through the index set  $\mathcal{I}$  corresponding to the threshold  $\varepsilon = 10^{-3}$  for the Vasicek problem with the Brownian bridge construction ( $d = 512$ ).

#### 7.4. Collateralized Mortgage Obligations

We finally deal with the problem to price a collateralized mortgage obligation (CMO). This model problem is described in detail in Cafilisch et al. (1997); Paskov and Traub (1995). There and in several further references, it is used to study the performance of quasi-Monte Carlo methods. It is also considered in Gerstner and Griebel (1998, 2003) to demonstrate the efficiency of SG methods. We use the same parameters as in Gerstner and Griebel (1998) and consider the case  $d = 256$ .

We first computed the superposition dimension in the anchored case for the CMO problem. We obtained again  $d_s \leq 2$  for all  $\alpha \in [0.9, 0.9999]$  and all path constructions, see Table 5(b). The truncation dimensions  $d_t$  of this problem are shown in Table 4(b). It is striking that the path construction has only a small impact on the truncation dimension in the anchored case, i.e., the advantage of BB, PCA and LT compared to RW is not so clear for the CMO problem. For  $\alpha = 0.9$  we have  $d_t = 123$  in case of RW and LT. This truncation dimension is reduced to  $d_t = 18$  and  $d_t = 13$  if BB and PCA is used, respectively. For higher accuracy requirements, however, i.e. for  $\alpha \geq 0.99$ , significantly less or even no reduction at all is achieved with these constructions. Note that for the CMO problem the effective dimensions in the classical case clearly differ from the truncation dimensions in the anchored case. There BB, PCA and LT lead to significant dimension reductions. LT even reduces the problem to the truncation dimension one in the classical case.

We next study the convergence behaviour of the different numerical methods for this problem. The respective numerical results are illustrated in Figure 6. One can see that the QMC method converges faster, less oscillatory and superior to MC if we switch from RW to BB, PCA or LT.

<sup>20</sup>Note that the results shown in Figure 2 also indicate that Algorithm 6.2 indeed correctly identifies the most important contributions to the integral value.

Table 4: Truncation dimensions of the Vasicek and CMO problem.

(a) Vasicek problem ( $d = 512$ )					(b) CMO problem ( $d = 256$ )												
$1-\alpha$	anchored-ANOVA				classical ANOVA				$1-\alpha$	anchored-ANOVA				classical ANOVA			
	RW	BB	PCA	LT	RW	BB	PCA	LT		RW	BB	PCA	LT	RW	BB	PCA	LT
1e-1	302	2	1	1	305	2	1	1	123	18	13	123	60	5	2	1	
1e-2	420	7	1	1	419	7	1	1	191	134	108	191	110	10	5	1	
1e-3	471	16	2	1	471	16	2	1	225	192	235	225	158	36	11	1	
1e-4	494	59	5	1	494	52	5	1	242	229	254	242	181	80	23	1	

Table 5: Superposition dimensions in the anchored case.

(a) Vasicek problem ( $d = 512$ )					(b) CMO problem ( $d = 256$ )				
$1-\alpha$	RW	BB	PCA	LT	$1-\alpha$	RW	BB	PCA	LT
1e-1	1	1	1	1	1e-1	1	1	1	1
1e-2	1	1	1	1	1e-2	2	2	1	2
1e-3	1	1	1	1	1e-3	2	2	2	2
1e-4	2	2	1	1	1e-4	2	2	2	2

SGP performs similar to QMC in case of BB and PCA and slightly worse in case of RW and LT. The mixed CUHRE/QMC methods COW, CPW and CAD attain the best results in combination with PCA. In this case they outperform (Q)MC and SGP. One can finally see that SGH combined with BB or PCA is by far the most efficient method for the CMO problem. It achieves the highest convergence rate and the most precise results. With  $10^4$  function evaluations SGH obtains a relative error which is about 100 times smaller than the relative error of the QMC method.

We next discuss the relation of the convergence behaviour of the numerical methods to the effective dimension of the CMO problem. We already showed that the path construction affects both the performance of the numerical methods (except for MC) and the truncation dimension of the integral. Since the truncation dimension in the classical case differs from the truncation dimension in the anchored case for this problem, it is interesting to see which of these two notions better predicts the convergence behaviour of the numerical methods. Remember that LT does not lead to an improved convergence of the dimension-wise quadrature methods compared to RW. This observation cannot be explained by the effective dimension in the classical case since LT obtains the optimal result  $d_t = 1$  for the CMO problem. The observation is, however, in clear correspondence with the fact that LT provides no reduction of the truncation dimension in the *anchored* case. This indicates that the performance of the dimension-wise quadrature methods depends on the truncation dimension in the anchored case, but not on the truncation dimension in the classical case. Note that the QMC method converges faster and less oscillatory with LT than with RW. This indicates that the convergence behaviour of the QMC method is rather related to the effective dimension  $d_t$  in the classical case than to the anchored one.

The different effective dimensions in the anchored and in the classical case are related to the fact that in the anchored-ANOVA decomposition the contributions  $If_{\mathbf{u}}$  are of different sign for different  $\mathbf{u}$  in the CMO problem. Summing the contributions thus leads to cancellation effects which are not seen in the anchored case since the absolute values  $|If_{\mathbf{u}}|$  are taken into account there. Nevertheless, also the error indicator of dimension-adaptive sparse grid methods and of other dimension-wise integration methods is based on the absolute values  $|\Delta_{\mathbf{k}}|$  and  $|q_{\mathbf{u}}|$ , respec-

tively, see Algorithm 6.2 and Algorithm 6.1. These methods can thus also not profit from such cancellation effects and their convergence behaviour therefore rather depends on the effective dimensions in the anchored case than on the effective dimension in the classical case.

Note finally that the truncation dimension  $d_t$  in the anchored case explains the impact of the path construction but not the high performance of the SGH methods since  $d_t$  is high for this problem. The fast convergence is explained by the low superposition dimension  $d_s \leq 2$  and by the smoothness of the integrand.

## 8. Concluding Remarks

In this article, we introduced a new general class of methods for the computation of high-dimensional integrals, which we referred to as dimension-wise quadrature methods. Our starting point was the anchored-ANOVA decomposition, which has (compared to the classical ANOVA decomposition) the advantage that only a finite number of function values is required for its computation. Our new methods then resulted from truncation of the anchored-ANOVA decomposition and from integration of the remaining terms using appropriate low-dimensional quadrature rules.

We discussed a priori (using function space weights) and a posteriori (using dimension-adaptivity) approaches for the truncation and derived bounds for the resulting modelling error. To this end, we introduced a the new notion of effective dimensions in the anchored case. We showed that the presented analysis also applies to sparse grid methods as they can be regarded as special cases of our general approach. We explained that sparse grid methods intertwine the truncation of the anchored-ANOVA series and the subsequent discretization which allows to balance modelling and discretization error in an optimal way.

We also presented dimension-wise quadrature methods which are not of sparse grid form but use the CUHRE method for the integration of the low-order anchored ANOVA terms and quasi-Monte Carlo for the higher order ones. This way, we obtained mixed CUHRE/QMC methods which are to our knowledge the first numerical quadrature methods which can profit from low effective dimension by dimension-adaptivity and can at the same time deal with low regularity by local adaptivity. A correct balancing of modeling and discretization errors is then more difficult. Numerical experiments for the Asian option as a test function from finance with discontinuous first derivatives demonstrate, however, that this disadvantage is more than compensated by the benefits of the local adaptivity. The numerical results showed the superiority of our new method to (quasi-) Monte Carlo methods and sparse grid methods for this model problem.

We considered further application problems from finance which lead to the integration of smooth functions with up to 512 dimensions. For these model problems the dimension-adaptive sparse grid method based on the Gauss-Hermite rule turned out to be most efficient. This method profits from the low effective dimension of the integral by its dimension-adaptive grid refinement and optimally exploits the smoothness of the integrand since it avoids the singular transformation to the unit cube. This way (quasi-) Monte Carlo methods were outperformed by several orders of magnitude even in hundreds of dimensions. We finally analysed the effective dimensions of our application problems in the classical and in the anchored case and showed that the results mainly explain the behaviour of our numerical quadrature methods.

Of course, our results could be extended into various directions. For instance, it would be interesting to identify the function classes for which the effective dimension in the anchored case coincides with the effective dimension in the classical case. We indicated that the pricing

problems of Asian options or zero coupon bonds belong to such a function class, but not the CMO problem. Note further that we always used the center of the integration domain as anchor point. Of course also other choices are possible. It would be of interest to analyse the impact of the anchor point in more detail and to understand how the anchor may be chosen best, see also Sobol (2003) or Wang (2008). Other possible areas for future research include improvements of our dimension-wise quadrature methods that are not of sparse grid form by, e.g., a more sophisticated balancing of modeling and discretization errors. For applications from finance, our mixed CUHRE/QMC methods can be further improved if, instead of the CUHRE method, a different locally adaptive method is employed which treats the integrals directly on  $\mathbb{R}^d$  such that the singular transformation to the unit cube can be avoided. The local error estimator of such a method could for example be based on Genz-Keister points, see Genz and Keister (1996).

Note furthermore that our dimension-wise approach can not only be used for integration but also for the representation and approximation of high-dimensional functions in the sense of Rabitz and Alis (1999). In this context, it would be interesting to study the efficient computation of further quantities, which can not be formulated as expected values or integrals, but which are also important for financial institutions, such as quantiles. Quasi-Monte Carlo methods can compute also quantiles very efficiently as shown by Papageorgiou and Paskov (1995) using the example of Value at Risk calculations. To our knowledge, it is not yet known if similar or even better results can be obtained with dimension-wise quadrature methods based on sparse grids. We finally remark that most of our methods and results are not restricted to applications from finance, but can also be used in other application areas such as chemistry or physics.

**Acknowledgement** The authors thank Ian Sloan from the University of New South Wales and two unknown referees for valuable comments and hints.

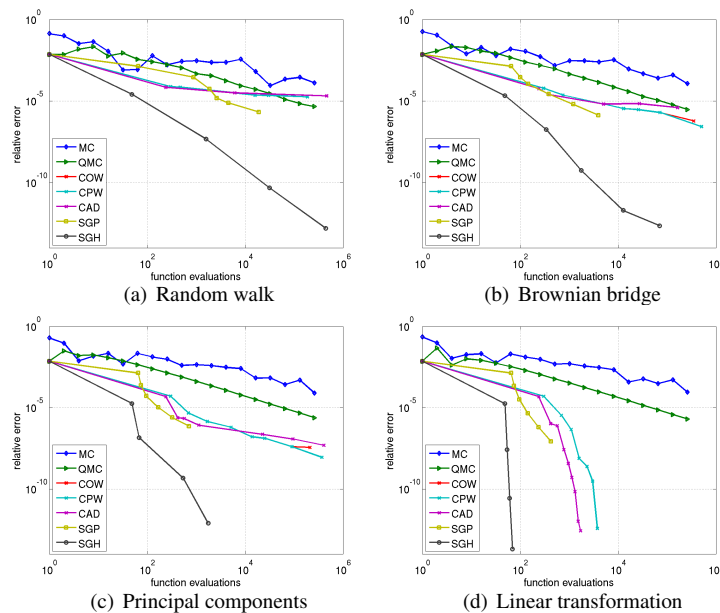


Figure 3: Convergence behaviour of the different methods for the Asian0 problem ( $d = 16$ ).

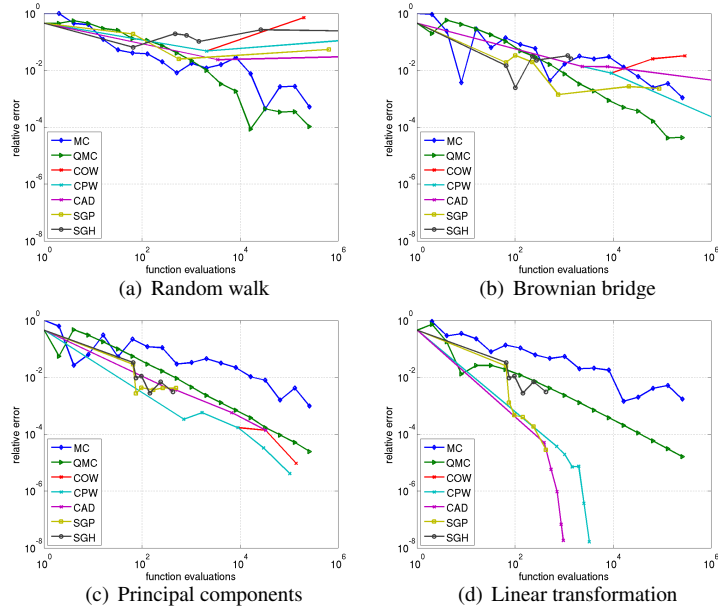


Figure 4: Convergence behaviour of the different methods for the Asian100 problem ( $d = 16$ ).

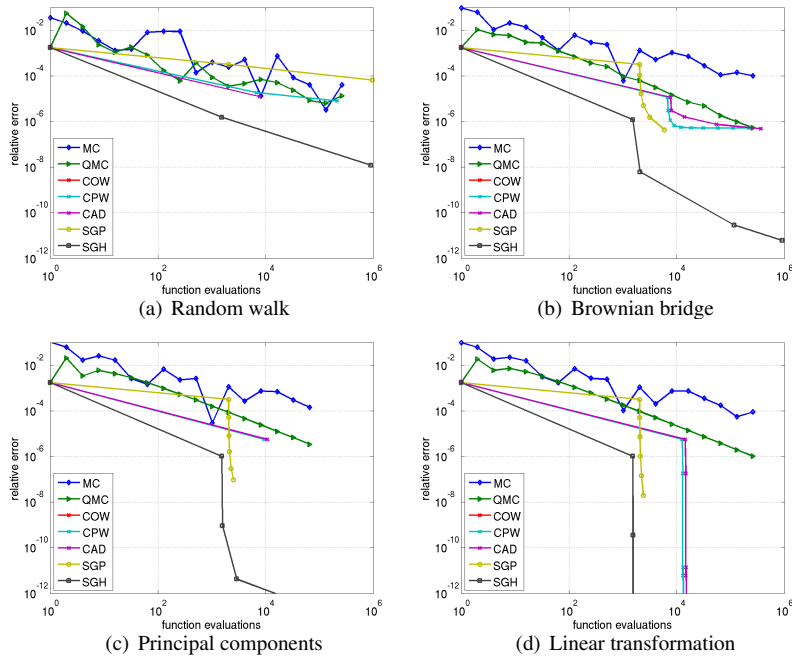


Figure 5: Convergence behaviour of the different methods for the Vasicek problem ( $d = 512$ ).

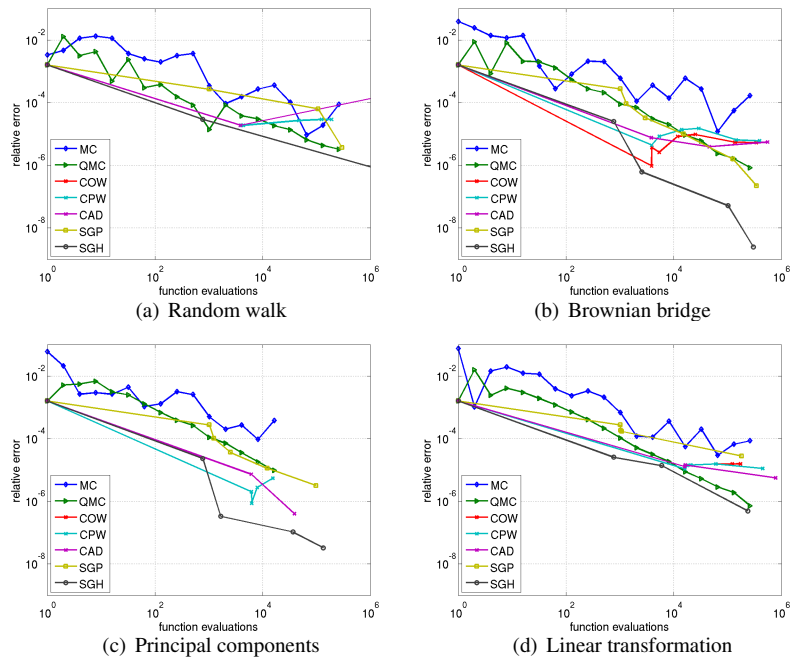


Figure 6: Convergence behaviour of the different methods for the CMO problem ( $d = 256$ ).

## References

- M. Griebel, Sparse grids and related approximation schemes for higher dimensional problems, in: L. Pardo, A. Pinkus, E. Suli, M. Todd (Eds.), *Foundations of Computational Mathematics (FoCM05)*, Santander, Cambridge University Press, 106–161, 2006.
- P. Davis, P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, 1984.
- J. Traub, H. Woźniakowski, G. Wasilkowski, *Information-Based Complexity*, Academic Press, New York, 1988.
- P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, 2003.
- H.-J. Bungartz, M. Griebel, Sparse grids, *Acta Numerica* 13 (2004) 1–123.
- T. Gerstner, M. Griebel, Numerical integration using sparse grids, *Numerical Algorithms* 18 (1998) 209–232.
- T. Gerstner, M. Griebel, Dimension-adaptive tensor-product quadrature, *Computing* 71 (1) (2003) 65–87.
- E. Novak, K. Ritter, High dimensional integration of smooth functions over cubes, *Numer. Math.* 75 (1996) 79–97.
- S. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk SSSR* 4 (1963) 240–243.
- R. Caflisch, W. Morokoff, A. Owen, Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension, *J. Comp. Finance* 1 (1) (1997) 27–46.
- S. Paskov, J. Traub, Faster valuation of financial derivatives, *J. Portfolio Management* 22 (1995) 113–120.
- X. Wang, I. Sloan, Why are high-dimensional finance problems often of low effective dimension?, *SIAM J. Sci. Comput.* 27 (1) (2005) 159 – 183.
- B. Moskowitz, R. Caflisch, Smoothness and dimension reduction in quasi-Monte Carlo methods, *J. Math. Comp. Modeling* 23 (1996) 37–54.
- P. Acworth, M. Broadie, P. Glasserman, A comparison of some Monte Carlo and quasi-Monte Carlo methods for option pricing, in: P. Hellekalek, H. Niederreiter (Eds.), *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Springer, 1–18, 1998.
- J. Imai, K. S. Tan, A general dimension reduction technique for derivative pricing, *Journal of Computational Finance* 10 (2) (2006) 129–155.
- I. Sloan, H. Woźniakowski, When are Quasi-Monte Carlo algorithms efficient for high-dimensional integrals?, *J. Complexity* 14 (1998) 1–33.
- M. Griebel, F. Kuo, I. Sloan, The smoothing effect of the ANOVA decomposition, technical report, University of New South Wales, 2008.
- R. Liu, A. B. Owen, Estimating mean dimensionality of analysis of variance decompositions, *Journal of the American Statistical Association* 101 (2006) 712–721.
- M. Hegland, Adaptive sparse grids, *ANZIAM J.* 44 (2003) C335–C353.
- L. Plaskota, The exponent of discrepancy of sparse grids is at least 2.1933, *Advances in Computational Mathematics* 12 (2000) 3–24.
- G. Wasilkowski, H. Woźniakowski, Weighted tensor product algorithms for linear multivariate problems, *J. Complexity* 15 (1999) 402–447.
- J. Bernsten, T. Espelid, A. Genz, Algorithm 698: DCUHRE – An adaptive multidimensional integration routine for a vector of integrals, *ACM Transactions on Mathematical Software* 17 (1991) 452–456.
- F. Kuo, I. Sloan, G. Wasilkowski, H. Woźniakowski, On decompositions of multivariate functions, technical report, University of New South Wales, 2008.
- H. Rabitz, O. Alis, General foundations of high-dimensional model representations, *Journal of Mathematical Chemistry* 25 (1999) 197–233.
- C. Lemieux, A. Owen, Quasi-regression and the relative importance of the ANOVA components of a function, in: K. T. Fang, F. J. Hickernell, H. Niederreiter (Eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2000*, Springer, 331–344, 2002.
- I. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulation* 55 (2001) 271–280.
- I. Sobol, Sensitivity estimates for nonlinear mathematical models, *Mathematical Modeling and Computational Experiment* 1 (1993) 407–414.
- I. Sobol, Theorems and examples on high dimensional model representation, *Reliability Engineering and System Safety* 79 (2003) 187–193.
- X. Wang, On the approximation error in high dimensional model representation, in: S. Mason et al. (Ed.), *Proceeding of the 2008 Winter Simulation Conference*, 453–462, 2008.
- I. H. Sloan, F. Y. Kuo, S. Joe, Constructing randomly shifted lattice rules in weighted Sobolev spaces, *SIAM J. Numer. Anal.* 40 (2002) 1650 – 1665.
- E. Novak, K. Ritter, A. Steinbauer, A multiscale method for the evaluation of Wiener integrals, in: *Approximation Theory IX, Volume 2: Computational Aspects*, C. K. Chui and L. L. Schumaker (eds.), Vanderbilt Univ. Press, 251–258, 1998.

- G. Wasilkowski, H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, *J. Complexity* 11 (1995) 1–56.
- V. Linetsky, The path integral approach to financial modeling and options pricing, *Computational Economics* 11 (1998) 129–163.
- L. Plaskota, G. W. Wasilkowski, The exact exponent of sparse grid quadratures in the weighted case, *J. Complexity* 17 (2001) 840–848.
- R. DeVore, Nonlinear approximation, *Acta Numerica* 7 (1998) 51–150.
- X. Wang, K.-T. Fang, The effective dimension and quasi-Monte Carlo integration, *J. Complexity* 19 (2) (2003) 101 – 124.
- X. Wang, On the effects of dimension reduction techniques on some high-dimensional problems in finance, *Operations Research* 54 (6) (2006) 1063–1078.
- T. Hahn, Cuba - a library for multidimensional numerical integration, *Computer Physics Communications* 168 (2005) 78–95.
- S. Ninomiya, S. Tezuka, Toward real-time pricing of complex financial derivatives, *Applied Mathematical Finance* 3 (1996) 1–20.
- A. Genz, B. Keister, Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight, *J. Comp. Appl. Math.* 71 (1996) 299–309.
- A. Papageorgiou, S. Paskov, Deterministic simulation for risk management, *J. Portfolio Management* 25 (5) (1995) 122–127.