1

Sparse grids and related approximation schemes for higher dimensional problems

Michael Griebel
Institut für Numerische Simulation
Universität Bonn
53113 Bonn, Germany
e-mail: griebel@iam.uni-bonn.de

Abstract

The efficient numerical treatment of high-dimensional problems is hampered by the curse of dimensionality. We review approximation techniques which overcome this problem to some extent. Here, we focus on methods stemming from Kolmogorov's theorem, the ANOVA decomposition and the sparse grid approach and discuss their prerequisites and properties. Moreover, we present energy-norm based sparse grids and demonstrate that, for functions with bounded mixed derivatives on the unit hypercube, the associated approximation rate in terms of the involved degrees of freedom shows no dependence on the dimension at all, neither in the approximation order nor in the order constant.

1.1 Introduction

The discretization of PDEs by conventional methods is limited to problems with up to three or four dimensions due to storage requirements and computational complexity. The reason is the so-called curse of dimensionality, a term coined in (Bellmann 1961). Here, the cost to compute and represent an approximation with a prescribed accuracy ε depends exponentially on the dimensionality d of the problem considered. We encounter complexities of the order $O(\varepsilon^{-d/r})$ with $r > 0$ depending on the respective approach, the smoothness of the function under consideration, the polynomial degree of the ansatz functions and the details of the implementation. If we consider simple uniform grids with piecewise d-polynomial functions over a bounded domain in a finite element or finite difference approach, this complexity estimate translates to $O(N^d)$ grid points or degrees of freedom for which approximation accuracies

of the order $O(N^{-r})$ are achieved.† Thus, the computational cost and storage requirements grow exponentially with the dimensionality of the problem, which is the reason of the dimensional restrictions mentioned above, even on the most powerful machines presently available.

The curse of dimensionality can be circumvented to some extent by restricting the class of functions under consideration. If we make a stronger assumption on the smoothness of the solution such that the order of accuracy depends on d as $O(N^{-c\cdot d})$ with $r = c \cdot d$, we directly see that the cost complexity is independent of d and that it is of the order $O(\varepsilon^{-d/(c\cdot d)}) = O(\varepsilon^{-1/c})$, for some c independent of d. This way, the curse of dimensionality can be broken easily.† In any case, such a smoothness assumption is somewhat unrealistic.

Nevertheless for practical applications in high(er) dimensions often a certain smoothness assumption on the function is implicitly present (e.g., in the data of the problem) which, in some way, relates to its dimensionality. Then, the curse of dimensionality is weakened or can even be broken completely. The problem on the one hand is to detect and classify applications where this may happen and on the other hand to develop and implement numerical schemes which then allow to exploit such a situation. This is the subject of this article. We intend to give an overview on recent approaches and results in this direction from the view of function approximation and solution of partial differential equations.

The remainder of this paper is organized as follows: In Section 1.2, we briefly consider applications in which high-dimensional partial differential equations appear. We then discuss the breaking of the curse of dimensionality from the theoretical point of view. Here, we collect known approaches of getting rid of the exponential dependence on d. Furthermore we consider the theorem of Kolmogorov in more detail and give a survey of approximation schemes which are related to it.

In Section 1.3 we consider dimension-wise decompositions of high-dimensional functions. Here we resort to ANOVA-type decompositions where a function is split into its contributions from different groups of subdimensions, an approach which is widely used in statistics. It basically involves the splitting of a one-dimensional function space into the constant subspace and the remainder space. A product construc-

† If the solution is not smooth but possesses singularities, the order r of accuracy deteriorates. Adaptive refinement/nonlinear approximation is employed with success. In the best case, the cost-benefit ratio of a smooth solution can be recovered.

† An example would be the p-version of the finite element method if we couple the polynomial degree p to the dimension d and consider functions from the Sobolev space H^{p+1}.

tion then gives the associated splitting for the d-dimensional case. This reveals the relative importance of different dimensions as well as their interactions and correlations. For certain applications it can be observed that an apparently high-dimensional function possesses a low effective dimension or that there is a certain decay for the component functions with their dimension. Then the curse of dimensionality may be avoided. We formalize this with the help of reproducing kernel Hilbert spaces. The importance of the different contributions in an ANOVA splitting can then be expressed by certain weights.

For practical computations the associated subspaces need to be further discretized. This leads to so-called sparse grids which are discussed in detail in Section 1.4. To this end, we refine the remainder space of the one-dimensional splitting, i.e., we equip it with a basis. We use the standard piecewise linear hierarchical basis in one dimension (Faber 1909, Yserentant 1986) as the simplest example of a one-dimensional multiscale series expansion which involves interpolation by piecewise linears. Then the tensor product construction generates a basis for the d-dimensional case. A proper truncation – that can be formally derived by solving an optimization problem closely related to M-term approximation which involves the error norm and the smoothness assumption – results in sparse grids. For functions with bounded mixed second derivatives, approximation schemes are gained which exhibit cost complexities of the order $O(N(\log N)^{d-1})$ and give an accuracy of $O(N^{-2}(\log N)^{d-1})$ if we measure the error in the L_2-norm. However if we consider the energy norm, optimality leads us to an energy-based sparse grid with cost complexity $O(N)$ and accuracy $O(N^{-1})$ only. Thus, the exponential dependence of the logarithmic terms on d is completely removed (but is still present in the constants). Finally we discuss the order constants in more detail. In one special case we are able to show that, for the best approximation $v_M^{(E)}$ in the energy-norm based sparse grid space with dimension M, the following error estimate holds:

$$\|u - v_M^{(E)}\|_E \leq c \cdot d^2 \cdot 0.97515^d \cdot M^{-1} \cdot |u|_{\mathbf{2},\infty}$$

where the regularity term $|u|_{\mathbf{2},\infty}$ involves mixed second derivatives of u.

The concluding remarks of Section 1.5 summarize the discussion and give an outlook on current developments with sparse grids.

1.2 High dimensional problems and the curse of dimensionality

Usually in classical physics most problems are formulated as systems of (nonlinear) partial differential equations in three space dimension and one time dimension.† Here, the geometry of the object under consideration can be quite complicated. As examples consider flow around a car or an airplane, combustion in an engine, structural analysis of mechanical machines and buildings in civil engineering or related multiphysics applications which involve coupled systems of partial differential equations. These problems can nowadays be well treated with parallel adaptive finite element methods involving multilevel solvers on large parallel computers. They are at the edge of today's applications of numerical simulation in science and engineering. An efficient geometry description, subsequent parallel mesh generation, reliable a-posteriori error estimators for adaptive finite element discretizations, robust parallel multilevel solvers and load-balancing techniques for the overall approach are subjects of ongoing research and form the mainstream in scientific computing.

However there are also problems which involve substantially more than just three spatial dimensions. Then, high-dimensionality often results from mathematical modelling. Besides pure integration problems stemming from physics and finance, typically models from the stochastics and data analysis world show up. For example, high-dimensional Laplace/diffusion problems and high-dimensional convection diffusion problems result from diffusion approximation techniques or the Fokker-Planck equation. Examples are the description of queueing networks (Mitzlaff 1997, Shen, Chen, Dai and Dai 2002), random excitations of mechanical structures (Johnson, Wojtkiewicz, Bergman and Spencer 1997, McWilliam, Knappett and Fox 2000, Wojtkiewicz and Bergman 2000), reaction mechanisms in molecular biology (Sjöberg 2002, Elf, Lötstedt and Sjöberg 2001), the viscoelasticity in polymer fluids (Rouse 1953, Prakash and Öttinger 1999, Prakash 2000, Venktiteswaran and Junk 2005*a*, Venktiteswaran and Junk 2005*b*, Lozinski and Chauviere 2003, Lozinski, Chauviere, Fang and Owens 2003, Chauviere and Lozinski 2004, Süli 2006), or various models for the pricing of financial derivatives

† Of course physical theories may involve more than just three spatial dimensions. For example the equations related to superstring theories which can be regarded as limits of the M-theory or the theory of supergravitation are formulated in 10 or 11 dimensions, respectively, see (Green, Schwarz and Witten 1998) and the references cited therein.

(Duffie 1996, Kwok 1998, Wilmott 1998, Reisinger 2003, Schwab 2003, Escobar and Seco 2005). Furthermore, homogenization with multiple scales (Allaire 1992, Cioranescu, Damlamian and Griso 2002, Matache 2002, Hoang and Schwab 2003) as well as stochastic elliptic equations (Schwab and Todor 2003*a*, Schwab and Todor 2003*b*) result in high-dimensional PDEs. Next, we find quite high-dimensional problems in quantum mechanics and particle physics. Here, the dimensionality of the Schrödinger equation (Messiah 2000) grows with the number of considered electrons and nuclei. Then, problems in statistical mechanics lead to the Liouville equation or the Langevin equation and related phase space models where the dimension depends on the number of particles (Balescu 1997). Furthermore, reinforcement learning and stochastic optimal control in continuous time give raise to the Hamilton-Jacobi-Bellman equation in high dimensions (Sutton and Barto 1998, Munos 2000, Munos and Moore 2002). Finally data mining problems involve differential operators as smoothing or regularization terms (priors) whose dimension grows with the number of features of the data (Girosi, Jones and Poggio 1995, Garcke, Griebel and Thess 2001, Schölkopf and Smola 2002, Hegland 2003, Garcke 2004).

Now, in higher dimensions, the question of the shape of the domain is not as important as in the two- and three-dimensional case, since complicated domains typically do not appear in applications. Conceptually, besides $\mathbb{R}^d$ itself, we use mainly hypercubes like $[-a, a]^d$, $a \in \mathbb{R}$, and their straightforward generalizations using different values of a for each coordinate direction as well as the corresponding structures in polar coordinates. These domains are of tensor product structure. This is an important prerequisite for numerical methods for higher-dimensional partial differential equations as we will see later.

1.2.1 Curse of dimensionality

Classical approximation schemes exhibit the curse of dimensionality (Bellmann 1961). We then have

$$||f - f_M|| = O(M^{-r/d}) \,,$$

where r and d denote the isotropic smoothness of the function f and the problem's dimensionality, respectively. This is one of the main obstacles in the numerical treatment of high-dimensional problems. Therefore, the question is whether we can find situations, i.e., either function spaces or error norms, for which the curse of dimensionality can be broken. At first

glance, there is an easy way out: if we make a stronger assumption on the smoothness of the function f such that $r = O(d)$, then, we directly obtain $||f - f_M|| = O(M^{-c})$ with constant $c > 0$. Of course, such an assumption is quite unrealistic.

However, about thirteen years ago, (Barron 1993) found an interesting result: Denote by $\mathcal{F}L_1$ the class of functions with Fourier transforms in L_1. Then, consider the class of functions of $\mathbb{R}^d$ with $\nabla f \in \mathcal{F}L_1$. We expect for the best M-term approximation f_M an approximation rate

$$||f - f_M|| = O(M^{-1/d})$$

since $\nabla f \in \mathcal{F}L_1 \approx r = 1$. However Barron was able to show that

$$||f - f_M|| = O(M^{-1/2}),$$

independent of d. Meanwhile, other function classes have been introduced with such properties. They comprise certain radial basis schemes, stochastic sampling techniques and approaches that work with spaces of functions with bounded mixed derivatives.

A better understanding of these results is possible with the help of harmonic analysis (Donoho 2000). Here, we resort to the approach of the L_1-combination of L_∞-atoms, see also (Triebel 1992, DeVore 1998). Consider the class of functions $\mathcal{F}(K)$ with integral representation

$$f(x) = \int A(x,t) d\mu(t) \quad \text{with} \quad \int d|\mu|(t) \leq K\,, \tag{1.1}$$

where for fixed t we call $A(x,t) = A_t(x)$ an L_∞-atom, if $|A_t(x)| \leq 1$ holds. Then, there are results from Maurey for Banach spaces and Stechkin in Fourier analysis which state that there exists an M-term sum

$$f_M(x) = \sum_{j=1}^{M} a_j A_{t_j}(x)$$

where

$$||f - f_M||_\infty \leq C \cdot M^{-1/2}$$

with C independent of d.

As a first example we consider superpositions of Gaussian bumps (radial basis schemes). These resemble the space $\mathcal{F}(K, Gaussians)$ with $t := (x_0, s)$ and Gaussian atoms $A(x,t) = \exp(-\|x - x_0\|^2/s^2)$. Now, if the sum of the height of all Gaussians is bounded by K, (Niyogi and Girosi 1998) showed that the resulting approximation rate is independent of d for the corresponding radial basis schemes. There is no further

condition on the widths or positions of the bumps. Note that this corresponds to a ball in Besov space $B^d_{1,1}(\mathbb{R}^d)$ which is just the bump algebra in (Meyer 1992). Thus, we have nothing but a restriction to smoother functions for higher dimensions such that the ratio r/d stays constant and, consequently, $M^{-r/d}$ does again not grow with d.

Another class of functions with an approximation rate independent of d is $\mathcal{F}(K, Orthant)$ which uses the parameter set of shifted orthants. Now $t = (x_0, k)$, and k is the orthant indicator. Furthermore, $A(x,t)$ is the indicator of orthant k with apex at x_0. Again, if the integral (1.1) is at most K, the resulting approximation rate is of order $O(M^{-1/2})$ independent of d. A typical and well-known example for such a construction is the cumulative distribution function in $\mathbb{R}^d$. This just results in the Monte Carlo method.

A more general class are the functions which are formed by any superposition of 2^d functions, each orthantwise monotone for a different orthant. Now, the condition $\int d|\mu|(t) \leq 1$ is the same as

$$\frac{\partial^d f}{\partial x_1 \cdots \partial x_d} \in L_1 \,, \tag{1.2}$$

i.e., we obtain the space of bounded mixed first variation. Again, this means to consider only functions which get smoother as the dimensionality increases, but, in contrast to the examples mentioned above, now, only an anisotropic smoothness assumption is involved. Note that this is just the prerequisite for sparse grids with the piecewise constant hierarchical basis.

Further results on high-dimensional (and even infinite-dimensional) problems and their tractability were given by (Wasilkovski and Woźniakowski 1995, Sloan and Woźniakowski 1998, Wasilkovski and Woźniakowski 1999, Sloan 2001, Hickernell, Sloan and Wasilkowski 2004, Dick, Sloan, Wang and Woźniakowski 2004, Sloan, Wang and Woźniakowski 2004). Here, especially in the context of numerical integration, the notion of *weighted* Sobolev spaces was introduced. Following the observation that for some problems the integrand becomes less and less variable in successive coordinate directions, a sequence of positive weights $\{\gamma_j\}$ with decreasing values is used, with the weight γ_j being associated with coordinate direction j. Then it can be shown that the integration problem in a particular Sobolev space setting becomes *strongly tractable* (Traub and Woźniakowski 1980, Traub, Wasilkowski and Woźniakowski 1983, Traub, Wasilkowski and Woźniakowski 1988), i.e., that the worst-case error for all functions in the unit ball of the weighted Sobolev

space is bounded independently of d and tends polynomially to zero if and only if the sum of the weights is asymptotically bounded from above. This corresponds to a decay of the kernel contributions in a reproducing kernel Hilbert space with increasing d. The original paper (Sloan and Woźniakowski 1998) assumes that the integrand belongs to a Sobolev space of functions with square-integrable mixed first derivatives with the weights built into the definition of the associated inner product. Note that this assumption is closely related to that of (1.2) above. Since then, more general assumptions on the weights and, thus, on the induced weighted function spaces have been found (Dick et al. 2004, Hickernell et al. 2004, Sloan et al. 2004, Hickernell and Woźniakowski 2000, Wasilkowski and Woźniakowski 2004).

In any case, we observe that a certain smoothness assumption on the function under consideration changes with d and leads to approximation rates which no longer depend exponentially on d. This raises the question what smoothness for changing d and smoothness for $d \to \infty$ mean at all.

To this end, let us note an interesting aspect, namely the concentration of measure phenomenon (Milman 1988, Milman and Schechtman 2001, Talagrand 1995, Gromov 1999, Ledoux 2001) for probabilities in normed spaces in high dimensions (also known as the geometric law of large numbers). This is an important development in modern analysis and geometry, manifesting itself across a wide range of mathematical sciences, particularly geometric functional analysis, probability theory, graph theory, diverse fields of computer science, and statistical physics. In the statistical setting it states the following: Let f be a Lipschitz function with Lipschitz constant L on the d-sphere. Let P be a normalized Lebesgue measure on the sphere and let X be a random variable uniformly distributed with respect to P. Then,

$$P\{|f(X) - Ef(X)| > t\} \leq c_1 \exp(-c_2 t^2/L^2)$$

with constants c_1, c_2 independent of f and d. In its simplest form, the phenomenon of concentration of measure just says that every Lipschitz function on a sufficiently high-dimensional domain Ω is well approximated by a constant function (Hegland and Pestov 1999, Baxter and Iserles 2003). Thus, there is some chance to treat high-dimensional problems despite the curse of dimensionality.

The relation of the concentration of measure phenomenon to approximation estimates was further elaborated upon in (Hegland and Pozzi 2005). There, with the help of a concentration function which

expresses the concentration effect of the underlying metric space, new inequalities for the error of function approximations have been derived. Besides estimates for the above mentioned approximation of a function by a constant (e.g., by its mean or by the evaluation at a random point) and radial basis functions, also piecewise constant approximation schemes and piecewise approximations of higher order by Hermite polynomials have been studied. The resulting approximation rates were the same as with conventional estimates based on finite elements. However the constants in the estimates were substantially better. They are independent of the dimension and in addition allow realistic bounds for multimodal distributions which is not the case for classical approaches based on interpolation theory. These techniques may be employed to obtain better estimates for the constants in the order estimates with respect to the dimension in e.g., sparse grid approximation schemes. This however is future work.

1.2.2 The theorem of Komogorov and related approximation schemes

One approach to develop efficient approximations which allow one to overcome the curse of dimensionality is to describe multivariate continuous functions as a superposition (Rassias and Simsa 1995, Khavinson 1997) of a number of continuous functions with fewer variables. This question is related to Hilbert's 13th problem, see (Vitushkin 2004) and the references cited therein. It was answered in (Kolmogorov 1957) who found that every continuous function of several variables can be represented by the superposition of continuous functions with only two variables. Kolmogorov even showed that every continuous function of several variables can be represented by the superposition of continuous functions with only one variable,† see also (Sprecher 1965, Lorentz, v. Golitschek and Makovoz 1996, Khavinson 1997) for improved versions.

Kolmogorov's famous result can be expressed as follows: Let f be a multivariate continuous function on the unit cube, i.e., $f(x_1, ..., x_d) : [0,1]^d \to \mathbb{R}$. Each function $f \in C([0,1]^d)$ has a representation

$$f(x_1, ..., x_d) = \sum_{i=1}^{2d+1} f_i(\sum_{j=1}^{d} \phi_{i,j}(x_j)) \tag{1.3}$$

† Kolmogorov's student Arnold showed even before in (Arnold 1957, Arnold 1958, Arnold 1959) that any $f \in C([0,1]^3)$ can be represented as a superposition of continuous functions in two variables, and thus refuted Hilbert's conjecture.

where all $\{f_i\}$ and $\{\phi_{i,j}\}$ are one-dimensional continuous functions defined on $\mathbb{R}$ and all $\{\phi_{i,j}\}$ are independent of the choice of f. An improvement was given in (Fridman 1967) where it was shown that the inner functions $\{\phi_{i,j}\}$ can be chosen to be Lipschitz continuous with exponent one.

There have been various refinements of this result. A version with just one outer function and $2d+1$ inner functions, see (Lorentz et al. 1996) Chapter 17, reads: There exist $2d+1$ continuous, strictly increasing functions $\phi_i : [0,1] \to [0,1]$ and d positive constants λ_i with $\sum_{i=1}^{d} \lambda_i \leq 1$ with the property that each function $f \in C([0,1]^d)$ has a representation

$$f(x_1, ..., x_d) = \sum_{i=1}^{2d+1} g(\sum_{j=1}^{d} \lambda_j \phi_i(x_j)) \tag{1.4}$$

for some non-smooth $g \in C([0,1])$ depending on f. Here, the functions ϕ_i together with their summation provide a one-to-one embedding of the unit cube $[0,1]^d$ into $\mathbb{R}^{2d+1}$, i.e., we have with $X_i := \sum_{j=1}^{d} \lambda_j \phi_i(x_j)$ the representation $f(x_1, ..., x_d) = \sum_{i=1}^{2d+1} g(X_i)$. Note that there is a close relation to d-dimensional topology: The theorem of Menger and Nöbeling, see (Hurewicz and Wallman 1948), page 84, tells us that any d-dimensional compact set can be homeomorphically embedded into $[0,1]^{2d+1}$. Thus, Kolmogorov's theorem (1.4) can be seen as just a special case of it.

Another version with $2d+1$ outer functions and one inner function is due to (Sprecher 1965). It reads

$$f(x_1, ..., x_d) = \sum_{i=1}^{2d+1} f_i(\sum_{j=1}^{d} \lambda_j \phi(x_j + i \cdot \alpha)) \tag{1.5}$$

with suitable constants λ_i, α and a continuous one-dimensional function ϕ. It also gives an embedding of $[0,1]^d$ into $\mathbb{R}^{2d+1}$ by $f(x_1, ..., x_d) = \sum_{i=1}^{2d+1} f_i(X_i)$ with $X_i := \sum_{j=1}^{d} \lambda_j \phi(x_j + i \cdot \alpha)$.

The proof of Kolmogorov's theorem is non-constructive and does not provide us with a way to choose the inner and outer functions in (1.3), (1.4) or (1.5), respectively. A first attempt to remedy this problem with merely an approximation of the inner function and an interpolation of the outer functions was made in (de Figueiredo 1980). Recently, however, algorithms were given to explicitly construct the functions in (1.5). The implementation of an inner function ϕ which does not depend on f was discussed in (Sprecher 1996). Here ϕ is pointwise defined on an everywhere dense set of rational numbers in $[0,1]$ from which

it can be uniquely extended to a continuous function on $[0,1]$. The resulting ϕ is non-continuous. There is also a close relation of ϕ to space-filling curves, see (Sprecher and Draghici 2002, Sagan 1994). An implementation of the outer functions f_i by an iterative method was presented in (Sprecher 1997). This established the first constructive proof of Kolmogorov's theorem. It furthermore allows to realize (1.5) as a feedforward neural network with a hidden layer that computes the variables X_i and therefore involves the embedding mapping only, and a single output layer in which f is computed by means of the functions $f_i(X_i)$, see also (Hecht-Nielsen 1987*a*, Hecht-Nielsen 1987*b*). In (Köppen 2002) the construction was improved to give a continuous inner function ϕ.

Thus, in view of Kolmogorov's result, it seems that there are no high-dimensional functions and thus no high-dimensional problems at all. However, it turns out that the representing functions are quite bad, i.e., they are at best only continuous and highly non-smooth. This limits their practical use for approximation and interpolation purposes (Girosi and Poggio 1989), like e.g., for the discretization of PDEs within the Galerkin approach. In particular, the representing functions cannot be chosen to be differentiable. This even holds if one wants to represent an analytic function f only, see (Vitushkin 1964).

Nevertheless Kolmogorov's theorem inspired many linear and nonlinear approximation schemes and there have been various attempts to generalize Kolmogorov's formula. Moreover, in (Kurkova 1991) it was noticed that in the proof of Kolmogorov's superposition theorem the fixed number of $2d+1$ basis functions can be replaced by a variable number m and the task of function representation can be replaced by the task of function approximation. In the following we give an (incomplete) list of approaches which are related to Kolmogorov's theorem.

- Popular approximation schemes in statistics are the so-called additive models, see (Hastie and Tibshirani 1986, Hastie and Tibshirani 1990). They resemble the approximation

$$f(x_1, ..., x_d) \approx \sum_{i=1}^{d} f_i(x_i). \tag{1.6}$$

  This form can be derived from (1.3) by choosing d instead of $2d+1$ and replacing the inner functions ϕ_{ij} trivially by the identity if $i = j$ and zero otherwise.
- The projection pursuit algorithm (Friedman and Stützle 1981, Stone

1985) approximates a function f by

$$f(x_1, ..., x_d) \approx f_0 + \sum_{i=1}^{m} f_i(\sum_{j=1}^{d} \beta_{ij} \cdot x_j)$$

with the so-called projection directions $\vec{\beta_i} := (\beta_{i1}, ..., \beta_{id})$ and with f_0 as the average of the function f. Here the parameter vectors $\vec{\beta_i}$ and the functions f_i are estimated from the data. This scheme can also be interpreted as a special case of (1.6) using linear combinations of the original coordinates.

- Closely related are multilayer perceptrons with a single hidden layer. They approximate f by

$$f(x_1, ..., x_d) \approx h(\sum_{i=1}^{m} \alpha_i g(\sum_{j=1}^{d} \beta_{ij} x_j))$$

where h and g are arbitrary nonlinear functions. Here, the network is trained with a given set of input and output values and the approximation is then determined by the values of α_i and the vectors $\vec{\beta_i}$ which are found by least squares minimization. In (Kurkova 1991) it was demonstrated how to approximate a Hecht-Nielsen network which implements Komogorov's superposition approach by such traditional neural networks.

- Also radial basis schemes belong to the class of approximation schemes which can be derived from (1.3). Here the dimension embedding takes place by a distance function, i.e., the sum of the inner functions gets replaced by the Euclidean norm. They can under certain assumptions be written as

$$f(x) \approx \sum_{i=1}^{m} \beta_i f_i(\|x - y_i\|) + P(x)$$

where $x = (x_1, \ldots, x_d)$, the f_i are a chosen set of radial basis functions, $y_i \in \mathbb{R}^d$ are their centers, the β_i are constants and $P(x)$ is a polynomial. The coefficients are then fitted to the data by means of least-squares minimization.

- Starting from the representation (1.4), (Igelnik and Parikh 2003) introduced so-called Kolmogorov spline networks. There, $2d+1$ is replaced by a general m and the outer function g and the inner functions ϕ_i are replaced by cubic splines $s(\cdot, \gamma_i)$ and $s(\cdot, \gamma_{i,j})$ where the parameters γ_i and $\gamma_{i,j}$ of the splines are adjusted to fit given data on f properly.

The approximation scheme is defined as

$$f_m(x_1, ..., x_d) = \sum_{i=1}^{m} s(\sum_{j=1}^{d} \lambda_j s(x_j, \gamma_{i,j}), \gamma_i) \tag{1.7}$$

with positive numbers $\lambda_1, \ldots, \lambda_d$ with $\sum \lambda_j \leq 1$ which can be chosen independent of f. It was shown that, for any function f from the class of continuously differentiable functions on $[0,1]^d$ with bounded gradient, there exists a function f_m of the form (1.7) such that $\|f - f_m\| = O(1/m)$. The number of degrees of freedom involved in the network is of the order $O(m^{2/3})$. This result compares favorably with the approximation order $O(1/\sqrt{m})$ and the number of degrees of freedom $O(m^2)$ usually achieved for general one-hidden layer feedforward networks for this class of functions, compare also (Barron 1993, Igelnik and Parikh 2003) and the references cited therein. A similar approach was also presented in (Coppejans 2004).

- Finally, Kronecker-product type approximations of the form

$$f(x) \approx \sum_{i=1}^{m} c_i \prod_{j=1}^{d} f_{ij}(x_j), \quad c_i \in \mathbb{R},$$

possess a structure similar to Kolmogorov's theorem. Here however the inner sum is replaced by a product of one-dimensional functions. To see the relation to (1.3) choose there $f_i(\cdot) = c_i \cdot \exp(\cdot)$ and $\phi_{ij}(\cdot) = \log(f_{ij}(\cdot))$. For numerical purposes, the one-dimensional functions f_{ij} are further expanded in a series with a suitable multilevel basis which is then properly truncated or, more often, they are just simply discretized on a uniform grid. For details and applications see (Beylkin and Molenkamp 2002, Beylkin and Molenkamp 2005, Tyrtyshnikov 2004, Hackbusch and Khoromskij 2004), the related developments on so-called H- and H^2-matrices (Hackbusch, Khoromskij and Sauter 2000, Grasedyck and Hackbusch 2003) and the references cited therein. A similar decomposition is used in the MCTDH approach (Beck, Jäckle, Worth and Meyer 2000).

The basic theory of this decomposition can be found in (Golomb 1959): In the case d=2 mainly the classical Hilbert-Schmidt theory appears, i.e., the functions f_{ij} are the unique solution of a system of two coupled linear integral equations which resemble the continuous analogue of the classical singular value decomposition. For the case $d > 2$ however, a system of nonlinear integral equations results for which the solution is no longer unique. Then heuristics must be used

to obtain some solution. Nevertheless it is observed in applications that good approximations can be obtained with an already relatively small number m.

For most of these approximation schemes the parameters are obtained by some kind of (least-squares) minimization. Here, however the objective functional may not be globally convex and can have many minima which results in non-unique representations. Thus, the associated approximation rates for these schemes for increasing m are not always fully understood and, moreover, it is not clear which representation to prefer over another for a particular application. In the following we therefore study a simpler linear decomposition of a d-dimensional function into its contributions from different (groups of) subdimensions which can be seen as a multivariate generalization of (1.6).

1.3 Dimension-wise space decomposition

We consider a decomposition of the d-dimensional function f as

$$f(x_1,...,x_d) = f_0 + \sum_{j_1}^{d} f_{j_1}(x_{j_1}) + \sum_{j_1<j_2}^{d} f_{j_1,j_2}(x_{j_1},x_{j_2}) \tag{1.8}$$
$$+ \sum_{j_1<j_2<j_3}^{d} f_{j_1,j_2,j_3}(x_{j_1},x_{j_2},x_{j_3}) + \cdots + f_{j_1,...,j_d}(x_{j_1},...,x_{j_d}).$$

Here, f_0 is a constant function, f_{j_1} are one-dimensional functions, f_{j_1,j_2} are two-dimensional functions, and so on. This type of decomposition goes back to (Hoeffding 1948) and is well known in statistics under the name ANOVA (analysis of variance), see also (Efron and Stein 1981). Note that (1.8) is a finite expansion of f into 2^d different terms. Such a decomposition can be gained by a tensor product construction of a splitting of the one-dimensional function space into its constant subspace and its remainder. This will be explained in more detail in the following.

1.3.1 ANOVA-like decompositions

Let $V^{(d)}$ denote the underlying space of d-dimensional functions $f(x) = f(x_1,..,x_d) : \bar{\Omega}^{(d)} \to \mathbb{R}$ with $\bar{\Omega}^{(d)} = [0,1]^d$. Let μ be a product measure with unit mass which has a density, i.e.,

$$d\mu(x) = \prod_{j=1}^{d} d\mu_j(x_j), \quad \int_{\bar{\Omega}^{(1)}} d\mu(x_j) = 1, \tag{1.9}$$

$$d\mu(x) = h(x) = \prod_{j=1}^{d} h_j(x_j) dx_j, \tag{1.10}$$

where $h_j(x_j)$ is the marginal density of the input variable x_j. Furthermore, let $V^{(d)}$ be a Hilbert space equipped with the inner product $(f,g) = \int_{\bar{\Omega}^{(d)}} f(x)g(x)d\mu(x)$ and associated norm $\|\cdot\|$.

First we will deal with the one-dimensional case. We decompose $V^{(1)}$ in a simple two-scale fashion by

$$V^{(1)} = \mathbf{1} \oplus W \tag{1.11}$$

where $\mathbf{1}$ denotes the one-dimensional subspace $span\{1\}$ which contains the constant functions. Associated to such a splitting is a mapping $P : V^{(1)} \to \mathbf{1}$ with

$$Pf(x) = \int_{\bar{\Omega}^{(1)}} f(x) d\mu(x). \tag{1.12}$$

Examples are the conventional Lebesque measure $d\mu(x) = dx$ which leads to the integral average

$$Pf(x) = \int_{\bar{\Omega}^{(1)}} f(x) dx$$

or the Dirac measure located at a point a, i.e., $d\mu(x) = \delta(x-a)dx$, which results in the simple evaluation at point a

$$Pf(x) = \int_{\bar{\Omega}^{(1)}} \delta(x-a) f(x) dx = f(a). \tag{1.13}$$

This introduces the decomposition

$$f(x) = f_0 + f_1(x) \tag{1.14}$$

with

$$f_0 = Pf(x) = \int_{\bar{\Omega}^{(1)}} f(x) d\mu(x) \in \mathbf{1} \text{ and} \tag{1.15}$$

$$f_1(x) = (I-P)f(x) = f(x) - \int_{\bar{\Omega}^{(1)}} f(x) d\mu(x) \in W. \tag{1.16}$$

Then W is the subspace of $V^{(1)}$ of functions which satisfy the relation $\int_{\bar{\Omega}^{(1)}} f(x) d\mu(x) = 0$. It is orthogonal to $\mathbf{1}$ and, with $\int_{\bar{\Omega}^{(1)}} f_1 d\mu(x) = 0$, it is easy to see that $\|f\|^2 = \|f_0\|^2 + \|f_1\|^2$ and $(f,g) = (f_0,g_0) + (f_1,g_1)$ with g split analogously to (1.14).

Note that for (1.13) with differentiable f there is a close relation to the Taylor expansion. The decomposition (1.14) is just the Taylor formula of first order and $f_1 = f(x) - f(a) = \int_a^x f'(t)dt$ is the remainder term.

Now we consider the d-dimensional case: The one-dimensional splitting introduces a natural decomposition of the d-dimensional function space $V^{(d)}$ by a tensor product construction

$$\begin{aligned} V^{(d)} &= \bigotimes_{j=1}^{d} (\mathbf{1}_j \oplus W_j) \\ &= \mathbf{1}_1 \otimes \cdots \otimes \mathbf{1}_d \\ &\oplus \bigoplus_{i=1}^{d} \mathbf{1}_1 \otimes \cdots \otimes W_i \otimes \cdots \otimes \mathbf{1}_d \\ &\oplus \bigoplus_{i=1}^{d} \bigoplus_{i<j} \mathbf{1}_1 \otimes \cdots \otimes W_i \otimes \cdots \otimes W_j \otimes \cdots \otimes \mathbf{1}_d \\ &\oplus \bigoplus_{i=1}^{d} \bigoplus_{i<j} \bigoplus_{i<j<k} \mathbf{1}_1 \otimes \cdots \otimes W_i \otimes \cdots \otimes W_j \otimes \cdots \otimes W_k \otimes \cdots \otimes \mathbf{1}_d \\ &\cdots \\ &\oplus W_1 \otimes \cdots \otimes W_d. \end{aligned} \tag{1.17}$$

Here $\mathbf{1}_j = \mathbf{1}$ and $W_j = W$; we use the index j merely to indicate the respective coordinate direction for explanatory reasons. Another notation which involves the subsets of the index set $\{1, 2, \ldots, d\}$ is †

$$V^{(d)} = \bigoplus_{\mathbf{u} \subset \{1,\ldots,d\}} \left(\bigotimes_{k \in \{1,\ldots,d\}/\mathbf{u}} \mathbf{1}_k \right) \otimes \left(\bigotimes_{j \in \mathbf{u}} W_j \right) =: \bigoplus_{\mathbf{u} \subset \{1,\ldots,d\}} W_{\mathbf{u}}. \tag{1.18}$$

Then, a function $f \in V^{(d)}$ is decomposed accordingly as

$$\begin{aligned} f(x_1, ..., x_d) &= f_0 + \sum_{j_1}^{d} f_{j_1}(x_{j_1}) + \sum_{j_1<j_2}^{d} f_{j_1,j_2}(x_{j_1}, x_{j_2}) \\ &+ \sum_{j_1<j_2<j_3}^{d} f_{j_1,j_2,j_3}(x_{j_1}, x_{j_2}, x_{j_3}) + \cdots + f_{j_1,\ldots,j_d}(x_{j_1}, ..., x_{j_d}) \\ &= \sum_{\mathbf{u} \subset \{1,\ldots,d\}} f_{\mathbf{u}}(x_{\mathbf{u}}) \end{aligned} \tag{1.19}$$

where $f_{\mathbf{u}} \in W_{\mathbf{u}}$ and $x_{\mathbf{u}}$ denotes the variables x_i of x with $i \in \mathbf{u}$. Note

† Note the obvious identity $\prod_{j \in \mathbf{v}} (b_j + c_j) = \sum_{\mathbf{u} \subset \mathbf{v}} \prod_{k \in \mathbf{v}/\mathbf{u}} b_k \prod_{j \in \mathbf{u}} c_j, \forall b_j, c_j \in \mathbb{R}$. It can be applied for products of sums of functions and products of sums of subspaces in an analogous way.

that due to the power set construct this is a finite expansion which involves 2^d different terms. The decomposition is unique for a fixed choice of the one-dimensional mapping $P : V^{(1)} \to \mathbf{1}$.

Associated is the identity

$$\begin{aligned} I^{(d)} &= \bigotimes_{j=1}^{d} (P_j + (I_j - P_j)) \\ &= \sum_{\mathbf{u} \subset \{1,\ldots,d\}} \left(\prod_{k \in \{1,\ldots,d\}/\mathbf{u}} P_k \right) \cdot \left(\prod_{j \in \mathbf{u}} (I_j - P_j) \right) =: \sum_{\mathbf{u} \subset \{1,\ldots,d\}} P_{\mathbf{u}} \end{aligned}$$

where P_j and I_j denote the one-dimensional projection operator (1.12) and the identity for the j-th coordinate direction, respectively. Here the projection $\prod_{j=1}^{d} P_j f$ is the unconditional mean of f (with respect to the measure μ) and the partial projection $\prod_{k \in \{1,\ldots,d\}/\mathbf{u}} P_k f$ is the conditional mean $\int .. \int f(x) \prod_{k \in \{1,\ldots,d\}/\mathbf{u}} d\mu_k(x_k)$.

For the example of the conventional Lebesque measure $d\mu(x) = dx$ we obtain the functions in (1.19) as

$$\begin{aligned} f_0 &= \prod_{j=1}^{d} P_j f = \int_{\bar{\Omega}^{(d)}} f(x) \prod_{i=1}^{d} dx_i, \\ f_{j_1}(x_{j_1}) &= \int_{\bar{\Omega}^{(d-1)}} f(x) \prod_{i \neq j_1} dx_i - f_0, \\ f_{j_1,j_2}(x_{j_1,j_2}) &= \int_{\bar{\Omega}^{(d-2)}} f(x) \prod_{i \notin \{j_1,j_2\}} dx_i - f_{j_1}(x_{j_1}) - f_{j_2}(x_{j_2}) - f_0, \\ \ldots \quad & \quad \ldots \\ f_{j_1,\ldots,j_k}(x_{j_1,\ldots,j_k}) &= \int_{\bar{\Omega}^{(d-k)}} f(x) \prod_{i \notin \{j_1,\ldots,j_k\}} dx_i \\ &\quad - \sum_{i_1 < \cdots < i_{k-1} \subset \{j_1,\ldots,j_k\}} f_{i_1,\ldots,i_{k-1}}(x_{i_1}, \ldots, x_{i_{k-1}}) \\ &\quad - \sum_{i_1 < \cdots < i_{k-2} \subset \{j_1,\ldots,j_k\}} f_{i_1,\ldots,i_{k-2}}(x_{i_1}, \ldots, x_{i_{k-2}}) \\ &\quad \ldots \\ &\quad - \sum_{j_1} f_{j_1}(x_{j_1}) - f_0, \\ \ldots \quad & \quad \ldots \end{aligned} \tag{1.20}$$

This is just the well known ANOVA decomposition used in statistics, see (Efron and Stein 1981, Wahba 1990) and the references cited therein.

There, if the input consists of independently distributed uniform random variables (with respect to the Lebesque measure) then the component functions are uncorrelated and the total variance D can be written as

$$D = E(f - f_0)^2 = \sum_{j_1} D_{j_1} + \sum_{j_1<j_2} D_{j_1,j_2} + \sum_{j_1<j_2<j_3} D_{j_1,j_2,j_3} + \dots D_{j_1,\dots j_d}$$

with the partial variances†

$$D_{j_1,\dots,j_k} = \int_{\bar{\Omega}^{(k)}} (f_{j_1,\dots,j_k})^2 dx_{j_1} \dots dx_{j_k}.$$

For the example of the Dirac measure located at a point a_j, i.e., with $d\mu(x_j) = \delta(x_j - a_j)dx_j$ and $d\mu(x) = \prod_{j=1}^d d\mu(x_j)$, we obtain the functions in (1.19) as

$$\begin{aligned}
f_0 &= f(x)|_{x=a} \\
f_{j_1}(x_{j_1}) &= f(x)|_{x=a\setminus x_{j_1}} - f_0, \\
f_{j_1,j_2}(x_{j_1,j_2}) &= f(x)|_{x=a\setminus\{x_{j_1},x_{j_2}\}} - f_{j_1}(x_{j_1}) - f_{j_2}(x_{j_2}) - f_0, \\
\dots &\quad \dots \\
f_{j_1,\dots,j_k}(x_{j_1,\dots,j_k}) &= f(x)|_{x=a\setminus\{x_{j_1},\dots,x_{j_k}\}} \\
&\quad - \sum_{\{i_1,\dots,i_{k-1}\}\subset\{j_1,\dots,j_k\}} f_{i_1,\dots,i_{k-1}}(x_{i_1},\dots,x_{i_{k-1}}) \\
&\quad - \sum_{\{i_1,\dots,i_{k-2}\}\subset\{j_1,\dots,j_k\}} f_{i_1,\dots,i_{k-2}}(x_{i_1},\dots,x_{i_{k-2}}) \\
&\quad \dots \\
&\quad - \sum_{j_1} f_{j_1}(x_{j_1}) - f_0, \\
\dots &\quad \dots
\end{aligned} \tag{1.21}$$

where now only (partial) point evaluations in the point $a = (a_1, .., a_d)$ are involved. Here we use the notation

$$f(x)|_{x=a\setminus x_i} = f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_d)$$

with its obvious generalization to $a \setminus \{x_{j_1}, \dots, x_{j_k}\}$. This approach is considered in (Rabitz and Alis 1999) under the name cut-HDMR and is closely related to the anchor spaces of (Sloan et al. 2004, Dick et al. 2004, Hickernell and Woźniakowski 2000, Wasilkowski and Woźniakowski 2004). Note that a component function vanishes if the value of one of

† The global sensitivity indices are then defined as $S_{j_1,\dots,j_k} = D_{j_1,\dots,j_k}/D$. They describe the contribution of the input $\{x_{j_1}, \dots, x_{j_k}\}$ to the variance of the output.

its input variables x_i is equal to the associated coordinate of the point a, i.e.,

$$f_{j_1,\ldots,j_k}(x_{j_1,\ldots,j_k})|_{x_i=a_i} = 0 \quad i \in \{j_1,\ldots,j_k\}.$$

Thus the decomposition (1.21) expresses f as a superposition of its values on lines, faces, hyperplanes etc. which pass through the point a. Also note that the component functions fulfill

$$f_{i_1,\ldots,i_p}(x_{i_1,\ldots,i_p}) f_{j_1,\ldots,j_q}(x_{j_1,\ldots,j_q})|_{x_k=a_k} = 0 \quad k \in \{i_1,\ldots,i_p\} \cup \{j_1,\ldots,j_p\},$$

which is a direct consequence of the orthogonality

$$\int_{\bar{\Omega}^{(d)}} f_{i_1,\ldots,i_p}(x_{i_1,\ldots,i_p}) f_{j_1,\ldots,j_q}(x_{j_1,\ldots,j_q}) d\mu(x) = 0$$

with $d\mu(x) = \prod_{j=1}^{d} \delta(x_j - a_j) dx_j$.

Note furthermore that for differentiable f there is a close relation to the multivariate Taylor expansion. The decomposition (1.21) is just the multivariate Taylor formula of first order in each coordinate direction with partial remainder terms. Moreover, the multivariate Taylor expansion of f around a (provided that f is sufficiently many times differentiable of course) and a short calculation shows the following: The component functions of first order, i.e., f_{j_1}, are the sum of all terms in the Taylor series which depend only on x_{j_1}, the component functions of second order, i.e., f_{j_1,j_2}, are the sum of all terms in the Taylor series which depend on x_{j_1} and x_{j_2}, and so on. Thus (1.21) resembles a rearrangement of the infinite number of terms in the full Taylor series into a finite number, i.e., 2^d, of different groups where each group corresponds to one component function (which still contains as series an infinite number of terms).

Note finally that there are various generalizations of (1.21). Instead of the Dirac measure at one point a we could also take the average of the Dirac measures at m different points and build an ANOVA-type decomposition on it. This approach is closely related to the multi-cut-HDMR method of (Li, Rosenthal and Rabitz 2001*b*, Li, Schoendorf, Ho and Rabitz 2004). Other variants (mp-cut-HDMR and lp-RS) can be found in (Li, Wang, Rosenthal and Rabitz 2001*c*) and (Li, Atramonov, Rabitz, Wang, Georgopoulos and Demiralp 2001*a*), respectively.

In summary an ANOVA-type decomposition of f into component functions reveals the relative importance of the different dimensions as well as their interactions and correlations. In general, an arbitrary function f may result in zero components except of its highest order term

$f_{j_1,\ldots,j_d}(x_{j_1},\ldots,x_{j_d})$ or might have all its components beeing relevant. Then, nothing is gained with respect to the curse of dimensionality when switching from f to its ANOVA decomposition. However in many practical applications it can be observed that the finite series (1.19) decays rapidly. In some cases it is even of *finite order* q, i.e., for the components $f_{\mathbf{u}}$ of the decomposition (1.19) there holds

$$f_{\mathbf{u}} = 0 \text{ with } |\mathbf{u}| > q$$

with $q << d$. This usually expresses the fact that reasonable, meaningful (observable) coordinates of the physical system under consideration had been chosen. Alternatively it may happen that the different dimensions are not of equal importance and we find a decay in the contribution of the dimensions (after sorting according to their relevance) and their associated higher order interactions.

Examples with such types of behavior of the expansion (1.19) are:

- In most molecular dynamics simulation codes only two-body (bonds), three-body (angle) and four-body (dihedral) potential functions are used to describe molecules, i.e., it holds that the finite order of the associated ANOVA decomposition is trivially $q \leq 4$. It seems that this mostly gives a sufficient representation of the potential energy hypersurface of a system, especially when macroscopic variables are sought.
- Closely related is the Mayer cluster expansion in statistical mechanics. Here, for pair potentials $U_{ij}(x_i, x_j)$ which express the interaction between two particles, the term

  $$\exp(-\sum_{i<j} U_{ij}(x_i, x_j)) = \prod_{i<j} \exp(-U_{ij}(x_i, x_j))$$

  gets transformed by $\exp(-U_{ij}(x_i, x_j)) =: 1 + \phi_{ij}(x_i, x_j)$ into

  $$1 + \sum_{i<j} \phi_{ij} + \sum_{i<j} \sum_{k<l} \phi_{ij}\phi_{kl} + \sum_{i<j} \sum_{k<l} \sum_{m<n} \phi_{ij}\phi_{kl}\phi_{mn} + \cdots$$

  which allows one to write the partition function of the canonical ensemble of a particle system by means of cluster integrals, for details see (Hill 1956), page 123 ff.
- In many statistical applications a statistics of second order is sufficient, i.e., the covariances of the input variables play an important role, but higher-order correlations are neglected. Again this means that we have finite order $q \leq 2$ in the associated ANOVA decomposition.

- In data mining it is found from multivariate adaptive regression splines (MARS), see (Friedman 1991), that even for really high-dimensional data there appear at most 5-7 dimensional interactions, i.e., $q \leq 7$, and higher-order interactions are practically not significant.
- The Brownian bridge representation of a Markov process results in a concentration of the total variation in the first few levels of the discretization since the variance decays with the factor $2^{-1/2}$ from level to level, see also (Caflisch, Morokoff and Owen 1997, Morokoff 1998, Gerstner and Griebel 2003, Gerstner and Griebel 1998) where the Brownian bridge was used in high-dimensional integration problems. A further analysis in view of reproducing kernel Hilbert spaces with weights is given in (Leobacher, Scheicher and Larcher 2003). Note that for the Karhunen-Loewe decomposition an even better decay may result than for the Brownian bridge.
- Many problems in mathematical finance can be formulated as high-dimensional integrals, where the large number of dimensions arises from small time steps in time discretization and/or a large number of state variables. Examples are option pricing, bond valuation or the pricing of collateral mortgage backed securities. There, it turns out that for the ANOVA decomposition of the integrand the importance of each dimension is naturally weighted by certain hidden weights where with the increase of dimension the lower-order terms continue to play a significant role and the higher-order terms tend to be negligible, see (Caflisch et al. 1997, Sloan and Wang 2005). This is a reason why Quasi-Monte-Carlo performs better than expected, especially for high-dimensional integrands.
- There is also a counterexample. In quantum chemistry, the solution of Schrödinger's equation for d fermions has to obey the antisymmetry condition due to Pauli's principle. It can be shown that for an ANOVA decomposition of an antisymmetric f all terms $f_{\mathbf{u}}$ with $|\mathbf{u}| < d-1$ are identically zero and all information of f is contained in the terms with order $d-1$ and d.†

In summary, in certain applications, i.e., for f from special function spaces, we know a-priori how the ANOVA-components decay and may resort to a truncation of (1.19) after the q-th order terms, where q is related to the needed accuracy. Or we may a-priori know if ANOVA-terms of order higher than some q are present at all or not. The question is

† For bosonic systems a symmetry condition is needed instead. Then all functions $f_{\mathbf{u}}$ with the same order $|\mathbf{u}|$ are the same.

how these situations and the associated function spaces can be characterized. A possibility are reproducing kernel Hilbert spaces. The associated multi-dimensional kernel function can be decomposed analogously to the ANOVA expansion into a sum of kernels. Then these partial kernels can be equipped with different individual weights. These weights allow one to model various behaviors of decay for the different contributions in the ANOVA decomposition as well as truncations to finite order. This will be dealt with in the following.

1.3.2 Reproducing kernel Hilbert spaces

The theory of reproducing kernel Hilbert spaces (RKHS) was introduced in (Aronzaijn 1950). It allows to describe function spaces in a concise and elegant way by means of so-called reproducing kernel functions. To this end, we assume that $f : [0,1]^d \to \mathbb{R}$ belongs to a Hilbert space H with associated inner product $\langle \cdot, \cdot \rangle_H$ and norm $\|f\|_H = \langle f, f \rangle_H^{1/2}$. We assume that H is continuously embedded into $L_2([0,1]^d)$. Thus, we consider integrable functions f with respect to the Lebesgue measure for which $\|f\|_{L_2} := \left(\int_{[0,1]^d} f^2(t) dt \right)^{1/2} < \infty$. Furthermore, there is a non-negative number $c(H)$ depending on the space H such that

$$\|f\|_{L_2} \leq c(H) \, \|f\|_H \qquad \text{for all } f \in H. \tag{1.22}$$

Finally we assume that the evaluation of the function f is well-defined and continuous, i.e., that the linear functional $f \in H \mapsto f(x)$ is continuous for any $x \in [0,1]^d$. These assumptions are equivalent to the requirement that H is a reproducing kernel Hilbert space, see (Aronzaijn 1950). Hence, H has an associated kernel $K^{(d)} : [0,1]^d \times [0,1]^d \to \mathbb{R}$ which is uniquely defined by the following three conditions:

- $K^{(d)}(\cdot, t) \in H$ for all $t \in [0,1]^d$,
- $\left(K^{(d)}(x_i, x_j) \right)_{i,j=1}^n$ is a symmetric and non-negative definite matrix for all n and points x_i from $[0,1]^d$,
- $f(t) = \left\langle f, K^{(d)}(\cdot, t) \right\rangle_H$ for all $f \in H$ and all $t \in [0,1]^d$ (reproducing kernel property).

Thus it is sufficient to give $K^{(d)}$ to uniquely characterize the associated function space H. The theory of reproducing kernel Hilbert spaces can be found in detail in (Aronzaijn 1950); further aspects are discussed in (Wahba 1990, Ritter 2000).

From the three properties of reproducing kernels it easily follows that

$$\begin{aligned} K^{(d)}(t,x) &= \left\langle K^{(d)}(\cdot,x), K^{(d)}(\cdot,t)\right\rangle_H \qquad \text{for all } t,x \in [0,1]^d, \\ \sqrt{K^{(d)}(t,t)} &= \|K^{(d)}(\cdot,t)\|_H \qquad \text{for all } t \in [0,1]^d, \\ |f(t)| &\le \|f\|_H \sqrt{K^{(d)}(t,t)} \qquad \text{for all } f \in H,\ t \in [0,1]^d. \end{aligned}$$

If H is separable, then for an arbitrary orthonormal basis $\{\eta_i\}$, we have $K^{(d)}(\cdot,x) = \sum_{i=1}^{\dim(H)} c_i \eta_i$ with $c_i = \left\langle \eta_i, K^{(d)}(\cdot,x)\right\rangle_H = \eta_i(x)$. Therefore

$$K^{(d)}(x,t) = \sum_{i=1}^{\dim(H)} \eta_i(x)\eta_i(t) \qquad \text{for all } x,t \in [0,1]^d. \tag{1.23}$$

In a way, the reverse of this argument is also true, see (Wahba 1990). To this end, let $\{\eta_i\}_{i=1}^{\infty}$ be a given arbitrary sequence of linearly independent functions defined on $[0,1]^d$ such that $\sum_{i=1}^{\infty} \eta_i^2(t) < \infty$ for all $t \in [0,1]^d$. Consider the space $H = \operatorname{span}\{\eta_1, \eta_2, \dots\}$ of functions $f(t) = \sum_{i=1}^{\infty} f_i \eta_i(t)$ with real numbers f_i such that $\sum_{i=1}^{\infty} f_i^2 < \infty$. Observe that $f(t)$ is well-defined. For $f \in H$ the coefficients f_i are uniquely determined since the η_i's are linearly independent. The inner product in H is given by requiring that the η_i's be orthonormal, $\langle \eta_i, \eta_j\rangle_H = \delta_{i,j}$. Hence, for $f, g \in H$ we have $\langle f, g\rangle_H = \sum_{i=1}^{\infty} f_i g_i$ with f_i and g_i being the coefficients of f and g, respectively. Then H is a Hilbert space. Furthermore, it can be easily shown that

$$K^{(d)}(x,t) = \sum_{i=1}^{\infty} \eta_i(x)\eta_i(t)$$

is its reproducing kernel.

Note that the Hilbert space $L_2([0,1]^d)$ does not have a reproducing kernel, since point evaluation $t \in [0,1]^d \mapsto f(t)$ is not well-defined for $L_2([0,1]^d)$ and thus can not be continuous. It is easy to see that H is continuously embedded in L_2 if we assume that

$$\int_{[0,1]^d} K^{(d)}(t,t)\, dt < \infty. \tag{1.24}$$

Indeed, $f^2(t) \le \|f\|_H^2 \cdot K^{(d)}(t,t)$, and therefore (1.22) holds with

$$c(H) = \left(\int_{[0,1]^d} K^{(d)}(t,t)\, dt \right)^{1/2}. \tag{1.25}$$

In this case, H is a proper subset of L_2, and $K^{(d)}(\cdot,t) \in L_2$ for arbitrary $t \in [0,1]^d$. Many examples of reproducing kernel Hilbert spaces can be found in the literature, see for example (Wahba 1990, Ritter 2000).

Remember now our approach in Subsection 1.3.1: We first split $V^{(1)} = \mathbf{1} \oplus W$ in (1.11) and then used the tensor product construction in (1.17) to gain the decomposition $V^{(d)} = \bigoplus_{\mathbf{u} \subset \{1,\ldots,d\}} W_{\mathbf{u}}$ in (1.18) for the d-dimensional case. From (Aronzaijn 1950) we know the following facts:

- The reproducing kernel for the direct sum of two orthogonal subspaces is the sum of the single reproducing kernels.
- The reproducing kernel for a tensor product of two RKHS is the product of the single reproducing kernels.

This allows us, depending on the one-dimensional splitting and the associated orthogonal norm, to build a d-dimensional RKHS as a product of the sum of one-dimensional RKHSs. If we have for the orthogonal splitting (1.11) the associated sum of reproducing kernels $K(x,y) = K^{\mathbf{1}}(x,y) + K^W(x,y)$ we obtain for the splitting (1.18) the corresponding kernel

$$\begin{aligned} K^{(d)}(x,y) &= \prod_{j=1}^{d} (K_j^{\mathbf{1}}(x_j,y_j) + K_j^W(x_j,y_j)) \\ &= \sum_{\mathbf{u} \subset \{1,\ldots,d\}} \prod_{k \in \{1,\ldots,d\} \setminus \mathbf{u}} K_k^{\mathbf{1}}(x_k,y_k) \cdot \prod_{j \in \mathbf{u}} K_j^W(x_j,y_j) \\ &=: \sum_{\mathbf{u} \subset \{1,\ldots,d\}} K_{\mathbf{u}}(x,y) \end{aligned}$$

with $K_{\mathbf{u}}(x,y) = \prod_{k \in \{1,\ldots,d\} \setminus \mathbf{u}} K_k^{\mathbf{1}}(x_k,y_k) \cdot \prod_{j \in \mathbf{u}} K_j^W(x_j,y_j)$. Here we again use the indices j and k to indicate the respective coordinate directions. In the special case of $K_j^{\mathbf{1}}(x_j,y_j) = 1$ we directly have

$$K^{(d)}(x,y) = \sum_{\mathbf{u} \subset \{1,\ldots,d\}} \prod_{j \in \mathbf{u}} K_j^W(x_j,y_j).$$

1.3.3 Weighted spaces

Now we are in the position to introduce weights into the splittings. We follow (Sloan et al. 2004, Dick et al. 2004, Wasilkowski and Woźniakowski 2004), see also (Kuo and Sloan 2005). First we consider the simple case where each dimension gets its own weight $\gamma_j \in \mathbb{R}_0^+, j = 1,\ldots,d$, i.e.,

where d different non-negative weights are involved. We then have $K_j^{\mathbf{1}} = 1$ and replace K_j^W by $\gamma_j \cdot K_j^W$.† We obtain with $\prod_{k\in\{1,\ldots,d\}\setminus\mathbf{u}} K_k^{\mathbf{1}} = 1$

$$\begin{aligned} K^{(d)}(x,y) &= \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \prod_{j\in\mathbf{u}} \gamma_j \cdot \prod_{j\in\mathbf{u}} K_j^W(x_j,y_j) \\ &=: \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \gamma_{d,\mathbf{u}} \cdot K_{d,\mathbf{u}}(x_{\mathbf{u}},y_{\mathbf{u}}) \end{aligned}$$

with $\gamma_{d,\mathbf{u}} = \prod_{j\in\mathbf{u}} \gamma_j$ and $K_{d,\mathbf{u}}(x_{\mathbf{u}},y_{\mathbf{u}}) = \prod_{j\in\mathbf{u}} K_j^W(x_j,y_j)$. The resulting weights $\gamma_{d,\mathbf{u}}$ are just products of the γ_j.

We can generalize this approach as follows: We set

$$\begin{aligned} K^{(d)}(x,y) &= \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \gamma_{d,\mathbf{u}} \cdot \prod_{j\in\mathbf{u}} K_j^W(x_j,y_j) \\ &=: \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \gamma_{d,\mathbf{u}} \cdot K_{d,\mathbf{u}}(x_{\mathbf{u}},y_{\mathbf{u}}) \end{aligned} \tag{1.26}$$

where we now allow 2^d general non-negative weights $\gamma_{d,\mathbf{u}} \in \mathbb{R}_0^+$ which need no longer be formed as products of one-dimensional weights γ_i but may be chosen arbitrarily. Here we use the convention $\gamma_{d,\{\}} = 1$ and $\prod_{j\in\{\}} K_j^W = 1$.

As an example we consider the reproducing kernels

$$K_j^W(x,y) = \frac{1}{2}B_2(x-y) + (x-\frac{1}{2})(y-\frac{1}{2}) + \mu_j(x) + \mu_j(y) + m_j$$

where $B_2(x) := x^2 - x + 1/6$ denotes the Bernoulli polynomial of degree 2, μ_j is a function with bounded derivative in $[0,1]$ such that

$$\int_0^1 \mu_j(x)dx = 0, \quad m_j := \int_0^1 (\mu'(x))^2 dx.$$

This kernel was presented for example in (Sloan et al. 2004). It allows to capture the two types of ANOVA-decompositions introduced in (1.20) and (1.21) as special cases and it also allows one to generalize them by means of the weights $\gamma_{d,\mathbf{u}}$.

The choice $\mu_j(x) = 0, j = 1,\ldots,d$, gives $m_j = 0$ and thus $K_j^W(x,y) = \frac{1}{2}B_2(x-y) + (x-\frac{1}{2})(y-\frac{1}{2})$. Note that $\int_0^1 K_j^W(x,y)dy = 0, \forall x \in [0,1]$. Then, the associated kernel (1.26) is called the ANOVA Sobolev kernel with general weights $\gamma_{d,\mathbf{u}}$. It can be shown that the associated inner

† Without loss of generality we choose here the weight one in front of the kernel $K^{\mathbf{1}}$ which is associated to the subspace $\mathbf{1}$ of constants.

product in $V^{(d)}$ is now

$$\langle f,g\rangle_{V^{(d)}} = \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \frac{1}{\gamma_{d,\mathbf{u}}} \int_{[0,1]^{|\mathbf{u}|}} \left(\int_{[0,1]^{d-|\mathbf{u}|}} \frac{\partial^{|\mathbf{u}|} f(x)}{\partial x_{\mathbf{u}}} dx_{-\mathbf{u}} \right. \\ \left. \int_{[0,1]^{d-|\mathbf{u}|}} \frac{\partial^{|\mathbf{u}|} g(x)}{\partial x_{\mathbf{u}}} dx_{-\mathbf{u}} \right) dx_{\mathbf{u}} \tag{1.27}$$

where we interpret the term associated to $\mathbf{u} = \{\}$ as the product of integrals $\int_{[0,1]^d} f(x)dx \int_{[0,1]^d} g(x)dx$. Here, $x_{\mathbf{u}}$ denotes the $|\mathbf{u}|$-dimensional vector of the components x_j with $j \in \mathbf{u}$ and $x_{-\mathbf{u}}$ denotes $x_{\{1,\ldots,d\}\setminus\mathbf{u}}$.

The choice

$$K_j^W(x,y) = \begin{cases} \min(|x-a_j|,|y-a_j|), & \text{if } (x-a_j)(y-a_j) > 0, \\ 0, & \text{else,} \end{cases}$$

leads for (1.26) to the so-called anchored ANOVA Sobolev kernel with point $a = (a_1,\ldots,a_d)$ and general weights $\gamma_{d,\mathbf{u}}$. The associated inner product is

$$\langle f,g\rangle_{V^{(d)}} = \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \frac{1}{\gamma_{d,\mathbf{u}}} \int_{[0,1]^{|\mathbf{u}|}} \frac{\partial^{|\mathbf{u}|} f(x_{\mathbf{u}},a_{-\mathbf{u}})}{\partial x_{\mathbf{u}}} \frac{\partial^{|\mathbf{u}|} g(x_{\mathbf{u}},a_{-\mathbf{u}})}{\partial x_{\mathbf{u}}} dx_{\mathbf{u}} \tag{1.28}$$

where $(x_{\mathbf{u}}, a_{-\mathbf{u}})$ denotes the d-dimensional vector whose j-th component is equal to x_j if $j \in \mathbf{u}$ and to a_j if $j \notin \mathbf{u}$, respectively. For the case $\mathbf{u} = \{\}$ we set $\int_{[0,1]^{|\{\}|}} f(x_{\{\}}, a_{-\{\}})dx_{\{\}} := f(a)$.

In both cases the associated weighted inner product can be written as

$$\langle f,g\rangle_{V^{(d)}} = \sum_{\mathbf{u}\subset\{1,\ldots,d\}} \frac{1}{\gamma_{d,\mathbf{u}}} (f_{\mathbf{u}},g_{\mathbf{u}})_{V_{\mathbf{u}}} \tag{1.29}$$

where

$$(f_{\mathbf{u}},g_{\mathbf{u}})_{V_{\mathbf{u}}} = \int_{[0,1]^{|\mathbf{u}|}} \frac{\partial^{|\mathbf{u}|} f_{\mathbf{u}}(x_{\mathbf{u}})}{\partial x_{\mathbf{u}}} \frac{\partial^{|\mathbf{u}|} g_{\mathbf{u}}(x_{\mathbf{u}})}{\partial x_{\mathbf{u}}} d\mu(x_{\mathbf{u}}) \tag{1.30}$$

and $f = \sum_{\mathbf{u}} f_{\mathbf{u}}$ and $g = \sum_{\mathbf{u}} g_{\mathbf{u}}$ are the ANOVA decompositions of f and g with respect to the chosen measure $d\mu(x)$. A straightforward calculation which uses the fact that $f_{\mathbf{u}}$ and $g_{\mathbf{u}}$ are the components of an ANOVA decomposition and thus possess orthogonality properties shows that (1.30) and the integrals in the sum (1.27) and (1.28) are indeed equivalent.†

† Plug the ANOVA decompositions $f = \sum_{\mathbf{v}} f_{\mathbf{v}}$ and $g = \sum_{\mathbf{v}} g_{\mathbf{v}}$ into (1.27), change the order of the sum, the integral and the derivative and use the orthogonality of the ANOVA decomposition and the fact that the partial derivative is non-zero only if $\mathbf{u} \subset \mathbf{v}$. An analogous argument holds for (1.28).

From (1.29) we see the effect of the weights on the inner product: The $\{\gamma_{d,\mathbf{u}}\}$ are non-negative numbers which measure the influence of the associated partial derivative of the function and, consequently, also the influence of the corresponding terms $f_{\mathbf{u}}$ of the decomposition (1.14). Note that for positive weights the associated weighted norm

$$\|f\|_{V^{(d)}} = \sqrt{\langle f, f\rangle_{V^{(d)}}} = \sqrt{\sum_{\mathbf{u}} \frac{1}{\gamma_{d,\mathbf{u}}} (f_{\mathbf{u}}, f_{\mathbf{u}})_{V_{\mathbf{u}}}}$$

is equivalent (up to a constant) to the conventional norm in $V^{(d)}$ (with just a weighting of the contributions to the overall norm). However, for any $\mathbf{u}$ with $\gamma_{d,\mathbf{u}} \to 0$ the associated contribution to the norm is forced to zero since

$$\frac{1}{\gamma_{d,\mathbf{u}}} (f_{\mathbf{u}}, f_{\mathbf{u}})_{V_{\mathbf{u}}} \leq \text{const.} \Rightarrow (f_{\mathbf{u}}, f_{\mathbf{u}})_{V_{\mathbf{u}}} \leq \text{const.} \cdot \gamma_{d,\mathbf{u}} \to 0.$$

Thus $f_{\mathbf{u}} = 0$, the associated subspace $W_{\mathbf{u}}$ is switched off and we obtain a true subspace of the overall space.

The weights $\{\gamma_{d,\mathbf{u}}\}$ therefore allow to explicitly prescribe the importance of different dimensions and of the correlations and interactions between (groups of) dimensions and thus allow to characterize the associated function spaces and the possibly low, hidden dimensionality of intrinsically high-dimensional functions.

An attempt in this direction was the concept of effective dimension introduced in (Caflisch et al. 1997). There, based on the ANOVA decomposition of a function, the distribution of the overall variance to the ANOVA components was considered. This lead to the definitions of the truncation dimension d_t and the superposition dimension d_s of a function. There, f has truncation dimension d_t if the sum of the partial variances of the ANOVA terms $f_{\mathbf{u}}$ with $\mathbf{u} \subset \{1, \ldots, d_t\}$ exceeds 99 percent of the total variance $\sigma(f)$. Alternatively, f has superposition dimension d_s if the sum of the partial variances of the ANOVA terms $f_{\mathbf{u}}$ with order $|\mathbf{u}| \leq d_s$ exceeds 99 percent of the total variance. It was argued that the success of Quasi Monte Carlo methods for high-dimensional problems from finance is due to the relatively low effective dimensions of the integrands involved. In particular, the example of a mortgage backed security with nominal 360 dimensions from (Paskov and Traub 1995) showed an effective dimension of about only 50 in the truncation sense and about 32 in the superposition sense, see (Caflisch et al. 1997) for details and (Sloan and Wang 2005) for a further discussion on this subject. With the help of the general weights $\{\gamma_{d,\mathbf{u}}\}$,

besides these two simple situations, more general situations can now be modeled and analyzed. In addition to the product weights mentioned above, also the case of order-dependent weights, i.e., the interaction between the variables in $x_{\mathbf{u}}$ depends only on $|\mathbf{u}|$, and the case of finite-order weights, i.e., there exists $q \in \mathbb{N}$ such that $\gamma_{d,\mathbf{u}} = 0$ for all $|\mathbf{u}| > q$, has been studied, see (Sloan and Woźniakowski 1998, Dick et al. 2004, Hickernell et al. 2004, Sloan et al. 2004, Hickernell and Woźniakowski 2000, Wasilkowski and Woźniakowski 2004) and the references cited therein. This was mainly done for the analysis of Quasi Monte Carlo methods and lattice rules for the numerical integration of high-dimensional functions.

A closely related approach with weighted kernels can be found in the area of data analysis, where the weights are called rescaling parameters. There, for so-called interaction spline models (Wahba 1990), page 129 ff., strategies are discussed to delete ANOVA-component subspaces driven by data fitting methods. The weights $\gamma_{d,\mathbf{u}}$ are not given a-priori but are determined in an adaptive fashion by statistical tests. Alternative techniques are the l_1-penalty method or the structured Multicategory Support Vector Machine where an updating algorithm is used for the tuning of the weights, see (Lee, Lin and Wahba 2004*b*, Lee, Kim, Lee and Koo 2004*a*).

Now, with a-priori knowledge, i.e., for the case of weights with finite order q, where $\gamma_{d,\mathbf{u}} = 0$ for $|\mathbf{u}| > q$, only a proper discretization of the remaining component functions is needed. Then, the curse of dimensionality is no longer present with respect to d but only with respect to q, as shown for quadrature and more general linear problems in (Wasilkowski and Woźniakowski 2004). In a similar fashion, for the case of sufficiently fast-decaying weights, the series expansion (1.19) may be truncated accordingly, which then results in an approximation to the true function.† Then, again, for a proper discretization of the remaining component functions, the curse of dimensionality may no longer be present with respect to d but only with respect to a smaller intrinsic dimension of the overall function. In this sense, the curse of dimensionality can be broken and extrinsically high-dimensional problems can be tackled.

Note that a truncation of the ANOVA series introduces a modelling error whereas the subsequent discretization of the remaining subspaces relates to a discretization error. This unnatural distinction between modelling error and subsequent discretization error can be overcome if

† Alternatively a closure by an approximation of the higher order terms by means of products of lower order terms or similar may be sought.

we intertwine the truncation of the ANOVA series and the discretization. However, how these two types of errors may be balanced, how this may be done in a purely adaptive fashion, what the smoothness assumptions, a-posteriori error indicators and refinement procedures, first, for the ANOVA parts and, second, within the discretization of the ANOVA parts have to be and how they relate to each other is presently not completely clear, especially for PDE-based problems.‡ Nevertheless, such a type of approach needs to be developed and applied in the area of partial differential equations in the future.

1.4 Sparse grids

So far we have seen how an ANOVA-type decomposition may be used to detect important and unimportant correlations and interactions between (groups of) dimensions. However the components in the decomposition (1.19) are still continuous functions and the corresponding subspaces $W_{\mathbf{u}}$ are in general infinite-dimensional. Moreover, for practical computations a choice of basis and a further discretization is needed for each of the subspaces. To this end, we can follow the same principle as in Subsection 1.3.1. First we equip the space W in the one-dimensional splitting (1.11) with a proper (infinite) basis $\{\phi_k\}$ (the constant, i.e., $\phi_0 = 1$ is excluded). Then we apply the tensor product construction (1.17) to come to the d-dimensional case. Here we just form the products of the respective one-dimensional basis functions. This results in an induced basis $\{\phi_{\mathbf{u},\mathbf{k}}\}$ for each of the ANOVA subspaces $W_{\mathbf{u}}$ with multi-index $\mathbf{k} = (k_{j_1}, \ldots, k_{j_{|\mathbf{u}|}}) \in \mathbb{N}^{|\mathbf{u}|}$ where $\phi_{\mathbf{u},\mathbf{k}}(x_{\mathbf{u}}) = \prod_{j\in\mathbf{u}} \phi_{k_j}(x_j) \cdot \prod_{j\in\{1,\ldots,d\}\setminus\mathbf{u}} 1$. This may be seen as an (infinite) refinement of the ANOVA decomposition by a further decomposition of the space W (e.g., by means of the span of certain basis functions).† Then, we can write each component function as linear combination of these basis functions. This overall expansion of the sum of

‡ If we assume, for example, $f_{\mathbf{u}} \in C^{|\mathbf{u}|\cdot k}([0,1]^{|\mathbf{u}|}), \forall \mathbf{u}$, we directly see that $f \in C^{k\cdot d}([0,1]^d)$. However there exist more partial derivatives like, for example, the $|\mathbf{u}|$-th mixed derivative of $f_{\mathbf{u}}$ and, therefore, f belongs to the space of bounded k-th mixed derivatives, compare also (Novak and Ritter 1998). Note that the inverse direction of this implication is in general not valid. Now, for functions whose ANOVA decomposition is of finite order q, only a prerequisite $f_{\mathbf{u}} \in C^{q\cdot k}([0,1]^{|\mathbf{u}|})$ is at most necessary to have f from the space of bounded k-th mixed derivatives.

† Note the close relation to the work in (Lemieux and Owen 2002, Liu and Owen 2005), to MARS (Friedman 1991), to the WARNAX model (Wei and Billings 2004) and to tensor product space ANOVA models (Gu and Wahba 1993, Lin 2000). For example in (Wahba 1990), page 130, the one-dimensional two-scale splitting is extended to more terms which correspond to higher derivatives. Orthogonality

component functions of the ANOVA decomposition must be truncated properly in each of its components to obtain a finite dimensional approximation to f and to its component functions f_{j_1}, $f_{j_1,j_2}, \ldots,$ $f_{j_1,\ldots j_d}$. This leads to so-called sparse grids. Depending on the smoothness assumed – here usually certain mixed derivatives have to be bounded – and depending on the specific one-dimensional basis chosen, such an approach allows to get rid of the curse of dimensionality to some extent.

This concept works for quite general systems of one-dimensional basis functions. Candidates are the eigenbasis of an associated one-dimensional differential operator (which may be chosen depending on the respective higher dimensional problem under consideration), classical Fourier bases, (hierarchical) global polynomial systems (Boyd 2000, Karniadakis and Sherwin 1999, Szabo and Babuska 1991, Bungartz 1998, Bungartz and Griebel 2004) or function families with localization properties like wavelets (Daubechies 1992), prewavelets (Chui and Wang 1992, Griebel and Oswald 1995*b*) or interpolets (Deslauriers and Dubuc 1989, Donoho and Yu 1999) and related wavelet-like constructs, see (Cohen 2003, Bungartz and Griebel 2004) for a survey. But also multiscale finite element systems and frames (Oswald 1994, Griebel 1994, Griebel and Oswald 1994, Griebel and Oswald 1995*a*, Griebel and Oswald 1995*b*) may be used.

In the following, we restrict ourselves for reasons of simplicity to the standard hat function and the associated hierarchical Faber basis. It is closely related to piecewise linear finite elements and thus suited for the discretization of elliptic PDEs of second order in weak form. This choice allows in a straightforward way to derive approximation orders and cost complexities by simple geometric series arguments and triangle inequalities. Moreover, for this special choice of basis we are able to also derive estimates of the constants involved and their dependence on the dimension d. We now closely follow (Bungartz 1992, Bungartz and Griebel 1999, Bungartz 1998, Bungartz and Griebel 2004).

1.4.1 Hierarchical multilevel subspace splitting

1.4.1.1 Subspace decomposition

Let $\bar{\Omega} := [0,1]^d$ denote the d-dimensional unit interval. We consider multivariate functions u, $u(\mathbf{x}) \in \mathbb{R}$, $\mathbf{x} := (x_1, \ldots, x_d) \in \bar{\Omega}$, with (in

is then given by a proper eigenbasis associated to this splitting. This results in so-called interaction splines.

some sense) bounded weak mixed derivatives

$$D^{\boldsymbol{\alpha}} u \;:=\; \frac{\partial^{|\boldsymbol{\alpha}|_1} u}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \tag{1.31}$$

up to some given order $r \in \mathbb{N}_0$. Here, $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ denotes a d-dimensional multi-index with the norms $|\boldsymbol{\alpha}|_1 := \sum_{j=1}^d \alpha_j$ and $|\boldsymbol{\alpha}|_\infty := \max_{1 \le j \le d} \alpha_j$. Furthermore, we use for multi-indices component-wise arithmetic operations, for example $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} := (\alpha_1 \beta_1, \ldots, \alpha_d \beta_d)$, $\gamma \cdot \boldsymbol{\alpha} := (\gamma \alpha_1, \ldots, \gamma \alpha_d)$, or $2^{\boldsymbol{\alpha}} := (2^{\alpha_1}, \ldots, 2^{\alpha_d})$, the relational operators $\boldsymbol{\alpha} \le \boldsymbol{\beta} :\Leftrightarrow \forall_{1 \le j \le d}\ \alpha_j \le \beta_j$ and $\boldsymbol{\alpha} < \boldsymbol{\beta} \quad :\Leftrightarrow \quad \boldsymbol{\alpha} \le \boldsymbol{\beta} \ \wedge\ \exists_{1 \le j \le d}\ \alpha_j < \beta_j$, and, finally, special multi-indices like $\mathbf{0} := (0, \ldots, 0)$ or $\mathbf{1} := (1, \ldots, 1)$, and so on.

In the following, for $q \in \{2, \infty\}$ and $r \in \mathbb{N}_0$, we study the spaces

$$\begin{aligned} X^{q,r}(\bar{\Omega}) \;&:=\; \left\{ u : \bar{\Omega} \to \mathbb{R} : D^{\boldsymbol{\alpha}} u \in L_q(\Omega), |\boldsymbol{\alpha}|_\infty \le r \right\}, \\ X_0^{q,r}(\bar{\Omega}) \;&:=\; \left\{ u \in X^{q,r}(\bar{\Omega}) : u|_{\partial\Omega} = 0 \right\}. \end{aligned} \tag{1.32}$$

Thus, $X^{q,r}(\bar{\Omega})$ denotes the space of all functions of bounded (with respect to the L_q-norm) mixed derivatives up to order r, and $X_0^{q,r}(\bar{\Omega})$ will be the subspace of $X^{q,r}(\bar{\Omega})$ consisting of those $u \in X^{q,r}(\bar{\Omega})$ which vanishes on the boundary $\partial\Omega$. Note that we first restrict ourselves to the case of homogeneous boundary conditions, i.e., to $X_0^{q,r}(\bar{\Omega})$. As smoothness parameter $r \in \mathbb{N}_0$, we need $r = 2$ for the case of piecewise linear approximations which will be in the following in the focus of our considerations. Finally, for functions $u \in X_0^{q,r}(\bar{\Omega})$ and multi-indices $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}|_\infty \le r$, we introduce the seminorm

$$|u|_{\boldsymbol{\alpha},\infty} \;:=\; \left\| D^{\boldsymbol{\alpha}} u \right\|_\infty . \tag{1.33}$$

Now, with the multi-index $\mathbf{l} = (l_1, \ldots, l_d) \in \mathbb{N}^d$ which indicates the level of refinement in a multivariate sense, we consider the family of d-dimensional standard rectangular grids

$$\left\{ \Omega_{\mathbf{l}}, \mathbf{l} \in \mathbb{N}^d \right\} \tag{1.34}$$

on $\bar{\Omega}$ with mesh size

$$\mathbf{h}_{\mathbf{l}} \;:=\; (h_{l_1}, \ldots, h_{l_d}) \;:=\; 2^{-\mathbf{l}}. \tag{1.35}$$

That is, the grid $\Omega_{\mathbf{l}}$ is equidistant with respect to each individual coordinate direction, but, in general, may have different mesh sizes in the different coordinate directions. The grid points $\mathbf{x}_{\mathbf{l},\mathbf{i}}$ of grid $\Omega_{\mathbf{l}}$ are just the points

$$\mathbf{x}_{\mathbf{l},\mathbf{i}} \;:=\; (x_{l_1,i_1}, \ldots, x_{l_d,i_d}) \;:=\; \mathbf{i} \cdot \mathbf{h}_{\mathbf{l}}, \qquad \mathbf{0} \le \mathbf{i} \le 2^{\mathbf{l}}. \tag{1.36}$$

Thus, here and in the following, the multi-index $\mathbf{l}$ indicates the level (of a grid, a point, or, later on, a basis function, respectively), whereas the multi-index $\mathbf{i}$ denotes the location of a given grid point $\mathbf{x}_{\mathbf{l},\mathbf{i}}$ in the respective grid $\Omega_{\mathbf{l}}$.

Next, we define discrete approximation spaces and sets of basis functions that span those discrete spaces. In a piecewise linear setting, the simplest choice of a 1 D basis function is the standard hat function $\phi(x)$,

$$\phi(x) \;:=\; \begin{cases} 1-|x|, & \text{if } x \;\in\; [-1,1]\,, \\ 0, & \text{else.} \end{cases} \tag{1.37}$$

This function can be used to generate an arbitrary $\phi_{l_j,i_j}(x_j)$ with support $[x_{l_j,i_j} - h_{l_j}, x_{l_j,i_j} + h_{l_j}] = [(i_j-1)h_{l_j}, (i_j+1)h_{l_j}]$ by dilation and translation, that is

$$\phi_{l_j,i_j}(x_j) \;:=\; \phi\left(\frac{x_j - i_j \cdot h_{l_j}}{h_{l_j}}\right). \tag{1.38}$$

The resulting 1 D basis functions are the input of the tensor product construction which provides a suitable piecewise d-linear basis function in each grid point $\mathbf{x}_{\mathbf{l},\mathbf{i}}$ (see Figure 1.1):

$$\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \;:=\; \prod_{j=1}^{d} \phi_{l_j,i_j}(x_j)\,. \tag{1.39}$$

Since we deal with homogeneous boundary conditions (i.e., with $X_0^{q,2}(\bar{\Omega})$),

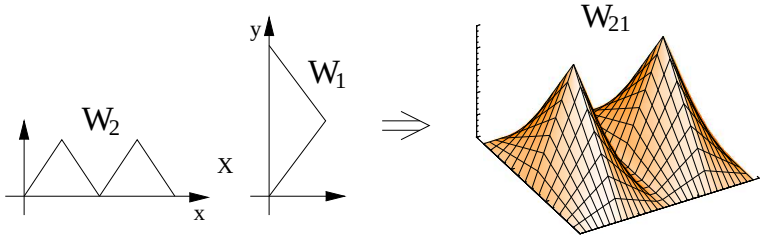

Fig. 1.1. Tensor product approach for piecewise bilinear basis functions.

only those $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$ that correspond to *inner* grid points of $\Omega_{\mathbf{l}}$ are taken into account for the definition of

$$V_{\mathbf{l}} \;:=\; \operatorname{span}\left\{\phi_{\mathbf{l},\mathbf{i}} :\; \mathbf{1} \le \mathbf{i} \le 2^{\mathbf{l}} - \mathbf{1}\right\}, \tag{1.40}$$

the space of piecewise d-linear functions with respect to the interior of

$\Omega_{\mathbf{l}}$. Obviously, the $\phi_{\mathbf{l},\mathbf{i}}$ form a basis of $V_{\mathbf{l}}$, with one basis function $\phi_{\mathbf{l},\mathbf{i}}$ of a support of the fixed size $2 \cdot \mathbf{h}_{\mathbf{l}}$ for each inner grid point $\mathbf{x}_{\mathbf{l},\mathbf{i}}$ of $\Omega_{\mathbf{l}}$, and this basis $\{\phi_{\mathbf{l},\mathbf{i}}\}$ is simply the standard nodal point basis of the finite dimensional space $V_{\mathbf{l}}$.

Additionally, we introduce the hierarchical increments $W_{\mathbf{l}}$,

$$W_{\mathbf{l}} \;:=\; \operatorname{span}\left\{\phi_{\mathbf{l},\mathbf{i}} :\; \mathbf{1} \le \mathbf{i} \le 2^{\mathbf{l}} - \mathbf{1},\; i_j \text{ odd for all } 1 \le j \le d\right\}, \quad (1.41)$$

for which the relation

$$V_{\mathbf{l}} \;=\; \bigoplus_{\mathbf{k} \le \mathbf{l}} W_{\mathbf{k}} \quad (1.42)$$

can be easily seen. Note that the supports of all basis functions $\phi_{\mathbf{l},\mathbf{i}}$ spanning $W_{\mathbf{l}}$ are mutually disjoint. Thus, with the index set

$$\mathbf{I}_{\mathbf{l}} \;:=\; \left\{\mathbf{i} \in \mathbb{N}^d : \mathbf{1} \le \mathbf{i} \le 2^{\mathbf{l}} - \mathbf{1},\; i_j \text{ odd for all } 1 \le j \le d\right\}, \quad (1.43)$$

we get another basis of $V_{\mathbf{l}}$, the *hierarchical basis*

$$\{\phi_{\mathbf{k},\mathbf{i}} : \mathbf{i} \in \mathbf{I}_{\mathbf{k}}, \mathbf{k} \le \mathbf{l}\} \quad (1.44)$$

which generalizes the well-known 1 D basis shown in Figure 1.2 to the d-dimensional case by means of a tensor product approach. With these

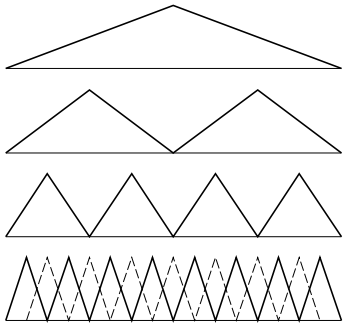

Fig. 1.2. Piecewise linear hierarchical basis (solid) vs. nodal point basis (dashed).

hierarchical difference spaces $W_{\mathbf{l}}$, we can define

$$V^{(d)} \;:=\; \sum_{l_1=1}^{\infty} \cdots \sum_{l_d=1}^{\infty} W_{(l_1,\ldots,l_d)} \;=\; \bigoplus_{\mathbf{l} \in \mathbb{N}^d} W_{\mathbf{l}} \quad (1.45)$$

with its natural hierarchical basis

$$\left\{\phi_{\mathbf{l},\mathbf{i}} : \mathbf{i} \in \mathbf{I}_{\mathbf{l}}, \mathbf{l} \in \mathbb{N}^d\right\}. \quad (1.46)$$

Except for completion with respect to the H^1-norm, $V^{(d)}$ is just the underlying Sobolev space $H_0^1(\bar{\Omega})$, i.e., $\bar{V}^{(d)} = H_0^1(\bar{\Omega})$.

Now it is easy to see that any function $u \in H_0^1(\bar{\Omega})$ and, consequently, any $u \in X_0^{q,2}(\bar{\Omega})$ can be uniquely split by

$$u(\mathbf{x}) \;=\; \sum_{\mathbf{l}} u_{\mathbf{l}}(\mathbf{x}), \qquad u_{\mathbf{l}}(\mathbf{x}) \;=\; \sum_{\mathbf{i}\in\mathbf{I}_{\mathbf{l}}} v_{\mathbf{l},\mathbf{i}} \cdot \phi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \;\in\; W_{\mathbf{l}}\,, \tag{1.47}$$

where the $v_{\mathbf{l},\mathbf{i}} \in \mathbb{R}$ are the coefficient values of the hierarchical product basis representation of u.

1.4.1.2 Basic properties of the subspaces

We summarize the most important properties of the hierarchical subspaces $W_{\mathbf{l}}$.

From (1.41) and (1.43), we immediately learn the dimension of $W_{\mathbf{l}}$, i. e. the number of degrees of freedom (grid points or basis functions, resp.) associated with $W_{\mathbf{l}}$:

$$|W_{\mathbf{l}}| \;=\; |\mathbf{I}_{\mathbf{l}}| \;=\; 2^{|\mathbf{l}-\mathbf{1}|_1}\,. \tag{1.48}$$

The question is now how important $W_{\mathbf{l}}$ is for the interpolation of some given $u \in X_0^{q,2}(\bar{\Omega})$. In the following we will discuss the contribution of a subspace $W_{\mathbf{l}}$ to the overall interpolant according to (1.47). Here, besides the L_p-norms, $p \in \{2, \infty\}$ we will concentrate on the energy norm

$$\|u\|_E \;:=\; \left(\int_\Omega \sum_{j=1}^{d} \left(\frac{\partial u(\mathbf{x})}{\partial x_j} \right)^2 \mathrm{d}\mathbf{x} \right)^{1/2}, \tag{1.49}$$

which is equivalent to the H^1-norm in $H_0^1(\bar{\Omega})$. For the Laplacian, (1.49) indeed indicates the energy norm in finite element terminology.†

First, we look at the different hierarchical basis functions $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$. A straightforward calculation gives

$$\|\phi_{\mathbf{l},\mathbf{i}}\|_E \;=\; \sqrt{2} \cdot \left(\frac{2}{3}\right)^{(d-1)/2} \cdot 2^{-|\mathbf{l}|_1/2} \cdot \left(\sum_{j=1}^{d} 2^{2l_j} \right)^{1/2}. \tag{1.50}$$

Next, we consider the hierarchical coefficient values $v_{\mathbf{l},\mathbf{i}}$ in more detail. They can be computed from the function values $u(\mathbf{x}_{\mathbf{l},\mathbf{i}})$ in the following

† Note that analogous results for the maximum norm $\|\cdot\|_\infty$ and the L_p-norm $\|\cdot\|_p$ (in general $p = 2$) can be found in e.g., (Bungartz and Griebel 2004).

way:

$$v_{\mathbf{l},\mathbf{i}} \;=\; \left(\prod_{j=1}^{d} \left[-\tfrac{1}{2}\; 1\; -\tfrac{1}{2} \right]_{x_{l_j,i_j},l_j} \right) u. \tag{1.51}$$

This is due to the definition of the spaces $W_{\mathbf{l}}$ and their basis functions (1.41), whose supports are mutually disjoint and do not contain coarse grid points $\mathbf{x_{k,j}}$, $\mathbf{k} < \mathbf{l}$, in their interior. The right-hand side term in (1.51), as usual in multigrid terminology (see, for example, (Hackbusch 1985)), denotes a d-dimensional stencil which gives the coefficients for a linear combination of nodal values of its argument u.

A straightforward calculation using partial integration twice and the product structure of (1.51), see (Bungartz and Griebel 2004) for details, gives the integral representation

$$v_{\mathbf{l},\mathbf{i}} \;=\; \int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \cdot D^{\mathbf{2}} u(\mathbf{x})\; \mathrm{d}\mathbf{x} \tag{1.52}$$

for any coefficient value $v_{\mathbf{l},\mathbf{i}}$ of the hierarchical representation (1.47) of $u \in X_0^{q,2}(\bar{\Omega})$. Here $\psi_{l_j,i_j}(x_j) := -2^{-(l_j+1)} \cdot \phi_{l_j,i_j}(x_j)$, and furthermore $\psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) := \prod_{j=1}^{d} \psi_{l_j,i_j}(x_j)$.

Starting from (1.52), we are now able to give bounds for the hierarchical coefficients with respect to the seminorm introduced in (1.33). For the detailed proof see again e.g., (Bungartz and Griebel 2004). We obtain

$$|v_{\mathbf{l},\mathbf{i}}| \;\leq\; 2^{-d} \cdot 2^{-2\cdot|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},\infty}. \tag{1.53}$$

We are now ready to state the following Lemma.

Lemma 1 *Let $u \in X_0^{q,2}(\bar{\Omega})$ be given in its hierarchical representation (1.47). Then, the following estimate holds for its components $u_{\mathbf{l}} \in W_{\mathbf{l}}$:*

$$||u_{\mathbf{l}}||_E \;\leq\; \frac{1}{2 \cdot 12^{(d-1)/2}} \cdot 2^{-2\cdot|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^{d} 2^{2\cdot l_j} \right)^{1/2} \cdot |u|_{\mathbf{2},\infty}. \tag{1.54}$$

Proof Note that the supports of all $\phi_{\mathbf{l},\mathbf{i}}$ contributing to $u_{\mathbf{l}}$ according to (1.47) are mutually disjoint. Then

$$||u_{\mathbf{l}}||_E^2 = \left\| \sum_{\mathbf{i}\in\mathbf{I_l}} v_{\mathbf{l},\mathbf{i}} \cdot \phi_{\mathbf{l},\mathbf{i}} \right\|_E^2 = \sum_{\mathbf{i}\in\mathbf{I_l}} |v_{\mathbf{l},\mathbf{i}}|^2 \cdot ||\phi_{\mathbf{l},\mathbf{i}}||_E^2$$

$$\leq \sum_{\mathbf{i}\in\mathbf{I}_\mathbf{l}} \frac{1}{4^d} \cdot 2^{-4\cdot|\mathbf{l}|_1} \cdot |u|^2_{\mathbf{2},\infty} \cdot 2 \cdot \left(\frac{2}{3}\right)^{d-1} \cdot 2^{-|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2\cdot l_j}\right)$$

$$= \frac{1}{2\cdot 6^{d-1}} \cdot 2^{-5\cdot|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2\cdot l_j}\right) \cdot \sum_{\mathbf{i}\in\mathbf{I}_\mathbf{l}} |u|^2_{\mathbf{2},\infty}$$

$$= \frac{1}{4\cdot 12^{d-1}} \cdot 2^{-4\cdot|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2\cdot l_j}\right) \cdot |u|^2_{\mathbf{2},\infty}.$$

This shows (1.54). □

1.4.2 Energy-norm based sparse grids

We will now construct finite-dimensional approximation spaces U for $V^{(d)}$ or $X_0^{q,2}(\bar{\Omega})$, respectively. Such a U is based on a subspace selection $\mathbf{I} \subset \mathbb{N}^d$,

$$U := \bigoplus_{\mathbf{l}\in\mathbf{I}} W_\mathbf{l}\,, \tag{1.55}$$

with corresponding interpolants

$$u_U := \sum_{\mathbf{l}\in\mathbf{I}} u_\mathbf{l}\,, \qquad u_\mathbf{l} \in W_\mathbf{l}\,. \tag{1.56}$$

The estimate

$$\|u - u_U\| = \left\|\sum_\mathbf{l} u_\mathbf{l} - \sum_{\mathbf{l}\in\mathbf{I}} u_\mathbf{l}\right\| \leq \sum_{\mathbf{l}\notin\mathbf{I}} \|u_\mathbf{l}\| \leq \sum_{\mathbf{l}\notin\mathbf{I}} b(\mathbf{l})\cdot|u| \tag{1.57}$$

will then allow the evaluation of the approximation space U with respect to a norm $\|\cdot\|$ and a corresponding seminorm $|\cdot|$ on the basis of the bounds from above indicating the *benefit* $b(\mathbf{l})$ of $W_\mathbf{l}$.

1.4.2.1 Construction of subspaces by optimization

We now address the question how to determine optimal subspace index sets $\mathbf{I}$ which optimize cost versus accuracy for the interpolation error for functions u from $X_0^{q,2}$. To this end, we look for an optimum $V^{(\mathrm{opt})}$ by solving a restricted optimization problem of the type

$$\max_{u\in X_0^{q,2}:\ |u|=1} \|u - u_{V^{(\mathrm{opt})}}\| = \min_{U\subset V^{(d)}:\ |U|=w} \max_{u\in X_0^{q,2}:\ |u|=1} \|u - u_U\| \tag{1.58}$$

for some prescribed work count w. The aim is to profit from a given work count as much as possible.† Of course, any potential solution $V^{(\mathrm{opt})}$ of (1.58) has to be expected to depend on the norm $\|\cdot\|$ as well as on the seminorm $|\cdot|$ used to measure the error of u's interpolant $u_U \in U$ or the smoothness of u, respectively. Note that this *a-priori optimization* strategy depends only on the problem class (i.e., on the space u has to belong to – here $X_0^{q,2}(\bar{\Omega})$), but not on u itself.‡

According to our hierarchical setting, we will allow discrete spaces of the type $U := \bigoplus_{\mathbf{l}\in\mathbf{I}} W_{\mathbf{l}}$ for an arbitrary finite index set $\mathbf{I} \subset \mathbb{N}^d$ as candidates for the optimization process only. Now, an approach like (1.58) selects certain $W_{\mathbf{l}}$ due to their importance and, thus, selects the respective grids and the underlying index sets $\mathbf{I} \subset \mathbb{N}^d$. This is done by using techniques known from combinatorial optimization as follows:

For the following, a grid and its representation $\mathbf{I}$ – formerly a finite set of multi-indices – is nothing but a bounded subset of $\mathbb{N}_+^d$, and a hierarchical subspace $W_{\mathbf{l}}$ just corresponds to a point $\mathbf{l} \in \mathbb{N}_+^d$. First, we have to reformulate the optimization problem (1.58). We define the local functions $c(\mathbf{l})$ and $b(\mathbf{l})$, for the multi-indices $\mathbf{l} \in \mathbb{N}^d$. According to (1.48), the local cost $c(\mathbf{l})$ is

$$c(\mathbf{l}) \;:=\; |W_{\mathbf{l}}| \;=\; 2^{|\mathbf{l}-\mathbf{1}|_1}. \tag{1.59}$$

Obviously, $c(\mathbf{l}) \in \mathbb{N}$ holds for all $\mathbf{l} \in \mathbb{N}^d$. Concerning the local benefit $b(\mathbf{l})$, we define

$$b(\mathbf{l}) \;:=\; \gamma \cdot \beta(\mathbf{l})\,, \tag{1.60}$$

where $\beta(\mathbf{l})$ is an upper bound for $\|u_{\mathbf{l}}\|^2$ according to (1.54), and γ is a factor which depends on the problem's dimensionality d and on the smoothness of the data, i.e., of u, but which is constant with respect to $\mathbf{l}$, such that $b(\mathbf{l}) \in \mathbb{N}$. The bound in (1.54) shows that such a choice of γ is indeed possible. At the moment, we do not yet fix the norm to be used here.

Now, the search for an optimal grid $\mathbf{I} \subset \mathbb{N}^d$ can be restricted to all $\mathbf{I} \subset \mathbf{I}^{(\max)} := \{1,\ldots,N\}^d$ for a sufficiently large N without loss of generality. Next, global cost and benefit functions $C(\mathbf{I})$ and $B(\mathbf{I})$ are to

† Note that an optimization the other way round could be done as well: Prescribe some desired accuracy ε and look for the discrete approximation scheme that achieves this with the smallest possible work count. This is in fact the point of view of computational complexity.

‡ This is in contrast to adaptive grid refinement which uses a-posteriori error estimators to approximate one given function u.

be defined. For $C(\mathbf{I})$, we set

$$C(\mathbf{I}) \;:=\; \sum_{\mathbf{l}\in\mathbf{I}} c(\mathbf{l}) \;=\; \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} x(\mathbf{l})\cdot c(\mathbf{l})\,, \tag{1.61}$$

where

$$x(\mathbf{l}) \;:=\; \begin{cases} 0 & : \quad \mathbf{l}\notin\mathbf{I}, \\ 1 & : \quad \mathbf{l}\in\mathbf{I}. \end{cases} \tag{1.62}$$

The interpolant to u on a grid $\mathbf{I}$ provides the global benefit $B(\mathbf{I})$:

$$\begin{aligned} \left\| u \;-\; \sum_{\mathbf{l}\in\mathbf{I}} u_{\mathbf{l}} \right\|^2 \;&\approx\; \left\| \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} u_{\mathbf{l}} \;-\; \sum_{\mathbf{l}\in\mathbf{I}} u_{\mathbf{l}} \right\|^2 \\ &\le \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}\setminus\mathbf{I}} \|u_{\mathbf{l}}\|^2 \\ &\le \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} (1-x(\mathbf{l}))\cdot\gamma\cdot\beta(\mathbf{l}) \\ &= \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} \gamma\cdot\beta(\mathbf{l}) \quad - \quad \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} x(\mathbf{l})\cdot\gamma\cdot\beta(\mathbf{l}) \\ &=: \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} \gamma\cdot\beta(\mathbf{l}) \quad - \quad B(\mathbf{I})\,. \end{aligned} \tag{1.63}$$

Of course, (1.63) gives only an upper bound for an approximation to the (squared) interpolation error, because it does not take into account all $\mathbf{l}\notin\mathbf{I}^{(\max)}$. However, since N and, consequently, $\mathbf{I}^{(\max)}$ can be chosen to be as large as is appropriate, this is not a serious restriction. Altogether, we get the following reformulation of (1.58):

$$\begin{aligned} &\max_{\mathbf{I}\subset\mathbf{I}^{(\max)}} B(\mathbf{I}) \quad \text{with} \quad C(\mathbf{I}) = w, \quad \text{i.e.,} \\ &\max_{\mathbf{I}\subset\mathbf{I}^{(\max)}} \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} x(\mathbf{l})\cdot\gamma\cdot\beta(\mathbf{l}) \quad \text{with} \quad \sum_{\mathbf{l}\in\mathbf{I}^{(\max)}} x(\mathbf{l})\cdot c(\mathbf{l}) \;=\; w\,. \end{aligned} \tag{1.64}$$

If we arrange the $\mathbf{l}\in\mathbf{I}^{(\max)}$ in some linear order (a lexicographical one, for instance) with local cost c_i and benefit b_i, $i=1,\ldots,N^d =: M$, (1.64) results in

$$\max_{\mathbf{x}} \; \mathbf{b}^T\mathbf{x} \qquad \text{with} \qquad \mathbf{c}^T\mathbf{x} \;=\; w\,, \tag{1.65}$$

where $\mathbf{b}\in\mathbb{N}^M$, $\mathbf{c}\in\mathbb{N}^M$, $\mathbf{x}\in\{0,1\}^M$, and, without loss of generality, $w\in\mathbb{N}$. In combinatorial optimization, a problem like (1.65) is called a *binary knapsack problem* (Martello and Toth 1990), which is known to be NP-hard. However, a slight change makes things much easier. If

rational solutions, i.e., $\mathbf{x} \in ([0,1] \cap \mathbb{Q})^M$, are allowed, too, there exists a very simple algorithm that provides an optimal solution vector $\mathbf{x} \in ([0,1] \cap \mathbb{Q})^M$:

(i) rearrange the order that $\frac{b_1}{c_1} \geq \frac{b_2}{c_2} \cdots \geq \frac{b_M}{c_M}$,

(ii) let $r := \max \left\{ j : \sum_{i=1}^{j} c_i \leq w \right\}$,

(iii) $x_1 := \cdots := x_r := 1$,
$x_{r+1} := \left(w - \sum_{i=1}^{r} c_i \right) / c_{r+1}$,
$x_{r+2} := \cdots := x_M := 0$.

Although there is only one potential non-binary coefficient x_{r+1}, the rational solution vector $\mathbf{x}$, generally, has nothing to do with its binary counterpart. But, fortunately, our knapsack is of variable size, since the global work count w is an arbitrarily chosen natural number. Therefore, it is possible to force the solution of the *rational* problem to be a *binary* one which is, of course, also a solution of the corresponding *binary* problem. Consequently, the *global* optimization problem (1.58) or (1.65), respectively, can be reduced to the discussion of the *local* cost-benefit ratios b_i/c_i or $b(\mathbf{l})/c(\mathbf{l})$ of the underlying subspaces $W_{\mathbf{l}}$. Those subspaces with the best cost-benefit ratios are taken into account first, and the smaller these ratios become, the more negligible the underlying subspaces turn out to be.

Now, if our cost-benefit approach is based on the L_p-norms, with $p \in \{1, \infty\}$ we showed in (Bungartz and Griebel 2004) that this results in the regular sparse grid spaces

$$V_n^{(1)} := \bigoplus_{|\mathbf{l}|_1 \leq n+d-1} W_{\mathbf{l}} \tag{1.66}$$

which have been introduced in (Zenger 1991). Note that they are the finite element analogon of the well-known hyperbolic cross or Korobov spaces which are based on the Fourier series expansion instead of the hierarchical Faber basis. An example of a regular sparse grid is given for the two- and three-dimensional case in Figure 1.3. The basic concept can be traced back to (Smolyak 1963, Babenko 1960), see also (Gordon 1969, Gordon 1971, Delvos 1982, Delvos and Schempp 1989, DeVore, Konyagin and Temlyakov 1998).

The dimension of the space $V_n^{(1)}$ fulfills

$$|V_n^{(1)}| \;=\; O(h_n^{-1} \cdot |\log_2 h_n|^{d-1}) \tag{1.67}$$

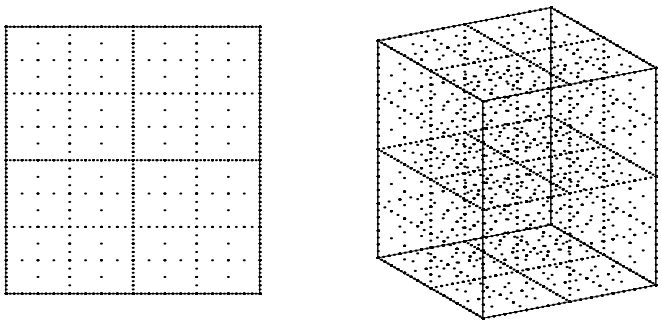

Fig. 1.3. L_p-norm based sparse grids: Two-dimensional example (left) and three-dimensional example (right), here including sparse grid points on the boundary.

with $h_n = 2^{-n}$, whereas, for the interpolation error of a function $u \in X_0^{q,2}(\bar{\Omega})$ in the sparse grid space $V_n^{(1)}$ there holds

$$||u - u_n^{(1)}||_p \;=\; O(h_n^2 \cdot n^{d-1}), \tag{1.68}$$

for the L_p-norms, and

$$||u - u_n^{(1)}||_E \;\leq\; \frac{d \cdot |u|_{\mathbf{2},\infty}}{2 \cdot 3^{(d-1)/2} \cdot 4^{d-1}} \cdot 2^{-n} \;=\; O(h_n), \tag{1.69}$$

for the energy-norm, see for example (Bungartz and Griebel 2004) for detailed proofs. Note that the conventional full grid space

$$V_n^{(\infty)} \;:=\; \bigoplus_{|\mathbf{l}|_\infty \leq n} W_\mathbf{l}$$

results in an error in the L_p-norm of the order $O(h_n^2)$ and an error in the energy-norm of the order $O(h_n)$. It however possesses a dimension $|V_n^{(\infty)}| \;=\; O(h_n^{-d})$ and thus exhibits the curse of dimensionality with respect to h_n. In comparison to that we now see a crucial improvement for $V_n^{(1)}$: The number of degrees of freedom is reduced significantly, whereas the accuracy deteriorates only slightly for the L_p-norm and stays of the same order for the energy-norm. The curse of dimensionality is now present in the $\log(h_n)$-term only. Since this result is optimal with respect to the L_p-norms, a further improvement can only be expected if we change the setting. Therefore, in the following, we turn to the energy norm.

1.4.2.2 Energy-based sparse grids

We now base our cost-benefit approach on the energy norm. According to (1.48) and (1.54) and following the discussion in Section 1.4.2.1, we define

$$\begin{aligned} cbr_E(\mathbf{l}) \;:=\; \frac{b_E(\mathbf{l})}{c(\mathbf{l})} \;&:=\; \frac{2^{-4\cdot|\mathbf{l}|_1}\cdot|u|^2_{\mathbf{2},\infty}}{4\cdot 12^{d-1}\cdot 2^{|\mathbf{l}-\mathbf{1}|_1}}\cdot\sum_{j=1}^{d}4^{l_j} \\ &=\; \frac{3}{6^d}\cdot 2^{-5\cdot|\mathbf{l}|_1}\cdot\sum_{j=1}^{d}4^{l_j}\cdot|u|^2_{\mathbf{2},\infty} \end{aligned} \tag{1.70}$$

as the local cost-benefit ratio. Note that, instead of $\|u_{\mathbf{l}}\|_E$ itself, only an upper bound for the squared energy norm of $u_{\mathbf{l}}$ is used. The resulting optimal grid $\mathbf{I}^{(\mathrm{opt})}$ will consist of all those multi-indices $\mathbf{l}$ or their respective hierarchical subspaces $W_{\mathbf{l}}$ that fulfill $cbr_E(\mathbf{l}) \geq \sigma_E(n)$ for some given constant threshold $\sigma_E(n)$. Here, $\sigma_E(n)$ is defined via the cost-benefit ratio of $W_{\bar{\mathbf{l}}}$ with $\bar{\mathbf{l}} := (n, 1, \ldots, 1)$:

$$\sigma_E(n) \;:=\; cbr_E(\bar{\mathbf{l}}) \;=\; \frac{3}{6^d}\cdot 2^{-5\cdot(n+d-1)}\cdot\big(4^n+4\cdot(d-1)\big)\cdot|u|^2_{\mathbf{2},\infty}\,. \tag{1.71}$$

Thus, applying the criterion $cbr_E(\mathbf{l}) \geq \sigma_E(n)$, we come to a sparse grid approximation space $V_n^{(E)}$ which is based on the energy norm:

$$V_n^{(E)} \;:=\; \bigoplus_{|\mathbf{l}|_1-\frac{1}{5}\cdot\log_2\left(\sum_{j=1}^{d}4^{l_j}\right)\;\leq\;(n+d-1)-\frac{1}{5}\cdot\log_2(4^n+4d-4)} W_{\mathbf{l}}\,. \tag{1.72}$$

For a comparison of the underlying subspace schemes of $V_n^{(1)}$ and $V_n^{(E)}$ in two dimensions, see Figure 1.4.

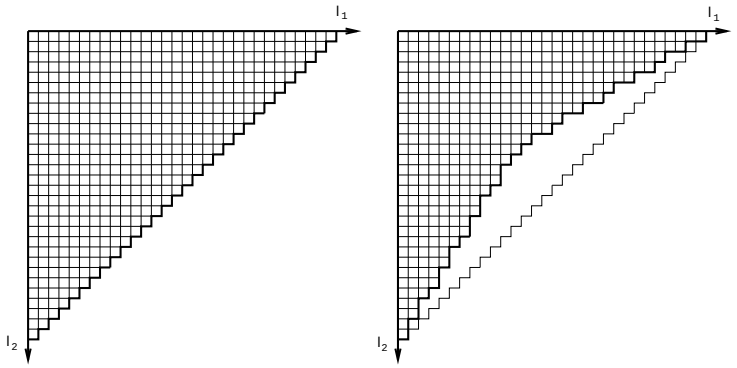

Fig. 1.4. Scheme of subspaces for $V_{30}^{(1)}$ (left) and $V_{30}^{(E)}$ (right), $d = 2$.

First, we look at the number of grid points of the underlying sparse grids.

Lemma 2 *The energy-based sparse grid space $V_n^{(E)}$ is a subspace of $V_n^{(1)}$, and its dimension fulfills*

$$|V_n^{(E)}| \;\le\; 2^n \cdot \frac{d}{2} \cdot \left(1 - 2^{-\frac{2}{3}}\right)^{-d} \;\le\; 2^n \cdot \frac{d}{2} \cdot e^d \;=\; O(h_n^{-1}) \,. \qquad (1.73)$$

Proof For subspaces $W_{\mathbf{l}}$ with $|\mathbf{l}|_1 = n + d - 1 + i,\ i \in \mathbb{N}$, we have

$$\begin{aligned} |\mathbf{l}|_1 - \frac{1}{5} \cdot \log_2 \left(\sum_{j=1}^{d} 4^{l_j} \right) &\ge n + d - 1 + i - \frac{1}{5} \cdot \log_2 \left(4^{n+i} + 4d - 4\right) \\ &\ge n + d - 1 + i - \frac{1}{5} \cdot \log_2 \left(4^i \left(4^n + 4d - 4\right)\right) \\ &> n + d - 1 - \frac{1}{5} \cdot \log_2 \left(4^n + 4d - 4\right) . \end{aligned}$$

Therefore, no $W_{\mathbf{l}}$ with $|\mathbf{l}|_1 > n+d-1$ can belong to $V_n^{(E)}$. Consequently, $V_n^{(E)}$ is a subspace of $V_n^{(1)}$ and $|V_n^{(E)}| \;\le\; |V_n^{(1)}|$ for all $n \in \mathbb{N}$. Starting from that, (1.48) provides

$$\begin{aligned} |V_n^{(E)}| &= \sum_{i=0}^{n-1} \sum_{\substack{|\mathbf{l}|_1 = n+d-1-i, \\ \sum_{j=1}^{d} 4^{l_j} \ge \frac{4^n+4d-4}{32^i}}} |W_{\mathbf{l}}| \\ &= 2^n \cdot \frac{1}{2} \cdot \sum_{i=0}^{n-1} 2^{-i} \cdot \sum_{\substack{|\mathbf{l}|_1 = n+d-1-i, \\ \sum_{j=1}^{d} 4^{l_j} \ge \frac{4^n+4d-4}{32^i}}} 1 \\ &\le 2^n \cdot \frac{1}{2} \cdot \lim_{n\to\infty} \sum_{i=0}^{n-1} 2^{-i} \cdot \sum_{\substack{|\mathbf{l}|_1 = n+d-1-i, \\ \sum_{j=1}^{d} 4^{l_j} \ge \frac{4^n+4d-4}{32^i}}} 1 \\ &= 2^n \cdot \frac{1}{2} \cdot \lim_{n\to\infty} \sum_{i=0}^{n-1} 2^{-i} \cdot d \cdot \binom{d-1-\lfloor 1.5i \rfloor}{d-1} , \end{aligned}$$

since it can be shown that, for $n \to \infty$, our energy-based sparse grid and the grid resulting from the second condition $|\mathbf{l}|_\infty \ge n - \lfloor 2.5i \rfloor$ for the inner sum instead of

$$\sum_{j=1}^{d} 4^{l_j} \ge \frac{4^n + 4d - 4}{32^i}$$

are the same, and since there exist

$$\binom{d-1+\lfloor 1.5i \rfloor}{d-1}$$

such subspaces $W_{\mathbf{l}}$ with $|\mathbf{l}|_\infty = l_1$. Consequently, we obtain

$$\begin{aligned} |V_n^{(E)}| &\leq 2^n \cdot \frac{d}{2} \cdot \sum_{i=0}^{\infty} 2^{-\frac{2}{3}i} \cdot \binom{d-1+i}{d-1} \\ &= 2^n \cdot \frac{d}{2} \cdot \left(1 - 2^{-\frac{2}{3}}\right)^{-d} \\ &\leq 2^n \cdot \frac{d}{2} \cdot e^d \,, \end{aligned}$$

since $\sum_{i=0}^{\infty} x^i \cdot \binom{k+i}{k} = (1-x)^{-k-1}$ for $k \in \mathbb{N}_0$ and $0 < x < 1$. □

Next, we have to deal with the interpolation accuracy of the energy-based sparse grid spaces $V_n^{(E)}$ and to study the sparse grid interpolant $u_n^{(E)} \in V_n^{(E)}$.

Theorem 1 *The energy norm of the interpolation error of some $u \in X_0^{q,2}(\bar{\Omega})$ in the energy-based sparse grid space $V_n^{(E)}$ is bounded by*

$$\|u - u_n^{(E)}\|_E \leq \frac{d \cdot |u|_{\mathbf{2},\infty}}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2}\right)^{d-1}\right) \cdot 2^{-n} \;=\; O(h_n). \quad (1.74)$$

Proof First, since

$$||u - u_n^{(E)}||_E \;\leq\; ||u - u_n^{(1)}||_E \;+\; ||u_n^{(1)} - u_n^{(E)}||_E \,,$$

and since we know that $||u - u_n^{(1)}||_E$ is of the order $O(h_n)$, we can restrict ourselves to $||u_n^{(1)} - u_n^{(E)}||_E$. For that, it can be shown that, for $i \in \mathbb{N}_0$, each $W_{\mathbf{l}}$ with $|\mathbf{l}|_1 = n + d - 1 - i$ and $|\mathbf{l}|_\infty \geq n - 2.5i$ is a subspace of $V_n^{(E)}$. Therefore, we obtain with (1.54)

$$\begin{aligned} ||u_n^{(1)} - u_n^{(E)}||_E &\leq \sum_{W_{\mathbf{l}} \subseteq V_n^{(1)} \ominus V_n^{(E)}} ||u_{\mathbf{l}}||_E \leq \sum_{i=0}^{i^*} \sum_{\substack{|\mathbf{l}|_1 = n+d-1-i, \\ |\mathbf{l}|_\infty < n-2.5i}} ||u_{\mathbf{l}}||_E \\ &\leq \frac{|u|_{\mathbf{2},\infty}}{2 \cdot 12^{(d-1)/2}} \cdot \sum_{i=0}^{i^*} \sum_{\substack{|\mathbf{l}|_1 = n+d-1-i, \\ |\mathbf{l}|_\infty < n-2.5i}} 4^{-|\mathbf{l}|_1} \cdot \Big(\sum_{j=1}^{d} 4^{l_j}\Big)^{1/2} \\ &\leq \frac{|u|_{\mathbf{2},\infty}}{2 \cdot 12^{(d-1)/2}} \cdot 4^{-n-d+1} \cdot \sum_{i=0}^{i^*} 4^i \cdot \sum_{\substack{|\mathbf{l}|_1 = n+d-1-i, \\ |\mathbf{l}|_\infty < n-2.5i}} \Big(\sum_{j=1}^{d} 2^{l_j}\Big) \end{aligned}$$

$$\leq \frac{|u|_{\mathbf{2},\infty}}{2 \cdot 12^{(d-1)/2}} \cdot 4^{-n-d+1} \cdot \sum_{i=0}^{i^*} 4^i \cdot \sum_{j=1}^{n-1-\lfloor 2.5i \rfloor} d \cdot \binom{n+d-2-i-j}{d-2} \cdot 2^j$$

$$= \frac{|u|_{\mathbf{2},\infty}}{2 \cdot 12^{(d-1)/2}} 4^{-n-d+1} \sum_{i=0}^{i^*} 4^i \sum_{k=1}^{n-1-\lfloor 2.5i \rfloor} d \binom{d-2+\lfloor 1.5i \rfloor + k}{d-2} 2^{n-\lfloor 2.5i \rfloor - k}$$

$$= \frac{d \cdot |u|_{\mathbf{2},\infty}}{2 \cdot 12^{(d-1)/2}} 4^{-(d-1)} 2^{-n} \sum_{i=0}^{i^*} 2^{-\lfloor \frac{i}{2} \rfloor} \sum_{k=1}^{n-1-\lfloor 2.5i \rfloor} \binom{d-2+\lfloor 1.5i \rfloor + k}{d-2} 2^{-k}$$

$$\leq \frac{d \cdot |u|_{\mathbf{2},\infty}}{2 \cdot 12^{(d-1)/2}} \cdot 4^{-(d-1)} \cdot 2^{-n} \cdot 2 \cdot 5^{d-1}$$

$$= \frac{d \cdot |u|_{\mathbf{2},\infty}}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{5}{2}\right)^{d-1} \cdot 2^{-n},$$

where $0 \leq i^* \leq n-1$ is the maximum value of i for which the set of indices $\mathbf{l}$ with $|\mathbf{l}|_1 = n+d-1-i$ and $|\mathbf{l}|_\infty < n - 2.5i$ is not empty. Together with (1.69) we get the result. □

The crucial result of this section is that, with the energy-based sparse grid spaces $V_n^{(E)}$, the curse of dimensionality can be overcome. In both (1.73) and (1.74), the n-dependent terms are free of any d-dependencies: There is an order of $O(2^n)$ for the dimension and $O(2^{-n})$ for the interpolation error. In particular, there is no longer any polynomial term in n like n^{d-1} as for the case of the space $V_1^{(1)}$. That is, apart from the factors that are constant with respect to n, there is no d-dependence in either $|V_n^{(E)}|$ or $||u - u_n^{(E)}||_E$ and, thus, no deterioration in complexity for higher dimensional problems. The curse of dimensionality has thus been completely overcome, at least with respect to n. However the constants in the order estimates are still dependent on the dimension d. This will be studied in more detail in the following section.

1.4.3 The constants and their dependence on d

So far, we derived the estimate

$$|V_n^{(E)}| \leq c_1(d) \cdot 2^n \tag{1.75}$$

for the degrees of freedom of the sparse grid spaces $V_n^{(E)}$ with the constant

$$c_1(d) = \frac{d}{2}(1 - 2^{-2/3})^{-d}$$

and the estimate

$$\|u - u_n^{(E)}\|_E \leq c_2(d) \cdot 2^{-n} \cdot |u|_{\mathbf{2},\infty} \;=\; O(h_n)\,, \tag{1.76}$$

for the accuracy of the achieved interpolation error with the constant

$$c_2(d) = \frac{d}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2} \right)^{d-1} \right).$$

Note that the upper bound (1.69) for the interpolant $u_n^{(E)}$ of u in $V_n^{(E)}$ also gives by virtue of Cea's lemma an upper bound for the best approximation in $V_n^{(E)}$. We thus have

$$\inf_{v_n \in V_n^{(E)}} \|u - v_n\|_E \leq \|u - u_n^{(E)}\|_E.$$

We are interested in casting these results in a form which is more common in approximation theory, i.e., we want to express the bound for the approximation error in terms of the amount of degrees of freedom involved. To this end we define the number of degrees of freedom for the best approximation of u in $V_n^{(E)}$ as

$$M := |V_n^{(E)}|. \tag{1.77}$$

We then express the estimate of the approximation error in terms of M.

Theorem 2 *For the best approximation of a function $u \in X_0^{q,2}(\bar{\Omega})$ in the space $V_n^{(E)}$ with respect to the energy norm, there holds*

$$\inf_{v_n \in V_n^{(E)}} \|u - v_n\|_E \leq \|u - u_n^{(E)}\|_E \leq c \cdot d^2 \cdot 0.97515^d \cdot M^{-1} \cdot |u|_{\mathbf{2},\infty}. \tag{1.78}$$

Proof First, with the definition (1.77) we solve (1.75) for 2^n. Taking the inverse we obtain $2^{-n} \leq c_1(d) \cdot M^{-1}$. Then

$$\|u - u_n^{(E)}\|_E \leq c_2(d) \cdot 2^{-n} \cdot |u|_{\mathbf{2},\infty} \leq c_1(d) \cdot c_2(d) \cdot M^{-1} \cdot |u|_{\mathbf{2},\infty}.$$

With

$$\begin{aligned} c_1(d) \cdot c_2(d) &= \frac{d}{2}(1 - 2^{-2/3})^{-d} \cdot \frac{d}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2} \right)^{d-1} \right) \\ &= \sqrt{3} \cdot d^2 \left(\frac{1}{(1 - 2^{-2/3}) \cdot \sqrt{3} \cdot 4} \right)^d \\ &+ \frac{4 \cdot \sqrt{3}}{5} \cdot d^2 \left(\frac{5}{(1 - 2^{-2/3}) \cdot \sqrt{3} \cdot 8} \right)^d \\ &\leq \sqrt{3} \cdot d^2 \cdot 0.39901^d + \frac{4\sqrt{3}}{5} \cdot d^2 \cdot 0.97515^d \end{aligned} \tag{1.79}$$

the estimate (1.78) results. □

Thus we see that we obtain for the constant a decay for $d \to \infty$ to zero.† The term $|u|_{\mathbf{2},\infty}$ however is also dependent on d and may grow exponentially with it. Already for the simple example $u(\mathbf{x}) = \prod_{j=1}^{d} \sin(2\pi k x_j)$ we see that $|u|_{\mathbf{2},\infty} = (4\pi^2 k^2)^d$ grows faster than 0.97515^d decays. Obviously it is sufficient to restrict ourselves to the approximation of functions $u \in X_0^{q,2}$ with $|u|_{\mathbf{2},\infty} = o(1/(d^2 \cdot 0.97515^d))$ to insure that $\|u - u_n^{(E)}\|_E$ is bounded for all d. But it is not clear how interesting this function class for large d is in practice. Nevertheless, the facts we know from the concentration of measure phenomenon (i.e., that the best approximation in very high dimensions is nearly constant) gives hope in this direction.

Note that if we use the seminorm

$$|u|_{\mathbf{2},2} \;:=\; \left\|D^{\mathbf{2}} u\right\|_2 \;=\; \left(\int_{\bar{\Omega}} \left|D^{\mathbf{2}} u\right|^2 \,\mathrm{d}\mathbf{x}\right)^{1/2}.$$

instead and rewrite all the above lemmata, theorems and their proofs in terms of the associated regularity assumption $|u|_{\mathbf{2},2}$ we are no longer able to derive a favorable estimate as in (1.79). We obtain slightly worse estimates where, for the c_1 and c_2 involved, we only get $c_1(d) \cdot c_2(d) \leq c \cdot \sqrt{3} \cdot d^2 \cdot 0.45041^d + \sqrt{3}\frac{4}{5} \cdot d^2 \cdot 1.12601^d$. Thus we see a blow-up to infinity for $d \to \infty$ for these estimates. Since the corresponding c_1 and c_2 are only upper bounds of the true d-dependent constants it is not clear if this also holds for them or not. Note furthermore that a rescaling of the size of the domain Ω of course also influences the constants c_1 and c_2 which has to be taken into account in the above discussion.

Let us finally consider the case of non-homogeneous boundary conditions, i.e., u from the space $X^{q,2}$. Now, to capture also functions living on the boundary of $\bar{\Omega}$, we generalize the two-scale splitting (1.11) to a three-scale decomposition $V^{(1)} = \mathbf{1} \oplus \mathbf{lin} \oplus \tilde{W}$ where $\mathbf{1}$ denotes the subspace of constants, $\mathbf{lin}$ denotes the subspace of linear functions (without the constants) and $\tilde{W}$ denotes the remainder, respectively. Note that $\mathbf{1} \oplus \mathbf{lin}$ is just the kernel of the second derivative. This augmented splitting is now used as the input of a tensor product construction to gain a splitting of the function space for the d-dimensional case. Analogously to (1.17) and (1.18) a decomposition of the d-dimensional space into now 3^d subspaces is introduced by $X^{q,2} \doteq V^{(d)} = \bigotimes_{j=1}^{d} (\mathbf{1}_j \oplus \mathbf{lin}_j \oplus \tilde{W}_j)$ and we can repeat the discussion of Section 1.3.1 in a similar way for the refined decomposition. Informally speaking, the space $X^{q,2}$ can then

† Note that this holds for the asymptotics with respect to M, i.e., the estimates for c_1 and c_2 were done for asymptotically large n.

be decomposed into $X^{q,2} \setminus X_0^{q,2}$ and $X_0^{q,2}$ consistent with this refined decomposition, and we can split a function $u \in X^{q,2}$ accordingly into

$$u = \tilde{u} + v \quad \text{where} \quad \tilde{u} \in X^{q,2} \setminus X_0^{q,2} \quad \text{and} \quad v \in X_0^{q,2}.$$

The regularity condition $|u|_{\mathbf{2},\infty} \leq c < \infty$ translates to $\|D^{\mathbf{2}}u\|_\infty \leq c < \infty$, with some d-dependent constant c. For the term $D^{\mathbf{2}}u$ we obtain

$$D^{\mathbf{2}}u = \frac{\partial^{2d}u}{\partial x_1^2 \cdots \partial x_d^2} = \frac{\partial^{2d}(\tilde{u}+v)}{\partial x_1^2 \cdots \partial x_d^2} = \frac{\partial^{2d}v}{\partial x_1^2 \cdots \partial x_d^2}$$

since $\frac{\partial^{2d}\tilde{u}}{\partial x_1^2 \cdots \partial x_d^2}$ vanishes due to the involved constant and linear subspaces. We are thus just in the situation of homogeneous boundary conditions as treated previously and the lemmata and theorems above apply. Note that the more general assumption $D^{\boldsymbol{\alpha}}u \in L_\infty, |\boldsymbol{\alpha}|_\infty \leq 2$, from (1.32) relates to the (second) variation of Hardy and Krause (Owen 2004) and involves in a dimension-recursive way different partial (mixed) derivatives up to second order of the contributions of u from the various boundary manifolds of $\bar{\Omega}$.

In an analogous way, we enlarge our basis function set to also capture functions living on the boundary of $\bar{\Omega}$. To this end we introduce two more functions into the one-dimensional hierarchical basis from Figure 1.2 which are associated to the left and right boundary point of $[0,1]$ and number their associated level l by -1 for the left point and 0 for the right point, respectively. As basis functions we use the constant function $\phi_{-1,1}(x) := 1$ for the left boundary point and the linear function $\phi_{0,1}(x) = x$ for the right boundary point. Then, $\{\phi_{-1,1}, \phi_{0,1}\}$ just spans the subspace of constant and linear functions.† This augmented system of basis functions is now used as the input of the tensor product construction (1.39) to gain a function system for the d-dimensional case. Moreover, analogously to (1.45) and (1.47), a function $u \in X^{q,2}$ can now be represented as $u(\mathbf{x}) = \sum_{\mathbf{l} \in (\mathbb{N} \cup \{-1,0\})^d} u_{\mathbf{l}}(\mathbf{x})$ and the space $X^{q,2}$ gets decomposed as $\bigotimes_{\mathbf{l} \in (\mathbb{N} \cup \{-1,0\})^d} W_{\mathbf{l}}$.‡

Our approach so far was focused on the space of bounded second mixed derivatives, the energy-norm as measure for the approximation error and (piecewise) linear hierarchical basis functions. It can be carried over to a more general setting where we measure the approximation error in the H^s-norm, $s \in (-\infty, \infty)$, assume a smoothness of the type $|u|_{H^{l,t}_{mix}}$, where

† We could also have used the linear function $1 - x$ at the left boundary point instead. Here, we use the constant one to be completely in sync with our splitting $V = \mathbf{1} \oplus W$ from (1.11).

‡ $X^{q,2} \setminus X_0^{q,2}$ is then (up to completion) just $\sum_{\mathbf{l} \in (\mathbb{N} \cup \{-1,0\})^d \setminus \mathbb{N}^d} W_{\mathbf{l}}$.

l denotes isotropic and t mixed smoothness, see (Griebel and Knapek 2000), and use wavelet-type multilevel systems with sufficient primal and dual regularity. Then, depending on these additional parameters, we can again derive optimal discrete approximation spaces, we may study their cost complexities and approximation properties for different regimes of s, l, t and we can identify situations where the curse of dimensionality can be broken. The approach is based on norm-equivalences and associated wavelet-type multilevel systems. This is explained in more detail in (Griebel and Knapek 2000, Knapek 2000*a*), see also (Knapek 2000*b*) for a variant using Fourier bases. Since the constants in these norm-equivalences depend on d, the constants in the resulting error estimates also depend on d and cannot, in contrast to our approach in Section 1.4.3, be estimated explicitly.

Another generalization of the sparse grid concept uses optimization not with respect to a whole class of functions involving error norm and smoothness prerequisite (a-priori knowledge) but, in the interpolation context, with respect to one single given function or, alternatively, in the context of PDEs, with respect to a given right-hand side or other data in the partial differential equation. This leads with proper a-posteriori error estimators to an adaptively refined sparse grid which adapts itself (hopefully in an optimal way) to the specific situation. The adaption and refinement process can be performed on the level of the subspaces $W_{\mathbf{l}}$ from (1.41). This leads to a so-called dimension-adaptive method for sparse grids, see (Gerstner and Griebel 2003). This approach is well suited for high-dimensional functions and detects important and unimportant dimensions and groups of dimensions of an ANOVA-decomposition in an automatic way (provided that the error indicators are sound and no premature termination of the adaption process occurs). The method was developed and used so far for integration problems, its application to partial differential equations is future work. Alternatively, the adaption and refinement process can be performed on the level of the single basis functions $\phi_{\mathbf{l},\mathbf{i}}$. We then obtain a method where, besides the detection of important and unimportant dimensions and groups of dimensions, singularities and similar local variations in a function are additionally found and resolved. Here, the development of sound local error estimators (via the dual problem), efficient refinement strategies and the associated complexities with respect to d are an area of active research (Griebel 1998, Bungartz and Griebel 2004, Bungartz 1998, Schneider 2000). Figure 1.5 gives examples of two- and three-dimensional adaptive sparse grids.

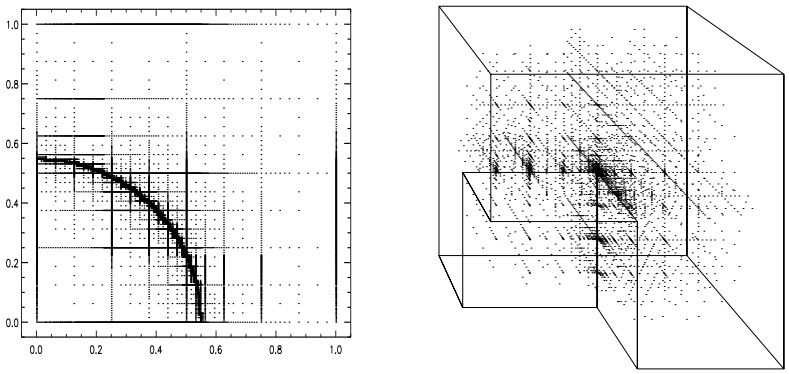

Fig. 1.5. Adaptively refined sparse grids: Two-dimensional example (left) and three-dimensional example (right).

1.5 Concluding remarks

We reviewed approximation techniques which have the potential to overcome the curse of dimensionality which is a main obstacle in the numerical treatment of most high-dimensional problems. After a survey on methods stemming from Kolmogorov's theorem, we focused on the ANOVA decomposition and the sparse grid approach and discussed their properties and prerequisites. Moreover, we presented energy-norm based sparse grids and demonstrated that, for functions with bounded mixed second derivatives on the unit hypercube, the associated approximation rate in terms of the involved degrees of freedom shows no dependence on the dimension at all, neither in the approximation order nor in the order constant. Important ingredients were the product structure of the underlying domain in high dimensions, a one-dimensional splitting of the space into the constant and the remainder subspaces and, as a refinement of the remainder subspace, a one-dimensional multilevel basis. Then a tensor-product approach leads to both, the ANOVA decomposition and a multilevel series expansion of the underlying function. Proper truncation may result in an ANOVA decomposition with finite order weights or, if the more elaborate one-dimensional multiscale splitting is employed, in sparse grids. In this sense sparse grids are closely related to the ANOVA approach and can be seen as a discretization and refined version of it. In the case of functions with low effective dimension or alternatively, bounded mixed second derivatives, the curse of dimensionality indeed can be broken.

These techniques can, together with a Galerkin approach or with a finite difference method, be applied successfully to higher-dimensional partial differential equations. Areas of actual research are here, besides elliptic partial differential equations, also parabolic problems, like e.g., the Fokker-Planck equation with many practical applications in the natural sciences and financial engineering, ranging from the modelling of mechanical systems with random oscillations to the pricing of financial derivatives. Also the Schrödinger equation is of utmost interest. Here, in (Griebel and Hamaekers 2006) we developed antisymmetric sparse grid spaces to cope with the antisymmetry condition stemming from Pauli's principle, see also (Yserentant 2004). Further actual work is the use of sparse grids in space-time (Griebel, Oeltz and Vassilevski 2005). There a product structure between space and time exists naturally and can be exploited.

To reach higher space dimensions the constants in the complexities must be kept as low as possible. Besides the theoretical results on the constants which we presented in the preceding section for approximations in the energy-norm, also the effect of the detailed implementation (data structures, fast solution of the discretized linear systems by e.g., multigrid) on the complexity constants has to be taken into account. Presently we are able to treat elliptic differential equations with up to about 120 dimensions on a modern workstation (provided that there are homogeneous boundary conditions and a product-type right hand side), see (Feuersänger 2005).

Further work has surely to be done to better relate ANOVA-type approaches from high-dimensional integration, data mining and statistics to the solution of partial differential equations by sparse grid techniques, especially with respect to adaptivity. Then certain classes of high-dimensional PDE problems with e.g., lower effective dimension or a decay in the interaction weights of the solution may be detected automatically and treated effectively. Finally there is hope to numerically deal with high-dimensional problems due to the concentration of measure phenomenon.

Acknowledgements: I am grateful to M. Arndt, H. Bungartz, P. Jones, F. Kuo, M. Molenkamp, A. Pinkus, M. Schweitzer, E. Süli, I. Sloan, G. Wasilkowski and H. Woźniakowski for valuable discussions and comments.

Bibliography

G. Allaire (1992), 'Homogenization and two-scale convergence', *SIAM J. Math. Anal.* **21**(6), 1482–1516.

V. Arnold (1957), 'On functions of three variables', *Dokl. Akad. Nauk SSSR* **114**, 679–681. English translation: American Math. Soc. Transl. (2), 28, 51-54, 1963.

V. Arnold (1958), 'On the representation of functions of several variables by superpositions of functions of fewer variables', *Mat. Prosveshchenie* **3**, 41–61.

V. Arnold (1959), 'On the representation of continuous functions of three variables by superpositions of continuous functions of two variables', *Mat. Sb.* **48**, 3–74. English translation: American Math. Soc. Transl. (2), 28, 61-147, 1963.

N. Aronzaijn (1950), 'Theory of reproducing kernels', *Trans. Amer. Math. Soc.* **68**, 337–404.

K. Babenko (1960), 'Approximation by trigonometric polynomials in a certain class of periodic functions of several variables', *Soviet Math. Dokl.* **1**, 672–675. Russian original in Dokl. Akad. Nauk SSSR, 132 (1960), pp. 982–985.

R. Balescu (1997), *Statistical Dynamics, Matter out of Equilibrium*, Imperial College Press, Imperial College, London.

A. Barron (1993), 'Universal approximation bounds for superpositions of a sigmoidal function', *IEEE Transactions on Information Theory* **39**(3), 930–945.

B. Baxter and A. Iserles (2003), On the foundations of computational mathematics, in *Foundations of Computational Mathematics, Special Volume of Handbook of Numerical Analysis Vol. XI* (F. Cucker, ed.), Elsevier Science, pp. 3–34.

M. Beck, A. Jäckle, G. Worth and H. Meyer (2000), 'The multiconfiguration time-dependent Hartree (MCTDH) method: A highly efficient algorithm for propagating wavepackets', *Phys. Reports* **324**, 1–105.

R. Bellmann (1961), *Adaptive Control Processes: A Guided Tour*, Princeton University Press.

G. Beylkin and M. Molenkamp (2002), 'Numerical operator calculus in higher dimensions', *PNAS* **99**(16), 10246–10251.

G. Beylkin and M. Molenkamp (2005), 'Algorithms for numerical analysis in high dimensions', *SIAM Journal on Scientific Computing* **26**(6), 2133–2159.

J. Boyd (2000), *Chebyshev and Fourier spectral methods*, Dover Publications, New York.

H. Bungartz (1992), *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*, Dissertation, Institut für Informatik, TU München.

H. Bungartz (1998), *Finite Elements of Higher Order on Sparse Grids*, Habilitationsschrift, Institut für Informatik, TU München, and Shaker Verlag, Aachen.

H. Bungartz and M. Griebel (1999), 'A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives', *J. Complexity* **15**, 167–199.

H. Bungartz and M. Griebel (2004), 'Sparse grids', *Acta Numerica* **13**, 147–269.

R. Caflisch, W. Morokoff and A. Owen (1997), 'Valuation of mortgage backed

securities using Brownian bridges to reduce effective dimension', *J. Comput. Finance* **1**(1), 27–46.
C. Chauviere and A. Lozinski (2004), 'Simulation of dilute polymer solutions using a Fokker-Planck equation', *Computers and Fluids* **33**, 687–696.
C. Chui and Y. Wang (1992), 'A general framework for compactly supported splines and wavelets', *J. Approx. Theory* **71**, 263–304.
D. Cioranescu, A. Damlamian and G. Griso (2002), 'Periodic unfolding and homogenization', *C. R. Acad. Sci. Paris, Ser. I* **335**, 99–104.
A. Cohen (2003), *Numerical Analysis of Wavelet Methods*, Studies in Mathematics and its Applications, Vol. 32, North Holland.
M. Coppejans (2004), 'On Kolmogorov's representation of functions of several variables by functions of one variable', *Journal of Econometrics* **123**, 1–31.
I. Daubechies (1992), *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Series in Appl. Math. 61, SIAM.
R. de Figueiredo (1980), 'Implications and applications of Kolmogorov's superposition theorem', *IEEE Transactions on Automatic Control* **25**(6), 1227–1231.
F. Delvos (1982), 'd-variate Boolean interpolation', *J. Approx. Theory* **34**, 99–114.
F. Delvos and W. Schempp (1989), *Boolean Methods in Interpolation and Approximation*, Vol. 230 of *Pitman Research Notes in Mathematics*, Longman Scientific and Technical, Harlow.
G. Deslauriers and S. Dubuc (1989), 'Symmetric iterative interpolation processes', *Constr. Approx.* **5**, 49–68.
R. DeVore (1998), 'Nonlinear approximation', *Acta Numerica* **7**, 51–150.
R. DeVore, S. Konyagin and V. Temlyakov (1998), 'Hyperbolic wavelet approximation', *Constr. Approx.* **14**, 1–26.
J. Dick, I. Sloan, X. Wang and H. Woźniakowski (2004), 'Liberating the weights', *Journal of Complexity* **20**(5), 593–623.
D. Donoho (2000), 'High-dimensional data analysis: The curses and blessings of dimensionality'. Aide-Memoire.
D. Donoho and P. Yu (1999), Deslauriers-Dubuc: Ten years after, in *CRM Proceedings and Lecture Notes Vol. 18* (G. Deslauriers and S. Dubuc, eds).
D. Duffie (1996), *Dynamic Asset Pricing Theory*, Princeton University Press.
B. Efron and C. Stein (1981), 'The jackknife estimator of variance', *Annals of Statist.* **9**, 586–596.
J. Elf, P. Lötstedt and P. Sjöberg (2001), Problems of high dimension in molecular biology, in *17th Gamm Seminar Leipzig 2001* (W. Hackbusch, ed.), pp. 1–10.
M. Escobar and L. Seco (2005), A partial differential equation for credit derivaties pricing, in *Workshop High-dimensional Partial Differential Equations in Science and Engineering, August 7-12, CRM, Montreal.*
G. Faber (1909), 'Über stetige Funktionen', *Mathematische Annalen* **66**, 81–94.
C. Feuersänger (2005), Dünngitterverfahren für hochdimensionale elliptische partielle Differentialgleichungen, Diplomarbeit, Institut für Numerische Simulation, Universität Bonn.
B. Fridman (1967), 'An improvement in the smoothness of the functions in Kolmogorov's theorem on superpositions', *Dokl. Akad. Nauk SSSR* **177**, 1019–1022. English translation: Soviet Math. Dokl. (8), 1550-1553,

1967.
J. Friedman (1991), 'Multivariate adaptive regression splines', *Ann. Statist.* **19**(1), 1–141.
J. Friedman and W. Stützle (1981), 'Projection pursuit regression', *J. Amer. Statist. Assoc.* **76**, 817–823.
J. Garcke (2004), Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern, Dissertation, Institut für Numerische Simulation, Universität Bonn.
J. Garcke, M. Griebel and M. Thess (2001), 'Data mining with sparse grids', *Computing* **67**(3), 225–253.
T. Gerstner and M. Griebel (1998), 'Numerical integration using sparse grids', *Numerical Algorithms* **18**, 209–232.
T. Gerstner and M. Griebel (2003), 'Dimension-adaptive tensor-product quadrature', *Computing* **71**, 65–87.
F. Girosi and T. Poggio (1989), 'Representation properties of networks: Kolmogorov's theorem is irrelevant', *Neural Comput.* **1**, 465–469.
F. Girosi, M. Jones and T. Poggio (1995), 'Regularization theory and neural networks architectures', *Neural Computation* **7**, 219–265.
M. Golomb (1959), Approximation by functions of fewer variables, in *On Numerical Approximation* (R. E. Langer, ed.), University of Wisconsin Press, Madison, pp. 275–327.
W. Gordon (1969), Distributive lattices and the approximation of multivariate functions, in *Approximation with Special Emphasis on Spline Functions* (I. Schoenberg, ed.), Academic Press, New York, pp. 223–277.
W. Gordon (1971), 'Blending function methods of bivariate and multivariate interpolation and approximation', *SIAM J. Numer. Anal.* **8**, 158–177.
L. Grasedyck and W. Hackbusch (2003), 'Construction and arithmetics of $\mathcal{H}$-matrices', *Computing* **70**, 295–334.
M. Green, J. Schwarz and E. Witten (1998), *Superstring Theory, Vol. 1 and 2*, Cambridge University Press.
M. Griebel (1994), 'Multilevel algorithms considered as iterative methods on semidefinite systems', *SIAM J. Sci. Stat. Comput.* **15**(3), 547–565.
M. Griebel (1998), 'Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences', *Computing* **61**(2), 151–179.
M. Griebel and J. Hamaekers (2006), 'Sparse grids for the Schrödinger equation', *Mathematical Modelling and Numerical Analysis.* to appear.
M. Griebel and S. Knapek (2000), 'Optimized tensor-product approximation spaces', *Constr. Approx.* **16**(4), 525–540.
M. Griebel and P. Oswald (1994), 'On additive Schwarz preconditioners for sparse grid discretizations', *Numer. Math.* **66**, 449–463.
M. Griebel and P. Oswald (1995*a*), 'On the abstract theory of additive and multiplicative Schwarz algorithms', *Numer. Math.* **70**(2), 161–180.
M. Griebel and P. Oswald (1995*b*), 'Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems', *Adv. Comput. Math.* **4**, 171–206.
M. Griebel, D. Oeltz and P. Vassilevski (2005), Space-time approximation with sparse grids, Preprint UCRL-JRNL-211686, Lawrence Livermore National Laboratory. Submitted to SIAM J. Sci. Comput. Also as Preprint No. 222, SFB 611, University of Bonn.
M. Gromov (1999), *Metric Structures for Riemannian and Non-Riemannian Spaces, Progress in Mathematics 152*, Birkhäuser Verlag.

C. Gu and G. Wahba (1993), 'Semiparametric analysis of variance with tensor product thin plate splines', *Journal of the Royal Statistical Society, Series B* **55**(2), 353–368.

W. Hackbusch (1985), *Multigrid Methods and Applications*, Springer, Berlin/Heidelberg.

W. Hackbusch and B. Khoromskij (2004), Kronecker tensor-product based approximation to certain matrix-valued functions in higher dimensions, Technical Report 16/2004, Max Planck Institute for Mathematics in the Sciences, Leipzig.

W. Hackbusch, B. Khoromskij and S. Sauter (2000), On $\mathcal{H}^2$-matrices, in *Lectures on Applied Mathematics* (H. Bungartz, R. Hoppe and C. Zenger, eds), Springer, Berlin, pp. 9–29.

T. Hastie and R. Tibshirani (1986), 'Generalized additive models (with discussion)', *Statist. Sci.* **1**, 297–318.

T. Hastie and R. Tibshirani (1990), *Generalized Additive Models*, Monographs on Statistics and Applied Probability 43, Chapman and Hall.

R. Hecht-Nielsen (1987*a*), 'Counter propagation networks', *Proceeedings IEEE Int. Conf. Neural Networks II* pp. 19–32.

R. Hecht-Nielsen (1987*b*), 'Kolmogorov's mapping neural network existence theorem', *Proceeedings IEEE Int. Conf. Neural Networks* pp. 11–13.

M. Hegland (2003), Adaptive sparse grids, in *Proc. of 10th Computational Techniques and Applications Conference CTAC-2001* (K. Burrage and R. Sidje, eds), Vol. 44 of *ANZIAM J.*, pp. C335–C353.

M. Hegland and V. Pestov (1999), Additive models in high dimensions, Technical Report 99-33, School of Mathematical and Computing Sciences, Victoria University of Wellington.

M. Hegland and P. Pozzi (2005), Concentration of measure and the approximation of functions of many variables, in *Mathematical Methods for Curves and Surfaces: Tromso 2004* (M. Daehlen, K. Morken and L. Schumaker, eds), Nashboro Press, Brentwood TN, pp. 101–114.

F. Hickernell and H. Woźniakowski (2000), 'Integration and approximation in arbitrary dimensions', *Adv. Comput. Math.* **12**, 25–58.

F. Hickernell, I. Sloan and G. Wasilkowski (2004), On tractability of weighted integration for certain Banach spaces of functions, in *Monte Carlo and Quasi-Monte Carlo Methods 2002* (H. Niederreiter, ed.), Springer, pp. 51–71.

T. Hill (1956), *Statistical Mechanics, Principles and Selected Applications*, McGraw-Hill, New York.

V. Hoang and C. Schwab (2003), High-dimensional finite elements for elliptic problems with multiple scales, Technical Report Nr. 2003-14, Seminar für Angewandte Mathematik, ETH Zürich.

W. Hoeffding (1948), 'A class of statistics with asympotically normal distributions', *Annals of Math. Statist.* **19**, 293–325.

W. Hurewicz and H. Wallman (1948), *Dimension Theory*, Princeton University Press, Princeton, NJ.

B. Igelnik and N. Parikh (2003), 'Kolmogorov's spline network', *IEEE Transactions on Neural Networks* **14**(4), 725–733.

E. Johnson, S. Wojtkiewicz, L. Bergman and B. Spencer (1997), Finite element and finite difference solutions to the transient Fokker-Planck equation, in *Nonlinear and Stochastic Beam Dynamics in Accelerators - A Challenge to Theoretical and Computational Physics* (A. Bazzani, J. El-

lison, H. Mais and G. Turchetti, eds), pp. 290–306. Proceedings of a Workshop, Lüneburg, Germany, 29. Sep.- 3. Oct. 1997.

G. Karniadakis and S. Sherwin (1999), *Spectral/hp Element Methods for CFD*, Oxford University Press.

S. Khavinson (1997), *Best Approximation by Linear Superpositions*, Translations of Mathematical Monographs, Vol. 159, AMS.

S. Knapek (2000*a*), *Approximation and Kompression mit Tensorprodukt-Multiskalen-Approximationsräumen*, Dissertation, Institut für Angewandte Mathematik, Universität Bonn.

S. Knapek (2000*b*), Hyperbolic cross approximation of integral operators with smooth kernel, Technical Report 665, SFB 256, Univ. Bonn.

A. Kolmogorov (1957), 'On the representation of continuous functions of several variables by superpositions of continuous functions of several variables and addition', *Dokl. Akad. Nauk SSSR* **114**, 953–956. English translation: American Math. Soc. Transl. (2), 28, 55-59, 1963.

M. Köppen (2002), 'On the training of a Kolmogorov network', *Artificial neural networks, ICANN2002, LNCS 2415* pp. 474–479.

F. Kuo and I. Sloan (2005), 'Lifting the curse of dimensionality', *Notices of the AMS*.

V. Kurkova (1991), 'Kolmogorov's theorem is relevant', *Neural Comput.* **3**, 617–622.

Y. Kwok (1998), *Mathematical Models of Financial Derivatives*, Springer.

M. Ledoux (2001), *The Concentration of Measure Phenomenon*, Math. surveys and monographs, Vol. 89, AMS.

Y. Lee, Y. Kim, S. Lee and J. Koo (2004*a*), Structured Multicategory Support Vector Machine with ANOVA decomposition, Technical Report 743, Department of Statistics, The Ohio State University.

Y. Lee, Y. Lin and G. Wahba (2004*b*), 'Multicategory Support Vector Machines, theory, and application to the classification of microarray data and satellite radiance data', *Journal of the American Statistical Association* **99**, 67 – 81.

C. Lemieux and A. Owen (2002), Quasi-regression and the relative importance of the ANOVA components of a function, in *Monte Carlo and Quasi-Monte Carlo Methods* (K. Fang, F. Hickernell and H. Niederreiter, eds), Springer-Verlag, Berlin, pp. 331–344.

G. Leobacher, K. Scheicher and G. Larcher (2003), 'On the tractability of the Brownian bridge algorithm', *Journal of Complexity* **19**, 511–528.

G. Li, M. Atramonov, H. Rabitz, S. Wang, P. Georgopoulos and M. Demiralp (2001*a*), 'Highdimensional model representations generated from low dimensional data samples: lp-RS-HDMR', *Journal of Comput. Chemistry* **24**, 647–656.

G. Li, C. Rosenthal and H. Rabitz (2001*b*), 'Highdimensional model representations', *Journal of Phys. Chem. A* **105**(33), 7765–7777.

G. Li, J. Schoendorf, T. Ho and H. Rabitz (2004), 'Multicut HDMR with an application to an ionospheric model', *Journal Comp. Chem.* **25**, 1149–1156.

G. Li, S. Wang, C. Rosenthal and H. Rabitz (2001*c*), 'Highdimensional model representations generated from low dimensional data samples: I. mp-cut-HDMR', *Journal of Mathematical Chemistry* **30**(1), 1–30.

B. Lin (2000), 'Tensor product space ANOVA models', *The Annals of Statistics* **28**(3), 734–755.

R. Liu and A. Owen (2005), Estimating mean dimensionality of ANOVA decompositions, Technical report, Stanford, Department of Statistics.
G. Lorentz, M. v. Golitschek and Y. Makovoz (1996), *Constructive Approximation*, Springer Verlag.
A. Lozinski and C. Chauviere (2003), 'A fast solver for Fokker-Planck equations applied to viscoelastic flow calculations: 2d FENE model', *Journal of Computational Physics* **189**, 607–625.
A. Lozinski, C. Chauviere, J. Fang and R. Owens (2003), 'Fokker-Planck simulations of fast flows of melts and concentrated polymer solutions in complex geometries', *Journal of Rheology* **47**(2), 535–561.
S. Martello and P. Toth (1990), *Knapsack Problems : Algorithms and Computer Implementations*, John Wiley & Sons, Chichester.
A. Matache (2002), 'Sparse two-scale FEM for homogenization problems', *SIAM Journal of Scientific Computing* **17**, 709–720.
S. McWilliam, D. Knappett and C. Fox (2000), 'Numerical solution of the stationary FPK equation using Shannon wavelets', *Journal of Sound and Vibration* **232**(2), 405–430.
A. Messiah (2000), *Quantum Mechanics*, Dover Publications, New York.
Y. Meyer (1992), *Wavelets and Operators*, Cambridge University Press.
V. Milman (1988), 'The heritage of P. Levy in geometrical functional analysis', *Asterisque* **157-158**, 273–301.
V. Milman and G. Schechtman (2001), *Asymptotic Theory of Finite Normed Spaces*, Lecture Notes in Mathematics Vol. 1200, Springer.
U. Mitzlaff (1997), *Diffusionsapproximation von Warteschlangensystemen*, Dissertation, Institut für Mathematik, Technische Universität Clausthal.
W. Morokoff (1998), 'Generating quasi-random paths for stochastic processes', *SIAM Review* **40**(4), 765–788.
R. Munos (2000), 'A study of reinforcement learning in the continuous case by the means of viscosity solutions', *Machine Learning Journal* **40**, 265–299.
R. Munos and A. Moore (2002), 'Variable resolution discretization in optimal control', *Machine Learning Journal* **49**, 291–323.
P. Niyogi and F. Girosi (1998), 'Generalization bounds for function approxmiation from scattered noisy data', *Advances in Computational Mathematics* **10**, 51–80.
E. Novak and K. Ritter (1998), The curse of dimension and a universal method for numerical integration, in *Multivariate Approximation and Splines* (G. Nürnberger, J. Schmidt and G. Walz, eds).
P. Oswald (1994), *Multilevel Finite Element Approximation*, Teubner Skripten zur Numerik, Teubner, Stuttgart.
A. Owen (2004), Multidimensional variation for quasi-Monte Carlo, Technical Report Nr. 2004-2, Dept. of Statistics, Stanford Univ.
S. Paskov and J. Traub (1995), 'Faster valuation of financial derivatives', *J. Portfolio Management* **22**, 113–120.
J. Prakash (2000), Rouse chains with excluded volume interactions: Linear viscoelasticity, Technical Report Nr. 221, Berichte der Arbeitsgruppe Technomathematik, Universität Kaiserslautern.
J. Prakash and H. Öttinger (1999), 'Viscometric functions for a dilute solution of polymers in a good solvent', *Macromolecules* **32**, 2028–2043.
H. Rabitz and O. Alis (1999), 'General foundations of high-dimensional model representations', *Journal of Mathematical Chemistry* **25**, 197–233.
T. Rassias and J. Simsa (1995), *Finite Sum Decompositions in Mathematical*

Analysis, Pure and Applied Mathematics, Wiley.
C. Reisinger (2003), *Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben*, Dissertation, Universität Heidelberg.
K. Ritter (2000), *Average-Case Analysis of Numerical Problems*, Lecture Notes in Mathematics, No. 1733, Springer, New York.
P. Rouse (1953), 'A theory of the linear viscoelastic properties of dilute solutions of coiling polymers', *J. Chem. Phys.* **21**(7), 1272–1280.
H. Sagan (1994), *Space-Filling Curves*, Springer Verlag.
S. Schneider (2000), *Adaptive solution of elliptic PDE by hierarchical tensor product finite elements*, Dissertation, Institut für Informatik, TU München.
B. Schölkopf and A. Smola (2002), *Learning with Kernels*, MIT Press.
C. Schwab (2003), Numerical solution of parabolic problems in high dimensions, in *Proceedings of the 19th GAMM-Seminar*, pp. 1–20.
C. Schwab and R. Todor (2003*a*), 'Sparse finite elements for stochastic elliptic problems', *Numerische Mathematik* **95**, 707–734.
C. Schwab and R. Todor (2003*b*), 'Sparse finite elements for stochastic elliptic problems – higher order moments', *Computing* **71**(1), 43–63.
X. Shen, H. Chen, J. Dai and W. Dai (2002), 'The finite element method for computing the stationary distribution on an SRBM in a hypercube with applications to finite buffer queueing networks', *Queuing Systems* **42**, 33–62.
P. Sjöberg (2002), *Numerical Solution of the Master Equation in Molecular Biology*, Master Thesis, Department for Scientific Computing, Uppsala Universität.
I. Sloan (2001), QMC integration – beating intractability by weighting the coordinate directions, Technical Report AMR01/12, University of New South Wales, Applied Mathematics report.
I. Sloan and X. Wang (2005), 'Why are high-dimensional finance problems often of low effective dimension', *SIAM Journal on Scientific Computing* **27**(1), 159–183.
I. Sloan and H. Woźniakowski (1998), 'When are quasi-Monte Carlo algorithms efficient for high dimensional integrals ?', *J. Complexity* **14**, 1–33.
I. Sloan, X. Wang and H. Woźniakowski (2004), 'Finite-order weights imply tractability of multivariate integration', *Journal of Complexity* **20**, 46–74.
S. Smolyak (1963), 'Quadrature and interpolation formulas for tensor products of certain classes of functions', *Soviet Math. Dokl.* **4**, 240–243. Russian original in Dokl. Akad. Nauk SSSR, 148 (1963), pp. 1042–1045.
D. Sprecher (1965), 'On the structure of continuous functions of several variables', *Transactions American Mathematical Society* **115**(3), 340–355.
D. Sprecher (1996), 'A numerical implementation of Kolmogorov's superpositions', *Neural Networks* **9**(5), 765–772.
D. Sprecher (1997), 'A numerical implementation of Kolmogorov's superpositions II', *Neural Networks* **10**(3), 447–457.
D. Sprecher and S. Draghici (2002), 'Space filling curves and Kolmogorov superposition-based neural networks', *Neural Networks* **15**, 57–67.
C. Stone (1985), 'Additive regression and other nonparametric models', *Ann. Statist.* **13**, 689–705.
E. Süli (2006), Finite element approximation of high-dimensional transport-dominated diffusion problems, in these proceedings.

R. Sutton and A. Barto (1998), *Reinforcement Learning. An Introduction*, MIT Press, Cambridge, London.
B. Szabo and I. Babuska (1991), *Finite Element Analysis*, Wiley.
M. Talagrand (1995), *Concentration of Measure and Isoperimetric Inequalities in Product Spaces*, Vol. 81, Publ. Math. IHES.
J. Traub and H. Woźniakowski (1980), *A General Theory of Optimal Algorithms*, Academic Press, New York.
J. Traub, G. Wasilkowski and H. Woźniakowski (1983), *Information, Uncertainty, Complexity*, Addison-Wesley, Reading.
J. Traub, G. Wasilkowski and H. Woźniakowski (1988), *Information-Based Complexity*, Academic Press, New York.
H. Triebel (1992), *Theory of function spaces II*, Birkhäuser Verlag.
E. Tyrtyshnikov (2004), 'Kronecker-product type approximations for some function-related matrices', *Linear Algebra and its Applications* **379**, 423–437.
G. Venktiteswaran and M. Junk (2005*a*), 'A QMC approach for high dimensional Fokker-Planck equations modelling polymeric liquids', *Math. Comp. Simul.* **68**, 23–41.
G. Venktiteswaran and M. Junk (2005*b*), 'Quasi-Monte Carlo algorithms for diffusion equations in high dimensions', *Math. Comp. Simul.* **68**, 43–56.
A. Vitushkin (1964), 'Proof of the existence of analytic functions of several variables not representable by linear superpositions of continuously differentiable functions of fewer variables', *Dokl. Akad. Nauk SSSR* **156**, 1258–1261. English translation: Soviet Math. Dokl. (5), 793-796, 1964.
A. Vitushkin (2004), 'On Hilbert's thirteenth problem and related questions', *Russian Math. Surveys* **51**(1), 11–25.
G. Wahba (1990), *Spline Models for Observational Data*, Vol. 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia.
G. Wasilkovski and H. Woźniakowski (1995), 'Explicit cost bounds of algorithms for multivariate tensor product problems', *J. Complexity* **11**, 1–56.
G. Wasilkovski and H. Woźniakowski (1999), 'Weighted tensor product algorithms for linear multivariate problems', *J. Complexity* **15**, 402–447.
G. Wasilkowski and H. Woźniakowski (2004), 'Finite-order weights imply tractability of linear multivariate problems', *Journal of Approximation Theory* **130**(1), 57–77.
H. Wei and A. Billings (2004), 'A unified wavelet-based modelling framework for non-linear system identification: The WARNARX model structure', *Int. J. Control* **77**(4), 351–366.
P. Wilmott (1998), *Derivatives, the Theory and Practice of Financial Engineering*, John Wiley & Sons.
S. Wojtkiewicz and L. Bergman (2000), Numerical solution of high-dimensional Fokker-Planck equations, in *8th ASCE Speciality Conference on Probabilistic Mechanics and Structural Reliability*. PCM2000-167.
H. Yserentant (1986), 'On the multi-level splitting of finite element spaces', *Numer. Math.* **49**, 379–412.
H. Yserentant (2004), 'On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives', *Numer. Math.* **98**(4), 731–759.
C. Zenger (1991), Sparse grids, in *Parallel Algorithms for Partial Differential Equations* (W. Hackbusch, ed.), Vol. 31 of *NNFM*, Vieweg, Braunschweig/Wiesbaden.