

Sparse grids

Hans-Joachim Bungartz

IPVS, Universität Stuttgart,

Universitätsstraße 38, D-70569 Stuttgart, Germany

E-mail: bungartz@informatik.uni-stuttgart.de

Michael Griebel

Institut für Numerische Simulation, Universität Bonn,

Wegelerstraße 6, D-53113 Bonn, Germany

E-mail: griebel@iam.uni-bonn.de

We present a survey of the fundamentals and the applications of sparse grids, with a focus on the solution of partial differential equations (PDEs). The sparse grid approach, introduced in Zenger (1991), is based on a higher-dimensional multiscale basis, which is derived from a one-dimensional multiscale basis by a tensor product construction. Discretizations on sparse grids involve $O(N \cdot (\log N)^{d-1})$ degrees of freedom only, where d denotes the underlying problem's dimensionality and where N is the number of grid points in one coordinate direction at the boundary. The accuracy obtained with piecewise linear basis functions, for example, is $O(N^{-2} \cdot (\log N)^{d-1})$ with respect to the L_2 - and L_∞ -norm, if the solution has bounded second mixed derivatives. This way, the curse of dimensionality, *i.e.*, the exponential dependence $O(N^d)$ of conventional approaches, is overcome to some extent. For the energy norm, only $O(N)$ degrees of freedom are needed to give an accuracy of $O(N^{-1})$. That is why sparse grids are especially well-suited for problems of very high dimensionality.

The sparse grid approach can be extended to nonsmooth solutions by adaptive refinement methods. Furthermore, it can be generalized from piecewise linear to higher-order polynomials. Also, more sophisticated basis functions like interpolants, prewavelets, or wavelets can be used in a straightforward way.

We describe the basic features of sparse grids and report the results of various numerical experiments for the solution of elliptic PDEs as well as for other selected problems such as numerical quadrature and data mining.

CONTENTS

1	Introduction	2
2	Breaking the curse of dimensionality	5
3	Piecewise linear interpolation on sparse grids	8
4	Generalizations, related concepts, applications	42
5	Numerical experiments	73
6	Concluding remarks	109
	References	110

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 of computing and representing an approximation with a prescribed accuracy ε depends exponentially on the dimensionality d of the considered problem. We encounter complexities of the order $O(\varepsilon^{-\alpha d})$ with $\alpha > 0$ depending on the respective approach, the smoothness of the function under consideration, 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, for instance, this complexity estimate translates to $O(N^d)$ grid points or degrees of freedom for which approximation accuracies of the order $O(N^{-\alpha'})$ are achieved, where α' depends on the smoothness of the function under consideration and the polynomial degree of the approximating functions.¹ Thus, the computational cost and storage requirements grow exponentially with the dimensionality of the problem, which is the reason for 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 in a *negative* exponential way, *i.e.*, it behaves like $O(N^{-\beta/d})$, we directly see that the cost–benefit ratio is independent of d and that it is of the order $O(N^d \cdot N^{-\beta/d}) = O(N^{-\beta})$, with some β independent of d . This way, the curse of dimensionality can be broken easily. However, such an assumption is somewhat unrealistic.

¹ If the solution is not smooth but possesses singularities, the order α 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.

In the sparse grid method, in principle, we follow this approach, but we only assume the functions to live in spaces of functions with bounded mixed derivatives instead. First, we need a 1D multilevel basis, preferably an H^1 - and L_2 -stable one. Then, if we express a 1D function as usual as a linear combination of these basis functions, the corresponding coefficients decrease from level to level with a rate which depends on the smoothness of the function and on the chosen set of basis functions. From this, by a simple tensor product construction, we obtain a multilevel basis for the higher-dimensional case. Note that here 1D bases living on different levels are used in the product construction, *i.e.*, we obtain basis functions with anisotropic support in the higher-dimensional case. Now, if the function to be represented has bounded second mixed derivatives and if we use a piecewise linear 1D basis as a starting point, it can be shown that the corresponding coefficients decrease with a factor proportional to $2^{-2|\mathbf{l}|_1}$ where the multi-index $\mathbf{l} = (l_1, \dots, l_d)$ denotes the different levels involved. If we then omit coefficients whose absolute values are smaller than a prescribed tolerance, we obtain sparse grids. It turns out that the number of degrees of freedom needed for some prescribed accuracy no longer depends (up to logarithmic factors) on d exponentially. This allows substantially faster solution of moderate-dimensional problems and can enable the solution of higher-dimensional problems.²

Of course, this sparse grid approach is not restricted to the piecewise linear case. Extensions of the standard piecewise linear hierarchical basis to general polynomial degree p as well as interpolets (see Section 4.3) or (pre-) wavelets (see Section 4.4) have been successfully used as the univariate ingredient for tensor product construction. Finally, the spectrum of applications of sparse grids ranges from numerical quadrature, via the discretization of PDEs, to fields such as data mining.

The remainder of this paper is organized as follows. In Section 2, we briefly discuss the breaking of the curse of dimensionality from the theoretical point of view. Here, we resort to harmonic analysis and collect known approaches for escaping the exponential dependence on d . It turns out that one example of such a method simply corresponds to the assumption of using a space of bounded mixed first variation, *i.e.*, an anisotropic multivariate smoothness assumption. Except for the degree of the derivative, this resembles the explicit assumption of Section 3, where we introduce the principles of the sparse grid technique and derive the interpolation properties of the resulting sparse grid spaces. As a starting point, we use the

² The constant in the corresponding complexity estimates, however, still depends exponentially on d . This still limits the approach for PDEs to moderate-dimensional problems. At present we are able to deal with 18D PDEs.

standard piecewise linear multiscale basis in one dimension, *i.e.*, the Faber basis (Faber 1909, Yserentant 1986), as the simplest example of a multiscale series expansion which involves interpolation by piecewise linears. To attain higher dimensions, we resort to a tensor product construction. Then, for functions with bounded mixed second derivatives, sparse grid approximation schemes are derived which exhibit cost complexities of the order $O(N(\log N)^{d-1})$ and give an error of $O(N^{-2}(\log N)^{d-1})$ in the L_2 -norm. We then show that, for our smoothness assumption, the sparse grid discretization scheme can also be formally derived by solving an optimization problem which is closely related to n -term approximation (DeVore 1998). Finally, we consider the energy norm, for which optimality leads us to an energy-based sparse grid with cost complexity $O(N)$ and accuracy $O(N^{-1})$. Thus the exponential dependence of the logarithmic terms on d is completely removed (although present in the constants).

In Section 4 we discuss generalizations of the above approach based on piecewise linear hierarchical bases. First, instead of these, higher-order polynomial hierarchical bases can be employed. Here we describe the construction of multilevel polynomial bases by means of a hierarchical Lagrangian interpolation scheme and analyse the resulting sparse grid approximation properties. A further example of an extension is based on the interpolets due to Deslauriers and Dubuc (1989). Such higher-order approaches allow us to take advantage of higher regularity assumptions concerning the solution, resulting in better approximation rates. However, all these hierarchical multiscale bases have a crucial drawback when $d > 1$. They are not stable, in the sense that the error can be estimated from above by a multilevel norm with constants independent of the level, but not from below. Lower bounds can be obtained by using wavelet-type multiscale bases, semi-orthogonal prewavelets, or related biorthogonal multiscale bases instead. Again, then, the tensor product construction and successive optimization lead to sparse grids. Finally, we close this section with a short overview of the state of the art of sparse grid research. Here, the focus is put on the numerical treatment of PDEs based on different discretization approaches, and we include a short discussion of adaptive grid refinement and fast solvers in the sparse grid context.

In Section 5 we present numerical results of selected experiments. First, to show the properties of the sparse grid approximation, we discuss some PDE model problems in two or three dimensions. Second, we apply sparse grids to the solution of flow problems via the Navier–Stokes equations. Finally, we present results for two non-PDE problem classes of high dimensionality: numerical quadrature and data mining.

The concluding remarks of Section 6 close this discussion of sparse grid methods.

2. Breaking the curse of dimensionality

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

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

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 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 $\|f - f_n\| = O(n^{-c})$ with constant $c > 0$. Of course, such an assumption is completely unrealistic.

However, about ten 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 an approximation rate

$$\|f - f_n\| = O(n^{-1/d})$$

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

$$\|f - f_n\| = O(n^{-1/2})$$

independent of d . Meanwhile, other function classes are known with such properties. These 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) and DeVore (1998). Consider the class of functions $\mathcal{F}(M)$ with integral representation

$$f(x) = \int A(x, t) d|\mu|(t)$$

with

$$\int d|\mu|(t) \leq M, \quad (2.1)$$

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 n -term sum

$$f_n(x) = \sum_{j=1}^n a_j A_{t_j}(x)$$

where

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

with C independent of d . In the following, we consider examples of such spaces.

Example 1. (Radial basis schemes) Consider superpositions of Gaussian bumps. These resemble the space $\mathcal{F}(M, \text{Gaussians})$ with $t := (x_0, s)$ and Gaussian atoms $A(x, t) = \exp(-\|x - x_0\|^2/s^2)$, where $\|\cdot\|$ denotes the Euclidean norm. Now, if the sum of the height of all Gaussians is bounded by M , 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 $B_{1,1}^d(\mathbb{R}^d)$ which is just Meyer's bump algebra (Meyer 1992). Thus, we have a restriction to smoother functions in higher dimensions such that the ratio r/d stays constant and, consequently, $n^{-r/d}$ does not grow with d .

Example 2. (Orthant scheme I) Another class of functions with an approximation rate independent of d is $\mathcal{F}(M, \text{Orthant})$. 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 (2.1) is at most M , the resulting approximation rate is of order $O(n^{-1/2})$ independent of d . A typical and well-known example for such a construction is the cumulative distribution function in \mathbb{R}^d . This simply results in the Monte Carlo method.

Example 3. (Orthant scheme II) A further interesting function class are the functions which are formed by any superposition of 2^d functions, each orthant-wise monotone for a different orthant. Now, the condition $\int d|\mu|(t) \leq 1$ is the same as

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

i.e., we obtain the space of bounded mixed first variation. Again, this means considering only functions that get smoother as the dimensionality grows, but, in contrast to the examples mentioned above, 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 recently given by Woźniakowski, Sloan,

and others (Wasilkowski and Woźniakowski 1995, Sloan and Woźniakowski 1998, Wasilkowski and Woźniakowski 1999, Sloan 2001, Hickernell, Sloan and Wasilkowski 2003, Dick, Sloan, Wang and Woźniakowski 2003). Here, especially in the context of numerical integration, they introduce the notion of *weighted* Sobolev spaces. Having observed that for some problems the integrand becomes less and less variable in successive coordinate directions, they introduce a sequence of positive weights $\{\gamma_j\}$ with decreasing values, with the weight γ_j being associated with coordinate direction j . They are able to show that the integration problem in a particular Sobolev space setting becomes *strongly tractable* (Traub and Woźniakowski 1980, Traub, Wasilkowski and Woźniakowski 1983, 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 goes polynomially to zero if and only if the sum of the weights is asymptotically bounded from above. This corresponds to the decay of the kernel contributions in a reproducing kernel Hilbert space with rising 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 Example 3 above. Since then, more general assumptions on the weights, and thus on the induced weighted function spaces, have been found (Dick *et al.* 2003, Hickernell *et al.* 2003).

In any case, we observe that a certain smoothness assumption on the function under consideration changes with d to obtain approximation rates which no longer depend exponentially on d . This raises the question of the very meaning of smoothness as the dimension changes and tends to ∞ .

To this end, let us finally note an interesting aspect, namely the concentration of measure phenomenon (Milman 1988, Talagrand 1995, Gromov 1999) for probabilities in normed spaces of high dimension (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, Lipschitz constant L , on the d -sphere. Let P be normalized Lebesgue measure on the sphere and let X be 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 and c_2 independent of f and d (see Milman and Schechtman (1986) or Baxter and Iserles (2003, Section 2.3)). In its simplest form, the concentration of measure phenomenon states that every Lipschitz function on a sufficiently high-dimensional domain Ω is well approximated by a

constant function (Hegland and Pestov 1999). Thus, there is some chance of treating high-dimensional problems despite the curse of dimensionality.

3. Piecewise linear interpolation on sparse grids

As a first approach to sparse grids and their underlying hierarchical multi-level setting, we discuss the problem of *interpolating* smooth functions with the help of piecewise d -linear hierarchical bases. For that, we introduce a tensor product-based subspace splitting and study the resulting subspaces. Starting from their properties, sparse grids are defined via an optimization process in a cost–benefit spirit closely related to the notion of n -term approximation. Out of the variety of norms with respect to which such an optimized discretization scheme can be derived, we restrict ourselves to the L_2 -, the L_∞ - and the energy norm, and thus to the respective types of sparse grids. After presenting the most important approximation properties of the latter, a short digression into recurrences and complexity will demonstrate their asymptotic characteristics and, consequently, their potential for problems of high dimensionality.

3.1. Hierarchical multilevel subspace splitting

In this section, the basic ingredients for our tensor product-based hierarchical setting are provided.

Subspace decomposition

Let us start with some notation and with the preliminaries that are necessary for a detailed discussion of sparse grids for purposes of interpolation or approximation, respectively. On the d -dimensional unit interval $\bar{\Omega} := [0, 1]^d$, we consider multivariate functions u , $u(\mathbf{x}) \in \mathbb{R}$, $\mathbf{x} := (x_1, \dots, x_d) \in \bar{\Omega}$, with (in some sense) bounded weak mixed derivatives

$$D^\alpha u := \frac{\partial^{|\alpha|_1} u}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \quad (3.1)$$

up to some given order $r \in \mathbb{N}_0$. Here, $\alpha \in \mathbb{N}_0^d$ denotes a d -dimensional multi-index with the two norms

$$|\alpha|_1 := \sum_{j=1}^d \alpha_j \quad \text{and} \quad |\alpha|_\infty := \max_{1 \leq j \leq d} \alpha_j.$$

In the context of multi-indices, we use component-wise arithmetic operations, for example

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} := (\alpha_1 \beta_1, \dots, \alpha_d \beta_d), \quad \gamma \cdot \boldsymbol{\alpha} := (\gamma \alpha_1, \dots, \gamma \alpha_d),$$

or

$$2^\alpha := (2^{\alpha_1}, \dots, 2^{\alpha_d}),$$

the relational operators

$$\alpha \leq \beta \Leftrightarrow \forall_{1 \leq j \leq d} \alpha_j \leq \beta_j$$

and

$$\alpha < \beta \Leftrightarrow \alpha \leq \beta \text{ and } \alpha \neq \beta,$$

and, finally, special multi-indices such as

$$\mathbf{0} := (0, \dots, 0) \quad \text{or} \quad \mathbf{1} := (1, \dots, 1),$$

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

$$\begin{aligned} X^{q,r}(\bar{\Omega}) &:= \{u : \bar{\Omega} \rightarrow \mathbb{R} : D^\alpha u \in L_q(\Omega), |\alpha|_\infty \leq r\}, \\ X_0^{q,r}(\bar{\Omega}) &:= \{u \in X^r(\bar{\Omega}) : u|_{\partial\Omega} = 0\}. \end{aligned} \quad (3.2)$$

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 $u \in X^r(\bar{\Omega})$ vanishing on the boundary $\partial\Omega$. Note that, for the theoretical considerations, we shall restrict ourselves to the case of homogeneous boundary conditions, *i.e.*, to $X_0^{q,r}(\bar{\Omega})$. Furthermore, note that we omit the ambient dimension d when clear from the context. Concerning the smoothness parameter $r \in \mathbb{N}_0$, we need $r = 2$ for the case of piecewise linear approximations which, for the moment, will be in the centre of interest. Finally, for functions $u \in X_0^{q,r}(\bar{\Omega})$ and multi-indices α with $|\alpha|_\infty \leq r$, we introduce the seminorms

$$\begin{aligned} |u|_{\alpha,\infty} &:= \|D^\alpha u\|_\infty, \\ |u|_{\alpha,2} &:= \|D^\alpha u\|_2 = \left(\int_{\bar{\Omega}} |D^\alpha u|^2 \, d\mathbf{x} \right)^{1/2}. \end{aligned} \quad (3.3)$$

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

$$\{\Omega_{\mathbf{l}} : \mathbf{l} \in \mathbb{N}^d\} \quad (3.4)$$

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

$$\mathbf{h}_{\mathbf{l}} := (h_{l_1}, \dots, h_{l_d}) := 2^{-\mathbf{l}}. \quad (3.5)$$

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

$$\mathbf{x}_{\mathbf{l},\mathbf{i}} := (x_{l_1,i_1}, \dots, x_{l_d,i_d}) := \mathbf{i} \cdot \mathbf{h}_{\mathbf{l}}, \quad \mathbf{0} \leq \mathbf{i} \leq 2^{\mathbf{l}}. \quad (3.6)$$

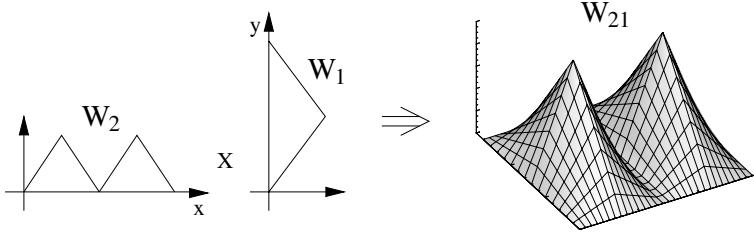


Figure 3.1. Tensor product approach for piecewise bilinear basis functions.

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

$$\phi(x) := \begin{cases} 1 - |x|, & \text{if } x \in [-1, 1], \\ 0, & \text{otherwise.} \end{cases} \quad (3.7)$$

This mother of all piecewise linear basis functions 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,

$$\phi_{l_j,i_j}(x_j) := \phi\left(\frac{x_j - i_j \cdot h_{l_j}}{h_{l_j}}\right). \quad (3.8)$$

The resulting 1D 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 3.1):

$$\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) := \prod_{j=1}^d \phi_{l_j,i_j}(x_j). \quad (3.9)$$

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

$$V_{\mathbf{l}} := \text{span}\{\phi_{\mathbf{l},\mathbf{i}} : \mathbf{i} \leq \mathbf{l} \leq \mathbf{l} - \mathbf{1}\}, \quad (3.10)$$

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 just the standard *nodal point basis* of the finite-dimensional space $V_{\mathbf{l}}$.

Additionally, we introduce the hierarchical increments W_1 ,

$$W_1 := \text{span}\{\phi_{\mathbf{l},\mathbf{i}} : \mathbf{1} \leq \mathbf{i} \leq 2^1 - \mathbf{1}, i_j \text{ odd for all } 1 \leq j \leq d\}, \quad (3.11)$$

for which the relation

$$V_1 = \bigoplus_{\mathbf{k} \leq \mathbf{l}} W_{\mathbf{k}} \quad (3.12)$$

can be easily seen. Note that the supports of all basis functions $\phi_{\mathbf{l},\mathbf{i}}$ spanning W_1 are mutually disjoint. Thus, with the index set

$$\mathbf{I}_1 := \{\mathbf{i} \in \mathbb{N}^d : \mathbf{1} \leq \mathbf{i} \leq 2^1 - \mathbf{1}, i_j \text{ odd for all } 1 \leq j \leq d\}, \quad (3.13)$$

we get another basis of V_1 , the *hierarchical basis*

$$\{\phi_{\mathbf{k},\mathbf{i}} : \mathbf{i} \in \mathbf{I}_{\mathbf{k}}, \mathbf{k} \leq \mathbf{l}\}, \quad (3.14)$$

which generalizes the well-known 1D basis shown in Figure 3.2 to the d -dimensional case by means of a tensor product approach. With these hierarchical difference spaces W_1 , we can define

$$V := \sum_{l_1=1}^{\infty} \cdots \sum_{l_d=1}^{\infty} W_{(l_1, \dots, l_d)} = \bigoplus_{\mathbf{l} \in \mathbb{N}^d} W_{\mathbf{l}} \quad (3.15)$$

with its natural hierarchical basis

$$\{\phi_{\mathbf{l},\mathbf{i}} : \mathbf{i} \in \mathbf{I}_{\mathbf{l}}, \mathbf{l} \in \mathbb{N}^d\}. \quad (3.16)$$

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

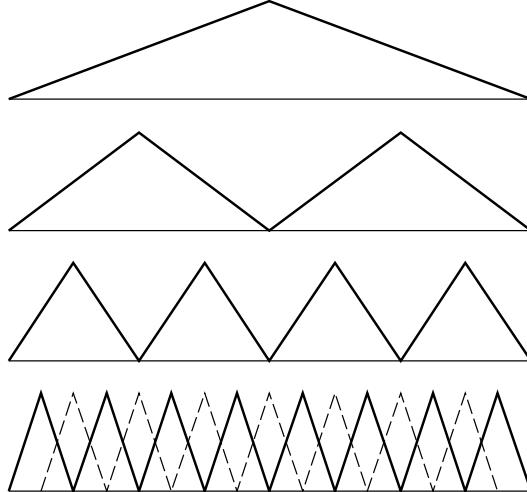


Figure 3.2. Piecewise linear hierarchical basis (solid)
vs. nodal point basis (dashed).

Later we shall deal with finite-dimensional subspaces of V . Note that, for instance, with the discrete spaces

$$V_n^{(\infty)} := \bigoplus_{|\mathbf{l}|_\infty \leq n} W_{\mathbf{l}}, \quad (3.17)$$

the limit

$$\lim_{n \rightarrow \infty} V_n^{(\infty)} = \lim_{n \rightarrow \infty} \bigoplus_{|\mathbf{l}|_\infty \leq n} W_{\mathbf{l}} := \bigcup_{n=1}^{\infty} V_n^{(\infty)} = V \quad (3.18)$$

exists due to $V_n^{(\infty)} \subset V_{n+1}^{(\infty)}$. Hence, 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

$$u(\mathbf{x}) = \sum_{\mathbf{l}} u_{\mathbf{l}}(\mathbf{x}), \quad u_{\mathbf{l}}(\mathbf{x}) = \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} v_{\mathbf{l},\mathbf{i}} \cdot \phi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \in W_{\mathbf{l}}, \quad (3.19)$$

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

Before we turn to finite-dimensional approximation spaces for $X_0^{q,2}(\bar{\Omega})$, we summarize the most important properties of the hierarchical subspaces $W_{\mathbf{l}}$ according to Bungartz (1992b) and Bungartz and Griebel (1999).

Basic properties of the subspaces

Concerning the subspaces $W_{\mathbf{l}}$, the crucial questions are how important $W_{\mathbf{l}}$ is for the interpolation of some given $u \in X_0^{q,2}(\bar{\Omega})$ and what computational and storage cost come along with it. From (3.11) and (3.13), we immediately learn the dimension of $W_{\mathbf{l}}$, *i.e.*, the number of degrees of freedom (grid points or basis functions, respectively) associated with $W_{\mathbf{l}}$:

$$|W_{\mathbf{l}}| = |\mathbf{I}_{\mathbf{l}}| = 2^{|\mathbf{l}-\mathbf{1}|_1}. \quad (3.20)$$

Equation (3.20) already answers the second question.

The following discussion of a subspace's contribution to the overall interpolant according to (3.19) will be based upon three norms: the maximum norm $\|\cdot\|_\infty$, the L_p -norm $\|\cdot\|_p$ ($p = 2$ in general), and the energy norm

$$\|u\|_E := \left(\int_{\Omega} \sum_{j=1}^d \left(\frac{\partial u(\mathbf{x})}{\partial x_j} \right)^2 d\mathbf{x} \right)^{1/2}, \quad (3.21)$$

which is equivalent to the H^1 -norm in $H_0^1(\bar{\Omega})$. For the Laplacian, (3.21) indeed indicates the energy norm in finite element terminology. First we look at the different hierarchical basis functions $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$.

Lemma 3.1. For any piecewise d -linear basis function $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$, the following equations hold:

$$\begin{aligned} \|\phi_{\mathbf{l},\mathbf{i}}\|_\infty &= 1, \\ \|\phi_{\mathbf{l},\mathbf{i}}\|_p &= \left(\frac{2}{p+1} \right)^{d/p} \cdot 2^{-|\mathbf{l}|_1/p}, \quad (p \geq 1), \\ \|\phi_{\mathbf{l},\mathbf{i}}\|_E &= \sqrt{2} \cdot \left(\frac{2}{3} \right)^{(d-1)/2} \cdot 2^{-|\mathbf{l}|_1/2} \cdot \left(\sum_{j=1}^d 2^{2l_j} \right)^{1/2}. \end{aligned} \tag{3.22}$$

Proof. All equalities result from straightforward calculations based on the definition of $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$ (see (3.7) to (3.9), and Bungartz (1992b) and Bungartz and Griebel (1999), for example). \square

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 way:

$$\begin{aligned} v_{\mathbf{l},\mathbf{i}} &= \left(\prod_{j=1}^d \begin{bmatrix} -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix}_{x_{l_j,i_j,l_j}} \right) u =: \left(\prod_{j=1}^d I_{x_{l_j,i_j,l_j}} \right) u \\ &=: I_{\mathbf{x}_{\mathbf{l},\mathbf{i}},\mathbf{l}} u. \end{aligned} \tag{3.23}$$

This is due to the definition of the spaces W_1 and their basis functions (3.11), whose supports are mutually disjoint and do not contain coarse grid points $\mathbf{x}_{\mathbf{k},\mathbf{j}}$, $\mathbf{k} < \mathbf{l}$, in their interior. Definition (3.23) illustrates why $v_{\mathbf{l},\mathbf{i}}$ is also called *hierarchical surplus*. In (3.23), as usual in multigrid terminology (*cf.* Hackbusch (1985, 1986), for instance), $I_{\mathbf{x}_{\mathbf{l},\mathbf{i}},\mathbf{l}}$ denotes a d -dimensional stencil which gives the coefficients for a linear combination of nodal values of its argument u . This operator-based representation of the hierarchical coefficients $v_{\mathbf{l},\mathbf{i}}$ leads to an integral representation of $v_{\mathbf{l},\mathbf{i}}$, as follows.

Lemma 3.2. Let $\psi_{l_j,i_j}(x_j) := -2^{-(l_j+1)} \cdot \phi_{l_j,i_j}(x_j)$. Further, let $\psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) := \prod_{j=1}^d \psi_{l_j,i_j}(x_j)$. For any coefficient value $v_{\mathbf{l},\mathbf{i}}$ of the hierarchical representation (3.19) of $u \in X_0^{q,2}(\bar{\Omega})$, the following relation holds:

$$v_{\mathbf{l},\mathbf{i}} = \int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \cdot D^2 u(\mathbf{x}) \, d\mathbf{x}. \tag{3.24}$$

Proof. First we look at the simplest case $d = 1$, where we can omit the index j for clarity. Partial integration provides

$$\begin{aligned} \int_{\Omega} \psi_{l,i}(x) \cdot \frac{\partial^2 u(x)}{\partial x^2} \, dx \\ = \int_{x_{l,i}-h_l}^{x_{l,i}+h_l} \psi_{l,i}(x) \cdot \frac{\partial^2 u(x)}{\partial x^2} \, dx \end{aligned}$$

$$\begin{aligned}
&= \left[\psi_{l,i}(x) \cdot \frac{\partial u(x)}{\partial x} \right]_{x_{l,i}-h_l}^{x_{l,i}+h_l} - \int_{x_{l,i}-h_l}^{x_{l,i}+h_l} \frac{\partial \psi_{l,i}(x)}{\partial x} \cdot \frac{\partial u(x)}{\partial x} dx \\
&= \int_{x_{l,i}-h_l}^{x_{l,i}} \frac{1}{2} \cdot \frac{\partial u(x)}{\partial x} dx - \int_{x_{l,i}}^{x_{l,i}+h_l} \frac{1}{2} \cdot \frac{\partial u(x)}{\partial x} dx \\
&= I_{x_{l,i},l} u,
\end{aligned}$$

since $\psi_{l,i}(x_{l,i} - h_l) = \psi_{l,i}(x_{l,i} + h_l) = 0$ and since $\partial \psi_{l,i}(x)/\partial x \in \{\frac{1}{2}, -\frac{1}{2}\}$ due to the construction of $\psi_{l,i}$ and $\phi_{l,i}$. Finally, the tensor product approach according to the operator product given in (3.23) leads to a straightforward generalization to $d > 1$. \square

The above lemma and its proof show the close relations of our hierarchical basis approach to integral transforms like wavelet transforms. Applying successive partial integration to (3.24), twice for $d = 1$ and $2d$ times for general dimensionality, we get

$$v_{\mathbf{l},\mathbf{i}} = \int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \cdot D^2 u(\mathbf{x}) d\mathbf{x} = \int_{\Omega} \hat{\psi}_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \cdot u(\mathbf{x}) d\mathbf{x}, \quad (3.25)$$

where $\hat{\psi}_{\mathbf{l},\mathbf{i}}(\mathbf{x})$ equals $D^2 \psi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$ in a weak sense (*i.e.*, in the sense of distributions) and is a linear combination of 3^d Dirac pulses of alternating sign. Thus, the hierarchical surplus $v_{\mathbf{l},\mathbf{i}}$ can be interpreted as the coefficient resulting from an integral transform with respect to a function $\hat{\psi}_{\mathbf{l},\mathbf{i}}(\mathbf{x})$ of an oscillating structure.

Starting from (3.24), we are now able to give bounds for the hierarchical coefficients with respect to the different seminorms introduced in (3.3).

Lemma 3.3. Let $u \in X_0^{q,2}(\bar{\Omega})$ be given in its hierarchical representation. Then, the following estimates for the hierarchical coefficients $v_{\mathbf{l},\mathbf{i}}$ hold:

$$\begin{aligned}
|v_{\mathbf{l},\mathbf{i}}| &\leq 2^{-d} \cdot 2^{-2 \cdot |\mathbf{l}|_1} \cdot |u|_{\mathbf{2},\infty}, \\
|v_{\mathbf{l},\mathbf{i}}| &\leq 2^{-d} \cdot \left(\frac{2}{3} \right)^{d/2} \cdot 2^{-(3/2) \cdot |\mathbf{l}|_1} \cdot |u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}|_{\mathbf{2},2},
\end{aligned} \quad (3.26)$$

where $\text{supp}(\phi_{\mathbf{l},\mathbf{i}})$ denotes the support of $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$.

Proof. With (3.22), (3.24), and with the definition of $\psi_{\mathbf{l},\mathbf{i}}$, we get

$$\begin{aligned}
|v_{\mathbf{l},\mathbf{i}}| &= \left| \int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \cdot D^2 u(\mathbf{x}) d\mathbf{x} \right| \leq \|\psi_{\mathbf{l},\mathbf{i}}\|_1 \cdot \|D^2 u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}\|_{\infty} \\
&= 2^{-d} \cdot 2^{-|\mathbf{l}|_1} \cdot \|\phi_{\mathbf{l},\mathbf{i}}\|_1 \cdot |u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}|_{\mathbf{2},\infty} \\
&\leq 2^{-d} \cdot 2^{-2 \cdot |\mathbf{l}|_1} \cdot |u|_{\mathbf{2},\infty}.
\end{aligned}$$

For $|\cdot|_{\mathbf{2},2}$, the Cauchy–Schwarz inequality provides

$$\begin{aligned} |v_{\mathbf{l},\mathbf{i}}| &= \left| \int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) \cdot D^2 u(\mathbf{x}) \, d\mathbf{x} \right| \leq \|\psi_{\mathbf{l},\mathbf{i}}\|_2 \cdot \|D^2 u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}\|_2 \\ &= 2^{-d} \cdot 2^{-|\mathbf{l}|_1} \cdot \|\phi_{\mathbf{l},\mathbf{i}}\|_2 \cdot |u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}|_{\mathbf{2},2}, \end{aligned}$$

which, with (3.22), is the desired result. \square

Finally, the results from the previous three lemmata lead to bounds for the contribution $u_{\mathbf{l}} \in W_{\mathbf{l}}$ of a subspace $W_{\mathbf{l}}$ to the hierarchical representation (3.19) of a given $u \in X_0^{q,2}(\bar{\Omega})$.

Lemma 3.4. Let $u \in X_0^{q,2}(\bar{\Omega})$ be given in its hierarchical representation (3.19). Then, the following estimates for its components $u_{\mathbf{l}} \in W_{\mathbf{l}}$ hold:

$$\|u_{\mathbf{l}}\|_{\infty} \leq 2^{-d} \cdot 2^{-2|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},\infty}, \quad (3.27)$$

$$\|u_{\mathbf{l}}\|_2 \leq 3^{-d} \cdot 2^{-2|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},2},$$

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

$$\|u_{\mathbf{l}}\|_E \leq \sqrt{3} \cdot 3^{-d} \cdot 2^{-2|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2 \cdot l_j} \right)^{1/2} \cdot |u|_{\mathbf{2},2}.$$

Proof. Since the supports of all $\phi_{\mathbf{l},\mathbf{i}}$ contributing to $u_{\mathbf{l}}$ according to (3.19) are mutually disjoint, the first estimate follows immediately from the respective statements in (3.22) and (3.26). For the estimate concerning the L_2 -norm, we get with the same argument of disjoint supports and with (3.22) and (3.26)

$$\begin{aligned} \|u_{\mathbf{l}}\|_2^2 &= \left\| \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} v_{\mathbf{l},\mathbf{i}} \cdot \phi_{\mathbf{l},\mathbf{i}} \right\|_2^2 = \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} |v_{\mathbf{l},\mathbf{i}}|^2 \cdot \|\phi_{\mathbf{l},\mathbf{i}}\|_2^2 \\ &\leq \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} \frac{1}{6^d} \cdot 2^{-3|\mathbf{l}|_1} \cdot |u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}|_{\mathbf{2},2}^2 \cdot \left(\frac{2}{3} \right)^d \cdot 2^{-|\mathbf{l}|_1} \\ &= 9^{-d} \cdot 2^{-4|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},2}^2. \end{aligned}$$

Finally, an analogous argument provides

$$\begin{aligned} \|u_{\mathbf{l}}\|_E^2 &= \left\| \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} v_{\mathbf{l},\mathbf{i}} \cdot \phi_{\mathbf{l},\mathbf{i}} \right\|_E^2 \\ &= \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} |v_{\mathbf{l},\mathbf{i}}|^2 \cdot \|\phi_{\mathbf{l},\mathbf{i}}\|_E^2 \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{\mathbf{i} \in \mathbf{I}_1} \frac{1}{4^d} \cdot 2^{-4 \cdot |\mathbf{l}|_1} \cdot |u|_{2,\infty}^2 \cdot 2 \cdot \left(\frac{2}{3}\right)^{d-1} \cdot 2^{-|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2 \cdot l_j}\right) \\
&= \frac{1}{2 \cdot 6^{d-1}} \cdot 2^{-5 \cdot |\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2 \cdot l_j}\right) \cdot \sum_{\mathbf{i} \in \mathbf{I}_1} |u|_{2,\infty}^2 \\
&= \frac{1}{4 \cdot 12^{d-1}} \cdot 2^{-4 \cdot |\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 2^{2 \cdot l_j}\right) \cdot |u|_{2,\infty}^2
\end{aligned}$$

as well as the second estimate for $\|u_{\mathbf{l}}\|_E$. \square

In the next section, the information gathered above will be used to construct finite-dimensional approximation spaces U for V or $X_0^{q,2}(\bar{\Omega})$, respectively. Such a U shall be based on a subspace selection $\mathbf{I} \subset \mathbb{N}^d$,

$$U := \bigoplus_{\mathbf{l} \in \mathbf{I}} W_{\mathbf{l}}, \quad (3.28)$$

with corresponding interpolants or approximants

$$u_U := \sum_{\mathbf{l} \in \mathbf{I}} u_{\mathbf{l}}, \quad u_{\mathbf{l}} \in W_{\mathbf{l}}. \quad (3.29)$$

The estimate

$$\|u - u_U\| = \left\| \sum_{\mathbf{l}} u_{\mathbf{l}} - \sum_{\mathbf{l} \notin \mathbf{I}} u_{\mathbf{l}} \right\| \leq \sum_{\mathbf{l} \notin \mathbf{I}} \|u_{\mathbf{l}}\| \leq \sum_{\mathbf{l} \notin \mathbf{I}} b(\mathbf{l}) \cdot |u| \quad (3.30)$$

will 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 $W_{\mathbf{l}}$.

3.2. Sparse grids

Interpolation in finite-dimensional spaces

The hierarchical multilevel splitting introduced in the previous section brings along a whole family of hierarchical subspaces $W_{\mathbf{l}}$ of V . However, for discretization purposes, we are more interested in decompositions of finite-dimensional subspaces of V or $X_0^{q,2}(\bar{\Omega})$ than in the splitting of V itself. Therefore, we now turn to finite sets \mathbf{I} of active levels \mathbf{l} in the summation (3.19). For some $n \in \mathbb{N}$, for instance, one possibility $V_n^{(\infty)}$ has already been mentioned in (3.17). The finite-dimensional $V_n^{(\infty)}$ is just the usual space of piecewise d -linear functions on the rectangular grid $\Omega_{n \cdot \mathbf{1}} = \Omega_{(n, \dots, n)}$ with equidistant mesh size $h_n = 2^{-n}$ in each coordinate direction. In a scheme of

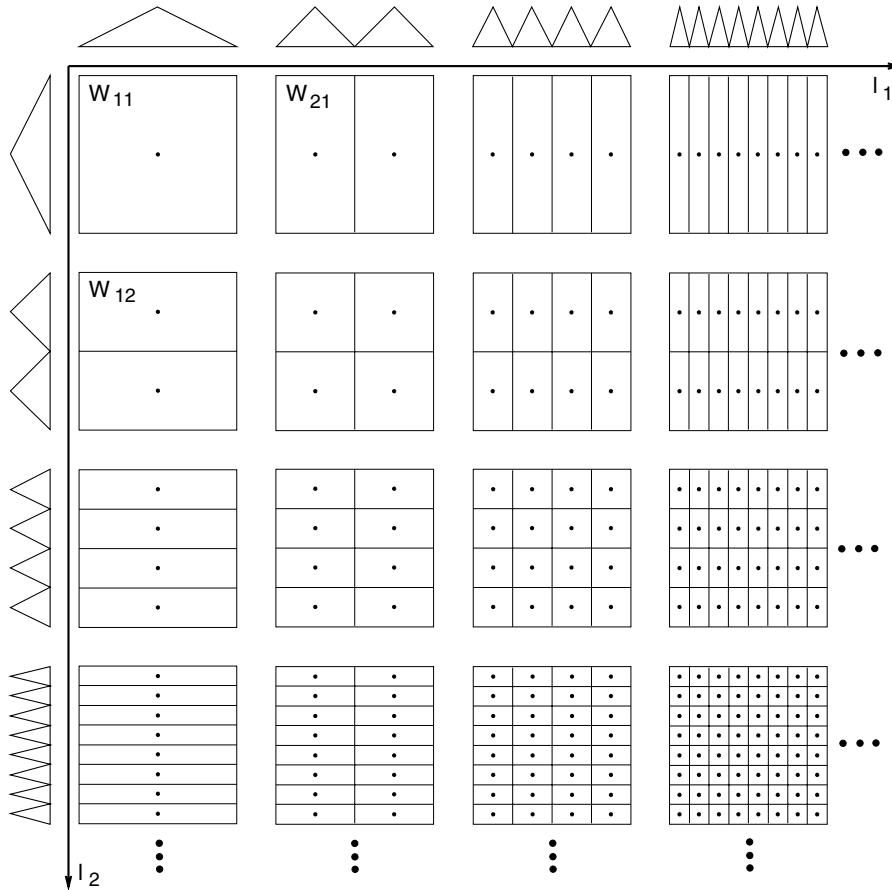


Figure 3.3. Scheme of subspaces for $d = 2$. Each square represents one subspace W_1 with its associated grid points. The supports of the corresponding basis functions have the same mesh size \mathbf{h}_1 and cover the domain Ω .

subspaces W_1 as shown in Figure 3.3 for the 2D case, $V_n^{(\infty)}$ corresponds to a square sector of subspaces: see Figure 3.4.

Obviously, the dimension of $V_n^{(\infty)}$ (*i.e.*, the number of inner grid points in the underlying grid) is

$$|V_n^{(\infty)}| = (2^n - 1)^d = O(2^{d \cdot n}) = O(h_n^{-d}). \quad (3.31)$$

For the error $u - u_n^{(\infty)}$ of the interpolant $u_n^{(\infty)} \in V_n^{(\infty)}$ of a given function $u \in X_0^{q,2}(\bar{\Omega})$ with respect to the different norms we are interested in, the following lemma states the respective results.

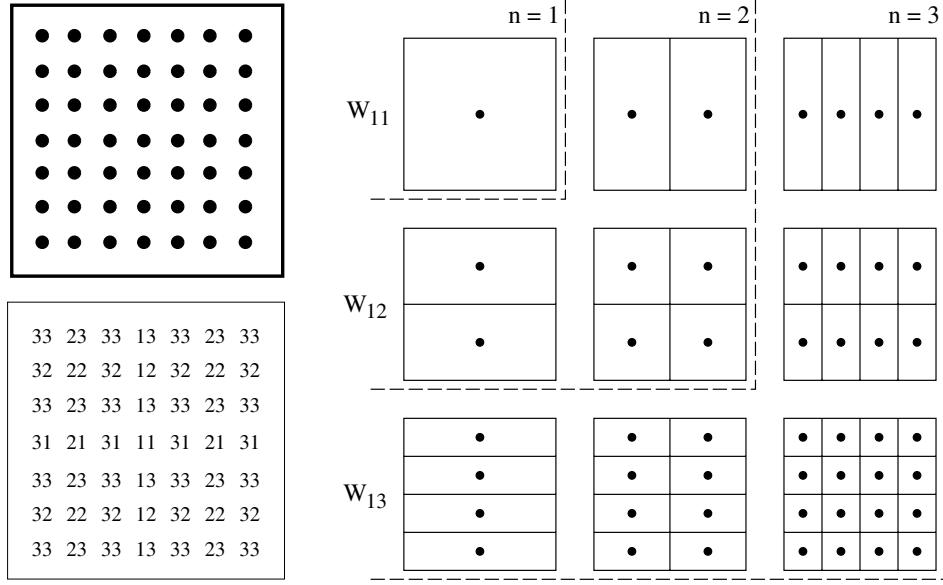


Figure 3.4. The (full) grid of $V_3^{(\infty)}$, $d = 2$, and the assignment of grid points to subspaces.

Lemma 3.5. For $u \in X_0^{q,2}(\bar{\Omega})$, the following estimates for the different norms of the interpolation error $u - u_n^{(\infty)}$, $u_n^{(\infty)} \in V_n^{(\infty)}$, hold:

$$\begin{aligned} \|u - u_n^{(\infty)}\|_\infty &\leq \frac{d}{6^d} \cdot 2^{-2n} \cdot |u|_{2,\infty} &= O(h_n^2), \\ \|u - u_n^{(\infty)}\|_2 &\leq \frac{d}{9^d} \cdot 2^{-2n} \cdot |u|_{2,2} &= O(h_n^2), \\ \|u - u_n^{(\infty)}\|_E &\leq \frac{d^{3/2}}{2 \cdot 3^{(d-1)/2} \cdot 6^{d-1}} \cdot 2^{-n} \cdot |u|_{2,\infty} = O(h_n), \\ \|u - u_n^{(\infty)}\|_E &\leq \frac{d^{3/2}}{\sqrt{3} \cdot 9^{d-1}} \cdot 2^{-n} \cdot |u|_{2,2} &= O(h_n). \end{aligned} \quad (3.32)$$

Proof. For the L_∞ -norm, (3.27) provides

$$\|u - u_n^{(\infty)}\|_\infty \leq \sum_{|\mathbf{l}|_\infty > n} \|u_{\mathbf{l}}\|_\infty \leq \frac{1}{2^d} \cdot |u|_{2,\infty} \cdot \sum_{|\mathbf{l}|_\infty > n} 2^{-2 \cdot |\mathbf{l}|_1},$$

from which we get

$$\begin{aligned}
\|u - u_n^{(\infty)}\|_\infty &\leq \frac{1}{2^d} \cdot |u|_{2,\infty} \cdot \left(\sum_{\mathbf{l}} 4^{-|\mathbf{l}|_1} - \sum_{|\mathbf{l}|_\infty \leq n} 4^{-|\mathbf{l}|_1} \right) \\
&= \frac{1}{2^d} \cdot |u|_{2,\infty} \cdot \left(\left(\frac{1}{3}\right)^d - \left(\sum_{i=1}^n 4^{-i}\right)^d \right) \\
&= \frac{1}{2^d} \cdot |u|_{2,\infty} \cdot \left(\frac{1}{3}\right)^d \cdot (1 - (1 - 4^{-n})^d) \\
&\leq \frac{d}{6^d} \cdot |u|_{2,\infty} \cdot 4^{-n}.
\end{aligned}$$

The respective result for the L_2 -norm can be obtained in exactly the same way. For the error with respect to the energy norm, (3.27) leads to

$$\begin{aligned}
\|u - u_n^{(\infty)}\|_E &\leq \frac{d^{1/2}}{2 \cdot 12^{(d-1)/2}} \cdot |u|_{2,\infty} \cdot \sum_{|\mathbf{l}|_\infty > n} 2^{-2 \cdot |\mathbf{l}|_1} \cdot \max_{1 \leq j \leq d} 2^{l_j} \\
&\leq \frac{d^{3/2}}{2 \cdot 12^{(d-1)/2}} \cdot |u|_{2,\infty} \cdot \sum_{|\mathbf{l}|_\infty = l_1 > n} 2^{-2 \cdot |\mathbf{l}|_1} \cdot 2^{l_1} \\
&= \frac{d^{3/2}}{2 \cdot 12^{(d-1)/2}} \cdot |u|_{2,\infty} \cdot \sum_{l_1 > n} 2^{-l_1} \cdot \left(\sum_{l_j=1}^{l_1} 4^{-l_j}\right)^{d-1} \\
&\leq \frac{d^{3/2}}{2 \cdot 12^{(d-1)/2}} \cdot |u|_{2,\infty} \cdot \frac{1}{3^{d-1}} \cdot 2^{-n},
\end{aligned}$$

and an analogous argument provides the second estimate. \square

It is important to note that we get the same order of accuracy as in standard approximation theory, although our regularity assumptions differ from those normally used there.

Equations (3.31) and (3.32) clearly reveal the crucial drawback of $V_n^{(\infty)}$, the *curse of dimensionality* discussed in detail in Section 2. With d increasing, the number of degrees of freedom that are necessary to achieve an accuracy of $O(h)$ or $O(h^2)$, respectively, grows exponentially. Therefore, we ask how to construct discrete approximation spaces that are better than $V_n^{(\infty)}$ in the sense that the same number of invested grid points leads to a higher order of accuracy. Hence, in the following, we look for an optimum $V^{(\text{opt})}$ by solving a restricted optimization problem of the type

$$\max_{u \in X_0^{q,2}: |u|=1} \|u - u_{V^{(\text{opt})}}\| = \min_{U \subset V: |U|=w} \max_{u \in X_0^{q,2}: |u|=1} \|u - u_U\| \quad (3.33)$$

for some prescribed cost or work count w . The aim is to profit from a given work count as much as possible. 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 work count possible. This is in fact the point of view of computational complexity. Of course, any potential solution $V^{(\text{opt})}$ of (3.33) 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. According to our hierarchical setting, we will allow discrete spaces of the type $U := \bigoplus_{I \in \mathbf{I}} W_I$ for an arbitrary finite index set $\mathbf{I} \subset \mathbb{N}^d$ as candidates for the optimization process only.

An approach such as (3.33) selects certain W_I due to their importance, and thus selects the respective underlying grid points. Depending on the invested work count w , we can expect to get some kind of regular structure or grid patterns. However, in contrast to adaptive grid refinement, which is highly problem-dependent, such a proceeding simply depends 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. Although such *a priori optimization* strategies are not very widespread in the context of PDEs, there is a long tradition in approximation theory and numerical quadrature. For example, think of the Gauss quadrature rules where the grid points are chosen as the roots of certain classes of orthogonal polynomials; *cf.* Krommer and Ueberhuber (1994), for instance. Compared with equidistant quadrature rules based on polynomial interpolants with the same number n of grid points, the degree of accuracy, *i.e.*, the maximum polynomial degree up to which a numerical quadrature rule provides exact results, can be augmented from at most n to $2n - 1$.

Another nice example of the usefulness of an *a priori* grid optimization in numerical quadrature is provided by the *Koksma–Hlawka inequality* (Hlawka 1961, Krommer and Ueberhuber 1994), which says that, for every quadrature formula Q_n based on a simple averaging of samples,

$$Q_n u := \frac{1}{n} \cdot \sum_{i=1}^n u(\mathbf{x}_i), \quad (3.34)$$

that is used to get an approximation to $Iu := \int_{\bar{\Omega}} u(\mathbf{x}) d\mathbf{x}$, a sharp error bound is given by

$$|Q_n u - Iu| \leq V(u) \cdot D_n^*(\mathbf{x}_1, \dots, \mathbf{x}_n). \quad (3.35)$$

Here, $V(u)$ is the so-called *variation of u in the sense of Hardy and Krause*, a property of u indicating the global smoothness of u : the smoother u is on $\bar{\Omega}$, the smaller the values of $V(u)$. $D_n^*(\mathbf{x}_1, \dots, \mathbf{x}_n)$ denotes the so-called *star discrepancy* of the grid $(\mathbf{x}_1, \dots, \mathbf{x}_n)$, which measures the deviation of a

finite part of the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ from the uniform distribution, and is defined by

$$D_n^*(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sup_{E \in \mathcal{E}} \left| \frac{1}{n} \cdot \sum_{i=1}^n \chi_E(\mathbf{x}_i) - \int_{\bar{\Omega}} \chi_E(\mathbf{x}) d\mathbf{x} \right|, \quad (3.36)$$

where $\mathcal{E} := \{[0, e_1[\times \dots \times [0, e_d[\subset \bar{\Omega}\}$ is the set of all subcubes of $\bar{\Omega}$ with $\mathbf{0}$ as a corner and χ_E is the characteristic function of $E \in \mathcal{E}$. Although we do not want to go into detail here, we emphasize the crucial point of (3.35). The quadrature error $|Q_n u - Iu|$ divides into two parts: a *problem-dependent* one (the variation of u) unaffected by the grid, and a *grid-dependent* one (the star discrepancy) uninfluenced by the actual problem (*i.e.*, u). This clearly shows the benefit of the two optimization strategies mentioned above: the construction of grids of low discrepancy (*low-discrepancy formulas*) reduces the second factor in (3.35), whereas adaptive grid refinement can help to concentrate further grid points on subregions of $\bar{\Omega}$ with a (locally) high variation. After this digression into numerical quadrature, we return to our actual topic, the solution of (3.33) in our hierarchical subspace setting.

Formal derivation and properties of sparse grids

As already mentioned, the candidates for $V^{(\text{opt})}$ are finite sets of $W_{\mathbf{l}}$. Therefore, spaces $U := \bigoplus_{\mathbf{l} \in \mathbf{I}} W_{\mathbf{l}}$, the respective grids, and the underlying index sets $\mathbf{I} \subset \mathbb{N}^d$ have to be identified. There are two obvious ways to tackle such problems: a continuous one based on an analytical approach where the multi-index \mathbf{l} is generalized to a nonnegative real one, and a discrete one which uses techniques known from combinatorial optimization; for details, we refer the reader to Bungartz (1998).

Continuous optimization

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{R}_+^d , and a hierarchical subspace $W_{\mathbf{l}}$ just corresponds to a point $\mathbf{l} \in \mathbb{R}_+^d$.

First we have to formulate the optimization problem (3.33). To this end, and inspired by (3.20), the *local cost function* $c(\mathbf{l})$ is defined as a straightforward generalization of the number of degrees of freedom involved:

$$c(\mathbf{l}) := 2^{|\mathbf{l}|_1 - d} = 2^{l_1 + \dots + l_d - d}. \quad (3.37)$$

For the *local benefit function* $b(\mathbf{l})$, we use the squared upper bounds for $\|u_{\mathbf{l}}\|$ according to (3.27). At the moment, we do not fix the norm to be used here. Obviously, the search for an optimal $\mathbf{I} \subset \mathbb{R}_+^d$ can be restricted to $\mathbf{I} \subset \mathbf{I}^{(\text{max})} := [0, N]^d$ for a sufficiently large N without loss of generality. Based on the two local quantities $c(\mathbf{l})$ and $b(\mathbf{l})$, the *global cost* $C(\mathbf{I})$ and the

global benefit $B(\mathbf{I})$ of a grid \mathbf{I} are defined by

$$C(I) := \int_{\mathbf{I}} c(\mathbf{l}) d\mathbf{l}, \quad B(I) := \int_{\mathbf{I}} b(\mathbf{l}) d\mathbf{l}. \quad (3.38)$$

This leads to the desired restricted optimization problem according to (3.33):

$$\max_{C(\mathbf{I})=w} B(\mathbf{I}). \quad (3.39)$$

For the solution of (3.39), we start from an arbitrary $\mathbf{I} \subset \mathbf{I}^{(\max)}$ that has a sufficiently smooth boundary $\partial\mathbf{I}$. With a sufficiently smooth mapping τ ,

$$\tau : \mathbb{R}_+^d \rightarrow \mathbb{R}_+^d, \quad \tau(\mathbf{l}) = \mathbf{0} \quad \text{for } \mathbf{l} \in \partial\mathbb{R}_+^d, \quad (3.40)$$

we define a small disturbance $\varphi_{\varepsilon,\tau}$ of the grid \mathbf{I} :

$$\varphi_{\varepsilon,\tau} : \mathbf{I} \rightarrow \mathbf{I}_{\varepsilon,\tau} \subset \mathbf{I}^{(\max)}, \quad \varphi_{\varepsilon,\tau}(\mathbf{l}) := \mathbf{l} + \varepsilon \cdot \tau(\mathbf{l}), \quad \varepsilon \in \mathbb{R}. \quad (3.41)$$

For the global cost of the disturbed grid $\mathbf{I}_{\varepsilon,\tau}$, we get

$$C(\mathbf{I}_{\varepsilon,\tau}) = \int_{\mathbf{I}_{\varepsilon,\tau}} c(\mathbf{k}) d\mathbf{k} = \int_{\mathbf{I}} c(\mathbf{l} + \varepsilon \cdot \tau(\mathbf{l})) \cdot |\det D\varphi_{\varepsilon,\tau}| d\mathbf{l}. \quad (3.42)$$

Taylor expansion of $c(\mathbf{l} + \varepsilon \cdot \tau(\mathbf{l}))$ in $\varepsilon = 0$ provides

$$c(\mathbf{l} + \varepsilon \cdot \tau(\mathbf{l})) = c(\mathbf{l}) + \varepsilon \cdot \nabla c(\mathbf{l}) \cdot \tau(\mathbf{l}) + O(\varepsilon^2), \quad (3.43)$$

where $\nabla c(\mathbf{l}) \cdot \tau(\mathbf{l})$ denotes the scalar product. Furthermore, a straightforward calculation shows

$$|\det D\varphi_{\varepsilon,\tau}| = 1 + \varepsilon \cdot \operatorname{div} \tau + O(\varepsilon^2). \quad (3.44)$$

Thus, since $\mathbf{I} \subset \mathbf{I}^{(\max)}$ with $\mathbf{I}^{(\max)}$ bounded, Gauss's theorem leads to

$$C(\mathbf{I}_{\varepsilon,\tau}) = C(\mathbf{I}) + \varepsilon \cdot \int_{\partial\mathbf{I}} c(\mathbf{l}) \cdot \tau(\mathbf{l}) d\vec{S} + O(\varepsilon^2). \quad (3.45)$$

Consequently, for the derivative with respect to ε , we get

$$\frac{\partial C(\mathbf{I}_{\varepsilon,\tau})}{\partial \varepsilon} \Big|_{\varepsilon=0} = \lim_{\varepsilon \rightarrow 0} \frac{C(\mathbf{I}_{\varepsilon,\tau}) - C(\mathbf{I})}{\varepsilon} = \int_{\partial\mathbf{I}} c(\mathbf{l}) \cdot \tau(\mathbf{l}) d\vec{S}. \quad (3.46)$$

Similar arguments hold for the global benefit $B(\mathbf{I})$ and result in

$$\frac{\partial B(\mathbf{I}_{\varepsilon,\tau})}{\partial \varepsilon} \Big|_{\varepsilon=0} = \lim_{\varepsilon \rightarrow 0} \frac{B(\mathbf{I}_{\varepsilon,\tau}) - B(\mathbf{I})}{\varepsilon} = \int_{\partial\mathbf{I}} b(\mathbf{l}) \cdot \tau(\mathbf{l}) d\vec{S}. \quad (3.47)$$

Now, starting from the *optimal* grid $\mathbf{I}^{(\text{opt})}$, Lagrange's principle for the optimization under a constraint can be applied, and we get

$$\lambda \cdot \int_{\partial\mathbf{I}^{(\text{opt})}} c(\mathbf{l}) \cdot \tau(\mathbf{l}) d\vec{S} = \int_{\partial\mathbf{I}^{(\text{opt})}} b(\mathbf{l}) \cdot \tau(\mathbf{l}) d\vec{S}. \quad (3.48)$$

Since τ vanishes on the boundary of \mathbb{R}_+^d , *i.e.*, $\tau(\mathbf{l}) = \mathbf{0}$ when any component of \mathbf{l} vanishes, (3.48) is equivalent to

$$\lambda \cdot \int_{\partial \mathbf{I}^{(\text{opt})} \setminus \partial \mathbb{R}_+^d} c(\mathbf{l}) \cdot \tau(\mathbf{l}) \, d\vec{S} = \int_{\partial \mathbf{I}^{(\text{opt})} \setminus \partial \mathbb{R}_+^d} b(\mathbf{l}) \cdot \tau(\mathbf{l}) \, d\vec{S}. \quad (3.49)$$

Finally, since (3.49) is valid for all appropriate smooth disturbances τ ,

$$\lambda \cdot c(\mathbf{l}) = b(\mathbf{l}) \quad (3.50)$$

holds for all $\mathbf{l} \in \partial \mathbf{I}^{(\text{opt})} \setminus \partial \mathbb{R}_+^d$.

This is a quite interesting result, because (3.50) says that the ratio of the local benefit $b(\mathbf{l})$ to the local cost $c(\mathbf{l})$ is constant on the boundary $\partial \mathbf{I}^{(\text{opt})} \setminus \partial \mathbb{R}_+^d$ of any grid $\mathbf{I}^{(\text{opt})}$ that is optimal in our sense. This means that the *global* optimization process (3.33) or (3.39), respectively, in which we look for an optimal grid can be reduced to studying the *local* cost–benefit ratios $b(\mathbf{l})/c(\mathbf{l})$ of the subspaces associated with \mathbf{l} . Therefore, if we come back to real hierarchical subspaces W_1 and to indices $\mathbf{l} \in \mathbb{N}^d$, all one has to do is to identify sets of subspaces W_1 with constant cost–benefit ratio in the subspace scheme of Figure 3.3. The grid $\mathbf{I}^{(\text{opt})}$, then, contains the region where the cost–benefit ratio is bigger than or equal to the constant value on the boundary $\partial \mathbf{I}^{(\text{opt})} \setminus \partial \mathbb{R}_+^d$.

Discrete optimization

Since the above continuous optimization process with its roundabout way of generalizing integer multi-indices to real ones is a bit unnatural, (3.33) is now formulated as a discrete optimization problem.

First of all, we redefine the local functions $c(\mathbf{l})$ and $b(\mathbf{l})$, now for multi-indices $\mathbf{l} \in \mathbb{N}^d$ only. According to (3.20), the local cost $c(\mathbf{l})$ is defined by

$$c(\mathbf{l}) := |W_1| = 2^{|\mathbf{l}-\mathbf{1}|_1}, \quad (3.51)$$

which is exactly the same as (3.37) restricted to $\mathbf{l} \in \mathbb{N}^d$. Obviously, $c(\mathbf{l}) \in \mathbb{N}$ holds for all $\mathbf{l} \in \mathbb{N}^d$. Concerning the local benefit function, we define

$$b(\mathbf{l}) := \gamma \cdot \beta(\mathbf{l}), \quad (3.52)$$

where $\beta(\mathbf{l})$ is an upper bound for $\|u_{\mathbf{l}}\|^2$ according to (3.27), and γ is a factor depending on the problem’s dimensionality d and on the smoothness of the data, *i.e.*, of u , but constant with respect to \mathbf{l} , such that $b(\mathbf{l}) \in \mathbb{N}$. The respective bounds in (3.27) show that such a choice of γ is possible for each of the three norms that are of interest in our context. Note that, as in the continuous case, we do not make any decision concerning the actual choice of norm to be used for $b(\mathbf{l})$ for the moment.

Again, the search for an optimal grid $\mathbf{I} \subset \mathbb{N}^d$ can be restricted to all $\mathbf{I} \subset \mathbf{I}^{(\text{max})} := \{1, \dots, N\}^d$ for a sufficiently large N without loss of generality.

Next, the global cost benefit functions are redefined as well. For $C(\mathbf{I})$, we define

$$C(\mathbf{I}) := \sum_{\mathbf{l} \in \mathbf{I}} c(\mathbf{l}) = \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} x(\mathbf{l}) \cdot c(\mathbf{l}), \quad (3.53)$$

where

$$x(\mathbf{l}) := \begin{cases} 0, & \mathbf{l} \notin \mathbf{I}, \\ 1, & \mathbf{l} \in \mathbf{I}. \end{cases} \quad (3.54)$$

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

$$\begin{aligned} \left\| u - \sum_{\mathbf{l} \in \mathbf{I}} u_{\mathbf{l}} \right\|^2 &\approx \left\| \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} u_{\mathbf{l}} - \sum_{\mathbf{l} \in \mathbf{I}} u_{\mathbf{l}} \right\|^2 \\ &\leq \sum_{\mathbf{l} \in \mathbf{I}^{(\max)} \setminus \mathbf{I}} \|u_{\mathbf{l}}\|^2 \\ &\leq \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} (1 - x(\mathbf{l})) \cdot \gamma \cdot \beta(\mathbf{l}) \\ &= \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} \gamma \cdot \beta(\mathbf{l}) - \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} x(\mathbf{l}) \cdot \gamma \cdot \beta(\mathbf{l}) \\ &=: \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} \gamma \cdot \beta(\mathbf{l}) - B(\mathbf{I}). \end{aligned} \quad (3.55)$$

Of course, (3.55) 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 big as appropriate, this is not a serious restriction. Altogether, we get the following reformulation of (3.33):

$$\max_{\mathbf{I} \subset \mathbf{I}^{(\max)}} \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} x(\mathbf{l}) \cdot \gamma \cdot \beta(\mathbf{l}) \quad \text{with} \quad \sum_{\mathbf{l} \in \mathbf{I}^{(\max)}} x(\mathbf{l}) \cdot c(\mathbf{l}) = w. \quad (3.56)$$

If we arrange the $\mathbf{l} \in \mathbf{I}^{(\max)}$ in some linear order (*e.g.*, a lexicographical one) with local cost c_i and benefit b_i , $i = 1, \dots, N^d =: M$, (3.56) reads as

$$\max_{\mathbf{x}} \mathbf{b}^T \mathbf{x} \quad \text{with} \quad \mathbf{c}^T \mathbf{x} = w, \quad (3.57)$$

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 (3.57) is called a *binary knapsack problem* (Martello and Toth 1990), which is known to be an NP-hard one. 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 $\mathbf{x} \in ([0, 1] \cap \mathbb{Q})^M$:

- (1) rearrange the order such that $\frac{b_1}{c_1} \geq \frac{b_2}{c_2} \dots \geq \frac{b_M}{c_M}$;
- (2) let $r := \max\{j : \sum_{i=1}^j c_i \leq w\}$;
- (3) $x_1 := \dots := x_r := 1$,
 $x_{r+1} := (w - \sum_{i=1}^r c_i)/c_{r+1}$,
 $x_{r+2} := \dots := x_M := 0$.

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, as in the continuous case before, the *global* optimization problem (3.33) or (3.57), 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. This is in the same spirit as n -term approximation (DeVore 1998).

*L*₂-based sparse grids

Owing to (3.27), the L_2 - and L_∞ -norm of $W_{\mathbf{l}}$'s contribution $u_{\mathbf{l}}$ to the hierarchical representation (3.19) of $u \in X_0^{q,2}(\bar{\Omega})$ are of the same order of magnitude. Therefore there are no differences in the character of the cost–benefit ratio, and the same optimal grids $\mathbf{I}^{(\text{opt})}$ will result from the optimization process described above. According to (3.20) and (3.27), we define

$$\begin{aligned} \text{cbr}_\infty(\mathbf{l}) &:= \frac{b_\infty(\mathbf{l})}{c(\mathbf{l})} := \frac{2^{-4|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},\infty}^2}{4^d \cdot 2^{|\mathbf{l}-\mathbf{1}|_1}} = \frac{1}{2^d} \cdot 2^{-5|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},\infty}^2, \\ \text{cbr}_2(\mathbf{l}) &:= \frac{b_2(\mathbf{l})}{c(\mathbf{l})} := \frac{2^{-4|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},2}^2}{9^d \cdot 2^{|\mathbf{l}-\mathbf{1}|_1}} = \left(\frac{2}{9}\right)^d \cdot 2^{-5|\mathbf{l}|_1} \cdot |u|_{\mathbf{2},2}^2 \end{aligned} \quad (3.58)$$

as the local cost–benefit ratios. Note that we use bounds for the squared norms of $u_{\mathbf{l}}$ for reasons of simplicity, but without loss of generality An optimal grid $\mathbf{I}^{(\text{opt})}$ will consist of all multi-indices \mathbf{l} or their corresponding subspaces $W_{\mathbf{l}}$ where $\text{cbr}_\infty(\mathbf{l})$ or $\text{cbr}_2(\mathbf{l})$ is bigger than some prescribed threshold $\sigma_\infty(n)$ or $\sigma_2(n)$, respectively. We choose those thresholds to be of the order of $\text{cbr}_\infty(\bar{\mathbf{l}})$ or $\text{cbr}_2(\bar{\mathbf{l}})$ with $\bar{\mathbf{l}} := (n, 1, \dots, 1)$:

$$\begin{aligned} \sigma_\infty(n) &:= \text{cbr}_\infty(\bar{\mathbf{l}}) = \frac{1}{2^d} \cdot 2^{-5(n+d-1)} \cdot |u|_{\mathbf{2},\infty}^2, \\ \sigma_2(n) &:= \text{cbr}_2(\bar{\mathbf{l}}) = \left(\frac{2}{9}\right)^d \cdot 2^{-5(n+d-1)} \cdot |u|_{\mathbf{2},2}^2. \end{aligned} \quad (3.59)$$

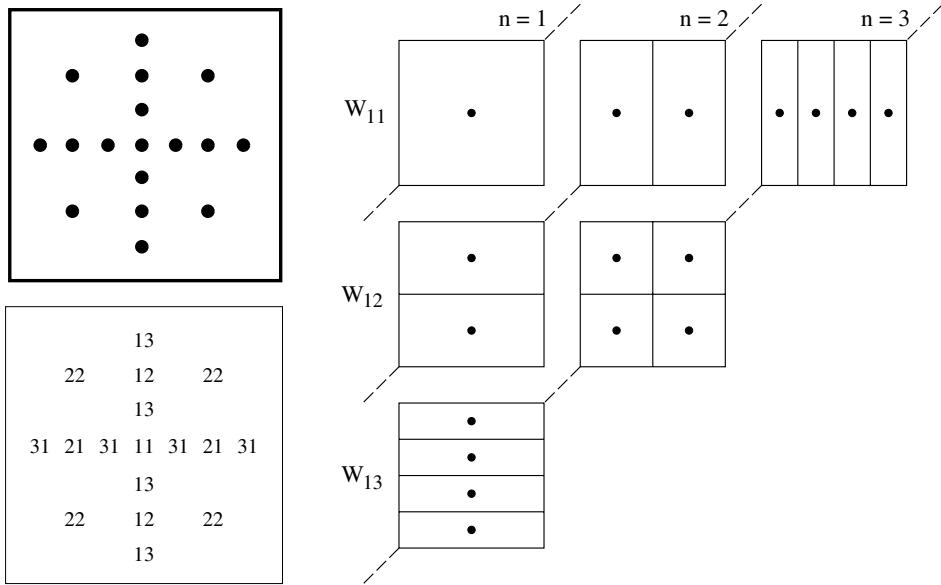


Figure 3.5. The sparse grid of $V_3^{(1)}$, $d = 2$, and the assignment of grid points to subspaces.

That is, we fix d subspaces on the axes in the subspace scheme of Figure 3.3 and search for all W_1 whose cost–benefit ratio is equal or better. Thus, applying the criterion $\text{cbr}_\infty(\mathbf{l}) \geq \sigma_\infty(n)$ or $\text{cbr}_2(\mathbf{l}) \geq \sigma_2(n)$, respectively, we get the relation

$$|\mathbf{l}|_1 \leq n + d - 1 \quad (3.60)$$

that qualifies a subspace W_1 to be taken into account. This result leads us to the definition of a new discrete approximation space $V_n^{(1)}$,

$$V_n^{(1)} := \bigoplus_{|\mathbf{l}|_1 \leq n+d-1} W_1, \quad (3.61)$$

which is L_∞ - and L_2 -optimal with respect to our cost–benefit setting. The grids that correspond to the spaces $V_n^{(1)}$ are just the standard *sparse grids* as were introduced in Zenger (1991), studied in detail in Bungartz (1992b), and discussed in a variety of other papers for different applications. In comparison with the standard *full grid* space $V_n^{(\infty)}$, we now have triangular or simplicial sectors of subspaces in the scheme of Figure 3.3: see Figure 3.5. Figure 3.6, finally, gives two examples of sparse grids: a regular 2D and an adaptively refined 3D one.

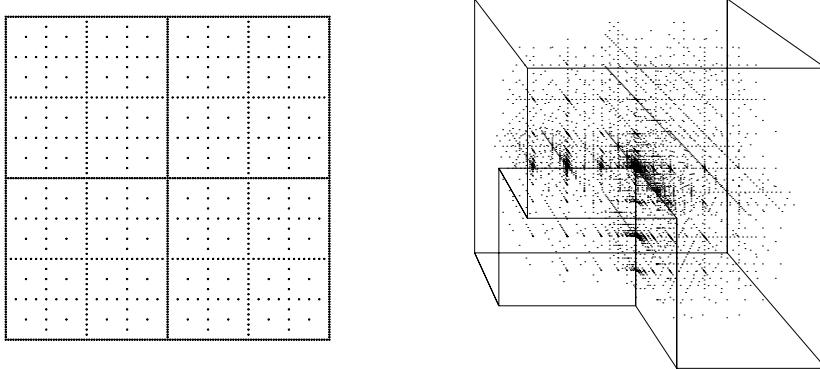


Figure 3.6. Sparse grids: regular example (left) and adaptive one (right).

Now, let us turn to the basic properties of the sparse grid approximation spaces $V_n^{(1)}$.

Lemma 3.6. The dimension of the space $V_n^{(1)}$, i.e., the number of degrees of freedom or inner grid points, is given by

$$\begin{aligned} |V_n^{(1)}| &= \sum_{i=0}^{n-1} 2^i \cdot \binom{d-1+i}{d-1} \\ &= (-1)^d + 2^n \cdot \sum_{i=0}^{d-1} \binom{n+d-1}{i} \cdot (-2)^{d-1-i} \\ &= 2^n \cdot \left(\frac{n^{d-1}}{(d-1)!} + O(n^{d-2}) \right). \end{aligned} \quad (3.62)$$

Thus, we have

$$|V_n^{(1)}| = O(h_n^{-1} \cdot |\log_2 h_n|^{d-1}). \quad (3.63)$$

Proof. With (3.20) and (3.61), we get

$$\begin{aligned} |V_n^{(1)}| &= \left| \bigoplus_{|\mathbf{l}|_1 \leq n+d-1} W_{\mathbf{l}} \right| = \sum_{|\mathbf{l}|_1 \leq n+d-1} 2^{|\mathbf{l}-\mathbf{1}|_1} = \sum_{i=d}^{n+d-1} 2^{i-d} \cdot \sum_{|\mathbf{l}|_1=i} 1 \\ &= \sum_{i=d}^{n+d-1} 2^{i-d} \cdot \binom{i-1}{d-1} \\ &= \sum_{i=0}^{n-1} 2^i \cdot \binom{d-1+i}{d-1}, \end{aligned}$$

since there are $\binom{i-1}{d-1}$ ways to form the sum i with d nonnegative integers. Furthermore,

$$\begin{aligned}
& \sum_{i=0}^{n-1} 2^i \cdot \binom{d-1+i}{d-1} \\
&= \frac{1}{(d-1)!} \cdot \sum_{i=0}^{n-1} (x^{i+d-1})^{(d-1)} \Big|_{x=2} \\
&= \frac{1}{(d-1)!} \cdot \left(x^{d-1} \cdot \frac{1-x^n}{1-x} \right)^{(d-1)} \Big|_{x=2} \\
&= \frac{1}{(d-1)!} \cdot \sum_{i=0}^{d-1} \binom{d-1}{i} \cdot (x^{d-1} - x^{n+d-1})^{(i)} \cdot \left(\frac{1}{1-x} \right)^{(d-1-i)} \Big|_{x=2} \\
&= (-1)^d + 2^n \cdot \sum_{i=0}^{d-1} \binom{n+d-1}{i} \cdot (-2)^{d-1-i},
\end{aligned}$$

from which the result concerning the order and the leading coefficient follows immediately. \square

The above lemma shows the order $O(2^n \cdot n^{d-1})$ or, with $h_n = 2^{-n}$, $O(h_n^{-1} \cdot |\log_2 h_n|^{d-1})$, which is a significant reduction of the number of degrees of freedom and, thus, of the computational and storage requirement compared with the order $O(h_n^{-d})$ of $V_n^{(\infty)}$.

The other question to be discussed concerns the interpolation accuracy that can be obtained on sparse grids. For that, we look at the interpolation error $u - u_n^{(1)}$ of the sparse grid interpolant $u_n^{(1)} \in V_n^{(1)}$ which, due to (3.19) and (3.61), can be written as

$$u - u_n^{(1)} = \sum_{\mathbf{l}} u_{\mathbf{l}} - \sum_{|\mathbf{l}|_1 \leq n+d-1} u_{\mathbf{l}} = \sum_{|\mathbf{l}|_1 > n+d-1} u_{\mathbf{l}}.$$

Therefore, for any norm $\|\cdot\|$, we have

$$\|u - u_n^{(1)}\| \leq \sum_{|\mathbf{l}|_1 > n+d-1} \|u_{\mathbf{l}}\|. \quad (3.64)$$

The following lemma provides a prerequisite for the estimates of the interpolation error with respect to the different norms we are interested in. For $d, n \in \mathbb{N}$, we define

$$A(d, n) := \sum_{k=0}^{d-1} \binom{n+d-1}{k} = \frac{n^{d-1}}{(d-1)!} + O(n^{d-2}). \quad (3.65)$$

Lemma 3.7. For purposes of summation over all grid points $\mathbf{x}_{\mathbf{l},\mathbf{i}}$ with corresponding basis functions $\phi_{\mathbf{l},\mathbf{i}} \notin V_n^{(1)}$, we obtain for arbitrary $s \in \mathbb{N}$

$$\begin{aligned} \sum_{|\mathbf{l}|_1 > n+d-1} 2^{-s \cdot |\mathbf{l}|_1} &= 2^{-s \cdot n} \cdot 2^{-s \cdot d} \cdot \sum_{i=0}^{\infty} 2^{-s \cdot i} \cdot \binom{n+i+d-1}{d-1} \\ &\leq 2^{-s \cdot n} \cdot 2^{-s \cdot d} \cdot 2 \cdot A(d, n). \end{aligned} \quad (3.66)$$

Proof. As for the proof of the previous lemma, we get

$$\begin{aligned} \sum_{|\mathbf{l}|_1 > n+d-1} 2^{-s \cdot |\mathbf{l}|_1} &= \sum_{i=n+d}^{\infty} 2^{-s \cdot i} \cdot \sum_{|\mathbf{l}|_1=i} 1 \\ &= \sum_{i=n+d}^{\infty} 2^{-s \cdot i} \cdot \binom{i-1}{d-1} \\ &= 2^{-s \cdot n} \cdot 2^{-s \cdot d} \cdot \sum_{i=0}^{\infty} 2^{-s \cdot i} \cdot \binom{n+i+d-1}{d-1}. \end{aligned}$$

Since

$$\begin{aligned} \sum_{i=0}^{\infty} x^i \cdot \binom{n+i+d-1}{d-1} &= \frac{x^{-n}}{(d-1)!} \cdot \left(\sum_{i=0}^{\infty} x^{n+i+d-1} \right)^{(d-1)} \\ &= \frac{x^{-n}}{(d-1)!} \cdot \left(x^{n+d-1} \cdot \frac{1}{1-x} \right)^{(d-1)} \\ &= \frac{x^{-n}}{(d-1)!} \cdot \sum_{k=0}^{d-1} \binom{d-1}{k} \cdot (x^{n+d-1})^{(k)} \cdot \left(\frac{1}{1-x} \right)^{(d-1-k)} \\ &= \sum_{k=0}^{d-1} \binom{n+d-1}{k} \cdot \left(\frac{x}{1-x} \right)^{d-1-k} \cdot \frac{1}{1-x}, \end{aligned} \quad (3.67)$$

we get with $x := 2^{-s}$:

$$\sum_{i=0}^{\infty} 2^{-s \cdot i} \cdot \binom{n+i+d-1}{d-1} \leq 2 \cdot \sum_{k=0}^{d-1} \binom{n+d-1}{k} = 2 \cdot A(d, n). \quad \square$$

With the above lemma, we obtain the desired result concerning the interpolation quality of standard sparse grid spaces $V_n^{(1)}$.

Theorem 3.8. For the L_∞ -, the L_2 -, and the energy norm, we have the following upper bounds for the interpolation error of a function $u \in X_0^{q,2}(\bar{\Omega})$ in the sparse grid space $V_n^{(1)}$:

$$\begin{aligned}\|u - u_n^{(1)}\|_\infty &\leq \frac{2 \cdot |u|_{2,\infty}}{8^d} \cdot 2^{-2n} \cdot A(d,n) = O(h_n^2 \cdot n^{d-1}), \\ \|u - u_n^{(1)}\|_2 &\leq \frac{2 \cdot |u|_{2,2}}{12^d} \cdot 2^{-2n} \cdot A(d,n) = O(h_n^2 \cdot n^{d-1}), \\ \|u - u_n^{(1)}\|_E &\leq \frac{d \cdot |u|_{2,\infty}}{2 \cdot 3^{(d-1)/2} \cdot 4^{d-1}} \cdot 2^{-n} = O(h_n), \\ \|u - u_n^{(1)}\|_E &\leq \frac{d \cdot |u|_{2,2}}{\sqrt{3} \cdot 6^{d-1}} \cdot 2^{-n} = O(h_n).\end{aligned}\quad (3.68)$$

Proof. With (3.27), (3.64), and (3.66) for $s = 2$, we get

$$\begin{aligned}\|u - u_n^{(1)}\|_\infty &\leq \sum_{|\mathbf{l}|_1 > n+d-1} \|u_{\mathbf{l}}\|_\infty \leq \frac{|u|_{2,\infty}}{2^d} \cdot \sum_{|\mathbf{l}|_1 > n+d-1} 2^{-2 \cdot |\mathbf{l}|_1} \\ &\leq \frac{2 \cdot |u|_{2,\infty}}{8^d} \cdot 2^{-2n} \cdot A(d,n)\end{aligned}$$

and, analogously, the corresponding result for the L_2 -norm. Concerning the first bound with respect to the energy norm, we have

$$\begin{aligned}\|u - u_n^{(1)}\|_E &\leq \sum_{|\mathbf{l}|_1 > n+d-1} \|u_{\mathbf{l}}\|_E \\ &\leq \frac{|u|_{2,\infty}}{2 \cdot 12^{(d-1)/2}} \cdot \sum_{|\mathbf{l}|_1 > n+d-1} 4^{-|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \\ &= \frac{|u|_{2,\infty}}{2 \cdot 12^{(d-1)/2}} \cdot \sum_{i=n+d}^{\infty} 4^{-i} \cdot \sum_{|\mathbf{l}|_1=i} \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \\ &\leq \frac{|u|_{2,\infty}}{2 \cdot 12^{(d-1)/2}} \cdot \sum_{i=n+d}^{\infty} d \cdot 2^{-i} \\ &= \frac{d \cdot |u|_{2,\infty}}{2 \cdot 3^{(d-1)/2} \cdot 4^{d-1}} \cdot 2^{-n},\end{aligned}$$

because

$$\sum_{|\mathbf{l}|_1=i} \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \leq d \cdot 2^i,$$

which can be shown by complete induction with respect to d . The last estimate can be obtained with analogous arguments. \square

This theorem shows the crucial improvement of the sparse grid space $V_n^{(1)}$ in comparison with $V_n^{(\infty)}$. The number of degrees of freedom is reduced significantly, whereas the accuracy is only slightly deteriorated – for the L_∞ - and the L_2 -norm – or even stays of the same order if the error is measured in the energy norm. This lessens the curse of dimensionality, but it does not overcome it completely. Since this result is optimal with respect to both the L_∞ - and the L_2 -norm, a further improvement can only be expected if we change the setting. Therefore, in the following, we study the optimization process with respect to the energy norm.

Energy-based sparse grids

Now, we base our cost–benefit approach on the energy norm. According to (3.20) and (3.27), we define

$$\begin{aligned} \text{cbr}_E(\mathbf{l}) := \frac{b_E(\mathbf{l})}{c(\mathbf{l})} &:= \frac{2^{-4\cdot|\mathbf{l}|_1} \cdot |u|_{2,\infty}^2}{4 \cdot 12^{d-1} \cdot 2^{|\mathbf{l}-\mathbf{1}|_1}} \cdot \sum_{j=1}^d 4^{l_j} \\ &= \frac{3}{6^d} \cdot 2^{-5\cdot|\mathbf{l}|_1} \cdot \sum_{j=1}^d 4^{l_j} \cdot |u|_{2,\infty}^2 \end{aligned} \quad (3.69)$$

as the local cost–benefit ratio. Again, 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 $\mathbf{I}^{(\text{opt})}$ will consist of all those multi-indices \mathbf{l} or their respective hierarchical subspaces $W_{\mathbf{l}}$ that fulfil $\text{cbr}_E(\mathbf{l}) \geq \sigma_E(n)$ for some given constant threshold $\sigma_E(n)$. As before, $\sigma_E(n)$ is defined via the cost–benefit ratio of $W_{\bar{\mathbf{l}}}$ with $\bar{\mathbf{l}} := (n, 1, \dots, 1)$:

$$\sigma_E(n) := \text{cbr}_E(\bar{\mathbf{l}}) = \frac{3}{6^d} \cdot 2^{-5\cdot(n+d-1)} \cdot (4^n + 4 \cdot (d-1)) \cdot |u|_{2,\infty}^2. \quad (3.70)$$

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

$$V_n^{(E)} := \bigoplus_{\substack{|\mathbf{l}|_1 - \frac{1}{5} \cdot \log_2 \left(\sum_{j=1}^d 4^{l_j} \right) \leq (n+d-1) - \frac{1}{5} \cdot \log_2 (4^n + 4d - 4)}} W_{\mathbf{l}}. \quad (3.71)$$

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

Lemma 3.9. The energy-based sparse grid space $V_n^{(E)}$ is a subspace of $V_n^{(1)}$, and its dimension fulfills

$$|V_n^{(E)}| \leq 2^n \cdot \frac{d}{2} \cdot e^d = O(h_n^{-1}). \quad (3.72)$$

Proof. For subspaces W_1 with $|\mathbf{l}|_1 = n + d - 1 + i$, $i \in \mathbb{N}$, we have

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

Therefore, no W_1 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 $|V_n^{(E)}| \leq |V_n^{(1)}|$ for all $n \in \mathbb{N}$. Starting from that, (3.20) provides

$$\begin{aligned} |V_n^{(E)}| &= \sum_{i=0}^{n-1} \sum_{|\mathbf{l}|_1=n+d-1-i, \sum_{j=1}^d 4^{l_j} \geq \frac{4^n+4d-4}{32^i}} |W_1| \\ &= 2^n \cdot \frac{1}{2} \cdot \sum_{i=0}^{n-1} 2^{-i} \cdot \sum_{|\mathbf{l}|_1=n+d-1-i, \sum_{j=1}^d 4^{l_j} \geq \frac{4^n+4d-4}{32^i}} 1 \\ &\leq 2^n \cdot \frac{1}{2} \cdot \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} 2^{-i} \cdot \sum_{|\mathbf{l}|_1=n+d-1-i, \sum_{j=1}^d 4^{l_j} \geq \frac{4^n+4d-4}{32^i}} 1 \\ &= 2^n \cdot \frac{1}{2} \cdot \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} 2^{-i} \cdot d \cdot \binom{d-1-\lfloor 1.5i \rfloor}{d-1}, \end{aligned}$$

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

Table 3.1. Dimension of $V_n^{(\infty)}$, $V_n^{(1)}$, and $V_n^{(E)}$ for different values of d and n .

	$d = 2$		$d = 3$		$d = 4$	
	$n = 10$	$n = 20$	$n = 10$	$n = 20$	$n = 10$	$n = 20$
$V_n^{(\infty)}$	$1.05 \cdot 10^6$	$1.10 \cdot 10^{12}$	$1.07 \cdot 10^9$	$1.15 \cdot 10^{18}$	$1.10 \cdot 10^{12}$	$1.21 \cdot 10^{24}$
$V_n^{(1)}$	9217	$1.99 \cdot 10^7$	47103	$2.00 \cdot 10^8$	$1.78 \cdot 10^5$	$1.41 \cdot 10^9$
$V_n^{(E)}$	3841	$4.72 \cdot 10^6$	10495	$1.68 \cdot 10^7$	24321	$5.27 \cdot 10^7$

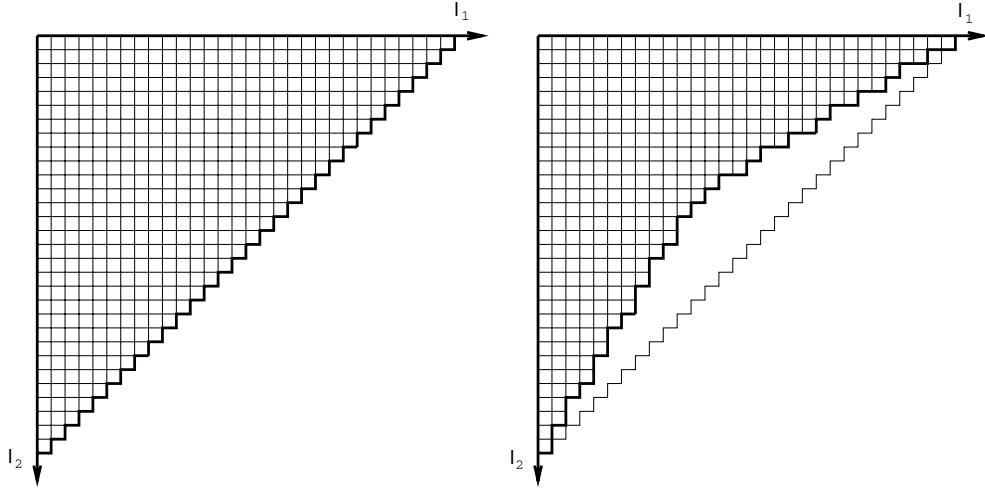


Figure 3.7. Scheme of subspaces for $V_{30}^{(1)}$ (left) and $V_{30}^{(E)}$ (right), $d = 2$.

sum instead of

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

are the same, and since there exist

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

such subspaces W_l with $|l|_\infty = l_1$. Consequently, we obtain

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

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

Table 3.1 compares the dimensions of the standard full grid approximation space $V_n^{(\infty)}$ with both sparse grid spaces $V_n^{(1)}$ and $V_n^{(E)}$ for different dimensionalities $d \in \{2, 3, 4\}$ and for the two resolutions $n = 10$ and $n = 20$. The sparse grid effect is already obvious for $V_n^{(1)}$. Especially for larger d , the advantages of $V_n^{(E)}$ become evident. For a comparison of the underlying subspace schemes of $V_n^{(1)}$ and $V_n^{(E)}$ in 2D, see Figure 3.7.

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 $u_n^{(E)} \in V_n^{(E)}$. Note that we look at the energy norm only, since, with respect to both the L_∞ - and the L_2 -norm, the spaces $V_n^{(1)}$ are already optimal in our cost–benefit setting. Thus, with the reduced number of grid points of $V_n^{(E)}$, a deterioration of the (L_∞ - and L_2 -) interpolation quality is to be expected.

Theorem 3.10. 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

$$\begin{aligned}\|u - u_n^{(E)}\|_E &\leq \frac{d \cdot |u|_{2,\infty}}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2} \right)^{d-1} \right) \cdot 2^{-n} = O(h_n), \quad (3.73) \\ \|u - u_n^{(E)}\|_E &\leq \frac{2 \cdot d \cdot |u|_{2,2}}{\sqrt{3} \cdot 6^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2} \right)^{d-1} \right) \cdot 2^{-n} = O(h_n).\end{aligned}$$

Proof. First, since

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

and since we already 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_1 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 (3.27)

$$\begin{aligned}\|u_n^{(1)} - u_n^{(E)}\|_E &\leq \sum_{W_1 \subseteq V_n^{(1)} \ominus V_n^{(E)}} \|u_1\|_E \\ &\leq \sum_{i=0}^{i^*} \sum_{|\mathbf{l}|_1=n+d-1-i, |\mathbf{l}|_\infty < n-2.5i} \|u_1\|_E \\ &\leq \frac{|u|_\infty}{2 \cdot 12^{(d-1)/2}} \cdot \sum_{i=0}^{i^*} \sum_{|\mathbf{l}|_1=n+d-1-i, |\mathbf{l}|_\infty < n-2.5i} 4^{-|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \\ &\leq \frac{|u|_\infty}{2 \cdot 12^{(d-1)/2}} \cdot 4^{-n-d+1} \cdot \sum_{i=0}^{i^*} 4^i \cdot \sum_{|\mathbf{l}|_1=n+d-1-i, |\mathbf{l}|_\infty < n-2.5i} \left(\sum_{j=1}^d 2^{l_j} \right) \\ &\leq \frac{|u|_\infty}{2 \cdot 12^{(d-1)/2}} \cdot 4^{-n-d+1} \cdot \sum_{i=0}^{i^*} 4^i \cdot \sum_{j=1}^{n-1-\lfloor 2.5i \rfloor} d \cdot \binom{n+d-2-i-j}{d-2} \cdot 2^j \\ &= \frac{|u|_\infty}{2 \cdot 12^{(d-1)/2}} 4^{-n-d+1} \sum_{i=0}^{i^*} 4^i \sum_{k=1}^{n-1-\lfloor 2.5i \rfloor} d \binom{d-2+\lfloor 1.5i \rfloor+k}{d-2} 2^{n-\lfloor 2.5i \rfloor-k}\end{aligned}$$

$$\begin{aligned}
&= \frac{d \cdot |u|_\infty}{2 \cdot 12^{(d-1)/2}} 4^{-(d-1)} 2^{-n} \sum_{i=0}^{i^*} 2^{-\lfloor \frac{i}{2} \rfloor} \sum_{k=1}^{n-1-\lfloor 2.5i \rfloor} \binom{d-2 + \lfloor 1.5i \rfloor + k}{d-2} 2^{-k} \\
&\leq \frac{d \cdot |u|_\infty}{2 \cdot 12^{(d-1)/2}} \cdot 4^{-(d-1)} \cdot 2^{-n} \cdot 2 \cdot 5^{d-1} \\
&= \frac{d \cdot |u|_\infty}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{5}{2}\right)^{d-1} \cdot 2^{-n},
\end{aligned}$$

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 (3.68), we get the first result and, in a completely analogous way, the second one, too. \square

Though we have only derived upper bounds for the energy norm of the interpolation error, it is helpful to compare the respective results (3.32), (3.68) and (3.73) for the three approximation spaces $V_n^{(\infty)}$, $V_n^{(1)}$ and $V_n^{(E)}$. Table 3.2 shows that there is no asymptotic growth with respect to d , either for the full grid case or for our two sparse grid spaces.

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 (3.72) and (3.73), 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. Especially, there is no longer any polynomial term in n like n^{d-1} . That is, apart from the factors that are constant with respect to n , there is no d -dependence of both $|V_n^{(E)}|$ and $\|u - u_n^{(E)}\|_E$, and thus no deterioration in complexity for higher-dimensional problems. Furthermore, the growth of the d -dependent terms in d is not too serious, since we have a factor

Table 3.2. d -depending constants in the bounds for $\|u - u_n^{(\cdot)}\|_E$ (multiply with $|u|_{2,\infty} \cdot 2^{-n}$ (first row) or $|u|_{2,2} \cdot 2^{-n}$ (second row) to get the respective bounds).

$V_n^{(\infty)}$	$V_n^{(1)}$	$V_n^{(E)}$
$\frac{d^{3/2}}{2 \cdot 3^{(d-1)/2} \cdot 6^{d-1}}$	$\frac{d}{2 \cdot 3^{(d-1)/2} \cdot 4^{d-1}}$	$\frac{d}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2}\right)^{d-1}\right)$
$\frac{d^{3/2}}{\sqrt{3} \cdot 9^{d-1}}$	$\frac{d}{\sqrt{3} \cdot 6^{d-1}}$	$\frac{2 \cdot d}{\sqrt{3} \cdot 6^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2}\right)^{d-1}\right)$

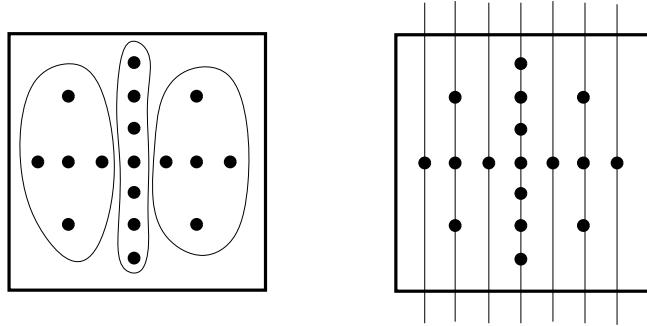


Figure 3.8. Recursive structure of $V_3^{(1)}$ for $d = 2$.

of $\frac{d}{2} \cdot e^d$ in the upper bound of $|V_n^{(E)}|$ and (in the best case; see Table 3.2) $\frac{d}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\frac{1}{2} + \left(\frac{5}{2}\right)^{d-1}\right)$ in the upper bound of $\|u - u_n^{(E)}\|_E$.

3.3. Recurrences and complexity

In this section, we make a short digression into recurrence formulas for sparse grids and into sparse grid complexity, in order to learn more about their asymptotic behaviour. We present the most interesting results only and refer the reader to Bungartz (1998) for details or proofs.

Recurrence formulas

For the following, we restrict ourselves to the sparse grid spaces $V_n^{(1)}$. In (3.61), they were introduced with the help of an explicit formula. Now we study their recursive character to obtain further results concerning their complexity and asymptotic properties. First, starting from (3.62), one can show a recurrence relation for $|V_n^{(1)}|$, the number of (inner) grid points of a sparse grid. Note that $|V_n^{(1)}|$ depends on two parameters: the *dimensionality* d and the *resolution* n . Defining

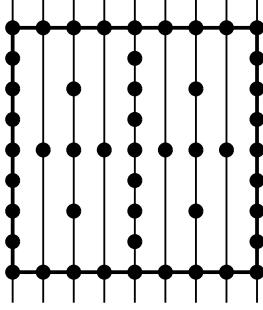
$$a_{n,d} := |V_n^{(1)}|, \quad (3.74)$$

we get

$$a_{n,d} = a_{n,d-1} + 2 \cdot a_{n-1,d}. \quad (3.75)$$

That is, the d -dimensional sparse grid of resolution (or depth) n consists of a $(d-1)$ -dimensional one of depth $n-1$ (separator) and of two d -dimensional sparse grids of depth $n-1$ (*cf.* Figure 3.8, left). If we continue with this decomposition in a recursive way, we finally obtain a full-history version of (3.75) with respect to n ,

$$a_{n,d} = \sum_{i=0}^{n-1} 2^i \cdot a_{n-i,d-1}, \quad (3.76)$$

Figure 3.9. Recursive structure of $b_{3,2}$.

since $a_{1,d} = 1$ for all $d \in \mathbb{N}$ due to (3.74). Thus, a sparse grid $V_n^{(1)}$ of dimensionality d and depth n can be completely reduced to sparse grids of dimensionality $d - 1$ and depth k , $k = 1, \dots, n$ (*cf.* Figure 3.8, right).

In addition to $a_{n,d}$, we shall now deal with sparse grids with grid points on the boundary. To this end, let $b_{n,d}$ be the number of overall grid points of the L_2 -based sparse grid of parameters d and n , *i.e.*, in the interior *and* on the boundary of $\bar{\Omega}$. On $\partial\bar{\Omega}$, we assume sparse grids of the same resolution n , but of a reduced dimensionality. Since the boundary of the d -dimensional unit interval $\bar{\Omega}$ consists of $\binom{d}{j} \cdot 2^{d-j}$ j -dimensional unit intervals ($j = 0, \dots, d$), we get

$$b_{n,d} := \sum_{j=0}^d \binom{d}{j} \cdot 2^{d-j} \cdot a_{n,j}, \quad (3.77)$$

where $a_{n,0} := 1$ for all $n \in \mathbb{N}$. With the help of (3.75) and (3.77), the following recurrence relation for the $b_{n,d}$ can be derived:

$$b_{n,d} = 2 \cdot b_{n-1,d} + 3 \cdot b_{n,d-1} - 4 \cdot b_{n-1,d-1} \quad (3.78)$$

with its full-history version with respect to n

$$\begin{aligned} b_{n,d} &= 3 \cdot b_{n,d-1} + \sum_{i=1}^{n-1} 2^i \cdot b_{n-i,d-1} \\ &= 2 \cdot b_{n,d-1} + \sum_{i=0}^{n-1} 2^i \cdot b_{n-i,d-1}, \end{aligned} \quad (3.79)$$

where the first term stands for the boundary faces $x_d \in \{0, 1\}$, whereas the sum denotes the inner part of the grid with respect to direction x_d . Figure 3.9 illustrates the recursive structure of $b_{n,d}$ in the 2D case.

Finally, a third quantity $c_{p,d}$ shall be introduced that motivates the sparse grid pattern from a perspective of approximation with polynomials, thus

anticipating, to some extent, the higher-order approaches to be discussed later. Starting from our hierarchical setting, we are looking for the minimum number of grid points that are necessary to realize a polynomial approximation of a certain degree p . In the simple 1D case, things are evident. First, with one degree of freedom available in the left point of the boundary ($x = 0$), a constant basis function is placed there, while on the right side ($x = 1$), a linear function is chosen, thus allowing linear interpolation ($p = 1$) in $\bar{\Omega}$ with two degrees of freedom. Afterwards, on each level $l := 1$ of inner grid points, we raise the degree p of the basis functions by one, and consequently get an approximation order of degree $p = l + 1$ for level l . Note that there is no overall interpolant of degree $p > 2$ on $\bar{\Omega}$, but, owing to the hierarchical subspace decomposition, there exists an interpolant, continuous on $\bar{\Omega}$, that is piecewise polynomial of degree p with respect to level $l - 1 = p - 2$.

For $d > 1$, things are a little bit more complicated. We discuss the definition of $c_{p,d}$ for $d = 2$. For constant interpolation, just one degree of freedom is necessary. Linear interpolation, *i.e.*, $p = 1$, can be obtained with three grid points, *e.g.*, with the three degrees of freedom $1, x$, and y living in three of the unit square's four corners. This proceeding is completely consistent with the tensor product approach: for $x = 0$, the 1D basis function is constant with respect to x . Thus, we need linear interpolation with respect to y (two degrees of freedom for $x = 0$). On the right-hand side of the unit square ($x = 1$), we are linear with respect to x and thus need only one degree of freedom with respect to y (see Figure 3.10). For a quadratic approximation, six degrees of freedom are necessary, and so on. This economic use of degrees of freedom leads to a certain asymmetry of the respective grid patterns and to a delayed generation of inner grid points: when the first inner grid point appears, the maximum depth on the boundary is already three in the 2D case. Nevertheless, the resulting grids for a given polynomial degree p are very closely related to the standard sparse grids described by $b_{n,d}$, since, obviously, the grid patterns resulting in the interior are just the standard patterns of $V_n^{(1)}$. Hence, here, the maximum degree p of the basis functions used takes the part of the resolution n as the second parameter besides d .

The principles of construction for the grids resulting from this polynomial approach lead us to a recurrence relation for $c_{p,d}$,

$$c_{p,d} = c_{p,d-1} + c_{p-1,d-1} + \sum_{i=0}^{p-2} 2^i \cdot c_{p-2-i,d-1}, \quad (3.80)$$

from which we get

$$c_{p,d} = 2 \cdot c_{p-1,d} + c_{p,d-1} - c_{p-1,d-1} - c_{p-2,d-1}. \quad (3.81)$$

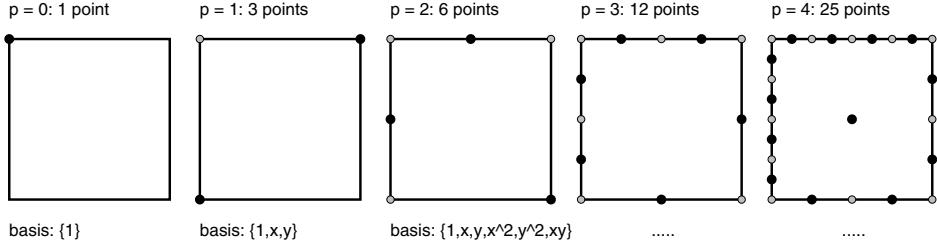


Figure 3.10. Minimum number of grid points for piecewise polynomial degree $p = 3$ in the hierarchical setting ($d = 2$) and corresponding local bases.

In the following, we present some properties of the $a_{n,d}$, $b_{n,d}$, and $c_{p,d}$. Though we do not want to go into detail here, note that, in contrast to many recurrences studied for the analysis of algorithms (*cf.* Graham, Knuth and Patashnik (1994), Sedgewick and Flajolet (1996), for example), these quantities, and thus the overall storage requirement and computational cost connected with sparse grids, depend on two parameters. Let $d \in \mathbb{N}$, $n \in \mathbb{N}$, and $p \in \mathbb{N}_0$. The initial conditions

$$\begin{aligned} a_{1,d} &= 1 \quad \forall d \in \mathbb{N}, & a_{n,1} &= 2^n - 1 \quad \forall n \in \mathbb{N}, \\ b_{1,d} &= 3^d \quad \forall d \in \mathbb{N}, & b_{n,1} &= 2^n + 1 \quad \forall n \in \mathbb{N}, \\ c_{0,d} &= 1 \quad \forall d \in \mathbb{N}, & c_{p,1} &= 2^{p-1} + 1 \quad \forall p \in \mathbb{N}, \\ c_{1,d} &= d + 1 \quad \forall d \in \mathbb{N} \end{aligned} \tag{3.82}$$

follow immediately from the semantics of $a_{n,d}$, $b_{n,d}$, and $c_{p,d}$. Owing to (3.76), (3.79), (3.80), and the initial conditions (3.82), all $a_{n,d}$, $b_{n,d}$, and $c_{p,d}$ are natural numbers. Next, we study the behaviour of our three quantities for increasing d , n , or p , respectively. The following lemma summarizes the asymptotic behaviour of $a_{n,d}$, $b_{n,d}$, and $c_{p,d}$.

Lemma 3.11. The following relations are valid for the n - and d -asymptotic behaviour of $a_{n,d}$ and $b_{n,d}$:

$$\begin{array}{ll} \frac{a_{n,d+1}}{a_{n,d}} \xrightarrow{d \rightarrow \infty} 1, & \frac{a_{n,d+1}}{a_{n,d}} \xrightarrow{n \rightarrow \infty} \infty, \\ \frac{a_{n+1,d}}{a_{n,d}} \xrightarrow{n \rightarrow \infty} 2, & \frac{a_{n+1,d}}{a_{n,d}} \xrightarrow{d \rightarrow \infty} \infty, \\ \frac{b_{n,d+1}}{b_{n,d}} \xrightarrow{d \rightarrow \infty} 3, & \frac{b_{n,d+1}}{b_{n,d}} \xrightarrow{n \rightarrow \infty} \infty, \\ \frac{b_{n+1,d}}{b_{n,d}} \xrightarrow{n \rightarrow \infty} 2, & \frac{b_{n+1,d}}{b_{n,d}} \xrightarrow{d \rightarrow \infty} \infty. \end{array} \tag{3.83}$$

For $c_{p,d}$, the respective limits are

$$\begin{aligned} \frac{c_{p,d+1}}{c_{p,d}} &\xrightarrow{d \rightarrow \infty} 1, & \frac{c_{p,d+1}}{c_{p,d}} &\xrightarrow{p \rightarrow \infty} \infty, \\ \frac{c_{p+1,d}}{c_{p,d}} &\xrightarrow{p \rightarrow \infty} 2, & \frac{c_{p+1,d}}{c_{p,d}} &\xrightarrow{d \rightarrow \infty} \infty. \end{aligned} \quad (3.84)$$

For $a_{n,d}$, $b_{n,d}$, and $c_{p,d}$, the results for the limits $n \rightarrow \infty$ are not too surprising. In the case of $a_{n,d} = |V_n^{(d)}|$, for instance, they can be derived from the order terms $2^n \cdot n^{d-1}$ with respect to n according to (3.63). However, the statements concerning the limits $d \rightarrow \infty$ provide some new and interesting information. First, the ratio $a_{n+1,d}/a_{n,d}$ of two sparse grids of resolution $n+1$ and n , for example, is not bounded for increasing d . That is, increasing the resolution does not entail only a certain bounded factor of increasing cost, but the relative increase in cost switching from resolution n to $n+1$ becomes bigger and bigger for higher dimensionality and is not bounded. Second, for a fixed resolution n , the relative difference between two sparse grids of dimensionality d and $d+1$ becomes more and more negligible for increasing d , which is, in fact, somewhat surprising. Note that this is a hard statement and not just one dealing with orders of magnitude.

However, after all those asymptotic considerations, it is important to note that, often, such limits are not excessively useful for numerical purposes, since practical computations do not always reach the region of asymptotic behaviour. Therefore, usually, the constant factors play a more predominant part than an asymptotic point of view suggests. Furthermore, it certainly makes sense to increase resolution during the numerical solution of a given problem, but a variable dimensionality is, of course, more of a theoretical interest.

ε -complexity

An approach that is closely related to the cost–benefit setting used for the derivation of the sparse grid approximation spaces is the concept of the ε -complexity (Traub and Woźniakowski 1980, Traub, Wasilkowski and Woźniakowski 1983, Traub, Wasilkowski and Woźniakowski 1988, Woźniakowski 1985). The ε -complexity of a numerical method or algorithm indicates the computational work that is necessary to produce an approximate solution of some prescribed accuracy ε . In particular, for the complexity of general multivariate tensor product problems, see Wasilkowski and Woźniakowski (1995). We consider the ε -complexity of the different discrete approximation spaces $V_n^{(\infty)}$, $V_n^{(1)}$, and $V_n^{(E)}$ for the problem of *representing* a function $u \in X_0^{q,2}(\bar{\Omega})$ on a grid, *i.e.*, the problem of constructing the interpolant $u_n^{(\infty)} \in V_n^{(\infty)}$, $u_n^{(1)} \in V_n^{(1)}$, or $u_n^{(E)} \in V_n^{(E)}$, respectively. To this end, the overall computational cost caused by the interpolation in one of the

three above discrete spaces will be estimated by the number of degrees of freedom (*i.e.*, grid points), or by an upper bound for this number. This does not constitute a restriction, since there are algorithms that can calculate the interpolant $u_n^{(\cdot)}$ in $O(|V_n^{(\cdot)}|)$ arithmetic operations, of course. Furthermore, as a measure for the accuracy, we use the upper bounds for the interpolation error $\|u - u_n^{(\infty)}\|$, $\|u - u_n^{(1)}\|$, and $\|u - u_n^{(E)}\|$ with respect to the different norms, as provided by (3.32), (3.68), and (3.73).

First, we deal with the well-known case of the regular full grid space $V_n^{(\infty)}$, where the curse of dimensionality is predominant and causes the problem to be *intractable* in the sense of Traub *et al.* (1988): the computational cost of obtaining an approximate solution of some given accuracy ε grows exponentially in the problem's dimensionality d . Note that, in the following, all occurring order terms have to be read with respect to ε or N , respectively, *i.e.*, for arbitrary, but fixed d .

Lemma 3.12. For the ε -complexities $N_\infty(\varepsilon)$, $N_2(\varepsilon)$, and $N_E(\varepsilon)$ of the problem of computing the interpolant $u_n^{(\infty)} \in V_n^{(\infty)}$ with respect to the L_∞ -, the L_2 -, and the energy norm for some prescribed accuracy ε , the following relations hold:

$$\begin{aligned} N_\infty(\varepsilon) &= O(\varepsilon^{-\frac{d}{2}}), & N_2(\varepsilon) &= O(\varepsilon^{-\frac{d}{2}}), \\ N_E(\varepsilon) &= O(\varepsilon^{-d}). \end{aligned} \quad (3.85)$$

Conversely, given a number N of grid points, the following accuracies can be obtained with respect to the different norms:

$$\begin{aligned} \varepsilon_{L_\infty}(N) &= O(N^{-\frac{2}{d}}), & \varepsilon_{L_2}(N) &= O(N^{-\frac{2}{d}}), \\ \varepsilon_E(N) &= O(N^{-\frac{1}{d}}). \end{aligned} \quad (3.86)$$

Proof. The statements follow directly from (3.31) and (3.32). \square

Next we turn to the L_2 -based sparse grid space $V_n^{(1)}$. As we have already seen, $V_n^{(1)}$ lessens the curse of dimensionality, but does not yet overcome it.

Lemma 3.13. For the ε -complexities $N_\infty(\varepsilon)$, $N_2(\varepsilon)$, and $N_E(\varepsilon)$ of the problem of computing the interpolant $u_n^{(1)} \in V_n^{(1)}$ with respect to the L_∞ -, the L_2 -, and the energy norm for some prescribed accuracy ε , the following relations hold:

$$\begin{aligned} N_\infty(\varepsilon), N_2(\varepsilon) &= O(\varepsilon^{-\frac{1}{2}} \cdot |\log_2 \varepsilon|^{\frac{3}{2} \cdot (d-1)}), \\ N_E(\varepsilon) &= O(\varepsilon^{-1} \cdot |\log_2 \varepsilon|^{d-1}). \end{aligned} \quad (3.87)$$

Conversely, given a number N of grid points, the following accuracies can be obtained with respect to the different norms:

$$\begin{aligned}\varepsilon_\infty(N), \varepsilon_2(N) &= O\left(N^{-2} \cdot |\log_2 N|^{3 \cdot (d-1)}\right), \\ \varepsilon_E(N) &= O\left(N^{-1} \cdot |\log_2 N|^{d-1}\right).\end{aligned}\quad (3.88)$$

Proof. See Bungartz (1998), for example. \square

Finally, for the energy-based sparse grid space $V_n^{(E)}$, the situation is evident and gratifying: the curse of dimensionality has disappeared.

Lemma 3.14. For the ε -complexity $N_E(\varepsilon)$ of the problem of computing the interpolant $u_n^{(E)} \in V_n^{(E)}$ with respect to the energy norm for some prescribed accuracy ε , the following relation holds:

$$N_E(\varepsilon) = O\left(\varepsilon^{-1}\right). \quad (3.89)$$

Thus, for a fixed number of N of grid points, the following accuracy can be obtained:

$$\varepsilon_E(N) = O\left(N^{-1}\right). \quad (3.90)$$

Proof. Both results are obvious consequences of (3.72) and (3.73). \square

With the above remarks on the ε -complexity of the problem of interpolating functions $u \in X_0^{q,2}(\bar{\Omega})$ in our two sparse grid spaces $V_n^{(1)}$ and $V_n^{(E)}$, we close the discussion of the piecewise d -linear case.

4. Generalizations, related concepts, applications

In the previous section we presented the main ideas of the sparse grid approach starting from the 1D piecewise linear hierarchical basis, which was then extended to the general piecewise d -linear hierarchical basis by the discussed tensor product construction. However, the discretization on sparse grids is, of course, not limited to this explanatory example, but can be directly generalized to other multiscale bases such as p -type hierarchical bases, prewavelets, or wavelets, for instance. To this end, another 1D multiscale basis must be chosen. Then the tensor product construction as well as the cut-off of the resulting series expansion will lead to closely related sparse grids.

In the following, we first give a short survey of the historic background of sparse grids. Then we generalize the piecewise linear hierarchical basis to hierarchical polynomial bases of higher order, and discuss hierarchical Lagrangian interpolation of Bungartz (1998) and the so-called interpolets of Deslauriers and Dubuc (1989). After that, we discuss the use of prewavelets and wavelets in a sparse grid context. A short overview of the current state

concerning sparse grid applications, with a focus on the discretization of PDEs including some remarks on adaptive refinement and fast solvers, will close this section.

4.1. Ancestors

The discussion of hierarchical finite elements (Peano 1976, Zienkiewicz, Kelly, Gago and Babuška 1982) and, in particular, the series of articles by Yserentant (1986, 1990, 1992) introducing the use of hierarchical bases for the numerical solution of PDEs, both for purposes of an explicit discretization and for the construction of preconditioners, was the starting point of Zenger's sparse grid concept (Zenger 1991). The generalization of Yserentant's hierarchical bases to a strict tensor product approach with its underlying hierarchical subspace splitting discussed in Section 3 allowed the *a priori* identification of more and of less important subspaces and grid points. As we have seen, it is this characterization of subspaces that the definition of sparse grids is based on. With the sparse grid approach, for the first time, *a priori* optimized and fully structured grid patterns were integrated into existing and well-established discretization schemes for PDEs such as finite elements, and were combined with a very straightforward access to adaptive grid refinement.

Even if the sparse grid concept was new for the context of PDEs, very closely related techniques had been studied for purposes of approximation, recovery, or numerical integration of smooth functions, before. For instance, the generalization of Archimedes' well-known hierarchical quadrature of $1 - x^2$ on $[-1, 1]$ to the d -dimensional case via Cavalieri's principle (see Figure 4.1) is a very prominent example of an (indeed early) hierarchical tensor product approach. Much later, Faber (1909) discussed the hierarchical representation of functions.

Finally, once more, the Russian literature turns out to be helpful for exploring the roots of a new approach in numerical mathematics. Two

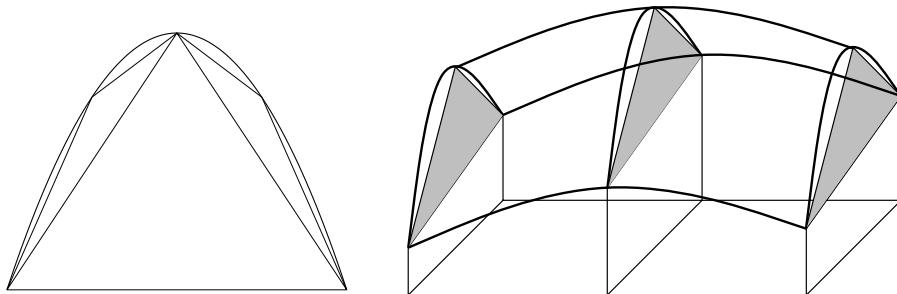


Figure 4.1. Hierarchical quadrature according to Archimedes (left) and application of Cavalieri's principle (right).

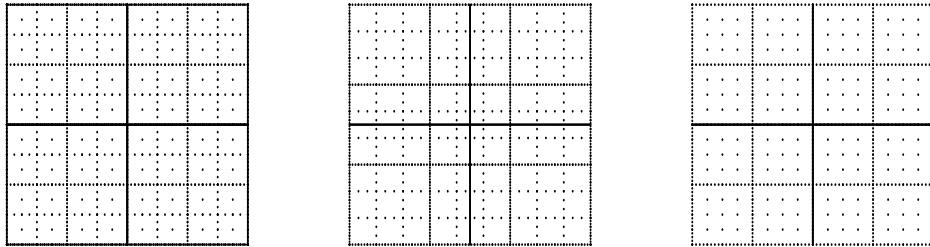


Figure 4.2. Smolyak quadrature patterns based on the trapezoidal rule as the one-dimensional algorithm: $p = 2$ and $n = 8$ (left), $p = 3$ and $n = 5$ (centre), $p = 4$ and $n = 4$ (right).

names that have to be mentioned here are those of Smolyak and Babenko. Smolyak (1963) studied classes of quadrature formulas of the type

$$Q_n^{(d)} f := \left(\sum_{i=0}^n (Q_i^{(1)} - Q_{i-1}^{(1)}) \otimes Q_{n-i}^{(d-1)} \right) f \quad (4.1)$$

that are based on a tensor product \otimes of lower-dimensional operators. In (4.1), $Q_n^{(d)}$ denotes a d -dimensional quadrature formula based on the 1D rule $Q_n^{(1)}$ that is, for $n \in \mathbb{N}$, usually chosen to be the compound formula resulting from the application of some simple formula Q on p^n subintervals $[i/p^n, (i+1)/p^n]$, $i = 0, \dots, p^n - 1$, of $[0, 1]$ for some natural number $p \geq 2$. Furthermore, the midpoint rule is usually taken as $Q_0^{(1)}$, and $Q_{-1}^{(1)} \equiv 0$ (see Frank and Heinrich (1996), Novak and Ritter (1996), and Smolyak (1963)). Figure 4.2 shows several examples of grids resulting from the application of 2D Smolyak quadrature for different n and p . Functions suitable for the Smolyak approach typically live in spaces of bounded (L_p -integrable) mixed derivatives which are closely related to our choice of $X_0^{q,r}(\bar{\Omega})$ in (3.2). A similar approach to the approximation of periodic multivariate functions of bounded mixed derivatives may be found in Babenko's *hyperbolic crosses* (Babenko 1960). Here, Fourier monomials or coefficients, respectively, are taken from sets of the type

$$\Gamma(n) := \left\{ \mathbf{k} \in \mathbb{Z}^d : \prod_{j=1}^d \max\{|k_j|, 1\} \leq n \right\}. \quad (4.2)$$

For a detailed discussion of those methods, we refer to the work of Temlyakov (1989, 1993, 1994). There are several other approaches that are more or less based on Smolyak's tensor product technique, for example the so-called *Boolean methods* of Delvos *et al.* (Delvos 1982, Delvos and Schempp 1989, Delvos 1990, Baszenski and Delvos 1993) or the *discrete blending methods*

of Baszenski, Delvos and Jester (1992) going back to work of Gordon (Gordon 1969, 1971, Gordon and Hall 1973). For the general analysis of such tensor product methods, see also Wasilkowski and Woźniakowski (1995). Concerning the computational cost, Smolyak's tensor product approach and its derivatives are characterized by terms of the order $O(N \cdot (\log_2 N)^{d-1})$, where N denotes the 1D cost, as we have shown them for the sparse grid spaces $V_n^{(1)}$ in (3.63). Furthermore, the tensor product approach of (4.1) itself, obviously, calls to mind the definition (3.61) of the L_2 -based sparse grid spaces $V_n^{(1)}$, which can be written as

$$V_n^{(d,1)} := V_n^{(1)} = \sum_{|\mathbf{l}|_1 \leq n+d-1} W_{\mathbf{l}}^{(d)} = \sum_{l=1}^n W_l^{(1)} \otimes V_{n+d-1-l}^{(d-1,1)}, \quad (4.3)$$

that is, in a tensor product form, too.

Finally, there are of course close relations of hierarchical bases and sparse grids on the one hand and wavelets on the other hand. These will be discussed in Section 4.4.

4.2. Higher-order polynomials

In this section, we discuss how to generalize the piecewise linear sparse grid method to higher-order basis functions.

The hierarchical Lagrangian interpolation

The first approach, the so-called hierarchical Lagrangian interpolation introduced in Bungartz (1998), uses a hierarchical basis of piecewise polynomials of arbitrary degree p , still working with just one degree of freedom per node. Before we discuss the underlying p -hierarchy, note that the piecewise constant case, *i.e.*, $p = 0$, as illustrated in Figure 4.3, is the natural starting point of such a hierarchy.

In accordance with the tensor product approach (3.9), we define basis functions of the (generalized) degree $\mathbf{p} := (p_1, \dots, p_d) \in \mathbb{N}^d$ as products

$$\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) := \prod_{j=1}^d \phi_{l_j,i_j}^{(p_j)}(x_j) \quad (4.4)$$

of d 1D basis polynomials of degree p_j with the respective supports $[x_{l_j,i_j} - h_{l_j}, x_{l_j,i_j} + h_{l_j}]$. Note that, since there is no change in the underlying grids $\Omega_{\mathbf{l}}$, the grid points $\mathbf{x}_{\mathbf{l},\mathbf{i}}$ or x_{l_j,i_j} and the mesh widths $\mathbf{h}_{\mathbf{l}}$ or h_{l_j} are defined exactly as in the linear case. This, again, allows the restriction to the 1D case. For reasons of clarity, we will omit the index j . As already mentioned, we want to preserve the 'character' of the elements, *i.e.*, we want to keep to \mathcal{C}^0 -elements and manage without increasing the number of degrees of freedom per element

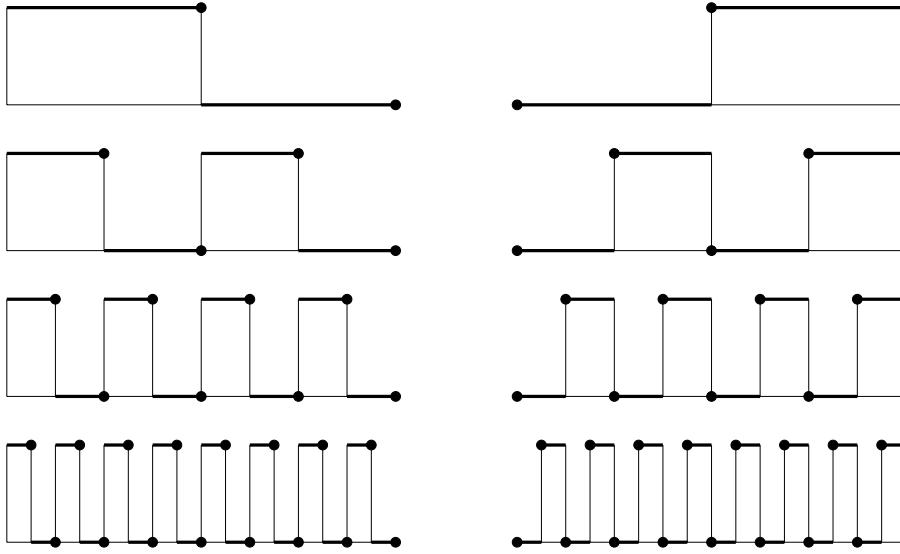


Figure 4.3. Piecewise constant hierarchical bases: continuous from the left (left) or continuous from the right (right).

or per grid point for a higher p . However, to determine a polynomial $u^{(p)}(x)$ of degree p uniquely on $[x_{l,i} - h_l, x_{l,i} + h_l]$, we need $p+1$ conditions $u^{(p)}(x)$ has to fulfil. In the linear case, the interpolant resulting from the hierarchically higher levels is defined by its values in the two boundary points $x_{l,i} \pm h_l$. For $p \geq 2$, these two conditions are no longer sufficient. Therefore, we profit from the hierarchical history of $x_{l,i}$. Figure 4.4 shows the hierarchical relations of the grid points according to the hierarchical subspace splitting of Section 3.1. Apart from $x_{l,i} \pm h_l$, which mark the boundary of the support of $\phi_{l,i}^{(p)}$, $x_{l,i}$ may have hierarchical ancestors that are all located outside this support. Consequently, for the definition of such a local interpolant $u^{(p)}(x)$, it is reasonable and, for the construction of a hierarchical basis, essential to take the values of u in $x_{l,i} \pm h_l$ (as in the linear case) and, in addition, in a sufficient number of hierarchically next ancestors of $x_{l,i}$. These considerations lead us to the following definition.

Let $u \in C^{p+1}([0, 1])$, $1 \leq p \leq l$, and let Ω_l denote the 1D grid of mesh width $h_l = 2^{-l}$ with grid points $x_{l,i}$ according to (3.4)–(3.6). Then, the *hierarchical Lagrangian interpolant* $u^{(p)}(x)$ of degree p of $u(x)$ with respect to Ω_l is defined on $[x_{l,i} - h_l, x_{l,i} + h_l]$, i odd, as the polynomial interpolant of $(x_k, u(x_k))$, $k = 1, \dots, p+1$, where the x_k are just $x_{l,i} \pm h_l$ and the $p-1$ next hierarchical ancestors of $x_{l,i}$. Note that $u^{(p)}(x)$ is continuous on $\bar{\Omega}$, piecewise of polynomial degree p with respect to the grid Ω_{l-1} , and

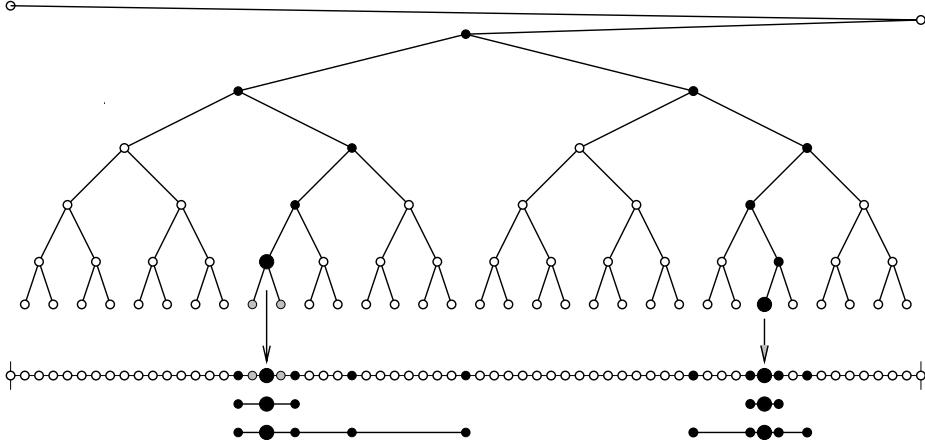


Figure 4.4. Ancestors (here: boundary points of the respective basis function's support and two more ($p = 4$); solid) and descendants (dotted) of two grid points.

it interpolates $u(x)$ on Ω_{l-1} . Nevertheless, it is defined *locally* on $[x_{l,i} - h_l, x_{l,i} + h_l]$. Thus, the width of the interval taken for the local *definition* of $u^{(p)}(x)$ is $2^p \cdot h_l$, but for any kind of further calculations, $u^{(p)}(x)$ is living only on the local interval of size $2 \cdot h_l$. The restriction $l \geq p$ is due to the fact that, with degree p , we need at least $p + 1$ ancestors for the interpolation. Now we study the local approximation properties of the hierarchical Lagrangian interpolation, that is, the interpolation error $u(x_{l,i}) - u^{(p)}(x_{l,i})$ or hierarchical surplus (*cf.* (3.19)).

Lemma 4.1. Let $u \in \mathcal{C}^{p+1}([0, 1])$, $1 \leq p \leq l$, and let $x_1 < \dots < x_{p+1}$ be the ancestors of $x_{l,i}$ on level l , i odd, taken for the construction of the hierarchical Lagrangian interpolant $u^{(p)}$ of u in $[x_{l,i} - h_l, x_{l,i} + h_l]$. Then the hierarchical surplus $v^{(p)}(x_{l,i})$ in $x_{l,i}$ fulfills

$$v^{(p)}(x_{l,i}) := u(x_{l,i}) - u^{(p)}(x_{l,i}) = \frac{1}{(p+1)!} \cdot D^{p+1}u(\xi) \cdot \prod_{k=1}^{p+1} (x_{l,i} - x_k) \quad (4.5)$$

for some $\xi \in [x_1, x_{p+1}]$. Moreover, the order of approximation is given by

$$|v^{(p)}(x_{l,i})| \leq \frac{1}{(p+1)!} \cdot |D^{p+1}u(\xi)| \cdot h_l^{p+1} \cdot 2^{p \cdot (p+1)/2 - 1}. \quad (4.6)$$

Proof. (4.5) is a standard remainder formula for the interpolation with polynomials. For (4.6), a careful look at the distances $x_{l,i} - x_k$ provides

$$\begin{aligned} |v^{(p)}(x_{l,i})| &= \frac{1}{(p+1)!} \cdot |D^{p+1}u(\xi)| \cdot \prod_{k=1}^{p+1} |x_{l,i} - x_k| \\ &\leq \frac{1}{(p+1)!} \cdot |D^{p+1}u(\xi)| \cdot h_l^{p+1} \cdot \prod_{k=1}^p (2^k - 1) \\ &\leq \frac{1}{(p+1)!} \cdot |D^{p+1}u(\xi)| \cdot h_l^{p+1} \cdot 2^{p(p+1)/2-1}, \end{aligned}$$

which is (4.6). \square

Hence, we obtain the desired increase in the order of approximation, for the price of a factor growing exponentially in p^2 . This is a hint that increasing p on each new level will not be the best strategy.

As an analogue to (3.24), an integral representation can be shown for the general order hierarchical surplus, too. For that, define $s_{l,i}^{(p)}(t)$ as the minimum support spline with respect to $x_{l,i}$ and its $p+1$ direct hierarchical ancestors (renamed x_0, \dots, x_{p+1} in increasing order):

$$s_{l,i}^{(p)}(t) := [x_0, \dots, x_{p+1}] (x-t)_+^p = \sum_{k=0}^{p+1} \frac{(x_k - t)_+^p}{w'_{l,i}(x_k)}. \quad (4.7)$$

Here, $[x_0, \dots, x_{p+1}] f(x)$ just denotes the divided differences of order p with respect to x ,

$$(x-t)_+^p := \begin{cases} (x-t)^p, & \text{for } x-t \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (4.8)$$

and

$$w_{l,i}(x) := \prod_{j=0}^{p+1} (x - x_j). \quad (4.9)$$

Lemma 4.2. With the above definitions, we get the following integral representation for the hierarchical surplus $v^{(p)}(x_{l,i})$:

$$v^{(p)}(x_{l,i}) = \frac{w'_{l,i}(x_{l,i})}{p!} \cdot \int_{-\infty}^{\infty} s_{l,i}^{(p)}(t) \cdot D^{p+1}u(t) dt. \quad (4.10)$$

Proof. See Bungartz (1998). \square

An immediate consequence of (4.10) is the relation

$$\int_{x_0}^{x_{p+1}} D^{p-1} s_{l,i}^{(p)}(t) \cdot f(t) dt = 0 \quad (4.11)$$

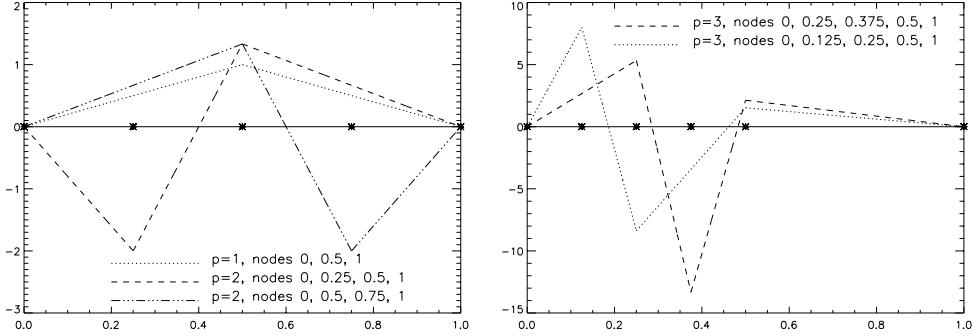


Figure 4.5. $D^{p-1} s_{l,i}^{(p)}(t)$ for $p = 1$ and the two constellations of $p = 2$ (left) and for two of the four cubic constellations (right).

for all $f \in \mathbb{P}_{p-2}$. Owing to (4.7),

$$D^{p-1} s_{l,i}^{(p)}(t) = (-1)^{p-1} \cdot \sum_{k=0}^{p+1} \frac{(x_k - t)_+}{w'_{l,i}(x_k)} \quad (4.12)$$

is piecewise linear and continuous. In wavelet terminology, (4.11) means that the $p - 1$ first moments of the $D^{p-1} s_{l,i}^{(p)}(t)$ vanish. Therefore we could construct $s_{l,i}^{(p)}(t)$ and, hence, the hierarchical surplus $v^{(p)}(x_{l,i})$ of degree p the other way round, too. Starting from x_0, \dots, x_{p+1} , look for the piecewise linear and continuous function $\sigma^{(p)}(t)$ with $\sigma^{(p)}(x_0) = \sigma^{(p)}(x_{p+1}) = 0$ and with vanishing first $p - 1$ moments, which is determined up to a constant factor, and integrate the resulting $\sigma^{(p)}(t)$ $p - 1$ times. The remaining degree of freedom can be fixed by forcing the coefficient of $u(x_m)$ to be 1 (as it is in $v^{(p)}(x_m)$). The left-hand part of Figure 4.5 shows this $(p - 1)$ st derivative of $s_{l,i}^{(p)}(t)$ for the possible constellations of ancestors for $p = 1, 2$. In the linear case, we have no vanishing moment; for $p = 2$, we get one. The cubic case is illustrated in the right-hand part of Figure 4.5.

Note that Figure 4.5 suggests that the higher-order case may be obtained without explicitly working with polynomials of arbitrary degree, but just by a superposition of the linear approximations in different ancestors. In fact, the spline's $(p - 1)$ st derivative, and thus the calculation of the hierarchical surplus $v^{(p)}(x_{l,i})$ of degree p , can be reduced to the linear case.

Lemma 4.3. Let p , l , and i be defined as before. Furthermore, let x_m denote the current grid point $x_{l,i}$, and let x_n be its hierarchical father. For any $u \in C^{p+1}([0, 1])$, the hierarchical surplus $v^{(p)}(x_m)$ of degree p in x_m can be calculated with the help of $v^{(p-1)}(x_m)$ and of $v^{(p-1)}(x_n)$:

$$v^{(p)}(x_m) = v^{(p-1)}(x_m) - \alpha(x_0, \dots, x_{p+1}) \cdot v^{(p-1)}(x_n), \quad (4.13)$$

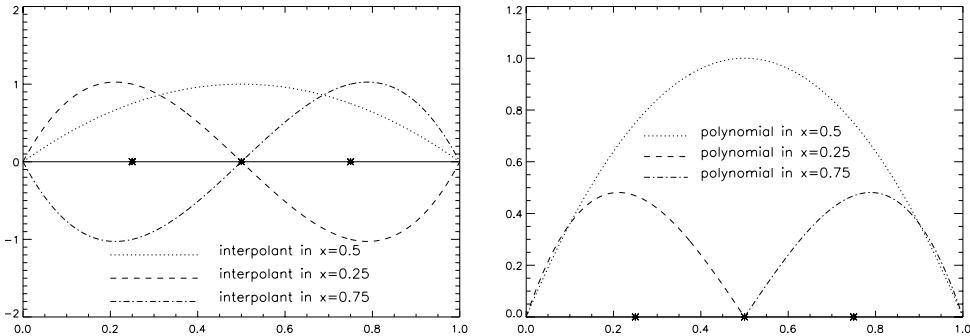


Figure 4.6. Hierarchical basis polynomials for $p = 2$ and $p = 3$ (different scaling for clarity): construction via hierarchical Lagrangian interpolation (left) and restriction to the respective hierarchical support (right).

where α depends on the relative position of x_m 's ancestors, but not on the interpolated values.

Proof. See Bungartz (1998). \square

Having introduced the hierarchical Lagrangian interpolation we now discuss the corresponding hierarchical basis polynomials of degree p . Such a $\phi_{l,i}^{(p)}$ in $x_{l,i}$ with i odd, $l \geq p-1$, and $p \geq 2$ is uniquely defined on $[x_{l,i} - h_l, x_{l,i} + h_l]$ by the following $p+1$ conditions:

$$\phi_{l,i}^{(p)}(x_{l,i}) := 1, \quad \phi_{l,i}^{(p)}(x_k) := 0, \quad (4.14)$$

where the x_k are just $x_{l,i} \pm h_l$ and the $p-2$ next hierarchical ancestors of $x_{l,i}$. Additionally, we force $\phi_{l,i}^{(p)}$ to vanish outside $[x_{l,i} - h_l, x_{l,i} + h_l]$:

$$\phi_{l,i}^{(p)}(x) := 0 \quad \text{for } x \notin [x_{l,i} - h_l, x_{l,i} + h_l]. \quad (4.15)$$

Note that the restriction $p \geq 2$ is due to the fact that $p = 1$ does not fit into this scheme. Since the typical basis function used for linear approximation is a *piecewise* linear one only, there are three degrees of freedom to determine it uniquely on $[x_{l,i} - h_l, x_{l,i} + h_l]$, as in the quadratic case.

At this point, three things are important to realize. First, this definition is fully consistent with the hierarchical Lagrangian interpolation: A global interpolant $u^{(p)}(x)$ built up with the help of these $\phi_{l,i}^{(p)}$ fulfills all requirements. Second, though the definition of the $\phi_{l,i}^{(p)}$ is based on points *outside* $[x_{l,i} - h_l, x_{l,i} + h_l]$ (for $p > 2$, at least, to be precise), we use only $[x_{l,i} - h_l, x_{l,i} + h_l]$ as support of $\phi_{l,i}^{(p)}$ according to (4.15). Thus, the support of the basis polynomials does not change in comparison with the piecewise linear case.

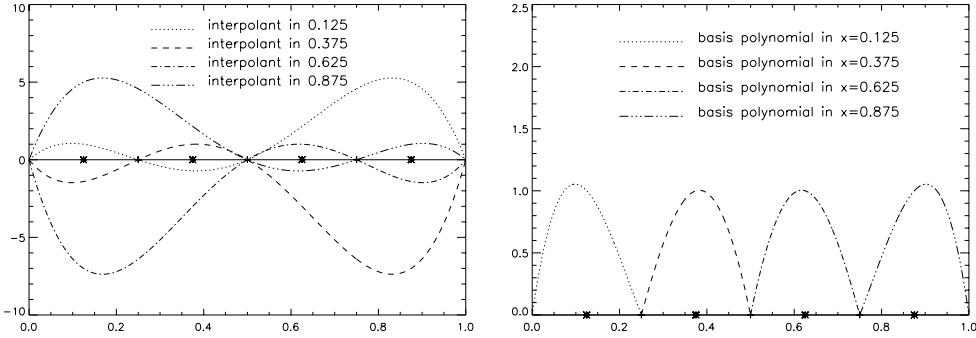


Figure 4.7. Hierarchical basis polynomials for $p = 4$: construction via hierarchical Lagrangian interpolation (left) and restriction to the respective hierarchical support (right).

Finally, since we need $p - 2$ ancestors outside $[x_{l,i} - h_l, x_{l,i} + h_l]$, a basis polynomial of degree p can not be used earlier than on level $p - 1$.

The quadratic basis polynomial is uniquely defined by its values in $x_{l,i}$ and $x_{l,i} \pm h_l$. For $p = 3$, however, the shape of the resulting polynomial depends on where the third zero outside $\phi_{l,i}^{(p)}$'s support is located. Owing to the hierarchical relations illustrated in Figure 4.4, there are two possibilities for this ancestor's position: to the left of $x_{l,i}$, *i.e.*, in $x = x_{l,i} - 3 \cdot h_l$, or to the right in $x = x_{l,i} + 3 \cdot h_l$. Thus we get two different types of cubic basis polynomials. Figure 4.6 illustrates this for the highest two levels $l \in \{1, 2\}$. In $x = 0.5$ (*i.e.*, $l = 1$), only $p = 2$ is possible owing to the lack of ancestors outside $\bar{\Omega}$. On the next level, in $x = 0.25$, the third ancestor $x = 1.0$ is situated to the right of 0.5. In $x = 0.75$, the third ancestor is $x = 0.0$, to the left of $x = 0.75$. Of course, both cubic functions are symmetric with respect to $x = 0.5$. If we continue with cubic basis polynomials on the lower levels $l > 2$, no new relationships will occur. Thus we can manage with these two types of cubic polynomials. Figure 4.7 shows the four types of quartic basis polynomials. Owing to the underlying hierarchy, these four (pairwise symmetric) polynomials cover all possible constellations. Obviously, in the general case of an arbitrary polynomial degree $p > 1$, our approach leads to 2^{p-2} different types of basis functions.

Moreover, Figure 4.7 illustrates that the four quartic basis polynomials do not differ that much. This effect even holds for all basis polynomials of arbitrary p . Figure 4.8 shows this similarity for all different types of basis polynomials up to degree 7 (all now plotted with respect to the common support $[-1, 1]$). This similarity is also reflected in the analytical properties of the basis polynomials. For details, we refer the reader to Bungartz (1998).

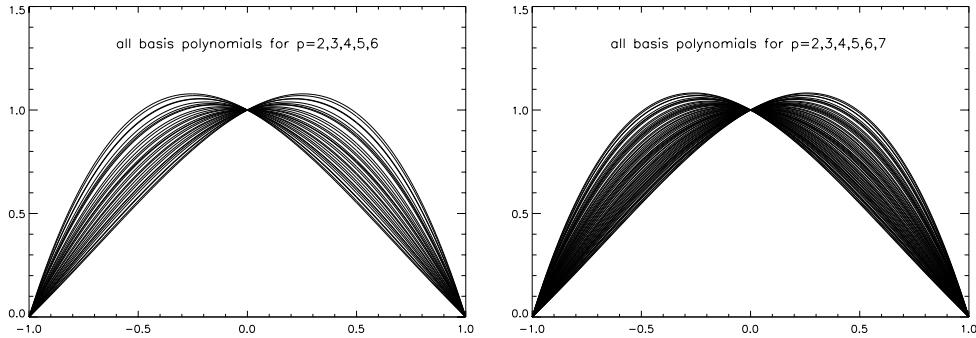


Figure 4.8. All hierarchical basis polynomials for $p \leq 6$ (31 different types; left) and $p \leq 7$ (63 different types; right) with respect to the support $[-1, 1]$.

Higher-order approximation on sparse grids

Finally, we deal with the generalization of the piecewise linear sparse grid approximation spaces $V_n^{(1)}$ and $V_n^{(E)}$ to the case of piecewise polynomials of degree $p \geq 2$ or $\mathbf{p} \geq \mathbf{2}$ according to the hierarchical Lagrangian interpolation. For the discussion of the approximation properties of the resulting sparse grid spaces of higher order, we restrict ourselves to the *p-regular* case. For some fixed maximum degree $p := p_{\max}$, the polynomials are chosen as described in the previous section until we reach level $p - 1$ or $(p - 1) \cdot \mathbf{1}$. On the further levels, we use this p . Note that the degree \mathbf{p} of a basis function or of a subspace, respectively, has to fulfil the condition

$$\mathbf{2} \leq \mathbf{p} = \min\{p \cdot \mathbf{1}, \mathbf{1} + \mathbf{1}\}, \quad (4.16)$$

where the minimum is taken component-wise. The lower bound reflects the fact that there are no linear hierarchical Lagrangian basis functions; the upper bound is caused by our choice of a maximum degree p and by the lack of basis polynomials of degree \mathbf{p} earlier than on level $\mathbf{p} - \mathbf{1}$.

The starting point is the space $X_0^{p+1,q}$ of functions of (in some sense) bounded weak mixed derivatives of the order less than or equal to $p + 1$ in each direction: *cf.* (3.2). For the approximation of some $u \in X_0^{p+1,q}$, we use product-type basis functions $\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x})$ according to (4.4). Without introducing nodal subspaces $V_1^{(\mathbf{p})}$, we directly turn to the hierarchical subspaces $W_1^{(\mathbf{p})}$:

$$W_1^{(\mathbf{p})} := \text{span}\{\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})} : \mathbf{i} \in \mathbf{I}_1\}. \quad (4.17)$$

The completion of the sum of all $W_1^{(\mathbf{p})}$ with respect to the energy norm

contains $X_0^{p+1,q}$. Thus, analogously to (3.19), the hierarchical subspace decomposition leads to a hierarchical representation of degree p of $u \in X_0^{p+1,q}$:

$$u(\mathbf{x}) = \sum_{\mathbf{l}} u_{\mathbf{l}}^{(\mathbf{p})}(\mathbf{x}), \quad u_{\mathbf{l}}^{(\mathbf{p})}(\mathbf{x}) = \sum_{\mathbf{i} \in \mathbf{I}_{\mathbf{l}}} v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})} \cdot \phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) \in u_{\mathbf{l}}^{(\mathbf{p})} \in W_{\mathbf{l}}^{(\mathbf{p})}, \quad (4.18)$$

where $v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})} \in \mathbb{R}$ is just the hierarchical coefficient or surplus. Note that, in (4.18), the degree \mathbf{p} is not constant because of (4.16).

Concerning the cost and benefit of the $W_{\mathbf{l}}^{(\mathbf{p})}$, remember that the number of degrees of freedom induced by $W_{\mathbf{l}}^{(\mathbf{p})}$ does not increase in comparison to the linear situation. Thus we still have

$$|W_{\mathbf{l}}^{(\mathbf{p})}| = |W_{\mathbf{l}}| = 2^{|\mathbf{l}-\mathbf{1}|_{\mathbf{l}}}. \quad (4.19)$$

On the other hand, the discussion of $W_{\mathbf{l}}^{(\mathbf{p})}$'s benefit, *i.e.*, of its contribution to the overall interpolant $u_{\mathbf{l}}^{(\mathbf{p})} \in W_{\mathbf{l}}^{(\mathbf{p})}$ of some $u \in X_0^{p+1,q}$, requires studying the different norms of our hierarchical Lagrangian basis polynomials $\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x})$.

Lemma 4.4. For any d -dimensional hierarchical Lagrangian basis polynomial $\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x})$ according to the above discussion, the following relations hold:

$$\begin{aligned} \|\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}\|_{\infty} &\leq 1.117^d, \\ \|\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}\|_q &\leq 1.117^d \cdot 2^{d/q} \cdot 2^{-|\mathbf{l}|_1/q}, \quad q \geq 1, \\ \|\phi_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}\|_E &\leq 3.257 \cdot \left(\frac{5}{2}\right)^{d/2} \cdot 2^{-|\mathbf{l}|_1/2} \cdot \left(\sum_{j=1}^d 2^{2l_j}\right)^{1/2}. \end{aligned} \quad (4.20)$$

Proof. The statements follow immediately from analytical properties of the basis polynomials: see Bungartz (1998) for details. \square

Next we consider $v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}$, the d -dimensional hierarchical surplus of degree \mathbf{p} . Inspired by the integral representation (4.10) of the 1D surplus, we define

$$\sigma_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) := \prod_{j=1}^d \frac{w'_{l_j,i_j}(x_{l_j,i_j})}{p_j!} \cdot s_{l_j,i_j}^{(p_j)}(x_j), \quad (4.21)$$

where $w_{l_j,i_j}(x_j)$ and $s_{l_j,i_j}^{(p_j)}(x_j)$ are defined exactly as in (4.9) and (4.7), but now for direction x_j and based on the respective hierarchical ancestors. With the help of (4.21) we obtain an integral representation analogous to (3.24).

Lemma 4.5. For $u \in X_0^{p+1,q}$, the hierarchical surplus $v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}$ fulfills

$$v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})} = \int_{\Omega} \sigma_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) \cdot D^{\mathbf{p}+\mathbf{1}} u(\mathbf{x}) \, d\mathbf{x}. \quad (4.22)$$

Proof. Analogously to (3.23) for $p = 1$, the d -dimensional interpolation operator or stencil can be written as an operator product of the d univariate operators. Thus we can proceed as in the proof of (3.24) and do the integration with respect to the d coordinate directions one after the other. According to (4.10) and (4.21), the 1D integral with respect to x_j leads to the 1D surplus with respect to x_j . Consequently, the d -dimensional integral equals the d -dimensional surplus, as asserted in (4.22). \square

Again, note the close relations between the hierarchical approach and integral transforms. Applying successive partial integration to (4.22), we get

$$v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})} = \int_{\Omega} \sigma_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) D^{\mathbf{p}+\mathbf{1}} u(\mathbf{x}) d\mathbf{x} = (-1)^{|\mathbf{l} \cdot (\mathbf{p}+\mathbf{1})|_1} \int_{\Omega} \hat{\sigma}_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) u(\mathbf{x}) d\mathbf{x}, \quad (4.23)$$

where $\hat{\sigma}_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x})$ equals $D^{\mathbf{p}+\mathbf{1}} \sigma_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x})$ in a weak sense and is a linear combination of $(p+2)^d$ Dirac pulses of alternating sign. Thus, again, the surplus can be interpreted as the coefficient resulting from an integral transform based on the function $\hat{\sigma}_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x})$.

Next, (4.22) leads us to upper bounds for the d -dimensional surplus of degree \mathbf{p} and for $W_1^{(\mathbf{p})}$'s contribution $u_1^{(\mathbf{p})}$ to the hierarchical representation of u .

Lemma 4.6. Let $u \in X_0^{p+1,q}$ be given in its hierarchical representation (4.18). Then, with

$$c(\mathbf{p}) := \prod_{j=1}^d \frac{2^{p_j \cdot (p_j+1)/2}}{(p_j + 1)!}, \quad (4.24)$$

and with the seminorms $|u|_{\alpha,\infty}$ and $|u|_{\alpha,2}$ defined in (3.3), the following estimates hold for the d -dimensional hierarchical surplus $v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}$:

$$\begin{aligned} |v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}| &\leq \left(\frac{1}{2}\right)^d \cdot c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+\mathbf{1})|_1} \cdot |u|_{\mathbf{p}+\mathbf{1},\infty}, \\ |v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}| &\leq \left(\frac{1}{6}\right)^{d/2} \cdot c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+\mathbf{1})|_1} \cdot 2^{|\mathbf{l}|_1/2} \cdot \|D^{\mathbf{p}+\mathbf{1}} u|_{\text{supp}(\phi_{\mathbf{l},\mathbf{i}})}\|_2. \end{aligned} \quad (4.25)$$

Proof. Owing to (4.21), (4.22), and $s_{l_j,i_j}^{(p_j)}(x_j) \geq 0$, we have

$$\begin{aligned} |v_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}| &= \left| \int_{\Omega} \sigma_{\mathbf{l},\mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) \cdot D^{\mathbf{p}+\mathbf{1}} u(\mathbf{x}) d\mathbf{x} \right| \\ &= \left| \int_{\Omega} \prod_{j=1}^d \frac{w'_{l_j,i_j}(x_{l_j,i_j})}{p_j!} \cdot s_{l_j,i_j}^{(p_j)}(x_j) \cdot D^{\mathbf{p}+\mathbf{1}} u(\mathbf{x}) d\mathbf{x} \right| \end{aligned}$$

$$\begin{aligned} &\leq \left| \prod_{j=1}^d \frac{w'_{l_j, i_j}(x_{l_j, i_j})}{p_j!} \cdot \int_{[0,1]} s_{l_j, i_j}^{(p_j)}(x_j) dx_j \right| \cdot \|D^{\mathbf{p}+1} u\|_\infty \\ &= \prod_{j=1}^d \left| \frac{w'_{l_j, i_j}(x_{l_j, i_j})}{p_j!} \cdot \int_{[0,1]} s_{l_j, i_j}^{(p_j)}(x_j) dx_j \right| \cdot |u|_{\mathbf{p}+1, \infty}. \end{aligned}$$

Because of (4.10), each of the d factors in the above product is the absolute value of the hierarchical surplus of $x_j^{p_j+1}/(p_j+1)!$, for $j = 1, \dots, d$, which is bounded by

$$\frac{1}{(p_j+1)!} \cdot h_{l_j}^{p_j+1} \cdot 2^{p_j \cdot (p_j+1)/2-1}$$

because of (4.6). Thus we obtain

$$\begin{aligned} |v_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}| &\leq \prod_{j=1}^d \frac{1}{(p_j+1)!} \cdot h_{l_j}^{p_j+1} \cdot 2^{p_j \cdot (p_j+1)/2-1} \cdot |u|_{\mathbf{p}+1, \infty} \\ &= 2^{-d} \cdot c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot |u|_{\mathbf{p}+1, \infty}. \end{aligned}$$

For the bound with respect to the L_2 -norm, we start from

$$|v_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}| = \left| \int_{\Omega} \sigma_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}(\mathbf{x}) \cdot D^{\mathbf{p}+1} u(\mathbf{x}) d\mathbf{x} \right| \leq \|\sigma_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}\|_2 \cdot \|D^{\mathbf{p}+1} u|_{\text{supp}(\phi_{\mathbf{l}, \mathbf{i}})}\|_2.$$

According to (4.21), and since $s_{l_j, i_j}^{(p_j)}(x_j) \geq 0$,

$$\begin{aligned} \|\sigma_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}\|_2 &= \prod_{j=1}^d \left\| \frac{w'_{l_j, i_j}(x_{l_j, i_j})}{p_j!} \cdot s_{l_j, i_j}^{(p_j)} \right\|_2 \\ &= \prod_{j=1}^d \left(\int_{[0,1]} \left| \frac{w'_{l_j, i_j}(x_{l_j, i_j})}{p_j!} \cdot s_{l_j, i_j}^{(p_j)}(x_j) \right|^2 dx_j \right)^{1/2} \\ &\leq \prod_{j=1}^d \left(\max_{x_j \in [0,1]} \left| \frac{w'_{l_j, i_j}(x_{l_j, i_j})}{p_j!} \cdot s_{l_j, i_j}^{(p_j)}(x_j) \right| \cdot \frac{1}{(p_j+1)!} h_{l_j}^{p_j+1} 2^{p_j \cdot (p_j+1)/2-1} \right)^{1/2} \\ &\leq \prod_{j=1}^d \left(\max_{x_j \in [0,1]} |s_{l_j, i_j}^{(p_j)}(x_j)| \cdot \frac{1}{p_j! \cdot (p_j+1)!} h_{l_j}^{2 \cdot (p_j+1)} 2^{p_j \cdot (p_j+1)-2} \right)^{1/2} \end{aligned}$$

holds because of (4.6) and (4.10). In (4.10), choose u such that

$$D^{p_j+1} u(x_j) = \frac{w'_{l_j, i_j}(x_{l_j, i_j})}{p_j!} \cdot s_{l_j, i_j}^{(p_j)}(x_j)$$

and apply (4.6). Finally, since

$$|s_{l_j, i_j}^{(p_j)}(x_j)| \leq \frac{2}{3} \cdot \frac{1}{(p_j + 1) \cdot h_{l_j}}$$

can be shown for all $x_j \in [0, 1]$ and for all possible l_j , i_j , and p_j , we get

$$\begin{aligned} \|\sigma_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}\|_2 &\leq 2^{-d} \cdot \left(\frac{2}{3}\right)^{d/2} \cdot c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot 2^{|\mathbf{l}|_1/2}, \\ |v_{\mathbf{l}, \mathbf{i}}^{(\mathbf{p})}| &\leq \left(\frac{1}{6}\right)^{d/2} \cdot c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot 2^{|\mathbf{l}|_1/2} \cdot \|D^{\mathbf{p}+1} u|_{\text{supp}(\phi_{\mathbf{l}, \mathbf{i}})}\|_2. \end{aligned}$$

Lemma 4.7. Let $u \in X_0^{p+1, q}$ be given in its representation (4.18). Then the following upper bounds hold for the contributions $u_{\mathbf{l}}^{(\mathbf{p})} \in W_{\mathbf{l}}^{(\mathbf{p})}$:

$$\begin{aligned} \|u_{\mathbf{l}}^{(\mathbf{p})}\|_{\infty} &\leq 0.5585^d c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot |u|_{\mathbf{p}+1, \infty}, \\ \|u_{\mathbf{l}}^{(\mathbf{p})}\|_2 &\leq 1.117^d \cdot \left(\frac{1}{3}\right)^{d/2} c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot |u|_{\mathbf{p}+1, 2}, \\ \|u_{\mathbf{l}}^{(\mathbf{p})}\|_E &\leq 3.257 \cdot \left(\frac{5}{8}\right)^{d/2} c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot \left(\sum_{j=1}^d 2^{2l_j}\right)^{1/2} \cdot |u|_{\mathbf{p}+1, \infty}, \\ \|u_{\mathbf{l}}^{(\mathbf{p})}\|_E &\leq 3.257 \cdot \left(\frac{5}{12}\right)^{d/2} c(\mathbf{p}) \cdot 2^{-|\mathbf{l} \cdot (\mathbf{p}+1)|_1} \cdot \left(\sum_{j=1}^d 2^{2l_j}\right)^{1/2} \cdot |u|_{\mathbf{p}+1, 2}. \end{aligned} \tag{4.26}$$

Proof. All results follow from the previous lemmata, with arguments completely analogous to the piecewise linear case. \square

Now, we are ready for the optimization process studied in detail for the piecewise linear case in Section 3. For the p -regular scenario, a slight simplification allows us to use the diagonal subspace pattern again, that is,

$$V_n^{(1,1)} := V_n^{(1)}, \quad V_n^{(p,1)} := \bigoplus_{|\mathbf{l}|_1 \leq n+d-1} W_{\mathbf{l}}^{(\mathbf{p})} \quad \text{for } p > 1. \tag{4.27}$$

As before, note that \mathbf{p} is not constant, owing to (4.16). The following theorem deals with the approximation quality of these sparse grid spaces $V_n^{(p,1)}$.

Theorem 4.8. For the L_{∞} -, L_2 -, and energy norm, the following bounds for the error of the interpolant $u_n^{(p,1)} \in V_n^{(p,1)}$ of $u \in X_0^{p+1, q}$ hold:

$$\begin{aligned} \|u - u_n^{(p,1)}\|_{\infty} &\leq \left(\frac{0.5585}{2^{p+1}}\right)^d c(p \cdot \mathbf{1}) \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot A(d, n) \cdot h_n^{p+1} + O(h_n^{p+1}) \\ &= O(h_n^{p+1} \cdot n^{d-1}), \end{aligned} \tag{4.28}$$

$$\begin{aligned}
\|u - u_n^{(p,1)}\|_2 &\leq \left(\frac{1.117}{\sqrt{3} \cdot 2^{p+1}} \right)^d \cdot c(p \cdot \mathbf{1}) \cdot |u|_{(p+1) \cdot \mathbf{1}, 2} \cdot A(d, n) \cdot h_n^{p+1} + O(h_n^{p+1}) \\
&= O(h_n^{p+1} \cdot n^{d-1}), \\
\|u - u_n^{(p,1)}\|_E &\leq 3.257 \cdot \left(\frac{\sqrt{5}}{\sqrt{2} \cdot 2^{p+1}} \right)^d \cdot c(p \cdot \mathbf{1}) \cdot \frac{d \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty}}{1 - 2^{-p}} \cdot h_n^p + O(h_n^p) \\
&= O(h_n^p), \\
\|u - u_n^{(p,1)}\|_E &\leq 3.257 \cdot \left(\frac{\sqrt{5}}{\sqrt{3} \cdot 2^{p+1}} \right)^d \cdot c(p \cdot \mathbf{1}) \cdot \frac{d \cdot |u|_{(p+1) \cdot \mathbf{1}, 2}}{1 - 2^{-p}} \cdot h_n^p + O(h_n^p) \\
&= O(h_n^p).
\end{aligned}$$

Proof. Actually, there is just one major difference compared to the proof of (3.68) dealing with the piecewise linear case. Now, owing to (4.16), the polynomial degree \mathbf{p} is not constant for all subspaces $W_1^{(\mathbf{p})}$ neglected in $V_n^{(p,1)}$, but depends on the respective level \mathbf{l} . However, the influence of all subspaces with $\mathbf{p} < p \cdot \mathbf{1}$ can be collected in a term of the order $O(h_n^{p+1})$ with respect to the L_∞ - and L_2 -norm or of the order $O(h_n^p)$ with respect to the energy norm, if $n \geq p - 1$: for sufficiently large n , each of those subspaces involves at least one coordinate direction x_j with $l_j = O(n)$ and $p_j = p$.

Therefore we can proceed as in the proof of (3.68) and assume a constant degree $\mathbf{p} = p \cdot \mathbf{1}$. With (4.26) and (3.66) for $s = p + 1$, we get

$$\begin{aligned}
\|u - u_n^{(p,1)}\|_\infty &\leq \sum_{|\mathbf{l}|_1 > n+d-1} \|u_{\mathbf{l}}^{(\mathbf{p})}\|_\infty \\
&\leq \sum_{|\mathbf{l}|_1 > n+d-1} 0.5585^d \cdot c(p \cdot \mathbf{1}) \cdot 2^{-(p+1) \cdot |\mathbf{l}|_1} \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty} + O(h_n^{p+1}) \\
&= 0.5585^d \cdot c(p \cdot \mathbf{1}) \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot \sum_{|\mathbf{l}|_1 > n+d-1} 2^{-(p+1) \cdot |\mathbf{l}|_1} + O(h_n^{p+1}) \\
&\leq 0.5585^d \cdot c(p \cdot \mathbf{1}) \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot 2^{-(p+1) \cdot (n+d)} \cdot A(d, n) + O(h_n^{p+1}) \\
&= \left(\frac{0.5585}{2^{p+1}} \right)^d \cdot c(p \cdot \mathbf{1}) \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot A(d, n) \cdot h_n^{p+1} + O(h_n^{p+1}) \\
&= O(h_n^{p+1} \cdot n^{d-1})
\end{aligned}$$

owing to the definition of $A(d, n)$ in (3.65), and, by analogy, the correspond-

ing result for the L_2 -norm. Concerning the energy norm, we have

$$\begin{aligned} \|u - u_n^{(p,1)}\|_E &\leq \sum_{|\mathbf{l}|_1 > n+d-1} \|u_{\mathbf{l}}^{(\mathbf{p})}\|_E \\ &\leq 3.257 \left(\frac{5}{8} \right)^{d/2} c(p \cdot \mathbf{1}) |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot \sum_{|\mathbf{l}|_1 > n+d-1} 2^{-(p+1) \cdot |\mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \\ &\quad + O(h_n^p) \end{aligned}$$

and, as for the linear case,

$$\begin{aligned} \|u - u_n^{(p,1)}\|_E &= 3.257 \left(\frac{5}{8} \right)^{d/2} c(p \cdot \mathbf{1}) |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot \sum_{i=n+d}^{\infty} 2^{-(p+1) \cdot i} \cdot \sum_{|\mathbf{l}|_1=i} \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \\ &\quad + O(h_n^p) \\ &\leq 3.257 \left(\frac{5}{8} \right)^{d/2} c(p \cdot \mathbf{1}) |u|_{(p+1) \cdot \mathbf{1}, \infty} \cdot \sum_{i=n+d}^{\infty} d \cdot 2^{-p \cdot i} + O(h_n^p) \\ &\leq 3.257 \left(\frac{\sqrt{5}}{\sqrt{2} \cdot 2^{p+1}} \right)^d c(p \cdot \mathbf{1}) \frac{d \cdot |u|_{(p+1) \cdot \mathbf{1}, \infty}}{1 - 2^{-p}} \cdot h_n^p + O(h_n^p) \\ &= O(h_n^p), \end{aligned}$$

because $\sum_{|\mathbf{l}|_1=i} \left(\sum_{j=1}^d 4^{l_j} \right)^{1/2} \leq d \cdot 2^i$ as in the proof of (3.68). The second energy estimate can be obtained in an analogous way. \square

This theorem shows that our approach, indeed, leads to a sparse grid approximation of higher order. For the space $V_n^{(p,1)}$, *i.e.*, for a maximum degree of p in each direction, we get an interpolation error of the order $O(h_n^{p+1} \cdot |\log_2 h_n|^{d-1})$ with respect to both the L_∞ - and the L_2 -norm. For the energy norm, the result is an error of the order $O(h_n^p)$.

Of course, the above optimization process can be based on the energy norm, too. For that, we start from (4.26) and define the local cost–benefit ratio, now with respect to the energy norm (*cf.* (3.69)):

$$\begin{aligned} \text{cbr}_E(\mathbf{l}) &:= \frac{b_E(\mathbf{l})}{c(\mathbf{l})} \\ &= 10.608 \cdot \left(\frac{5}{4} \right)^d \cdot c^2(\mathbf{p}) \cdot 2^{-|2 \cdot \mathbf{l} \cdot \mathbf{p} + 3 \cdot \mathbf{l}|_1} \cdot \left(\sum_{j=1}^d 4^{l_j} \right) \cdot |u|_{\mathbf{p}+\mathbf{1}, \infty}^2. \end{aligned} \tag{4.29}$$

Although we do not want to study the energy-based sparse grid spaces of higher order $V_n^{(p,E)}$ resulting from (4.29) in detail, we shall nevertheless

mention their most important properties. First, we can show that

$$V_n^{(E)} \subseteq V_n^{(p,E)} \subseteq V_n^{(1)} \quad (4.30)$$

is valid for arbitrary n . Second, as long as $V_n^{(p,E)}$ is not the same space as $V_n^{(1)}$, we are rid of the log-factors, again – both concerning the overall cost (which is of the order $O(h_n^{-1})$) and the approximation quality with respect to the energy norm (which is of the order $O(h_n^p)$).

We close this discussion by briefly returning to the notion of ε -complexity. As in the discussion of the piecewise linear counterpart, all occurring order terms have to be read with respect to ε or N . That is, for the following, both d and p are supposed to be arbitrary, but fixed.

Lemma 4.9. For the ε -complexities $N_\infty^{(p)}(\varepsilon)$, $N_2^{(p)}(\varepsilon)$, and $N_E^{(p)}(\varepsilon)$ of the problem of computing the interpolant $u_n^{(p,1)} \in V_n^{(p,1)}$ with respect to the L_∞ -, L_2 -, and the energy norm for given accuracy ε , the following relations hold:

$$\begin{aligned} N_\infty^{(p)}(\varepsilon) &= O\left(\varepsilon^{-\frac{1}{p+1}} \cdot |\log_2 \varepsilon|^{\frac{p+2}{p+1} \cdot (d-1)}\right), \\ N_2^{(p)}(\varepsilon) &= O\left(\varepsilon^{-\frac{1}{p+1}} \cdot |\log_2 \varepsilon|^{\frac{p+2}{p+1} \cdot (d-1)}\right), \\ N_E^{(p)}(\varepsilon) &= O\left(\varepsilon^{-\frac{1}{p}} \cdot |\log_2 \varepsilon|^{d-1}\right). \end{aligned} \quad (4.31)$$

Conversely, given a number of N of grid points, the following accuracies can be obtained with respect to the different norms:

$$\begin{aligned} \varepsilon_\infty^{(p)}(N) &= O\left(N^{-(p+1)} \cdot |\log_2 N|^{(p+2) \cdot (d-1)}\right), \\ \varepsilon_2^{(p)}(N) &= O\left(N^{-(p+1)} \cdot |\log_2 N|^{(p+2) \cdot (d-1)}\right), \\ \varepsilon_E^{(p)}(N) &= O\left(N^{-p} \cdot |\log_2 N|^{p \cdot (d-1)}\right). \end{aligned} \quad (4.32)$$

Finally, for the same problem tackled by the energy-based sparse grid approximation space $V_n^{(p,E)}$, we get

$$N_E^{(p)}(\varepsilon) = O\left(\varepsilon^{-\frac{1}{p}}\right) \quad \text{and} \quad \varepsilon_E^{(p)}(N) = O\left(N^{-p}\right). \quad (4.33)$$

Proof. The proof follows exactly the argumentation of the linear case. \square

In comparison with the full grid case, where we get $N_\infty^{(p)}(\varepsilon) = O(\varepsilon^{-d/(p+1)})$ and $\varepsilon_\infty^{(p)}(N) = O(N^{-(p+1)/d})$ with respect to the L_∞ - or L_2 -norm and $N_E^{(p)}(\varepsilon) = O(\varepsilon^{-d/p})$ and $\varepsilon_E^{(p)}(N) = O(N^{-p/d})$ with respect to the energy norm, as before, the sparse grid space $V_n^{(p,1)}$ lessens the curse of dimensionality in a significant manner; $V_n^{(p,E)}$, however, completely overcomes it.

4.3. Interpolets

Another hierarchical multiscale basis with higher-order functions is given by the interpolet family (Deslauriers and Dubuc 1989, Donoho and Yu 1999). These functions are obtained from a simple but powerful interpolation process. For given data $y(s)$, $s \in \mathbb{Z}$, we seek an interpolating function $y : \mathbb{R} \rightarrow \mathbb{R}$ which is as smooth as possible. To this end, in a first step, the interpolated values in $\mathbb{Z} + \frac{1}{2}$ are determined. Here, for the determination of $y(s + \frac{1}{2})$, the Lagrangian polynomial $p(x)$ of degree $2n-1$ is calculated which interpolates the data in $s - n + 1, \dots, s + n$. Then we set $y(s + \frac{1}{2}) := p(s + \frac{1}{2})$. The parameter n later determines the smoothness of the interpolant and the degree of polynomial exactness. Since the values $y(\mathbb{Z})$ and $y(\mathbb{Z} + \frac{1}{2})$ are now known, we can compute the values $y(\mathbb{Z} + \frac{1}{4})$ and $y(\mathbb{Z} + \frac{3}{4})$ using the same scheme. Here, for example, the values $y(s + \frac{1}{2}(-n+1)), \dots, y(s + \frac{n}{2})$ are used for $y(s + \frac{1}{4})$. This way, interpolation values for y can be found on a set which is dense in \mathbb{R} . Since the interpolant depends linearly on the data, there exists a fundamental function F with

$$y(x) = \sum_{s \in \mathbb{Z}} y(s)F(x - s).$$

F is the interpolant for the data $y(s) = \delta_{0,s}, s \in \mathbb{Z}$. Now, the interpolet mother function of the order $2n$ is just $\phi := F$. A hierarchical multiscale basis is formed from that ϕ by dilation and translation as in (3.8). The function ϕ has the following properties.

Scaling equation. ϕ is the solution of

$$\phi(x) = \sum_{s \in \mathbb{Z}} h_s \phi(2x - s). \quad (4.34)$$

The mask coefficients $\mathbf{h} := \{h_s\}_{s \in \mathbb{Z}}$ are given by $h_0 = 1$, $h_s = h_{-s}$ ($s \in \mathbb{Z}$), and

$$\begin{pmatrix} 1^0 & 3^0 & \dots & (2n-1)^0 \\ 1^2 & 3^2 & \dots & (2n-1)^2 \\ \vdots & \vdots & \ddots & \vdots \\ 1^{2n-2} & 3^{2n-2} & \dots & (2n-1)^{2n-2} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_3 \\ \vdots \\ h_{2n-1} \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (4.35)$$

All other h_s are zero.

Compact support.

$$\text{supp } \phi = [-2n+1, 2n-1]. \quad (4.36)$$

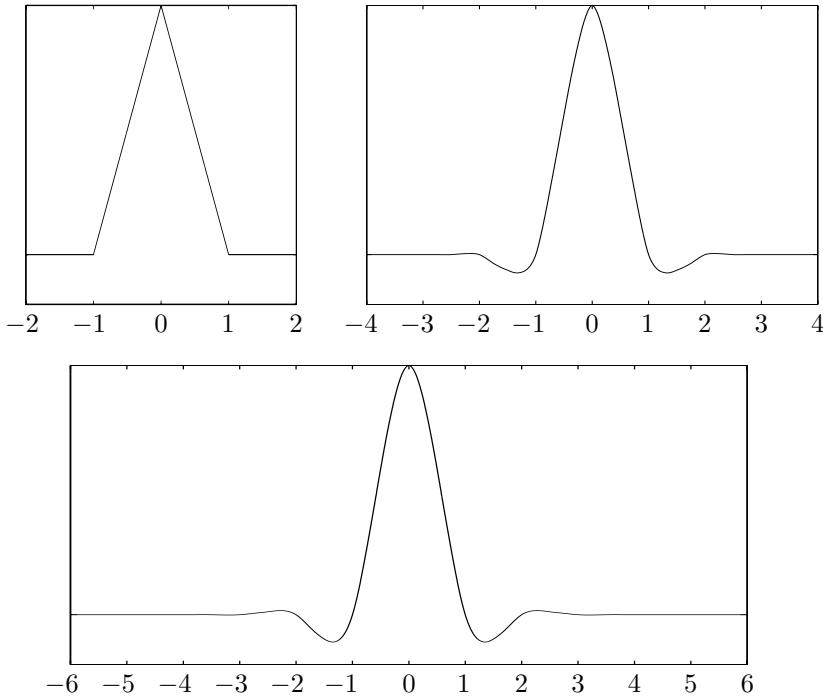


Figure 4.9. The interpolets ϕ for $N = 2, 4$ and 6 .

Polynomial exactness. In a pointwise sense, polynomials of degree *less* than $N = 2n$ can be written as linear combinations of translates of ϕ , *e.g.*,

$$\forall 0 \leq i < N : \forall x \in \mathbb{R} : x^i = \sum_{s \in \mathbb{Z}} s^i \phi(x - s). \quad (4.37)$$

Interpolation property. With $\delta(x)$ denoting the Dirac functional, we get

$$\forall s \in \mathbb{Z} : \phi(s) = \delta(s). \quad (4.38)$$

This is the main property which distinguishes interpolating multiscale bases from other non-interpolating multiscale bases. It allows particularly simple multilevel algorithms for the evaluation of nonlinear terms.

The functions ϕ for different values of the parameter $N = 2n$ are given in Figure 4.9. Note that there is not much difference between the interpolets with $N = 4$ and $N = 6$. This behaviour is quite similar to that of the higher-order polynomials of Section 4.2, which were created by hierarchical Lagrangian interpolation. Note, furthermore, that whilst interpolets are defined on the whole of \mathbb{R} , their construction can easily be adapted to a bounded interval; see Koster (2002) for the respective construction.

Now, we can use such a higher-order interpolet function ϕ as the mother function in (3.7) for the definition of the 1D hierarchical basis $\{\phi_{l_j,i_j}(x_j)\}$ in (3.8). Again, as in our introductory example of piecewise linear hierarchical bases of Section 3.1, we use these 1D functions as the input of the tensor product construction, which provides a suitable piecewise d -dimensional basis function $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) := \prod_{j=1}^d \phi_{l_j,i_j}(x_j)$ in each grid point $\mathbf{x}_{\mathbf{l},\mathbf{i}}$. In an analogous way to that of Sections 3 and 4.2, we can derive sparse grids based on interpolets. Also, all other considerations regarding the estimates for the cost and error complexities carry over accordingly.

4.4. Prewavelets and wavelets

Note again that the presented hierarchical multiscale bases with higher-order polynomials or interpolets, after the tensor product construction, allow for a relatively cheap evaluation of differential operators and discretization schemes on sparse grids in a Galerkin approach owing to their interpolating properties. Also, upper error estimates can be easily derived for various sparse grid spaces following the arguments in Sections 3.2 and 4.2. However, they all exhibit a main drawback: there is no lower error estimate with constants independent of the number of levels involved, that is, they form no *stable* multiscale splitting (Oswald 1994). The consequences are twofold. First, the (absolute) value of the hierarchical coefficient is just a local error indicator and no true error estimator. We obtain sufficiently refined sparse grids (compare the associated numerical experiments in Section 5) on which the error is properly reduced, but it may happen that too many grid points are employed for a prescribed error tolerance in an adaptive procedure. Efficiency is thus not guaranteed. Second, the condition number of the linear system which results from a symmetric elliptic partial differential operator in multiscale basis representation is, after diagonal scaling, in general not independent of the finest mesh size involved.³ To obtain a fast mesh independent solver, additional lifting tricks (Sweldens 1997, Daubechies and Sweldens 1998, Koster 2002) or multigrid-type extensions (Oswald 1994, Griebel 1994a, Griebel and Oswald 1994, Griebel and Oswald 1995a, Griebel and Oswald 1995b) are necessary. These difficulties are avoided if we employ stable multiscale splittings (Oswald 1994, Dahmen and Kunoth 1992, Carnicer, Dahmen and Pena 1996, Cohen 2003) and the respective L_2 - and H_1 -stable multiscale bases, for which two-sided estimates exist.

³ This is the case for stable splittings with wavelets. Then a simple fast solver results from the diagonal scaling preconditioner: see Dahmen and Kunoth (1992) and Oswald (1994).

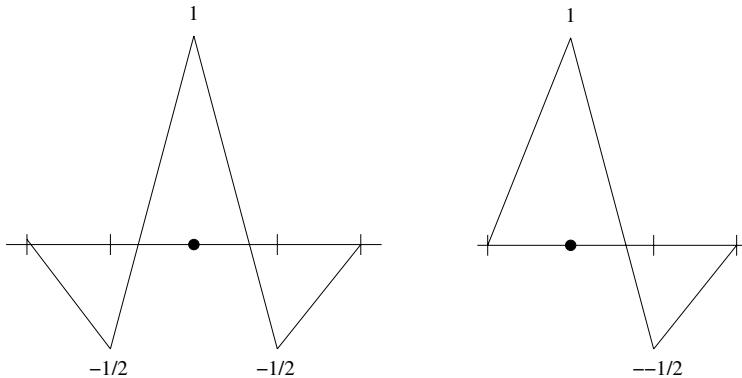


Figure 4.10. Non-orthogonal linear spline wavelet (left: interior; right: boundary).

Since Daubechies' fundamental discovery of orthogonal multiscale bases with local support, *i.e.*, the classical wavelets (Daubechies 1988), an enormous literature has arisen. At present, there exists a whole zoo of such bases from the wavelet family. Here, we stick to simple hierarchical spline-like function systems. Within this class, we have orthonormal wavelets (Lemarié 1988), biorthogonal spline wavelets (Cohen, Daubechies and Feauveau 1992), semi-orthogonal spline wavelets (Chui and Wang 1992), fine grid correction spline wavelets (Cohen and Daubechies 1996, Lorentz and Oswald 1998) and multiwavelets derived from splines. The construction principles of such functions are highly developed, and there is abundant literature on them. Most theory can be found in Daubechies (1992), Chui (1992) and Meyer (1992), and the references cited therein. Further reading is Daubechies (1988, 1993), and Cohen *et al.* (1992, 2001). A nice introduction, similar in spirit to this paper, is given in Cohen (2003).

In the following, let us briefly mention the simplest but also cheapest mother wavelets which are made up from piecewise linears by stable completion procedures (Carnicer *et al.* 1996). They are sufficient to be used for a second-order PDE within the Galerkin method.

These are the so-called linear pre-wavelets, which are a special case of non-orthogonal spline wavelets (Cohen and Daubechies 1996, Lorentz and Oswald 1998, Stevenson 1996) of linear order. The corresponding mother function ϕ is shown in Figure 4.10 (left). Again, by translation and dilation in an analogous way to (3.8), we get a 1D multilevel basis from this ϕ . Note that a modification is necessary near the boundary. For homogeneous Dirichlet conditions, the scaled and dilated function in Figure 4.10 (right) can be used in points next to the boundary.

Another, still reasonably cheap approach involves linear prewavelets, which are a special case of semi-orthogonal spline wavelets (Chui and Wang 1992,

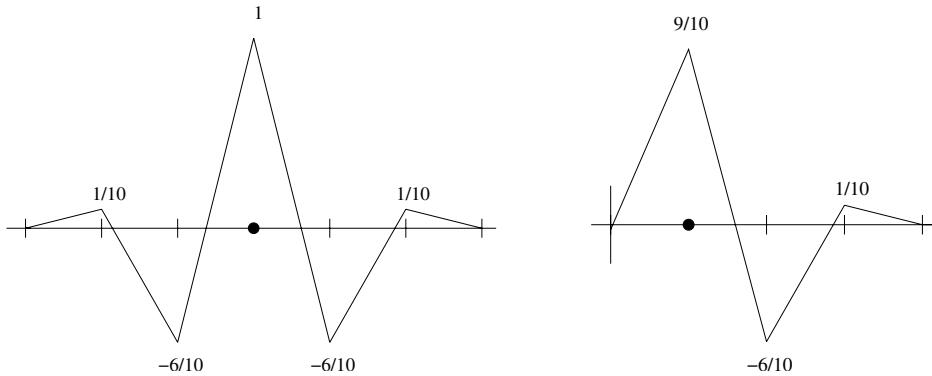


Figure 4.11. Semi-orthogonal linear spline wavelet (left: interior; right: boundary).

Griebel and Oswald 1995b). The corresponding mother function ϕ is shown in Figure 4.11 (left). Again, a modification is necessary near the boundary. For homogeneous Dirichlet conditions, this is shown in Figure 4.11 (right). A construction for Neumann conditions can be found in Griebel and Oswald (1995b).

Finally, let us mention the so-called lifting wavelets. They are the result of the application of Sweldens' lifting scheme (Carnicer *et al.* 1996, Sweldens 1997) to the interpolet basis $\{\phi_{l_j,i_j}(x_j)\}$ or other interpolatory multiscale bases, for example. Here, the hierarchical (*i.e.*, if we consider odd indices only; see (3.11)) lifting wavelets are defined by

$$\hat{\phi}_{l_j,2i_j+1} := \phi_{(l_j,2i_j+1)} + \sum_{s_j} Q_{s_j,2i_j+1}^{l_j} \phi_{(l_j-1,s_j)}$$

on finer levels $l_j > 0$. The basic idea is to choose the weights $Q_{s,i_j}^{l_j}$ in this linear combination in such a way that $\hat{\phi}_{l_j,i_j}$ has more vanishing moments than ϕ_{l_j,i_j} , and thus to obtain a stabilization effect. If we apply this approach to the hierarchical interpolet basis of Section 4.3 so that we achieve two vanishing moments in the lifting wavelet basis, we end up with

$$\hat{\phi}_{l_j,2i_j+1} = \phi_{(l_j,2i_j+1)} - \frac{1}{4}(\phi_{(l_j-1,i_j)} + \phi_{(l_j-1,i_j+1)})$$

for the odd indices $2i_j + 1$. The corresponding mother function $\hat{\phi}$ is shown in Figure 4.12.

Again, as in our introductory example of piecewise linear hierarchical bases of Section 3.1, we can use these 1D multilevel basis functions as the input of the tensor product construction which provides a suitable piecewise d -dimensional basis function $\phi_{\mathbf{l},\mathbf{i}}(\mathbf{x})$ in each grid point $\mathbf{x}_{\mathbf{l},\mathbf{i}}$. As in Sections 3

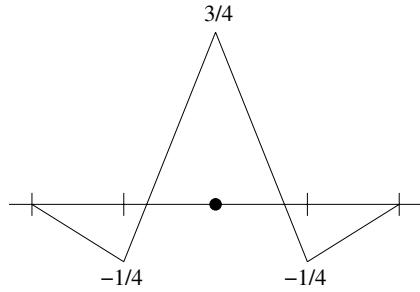


Figure 4.12. Lifting interpolant with $N = 2$ and two vanishing moments.

and 4.2, we can derive sparse grids based on wavelets. Also, most other considerations regarding the estimates for the cost and error complexities carry over accordingly. Further information on wavelets and sparse grids can be found in Griebel and Oswald (1995b), Hochmuth (1999), DeVore, Konyagin and Temlyakov (1998), Koster (2002), Knapek (2000a), Griebel and Knapek (2000) and the references cited therein.

At the end of this subsection, let us dwell a bit into theory, and let us give an argument why there is a difference between hierarchical polynomials and interpolants on the one hand and prewavelets and pre-prewavelets on the other hand.

For all mother functions and the resulting product multiscale bases and sparse grid subspaces, we can, in principle, follow the arguments and proofs of Sections 3 and 4.2, respectively, to obtain cost complexity estimates and upper error bounds with relatively sharp estimates for the order constants. The main tools in the proofs are the simple triangle inequality and geometric series arguments. However, as already mentioned, a lower bound for the error can not be obtained so easily. An alternative approach is that of Griebel and Oswald (1995b) and Knapek (2000a), where we developed a technique for which two-sided error norm estimates for the 1D situation can be carried over to the higher-dimensional case. The approach is based on the representation of Sobolev spaces $H^s([0, 1]^d)$, $s \geq 0$, as

$$H^s([0, 1]^d) = \bigcap_{i=1}^d \underbrace{L_2([0, 1]) \otimes \cdots \otimes L_2([0, 1])}_{(i-1) \text{ times}} \otimes H^s([0, 1]) \otimes \underbrace{L_2([0, 1]) \otimes \cdots \otimes L_2([0, 1])}_{(d-i) \text{ times}}. \quad (4.39)$$

The Sobolev space $H_{\text{mix}}^s([0, 1]^d)$, $s \geq 0$, on the other hand is defined as the

simple tensor product

$$H_{\text{mix}}^s([0, 1]^d) = \underbrace{H^s([0, 1]) \otimes \cdots \otimes H^s([0, 1])}_{d \text{ times}}.$$

Now, for the different components of the intersection, we obtain two-sided norm estimates, if the univariate multiscale functions ϕ_{l_j, i_j} of (3.8) allow two-sided norm estimates for both H^s and L_2 .

Theorem 4.10. Let the univariate multiscale basis $\{\phi_{l_j, i_j}\}$ satisfy the norm equivalence

$$\|u\|_{H^s}^2 \sim \sum_{l_j} 2^{-l_j} \sum_{i_j} 2^{2l_j s} |u_{l_j, i_j}|^2, \quad (4.40)$$

where $u(x_j) = \sum_{l_j, i_j} u_{l_j, i_j} \phi_{l_j, i_j}(x_j)$, and for $-\gamma_1 < s < \gamma_2$, $\gamma_1, \gamma_2 > 0$. Then, the multivariate basis functions $\{\phi_{\mathbf{l}, \mathbf{i}}(\mathbf{x})\}$ fulfil the norm equivalences

$$\begin{aligned} \|u\|_{H^s}^2 &= \left\| \sum_{\mathbf{l}, \mathbf{i}} u_{\mathbf{l}, \mathbf{i}} \phi_{\mathbf{l}, \mathbf{i}} \right\|_{H^s}^2 \sim \sum_{\mathbf{l}, \mathbf{i}} 2^{2s|\mathbf{l}|_\infty} |u_{\mathbf{l}, \mathbf{i}}|^2 2^{-|\mathbf{l}|_1}, \\ \|u\|_{H_{\text{mix}}^s}^2 &= \left\| \sum_{\mathbf{l}, \mathbf{i}} u_{\mathbf{l}, \mathbf{i}} \phi_{\mathbf{l}, \mathbf{i}} \right\|_{H_{\text{mix}}^s}^2 \sim \sum_{\mathbf{l}, \mathbf{i}} 2^{2s|\mathbf{l}|_1} |u_{\mathbf{l}, \mathbf{i}}|^2 2^{-|\mathbf{l}|_1}, \end{aligned}$$

where $u(\mathbf{x}) = \sum_{\mathbf{l}, \mathbf{i}} u_{\mathbf{l}, \mathbf{i}} \phi_{\mathbf{l}, \mathbf{i}}(\mathbf{x})$.

Here, \sim denotes a two-sided equivalence, *i.e.*, $a \sim b$ means that there exist positive constants c_1, c_2 such that $c_1 \cdot b \leq a \leq c_2 \cdot b$. Note the distinct difference in the quality of the two estimates. It is exactly the difference between $|\mathbf{l}|_\infty$ and $|\mathbf{l}|_1$ which leads to a significant reduction in ε -complexity, *i.e.*, it allows us to use substantially fewer degrees of freedom to reach the same truncation error in the H_{mix}^s -norm than in the H^s -norm.

For the proof and further details, see Griebel and Oswald (1995b), Knapek (2000a) and Koster (2002). Here, the bounds γ_1, γ_2 for the range of the regularity parameter s depend on the specific choice of the mother function ϕ . The value γ_2 is determined by the Sobolev regularity of ϕ , *i.e.*, $\gamma_2 = \sup\{s : \phi \in H^s\}$. The theory here works with biorthogonality arguments and dual spaces: our univariate multiscale basis $\{\phi_{l_j, i_j}\}$ possesses a *dual basis* $\{\tilde{\phi}_{l_j, i_j}\}$, in the sense that (for our hierarchical functions⁴) the

⁴ In the general setting, there is a multiresolution analysis with spaces spanned by scaling functions and difference spaces spanned by wavelets. Then, biorthogonality conditions must be fulfilled between all these function systems. In the hierarchical approach, certain scaling functions are simply wavelets, and the biorthogonality conditions are reduced.

biorthogonality relations (Cohen *et al.* 1992)

$$\langle \tilde{\phi}_{l_j, i_j}, \phi_{l_j, k_j} \rangle = \delta_{i_j, k_j}, \quad \langle \tilde{\phi}_{l_j-1, i_j}, \phi_{l_j, k_j} \rangle = \delta_{i_j, k_j}.$$

If the primal and dual functions are at least in L_2 , then $\langle \cdot, \cdot \rangle$ is the usual L_2 -inner product. The value of γ_1 is here just the Sobolev regularity of the dual mother function $\tilde{\phi}$.

Note that, with $\gamma_1, \gamma_2 > 0$, the 1D norm equivalence for $s = 0$, and thus the L_2 -stability of the 1D hierarchical basis $\{\phi_{l_j, i_j}\}$, is a prerequisite in Theorem 4.10. However, the linear hierarchical bases of Section 3, the higher-order Lagrangian interpolation functions of Section 4.2 and the interpolants of Section 4.3 do not fulfil this prerequisite, so they are not L_2 -stable.⁵ Then we obtain simply one-sided estimates. For these function systems, we get the same upper estimates with our simple tools as the triangle inequality and the geometric series arguments, but we can pinpoint, additionally, quite sharp bounds for the constants. The one-sided estimates still allow us to use the hierarchical coefficients $|u_{1,i}|$ (*e.g.*, after suitable weighting) as error indicators in a refinement procedure, but give us no true error estimator. To this end, grids are constructed in an adaptation process steered by the indicator such that a prescribed global error tolerance is reached. This aim, however, might be reached by using more points than necessary. For the function systems built from the wavelet-type mother functions, the prerequisite of L_2 -stability is fulfilled, and we obtain two-sided estimates. Thus, the wavelet coefficients $|u_{1,i}|$ can serve as local error estimators, and the lower estimate part then gives us efficiency of the corresponding adaptive scheme, which means that only the necessary number of points (up to a constant) to reach a prescribed error tolerance is employed in a grid adaptation process. Furthermore, a fast solver can now be gained with level-independent convergence rate by simple diagonal scaling as a preconditioner in the multiscale system.

Note that from relation (4.39) it also becomes clear that the constants in the multivariate norm equivalence are more or less given by the d th power of the constants in the univariate equivalence. This causes rather large constants for higher dimensionality d .

4.5. Sparse grid applications

PDE discretization techniques

Since its introduction, the sparse grid concept has been applied to most of the relevant discretization schemes for PDEs. These are finite element

⁵ The reason is that the dual ‘functions’ are Dirac functionals. Then $\gamma_1 = -\frac{1}{2}$ and the norm equivalency (4.40) is only valid for $s \in]\frac{1}{2}, \gamma_2[$, *i.e.*, it does not hold for $s = 0$.

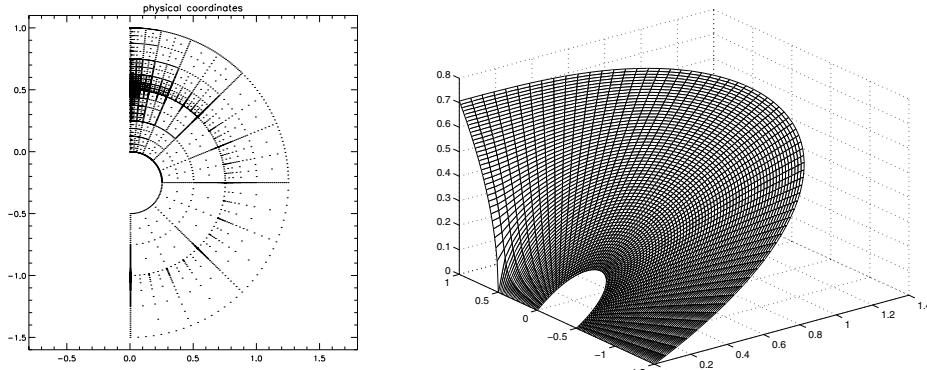


Figure 4.13. Laplace equation with a singularity on the boundary:
adaptive sparse grid (left) and corresponding solution (right).

methods and Galerkin techniques, finite differences, finite volumes, spectral methods, and splitting extrapolation, which leads to the combination technique.

The first main focus of the development of sparse grid discretization methods was on piecewise linear *finite elements*. In the pioneering work of Zenger (1991) and Griebel (1991b), the foundations for adaptive refinement, multilevel solvers, and parallel algorithms for sparse grids were laid. Subsequent studies included the solution of the 3D Poisson equation Bungartz (1992a, 1992b), the generalization to arbitrary dimensionality d (Balder 1994) and to more general equations (the Helmholtz equation (Balder and Zenger 1996), parabolic problems using a time–space discretization (Balder, Rüde, Schneider and Zenger 1994), the biharmonic equation (Störkkuhl 1995), and general linear elliptic operators of second order in 2D (Pflaum 1996, Dornseifer and Pflaum 1996). As a next step, the solution of general linear elliptic differential equations and, via mapping techniques, the treatment of more general geometries was implemented (Bungartz and Dornseifer 1998, Dornseifer 1997) (see Figure 4.13). Since then, algorithmic improvements for the general linear elliptic operator of second order have been studied in detail (Schneider 2000, Achatz 2003a, Achatz 2003b).

In the first experiments with polynomial bases of higher-order, bicubic Hermite bases for the biharmonic equation (Störkkuhl 1995) and hierarchical Lagrange bases of degree two (Bungartz 1996) as well as of arbitrary degree (Bungartz and Dornseifer 1998, Bungartz 1997) were studied. The general concept of the hierarchical Lagrangian interpolation on sparse grids was introduced in Bungartz (1998). Afterwards, the higher-order concept was combined with a duality-based adaptivity approach (Schneider 2000) and with operator extension (Achatz 2003a). In addition to that, aspects of complexity, especially for the solution of the Poisson equation (Bungartz

and Griebel 1999), were discussed with regard to Werschulz (1995). Furthermore, besides the classical hierarchical bases according to Section 3, prewavelet bases (Oswald 1994, Griebel and Oswald 1995b) and wavelet decompositions (Sprengel 1997a) were considered. Also, the parallelization of the sparse grid algorithms was pursued: see Pfaffinger (1997), Hahn (1990), Griebel (1991a, 1991b).

Another development in the discretization of PDEs on sparse grids using second-order *finite differences* was derived and implemented in Griebel (1998), Griebel and Schiekofer (1999), Schiekofer (1998), and Schiekofer and Zumbusch (1998). There, based on Taylor expansions and sufficient regularity assumptions, the same orders of consistency can be shown as they are known from the comparable full grid operators. This means, in particular, that the typical log-term does not occur. Generalizations to higher-order finite difference schemes using conventional Taylor expansion (Schiekofer 1998) and interpolets (Koster 2002) as discussed in Section 4.3 were also developed. The conceptual simplicity of finite difference schemes allowed for straightforward adaptive refinement strategies based on hashing (Griebel 1998, Schiekofer 1998), for the application of sparse grids to elliptic problems with general coefficient functions, and for the handling of general domains and nonlinear systems of parabolic PDEs of reaction–diffusion type (Schiekofer 1998). This opened the way to a first solution of the Navier–Stokes equations on sparse grids: see Schiekofer (1998), Griebel and Koster (2000), Koster (2002) and Griebel and Koster (2003).

Furthermore, the so-called *combination technique* (Griebel, Schneider and Zenger 1992), an extrapolation-type sparse grid variant, has been discussed in several papers (Bungartz, Griebel, Röschke and Zenger 1994b, 1994c, Bungartz, Griebel and Rüde 1994a, Pflaum and Zhou 1999). Because a sparse grid can be represented as a superposition of several (much coarser) full grids Ω_l (see Figure 4.14 for the 2D case), a sparse grid solution for some PDEs can be obtained from the linear combination of solutions u_l computed on the respective coarse grids. Since the latter ones are regular full grids, existing solvers can be used without any need for an explicit discretization on a sparse grid. For the 2D case, we get

$$u_n^{(c)}(\mathbf{x}) := \sum_{|\mathbf{l}|_1=n+1} u_l(\mathbf{x}) - \sum_{|\mathbf{l}|_1=n} u_l(\mathbf{x}), \quad (4.41)$$

and the 3D combination formula is given by

$$u_n^{(c)}(\mathbf{x}) := \sum_{|\mathbf{l}|_1=n+2} u_l(\mathbf{x}) - 2 \cdot \sum_{|\mathbf{l}|_1=n+1} u_l(\mathbf{x}) + \sum_{|\mathbf{l}|_1=n} u_l(\mathbf{x}). \quad (4.42)$$

Note that, in contrast to (3.19), where $u_l \in W_l$, u_l denotes the approximate solution of the underlying PDE on the regular full grid Ω_l . Compared with the direct discretization on sparse grids, the combination technique

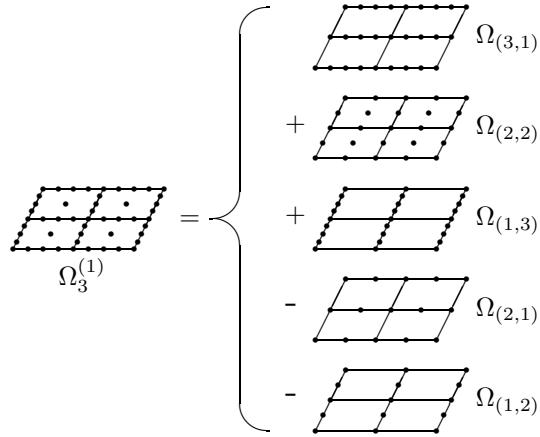


Figure 4.14. Combination technique for $d = 2$:
combine coarse full grids Ω_l , $|\mathbf{l}|_1 \in \{n, n+1\}$, with
mesh widths 2^{-l_1} and 2^{-l_2} to get a sparse grid $\Omega_n^{(1)}$
corresponding to $V_n^{(1)}$.

turns out to be advantageous in two respects. First, the possibility of using existing codes allows the straightforward application of the combination technique to complicated problems. Second, since the different subproblems can be solved fully in parallel, there is a very elegant and efficient inherent coarse-grain parallelism that makes the combination technique perfectly suited to modern high-performance computers (Griebel 1992, 1993). This has been shown in a series of papers dealing with problems from computational fluid dynamics, including turbulence simulation (Bungartz and Huber 1995, Huber 1996b, Griebel and Thurner 1993, Griebel, Huber and Zenger 1993a, Griebel 1993, Griebel and Thurner 1995, Griebel and Huber 1995, Griebel, Huber and Zenger 1996, Huber 1996a). Furthermore, note that there are close relations to the above-mentioned Boolean and blending methods (Delvos 1982, Delvos and Schempp 1989, Gordon 1969, Gordon and Hall 1973, Baszenski *et al.* 1992, Hennart and Mund 1988) as well as to the so-called *splitting extrapolation method* (Liem, Lu and Shih 1995).

Concerning alternative discretization schemes, there are ongoing investigations on sparse grid finite volume methods for the Euler equations (Hemker 1995, Hemker and de Zeeuw 1996, Koren, Hemker and de Zeeuw 1996). In this context, a variant – the so-called *sets of grids* or *grids of grids* – has been developed (Hemker and Pflaum 1997, Hemker, Koren and Noordmans 1998). This approach goes back to the notion of semi-coarsening (Mulder 1989, Naik and van Rosendale 1993). Furthermore, *spectral methods* have been im-

plemented on sparse grids and studied for PDEs with periodic boundary conditions (Kupka 1997).

Adaptivity and fast solvers

Especially in the context of PDE, two ingredients are essential for each kind of discretization scheme: adaptive mesh refinement and fast linear solvers.

Concerning the first, grid adaptation based on local error estimators (typically closely related to the hierarchical surplus, *i.e.*, the hierarchical basis coefficients) is a very widely used approach (Zenger 1991, Bungartz 1992a, 1992b, Griebel 1998, Koster 2002, Griebel and Koster 2000). However, in Schneider (2000), the *dual problem approach* (Becker and Rannacher 1996, Eriksson, Estep, Hansbo and Johnson 1996) has been successfully applied to sparse grids, allowing for global error estimation too. As a result, the error control can be governed by some appropriate linear error functional, and the respective refinement strategies include l_2 -driven as well as point-driven adaptation.

Concerning the efficient multilevel solution of the resulting linear systems, a number of contributions using additive and multiplicative subspace correction schemes have been made (Griebel 1991a, 1991b, Pflaum 1992, Griebel, Zenger and Zimmer 1993b, Griebel 1994b, Griebel and Oswald 1994, 1995a, 1995b, Bungartz 1997, Pflaum 1998), showing the availability of fast algorithms for the solution of the linear systems resulting from sparse grid discretizations.

Applications

Flow problems were the first focus for the use of sparse grid PDE solvers. Now, however, sparse grids are also used for problems from quantum mechanics (Garcke 1998, Garcke and Griebel 2000, Yserentant 2004, Hackbusch 2001), for problems in the context of stochastic differential equations (Schwab and Todor 2002, 2003), or for the discretization of differential forms arising from Maxwell's equations (Hiptmair and Gradinaru 2003).

Aside from the field of PDEs, sparse grids are being applied to a variety of problems that will not be forgotten here. Among these problems are integral equations (Frank, Heinrich and Pereverzev 1996, Griebel, Oswald and Schiekofer 1999, Knapek and Koster 2002, Knapek 2000b), general operator equations (Griebel and Knapek 2000, Knapek 2000a, Hochmuth, Knapek and Zumbusch 2000), eigenvalue problems (Garcke 1998, Garcke and Griebel 2000), periodic interpolation (Pöplau and Sprengel 1997), interpolation on Gauss–Chebyshev grids (Sprengel 1997b), Fourier transforms (Hallatschek 1992), tabulation of reduced chemical systems (Heroth 1997), digital elevation models and terrain representation Gerstner (1995, 1999), audio and image compression (Frank 1995, Paul 1995) (see Figure 4.15), and possibly others not listed here.

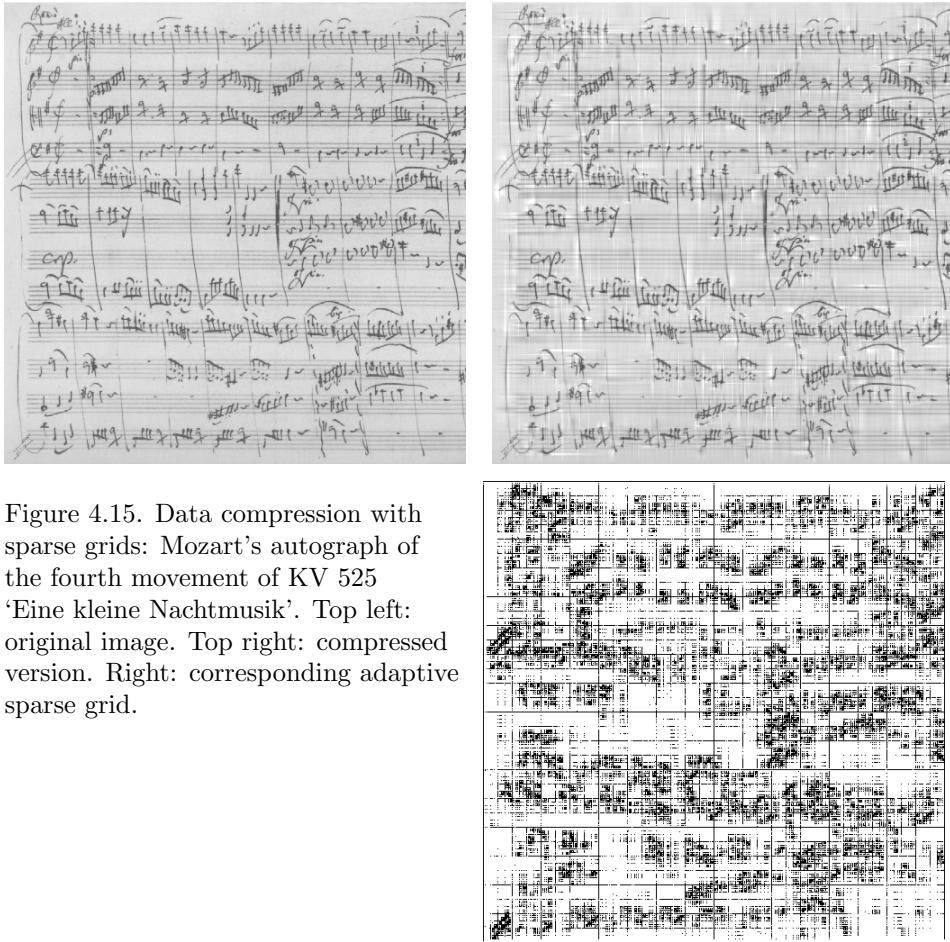


Figure 4.15. Data compression with sparse grids: Mozart’s autograph of the fourth movement of KV 525 ‘Eine kleine Nachtmusik’. Top left: original image. Top right: compressed version. Right: corresponding adaptive sparse grid.

Of course, owing to the historic background described in Section 4.1, numerical quadrature has always been a hot topic in sparse grid research. Starting from the explicit use of the piecewise linear basis functions to calculate integrals Bonk (1994*b*, 1994*a*), quadrature formulas based on the midpoint rule (Bazsensi and Delvos 1993), the rectangle rule (Paskov 1993), the trapezoidal rule (Bonk 1994*a*), the Clenshaw–Curtis rule (Cools and Maerten 1997, Novak and Ritter 1998), the Gauss rules (Novak and Ritter 1997), and the Gauss–Patterson rules (Gerstner and Griebel 1998, Petras 2000) were used as 1D input for Smolyak’s principle according to (4.1). Furthermore, the higher-order basis polynomials from Section 4.2 were used for an adaptive quadrature algorithm (Bungartz and Dirnstorfer 2003). These techniques can be used advantageously for the computation of stiffness matrix entries in conventional finite element methods and especially in partition of unity methods (Griebel and Schweitzer 2002, Schweitzer 2003). Gerstner

and Griebel (2003) developed a generalization of the conventional sparse grid approach which is able to adaptively assess the dimensions according to their importance. For further information on numerical quadrature on sparse grids and on data mining, another interesting field of application, we refer to the respective numerical examples in Section 5.3.

5. Numerical experiments

In this section, we present a collection of numerical results for different problems solved on sparse grids. We start with the discussion of the basic features of sparse grid methods for PDEs, applied to simpler 2D and 3D model problems. Then we turn to the solution of the Navier–Stokes equations on sparse grids. Finally, we illustrate the potential of sparse grids for problems of a higher dimensionality, here in the context of numerical quadrature and data mining.

5.1. PDE model problems

We start with some PDE model problems to demonstrate the interpolation properties of the L_2 -based sparse grid spaces presented as well as their behaviour concerning the approximation of a PDE’s solution. We omit examples with energy-based sparse grids, since the situation with respect to the energy norm in a finite element context is completely clear owing to Céa’s lemma, whereas the L_2 -quality of the finite element solution is still an open question.

Hence we choose a standard Ritz–Galerkin finite element setting and use the hierarchical Lagrangian basis polynomials from Section 4.2 in the p -regular way, *i.e.*, choosing the local degree as high as possible according to (4.16) up to some given maximum value for p . Note that a Petrov–Galerkin approach with the higher-order basis functions appearing in the approximation space only leads to the same approximation properties. We will not discuss the algorithmic scheme which is applied to ensure the linear complexity of the matrix–vector product, the so-called *unidirectional principle* (Bungartz 1998), allowing us to treat the single dimensions separately. However, note that, especially in the case of more general operators, this is actually the challenging part of a sparse grid finite element implementation; see Dornseifer (1997), Bungartz and Dornseifer (1998), Schneider (2000), and Achatz (2003a, 2003b). Concerning the iterative solution of the resulting linear systems, several multilevel schemes are available, most of them based upon the semi-coarsening approach of Mulder (1989) and Naik and van Rosendale (1993) and its sparse grid extension (Griebel 1991b, 1991a, Griebel *et al.* 1993b, Griebel and Oswald 1995b, Pflaum 1992, 1998, Bungartz 1997). In the context of solvers, it is important that the influence

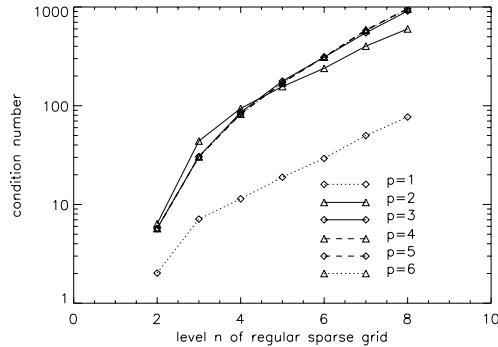


Figure 5.1. Condition numbers of the (diagonally preconditioned) stiffness matrix for the sparse grid spaces $V_n^{(p,1)}$.

of the polynomial degree starting from $p = 2$ be only moderate, which is in contrast to the behaviour known from hierarchical polynomials in a p - or $h\text{-}p$ -version context (*cf.* Zumbusch (1996), for example). Figure 5.1 hints this, showing spectral condition numbers of the stiffness matrix.

Our model problems cover regular and adaptively refined sparse grids in two and three dimensions, mainly using the Laplacian as the operator, but also the second-order linear operator with more general coefficients. For measuring the error, we consider the error's discrete maximum norm, the discrete l_2 -norm, and the error in the centre point $P = (0.5, \dots, 0.5)$. Where grid adaptation is used, it is based on a sparse grid extension of the dual problem approach (Becker and Rannacher 1996, Eriksson *et al.* 1996); see Schneider (2000).

The 2D Laplace equation

On $\bar{\Omega} = [0, 1]^2$, let

$$\begin{aligned} \Delta u(\mathbf{x}) &= 0 && \text{in } \Omega, \\ u(\mathbf{x}) &= \sin(\pi x_1) \cdot \frac{\sinh(\pi x_2)}{\sinh(\pi)} && \text{on } \partial\Omega, \end{aligned} \tag{5.1}$$

Figure 5.2 shows the smooth solution $u(\mathbf{x})$ of (5.1). We study the accuracy of the hierarchical Lagrangian approach for the regular sparse grid spaces $V_n^{(p,1)}$ with polynomial degrees $p \in \{1, \dots, 6\}$. Figure 5.3 illustrates the maximum and the l_2 -error. The effect of the higher-order approximation can be seen clearly. For $p = 6$, we already get troubles with machine accuracy because of the use of standard double precision floating point numbers.

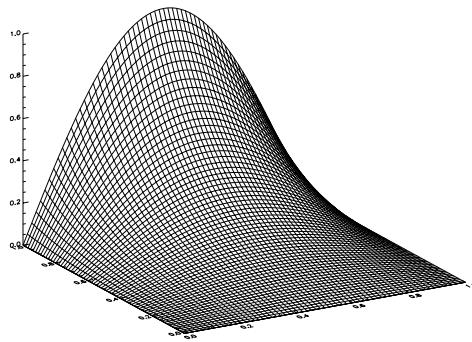


Figure 5.2. Solution $u(\mathbf{x}) = \sin(\pi x_1) \cdot \frac{\sinh(\pi x_2)}{\sinh(\pi)}$ of (5.1).

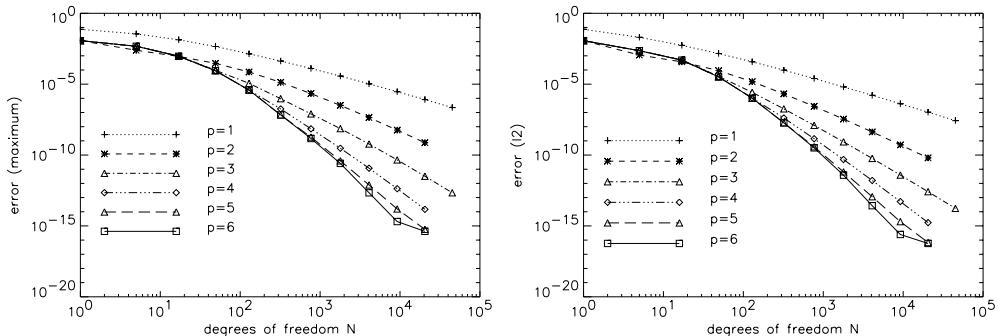


Figure 5.3. Example (5.1): maximum and l_2 -error for regular sparse grid spaces $V_n^{(p,1)}$, $p \in \{1, \dots, 6\}$.

In Figure 5.4, a summary of the convergence behaviour with respect to all error norms regarded and for $p \in \{1, \dots, 6\}$ is provided. In addition to the error plots, we show the curves of expected sparse grid convergence (sgc) due to the results (3.88) and (4.32), stating

$$\begin{aligned}\varepsilon_\infty(N) &= O(N^{-2} \cdot |\log_2 N|^{3 \cdot (d-1)}), \\ \varepsilon_\infty^{(p)}(N) &= O(N^{-(p+1)} \cdot |\log_2 N|^{(p+2) \cdot (d-1)}).\end{aligned}\quad (5.2)$$

These curves indicate the (asymptotic) behaviour of the ε -complexity of sparse grids with respect to the problem of interpolating a given function. For all polynomial degrees presented, we observe a rather striking correspondence of theoretical results concerning the mere quality of interpolation and experimental results for the accuracy of calculating approximate solutions of PDEs. Note that this correspondence is still an open question with respect to the L_2 - or maximum norm, in contrast to the energy norm for which the answer is provided by Céa's lemma.

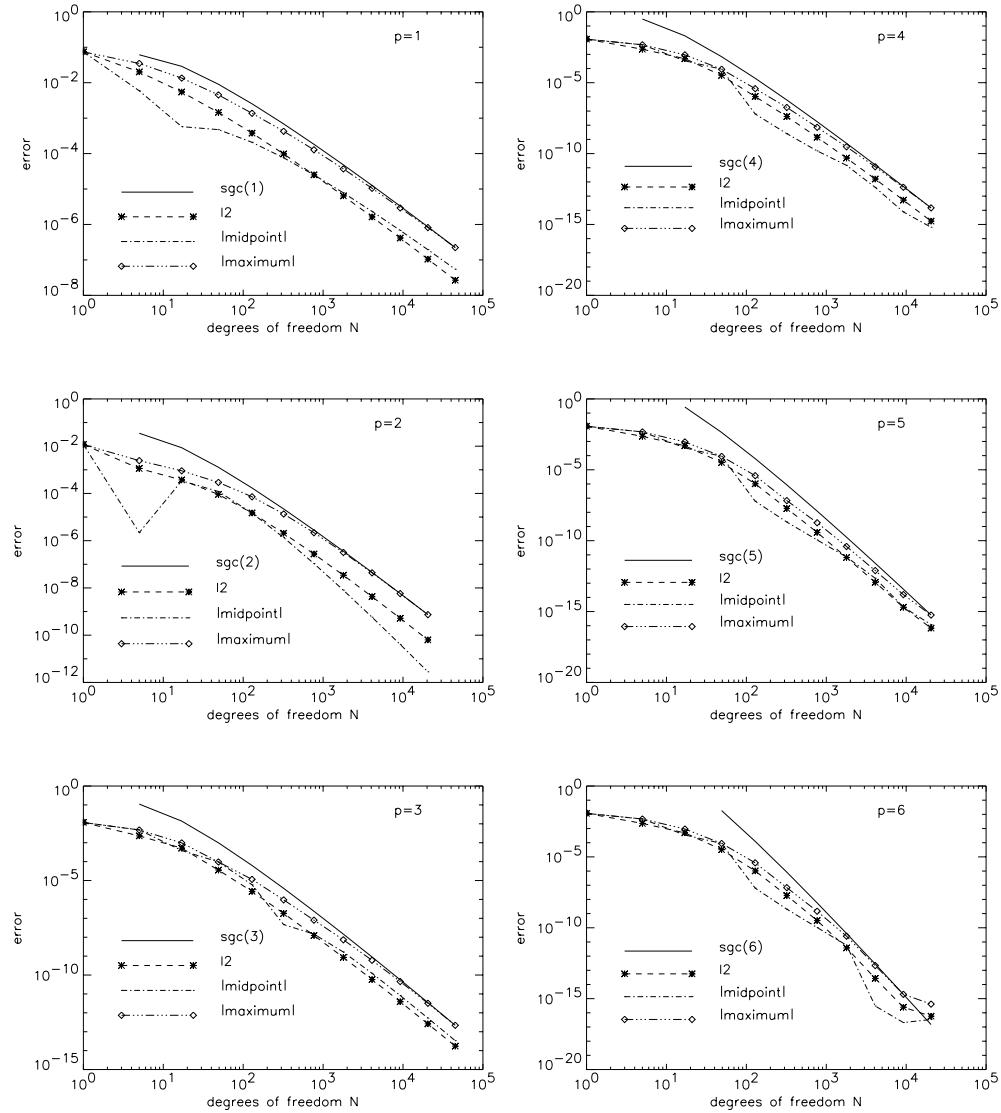


Figure 5.4. Example (5.1): convergence on regular sparse grid spaces $V_n^{(p,1)}$ for $p \in \{1, 2, 3\}$ (left) and $p \in \{4, 5, 6\}$ (right); the solid lines indicate the respective expected sparse grid convergence (sgc) due to the interpolation accuracy.

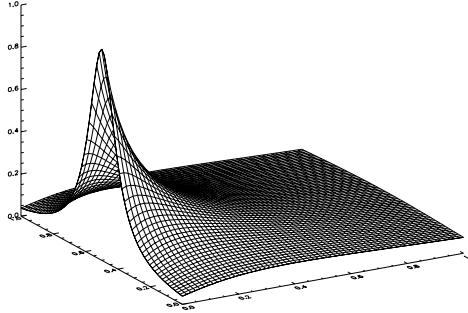


Figure 5.5. Solution $u(\mathbf{x}) = \frac{0.1 \cdot (x_1 + 0.1)}{(x_1 + 0.1)^2 + (x_2 - 0.5)^2}$ of example (5.3).

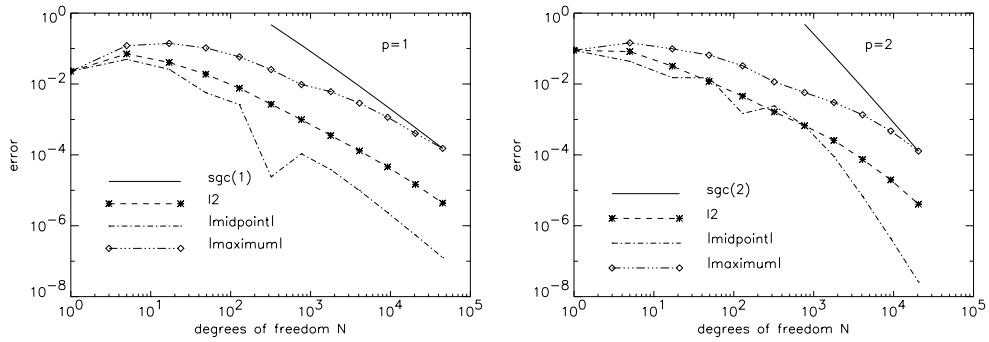


Figure 5.6. Example (5.3): convergence on $V_n^{(p,1)}$ for $p = 1$ (left) and $p = 2$ (right); the solid lines indicate the respective expected sparse grid convergence (sgc; position of curve chosen for clarity).

Next, we turn to adaptively refined sparse grids. On $\bar{\Omega} = [0, 1]^2$, let

$$\begin{aligned} \Delta u(\mathbf{x}) &= 0 && \text{in } \Omega, \\ u(\mathbf{x}) &= \frac{0.1 \cdot (x_1 + 0.1)}{(x_1 + 0.1)^2 + (x_2 - 0.5)^2} && \text{on } \partial\Omega. \end{aligned} \quad (5.3)$$

In $\mathbf{x} = (-0.1, 0.5)$, $u(\mathbf{x})$ has a singularity which is outside $\bar{\Omega}$, but close to the boundary $\partial\Omega$. Figure 5.5 shows the solution $u(\mathbf{x})$.

Figure 5.6 illustrates the case of the regular sparse grid spaces $V_n^{(1,1)}$ (left) and $V_n^{(2,1)}$ (right). Obviously, things are not too convincing for reasonable N . However, decisive progress can be made if we apply our l_2 -adaptivity.

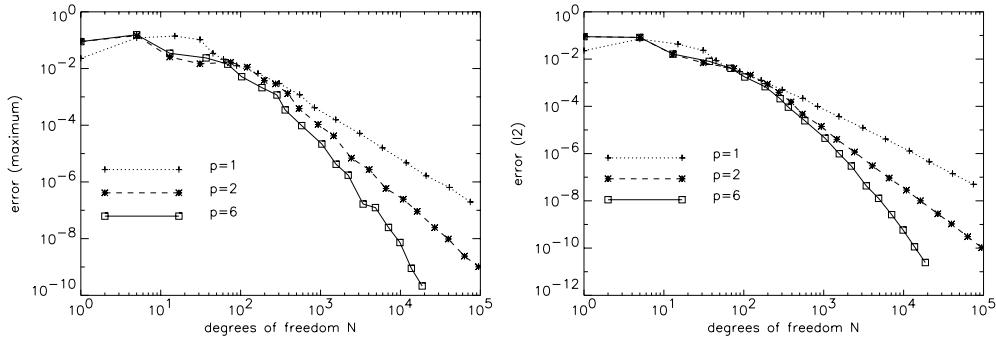


Figure 5.7. Example (5.3): maximum and l_2 -error for adaptive refinement, $p \in \{1, 2, 6\}$.

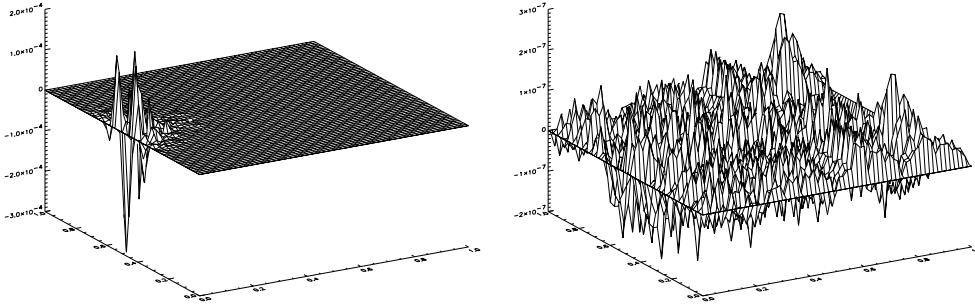


Figure 5.8. Example (5.3): error on regular (left) and adaptive sparse grid (right).

Figure 5.7 shows a gain in accuracy with higher p that is comparable to the smooth situation of (5.1).

In Figure 5.8, we compare the error on a regular sparse grid and on an adaptively refined one. As expected, the l_2 -adaptation process reduces the error equally over the whole domain. In contrast to that, regular sparse grids show large errors near the singularity.

As for (5.1), the achieved accuracy will be compared to the theoretical results concerning the ε -complexity of interpolation on sparse grids. In Figure 5.9, again for $p = 1$, the correspondence is striking. For $p = 6$, it seems that the asymptotic behaviour needs bigger values of N to appear. This was to be expected, since our hierarchical Lagrangian basis polynomials of degree $p = 6$ need at least level 5 to enter the game. However, the adaptation process causes a delayed creation of new grid points in the smooth areas.

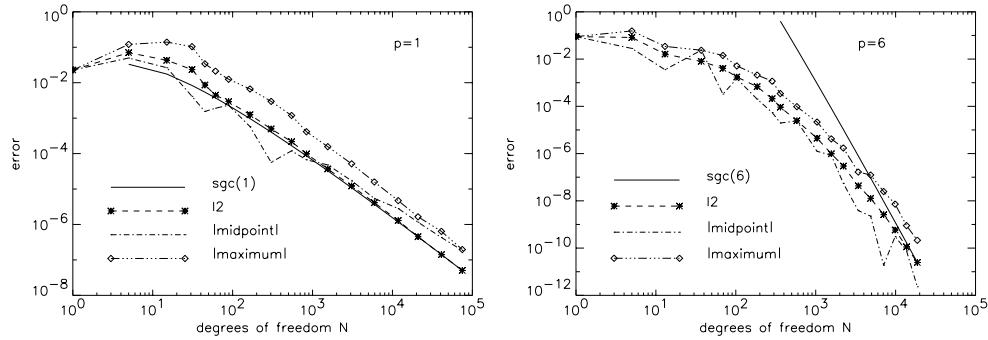


Figure 5.9. Example (5.3): convergence with adaptive mesh refinement for $p = 1$ (left) and $p = 6$ (right); the solid lines indicate the respective expected sparse grid convergence (sgc ; position of curve chosen for clarity).

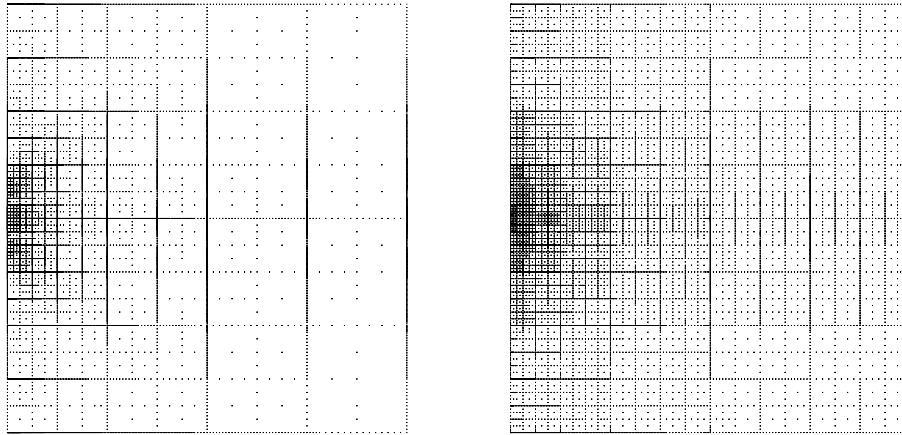


Figure 5.10. Example (5.3): adaptive grids for $p = 1$ (left) and $p = 6$ (right).

Nevertheless, with the adaptive refinement advancing, the higher-order accuracy comes to fruition.

Finally, to get an impression of the adaptation process, Figure 5.10 shows three adaptively refined grids with 7641 grid points ($p = 1$, left), 12853 grid points ($p = 2$, middle), and 10965 grid points ($p = 6$, right).

With these results supporting the efficiency of the combination of higher-order approximation and adaptive mesh refinement on sparse grids, we close the discussion of the 2D Laplacian and turn to 3D problems.

The 3D Laplace equation

For the 3D case, we restrict ourselves to a smooth model problem. On $\bar{\Omega} = [0, 1]^3$, let

$$\begin{aligned}\Delta u(\mathbf{x}) &= 0 && \text{in } \Omega, \\ u(\mathbf{x}) &= \frac{\sinh(\sqrt{2}\pi x_1)}{\sinh(\sqrt{2}\pi)} \cdot \sin(\pi x_2) \cdot \sin(\pi x_3) && \text{on } \partial\Omega.\end{aligned}\quad (5.4)$$

For the polynomial degrees $p \in \{1, \dots, 4\}$, Figure 5.11 compares the accuracy with respect to the error's maximum norm or its l_2 -norm, respectively. Again, the effects of the improved approximation properties of our hierarchical polynomial bases are evident.

Figure 5.12 shows that we do already come quite close to the asymptotic behaviour predicted for the quality of mere interpolation. This is true in spite of the fact that up to 100 000 degrees of freedom are not excessive for a 3D problem. Remember that, although the curse of dimensionality is lessened significantly by the L_2 - or L_∞ -based sparse grid spaces $V_n^{(1)}$ and $V_n^{(p,1)}$, there is still some d -dependence. Evaluating the respective order terms of the ε -complexity derived before and given once more in (5.2) for the 3D case, we observe an exponent of $(p+2) \cdot (d-1) = 2p+4$ in the log-factor, *i.e.*, an exponent of 12 for polynomial degree $p=4$, for example. Nevertheless, as in the 2D case of (5.1) and (5.3), the benefit caused by higher polynomial degrees in combination with sparse grid discretization is again rather impressive.

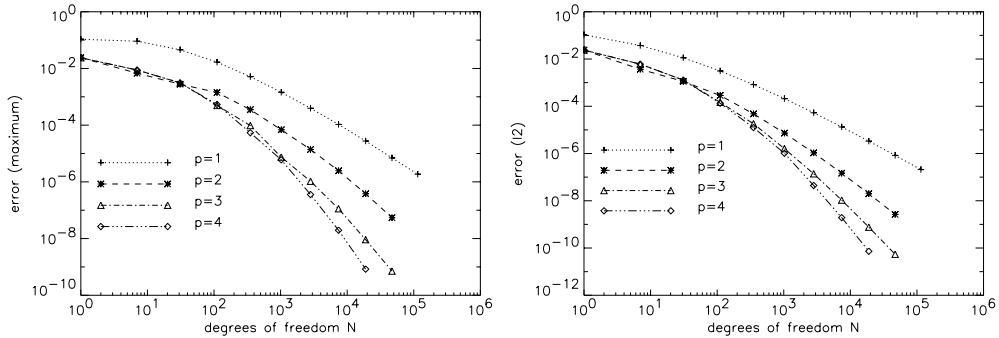


Figure 5.11. Example (5.4): maximum and l_2 -error for regular sparse grid spaces $V_n^{(p,1)}$, $p \in \{1, 2, 3, 4\}$.

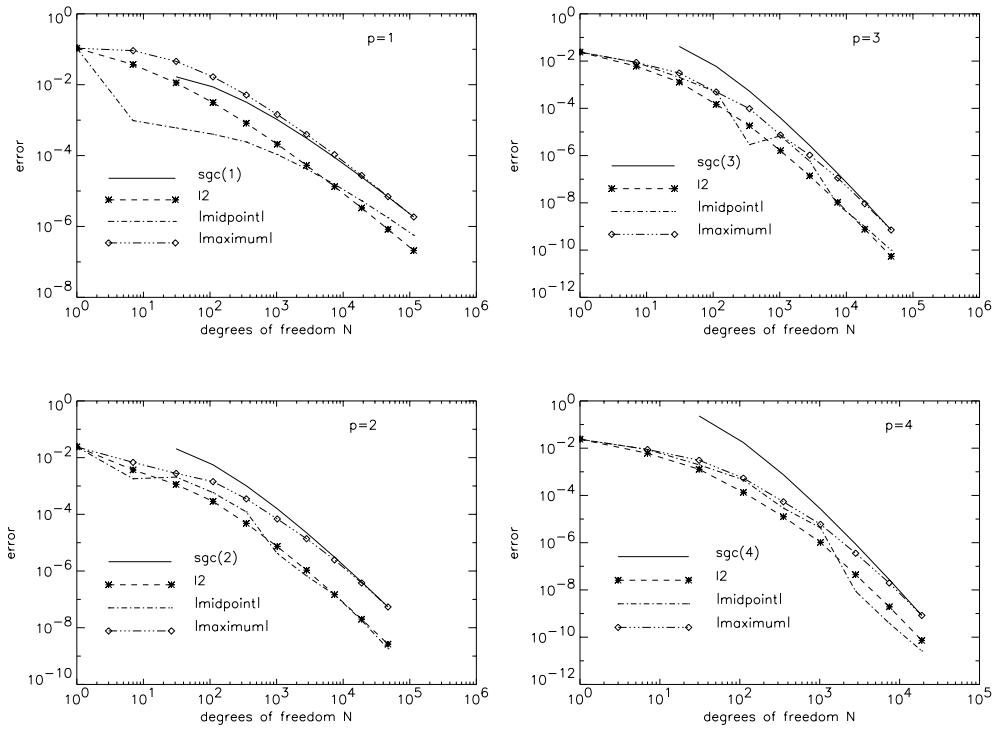


Figure 5.12. Example (5.4): convergence on $V_n^{(p,1)}$, $p \in \{1, 2\}$ (left) and $p \in \{3, 4\}$ (right); the solid lines indicate the respective expected sparse grid convergence (sgc; position of curve chosen for clarity).

Towards greater generality

Figure 5.13 illustrates the effect of using the point functional with respect to the midpoint P of Ω for the dual problem-based control of adaptive refinement instead of the l_2 -adaptation regarded up to now. It can be shown that the resulting error estimator is strongly related to the energy of the basis function living in P . Since we have an $O(h^p)$ -behaviour of the energy norm, the energy decreases of the order $O(h^{2p})$. Consequently, a convergence behaviour of $O(h^{2p})$ for the error in the midpoint can be expected. Actually, for the quadratic case, this can be seen from Figure 5.13 for both example (5.1) and even for the root singularity $\text{Re}(z^{1/2})$ on $\Omega :=]0, 1[\times]-\frac{1}{2}, \frac{1}{2}[$. Note that the solid line now indicates an N^{-4} -behaviour. In addition to (5.3), the results for the root singularity show that we can tackle really singular problems too.

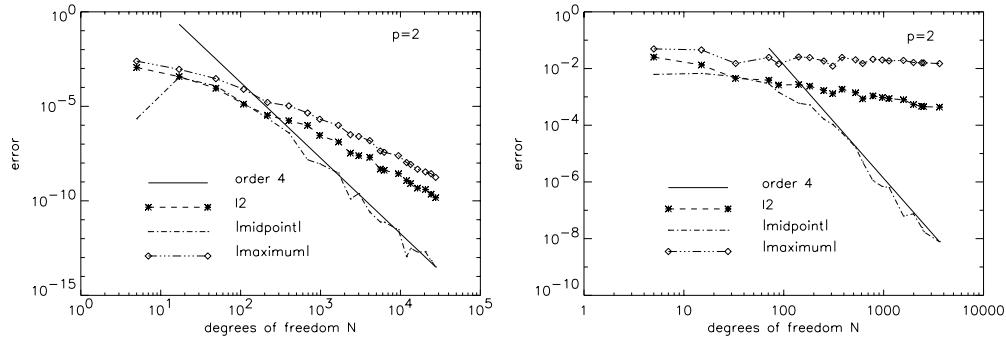


Figure 5.13. Reducing the error in the midpoint for $p = 2$: example (5.1) (left) and root singularity $\text{Re}(z^{1/2})$ (right); the solid lines indicate the expected convergence N^{-4} (position of curve chosen for clarity).

In Figure 5.14, we present the error resulting from adaptive refinement based on the point functional for example (5.1). Obviously, the error is primarily reduced near the midpoint. The right-hand side of Figure 5.14 shows the underlying adaptive grid consisting of 12 767 grid points.

Since all examples have so far been treated with a p -regular approach, we want to present at least one result for a p -asymptotic proceeding, *i.e.*, for the scenario of increasing p for each new level l of grid points occurring. On $\Omega = [0, 1]^2$, let

$$\begin{aligned} -\Delta u(\mathbf{x}) &= -2 \cdot \frac{x_1 \cdot (x_1^2 x_2^2 - 3x_2^2 + x_1^4 + 2x_1^2 + 1)}{(x_1^2 + 1)^3} && \text{in } \Omega, \\ u(\mathbf{x}) &= \frac{x_1 x_2^2}{x_1^2 + 1} && \text{on } \partial\Omega. \end{aligned} \quad (5.5)$$

According to the p -asymptotic strategy, in Figure 5.15, the polynomial degree is chosen as $\mathbf{p} := \mathbf{l} + \mathbf{1}$ up to $p = 12$. In order to avoid trouble with machine accuracy, we switch to quadruple precision. Of course, striving for twenty-five or more correct digits is not particularly useful. Nevertheless, we obtain a convincing diagram that shows the sub-exponential behaviour of the p -asymptotic strategy which, by the way, performs better than theory had suggested.

The last finite element example of this section demonstrates that our approach is not limited to simple operators such as the Laplacian. Here we turn to a variable diffusion coefficient. For a detailed discussion of the general linear elliptic operator of second order, see Achatz (2003b), for instance.

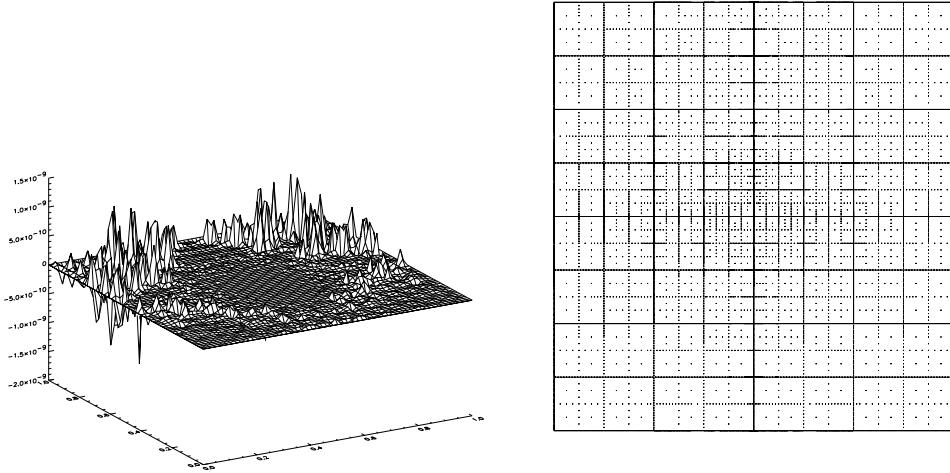


Figure 5.14. Example (5.1): reducing the error in the midpoint for $p = 2$; error (left) and adaptively refined grid (right).

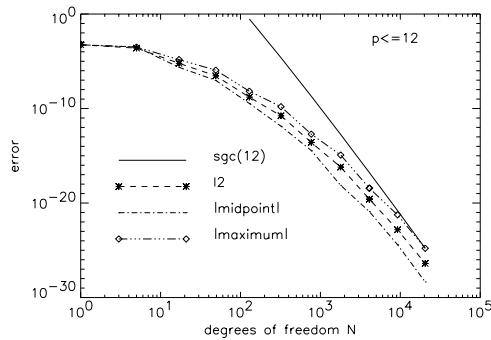


Figure 5.15. Example (5.5): regular sparse grid spaces $V_n^{(p,1)}$, p -asymptotic proceeding up to $p = 12$ with quadruple precision.

On the unit square, let

$$-\nabla \cdot (A(\mathbf{x}) \nabla u(\mathbf{x})) = f(\mathbf{x}) \quad \text{in } \Omega, \quad (5.6)$$

where

$$A(\mathbf{x}) := \begin{pmatrix} 4 + \sin(2\pi x_1) + \sin(2\pi x_2) & 1 + 4x_2 \cdot (1 - x_2) \cdot \sin(2\pi x_1) \\ 1 + 4x_2 \cdot (1 - x_2) \cdot \sin(2\pi x_1) & 4 + \sin(2\pi x_1 x_2) \end{pmatrix},$$

and where $f(\mathbf{x})$ and the Dirichlet boundary conditions are chosen such that

$$u(\mathbf{x}) := \sin(\pi x_1) \cdot \sin(\pi x_2) \quad (5.7)$$

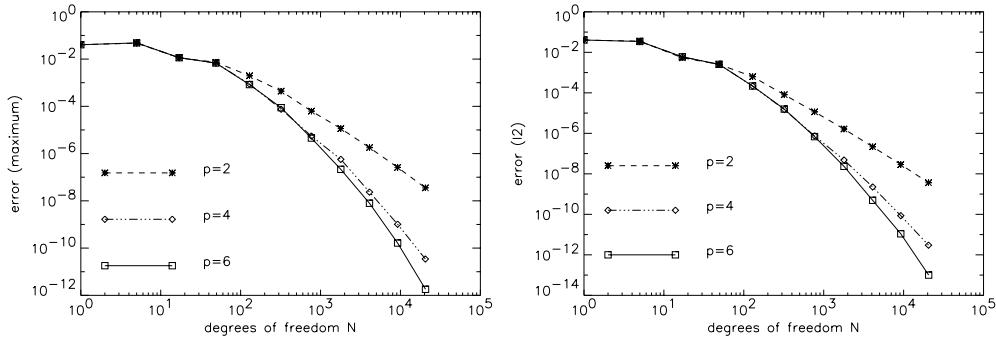


Figure 5.16. Example (5.6): maximum and l_2 -error for the regular sparse grid spaces $V_n^{(p,1)}$, $p \in \{2, 4, 6\}$.

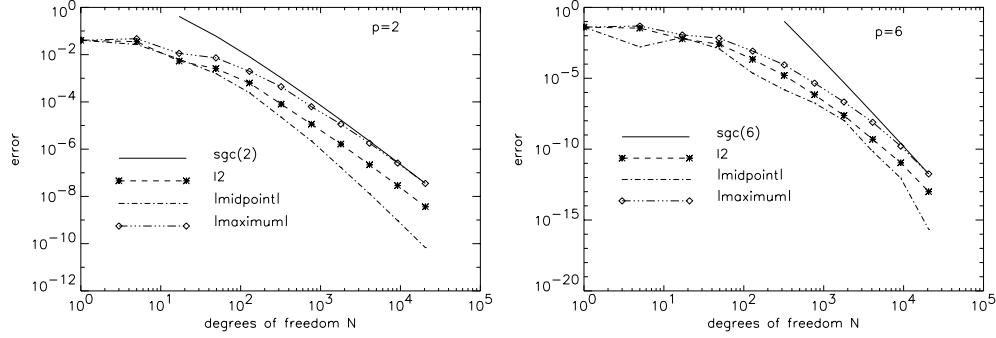


Figure 5.17. Example (5.6): convergence on $V_n^{(p,1)}$ for $p = 2$ (left) and $p = 6$ (right); the solid lines indicate the respective expected sparse grid convergence (sgc; position of curve chosen for clarity).

is the solution. We present results for $p \in \{2, 4, 6\}$. Note that, owing to the smoothness of $u(\mathbf{x})$, we restrict ourselves to the regular sparse grid spaces $V_n^{(p,1)}$. Figure 5.16 shows the maximum and the l_2 -error.

Obviously, the higher-order approximation of our hierarchical Lagrangian basis polynomials comes to fruition in the more general situation of example (5.6) too. Figure 5.17 illustrates that we come close to the expected asymptotic behaviour already for moderate values of N .

So far, all problems have been tackled with a finite element approach. However, finite differences have been successfully used for the numerical solution of PDEs too (see Griebel and Schiekofer (1999), Schiekofer (1998), Schiekofer and Zumbusch (1998), for example). Hence we want to present one model problem that has been solved with finite differences on sparse grids, *i.e.*, on $V_n^{(1)}$, to be precise. On the unit cube, define the matrix

$$\hat{A}(\mathbf{x}) := \begin{pmatrix} x_1^2 \exp(\pi x_2 - 1) & x_1^2 & \pi^{-3} \\ 0 & \pi & 0 \\ 0 & \sin(x_1) & 0.1 \cdot (x_1^2 + 2x_2) \end{pmatrix},$$

the vector

$$\hat{b}(\mathbf{x}) := \begin{pmatrix} \hat{b}_1(\mathbf{x}) \\ \hat{b}_2(\mathbf{x}) \\ \hat{b}_3(\mathbf{x}) \end{pmatrix} := \begin{pmatrix} x_1/(x_1 x_3 + 0.1) \\ \cos(\exp(x_1 + x_2 x_3)) \\ x_1 x_2^2 \end{pmatrix},$$

and the scalar function

$$\hat{c}(\mathbf{x}) := \pi \cdot (x_1 + x_2 + x_3).$$

Furthermore, choose the right-hand side $f(\mathbf{x})$ such that

$$u(\mathbf{x}) = \arctan \left(100 \left(\frac{x_1 + x_2 + x_3}{\sqrt{3}} - \frac{4}{5} \right) \prod_{i=1}^3 (x_i - x_i^2) \right) \quad (5.8)$$

solves

$$-\nabla \cdot (\hat{A}(\mathbf{x}) \nabla u(\mathbf{x})) + \hat{b}(\mathbf{x}) \cdot \nabla u(\mathbf{x}) + \hat{c}(\mathbf{x}) u(\mathbf{x}) = f(\mathbf{x}) \quad (5.9)$$

on the unit cube with Dirichlet boundary conditions. Figure 5.19 (overleaf) illustrates the solution $u(\mathbf{x})$ due to (5.8) for the two x_3 -values 0.5 and 0.875. Numerical results are presented in Figure 5.18.

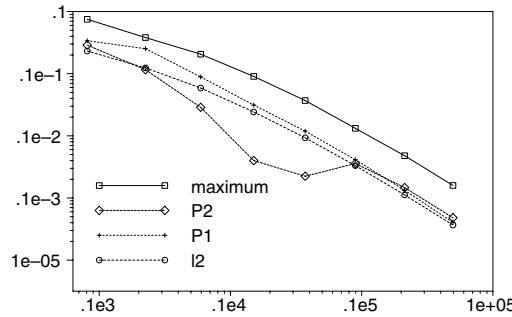


Figure 5.18. Example (5.9), finite differences, regular sparse grid spaces $V_n^{(1)}$: maximum l_2 -error as well as errors in $P_1 = (0.5, 0.5, 0.5)$ and $P_2 = (1/\pi, 1/\pi, 1/\pi)$.

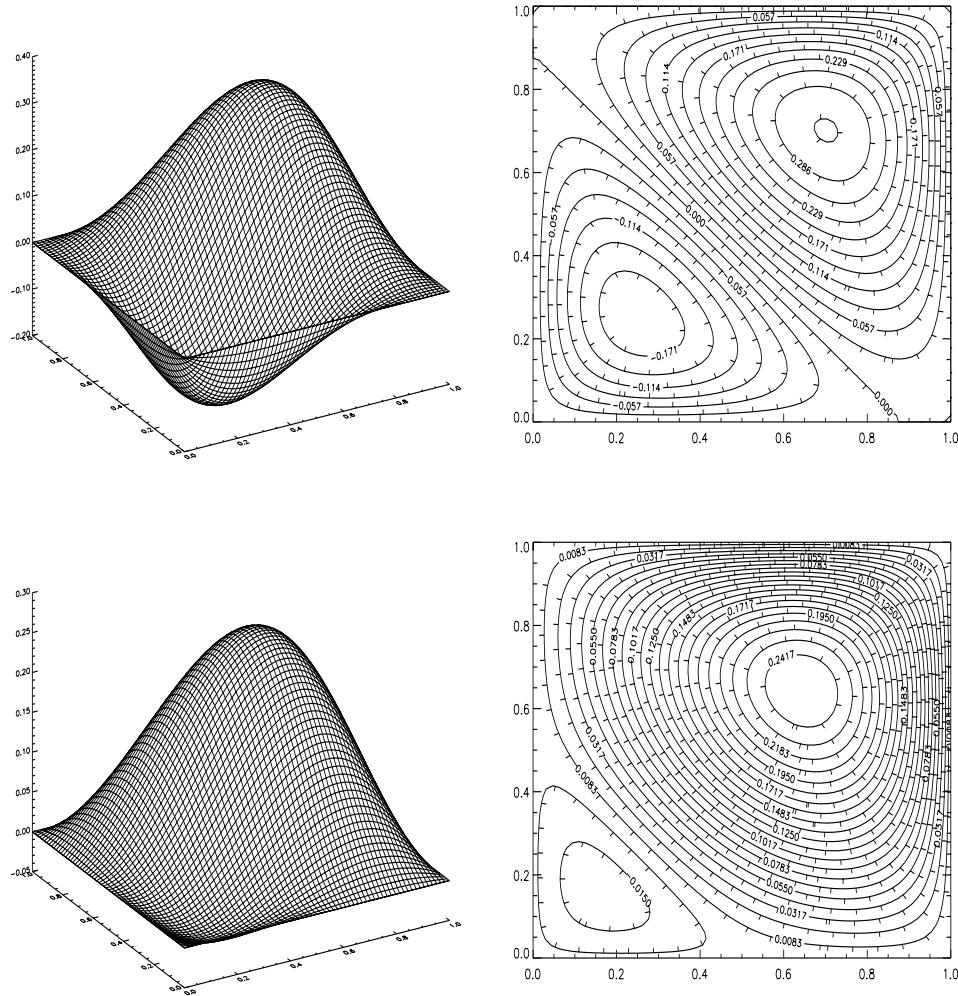


Figure 5.19. Example (5.9), finite differences, regular sparse grid spaces $V_n^{(1)}$: solution $u(\mathbf{x})$ and its contour lines for $x_3 = 0.5$ (top) and $x_3 = 0.875$ (bottom).

5.2. Flow problems and the Navier–Stokes equations

We now consider the application of an adaptive sparse grid method to the solution of incompressible flow problems. The governing relations are given by the Navier–Stokes equations

$$\begin{aligned}\partial_t \mathbf{u} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) &= \mathbf{f} - \nabla p + \nu \Delta \mathbf{u}, \quad (\mathbf{x}, t) \in \Omega \times [0, T], \\ \nabla \cdot \mathbf{u} &= 0, \\ \mathbf{u}(\mathbf{x}, 0) &= u^0(\mathbf{x}),\end{aligned}\tag{5.10}$$

here in primitive variables, with velocity $\mathbf{u} = (u, v, w)^T$ and pressure p . The parameter ν denotes the kinematic viscosity; \mathbf{f} is a prescribed volume force. The system has to be completed by proper initial and boundary conditions.

We apply the the pressure correction scheme (Chorin 1968, Temam 1969, Bell, Colella and Glaz 1989). To this end, the discretization of the Navier–Stokes equations is split into two subproblems.

Transport step. Calculate the auxiliary velocity $\hat{\mathbf{u}}^{n+1}$ by

$$\frac{\hat{\mathbf{u}}^{n+1} - \mathbf{u}^n}{k} + \mathbf{C}(\mathbf{u}^n, \mathbf{u}^{n-1}, \dots) = \mathbf{f}^{n+1} - \nabla p^n + \nu \Delta \hat{\mathbf{u}}^{n+1}.\tag{5.11}$$

Here, $\mathbf{C}(\mathbf{u}^n, \mathbf{u}^{n-1}, \dots)$ resembles a stabilized, space-adaptive discretization of $\nabla \cdot (\mathbf{u} \otimes \mathbf{u})$ by means of a Petrov–Galerkin/collocation approach using sparse grid interpolets. Adaptive refinement or coarsening takes place in every time-step steered by the size of the actual coefficients of the multiscale representation of the current solution approximation; see Griebel (1998), Koster (2002) and Griebel and Koster (2000) for further details.

Projection step. Calculate \mathbf{u}^{n+1} , p^{n+1} as the solution of

$$\begin{aligned}\frac{\mathbf{u}^{n+1} - \hat{\mathbf{u}}^{n+1}}{k} &= -\nabla(p^{n+1} - p^n), \\ \nabla \cdot \mathbf{u}^{n+1} &= 0.\end{aligned}\tag{5.12}$$

Instead of the simple Euler scheme, higher-order time discretization techniques can be applied analogously. Of course, both subproblems have to be augmented by boundary conditions – not only for \mathbf{u} and $\hat{\mathbf{u}}$, but also for the pressure p . The (in general non-physical) pressure boundary conditions especially are the subject of controversy (*cf.* Gresho and Sani (1987), Karniadakis and Sherwin (1999), Prohl (1997)), which is beyond the scope of this paper. In the following, we will assume periodic boundary conditions for \mathbf{u} and p . The saddle point problem (5.12) is treated by solving the Schur complement equation

$$\nabla \cdot \nabla(p^{n+1} - p^n) = \frac{1}{k} \nabla \cdot \hat{\mathbf{u}}^{n+1},\tag{5.13}$$

followed by the correction of the velocity

$$\mathbf{u}^{n+1} = \hat{\mathbf{u}}^{n+1} - k\nabla(p^{n+1} - p^n). \quad (5.14)$$

To this end, a weakly divergence-free adaptive discretization of $\nabla \cdot \nabla$ is applied together with sparse grid interpolants. The solver involves a transform to the lifting interpolants of Section 4.4, a diagonal scaling, and a backtransform; see Koster (2002) and Griebel and Koster (2000) for further details.

Merging of modons

Now, we apply the adaptive version (Griebel and Koster 2000, Koster 2002, Griebel and Koster 2003) of this sparse grid interpolant solver to the model problem of the interaction of three vortices in a 2D flow. Here we use the interpolants of Section 4.3 with $N = 6$.

The initial velocity is induced by three vortices, each with a Gaussian vorticity profile

$$\omega(\mathbf{x}, 0) = \omega_0 + \sum_{i=1}^3 \omega_i \exp\left(\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma_i^2}\right), \quad \mathbf{x} \in [0, 1]^2.$$

The first two vortices have the same positive sign $\omega_1, \omega_2 > 0$, and the third has a negative sign; see Figure 5.20.

The different parameters ω_0 , ω_i , and σ_i are chosen such that the mean value of $\omega(\cdot, 0)$ vanishes and that $\omega(\cdot, 0)|_{\partial[0,1]^2}$ is almost ω_0 to allow for periodic boundary conditions.

Owing to the induced velocity field, the three vortices start to rotate around each other. In a later stage, the two same-sign vortices merge, which leads to a configuration of two counter-rotating vortices. This process is the basic mechanism in 2D turbulence, and it takes place, *e.g.*, in the shear layer problem of the next subsection or during the convergence of ω to the solution of the Joyce–Montgomery equation (Chorin 1998) for random initial vorticity fields.

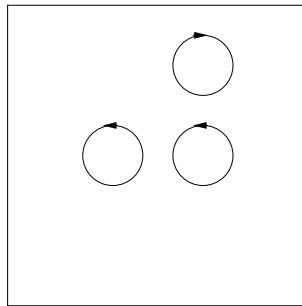


Figure 5.20. Initial configuration for the three vortices' interaction.

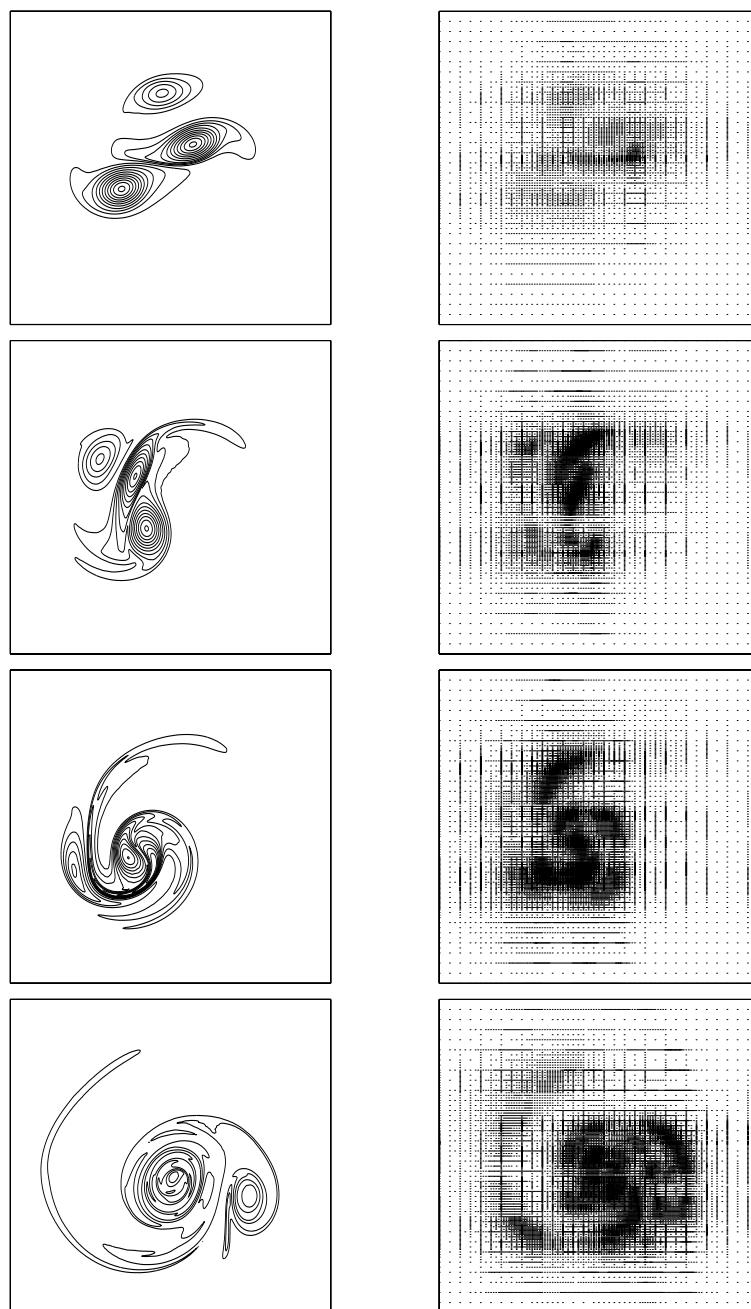


Figure 5.21. Isolines of ω at $t = 5, 10, 15, 35$ (from top to bottom). Left: adaptive wavelet solver ($\epsilon = 4 \cdot 10^{-4}$). Right: adaptive sparse grid.

In our numerical experiments, the viscosity was set to $\nu = 3.8 \cdot 10^{-6}$. The maximum velocity of the initial condition is $7 \cdot 10^{-2}$, which corresponds to a Reynolds number of ≈ 55200 . For the time discretization in the Chorin projection scheme, we used the third-order Adams–Bashforth scheme. The time-step used was $dt = 10^{-2}$. The values of the threshold parameter ϵ were set to $\{8, 4, 1\} \cdot 10^{-4}$, and the finest level 1 was limited to $(10, 10)$ to avoid very fine time-steps due to the CFL-condition.

In Figure 5.21 (page 89), the contour lines of the vorticity of the adaptive solutions are shown. The contour levels are equally spaced from -1.5 to 3 , which is approximately the minimum/maximum value of $\omega(., 0)$. Obviously, our method recognizes the arising complicated flow patterns in space and time (see Figure 5.21) and devotes many more degrees of freedom to these than to smooth regions. Note that the left-hand pictures of Figure 5.21 are enhanced by 20%. Further comparisons of these results with a conventional Fourier spectral solution showed that the results of the adaptive wavelet solver are quite accurate. But to achieve the same accuracy, only a small number of degrees of freedom was needed by our adaptive sparse grid approach – less than 1% compared with the Fourier spectral technique.

2D shear layer flow

Now we apply the adaptive wavelet solver to a shear layer model problem. The initial configuration of the temporally developing shear layer is a velocity field with a hyperbolic-tangent profile $u(y) = U \tanh(2y/\delta)$, $v = 0$, where δ is the vorticity shear layer thickness $\delta = 2U / \max(dy/du)$; see Figure 5.22. This initial condition is an exact stationary solution of the Navier–Stokes equations for the case $\nu = 0$, *i.e.*, for inviscid fluids. However, from linear stability analysis this initial configuration is known to be

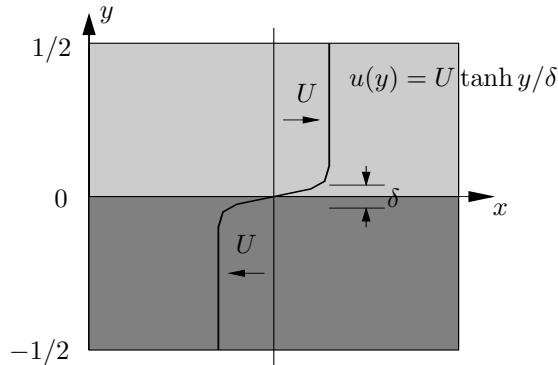


Figure 5.22. Velocity profile of the initial condition for the shear flow model problem. The thickness of the boundary layer is defined as $\delta = 2U / \max_y |\partial_y u(y)|$.

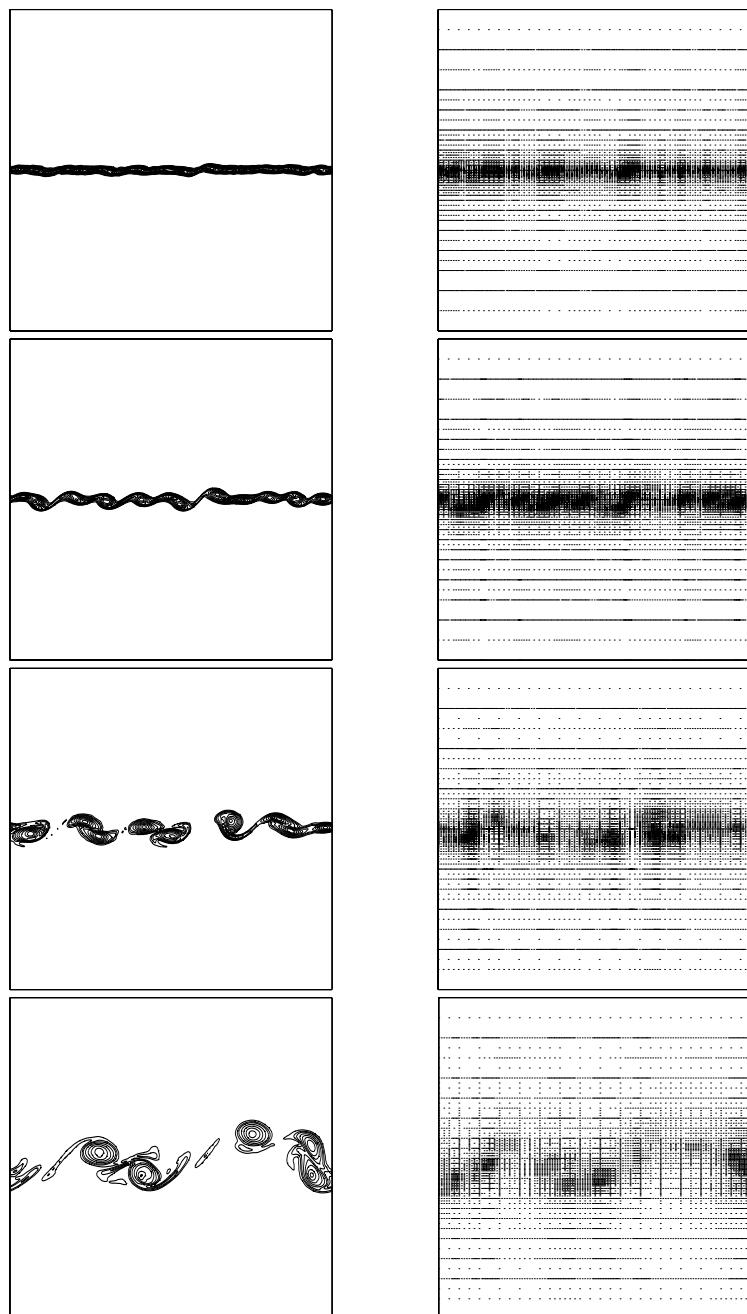


Figure 5.23. Isolines of the rotation for the values $\{0.1, 0.2, 0.3, 0.5, 0.75, 1.0, \dots, 2.25, 2.5\}$ at time $t = 4, 8, 16, 36$.
Left: isolines of computed rotation. Right: adaptive sparse grid.

inviscidly unstable. Slight perturbations are amplified by Kelvin–Helmholtz instabilities and vortex roll-up occurs (Michalke 1964).

In our numerical experiments, the vorticity thickness δ was chosen so that ten vortices should develop in the computational domain $[0, 1]^2$. The maximum velocity was $U = 1.67 \cdot 10^{-2}$, and the viscosity was $\nu = 3.8 \cdot 10^{-6}$. The instabilities were triggered by a superimposed weak white noise in the shear layer.

The numerical simulations were applied to the periodized version of the problem with two shear layers on $[0, 1] \times [0, 2]$. We used the same adaptive sparse grid solver as in the previous experiment, based on interpolets with $N = 6$, a stabilized Petrov–Galerkin collocation, and Chorin’s projection method. The values of the threshold parameter ϵ were set to $\{12, 8, 4\} \cdot 10^{-4}$ and the finest level 1 was limited to $(10, 10)$. The time-step was $5 \cdot 10^{-3}$.

On the left-hand side of Figure 5.23 (page 91), the resulting rotation $\omega = \nabla \times \mathbf{u}$ for $\epsilon = 8 \cdot 10^{-4}$ is given. The evolution of the sparse grid points is shown on the right-hand side of Figure 5.23. The initial velocity is not very smooth owing to the white noise added for triggering the instability. Therefore, a large number of degrees of freedom (DOF) is required to resolve the initial velocity in the shear layer sufficiently well. Then, in a first phase, the diffusion makes the velocity smooth very fast, which leads to the strong decay of the number of DOF. This process is stopped by the development of the Kelvin–Helmholtz instabilities leading to an increase of the number of DOF ($4 < t < 10$). In the last phase ($t > 10$), the number of coherent vortices constantly decreases by successive merging. The final state comprises two counter-rotating vortices which dissipate.

Besides the generation and the roll-up of Kelvin–Helmholtz vortices, we also see that the vortices merge with time. This process, in which a few large vortices are created from many small ones, is typical for 2D flows; *cf.* Chapter 4.6 of Chorin (1998). It comes with an energy transfer from fine to coarse structures, *i.e.*, from fine to coarse levels in the sense of an isotropic multiscale representation (Novikov 1976). Thus 2D turbulence behaves fundamentally differently to 3D turbulence. Here, predominantly an energy transfer from coarse to fine levels occurs, that is, coarse structures decay to fine structures. The maximal level was limited to 10 to avoid overly small time-steps resulting from the CFL condition.

3D shear layer flow

As seen in the previous section, vortices with the same rotation direction merge successively in 2D flows. In 3D flows, however, this effect is no longer present. Here the additional third dimension allows for vortex tubes. They can be bent, folded, and stretched over time until they break apart. Their diameter reduces during stretching, and therefore smaller-scale structures

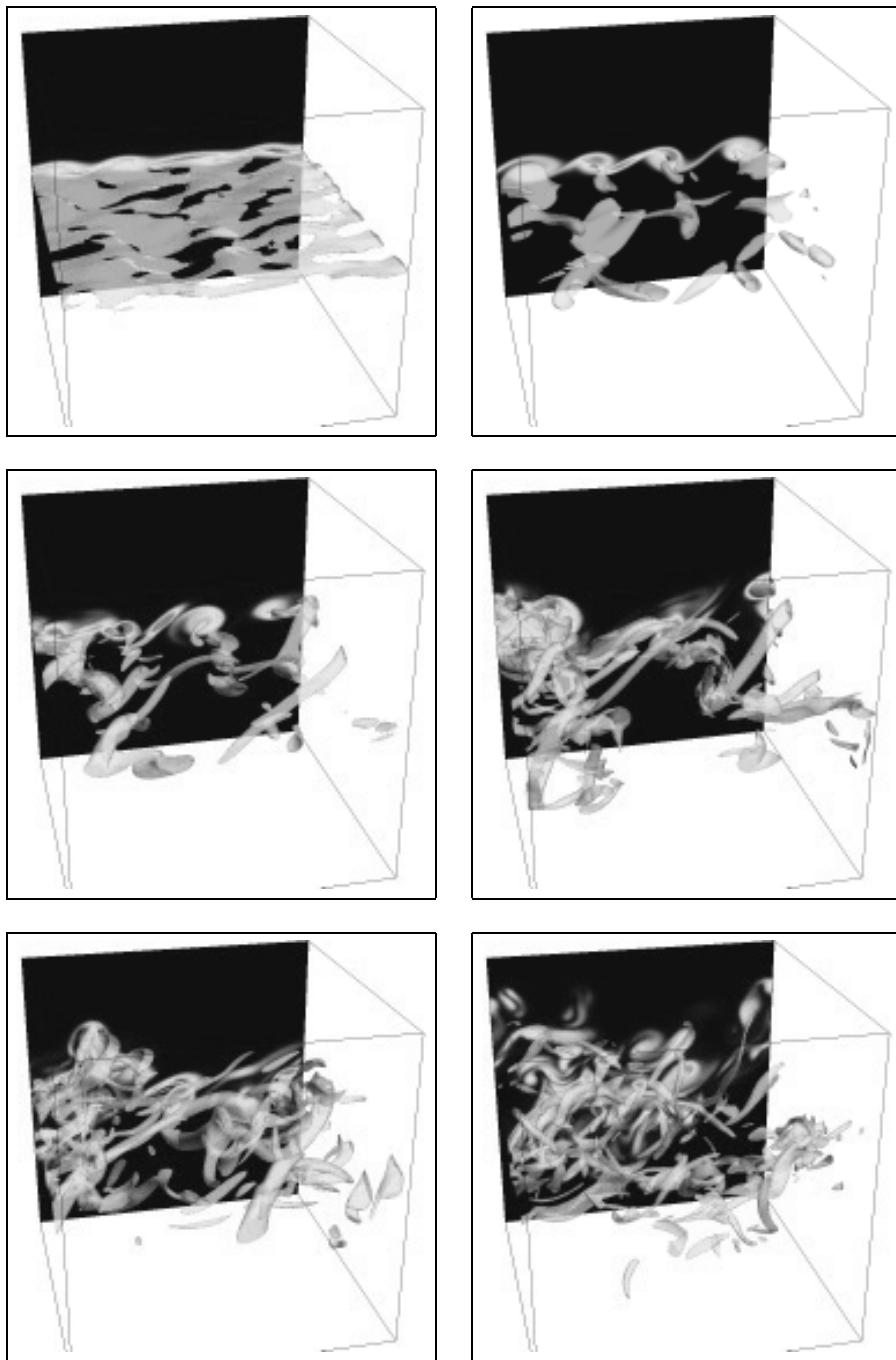


Figure 5.24. Iso-surfaces of the rotation for the 3D turbulent shear layer problem at time 0, 15, 30, 60, 90, 150.

are created. In this way energy is transported to finer scales. This energy cascade proceeds recursively until the structures are so tiny and the associated relative Reynolds number is so small that energy is annihilated by dissipation. The different phases of a developing 3D shear layer are shown in Figure 5.24 (page 93), from which this increase of complexity in the flow becomes clear. The initial conditions for this calculation are the 3D analogues of the initial conditions for the previous 2D example. However, viscosity was increased, *i.e.*, we have set $\nu = 5.04 \cdot 10^{-6}$. The remaining parameters were the same as for the 2D calculation.

5.3. Problems of high dimensionality

We now turn to higher-dimensional problems. To this end, a few general remarks are helpful. First, the sparse grid approach is limited to (topologically) quadrilateral domains due to its tensor product nature. Complicated domains must be built up by gluing together quadrilateral patches on which locally sparse grids are employed, or by appropriate mapping techniques. Now, in higher dimensionalities, this question of the shape of the domain is not as important as in the 2D and 3D case, since complicated domains typically do not appear in applications. Conceptually, besides \mathbb{R}^d itself, we use mainly hypercubes $[-a, a]^d$, $a > 0$, 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 and cause no difficulties to the sparse grid approach. Second, complexity issues are still crucial in higher-dimensional applications. In the simplest case of a sufficiently smooth function with bounded mixed derivatives, the cost and error estimates for the L_p -norms still possess $(\log N)^{d-1}$ -terms which depend exponentially on the dimensionality d . This limits the practical applications of the sparse grid method for PDEs to at most 18 dimensions at present. Furthermore, in the nonsmooth case, adaptivity towards a singularity is a difficult task in higher dimensions. In principle, it can be done, as demonstrated in the examples of the previous sections. However, an implementation of refinement strategies, error indicators, and solvers with optimal complexity is more difficult than in the low-dimensional case. Hence, utmost care has to be taken in the implementation not to obtain order constants in the work count that depend exponentially on d . For a simple example of this problem, just consider the stiffness matrix for d -linear finite elements on a uniform grid. There, the number of nonzero entries grows like 3^d for a second-order elliptic PDE with general coefficient functions. Also the work count for the computation of each entry by a conventional quadrature scheme involves terms that grow exponentially with d . Here, at least in certain simple cases like the Laplacian, an algorithmic scheme can be applied, which is based on the

so-called *unidirectional principle* (Bungartz 1998). It allows us to treat the single dimensions separately and results in linear complexity in both the degrees of freedom *and d*, for the matrix–vector product in a Galerkin sparse grid approach. However, note that, especially in the case of more general nonseparable operators, this is the challenging part of a sparse grid finite element implementation.

The next question concerns the sources of high-dimensional problems and PDEs. Here, besides pure integration problems stemming from physics and finance, typically models from the stochastic and data analysis world show up. For example, high-dimensional Laplace/diffusion problems and high-dimensional convection–diffusion problems result from diffusion approximation techniques and the Fokker–Planck equation. Examples are the description of queueing networks (Mitzlaff 1997, Shen, Chen, Dai and Dai 2002), 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), or various models for the pricing of financial derivatives (Reisinger 2003). Furthermore, homogenization with multiple scales (Allaire 1992, Cioranescu, Damlamian and Griso 2002, Matache 2001, Matache and Schwab 2001, Hoang and Schwab 2003) as well as stochastic elliptic equations (Schwab and Todor 2002, 2003) result in high-dimensional PDEs. Next, we find quite high-dimensional problems in quantum mechanics. Here, the dimensionality of the Schrödinger equation grows with the number of considered atoms. Sparse grids have been used in this context by Garcke (1998), Garcke and Griebel (2000), Yserentant (2004) and Hackbusch (2001).

In the following, we illustrate the potential of sparse grids for problems of a higher dimensionality in the context of numerical quadrature and data mining.

Classification and regression in data mining

Data mining is the process to find hidden patterns, relations, and trends in large data sets. It plays an increasing role in commerce and science. Typical scientific applications are the post-processing of data in medicine (*e.g.*, CT data), the evaluation of data in astrophysics (*e.g.*, telescope and observatory data), and the grouping of seismic data, or the evaluation of satellite pictures (*e.g.*, NASA earth observing system). Financial and commercial applications are perhaps of greater importance. With the development of the internet and e-commerce, there are huge data sets collected, more or less automatically, which can be used for business decisions and further strategic planning. Here, applications range from contract management to risk assessment, from the segmentation of customers for marketing to fraud detection, stock analysis and turnover prediction.

Usually, the process of data mining (or knowledge discovery) can be separated into the planning step, the preparation phase, the mining phase (*i.e.*, machine learning) and the evaluation. To this end, association-analysis classification, clustering, and prognostics are to be performed. For a thorough overview of the various tasks arising in the data mining process, see Berry and Linoff (2000) and Cios, Pedrycz and Swiniarski (1998).

In the following, we consider the classification problem in detail. Here, a set of data points in d -dimensional feature space is given together with a class label in $\{-1, 1\}$, for example. From these data, a classifier must be constructed which allows us to predict the class of any newly given data point for future decision making. Widely used approaches are nearest neighbour methods, decision tree induction, rule learning, and memory-based reasoning. There are also classification algorithms based on adaptive multivariate regression splines, neural networks, support vector machines, and regularization networks. Interestingly, these latter techniques can be interpreted in the framework of regularization networks (Girosi, Jones and Poggio 1995). This approach allows a direct description of the most important neural networks, and it also allows for an equivalent description of support vector machines and n -term approximation schemes (Girosi 1998).

We follow Garcke, Griebel and Thess (2001) and consider a given set of already classified data (the training set):

$$S = \{(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}\}_{i=1}^M.$$

We now assume that these data have been obtained by sampling of an unknown function f which belongs to some function space V defined over \mathbb{R}^d . The sampling process was disturbed by noise. The aim is now to recover the function f from the given data as well as possible. This is clearly an ill-posed problem since infinitely many solutions are possible. To get a well-posed, uniquely solvable problem, we have to restrict f . To this end, regularization theory (Tikhonov and Arsenin 1977, Wahba 1990) imposes an additional smoothness constraint on the solution of the approximation problem, and the regularization network approach considers the variational problem

$$\min_{f \in V} R(f)$$

with

$$R(f) = \frac{1}{M} \sum_{i=1}^m C(f(\mathbf{x}_i), y_i) + \lambda \Phi(f). \quad (5.15)$$

Here, $C(\cdot, \cdot)$ denotes an error cost function which measures the interpolation error and $\Phi(f)$ is a smoothness functional which must be well defined for $f \in V$. The first term enforces closeness of f to the data, the second term enforces smoothness of f , and the regularization parameter λ balances these

two terms. We consider the case

$$C(x, y) = (x - y)^2 \quad \text{and} \quad \Phi(f) = \|Pf\|_2^2 \quad \text{with} \quad Pf = \nabla f.$$

The value of λ can be chosen according to cross-validation techniques (Allen 1972, Golub, Heath and Wahba 1979, Utreras 1979, Wahba 1985) or to some other principle, such as structural risk minimization (Vapnik 1982). We find exactly this type of formulation in the case $d = 2, 3$ in many scattered data approximation methods (see Arge, Dæhlen and Tveito (1995) and Hoschek and Lasser (1992)), where the regularization term is usually physically motivated.

We now restrict the problem to a finite-dimensional subspace $V_N \subset V$. The function f is then replaced by

$$f_N = \sum_{j=1}^N \alpha_j \varphi_j(\mathbf{x}). \quad (5.16)$$

Here $\{\psi_j\}_{j=1}^N$ should span V_N and preferably should form a basis for V_N . The coefficients $\{\alpha_j\}_{j=1}^N$ denote the degrees of freedom. Note that the restriction to a suitably chosen finite-dimensional subspace involves some additional regularization (regularization by discretization) which depends on the choice of V_N . In this way we obtain from the minimization problem a feasible linear system. We thus have to minimize

$$R(f_N) = \frac{1}{M} \sum_{i=1}^M (f_N(\mathbf{x}_i) - y_i)^2 + \lambda \|Pf_N\|_{L_2}^2, \quad f_N \in V_N \quad (5.17)$$

in the finite-dimensional space V_N . We plug (5.16) into (5.17) and obtain

$$R(f_N) = \frac{1}{M} \sum_{i=1}^M \left(\sum_{j=1}^N \alpha_j \varphi_j(\mathbf{x}_i) - y_i \right)^2 + \lambda \|P \sum_{j=1}^N \alpha_j \varphi_j\|_{L_2}^2 \quad (5.18)$$

$$= \frac{1}{M} \sum_{i=1}^M \left(\sum_{j=1}^N \alpha_j \varphi_j(\mathbf{x}_i) - y_i \right)^2 + \lambda \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j (P\varphi_i, P\varphi_j)_{L_2}. \quad (5.19)$$

Differentiation with respect to α_k , $k = 1, \dots, N$, gives

$$0 = \frac{\partial R(f_N)}{\partial \alpha_k} = \frac{2}{M} \sum_{i=1}^M \left(\sum_{j=1}^N \alpha_j \varphi_j(\mathbf{x}_i) - y_i \right) \cdot \varphi_k(\mathbf{x}_i) + 2\lambda \sum_{j=1}^N \alpha_j (P\varphi_j, P\varphi_k)_{L_2}. \quad (5.20)$$

This is equivalent to ($k = 1, \dots, N$)

$$\lambda \sum_{j=1}^N \alpha_j (P\varphi_j, P\varphi_k)_{L_2} + \frac{1}{M} \sum_{j=1}^N \alpha_j \sum_{i=1}^M \varphi_j(\mathbf{x}_i) \cdot \varphi_k(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^M y_i \varphi_k(\mathbf{x}_i) \quad (5.21)$$

and we obtain ($k = 1, \dots, N$)

$$\sum_{j=1}^N \alpha_j \left[M \lambda (P\varphi_j, P\varphi_k)_{L_2} + \sum_{i=1}^M \varphi_j(\mathbf{x}_i) \cdot \varphi_k(\mathbf{x}_i) \right] = \sum_{i=1}^M y_i \varphi_k(\mathbf{x}_i). \quad (5.22)$$

In matrix notation, we end up with the linear system

$$(\lambda C + B \cdot B^T) \alpha = B y. \quad (5.23)$$

Here, C is a square $N \times N$ matrix with entries $C_{j,k} = M \cdot (P\varphi_j, P\varphi_k)_{L_2}$, $j, k = 1, \dots, N$, and B is a rectangular $N \times M$ matrix with entries $B_{j,i} = \varphi_j(\mathbf{x}_i)$, $i = 1, \dots, M$, $j = 1, \dots, N$. The vector y contains the data y_i and has length M . The unknown vector α contains the degrees of freedom α_j and has length N . With the gradient $P = \nabla$ in the regularization expression in (5.15), we obtain a Poisson problem with an additional term that resembles the interpolation problem. The natural boundary conditions for such a differential equation in $\Omega = [0, 1]^d$, for instance, are Neumann conditions. The discretization (5.16) gives us then the linear system (5.23) where C corresponds to a discrete Laplacian. To obtain the classifier f_N , we now have to solve this system.

Again, the curse of dimensionality prohibits us from using for V_N conventional finite element spaces living on a uniform grid. The complexity would grow exponentially with d . Instead, we used a sparse grid approach, namely the combination method, as already described for the 2D and 3D case in (4.41), (4.42), and Figure 4.14. These formulae can easily be generalized to the d -dimensional case. Here we consider the minimization problem on a sequence of grids, solve the resulting linear systems (5.23), and combine the solution accordingly. The complexity of the method is with respect to the number of levels n , as usual for regular sparse grids of the order $O(n^{d-1}2^n)$. With respect to the number M of training data, the cost complexity scales linearly in M for a clever implementation. This is a substantial advantage in comparison to most neural networks and support vector machines, which generally scale at least quadratically in M and are therefore not suited to problems with very large data sets.

We now apply our approach to different test data sets. Here we use both synthetic data generated by DatGen (Melli 2003) and real data from practical data mining applications. All the data sets are rescaled to $[0, 1]^d$.

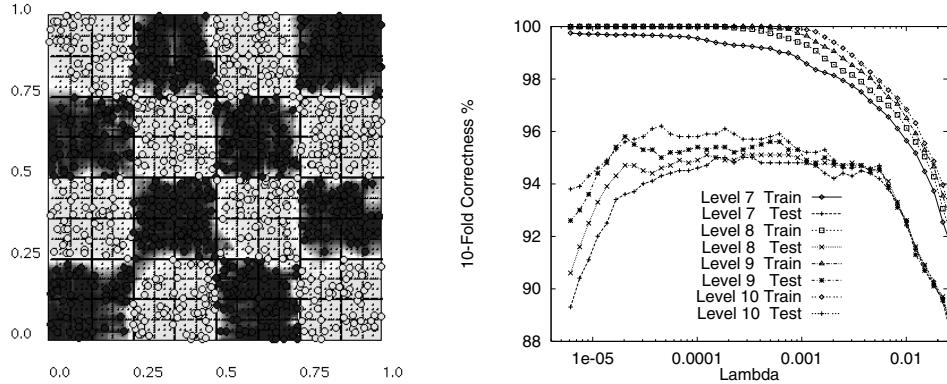


Figure 5.25. Left: chessboard data set, combination technique with level 10, $\lambda = 4.53999 \cdot 10^{-5}$. Right: plot of the dependence on λ (in logscale) and level.

To evaluate our method we give the correctness rates on testing data sets if available and the 10-fold cross-validation results on training and testing data sets. For the 10-fold cross-validation we proceed as follows. We divide the training data into 10 equally sized disjoint subsets. For $i = 1$ to 10, we pick the i th of these subsets as a further testing set and build the sparse grid combination classifier with the data from the remaining nine subsets. We then evaluate the correctness rates of the current training and testing set. In this way we obtain ten different training and testing correctness rates. The 10-fold cross-validation result is just the average of these ten correctness rates. For further details, see Stone (1974). For a critical discussion on the evaluation of the quality of classifier algorithms, see Salzberg (1997).

We first consider 2D problems with small sets of data that correspond to certain structures. Then we treat problems with huge sets of synthetic data with up to 5 million points.

The first example is taken from Ho and Kleinberg (1996) and Kaufman (1999). Here, 1000 training data points were given which are more or less uniformly distributed in $\Omega = [0, 1]^2$. The associated data values are plus one or minus one depending on their location in Ω such that a 4×4 chessboard structure appears: see Figure 5.25 (left). We computed the 10-fold cross-validated training and testing correctness with the sparse grid combination method for different values of the regularization parameter λ and different levels n . The results are shown in Figure 5.25 (right).

We see that the 10-fold testing correctness is well around 95% for values of λ between $3 \cdot 10^{-5}$ and $5 \cdot 10^{-3}$. Our best 10-fold testing correctness was 96.20% on level 10 with $\lambda = 4.54 \cdot 10^{-5}$. The chessboard structure is thus reconstructed with less than 4% error.

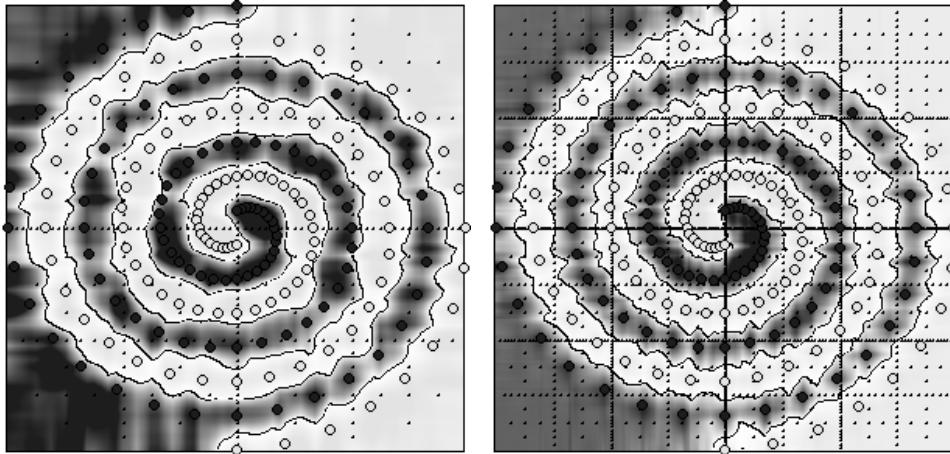


Figure 5.26. Spiral data set, sparse grid with level 6 (left) and 8 (right), $\lambda = 0.001$.

Another 2D example with structure is the spiral data set, first proposed by Wieland; see also Fahlmann and Lebiere (1990). Here, 194 data points describe two intertwined spirals: see Figure 5.26. This is surely an artificial problem, which does not appear in practical applications. However it serves as a hard test case for new data mining algorithms. It is known that neural networks can have severe problems with this data set, and some neural networks can not separate the two spirals at all. In Figure 5.26 we give the results obtained with our sparse grid combination method with $\lambda = 0.001$ for $n = 6$ and $n = 8$. Already for level 6, the two spirals are clearly detected and resolved. Note that here only 577 grid points are contained in the sparse grid. For level 8 (2817 sparse grid points), the shape of the two reconstructed spirals gets smoother and the reconstruction gets more precise.

The BUPA Liver Disorders data set from the Irvine Machine Learning Database Repository (Blake and Merz 1998) consists of 345 data points with six features plus a selector field used to split the data into two sets with 145 instances and 200 instances, respectively. Here we only have training data and can therefore only report our 10-fold cross-validation results. No comparison with unused test data is possible.

We compare with the two best results from Lee and Mangasarian (2001), the smoothed support vector machine (SSVM) introduced therein, and the feature selection concave minimization (FSV) algorithm due to Bradley and Mangasarian (1998). Table 5.1 gives the results for the 10-fold correctness.

Our sparse grid combination approach performs on level 3 with $\lambda = 0.0625$ at 69.23% 10-fold testing correctness. But our other results were also in this range. Our method performs only slightly worse here than the SSVM but

Table 5.1. Results for the BUPA liver disorders data set.

	SSVM	FSV	sparse grid combination method			
			level 1 $\lambda = 0.0001$	level 2 $\lambda = 0.1$	level 3 $\lambda = 0.0625$	level 4 $\lambda = 0.625$
train. %	70.37	68.18	83.28	79.54	90.20	88.66
test. %	70.33	65.20	66.89	66.38	69.23	68.74

Table 5.2. Results for a 6D synthetic massive data set, $\lambda = 0.01$.

	# of points	training corr.	testing corr.	total time (sec)	data matrix time (sec)	# of iterat.
level 1	50 000	87.9	88.1	158	152	41
	500 000	88.0	88.1	1570	1528	44
	5 million	88.0	88.0	15933	15514	46
level 2	50 000	89.2	89.3	1155	1126	438
	500 000	89.4	89.2	11219	11022	466
	5 million	89.3	89.2	112656	110772	490

clearly better than FSV. Note that the results for the robust linear program (RLP) algorithm (Bennett and Mangasarian 1992), the support vector machine using the 1-norm approach ($\text{SVM}_{\|\cdot\|_1}$), and the classical support vector machine ($\text{SVM}_{\|\cdot\|_2}$) (Bradley and Mangasarian 1998, Cherkassky and Mulier 1998, Vapnik 1995) were reported to be somewhat worse in Lee and Mangasarian (2001).

Next, we produced a 6D data set with 5 million training points and 20 000 points with DatGen (Melli 2003) for testing. We used the call

```
datgen -r1 -X0/100,R,0 :0/100,R,0:0/100,R,0:0/100,R,0:0/200,
R,0:0/200,R,0 -R2 -C2/4 -D2/5 -T10/60 -0502 0000 -p -e0.15.
```

The results are given in Table 5.2. On level one, a testing correctness of 88% was achieved already, which is quite satisfying for this data. We see that really huge data sets of 5 million points could be handled. We also give the CPU time which is needed for the computation of the matrices $G_1 = B_1 \cdot B_1^T$. Here, more than 96% of the computing time is spent on the matrix assembly. Again, the execution times scale linearly with the number of data points.

Analogously, we can at present deal with up to 18-dimensional problems before storage limitations on our available computers stop us. This moderate number of dimensions, however, is still sufficient for many practical applications. In very high-dimensional data, there exist mostly strong correlations and dependencies between the dimensions. Then, in a preprocessing step, the effective dimensions can be determined, for instance, by means of principal component analysis, and the dimension can be reduced in many cases to 18 or less.

To reach 18 dimensions, a generalization of the sparse grid combination technique to simplicial basis functions (Garcke and Griebel 2001*a*, 2002) is needed. Note that the sparse grid combination technique can be parallelized in a straightforward way (Garcke and Griebel 2001*b*, Garcke, Hegland and Nielsen 2003). Finally, the sparse grid technique can be used in a dimension-adaptive fashion (Hegland 2002, 2003), which further enhances the method's capabilities. This approach will be discussed for the example of numerical integration now.

Integration

The computation of high-dimensional integrals is a central part of computer simulations in many application areas such as statistical mechanics, financial mathematics, and computational physics. Here, the arising integrals usually cannot be solved analytically, and thus numerical approaches are required. Furthermore, often a high-accuracy solution is needed, and thus such problems can be computationally quite challenging even for parallel supercomputers. Conventional algorithms for the numerical computation of such integrals are usually limited by the curse of dimensionality. However, for special function classes, such as spaces of functions which have bounded mixed derivatives, Smolyak's construction (Smolyak 1963) (see (4.1)) can overcome this curse to a certain extent. In this approach, multivariate quadrature formulas are constructed using combinations of tensor products of appropriate 1D formulas. In this way, the number of function evaluations and the numerical accuracy become independent of the dimension of the problem up to logarithmic factors. Smolyak's construction is simply our sparse grid approach. It has been applied to numerical integration by several authors, using the midpoint rule (Baszenski and Delves 1993), the rectangle rule (Paskov 1993), the trapezoidal rule (Bonk 1994*a*), the Clenshaw–Curtis rule (Cools and Maerten 1997, Novak and Ritter 1998), the Gauss rules (Novak and Ritter 1997), and the Gauss–Patterson rules (Gerstner and Griebel 1998, Petras 2000) as the 1D basis integration procedure. The latter approach, in particular, achieves the highest polynomial exactness of all nested quadrature formulas and shows very good results for sufficiently smooth multivariate integrands. Further studies have been

made concerning extrapolation methods (Bonk 1994a), discrepancy measures (Frank and Heinrich 1996), and complexity questions (Wasilkowski and Woźniakowski 1999).

There is also a large variety of other methods for the numerical integration of multivariate functions such as Monte Carlo and quasi-Monte Carlo methods (Niederreiter 1992), lattice rules (Sloan and Joe 1994), adaptive subdivision methods (Genz and Malik 1980, Dooren and Ridder 1976), and approximation methods based on neural networks (Barron 1994, Mhaskar 1996). Each of these methods is particularly suitable for functions from a certain function class and has a complexity which is then also independent or nearly independent of the problem's dimensionality.

Despite the large improvements of the quasi-Monte Carlo and sparse grid methods over the Monte Carlo method, their convergence rates will suffer more and more with rising dimension owing to their respective dependence on the dimension in the logarithmic terms. Therefore, one aim of recent numerical approaches has been to reduce the dimension of the integration problem without affecting the accuracy unduly.

In some applications, the different dimensions of the integration problem are not equally important. For example, in path integrals the number of dimensions corresponds to the number of time-steps in the time discretization. Typically, the first steps in the discretization are more important than the last steps since they determine the outcome more substantially. In other applications, although the dimensions seem to be of the same importance at first sight, the problem can be transformed into an equivalent one where the dimensions are not. Examples are the Brownian bridge discretization or the Karhunen–Loeve decomposition of stochastic processes.

Intuitively, problems where the different dimensions are not of equal importance might be easier to solve: numerical methods could concentrate on the more important dimensions. Interestingly, complexity theory also reveals that integration problems with weighted dimensions can become tractable even if the unweighted problem is not (Wasilkowski and Woźniakowski 1999). Unfortunately, classical adaptive numerical integration algorithms (Genz and Malik 1980, Dooren and Ridder 1976) cannot be applied to high-dimensional problems, since the work overhead in order to find and adaptively refine in important dimensions would be too large.

To this end, a variety of algorithms have been developed that try to find and quantify important dimensions. Often, the starting point of these algorithms is Kolmogorov's superposition theorem: see Kolmogorov (1956, 1957). Here, a high-dimensional function is approximated by sums of lower-dimensional functions. A survey of this approach from the perspective of approximation theory is given in Khavinson (1997). Further results can be found in Rassias and Simsa (1995) and Simsa (1992). Analogous ideas are followed in statistics for regression problems and density estimation.

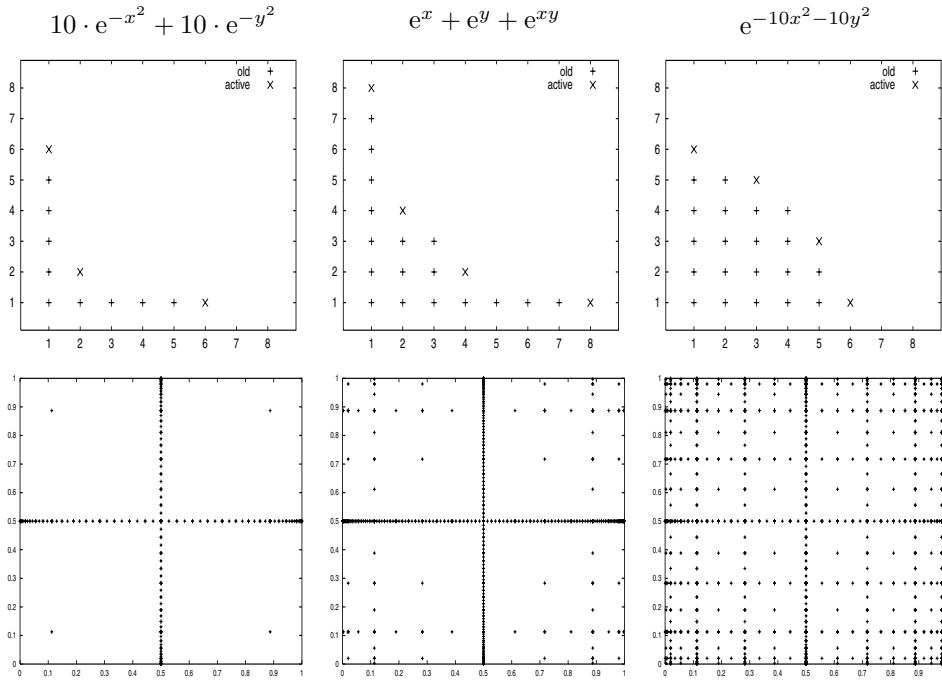


Figure 5.27. The resulting index sets and corresponding sparse grids for $\text{TOL} = 10^{-15}$ for some isotropic test functions.

Here, examples are so-called additive models (Hastie and Tibshirani 1990), multivariate adaptive regression splines (MARS) (Friedman 1991), and the ANOVA decomposition (Wahba 1990, Yue and Hickernell 2002); see also Hegland and Pestov (1999). Other interesting techniques for dimension reduction are presented in He (2001). If the importance of the dimensions is known *a priori*, techniques such as importance sampling can be applied in Monte Carlo methods (Kalos and Whitlock 1986). For the quasi-Monte Carlo method, a sorting of the dimensions according to their importance leads to a better convergence rate (yielding a reduction of the effective dimension). The reason for this is the better distributional behaviour of low-discrepancy sequences in lower dimensions than in higher ones (Caflisch, Morokoff and Owen 1997). The sparse grid method, however, *a priori* treats all dimensions equally and thus gains no immediate advantage for problems where dimensions are of different importance.

In Gerstner and Griebel (2003), we developed a generalization of the conventional sparse grid approach which is able to adaptively assess the dimensions according to their importance, and thus reduces the dependence of the computational complexity on the dimension. This is quite in the spirit of Hegland (2002, 2003). The dimension-adaptive algorithm tries to

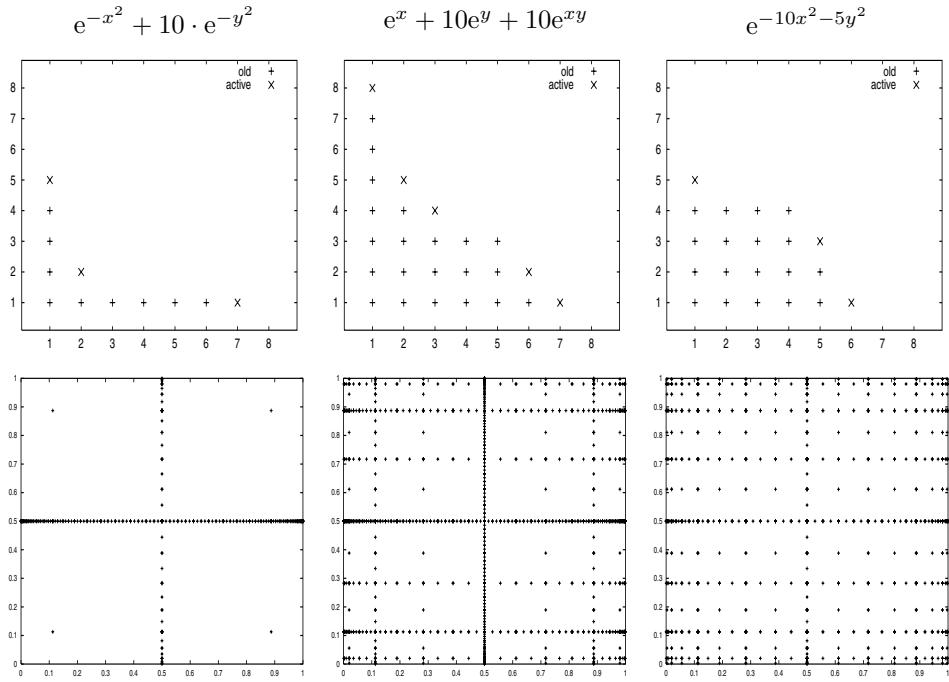


Figure 5.28. The resulting index sets and corresponding sparse grids for $\text{TOL} = 10^{-15}$ for some anisotropic test functions.

find important dimensions automatically and adapts (places more integration points) in those dimensions. To achieve this efficiently, a data structure for a fast bookkeeping and searching of generalized sparse grid index sets is necessary.

We will now show the performance of the dimension-adaptive algorithm in numerical examples. First, we consider some 2D test functions, which allows us to show the resulting grids and level index sets. In these cases, the exact value of the integral is known (or can be computed quickly). The second example is a path integral of 32 dimensions in which the integral value is also known beforehand. The third example is a 256-dimensional application problem from finance where the exact value is unknown. We use the well-known Patterson formulas (Patterson 1986) for univariate quadrature in all examples. These were shown to be a good choice for the sparse grid construction by Gerstner and Griebel (1998).

Let us first consider simple combinations of exponential functions defined over $[0, 1]^2$. In Figures 5.27 and 5.28 we depict the level index sets used by the algorithm as well as the resulting dimension-adapted sparse grids for some isotropic and some anisotropic functions, respectively. In these examples, the selected error threshold is $\text{TOL} = 10^{-15}$.

The first example is a sum of 1D functions. The dimension-adaptive algorithm correctly selects no indices in joint dimensions. Also, more points are placed in the x -direction than in the y -direction in the anisotropic case. Clearly, here the conventional sparse grid would use too many points in joint directions.

The second example is not separable, nor does it have product structure. The resulting level index set is almost triangular, like the conventional sparse grid. However, the dimension-adaptive algorithm chooses to select more points on the axes, while the conventional sparse grid would have used too many points in the interior. In our experience, many application problems fall in this category, which we would call *nearly additive*.

The third example is the well-known Gaussian hat function, and has product structure. In this example, many points in joint dimensions are required, and the conventional sparse grid would have placed too few points there. At first sight, this is a surprising result, since product functions should be more easily integrable by a tensor product approach. However, the mixed derivatives of the Gaussian can become large even if they are bounded, which reduces the efficiency of both the conventional sparse grid and the dimension-adaptive approaches.

Let us now approach some higher-dimensional problems. We will first consider an initial value problem given by the linear PDE

$$\frac{\partial u}{\partial t} = \frac{1}{2} \cdot \frac{\partial^2 u}{\partial x^2}(x, t) + v(x, t) \cdot u(x, t),$$

with initial condition $u(x, 0) = f(x)$. The solution of this problem can be obtained with the Feynman–Kac formula as

$$u(x, t) = E_{x,0} \left(f(\xi(t)) \cdot e^{\int_0^t v(\xi(r), t-r) dr} \right),$$

where ξ represents a Wiener path starting at $\xi(0) = x$. The expectation $E_{x,0}$ can be approximated by a discretization of time using a finite number of time-steps $t_i = i \cdot \Delta t$ with $\Delta t = t/d$. The integral in the exponent is approximated by a 1D quadrature formula such as a sufficiently accurate trapezoidal rule.

The most natural way to discretize the Wiener path is by a random walk, *i.e.*, by the recursive formula

$$\xi_k = \xi_{k-1} + \sqrt{\Delta t} z_k,$$

where $\xi_0 = x$ and z_k are normally distributed random variables with mean zero and variance one. The dimensions in the random walk discretization are all of the same importance since all the variances are identical to Δt .

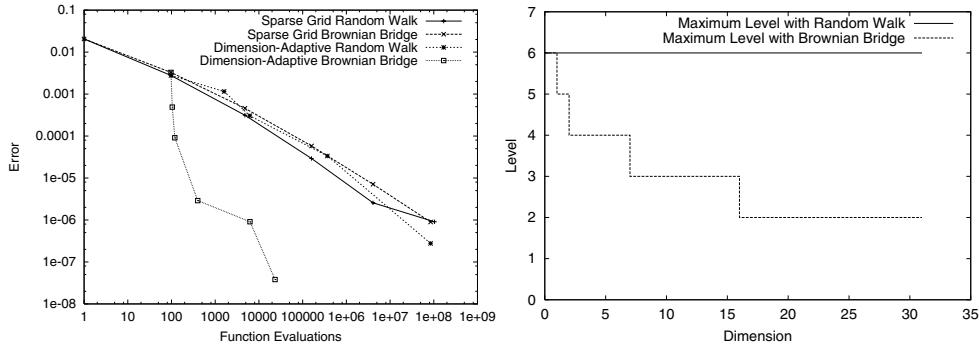


Figure 5.29. Computational results for the path integral ($d = 32$): integration error vs. number of function evaluations (left) and maximum level over all dimensions (sorted) for the dimension-adaptive algorithm with Brownian bridge discretization (right).

In the Brownian bridge discretization (Caflisch *et al.* 1997), however, the path is discretized using a future and a past value

$$\xi_k = \frac{1}{2}(\xi_{k-h} + \xi_{k+h}) + \sqrt{\frac{h \cdot \Delta t}{2}} \cdot z_k.$$

Starting with $\xi_0 := x$ and $\xi_d := x + \sqrt{t} z_d$, the subsequent values to be computed are $\xi_{d/2}, \xi_{d/4}, \xi_{3d/4}, \xi_{d/8}, \xi_{3d/8}, \xi_{5d/8}, \xi_{7d/8}, \xi_{1d/16}, \xi_{3d/16}, \dots$ with corresponding $h = 1/2, 1/4, 1/4, 1/4, 1/8, 1/8, 1/8, 1/8, 1/16, 1/16, \dots$. The Brownian bridge leads to a concentration of the total variance in the first few steps of the discretization and thus to a weighting of the dimensions.

Let us now consider the concrete example (Morokoff and Caflisch 1995)

$$v(x, t) = \left(\frac{1}{t+1} + \frac{1}{x^2+1} - \frac{4x^2}{(x^2+1)^2} \right),$$

with initial condition $u(x, 0) = \frac{1}{x^2+1}$. The exact solution is then

$$u(x, t) = \frac{t+1}{x^2+1}.$$

The results for $d = 32$, $t = 0.02$ and $x = 0$ are shown on the left-hand side of Figure 5.29. We see the integration error plotted against the number of function evaluations in a log-log scale. Here, the conventional sparse grid method is compared with the dimension-adaptive algorithm for the random walk and Brownian bridge discretizations. In this example, the conventional sparse grid is for the random walk discretization obviously close to the optimum, since the dimension-adaptive method cannot improve on the performance. The conventional sparse grid gains no advantage from the Brownian bridge discretization, but the convergence rate of the dimension-adaptive algorithm

is dramatically improved. Note that the convergence rate of the quasi-Monte Carlo method (with Brownian bridge) is comparable to that of the conventional sparse grid approach (Morokoff and Caflisch 1995, Gerstner and Griebel 1998). On the right-hand side of Figure 5.29, we plot the maximum level per dimension of the final level index set of the dimension-adaptive method with and without the Brownian bridge discretization. Here the dimensions are sorted according to this quantity. For the Brownian bridge discretization, the maximum level decays with the dimension. This shows that only a few dimensions are important and thus contribute substantially to the total integral while the other dimensions add significantly less.

Let us now consider a typical collateralized mortgage obligation problem, which involves several tranches, which in turn derive their cash flows from an underlying pool of mortgages (Caflisch *et al.* 1997, Paskov and Traub 1995). The problem is to estimate the expected value of the sum of present values of future cash flows for each tranche. Let us assume that the pool of mortgages has a $21\frac{1}{3}$ year maturity and cash flows are obtained monthly. Then the expected value requires the evaluation of an integral of dimension $d = 256$ for each tranche,

$$\int_{\mathbb{R}^d} v(\xi_1, \dots, \xi_d) \cdot g(\xi_1) \cdot \dots \cdot g(\xi_d) d\xi_1 \dots d\xi_d,$$

with Gaussian weights $g(\xi_i) = (2\pi\sigma^2)^{-1/2} e^{-\xi_i^2/2\sigma^2}$. The sum of the future cash flows v is basically a function of the interest rates i_k (for month k),

$$i_k := K_0 e^{\xi_1 + \dots + \xi_k} i_0$$

with a certain normalizing constant K_0 and an initial interest rate i_0 (for details see the first example in Caflisch *et al.* (1997) and compare Gerstner and Griebel (1998) and Paskov and Traub (1995)). Again the interest rates can be discretized using either a random walk or the Brownian bridge construction. For the numerical computation, the integral over \mathbb{R}^d is transformed into an unweighted integral on $[0, 1]^d$ with the help of the inverse normal distribution.

In Figure 5.30, we again compare the conventional sparse grid method with the dimension-adaptive method for the random walk and the Brownian bridge discretization. The error is computed against an independent quasi-Monte Carlo calculation. Note also that in this example the convergence rate of the conventional sparse grid approach is comparable to the quasi-Monte Carlo method (Gerstner and Griebel 1998).

We see again that a weighting of the dimensions does not influence the convergence of the conventional sparse grid method. But for the dimension-adaptive method the amount of work is again substantially reduced (by several orders of magnitude) for the same accuracy when the Brownian bridge discretization is used, and thus higher accuracies can be obtained. In this

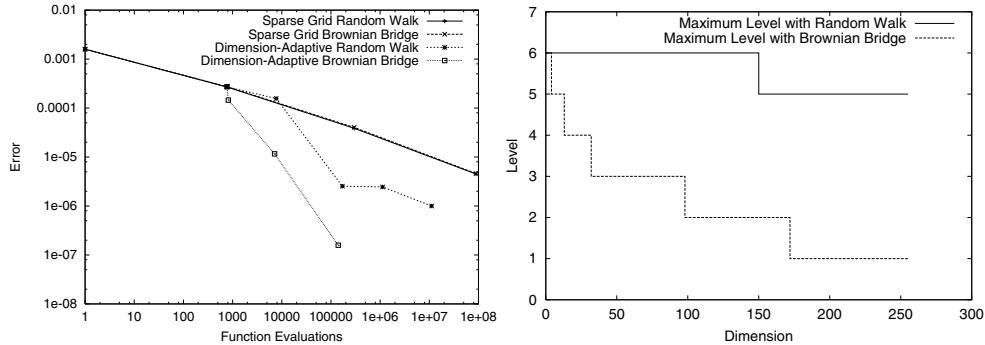


Figure 5.30. Computational results for the CMO problem ($d = 256$): integration error vs. number of function evaluations (left) and maximum level over all dimensions (sorted) for the dimension-adaptive algorithm with and without Brownian bridge discretization (right).

example the dimension-adaptive method also gives better results than the conventional sparse grid method for the random walk discretization. This implies that the conventional sparse grid uses too many points in mixed dimensions for this problem. The problem seems to be intrinsically lower-dimensional and nearly additive (Caflisch *et al.* 1997).

At present, we are working to carry this dimension-adaptive approach over to PDEs.

6. Concluding remarks

In this contribution we have given an overview of the basic principles and properties of sparse grids as well as a report on the state of the art concerning sparse grid applications. Starting from the dominant motivation – breaking the curse of dimensionality – we discussed the underlying tensor product approach, based upon different 1D multiscale bases such as the classical piecewise linear hierarchical basis, general hierarchical polynomial bases, interpolets, or wavelets. We then presented various resulting sparse grid constructions and discussed their properties with respect to computational complexity, discretization error, and smoothness requirements. The approach can be extended to nonsmooth solutions by adaptive refinement methods. We demonstrated the effectiveness of sparse grids in a series of applications. The presented numerical results include 2D and 3D PDE model problems, flow problems, and even two non-PDE applications in higher dimensions, namely numerical quadrature and data mining.

Since their introduction slightly more than a decade ago, sparse grids have seen a very successful development and a variety of different applications.

Especially for higher-dimensional scenarios, we are convinced that sparse grids, together with dimension adaptivity, will also have a thriving future.

For readers who want to stay up-to-date on sparse grid research, we refer to the sparse grid bibliography at www.ins.uni-bonn.de/info/sgbib, which gives roughly 300 articles from the past 40 years.

REFERENCES

- S. Achatz (2003a), Adaptive finite Dünngitter-Elemente höherer Ordnung für elliptische partielle Differentialgleichungen mit variablen Koeffizienten, Dissertation, Institut für Informatik, TU München.
- S. Achatz (2003b), ‘Higher order sparse grids methods for elliptic partial differential equations with variable coefficients’, *Computing* **71**, 1–15.
- G. Allaire (1992), ‘Homogenization and two-scale convergence’, *SIAM J. Math. Anal.* **21**, 1482–1516.
- A. Allen (1972), *Regression and the Moore–Penrose Pseudoinverse*, Academic Press, New York.
- E. Arge, M. Dæhlen and A. Tveito (1995), ‘Approximation of scattered data using smooth grid functions’, *J. Comput. Appl. Math.* **59**, 191–205.
- 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), 982–985.
- R. Balder (1994), Adaptive Verfahren für elliptische und parabolische Differentialgleichungen auf dünnen Gittern, Dissertation, Institut für Informatik, TU München.
- R. Balder and C. Zenger (1996), ‘The solution of multidimensional real Helmholtz equations on sparse grids’, *SIAM J. Sci. Comp.* **17**, 631–646.
- R. Balder, U. Rüde, S. Schneider and C. Zenger (1994), Sparse grid and extrapolation methods for parabolic problems, in *Proc. International Conference on Computational Methods in Water Resources, Heidelberg 1994* (A. Peters et al., eds), Kluwer Academic, Dordrecht, pp. 1383–1392.
- A. Barron (1993), ‘Universal approximation bounds for superpositions of a sigmoidal function’, *IEEE Trans. Inform. Theory* **39**, 930–945.
- A. Barron (1994), ‘Approximation and estimation bounds for artificial neural networks’, *Machine Learning* **14**, 115–133.
- G. Baszenski and F. Delvos (1993), Multivariate Boolean midpoint rules, in *Numerical Integration IV* (H. Brass and G. Hämerlin, eds), Vol. 112 of *International Series of Numerical Mathematics*, Birkhäuser, Basel, pp. 1–11.
- G. Baszenski, F. Delvos and S. Jester (1992), Blending approximations with sine functions, in *Numerical Methods of Approximation Theory 9* (D. Braess and L. Schumaker, eds), Vol. 105 of *International Series of Numerical Mathematics*, Birkhäuser, Basel, pp. 1–19.
- B. J. C. Baxter and A. Iserles (2003), ‘On the foundations of computational mathematics’, in *Handbook of Numerical Analysis*, Vol. 11 (F. Cucker, ed.), Elsevier, pp. 3–35.

- R. Becker and R. Rannacher (1996), ‘A feed-back approach to error control in finite element methods: Basic analysis and examples’, *East-West J. Numer. Math.* **4**, 237–264.
- J. Bell, P. Colella and H. Glaz (1989), ‘A second order projection method for the incompressible Navier–Stokes equations’, *J. Comput. Phys.* **85**, 257–283.
- R. Bellmann (1961), *Adaptive Control Processes: A Guided Tour*, Princeton University Press.
- K. Bennett and O. Mangasarian (1992), ‘Robust linear programming discrimination of two linearly inseparable sets’, *Optimiz. Methods and Software* **1**, 23–34.
- M. Berry and G. Linoff (2000), *Mastering Data Mining*, Wiley.
- C. Blake and C. Merz (1998), ‘UCI repository of machine learning databases’.
www.ics.uci.edu/~mlearn/MLRepository.html
- T. Bonk (1994a), Ein rekursiver Algorithmus zur adaptiven numerischen Quadratur mehrdimensionaler Funktionen, Dissertation, Institut für Informatik, TU München.
- T. Bonk (1994b), A new algorithm for multi-dimensional adaptive numerical quadrature, in *Adaptive Methods: Algorithms, Theory, and Applications* (W. Hackbusch and G. Wittum, eds), Vol. 46 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden, pp. 54–68.
- P. Bradley and O. Mangasarian (1998), Feature selection via concave minimization and support vector machines, in *Machine Learning: Proc. 15th International Conference; ICML '98* (J. Shavlik, ed.), Morgan Kaufmann, pp. 82–90.
- H.-J. Bungartz (1992a), An adaptive Poisson solver using hierarchical bases and sparse grids, in *Iterative Methods in Linear Algebra* (P. de Groen and R. Beauwens, eds), Elsevier, Amsterdam, pp. 293–310.
- H.-J. Bungartz (1992b), Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung, Dissertation, Institut für Informatik, TU München.
- H.-J. Bungartz (1996), Concepts for higher order finite elements on sparse grids, in *Proc. International Conference on Spectral and High Order Methods, Houston 1995* (A. Ilin and L. Scott, eds), *Houston J. Math.*, pp. 159–170.
- H.-J. Bungartz (1997), ‘A multigrid algorithm for higher order finite elements on sparse grids’, *ETNA* **6**, 63–77.
- H.-J. Bungartz (1998), Finite elements of higher order on sparse grids, Habilitationsschrift, Institut für Informatik, TU München, and Shaker Verlag, Aachen.
- H.-J. Bungartz and S. Dirnstorfer (2003), ‘Multivariate quadrature on adaptive sparse grids’, *Computing* **71**, 89–114.
- H.-J. Bungartz and T. Dornseifer (1998), Sparse grids: Recent developments for elliptic partial differential equations, in *Multigrid Methods V* (W. Hackbusch and G. Wittum, eds), Vol. 3 of *Lecture Notes in Computational Science and Engineering*, Springer, Berlin/Heidelberg.
- H.-J. 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.-J. Bungartz and W. Huber (1995), First experiments with turbulence simulation on workstation networks using sparse grid methods, in *Computational Fluid*

- Dynamics on Parallel Systems* (S. Wagner, ed.), Vol. 50 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden.
- H.-J. Bungartz, M. Griebel and U. Rüde (1994a), ‘Extrapolation, combination, and sparse grid techniques for elliptic boundary value problems’, *Comput. Meth. Appl. Mech. Eng.* **116**, 243–252.
- H.-J. Bungartz, M. Griebel, D. Röschke and C. Zenger (1994b), ‘Pointwise convergence of the combination technique for the Laplace equation’, *East-West J. Numer. Math.* **2**, 21–45.
- H.-J. Bungartz, M. Griebel, D. Röschke and C. Zenger (1994c), Two proofs of convergence for the combination technique for the efficient solution of sparse grid problems, in *Domain Decomposition Methods in Scientific and Engineering Computing* (D. Keyes and J. Xu, eds), Vol. 180 of *Contemporary Mathematics*, AMS, Providence, RI, pp. 15–20.
- H.-J. Bungartz, M. Griebel, D. Röschke and C. Zenger (1996), ‘A proof of convergence for the combination technique using tools of symbolic computation’, *Math. Comp. Simulation* **42**, 595–605.
- R. Caflisch, W. Morokoff and A. Owen (1997), ‘Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension’, *J. Comput. Finance* **1**, 27–46.
- J. Carnicer, W. Dahmen and J. Pena (1996), ‘Local decomposition of refinable spaces’, *Appl. Comp. Harm. Anal.* **3**, 127–153.
- V. Cherkassky and F. Mulier (1998), *Learning from Data: Concepts, Theory and Methods*, Wiley.
- A. Chorin (1968), ‘Numerical solution of the Navier–Stokes equations’, *Math. Comp.* **22**, 745–762.
- A. Chorin (1998), *Vorticity and Turbulence*, Springer.
- C. Chui (1992), *An Introduction to Wavelets*, Academic Press, Boston.
- 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’, *CR Acad. Sci. Paris, Ser. I* **335**, 99–104.
- K. Cios, W. Pedrycz and R. Swiniarski (1998), *Data Mining Methods for Knowledge Discovery*, Kluwer.
- A. Cohen (2003), *Numerical Analysis of Wavelet Methods*, Vol. 32 of *Studies in Mathematics and its Applications*, North-Holland.
- A. Cohen and I. Daubechies (1996), ‘A new technique to estimate the regularity of refinable functions’, *Rev. Mat. Iberoamer.* **12**, 527–591.
- A. Cohen, W. Dahmen and R. DeVore (2001), ‘Adaptive wavelet methods for elliptic operator equations’, *Math. Comp.* **70**, 27–75.
- A. Cohen, I. Daubechies and J. Feauveau (1992), ‘Biorthogonal bases of compactly supported wavelets’, *Comm. Pure Appl. Math.* **45**, 485–560.
- R. Cools and B. Maerten (1997), Experiments with Smolyak’s algorithm for integration over a hypercube, Technical Report, Department of Computer Science, Katholieke Universiteit Leuven.
- W. Dahmen and A. Kunoth (1992), ‘Multilevel preconditioning’, *Numer. Math.* **63**, 315–344.

- I. Daubechies (1988), ‘Orthogonal bases of compactly supported wavelets’, *Comm. Pure Appl. Math.* **41**, 909–996.
- I. Daubechies (1992), *Ten Lectures on Wavelets*, Vol. 61 of *CBMS–NSF Regional Conf. Series in Appl. Math.*, SIAM.
- I. Daubechies (1993), ‘Orthonormal bases of compactly supported wavelets II’, *SIAM J. Math. Anal.* **24**, 499–519.
- I. Daubechies and W. Sweldens (1998), ‘Factoring wavelet transforms into lifting steps’, *J. Fourier Anal. Appl.* **4**, 245–267.
- F. Delvos (1982), ‘ d -variate Boolean interpolation’, *J. Approx. Theory* **34**, 99–114.
- F. Delvos (1990), ‘Boolean methods for double integration’, *Math. Comp.* **55**, 683–692.
- 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, in *Acta Numerica*, Vol. 7, Cambridge University Press, pp. 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 (2003), Liberating the weights, Technical Report AMR03/04, University of New South Wales.
- 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 Vol. 18 of *CRM Proceedings and Lecture Notes* (G. Deslauriers and S. Dubuc, eds).
- P. V. Dooren and L. D. Ridder (1976), ‘An adaptive algorithm for numerical integration over an n -dimensional cube’, *J. Comp. Appl. Math.* **2**, 207–217.
- T. Dornseifer (1997), Diskretisierung allgemeiner elliptischer Differentialgleichungen in krummlinigen Koordinatensystemen auf dünnen Gittern, Dissertation, Institut für Informatik, TU München.
- T. Dornseifer and C. Pflaum (1996), ‘Elliptic differential equations on curvilinear bounded domains with sparse grids’, *Computing* **56**, 607–615.
- 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.
- K. Eriksson, D. Estep, P. Hansbo and C. Johnson (1996), *Adaptive Finite Elements*, Springer, Berlin/Heidelberg.
- G. Faber (1909), ‘Über stetige Funktionen’, *Mathematische Annalen* **66**, 81–94.
- S. Fahrmann and C. Lebiere (1990), The cascade-correlation learning architecture, in *Advances in Neural Information Processing Systems*, Vol. 2 (Touretzky, ed.), Morgan-Kaufmann.
- A. Frank (1995), Hierarchische Polynombasen zum Einsatz in der Datenkompression mit Anwendung auf Audiodaten, Diplomarbeit, Institut für Informatik, TU München.
- K. Frank and S. Heinrich (1996), ‘Computing discrepancies of Smolyak quadrature rules’, *J. Complexity* **12**, 287–314.

- K. Frank, S. Heinrich and S. Pereverzev (1996), ‘Information complexity of multivariate Fredholm integral equations in Sobolev classes’, *J. Complexity* **12**, 17–34.
- J. Friedman (1991), ‘Multivariate adaptive regression splines’, *Ann. Statist.* **19**, 1–141. With discussion and a rejoinder by the author.
- J. Garcke (1998), Berechnung von Eigenwerten der stationären Schrödinger-Gleichung mit der Kombinationstechnik, Diplomarbeit, Institut für Angewandte Mathematik, Universität Bonn.
- J. Garcke and M. Griebel (2000), ‘On the computation of the eigenproblems of hydrogen and helium in strong magnetic and electric fields with the sparse grid combination technique’, *J. Comput. Phys.* **165**, 694–716.
- J. Garcke and M. Griebel (2001a), Data mining with sparse grids using simplicial basis functions, in *Proc. 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA* (F. Provost and R. Srikant, eds), pp. 87–96.
- J. Garcke and M. Griebel (2001b), On the parallelization of the sparse grid approach for data mining, in *Large-Scale Scientific Computations, Third International Conference, LSSC 2001, Sozopol, Bulgaria* (S. Margenov, J. Wasniewski and P. Yalamov, eds), Vol. 2179 of *Lecture Notes in Computer Science*, pp. 22–32.
- J. Garcke and M. Griebel (2002), ‘Classification with sparse grids using simplicial basis functions’, *Intelligent Data Analysis* **6**, 483–502.
- J. Garcke, M. Griebel and M. Thess (2001), ‘Data mining with sparse grids’, *Computing* **67**, 225–253.
- J. Garcke, M. Hegland and O. Nielsen (2003), Parallelisation of sparse grids for large scale data analysis, in *Proc. International Conference on Computational Science 2003 (ICCS 2003), Melbourne, Australia* (P. Sloot, D. Abramson, A. Bogdanov, J. Dongarra, A. Zomaya and Y. Gorbachev, eds), Vol. 2659 of *Lecture Notes in Computer Science*, Springer, pp. 683–692.
- A. Genz and A. Malik (1980), ‘An adaptive algorithm for numerical integration over an n -dimensional rectangular region’, *J. Comp. Appl. Math.* **6**, 295–302.
- T. Gerstner (1995), Ein adaptives hierarchisches Verfahren zur Approximation und effizienten Visualisierung von Funktionen und seine Anwendung auf digitale 3D Höhenmodelle, Diplomarbeit, Institut für Informatik, TU München.
- T. Gerstner (1999), Adaptive hierarchical methods for landscape representation and analysis, in *Process Modelling and Landform Evolution*, Vol. 78 of *Lecture Notes in Earth Sciences*, Springer.
- T. Gerstner and M. Griebel (1998), ‘Numerical integration using sparse grids’, *Numer. Alg.* **18**, 209–232.
- T. Gerstner and M. Griebel (2003), ‘Dimension-adaptive tensor-product quadrature’, *Computing* **71**, 65–87.
- F. Girosi (1998), ‘An equivalence between sparse approximation and support vector machines’, *Neural Computation* **10**, 1455–1480.
- F. Girosi, M. Jones and T. Poggio (1995), ‘Regularization theory and neural networks architectures’, *Neural Computation* **7**, 219–265.
- G. Golub, M. Heath and G. Wahba (1979), ‘Generalized cross validation as a method for choosing a good ridge parameter’, *Technometrics* **21**, 215–224.

- 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.
- W. Gordon and C. Hall (1973), ‘Transfinite element methods: Blending-function interpolation over arbitrary curved element domains’, *Numer. Math.* **21**, 109–129.
- R. Graham, D. Knuth and O. Patashnik (1994), *Concrete Mathematics*, Addison-Wesley, Reading.
- P. Gresho and R. Sani (1987), ‘On pressure boundary conditions for the incompressible Navier–Stokes equations’, *Int. J. Numer. Meth. Fluids* **7**, 371–394.
- M. Griebel (1991a), Parallel multigrid methods on sparse grids, in *Multigrid Methods III* (W. Hackbusch and U. Trottenberg, eds), Vol. 98 of *International Series of Numerical Mathematics*, Birkhäuser, Basel, pp. 211–221.
- M. Griebel (1991b), A parallelizable and vectorizable multi-level algorithm on sparse grids, in *Parallel Algorithms for Partial Differential Equations* (W. Hackbusch, ed.), Vol. 31 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden, pp. 94–100.
- M. Griebel (1992), ‘The combination technique for the sparse grid solution of PDEs on multiprocessor machines’, *Parallel Processing Letters* **2**, 61–70.
- M. Griebel (1993), Sparse grid multilevel methods, their parallelization and their application to CFD, in *Proc. Parallel Computational Fluid Dynamics 1992* (J. Häser, ed.), Elsevier, Amsterdam, pp. 161–174.
- M. Griebel (1994a), ‘Multilevel algorithms considered as iterative methods on semi-definite systems’, *SIAM J. Sci. Statist. Comput.* **15**, 547–565.
- M. Griebel (1994b), *Multilevelmethoden als Iterationsverfahren über Erzeugendensystemen*, Teubner Skripten zur Numerik, Teubner, Stuttgart.
- M. Griebel (1998), ‘Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences’, *Computing* **61**, 151–179.
- M. Griebel and W. Huber (1995), Turbulence simulation on sparse grids using the combination method, in *Parallel Computational Fluid Dynamics, New Algorithms and Applications* (N. Satofuka, J. Périaux and A. Ecer, eds), Elsevier, Amsterdam, pp. 75–84.
- M. Griebel and S. Knapik (2000), ‘Optimized tensor-product approximation spaces’, *Constr. Approx.* **16**, 525–540.
- M. Griebel and F. Koster (2000), Adaptive wavelet solvers for the unsteady incompressible Navier–Stokes equations, in *Advances in Mathematical Fluid Mechanics* (J. Malek, J. Nečas and M. Rokyta, eds), Springer, pp. 67–118.
- M. Griebel and F. Koster (2003), Multiscale methods for the simulation of turbulent flows, in *Numerical Flow Simulation III* (E. Hirschel, ed.), Vol. 82 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, Springer, pp. 203–214.
- M. Griebel and P. Oswald (1994), ‘On additive Schwarz preconditioners for sparse grid discretizations’, *Numer. Math.* **66**, 449–463.
- M. Griebel and P. Oswald (1995a), ‘On the abstract theory of additive and multiplicative Schwarz algorithms’, *Numer. Math.* **70**, 161–180.

- M. Griebel and P. Oswald (1995b), ‘Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems’, *Adv. Comput. Math.* **4**, 171–206.
- M. Griebel and T. Schiekofer (1999), An adaptive sparse grid Navier–Stokes solver in 3D based on the finite difference method, in *Proc. ENUMATH97* (H. Bock, G. Kanschat, R. Rannacher, F. Brezzi, R. Glowinski, Y. Kuznetsov and J. Periaux, eds), World Scientific, Heidelberg.
- M. Griebel and M. Schweitzer (2002), ‘A particle-partition of unity method, part II: Efficient cover construction and reliable integration’, *SIAM J. Sci. Comp.* **23**, 1655–1682.
- M. Griebel and V. Thurner (1993), ‘Solving CFD-problems efficiently by the combination method’, *CFD-News* **3**, 19–31.
- M. Griebel and V. Thurner (1995), ‘The efficient solution of fluid dynamics problems by the combination technique’, *Int. J. Numer. Meth. Heat Fluid Flow* **5**, 251–269.
- M. Griebel, W. Huber and C. Zenger (1993a), A fast Poisson solver for turbulence simulation on parallel computers using sparse grids, in *Flow Simulation on High-Performance Computers I* (E. Hirschel, ed.), Vol. 38 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden.
- M. Griebel, W. Huber and C. Zenger (1996), Numerical turbulence simulation on a parallel computer using the combination method, in *Flow Simulation on High-Performance Computers II* (E. Hirschel, ed.), Vol. 52 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden, pp. 34–47.
- M. Griebel, P. Oswald and T. Schiekofer (1999), ‘Sparse grids for boundary integral equations’, *Numer. Math.* **83**, 279–312.
- M. Griebel, M. Schneider and C. Zenger (1992), A combination technique for the solution of sparse grid problems, in *Iterative Methods in Linear Algebra* (P. de Groen and R. Beauwens, eds), Elsevier, Amsterdam, pp. 263–281.
- M. Griebel, C. Zenger and S. Zimmer (1993b), ‘Multilevel Gauss–Seidel-algorithms for full and sparse grid problems’, *Computing* **50**, 127–148.
- M. Gromov (1999), *Metric Structures for Riemannian and Non-Riemannian Spaces*, Vol. 152 of *Progress in Mathematics*, Birkhäuser.
- W. Hackbusch (1985), *Multigrid Methods and Applications*, Springer, Berlin/Heidelberg.
- W. Hackbusch (1986), *Theorie und Numerik elliptischer Differentialgleichungen*, Teubner, Stuttgart.
- W. Hackbusch (2001), ‘The efficient computation of certain determinants arising in the treatment of Schrödinger’s equation’, *Computing* **67**, 35–56.
- W. Hahn (1990), Parallelisierung eines adaptiven hierarchischen Dünngitterverfahrens, Diplomarbeit, Institut für Informatik, TU München.
- K. Hallatschek (1992), ‘Fouriertransformation auf dünnen Gittern mit hierarchischen Basen’, *Numer. Math.* **63**, 83–97.
- T. Hastie and R. Tibshirani (1990), *Generalized Additive Models*, Chapman and Hall.
- T. He (2001), *Dimensionality Reducing Expansion of Multivariate Integration*, Birkhäuser.

- M. Hegland (2002), Additive sparse grid fitting, in *Proc. 5th International Conference on Curves and Surfaces, Saint-Malo, France 2002*. Submitted.
- M. Hegland (2003), Adaptive sparse grids, in *Proc. 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, MCS-VUW research report.
- P. Hemker (1995), ‘Sparse-grid finite-volume multigrid for 3D problems’, *Adv. Comput. Math.* **4**, 83–110.
- P. Hemker and P. de Zeeuw (1996), BASIS3: A data structure for 3-dimensional sparse grids, in *Euler and Navier-Stokes Solvers Using Multi-Dimensional Upwind Schemes and Multigrid Acceleration* (H. Deconinck and B. Koren, eds), Vol. 56 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden, pp. 443–484.
- P. Hemker and C. Pflaum (1997), ‘Approximation on partially ordered sets of regular grids’, *Appl. Numer. Math.* **25**, 55–87.
- P. Hemker, B. Koren and J. Noordmans (1998), 3D multigrid on partially ordered sets of grids. *Multigrid Methods V*, Vol. 3 of *Lecture Notes in Computational Science and Engineering*, Springer, Berlin/Heidelberg, pp. 105–124.
- J. Hennart and E. Mund (1988), ‘On the h - and p -versions of the extrapolated Gordon’s projector with applications to elliptic equations’, *SIAM J. Sci. Statist. Comput.* **9**, 773–791.
- J. Heroth (1997), Are sparse grids suitable for the tabulation of reduced chemical systems?, Technical Report, TR 97-2, Konrad-Zuse-Zentrum für Informationstechnik Berlin.
- F. Hickernell, I. Sloan and G. Wasilkowski (2003), On tractability of weighted integration for certain Banach spaces of functions, Technical Report AMR03/08, University of New South Wales.
- R. Hiptmair and V. Gradinaru (2003), ‘Multigrid for discrete differential forms on sparse grids’, *Computing* **71**, 17–42.
- E. Hlawka (1961), ‘Funktionen von beschränkter Variation in der Theorie der Gleichverteilung’, *Ann. Math. Pure Appl.* **54**, 325–333.
- T. Ho and E. Kleinberg (1996), ‘Checkerboard dataset’.
www.cs.wisc.edu/math-prog/mpml.html
- V. Hoang and C. Schwab (2003), High-dimensional finite elements for elliptic problems with multiple scales, Technical Report 2003-14, Seminar für Angewandte Mathematik, ETH Zürich.
- R. Hochmuth (1999), Wavelet bases in numerical analysis and restricted nonlinear approximation, Habilitationsschrift, Freie Universität Berlin.
- R. Hochmuth, S. Knapek and G. Zumbusch (2000), Tensor products of Sobolev spaces and applications, Technical Report 685, SFB 256, Universität Bonn.
- J. Hoschek and D. Lasser (1992), *Grundlagen der Geometrischen Datenverarbeitung*, Teubner, chapter 9.
- W. Huber (1996a), Numerical turbulence simulation on different parallel computers using the sparse grid combination method, in *Proc. EuroPar ’96, Lyon 1996* (L. Bougé, P. Fraigniaud, A. Mignotte and Y. Robert, eds), Vol. 1124 of *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, pp. 62–65.

- W. Huber (1996b), Turbulenzsimulation mit der Kombinationsmethode auf Workstation-Netzen und Parallelrechnern, Dissertation, Institut für Informatik, TU München.
- M. Kalos and P. Whitlock (1986), *Monte Carlo Methods*, Wiley.
- G. Karniadakis and S. Sherwin (1999), *Spectral/hp Element Methods for CFD*, Oxford University Press.
- L. Kaufman (1999), Solving the quadratic programming problem arising in support vector classification, in *Advances in Kernel Methods: Support Vector Learning* (B. Schölkopf, C. Burges and A. Smola, eds), MIT Press, pp. 146–167.
- S. Khavinson (1997), *Best Approximation by Linear Superposition (Approximate Nomography)*, Vol. 159 of *AMS Translations of Mathematical Monographs*, AMS, Providence, RI.
- S. Knapek (2000a), Approximation and Kompression mit Tensorprodukt-Multiskalen-Approximationsräumen, Dissertation, Institut für Angewandte Mathematik, Universität Bonn.
- S. Knapek (2000b), Hyperbolic cross approximation of integral operators with smooth kernel, Technical Report 665, SFB 256, Universität Bonn.
- S. Knapek and F. Koster (2002), ‘Integral operators on sparse grids’, *SIAM J. Numer. Anal.* **39**, 1794–1809.
- A. Kolmogorov (1956), ‘On the representation of continuous functions of several variables by superpositions of continuous functions of fewer variables’, *Dokl. Akad. Nauk SSSR* **108**, 179–182. In Russian; English Translation in *Amer. Math. Soc. Transl. (2)* **17** (1961), 369–373.
- A. Kolmogorov (1957), ‘On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition’, *Dokl. Akad. Nauk SSSR* **114**, 953–956. In Russian; English Translation in *Amer. Math. Soc. Transl. (2)* **28** (1963), 55–59.
- B. Koren, P. Hemker and P. de Zeeuw (1996), Semi-coarsening in three directions for Euler-flow computations in three dimensions, in *Euler and Navier–Stokes Solvers Using Multi-Dimensional Upwind Schemes and Multigrid Acceleration* (H. Deconinck and B. Koren, eds), Vol. 56 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden, pp. 547–567.
- F. Koster (2002), Multiskalen-basierte Finite Differenzen Verfahren auf adaptiven dünnen Gittern, Dissertation, Institut für Angewandte Mathematik, Universität Bonn.
- A. Krommer and C. Ueberhuber (1994), *Numerical Integration on Advanced Computer Systems*, Vol. 848 of *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg.
- F. Kupka (1997), Sparse Grid Spectral Methods for the Numerical Solution of Partial Differential Equations with Periodic Boundary Conditions, Dissertation, Institut für Mathematik, Universität Wien.
- Y. Lee and O. Mangasarian (2001), ‘SSVM: A smooth support vector machine for classification’, *Comput. Optimiz. Appl.* **20**, 5–22.
- P. Lemarié (1988), ‘Ondelettes à localisation exponentielle’, *J. Math. Pures Appl.* **67**, 222–236.
- C. Liem, T. Lu and T. Shih (1995), *The Splitting Extrapolation Method*, World Scientific, Singapore.

- R. Lorentz and P. Oswald (1998), Multilevel finite element Riesz bases in Sobolev spaces, in *Proc. 9th International Conference on Domain Decomposition* (P. Bjoerstad *et al.*, eds), Domain Decomposition Press, Bergen, pp. 178–187.
- S. Martello and P. Toth (1990), *Knapsack Problems: Algorithms and Computer Implementations*, Wiley, Chichester.
- A. Matache (2001), Sparse two-scale FEM for homogenization problems, Technical Report 2001-09, Seminar für Angewandte Mathematik, ETH Zürich.
- A. Matache and C. Schwab (2001), Two-scale FEM for homogenization problems, Technical Report 2001-06, Seminar für Angewandte Mathematik, ETH Zürich.
- G. Melli (2003), ‘DatGen: A program that creates structured data’, web site: www.datasetgenerator.com
- Y. Meyer (1992), *Wavelets and Operators*, Cambridge University Press.
- H. Mhaskar (1996), ‘Neural networks and approximation theory’, *Neural Networks* **9**, 711–722.
- A. Michalke (1964), ‘On the inviscid instability of the hyperbolic tangent velocity profile’, *J. Fluid Mech.* **19**, 543–556.
- V. Milman (1988), ‘The heritage of P. Levy in geometrical functional analysis’, *Asterisque* **157–158**, 273–301.
- V. Milman and G. Schechtman (1986), *Asymptotic Theory of Finite-Dimensional Normed Spaces*, Vol. 1200 of *Lecture Notes in Mathematics*, Springer.
- U. Mitzlaff (1997), Diffusionsapproximation von Warteschlangensystemen, Dissertation, Institut für Mathematik, Technische Universität Clausthal.
- W. Morokoff and R. Caflisch (1995), ‘Quasi-Monte Carlo integration’, *J. Comput. Phys.* **122**, 218–230.
- W. Mulder (1989), ‘A new multigrid approach to convection problems’, *J. Comput. Phys.* **83**, 303–323.
- N. Naik and J. van Rosendale (1993), ‘The improved robustness of multigrid elliptic solvers based on multiple semicoarsened grids’, *SIAM J. Numer. Anal.* **30**, 215–229.
- H. Niederreiter (1992), *Random Number Generation and Quasi-Monte-Carlo Methods*, SIAM, Philadelphia.
- P. Niyogi and F. Girosi (1998), ‘Generalization bounds for function approximation from scattered noisy data’, *Adv. Comput. Math.* **10**, 51–80.
- E. Novak and K. Ritter (1996), ‘High dimensional integration of smooth functions over cubes’, *Numer. Math.* **75**, 79–98.
- E. Novak and K. Ritter (1997), Simple cubature formulas for d -dimensional integrals with high polynomial exactness and small error, Technical Report, Institut für Mathematik, Universität Erlangen–Nürnberg.
- 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), International Series in Numerical Mathematics, Birkhäuser, Basel, pp. 177–188.
- E. Novikov (1976), ‘Dynamics and statistics of a system of vortices’, *Sov. Phys. JETP* **41**, 937–943.
- P. Oswald (1994), *Multilevel Finite Element Approximation*, Teubner Skripten zur Numerik, Teubner, Stuttgart.

- S. Paskov (1993), ‘Average case complexity of multivariate integration for smooth functions’, *J. Complexity* **9**, 291–312.
- S. Paskov and J. Traub (1995), ‘Faster valuation of financial derivatives’, *J. Portfolio Management* **22**, 113–120.
- T. Patterson (1986), ‘The optimum addition of points to quadrature formulae’, *Math. Comp.* **22**, 847–856.
- A. Paul (1995), Kompression von Bildfolgen mit hierarchischen Basen, Diplomarbeit, Institut für Informatik, TU München.
- A. Peano (1976), ‘Hierarchies of conforming finite elements for plane elasticity and plate bending’, *Comp. Math. Appl.* **2**, 211–224.
- K. Petras (2000), ‘On the Smolyak cubature error for analytic functions’, *Adv. Comput. Math.* **12**, 71–93.
- A. Pfaffinger (1997), Funktionale Beschreibung und Parallelisierung von Algorithmen auf dünnen Gittern, Dissertation, Institut für Informatik, TU München.
- C. Pflaum (1992), Anwendung von Mehrgitterverfahren auf dünnen Gittern, Diplomarbeit, Institut für Informatik, TU München.
- C. Pflaum (1996), Diskretisierung elliptischer Differentialgleichungen mit dünnen Gittern, Dissertation, Institut für Informatik, TU München.
- C. Pflaum (1998), ‘A multilevel algorithm for the solution of second order elliptic differential equations on sparse grids’, *Numer. Math.* **79**, 141–155.
- C. Pflaum and A. Zhou (1999), ‘Error analysis of the combination technique’, *Numer. Math.* **84**, 327–350.
- G. Pöplau and F. Sprengel (1997), Some error estimates for periodic interpolation on full and sparse grids, in *Curves and Surfaces with Applications in CAGD* (A. Le Méhauté, C. Rabut and L. Schumaker, eds), Vanderbilt University Press, Nashville, Tennessee, pp. 355–362.
- J. Prakash (2000), Rouse chains with excluded volume interactions: Linear viscoelasticity, Technical Report 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.
- A. Prohl (1997), *Projection and Quasi-Compressibility Methods for Solving the Incompressible Navier–Stokes Equations*, Advances in Numerical Mathematics, B. G. Teubner.
- T. Rassias and J. Sims (1995), *Finite Sums Decompositions in Mathematical Analysis*, Wiley.
- C. Reisinger (2003), Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben, Dissertation, Universität Heidelberg.
- P. Rouse (1953), ‘A theory of the linear viscoelastic properties of dilute solutions of coiling polymers’, *J. Chem. Phys.* **21**, 1272–1280.
- S. Salzberg (1997), ‘On comparing classifiers: Pitfalls to avoid and a recommended approach’, *Data Mining and Knowledge Discovery* **1**, 317–327.
- T. Schiekofer (1998), Die Methode der finiten Differenzen auf dünnen Gittern zur Lösung elliptischer und parabolischer partieller Differentialgleichungen, Dissertation, Institut für Angewandte Mathematik, Universität Bonn.

- T. Schiekofer and G. Zumbusch (1998), Software concepts of a sparse grid finite difference code, in *Proc. 14th GAMM-Seminar on Concepts of Numerical Software* (W. Hackbusch and G. Wittum, eds), *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden.
- S. Schneider (2000), Adaptive Solution of Elliptic PDE by Hierarchical Tensor Product Finite Elements, Dissertation, Institut für Informatik, TU München.
- C. Schwab and R. Todor (2002), Sparse finite elements for stochastic elliptic problems, Technical Report 2002-05, Seminar für Angewandte Mathematik, ETH Zürich.
- C. Schwab and R. Todor (2003), ‘Sparse finite elements for stochastic elliptic problems: Higher order moments’, *Computing* **71**, 43–63.
- M. Schweitzer (2003), *A Parallel Multilevel Partition of Unity Method for Elliptic Partial Differential Equations*, Vol. 29 of *Lecture Notes in Computational Science and Engineering*, Springer.
- R. Sedgewick and P. Flajolet (1996), *Analysis of Algorithms*, Addison-Wesley, Reading.
- 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’, *Queueing Systems* **42**, 33–62.
- J. Simsa (1992), ‘The best L^2 -approximation by finite sums of functions with separable variables’, *Aequationes Mathematicae* **43**, 284–263.
- P. Sjöberg (2002), Numerical solution of the master equation in molecular biology, Master’s Thesis, Department of 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.
- I. Sloan and S. Joe (1994), *Lattice Methods for Multiple Integration*, Oxford University Press.
- I. Sloan and H. Woźniakowski (1998), ‘When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?’, *J. Complexity* **14**, 1–33.
- 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), 1042–1045.
- F. Sprengel (1997a), Interpolation and Wavelet Decomposition of Multivariate Periodic Functions, Dissertation, FB Mathematik, Universität Rostock.
- F. Sprengel (1997b), ‘A unified approach to error estimates for interpolation on full and sparse Gauss–Chebyshev grids’, *Rostocker Math. Kolloq.* **51**, 51–64.
- R. Stevenson (1996), Piecewise linear (pre)-wavelets on non-uniform meshes, in *Multigrid Methods IV* (W. Hackbusch and G. Wittum, eds), Vol. 3 of *Lecture Notes in Computational Science and Engineering*, Springer.
- M. Stone (1974), ‘Cross-validatory choice and assessment of statistical predictions’, *J. Royal Statist. Soc. Ser. B* **36**, 111–147.
- T. Störkkuhl (1995), Ein numerisches adaptives Verfahren zur Lösung der biharmonischen Gleichung auf dünnen Gittern, Dissertation, Institut für Informatik, TU München.
- W. Sweldens (1997), ‘The lifting scheme: A construction of second generation wavelets’, *SIAM J. Math. Anal.* **29**, 511–546.

- M. Talagrand (1995), ‘Concentration of measure and isoperimetric inequalities in product spaces’, *Publ. Math. IHES* **81**, 73–205.
- R. Temam (1969), ‘Sur l’approximation de la solution des équations de Navier–Stokes par la méthode des fractionnaires (II)’, *Arch. Rat. Mech. Anal.* **33**, 377–385.
- V. Temlyakov (1989), *Approximation of Functions with Bounded Mixed Derivative*, Vol. 178 of *Proc. Steklov Inst. of Math.*, AMS, Providence, RI.
- V. Temlyakov (1993), ‘On approximate recovery of functions with bounded mixed derivative’, *J. Complexity* **9**, 41–59.
- V. Temlyakov (1994), *Approximation of Periodic Functions*, Nova Science, Commack, New York.
- A. Tikhonov and V. Arsenin (1977), *Solutions of Ill-Posed Problems*, W. H. Winston, Washington DC.
- 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.
- F. Utreras (1979), Cross-validation techniques for smoothing spline functions in one or two dimensions, in *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds), Springer, Heidelberg, pp. 196–231.
- V. Vapnik (1982), *Estimation of Dependences Based on Empirical Data*, Springer, Berlin.
- V. Vapnik (1995), *The Nature of Statistical Learning Theory*, Springer.
- G. Wahba (1985), ‘A comparison of GCV and GML for choosing the smoothing parameter in the generalized splines smoothing problem’, *Ann. Statist.* **13**, 1378–1402.
- G. Wahba (1990), *Spline Models for Observational Data*, Vol. 59 of *Series in Applied Mathematics*, SIAM, Philadelphia.
- G. Wasilkowski and H. Woźniakowski (1995), ‘Explicit cost bounds of algorithms for multivariate tensor product problems’, *J. Complexity* **11**, 1–56.
- G. Wasilkowski and H. Woźniakowski (1999), ‘Weighted tensor product algorithms for linear multivariate problems’, *J. Complexity* **15**, 402–447.
- A. Werschulz (1995), The complexity of the Poisson problem for spaces of bounded mixed derivatives, Technical Report CUCS-016-95, Columbia University.
- H. Woźniakowski (1985), ‘A survey of information-based complexity’, *J. Complexity* **1**, 11–44.
- H. Yserentant (1986), ‘On the multi-level splitting of finite element spaces’, *Numer. Math.* **49**, 379–412.
- H. Yserentant (1990), ‘Two preconditioners based on the multi-level splitting of finite element spaces’, *Numer. Math.* **58**, 163–184.
- H. Yserentant (1992), Hierarchical bases, in *Proc. ICIAM ’91, Washington 1991* (R. O’Malley *et al.*, eds), SIAM, Philadelphia.
- H. Yserentant (2004), On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives. *Numer. Math.*, in press.

- R. Yue and F. Hickernell (2002), ‘Robust designs for smoothing spline ANOVA models’, *Metrika* **55**, 161–176.
- C. Zenger (1991), Sparse grids, in *Parallel Algorithms for Partial Differential Equations* (W. Hackbusch, ed.), Vol. 31 of *Notes on Numerical Fluid Mechanics*, Vieweg, Braunschweig/Wiesbaden.
- O. Zienkiewicz, D. Kelly, J. Gago and I. Babuška (1982), Hierarchical finite element approaches, error estimates, and adaptive refinement, in *The Mathematics of Finite Elements and Applications IV* (J. Whiteman, ed.), Academic Press, London.
- G. Zumbusch (1996), *Simultaneous hp Adaptation in Multilevel Finite Elements*, Shaker, Aachen.