Acta Numerica (1993), *pp.* 285-326

Old and New Convergence Proofs for Multigrid Methods

Harry Yserentant
Mathematisches Institut
Universität Tübingen
D-7400 Tübingen, Germany
E-mail: harry@tue-num2.mathematik.uni-tuebingen.de

Multigrid methods are the fastest known methods for the solution of the large systems of equations arising from the discretization of partial differential equations. For self-adjoint and coercive linear elliptic boundary value problems (with Laplace's equation and the equations of linear elasticity as two typical examples), the convergence theory reached a mature, if not its final state. The present article reviews old and new developments for this type of equation and describes the recent advances.

CONTENTS

1. Introduction

The discretization of partial differential equations leads to very large systems of equations. For two-dimensional problems, several ten thousand unknowns are not unusual, and in three space dimensions, more than one million unknowns can be reached very easily. The direct solution of systems of this size is prohibitively expensive, both with respect to the amount of storage and to the computational work. Therefore iterative methods like the Gauß–Seidel or the Jacobi iteration have been used from the beginning of the numerical treatment of partial differential equations.

An important step was Young's successive over-relaxation method (1950) which is much faster than the closely related Gauß–Seidel iteration. Nevertheless, this method shares with direct elimination methods the disadvantage that the amount of work does not remain proportional to the number of unknowns; the computer time needed to solve a problem grows more rapidly than the size of the problem. The standard reference on iterative methods is Varga (1962). For a recent treatment, see Hackbusch (1991).

Multigrid methods were the first to overcome this complexity barrier. Multigrid methods are composed of simple basic iterations. Probably the first working multigrid method was developed and analysed by Fedorenko (1964) for the Laplace equation on the unit square. Bachvalov (1966) considered the theoretically much more complex case of variable coefficients. Although the basic idea of combining discretizations on different grids in an iterative scheme appears to be very natural, the potential of this idea was not recognized before the middle of the 1970s. At this time, the multigrid idea began to spread.

The report of Hackbusch (1976) and the paper of Brandt (1977) were the historical break through. The first big multigrid conference in 1981 in Köln was a culmination point of the development; the conference proceedings edited by Hackbusch and Trottenberg (1982) are still a basic reference. With Hackbusch's 1985 monograph, the first stage in multigrid theory came to an end.

Today, multigrid methods are used in nearly every field where partial differential equations are solved by numerical methods. They are applied in computational fluid dynamics as well as in semiconductor simulations. The bibliographies in McCormick (1987) and Wesseling (1992) each contain several hundred references.

The field of multigrid methods has became too large to review in a single article. Therefore, in this paper, we restrict our attention to the class of problems which is best understood, namely to self-adjoint and coercive linear elliptic boundary value problems. For mathematicians, the typical equation in this class is the Laplace equation. People, who are more oriented to real life, would probably think of the partial differential equations of structural mechanics.

Hackbusch (1982) and Braess and Hackbusch (1983) gave the first really satisfying convergence proofs for multigrid methods applied to this class of problem. The main problem with these proofs, and with all other convergence proofs appearing until the beginning of the 1990s, is that they are based on regularity properties of the boundary value problem which are rarely satisfied in practice. In addition, the underlying finite element or finite difference meshes have to be quasi-uniform, i.e. all discretization cells have to be of approximately the same size. Although these assumptions are common in the theory of finite element methods, they are unrealistic.

These problems led Bank and Dupont (1980) and Axelsson and Gustafsson (1983) to the development of the two-level hierarchical basis methods. Yserentant (1986b) and Bank, Dupont and Yserentant (1988) extended this idea to the multilevel case. These methods have a simpler structure than the usual multigrid methods and do not depend, by their construction, on the restrictive assumptions mentioned earlier. Hierarchical basis methods have been shown to be very efficient in adaptive finite element codes; see Bank (1990) and Deuflhard, Leinen and Yserentant (1989).

Another development of the 1980s were the domain decomposition methods with the Schwarz alternating method as an early example; see Chan *et al.* (1989), for example. Recently these independent fields merged in a joint abstract theory which is flexible enough to treat many, at first sight completely different iteration schemes. The basic references are Bramble, Pasciak, Wang and Xu (1991a, 1991b), Bramble and Pasciak (1991), Dryja and Widlund (1991) and, especially in regard to terminology, Xu (1992b). This unified theory is one of the main topics of the present review article.

Fast iterative methods for the systems of equations

$$Au = f \tag{1.1}$$

resulting from the discretization of self-adjoint coercive linear elliptic boundary value problems are not only of interest in their own field but also of interest elsewhere.

For example, such methods can be utilized for the efficient solution of saddle point problems

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{1.2}$$

as they arise from the discretization of the Stokes equation. Such approaches are described and analysed in the papers of Bramble and Pasciak (1988) and Bank, Welfert and Yserentant (1990).

Fast iterative methods for the equation (1.1) can also be used to construct comparably fast methods for the solution of systems

$$(A + M)\, u = f \tag{1.3}$$

arising from the discretization of boundary value problems with lower order terms (here represented by M) making the system indefinite and/or unsymmetric. Helmholtz type and convection–diffusion equations fall into this class. Methods of this type are described in Yserentant (1986c), Vassilevski (1992) and Xu (1992a). The mathematical background of these papers is an observation concerning the finite element discretization of perturbed elliptic boundary value problems which has been made by Schatz (1974). The analysis of multigrid methods, which can be directly applied to such boundary value problems, is also based on such perturbation arguments; see

Bank (1981), for example. A completely different approach can be found in Yserentant (1988).

If M is not self-adjoint and becomes the dominating part in (1.3), i.e. for convection-dominated problems, for example, the construction of appropriate fast solvers becomes more difficult and is still in its infancy. Until the present day, most iterative methods for such problems were constructed on a more or less heuristic basis. Multigrid methods based on incomplete factorizations turned out to be very efficient. For certain model problems, a rigorous analysis is possible; see Hackbusch (1985), Wittum (1989b) or Stevenson (1992). With his frequency decomposition multigrid methods, Hackbusch (1989a,b) presented a very promising approach. These methods are also well suited to boundary value problems with strong anisotropies. It should be mentioned that conjugate-gradient-like algorithms play an important role in the solution of nonsymmetric linear algebraic equations like (1.3). For a survey of recent developments, see Freund, Golub and Nachtigal (1992).

Last, but not least, fast iterative solvers for standard symmetric, positive definite finite element equations can be applied to related nonlinear boundary value problems via approximate Newton techniques; see Bank and Rose (1982) and Deuflhard (1992). Often, such methods based on inner–outer iterations present an alternative to nonlinear multigrid methods which treat the boundary value problem directly. Information about nonlinear multigrid methods can be found in Hackbusch's 1985 book. A very elaborate convergence analysis is given in Hackbusch and Reusken (1989).

The rest of this paper is organized as follows. In Section 2, we introduce a very general class of approximate subspace correction methods for the solution of abstract linear equations $Au = f$ with self-adjoint and positive definite operators A replacing the usual matrices. The classical multigrid or, as we often prefer to say, multilevel methods as well as many domain decomposition methods are such subspace correction methods. How multigrid methods can be interpreted in this sense, will be discussed in detail. In addition to the classical multigrid algorithms, we present the hierarchical basis methods, which are extremely well suited to adaptively generated, nonuniform finite element meshes.

In Section 3, a first convergence proof for two-grid methods is given. This proof follows the lines given in Bank and Dupont (1981). The two-grid convergence result can be utilized to prove the convergence of the so called multigrid W-cycle. All early convergence proofs for multigrid methods followed this strategy.

A more sophisticated multigrid convergence proof based on the ideas of Braess and Hackbusch (1983) will be presented in Section 4. Contrary to the two grid–multigrid analysis, this convergence proof also applies to the multigrid V-cycle which is simpler than the W-cycle.

In Section 5, a general convergence theory for the recursively defined, multiplicative subspace correction methods introduced in Section 2 is developed. In Section 6, this abstract theory is applied to multigrid methods for the solution of finite element equations and to the hierarchical basis multigrid method of Bank, Dupont and Yserentant (1988). In addition to a regularity dependent convergence result, which is closely related to the result of Wittum (1989a), a first regularity free convergence estimate is derived.

Utilizing the results presented in Section 7, one can show that multigrid methods reach an optimal complexity which is independent of the regularity properties of the boundary value problem. These results are mainly due to Oswald (1990, 1991) and to the forthcoming paper of Dahmen and Kunoth (1992). Many of the tools employed in these papers were taken from the classical approximation theory and the theory of function spaces. A self-contained presentation for the case of second order problems can be found in Bornemann and Yserentant (1992).

Finally, in Section 8, we devote our attention to additive multilevel methods which are a special case of the additive subspace correction methods already introduced in Section 2. The most prominent examples of these methods are the hierarchical basis solver (Yserentant, 1986b) and the recent multilevel nodal basis method of Bramble, Pasciak and Xu (1990) and Xu (1989). Our presentation follows Xu (1992b) and Yserentant (1990). These methods are more flexible and simpler than the usual recursively defined multilevel methods and fit very well to nonuniformly refined grids. In addition, they present advantages for implementation on parallel computers.

The present survey article is strongly influenced by the recent work of Bramble, Pasciak, Wang and Xu. Although not explicitly stated at every place, often we follow their argumentation very closely, especially in Sections 2, 5 and 8. The merits of these authors are herewith explicitly acknowledged.

Special thanks also to Randy Bank and Wolfgang Hackbusch who laid the foundations of multigrid convergence theory. They have supported me in many respects.

2. Subspace correction- and multilevel-methods

We begin this section with a very abstract formulation of a discrete elliptic boundary value problem. Let $\mathcal{S}$ be a finite dimensional space. We assume that $\mathcal{S}$ is equipped with an inner product $a(u, v)$ inducing the norm

$$\|u\| = a(u, u)^{1/2} \tag{2.1}$$

and a second inner product (u, v) inducing the norm

$$\|u\|_0 = (u, u)^{1/2}. \tag{2.2}$$

We introduce a symmetric and positive definite operator $A : \mathcal{S} \to \mathcal{S}$ by the relation

$$(Au, v) = a(u, v), \quad v \in \mathcal{S}, \tag{2.3}$$

where symmetric and positive definite here is always understood to be symmetric and positive definite with respect to the inner product (u, v). Our aim is the construction and analysis of a general class of fast solvers for the abstract linear equation

$$Au = f. \tag{2.4}$$

This equation is equivalent to the problem finding an $u \in \mathcal{S}$ satisfying the relation

$$a(u, v) = (f, v) \tag{2.5}$$

for all elements $v \in \mathcal{S}$. We remark that the inner product (2.2) does not enter the final form of the algorithms as they are implemented on the computer, and that the constants in our central abstract convergence theorems will be invariant under a change in this inner product.

In the applications that we have in mind, (2.1) is the norm induced by the elliptic boundary value problem under consideration whereas (2.2) is chosen to be a L_2-like inner product. To give an example, let $\Omega \subseteq \mathbb{R}^2$ be a bounded polygonal domain. As a model problem, we consider the differential equation

$$-\sum_{i,j=1}^{2} D_j(a_{ij} D_i u) = f \tag{2.6}$$

on Ω with homogeneous boundary conditions $u = 0$ on the boundary of Ω. The weak formulation of this boundary value problem is to find a function $u \in H_0^1(\Omega)$ satisfying the relation

$$a(u, v) = \int_\Omega f v \, \mathrm{d}x \tag{2.7}$$

for all $v \in H_0^1(\Omega)$ where, in this example, the bilinear form $a(u, v)$ is given by the integral expression

$$a(u, v) = \int_\Omega \sum_{i,j=1}^{2} a_{ij} D_i u D_j v \, \mathrm{d}x. \tag{2.8}$$

We assume that the a_{ij} are continuously differentiable functions, that

$$a_{ij} = a_{ji}, \tag{2.9}$$

and that there are positive constants M and δ with

$$\delta \sum_{i=1}^{2} \xi_i^2 \;\le\; \sum_{i,j=1}^{2} a_{ij}(x) \xi_i \xi_j \;\le\; M \sum_{i=1}^{2} \xi_i^2 \tag{2.10}$$

for all $x \in \Omega$ and all $\xi_1, \xi_2 \in \mathbb{R}$. These conditions guarantee that (2.8) defines an inner product on $H_0^1(\Omega)$ which is equivalent to the usual inner product on this space.

By a triangulation $\mathcal{T}$ of Ω, we mean a set of triangles such that the intersection of two such triangles is either empty or consists of a common edge or a common vertex. Here we start with an intentionally coarse initial triangulation $\mathcal{T}_0$ of Ω. The triangulation $\mathcal{T}_0$ is refined several times, giving a family of nested triangulations $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \ldots$. For ease of presentation, we consider only uniformly refined families of triangulations in this article, at least from a rigorous point of view. Thus a triangle of $\mathcal{T}_{k+1}$ is generated by subdividing of triangle of $\mathcal{T}_k$ into four congruent subtriangles.

Nevertheless one should keep in mind that nonuniformly refined meshes are absolutely necessary to approximate solutions with singularities arising from corners, cracks, interfaces or nonlinearities. On an informal basis, we will discuss whether and in which way the presented results can be generalized to such sequences of grids.

For triangular grids, the most successful nonuniform refinement scheme is due to Bank and Weiser (1985). It is also described in Bank *et al.* (1988). The scheme is based on the regular subdivision of triangles as described earlier and on carefully chosen additional bisections of triangles. Refinement schemes, which are based exclusively on the bisection of triangles, are discussed in Bänsch (1991) and Rivara (1984). The nonuniform refinement of tetrahedral meshes in three space dimensions is a harder challenge. In Bänsch (1991), the bisection of tetrahedra is utilized. The refinement strategy of Bank and Weiser can also be generalized to three dimensions.

Corresponding to the triangulations $\mathcal{T}_k$ we have finite element spaces $\mathcal{S}_k$. In our example, $\mathcal{S}_k$ consists of all functions which are continuous on Ω and linear on the triangles in $\mathcal{T}_k$ and which vanish on the boundary of Ω. By construction, $\mathcal{S}_k$ is a subspace of $\mathcal{S}_l$ for $k \leq l$. The extension of the presented results to higher order spaces is more or less obvious.

For the rest of this paper, we fix a final level j and the corresponding finite element space $\mathcal{S} = \mathcal{S}_j$. The discrete boundary value problem corresponding to the abstract linear problem (2.4), (2.5) is to find a function $u \in \mathcal{S}$ satisfying the relation

$$a(u, v) = \int_\Omega f v \, \mathrm{d}x \tag{2.11}$$

for all functions $v \in \mathcal{S}$.

As mentioned earlier, the inner product (2.2) is usually a L_2-like inner product with an appropriate weight function. Our choice is

$$(u, v) = \sum_{T \in \mathcal{T}_0} \frac{1}{\mathrm{area}(T)} \int_T uv \, \mathrm{d}x. \tag{2.12}$$

The task of the weights here is to make our estimates independent of the size of the triangles in the initial triangulation. In the three-dimensional case, these factors have to be replaced by other factors behaving like $1/\mathrm{diam}(T)^2$.

After this illustrating example, which will accompany the whole paper, we return to the general theory. Let $\mathcal{W}_0, \mathcal{W}_1, \ldots, \mathcal{W}_J$ be subspaces of $\mathcal{S}$. We assume that every $u \in \mathcal{S}$ can be written as

$$u = w_0 + w_1 + \ldots + w_J, \quad w_l \in \mathcal{W}_l. \tag{2.13}$$

We neither assume that this representation is unique, nor that the spaces $\mathcal{W}_l$ are nested.

We need two kinds of orthogonal projections onto the spaces $\mathcal{W}_l$. The projections $Q_l : \mathcal{S} \to \mathcal{W}_l$ are defined by

$$(Q_l u, w_l) = (u, w_l), \quad w_l \in \mathcal{W}_l, \tag{2.14}$$

and the projections $P_l : \mathcal{S} \to \mathcal{W}_l$ by

$$a(P_l u, w_l) = a(u, w_l), \quad w_l \in \mathcal{W}_l. \tag{2.15}$$

If $u \in \mathcal{S}$ is the solution of (2.4), and (2.5), respectively, $P_l u \in \mathcal{W}_l$ is the Ritz approximation of this solution in $\mathcal{W}_l$.

The basic building block of the iterative methods considered here are the *subspace corrections*

$$\widetilde{u} \leftarrow \widetilde{u} + P_l(u - \widetilde{u}) \tag{2.16}$$

with respect to the spaces $\mathcal{W}_l$. The subspace correction (2.16) makes the error $u - \widetilde{u}$ between the exact solution and the new approximation a-orthogonal to the space $\mathcal{W}_l$.

To express these subspace corrections in terms of the right-hand side f and the approximations $\widetilde{u}$, we introduce the Ritz approximations $A_l : \mathcal{W}_l \to \mathcal{W}_l$ of the operator A with respect to the spaces $\mathcal{W}_l$. They are defined by

$$(A_l u, v) = (Au, v), \quad u, v \in \mathcal{W}_l, \tag{2.17}$$

or equivalently by

$$(A_l u, v) = a(u, v), \quad u, v \in \mathcal{W}_l. \tag{2.18}$$

The operators A, A_l, P_l and Q_l are connected by the relation

$$A_l P_l = Q_l A. \tag{2.19}$$

(2.19) easily follows from

$$(A_l P_l u, w_l) = a(P_l u, w_l) = a(u, w_l) = (Au, w_l) = (Q_l A u, w_l).$$

By (2.19), the Ritz approximation $P_l u$ of the solution u of (2.4) satisfies the equation

$$A_l P_l u = Q_l f, \tag{2.20}$$

and the subspace correction (2.16) can be written as

$$\widetilde{u} \leftarrow \widetilde{u} + A_l^{-1} Q_l (f - A\widetilde{u}). \tag{2.21}$$

It requires the computation of the Ritz approximation

$$P_l(u - \widetilde{u}) = A_l^{-1} Q_l (f - A\widetilde{u}) \tag{2.22}$$

of the error $u - \widetilde{u}$.

The problem is that, for sufficiently large and complicated subspaces $\mathcal{W}_l$, the computation of this approximate defect is far too expensive to lead to a reasonable method. Therefore we replace the subspace corrections (2.21) by *approximate subspace corrections*

$$\widetilde{u} \leftarrow \widetilde{u} + B_l^{-1} Q_l (f - A\widetilde{u}) \tag{2.23}$$

with symmetric and positive definite operators $B_l : \mathcal{W}_l \to \mathcal{W}_l$. The operators B_l should have the property that the correction term

$$d_l = B_l^{-1} Q_l (f - A\widetilde{u}) \tag{2.24}$$

can easily be computed as the solution of the linear system

$$(B_l d_l, w_l) = (f - A\widetilde{u}, w_l), \quad w_l \in \mathcal{W}_l. \tag{2.25}$$

Here we should remark that the computation of the right-hand side of (2.25)

$$(f - A\widetilde{u}, w_l) = (f, w_l) - a(\widetilde{u}, w_l) \tag{2.26}$$

does not require an explicit knowledge of the abstract operator A but only of the bilinear form $a(u, v)$ and of the linear functional representing the right-hand side of the equation.

A common situation is that the computation of the correction term in (2.23) consists of, say, m steps of a given convergent iterative procedure

$$\widetilde{w}_l \leftarrow \widetilde{w}_l + \widehat{B}_l^{-1} (r_l - A_l \widetilde{w}_l) \tag{2.27}$$

for the solution of the equation

$$A_l w_l = r_l, \quad r_l = Q_l (f - A\widetilde{u}), \tag{2.28}$$

with a symmetric positive definite operator $\widehat{B}_l : \mathcal{W}_l \to \mathcal{W}_l$, that means of m Jacobi steps, for example, where one starts with $w_l = 0$. Then the operator B_l is given by

$$B_l^{-1} = (I - (I - \widehat{B}_l^{-1} A_l)^m) A_l^{-1}, \tag{2.29}$$

and is automatically symmetric and positive definite.

If one combines the single subspaces corrections

$$\widetilde{u} \leftarrow \widetilde{u} + B_l^{-1} Q_l (f - A\widetilde{u}) \tag{2.30}$$

sequentially in the order $l = 0, 1, \ldots, J$, one obtains the *multiplicative subspace correction method* corresponding to the subspaces $\mathcal{W}_0, \ldots, \mathcal{W}_J$ of $\mathcal{S}$.

These method generalize the classical Gauß–Seidel iteration where the subspaces are one-dimensional and are spanned by basis functions.

Sufficient for the convergence of this composed method is that the iterations

$$w_l \leftarrow w_l + B_l^{-1}(f_l - A_l w_l) \tag{2.31}$$

for the solution of the equations $A_l w_l = f_l$ on $\mathcal{W}_l$ converge for arbitrarily chosen right-hand sides f_l and the initial approximation $w_l = 0$; this follows from Theorem 5.1 and the finite dimension of $\mathcal{S}$. If one assumes that

$$(A_l w_l, w_l) \leq \omega\,(B_l w_l, w_l), \quad w_l \in \mathcal{W}_l, \tag{2.32}$$

this condition is equivalent to

$$0 < \omega < 2. \tag{2.33}$$

We remark that the condition (2.33) is automatically satisfied, if the B_l themselves represent multiplicative subspace correction methods with exact subspace solvers, for example Gauß–Seidel iterations for the approximate solutions of the linear systems involving the operator A_l. For this particular choice, we have $\omega = 1$.

Classical multigrid methods for the solution of finite element equations like (2.11) fall into the category of such multiplicative subspace correction methods. The *multigrid V-cycle* is a multiplicative subspace correction method with the coarse grid spaces $\mathcal{S}_l$ as subspaces $\mathcal{W}_l$.

The V-cycle is usually defined by recursion on the number j of refinement levels. For the initial level 0, when only one grid is present, the equations are solved exactly. For two or more levels, one proceeds as follows.

Beginning with an approximation $u_0 = \widetilde{u}$ of the finite element equation $A_j u = f$, first a *coarse grid correction* is performed. For the two-level case, one computes the approximate defect $d = P_{j-1}(u - \widetilde{u}) \in \mathcal{S}_{j-1}$ as the solution of the level $j-1$ equation

$$A_{j-1} d = Q_{j-1}(f - A_j u) \tag{2.34}$$

and sets

$$u_1 = u_0 + d. \tag{2.35}$$

Then further intermediate approximations $u_2, \ldots, u_{m+1} \in \mathcal{S}_j$ are determined by m so called *smoothing steps*

$$u_{i+1} = u_i + \widehat{B}_j^{-1}(f - A_j u_i). \tag{2.36}$$

One ends with $\widetilde{u} = u_{m+1}$ as the new approximation for the solution of the equation $A_j u = f$. For more than two levels, the coarse grid equation (2.34) is approximately solved by a call of the method for the level $j-1$.

If the coarse level equations (2.34) are not solved by *one* but by *two* calls

of the method for the preceding level, one speaks of a *W-cycle multigrid method*. Other cycling strategies are possible but will not be discussed here.

For the multigrid V-cycle, the number $j+1$ of levels and the number $J+1$ of subspace corrections coincide. Compared with the V-cycle, additional subspace corrections are added in the W-cycle. The number of subspace corrections exceeds the number of levels, although each of the subspaces $\mathcal{W}_l, l = 0, \ldots, J$, is one of the spaces $\mathcal{S}_k, k = 0, \ldots, j$.

The reason for the extremely fast convergence of multigrid methods in comparison to the underlying smoothers is that these iterations are very selective for the different components of the error. Fast oscillating components (with respect to the given level) are strongly reduced whereas the remaining components are nearly unaffected. The error is *smoothed*, as the term 'smoothing step' indicates. As the smooth components of the error are already small because of the preceding coarse grid correction, the composed method can be very efficient.

For simple model problems (constant coefficients, square grids, periodic boundary conditions, etc.) the interaction of the smoothing steps and the coarse level corrections can be quantitatively studied using a *Fourier* or *local mode analysis*. We refer to Hackbusch's 1985 book, to Stüben and Trottenberg (1982) and to Brandt (1982).

For any good iterative solver for the solution of finite element equations, the amount of work per iteration step should be proportional to the number of unknowns. Next we check whether this condition is satisfied for the multigrid methods introduced earlier.

Without regarding the algorithmic realization in detail (recall only (2.26)), we get the recursion formula

$$W_j = p\, W_{j-1} + C n_j \tag{2.37}$$

for the work W_j necessary to perform one step of the multigrid method for the solution of a equation in $\mathcal{S}_j$. n_j denotes the dimension of $\mathcal{S}_j$. $p = 1$ corresponds to the V-cycle and $p = 2$ to the W-cycle. Here we have assumed that the amount of work per cycle, except for the approximate solution of the coarse level equation (2.34), behaves like Cn_j, which is the case for all reasonable smoothers. The recursion formula (2.37) yields

$$W_j = p^j W_0 + C \sum_{k=1}^{j} p^{j-k} n_k. \tag{2.38}$$

If one disregards the work for the solution of the equations on the level 0, a simple analysis shows that the operation count W_j for the single multigrid cycle behaves like $\mathcal{O}(n_j)$ if and only if the dimensions n_k are related by

$$n_k \le c q^{j-k} n_j, \quad k = 1, \ldots, j, \tag{2.39}$$

where $q < 1$ for the V-cycle and $q < 1/2$ for the W-cycle. This means that the dimensions of the spaces $\mathcal{S}_k$ have to increase geometrically. These conditions are satisfied for our model problem (where the dimension essentially grows by the factor 4 from one level to the next), but they can cause problems for adaptively generated, nonuniformly refined meshes.

An interesting modification of the classical multigrid methods, especially as it concerns the application to such adaptively generated, highly nonuniform meshes, is the *hierarchical basis multigrid method* introduced by Bank *et al.* (1988).

Compared with classical multigrid methods, it works with smaller spaces $\mathcal{W}_l$. Nevertheless, for two-dimensional problems, it reaches a similar efficiency as those of classical multigrid methods. Its structure fits very well to nonuniformly refined meshes and allows the use of simple data structures. Under some mild restrictions, the W-cycle version also works in three space dimensions as can be shown along the lines given in Bank and Dupont (1980), Braess (1981) and Axelsson and Vassilevski (1989).

To describe this method, we have to realize that a function in the finite element space $\mathcal{S}_k$ is uniquely determined by its values at the nodes $x \in \mathcal{N}_k$ which are the vertices of the triangles in the triangulation $\mathcal{T}_k$ which do not lie on the boundary of Ω. Therefore we can define an interpolation operator $\mathcal{I}_k : \mathcal{S} \to \mathcal{S}_k$ by

$$(\mathcal{I}_k u)(x) = u(x), \quad x \in \mathcal{N}_k. \tag{2.40}$$

Utilizing these interpolation operators, we can define the subspace

$$\mathcal{W}_k \;=\; \{\mathcal{I}_k u - \mathcal{I}_{k-1} u \,|\, u \in \mathcal{S}\} \tag{2.41}$$

of $\mathcal{S}_k$ as the image of $\mathcal{S}$ (or of $\mathcal{S}_k$) under the operator $\mathcal{I}_k - \mathcal{I}_{k-1}$. The functions in this space $\mathcal{W}_k$ vanish at the nodes $x \in \mathcal{N}_{k-1}$. Therefore they are given by their values at the nodes $x \in \mathcal{N}_k \setminus \mathcal{N}_{k-1}$.

In the hierarchical basis methods, the spaces $\mathcal{S}_k$ are replaced by the spaces (2.41). As with classical multigrid methods, approximate solvers B_k of a very simple structure can be used. For a survey, we refer to Yserentant (1992).

In comparison with the recursion

$$W_j = W_{j-1} + Cn_j \tag{2.42}$$

for the work necessary to perform one multigrid V-cycle, the corresponding recursion formula

$$W_j = W_{j-1} + C(n_j - n_{j-1}) \tag{2.43}$$

for the hierarchical basis multigrid method has a different quality. It yields the operation count

$$W_j \le W_0 + Cn_j \tag{2.44}$$

independently of any assumption on the distribution of the unknowns among the levels.

For nonuniformly refined families of grids, the only alternative to the hierarchical basis multigrid method are multigrid methods in which the spaces (2.41) are enriched by those basis functions of $\mathcal{S}_k$ which are associated with the nodes in $\mathcal{N}_{k-1}$ having neighbours in $\mathcal{N}_k \setminus \mathcal{N}_{k-1}$. From a computational point of view, this corresponds to *local smoothing procedures*. The theoretical understanding of such methods began with Bramble *et al.* (1991b).

The recursively defined multiplicative subspace correction methods can be seen as generalizations of the Gauß–Seidel method. The corresponding Jacobi-type iterations have recently been the focus of much interest. Because of their simpler structure, these *additive subspace correction methods*

$$\widetilde{u} \leftarrow \widetilde{u} + \sum_{l=0}^{J} B_l^{-1} Q_l (f - A\widetilde{u}) \tag{2.45}$$

offer many advantages as preconditioners for the conjugate gradient method.

With subspaces $\mathcal{W}_l$ as in the hierarchical basis multigrid method and Jacobi type methods as approximate solvers B_l, the additive subspace correction method for the solution of the finite element equations (2.11) becomes the hierarchical basis solver; see Yserentant (1986b, 1990, 1992). For the choice $\mathcal{W}_l = \mathcal{S}_l$, one obtains the multilevel nodal basis preconditioner of Bramble *et al.* (1990) and Xu (1989); see also Yserentant (1990).

3. An analysis for the two-level case

The first general convergence proofs for classical, recursively defined multigrid methods for finite element equations like (2.11) stem from the end of the 1970s. They are mainly the work of Wolfgang Hackbusch and of Randolph E. Bank and Todd Dupont; see Hackbusch (1981, 1985) and Bank and Dupont (1981).

It is Hackbusch's merit to have identified and clearly separated the two main building blocks which lay the foundation to all standard convergence proofs and which became the basis of a countless number of articles appearing until the present day. These properties are the *smoothing property*, which essentially describes the necessary relations between the approximate subspace solvers and the finite element equations, and the *approximation property*, which describes the interaction of the different levels. Both properties will be discussed in this section.

In this and the next section, we assume that the eigenvalues of the error propagation operators

$$I - \widehat{B}_k^{-1} A_k \tag{3.1}$$

of the smoothing iterations

$$w_k \leftarrow w_k + \widehat{B}_k^{-1}(f_k - A_k w_k) \tag{3.2}$$

are nonnegative. This condition is equivalent to

$$(w_k, A_k w_k) \le (w_k, \widehat{B}_k w_k), \quad w_k \in \mathcal{S}_k, \tag{3.3}$$

and guarantees the convergence of the iteration (3.2). The condition (3.3) is stronger than (2.33). (3.3) is our version of the smoothing property.

(3.3) holds, if the operators $\widehat{B}_k$ are properly scaled, i.e. if the iteration (3.2) is sufficiently damped. If (3.2) represents a symmetric (block) Gauß–Seidel iteration, (3.3) is automatically satisfied.

Our second assumption concerns the spaces

$$\mathcal{V}_k = \{P_k u - P_{k-1} u \,|\, u \in \mathcal{S}\} \subseteq \mathcal{S}_k, \tag{3.4}$$

i.e. the a-orthogonal complements of the spaces $\mathcal{S}_{k-1}$ in $\mathcal{S}_k$. We assume that there exists a constant K with

$$(v_k, \widehat{B}_k v_k) \le K(v_k, A_k v_k), \quad v_k \in \mathcal{V}_k. \tag{3.5}$$

(3.5) is the counterpart to (3.3). As the norms induced by the $\widehat{B}_k$ are generally much stronger than the energy norm induced by the A_k, (3.5) can be interpreted as an approximation property of the functions in $\mathcal{S}_k$ by functions from the subspace $\mathcal{S}_{k-1}$.

Without any doubt, (3.5) is a much more critical assumption than (3.3). In the finite element case, and especially for our model problem, (3.5) is essentially equivalent to the Aubin–Nitsche Lemma; see Ciarlet (1978), for example. Assume that, for $k = 1, \ldots, j$ and all $u_k \in \mathcal{S}_k$,

$$c_1 4^k \|u_k\|_0^2 \le (u_k, \widehat{B}_k u_k) \le c_2 4^k \|u_k\|_0^2. \tag{3.6}$$

Because of the scaling of the L_2-like norm (2.12) by the areas of the triangles in the initial triangulation and the fact that the diameter of the triangles shrinks by the factor 2 from one level to the next, smoothers like the Jacobi iteration or the symmetric point Gauß–Seidel iteration have this property. With (3.6), the condition (3.5) is equivalent to the estimate

$$\|P_k u - P_{k-1} u\|_0 \le c\, 2^{-k} \|P_k u - P_{k-1} u\| \tag{3.7}$$

for the functions $u \in \mathcal{S}$, or, with an infinite sequence of spaces $\mathcal{S}_k$, even equivalent to the estimate

$$\|u - P_k u\|_0 \le \tfrac{1}{3} c\, 2^{-k} \|u - P_k u\| \tag{3.8}$$

for the functions u in the continuous solution space $H_0^1(\Omega)$. (3.8) and (3.7), respectively, imply (3.5) if the upper estimate in (3.6) holds.

It is well known that (3.8) holds only for H^2-regular problems, i.e. if the

solution u of the continuous problem (2.6) belongs to H^2 for right-hand sides $f \in L_2$ and satisfies an estimate

$$\|u\|_{H^2} \leq \widetilde{c}\|f\|_{L_2}. \tag{3.9}$$

This holds only if the domain Ω has a C^2-boundary or if Ω is convex. For domains Ω with re-entrant corners, (3.9) is wrong.

This fact restricts the applicability of the classical multigrid convergence theory, although, using differently weighted L_2-norms (with weights depending on the interior angles of the domain) and properly refined triangulations, the algebraic estimates derived in this and the next section can also be applied to certain problems on domains with re-entrant corners; see Yserentant (1986a, 1983) and S. Zhang (1990).

The basic result of this section is an estimate for the convergence rate of the algorithm described by (2.34), (2.35) and (2.36), i.e. for the case in which the coarse grid correction is exactly determined and only two levels are present.

Theorem 3.1 If $u_0 \in \mathcal{S}$ is the initial approximation for the solution $u \in \mathcal{S}$ of the finite element equation $Au = f$, the new approximation $u_{m+1} \in \mathcal{S}$ (after a full two-grid cycle (2.34), (2.35), (2.36)) satisfies the estimate

$$\|u - u_{m+1}\|^2 \leq K\gamma(m)\|u - u_0\|^2, \tag{3.10}$$

where the generic constant $\gamma(m)$ is given by

$$\gamma(m) = \frac{1}{2m+1}\left(1 - \frac{1}{2m+1}\right)^{2m} \tag{3.11}$$

and K is the constant from the approximation property (3.5).

Proof. The main ingredient of the proof is a biorthogonal basis $\psi_1, \ldots, \psi_n$ of $\mathcal{S}$ with

$$(\psi_i, \widehat{B}\psi_l) = \delta_{il}, \quad (\psi_i, A\psi_l) = \lambda_i\delta_{il}, \tag{3.12}$$

where we suppress the subscript j for a while; the existence of such a basis follows from basic facts of linear algebra. Then, for $u = \sum_{i=1}^n a_i\psi_i \in \mathcal{S}$, the norm (2.1) is given by

$$\|u\|^2 = \sum_{i=1}^n \lambda_i a_i^2. \tag{3.13}$$

If we introduce the discrete norm

$$|\!|\!|u|\!|\!| = (u, \widehat{B}u)^{1/2}, \quad u \in \mathcal{S}, \tag{3.14}$$

this norm has the representation

$$|\!|\!|u|\!|\!|^2 = \sum_{i=1}^n a_i^2. \tag{3.15}$$

The proof of Theorem 3.1 is based on the eigenfunction expansion

$$Gv = \sum_{i=1}^{n}(1-\lambda_i)a_i\psi_i \tag{3.16}$$

for $v = \sum_{i=1}^{n} a_i\psi_i$ of the error propagation operator

$$G = I - \widehat{B}^{-1}A \tag{3.17}$$

for the smoothing process. Introducing the errors

$$e_k = u - u_k, \quad k = 0, \dots, m+1, \tag{3.18}$$

between the exact solution $u \in \mathcal{S}$ and the intermediate approximations $u_0, u_1, \dots, u_{m+1}$ of u, the errors $e_2, \dots, e_m$ can be expressed as

$$e_{k+1} = G^k e_1, \quad k = 0, \dots, m, \tag{3.19}$$

in terms of the error e_1 after the coarse grid correction.

If $e_1 = \sum_{i=1}^{n} a_i\psi_i$, one finds

$$\|e_{m+1}\|^2 = \|\sum_{i=1}^{n}(1-\lambda_i)^m a_i\psi_i\|^2 = \sum_{i=1}^{n}\lambda_i(1-\lambda_i)^{2m}a_i^2. \tag{3.20}$$

The smoothing property (3.3) is equivalent to the bound

$$\lambda_i \le 1, \quad i = 1, \dots, n, \tag{3.21}$$

for the eigenvalues λ_i. Therefore

$$\max_{i=1,\dots,n}\lambda_i(1-\lambda_i)^{2m} \le \max_{0\le\lambda\le1}\lambda(1-\lambda)^{2m} = \gamma(m) \tag{3.22}$$

and

$$\|e_{m+1}\|^2 \le \gamma(m)\sum_{i=1}^{n}a_i^2 = \gamma(m)|\!|\!|e_1|\!|\!|^2. \tag{3.23}$$

As, by the approximation property (3.5),

$$|\!|\!|e_1|\!|\!|^2 = |\!|\!|e_0 - P_{j-1}e_0|\!|\!|^2 \le K\|e_0 - P_{j-1}e_0\|^2 \le K\|e_0\|^2, \tag{3.24}$$

the proposition

$$\|e_{m+1}\|^2 \le K\gamma(m)\|e_0\|^2 \tag{3.25}$$

of Theorem 3.1 follows. □

The theorem states that the two-grid method converges and that its convergence rate $K\gamma(m)$ becomes even arbitrarily small as soon as the number m of smoothing steps is sufficiently large. As the constant K in the approximation property (3.5) does not depend on j, the convergence rate is independent of the grid size.

This fact can be utilized to prove the convergence of the W-cycle by a relatively simple recursion argument, which is the idea behind all early multigrid convergence proofs. A detailed discussion can be found in Hackbusch's 1985 book or in Stüben and Trottenberg (1982).

Although the two-grid–multigrid analysis is very suggestive and has a broad range of application, the results obtained in this way do not completely satisfy for the given case of self-adjoint, coercive elliptic boundary value problems. Experience says that, applied to problems of this class, not only the W-cycle but also the much simpler and cheaper V-cycle converges very fast. In addition, *one* smoothing step per level turns out to be sufficient.

4. A convergence proof for the V-cycle

The much more sophisticated convergence analysis of Braess and Hackbusch (1983) supports these observations theoretically. In this section, we derive their result in an algebraic language as in Yserentant (1983). Closely related estimates are proven in Bank and Douglas (1985). The assumptions in this section are the same as in the previous section.

Following the original work and contrary to the definition given in Section 2, in this section we assume that the order of the coarse grid correction (2.34), (2.35) and of the smoothing steps (2.36) is reversed. This is no essential change because it is easy to see that the convergence rates of both versions are equal. The order that we chose in Section 2, seems to be more natural from the point of view of subspace correction methods. This version will be analysed in Section 6.

Theorem 4.1 If $u \in \mathcal{S}$ denotes the exact solution of the equation to be solved and if $u_0 \in \mathcal{S}$ is the given initial approximation of u, the new approximation $u_{m+1} \in \mathcal{S}$, obtained by a multigrid V-cycle or W-cycle, satisfies the estimate

$$\|u - u_{m+1}\|^2 \;\leq\; \frac{c}{c+2m}\,\|u - u_0\|^2, \tag{4.1}$$

where $c = K^2$ and K is the constant from the approximation property (3.5).

As in the two-level proof of the previous section, the proof is based on the eigenfunction expansion (3.16) of the error propagation operator G for the smoothing process which is given by (3.17).

As the eigenvalues λ_i are not greater than 1, we have $1 - \lambda_i \geq 0$ for all i and can define the powers G^α, $\alpha \geq 0$, of G by

$$G^\alpha v \;=\; \sum_{i=1}^{n} (1-\lambda_i)^\alpha a_i \psi_i, \tag{4.2}$$

where $v = \sum_{i=1}^{n} a_i \psi_i$ is the eigenfunction expansion of the function $v \in \mathcal{S}$.

The main ingredient of the proof is the functional

$$\rho(v) = \begin{cases} \|G^{1/2}v\|^2/\|v\|^2 & , \quad v \neq 0 \\ 0 & , \quad v = 0 \end{cases} \tag{4.3}$$

for the elements $v \in \mathcal{S}$. Note that always $\rho(v) \leq 1$. $\rho(v)$ can be seen as a measure for the smoothness of v. If $\rho(v)$ is small compared with 1, the smoothed element Gv has a small norm compared with v.

Our first lemma describes the success of the coarse grid correction in terms of this kind of smoothness of the error.

Lemma 4.2 For all functions $v \in \mathcal{S}$, $v - P_{j-1}v$ satisfies the estimate

$$\|v - P_{j-1}v\|^2 \;\leq\; \min\{1, K(1-\rho(v))\}\,\|v\|^2. \tag{4.4}$$

Proof. Let $v \in \mathcal{S}$ and $v - P_{j-1}v$ have the eigenfunction expansions

$$v = \sum_{i=1}^{n} a_i\psi_i, \quad v - P_{j-1}v = \sum_{i=1}^{n} b_i\psi_i.$$

Then

$$\|v - P_{j-1}v\|^2 = a(v - P_{j-1}v, v) = \sum_{i=1}^{n} \lambda_i a_i b_i.$$

With the Schwarz inequality, this yields

$$\begin{aligned} \|v - P_{j-1}v\|^2 \;&\leq\; \Big(\sum_{i=1}^{n} b_i^2\Big)^{1/2}\Big(\sum_{i=1}^{n} \lambda_i^2 a_i^2\Big)^{1/2} \\ &= \Big(\sum_{i=1}^{n} b_i^2\Big)^{1/2}\Big\{\sum_{i=1}^{n} \lambda_i a_i^2 - \sum_{i=1}^{n} \lambda_i(1-\lambda_i)a_i^2\Big\}^{1/2} \\ &= \;\; |\!|\!| v - P_{j-1}v |\!|\!|\,\{\,\|v\|^2 - \|G^{1/2}v\|^2\,\}^{1/2} \\ &= \;\; |\!|\!| v - P_{j-1}v |\!|\!|\,\{1-\rho(v)\}^{1/2}\,\|v\|. \end{aligned}$$

Inserting (3.5), this means

$$|\!|\!| v - P_{j-1}v |\!|\!|^2 \leq K\|v - P_{j-1}v\|^2,$$

one obtains

$$\|v - P_{j-1}v\|^2 \leq K(1-\rho(v))\,\|v\|^2.$$

This proves the proposition. □

Together with the next lemma describing the effect of the smoothing iterations, Lemma 4.2 forms the backbone of the proof of Theorem 4.1.

Lemma 4.3 For all functions $v \in \mathcal{S}$, $G^k v$ satisfies the estimate

$$\|G^k v\| \;\leq\; \rho(G^k v)^k \|v\|. \tag{4.5}$$

Proof. Let $\mu_i = 1 - \lambda_i$. Because of $\mu_i \geq 0$ and utilizing Hölder's inequality, one obtains, for all $v = \sum_{i=1}^{n} a_i \psi_i$,

$$\begin{aligned}
\|G^k v\|^2 &= \sum_{i=1}^{n} \lambda_i (\mu_i^k a_i)^2 \\
&= \sum_{i=1}^{n} (\lambda_i \mu_i^{2k+1} a_i^2)^{\frac{2k}{2k+1}} (\lambda_i a_i^2)^{\frac{1}{2k+1}} \\
&\leq \Big(\sum_{i=1}^{n} \lambda_i \mu_i^{2k+1} a_i^2 \Big)^{\frac{2k}{2k+1}} \Big(\sum_{i=1}^{n} \lambda_i a_i^2 \Big)^{\frac{1}{2k+1}} \\
&= \|G^{k+1/2} v\|^{\frac{4k}{2k+1}} \|v\|^{\frac{2}{2k+1}}.
\end{aligned}$$

This estimate is equivalent to

$$\|G^k v\| \|G^k v\|^{2k} \leq \|G^{1/2}(G^k v)\|^{2k} \|v\|.$$

This is the proposition. □

Now we can prove the theorem. Denoting by $d = P_{j-1} e_m$ the exact coarse grid correction and by $\widetilde{d} \in \mathcal{S}_{j-1}$ the approximate coarse grid correction computed by p steps of the method for the level $j-1$, one obtains, utilizing

$$e_{m+1} = (e_m - P_{j-1} e_m) + (d - \widetilde{d}), \tag{4.6}$$

the relation

$$\|e_{m+1}\|^2 = \|e_m - P_{j-1} e_m\|^2 + \|d - \widetilde{d}\|^2. \tag{4.7}$$

With an upper bound δ_{j-1} for the convergence rate of the method on the preceding level $j-1$,

$$\|e_{m+1}\|^2 \leq \|e_m - P_{j-1} e_m\|^2 + \delta_{j-1}^{2p} \|P_{j-1} e_m\|^2 \tag{4.8}$$

follows. This equation can be rewritten as

$$\|e_{m+1}\|^2 \leq (1 - \delta_{j-1}^{2p}) \|e_m - P_{j-1} e_m\|^2 + \delta_{j-1}^{2p} \|e_m\|^2. \tag{4.9}$$

Lemma 4.2 yields

$$\|e_m - P_{j-1} e_m\|^2 \leq \min\{1, K(1 - \rho(e_m))\} \|e_m\|^2, \tag{4.10}$$

and Lemma 4.3

$$\|e_m\|^2 \leq \rho(e_m)^{2m} \|e_0\|^2. \tag{4.11}$$

If we insert these relations, we have proven the estimate

$$\|u - u_{m+1}\|^2 \leq \delta_j^2 \|u - u_0\|^2, \tag{4.12}$$

where δ_j is given by the relatively complicated expression

$$\delta_j^2 = \max_{0 \leq \rho \leq 1} \rho^{2m} [(1 - \delta_{j-1}^{2p}) \min\{1, K(1 - \rho)\} + \delta_{j-1}^{2p}]. \tag{4.13}$$

Together with

$$\delta_0 = 0, \tag{4.14}$$

this recursion leads to an estimate for the convergence rates δ_j.

For a fixed $\rho \in [0,1]$, the function

$$\varepsilon \to \rho^{2m}[(1-\varepsilon)\min\{1, K(1-\rho)\} + \varepsilon] \tag{4.15}$$

increases monotonically. Introducing the abbreviation $c(m) = c/(c+2m)$, the assumption

$$\delta_{j-1}^2 \le c(m) \tag{4.16}$$

leads therefore to

$$\delta_j^2 \le \max_{0\le\rho\le 1} R(\rho), \tag{4.17}$$

where the function $R(\rho)$ is given by

$$R(\rho) \;=\; \rho^{2m}[(1-c(m))\min\{1, K(1-\rho)\} + c(m)]. \tag{4.18}$$

As necessarily $K \ge 1$, $R(\rho)$ is monotonically increasing on the interval $[0,1]$. This proves the estimate

$$\delta_j^2 \le R(1) = c(m) \tag{4.19}$$

for the convergence rate of the multigrid method.

Later, in Wittum (1989a), Theorem 4.1 has been generalized to the case that the error propagation operator (3.1) can have negative eigenvalues. Both the analysis of Braess and Hackbusch and that of Wittum do not take the internal structure of the smoothers into account. A result refined in this respect has been proven by Stevenson (1992). Reusken (1992) examined the convergence of multigrid methods with respect to the maximum norm. He shows that, up to a logarithmic factor, one can obtain the same convergence estimates as for the energy norm studied here. Another convergence proof, based on projection arguments and norm estimates instead of eigenfunction expansions, can be found in Mandel, McCormick and Bank (1987).

5. General multiplicative methods

A main drawback of all these approaches is their strong dependence on the regularity properties of the boundary value problem and of the considered family of grids which is reflected in the assumption (3.5). This fact makes it extremely difficult to apply these theories in a rigorous sense to problems with singularities caused by re-entrant corners, jumps in the boundary conditions, by interfaces, and so on.

Recently Bramble *et al.* (1991a, 1991b) developed an alternative convergence theory which overcomes these difficulties to a large extent.

This theory can be formulated in the abstract framework of the multiplicative subspace correction methods introduced in Section 2. We remark that the case of non overlapping subspaces $\mathcal{W}_k$ (which covers the hierarchical basis multigrid method) has implicitly been treated in Bank *et al.* (1988).

With some slight modifications as presented in Bramble and Pasciak (1991), Xu (1992b), or in the present paper, the theory shows that the convergence rate of multigrid methods is uniformly bounded independently of any regularity of the boundary value under consideration. It does not show that the convergence rate tends to zero if one increases the number of smoothing steps per level.

In this section, we develop the abstract theory for the multiplicative subspace correction methods introduced in Section 2. The application to our model problem and to other elliptic boundary value problems will be discussed in the next section.

The theory is based on the decomposition of the space $\mathcal{S}$ into a direct sum

$$\mathcal{S} = \mathcal{V}_0 \oplus \mathcal{V}_1 \oplus \ldots \oplus \mathcal{V}_J \tag{5.1}$$

of subspaces $\mathcal{V}_k \subseteq \mathcal{W}_k$. These subspaces $\mathcal{V}_k$ are only a tool for the theoretical analysis, they do not enter the practical computation. Often, this fact gives a lot of freedom in the choice of these subspaces and makes the convergence theory very flexible.

Two assumptions have to be fulfilled to apply the theory. The first assumption concerns the *stability of the decomposition*. We require that there exists a constant K_1 such that, for all $v_k \in \mathcal{V}_k$,

$$\sum_{k=0}^{J} (B_k v_k, v_k) \;\leq\; K_1 \,\| \sum_{k=0}^{J} v_k \|^2. \tag{5.2}$$

The second assumption is a *Cauchy–Schwarz type inequality*. We assume that there exist constants $\gamma_{kl} = \gamma_{lk}$ with

$$a(w_k, v_l) \;\leq\; \gamma_{kl}\, (B_k w_k, w_k)^{1/2} (B_l v_l, v_l)^{1/2} \tag{5.3}$$

for $k \leq l$, all $w_k \in \mathcal{W}_k$, and all $v_l \in \mathcal{V}_l$ such that

$$\sum_{k,l=0}^{J} \gamma_{kl} x_k y_l \;\leq\; K_2 (\sum_{k=0}^{J} x_k^2)^{1/2} (\sum_{l=0}^{J} y_l^2)^{1/2} \tag{5.4}$$

holds for all $x_k, y_l \in \mathbb{R}$. That means, we require that the spectral radius of the matrix (γ_{kl}) is bounded by a constant K_2.

In addition, we assume that the constant ω in (2.32) satisfies the condition

$$\omega < 2 \tag{5.5}$$

which is equivalent to the convergence of the basic iterations (2.31).

(5.3) includes the Cauchy–Schwarz type inequality

$$a(v_k, v_l) \;\le\; \gamma_{kl}\,(B_k v_k, v_k)^{1/2} (B_l v_l, v_l)^{1/2} \tag{5.6}$$

for $v_k \in \mathcal{V}_k, v_l \in \mathcal{V}_l$, and $k, l = 0, \dots, J$. (5.6) and (5.4) imply that, corresponding to (5.2),

$$\| \sum_{k=0}^{J} v_k \|^2 \;\le\; K_2 \sum_{k=0}^{J} (B_k v_k, v_k) \tag{5.7}$$

for all $v_k \in \mathcal{V}_k$. Therefore the expression

$$\| \!| \sum_{k=0}^{J} v_k | \!\|^2 \;=\; \sum_{k=0}^{J} (B_k v_k, v_k) \tag{5.8}$$

defines a norm on $\mathcal{S}$ which is, up to the constants K_1 and K_2, equivalent to the norm (2.1) induced by the abstract boundary value problem itself.

With the orthogonal projections P_k onto the spaces $\mathcal{W}_k$, the exact subspace corrections (2.16) are

$$\widetilde{u} \;\leftarrow\; \widetilde{u} + P_k (u - \widetilde{u}). \tag{5.9}$$

If we define the operators

$$T_k := B_k^{-1} A_k P_k = B_k^{-1} Q_k A, \tag{5.10}$$

the approximate subspace corrections (2.23) are correspondingly given by

$$\widetilde{u} \;\leftarrow\; \widetilde{u} + T_k (u - \widetilde{u}). \tag{5.11}$$

After the substep (2.23), the new error is

$$\widetilde{u} - u \;\leftarrow\; (I - T_k)(\widetilde{u} - u). \tag{5.12}$$

Thus the convergence rate of the multiplicative subspace correction method with respect to the norm (2.1) is the induced norm of the operator

$$E = (I - T_J) \dots (I - T_0). \tag{5.13}$$

Theorem 5.1 Every cycle of the abstract multiplicative subspace correction method introduced in Section 2 reduces the norm (2.1) of the error at least by the factor $\|E\|$ where

$$\|E\|^2 \;\le\; 1 - \frac{2 - \omega}{K_1 (1 + K_2)^2}. \tag{5.14}$$

This factor depends only on the constant K_1 from the stability assumption (5.2), on the constant K_2 from (5.4), and on the constant $\omega < 2$ from equation (2.32).

There are several, in principle, very closely related versions of this theorem

in the papers of Bramble, Pasciak, Wang and Xu. The present version bears at most resemblance to that of Xu (1992b).

The proof of the theorem is somewhat technical. It is not so easy to detect the idea hidden behind it except that the terms considered are cleverly arranged and split up. The following two lemmas are the main tools:

Lemma 5.2 For all $v_k \in \mathcal{V}_k$ and all $u_k \in \mathcal{S}$ (!),

$$\sum_{k=0}^{J} a(v_k, u_k) \;\le\; \sqrt{K_1}\,\|\sum_{k=0}^{J} v_k\| \Big(\sum_{k=0}^{J} a(T_k u_k, u_k)\Big)^{1/2}. \tag{5.15}$$

Proof. One has

$$\begin{aligned}
\sum_{k=0}^{J} a(v_k, u_k) \;&=\; \sum_{k=0}^{J} (B_k^{1/2} v_k, B_k^{-1/2} A_k P_k u_k) \\
&\le\; \Big(\sum_{k=0}^{J} \|B_k^{1/2} v_k\|_0^2\Big)^{1/2} \Big(\sum_{k=0}^{J} \|B_k^{-1/2} A_k P_k u_k\|_0^2\Big)^{1/2} \\
&=\; \Big(\sum_{k=0}^{J} (B_k v_k, v_k)\Big)^{1/2} \Big(\sum_{k=0}^{J} (B_k^{-1} A_k P_k u_k, A_k P_k u_k)\Big)^{1/2} \\
&=\; \Big(\sum_{k=0}^{J} (B_k v_k, v_k)\Big)^{1/2} \Big(\sum_{k=0}^{J} a(T_k u_k, u_k)\Big)^{1/2}.
\end{aligned}$$

With the stability assumption (5.2), the proposition follows. □

Lemma 5.3 For all $u \in \mathcal{S}$,

$$\|T_k u\|^2 \;\le\; \omega\, a(T_k u, u). \tag{5.16}$$

Proof. Because

$$\begin{aligned}
\|T_k u\|^2 &= (T_k u, A_k T_k u) \le \omega\,(T_k u, B_k T_k u) \\
&= \omega\,(T_k u, B_k B_k^{-1} A_k P_k u) = \omega\, a(T_k u, P_k u) = \omega\, a(T_k u, u),
\end{aligned}$$

the proposition is a simple consequence of (2.32). □

Now we are ready to prove Theorem 5.1. The proposition (5.14) is equivalent to the estimate

$$(2-\omega)\|v\|^2 \;\le\; K_1(1+K_2)^2(\|v\|^2 - \|Ev\|^2) \tag{5.17}$$

for all $v \in \mathcal{S}$. With $E_{-1} = I$ and

$$E_k = (I - T_k)\ldots(I - T_0), \quad k = 1, \ldots J,$$

one obtains

$$\|E_{k-1}v\|^2 - \|E_kv\|^2 \;=\; 2\,a(T_kE_{k-1}v, E_{k-1}v) - \|T_kE_{k-1}v\|^2.$$

With Lemma 5.3,

$$\|E_{k-1}v\|^2 - \|E_kv\|^2 \;\geq\; (2-\omega)\,a(T_kE_{k-1}v, E_{k-1}v)$$

follows. Because of $E_J = E$, summation yields

$$\|v\|^2 - \|Ev\|^2 \;\geq\; (2-\omega)\sum_{k=0}^{J} a(T_kE_{k-1}v, E_{k-1}v).$$

Because $\omega < 2$, (5.17) therefore follows from

$$\|v\|^2 \;\leq\; K_1(1+K_2)^2\sum_{k=0}^{J} a(T_kE_{k-1}v, E_{k-1}v). \tag{5.18}$$

For the proof of (5.18), let

$$v = \sum_{l=0}^{J} v_l, \quad v_l \in \mathcal{V}_l.$$

Then

$$\|v\|^2 \;=\; \sum_{l=0}^{J} a(E_{l-1}v, v_l) + \sum_{l=1}^{J} a((I-E_{l-1})v, v_l). \tag{5.19}$$

By Lemma 5.2,

$$\sum_{l=0}^{J} a(E_{l-1}v, v_l) \;\leq\; \sqrt{K_1}\,\|v\|\,\Big(\sum_{k=0}^{J} a(T_kE_{k-1}v, E_{k-1}v)\Big)^{1/2}. \tag{5.20}$$

For the second term on the right-hand side of (5.19), because

$$I - E_{l-1} = \sum_{k=0}^{l-1} T_kE_{k-1},$$

and utilizing (5.3), one gets the estimate

$$\begin{aligned}
\sum_{l=1}^{J} a((I-E_{l-1})v, v_l) &= \sum_{l=1}^{J}\sum_{k=0}^{l-1} a(T_k E_{k-1} v, v_l) \\
&\le \sum_{l=1}^{J}\sum_{k=0}^{l-1} \gamma_{kl}\,(B_k T_k E_{k-1} v, T_k E_{k-1} v)^{1/2} (B_l v_l, v_l)^{1/2} \\
&\le \sum_{l=0}^{J}\sum_{k=0}^{J} \gamma_{kl}\,(B_k T_k E_{k-1} v, T_k E_{k-1} v)^{1/2} (B_l v_l, v_l)^{1/2} \qquad (5.21) \\
&\le K_2 \Big(\sum_{k=0}^{J} (B_k T_k E_{k-1} v, T_k E_{k-1} v)\Big)^{1/2} \Big(\sum_{l=0}^{J} (B_l v_l, v_l)\Big)^{1/2} \\
&= K_2 \Big(\sum_{k=0}^{J} (B_k v_k, v_k)\Big)^{1/2} \Big(\sum_{k=0}^{J} a(T_k E_{k-1} v, E_{k-1} v)\Big)^{1/2}.
\end{aligned}$$

By the assumption (5.2),

$$\sum_{l=1}^{J} a((I-E_{l-1})v, v_l) \le K_2 \sqrt{K_1}\, \|v\| \Big(\sum_{k=0}^{J} a(T_k E_{k-1} v, E_{k-1} v)\Big)^{1/2} \qquad (5.22)$$

follows. Combined with (5.20), one obtains (5.18). This finishes the proof of Theorem 5.1.

Sometimes (for multigrid methods with simple smoothers, for example) it is possible to prove the stronger estimate

$$a(w_k, w_l') \le \gamma_{kl}\,(B_k w_k, w_k)^{1/2} (B_l w_l', w_l')^{1/2} \qquad (5.23)$$

for $w_k \in \mathcal{W}_k, w_l' \in \mathcal{W}_l, k, l = 0, \ldots, J$, which implies (5.3). In this case, the estimate in Theorem 5.1 holds *independently of the order in which the subspace corrections are performed.*

There are situations in which the proof of the Cauchy–Schwarz type inequality (5.3) causes problems, especially if the coefficients functions of the differential operator under consideration are not smooth or even not differentiable. In such cases, it is often still possible to prove the norm estimate (5.7), provided that the energy norm (2.1) behaves like the energy norm induced by a boundary value problem for which one can prove estimates like (5.3), (5.4). The norm estimate (5.7) is sufficient to derive an estimate for the norm of the error propagation operator (5.13) which does not deteriorate too rapidly in terms of the number of subspaces $\mathcal{W}_k$.

Theorem 5.4 Assuming only (5.7) instead of (5.3) and (5.4), the norm of the error propagation operator (5.13) satisfies the estimate

$$\|E\|^2 \le 1 - \frac{2-\omega}{K_1(1+\sqrt{\omega K_2 J}\,)^2}. \qquad (5.24)$$

Proof. Instead of (5.21), one obtains

$$\begin{aligned}\sum_{l=1}^{J} a((I-E_{l-1})v,v_l) &= \sum_{l=1}^{J}\sum_{k=0}^{l-1} a(T_kE_{k-1}v,v_l)\\ &= \sum_{k=0}^{J-1}\sum_{l=k+1}^{J} a(T_kE_{k-1}v,v_l)\\ &\le \Big(\sum_{k=0}^{J-1}\|T_kE_{k-1}v\|^2\Big)^{1/2}\Big(\sum_{k=0}^{J-1}\|\sum_{l=k+1}^{J} v_l\|^2\Big)^{1/2}.\end{aligned}$$

For the first factor on the right-hand side, one gets, by Lemma 5.3,

$$\sum_{k=0}^{J-1}\|T_kE_{k-1}v\|^2 \le \omega\sum_{k=0}^{J} a(T_kE_{k-1}v,E_{k-1}v).$$

Using (5.7), the second factor can be estimated as follows.

$$\begin{aligned}\sum_{k=0}^{J-1}\|\sum_{l=k+1}^{J} v_l\|^2 &\le K_2\sum_{k=0}^{J-1}\sum_{l=k+1}^{J}(B_lv_l,v_l)\\ &\le K_2J\sum_{l=0}^{J}(B_lv_l,v_l) \le K_1K_2J\|v\|^2.\end{aligned}$$

This yields the estimate

$$\sum_{l=1}^{J} a((I-E_{l-1})v,v_l) \le \sqrt{\omega K_1K_2J}\,\|v\|\Big(\sum_{k=0}^{J} a(T_kE_{k-1}v,E_{k-1}v)\Big)^{1/2}$$

which replaces (5.22). □

6. The application to multilevel algorithms

In this section we apply the abstract theory presented in the last section to the model problem of Section 2. We prove convergence results for the multigrid methods introduced there.

We begin with the classical multigrid method where the subspaces $\mathcal{W}_k$ are the finite element spaces $\mathcal{S}_k$. In the notation of the previous section, the error propagation operator of the V-cycle is

$$E_V = E_V^{(j)}, \quad E_V^{(k)} = (I-T_k)\dots(I-T_0). \tag{6.1}$$

Because $A_0 = B_0$, i.e. $T_0 = P_0$, the $E_V^{(k)}$ satisfy the recursion

$$E_V^{(0)} = I - P_0, \quad E_V^{(k+1)} = (I-T_{k+1})E_V^{(k)}. \tag{6.2}$$

The corresponding recursion for the W-cycle version is

$$E_W^{(0)} = I - P_0, \quad E_W^{(k+1)} = (I-T_{k+1})E_W^{(k)}E_W^{(k)}. \tag{6.3}$$

It follows by induction that

$$E_W^{(k)} = E_V^{(k)} R_W^{(k)}, \quad \|R_W^{(k)}\| \le 1. \tag{6.4}$$

Therefore one gets, for the energy norm of the error propagation operator $E_W = E_W^{(j)}$ of the W-cycle,

$$\|E_W\| \le \|E_V\|. \tag{6.5}$$

Moreover, if the order of the coarse grid corrections and the smoothing steps is again reversed, the W-cycle always reduces the energy norm of the error by at least the same factor as the V-cycle.

For the analysis of the V-cycle multigrid method, we assume that, for $k \ge 1$, the operators $B_k : \mathcal{S}_k \to \mathcal{S}_k$, the smoothers, satisfy the estimate

$$c_1\|u_k\|^2 \le (u_k, B_k u_k) \le c_2 4^k \|u_k\|_0^2 \tag{6.6}$$

for all $u_k \in \mathcal{S}_k$. This condition is less restrictive than (3.6) and also covers certain symmetric block Gauß–Seidel schemes, for example.

The crucial point for the application of Theorem 5.1 is the choice of the spaces $\mathcal{V}_k \subseteq \mathcal{S}_k$. Recall that these subspaces do not enter into the computational process.

The most obvious choice is the a-orthogonal decomposition of $\mathcal{S}$, i.e. the decomposition of $\mathcal{S}$ into $\mathcal{V}_0 = \mathcal{S}_0$ and

$$\mathcal{V}_k = \{P_k u - P_{k-1} u \,|\, u \in \mathcal{S}\} \subseteq \mathcal{S}_k \tag{6.7}$$

for $k = 1, \ldots, j$. As it has already been discussed in Section 3, for H^2-regular problems,

$$4^k\|v_k\|_0^2 \;\le\; C\|v_k\|^2, \quad v_k \in \mathcal{V}_k \tag{6.8}$$

holds. Because

$$\|P_0 u\|^2 + \sum_{k=1}^{j} \|P_k u - P_{k-1} u\|^2 \;=\; \|u\|^2 \tag{6.9}$$

for the functions $u \in \mathcal{S}$, (6.8) is equivalent to

$$\|v_0\|^2 + \sum_{k=1}^{j} 4^k \|v_k\|_0^2 \;\le\; C\,\|\sum_{k=0}^{j} v_k\|^2, \quad v_k \in \mathcal{V}_k. \tag{6.10}$$

With assumption (6.6), this yields (5.2), i.e.

$$\sum_{k=0}^{j} (B_k v_k, v_k) \;\le\; K_1\,\|\sum_{k=0}^{j} v_k\|^2. \tag{6.11}$$

The Cauchy–Schwarz type inequality (5.3) is trivial because

$$a(w_k, v_l) = 0, \quad w_k \in \mathcal{W}_k,\ v_l \in \mathcal{V}_l\ (k < l).$$

Thus we have shown that, for H^2-regular problems, every V-cycle (and every W-cycle) reduces the energy norm of the error at least by a factor

$$1 - \mathcal{O}(1)$$

which is uniformly less than 1 regardless of the number of refinement levels. This has already been proven in Wittum (1989a) and, with the restriction that the smoothing iterations are sufficiently damped, in Braess and Hackbusch (1983) and Bank and Douglas (1985); see Section 4. Therefore this estimate is surely not the most spectacular application of the abstract theory developed in the last section.

But before we discuss other choices of the spaces $\mathcal{V}_k$ leading to improved convergence estimates, we turn to the hierarchical basis multigrid method also described in Section 2.

For the hierarchical basis multigrid method, the finite element space $\mathcal{S}$ is already the direct sum of the spaces $\mathcal{W}_0 = \mathcal{S}_0$ and

$$\mathcal{W}_k \;=\; \{\mathcal{I}_k u - \mathcal{I}_{k-1} u \,|\, u \in \mathcal{S}\}, \quad k = 1, \ldots, j, \tag{6.12}$$

introduced in Section 2. Therefore the only possible choice for the subspaces $\mathcal{V}_k$ here are the spaces $\mathcal{W}_k$ itself.

It has been shown by Yserentant (1986b) that the decomposition of $\mathcal{S}$ into these spaces $\mathcal{V}_k$ is stable in the sense that, for all $v_k \in \mathcal{V}_k$,

$$\|v_0\|^2 + \sum_{k=1}^{j} 4^k \|v_k\|_0^2 \;\le\; C_1 (j+1)^2 \, \| \sum_{k=0}^{j} v_k \|^2. \tag{6.13}$$

If we assume that the level 0 equations are again solved exactly, i.e. that $B_0 = A_0$, and that, for $k \ge 1$, the operators $B_k : \mathcal{W}_k \to \mathcal{W}_k$ satisfy an estimate

$$c_1 4^k \|w_k\|_0^2 \le (w_k, B_k w_k) \le c_2 4^k \|w_k\|_0^2 \tag{6.14}$$

for all $w_k \in \mathcal{W}_k$, a stability condition like (5.2), namely

$$\sum_{k=0}^{j} (B_k v_k, v_k) \;\le\; K_1^* (j+1)^2 \, \| \sum_{k=0}^{j} v_k \|^2. \tag{6.15}$$

follows. The constant $K_1 = K_1^*(j+1)^2$ depends here on the number j of refinement levels.

The proof of (6.13) is based on the estimate

$$\|\mathcal{I}_k u\|^2 \le C(j-k+1)\|u\|^2, \quad u \in \mathcal{S}, \tag{6.16}$$

for the energy norm of the interpolation operators $\mathcal{I}_k : \mathcal{S} \to \mathcal{S}_k$. On one hand, this is a very robust estimate which is not affected by arbitrarily large jumps in the coefficient functions across the boundaries of the triangles in the initial triangulation. Unfortunately, on the other hand, it is dimension

dependent. For three space dimensions, the logarithmic factor has to be replaced by a factor which grows exponentially in the number $j-k$ of the remaining refinement levels. Details can be found in Yserentant (1986b, 1992) and Bank *et al.* (1988).

The Cauchy–Schwarz type inequality (5.3) follows from (6.14) and Lemma 6.1 which has essentially been proven in Yserentant (1986b). Related results can be found in Xu (1992b), Bramble and Pasciak (1991), and in X. Zhang (1991).

Lemma 6.1 There is a constant C, depending only on the constants in (2.10) describing the ellipticity of the boundary value problem, on the variation of the coefficient functions, and on the shape regularity of the triangles, such that, for $k \le l$ and all functions $u \in \mathcal{S}_k$ and $v \in \mathcal{S}_l$,

$$a(u,v) \;\le\; C\,(\frac{1}{\sqrt{2}})^{l-k}\,\|u\|\,2^l\|v\|_0. \tag{6.17}$$

Proof. For $l > k+1$, we fix a triangle $T \in \mathcal{T}_k$ and prove the local estimate

$$a(u,v)|_T \;\le\; C\,(\frac{1}{\sqrt{2}})^{l-k}\,|u|_{1;T}\,2^l\|v\|_{0;T}. \tag{6.18}$$

This estimate implies the global estimate (6.17). The basic idea is to split v into the function $v_0 \in \mathcal{S}_l$ given by

$$v_0(x) = \begin{cases} v(x) & , \quad x \in \mathcal{N}_l \cap \partial T \\ 0 & , \quad x \in \mathcal{N}_l \setminus \partial T \end{cases}$$

and into $v_1 = v - v_0$. Then the inner product $a(u,v)|_T$ can be written as

$$a(u,v)|_T = a(u,v_0)|_T + a(u,v_1)|_T.$$

The essential point is that v_1 vanishes on the boundary of T. Therefore we obtain, by partial integration and the product rule,

$$a(u,v_1)|_T \;=\; -\sum_{i,j=1}^{2}\int_T D_j a_{ij} D_i u\, v_1\,\mathrm{d}x \;-\; \sum_{i,j=1}^{2}\int_T a_{ij} D_j D_i u\, v_1\,\mathrm{d}x. \tag{6.19}$$

As u is linear on T, the second term on the right-hand side of the equation (6.19) vanishes. Assuming $T \subseteq T' \in \mathcal{T}_0$ and

$$|(D_j a_{ij})(x)| \le M_1 \mathrm{diam}(\mathrm{T}')^{-1}, \quad x \in T',$$

the first term on the right-hand side of equation (6.19) and therefore $a(u,v_1)$ can be estimated to be

$$a(u,v_1)|_T \;\le\; c_1|u|_{1;T}\|v_1\|_{0;T}. \tag{6.20}$$

The function v_0 vanishes outside a boundary strip S of T with

$$\frac{\mathrm{area}(S)}{\mathrm{area}(T)} = 1 - (1 - 3(\tfrac{1}{2})^{l-k})^2 \le 6(\tfrac{1}{2})^{l-k}.$$

Therefore

$$a(u, v_0)|_T \leq M|u|_{1;S}|v_0|_{1;S}.$$

As the restriction of u to T is linear,

$$|u|^2_{1,2;S} = \text{area}(S)|u|^2_{1,\infty;T} = \frac{\text{area}(S)}{\text{area}(T)}\,|u|^2_{1,2;T}.$$

Utilizing the inverse inequality

$$|v_0|_{1;S} \leq c_2 2^l \|v_0\|_{0;S},$$

the inner product $a(u, v_0)|_T$ can be estimated to be

$$a(u, v_0)|_T \leq c_3(\frac{1}{\sqrt{2}})^{l-k}|u|_{1;T}2^l\|v_0\|_{0;S}. \tag{6.21}$$

We remark that the factor 2^l enters because we normalized the L_2-like norm $\|\cdot\|_{0;T}$ according to (2.12) by the areas of the triangles in the initial triangulation. Because

$$\|v_0\|_{0;S} \leq c_4\|v\|_{0;T}, \quad \|v_1\|_{0;T} \leq c_4\|v\|_{0;T}$$

one obtains the proposition combining (6.20) and (6.21).

For $l = k, k+1$, the proposition follows from the usual Cauchy–Schwarz inequality and the inverse estimate given earlier. □

As result, every step of the hierarchical basis multigrid method reduces the energy norm of the error by at least a factor behaving like

$$1 - \mathcal{O}(1/j^2).$$

Thus Theorem 5.1 leads to an alternative proof of the main convergence theorem in Bank *et al.* (1988) for the special case that the coefficient functions of the differential operator are continuously differentiable. Note that the diameter of the triangles shrinks by the factor 2^{-j} in the transition from level 0 to level j. Therefore j grows logarithmically in the gridsize, which means very slowly. If the subspace corrections are repeated in the reversed order after every cycle, one gets a symmetrized iterative procedure which can be accelerated by the conjugate gradient method. Usually the hierarchical basis multigrid method is applied in this form, so that every step reduces the error in fact by a factor behaving like

$$1 - \mathcal{O}(1/j).$$

We remark that the fact, that the considered finite element functions are piecewise linear, is not essential for the Cauchy–Schwarz type inequality (6.17). With an additional factor 2^k on the right-hand side of (6.20) arising from the the second term on the right-hand side of equation (6.19) and a new

constant in (6.21), the proof of Lemma 6.1 transfers to the case of higher order polynomials of fixed degree.

The subspace decomposition needed in the construction and analysis of the hierarchical basis multigrid method can also be used to derive an alternative convergence result for the usual multigrid method. How, is described in the next section by the example of L_2-like decompositions. One finds that every multigrid cycle reduces the energy norm of the error at least by a factor behaving like

$$1 - \mathcal{O}(1/j^2),$$

as for the hierarchical basis multigrid method. For a symmetrized version accelerated by the conjugate gradient method, one again obtains a better reduction factor

$$1 - \mathcal{O}(1/j).$$

Contrary to the asymptotically better estimates derived earlier, these estimates (as well as the estimates for the hierarchical basis multigrid method) do not depend on the regularity of the boundary value problem and are even independent of jumps of the coefficient functions across the boundaries of the triangles in the initial triangulation. On the other hand, contrary to the estimate earlier, they are restricted to two space dimensions.

Without any essential change, the analysis of the hierarchical basis multigrid method can be transferred to the case of nonuniformly refined grids. Utilizing the same splitting of $\mathcal{S}$, one can also analyse multigrid methods which are based on local smoothing procedures.

7. L_2-like subspace decompositions

The best, in a certain sense, subspace decomposition is the orthogonal decomposition of $\mathcal{S}$ into $\mathcal{V}_0 = \mathcal{S}_0$ and the orthogonal complements

$$\mathcal{V}_k = \{Q_k u - Q_{k-1} u \,|\, u \in \mathcal{S}\} \subseteq \mathcal{S}_k \tag{7.1}$$

for $k = 1, \ldots, j$. This decomposition has been used for the first time in the analysis of multigrid methods in Bramble *et al.* (1991b) and, for the analysis of closely related additive multilevel methods as discussed in the next section, in Bramble *et al.* (1990) and Xu (1989).

The stability (5.2) of this decomposition can be essentially derived from the error estimate

$$\|u - Q_k u\|_0 \leq C_1 2^{-k} \|u\|, \tag{7.2}$$

which holds for all functions in $H^1(\Omega)$. For H^2-regular problems, this error estimate follows from the Aubin–Nitsche Lemma which is also the basis for the classical proofs in Sections 3 and 4. Here we have less regular boundary value problems in mind. An elementary proof of (7.2), which does not rely on

such regularity assumptions and which is based on local quasi-interpolants, can be found in Yserentant (1990), for example. Utilizing (7.2), one can show very easily that the discrete norm

$$|\!|\!| u |\!|\!|^2 = \|Q_0 u\|^2 + \sum_{k=1}^{j} 4^k \|Q_k u - Q_{k-1} u\|_0^2 \tag{7.3}$$

satisfies the estimate

$$|\!|\!| u |\!|\!|^2 \leq K_1^*(j+1)\|u\|^2, \tag{7.4}$$

which contains an additional logarithmic factor compared with (6.10).

Utilizing the equivalence of certain Besov and Sobolev spaces, Oswald (1990, 1991, 1992) and Dahmen and Kunoth (1992) recently developed a very general framework to compare norms like (7.3) with Sobolev norms. Especially, they could improve (7.4) to

$$|\!|\!| u |\!|\!|^2 \leq K_1 \|u\|^2 \tag{7.5}$$

with a constant K_1 neither depending on the number of refinement levels nor on regularity properties of the boundary value problem. In Bornemann and Yserentant (1992), a more specialized, but relatively elementary proof of (7.5) is given. The influence of boundary conditions and nonuniform refinements is discussed very carefully in this article.

Results, which are related to (7.5), have been proven in Bramble and Pasciak (1991), Xu (1992b), and X. Zhang (1992). These articles are based on the regularity theory of elliptic equations, although the degree of regularity finally enters only in the size of the constants.

Supposing again the property (6.6) of the smoothers, (7.5) yields the first basic assumption (5.2). In addition, with (7.5) (or also with (7.2)) one obtains

$$4^k \|v_k\|_0^2 \leq C\|v_k\|^2, \quad v_k \in \mathcal{V}_k, \tag{7.6}$$

so that, on $\mathcal{V}_k$, the energy norm $\|\cdot\|$ induced by the boundary value problem under consideration is equivalent to the scaled L_2-like norm $2^k\|\cdot\|_0$.

For the proof of the second basic assumption (5.3), (5.4) of the general theory in Section 5, we can again utilize Lemma 6.1, i.e.

$$a(u_k, v_l) \;\leq\; C\Big(\frac{1}{\sqrt{2}}\Big)^{l-k} \|u_k\|\, 2^l \|v_l\|_0. \tag{7.7}$$

for $k \leq l$ and all functions $u_k \in \mathcal{S}_k$ and $v_l \in \mathcal{S}_l$. With (7.6), one obtains, for functions $v_l \in \mathcal{V}_l$, the strengthened Cauchy–Schwarz inequality

$$a(u_k, v_l) \;\leq\; \widehat{C}\Big(\frac{1}{\sqrt{2}}\Big)^{l-k} \|u_k\|\, \|v_l\|. \tag{7.8}$$

With (6.6), (7.8) yields the desired Cauchy–Schwarz type inequality (5.3)

$$a(w_k, v_l) \leq \widetilde{C}(\frac{1}{\sqrt{2}})^{l-k}(B_k w_k, w_k)^{1/2}(B_l v_l, v_l)^{1/2} \tag{7.9}$$

for $k \leq l$ and the functions $w_k \in \mathcal{W}_k$ and $v_l \in \mathcal{V}_l$.

Thus we have proven the convergence of classical multigrid methods with a convergence rate

$$1 - \mathcal{O}(1)$$

which does not deteriorate with the number of refinement levels and without utilizing regularity properties of the boundary value problem.

If the smoothers $B_k : \mathcal{S}_k \to \mathcal{S}_k$ satisfy the stronger and somewhat more restrictive condition (as compared with the condition (6.6))

$$c_1 4^k \|u_k\|_0^2 \leq (u_k, B_k u_k) \leq c_2 4^k \|u_k\|_0^2 \tag{7.10}$$

for all $u_k \in \mathcal{S}_k$, the proof of the Cauchy–Schwarz type inequality (5.3) can be based directly on (7.7), and (7.9) holds even for all functions $v_l \in \mathcal{S}_l$. According to the remark in Section 5, for this case, interestingly the optimality of the multigrid method does not depend on the order in which the subspace corrections are performed.

If the coefficient functions of the differential operator in (2.6) are no longer differentiable, or if the derivatives are large, one can still apply Theorem 5.4 and gets a nearly optimal convergence rate.

For nonuniformly refined grids, one can work with L_2-like decompositions which are based on *local projections* as introduced in Dahmen and Kunoth (1992) or Bornemann and Yserentant (1992). Such decompositions can be analysed on the basis of the equivalence of the energy norm to the discrete norm (7.3).

8. Additive multilevel methods

Stimulated by the development of domain decomposition and of hierarchical basis methods, the interest has recently shifted from the recursively defined classical multilevel algorithms to additive multilevel methods. The most prominent new example in this class of algorithms is the multilevel nodal basis method of Bramble, Pasciak and Xu.

A main reason for this development is that additive multilevel algorithms fit much better to nonuniformly refined grids (as they are absolutely necessary for the solution of complicated real-life problems) because these algorithms allow the use of simpler, more natural data structures. Another reason is that the higher flexibility of these algorithms simplifies the use of parallel computers, although this should not be viewed too naively. It is fair to mention that additive methods usually need slightly more iteration

steps than their multiplicative counterparts, although the single iteration step tends to be cheaper.

Additive multilevel methods fall into the class of the additive subspace correction methods already introduced in Section 2. In the same way, as the multiplicative subspace correction methods introduced in Section 2 correspond to the Gauß–Seidel method, the additive subspace correction methods discussed in this section are associated with the Jacobi iteration. In the notation of Section 2, for given subspaces $\mathcal{W}_0, \ldots, \mathcal{W}_J$ of $\mathcal{S}$ and given approximations B_k, the additive subspace correction method for the solution of the abstract equation (2.4) is

$$\widetilde{u} \;\leftarrow\; \widetilde{u} + \alpha \sum_{k=0}^{J} B_k^{-1} Q_k (f - A\widetilde{u}). \tag{8.1}$$

This means that the single subspace corrections are not applied in a sequential order but in parallel. The iteration (8.1) can be rewritten as

$$\widetilde{u} \;\leftarrow\; \widetilde{u} + \alpha\, C(f - A\widetilde{u}) \tag{8.2}$$

with the approximate inverse

$$C \;=\; \sum_{k=0}^{J} B_k^{-1} Q_k \tag{8.3}$$

of the operator $A : \mathcal{S} \to \mathcal{S}$.

As in the convergence theory for the multiplicative variant, the convergence theory for the additive subspace correction method is based on splitting $\mathcal{S}$ into subspaces $\mathcal{V}_k$ of the spaces $\mathcal{W}_k$. The convergence estimates are based on two assumptions. The first is again the stability assumption (5.2) that, for all $v_k \in \mathcal{V}_k$,

$$\sum_{k=0}^{J} (B_k v_k, v_k) \;\le\; K_1 \,\|\sum_{k=0}^{J} v_k\|^2. \tag{8.4}$$

The second assumption is that there exists a (new) constant K_2 with

$$\|\sum_{k=0}^{J} w_k\|^2 \;\le\; K_2 \sum_{k=0}^{J} (B_k w_k, w_k) \tag{8.5}$$

for all $w_k \in \mathcal{W}_k$. The assumption (8.5) can be deduced from the Cauchy–Schwarz type estimate (5.23) which is stronger than the assumption (5.3). The related condition (5.8)

$$\|\sum_{k=0}^{J} v_k\|^2 \;\le\; K_2 \sum_{k=0}^{J} (B_k v_k, v_k)$$

for the elements $v_k \in \mathcal{V}_k$ is *not* sufficient.

The main result for additive subspace correction methods is a generalization of well known theorems from the theory of domain decomposition methods; see Widlund (1989) and Björstad and Mandel (1991).

Theorem 8.1 The operator C is symmetric and positive definite with respect to the inner product (2.2) on $\mathcal{S}$. Therefore the eigenvalues λ of the operator CA are real and positive. They range in the interval

$$1/K_1 \le \lambda \le K_2. \tag{8.6}$$

Proof. As the B_k are symmetric operators, one has, for all $u, v \in \mathcal{S}$,

$$(Cu, v) \;=\; \sum_{k=0}^{J} (B_k^{-1} Q_k u, Q_k v) \;=\; (u, Cv)$$

so that C is symmetric and also positive definite. Let $v \in \mathcal{S}$ have the decomposition $v = \sum_{k=0}^{J} v_k$ with elements $v_k \in \mathcal{V}_k$. Then, by Lemma 5.2,

$$\begin{aligned} \|v\|^2 \;&=\; \sum_{k=0}^{J} a(v_k, v) \;=\; \sum_{k=0}^{J} a(v_k, P_k v) \\ &\le\; \sqrt{K_1}\, \|v\| \Big(\sum_{k=0}^{J} a(T_k P_k v, P_k v) \Big)^{1/2} \\ &=\; \sqrt{K_1}\, \|v\| \Big(\sum_{k=0}^{J} a(T_k v, v) \Big)^{1/2} \end{aligned}$$

where, as in Section 5, $T_k = B_k^{-1} Q_k A$. Because $\sum_{k=0}^{J} T_k = CA$, one gets

$$a(v, v) \;\le\; K_1\, a(CAv, v).$$

Therefore the eigenvalues of CA cannot be less than $1/K_1$. By the new assumption (8.5), one obtains, for all $v \in \mathcal{S}$,

$$\| \sum_{k=0}^{J} T_k v \|^2 \le K_2 \sum_{k=0}^{J} (B_k T_k v, T_k v) = K_2 \sum_{k=0}^{J} a(T_k v, v)$$

or, again with $\sum_{k=0}^{J} T_k = CA$, the estimate

$$\|CAv\|^2 \le K_2\, a(CAv, v).$$

Therefore the eigenvalues of CA are not greater than K_2. □

By Theorem 8.1, the iteration (8.2) can be accelerated by the conjugate gradient method. In fact, additive subspace correction methods are nearly exclusively used in this way so that the proper choice of the damping parameter α is no longer a question of practical interest. The quality of C as a preconditioner for A is essentially described by the spectral condition

number

$$\kappa = \kappa(C^{1/2} A C^{1/2}) \tag{8.7}$$

which is defined as the ratio of the maximum and the minimum eigenvalue of the operator $C^{1/2} A C^{1/2}$. As this operator is similar to CA, Theorem 8.1 says that

$$\kappa \leq K_1 K_2 \tag{8.8}$$

is an upper bound for this condition number.

For additive multilevel methods for the solution of finite element equations like (2.11), the spaces $\mathcal{W}_k \subseteq \mathcal{S}$ are subspaces of the coarse level spaces $\mathcal{S}_k$; for the *multilevel nodal basis method* of Bramble *et al.* (1990) and Xu (1989) one has $\mathcal{W}_k = \mathcal{S}_k$, and for the *hierarchical basis method* (Yserentant 1986b, 1990, 1992), the spaces $\mathcal{W}_k$ are the hierarchical complements (2.41). Therefore the multilevel nodal basis method can be seen as the additive version of the multigrid V-cycle whereas the hierarchical basis method is the additive version of the hierarchical basis multigrid method.

For both methods, we again require that $B_0 = A_0$ and that, for $k \geq 1$, the operators $B_k : \mathcal{W}_k \to \mathcal{W}_k$ satisfy an estimate

$$c_1 4^k \|w_k\|_0^2 \leq (w_k, B_k w_k) \leq c_2 4^k \|w_k\|_0^2 \tag{8.9}$$

for all $w_k \in \mathcal{W}_k$. This is essentially the condition (3.6) which is somewhat more restrictive than the condition (6.6) used in the analysis of the multiplicative variants. However, remember that simple point Jacobi and Gauß–Seidel smoothers are covered by (8.9).

Then, for arbitrary subspaces $\mathcal{W}_k \subseteq \mathcal{S}_k$, the Cauchy–Schwarz type inequality

$$a(u, v) \;\leq\; C \,(\frac{1}{\sqrt{2}})^{l-k} \,\|u\|\, 2^l \|v\|_0 \tag{8.10}$$

from Lemma 6.1 for the functions $u \in \mathcal{S}_k$ and $v \in \mathcal{S}_l$, $k \leq l$, and (8.9) yield the new condition (8.5).

The subspaces $\mathcal{V}_k \subseteq \mathcal{W}_k$ are chosen as for the corresponding multiplicative schemes. Therefore the stability condition (8.4) has already been derived in the last section. With Theorem 8.1, we can conclude that the additive multilevel methods have qualitatively the same convergence behaviour as their multiplicative counterparts.

In order to exhibit the advantages of additive multilevel methods, the approximations B_k for the operators A_k should be chosen as simple as possible. The best possible choice is probably the Jacobi method.

In the following we discuss the realization of the multilevel nodal basis method in conjunction with the Jacobi method. Let $\mathcal{N}_k = \{x_1, \ldots, x_{n_k}\}$ be the set of vertices of the triangles in $\mathcal{T}_k$ not lying on the boundary of Ω. Then $\mathcal{S}_k$ is spanned by the nodal basis functions $\psi_i^{(k)}$, $i = 1, \ldots, n_k$, which

are defined by

$$\psi_i^{(k)}(x_l) = \delta_{il}, \quad x_l \in \mathcal{N}_k. \tag{8.11}$$

Then, for the given B_k, the operator

$$C \;=\; A_0^{-1}Q_0 + \sum_{k=1}^{j} B_k^{-1}Q_k \tag{8.12}$$

can be written as

$$Cr \;=\; A_0^{-1}Q_0 r + \sum_{k=1}^{j}\sum_{i=1}^{n_k} \frac{(r,\psi_i^{(k)})}{a(\psi_i^{(k)},\psi_i^{(k)})}\,\psi_i^{(k)}. \tag{8.13}$$

To realize the iteration

$$\widetilde{u} \;\leftarrow\; \widetilde{u} + \alpha\, Cr, \quad r = f - A\widetilde{u}, \tag{8.14}$$

or its conjugate gradient accelerated version efficiently, the functions $\widetilde{u}$ and the residuals r have to be represented differently. We store $\widetilde{u}$ by the values

$$\widetilde{u}(x_i), \quad i = 1,\ldots,n, \tag{8.15}$$

whereas r is represented by

$$(r,\psi_i), \quad i = 1,\ldots,n, \tag{8.16}$$

where, for simplicity, $n = n_j$ and $\psi_i = \psi_i^{(j)}$. The inner products (8.16) are given by

$$(r,\psi_i) = (f,\psi_i) - \sum_{l=1}^{n} a(\psi_i,\psi_l)\,u(x_l), \tag{8.17}$$

so that only the usual residual has to be computed; an explicit representation of the operator A is not needed. Note that the values $(r,\psi_i^{(k)})$ can be recursively computed beginning with the values $(r,\psi_i) = (r,\psi_i^{(j)})$, and that the summation of the single terms in (8.13) can be formulated as a recursive process, too. The function

$$u_0 = A_0^{-1}Q_0 r \in \mathcal{S}_0 \tag{8.18}$$

satisfies the relation

$$a(u_0,v) = (r,v), \quad v \in \mathcal{S}_0. \tag{8.19}$$

To compute u_0, therefore one needs only $(r,\psi_i^{(0)})$, $i = 1,\ldots,n_0$, but not $Q_0 r$ itself, and one has to solve a linear system with the level 0 discretization matrix.

The appropriate modification of the multilevel nodal basis preconditioner

to nonuniformly refined grids is

$$Cr \;=\; A_0^{-1}Q_0 r + \sum_{k=1}^{j} \sum_{\psi_i^{(k)} \neq \psi_i^{(k-1)}} \frac{(r, \psi_i^{(k)})}{a(\psi_i^{(k)}, \psi_i^{(k)})}\; \psi_i^{(k)} \tag{8.20}$$

where the inner sum stands for

$$\sum_{\psi_i^{(k)} \neq \psi_i^{(k-1)}} \;=\; \sum_{\substack{i=1 \\ \psi_i^{(k)} \neq \psi_i^{(k-1)}}}^{n_{k-1}} \;+\; \sum_{i=n_{k-1}+1}^{n_k} . \tag{8.21}$$

Only those basis functions $\psi_i^{(k)}$ of $\mathcal{S}_k$ are still taken into account in the inner sum which are associated with the new nodes $x_i, i = n_{k-1}+1, \ldots, n_k$, on the level k and with the neighbours of these nodes. With this modification, the operation count for the single iteration step (8.14) remains strictly proportional to the number of unknowns independent of the distribution of the unknowns among the levels.

As the corresponding multigrid methods based on local smoothing procedures, this version of the multilevel nodal basis method can be analysed utilizing the local L_2-decompositions introduced in Bornemann and Yserentant (1992). It turns out that the condition number (8.7) behaves like $\mathcal{O}(1)$, as in a uniform refinement.

The hierarchical basis method goes one step further. In Xu's formulation (Xu, 1989) it is given by

$$Cr \;=\; A_0^{-1}Q_0 r + \sum_{k=1}^{j} \sum_{i=n_{k-1}+1}^{n_k} \frac{(r, \psi_i^{(k)})}{a(\psi_i^{(k)}, \psi_i^{(k)})}\; \psi_i^{(k)}. \tag{8.22}$$

As every term in the double sum can be associated with a node of the final level, the algorithmic realization of this method becomes extremely simple; see Yserentant (1986b, 1990).

If we introduce the hierarchical basis functions $\widehat{\psi}_i, i = 1, \ldots, n$, of $\mathcal{S}$ by

$$\widehat{\psi}_i = \psi_i^{(0)}, \quad x_i \in \mathcal{N}_0, \tag{8.23}$$

and by

$$\widehat{\psi}_i = \psi_i^{(k)}, \quad x_i \in \mathcal{N}_k \setminus \mathcal{N}_{k-1}, \tag{8.24}$$

the hierarchical basis preconditioner takes the form

$$Cr \;=\; A_0^{-1}Q_0 r + \sum_{i=n_0+1}^{n} \frac{(r, \widehat{\psi}_i)}{a(\widehat{\psi}_i, \widehat{\psi}_i)}\; \widehat{\psi}_i. \tag{8.25}$$

Thus it is, up to a small block of the dimension n_0 of the initial finite element space $\mathcal{S}_0$, Jacobi's old method, now with respect to the hierarchical basis

formulation of the discrete elliptic boundary value problem. In this sense, it is the most simple multigrid method.

REFERENCES

O. Axelsson and I. Gustafsson (1983), 'Preconditioning and two-level multigrid methods of arbitrary degree of approximation', *Math. Comput.* **40**, 219–242.

O. Axelsson and P. Vassilevski (1989), 'Algebraic multilevel preconditioning methods I', *Numer. Math.* **56**, 157–177.

N.S. Bachvalov (1966), 'On the convergence of a relaxation method with natural constraints on the elliptic operator' (in Russian), *USSR Comput. Math. and Math. Phys.* **6.5**, 101–135.

E. Bänsch (1991), 'Local mesh refinement in 2 and 3 dimensions', *Impact of Computing in Sci. Engrg.* **3**, 181–191.

R.E. Bank (1981), 'A comparison of two multilevel iterative methods for nonsymmetric and indefinite elliptic finite element equations', *SIAM J. Numer. Anal.* **18**, 724–743.

R.E. Bank (1990), *PLTMG: A Software Package for Solving Elliptic Partial Differential Equations*, SIAM (Philadelphia).

R.E. Bank and C.C. Douglas (1985), 'Sharp estimates for multigrid convergence rates of convergence with general smoothing and acceleration', *SIAM J. Numer. Anal.* **22**, 617–633.

R.E. Bank and T. Dupont (1980), 'Analysis of a two-level scheme for solving finite element equations', Report CNA-159, Center for Numerical Analysis, University of Texas at Austin.

R.E. Bank and T. Dupont (1981), 'An optimal order process for solving finite element equations', *Math. Comput.* **36**, 35–51.

R.E. Bank, T. Dupont and H. Yserentant (1988), 'The hierarchical basis multigrid method', *Numer. Math.* **52**, 427–458.

R.E. Bank and D.J. Rose (1982), 'Analysis of a multilevel iterative method for nonlinear finite element equations', *Math. Comput.* **39**, 453–465.

R.E. Bank and A. Weiser (1985), 'Some a-posteriori estimators for elliptic partial differential equations', *Math. Comput.* **44**, 283–301.

R.E. Bank, B. Welfert and H. Yserentant (1990), 'A class of iterative methods for solving saddle point problems', *Numer. Math.* **56**, 645–666.

P.E. Björstad and J. Mandel (1991), 'On the spectra of sums of orthogonal projections with applications to parallel computing', *BIT* **31**, 76–88.

F.A. Bornemann and H. Yserentant (1992), 'A basic norm equivalence for the theory of multilevel methods', *Numer. Math.*, to appear.

D. Braess (1981), 'The contraction number of a multigrid method for solving the Poisson equation', *Numer. Math.* **37**, 387–404.

D. Braess and W. Hackbusch (1983), 'A new convergence proof for the multigrid method including the V-cycle', *SIAM J. Numer. Anal.* **20**, 967–975.

J.H. Bramble and J.E. Pasciak (1988), 'A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems', *Math. Comput.* **181**, 1–17.

J.H. Bramble and J.E. Pasciak (1991), 'New estimates for multilevel algorithms including the V-cycle', Technical Report '91-47, Mathematical Sciences Institute, Cornell University.

J.H. Bramble, J.E. Pasciak, J. Wang and J. Xu (1991a), 'Convergence estimates for product iterative methods with applications to domain decomposition', *Math. Comput.* **57**, 1–21.

J.H. Bramble, J.E. Pasciak, J. Wang and J. Xu (1991b), 'Convergence estimates for multigrid algorithms without regularity assumptions', *Math. Comput.* **57**, 23–45.

J.H. Bramble, J.E. Pasciak and J. Xu (1990), 'Parallel multilevel preconditioners', *Math. Comput.* **55**, 1–22.

A. Brandt (1977), 'Multi-level adaptive solutions to boundary-value problems', *Math. Comput.* **31**, 333–390.

A. Brandt (1982), 'Guide to multigrid development', in *Multigrid Methods* (W. Hackbusch and U. Trottenberg, eds), Lecture Notes in Mathematics 960, Springer (Berlin, Heidelberg, New York).

T. Chan, R. Glowinski, J. Periaux and O.B. Widlund, eds (1989), *Domain Decomposition Methods, Proceedings, Los Angeles 1988*, SIAM (Philadelphia).

P.G. Ciarlet (1978), *The Finite Element Method for Elliptic Problems*, North Holland (Amsterdam).

W. Dahmen and A. Kunoth (1992), 'Multilevel preconditioning', *Numer. Math.*, to appear.

P. Deuflhard (1992), *Newton-Techniques for Highly Nonlinear Problems – Theory and Algorithms*, Academic Press (New York) in preparation.

P. Deuflhard, P. Leinen and H. Yserentant (1989), 'Concepts of an adaptive hierarchical finite element code', *Impact of Computing in Sci. Engrg.* **1**, 3–35.

M. Dryja and O. Widlund (1991), 'Multilevel additive methods for elliptic finite element problems', in *Parallel Algorithms for Partial Differential Equations, Proceedings, Kiel 1990* (W. Hackbusch, ed.), Vieweg (Braunschweig, Wiesbaden).

R.P. Fedorenko (1964), 'The speed of convergence of an iterative process' (in Russian), *USSR Comput. Math. and Math. Phys.* **4.3**, 227–235.

R.W. Freund, G.H. Golub and N.M. Nachtigal (1992), 'Iterative solution of linear systems', *Acta Numerica 1992*, 1–44.

W. Hackbusch (1976), 'Ein iteratives Verfahren zur schnellen Auflösung elliptischer Randwertprobleme', Report 76-12, Mathematisches Institut der Universität zu Köln.

W. Hackbusch (1981), 'On the convergence of multi-grid iterations', *Beiträge Numer. Math.* **9**, 213–329.

W. Hackbusch (1982), 'Multi-grid convergence theory', in *Multigrid Methods* (W. Hackbusch and U. Trottenberg, eds), Lecture Notes in Mathematics 960, Springer (Berlin, Heidelberg, New York).

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

W. Hackbusch (1989a), 'The frequency decomposition multi-grid algorithm', in *Robust Multi-Grid Methods, Proceedings Kiel 1988* (W. Hackbusch, ed.). Vieweg (Braunschweig, Wiesbaden).

W. Hackbusch (1989b), 'The frequency decomposition multi-grid method. Part I: Application to anisotropic equations', *Numer. Math.* **56**, 229–245.
W. Hackbusch (1991), *Iterative Lösung großer schwachbesetzter Systeme*, Teubner (Stuttgart) (English translation in preparation).
W. Hackbusch and A. Reusken (1989), 'Analysis of a damped nonlinear multilevel method', *Numer. Math.* **55**, 225–246.
W. Hackbusch and U. Trottenberg, eds (1982), *Multigrid Methods, Proceedings, Köln 1981*, Lecture Notes in Mathematics 960, Springer (Berlin, Heidelberg, New York).
J. Mandel, S. McCormick and R. Bank (1987), 'Variational multigrid theory', in *Multigrid Methods* (S. McCormick, ed.), SIAM (Philadelphia).
S. McCormick, ed. (1987), *Multigrid Methods*, SIAM (Philadelphia).
P. Oswald (1990), 'On function spaces related to finite element approximation theory', *Zeitschrift für Analysis und ihre Anwendungen* **9**, 43–64.
P. Oswald (1991), 'On discrete norm estimates related to multilevel preconditioners in the finite element method', in *Proceedings of the International Conference on the Constructive Theory of Functions, Varna 1991.*
P. Oswald (1992), 'Stable splittings of Sobolev spaces and fast solution of variational problems', Forschungsergebnisse der Friedrich-Schiller-Universität Jena, Math/92/5.
A. Reusken (1992), 'On maximum norm convergence of multigrid methods for elliptic boundary value problems', Report RANA 92-03, Department of Mathematics and Computer Science, Eindhoven University.
M.C. Rivara (1984), 'Algorithms for refining triangular grids suitable for adaptive and multigrid techniques', *Int. J. Numer. Meth. Engrg.* **20**, 745–756.
A. Schatz (1974), 'An observation concerning Ritz–Galerkin methods with indefinite bilinear forms', *Math. Comput.* **28**, 959–962.
R. Stevenson (1992), 'New estimates of the contraction number of V-cycle multigrid with applications to anisotropic equations', Preprint Nr. 713, Department of Mathematics, University of Utrecht.
K. Stüben and U. Trottenberg (1982), 'Multigrid methods: fundamental algorithms, model problem analysis and applications', in *Multigrid Methods*, (W. Hackbusch and U. Trottenberg, eds.), Lecture Notes in Mathematics 960, Springer (Berlin, Heidelberg, New York).
R.S. Varga (1962), *Matrix Iterative Analysis*, Prentice Hall (Englewood Cliffs, NJ).
P. Vassilevski (1992), 'Preconditioning nonsymmetric and indefinite finite element matrices', *J. Numer. Linear Alg. with Appl.* **1**, 59–76.
P. Wesseling (1992), *An Introduction to Multigrid Methods*, John Wiley & Sons (Chichester).
O. Widlund (1989), 'Optimal iterative refinement methods', in *Domain Decomposition Methods* (T. Chan et al., eds), SIAM (Philadelphia).
G. Wittum (1989a), 'Linear iterations as smoothers in multigrid methods: theory with applications to incomplete decompositions', *Impact of Computing in Sci. Engrg.* **1**, 180–215.
G. Wittum (1989b), 'On the robustness of ILU smoothing', *SIAM J. Sci. Statist. Comput.* **10**, 699–717

J. Xu (1989), 'Theory of multilevel methods', Report AM48, Department of Mathematics, Pennsylvania State University.
J. Xu (1992a), 'A new class of iterative methods for nonsymmetric and indefinite problems', *SIAM J. Numer. Anal.* **29**, 303–319.
J. Xu (1992b),'Iterative methods by space decomposition and subspace correction', *SIAM Review* **34**.
D.M. Young (1950), 'Iterative Methods for Solving Partial Differential Equations of Elliptic Type', Thesis, Harvard University.
H. Yserentant (1983), 'On the convergence of multi-level methods for strongly nonuniform families of grids and any number of smoothing steps per level', *Computing* **30**, 305–313.
H. Yserentant (1986a), 'The convergence of multi-level methods for solving finite element equations in the presence of singularities', *Math. Comput.* **47**, 399–409.
H. Yserentant (1986b), 'On the multi-level splitting of finite element spaces', *Numer. Math.* **49**, 379–412.
H. Yserentant (1986c), 'On the multi-level splitting of finite element spaces for indefinite elliptic boundary value problems', *SIAM J. Numer. Anal.* **23**, 581-595.
H. Yserentant (1988), 'Preconditioning indefinite discretization matrices', *Numer. Math.* **54**, 719–734.
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 *ICIAM 91* (R.E. O'Malley, ed.), SIAM (Philadelphia).
S. Zhang (1990), 'Optimal order non-nested multigrid methods for solving finite element equations II: on non-quasiuniform meshes', *Math. Comput.* **55**, 439–450.
X. Zhang (1991), 'Multilevel additive Schwarz methods', Technical Report 582, Courant Institute, New York.
X. Zhang (1992), 'Multilevel Schwarz methods', *Numer. Math.*, to appear.

some references completed after publication:

F.A. Bornemann and H. Yserentant (1992), 'A basic norm equivalence for the theory of multilevel methods', *Numer. Math.* **64**, 455–476 (1993).
J.H. Bramble and J.E. Pasciak (1991), 'New estimates for multilevel algorithms including the V-cycle', *Math. Comput.* **60**, 959–962 (1993).
W. Dahmen and A. Kunoth (1992), 'Multilevel preconditioning', *Numer. Math.* **63**, 315–344.
J. Xu (1992b),'Iterative methods by space decomposition and subspace correction', *SIAM Review* **34**, 581–613.
X. Zhang (1992), 'Multilevel Schwarz methods', *Numer. Math.* **63**, 521–539.