Chapter V
Multigrid Methods

The multigrid method is one of the fastest methods for solving systems of equations involving a large number of unknowns. The method is due to Fedorenko, who formulated it first as a two-grid method [1961], and then later as a multigrid method [1964]. He showed that the algorithm requires only $\mathcal{O}(n)$ operations, where n is the number of unknowns. Bachvalov [1966] continued the study for difference equations, and allowed nonconstant coefficients. A. Brandt was the first to discover in the mid-seventies that the multigrid method is considerably better than other known methods, even for values of n which occur in actual problems. Nearly at the same time, the multigrid method was discovered independently by Hackbusch [1976], whose approach also led to a simplification of the concepts involved.

Our starting point is the observation that in solving a system of equations, we should use different methods for the high frequency (oscillating) and low frequency (smooth) parts. The idea of the multigrid method is to combine two different methods to get an algorithm which will be effective on the entire spectrum.

Classical iterative methods work essentially by smoothing, i.e. they quickly eliminate the high-frequency parts of the error function. The low frequency parts of the functions can then be computed relatively well on a coarser grid. Although we cannot strictly separate the low and high frequency parts, we are able to get iterative methods whose rate of convergence (i.e. error reduction factors) are in the range from $\frac{1}{20}$ to $\frac{1}{4}$; see Table 7. If we use the multigrid idea in combination with good starting values, we will get convergence in just one or two iterations, and the iterative character of the method almost disappears.

In developing a convergence theory, we have to take account of the smoothness as well as the absolute size of the error. This means that we need to work with (at least) two norms. Natural candidates are the Sobolev norms and their discrete analogs.

The algorithmic aspects discussed in §§1, 4 and 5 are essentially independent of the convergence theory presented in §§2 and 3. For more on multigrid methods, see the books of Hackbusch [1985], Hackbusch and Trottenberg [1982], Briggs [1987], McCormick [1989], and Wesseling [1992].

§ 1. Multigrid Methods for Variational Problems

Smoothing Properties of Classical Iterative Methods

Multigrid methods are based on the observation that the classical iterative methods result in smoothing. This is most easily seen by examining the model Example II.4.3 involving the Poisson equation on a rectangle. This example was also discussed in connection with the Gauss–Seidel and Jacobi methods; cf. (IV.1.11).

1.1 Example. The discretization of the Poisson equation on the unit square using the standard five-point stencil leads to the system of equations

$$4x_{i,j} - x_{i+1,j} - x_{i-1,j} - x_{i,j+1} - x_{i,j-1} = b_{ij} \quad \text{for } 1 \le i, j \le n-1. \tag{1.1}$$

Here $x_{i,0} = x_{i,n} = x_{0,j} = x_{n,j} = 0$. We consider the iterative solution of (1.1) using the Jacobi method with relaxation parameter ω:

$$x_{i,j}^{\nu+1} = \frac{\omega}{4}(x_{i+1,j}^{\nu} + x_{i-1,j}^{\nu} + x_{i,j+1}^{\nu} + x_{i,j-1}^{\nu}) + \frac{\omega}{4} b_{ij} + (1-\omega) x_{i,j}^{\nu}. \tag{1.2}$$

By (IV.1.12), the eigenvectors $z^{k,m}$ of the iteration matrix defined implicitly in (1.2) can be thought of as the discretizations of the eigenfunctions

$$(z^{k,m})_{i,j} = \sin \frac{ik\pi}{n} \sin \frac{jm\pi}{n}, \quad 1 \le i, j, k, m \le n-1, \tag{1.3}$$

of the Laplace operator, with corresponding eigenvalues

$$\lambda^{km} = \frac{1}{2}\cos\frac{k\pi}{n} + \frac{1}{2}\cos\frac{m\pi}{n}, \qquad \text{if } \omega = 1,$$

and

$$\lambda^{km} = \frac{1}{4}\cos\frac{k\pi}{n} + \frac{1}{4}\cos\frac{m\pi}{n} + \frac{1}{2}, \quad \text{if } \omega = \frac{1}{2}.$$

In each step of the iteration, the individual spectral parts of the errors are multiplied by the corresponding factors λ^{km}. Thus, those terms corresponding to eigenvalues whose moduli are near 1 are damped the least.

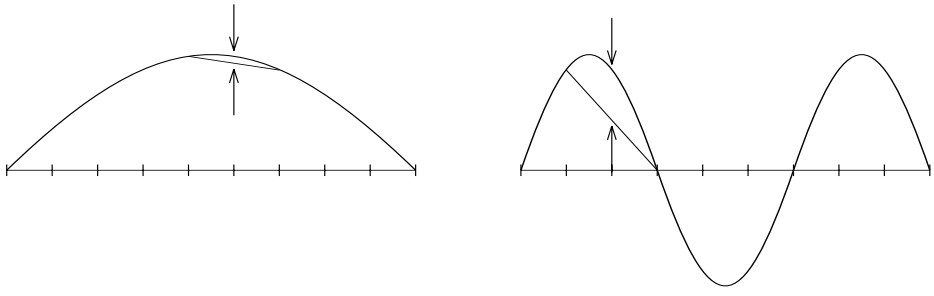

Fig. 48. Replacing values on a one-dimensional grid by the average values of their neighbors has only a minor effect on low frequency terms, but a substantial effect on high frequency ones.

1.2 Corollary. *After a few iterations of the Jacobi method (with $\omega = 1$), the error essentially contains only terms for which*

k and m are both small, or
k and m are both close to n.

Terms of the latter type are strongly reduced if we use $\omega = 1/2$ (instead of $\omega = 1$.) This leaves only the low frequency terms; see Fig. 48.

Accordingly, in carrying out the iteration we will get a clear reduction in the error as long as the error still contains highly oscillatory parts. However, as soon as the error becomes smooth, the iteration will essentially stop, and subsequently the error reduction per step will only be of order $1 - \mathcal{O}(n^{-2})$.

We will show later in connection with the convergence theory that the Jacobi method has a smoothing effect in general, and not just for the model problem, see Lemma 2.4. The following methods are used in practice for smoothing:

the Jacobi method and the Richardson iteration,
the successive overrelaxation method (SOR),
the symmetric successive overrelaxation method (SSOR),
iteration with the incomplete Cholesky decomposition (ICC).

We now list a few suggestions for the choice of the smoother. If the entire matrix of the system is to be stored, either the SOR or the SSOR method (with a small amount of underrelaxation) is useful in the standard case. For parallel computations, Jacobi relaxation is appropriate. Smoothing via ICC requires more work than the other methods, but turns out to be more robust for anisotropic problems; see Hemker [1980] and Wittum [1989b]. For example, we have a strongly anisotropic problem if one direction in space is preferred, such as for the differential equation

$$100\, u_{xx} + u_{yy} = f.$$

The Multigrid Idea

The above discussion suggests the following approach.

First we carry out several relaxation steps in order to strongly damp all oscillating components of the error. Then we go to a coarser grid, and approximate the remaining smooth part. This is possible because smooth functions can be approximated well on coarse grids.

We then alternately repeat the *smoothing step* on the fine grid and the *coarse-grid correction*. This results in an iterative method.

The system of equations corresponding to the problem on the coarse grid is usually simpler to solve than the original problem. In particular, in the planar case, if we go from a grid with mesh size h to one with mesh size $2h$, the number

of unknowns is reduced by a factor of about 4. Moreover, the bandwidth of the matrix for the coarse grid is about half as large. This means that the number of operations needed for Gauss elimination is reduced by a factor of about 16. This leads to major savings after just two or three iterations.

In general, the number of operations needed to solve the system corresponding to mesh size $2h$ will still be too large, and so we repeat the process. We continue to double the size of the grid until we get a sufficiently small system of equations.

The Algorithm

For the sake of simplicity, we first describe the two-grid algorithm for conforming finite elements.

To formulate multigrid algorithms, we need some notation. It is standard to use the letter $\mathcal{S}$ for the smoothing operator. For example, when smoothing via Richardson relaxation, we have

$$x \longmapsto \mathcal{S}x := x - \omega(A_h x - b_h). \tag{1.4}$$

In order to avoid confusion with the finite element spaces S_h, we suppress the subscript h in our notation for smoothing operators.

Let $\{\psi_i\}$ be a basis for S_h. Each vector $x \in \mathbb{R}^N$ with $N = N_h = \dim S_h$ is associated with the function $u = \sum_i x_i \psi_i \in S_h$. The indices are inherited in the correspondence.

The iterates u_h^k as well as the intermediate values $u_h^{k,1}$ are associated with the fine grid, while the quantities v_{2h} (which actually depend on k) correspond to the coarser grid with double mesh size.

1.3 Two-Level Iteration *(k-th cycle):*

Let u_h^k be a given approximation in S_h.

1. *Smoothing Step.* Perform ν smoothing steps:

$$u_h^{k,1} = \mathcal{S}^\nu u_h^k.$$

2. *Coarse-Grid Correction.* Compute the solution $\hat{v}_{2h}$ of the variational problem at level $2h$:

$$J(u_h^{k,1} + v) \longrightarrow \min_{v \in S_{2h}}!$$

Set

$$u_h^{k+1} = u_h^{k,1} + \hat{v}_{2h}. \qquad \square$$

1.4 Remarks. (1) The parameter ν controls the number of smoothing steps. In the standard case,

$$1 \leq \nu \leq 3.$$

The results of Ries, Trottenberg, and Winter [1983] presented in Table 7 show that for well-behaved problems such as the Poisson equation on a rectangle, it does not pay to do more than 2 smoothing steps. For more complicated problems (such as those involving nonconforming elements or mixed methods), it can happen that a much larger number of smoothing steps are necessary.

(2) In principle, the question of which bases to choose for the finite element spaces depends on the smoothing process. We can use the usual nodal basis functions, as will become clear from the convergence theory in §2.

Table 7. Bounds for the spectral radius ρ of the iteration matrix for the two-grid method for the Poisson equation. Here ν steps of the Gauss–Seidel method with the checkerboard order are used for smoothing.

ν	1	2	3	4
ρ	0.25	0.074	0.053	0.041

We turn now to the complete algorithm for several levels. It is easiest to give a precise formulation for conforming finite elements based on nested grids.

First choose a coarse triangulation $\mathcal{T}_{h_0}$. Let $\mathcal{T}_{h_1}$ be the triangulation which arises if we subdivide each triangle of $\mathcal{T}_{h_0}$ into four congruent subtriangles. Further subdivision leads to the grids[11] $\mathcal{T}_{h_2}, \mathcal{T}_{h_3}, \ldots, \mathcal{T}_{h_q}$ (see Fig. 49). We write $\mathcal{T}_\ell$ in place of $\mathcal{T}_{h_\ell}$ for $0 \leq \ell \leq \ell_{\max} =: q$. Suppose the finite element spaces corresponding to the triangulations $\mathcal{T}_\ell$ are S_{h_ℓ} (S_ℓ for short). Then

$$S_0 \subset S_1 \subset \ldots \subset S_{\ell_{\max}}. \tag{1.5}$$

Our goal is to compute the finite element solution of the boundary-value problem on the finest grid. The spaces S_ℓ corresponding to coarser grids arise only in the intermediate computations.

The variables in the multigrid algorithm involve up to three indices. Here

ℓ denotes the grid level,

k counts the iterations,

m counts the substeps inside each iteration.

We will usually write $u^{\ell,k}$ instead of $u^{\ell,k,0}$.

[11] Not all elements have to be refined as long as we take account of the rules discussed in Ch. II, §8 and in Ch. III, §8 in connection with local mesh refinement.

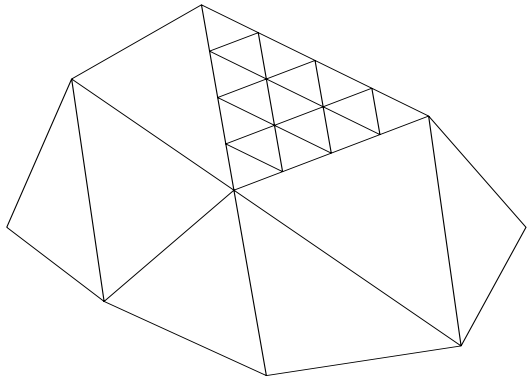

Fig. 49. A coarse triangulation for which one of the (coarse) triangles has been decomposed into 16 subtriangles in two steps. The other triangles should be decomposed in the analogous way.

We now define the multigrid method for conforming elements *recursively*.

1.5 Multigrid Iteration $\mathbf{MGM}_\ell$ (k-th cycle at level $\ell \geq 1$):

Let $u^{\ell,k}$ be a given approximation in S_ℓ.

1. *Pre-Smoothing.* Carry out ν_1 smoothing steps:

$$u^{\ell,k,1} = \mathcal{S}^{\nu_1} u^{\ell,k}. \tag{1.6}$$

2. *Coarse-Grid Correction.* Let $\hat{v}^{\ell-1}$ denote the solution of the variational problem at level $\ell - 1$,

$$J(u^{\ell,k,1} + v) \longrightarrow \min_{v \in S_{\ell-1}} ! \tag{1.7}$$

If $\ell = 1$, find the solution and set $v^{\ell-1} = \hat{v}^{\ell-1}$.
If $\ell > 1$, compute an approximation $v^{\ell-1}$ of $\hat{v}^{\ell-1}$ by carrying out μ steps of $\mathbf{MGM}_{\ell-1}$ with the starting value $u^{\ell-1,0} = 0$.
Set

$$u^{\ell,k,2} = u^{\ell,k,1} + v^{\ell-1}. \tag{1.8}$$

3. *Post-Smoothing.* Carry out ν_2 smoothing steps,

$$u^{\ell,k,3} = \mathcal{S}^{\nu_2} u^{\ell,k,2},$$

and set $u^{\ell,k+1} = u^{\ell,k,3}$. □

1.6 Remarks. (1) If only two levels are being used, then we have only the case $\ell = 1$, and the coarse-grid correction will be done exactly. For more than two levels, we compute the solution on the coarse grid only approximately, and for the convergence theory we treat the multigrid iteration as a perturbed two-grid iteration.

(2) For more than two levels, it makes a difference whether we perform the smoothing before or after the coarse-grid correction. This is controlled by the parameters ν_1 and ν_2. For simplicity, frequently only the pre-smoothing is performed. However, for the V-cycle, it is better to do an equal amount of smoothing both before and after, i.e. to choose $\nu_1 = \nu_2$.

(3) Choosing $\mu = 1$ or $\mu = 2$ leads to either a *V-cycle* or a *W-cycle*. The reason for this terminology is clear from the shape of the corresponding schemes shown in Fig. 50. Obviously a W-cycle is more expensive than a V-cycle.

In order to ensure that in running through several levels the error does not build up too much, in the early use of the multigrid method most people chose W-cycles. However, most problems are so well-behaved that multigrid algorithms with the V-cycle are faster. (For more than four levels, it is better to insert one W-cycle after every three V-cycles; see Problem 3.12.)

(4) We solve the system of equations corresponding to the variational problem on the coarsest grid using Gauss elimination or some other direct method.

(5) In practice, an auxiliary grid can be so coarse that it would never be used as the final grid. For the Poisson equation on the unit square, it is even possible that the grid is coarsened so much that the coarsest grid contains only one (interior) point. This does not ruin the convergence rate of multigrid algorithms.

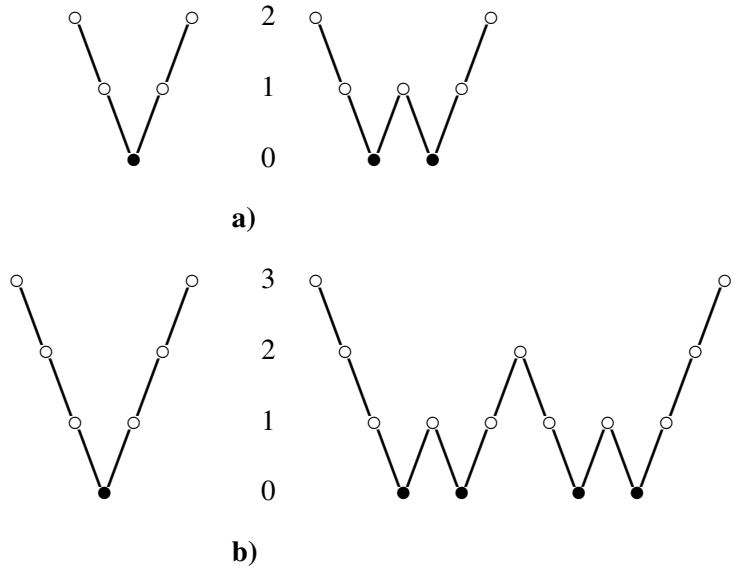

Fig. 50. V-cycle and W-cycle on three and four levels

Transfer Between Grids

The formulation (1.7) of the variational problem corresponding to a coarse-grid correction requires moving from the grid S_ℓ to $S_{\ell-1}$. The calculation uses the matrix-vector form

$$A_{\ell-1} y_{\ell-1} = b_{\ell-1}$$

of the system of equations which gives the solution of the auxiliary problem (1.7) on the coarser grid. The matrix $A_{\ell-1}$, which is of smaller dimension than A_ℓ, and the current right-hand side $b_{\ell-1}$ have to be computed.

All multigrid methods for linear systems of equations involve formulas of the form

$$\begin{aligned} A_{\ell-1} &:= r A_\ell p, \\ b_{\ell-1} &:= r d_\ell \qquad \text{with } d_\ell := b_\ell - A_\ell u^{\ell,k,1}. \end{aligned} \tag{1.9}$$

The matrix $r = r_\ell$ is called the *restriction*, and the matrix $p = p_\ell$ is called the *prolongation*. The choice of p and r has a major influence on the rate of convergence.

There is a canonical choice for p and r when using conforming Lagrange elements. Then the spaces are nested, i.e., $S_{\ell-1} \subset S_\ell$. In these cases we take p to be the matrix representation of the injection $j : S_{\ell-1} \hookrightarrow S_\ell$, and $r := p^t$ to be the matrix of the adjoint operator $j^* : S'_\ell \hookrightarrow S'_{\ell-1}$.

To express the right-hand side, let $\{\psi_i^\ell\}_{i=1}^{N_\ell}$ be a basis for S_ℓ, and let $\{\psi_j^{\ell-1}\}_{j=1}^{N_{\ell-1}}$ be a basis for $S_{\ell-1}$. Since $S_{\ell-1} \subset S_\ell$, there exists an $N_{\ell-1} \times N_\ell$ matrix r with

$$\psi_j^{\ell-1} = \sum_i r_{ji} \psi_i^\ell, \quad j = 1, 2, \ldots, N_{\ell-1}. \tag{1.10}$$

Consider again the weak formulation of the variational problem (1.7):

$$a(u^{\ell,k,1} + v, w) = (f, w)_0 \quad \text{for } w \in S_{\ell-1}, \tag{1.7}'$$

or[12]

$$a(v, w) = (d, w)_0 \quad \text{for } w \in S_{\ell-1}.$$

[12] The connection with the following calculation will be somewhat clearer if we make use of a little (unneeded) formalism. It is important for nonconforming problems because the bilinear forms on S_ℓ and $S_{\ell-1}$ can be different. For the iteration at level ℓ, the bilinear form $a(\cdot, \cdot)$ is first defined on S_ℓ, and then passing to $S_{\ell-1}$, we can formally include the injection

$$a(jv, jw) = \langle j^* d, jw \rangle \quad \text{for } w \in S_{\ell-1}.$$

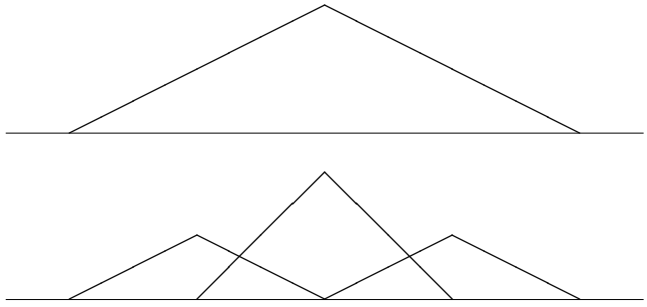

Fig. 51. Decomposition of a nodal basis function on the coarse grid (top) in terms of nodal basis functions on the fine grid (bottom)

Here d is defined by $(d, w)_0 := (f, w)_0 - a(u^{\ell,k,1}, w)$. In particular, $d = 0$ if $u^{\ell,k,1}$ is a solution at level ℓ. As in the derivation of (II.4.4), we successively insert $w = \psi_j^{\ell-1}$ for $j = 1, 2, \ldots, N_{\ell-1}$, and immediately take into account (1.10):

$$\sum_i r_{ji}\, a(u^{\ell,k,1} + v,\ \psi_i^\ell) = \sum_i r_{ji}\, (f, \psi_i^\ell)_0, \quad j = 1, 2, \ldots N_{\ell-1}.$$

We recall that $u^{\ell,k,1} = \sum_t x_t^{\ell,k,1} \psi_t^\ell$. Next we set $v = \sum_s y_s^{\ell-1} \psi_s^{\ell-1}$, and return to the basis of S_ℓ:

$$\sum_t \sum_{i,s} r_{ji}\, a(\psi_s^\ell, \psi_i^\ell)\, r_{st}\, y_t^{\ell-1} = \sum_i r_{ji} \Big[(f, \psi_i^\ell)_0 - \sum_t a(\psi_t^\ell, \psi_i^\ell) x_t^{\ell,k,1} \Big],$$
$$j = 1, 2, \ldots, N_{\ell-1}. \qquad (1.11)$$

The expression in the square brackets is just the i-th component of the residue d_ℓ defined in (1.9). Thus, (1.11) is the componentwise version of the equation $r A_\ell r^t\, y^{\ell-1} = r d_\ell$, and (1.9) follows with $p = r^t$. □

For completeness we note that the vector representation of the approximate solution after the coarse-grid correction (1.8) is

$$x^{\ell,k,2} = x^{\ell,k,1} + p\, y^{\ell-1}. \qquad (1.12)$$

In practice we usually compute the prolongation and restriction matrices via interpolation. Let $\{\psi_i^\ell\}$ be a nodal basis for S_ℓ. Then we have N_ℓ points z_t^ℓ with

$$\psi_i^\ell(z_t^\ell) = \delta_{it}, \quad i, t = 1, 2, \ldots, N_\ell.$$

For every $v \in S_\ell$, $v = \sum_i v(z_i^\ell)\, \psi_i^\ell$, and so $\psi_j^{\ell-1} = \sum_i \psi_j^{\ell-1}(z_i^\ell)\, \psi_i^\ell$ for the basis functions of $S_{\ell-1}$. Comparing coefficients with (1.10), we get

$$r_{ji} = \psi_j^{\ell-1}(z_i^\ell). \qquad (1.13)$$

This is a convenient description of the restriction matrix.

The matrix r is simplest and easiest to find for variational problems in $\mathbb{R}^1$; see Fig. 51. For piecewise linear functions,

$$r = \begin{pmatrix} \frac{1}{2} & 1 & \frac{1}{2} & & & & & \\ & \frac{1}{2} & 1 & \frac{1}{2} & & & & \\ & & \frac{1}{2} & 1 & \frac{1}{2} & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & \frac{1}{2} & 1 & \frac{1}{2} & \end{pmatrix}. \tag{1.14}$$

For affine families of finite elements, we only need to compute the coefficients for one (reference) element. In particular, for piecewise linear triangular elements,

$$r_{ji} = \begin{cases} 1 & \text{if } z_j^{\ell-1} = z_i^\ell, \\ \frac{1}{2} & \text{if } z_i^\ell \text{ is not a grid point of } \mathcal{T}_{\ell-1}, \\ & \text{but is a neighbor in } \mathcal{T}_\ell \text{ of } z_j^{\ell-1}, \\ 0 & \text{otherwise.} \end{cases}$$

If the variables are numbered as in the model Example II.4.3 on a rectangular grid, then the operators can be expressed as stencils. For the example we have

$$p = \begin{bmatrix} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \end{bmatrix}_*. \tag{1.15}$$

Note that r is always a sparse matrix, and thus there is never a need to store a full matrix. Frequently, it is given only in operator form, i.e., we have a procedure for computing the vector $rx^\ell \in \mathbb{R}^{N_{\ell-1}}$ for any given $x^\ell \in \mathbb{R}^{N_\ell}$. The way in which the nodes of the grid $\mathcal{T}_{\ell-1}$ are related to those of $\mathcal{T}_\ell$, and the way they are numbered, are critical to the efficiency of an algorithm.

For completeness, we describe the multigrid algorithm once again, paying more attention to the computational details. This formulation can also be used for difference methods. Suppose we are given the smoothing $\mathcal{S} = \mathcal{S}_\ell$, the restriction $r = r_\ell$, and the prolongation $p = p_\ell$.

1.7 Multigrid Iteration MGM$_\ell$ *(k-th cycle at level $\ell \geq 1$ in matrix-vector form):* Let $x^{\ell,k}$ be a given approximation in S_ℓ.

1. *Pre-Smoothing.* Carry out ν_1 smoothing steps:

$$x^{\ell,k,1} = \mathcal{S}^{\nu_1} x^{\ell,k}. \tag{1.16}$$

2. *Coarse-Grid Correction.* Compute the residue $d_\ell = b_\ell - A_\ell x^{\ell,k,1}$ and the restriction $b_{\ell-1} = rd_\ell$. Let

$$A_{\ell-1}\hat{y}^{\ell-1} = b_{\ell-1}.$$

If $\ell = 1$, find the solution, and set $y^{\ell-1} = \hat{y}^{\ell-1}$.
If $\ell > 1$, compute an approximation $y^{\ell-1}$ of $\hat{y}^{\ell-1}$ by carrying out μ steps of $\mathbf{MGM}_{\ell-1}$ with the starting value $x^{\ell-1,0} = 0$.
Set

$$x^{\ell,k,2} = x^{\ell,k,1} + p\, y^{\ell-1}.$$

3. *Post-Smoothing.* Carry out ν_2 smoothing steps

$$x^{\ell,k,3} = \mathcal{S}^{\nu_2} x^{\ell,k,2},$$

and set $x^{\ell,k+1} = x^{\ell,k,3}$. □

The development of the multigrid method for nonconforming elements and for saddle point methods is more complicated. For nonconforming elements, usually $S_{2h} \not\subset S_h$. Then the prolongation and restriction operators have to be specially computed. These elements have frequently been used as a model for the nonconforming P_1 element; see Brenner [1989] and Braess and Verfürth [1990]. On the other hand, for mixed methods, it is not clear at the outset how to select suitable smoothing operators. Previously, Jacobi smoothing was frequently applied to the squared system, as, e.g., in Verfürth [1988]. More recently, so-called transforming smoothing has provided a completely different approach, see Brandt and Dinar [1979], Wittum [1989], Braess and Sarazin [1997]. Moreover Bank, Welfert, and Yserentant [1990] propose a smoother which is based on the Uzawa algorithm.

Problems

1.8 Suppose we apply the Jacobi method in Example 1.1 with $\omega = \frac{2}{3}$. With $n = 32$, what frequencies are damped by less than a factor of 2 after ten iterations?

1.9 Suppose we are given a series of data points $\{y_i\}_{i=1}^{2n+1}$ which are the values of a function at equally spaced grid points, but suppose we only need data on a grid with double the mesh size. Suppose that in carrying out the elimination process we want to eliminate as much measurement error as possible. Which of the following three procedures does the best job of smoothing (here $i = 1, 2, \ldots, n$):

(a) $z_i = y_{2i}$,
(b) $z_i = \frac{1}{2}(y_{2i-1} + y_{2i+1})$,
(c) $z_i = \frac{1}{4}(y_{2i-1} + 2y_{2i} + y_{2i+1})$?

In performing the analysis, express the values cyclically, and compare the Fourier coefficients of $\{z_i\}$ and $\{y_i\}$.

1.10 Write Algorithm 1.7 as a formal procedure $\mathbf{MGM}_\ell(A_\ell, b_\ell, x^{\ell,k})$ in PASCAL or some other programming language.

1.11 What is the (one-dimensional) stencil of the restriction operator with the matrix representation (1.14)?

1.12 Dividing the squares into triangles in the model Example II.4.3 results in an unsymmetric stencil for the prolongation in (1.15). The symmetric form

$$p = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}_* \tag{1.17}$$

appears more natural. What are the stencils for the restriction operators which arise from (1.15) and (1.17) via the matrix equation $r = p^t$?

§ 2. Convergence of Multigrid Methods

A multigrid method is said to have *multigrid convergence* if the error is reduced by a factor of at least $\rho < 1$ in each iteration cycle, where ρ is *independent of* h. In this case the convergence as $h \to 0$ cannot be arbitrarily slow, in contrast to classical iterative methods. The factor ρ is called the *convergence rate*. Clearly, it is a measure of the speed of convergence.

Independently from Fedorenko [1961], Hackbusch [1976] and Nicolaides [1977] also presented convergence proofs. Here we make use of an idea employed by Bank and Dupont [1981] in their proof. A general framework due to Hackbusch [1985] admits to break convergence proofs into two separate parts. In this way they become very transparent. A *smoothing property*

$$\|\mathcal{S}^\nu v_h\|_X \le c\, h^{-\beta} \frac{1}{\nu^\gamma} \|v_h\|_Y \tag{2.1}$$

is combined with an *approximation property*

$$\|v_h - u_{2h}\|_Y \le c\, h^\beta \|v_h\|_X, \tag{2.2}$$

where u_{2h} is the coarse-grid approximation of v_h. Then for large ν, the product of the two factors is smaller than 1 and independent of h. In particular, it follows that the convergence rate tends to zero for large numbers of smoothing steps.

The various proofs differ in the choice of the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, where $\|\cdot\|_X$ generates a stronger topology than $\|\cdot\|_Y$. The pair (2.1) and (2.2) have to fit together in exactly the same way as the approximation property (II.6.20) and the inverse estimate (II.6.21). It is clear that we need two norms, or more generally two measures, for specifying the error. In addition to measuring the *size* of the error (w.r.t. whichever norm), we also have to measure the *smoothness* of the error function.

It is the goal of this section to establish convergence of the two-level iteration under the following hypotheses:

2.1 Hypotheses.

(1) The boundary-value problem is H^1- or H_0^1-elliptic.
(2) The boundary-value problem is H^2-regular.
(3) The spaces S_ℓ belong to a family of conforming finite elements with uniform triangulations, and the spaces are nested, i.e., $S_{\ell-1} \subset S_\ell$.
(4) We use nodal bases.

For most of our discussion we can get by with weaker hypotheses, but in that case the proofs become more technical. For example, $H^{1+\alpha}$-regularity with $\alpha > 0$ would suffice for this section; see Problem 2.12.

There is a different theory due to Bramble, Pasciak, Wang, and Xu [1991] and Xu [1992]. Multigrid algorithms are connected with a decomposition of finite element spaces. The space decomposition method has the advantage that it does not require regularity assumptions. On the other hand, it does not model the fact that the convergence rate is improved by increasing the number of smoothing steps and it applies only to the energy norm. Since the arguments of the theory are far away from the finite element theory in this book, we restrict ourselves in §5 to the motivation and the proof of the central tool.

We will also not discuss quantitative results on the convergence rate obtained by Fourier methods [Brandt 1977, Ries, Trottenberg, and Winter 1983].

Discrete Norms

So far, the quality of the approximation of a function by finite elements has been expressed in terms of higher Sobolev norms. This no longer works in dealing with the approximation of a function $v_h \in S_h$ by a function in the space S_{2h} on the coarse grid. In particular, if S_h consists of C^0 elements, then in general $S_h \not\subset H^2(\Omega)$, and we cannot employ estimates in the H^2 norm.

This leads us to assign another Hilbert scale to the N_h-dimensional space S_h. The new scale should be connected with the scale of the Sobolev spaces as closely as possible. – It is easiest to discuss how to do this in the following abstract form.

For a symmetric positive definite matrix A, the powers A^s are well defined for all values of s, and not just for integers. Thus, we go back to the spectral decomposition. The matrix A has a complete system of orthonormal eigenvectors $\{z_i\}_{i=1}^N$:

$$\begin{aligned} Az_i &= \lambda_i z_i, \qquad i = 1, 2, \ldots, N, \\ (z_i, z_j) &= \delta_{ij}, \qquad i, j = 1, 2, \ldots, N. \end{aligned}$$

Every vector $x \in \mathbb{R}^N$ can be written in the form

$$x = \sum_{i=1}^{N} c_i z_i, \tag{2.3}$$

and

$$A^s x = \sum_{i} c_i \lambda_i^s z_i \tag{2.4}$$

is well defined.

2.2 Definition. Let A be a symmetric positive definite $N \times N$ matrix, and suppose $s \in \mathbb{R}$. Then

$$|||x|||_s := (x, A^s x)^{1/2} \tag{2.5}$$

defines a norm, where $(\cdot, \cdot)$ is the Euclidean scalar product in $\mathbb{R}^N$.

Using (2.3), (2.4), and the orthogonality relation, we have

$$(x, A^s x) = \Big(\sum_k c_k z_k, \sum_i c_i \lambda_i^s z_i\Big) = \sum_i \lambda_i^s c_i^2.$$

Thus the norm (2.5) has the following alternative representation:

$$|||x|||_s = \Big(\sum_{i=1}^N \lambda_i^s |c_i|^2\Big)^{1/2} = \|A^{s/2} x\|. \tag{2.6}$$

2.3 Properties of the Norm (2.5).

(1) *Connection with the Euclidean norm:* $|||x|||_0 = \|x\|$, where $\|\cdot\|$ is the Euclidean norm.

(2) *Logarithmic convexity:* For $r, t \in \mathbb{R}$ and $s = \frac{1}{2}(r + t)$,

$$|||x|||_s \le |||x|||_r^{1/2} \cdot |||x|||_t^{1/2},$$
$$|(x, A^s y)| \le |||x|||_r \cdot |||y|||_t.$$

Indeed, with the help of the Cauchy–Schwarz inequality, it follows that

$$|(x, A^s y)| = |(A^{r/2} x, A^{t/2} y)| \le \|A^{r/2} x\| \, \|A^{t/2} y\| = |||x|||_r \, |||y|||_t .$$

This is the second inequality. The first follows if we choose $x = y$. Taking the logarithm of both sides and using the continuity, we see that the function

$$s \longmapsto \log |||x|||_s$$

is convex provided that $x \neq 0$.

(3) *Monotonicity:* Let α be the constant of ellipticity, i.e., $(x, Ax) \ge \alpha(x, x)$. Then

$$\alpha^{-t/2} |||x|||_t \ge \alpha^{-s/2} |||x|||_s, \quad \text{for } t \ge s.$$

For the special case $\alpha = 1$, this follows immediately from (2.6) and $\lambda_i \ge \alpha = 1$. Otherwise we have the monotonicity property for the normalized matrix $\alpha^{-1} A$ which implies the monotonicity as stated for A.

(4) *Shift theorem.* The solution of $Ax = b$ satisfies

$$|||x|||_{s+2} = |||b|||_s$$

for all $s \in \mathbb{R}$. This follows from $(x, A^{s+2} x) = (Ax, A^s Ax) = (b, A^s b)$. □

Using the scale defined with (2.5), we immediately get the following property of Richardson relaxation without any additional hypotheses. It can be thought of as a smoothing property, as we shall see later.

2.4 Lemma. *Let $\omega \geq \lambda_{\max}(A)$, $s \in \mathbb{R}$ and $t > 0$, and consider the iteration*

$$x^{\nu+1} = (1 - \frac{1}{\omega}A)x^{\nu}.$$

Then

$$|||x^{\nu}|||_{s+t} \leq c\nu^{-t/2}|||x^0|||_s,$$

where $c = \left(\frac{t\omega}{2e}\right)^{t/2}$.

Proof. If x^0 is expanded as in (2.3), then $x^{\nu} = \sum_i (1 - \lambda_i/\omega)^{\nu} c_i z_i$, and since $0 < \lambda_i/\omega \leq 1$ we have

$$\begin{aligned} |||x^{\nu}|||_{s+t}^2 &= \sum_i \lambda_i^{s+t} \left[(1 - \frac{\lambda_i}{\omega})^{\nu} c_i\right]^2 \\ &= \omega^t \sum_i \left(\frac{\lambda_i}{\omega}\right)^t \left(1 - \frac{\lambda_i}{\omega}\right)^{2\nu} \lambda_i^s c_i^2 \\ &\leq \omega^t \max_{0\leq\zeta\leq 1}\{\zeta^t(1-\zeta)^{2\nu}\} \sum_i \lambda_i^s c_i^2. \end{aligned} \tag{2.7}$$

To compute the maximum appearing in (2.7), we examine the function $\zeta(1-\zeta)^p$ in the interval [0,1] for $p > 0$. It attains its maximum at $\zeta = 1/(p+1)$. Thus,

$$\zeta(1-\zeta)^p \leq \frac{1}{p+1}(\frac{p}{p+1})^p = \frac{1}{p}\frac{1}{(1+\frac{1}{p})^{p+1}} \leq \frac{1}{p}\frac{1}{e}$$

in [0,1]. With $p = 2\nu/t$, it follows that $\max\{\zeta^t(1-t)^{2\nu}\} \leq (t/[2e\nu])^t$. Since the sum appearing in (2.7) is exactly $|||x^0|||_s^2$, the proof is complete. □

The assertion of Lemma 2.4 is independent of the choice of basis. For the following results, this happens only under some additional hypotheses.

Connection with the Sobolev Norm

In Example 1.1, smooth eigenfunctions are associated with the small eigenvalues of A, while eigenfunctions with strongly oscillating terms are associated with the large eigenvalues. This fact naturally depends on the choice of basis. As we will see, this holds in general if we select the nodal basis for affine families, because in this case, the norm $|||\cdot|||_0$ is equivalent to the Sobolev norm $\|\cdot\|_{0,\Omega}$.

2.5 Lemma. *Let $\mathcal{T}_h$ be a family of uniform partitions of $\Omega \subset \mathbb{R}^n$, and suppose S_h belongs to an affine family of finite elements. Suppose we normalize the functions in the nodal basis so that*

$$\psi_i(z_j) = h^{-n/2}\delta_{ij}. \tag{2.8}$$

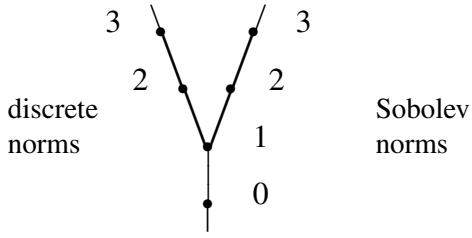

Fig. 52. Connection between the scales defined by discrete norms and Sobolev norms

Then for $v_h \in S_h$, $|||v_h|||_0 = [h^n \sum_i v_h(z_i)^2]^{1/2}$, *which is just the Euclidean norm of the coefficient vector relative to the basis* $\{\psi_i\}$. *Moreover, there exists a constant* c *independent of* h *such that*

$$c^{-1}\|v_h\|_{0,\Omega} \le |||v_h|||_0 \le c\|v_h\|_{0,\Omega}. \tag{2.9}$$

To establish (2.9), we make use of Problem II.6.12: On the reference element $T_{\text{ref}} \subset \mathbb{R}^n$, $\|p\|^2_{0,T_{\text{ref}}}$ and $\sum_i p(z_i)^2$ are equivalent. Now from the properties of affine transformations, we conclude that for every $T \in \mathcal{T}_h$, the quantities $\|v_h\|^2_{0,T}$ and $h^n \sum_i v_h(z_i)^2$ are equivalent, where the sum runs over the nodes belonging to T. The factor h^n enters through the transformation of the domain. Summing over the elements gives (2.9). □

In practice, we always normalize the basis functions so that $\psi_i(z_j) = \delta_{ij}$. We can ignore the difference in normalization factors since there is no essential change in the following results if the system matrix is multiplied by a constant factor.

We now further restrict ourselves to second order elliptic problems. For $v_h \in S_h$, we define $|||v_h|||_s$ by associating with v_h its coefficients relative to a nodal basis (2.8), and define the scale according to Definition 2.2 using the stiffness matrix A_h. Then in particular,

$$|||v_h|||_1^2 = (v_h, A_h v_h) = a(v_h, v_h),$$

and because of the ellipticity of the bilinear form a, it follows that

$$c^{-1}\|v_h\|_{1,\Omega} \le |||v_h|||_1 \le c\|v_h\|_{1,\Omega}. \tag{2.10}$$

More precisely, every function $v_h \in S_h$ is identified with its coefficient vector relative to the nodal basis. This makes sense since $\|\cdot\|_{s,\Omega}$ and $|||\cdot|||_s$ are equivalent for $s = 0$ and $s = 1$ by (2.9) and (2.10), respectively. For $s > 1$, this is no longer the case (see Fig. 52).

As a consequence of the equivalence, we get estimates of the largest and smallest eigenvalues, as well as the condition number of the stiffness matrix.

2.6 Lemma. *Suppose that the hypotheses of Lemma 2.5 hold. Then there is a constant c independent of h such that*

$$\lambda_{\min}(A_h) \geq c^{-1}, \quad \lambda_{\max}(A_h) \leq ch^{-2}, \quad \kappa(A_h) \leq c^2 h^{-2} \tag{2.11}$$

where A_h is the system matrix corresponding to any H^1- or H_0^1-elliptic problem.

Proof. For positive definite matrices, the eigenvalues can be estimated in terms of the Rayleigh quotients. Using the inverse estimate II.6.8, we get $\|v_h\|_{1,\Omega} \leq ch^{-1}\|v_h\|_{0,\Omega}$, and

$$\lambda_{\max}(A_h) = \sup_x \frac{(x, A_h x)}{(x, x)} = \sup_{v_h \in S_h} \frac{|||v_h|||_1^2}{|||v_h|||_0^2} \leq c \sup_{v_h \in S_h} \frac{\|v_h\|_{1,\Omega}^2}{\|v_h\|_{0,\Omega}^2} \leq ch^{-2}.$$

Similarly, using $\|v_h\|_{1,\Omega} \geq \|v_h\|_{0,\Omega}$, we get

$$\lambda_{\min}(A_h) = \inf_x \frac{(x, A_h x)}{(x, x)} = \inf_{v_h \in S_h} \frac{|||v_h|||_1^2}{|||v_h|||_0^2} \geq c^{-1} \inf_{v_h \in S_h} \frac{\|v_h\|_{1,\Omega}^2}{\|v_h\|_{0,\Omega}^2} \geq c^{-1}.$$

Finally, the third assertion follows from $\kappa(A_h) = \lambda_{\max}(A_h)/\lambda_{\min}(A_h)$. □

These estimates are sharp. The exponent of h^{-2} in (2.11) cannot be improved, since by Remark II.6.10, there exist functions in S_h for which $\|v_h\|_1 \approx ch^{-1}\|v_h\|_0$.

We also note that by the proof of the lemma, for every eigenfunction ϕ_h, the ratio $\|\phi_h\|_1^2/\|\phi_h\|_0^2$ gives the corresponding eigenvalue up to a constant. This shows that the oscillating eigenfunctions correspond to the large eigenvalues. In this connection, the situation is exactly the same as for the model Example 1.1.

Using Lemma 2.4 for $s = 0$, $t = 2$ and substituting the estimate (2.11) for $\lambda_{\max}$, we immediately get

2.7 Corollary. *(Smoothing Property) The iteration $x^{\nu+1} = (1 - \frac{1}{\omega}A_h)x^\nu$ with $\omega = \lambda_{\max}(A_h)$ satisfies*

$$|||x^\nu|||_2 \leq \frac{c}{\nu} h^{-2}\|x^0\|_0. \tag{2.12}$$

Approximation Property

The quality of the coarse-grid correction in S_{2h} can be expressed in terms of the $|||\cdot|||_2$ norm, and the results resemble those in Ch. II, §7 if we replace the Sobolev norm $\|\cdot\|_{2,\Omega}$ by $|||\cdot|||_2$. The essential tool here is the duality argument of Aubin–Nitsche. By the approximation results of Ch. II, §6, in estimating the $\|\cdot\|_1$ norm of the error in terms of the $\|\cdot\|_2$ norm, we gain one power of h. In Corollary II.7.7, this gain was "propagated downwards" on the right-hand branch of the scale in Fig. 52. Specifically, it was shown that the same improvement occurs in estimating the $\|\cdot\|_0$ norm by the $\|\cdot\|_1$ norm. We now carry out the same process in the reverse direction, moving to the scale with the discrete norm, see Braess and Hackbusch [1983].

2.8 Lemma. *Given $v_h \in S_h$, let u_{2h} be the solution of the weak equation*

$$a(v_h - u_{2h}, w) = 0 \quad \textit{for all } w \in S_{2h}.$$

In addition, let Ω be convex or have a smooth boundary. Then

$$\|v_h - u_{2h}\|_{1,\Omega} \le c\,2h|||v_h|||_2, \tag{2.13}$$

$$\|v_h - u_{2h}\|_{0,\Omega} \le c\,2h\|v_h - u_{2h}\|_{1,\Omega}. \tag{2.14}$$

Proof. By hypothesis, the problem is H^2-regular, and by Corollary II.7.7, (2.14) holds. In addition, recalling property 2.3(2), and a well-known argument from the proof of Céa's lemma, we have

$$\begin{aligned}\alpha\|v_h - u_{2h}\|_1^2 &\le a(v_h - u_{2h}, v_h - u_{2h}) = a(v_h - u_{2h}, v_h) = (v_h - u_{2h}, A_h v_h)\\ &\le |||v_h - u_{2h}|||_0\,|||v_h|||_2 \le c_1\|v_h - u_{2h}\|_0\,|||v_h|||_2\\ &\le c_1 c\,2h\,\|v_h - u_{2h}\|_1\,|||v_h|||_2\,.\end{aligned}$$

Dividing by $\|v_h - u_{2h}\|_1^2$, we get (2.13).

Second proof. Let $g \in S_h$ be defined by $(g, w)_{0,\Omega} = a(v_h, w)$ for all $w \in S_h$. Consider the auxiliary variational problem. Find $z \in H_0^1$ such that

$$a(z, w) = (g, w)_{0,\Omega} \quad \text{for all } w \in H_0^1(\Omega).$$

Obviously, v_h and u_{2h} are the finite element approximations of z in S_h an S_{2h}, respectively. From the H^2-regularity we conclude that

$$\|v_h - u_{2h}\|_{0,\Omega} \le \|v_h - z\|_{0,\Omega} + \|z - u_{2h}\|_{0,\Omega} \le ch^2\|u\|_{2,\Omega} \le ch^2\|g\|_{0,\Omega}.$$

On the other hand, we have $(A_h v_h, w) = a(v_h, w) = (g, w)_{0,\Omega}$ for $w \in S_h$, and the equivalence of the Euclidean norm and the L_2-norm yields

$$\|g\|_{0,\Omega}^2 = (A_h v_h, g) \le \|A_h v_h\|\,\|g\| \le |||v_h|||_2\,c\|g\|_{0,\Omega}.$$

After dividing by $\|g\|_{0,\Omega}$ and inserting the bound of $\|g\|_{0,\Omega}$ into the preceding estimate, we have

$$\|v_h - u_{2h}\|_{0,\Omega} \le ch^2|||v_h|||_2$$

which is the desired estimate. □

We note that the second proof can be extended more easily to the nonconforming case, since v and u_{2h} are separated early in the proof by the application of the triangle inequality; cf. Braess and Verfürth [1992], Brenner [1991], and the axiomatic considerations by Braess, Dryja, and Hackbusch [1999].

Convergence Proof for the Two-Grid Method

The last two lemmas show that the smoothing property and the approximation property given in (2.1) and (2.2) in an abstract (general) form hold with the choices

$$\|\cdot\|_X = |||\cdot|||_2\,, \quad \|\cdot\|_Y = |||\cdot|||_0\,, \quad \beta = 2, \quad \text{and } \gamma = 1. \tag{2.15}$$

The following proof makes clear the fundamental importance of the properties (2.1) and (2.2). To keep the formalism from obscuring the key ideas, we give a concrete proof. – Note that for the two-grid method, the finer grid corresponds to the level $\ell = 1$, where $u_1 = u_{\ell_1}$ is the desired solution.

2.9 Convergence Theorem. *Under Hypotheses 2.1, the two-grid method using Jacobi relaxation (with* $\lambda_{\max}(A_h) \le \omega \le c'\,\lambda_{\max}(A_h)$*) satisfies*

$$\|u^{1,k+1} - u_1\|_{0,\Omega} \le \frac{c}{\nu_1}\|u^{1,k} - u_1\|_{0,\Omega}\,,$$

where c is a constant independent of h, and ν_1 *is the number of pre-smoothings.*

Proof. For smoothing with Richardson relaxation,

$$u^{1,k,1} - u_1 = (1 - \frac{1}{\omega}A_h)^{\nu_1}(u^{1,k} - u_1).$$

By Lemma 2.6,

$$|||u^{1,k,1} - u_1|||_2 \le \frac{c}{\nu_1}h^{-2}\|u^{1,k} - u_1\|_{0,\Omega}. \tag{2.16}$$

By definition of the coarse-grid correction, $u^{1,k,2} = u^{1,k,1} + u_{2h}$ is characterized by

$$a(u^{1,k,1} + u_{2h}, w) = (f, w)_{0,\Omega} \quad \text{for all } w \in S_{2h}.$$

Moreover, the solution on level 1 satisfies the equation $a(u_1, w) = (f, w)_{0,\Omega}$ for $w \in S_h$. Since $S_{2h} \subset S_h$, subtracting the two equations gives

$$a(u^{1,k,1} - u_1 + u_{2h}, w) = 0 \quad \text{for all } w \in S_{2h}.$$

Now applying Lemma 2.8 to $v := u^{1,k,1} - u_1$, we get

$$\|u^{1,k,2} - u_1\|_{0.\Omega} \le c\,h^2|||u^{1,k,1} - u_1|||_2. \tag{2.17}$$

We can deal with the post-smoothing in a very rough way. Clearly, $|||(1 - \frac{1}{\omega}A_h)x|||_s \le |||x|||_s$. This implies that $|||u^{1,k,3} - u_1|||_0 \le |||u^{1,k,2} - u_1|||_0$, and because of the equivalence of the norms, we have

$$\|u^{1,k,3} - u_1\|_{0,\Omega} \le c\|u^{1,k,2} - u_1\|_{0,\Omega}. \tag{2.18}$$

Now combining (2.16)–(2.18) and taking into consideration $u^{1,k+1} = u^{1,k,3}$, we get the assertion. □

An Alternative Short Proof

If we are content with a convergence rate which is $\mathcal{O}(\nu^{-1/2})$, there is a much shorter proof. Then we do not need to introduce the scale of discrete norms (2.3), but the smoothing property of the matrix operations is less transparent without this scale. We refer to that proof for completeness; we need only rearrange some results known from the previous investigation.

The smoothing property and the approximation property also hold with

$$\|\cdot\|_X = \|\cdot\|_1, \quad \|\cdot\|_Y = \|\cdot\|_0, \quad \beta = 1, \text{ and } \gamma = \frac{1}{2}.$$

Indeed, the approximation property (2.13)

$$\|v - u_{2h}\|_0 \le ch\|v - u_{2h}\|_1 \le ch\|v\|_1$$

is immediate from the Aubin–Nitsche lemma. Moreover, when we apply Lemma 2.4 to the case $s = 0$, $s + t = 1$, we may prove the smoothing property

$$\|x^\nu\|_1 \le c\,\nu^{-1/2}\|x^0\|_0$$

without reference to the discrete norms. The rest of proof proceeds as for Theorem 2.9. □

Some Variants

It is easy to see the connection with the somewhat different terminology of Hackbusch [1985]. The matrix representation of the two-grid iteration is

$$u^{1,k+1} - u_1 = M(u^{1,k} - u_1) \tag{2.19}$$

with

$$\begin{aligned} M &= \mathcal{S}^{\nu_2}(I - pA_{2h}^{-1}rA_h)\mathcal{S}^{\nu_1} \\ &= \mathcal{S}^{\nu_2}(A_h^{-1} - pA_{2h}^{-1}r)A_h\mathcal{S}^{\nu_1}. \end{aligned} \tag{2.20}$$

In particular, $pA_{2h}^{-1}rA_h$ describes the coarse-grid correction for a two-grid method. Writing the smoothness and approximation properties in the form

$$\|A_h\mathcal{S}^\nu\| \le \frac{c}{\nu}h^{-2}, \quad \|(A_h^{-1} - pA_{2h}^{-1}r)\| \le ch^2 \tag{2.21}$$

and using $\|\mathcal{S}\| \le 1$, we get the contraction property $\|M\| \le c/\nu < 1$ for sufficiently large ν. Since $|||Ax|||_0 = |||x|||_2$, the smoothing and approximation properties follow from (2.21) with the same norms and parameters as in (2.15).

The smoothing property is usually established as in Lemma 2.4 whenever the convergence of the multigrid method is carried out in Hilbert spaces. The proof of convergence w.r.t. the maximum norm by Reusken [1992] is different.

2.10 Reusken's Lemma. *Let $\|\cdot\|$ be a matrix norm which is associated with a vector norm. Moreover, assume that $B = I - G^{-1}A$ satisfies*

$$\|B\| \le 1. \tag{2.22}$$

Then for $\mathcal{S} = \frac{1}{2}(I + B)$,

$$\|A\mathcal{S}^\nu\| \le \sqrt{\frac{8}{\pi\nu}}\,\|G\|.$$

Thus, the smoothing property is obtained if we can find a matrix G such that (2.22) holds and $\|G\|$ can be estimated by the same power of h as $\|A\|$. It is interesting to note that in view of (2.22), B almost defines a convergent iterative process. However, for the purposes of smoothing, we take the average with the old vector. Note that $A\mathcal{S}^\nu = 2^{-\nu}G(I - B)(I + B)^\nu$. The proof makes use of the formula $\|(I-B)(I+B)^\nu\| \le 2^{\nu+1}\sqrt{\frac{2}{\pi\nu}}$, which in turn is verified via the Binomial equation. □

Finally, we note that Lemma 2.8 cannot be applied to nonconforming elements or to nonnested spaces (i.e. if $S_{2h} \not\subset S_h$). In order to establish the approximation property in these cases, the duality argument has to be modified; see Brenner [1989], Braess and Verfürth [1990]. Let $r_h \in S_h$ be a representation of the residual defined by $(r_h, w)_0 = \langle d, w\rangle := (f, w)_0 - a(u^{\ell,k,1}, w)$ for $w \in S_h$. Then $u_\ell - u^{\ell,k,1}$ and $v_{\ell-1}$ are the finite element approximations in S_ℓ and $S_{\ell-1}$, respectively, of

$$a(z, w) = (r_h, w)_0 \quad \text{for all } w \in H_0^1(\Omega). \tag{2.23}$$

Note that $\|u_\ell - v_{\ell-1} - u^{\ell,k,1}\| \le \|u_\ell - u^{\ell,k,1} - z\| + \|v_{\ell-1} - z\|$, and the terms on the right-hand side are obtained from standard error estimates. The duality technique is needed to get sharp estimates. The representation of the residual in terms of $r_h \in S_\ell$ is chosen in order to use the discrete scale.

Problems

2.11 Show that for the scale of the Sobolev spaces, the analog

$$\|v\|_{s,\Omega}^2 \le \|v\|_{s-1,\Omega}\,\|v\|_{s+1,\Omega}$$

of (2.5) holds for $s = 0$ and $s = 1$.

2.12 Hypothesis 2.1 requires H^2-regularity. For many problems with reentrant corners, we only have $H^{3/2}$-regularity, and instead of (2.13),

$$\|v - u_{2h}\|_{1,\Omega} \le c\,h^{1/2}\,|||v|||_{3/2}.$$

Verify the smoothing and approximation properties for $\|\cdot\|_X = |||\cdot|||_{3/2}$ and $\|\cdot\|_Y = |||\cdot|||_{1/2}$.

2.13 The iteration in the two-grid algorithm 1.3 is a linear process with the iteration matrix M given in (2.20). Show that the spectral radius of M depends only on the sum $\nu_1 + \nu_2$, and not on how many a priori or a posteriori smoothings are carried out. – Does this also hold for several levels?

2.14 For every $\omega \le 1/\lambda_{\max}$, the Jacobi iteration (1.4) leads to smoothing. Show that for $\omega = 1.9/\lambda_{\max}$, it may converge faster as a stand-alone iteration [if $\kappa(A) > 20$], but is less effective as a smoother.

2.15 Let A be positive definite, and suppose $B = (I + A)^{-1}A$. Show that in the scale generated by A,

$$|||Bx|||_s \le |||x|||_s$$

for all $s \in \mathbb{R}$.

2.16 Suppose that for $s = 2$, two positive definite matrices A and B generate equivalent norms, i.e. norms which differ by at most a factor c. [For $s = 0$ they are trivially equivalent.] Show that the norms are equivalent for all s with $0 < s < 2$. – For $s > 2$ the assertion does not hold in general.

2.17 The smoothing property (2.12) for the Richardson iteration can be generalized to the solution of the saddle point problem

$$\begin{pmatrix} A & B^T \\ B & \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}.$$

There is an iteration such that the successive approximants stay in $ker\, B$. Let $\omega \ge \lambda_{\max}(A)$ and consider the iteration

$$\begin{pmatrix} u^{\nu+1} \\ p^{\nu+1} \end{pmatrix} = \begin{pmatrix} u^{\nu} \\ p^{\nu} \end{pmatrix} + \begin{pmatrix} \omega I & B^T \\ B & \end{pmatrix}^{-1} \begin{pmatrix} f - Au^{\nu} - B^T p^{\nu} \\ -Bu^{\nu} \end{pmatrix} \tag{2.24}$$

for $\nu = 0, 1, \ldots$ Let Q be the projector $Q := I - B^T(BB^T)^{-1}B$ and $M := Q(1 - \frac{1}{\omega}A)Q$. Show the following properties:

(1) $Bu^{\nu} = 0$ for $\nu > 0$.

(2) $u^{\nu+1}$ and $p^{\nu+1}$ depend only on u^{ν} but do not depend on p^{ν}.

(3) $u^{\nu+1} - u = Q(I - \frac{1}{\omega}A)\,(u^{\nu} - u)$.

(4) $A(u^{\nu} - u) + B^T(p^{\nu} - u) = \omega(I - \frac{1}{\omega}A)M(I - M)^{\nu-2}(I - \frac{1}{\omega}A)\,(u^0 - u)$.

Since $\|(I - \frac{1}{\omega}A)u\| \le \|u\|$ and the spectrum of M is contained in $[-1, +1]$, the last equation provides the smoothing property for the iteration (2.24).

The application of (2.24) to a multigrid algorithm for the Stokes problem was proposed by Braess and Sarazin [1996]; see also Zulehner [2000].

§ 3. Convergence for Several Levels

In the convergence theorem in the previous section we assumed that the coarse-grid correction was computed exactly. For algorithms using more than two levels, this is no longer the case. In this case we can think of the multigrid algorithm as a *perturbed two-grid algorithm*. It suffices to estimate the *size* of the perturbation; we don't need to know its details.

The goal of this section is to compute the convergence rate ρ_ℓ for the algorithm with ℓ changes of grid levels, i.e., with $\ell + 1$ levels:

$$\|u^{\ell,k+1} - u_\ell\| \leq \rho_\ell \, \|u^{\ell,k} - u_\ell\|. \tag{3.1}$$

Here u_ℓ is the solution in S_ℓ. We assume that we know the convergence rate ρ_1 for the two-grid method. We will find the rate ρ_ℓ from the rate ρ_1 by induction. First, we let $\|\cdot\|$ be an arbitrary norm.

Later, we sharpen the results by specializing to the energy norm, and in particular obtain convergence of the multigrid method already for a single smoothing step. Thus, we can dispense with the hypothesis that sufficiently many smoothing steps are carried out.

A Recurrence Formula for the W-Cycle

For smoothing with the Richardson method, clearly

$$\|u^{\ell,k,1} - u_\ell\| \leq \|u^{\ell,k} - u_\ell\|, \tag{3.2}$$

assuming an underlying discrete norm $\|\cdot\| := |||\cdot|||_s$. In the following, we will always assume that (3.2) is satisfied, since this property also holds in other important cases.

We compare the result $u^{\ell,k,2}$ of the actual coarse-grid correction with the exact coarse-grid correction $\hat{u}^{\ell,k,2}$. By (3.1), the two-grid rate is ρ_1, i.e.,

$$\|\hat{u}^{\ell,k,2} - u_\ell\| \leq \rho_1 \|u^{\ell,k} - u_\ell\|. \tag{3.3}$$

Together with (3.2), the triangle inequality yields

$$\|u^{\ell,k,1} - \hat{u}^{\ell,k,2}\| \leq (1 + \rho_1) \, \|u^{\ell,k} - u_\ell\|. \tag{3.4}$$

The left-hand side of (3.4) gives the size of the coarse-grid correction with exact computations. The real correction differs from the exact one by the error at level

$\ell - 1$. Thus, by the induction hypothesis, the relative error is at most $\rho_{\ell-1}^{\mu}$. Here $\mu = 1$ for the V-cycle, and $\mu = 2$ for the W-cycle, as usual. Hence,

$$\|u^{\ell,k,2} - \hat{u}^{\ell,k,2}\| \leq \rho_{\ell-1}^{\mu} \, \|u^{\ell,k,1} - \hat{u}^{\ell,k,2}\|. \tag{3.5}$$

We now substitute (3.4) into (3.5). Using (3.3), we see that for the W-cycle,

$$\|u^{\ell,k,2} - u_\ell\| \leq [\rho_1 + \rho_{\ell-1}^2(1 + \rho_1)] \, \|u^{\ell,k} - u_\ell\|.$$

Without post-smoothing, we have $u^{\ell,k+1} = u^{\ell,k,2}$, and (3.1) holds with rate ρ_ℓ, which can be estimated by (3.6).

3.1 Recurrence Formula. *For the multigrid method using the W-cycle, at level $\ell \geq 2$ we have*

$$\rho_\ell \leq \rho_1 + \rho_{\ell-1}^2(1 + \rho_1). \tag{3.6}$$

Formula (3.6) leads to an estimate of the convergence rate independent of ℓ, provided ρ_1 is sufficiently small.

3.2 Theorem. *Suppose the two-grid rate is $\rho_1 \leq \frac{1}{5}$. Then for the W-cycle,*

$$\rho_\ell \leq \frac{5}{3}\rho_1 \leq \frac{1}{3}, \quad \textit{for } \ell = 2, 3, \ldots \tag{3.7}$$

Proof. For $\ell = 1$ the assertion is clear. It follows from the assertion for $\ell - 1$ and the recurrence formula (3.6) that

$$\rho_\ell \leq \rho_1 + \frac{1}{3}\,(\frac{5}{3}\rho_1)\,(1 + \frac{1}{5}) = \frac{5}{3}\rho_1 \leq \frac{1}{3}. \qquad \square$$

By Theorem 2.9, the convergence rate for the two-grid method is indeed smaller than 1/5 for sufficiently many smoothing steps. Now Theorem 3.2 implies the convergence of the multigrid method under the same hypothesis.

An Improvement for the Energy Norm

When referring to the energy norm, the recurrence formula (3.6) for the convergence rate can be replaced by a significantly better one. With respect to this norm, the exact coarse-grid correction yields the orthogonal projection of $u^{\ell,k,1} - u_\ell$ onto the subspace $S_{\ell-1}$. The error $\hat{u}^{\ell,k,2} - u_\ell$ after the exact coarse-grid correction is therefore orthogonal to $S_{\ell-1}$. In particular, it is then orthogonal to $u^{\ell,k,1} - \hat{u}^{\ell,k,2}$ and to $u^{\ell,k,2} - \hat{u}^{\ell,k,2}$ (see Fig. 53).

Thus, the estimate (3.4) can be replaced by

$$\|u^{\ell,k,1} - \hat{u}^{\ell,k,2}\|^2 = \|u^{\ell,k,1} - u_\ell\|^2 - \|\hat{u}^{\ell,k,2} - u_\ell\|^2.$$

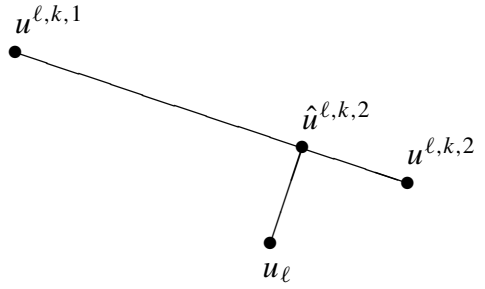

Fig. 53. The coarse-grid correction as an orthogonal projection

In addition, it follows from the orthogonality and (3.5) that

$$\begin{aligned}\|u^{\ell,k,2} - u_\ell\|^2 &= \|\hat{u}^{\ell,k,2} - u_\ell\|^2 + \|u^{\ell,k,2} - \hat{u}^{\ell,k,2}\|^2 \\ &\le \|\hat{u}^{\ell,k,2} - u_\ell\|^2 + \rho_{\ell-1}^{2\mu}\|u^{\ell,k,1} - \hat{u}^{\ell,k,2}\|^2 \\ &= (1 - \rho_{\ell-1}^{2\mu})\|\hat{u}^{\ell,k,2} - u_\ell\|^2 + \rho_{\ell-1}^{2\mu}\|u^{\ell,k,1} - u_\ell\|^2. \qquad (3.8)\end{aligned}$$

Now we make use of our knowledge of the two-grid rate. By (3.2),

$$\|u^{\ell,k,2} - u_\ell\|^2 \le [(1 - \rho_{\ell-1}^{2\mu})\rho_1^2 + \rho_{\ell-1}^{2\mu}]\|u^{\ell,k} - u_\ell\|^2.$$

Thus, (3.1) holds with a rate which can be estimated by (3.9).

3.3 A Recurrence Formula. *The multigrid method with $\mu = 1$ for the V-cycle and $\mu = 2$ for the W-cycle satisfies*

$$\rho_\ell^2 \le \rho_1^2 + \rho_{\ell-1}^{2\mu}(1 - \rho_1^2) \qquad (3.9)$$

at level $\ell \ge 2$ with respect to the energy norm.

3.4 Theorem. *If the two-grid rate with respect to the energy norm satisfies $\rho_1 \le \frac{1}{2}$, then*

$$\rho_\ell \le \frac{6}{5}\rho_1 \le 0.6, \quad \text{for } \ell = 2, 3, \ldots \qquad (3.10)$$

for the W-cycle.

Proof. For $\ell = 1$ there is nothing to prove. By the assertion for $\ell - 1$, it follows from the recurrence formula (3.9) that

$$\begin{aligned}\rho_\ell^2 &\le \rho_1^2 + (\frac{6}{5}\rho_1)^4(1 - \rho_1^2) = \rho_1^2\{1 + (\frac{6}{5})^4[\rho_1^2(1 - \rho_1^2)]\} \\ &\le \rho_1^2\{1 + (\frac{6}{5})^4 \ \frac{1}{4} \ \frac{3}{4}\} \le \frac{36}{25}\rho_1^2 \le 0.36.\end{aligned}$$

Taking the square root, we get the desired result. □

The recurrence formula (3.9) gives unsatisfactory results for the V-cycle, as can be seen from the rapidly growing values of ρ_ℓ in Table 8.

Table 8. Convergence rates ρ_ℓ as in (3.9) for $\rho_1 = 1/5$.

	$\ell =$ 1	2	3	4	5	$\sup_\ell \rho_\ell$
W-cycle	0.2	0.2038	0.2041	0.2042	0.2042	0.2042
V-cycle	0.2	0.280	0.340	0.389	0.430	1.0

The Convergence Proof for the V-cycle

It is possible to establish a bound smaller than 1 for the convergence rate for the V-cycle, independent of the number of levels. To this end, we need a refinement of the method of proof which has an additional advantage: it shows that just one smoothing step suffices. Of course, the result also applies to the W-cycle.

The analysis of the two-grid method in §2 was based in essence on the scale of $|||\cdot|||_s$ norms for s in $[0, 2]$. Since we intend to make use of the energy norm, in the following we have to stay between $s = 1$ and $s = 2$. This halves the span between the maximal and minimal s. Here we present a simplified version of the original proof of Braess and Hackbusch [1983], although we obviously get weaker results with larger numbers of smoothing steps. (See, however, the remark at the end of this section.)

As before, we write $\|\cdot\|$ instead of $|||\cdot|||_1$ for the energy norm.

3.5 Convergence Theorem. *Under Hypotheses 2.1, the multigrid method with V-cycles or W-cycles satisfies*

$$\begin{aligned} &\|u^{\ell,k+1} - u_\ell\| \le \rho_\ell \, \|u^{\ell,k} - u_\ell\|, \\ &\rho_\ell \le \rho_\infty := \Big(\frac{c}{c+2\nu}\Big)^{1/2}, \quad \ell = 0, 1, 2, \ldots, \end{aligned} \tag{3.11}$$

w.r.t. the energy norm, assuming that Jacobi relaxation with $\lambda_{\max}(A) \le \omega \le c_0 \, \lambda_{\max}(A)$ *is performed. Here* c *is a constant independent of* ℓ *and* ν.

Before presenting the proof, we establish some lemmas. For abbreviation, let

$$w^m := u^{\ell,k,m} - u_\ell, \quad m = 0, 1, 2, \tag{3.12}$$

and define $\hat{w}^2$ in a similar way.

We now introduce a measure of the smoothness of the functions in the finite element space S_h:

$$\beta = \beta(v_h) := \begin{cases} 1 - \lambda_{\max}^{-1}(A) \dfrac{|||v_h|||_2^2}{|||v_h|||_1^2}, & \text{if } v_h \neq 0, \\ 0, & \text{if } v_h = 0. \end{cases} \tag{3.13}$$

Clearly, $0 \le \beta < 1$. Smooth functions correspond to a number β near 1, and functions with a large oscillating part correspond to a small β. The factor β determines the amount of improvement for each smoothing step. Since the improvement always becomes successively smaller during the smoothing process, the size of the factor β *after* smoothing is decisive.

3.6 Lemma. *Smoothing using Jacobi relaxation satisfies*

$$\|\mathcal{S}^\nu v\| \le [\beta(\mathcal{S}^\nu v)]^\nu \, \|v\| \quad \textit{for all } v \in S_h.$$

Proof. Let $v = \sum_i c_i \phi_i$, where $\phi_1, \phi_2, \ldots$ are orthonormal eigenfunctions of A. In addition, let $\mu_i = 1 - \lambda_i/\lambda_{\max}$. By Hölder's inequality,

$$\sum_i \lambda_i \mu_i^{2\nu} |c_i|^2 \le (\sum_i \lambda_i \mu_i^{2\nu+1} |c_i|^2)^{\frac{2\nu}{2\nu+1}} (\sum_i \lambda_i |c_i|^2)^{\frac{1}{2\nu+1}}.$$

In view of (2.6), this inequality is equivalent to

$$\|\mathcal{S}^\nu v\|^{2\nu+1} \le \|\mathcal{S}^{\nu+\frac{1}{2}} v\|^{2\nu} \, \|v\|. \tag{3.14}$$

We abbreviate $w := \mathcal{S}^\nu v$, and divide (3.14) by $\|w\|^{2\nu}$. This gives

$$\|\mathcal{S}^\nu v\| \le \left(\frac{\|\mathcal{S}^{1/2} w\|}{\|w\|} \right)^{2\nu} \|v\|. \tag{3.15}$$

Since $\mathcal{S}$ is self-adjoint and commutes with A, we also have

$$\begin{aligned} \|\mathcal{S}^{1/2} w\|^2 &= (\mathcal{S}^{1/2} w, A\mathcal{S}^{1/2} w) = (w, ASw) \\ &= (w, Aw) - \frac{1}{\lambda_{\max}} (w, A^2 w) = \beta(w) \, \|w\|^2. \end{aligned}$$

The assertion follows after substitution in (3.15). □

The quality of the coarse-grid correction can also be estimated in terms of the parameter β.

3.7 Lemma. *For the exact coarse-grid correction, we have*

$$\begin{aligned} \|\hat{w}^2\| &\le \min\{c_1 \, \lambda_{\max}^{-1/2} |||w^1|||_2, \, \|w^1\|\} \\ &= \min\{c_1 \sqrt{1 - \beta(w^1)}, \, 1\} \, \|w^1\|. \end{aligned} \tag{3.16}$$

Proof. The inequality (2.13) asserts that $\|\hat{w}^2\| \le ch \, |||w^1|||_2$. Then using $\lambda_{\max} \le ch^{-2}$, we can eliminate the factor h to get the bound $c_1 \lambda_{\max}^{-1/2} |||w^1|||_2$. Moreover, the energy norm of the error is not increased by the coarse-grid correction, and the first assertion is proved. If we use (3.13) to eliminate $|||w^1|||_2$, we get the second assertion. □

The following formula is of central importance for the proof of the convergence theorem:

3.8 Recurrence Formula. *Suppose the hypotheses of Theorem 3.5 are satisfied. Then*

$$\rho_\ell^2 \le \max_{0\le\beta\le 1} \beta^{2\nu}\,[\rho_{\ell-1}^{2\mu} + (1-\rho_{\ell-1}^{2\mu})\,\min\{1, c_1^2(1-\beta)\}]. \tag{3.17}$$

Here $\mu = 1$ for the V-cycle and $\mu = 2$ for the W-cycle, respectively, and c_1 is the constant in Lemma 3.7.

Proof. By Lemma 3.6,

$$\|u^{\ell,k,1} - u_\ell\| \le \beta^\nu \|u^{\ell,k} - u_\ell\|,$$

where $\beta = \beta(u^{\ell,k,1} - u_\ell)$. By Lemma 3.7 with the same β, we have

$$\begin{aligned}\|\hat{u}^{\ell,k,2} - u_\ell\| &\le \min\{c_1\sqrt{1-\beta}, 1\}\,\|u^{\ell,k,1} - u_\ell\| \\ &\le \beta^\nu \min\{c_1\sqrt{1-\beta}, 1\}\,\|u^{\ell,k} - u_\ell\|.\end{aligned}$$

We now insert this estimate in (3.8) to get

$$\|u^{\ell,k,2} - u_\ell\|^2 \le \beta^{2\nu}[(1-\rho_{\ell-1}^{2\mu})\,\min\{c_1^2(1-\beta), 1\} + \rho_{\ell-1}^{2\mu}]\,\|u^{\ell,k} - u_\ell\|^2.$$

Since $0 \le \beta < 1$, this proves the recurrence formula. □

Table 9. Convergence rate ρ_ℓ according to the recurrence formula (3.17) for $\nu = 2$

	V-cycle									W-cycle
c	$\ell =$ 1	2	3	4	5	6	7	8	∞	
0.5	.1432	.174	.189	.199	.205	.210	.214	.217	.243	.1437
1	.2862	.340	.366	.382	.392	.400	.406	.410	.448	.2904

Proof of Theorem 3.5. We will verify (3.11) with $c := c_1^2$ and c_1 from (3.17). Since $\rho_0 = 0$, (3.11) holds for $\ell = 0$. To show how to get from $\ell - 1$ to ℓ, we insert $\rho_{\ell-1}^2 \le c/(c+2\nu)$ in the recurrence formula (3.17) and note that the function $\rho \mapsto \rho^2 + a(1-\rho^2)$ is nondecreasing on $[0, 1]$ whenever $0 \le a \le 1$:

$$\begin{aligned}\rho_\ell^2 &\le \max_{0\le\beta\le 1}\{\beta^{2\nu}\,[\frac{c}{c+2\nu} + (1-\frac{c}{c+2\nu})\ \min\{1, c_1^2(1-\beta)\}]\} \\ &\le \max_{0\le\beta\le 1}\{\beta^{2\nu}\,[\frac{c}{c+2\nu} + (1-\frac{c}{c+2\nu})\ c(1-\beta)]\} \\ &= \frac{c}{c+2\nu}\max_{0\le\beta\le 1}\{\beta^{2\nu}\,[1+2\nu(1-\beta)]\}.\end{aligned} \tag{3.18}$$

Simple differentiation with respect to β shows that the maximum in (3.18) is attained for $\beta = 1$. Setting $\beta = 1$ we see that $\rho_\ell^2 \le \frac{c}{c+2\nu}$. □

In the multigrid method, the low frequency parts are handled more efficiently in the W-cycle than in the V-cycle. Hence, it is not surprising that in (3.18) the maximum is attained for $\beta = 1$. Thus, it makes sense to insert W-cycles once in a while if the number of levels is very large.

The proof shows that for large ν, the contraction number decreases only like

$$\nu^{-1/2}.$$

If both pre-smoothing and post-smoothing are used, the rate of decrease is ν^{-1}, as shown by the duality technique of Braess and Hackbusch [1983]. We remark that later convergence proofs for the V-cycle usually make use of an algebraic hypothesis instead of the H^2-regularity; cf. Bramble, Pasciak, Wang, and Xu [1991]. The question of whether appropriate algebraic hypotheses are really independent of H^2-regularity remains open, despite the paper of Parter [1987].

If the regularity hypothesis 2.1(2) is not satisfied, we have to expect a less favorable convergence rate. Then as suggested by Bank, Dupont, and Yserentant [1988], it makes more sense to use the multigrid method as a preconditioner for a CG method rather than as a stand-alone iteration. In fact these authors go one step further, and use the multigrid idea only for the construction of a so-called *hierarchical basis*. Then the convergence rate behaves like $1 - \mathcal{O}((\log \frac{1}{h})^{-p})$, which is still quite good for practical computations.

Problems

3.9 Show that for the W-cycle,

$$\sup_{\ell} \rho_\ell^2 \le \frac{\rho_1^2}{1-\rho_1^2}, \quad \text{provided} \quad \rho_1 < \sqrt{\frac{1}{2}}.$$

Hint: Use (3.9) to derive a recurrence formula for $1 - \rho_\ell^2$.

3.10 Show that for large c, the recurrence formula (3.17) gives the two-grid rate

$$\rho_1 \le (1 - \frac{1}{c^2})^\nu,$$

and compare with (3.11).

3.11 The amount of computation required for the W-cycle is approximately 50 % larger than for the V-cycle. Use the tables to compare the error reduction of three V-cycles and two W-cycles.

§ 4. Nested Iteration

So far we have treated the multigrid method as a pure iterative method. However, it turns out that the multigrid idea can also be used to find good starting values, so that one or two cycles of the multigrid iteration suffice. For this purpose there are two essential ideas.

1. The solution at level $\ell - 1$ is a good starting point for the iteration at level ℓ. This idea can be carried still further: we do not need to compute the function exactly since an approximation already provides a reasonable starting value.

2. The finite element approximation u_h is subject to the discretization error $\|u_h - u\|$. Thus, it makes little sense to carry out a lot of steps of the multigrid iteration to get an accuracy corresponding to the roundoff error. Instead, we should stop the iteration when

$$\|u^{h,k} - u_h\| \leq \frac{1}{2}\|u_h - u\|. \tag{4.1}$$

We will get only a marginal improvement of the total error $\|u^{h,k} - u\|$ by going any further.

Using the above ideas, we will create algorithms such that the amount of computation grows only linearly with the number of unknowns.

The starting value calculation is based on certain precursors of the multigrid method. The solution on the $2h$ grid was used as a starting value for classical iterative methods. Although the corresponding starting error has relatively small low frequency terms, classical relaxation methods still require too many steps; cf. Problem 4.6.

Computation of Starting Values

The following method for computing a starting value based on the multilevel concept is called *nested iteration*.

4.1 Algorithm NI$_\ell$ *for computing a starting value* v^ℓ *at level* $\ell \geq 0$.
If $\ell = 0$, find $v^0 = u_0 = A_0^{-1} b_0$, and exit the procedure.
Let $\ell > 0$.
Find an approximate solution $v^{\ell-1}$ of the equation $A_{\ell-1} u_{\ell-1} = b_{\ell-1}$ by applying **NI**$_{\ell-1}$.
Compute the prolongation of $v^{\ell-1}$, and set $v^{\ell,0} = p\, v^{\ell-1}$.

Using $v^{\ell,0}$ as a starting value, carry out one step (in general $q \geq 1$ steps) of the multigrid iteration $\mathbf{MGM}_\ell$ (see Fig. 54), and set

$$v^\ell = v^{\ell,1}. \qquad \square$$

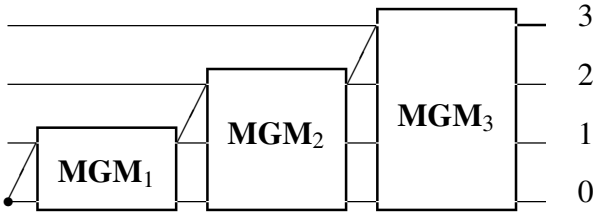

Fig. 54. Nested iteration $\mathbf{NI}_3$

For simplicity, in the following we assume that in Algorithm 4.1 only one cycle of the multigrid method is carried out. This is actually done for convergence rates $\rho < \frac{1}{4}$. Otherwise, we formally identify q cycles with one cycle having the convergence rate ρ^q, where q is such that $\rho^q < \frac{1}{4}$.

The accuracy of the resulting starting value can be computed easily. We restrict ourselves to the common case where the discretization error is of order $\mathcal{O}(h^2)$.

4.2 Theorem. *Assume that the finite element approximation $u_h \in S_h$ satisfies $\|u_h - u\| \leq c\,h^2$ for some constant $c > 0$. In addition, suppose the convergence rate ρ of the multigrid method w.r.t. the norm $\|\cdot\|$ is smaller than $1/4$. Then Algorithm 4.1 gives*

$$\|v^\ell - u_\ell\| \leq \frac{5\rho}{1-4\rho}\, c\, h_\ell^2. \tag{4.2}$$

Proof. Since $v^0 = u_0$, the formula (4.2) is immediate for $\ell = 0$.

Assuming (4.2) holds for $\ell - 1$, the fact that $h_{\ell-1} = 2h_\ell$ implies

$$\|v^{\ell-1} - u_{\ell-1}\| \leq \frac{5\rho}{1-4\rho}\, c(2h_\ell)^2.$$

Moreover, by the hypothesis on the discretization error,

$$\|u_{\ell-1} - u\| \leq c\,(2h_\ell)^2, \qquad \|u_\ell - u\| \leq c\,h_\ell^2\,.$$

Now the triangle inequality and $v^{\ell,0} = v^{\ell-1}$ give

$$\begin{aligned}\|v^{\ell,0} - u_\ell\| &\leq \|v^{\ell-1} - u_{\ell-1}\| + \|u_{\ell-1} - u\| + \|u - u_\ell\| \\ &\leq \frac{5\rho}{1-4\rho} 4c\,h_\ell^2 + 5c\,h_\ell^2 = \frac{5}{1-4\rho}\, c\,h_\ell^2. \end{aligned} \tag{4.3}$$

The multigrid cycle reduces the error by the factor ρ, i.e., $\|v^{\ell,1} - u_\ell\| \leq \rho\|v^{\ell,0} - u_\ell\|$. Using (4.3), we get the assertion (4.2). $\square$

4.3 Remark. In many cases a convergence rate of $\rho \leq 1/6$ is realistic. Then Algorithm 4.1 produces an approximation with error $(5/2)c\,h_\ell^2$, and one additional cycle suffices to reduce the error to less than $\frac{1}{2}c\,h_\ell^2$.

Complexity

To estimate the computational complexity, we can assume that the amount of computation for

$$\left.\begin{array}{l}\text{smoothing in } S_\ell,\\ \text{prolongation of } S_{\ell-1} \text{ to } S_\ell, \text{ and}\\ \text{restriction of the residue } d_\ell\end{array}\right\} \tag{4.4}$$

all involve cN_ℓ operations, where $N_\ell = \dim S_\ell$. The number of arithmetic operations for one smoothing step is proportional to the number of nonzero elements in the system matrix. For affine families of finite elements on uniform grids, this number is proportional to the number of unknowns. The prolongation and restriction matrices are even sparser. Thus, the operation count at level ℓ is

$$\leq (\nu + 1)\, c\, N_\ell, \tag{4.5}$$

where ν is the number of smoothings.

For grids in $\mathbb{R}^2$, we collect the operations (4.4) from all levels. Each time we move to a coarser grid the number of unknowns decreases by approximately the factor 4. Summing the terms (4.5) for the multigrid iteration $\mathbf{MGM}_\ell$ gives

$$\begin{aligned}(\nu+1)\,c\,(N_\ell + N_{\ell-1} + N_{\ell-2} + \cdots) &\leq \frac{4}{3}(\nu+1)\,c\,N_\ell \quad \text{for the V-cycle,}\\ (\nu+1)\,c\,(N_\ell + 2N_{\ell-1} + 4N_{\ell-2} + \cdots) &\leq 2(\nu+1)\,c\,N_\ell \quad \text{for the W-cycle.}\end{aligned} \tag{4.6}$$

In addition, we must include the work needed to solve the system of equations on the coarsest grid. We ignore this additional computational effort for the moment. This is justified if the number of levels is large. The special case of a small number of levels will be treated separately later.

The computation of starting values can be analyzed in the same way. Because of the increase in dimension, each cycle of Algorithm 4.1 requires four times as much work as its predecessor. Since $\sum_k 4^{-k} = 4/3$, we have that

The operation count for computing the starting value is $\frac{4}{3}$ times the operation count for one multigrid cycle on the finest grid.

In view of Remark 4.3, the operation count for the complete calculation is $\frac{7}{3}$ times the count for one multigrid cycle on the finest grid, provided the convergence rate is 1/6 or better. Thus, in particular, the complexity increases only linearly with the number of unknowns.

As mentioned before, in the V-cycle we distribute the smoothing steps equally between the phases before and after the coarse-grid correction. Thus, an analogous symmetric variant is recommended in calculating the starting values.

4.4 Algorithm $\mathbf{NI}_\ell$ *to compute a starting value v^ℓ at the level $\ell \geq 0$ (symmetric version).*
If $\ell = 0$, find $v^0 = u_0 = A_0^{-1} b_0$, and exit the procedure.
Let $\ell > 0$.
With $v^{\ell,0} = 0$ carry out one step of the multigrid iteration $\mathbf{MGM}_\ell$, and set $v^{\ell,1} = v^{\ell,0} + pv^{\ell-1}$.
Set $b_{\ell-1} = p(b_\ell - A_\ell v^{\ell,1})$.
Compute an approximate solution $v^{\ell-1}$ of the equation $A_{\ell-1}u_{\ell-1} = b_{\ell-1}$ using $\mathbf{NI}_{\ell-1}$.
Compute the prolongation of $v^{\ell-1}$, and set $v^{\ell,2} = v^{\ell,1} + p\, v^{\ell-1}$.
With $v^{\ell,2}$ carry out one step (in general $p \geq 1$ steps) of the multigrid iteration $\mathbf{MGM}_\ell$, and set

$$v^\ell = v^{\ell,3}. \qquad \square$$

This variant compensates for the effect that in the V-cycle, the oscillating parts are handled better than the smoother ones. A problem whose smooth parts are relatively large is given to the block $\mathbf{NI}_{\ell-1}$, so that the efficiency is increased.

Multigrid Methods with a Small Number of Levels

For many grids it is difficult to carry out more than two levels of coarsening. As we shall see, however, it still pays to use multigrid methods. In contrast to the case of a large number of levels, here the solution on the coarsest grid is a major part of the computational effort. To this effort we have to add the work required for smoothing, restrictions, and prolongations. This is, of course, less than with a large number of levels, and has already been estimated in (4.6) above.

For three levels, the number of unknowns on the coarsest grid is ca. 1/16 of that on the finest. In addition, the average bandwidth of the system matrix is reduced by a factor of about 4. Thus, the amount of work required for the Cholesky method is reduced by a factor $16 \cdot 4^2 = 256$. Now if we compute the starting value with Algorithm 4.1 and add a multigrid cycle, we have to solve

4 systems of equations for the V-cycle,
6 systems of equations for the W-cycle

on the coarsest grid (all with the same matrix). Even if the LU decomposition is recalculated each time, we still get a saving of a factor of 40 to 64, if we compare

just the effort of the exact solver.[13]

Since the other transfer operations discussed earlier hardly affect the result, we clearly see the advantages of the multigrid method, even in those cases where at first glance it appears to be of marginal use.

The CASCADE Algorithm

We also mention the CASCADE algorithm of Bornemann and Deuflhard [1996]. It uses a different strategy. We begin with a CG method on the coarsest grid, and proceed from there successively from level to level to the finest, without ever going back to coarser grids. We choose a much larger number of iteration steps on the coarser grids (with the smaller dimensions) than on the finer grids.

The justification for this approach is the following result:

4.5 Recursion Relation. *Let u_ℓ denote the solution of the variational problem in S_ℓ and v_ℓ be the result of the* CASCADE *algorithm. If m_ℓ steps of the cg-algorithm are performed on the level ℓ, then*

$$\|v_\ell - u_\ell\|_1 \le \|v_{\ell-1} - u_{\ell-1}\|_1 + c\frac{h_\ell}{m_\ell}\|f\|_0. \tag{4.7}$$

From (4.7) we conclude that the error on the finest level is of the order of the discretization error provided that sufficiently many steps are performed on the coarse levels, say $m_\ell = 3^{\ell_{\max}-\ell}\, m_{\ell_{\max}}$. Nevertheless, the main part of the computing time is consumed on the finest level, and the larger number of steps for $\ell < \ell_{\max}$ does not spoil the efficiency.

The recursion relation can be established via the construction of polynomials with approximation properties that differ slightly from those in Ch. IV, §3. For details, see Bornemann and Deuflhard [1996] or Shaidurov [1996]. The investigations have been extended to saddle point problems by Braess and Dahmen [1999] and to nonconforming elements by Stevenson [1999].

For the computation of starting values the CASCADE algorithm is simpler than nested iteration because grid changes occur only in one direction. The algorithm can often replace even the complete multigrid procedure. A clear advantage has the cascadic version of multigrid algoritms for treating *variational inequalities* since the return to coarser grids is more involved there; see Blum, Braess, and Suttmeier [2004].

[13] Another advantage is that we can store the resulting small system of equations in fast memory, while larger systems have to be stored in external memory.

Problems

4.6 Suppose the finite element approximation $u_h \in S_h$ is such that $\|u_h - u\| \le c\, h^2$ for some constant $c > 0$. In addition, suppose the convergence rate for the two-grid method with $\nu = \nu_2$ post-smoothings satisfies $\rho < 1/10$.

Suppose a user (who perhaps is not familiar with the multigrid method) applies a classical relaxation method, starting with the solution on the $2h$ grid. Show that after ν steps,

$$\|u^{h,\nu} - u\| \le \frac{3}{2}\, c \cdot h^2.$$

Why doesn't a corresponding assertion hold for more than two grids?

4.7 Compare the operation counts of
(a) $\mathbf{NI}_\ell$ with the V-cycle,
(b) $\mathbf{NI}_\ell$ with the W-cycle,
(c) the symmetric version of $\mathbf{NI}_\ell$ with the V-cycle.

4.8 Suppose we want to insert a so-called F-cycle (see Fig. 55) between the V-cycle and the W-cycle as follows:

For $\ell = 2$, the F-cycle and W-cycle coincide.

For $\ell \ge 3$, perform both an F-cycle and a V-cycle at level $\ell - 1$.

Find the recurrence formula analogous to (3.17), and determine the rates numerically for $c = 1$, $c = \frac{1}{2}$, $\nu = 2$, and $\ell \le 8$.

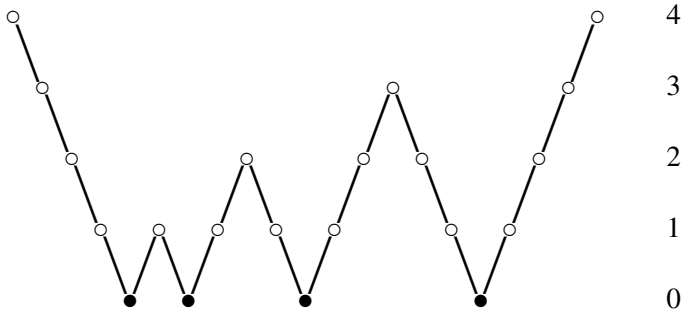

Fig. 55. F-cycle on five levels

4.9 Compare nested iteration $\mathbf{NI}_\ell$ using inner V-cycles with the F-cycle of $\mathbf{MGM}_\ell$. What is the difference?

§ 5. Multigrid Analysis via Space Decomposition

The multigrid analysis in §§2 and 3 is based on a smoothing property and an approximation property. The latter heavily depends on regularity assumptions. There are many variants and generalizations of Lemma 2.8, but one needs $H^{1+\alpha}$ regularity for some $\alpha > 0$.

This is different in the theory of Bramble, Pasciak, Wang, and Xu [1991]. For getting a connection with the previous theory we emphasize a decomposition property in the H^2 regular case although we did not state it explicitly. Let S_ℓ, $\ell = 0, 1, \ldots, L$, be a nested sequence of finite element spaces as in (1.5). Given $v \in S := S_L$, we may decompose it

$$v = \sum_{k=0}^{L} v_k, \quad v_k \in S_k \tag{5.1}$$

such that the partial sum $\sum_{k=0}^{\ell} v_k$ is the finite element solution to v in S_ℓ. It follows from Lemma 2.8 that

$$\|v_k\|_0 \le c h_k \|v_k\|_1, \quad k = 1, 2, \ldots, L.$$

On the other hand, an inverse inequality yields

$$\|v_k\|_1 \le c h_k^{-1} \|v_k\|_0, \quad k = 1, 2, \ldots, L.$$

From these inequalities and the orthogonality of the $v_k's$ we obtain the equivalence

$$\|v\|_1^2 = \sum_{k=0}^{L} \|v_k\|_1^2 \approx \|v_0\|_1^2 + \sum_{k=1}^{L} h_k^{-2} \|v_k\|_0^2. \tag{5.2}$$

A variational problem with a quadratic form that equals the right-hand side of (5.2) can be easily solved by successive solution in the subspaces because of the additive structure. Moreover, the smoothing procedures are efficient approximate solvers for the subspaces since the norms $\|\cdot\|_1$ and $h_k^{-1}\|\cdot\|_0$ are equivalent there. The multigrid method may be interpreted in this way.

Now Bramble, Pasciak, Wang, and Xu [1991] have observed that these and other similar relations may be derived from *properties of the function spaces* and that one does not need *regularity of the solution of the elliptic equation*. Oswald [1994] pointed out that Besov space properties are helpful to understand multilevel

methods in this context; see below, In contrast to §3 the *multilevel iteration* is not treated as a perturbation of the 2-grid procedure.

We want to present the basic ideas of the theory of Bramble, Pasciak, Wang, and Xu. A complete theory without regularity is beyond the scope of this book; rather we demonstrate its advantage for another application which is not covered by the standard theory. The extension to locally refined meshes will be illustrated.

For convenience, we restrict ourselves to symmetric smoothing operators and to nested spaces. For more general results see Xu [1992], Wang [1994], and Yserentant [1993]. All these theories do not reflect the improvement of the convergence rate when the number of smoothing steps is increased. This was only achieved via more involved considerations by Brenner [2000].

The norm without subscript refers to the energy norm $\|\cdot\| := (a(\cdot,\cdot))^{1/2}$ since the theory applies only to this norm.

Schwarz' Alternating Method

For a better understanding of space decomposition methods we first consider the *alternating method* which goes back to H.A. Schwarz [1869]. There is a simple geometrical interpretation, cf. Fig. 56, when the abstract formulation for the case of two subspaces is considered.

We are interested in the variational problem

$$a(u, v) = \langle f, v\rangle \quad \text{for } v \in H.$$

Here $a(.,.)$ is the inner product of the Hilbert space H and $\|\cdot\|$ is the corresponding norm. Let H be the direct sum of two subspaces

$$H = V \oplus W,$$

and the determination of a solution in the subspaces V or W is assumed to be easy. Then an alternating iteration in the two subspaces is natural.

5.1 Schwarz Alternating Method. Let $u_0 \in H$.

When u_{2i} is already determined, find $v_{2i} \in V$ such that

$$a(u_{2i} + v_{2i}, v) = \langle f, v\rangle \quad \text{for } v \in V.$$

Set $u_{2i+1} = u_{2i} + v_{2i}$.

When u_{2i+1} is already determined, find $w_{2i+1} \in W$ such that

$$a(u_{2i+1} + w_{2i+1}, w) = \langle f, w\rangle \quad \text{for } w \in W.$$

Set $u_{2i+2} = u_{2i+1} + w_{2i+1}$. □

Obviously, projections onto the two subspaces alternate during the iteration. The *strengthened Cauchy inequality* (5.3) is crucial in the analysis.

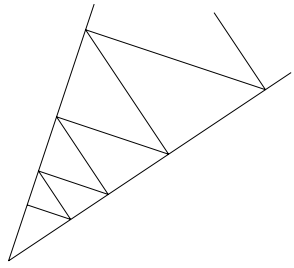

Fig. 56. Schwarz alternating iteration with one-dimensional subspaces V and W in Euclidean 2-space. The iterates $u_1, u_3, u_5, \ldots$ lie in $V^\perp$ and $u_2, u_4, \ldots$ in $W^\perp$. The angle between $V^\perp$ and $W^\perp$ is the same as between V and W.

5.2 Convergence Theorem. *Assume that there is a constant $\gamma < 1$ such that for the inner product in H*

$$|a(v, w)| \leq \gamma \|v\| \, \|w\| \quad \textit{for } v \in V, \ w \in W. \tag{5.3}$$

Then we have for the iteration with the Schwarz alternating method the error reduction

$$\|u_{k+1} - u\| \leq \gamma \|u_k - u\| \quad \textit{for } k \geq 1. \tag{5.4}$$

Proof. Because of the symmetry of the problem we may confine ourselves to even k. Since u_k is constructed by a minimization in the subspace W, we have

$$a(u_k - u, w) = 0 \quad \text{for } w \in W \tag{5.5}$$

We decompose $u_k - u = \hat{v} + \hat{w}$ with $\hat{v} \in V$, $\hat{w} \in W$. From (5.5) it follows with $w = \hat{w}$ that

$$a(\hat{v}, \hat{w}) = -\|\hat{w}\|^2. \tag{5.6}$$

By the strengthened Cauchy inequality (5.3) we have $a(\hat{v}, \hat{w}) = -\alpha_k \|\hat{v}\| \, \|\hat{w}\|$ with some $\alpha_k \leq \gamma$. Without loss of generality let $\alpha_k \neq 0$. It follows from (5.6) that $\|\hat{v}\| = \alpha_k^{-1} \|\hat{w}\|$ and $\|u_k - u\|^2 = \|\hat{v} + \hat{w}\|^2 = \|\hat{v}\|^2 - 2\|\hat{w}\|^2 + \|\hat{w}\|^2 = (\alpha_k^{-2} - 1)\|\hat{w}\|^2$.

Since u_{k+1} is the result of an optimization in V, we obtain an upper estimate from the simple test function $u_k + (\alpha_k^2 - 1)\hat{v}$.

$$\begin{aligned} \|u_{k+1} - u\|^2 &\leq \|u_k + (\alpha_k^2 - 1)\hat{v}\|^2 \\ &= \|\alpha_k^2 \hat{v} + \hat{w}\|^2 = (1 - \alpha_k^2)\|\hat{w}\|^2 = \alpha_k^2 \|u_k - u\|^2. \end{aligned}$$

Noting that $\alpha_k \leq \gamma$, the proof is complete. □

The bound in (5.4) is sharp. This becomes obvious from an example with one-dimensional spaces V and W depicted in Fig. 56 and also from Problem 5.8.

Algebraic Description of Space Decomposition Algorithms

The finite element spaces S_ℓ may be recursively constructed

$$\begin{aligned} S_0 &= W_0, \\ S_\ell &= S_{\ell-1} \oplus W_\ell, \quad \ell \geq 1, \\ S &= S_L. \end{aligned} \tag{5.7}$$

The finite element solution on the level ℓ is related to the operator $A_\ell : S_\ell \to S_\ell$ defined by

$$(A_\ell u, w) = a(u, w) \quad \text{for all } w \in S_\ell. \tag{5.8}$$

Moreover $A := A_L$. The corresponding Ritz projector $P_\ell : S \to S_\ell$ satisfies

$$a(P_\ell u, w) = a(u, w) \quad \text{for all } w \in S_\ell. \tag{5.9}$$

We note that the discussion below holds for any inner product $(\cdot, \cdot)$ in the Hilbert space S, but we will refer to the L_2 inner product or the ℓ_2 inner product when we deal with concrete examples. We recall that the L_2-norm is equivalent to the ℓ_2-norm of the associated vector representations, and the smoothing procedures refer to L_2-like operators. Therefore, we will use also the L_2-orthogonal projectors $Q_\ell : S \to S_\ell$,

$$(Q_\ell u, w) = (u, w) \quad \text{for all } w \in S_\ell. \tag{5.10}$$

It follows that

$$A_\ell P_\ell = Q_\ell A. \tag{5.11}$$

Indeed, for all $w \in S_\ell$ we obtain from (5.8)–(5.10) the equations $(A_\ell P_\ell u, w) = a(P_\ell u, w) = a(u, w) = a(u, Q_\ell w) = (Au, Q_\ell w) = (Q_\ell Au, w)$. Since $A_\ell P_\ell$ and $Q_\ell A$ are mappings into S_ℓ, this proves (5.11).

Assume that $\tilde{u}$ is an approximate solution of the variational problem in S, and let $A\tilde{u} - f$ be the residue. The solution of the variational problem in the subset $\tilde{u} + S_\ell$ is $\tilde{u} + A_\ell^{-1} Q_\ell (f - A\tilde{u})$. Therefore, the correction by the exact solution of the subproblem for the level ℓ is

$$A_\ell^{-1} Q_\ell (f - A\tilde{u}).$$

Since its computation is too expensive in general, the actual correction will be obtained from a computation with an approximate inverse B_ℓ^{-1}, i.e., the real correction will be

$$B_\ell^{-1} Q_\ell (f - A\tilde{u}). \tag{5.12}$$

The correction turns $\tilde{u}$ into $\tilde{u} + B_\ell^{-1} Q_\ell (f - A\tilde{u})$. For convenience, we will assume that

$$A_\ell \leq B_\ell, \tag{5.13}$$

i.e., $B_\ell - A_\ell$ is assumed to be positive semidefinite and tacitly B_ℓ is assumed to be symmetric. Often only the weaker condition $A_\ell \le \omega B_\ell$ with $\omega < 2$ is required, but we prefer to have the assumption without an extra factor in order to avoid some inconvenient factors in the estimates. Some standard techniques for dealing with the approximate solution above are found in Problems IV.4.14–17.

We recall (5.11) and following the standard notation, we define the linear mapping

$$T_\ell := B_\ell^{-1} Q_\ell A = B_\ell^{-1} A_\ell P_\ell. \tag{5.14}$$

From (5.12) we know that the correction of $\tilde{u}$ in the subspace S_ℓ yields the new iterate $\tilde{u} + T_\ell(u - \tilde{u})$, and its error is

$$(I - T_\ell)\ (u - \tilde{u}).$$

We consider the multigrid V-cycle with post-smoothing only. Consequently the error propagation operator for one complete cycle is

$$E := E_L$$

where

$$E_\ell := (I - T_\ell)\ (I - T_{\ell-1}) \dots (I - T_0), \quad \ell = 0, 1, \dots, L, \tag{5.15}$$

and $E_{-1} := I$. This representation elucidates that the subspace corrections are applied in a multiplicative way.

Assumptions

The assumptions refer to the family of finite element spaces S_ℓ and the complementary spaces W_ℓ specified in (5.7).

Assumption A1. There exists a constant K_1 such that for all $v_\ell \in W_\ell$, $\ell = 0, 1, \dots, L$,

$$\sum_{\ell=0}^{L} (B_\ell v_\ell, v_\ell) \le K_1 \| \sum_{\ell=0}^{L} v_\ell \|^2. \tag{5.16}$$

Assumption A2 *(Strengthened Cauchy–Schwarz Inequality).* There exist constants $\gamma_{k\ell} = \gamma_{\ell k}$ with

$$a(v_k, w_\ell) \le \gamma_{k\ell}\, (B_k v_k, v_k)^{1/2} (B_\ell w_\ell, w_\ell)^{1/2} \quad \text{for all } v_k \in S_k,\ w_\ell \in W_\ell \tag{5.17}$$

if $k \le \ell$. Moreover, there is a constant K_2 such that

$$\sum_{k,l=0}^{L} \gamma_{k\ell} x_k y_\ell \le K_2 \Big(\sum_{k=0}^{L} x_k^2 \Big)^{1/2} \Big(\sum_{\ell=0}^{L} y_\ell^2 \Big)^{1/2} \quad \text{for } x, y \in \mathbb{R}^{L+1}. \tag{5.18}$$

We postpone the verification of **A1**. – The verification of **A2** with a constant K_2 that is independent of the number of levels is not trivial. Therefore we provide a short proof of an estimate with a bound that increases only logarithmically. The standard Cauchy–Schwarz inequality and $A_\ell \le B_\ell$ imply that we have $\gamma_{k\ell} \le 1$ for all k, ℓ. Hence,

$$\sum_{k,l} \gamma_{k\ell} x_k y_\ell \le \Big(\sum_k |x_k|\Big)\Big(\sum_\ell |y_\ell|\Big) \le (L+1)\Big(\sum_k x_k^2\Big)^{1/2}\Big(\sum_\ell y_\ell^2\Big)^{1/2},$$

and (5.18) is obvious for

$$K_2 \le L + 1 \le c\,|\log h_L|. \tag{5.19}$$

Direct Consequences

From **A2** we conclude immediately that

$$\begin{aligned}\|\sum_{\ell=0}^{L} v_\ell\|^2 &= \sum_{k,\ell} a(v_k, v_\ell)\\ &\le \sum_{k,\ell} \gamma_{k\ell}(B_k v_k, v_k)^{1/2}\,(B_\ell v_\ell, v_\ell)^{1/2} \le K_2 \sum_{\ell=0}^{L}(B_\ell v_\ell, v_\ell).\end{aligned} \tag{5.20}$$

Hence, the norms encountered in (5.16) and (5.20), are equivalent provided that **A1** and **A2** hold.

A direct consequence of **A1** is an analogue of an inequality which we considered in §2 in connection with logarithmic convexity. Note the asymmetry in the occurrence of the spaces in (5.21).

5.3 Lemma. *Let $w_\ell \in W_\ell$ and $u_\ell \in S = S_L$ for $\ell = 0, 1, \dots, L$. Then we have*

$$\sum_{\ell=0}^{L} a(w_\ell, u_\ell) \le \sqrt{K_1}\,\|\sum_{\ell=0}^{L} w_\ell\|\,\Big(\sum_{\ell=0}^{L} a(T_\ell u_\ell, u_\ell)\Big)^{1/2}. \tag{5.21}$$

Proof. Since $P_\ell w_\ell = w_\ell$, it follows from the Cauchy–Schwarz inequality in Euclidean space that

$$\begin{aligned}\sum_{\ell=0}^{L} a(w_\ell, u_\ell) &= \sum_{\ell=0}^{L} a(w_\ell, P_\ell u_\ell)\\ &= \sum_{\ell=0}^{L} (B_\ell^{1/2} w_\ell, B_\ell^{-1/2} A_\ell P_\ell u_\ell)\\ &\le \Big(\sum_{\ell=0}^{L}(B_\ell w_\ell, w_\ell)\Big)^{1/2}\,\Big(\sum_{\ell=0}^{L}(A_\ell P_\ell u_\ell, B_\ell^{-1} A_\ell P_\ell u_\ell)\Big)^{1/2}.\end{aligned} \tag{5.22}$$

Next we derive an equality that is useful also in other contexts

$$\begin{aligned}(B_\ell T_\ell w, T_\ell w) &= (T_\ell w, B_\ell B_\ell^{-1} A_\ell P_\ell w) \\ &= (T_\ell w, A_\ell P_\ell w) = a(T_\ell w, P_\ell w) = a(T_\ell w, w).\end{aligned} \tag{5.23}$$

The first factor on the right-hand side of (5.22) can be estimated by **A1**. Since $T_\ell = B_\ell^{-1} A_\ell P_\ell$, we can insert (5.23) into the summands of the second factor, and the proof of the lemma is complete. □

It is more than a coincidence that $a(T_\ell w, w)$ is a multiple of the discrete norm $|||P_\ell w|||_2$ that we encountered in §2 if B_ℓ is a multiple of the identity on the subspace S_ℓ.

Convergence of Multiplicative Methods

First we estimate the reduction of the error by the multigrid algorithm on the level ℓ from below.

5.4 Lemma. *Let $\ell \geq 1$. Then*

$$\|v\|^2 - \|(I - T_\ell)v\|^2 \geq a(T_\ell v, v). \tag{5.24}$$

Proof. From the binomial formula we obtain that the left-hand side of (5.24) equals

$$2a(T_\ell v, v) - a(T_\ell v, T_\ell v). \tag{5.25}$$

Next, we consider the second term using $A_\ell \leq B_\ell$ and (5.23)

$$a(T_\ell v, T_\ell v) \leq (B_\ell T_\ell v, T_\ell v) = a(T_\ell v, v).$$

Therefore the negative term in (5.25) can be absorbed by the term $a(T_\ell v, v)$ by subtracting 1 from the factor 2. and the proof is complete. □

Now we turn to the central result of this §. It yields the convergence rate of the multigrid iteration in terms of the constants in the assumptions **A1** and **A2**.

5.5 Theorem. *Assume that* **A1** *and* **A2** *hold. Then the energy norm of the error propagation operator E of the multigrid iteration satisfies*

$$\|E\|^2 \leq 1 - \frac{1}{K_1(1 + K_2)^2}.$$

Proof. By applying Lemma 5.4 to $E_{\ell-1}v$ and noting that $E_\ell = (I - T_\ell)E_{\ell-1}$ we obtain

$$\|E_{\ell-1}v\|^2 - \|E_\ell v\|^2 \geq a(T_\ell E_{\ell-1}v, E_{\ell-1}v).$$

A summation over all levels is performed with telescoping

$$\|v\|^2 - \|Ev\|^2 \geq \sum_{\ell=0}^{L} a(T_\ell E_{\ell-1} v, E_{\ell-1} v). \tag{5.26}$$

Therefore the statement of the theorem will be clear if we verify

$$\|v\|^2 \leq K_1 (1 + K_2)^2 \sum_{\ell=0}^{L} a(T_\ell E_{\ell-1} v, E_{\ell-1} v). \tag{5.27}$$

Indeed, (5.26) and (5.27) yield $\|v\|^2 \leq K_1(1 + K_2)^2(\|v\|^2 - \|Ev\|^2)$, and the rest of the proof is concerned with establishing this inequality.

To this end, let

$$v = \sum_{\ell=0}^{L} v_\ell, \quad v_\ell \in W_\ell,$$

be a (stable) decomposition. Obviously,

$$\|v\|^2 = \sum_{\ell=0}^{L} a(E_{\ell-1} v, v_\ell) + \sum_{\ell=1}^{L} a((I - E_{\ell-1})v, v_\ell). \tag{5.28}$$

Lemma 5.3 is used to deal with the first term

$$\sum_{\ell=0}^{L} a(E_{\ell-1} v, v_\ell) \leq \sqrt{K_1}\, \|v\| \left(\sum_{\ell=0}^{L} a(T_\ell E_{\ell-1} v, E_{\ell-1} v) \right)^{1/2}. \tag{5.29}$$

Next from $E_\ell - E_{\ell-1} = -T_\ell E_{\ell-1}$ it follows by induction that

$$I - E_{\ell-1} = \sum_{k=0}^{\ell-1} T_k E_{k-1}.$$

With the bound of the second term on the right-hand side of (5.28), we verify that the conditions for obtaining an improvement on the level ℓ are not affected much by the corrections in the previous steps. Here **A2** enters and

$$\begin{aligned}
&\sum_{\ell=1}^{L} a((I - E_{\ell-1})v, v_\ell) \\
&\quad = \sum_{\ell=1}^{L} \sum_{k=0}^{\ell-1} a(T_k E_{k-1} v, v_\ell) \\
&\quad \leq \sum_{\ell=1}^{L} \sum_{k=0}^{\ell-1} \gamma_{k\ell} (B_k T_k E_{k-1} v, T_k E_{k-1} v)^{1/2}\, (B_\ell v_\ell, v_\ell)^{1/2} \\
&\quad \leq K_2 \left(\sum_{k=0}^{L} (B_k T_k E_{k-1} v, T_k E_{k-1} v) \right)^{1/2} \left(\sum_{\ell=0}^{L} (B_\ell v_\ell, v_\ell) \right)^{1/2}.
\end{aligned}$$

From (5.23) and **A1** it follows that

$$\sum_{\ell=1}^{L} a((I - E_{\ell-1})v, v_\ell) \le \sqrt{K_1}\, K_2\, \|v\| \left(\sum_{k=0}^{L} a_k(T_k E_{k-1} v, E_{k-1} v)\right)^{1/2}. \quad (5.30)$$

Adding (5.29) and (5.30) and dividing by $\|v\|$ we obtain (5.27) completing the proof. □

Verification of A1

In the case of full H^2 regularity and quasi-uniform triangulations optimal estimates are easily derived. We have an ideal case. Given $v \in S$, let u_ℓ be the finite element solution of v in S_ℓ, i.e., $u_\ell = P_\ell v$. Set

$$\begin{aligned} v &= \sum_{\ell=0}^{L} v_\ell, \\ v_0 = P_0 v, \quad v_\ell &= P_\ell v - P_{\ell-1} v = u_\ell - u_{\ell-1} \quad \text{for } \ell = 1, 2, \ldots, L. \end{aligned} \quad (5.31)$$

From the Galerkin orthogonality of finite element solutions we conclude that

$$\|v\|^2 = \sum_{\ell=0}^{L} \|v_\ell\|^2. \quad (5.32)$$

Since $u_{\ell-1}$ is also the finite element solution to u_ℓ in $S_{\ell-1}$ and $v_\ell = u_\ell - u_{\ell-1}$, it follows from the Aubin–Nitsche lemma that

$$\|v_\ell\|_0 \le c\, h_{\ell-1} \|v_\ell\| \quad \text{for } \ell = 1, 2, \ldots, L. \quad (5.33)$$

The approximate inverses for a multigrid algorithm with Richardson iteration as a smoother are given by

$$B_0 := A_0, \quad B_\ell := c\, \lambda_{\max}(A_\ell) I, \quad \ell = 1, 2, \ldots, L. \quad (5.34)$$

The inverse estimates yield $\lambda_{\max}(A_\ell) \le c h_\ell^{-2}$. Combining these facts and noting $h_{\ell-1} \le c h_\ell$ we obtain

$$\begin{aligned} \sum_{\ell=0}^{L} (B_\ell v_\ell, v_\ell) &\le (A_0 v_0, v_0) + \sum_{\ell=1}^{L} c h_\ell^{-2} (v_\ell, v_\ell) \\ &= \|v_0\|^2 + c \sum_{\ell=1}^{L} h_\ell^{-2} \|v_\ell\|_0^2 \\ &\le c \sum_{\ell=0}^{L} \|v_\ell\|^2 = c\, \|v\|^2. \end{aligned} \quad (5.35)$$

This proves **A1** with a constant $K_1 = c$ that is independent of the number of levels.

In the cases with less regularity we perform the decomposition by applying the L_2-orthogonal projectors Q_ℓ instead of P_ℓ

$$\begin{aligned} v &= \sum_{\ell=0}^{L} v_\ell, \\ v_0 = Q_0 v, \quad v_\ell &= Q_\ell v - Q_{\ell-1} v \quad \text{for } \ell = 1, 2, \ldots, L. \end{aligned} \tag{5.36}$$

From Lemma II.7.9 we have $\|Q_0 v\| \le c\|v\|$. Next from (II.7.15) it follows that $\|v_\ell\|_0 \le \|v - Q_\ell v\|_0 + \|v - Q_{\ell-1} v\|_0 \le c h_\ell \|v\|$. Recalling the approximate solvers from (5.34) we proceed as in the derivation of (5.35)

$$\begin{aligned} \sum_{\ell=0}^{L} (B_\ell v_\ell, v_\ell) &\le \|v_0\|_1^2 + c \sum_{\ell=1}^{L} h_\ell^{-2} \|v_\ell\|_0^2 \\ &\le c\,(L+1)\|v\|^2. \end{aligned} \tag{5.37}$$

This proves **A1** with a constant $K_1 \le c(L+1)^{1/2} \le c|\log h_L|^{1/2}$. Although this result is only suboptimal, it has the advantage that no regularity assumptions are required. As mentioned above, the logarithmic factor arises since we stay in the framework of Sobolev spaces. An analysis with the theory of Besov spaces shows that the factor can be dropped, see Oswald [1994].

Local Mesh Refinements

An inspection of the proof of Lemma II.7.9 shows that the estimate (5.37) remains true if the orthogonal projector Q_ℓ is replaced by an operator of Clément type, e.g., we may choose $I_\ell = I_{h_\ell}$ from (II.6.19). That interpolation operator is nearly local.

This has a big advantage when we consider finite element spaces which arise from local mesh refinements. Assume that the refinement of the triangulation on the level ℓ is restricted to a subdomain $\Omega_\ell \subset \Omega$ and that

$$\Omega_L \subset \Omega_{L-1} \subset \ldots \subset \Omega_0 = \Omega. \tag{5.38}$$

Given $v \in S_L$, its restriction to $\Omega \setminus \Omega_\ell$ coincides there with some finite element function in S_ℓ. Now we modify $I_\ell v$ at the nodes outside Ω_ℓ and set

$$(I_\ell v)(x_j) := v(x_j) \quad \text{if } x_j \notin \Omega_\ell.$$

Specifically, when defining I_ℓ, the construction of the operator $\tilde{Q}_j$ in (II.6.17) is augmented by the rule (II.6.23). We have

$$(I_\ell v)(x) = v(x) \tag{5.39}$$

for x outside a neighborhood of Ω_ℓ, and from problem II.6.17 we know that the modification changes only the constants in the estimates of the L_2-error. The strip of $\Omega \setminus \Omega_\ell$, in which (5.39) does not hold, is small if rule II.8.1(1) is observed during the refinement process. Hence,

$$\|v - I_\ell v\|_0 \le c h_\ell \|v\|_1. \tag{5.40}$$

We note that an estimate of this kind cannot be guaranteed for the finite element solution in S_ℓ. So by using interpolation of Clément type we also obtain multigrid convergence in cases with local mesh refinements.

There is also a consequence for computational aspects of the multigrid method. Since

$$v_{\ell+1} = I_{\ell+1} v - I_\ell v = 0 \ \text{ outside a neighborhood of } \Omega_\ell,$$

the smoothing procedure on the levels $\ell+1,\ \ell+2, \dots, L$ may be restricted to the nodes in a neighborhood of Ω_ℓ. It is not necessary to perform the smoothing iteration at each level on the whole domain. For this reason the computing effort only increases linearly with the dimension of S_L. As was pointed out by Xu [1992] and Yserentant [1993], local refinements induce a faster increase of the computational complexity.

Problems

5.7 Let V, W be subspaces of a Hilbert space H. Denote the projectors onto V and W by P_V, P_W, respectively. Show that the following properties are equivalent:

(1) A strengthened Cauchy inequality (5.3) holds with $\gamma < 1$.
(2) $\|P_W v\| \le \gamma \|v\|$ holds for all $v \in V$.
(3) $\|P_V w\| \le \gamma \|w\|$ holds for all $w \in W$.
(4) $\|v + w\| \ge \sqrt{1 - \gamma^2}\, \|v\|$ holds for all $v \in V$, $w \in W$.
(5) $\|v + w\| \ge \sqrt{\frac{1}{2}(1 - \gamma)}\, (\|v\| + \|w\|)$ holds for all $v \in V$, $w \in W$.

5.8 Consider a sequence obtained by the Schwarz alternating method. Let α_k be the factor in the Cauchy inequality for the decomposition of the error in the iteration step k as in the proof of Theorem 5.2. Show that (α_k) is a nondecreasing sequence.

§ 6. Nonlinear Problems

Multigrid methods are also very useful for the numerical solution of nonlinear differential equations. We need only make some changes in the multigrid method for linear equations. These changes are typical for the efficient treatment of nonlinear problems. However, there is one essential idea involved which we might not otherwise encounter. We have to correct the right-hand side of the nonlinear equation on the coarse grid in order to compensate for the error which arises in moving between grids.

As an example of an important nonlinear differential equation, consider the Navier–Stokes equation

$$\begin{aligned} -\Delta u + \mathrm{Re}\,(u\nabla)u - \mathrm{grad}\, p &= f \quad \text{in } \Omega, \\ \mathrm{div}\, u &= 0 \quad \text{in } \Omega, \\ u &= u_0 \quad \text{on } \partial\Omega. \end{aligned} \tag{6.1}$$

If we drop the quadratic term in the first equation, we get the Stokes problem (III.6.1). Another typical nonlinear differential equation is

$$\begin{aligned} -\Delta u &= e^{\lambda u} \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{6.2}$$

It arises in the analysis of explosive processes. The parameter λ specifies the relation between the reaction heat and the diffusion constant. – Nonlinear boundary conditions are also of interest, in particular for problems in (nonlinear) elasticity.

We write a nonlinear boundary-value problem as an equation of the form $\mathcal{L}(u) = 0$. Suppose that for each $\ell = 0, 1, \ldots, \ell_{\max}$, the discretization at level ℓ leads to the nonlinear equation

$$\mathcal{L}_\ell(u_\ell) = 0 \tag{6.3}$$

with $N_\ell := \dim S_\ell$ unknowns. In the sequel it is often more convenient to consider the formally more general equation

$$\mathcal{L}_\ell(u_\ell) = f_\ell \tag{6.4}$$

with given $f_\ell \in \mathbb{R}^{N_\ell}$.

Within the framework of multigrid methods, there are two fundamentally different approaches:

1. The *multigrid Newton method* (MGNM), which solves the linearized equation using the multigrid method.
2. The *nonlinear multigrid method* (NMGM), which applies the multigrid method directly to the given nonlinear equation.

The Multigrid Newton Method

Newton iteration requires solving a linear system of equations for every step of the iteration. However, it suffices to compute an approximate solution in each step.

The following algorithm is a variant of the damped Newton method. We denote the derivative of the (nonlinear) mapping $\mathcal{L}$ by $D\mathcal{L}$.

6.1 Multigrid Newton Method.
Let $u^{\ell,0}$ be an approximation to the solution of the equation $\mathcal{L}_\ell(u_\ell) = f_\ell$.
For $k = 0, 1, \ldots$, carry out the following calculation:

1. *(Determine the direction)* Set $d^k = f_\ell - \mathcal{L}_\ell(u^{\ell,k})$. Perform one cycle of the algorithm $\mathbf{MGM}_\ell$ to solve

$$D\mathcal{L}_\ell(u^{\ell,k})\, v = d^k$$

with the starting value $v^{\ell,0} = 0$. Call the result $v^{\ell,1}$.

2. *(Line search)* For $\lambda = 1, \frac{1}{2}, \frac{1}{4}, \ldots$, test if

$$\|\mathcal{L}_\ell(u^{\ell,k} + \lambda\, v^{\ell,1}) - f_\ell\| \le (1 - \frac{\lambda}{2})\, \|\mathcal{L}_\ell(u^{\ell,k}) - f_\ell\|. \tag{6.5}$$

As soon as (6.5) is satisfied, stop testing and set

$$u^{\ell,k+1} = u^{\ell,k} + \lambda\, v^{\ell,1}. \qquad \square$$

The direction to the next approximation is determined in the first step, and the distance to go in that direction is determined in the second step. If the approximations are sufficiently close to the solution, then we get $\lambda = 1$. In this case, step 2 can be replaced by the simpler classical method:

2′. Set $u^{\ell,k+1} = u^{\ell,k} + v^{\ell,1}$.

The introduction of the damping parameter λ and the associated test results in a stabilization; see Hackbusch and Reusken [1989]. Thus, the method is less sensitive to the choice of the starting value $u^{\ell,0}$.

It is known that the classical Newton method converges quadratically for sufficiently good starting values, provided the derivative $D\mathcal{L}_\ell$ is invertible at the solution. In Algorithm 6.1 we also have an extra linear error term, and the error $e^k := u^{\ell,k} - u_\ell$ satisfies the following recurrence formula:

$$\|e^{k+1}\| \le \rho\|e^k\| + c\|e^k\|^2.$$

Here ρ is the convergence rate of the multigrid algorithm.

This implies only linear convergence. At first glance this is a disadvantage, but quadratic convergence only happens in a neighborhood of the solution, and in particular only when the error $\|e^k\|$ is smaller than the discretization error. In view of the discussion in the previous section, this is no essential disadvantage.

The Nonlinear Multigrid Method

Methods based on applying the multigrid method directly to the nonlinear equation are often used instead of the multigrid Newton method. Here the calculation again involves smoothing steps and coarse-grid corrections. However, here the latter have a nonlinear character.

The simplest smoothing corresponds to the Jacobi method

$$v^\ell \longmapsto \mathcal{S}_\ell\, v^\ell := v^\ell + \omega[f_\ell - \mathcal{L}_\ell(v^\ell)]. \tag{6.6}$$

As in the linear case, the parameter ω is computed by estimating the largest eigenvalue of $D\mathcal{L}_\ell$.

The so-called *nonlinear Gauss–Seidel method* can be used to perform the smoothing. In order to reduce the amount of formalism, we restrict ourselves to an example, and consider the difference method for the equation (6.2) on a square grid. For the interior points, we have

$$u_i - h^2 e^{\lambda u_i} = G_i(u), \quad i = 1, 2, \ldots N_\ell, \tag{6.7}$$

where $G_i(u)$ is $\frac{1}{4}$ the sum of the values at the neighboring nodes. For each $i = 1, 2, \ldots$, we successively compute a refined value u_i by solving the i-th equation in (6.7) for u_i. This involves solving simple scalar nonlinear equations. More generally, we have

$$[\mathcal{L}_\ell(u_1^{k+1}, \ldots, u_i^{k+1}, u_{i+1}^k, u_{i+2}^k, \ldots)]_i = f_i, \quad i = 1, 2, \ldots$$

The Gauss–Seidel method can also be used as a smoother in the nonlinear case.

The computation of the coarse-grid correction has to be done differently than in the linear case. We emphasize that in general, $u_{\ell-1} \neq r u_\ell$, where u_ℓ and $u_{\ell-1}$ are the finite element solutions in S_ℓ and $S_{\ell-1}$, respectively. This is why so far the coarse-grid correction has only been applied to the defect equation. Here we do something different. In passing between grids, *we compensate for the deviation of $u_{\ell-1}$ from $r u_\ell$ by including an additive term on the right-hand side.* This correction is the reason why we replaced the original equation (6.3) by the more general equation (6.4).

We also need the restriction of $u^{\ell,k,1}$ at the level $\ell-1$. This could be done with a different operator from the one used to evaluate the restriction of the residue. Let r and $\tilde{r}$ be restriction operators for the residue and for the approximations, respectively, and let p be a prolongation. The operator $\mathcal{L}_{\ell-1}$ corresponds to the discretization of (6.3) at level $\ell - 1$.

6.2 Nonlinear Multigrid Iteration NMGM$_\ell$ *(k-th cycle at level $\ell \geq 1$):*
Let $u^{\ell,k}$ be a given approximation in S_ℓ.

1. *Pre-smoothing.* Perform ν_1 smoothing steps:

$$u^{\ell,k,1} = \mathcal{S}^{\nu_1} u^{\ell,k}.$$

2. *Coarse-grid correction.* Set

$$\begin{aligned} d_\ell &= f_\ell - \mathcal{L}_\ell(u^{\ell,k,1}), \\ u^{\ell-1,0} &= \tilde{r} u^{\ell,k,1}, \\ f_{\ell-1} &= \mathcal{L}_{\ell-1}(u^{\ell-1,0}) + r d_\ell, \end{aligned} \tag{6.8}$$

and let $\hat{v}_{\ell-1}$ be the solution of

$$\mathcal{L}_{\ell-1}(v) = f_{\ell-1}. \tag{6.9}$$

If $\ell = 1$, find the solution, and set $v^{\ell-1} = \hat{v}_{\ell-1}$.
If $\ell > 1$, determine an approximation $v^{\ell-1}$ of $\hat{v}_{\ell-1}$ by carrying out μ steps of **NMGM**$_{\ell-1}$ with the starting value $u^{\ell-1,0}$.
Set

$$u^{\ell,k,2} = u^{\ell,k,1} + p(v^{\ell-1} - u^{\ell-1,0}). \tag{6.10}$$

3. *Post-smoothing.* Perform ν_2 smoothing steps using

$$u^{\ell,k,3} = \mathcal{S}^{\nu_2} u^{\ell,k,2},$$

and set $u^{\ell,k+1} = u^{\ell,k,3}$. □

The reader can verify that in the linear case we get Algorithm 1.7, independent of the choice of the restriction operator $\tilde{r}$.

Note that the following diagram is *not commutative*:

$$\begin{array}{ccc} S_\ell & \xrightarrow{\mathcal{L}_\ell} & S_\ell \\ \tilde{r}\big\downarrow & & \big\downarrow r \\ S_{\ell-1} & \xrightarrow{\mathcal{L}_{\ell-1}} & S_{\ell-1}. \end{array}$$

More specifically, $f_{\ell-1} \neq r f_\ell$ in general. In fact,

$$f_{\ell-1} = r f_\ell + [\mathcal{L}_{\ell-1}(\tilde{r} u^{\ell,k,1}) - r\mathcal{L}_\ell(u^{\ell,k,1})]. \tag{6.11}$$

The *shift* by the extra term in the square brackets ensures that the solution u_ℓ of the equation (6.4) is a fixed point for the iteration. Assuming that $u^{\ell,k,1} = u_\ell$, it follows that $d_\ell = 0$ and $f_{\ell-1} = \mathcal{L}_{\ell-1}(u^{\ell-1,0})$. Thus, $v = u^{\ell-1,0}$ is a solution of (6.9), and $u^{\ell,k,2} = u^{\ell,k,1}$.

Starting Values

As explained in §4, for linear problems we can start on the coarsest grid and work toward the finest one. This is also possible for many nonlinear problems, but not for all. In particular, it can happen that the nonlinear problem only has the right number of solutions when the discretization is sufficiently fine.

For this reason, we now assume that we have a starting value which belongs to the domain of attraction of the desired solution. However, the error may still be much larger than the discretization error, and in fact by several orders of magnitude.

This is the usual case in practice, and we suggest proceeding as in Algorithm NI_ℓ. However, we have first to compute an appropriate right-hand side for the problems on the coarse grids.

In the following we use the notation of Algorithm 6.2.

6.3 Algorithm NLNI$_\ell$ $(\mathcal{L}_\ell, f_\ell, u^{\ell,0})$ *for improving a starting value* $u^{\ell,0}$ *for the equation* $\mathcal{L}_\ell(u_\ell) = f_\ell$ *at level* $\ell \geq 0$, *(such that the error of the result* $\hat{u}^\ell$ *is of the order of the discretization error).*

If $\ell = 0$, compute the solution $\hat{u}^\ell$ of the equation $\mathcal{L}_0(v) = f^0$, and exit the procedure.
Let $\ell > 0$.
Set $u^{\ell-1,0} = \tilde{r}u^{\ell,0}$ and

$$f_{\ell-1} = rf_\ell + [\mathcal{L}_{\ell-1}(u^{\ell-1,0}) - r\mathcal{L}_\ell(u^{\ell,0})]. \tag{6.12}$$

Find an approximate solution $\hat{u}_{\ell-1}$ of the equation $\mathcal{L}_{\ell-1}(v) = f_{\ell-1}$ by applying **NLNI**$_{\ell-1}$ $(\mathcal{L}_{\ell-1}, f_{\ell-1}, u^{\ell-1,0})$.
Determine the prolongation $u^{\ell,1} = p\hat{u}^{\ell-1}$.
Using $u^{\ell,1}$ as a starting value, carry out one step of the iteration **NLMG**$_\ell$. Denote the result as $u^{\ell,2}$. Set

$$\hat{u}^\ell = u^{\ell,2}. \qquad \square$$

Note that equation (6.12) has the same structure as (6.11).

Since we cannot proceed without reasonable starting values, for complicated problems, nonlinear multigrid methods are usually combined with continuation methods (also called incremental methods).

Problems

6.4 Verify that for linear problems, Algorithm 6.2 is equivalent to Algorithm 1.7.

6.5 The nonlinear equation (6.2) characterizes a solution of the nonquadratic variational problem

$$\int_\Omega [\frac{1}{2}(\nabla v)^2 - F(v)]dx \longrightarrow \min_{v \in H_0^1}!$$

(assuming a solution exists). Find a suitable function F on $\mathbb{R}$ by formally calculating the Euler equation corresponding to the variational problem.