Chapter 6

Many-Level Cycles

Having obtained satisfactorily performing two-level cycling algorithms, one needs next to turn on the complete sequence of grids, using now the two-level techniques in recursion. The new algorithmic questions which arise are discussed below. Some of them could theoretically be investigated by three-level mode analysis, but this trouble is neither needed nor normally taken.

6.1 Multigrid cycles. Initial and terminal relaxation

For any grid h, finest or intermediate, a multigrid h-**cycle** can recursively be defined as follows: make ν_1 relaxation sweeps on grid h; transfer the residual problem to the next coarser grid $H(=2h)$ and solve it there approximately, using γ H-cycles (unless H is the coarsest grid); interpolate the grid-H solution and add it as a correction to the grid-h solution; finally, make ν_2 more sweeps on grid h. On the coarsest grid the problem is solved either directly or by ν_0 relaxation sweeps (cf. §6.3).

In two-level cycles only the sum $\nu = \nu_1 + \nu_2$ matters. When h is an intermediate grid the separate values of ν_1 and ν_2 do make some difference, although not a big one. In regular elliptic solvers experience shows that $\nu_2 = [\nu/2]$ is probably the best prescription (see for example [ST82, Tables 3.3]). In double-discretization schemes (§10.2) it is important to use $\nu_2 = 0$. In "accommodative" algorithms (see §4.1, §6.2), the values of ν_1 and ν_2 vary and they are determined internally.

Note also that the several passes of a complex relaxation sweep (such as ADZ) can be divided between the initial and the terminal stages of the cycle [ST82, §8.2].

6.2 Switching criteria. Types of cycles

The criteria when to switch from a fine grid h to the next coarser grid $H = 2h$ were examined in a previous stage (§4.1). These same criteria can be used recursively, i.e., not only when h is the finest grid. We need in addition some criteria for switching from any grid H back to the next finer grid h. Two kinds of switches are used: Fixed and accommodative.

Fixed algorithms switch from H back to h after a preassigned number γ of H-cycles. The h-cycle is recursively defined as type $\mathrm{C}(\nu_1,\nu_2)^\gamma$ if all these H-cycles are of this same type; γ is then called the *cycle index*. The cycle is defined as type $\mathrm{F}(\nu_1,\nu_2)$ if $\gamma = 2$ and the first H-cycle is itself an $\mathrm{F}(\nu_1,\nu_2)$ cycle, while the second H-cycle is a $\mathrm{C}(\nu_1,\nu_2)^1$ cycle. See flowcharts and operation counts in [Bra81a, §6.1]. The cycle $\mathrm{C}(\nu_1,\nu_2)^1$ is also called a V cycle and denoted $\mathrm{V}(\nu_1,\nu_2)$; see Fig. 1.1. The cycle $\mathrm{C}(\nu_1,\nu_2)^2$ is also called a W cycle and denoted $\mathrm{W}(\nu_1,\nu_2)$.

V cycles require considerably less work than W cycles (1/3 in two-dimensional problems with full coarsening). F cycles are somewhat less expensive than W cycles in one-dimensional problems with many levels, and in higher dimensions when semi coarsening is used; but otherwise perform practically the same as W cycles.

Fractional γ. For extra flexibility, a $\mathrm{C}(\nu_1,\nu_2)^\gamma$ cycle can even be defined for non-integer γ. At each level h of the cycle, the number of visits N_c^h from the next-finer level to h is required to be *on average* as close as possible to γN_c^H. The larger γ, the more accurately solved is the H-grid problem, relative to the accuracy of the h-cycle it serves.

When should $\gamma > 1$ *be used?* Except for simulating two-level algorithms (§5.1), a large cycle index should only be used if the *smoothest component's two-level convergence factor* χ *is large.* Normally $\chi \to 0$ as the finest level meshsize $h \to 0$, and the V cycle convergence will be dominated by high-frequency components, leading to a convergence factor close to the two-level factor $\overline{\chi}$. On the other hand, when χ is constant in h a V cycle will not attain $\overline{\chi}$, because the coarse grid equations themselves will only be crudely solved, and the error will accumulate through the levels. In this case, $\gamma > 1$ is mandatory for obtaining good asymptotic factors. Indeed, the asymptotic convergence factor χ_m of an m-level $\mathrm{C}(\nu_1,\nu_2)^\gamma$ cycle with fixed γ is

$$\chi_m = 1 - (1-\chi)\left(1 - \chi_{m-1}^\gamma\right), \quad \chi_1 = 0. \tag{6.1}$$

For $\gamma(1-\chi) < 1$, $\chi_m \approx 1 - (\gamma(1-\chi))^{m-1}$ tends to 1 for $m \to \infty$, but has a finite limit $\overline{\chi}$ for $\gamma(1-\chi) \geq 1$ that satisfies

$$\frac{1-\overline{\chi}}{1-\overline{\chi}^\gamma} = 1 - \chi. \tag{6.2}$$

For a d-dimensional problem and full coarsening, the multilevel cycle work is linearly proportional to the finest grid size for $\gamma < 2^d$, therefore the smoothest component must be approximated to at least $\chi > 1 - 2^{-d}$. The

optimal γ can be determined from (6.2) by maximizing the accuracy per unit work (cf. §9.5).

Examples of large χ are (a) severe singularities (cf. §5.7); (b) intergrid transfers whose orders are too low (cf. §4.3). If for instance first-order transfers and Galerkin coarsening (§4.5) are employed for the d-dimensional Poisson equation, then $\chi = .5$. The optimal parameters are then $\gamma = 2.7, \overline{\chi} = .67$ for $d = 2$, $\gamma = 3.5, \overline{\chi} = .57$ for $d = 3$, and $\gamma \approx d, \overline{\chi} \approx .5$ for $d \gg 1$. (c) non-elliptic and singular perturbation problems. The artificial viscosity on grid kh is k times larger than on grid h, hence visiting grid kh only once per cycle would yield an asymptotic convergence factor no better than $1 - 1/k$ [Bra81a, §5.1]. Since $k = O(h^{-1})$ on the coarsest grid, the V cycle asymptotic factor will be $1 - O(h)$, which is very poor indeed. h-independent asymptotic convergence may be restored by using $\gamma > 1$; V cycles may however perform at a satisfactory level within FMG algorithms (see §7.4 and the numerical experiments in [Bra81a, §7.1]).

Accommodative algorithms switch from grid H back to grid h when a certain norm of the residuals on grid H drops below some factor η times the latest value of the corresponding norm on grid h. The parameter η is not sensitive one; a good general prescription seems to be $\eta = 1.1\overline{\lambda}$. If $\overline{\lambda}$ is not approximately known, use $\eta = 2^{-d}$, a value related to accuracy-to-work exchange-rate considerations (cf. §9.6).

Generally, accommodative algorithms may be troublesome at program development stages, since they may cause more complex interactions between the internal checks and the real questions one likes to examine. Their flexibility may prevent us from seeing some of the troubles, and they are not suitable for precise comparisons. In the production stages, accommodative algorithms have the disadvantage that they require the extra work of calculating the residual norms. On the other hand, accommodative algorithms can be more robust. Also, in complicated problems (which is where this robustness is needed), the residual norm calculation is inexpensive relatively to other calculations, assuming *dynamic* residuals (those calculated anyway in the relaxation process) are used.

6.3 Coarsest grids. Inhomogeneous and indefinite operators

When the multigrid h-cycle performs considerably poorer than expected, it is first important to distinguish between fine-grid and coarse-grid troubles. This distinction is easy to make, by simulating two-level algorithms (taking large γ or small η) and examining whether this improves the convergence factor (per h-cycle), and how much this improvement depends on the size of h. Also examine whether reasonable convergence is obtained on your coarsest grid. If not, or if the trouble is confined to coarse h, the following remarks may be relevant.

Inhomogeneous operators are the main source for the special troubles appearing only on sufficiently coarse grids. On such grids, lower order terms of the operator start to affect, or even dominate, the smoothing and convergence factors. If we have neglected them in designing the fine-grid relaxation, we should now take them into account. Generally, on every grid h, the important terms for designing relaxation are the h-principal terms (see §§2.1, 3.1, 3.4).

Another type of coarse level trouble is exhibited for example by the equation $-\Delta u + \sigma u = f$ with purely Neumann boundary conditions. If σ is positive but very small, the smoothing factor of a GS relaxation is essentially the same as for Poisson equation, but the convergence factor is roughly $4/(4 + h^2\sigma)$, which may be very slow even on the coarsest grid. Hence the coarsest-grid equations should be solved either directly (e.g., by elimination, which is inexpensive since the coarsest grid can contain just few points), or by relaxation, where after each sweep a suitable constant is subtracted from the approximate solution [ABDP81, §4]. If $\sigma = 0$ everywhere except in some small subdomain, that constant subtraction should be employed on all grids which are not fine enough to resolve that small subdomain.

Indefinite case. If σ is negative, the situation is much worse, whatever the boundary conditions: For the coarse grid to approximate the slowly converging fine-grid component, its meshsize must be fine enough: For large $|\sigma|$, the coarsest meshsize must satisfy $H \leq O(R^{-1/p}(-\sigma)^{-.5(p+1)/p})$, where R is the radius of the domain and p is the approximation order. In many cases this H is smaller than the *finest* affordable meshsize. To use coarser H, different intergrid transfers should be employed (cf. §4.2.2 and [BL97]).

In designing fine-level relaxation schemes for complex systems of equations, e.g., in fluid dynamics, we can take only subprincipal terms into account (§2.1, 3.1). On very coarse grids, however, this is no longer fully justified, and if the same relaxation schemes are still used, there the smoothing factors may deteriorate. We may then have to use either more sweeps (by increasing ν and/or γ, or by using accommodative algorithms), or more sophisticated relaxation. In solving Navier-Stokes equations, for example, improved results were obtained by using the high-speed DGS scheme (see §19.3) on all finer grids, while employing the most robust BGS (see §3.4) on the two coarsest grids.

Even for homogeneous operators, convergence of h-cycles can sometimes be slower on very coarse grids, because the convergence factor $\overline{\lambda}$ cannot be smaller than $O(h^{\tilde{m}})$; see (C) in §4.3. In such cases one can make more h-cycles, by increasing γ or switching accommodatively, which is inexpensive since h is coarse.

Sometimes troubles seen on coarse grids are only indications of bad procedures at some special, restricted regions, such as boundaries (see §5.1), or they may signal the need to operate some global conditions, which are not enforced on finer grids (see §5.6).

Of special concern is the **coarsest grid** itself. Relaxation there should be *converging*, not just smoothing as on other grids. various conditions not enforced on finer grids must be enforced on the coarsest one, calling for special procedures. If nothing better is known, one can always use either a direct solver or a slow but safe iterative process such as Kaczmarz relaxation (cf. §1.1); on the coarsest grid they are affordable. Finally note that the coarsest grid cannot efficiently contribute to convergence if all its points happen to lie too close (much closer than a meshsize) to boundaries with Dirichlet boundary conditions.