A MULTIGRID PRIMER

STEPHEN F. MCCORMICK* AND RASMUS TAMSTORF†

Abstract. While several tutorials and texts exists that introduce basic and advanced aspects of multigrid methodology, the aim here is a relatively short exposition that is simplified by focusing on the fundamental concepts in a *variational* setting: a quadratic *energy* functional minimization principle formulation of a symmetric positive definite matrix equation. This focus enables a simplification of the development and supporting theory, and it clarifies the principles that lead to effective algorithms. The development begins with the iterative form of multigrid solvers, exemplified by the so-called V-cycle. It then introduces the full multigrid method as a *direct solver* in the sense that it is aimed directly at the partial differential equation from which the matrix equation presumably arises. The aim of full multigrid is to achieve discretization-level accuracy at a minimal cost comparable to that of a few matrix multiplies on the finest level. This primer presents these basic multigrid methods along with some theory and practical principles and heuristics that can guide their development. The focus here is on understanding the general principles that lead to multigrid methods that work rather than the details of specific algorithms.

1. Introduction. The purpose here is to provide a brief introduction to multigrid (MG) methods together with some basic theory and practical principles and heuristics that can be used to guide their design. Some of this presentation has been taken from the supplemental material available with [11], while most of the following concepts can be studied further in [6], [8], and [12]. Our focus in what follows is on symmetric positive definite (SPD) matrix equations that arise from the discretization of elliptic partial differential equations (PDEs). Although multilevel methods have much broader applicability, this case allows for a much simpler development and clarification of the basic principles underlying multigrid methodology.

It is important to keep in mind that this primer focuses on general principles as opposed to detailed prescriptions for specific problems. The point is that a good understanding of these principles can guide to the development of effective multigrid methods for many diverse applications.

Multigrid methods have two essential components: *smoothing* and *coarsening*. Typically, the smoother is responsible for removing oscillatory errors in the solution, while the coarsening strategy effectively provides a way to deal with smooth errors. For multigrid to work optimally, these two components must be carefully designed to complement each other. An underlying theme of this primer is the focus on design conditions that guarantee such complementarity.

To better understand this interplay, we start in section 2 with the assumptions on the hierarchy of discrete matrix equations that are made available to the multigrid algorithms. We continue in section 3 by characterizing the smoothers that we have in mind together with their basic properties. In section 4, we begin by introducing a two-grid cycle. After a clarification of the notion of residuals in section 5, the two-grid development then leads into the construction of so-called V-cycles and W-cycles in section 6. These multilevel methods aim to provide an optimal *iterative* solver in the sense of achieving uniformly bounded convergence factors per cycle at a theoretical cost proportional to the number of degrees of freedom on the finest level. In section 7, we introduce full multigrid (FMG), a *direct* method in the sense that its target is the PDE: it uses a special fixed cycle that attempts to achieve discretization-level accuracy at the truly optimal total cost equivalent to that of just a few finest-level relaxation sweeps. In section 8, we review the analysis of two coarse-level approximation properties that confirm theoretical optimality

*University of Colorado, Boulder
†Independent Researcher

of the multigrid solvers that have been developed. Based on this analysis, we discuss how to design effective coarse grids and interpolation operators in subsection 9.1 and how the discretization of the underlying problem can affect multigrid performance in section 9.2. We end the primer with a short summary in subsection 10.

2. Preliminaries. We use bold font for matrices and vectors, typewriter font for algorithms, and roman font for functions and acronyms. For example, $\boldsymbol{G}$ signifies the error propagation matrix for a generic stationary linear relaxation method, $\mathtt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, h\right)$ denotes the relaxation algorithm, and G is shorthand for relaxation. For consistency with the PDE equation that uses L for the operator, u for the unknown solution, and f for the source term, our matrix equations appear as expressions like $\boldsymbol{Lu} = \boldsymbol{f}$ and $\boldsymbol{Lv} = \boldsymbol{g}$ with mesh size superscripts to denote the grid level. Note that this use is in contrast to the more traditional expressions like $\boldsymbol{Ax} = \boldsymbol{b}$.

We assume up through section 8 that a hierarchy of SPD matrices denoted by $\boldsymbol{L}^h$ and intergrid transfer operators called *interpolation* or *prolongation* $\boldsymbol{P}_{2h}^h$ and *restriction* $\boldsymbol{R}_h^{2h}$ have been determined on a sequence of coarse-to-fine grid levels that reflect increasingly accurate discretizations of a PDE. The assumption about these discretizations is that they are based on the simple case of a finest-level logically rectangular grid in two dimension with uniform mesh size h in both coordinate directions. A further simplification is that the coarse levels are constructed by eliminating every other grid line in both coordinate directions. Accordingly, we use $2h$ to denote the mesh size of a coarsening of a given grid h, where H represents the coarsest grid in the hierarchy. We do not explicitly specify here how these discretizations are to be constructed so that the discussion applies to virtually any procedure. However, it is probably best to think in terms of the Rayleigh-Ritz variational approach based on lowest-order finite elements as discussed further in section 8.2. In what follows, we assume that these constructs on all levels of the hierarchy are internally available to the algorithms that we introduce.

It is important to keep in mind that these assumptions are only to simplify the discussion and that multigrid and, more generally, multilevel methods (that might not even be based on grids) have much broader applicability.

3. Smoothing. Assume that the matrix $\boldsymbol{L}^h \in \mathbb{R}^{n\times n}$ is SPD, where h is the mesh width of the finest grid. The target problem is to find $\boldsymbol{u}^h \in \mathbb{R}^n$ such that

$$\boldsymbol{L}^h\boldsymbol{u}^h = \boldsymbol{f}^h, \tag{1}$$

where $\boldsymbol{f}^h \in \mathbb{R}^n$ is a given source term. The coarse-grid correction form of multigrid described later involves an equation on grid $2h$ where the source term is different than $\boldsymbol{f}^{2h}$. To accommodate this change, we generalize the notation of (1) by writing

$$\boldsymbol{L}^h\boldsymbol{v}^h = \boldsymbol{g}^h. \tag{2}$$

While (1) is our ultimate target, we focus first on (2) in our discussions on relaxation and basic multigrid solvers. This is of course just a change in notation, but it is useful because it allows us to emphasize the difference between treating (1) directly and treating it indirectly by way of (2). This indirect approach involves setting up (2) so that its solution is a correction to the current approximate solution of (1). In any case, we return to (1) when we introduce FMG that treats it directly.

The matrix equation in (2) can be solved using any of a number of simple iterative methods like Jacobi or Gauss-Seidel relaxation (cf. [7, Section 10.1]). However, in all

that follows, we restrict our attention to Richardson's iteration[1] given by

$$\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \frac{1}{\|\boldsymbol{L}^h\|}\left(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h\right), \tag{3}$$

where $\|\cdot\|$ denotes the matrix norm induced by the Euclidean vector norm $\|\cdot\|$. (By the notation in (3), we mean that the current approximation, $\boldsymbol{v}^h$, to the solution of (2) is replaced by the expression involving the old $\boldsymbol{v}^h$ to the right of the arrow.) As with many other methods, Richardson relaxation corrects the current approximation by a term involving the *residual* $\boldsymbol{r}^h \equiv \boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h$. Algorithm 1 illustrates μ sweeps of Richardson iteration applied to (2).

Algorithm 1 $\mathtt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, h\right)$; relaxation

Given: $\boldsymbol{L}^h$

$i \leftarrow 0$
while $i < \mu$ **do**
  $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \frac{1}{\|\boldsymbol{L}^h\|}\left(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h\right)$ ▷ Richardson iteration
  $i \leftarrow i + 1$
end while
return $\boldsymbol{v}^h$

Unfortunately, all simple relaxation methods applied to discretizations of partial differential equations, including Richardson's iteration, typically stall well before they reach an acceptable approximation to the solution. The cause of stalling comes from the residual's inability to see *smooth* error, by which we mean that the matrix applied to such an error yields a scaled residual $\frac{1}{\|\boldsymbol{L}^h\|}(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h)$ that is small compared to the *algebraic* error $\boldsymbol{e}^h \equiv \boldsymbol{v}^h - (\boldsymbol{L}^h)^{-1}\boldsymbol{g}^h$. For many discretized elliptic equations, this algebraic sense of smoothness of the error usually corresponds to the geometric sense, where errors vary slowly across the grid. We adopt the algebraic sense here instead because it applies to more general matrix equations that may not even have a geometric basis, and it sheds important light on how to treat these errors. In any event, correcting such smooth errors by way of the residual, as simple relaxation methods do, would therefore accomplish very little. Worse yet, while these methods may work well for a couple of iterations when the initial error has oscillatory components, this fast elimination of oscillatory error exposes the remaining smooth error that works to stall all subsequent iterations. While this limitation is a common smoothing property of most conventional iterative methods applied to discretizations of partial differential equations, we shall see that it provides the motivation and foundation for coarse-grid correction.

To clarify the source of this difficulty, we begin by studying the algebraic error $\boldsymbol{e}^h$ in more detail and by introducing ways to measure its smoothness. To this end, consider first the following simple relationship between the error and the residual in terms of the target equation in (2):

$$\boldsymbol{L}^h\boldsymbol{e}^h = \boldsymbol{L}^h\left(\boldsymbol{v}^h - (\boldsymbol{L}^h)^{-1}\boldsymbol{g}^h\right) = \boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h.$$

If we use this relationship in (3) and subtract $(\boldsymbol{L}^h)^{-1}\boldsymbol{g}^h$ from both sides of the result, we

[1] The reader should be warned that this primer rests on the simplicity of Richardson and other so-called Krylov methods (whose error propagation matrices as defined in (4) below are polynomials in $\boldsymbol{L}^h$). For other smoothers, the analysis and attendant algebra can be much more involved, although most of the basic principles are essentially the same. It is also important to point out that dividing the residual by $\|\boldsymbol{L}^h\|$ ensures that the iteration converges, but it is not necessarily the best choice for relaxation. See, for example, [6] for somewhat better scale factors.

end up with the error propagation expression

$$\boldsymbol{e}^h \leftarrow \boldsymbol{G}^h \boldsymbol{e}^h, \tag{4}$$

where $\boldsymbol{G}^h = \boldsymbol{I}^h - \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h$, with $\boldsymbol{I}^h \in \mathbb{R}^{n\times n}$ denoting the identity on grid h. Matrix $\boldsymbol{G}^h$ is called the error propagation matrix because it governs how the iteration transforms the error. To show that Richardson's iteration is convergent in the Euclidean norm $\|\cdot\|$, let $\langle\cdot,\cdot\rangle$ denote the Euclidean inner product and assume that $\boldsymbol{e}^h$ is any nonzero error component. Then

$$0 < \langle \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle \leq \frac{\|\boldsymbol{L}^h\|\langle \boldsymbol{e}^h, \boldsymbol{e}^h\rangle}{\|\boldsymbol{L}^h\|} = \langle \boldsymbol{e}^h, \boldsymbol{e}^h\rangle, \tag{5}$$

which can be rewritten as

$$0 \leq \langle (\boldsymbol{I}^h - \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h)\boldsymbol{e}^h, \boldsymbol{e}^h\rangle < \langle \boldsymbol{I}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle = \langle \boldsymbol{e}^h, \boldsymbol{e}^h\rangle.$$

This bound means that $\|\boldsymbol{G}^h\| < 1$, which proves that Richardson's iteration is convergent in the Euclidean norm because it implies that the new error in (4) must be smaller than the old error.

For a given error, $\boldsymbol{e}^h$, the following two quantities are particularly useful for measuring its algebraic smoothness:

$$\mathcal{M}_w(\boldsymbol{e}^h) = \frac{\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle}{||\boldsymbol{L}^h||\langle \boldsymbol{e}^h, \boldsymbol{e}^h\rangle} \text{ and } \mathcal{M}_s(\boldsymbol{e}^h) = \frac{\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{L}^h\boldsymbol{e}^h\rangle}{||\boldsymbol{L}^h||\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle}.$$

The *weak* measure, $\mathcal{M}_w(\boldsymbol{e}^h)$, is the relative Rayleigh quotient of $\boldsymbol{e}^h$, and the *strong* measure, $\mathcal{M}_s(\boldsymbol{e}^h)$, is the relative Rayleigh quotient of $\left(\boldsymbol{L}^h\right)^{\frac{1}{2}}\boldsymbol{e}^h$, both relative to $\|\boldsymbol{L}^h\|$. In either case, we call errors with a large measure *algebraically oscillatory* (or simply oscillatory), while errors with small measure are called *algebraically smooth* (smooth).

When $\boldsymbol{L}^h$ is symmetric, as we have assumed here, any error can be written as a linear combination of the eigenvectors, $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$, of $\boldsymbol{L}^h$: $\boldsymbol{e}^h = \sum_{i=1}^n \alpha_i \boldsymbol{w}_i$. Smooth vectors are rich in eigenvectors associated with the low part of the spectrum, while oscillatory vectors are rich in the high part. This is the basis for using the terms *smooth* and *oscillatory*. For example, if $\boldsymbol{e}^h$ is an eigenvector of $\boldsymbol{L}$ associated with eigenvalue λ, then $\mathcal{M}_w(\boldsymbol{e}^h) = \mathcal{M}_s(\boldsymbol{e}^h) = \frac{\lambda}{\|\boldsymbol{L}^h\|}$. By virtue of our SPD assumption, $\|\boldsymbol{L}^h\| = \max_{1\leq i\leq n}\lambda_i$, where the λ_i are the eigenvalues of $\boldsymbol{L}^h$. Thus, $\mathcal{M}_w(\boldsymbol{e}^h)$ assesses how large λ is relative to its largest possible value. For general linear combinations of eigenvectors, this measure assesses where the error lives in terms of the spectrum of $\boldsymbol{L}^h$. For many applications, this value can be very small at the lower end of the spectrum (typically $O(h^2)$ for discrete second-order elliptic operators).

These weak and strong measures are important because they identify errors that cannot be adequately reduced by relaxation and therefore must be reduced by coarse-grid correction. To begin to see this, note from (4) that relaxation reduces the nonzero current error, $\boldsymbol{e}^h$, in the Euclidean norm by the factor $\|\boldsymbol{G}^h\boldsymbol{e}^h\|/\|\boldsymbol{e}^h\|$. Using (5), we have that

$$\frac{\|\boldsymbol{G}^h\boldsymbol{e}^h\|^2}{\|\boldsymbol{e}^h\|^2} = 1 - 2\frac{\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle}{\|\boldsymbol{L}^h\|\cdot\|\boldsymbol{e}^h\|^2} + \frac{\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{L}^h\boldsymbol{e}^h\rangle}{\|\boldsymbol{L}^h\|^2\|\boldsymbol{e}^h\|^2} = 1 - (1+\xi)\mathcal{M}_w(\boldsymbol{e}^h), \tag{6}$$

where $\xi \in [0, 1)$. The last equality follows because

$$0 < \frac{\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{L}^h\boldsymbol{e}^h\rangle}{\|\boldsymbol{L}^h\|^2\|\boldsymbol{e}^h\|^2} \leq \frac{\|\boldsymbol{L}^h\|\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle}{\|\boldsymbol{L}^h\|^2\cdot\|\boldsymbol{e}^h\|^2} = \frac{\langle \boldsymbol{L}^h\boldsymbol{e}^h, \boldsymbol{e}^h\rangle}{\|\boldsymbol{L}^h\|\cdot\|\boldsymbol{e}^h\|^2} = \mathcal{M}_w(\boldsymbol{e}^h).$$

Estimate (6) confirms that relaxation slows *in the Euclidean norm* if and only if the *weak measure* is small. Let $\langle \cdot, \cdot \rangle_{\boldsymbol{L}^h} = \langle \boldsymbol{L}^h \cdot, \cdot \rangle$ and $\| \cdot \|_{\boldsymbol{L}^h} \equiv \sqrt{\langle \cdot, \cdot \rangle_{\boldsymbol{L}^h}}$ denote the *energy* inner product and its induced *energy* norm, respectively. Then an argument similar to the above shows that

$$\frac{\|\boldsymbol{G}^h \boldsymbol{e}^h\|^2_{\boldsymbol{L}^h}}{\|\boldsymbol{e}^h\|^2_{\boldsymbol{L}^h}} = 1 - (1+\chi)\mathcal{M}_s(\boldsymbol{e}^h),$$

where $\chi \in [0, 1)$. Thus, relaxation slows *in the energy norm* if and only if the *strong measure* is small.

This correspondence between weak vs. strong measures and Euclidean vs. energy norms carries over to the analysis of coarse-grid correction. As shown in what follows (see Subsection 8.1), if the coarse grid adequately approximates errors for which $\mathcal{M}_s(\boldsymbol{e}^h)$ is small, then the so-called multigrid *V-cycles* converge well in the energy norm. For the so-called *two-grid* or *W-cycles* (described in what follows) in the Euclidean norm, it is enough to have the coarse grid adequately approximate errors for which $\mathcal{M}_w(\boldsymbol{e}^h)$ is small. This latter requirement is weaker in part because small $\mathcal{M}_s(\boldsymbol{e}^h)$ implies small $\mathcal{M}_w(\boldsymbol{e}^h)$, which follows because

$$\langle \boldsymbol{L}^h \boldsymbol{e}^h, \boldsymbol{e}^h \rangle^2 \leq \|\boldsymbol{L}^h \boldsymbol{e}^h\|^2 \|\boldsymbol{e}^h\|^2 = \langle \boldsymbol{L}^h \boldsymbol{e}^h, \boldsymbol{L}^h \boldsymbol{e}^h \rangle \langle \boldsymbol{e}^h, \boldsymbol{e}^h \rangle,$$

but the converse is not true. More importantly, what we meant by adequate approximation for the strong measure, $\mathcal{M}_s(\boldsymbol{e}^h)$, involves the energy norm, which is a stronger requirement than that for the Euclidean norm used with the weak measure.

4. Coarse-grid correction. The coarse-grid correction phase of multigrid enters as a way to exploit the smoothing property of relaxation. The basic idea is that smooth error varies so slowly from one neighboring grid point to the next that it can be adequately approximated by fewer grid points. The error components thus computed on these coarse levels are then interpolated back to the fine grid to correct the approximation there.

To see this more clearly, consider the grid h *energy functional* given by

$$F^h(\boldsymbol{v}^h) \equiv \langle \boldsymbol{L}^h \boldsymbol{v}^h, \boldsymbol{v}^h \rangle - 2\langle \boldsymbol{v}^h, \boldsymbol{g}^h \rangle.$$

To show that minimizing $F^h(\boldsymbol{v}^h)$ is equivalent to solving (2), let $C \equiv \langle \boldsymbol{g}^h, (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \rangle$ and note that

$$\begin{aligned}
F^h(\boldsymbol{v}^h) &= \langle \boldsymbol{L}^h \boldsymbol{v}^h, \boldsymbol{v}^h \rangle - 2\langle \boldsymbol{v}^h, \boldsymbol{g}^h \rangle + \langle \boldsymbol{g}^h, (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \rangle - \langle \boldsymbol{g}^h, (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \rangle \\
&= \langle \boldsymbol{L}^h \boldsymbol{v}^h, \boldsymbol{v}^h \rangle - 2\langle \boldsymbol{L}^h \boldsymbol{v}^h, (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \rangle + \langle \boldsymbol{L}^h (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h, (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \rangle - C \\
&= \langle \boldsymbol{L}^h \left(\boldsymbol{v}^h - (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \right), \boldsymbol{v}^h - (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h \rangle - C \\
&= \|\boldsymbol{e}^h\|^2_{\boldsymbol{L}^h} - C.
\end{aligned}$$

Thus, $F^h(\boldsymbol{v}^h)$ differs from the square of the energy norm of the error by a constant (because $\boldsymbol{g}^h$ is fixed), which means that $\boldsymbol{v}^h$ solves (2) if and only if it is the minimizer of $F^h(\boldsymbol{v}^h)$. (That minimum is unique because the above equation shows that the minimum value of $-C$ is attained only when $\boldsymbol{e}^h \equiv \boldsymbol{v}^h - (\boldsymbol{L}^h)^{-1} \boldsymbol{g}^h = \boldsymbol{0}^h$.)

Assuming as we have that grid $2h$ and an interpolation operator $\boldsymbol{P}^h_{2h}$ from grid $2h$ to grid h are available, we can exploit this minimization principle further by noting that the best coarse-grid correction to a fixed approximation, $\boldsymbol{v}^h$, in the sense of minimizing $F^h(\boldsymbol{v}^h - \boldsymbol{P}^h_{2h} \boldsymbol{v}^{2h})$ over all possible $\boldsymbol{v}^{2h}$, is expressed by

$$\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \boldsymbol{P}^h_{2h} \left(\left(\boldsymbol{P}^h_{2h} \right)^t \boldsymbol{L}^h \boldsymbol{P}^h_{2h} \right)^{-1} \left(\boldsymbol{P}^h_{2h} \right)^t (\boldsymbol{L}^h \boldsymbol{v}^h - \boldsymbol{g}^h). \tag{7}$$

(We use superscript t to denote matrix transpose.) To verify this form of the correction, note that

$$\langle \boldsymbol{L}^h \boldsymbol{P}^h_{2h} \boldsymbol{v}^{2h}, \boldsymbol{v}^h \rangle = \langle \boldsymbol{v}^{2h}, \left(\boldsymbol{P}^h_{2h} \right)^t \boldsymbol{L}^h \boldsymbol{v}^h \rangle,$$

which shows that

$$
\begin{aligned}
&F^h(\boldsymbol{v}^h - \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}) \\
&= \langle \boldsymbol{L}^h \left(\boldsymbol{v}^h - \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h} \right), \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h} \rangle - 2\langle \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}), \boldsymbol{g}^h \rangle \\
&= \langle \boldsymbol{L}^h \boldsymbol{v}^h, \boldsymbol{v}^h \rangle - 2\langle \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{v}^h \rangle + \langle \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h} \rangle - 2\langle \boldsymbol{v}^h, \boldsymbol{g}^h \rangle + 2\langle \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{g}^h \rangle \\
&= \langle \boldsymbol{L}^h \boldsymbol{v}^h, \boldsymbol{v}^h \rangle - 2\langle \boldsymbol{v}^h, \boldsymbol{g}^h \rangle + \langle \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h} \rangle - 2\langle \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{v}^h \rangle + 2\langle \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{g}^h \rangle \\
&= \langle \boldsymbol{L}^h \boldsymbol{v}^h, \boldsymbol{v}^h \rangle - 2\langle \boldsymbol{v}^h, \boldsymbol{g}^h \rangle + \langle \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h} \rangle - 2\langle \boldsymbol{v}^{2h}, \left(\boldsymbol{P}_{2h}^h \right)^t \left(\boldsymbol{L}^h \boldsymbol{v}^h - \boldsymbol{g}^h \right) \rangle \\
&= F^h(\boldsymbol{v}^h) + F^{2h}(\boldsymbol{v}^{2h}),
\end{aligned}
$$

where F^{2h} is the following grid $2h$ version of F^h:

$$
F^{2h}(\boldsymbol{v}^{2h}) = \langle \left(\boldsymbol{P}_{2h}^h \right)^t \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h}, \boldsymbol{v}^{2h} \rangle - 2\langle \boldsymbol{v}^{2h}, \left(\boldsymbol{P}_{2h}^h \right)^t \left(\boldsymbol{L}^h \boldsymbol{v}^h - \boldsymbol{g}^h \right) \rangle.
$$

This result implies that $\boldsymbol{v}^{2h}$ minimizes $F^h(\boldsymbol{v}^h - \boldsymbol{P}_{2h}^h \boldsymbol{v}^{2h})$ if and only if minimizes $F^{2h}(\boldsymbol{v}^{2h})$. This grid $2h$ energy functional corresponds to the following grid $2h$ version of the matrix equation in (2):

$$
\boldsymbol{L}^{2h} \boldsymbol{v}^{2h} = \boldsymbol{g}^{2h}, \tag{8}
$$

where

$$
\boldsymbol{L}^{2h} = \left(\boldsymbol{P}_{2h}^h \right)^t \boldsymbol{L}^h \boldsymbol{P}_{2h}^h \text{ and } \boldsymbol{g}^{2h} = \left(\boldsymbol{P}_{2h}^h \right)^t \left(\boldsymbol{L}^h \boldsymbol{v}^h - \boldsymbol{g}^h \right).
$$

Thus, the grid $2h$ correction is given by (7) as asserted.

This is the form of the coarse-grid correction that we use here and it gives rise to the following so-called *variational conditions*:

$$
\boldsymbol{R}_h^{2h} \equiv \alpha^h \left(\boldsymbol{P}_{2h}^h \right)^t \text{ and } \boldsymbol{L}^{2h} \equiv \boldsymbol{R}_h^{2h} \boldsymbol{L}^h \boldsymbol{P}_{2h}^h,
$$

where $\boldsymbol{R}_h^{2h}$ is the restriction (or transfer) operator from the fine to the coarse grid and $\boldsymbol{L}^{2h}$ is the coarse-grid matrix. The second definition is called the *Galerkin condition*, while the first allows for any desired scaling $\alpha^h > 0$. (The reasoning here is that the correction in (7) is independent of scale, which allows these transfer operators to be chosen so that they approximately preserve constants, for example.) The basic issue in designing such a *variational* multigrid method is therefore to choose $\boldsymbol{P}_{2h}^h$, which amounts to determining a coarse set of grid points (or variables) and a method for interpolating functions defined on those points to functions defined on the fine-grid points. This energy-minimization formulation is convenient in the sense that, once this is done, the coarse-grid matrix and the restriction operator are then determined automatically from the variational conditions.

With these constructions in hand, a two-grid version of multigrid, starting with initial guess $\boldsymbol{v}^h$ and using μ relaxation sweeps, proceeds as follows:

- Apply coarse-grid correction: $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h (\boldsymbol{L}^{2h})^{-1} \boldsymbol{R}_h^{2h} \left(\boldsymbol{L}^h \boldsymbol{v}^h - \boldsymbol{g}^h \right)$.
- Relax on $\boldsymbol{v}^h$: apply $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \frac{1}{\|\boldsymbol{L}^h\|} \left(\boldsymbol{L}^h \boldsymbol{v}^h - \boldsymbol{g}^h \right)$ μ times.

Note that the error propagation matrix for this scheme is given by

$$
\boldsymbol{TG}^h \equiv \left(I - \frac{1}{\|\boldsymbol{L}^h\|} \boldsymbol{L}^h \right) \boldsymbol{T}^h, \text{ where } \boldsymbol{T}^h \equiv I - \boldsymbol{S}^h \text{ and } \boldsymbol{S}^h \equiv \boldsymbol{P}_{2h}^h (\boldsymbol{L}^{2h})^{-1} \boldsymbol{R}_h^{2h} \boldsymbol{L}^h. \tag{9}
$$

This version begins with coarse-grid correction because it simplifies the theory, but multigrid can be constructed with these two phases reversed. The reverse form may be more intuitive because one can then think of relaxation as smoothing the error so that coarse-grid correction can then work effectively, but both forms actually work equally well in that they have the same convergence bounds. Algorithm 2 illustrates the case of using both pre- and post-relaxation with μ and ν sweeps, respectively.

Algorithm 2 $\texttt{TG}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right)$; Two-grid cycle

Given: $\boldsymbol{L}^h$, $\boldsymbol{L}^{2h}$, $\boldsymbol{R}_h^{2h}$, $\boldsymbol{P}_{2h}^h$

1: $\boldsymbol{v}^h \leftarrow \texttt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, h\right)$ ▷ Relax μ times.
2: **if** $h < H$ **then** ▷ Keep coarsening
3: $\quad \boldsymbol{g}^{2h} = \boldsymbol{R}_h^{2h}\left(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h\right)$ ▷ Transfer grid h residual to grid 2h.
4: $\quad \boldsymbol{v}^{2h} \leftarrow \left(\boldsymbol{L}^{2h}\right)^{-1}\boldsymbol{g}^{2h}$ ▷ Solve the grid $2h$ equation exactly.
5: $\quad \boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}$ ▷ Correct fine-grid iterate.
6: $\quad \boldsymbol{v}^h \leftarrow \texttt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \nu, h\right)$ ▷ Relax ν times.
7: **end if**
8: **return** $\boldsymbol{v}^h$

5. A word about residuals. Our next task is to extend this two-grid method to multiple levels. To do so, the key is to recognize that the grid $2h$ equation (8) is of the same form as the grid h equation (2). This suggests that all we have to do is apply the two-grid solver to grid $2h$ to get started with three levels that includes grid $4h$. This approach provides a natural way to recursively define multilevel solvers, but it also introduces potential confusion. Grid h relaxation applies directly to (2), but (8) is indirectly applied to finding a correction that solves an averaged residual equation for (2). This recursion means that grid $2h$ ends up passing an averaged residual for its averaged residual equation to grid $4h$. We are talking here about averaged residuals of averaged residual equations and so on down the grid hierarchy. None but the finest grid involves the original source term $\boldsymbol{g}^h$ directly. This is why we have introduced (2) here to allow the source term to change within the multigrid cycle, while the notation in (1) is fixed. It is perhaps advisable for the reader to begin by thinking carefully about how multigrid works with just two grids, hopefully to better understand its recursive extension to many levels in what follows.

6. Multilevel solvers. Two-grid versions usually only achieve true multigrid efficiency when grid h is very coarse and therefore relatively easy to solve by relaxation alone. However, in this case, grid h is probably fairly easy also to solve by relaxation alone. The point here is that multiple levels are usually needed for multigrid to realize its full potential. The key to the extension to a potentially fast multilevel version is to notice that coarse-grid correction involves the solution of $\boldsymbol{L}^{2h}\boldsymbol{v}^{2h} = \boldsymbol{R}_h^{2h}\left(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h\right)$ (followed by correction to the fine grid via $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}$). Given a hierarchy of coarse grids from h to $2h$ to $4h$ and on down to a very coarse grid H, the idea is to simply replace this exact solution step by a scheme that first improves the grid $2h$ initial approximation, $\boldsymbol{v}^{2h} = \boldsymbol{0}^{2h}$, by μ pre-relaxation sweeps, a subsequent correction from grid $4h$ (obtained by yet coarser levels), and ν post-relaxation sweeps on grid $2h$. Continuing recursively, this process proceeds in this way to the coarsest grid H where the equation there is subjected to $\mu+\nu$ relaxation sweeps. (In some cases, relaxation on the very coarsest grid is replaced by a more aggressive solver, such as Gaussian elimination. This is especially important when grid H cannot be easily chosen coarse enough for relaxation to converge quickly, but still coarse enough that such solvers are feasible and relatively inexpensive.) This process is called a $V(\mu,\nu)$-cycle because it starts from the fine grid and passes source terms down through the coarser grids to the coarsest, doing μ relaxations down along the way, and then proceeds back up to the finest, performing ν relaxations on each grid up along the way. This $V(\mu,\nu)$ cycle applied to (2) is represent by the following expression:

$$\boldsymbol{v}^h \leftarrow \texttt{MV}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right), \tag{10}$$

where the $\boldsymbol{v}^h$ on the right is a current grid h approximation and $\boldsymbol{g}^h$ is a given source term. We define $\texttt{MV}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right)$ recursively as in Algorithm 3.

Algorithm 3 $\texttt{MV}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right)$; V-cycle

Given: $\boldsymbol{L}^h$, $\boldsymbol{R}_h^{2h}$, $\boldsymbol{P}_{2h}^h$

$\boldsymbol{v}^h \leftarrow \texttt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, h\right)$ ▷ Relax μ times.
if $h < H$ **then** ▷ Keep coarsening
  $\boldsymbol{g}^{2h} = \boldsymbol{R}_h^{2h}\left(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h\right)$ ▷ Transfer grid h residual to grid 2h.
  $\boldsymbol{v}^{2h} \leftarrow \texttt{MV}^{2h}\left(\boldsymbol{0}^{2h}, \boldsymbol{g}^{2h}, \mu, \nu, 2h\right)$ ▷ Apply $\texttt{MV}^{2h}$ with zero initial guess.
  $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}$ ▷ Correct fine-grid iterate.
end if
$\boldsymbol{v}^h \leftarrow \texttt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \nu, h\right)$ ▷ Relax ν times.
return $\boldsymbol{v}^h$

In some cases in practice, the approximation from the coarse grid is not accurate enough to achieve uniform V-cycle convergence, as when piecewise constant interpolation is used in coarsening. For such cases, more work in the form of a stronger cycling scheme would be required, which brings us to the $W(\mu, \nu)$-cycle that coarsens twice between relaxation steps. It is represented by the expression

$$\boldsymbol{v}^h \leftarrow \texttt{MW}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right) \tag{11}$$

and defined recursively as in Algorithm 4.

Algorithm 4 $\texttt{MW}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right)$; W-cycle

Given: $\boldsymbol{L}^h$, $\boldsymbol{R}_h^{2h}$, $\boldsymbol{P}_{2h}^h$

$\boldsymbol{v}^h \leftarrow \texttt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, h\right)$ ▷ Relax μ times.
if $h < H$ **then** ▷ Keep coarsening
  $\boldsymbol{g}^{2h} = \boldsymbol{R}_h^{2h}\left(\boldsymbol{L}^h\boldsymbol{v}^h - \boldsymbol{g}^h\right)$ ▷ Transfer grid h residual to grid 2h.
  $\boldsymbol{v}^{2h} \leftarrow \texttt{MW}^{2h}\left(\boldsymbol{0}^{2h}, \boldsymbol{g}^{2h}, \mu, \nu, 2h\right)$ ▷ Apply $\texttt{MW}^{2h}$ with zero initial guess.
  $\boldsymbol{v}^{2h} \leftarrow \texttt{MW}^{2h}\left(\boldsymbol{v}^{2h}, \boldsymbol{g}^{2h}, \mu, \nu, 2h\right)$ ▷ Apply $\texttt{MW}^{2h}$ to grid $2h$ iterate.
  $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h - \boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}$ ▷ Correct fine-grid iterate.
end if
$\boldsymbol{v}^h \leftarrow \texttt{G}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \nu, h\right)$ ▷ Relax ν times.
return $\boldsymbol{v}^h$

The respective V- and W-cycling schemes are depicted schematically of Figure 1 (a) and (b). The FMG-cycling scheme depicted in Figure 1 (c) is described next in some detail.

7. FMG. The aim of multigrid iterative solvers like the V-cycle is to achieve optimality in the sense that they converge uniformly well at a cost equivalent to just a few relaxation sweeps on the finest level. This optimality is often expressed *theoretically* as $O(n)$, where n is the number of grid points on the finest level[2]. However, this does not mean that these solvers achieve acceptable results at such a cost because the required accuracy increases with n (i.e., as h decreases). That is after all why one would want to

[2]This statement of optimality is only correct in theory, that is, in infinite precision. In practice, finite precision must be taken into account. Since precision must increase with the higher demands on accuracy that come with increasingly finer levels, so too must the cost of the finer-level operations in a V-cycle. It is therefore more accurate to interpret optimality as the cost equivalent to just a few relaxation sweeps on the finest level'. For simplicity, we nevertheless use $O(n)$ here, but it should be kept in mind as a theoretical statement.

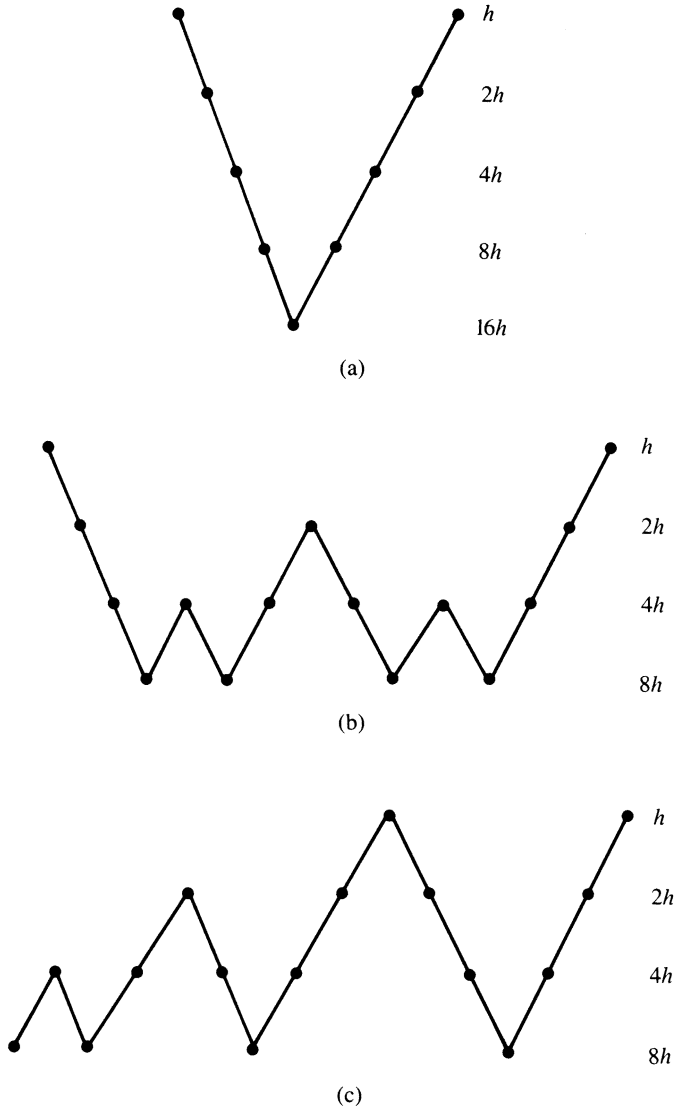

FIG. 1. *Schematic of three basic multigrid cycles: (a) MV; (b) MW; (c) FMG.*

have a larger n. In other words, the goal of refinement is to increase the accuracy of the discrete approximation to the solution of the PDE, so the number of multigrid cycles must increase accordingly. This typically leads to a total theoretical cost of at least $O(n \log n)$. This extra cost comes from starting the cycles with a naive initial guess. After all, with the solver starting its cycles on a given level, what better initial guess is there than zero? The answer is that the cycles could start by computing approximations on the coarsest up through the hierarchy so that, by the time the finest is reached, a fairly accurate approximation has already been computed at minimal cost. This idea is the basis for FMG, which often obtains accurate results at a truly optimal $O(n)$ total theoretical cost. (Theory is developed in the next section to clarify this sense of optimality.)

Returning now to the original problem of solving (1), the FMG algorithm for this purpose is based on q cycles of a generic multigrid method denoted by $\texttt{MG}^h\left(\boldsymbol{v}^h, \boldsymbol{g}^h, \mu, \nu, h\right)$, which could either be $\texttt{MV}, \texttt{MW}$, or any other basic multigrid solver. We denote the FMG algorithm by the expression

$$\boldsymbol{u}^h \leftarrow \texttt{FMG}^h\left(\mu, \nu, q, h\right) \tag{12}$$

and define it recursively in Algorithm 5.

Algorithm 5 $\texttt{FMG}^h(\mu, \nu, q, h)$; full multigrid

Given: $\boldsymbol{P}_{2h}^h$, $\boldsymbol{f}^h$

1: **if** $h < H$ **then**
2: $\quad \boldsymbol{u}^{2h} \leftarrow \texttt{FMG}^{2h}(\mu, \nu, q, h)$ ▷ Apply $\texttt{FMG}^{2h}$.
3: $\quad \boldsymbol{u}^h \leftarrow \boldsymbol{P}_{2h}^h \boldsymbol{u}^{2h}$ ▷ Start $\texttt{FMG}^h$ with grid $2h$ result.
4: **else**
5: $\quad \boldsymbol{u}^h \leftarrow \boldsymbol{0}^h$ ▷ Set initial grid H guess to zero.
6: **end if**
7: $\boldsymbol{u}^h \leftarrow \texttt{MG}^h\left(\boldsymbol{u}^h, \boldsymbol{f}^h, \mu, \nu, h\right)$ ▷ Apply $\texttt{MG}^h$ q times.
8: **return** $\boldsymbol{u}^h$

See Figure 1 (c) for a schematic of FMG using a single V-cycle on each level.

The main goal of FMG is to obtain discretization-level accuracy by using the coarse grids of a multilevel hierarchy to provide a good initial guess to a multigrid solver on the finest grid. More precisely, it attempts to achieve a *bound* on the accuracy of the final approximation to the exact discrete solution in energy that is of the same order as the *bound* on the energy error in that exact discrete solution compared to the exact PDE solution. The FMG processes for achieving this objective can be determined by theoretical analyses as developed in the next section, with a typical theoretically optimal $O(n)$ cost. But a key word here is *bound*. FMG does not necessarily reach an approximation to the exact discrete solution that is comparable to the discretization error simply because that discrete solution may be much more accurate than analysis predicts. For example, nothing prevents a discrete solution from being the exact PDE solution, and one would not expect FMG to compute the discrete solution exactly. However, such a case is perhaps not so common in practice, and accuracy at the level of the estimated discretization error might be all that one can expect in any case.

On the other hand, discrete solutions often behave in a more consistent pattern in the sense that their errors (not just their bounds) are proportional to a power of the mesh size. This may not necessarily occur on coarse levels, but in many cases it shows up asymptotically with decreasing mesh size, what has been called *saturation*. The hierarchy of grids that FMG uses provides an opportunity for estimating such occurrences so that discretization-level accuracy can be obtained. The assumption here is that coarse-grid computation is very inexpensive compared to the cost of multigrid solvers on the finest levels, so substantially more effort can be applied at negligible cost there. (Such is not the case in a parallel processing mode, however, when the number of processors is on the same order as the number of finest-level unknowns.) This effort can be used to develop a strategy for deciding whether to apply more multigrid cycles on a given level or to proceed to the next finer grid. In other words, instead of relying on error bounds that may or may not be sharp, relatively inexpensive additional cycles can be used with cost estimates to decide if they were worth it and to proceed ahead with the strategy adjusted accordingly. This refinement strategy does not of course guarantee that the process is fully able to accurately approximate the actual solutions on finer levels, but it may offer better results in many circumstances. This strategy might be further enhanced by analysis beforehand of the data (e.g., coefficients, source terms, boundary conditions, and domain) to estimate when saturation might occur.

Another important benefit of FMG that is often overlooked is its ability to avoid the pitfalls of practical error estimation. It is typical in practice for many methods to use the residuals or similar quantities as error measures. Unfortunately, they hide smooth errors that must be accounted for at an additional cost that grows with the matrix condition number. Single-grid methods applied to large-scale problems typically

exacerbate this difficulty because they tend to produce errors that are predominantly smooth. By contrast, multigrid solvers tend to produce errors that are better balanced in all error components, so their use of the residual is usually a better estimate of the actual error. But FMG is in an even better position: when properly designed, it can deliver approximations with errors that are comparable to discretization-error bounds without appealing to any convergence estimate at all. It might be argued that solvers should provide error measures, and that is certainly an important goal. However, one can think of FMG as being integral to the discretization in that it aims to achieve the bound that the discretization promises. Since discretization-error trends are usually determined a priori and are not always estimated in the computation itself, the *FMG-discretization* process is in concert with such an approach. If, however, an error measure is desired at execution, then the errors due to the solver and the discretization can be estimated according to the coarse-grid process described in the previous paragraph.

In terms of saturation, it should be noted that large-scale eigenproblems have historically been considered to be somewhat harder to solve than matrix equations. After all, eigenvalues and their vectors are typically computed by solving many matrix equations (e.g., by inverse iteration). But this is not so with multigrid for the minimal eigenvalues of elliptic PDEs. In fact, in theory, they are generally easier than matrix equations to solve by multilevel methods, in part because the solutions are as smooth as possible. But there is more to it than that because the accuracy on the levels generally behaves very regularly in the sense that it is more or less equal to some constant times a power of h, so that sharp error bounds might be more easily obtained. This would make the optimality of the FMG-discretization process very predictable.

Another benefit of FMG that is often overlooked in practice is its effectiveness for many nonlinear problems. The key point here is that the solution to the grid $2h$ problem is often (albeit not always) well inside the region of attraction of linearization methods, so FMG works virtually as well in these cases as it does for linear problems. With FAS and other multigrid ways to treat nonlinearity, it may be possible to avoid linearization entirely. But full multigrid often makes that concern secondary in the sense that the various ways to treat the nonlinearity often perform equally well when coupled with FMG. This of course assumes that the coarse levels can be efficiently formed and processed in practice, which is admittedly an issue with multigrid iterative solvers as well.

We began the discussion in this primer by providing motivation for using relaxation on coarse levels to improve the solver on the fine grid. We followed that by showing how to construct a multigrid solver based on this concept. We then described how FMG can provide accurate results at a cost proportional to the degrees of freedom in the finest-level discretization. Next, we develop some results that ensures theoretical efficiency based on these constructs.

8. Basic theory. We begin this section by describing properties of the operators introduced in (9) that lay the theoretical groundwork for multigrid. These properties result from $\boldsymbol{S}^h$ and $\boldsymbol{T}^h$ being energy-orthogonal projections onto $R(\boldsymbol{P}_{2h}^h)$, the *range* of $\boldsymbol{P}_{2h}^h$, and $N(\boldsymbol{R}_h^{2h}\boldsymbol{L}^h)$, the *null space* or *kernel* of $\boldsymbol{R}_h^{2h}\boldsymbol{L}^h$, respectively. Clearly, $\boldsymbol{S}^h$ maps onto $R(\boldsymbol{P}_{2h}^h)$ because its lead operator (i.e., what is applied last) is $\boldsymbol{P}_{2h}^h$. $\boldsymbol{S}^h$ is also a projection because

$$\begin{aligned}\left(\boldsymbol{S}^h\right)^2 &= \boldsymbol{P}_{2h}^h(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{P}_{2h}^h(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\\ &= \boldsymbol{P}_{2h}^h(\boldsymbol{L}^{2h})^{-1}\boldsymbol{L}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\\ &= \boldsymbol{S}^h.\end{aligned}$$

Also, $\boldsymbol{T}^h$ maps onto $N(\boldsymbol{R}_h^{2h}\boldsymbol{L}^h)$ because

$$\begin{aligned}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{T}^h &= \boldsymbol{R}_h^{2h}\boldsymbol{L}^h(\boldsymbol{I}^h - \boldsymbol{S}^h)\\ &= \boldsymbol{R}_h^{2h}\boldsymbol{L}^h(\boldsymbol{I}^h - \boldsymbol{P}_{2h}^h(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h)\\ &= \boldsymbol{R}_h^{2h}\boldsymbol{L}^h - \boldsymbol{L}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\\ &= \boldsymbol{0}.\end{aligned}$$

$\boldsymbol{T}^h$ is also a projection because

$$\left(\boldsymbol{T}^h\right)^2 = \left(\boldsymbol{I}^h - \boldsymbol{S}^h\right)^2 = \left(\boldsymbol{I}^h\right)^2 - 2\boldsymbol{S}^h + \left(\boldsymbol{S}^h\right)^2 = \boldsymbol{I}^h - 2\boldsymbol{S}^h + \boldsymbol{S}^h = \boldsymbol{T}^h.$$

Finally, these projections are orthogonal to each other because

$$\boldsymbol{T}^h\boldsymbol{S}^h = (\boldsymbol{I}^h - \boldsymbol{S}^h)\boldsymbol{S}^h = \boldsymbol{S}^h - \left(\boldsymbol{S}^h\right)^2 = \boldsymbol{0}.$$

Keep in mind that the range of $\boldsymbol{P}_{2h}^h$ consists of vectors that can be exactly represented by interpolation from the coarse grid, while the kernel of $\boldsymbol{R}_h^{2h}\boldsymbol{L}^h$ consists of error vectors that cannot be represented at all by passing their residuals to the coarse grid. That is, $\boldsymbol{S}^h$ and $\boldsymbol{T}^h$ provide a decomposition into errors that *can* be eliminated by the coarse grid and those that *cannot* (and must therefore be dealt with on the fine grid). More specifically, we have the following:

- Any grid h vector $\boldsymbol{e}^h$ can be expressed as $\boldsymbol{e}^h = \boldsymbol{s}^h + \boldsymbol{t}^h$, where $\boldsymbol{s}^h \in R(\boldsymbol{P}_{2h}^h)$, $\boldsymbol{t}^h \in N(\boldsymbol{R}_h^{2h}\boldsymbol{L}^h)$, and $\langle \boldsymbol{s}^h, \boldsymbol{t}^h\rangle_{\boldsymbol{L}^h} = 0$ (that is, $\boldsymbol{s}^h$ and $\boldsymbol{t}^h$ are orthogonal in the energy inner product).
- $\boldsymbol{T}^h\boldsymbol{s}^h = \boldsymbol{0}$, $\boldsymbol{S}^h\boldsymbol{t}^h = \boldsymbol{0}$, $\boldsymbol{T}^h\boldsymbol{t}^h = \boldsymbol{t}^h$, $\boldsymbol{S}^h\boldsymbol{s}^h = \boldsymbol{s}^h$, $\boldsymbol{T}^h\boldsymbol{e}^h = \boldsymbol{t}^h$, $\boldsymbol{S}^h\boldsymbol{e}^h = \boldsymbol{s}^h$, $\langle \boldsymbol{T}^h\boldsymbol{v}^h, \boldsymbol{S}^h\boldsymbol{z}^h\rangle_{\boldsymbol{L}^h} = 0\ \forall \boldsymbol{v}^h, \boldsymbol{z}^h$, and $\|\boldsymbol{S}^h\|_{\boldsymbol{L}^h} = \|\boldsymbol{T}^h\|_{\boldsymbol{L}^h} = 1$.
- $\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h} = \min_{\boldsymbol{u}^{2h}} \|\boldsymbol{e}^h - P_{2h}^h\boldsymbol{u}^{2h}\|_{\boldsymbol{L}^h}$ (that is, coarse-grid correction minimizes the energy norm of the error over all possible corrections in the range of interpolation).

8.1. Basic theory for multigrid cycling as an iterative method. The convergence theory developed here begins with the two-grid scheme and its error propagation matrix $\boldsymbol{TG}^h = \boldsymbol{G}^h\boldsymbol{T}^h$, where $\boldsymbol{G}^h = \boldsymbol{I}^h - \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h$. Each cycle converges with uniform energy bound for any initial guess if and only if

$$\|\boldsymbol{TG}^h\|_{\boldsymbol{L}^h}^2 = \|\boldsymbol{G}^h\boldsymbol{T}^h\|_{\boldsymbol{L}^h}^2 = \left\|(I - \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h)\boldsymbol{T}^h\right\|_{\boldsymbol{L}^h}^2 \le \gamma \tag{13}$$

for some fixed $\gamma \in [0, 1)$. This condition is the same as asking that the strong measure be bounded away from zero:

$$\mathcal{M}_s(\boldsymbol{t}^h) \ge \delta \tag{14}$$

for some fixed $\delta > 0$ and for all $\boldsymbol{t}^h$ in the range of $\boldsymbol{T}^h$. To see this, note first that

$$\begin{aligned}&\|\boldsymbol{TG}^h\|_{\boldsymbol{L}^h}^2\\ &= \max_{\boldsymbol{u}^h \ne \boldsymbol{0}^h} \frac{\langle \boldsymbol{L}^h(I - \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h)\boldsymbol{T}^h\boldsymbol{u}^h, (I - \frac{1}{\|\boldsymbol{L}^h\|}\boldsymbol{L}^h)\boldsymbol{T}^h\boldsymbol{u}^h\rangle}{\langle \boldsymbol{L}^h\boldsymbol{u}^h, \boldsymbol{u}^h\rangle}\\ &= \max_{\boldsymbol{t}^h = \boldsymbol{T}^h\boldsymbol{t}^h \ne \boldsymbol{0}^h, \boldsymbol{s}^h = \boldsymbol{S}^h\boldsymbol{s}^h} \frac{\langle \boldsymbol{L}^h\boldsymbol{t}^h, \boldsymbol{t}^h\rangle - \frac{2}{\|\boldsymbol{L}^h\|}\langle \boldsymbol{L}^h\boldsymbol{t}^h, \boldsymbol{L}^h\boldsymbol{t}^h\rangle + \frac{1}{\|(\boldsymbol{L}^h)^2\|}\langle \left(\boldsymbol{L}^h\right)^2\boldsymbol{t}^h, \boldsymbol{L}^h\boldsymbol{t}^h\rangle}{\langle \boldsymbol{L}^h\boldsymbol{t}^h, \boldsymbol{t}^h\rangle + \langle \boldsymbol{L}^h\boldsymbol{s}^h, \boldsymbol{s}^h\rangle}\\ &= \max_{\boldsymbol{t}^h = \boldsymbol{T}^h\boldsymbol{t}^h \ne \boldsymbol{0}^h} \frac{\langle \boldsymbol{L}^h\boldsymbol{t}^h, \boldsymbol{t}^h\rangle - \frac{1+\xi}{\|\boldsymbol{L}^h\|}\langle \boldsymbol{L}^h\boldsymbol{t}^h, \boldsymbol{L}^h\boldsymbol{t}^h\rangle}{\langle \boldsymbol{L}^h\boldsymbol{t}^h, \boldsymbol{t}^h\rangle}\\ &= 1 - (1+\xi)\mathcal{M}_s(\boldsymbol{t}^h),\end{aligned}$$

where ξ is the value of $\frac{\frac{1}{\|\boldsymbol{L}^h\|}\langle(\boldsymbol{L}^h)^2\boldsymbol{t}^h,\boldsymbol{L}^h\boldsymbol{t}^h\rangle}{\langle\boldsymbol{L}^h\boldsymbol{t}^h,\boldsymbol{L}^h\boldsymbol{t}^h\rangle}$ at the maximizing $\boldsymbol{t}^h$. Note further that

$$0 < \frac{\frac{1}{\|\boldsymbol{L}^h\|}\langle(\boldsymbol{L}^h)^2\boldsymbol{t}^h,\boldsymbol{L}^h\boldsymbol{t}^h\rangle}{\langle\boldsymbol{L}^h\boldsymbol{t}^h,\boldsymbol{L}^h\boldsymbol{t}^h\rangle} \leq 1,$$

so $\xi \in [0,1)$. As asserted, we thus have that (13) holds if and only if $1-(1+\xi)\mathcal{M}_s(\boldsymbol{t}^h) \leq \gamma$, that is, $\mathcal{M}_s(\boldsymbol{t}^h) \geq \delta \equiv \frac{1-\gamma}{1+\xi} > 0$. In fact, $\delta \in (\frac{1-\gamma}{2}, 1-\gamma]$.

To understand the two-grid condition (14) a little better, we make a (generally unrealistic) assumption that $\boldsymbol{t}^h$ is an eigenvector of $\boldsymbol{L}^h$ with corresponding eigenvalue λ. Then

$$\mathcal{M}_s(\boldsymbol{t}^h) = \frac{\langle\boldsymbol{L}^h\boldsymbol{t}^h,\boldsymbol{L}^h\boldsymbol{t}^h\rangle}{\|\boldsymbol{L}^h\|\langle\boldsymbol{L}^h\boldsymbol{t}^h,\boldsymbol{t}^h\rangle} = \frac{\lambda^2}{\lambda\|\boldsymbol{L}^h\|} = \frac{\lambda}{\|\boldsymbol{L}^h\|},$$

so the two-grid condition requires that λ be theoretically comparable (up to δ) to $\|\boldsymbol{L}^h\|$, the largest eigenvalue of $\boldsymbol{L}^h$. Of course, the practical quality of convergence depends on δ, and a value of δ near zero means that γ is near 1, so it signals poor expected numerical performance. Generally, we want error components that $\boldsymbol{T}^h$ cannot eliminate to lie primarily in the upper spectrum of $\boldsymbol{L}^h$.

Unfortunately, counterexamples (such as standard cell-centered coarsening of a discrete Poisson equation; c.f., [1]) show that this two-grid condition does not suffice to establish uniform convergence of the multilevel V-cycle. A condition that is sufficient, however, is the following property:

DEFINITION 8.1. ***Strong Approximation Property (SAP)***
A coarse-grid correction scheme satisfies the Strong Approximation Property if and only if

$$\min_{\boldsymbol{u}^{2h}} \|\boldsymbol{e}^h - \boldsymbol{I}_{2h}^h\boldsymbol{u}^{2h}\|_{\boldsymbol{L}^h}^2 \leq \frac{C}{\|\boldsymbol{L}^h\|}\langle\boldsymbol{L}^h\boldsymbol{e}^h,\boldsymbol{L}^h\boldsymbol{e}^h\rangle$$

for some fixed constant $C < \infty$ and for all grid h vectors $\boldsymbol{e}^h$.

The SAP can be written a little more compactly as

$$\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 \leq \frac{C}{\|\boldsymbol{L}^h\|}\langle\boldsymbol{L}^h\boldsymbol{e}^h,\boldsymbol{L}^h\boldsymbol{e}^h\rangle. \tag{15}$$

Another way to write this condition is in terms of the multigrid correction matrix and the strong measure:

$$\frac{\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2}{\|\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2} \leq C\mathcal{M}_s(\boldsymbol{e}^h).$$

Note that this property reduces to the two-grid condition (with $C = \frac{1}{\delta}$) when $\boldsymbol{e}^h$ is in the range of $\boldsymbol{T}^h$, that is, $\boldsymbol{e}^h = \boldsymbol{T}^h\boldsymbol{e}^h$. (Remember that vectors in the range of $\boldsymbol{T}^h$ are precisely those that the coarse-grid correction leaves untouched and therefore undiminished.)

THEOREM 1. (V-cycle convergence) *The strong approximation property is a sufficient condition for uniformly bounded convergence in the energy norm of the $V(0,1)$-cycle in that its error propagation matrix $\boldsymbol{M}\boldsymbol{V}^h$ satisfies*

$$\|\boldsymbol{M}\boldsymbol{V}^h\|_{\boldsymbol{L}^h}^2 \leq 1 - \frac{1}{C}.$$

Proof. Remembering that $\boldsymbol{e}^h = \boldsymbol{S}^h\boldsymbol{e}^h + \boldsymbol{T}^h\boldsymbol{e}^h$ and that $\boldsymbol{S}^h\boldsymbol{e}^h$ is in the range of $\boldsymbol{P}_{2h}^h$, then we can write $\boldsymbol{e}^h = \boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h} + \boldsymbol{T}^h\boldsymbol{e}^h$ for some grid $2h$ vector $\boldsymbol{v}^{2h}$. With (15), we thus

have that

$$\begin{aligned}\left\|\boldsymbol{G}^h\boldsymbol{e}^h\right\|_{\boldsymbol{L}^h}^2 &= \|\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 - \frac{2}{\|\boldsymbol{L}^h\|}\|\boldsymbol{L}^h\boldsymbol{e}^h\|^2 + \frac{1}{\|\boldsymbol{L}^h\|^2}\langle\left(\boldsymbol{L}^h\right)^2\boldsymbol{e}^h, \boldsymbol{L}^h\boldsymbol{e}^h\rangle \\ &\leq \|\boldsymbol{S}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 + \|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 - \frac{1}{\|\boldsymbol{L}^h\|}\|\boldsymbol{L}^h\boldsymbol{e}^h\|^2 \\ &\leq \|\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}\|_{\boldsymbol{L}^h}^2 + \left(1 - \frac{1}{C}\right)\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2. \end{aligned} \tag{16}$$

To obtain a recursion formula for $\mathbf{MV}^h$ in terms of its grid $2h$ counterpart $\mathbf{MV}^{2h}$, consider the following important relationship between the residual and the error for $\boldsymbol{u}^h = (\boldsymbol{L}^h)^{-1}\boldsymbol{f}^h + \boldsymbol{e}^h$ an approximation to the solution $(\boldsymbol{L}^h)^{-1}\boldsymbol{f}^h$ of (1):

$$\boldsymbol{L}^h\boldsymbol{u}^h - \boldsymbol{f}^h = \boldsymbol{L}^h((\boldsymbol{L}^h)^{-1}\boldsymbol{f}^h + \boldsymbol{e}^h) - \boldsymbol{f}^h = \boldsymbol{L}^h\boldsymbol{e}^h.$$

Thus, the solution to the grid $2h$ equation is

$$(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}(\boldsymbol{L}^h\boldsymbol{u}^h - \boldsymbol{f}^h) = (\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{e}^h.$$

Letting $\boldsymbol{e}^{2h}$ denote the error in the zero initial guess on grid $2h$, we thus have the error expansion

$$\mathbf{0}^{2h} = (\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{e}^h + \boldsymbol{e}^{2h}.$$

Now, since $\mathbf{MV}^{2h}$ is the error propagator for the grid $2h$ V-cycle, we must then have that the new approximate grid $2h$ solution is

$$\boldsymbol{u}^{2h} = (\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{e}^h + \mathbf{MV}^{2h}\boldsymbol{e}^{2h} = \boldsymbol{v}^{2h} - \mathbf{MV}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{e}^h,$$

where we used the fact that

$$(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{e}^h = (\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h} = \boldsymbol{v}^{2h}.$$

Finally, since $\boldsymbol{P}_{2h}^h\boldsymbol{u}^{2h}$ is subtracted from $\boldsymbol{u}^h$, the exact component $\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}$ cancels the smooth component $\boldsymbol{s}^h$ of $\boldsymbol{u}^h$, leaving the error $\mathbf{MV}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{e}^h$ in its place. This long and rather intricate argument has led us to the conclusion that the error propagations of the fine and coarse V-cycles are related according to the recursion[3]

$$\mathbf{MV}^h = \boldsymbol{G}^h\left(\left(\boldsymbol{P}_{2h}^h\mathbf{MV}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\right) + \boldsymbol{T}^h\right). \tag{17}$$

Note that

$$\begin{aligned}&\left(\left(\boldsymbol{P}_{2h}^h\mathbf{MV}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\right) + \boldsymbol{T}^h\right)\boldsymbol{e}^h \\ &\qquad = \left(\left(\boldsymbol{P}_{2h}^h\mathbf{MV}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\right) + \boldsymbol{T}^h\right)\left(\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h} + \boldsymbol{T}^h\boldsymbol{e}^h\right) \\ &\qquad = \boldsymbol{P}_{2h}^h\mathbf{MV}^{2h}(\boldsymbol{L}^{2h})^{-1}\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h} + \boldsymbol{T}^h\boldsymbol{e}^h \\ &\qquad = \boldsymbol{P}_{2h}^h\mathbf{MV}^{2h}\boldsymbol{v}^{2h} + \boldsymbol{T}^h\boldsymbol{e}^h. \end{aligned} \tag{18}$$

[3]It is useful to consider the two extreme cases of $\mathbf{MV}^{2h}$ here by reflecting on the form of (17). If the V-cycle on grid $2h$ happens to be an exact solver, then $\mathbf{MV}^{2h}$ would be the zero matrix, so the grid h V-cycle becomes an optimally convergent two-grid cycle. At the other extreme, if the grid $2h$ V-cycle makes no change to the grid $2h$ approximation (i.e., $\mathbf{0}^{2h}$), then $\mathbf{MV}^{2h}$ would be the grid $2h$ identity and the grid h V-cycle would devolve effectively into fine-grid relaxation only. It is the intermediate case that can be expected (and is assumed in (19) below), where the grid $2h$ V-cycle reduces a significant fraction of grid $2h$ error.

If we substitute into (16) the new error $\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}+\boldsymbol{T}^h\boldsymbol{e}^h$ and its smooth component $\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}$ in place of $\boldsymbol{e}^h$ and $\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}$, respectively, then (16)-(18) together with the induction hypothesis

$$\|\mathbf{M}\mathbf{V}^{2h}\|_{\boldsymbol{L}^{2h}}^2 \leq 1-\frac{1}{C} \tag{19}$$

shows that

$$\begin{aligned}
\|\mathbf{M}\mathbf{V}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 &= \|\boldsymbol{G}^h\left(\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}+\boldsymbol{T}^h\boldsymbol{e}^h\right)\|_{\boldsymbol{L}^h}^2 \\
&\leq \|\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}\|_{\boldsymbol{L}^h}^2+\left(1-\frac{1}{C}\right)\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 \\
&= \langle\boldsymbol{L}^h\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h},\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}\rangle+\left(1-\frac{1}{C}\right)\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 \\
&= \langle\boldsymbol{R}_h^{2h}\boldsymbol{L}^h\boldsymbol{P}_{2h}^h\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h},\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}\rangle+\left(1-\frac{1}{C}\right)\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 \\
&= \|\mathbf{M}\mathbf{V}^{2h}\boldsymbol{v}^{2h}\|_{\boldsymbol{L}^{2h}}^2+\left(1-\frac{1}{C}\right)\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2 \\
&\leq \left(1-\frac{1}{C}\right)\left(\|\boldsymbol{v}^{2h}\|_{\boldsymbol{L}^{2h}}^2+\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2\right) \\
&= \left(1-\frac{1}{C}\right)\left(\|\boldsymbol{P}_{2h}^h\boldsymbol{v}^{2h}\|_{\boldsymbol{L}^h}^2+\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2\right) \\
&= \left(1-\frac{1}{C}\right)\|\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2.
\end{aligned}$$

This proves our assertion that $\|\mathbf{M}\mathbf{V}^h\|_{\boldsymbol{L}^h}^2 \leq 1-\frac{1}{C}$. ∎

The result of this theorem holds for any relaxation method governed by an error propagation matrix $\boldsymbol{G}^h$ as long as (16) holds. This bound states that one relaxation sweep reduces the square of the energy of the error by $\frac{1}{C}$ times the square of the oscillatory part of the error. In this very general form, any relaxation scheme for which this holds (e.g., steepest descent or conjugate gradients) guarantees optimal performance when used in a V-cycle.

The strong approximation property requires that vectors in the kernel of $\boldsymbol{L}^h$ be exactly eliminated by coarse-grid correction. This follows from noting that if $\boldsymbol{e}^h$ is in the kernel of $\boldsymbol{L}^h$, then the right-hand side of the SAP vanishes, and thus so too must the left-hand side. However, we are assuming that $\boldsymbol{L}^h$ is SPD, so its kernel consists only of $\mathbf{0}^h$ and this requirement is, therefore, trivially satisfied by setting $\boldsymbol{u}^{2h}=\mathbf{0}^{2h}$. On the other hand, the strong approximation property also clearly shows that any vector that is in the near kernel (in the sense of yielding a relatively small residual, $\boldsymbol{L}^h\boldsymbol{e}^h$) must be nearly eliminated by coarse-grid correction. This again follows from noting that if $\boldsymbol{e}^h$ is in the near kernel of $\boldsymbol{L}^h$, then the right-hand side of the SAP is small, and thus so too must the left-hand side be. It is this observation that provides our motivation for constructing an interpolation operator that adequately approximates near-kernel components. This focus on the near kernel is generally sufficient because algebraically oscillatory vectors automatically satisfy the SAP: by definition, an algebraically oscillatory error $\boldsymbol{e}^h$ satisfies $\mathcal{M}_s(\boldsymbol{e}^h)\geq\delta$ for some $\delta\gg 0$, so it follows that

$$\frac{\|\boldsymbol{T}^h\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2}{\|\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2} \leq \frac{\|\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2}{\|\boldsymbol{e}^h\|_{\boldsymbol{L}^h}^2} \leq C\mathcal{M}_s(\boldsymbol{e}^h),$$

where $C=\frac{1}{\delta}\ll\infty$. Finally, since the SAP involves the choice of coarse-grid points and interpolation to the fine grid, it gives us a way to assess the quality of the coarse-grid correction process.

As we have just seen, the strong measure is related to the strong approximation property in the energy inner product. We have a similar relationship for the weak case in the Euclidean norm.

DEFINITION 8.2. ***Weak Approximation Property (WAP)*** *A coarse-grid correction scheme satisfies the Weak Approximation Property if and only if*

$$\min_{\boldsymbol{u}^{2h}} \|\boldsymbol{e}^h - P_{2h}^h \boldsymbol{u}^{2h}\|^2 \leq \frac{C}{\|\boldsymbol{L}^h\|} \langle \boldsymbol{L}^h \boldsymbol{e}^h, \boldsymbol{e}^h \rangle.$$

At this point, it should be no surprise that the WAP can be written in terms of the weak measure:

$$\frac{\|\boldsymbol{T}^h \boldsymbol{e}^h\|^2}{\|\boldsymbol{e}^h\|^2} \leq C \mathcal{M}_w(\boldsymbol{e}^h).$$

By arguments similar to those for the strong approximation property, the weak approximation property confirms uniform convergence of the two-grid form of multigrid in the Euclidean norm. It can also confirm convergence of multigrid cycling schemes like the W-cycle that is more aggressive than the V-cycle, provided the effort spent on the coarse grids is commensurate with the size of C.

Some relationships between the SAP and the WAP, together with an additional approximation property, are developed in Appendix A. They can be useful for designing geometrically based multilevel algorithms, but are especially so for algebraic multigrid methods (e.g., AMG [2]).

8.2. Basic theory for FMG as a direct method for solving PDEs. With the theory of convergence of multigrid cycles as an iterative solver at hand, we now turn to the more direct FMG approach that uses these cycles as basic building blocks to achieve accuracy comparable to that of the discretization at optimal cost. To explain this optimality in some detail, we first need to delve into the world of the continuum. To avoid too much abstraction, we settle with illustrating the general case by way of the following model two-dimensional Poisson problem:

$$-\Delta u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{20}$$

where $\Delta = \nabla \cdot \nabla = \partial_{xx} + \partial_{yy}$, Ω is the unit square with boundary $\partial\Omega$, and f is a square-integrable function on Ω. It is tempting to define the energy functional for (20) by $\langle -\Delta v, v \rangle - 2\langle v, f \rangle$, with angle brackets here denoting the $L^2(\Omega)$ inner product expressed by $\langle w, z \rangle = \int_\Omega wv \, dx \, dy$. However, simple functions in the discretization that are pieced together across the elements are generally not twice differentiable in any sense and are therefore not admissible arguments for such a definition. Instead, this definition is *weakened* using the Gauss Divergence Theorem, which is given by

$$\langle -\nabla \cdot \nabla u, u \rangle = \langle \nabla u, \nabla u \rangle. \tag{21}$$

This identity is valid because so-called *full regularity* of (20) and square integrability of f guarantee that u is sufficiently smooth. So (21) leads to the following definitions of the respective energy functional and energy norm for the weak form of (20):

$$F(u) \equiv \langle \nabla u, \nabla u \rangle - 2\langle u, f \rangle \text{ and } \|u\|_L \equiv \langle \nabla u, \nabla u \rangle^{\frac{1}{2}}.$$

We use subscript L here to suggest that it is derived from the strong form (20). We can thus take the formal view that L and $\|u\|_L$ correspond to weak forms of $-\Delta$ and $\langle Lu, u \rangle^{\frac{1}{2}}$, respectively.

The discretization of (20) is now a matter of choosing a finite-dimensional subspace of the admissible function space for F. To be specific, suppose that Ω is partitioned by

a hierarchy of triangles on a uniform grid, starting with a coarsest single-point mesh size of $\frac{1}{2}$ and continuing with uniform refinement by a factor of 2 to the finest grid with a sufficiently small mesh size. For each h in the hierarchy, let $\mathcal{H}^h$ denote the space of the finite element *functions* u^h that are linear in each grid h element, continuous across the domain, and zero on the boundary. Then the discretization on each grid h is derived by minimizing $F(u^h)$ over $u^h \in \mathcal{H}^h$, leading to the matrix equation in (1) defined in terms of the vector $\boldsymbol{u}^h$ whose entries are u^h's nodal values[4]. This *Rayleigh-Ritz* derivation is straightforward though lengthy enough that we leave it to the interested reader to find it in any of the available finite element texts such as [10]. A consequence of this formulation is that the equations and induced intergrid transfers satisfy the variational conditions on every level.

Since FMG has the PDE in mind, we need to have a measurement of the error in its result $\boldsymbol{v}^h$ compared to the function u that minimizes F. Another advantage of this variational formulation is that, instead of the vector $\boldsymbol{v}^h$, we can compare its equivalent finite element function v^h to u as measured in the energy norm, namely, $\|v^h - u\|_L$. (By equivalent we mean that $\boldsymbol{v}^h$ and v^h have the same nodal values.) This finite element view simplifies the discussion because all of the approximations and the exact weak solution of the PDE are in the same space, meaning for one that we are able to avoid using intergrid transfer operators. We are also able to avoid worrying about relative errors because all of the errors we are about to consider are measured in terms of the fixed solution u, whose energy norm is absorbed in the constant C introduced by assuming that

$$\|u^h - u\|_L \leq Ch^m, \tag{22}$$

where u^h is the exact minimizer of the functional on level h, C is some fixed constant, and $m > 0$ is the order of the energy discretization error. For standard continuous piecewise linear finite element discretization of (20) on a uniform triangulation and sufficient smoothness of u, this estimate holds with $m = 1$.

The goal of FMG is to produce a grid h function that approximates the discrete solution to within the same accuracy that the discrete solution approximates the continuum solution, that is,

$$\|v^h - u^h\|_L \leq Ch^m. \tag{23}$$

To see what this requires of the multigrid cycles, suppose for induction purposes that (23) is true for $2h$:

$$\|v^{2h} - u^{2h}\|_L \leq C(2h)^m. \tag{24}$$

This holds trivially for the coarsest grid H because the matrix problem there is the scalar equation $\boldsymbol{L}^H_{1,1}\boldsymbol{u}^H_1 = \boldsymbol{f}^H_1$ at the single grid point, and that is solved exactly by one sweep of Richardson's iteration with the particular scale $\frac{1}{\|\boldsymbol{L}^H\|} = \frac{1}{\boldsymbol{L}^H_{1,1}}$ that we have chosen.

Next, we need to appeal to an orthogonality property afforded by our variational setting. To this end, first note that a function w minimizes $F(w)$ over an admissible subspace $\mathcal{S}$ if and only if it minimizes the energy error $\|w - u\|_L$ over $\mathcal{S}$. (This follows directly from the same logic that we used for the discrete case to show that the energy functional and the square of the energy error differ by a constant.) So we can think of any subspace minimizer w of the energy functional also as a minimizer of the energy error $\|w - u\|_L$. But, for any such w, it must also be that $w - u$ is energy orthogonal to $\mathcal{S}$, that is, $\langle \nabla(w - u), \nabla z\rangle = 0$ for any $z \in \mathcal{S}$. (The closest energy approximation to $u \in \mathcal{S}$ is its energy-orthogonal projection onto that subspace.) To see this, let z be any nonzero

[4]This raises a subtle point. Note, in general, that u^h does not satisfy (21) because the left side of the identity is usually undefined. But (21) is valid for u and the right side is well-defined for both u^h and u, so both are admissible functions for F, meaning that it is valid to seek u by minimizing $F(u^h)$ over $\mathcal{H}^h$.

function in $\mathcal{S}$ and s be any nonzero scalar in $\mathbb{R}$. Then adding sz to w yields

$$\|w + sz - u\|_L^2 = \|w - u\|_L^2 + 2s\langle\nabla(w-u), \nabla z\rangle + s^2\|z\|_L^2.$$

If we assume that $w - u$ is not energy orthogonal to $\mathcal{S}$, then we must be able to choose z so that $\langle\nabla(w-u), \nabla z\rangle \neq 0$. However, if we choose $s = -\|z\|_L^{-2}\langle\nabla(w-u), \nabla z\rangle$, then

$$\begin{aligned}\|w + sz - u\|_L^2 &= \|w - u\|_L^2 - 2\|z\|_L^{-2}\langle\nabla(w-u), \nabla z\rangle^2 + \|z\|_L^{-4}\langle\nabla(w-u), \nabla z\rangle^2\|z\|_L^2 \\ &= \|w - u\|_L^2 - \|z\|_L^{-2}\langle\nabla(w-u), \nabla z\rangle^2 \\ &< \|w - u\|_L^2.\end{aligned}$$

This is a contradiction because w is assumed to be the minimizer over $\mathcal{S}$. So it must be that $\langle\nabla(w-u), \nabla z\rangle = 0$ for all $z \in \mathcal{S}$, as asserted.

Now, since u^h is the best energy approximation to u from $\mathcal{S} \equiv \mathcal{H}^h$, we can thus conclude that $u^h - u$ is energy-orthogonal to $u^{2h} - u^h \in \mathcal{S}$, which in turn implies that

$$\|u^{2h} - u\|_L^2 = \|u^{2h} - u^h\|_L^2 + \|u^h - u\|_L^2 \tag{25}$$

and, hence, that

$$\|u^{2h} - u^h\|_L^2 \leq \|u^{2h} - u\|_L^2. \tag{26}$$

But (25) also implies that u^{2h} minimizes $\|u^{2h} - u^h\|_L^2$. Thus, by an energy-orthogonal argument analogous to the above, together with (26), we thus have that

$$\|v^{2h} - u^h\|_L^2 \leq \|v^{2h} - u^{2h}\|_L^2 + \|u^{2h} - u^h\|_L^2 \leq \|v^{2h} - u^{2h}\|_L^2 + \|u^{2h} - u\|_L^2. \tag{27}$$

Finally, (27), (24), and (22) (with $2h$ replacing h) combine to yield

$$\|v^{2h} - u^h\|_L \leq \left((C(2h)^m)^2 + (C(2h)^m)^2\right)^{\frac{1}{2}} = 2^{m+\frac{1}{2}}Ch^m. \tag{28}$$

The conclusion is that, for FMG to achieve accuracy in energy comparable to that of the discretization itself, we need only apply enough multigrid cycles on each level to reduce the error by a factor of $2^{-(m+\frac{1}{2})}$. For the model problem with $m = 1$, the requirement is enough V-cycles to reduce the error by about a third. A $V(2,1)$-cycle is in common use, most likely due to the recommendation by Achi Brandt in the early days of multigrid, and it more than delivers that reduction factor for the model problem.

Other solvers applied to discrete elliptic PDEs have been able to achieve optimal iteration cost in some cases, but it seems improbable that the true optimality that FMG has been able to achieve could be obtained in general without appealing to the relatively inexpensive coarser levels for reasonably accurate initial guesses. In any case, this optimality of FMG came at a welcomed surprise in the early days of multigrid research, and it ensured that multigrid would become the method of choice for many applications.

9. Development. The assertions made thus far in this primer are supported by many theoretical and numerical results that can be found in the extensive multigrid literature. On the other hand, several remarks that we make in this section are not yet so well supported. The reader should keep in mind that some of the claims made here are merely opinions of the authors, albeit based on intuition gained over several combined decades of experience with multigrid theory and practice.

The purpose of this section stems from the limitation of the above theory in that it does not really assure that a particular multigrid method is very efficient in practice. This limitation is a common weakness of convergence estimates for numerical solution methods because they typically involve indeterminant constants such as the C in the WAP and SAP. Nevertheless, the theory is useful because it suggests certain properties

of the algorithm components that make for a good multigrid approach. Our specific aim is therefore to show how one might use the theory developed above to choose effective coarse grids and interpolation operators. We also suggest how this theory might guide the formulation of a suitable discretization method and even perhaps the PDE itself, provided one is allowed access to this assumed source of the matrix problem. Finally, we note that the objective of this section is to provide a broad view of the relevant methodology, thereby omitting the many details that would be beyond the scope of this primer.

9.1. Algorithm. To suggest how to use the theory to guide multigrid development, we focus here on the classical development of Smoothed Aggregation (SA [13]) because it provides a natural setting for this discussion in that it deals directly with the near-kernel components whose treatments are critical to achieving good approximation properties. SA is an *algebraic* multigrid method because it is constructed with little or no knowledge or use of a connection between the matrix equation and any geometric entity such as a grid. Our focus here is admittedly a departure from the *geometric* approach that we have so far had in mind. However, this focus not only introduces algebraic methods, but it also has the advantage that understanding the principles that arise in this setting gives a deeper sense of what makes multigrid work in general.

To avoid potential confusion, we need to be clear about terminology. While algebraic methods do not necessarily involve geometric information directly, the literature typically uses some geometric notions in the development. We have already used *smooth* in an algebraic context, which we will continue to do. But we will also use the standard convention of *nodes* to refer to the unknowns in the matrix equation, that is, the vector index or the degree of freedom. We also use *neighborhood* to refer to points connected directly in the matrix $\boldsymbol{L}^h$.

We first need to point out that, while V-cycle convergence is usually the desired target in applications because of its superior efficiency, the nature of the energy norm that the SAP is based on makes it more difficult to use as a design tool. The problem stems from the need to determine interpolation *locally*, that is, in an algebraic neighborhood. Using the energy norm to measure how well interpolation approximates local errors is problematic because $\boldsymbol{L}^h$, which is present in that norm, generally reaches to nodes outside of the neighborhood. This lack of locality inhibits the ability to isolate the design of interpolation. The problem comes from the need to determine what is smooth, that is, what is locally in the null space or near null space of $\boldsymbol{L}^h$. For this reason, it is more common in practice to develop interpolation schemes based on the WAP because estimates involving the Euclidean norm can be wholly restricted to individual neighborhoods. This locality provides the foundation for smoothed aggregation, as we show below.

The practical importance of determining the coarsening process locally cannot be overemphasized. Just as finite elements realized its true potential when it transitioned from a global spectral approach to a local piecewise polynomial methodology, so too is multigrid most effective when coarse-grid corrections are constructed from local approximations of errors based on their local character. The dimension of the space of smooth errors that relaxation cannot effectively reduce is typically a significant fraction of the dimension of the fine-level space. This is generally much too large to permit an exact computation of a global basis for that space. This reasoning is analogous to saying that global spectral discretizations based on eigenvectors or Fourier modes are generally impractical. So the approach taken by algebraic multigrid methods is to attempt to approximate algebraically smooth errors over small sets dofs that are interconnected in the matrix by using a significantly smaller number of nodes. The intent is that these approximations can then be pieced together to provide adequate global approximation to all algebraically smooth errors.

To get a glimpse of how this is done for smoothed aggregation, we first note that

SA begins by partitioning the fine nodes into groups or *aggregates* of nodes. This is accomplished by aggregating several nodes that are strongly interconnected in $\boldsymbol{L}^h = \left(a_{ij}^h\right)$. The basic idea is that $|a_{ij}^h|$ is used as a measure to determine the dependency of the values at points i and j of an algebraically smooth error, $\boldsymbol{e}^h = \left(e_i^h\right)$. The motivation here is that, for scalar elliptic equations, a large value of $|a_{ij}^h|$ relative to $\sqrt{a_{ii}^h a_{jj}^h}$ implies that e_i^h and e_j^h are approximately equal. Once a full set of aggregates have been determined, SA then uses basis functions defined in each aggregate to form linear combinations that approximate smooth error. The coefficients in these linear combinations constitute the coarse-grid unknowns. Often, local analysis can be used to guarantee that the WAP holds for a given proper choice of these basis functions, which would then guarantee that the method as constructed would hopefully provide an effective two-grid solver and possibly an effective W-cycle solver. But this *unsmoothed* version of SA is not likely to exhibit optimal V-cycle performance. This suggests that the next step would be to improve this *tentative* interpolation operator. The fundamental approach that SA takes is to apply a smoother to $\boldsymbol{P}_{2h}^h$, whence the moniker *smoothed aggregation*.

To understand how SA might construct interpolation based on the WAP, assume that the fine nodes have been partitioned into a set of aggregates, $\mathcal{A} = \{a\}$. For each aggregate, the aim now is to satisfy the *local* WAP: there exists a constant $C \ll \infty$ such that, for all fine-grid error $\boldsymbol{e}^h$ and every aggregate $a \in \mathcal{A}$, a coarse representative, $\boldsymbol{u}^{2h}$, exists that satisfies

$$\|\boldsymbol{e}^h - P_{2h}^h \boldsymbol{u}^{2h}\|_a^2 \leq \frac{C|a|}{\|\boldsymbol{L}^h\|} \langle \boldsymbol{L}^h \boldsymbol{e}^h, \boldsymbol{e}^h \rangle,$$

where $|\cdot|_a$ denotes the local Euclidean norm and $|a|$ is the relative size of a (so that $\sum_{a \in \mathcal{A}} |a| = 1$). Note that the sum of this local WAP over all $a \in \mathcal{A}$ yields the global WAP.

Now, it would be impractical to test the WAP to see if it holds for all $\boldsymbol{e}^h$, so the idea is instead to choose a set of vectors that hopefully represents all near-kernel components locally. As we noted above, focusing on such components is appropriate because those that are algebraically oscillatory automatically satisfy the WAP, just as they do the SAP. Fortunately, the local nature of near-kernel components are known for many problems. As examples, for scalar elliptic equations, they are approximately constant locally and, for linear elasticity, they look locally like *rigid body modes* (constant displacements and rigid rotations). In such cases, we can take these few global vectors, normalize them in energy, and then restrict them to each aggregate to determine an effective basis there. This can be done by finding a minimal local basis that adequately (according to the WAP) approximates all linear combinations of these restricted near-kernel components. The actual computation amounts to forming a local matrix, $\boldsymbol{T}$, whose columns consist of these restricted near-kernel components, solving the eigenproblem for $\boldsymbol{T}^t\boldsymbol{T}$, and then letting the basis be the resulting eigenvectors whose eigenvalues are less than $\frac{C|a|}{\|\boldsymbol{L}\|}$. (Remember that the near-kernel components were normalized in energy.) It is important to realize that SA computes a minimal basis, which not only controls complexity, but also ensures that redundancy does not creep into the coarse-grid problem. Any contamination by an exact or near linear dependence of a local basis would lead to artificial singularity or ill-conditioning of the coarse-grid matrix, making the development of the solver on coarser levels very problematic.

As noted above, to improve interpolation approximation properties, SA applies a smoother to the interpolation operator [14]. To understand the role of the smoothing process in SA, consider the one-dimensional example of a uniform grid of n points on the unit interval. If you consider three neighboring interior points and the vector function that is 1 at these points and 0 elsewhere, such a vector might make up a typical basis element for the coarse level that has about $\frac{n}{3}$ points. That is, you can define the coarse-level space

as the set of coefficients of these *piecewise-constant* basis elements. Piecewise-constant functions by themselves are generally unsuitable for *discretization* of elliptic equations. Analogously, piecewise-constant vectors offer only a low-order approximation to finer levels, so they are a generally poor *coarsening* option. But a simple averaging of the basis elements given by the stencil $(\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4})$ yields *piecewise-linear* basis elements, which provide higher-order approximation. Averaging of the basis elements is typically what a properly constructed smoother might do. This reasoning provides the motive for smoothing $\boldsymbol{P}_{2h}^{h}$ in an attempt to improve its approximation order. In other words, for the model 1D problem, piecewise constants satisfy the WAP but not the SAP; yet, a specific smoothing can result at least approximately in piecewise linears that do satisfy the SAP; and so the motivation is that smoothing interpolation might provide the approximation properties that we need to confirm good V-cycle convergence for other problems.

While the illustrations of this section have focused on smoothed aggregation, the message is much more general: development of an effective multigrid algorithm should pay special attention to the smooth or near-kernel errors if it is to be effective.

9.2. Problem formulation. As already noted in the previous section, there are similarities between multigrid and finite element methods. These similarities extend further in the sense that, just as the appropriate choice of finite elements can improve convergence of the discretization, so too does a good PDE formulation and discretization help convergence of the multigrid solver.

To be a little more specific, note that discretization can be viewed as an attempt to approximate an infinite-dimensional continuum by a finite-dimensional grid h space. This would only make sense if the target function could be pinned down somehow, which most discretizations do by relying on some sense of smoothness. For example, continuous piecewise-(bi)linear finite element spaces target solutions of elliptic partial differential equations by relying on boundedness of their second derivatives. This is a sensible goal when the corresponding operator exhibits full regularity because it means that the second derivatives of the solution are bounded by the source term. The success of discretization thus depends heavily on the behavior of the differential operator. Operators that exhibit reduced regularity, such as those with difficulties introduced by discontinuous coefficients or re-entrant corners, can certainly be treated effectively by discretization schemes, but they usually require special handling that carefully addresses these difficulties. Convergence of the exact solution of the discretized equations otherwise suffers.

So it is with multigrid. Effective coarse-grid correction depends heavily on the behavior of the matrix: if it comes from standard discretization of a fully regular elliptic differential operator, then standard coarse-grid correction schemes that mimic the discretization itself should be effective. In fact, multigrid aims to approximate the grid h space by the grid $2h$ space. This it does by pinning down the error using relaxation to obtain small relative residuals. For the fully regular case, small residuals mean that these errors vary slowly in any given neighborhood, that is, that the error is locally almost constant. It is this "discrete regularity" property that enables the use of standard coarse-grid correction schemes for full effectiveness. An important point here is that, while we expect good approximation by standard discretization of fully regular differential operators, we can also expect good overall performance of standard multigrid solvers applied to them.

On the other hand, it may be very challenging to develop effective multigrid methods for problems with discretizations that are less accurate or operators with reduced regularity. Special treatment in the multigrid approach may be necessary to handle these difficulties. In such cases, it may be necessary to pay attention to more than just the nature of the matrix problem. While it may not always be practical for a given application, it might otherwise be helpful to have some influence on the discretization and possibly the PDE formulation itself. For example, some applications use schemes that

provide stabilization, regularization, or penalty terms at the discretization or even PDE level to improve effectiveness of classical solvers. While these approaches may be helpful in that context, they can at times impede multigrid performance. Just a simple scaling of the matrix to provide a better condition number can make a straightforward multigrid solver stall because it changes the nature of the near kernel. Disabling such a scaling is sometimes all it take to restore efficiency. The point here is that multigrid development can be very difficult at times, but taking in the whole picture of the problem origin may provide important insight along the way. It has been said that the difficulty in design of effective multigrid methods is commensurate with the overall problem formulation. The corollary is perhaps that the development of an application in terms of formulating the PDE and its discretization should be done in concert with development of the multigrid method.

The discussion here has focussed primarily on elliptic PDEs, specifically linear second-order elliptic equations discretized by continuous piecewise linear or bilinear finite elements (or other traditional discretization methods with similar approximation properties). It is well known that multigrid *methods* have been very effective for solving problems in this model elliptic class. When regular grids are used, a geometric multigrid method using standard components (linear interpolation, full weighting, and simple relaxation schemes like Richardson iteration) is often able to solve such problems at optimal cost. When the problem begins to stray from the classical elliptic regime (as in the presence of strong convection, strong anisotropies, or jump-discontinuous coefficients) or when irregular grids are used, standard algebraic multigrid methods that do not necessarily have any knowledge of the matrix origin might be an option for automatic design of a multigrid algorithm.

What is less well known is that the multigrid *methodology* has much broader applicability than this model elliptic class. For example, multigrid has been applied successfully to a wide ranges of problems since its practical origin in Brandt's seminal paper [3]. In fact, multilevel methods have been extended well beyond the elliptic case to Navier-Stokes equations and other systems in fluid flow, as well as many problems in structural mechanics (c.f., Brandt [4]). The multilevel methodology has also been extended successfully to previously intractable problems that are beyond the realm of differential equations (e.g., geodetic equations [2], quantum chromodynamics [5], and Markov chains [9]). However, these applications require special multigrid treatment that can lead to algorithms that are much more sophisticated than the standard multigrid methods in more common use.

Our intention in this basic primer is not to delve further into the advanced topics mentioned in this or the previous sections, but rather to provide a glimpse of what may lie ahead in a deeper study of multigrid methodology.

10. Summary. With linear elliptic PDEs in mind, to achieve optimal multigrid performance, the following properties might be considered:

- The relaxation process should effectively reduce oscillatory components of the error. More precisely, the errors that relaxation cannot quickly eliminate should provide a pattern that can be effectively exploited in the local construction of interpolation.
- There should be some sense of locality, such as that afforded by the WAP, since that is what makes it possible and practical to determine an adequate representation of the near-kernel components.
- Coarse spaces should be constructed in neighborhoods of strongly interconnected points and they should carefully avoid (near) redundancy. The aim is that strong connections would guarantee some sense of a smoothing property of relaxation and a lack of any serious redundancy would avoid artificial ill-conditioning.
- For Krylov smoothers like Richardson's iteration, the range of the interpolation

operator (from the coarse grid to the fine grid) should represent near-kernel components of the fine-grid operator accurately according to the SAP. More generally, interpolation should approximate components that relaxation cannot adequately attenuate.

- For Krylov smoothers, the discretization of the PDE that that creates the matrix problems should be accurate enough to satisfy the continuum version of the SAP. When practical, attention to the PDE and its discretization may be advisable.
- With the proper design of multigrid as an efficient iterative solver and of the discretization as an accurate approximation to the PDE, FMG should be effective for delivering accuracy at the level of the discretization itself.

REFERENCES

[1] James H Bramble, Richard E Ewing, Joseph E Pasciak, and Jian Shen. The analysis of multigrid algorithms for cell centered finite difference methods. *Advances in Computational Mathematics*, 5(1):15–29, 1996. doi:10.1007/BF02124733.

[2] A. Brandt, S. McCormick, and J. Ruge. Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations. Technical report, Institute for Computational Studies, Colorado State University, POB 1852, Fort Collins, Colorado, 1982.

[3] Achi Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333–390, 1977. doi:10.1090/S0025-5718-1977-0431719-X.

[4] Achi Brandt and Oren E. Livne. *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics, Revised Edition*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2011. doi:10.1137/1.9781611970753.

[5] James Brannick, Marian Brezina, David Keyes, Oren Livne, Irene Livshits, Scott MacLachlan, Tom Manteuffel, Steve McCormick, John Ruge, and Ludmil Zikatanov. Adaptive Smoothed Aggregation in Lattice QCD. In Olof B. Widlund and David E. Keyes, editors, *Domain Decomposition Methods in Science and Engineering XVI*, volume 55 of *Lecture Notes in Computational Science and Engineering*, pages 505–512. Springer Berlin Heidelberg, 2007. doi:10.1007/978-3-540-34469-8_63.

[6] W. Briggs, V. Henson, and S. McCormick. *A Multigrid Tutorial, Second Edition*. Society for Industrial and Applied Mathematics, 2000. doi:10.1137/1.9780898719505.

[7] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1983.

[8] Stephen McCormick. Multigrid Methods for Variational Problems: Further Results. *SIAM Journal on Numerical Analysis*, 21(2):255–263, 1984. doi:10.1137/0721018.

[9] H. De Sterck, Thomas A. Manteuffel, Stephen F. McCormick, Quoc Nguyen, and John Ruge. Multilevel Adaptive Aggregation for Markov Chains, with Application to Web Ranking. *SIAM Journal on Scientific Computing*, 30(5):2235–2262, 2008. doi:10.1137/070685142.

[10] G. Strang and G. J. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, Englewood Cliffs, NJ, 1973.

[11] Rasmus Tamstorf, Toby Jones, and Stephen F. McCormick. Smoothed aggregation multigrid for cloth simulation. *ACM Transactions on Graphics*, 34(6), Oct 2015. doi:10.1145/2816795.2818081.

[12] Ulrich Trottenberg, Cornelius W Oosterlee, and Anton Schuller. *Multigrid*. Academic Press, 2000.

[13] P. Vaněk, J. Mandel, and M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56(3):179–196, 1996. doi:10.1007/BF02238511.

[14] Petr Vaněk. Acceleration of convergence of a two-level algorithm by smoothing transfer operator. *Applications of Mathematics*, 37:265–274, 1992. URL: http://dml.cz/dmlcz/104515.

A. Approximation properties. Following is a list of three approximation properties that an interpolation operator $\boldsymbol{P} \in \mathbb{R}^{n\times n_c}$ might posses in the context of approximating the solution of an equation in the SPD matrix $\boldsymbol{L} \in \mathbb{R}^{n\times n}$. (For simplicity here, n and n_c denote the respective fine and coarse grid mesh sizes, and the matrices and vectors are used without superscripts or subscripts.)

WAP$_{\boldsymbol{L}^2}$: there exists a coarse-grid vector $\boldsymbol{v}$ such that $\|\boldsymbol{u} - \boldsymbol{P}\boldsymbol{v}\|^2 \leq \frac{C}{\|\boldsymbol{L}\|^2}\|\boldsymbol{L}\boldsymbol{u}\|^2$.

WAP$_{\boldsymbol{L}}$: there exists a coarse-grid vector $\boldsymbol{v}$ such that $\|\boldsymbol{u} - \boldsymbol{P}\boldsymbol{v}\|^2 \leq \frac{C}{\|\boldsymbol{L}\|}\|\boldsymbol{u}\|_{\boldsymbol{L}}^2$.

SAP$_{\boldsymbol{L}}$: there exists a coarse-grid vector $\boldsymbol{v}$ such that $\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}}^2 \leq \frac{C}{\|\boldsymbol{L}\|}\|\boldsymbol{L}\boldsymbol{u}\|^2$.

The following results show that the WAP for $\boldsymbol{L}^2$ (WAP$_{\boldsymbol{L}^2}$) implies the WAP for $\boldsymbol{L}$ (WAP$_{\boldsymbol{L}}$) and is *equivalent* to the SAP for $\boldsymbol{L}$ (SAP$_{\boldsymbol{L}}$). We first show for completeness of proof that SAP$_{\boldsymbol{L}}$ implies WAP$_{\boldsymbol{L}}$. These properties are assumed to hold for any fine-grid vector $\boldsymbol{u}$ with C a fixed constant that may vary in meaning with each occurrence.

- SAP$_{\boldsymbol{L}}$ $\implies$ WAP$_{\boldsymbol{L}}$: Let $\boldsymbol{P}\boldsymbol{v}$ be the best approximation to $\boldsymbol{u}$ in energy and let $\boldsymbol{P}\boldsymbol{z}$ be the best approximation to $\boldsymbol{L}^{-1}(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})$ in energy. Then, because $\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}$ is energy orthogonal to the range of $\boldsymbol{P}$, we have that

$$\begin{aligned}\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|^2 &= \langle \boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}, \boldsymbol{L}^{-1}(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})\rangle_{\boldsymbol{L}} \\ &= \langle \boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}, \boldsymbol{L}^{-1}(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})-\boldsymbol{P}\boldsymbol{z}\rangle_{\boldsymbol{L}} \\ &\leq \|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}} \cdot \|\boldsymbol{L}^{-1}(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})-\boldsymbol{P}\boldsymbol{z}\|_{\boldsymbol{L}}.\end{aligned}$$

  The SAP$_{\boldsymbol{L}}$ for $\boldsymbol{L}^{-1}(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})$ then implies that

$$\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|^2 \leq \|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}} \cdot \sqrt{\frac{C}{\|\boldsymbol{L}\|}}\|\boldsymbol{L}\boldsymbol{L}^{-1}(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})\| = \|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}} \cdot \sqrt{\frac{C}{\|\boldsymbol{L}\|}}\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|.$$

  Dividing both sides by $\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|$, squaring, and applying the SAP$_{\boldsymbol{L}}$ for $\boldsymbol{u}$ yields the WAP$_{\boldsymbol{L}}$ (note the C^2):

$$\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|^2 \leq \frac{C}{\|\boldsymbol{L}\|}\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}}^2 \leq \frac{C^2}{\|\boldsymbol{L}\|^2}\|\boldsymbol{L}\boldsymbol{u}\|^2 \leq \frac{C^2}{\|\boldsymbol{L}\|}\|\boldsymbol{u}\|_{\boldsymbol{L}}^2.$$

- WAP$_{\boldsymbol{L}^2}$ $\implies$ WAP$_{\boldsymbol{L}}$: For the $\boldsymbol{v}$ guaranteed by WAP$_{\boldsymbol{L}^2}$, we have that

$$\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|^2 \leq \frac{C}{\|\boldsymbol{L}\|^2}\|\boldsymbol{L}\boldsymbol{u}\|^2 \leq \frac{C}{\|\boldsymbol{L}\|}\|\boldsymbol{u}\|_{\boldsymbol{L}}^2.$$

- WAP$_{\boldsymbol{L}^2}$ $\implies$ SAP$_{\boldsymbol{L}}$: For the $\boldsymbol{v}$ guaranteed by WAP$_{\boldsymbol{L}^2}$, we have that

$$\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}}^2 \leq \|\boldsymbol{L}\| \cdot \|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|^2 \leq \frac{C}{\|\boldsymbol{L}\|}\|\boldsymbol{L}\boldsymbol{u}\|^2.$$

- SAP$_{\boldsymbol{L}}$ $\implies$ WAP$_{\boldsymbol{L}^2}$: For the $\boldsymbol{v}$ guaranteed by SAP$_{\boldsymbol{L}}$, we have that

$$\frac{C}{\|\boldsymbol{L}\|^2}\|\boldsymbol{L}\boldsymbol{u}\|^2 \geq \frac{\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}}^2}{\|\boldsymbol{L}\|}.$$

  But the WAP$_{\boldsymbol{L}}$, which is implied by the SAP$_{\boldsymbol{L}}$, in turn implies that there exists an $\boldsymbol{u}$ for $\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}$ such that

$$\frac{\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|_{\boldsymbol{L}}^2}{\|\boldsymbol{L}\|} \geq \frac{1}{C}\|(\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v})-\boldsymbol{P}\boldsymbol{u}\|^2 = \frac{1}{C}\|\boldsymbol{u}-\boldsymbol{P}(\boldsymbol{v}+\boldsymbol{u})\|^2.$$

  Putting these together, with $\boldsymbol{v} \equiv \boldsymbol{v}+\boldsymbol{u}$, implies the WAP$_{\boldsymbol{L}^2}$ with C^2 instead of C: $\|\boldsymbol{u}-\boldsymbol{P}\boldsymbol{v}\|^2 \leq \frac{C^2}{\|\boldsymbol{L}\|^2}\|\boldsymbol{L}\boldsymbol{u}\|^2$.