NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2016; **23**:746–771
Published online 14 June 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nla.2053

A two-grid SA-AMG convergence bound that improves when increasing the polynomial degree

Xiaozhe Hu[1,*,†], Panayot S. Vassilevski[2] and Jinchao Xu[3]

[1]*Department of Mathematics, Tufts University, Medford, MA 02155, USA*
[2]*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA*
[3]*Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA*

SUMMARY

In this paper, we consider the convergence rate of a smoothed aggregation algebraic multigrid method, which uses a simple polynomial $(1-t)^{\nu}$ or an optimal Chebyshev-like polynomial to construct the smoother and prolongation operator. The result is purely algebraic, whereas a required main weak approximation property of the tentative interpolation operator is verified for a spectral element agglomeration version of the method. More specifically, we prove that, for partial differential equations (PDEs), the two-grid method converges uniformly without any regularity assumptions. Moreover, the convergence rate improves uniformly when the degree of the polynomials used for the smoother and the prolongation increases. Such a result, as is well-known, would imply uniform convergence of the multilevel W-cycle version of the algorithm. Numerical results, for both PDE and non-PDE (graph Laplacian) problems are presented to illustrate the theoretical findings. Published 2016. This article is a U.S. Government work and is in the public domain in the USA.

Received 19 September 2014; Revised 27 April 2016; Accepted 27 April 2016

KEY WORDS: algebraic multigrid (AMG); smoothed aggregation; polynomial smoothers; smoothing property

1. INTRODUCTION

In this paper, we study two-grid iterative methods for solving large-scale linear system of equations. We design a two-grid method in the framework of the smoothed aggregation (SA) spectral element agglomeration algebraic multigrid (AMG) method. Our aim is to prove that the studied method converges uniformly without any regularity assumption in a way that the convergence rate improves when we increase the degree of the polynomial. In principle, based on such two-grid algorithm, as it is well known, we can construct relatively simpler multilevel cycles (such as combined V/W-cycle) instead of the more involved (linear and nonlinear) algebraic multilevel iteration (AMLI) cycles with guaranteed uniform convergence.

There are many, including classical, convergence results for geometric and algebraic two-level algorithms. For example, more recent ones include [1, 2], and [3], which deal with the construction of the coarse spaces suitable for the domain decomposition method and the SA-AMG methods, respectively. Similar spectral construction of the coarse space goes back to the element-based algebraic multigrid (AMGe) in [4] and its modified version suited for multilevel extension in [5] and [6]. A general analysis for two-grid AMG method can be found in [7] where (non-sharp) theory was derived and in [8] where the exact convergence factor was characterized. In [9], two-sided bounds on the convergence of two-level methods were obtained. Earlier results on two-level hierarchical basis type methods are found in [10]. We refer interested readers to [11] for more details on both two-level and two-grid methods.

*Correspondence to: Xiaozhe Hu, Department of Mathematics, Tufts University, Medford, MA 02155, USA.
†E-mail: Xiaozhe.Hu@tufts.edu

Although uniform AMG convergence results have been established in the aforementioned references, the relationship between the convergence rate and the number of smoothing steps was beyond the available regularity-free convergence theory. For geometric multigrid method, it is well known that the convergence rate improves uniformly when we increase the number of smoothing steps (e.g. [12, 13]), which is possible by exploiting some regularity of the underlying PDEs. Until now, it was an open question if one can develop an AMG method that has similar properties, that is, with provable convergence rate that improves when the number of smoothing steps increases. In this paper, we develop such a two-grid AMG method based on the smoothed aggregation AMG method. Our theory is general and purely algebraic. The only assumption is that a tentative prolongator, which exhibits a weak approximation property, is available. The construction of the actual prolongator and the smoother are based on appropriate matrix polynomials. To construct a tentative prolongator with the desired weak approximation property, we resort to the spectral element agglomeration AMG method in the form proposed in [3], which is also extendable for graph Laplacian problems. Following the approach in [3], we construct a polynomial smoother, which satisfies certain smoothing property to be detailed later on. The tentative prolongator is constructed based on solving local generalized eigenvalue problems on each subdomains, agglomerates of fine-grid elements (or fine edges of the underlined graph in the graph Laplacian case). To improve the energy stability of the coarse basis, the polynomial smoothing idea from the SA-AMG method is applied and a simple polynomial $(1-t)^{\nu}$ or an optimal Chebyshev-like polynomial are used. Because we need the polynomial to be non-negative, we use the square of the original SA-polynomial. We show that by increasing the degree of the polynomial used to smooth the tentative prolongation, we can make the A-norm of the smoothed prolongation as close to unity as we want. We also demonstrate that the constructed smoothed prolongator still maintains a weak approximation property. Again, no regularity assumption is imposed on the original problem (in the PDE case). Although the constant in the weak approximation property estimate deteriorates, this can be compensated for by increasing the degree of the polynomial smoother (equivalent to increasing the number of smoothing steps). As a result, we are able to show that the overall convergence rate of the two-grid algorithm is uniform and improves uniformly when we increase the degree of the polynomials involved.

The remainder of the paper is structured as follows. In Section 2, we briefly review the two-grid method and introduce the main Chebyshev-like polynomial used in the SA-AMG. The smoothed aggregation spectral element agglomeration AMG method is recalled in Section 3. A main smoothing property and the weak approximation property are discussed there as well. In Section 4, we derive the two-grid convergence estimate and show that it improves when we increase the degree of the polynomials. Finally, in Section 5, we present some numerical experiments that illustrate our theoretical results.

2. PRELIMINARIES

We consider the system of linear equations

$$A\mathbf{u} = \mathbf{f}, \tag{2.1}$$

where $A \in \mathbb{R}^{N\times N}$ is a symmetric positive definite (SPD) matrix that comes from linear finite element discretization of the model second-order elliptic PDE

$$-\nabla \cdot a(x)\nabla u = f \tag{2.2}$$

on a polygonal domain Ω with suitable boundary conditions. The diffusion coefficient $a(x)$ satisfies $0 < a_{\min} \leqslant a(x) \leqslant a_{\max}$ on Ω and the right-hand side f is in $L_2(\Omega)$. We want to remark that A could come from the discretization of other types of second-order elliptic PDEs on the triangulated domain or graph-related problems. The algorithms and analysis developed later can be applied to other choice of A. In general, we assume that A is a sparse, that is, it has overall $\mathcal{O}(N)$ non-zero entries so that iterative methods are appropriate to apply.

Here, we focus on solving (2.1) by an (algebraic) multigrid method. More precisely, we consider a two-grid method and develop a special algorithm whose convergence rate improves when we increase the smoothing steps (for both smoother and the smoothed prolongator). We stress upon the fact that we do not rely on any regularity assumptions on the original problem (in the PDE case).

2.1. *The two-grid method*

First, we briefly recall the standard two-grid method. Assume we have a smoother M such that both the actions of M^{-1} and M^{-T} are computable (in $\mathcal{O}(N)$ operations), an interpolation operator (prolongator) P, the respective restriction $R = P^T$, and the coarse-grid problem $A_c = P^T AP$ are available. The specific construction of the smoother and prolongation will be introduced later. Based on these standard components, we define the standard (symmetrized) two-grid method in Algorithm 1.

Algorithm 1 Two-grid method

For a current iterate $\mathbf{u}$, we perform:

1: Presmoothing: $\mathbf{u} \leftarrow \mathbf{u} + M^{-1}(\mathbf{f} - A\mathbf{u})$
2: Restriction: $\mathbf{r}_c \leftarrow P^T(\mathbf{f} - A\mathbf{u})$
3: Coarse-grid correction: $\mathbf{e}_c \leftarrow A_c^{-1}\mathbf{r}_c$
4: Prolongation: $\mathbf{u} \leftarrow \mathbf{u} + P\mathbf{e}_c$
5: Postsmoothing: $\mathbf{u} \leftarrow \mathbf{u} + M^{-T}(\mathbf{f} - A\mathbf{u})$

It is well known that the two-grid method (Algorithm 1) leads to the composite iteration matrix E_{TG} based on which we define the two-grid operator B_{TG}

$$I - B_{TG}^{-1}A = E_{TG} = \left(I - M^{-T}A\right)\left(I - PA_c^{-1}P^TA\right)\left(I - M^{-1}A\right).$$

The actions of B_{TG}^{-1} are computable using Algorithm 1 starting with initial iterate $\mathbf{u} = 0$. Then $B_{TG}^{-1}\mathbf{f}$ equals the final $\mathbf{u}$. More explicitly, we have

$$B_{TG}^{-1} = \overline{M}^{-1} + \left(I - M^{-T}A\right)PA_c^{-1}P^T\left(I - AM^{-1}\right).$$

From the preceding, $\overline{M}$ is the symmetrized smoother given by the expression $M(M^T + M - A)^{-1}M^T$.

In general, B_{TG} provides an approximate inverse (or preconditioner) to A. If the smoother M is convergent in A-norm, that is, $\|I - M^{-1}A\|_A < 1$, or equivalently $M + M^T - A$ is SPD, then B_{TG}^{-1} is SPD and provides a convergent iterative method for A, that is, we have

$$(\mathbf{v}, \mathbf{v})_A \leqslant (B_{TG}\mathbf{v},\ \mathbf{v}).$$

For the convergence rate of the two-grid method, we have the following two-grid estimates, which can be found in Theorem 4.3, [8]. As shown in [9], such results can be obtained from a two-level version of the XZ-identity [14].

Theorem 2.1

For B_{TG} and the two-grid error propagation operator E_{TG}, we have the sharp estimates

$$(B_{TG}\mathbf{v}, \mathbf{v}) \leqslant K_{TG}(\mathbf{v}, \mathbf{v})_A \quad \text{or equivalently} \quad \|E_{TG}\|_A = \rho_{TG} := 1 - \frac{1}{K_{TG}},$$

where

$$K_{TG} = \max_{\mathbf{v}} \min_{\mathbf{v}_c} \frac{\|\mathbf{v} - P\mathbf{v}_c\|_{\overline{M}}^2}{\|\mathbf{v}\|_A^2}.$$

We can see that different choices of the smoother M and prolongator P define different two-grid methods. Our goal is to construct proper smoother M and prolongator P such that the constant K_{TG} is uniformly bounded. Moreover, the goal is to be able to make K_{TG} as close to unity as we want when we not only increase the number of smoothing steps but also improve the energy stability of the prolongator by using special polynomials for both, the smoother and prolongator. Then, by increasing the degree of the polynomials, we achieve the decreasing of the constant K_{TG}.

2.2. The smoothed aggregation algebraic multigrid optimal Chebyshev-like polynomial

In this section, we review the polynomial used in the SA-AMG (e.g., [15]). It is constructed on the basis of the classical Chebyshev polynomials. This polynomial possesses certain optimal properties, which we review next (details are found, e.g., in [11]). We will use this polynomial to construct the smoother M and prolongator P in one of the cases. We would like to mention that it is also possible to use other Chebyshev-type polynomials in our method, for example, the polynomial of best uniform approximation to $1/x$ studied in [16], as long as certain smoothing property and approximating property can be established (Section 3). However, to illustrate our approach and for the sake of simplicity, we only use the SA-AMG optimal Chebyshev-like polynomial in this paper.

We use the classical Chebyshev polynomials $T_k(t)$ defined by $T_0(t) = 1$, $T_1(t) = t$, and $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$, $k \geqslant 1$. Letting $t = \cos\alpha$, we have $T_k(t) = \cos k\alpha$. Given the Chebyshev polynomials $T_k(t)$, for any $b > 0$, we construct a polynomial $\varphi_\nu(t)$ for $t \in [0,b]$ as follows

$$\varphi_\nu(t) = (-1)^\nu \frac{1}{2\nu+1}\frac{\sqrt{b}}{\sqrt{t}} T_{2\nu+1}\left(\frac{\sqrt{t}}{\sqrt{b}}\right). \tag{2.3}$$

One can actually show that $\varphi_\nu(t)$ is a polynomial of degree ν such that $\varphi_\nu(0) = 1$. In addition, we have the following properties of $\varphi_\nu(t)$.

Proposition 2.2

(1) The polynomial $\varphi_\nu(t)$ has the following optimal property

$$\min_{p_\nu(0)=1} \max_{t\in[0,b]} \left|\sqrt{t}\, p_\nu(t)\right| = \max_{t\in[0,b]} \left|\sqrt{t}\varphi_\nu(t)\right| = \frac{\sqrt{b}}{2\nu+1}. \tag{2.4}$$

(2) $\varphi_\nu(0) = 1$ and

$$\max_{t\in[0,b]} |\varphi_\nu(t)| = 1. \tag{2.5}$$

(3) There is a constant C_ν independent of b such that

$$\sup_{t\in(0,b)} \frac{|1-\varphi_\nu(t)|}{\sqrt{t}} \leqslant C_\nu \frac{1}{b^{1/2}}, \tag{2.6}$$

where

$$\frac{C_\nu}{2\nu+1} \leqslant 2. \tag{2.7}$$

The proofs of (2.4) and (2.5) are found in [11], whereas (2.7) is proved in [15].

We also define a normalized version of $\varphi_\nu(t)$ for $t \in [0,1]$,

$$s_\nu(t) := (-1)^\nu \frac{1}{2\nu+1} \frac{T_{2\nu+1}\left(\sqrt{t}\right)}{\sqrt{t}}. \tag{2.8}$$

It has the following properties:

DOI: 10.1002/nla

Proposition 2.3

(1) $s_\nu^2(t) \in [0, 1]$ for $t \in (0, 1]$,
(2) $\max_{t\in(0,1]} \left|\sqrt{t} s_\nu(t)\right| \leqslant \frac{1}{2\nu+1}$.

3. SMOOTHED AGGREGATION SPECTRAL ELEMENT AGGLOMERATION ALGEBRAIC MULTIGRID

The two-grid method in this paper can be viewed as a version of the smoothed aggregation spectral element agglomeration AMG (SA-ρAMG) method. The construction of the smoother inverse, M^{-1}, and interpolation operator, P, is based on the SA-ρAMG method in the form described in [3]. Without any regularity assumption, the proposed two-grid method's convergence rate decreases when the number of smoothing steps (the degree of the polynomials) increases. Such property allows for cheaper multilevel extension of the algorithm. More precisely, large two-grid convergence rate can be compensated for by more recursive calls on the coarse levels when we consider multilevel cycles, such as n-fold V-cycle with $n > 1$ or (linear or nonlinear) AMLI-cycle with large polynomial degree. This may result in multilevel algorithms that are not optimal in terms of the computational cost. However, if we can make the two-level convergence factor as small as we want by increasing the number of smoothing steps, then we can in principle use W-cycle or even V-cycle method in the multilevel case, which is less expensive computationally and also more suitable for parallel computing, without deteriorating the overall convergence too much. Note that we need that the two-grid convergence factor be reduced independently of the levels, that is, it should be decreased uniformly on each level. Therefore, although we can observe in practice a decrease in the two-grid convergence rate when we increase the number of smoothing steps, such uniform behavior of this property is nontrivial to achieve and also not easy to prove theoretically.

3.1. *Construction of the smoother*

Let D be the diagonal or diagonal block of A (in general, D could be an SPD matrix with easy to compute inverse action and sparse inverse). We are interested in the smoother of the following form

$$M^{-1} = \left(I - p_{\nu_r}(b^{-1}D^{-1}A)\right) A^{-1},$$

for a polynomial $p_{\nu_r}(t)$ of degree $\nu_r \geqslant 1$ such that $p_{\nu_r}(0) = 1$ and $p_{\nu_r}(t) > 0$ for $t \in (0, 1]$. Here, $b : \left\| D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right\| \leqslant b = \mathcal{O}(1)$. We consider the simple choice $p_{\nu_r}(t) = (1-t)^{\nu_r}$ and also the Chebyshev-like polynomial. Note that the simple polynomial case includes traditional smoothers such as Richardson smoother, Jacobi smoother, ℓ-1 smoother, and symmetric Gauss-Seidel smoother when we choose D and b properly.

3.1.1. Simple smoother. First, we consider the simple choice $p_{\nu_r}(t) = (1-t)^{\nu_r}$. Traditional smoothers can usually be presented in the form by choosing different b and D. For example, $D = I$ and $b = \lambda_{\max}(A)$ gives the Richardson smoother, whereas letting $D = \operatorname{diag}(A)$ with appropriate weight b gives the weighted Jacobi smoother. Based on the fact that M is symmetric, we have

$$\overline{M}^{-1} = \left(I - p_{\nu_r}^2\left(b^{-1}D^{-1}A\right)\right) A^{-1} = \left(I - \left(I - b^{-1}D^{-1}A\right)^{2\nu_r}\right) A^{-1}. \tag{3.1}$$

The next lemma provides a smoothing property of this simple case, which is in the form of the smoothing property originally proposed in [15] (Lemma 6.1, estimate (43)).

Lemma 3.1
For any given parameter $q \in (0, 1)$ and $\overline{M}$ defined by (3.1), the estimate

$$\|\mathbf{v}\|_{\overline{M}}^2 \leqslant \frac{1}{1-q^2}\|\mathbf{v}\|_A^2 + \frac{1}{q^2}\eta_S(\nu_r)\|\mathbf{v}\|_D^2, \quad \forall \mathbf{v} \in \mathbb{R}^N$$

holds uniformly with $\eta_S(\nu_r) \simeq \frac{1}{2\nu_r}$.

Proof
We consider

$$\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \text{ and } \tilde{M} = D^{-\frac{1}{2}} \overline{M} D^{-\frac{1}{2}}. \tag{3.2}$$

Note that

$$\begin{aligned} \tilde{M}^{-1} = D^{\frac{1}{2}} \overline{M}^{-1} D^{\frac{1}{2}} &= \left(I - \left(I - b^{-1} D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right)^{2\nu_r} \right) \\ D^{\frac{1}{2}} A^{-1} D^{\frac{1}{2}} &= \left(I - \left(I - b^{-1} \tilde{A} \right)^{2\nu_r} \right) \tilde{A}^{-1}. \end{aligned}$$

Therefore, $\tilde{A}$ and $\tilde{M}$ are both symmetric, mutually commute, and have common eigenvectors. We partition the vector space $\mathbb{R}^N$ using the eigenpairs $(\lambda_i, \mathbf{v}_i)$ of $b^{-1}\tilde{A}$ as follows

$$U_1 = \left\{ \text{span}\{\mathbf{v}_i\} : \lambda_i \leqslant \frac{\theta}{\nu_r} < 1 \right\} \quad \text{and} \quad U_2 = \left\{ \text{span}\{\mathbf{v}_i\} : \lambda_i > \frac{\theta}{\nu_r} > 0 \right\}.$$

We will determine $\theta \in (0, \nu_r)$ later. Note that U_1 and U_2 are invariant subspaces for both $\tilde{A}$ and $\tilde{M}$.

For $\mathbf{w} \in U_1$, we have

$$\begin{aligned} \|\mathbf{w}\|^2_{\tilde{M}^{-1}} &= \left(\left(I - \left(I - b^{-1} \tilde{A} \right)^{2\nu_r} \right) \tilde{A}^{-1} \mathbf{w}, \mathbf{w} \right) \\ &\geqslant \min_{t \in \left(0, \frac{\theta}{\nu_r}\right)} \frac{1 - (1-t)^{2\nu_r}}{t} b^{-1} \|\mathbf{w}\|^2 \\ &= \min_{t \in \left(0, \frac{\theta}{\nu_r}\right)} \left(1 + (1-t) + \cdots + (1-t)^{2\nu_r - 1} \right) b^{-1} \|\mathbf{w}\|^2 \\ &\geqslant 2\nu_r \left(1 - \frac{\theta}{\nu_r} \right)^{2\nu_r - 1} b^{-1} \|\mathbf{w}\|^2. \end{aligned}$$

Therefore, we have

$$\|\mathbf{w}\|^2_{\tilde{M}} \leqslant \frac{1}{2\nu_r} \left(1 - \frac{\theta}{\nu_r} \right)^{-2\nu_r + 1} b \|\mathbf{w}\|.$$

For $\mathbf{w} \in U_2$, we have

$$\begin{aligned} \|\mathbf{w}\|^2_{\tilde{M}^{-1}} &= \left(\left(I - \left(I - b^{-1} \tilde{A} \right)^{2\nu_r} \right) \tilde{A}^{-1} \mathbf{w}, \mathbf{w} \right) \\ &= \left(\left(I - \left(I - b^{-1} \tilde{A} \right)^{2\nu_r} \right) \tilde{A}^{-\frac{1}{2}} \mathbf{w}, \tilde{A}^{-\frac{1}{2}} \mathbf{w} \right) \\ &\geqslant \min_{t \in \left(\frac{\theta}{\nu_r}, 1\right)} \left(1 - (1-t)^{2\nu_r} \right) \|\mathbf{w}\|^2_{\tilde{A}^{-1}} \\ &= \left(1 - \left(1 - \frac{\theta}{\nu_r} \right)^{2\nu_r} \right) \|\mathbf{w}\|^2_{\tilde{A}^{-1}}. \end{aligned}$$

Therefore,

$$\|\mathbf{w}\|^2_{\tilde{M}} \leqslant \frac{1}{1 - \left(1 - \frac{\theta}{\nu_r} \right)^{2\nu_r}} \|\mathbf{w}\|^2_{\tilde{A}}.$$

For any $\mathbf{w} \in \mathbb{R}^N$, we can use the ℓ_2- and $\tilde{M}$-orthogonal decomposition $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$ with $\mathbf{w}_1 \in U_1$ and $\mathbf{w}_2 \in U_2$ and have

$$\begin{aligned}
\|\mathbf{w}\|_{\tilde{M}}^2 &= \|\mathbf{w}_1 + \mathbf{w}_2\|_{\tilde{M}}^2 = \|\mathbf{w}_2\|_{\tilde{M}}^2 + \|\mathbf{w}_1\|_{\tilde{M}}^2 \\
&\leqslant \frac{1}{1 - \left(1 - \frac{\theta}{\nu_r}\right)^{2\nu_r}} \|\mathbf{w}_2\|_{\tilde{A}}^2 + \frac{\left(1 - \frac{\theta}{\nu_r}\right)^{-2\nu_r+1}}{2\nu_r} b\|\mathbf{w}_1\|^2 \\
&\leqslant \frac{1}{1 - \left(1 - \frac{\theta}{\nu_r}\right)^{2\nu_r}} \|\mathbf{w}\|_{\tilde{A}}^2 + \frac{\left(1 - \frac{\theta}{\nu_r}\right)^{-2\nu_r}}{2\nu_r} b\|\mathbf{w}\|^2 \\
&= \frac{1}{1-q^2}\|\mathbf{w}\|_{\tilde{A}}^2 + \frac{1}{q^2}\eta_S(\nu_r)\|\mathbf{w}\|^2,
\end{aligned}$$

for any $q \in (0,1)$ by choosing $\theta = \nu_r\left(1 - q^{\frac{1}{\nu_r}}\right) \in (0, \nu_r)$. Now, let $\mathbf{v} = D^{-\frac{1}{2}}\mathbf{w}$. Using the relations (3.2), we obtain

$$\|\mathbf{w}\|_{\tilde{M}}^2 = \|\mathbf{v}\|_{\overline{M}}^2, \quad \|\mathbf{w}\|_{\tilde{A}}^2 = \|\mathbf{v}\|_A^2, \quad \text{and } \|\mathbf{w}\|^2 = \|\mathbf{v}\|_D^2,$$

which completes the proof. □

Of our interest is to choose $q \in (0,1)$ small enough such that for a given tolerance $\tau \in (0,1)$, we can let $1 + \tau = \frac{1}{1-q^2}$, that is $\tau = \frac{q^2}{1-q^2} \in (0,1)$. Then, we have the following estimate to be used later on.

$$\|\mathbf{v}\|_{\overline{M}}^2 \leqslant (1+\tau)\left[\|\mathbf{v}\|_A^2 + \frac{1}{\tau}\,\eta_S(\nu_r)\,\|\mathbf{v}\|_D^2\right], \quad \forall\, \mathbf{v} \in \mathbb{R}^N. \tag{3.3}$$

Here, $\eta_S(\nu_r) \simeq \frac{1}{2\nu_r}$. Note that the second term from the preceding equation deteriorates with $\tau \mapsto 0$, which, however, can be controlled if $\nu_r \mapsto \infty$. We refer sometimes to this estimate as a '*smoothing property*' (which is different from the classical one due to W. Hackbusch, [12]).

3.1.2. Chebyshev-like polynomial smoother. From the proof of Lemma 3.1, we can see that the smoothing property is a result from an estimate of the minimum of a polynomial over a given interval. This motivates us to consider Chebyshev-like polynomials because of their minimization properties. Next, we discuss the Chebyshev-like polynomial smoother and its smoothing property [15]. More specifically, based on the polynomial introduced in Section 2.2, we consider the following polynomial:

$$p_{\nu_r}(t) = \left(1 - T_{2\nu_r+1}^2\left(\sqrt{t}\right)\right) s_{\nu_r}(t). \tag{3.4}$$

The following smoothing property of $p_{\nu_r}(t)$ has been shown in [15] (Lemma 6.1, estimate (43)) and the proof is very similar to the proof of Lemma 3.1. Therefore, it is omitted.

Lemma 3.2
For any given parameter $q \in (0,1)$ and $\overline{M}$ is defined by (3.4), the estimate

$$\|\mathbf{v}\|_{\overline{M}}^2 \leqslant \frac{1}{1-q^2}\,\|\mathbf{v}\|_A^2 + \frac{1}{q^2}\,\eta_S(\nu_r)\,\|\mathbf{v}\|_D^2, \quad \forall\, \mathbf{v} \in \mathbb{R}^N,$$

holds uniformly with $\eta_S(\nu_r) \simeq \frac{1}{(2\nu_r+1)^2}$.

Again, we can let $1+\tau = \frac{1}{1-q^2}$, that is $\tau = \frac{q^2}{1-q^2} \in (0, 1)$. Then, we have the following estimate

$$\|\mathbf{v}\|_{\overline{M}}^2 \leqslant (1+\tau)\left[\|\mathbf{v}\|_A^2 + \frac{1}{\tau}\,\eta_S(\nu_r)\,\|\mathbf{v}\|_D^2\right], \quad \forall\,\mathbf{v} \in \mathbb{R}^N. \tag{3.5}$$

Compare with the smoothing property for the simple polynomial case; here, $\eta_S(\nu_r) \simeq \frac{1}{(2\nu_r+1)^2}$, which decays much faster than the simple polynomial case.

3.2. *Coarse space construction*

In [3], the SA-ρAMG method was studied where the coarse space was constructed based on a combination of SA-AMG and spectral AMGe. The construction shows that the resulting coarse space satisfies a weak approximation property without any regularity assumptions (in the PDE case). In this section, by small modification of the construction in [3], we show that the resulting coarse grid interpolant has an A-norm as close to unity as we require and still satisfies a weak approximation property. Such coarse grid space or respective prolongator is one of the key ingredients in our analysis.

This section focuses on the PDE case, that is, the model PDE example given by (2.2). The construction translates to the graph-Laplacian problem in a straightforward manner (without assuming any geometry and respective mesh sizes).

We consider a quasi-uniform fine-grid triangulation $\mathcal{T}_h$ and assume a coarse triangulation $\mathcal{T}_H$ of non-overlapping agglomerated elements $\{T\}$ has been constructed (by construction, non-overlapping means that any fine-grid element $\tau \in \mathcal{T}$ only belongs to one agglomerated element $T \in \mathcal{T}_H$ and the intersection of any two agglomerated elements is either empty or union of boundaries (edges in 2D or faces in 3D)). Let $h = \operatorname{diam}(\tau)$, $\tau \in \mathcal{T}_h$, and $H = \operatorname{diam}(T)$, $T \in \mathcal{T}_H$. We do not assume that H is comparable with h, which means we allow aggressive coarsening. In addition to the $\mathcal{T}_H$, we assume that we have respective aggregates $\{\mathcal{A}_i\}_{i=1}^{N_A}$ where each aggregate $\mathcal{A} = \mathcal{A}_i$ is contained in a unique $T \in \mathcal{T}_H$. The set $\{\mathcal{A}_i\}$ provides a non-overlapping partition of the set $\{1, 2, \cdots, N\}$ of the fine-degrees of freedom. Such aggregates can be constructed based on $\mathcal{T}_H$ easily. For nodes that belong to unique $T \in \mathcal{T}_H$, they are assigned to the corresponding aggregate $\mathcal{A}$ and, for nodes that shared by multiple agglomerated elements T, they are assigned to only one of the corresponding aggregates. In theory, it does not matter which one as long as the interface nodes are not shared by multiple aggregates.

3.2.1. Construction of tentative prolongator. The construction of the tentative prolongator follows the construction in [3]. First, we need to solve local generalized eigenvalue problems associated with each subdomain T (a union of fine-grid elements) that contains exactly one aggregate $\mathcal{A}$ of fine-degrees of freedom. Let A_T be the local stiffness matrix assembled on T and D_T be any local matrix on T that is spectrally equivalent to D restricted to T. For example, if $D = \operatorname{diag}(A)$ then $D_T = \operatorname{diag}(A_T)$. For other choice of D, D_T could be just the restriction of D on T. We solve

$$A_T\mathbf{q}_k = \lambda_k D_T \mathbf{q}_k, \quad k = 1, \ldots, N_T, \tag{3.6}$$

where N_T is the number of degrees of freedom in T. Second, by choosing the first $m_{\mathcal{A}}$ eigenvectors corresponding to the first $m_{\mathcal{A}}$ smallest eigenvalues, we form rectangular matrix $Q_T = \left[\mathbf{q}_1, \ldots, \mathbf{q}_{m_{\mathcal{A}}}\right]$. Then, we extract the rows from Q_T with row-indices from the aggregate $\mathcal{A}$ to form $Q_{\mathcal{A}}$. Using proper orthogonalization, we can get linearly independent set from $Q_{\mathcal{A}}$ and form $\hat{P}_{\mathcal{A}}$. Finally, we can construct a tentative prolongation $\hat{P}$, which is a block-diagonal and orthogonal matrix (with block corresponding to the set of aggregates $\{\mathcal{A}\}$) as follows

$$\hat{P} = \begin{pmatrix} \hat{P}_{\mathcal{A}_1} & 0 & \ldots & 0 \\ 0 & \hat{P}_{\mathcal{A}_2} & \ldots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & \ldots & 0 & \hat{P}_{\mathcal{A}_{N_A}} \end{pmatrix}.$$

If $\hat{P}_T = \left[\mathbf{q}_1, \ \ldots, \ \mathbf{q}_{m_{\mathcal{A}}}\right]$, it is straightforward to see that the following local estimate holds: for any $\mathbf{v}$ restricted to T, there exists $\mathbf{v}_c$, such that

$$\left\| \mathbf{v} - \hat{P}_T \mathbf{v}_c \right\|_{D_T}^2 \leqslant \frac{1}{\lambda_{m_{\mathcal{A}}+1}} \|\mathbf{v}\|_{A_T}^2.$$

Here, restricted to T, $\mathbf{v}$ can be represented by the linear combination of eigenvectors $\mathbf{q}_k$, $k = 1, 2, \ldots, N_T$, that is, $\mathbf{v} = \sum_{k=1}^{N_T} \alpha_k \mathbf{q}_k$, and we can choose $\hat{P}_T \mathbf{v}_c := \sum_{k=1}^{m_{\mathcal{A}}} \alpha_k \mathbf{q}_k$. Then, the aforementioned local estimate holds. Based on such local estimates, summing them up, it is proven in [3], that the tentative prolongation $\hat{P}$ satisfies the (global) weak approximation property

$$\left\| \mathbf{v} - \hat{P} \mathbf{v}_c \right\|_D^2 \leqslant \sigma \left(\frac{H}{h}\right)^2 \left(\max_{\mathcal{A}} \frac{h^2}{H^2 \lambda_{m_{\mathcal{A}}+1}}\right) \|\mathbf{v}\|_A^2 \leqslant \hat{\eta}_w \left(\frac{H}{h}\right)^2 \|\mathbf{v}\|_A^2, \tag{3.7}$$

for a uniform constant $\hat{\eta}_w$. Here, H stands for the characteristic diameter of the subdomains T, and h is the fine-grid mesh size. It is clear that the more eigenmodes (starting from the lower part of the spectrum) we include in the coarse space, the better constant in the approximation property is obtained. That is, with the expense of possibly large dimensional coarse space, we can ensure uniform constant $\hat{\eta}_w$. The aforementioned estimate is our main assumption in the following analysis.

Remark 3.1
Based on the choice of D and its local version D_T previously, we can have different tentative prolongators. The simplest choice is to have D be the diagonal of A and D_T being the diagonal of local stiffness matrix A_T (without imposing essential boundary conditions on ∂T).

Remark 3.2
Estimate (3.7) follows from the fact that $\frac{1}{\lambda_{m_{\mathcal{A}}+1}}$ scales as $\left(\frac{H}{h}\right)^2$. Therefore, by choosing $m_{\mathcal{A}}$ appropriately (sufficiently large), a global constant $\hat{\eta}_w$ is guaranteed. In the simple model case of second-order elliptic PDEs, this can be made more precise, that is, it can be uniformly bounded (cf., [2]). For more complicated PDEs, for example, coming from $H(\text{curl})$ or $H(\text{div})$ bilinear forms, the element agglomeration AMGe approach can be extended; actually, it has been done for the whole de Rham sequence of spaces [17]. However, because our focus here is to derive a new two-grid convergence result, we do not address (3.7) any further because it is outside the scope of the present paper.

3.2.2. *Smoothed prolongator by simple smoother.* In this section, we consider $S = p_\nu(b^{-1} D^{-1} A)$ with $p_\nu(t) = (1-t)^\nu$ and $b : \left\| D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right\| \leqslant b = \mathcal{O}(1)$ to construct the smoothed prolongator. Based on the tentative $\hat{P}$, the smoothed prolongator P is defined as

$$P = S\hat{P}. \tag{3.8}$$

We note that $p_\nu(t)$ is positive on $(0, 1)$, which is needed in the analysis to follow.

We first have the following lemma about the matrix polynomial S.

Lemma 3.3
Let $S = p_\nu(b^{-1} D^{-1} A)$ and $p_\nu(t) = (1-t)^\nu$, we have

$$\begin{aligned} \|S\mathbf{v}\|_D &\leqslant \|\mathbf{v}\|_D, \\ \|S\mathbf{v}\|_A &\leqslant \|\mathbf{v}\|_A, \\ \|(I-S)\mathbf{v}\|_A &\leqslant \|\mathbf{v}\|_A. \end{aligned}$$

Proof
Note that $(1-t)^\nu \in (0, 1]$ for $t \in (0, 1]$, we have

DOI: 10.1002/nla

$$\begin{aligned}\|S\mathbf{v}\|_D &= \left\| D^{\frac{1}{2}} p_\nu(b^{-1}D^{-1}A)D^{-\frac{1}{2}} \left(D^{\frac{1}{2}}\mathbf{v}\right)\right\| \\ &= \left\| p_\nu\left(b^{-1}D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\left(D^{\frac{1}{2}}\mathbf{v}\right)\right\| \leqslant \left\|D^{\frac{1}{2}}\mathbf{v}\right\| = \|\mathbf{v}\|_D.\end{aligned}$$

Similarly,

$$\|S\mathbf{v}\|_A = \left\|A^{\frac{1}{2}} p_\nu(b^{-1}D^{-1}A)A^{-\frac{1}{2}}\left(A^{\frac{1}{2}}\mathbf{v}\right)\right\| = \left\|p_\nu\left(b^{-1}A^{\frac{1}{2}}D^{-1}A^{\frac{1}{2}}\right)\left(A^{\frac{1}{2}}\mathbf{v}\right)\right\| \leqslant \left\|A^{\frac{1}{2}}\mathbf{v}\right\| = \|\mathbf{v}\|_A.$$

In the latter two estimates, we used the fact that the SPD matrices $b^{-1}D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $b^{-1}A^{\frac{1}{2}}D^{-1}A^{\frac{1}{2}}$ have their spectra contained in (0, 1]. In the same way, we have

$$\|(I-S)\mathbf{v}\|_A = \left\|\left(I - p_\nu\left(b^{-1}A^{\frac{1}{2}}D^{-1}A^{\frac{1}{2}}\right)\right)\left(A^{\frac{1}{2}}\mathbf{v}\right)\right\| \leqslant \left\|A^{\frac{1}{2}}\mathbf{v}\right\| = \|\mathbf{v}\|_A,$$

which completes the proof. □

Now, letting $Q = \hat{P}\hat{P}^T D$ be the D-orthogonal projection onto the space Range$\left(\hat{P}\right)$, using the above properties, we are ready to prove the following lemma about the A-norm of the smoothed prolongator P. In what follows, for any $\mathbf{v}$, we let $\mathbf{v}_c = \hat{P}^T D\mathbf{v}$. We note that we use two different values of ν; one, ν, is used in the construction of the smoothed interpolant P, whereas another value, ν_r, is used in the construction of the smoother M with the subscript 'r' here referring to the relaxation (smoothing) process of the two-grid algorithm.

Lemma 3.4
Let P be the smoothed prolongator defined by (3.8) and ν is the polynomial degree. We have

$$\|\mathbf{v} - P\mathbf{v}_c\|_A \leqslant (1 + \eta_P(\nu_P))\|\mathbf{v}\|_A, \quad \forall\ \mathbf{v}, \tag{3.9}$$

where ν_P are chosen such that $(2\nu_P + 1)\frac{H^2}{h^2} \simeq 2\nu + 1$ for sufficiently large ν and

$$\eta_P(\nu_P) := \frac{\sqrt{b\hat{\eta}_w}}{\sqrt{2\nu_P + 1}} \to 0 \text{ as } \nu_P \to \infty$$

Proof

$$\begin{aligned}\|\mathbf{v} - P\mathbf{v}_c\|_A = \|\mathbf{v} - SQ\mathbf{v}\|_A &\leqslant \|(I-S)\mathbf{v}\|_A + \|S(\mathbf{v} - Q\mathbf{v})\|_A \\ &\leqslant \|\mathbf{v}\|_A + \|S(\mathbf{v} - Q\mathbf{v})\|_A \\ &\leqslant \|\mathbf{v}\|_A + \left\|A^{\frac{1}{2}}D^{-\frac{1}{2}}\left(I - b^{-1}D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)^\nu D^{\frac{1}{2}}(I-Q)\mathbf{v}\right\| \\ &\leqslant \|\mathbf{v}\|_A + \sqrt{b}\max_{t\in(0,1)} t^{\frac{1}{2}}(1-t)^\nu \|(I-Q)\mathbf{v}\|_D \\ &\leqslant \|\mathbf{v}\|_A + \frac{\sqrt{b}}{\sqrt{2\nu+1}}\|\mathbf{v} - Q\mathbf{v}\|_D \\ &\leqslant \left(1 + \sqrt{\hat{\eta}_w}\frac{H}{h}\frac{\sqrt{b}}{\sqrt{2\nu+1}}\right)\|\mathbf{v}\|_A.\end{aligned}$$

Here, the weak approximation property (3.7) is used in the last inequality. Thus, letting $2\nu + 1 \simeq (2\nu_P + 1)\frac{H^2}{h^2}$ for sufficiently large ν_P, the desired result (3.9) holds with $\eta_P(\nu_P) := \frac{\sqrt{b\hat{\eta}_w}}{\sqrt{2\nu_P+1}}$. □

The next lemma shows that the smoothed prolongation P (3.8) also exhibits a weak approximation property without any regularity assumptions.

Lemma 3.5
Let P be the smoothed prolongator defined by (3.8). We have

$$\|\mathbf{v} - P\mathbf{v}_c\|_D^2 \leqslant \eta_w \|\mathbf{v}\|_A^2, \quad \forall\ \mathbf{v}, \tag{3.10}$$

where $\eta_w = (2\nu + 1)\left(\eta'_w\right)^2$ and η'_w is a uniform constant that is independent of discretization parameters.

Proof
Similarly as before, letting $X = b^{-1}A^{1/2}D^{-1}A^{1/2}$, we have

$$\begin{aligned}
\|\mathbf{v} - P\mathbf{v}_c\|_D = \|\mathbf{v} - SQ\mathbf{v}\|_D &\leqslant \|S(\mathbf{v} - Q\mathbf{v})\|_D + \|(I - S)\mathbf{v}\|_D \\
&= \|(I - Q)\mathbf{v}\|_D + \left\| D^{\frac{1}{2}}\left(I - p_\nu(b^{-1}D^{-1}A)\right)A^{-\frac{1}{2}}\right\| \|\mathbf{v}\|_A \\
&\leqslant \|(I - Q)\mathbf{v}\|_D + \frac{1}{\sqrt{b}}\|X^{-1/2}(I - (I - X)^\nu)\|\|\mathbf{v}\|_A \\
&\leqslant \|(I - Q)\mathbf{v}\|_D + \frac{1}{\sqrt{b}}\sup_{t\in(0,1]}\frac{1 - (1 - t)^\nu}{\sqrt{t}}\|\mathbf{v}\|_A
\end{aligned}$$

Now note that, for $\frac{1}{\nu} \leqslant t \leqslant 1$, we have

$$\frac{1 - (1 - t)^\nu}{\sqrt{t}} \leqslant \frac{1}{\sqrt{t}} \leqslant \sqrt{\nu}.$$

For $0 < t < \frac{1}{\nu}$, we have

$$\frac{1 - (1 - t)^\nu}{\sqrt{t}} = \sqrt{t}\left(1 + (1 - t) + (1 - t)^2 + \cdots + (1 - t)^{\nu - 1}\right) \leqslant \frac{1}{\sqrt{\nu}}\nu = \sqrt{\nu}.$$

Therefore, we have

$$\begin{aligned}
\|\mathbf{v} - P\mathbf{v}_c\|_D &\leqslant \|(I - Q)\mathbf{v}\|_D + \frac{\sqrt{\nu}}{\sqrt{b}}\|\mathbf{v}\|_A \\
&\leqslant \left(\frac{H}{h}\sqrt{\hat{\eta}_w} + \frac{\sqrt{\nu}}{\sqrt{b}}\right)\|\mathbf{v}\|_A \\
&\leqslant \left(\sqrt{2\nu + 1}\right)\left(\sqrt{\hat{\eta}_w}\frac{H}{h}\frac{1}{\sqrt{2\nu + 1}} + \frac{1}{\sqrt{b}}\frac{\sqrt{\nu}}{\sqrt{2\nu + 1}}\right)\|\mathbf{v}\|_A \\
&\leqslant \left(\sqrt{2\nu + 1}\right)\eta'_w\|\mathbf{v}\|_A
\end{aligned}$$

In the latter inequality, η'_w is independent of the discretization parameters. Thus, (3.10) holds with

$$\eta_w = (2\nu + 1)\left(\eta'{}_w\right)^2.$$

This completes the proof. □

Remark 3.3
We note that η_w deteriorates like $2\nu + 1 \simeq (2\nu_P + 1)\ \frac{H^2}{h^2}$ when increasing ν_P. Therefore, for the overall convergence rate of the two-grid method, we need to carefully choose ν_P to balance the deterioration of η_w. Moreover, the deterioration rate is worse than in the Chebyshev-like polynomial case as we will show next.

3.2.3. Smoothed prolongator by Chebyshev-like polynomial. Based on the tentative $\hat{P}$, the matrix polynomial $S = s_\nu(b^{-1}D^{-1}A)$ and the polynomial $s_\nu(t)$ defined in (2.8) (usually is refereed to as the SA-polynomial), the smoothed prolongator P we consider, reads

$$P = S^2\hat{P}. \tag{3.11}$$

We note that for the analysis to follow, we need to use non-negative polynomial (see third estimate in Lemma 3.6), which motivates the choice S^2 versus the commonly used polynomial $s_\nu(t)$ in SA-AMG. The role of this polynomial is to make the interpolation matrix P more stable in energy norm, while maintaining the weak approximation property of the tentative interpolant $\hat{P}$ with some deterioration, which we can control.

Based on Proposition 2.3, we have the following main estimates for the matrix polynomial S. Because the proof is similar to the proof of Lemma 3.3, thus, it is omitted.

Lemma 3.6
Let $S = s_\nu(b^{-1}D^{-1}A)$ where $b = \mathcal{O}(1)$ is such that $\left\| D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \right\| \leqslant b$ and $s_\nu(t)$ is defined in (2.8). We have,

$$\|S\mathbf{v}\|_D \leqslant \|\mathbf{v}\|_D, \tag{3.12}$$

$$\|S\mathbf{v}\|_A \leqslant \|\mathbf{v}\|_A, \tag{3.13}$$

$$\|(I - S^2)\mathbf{v}\|_A \leqslant \|\mathbf{v}\|_A. \tag{3.14}$$

Next, we show that for this Chebyshev-like polynomial, we can also have the following lemma about the A-norm of the smoothed prolongation P.

Lemma 3.7
Let P be the smoothed prolongator defined in (3.11), then

$$\|\mathbf{v} - P\mathbf{v}_c\|_A \leqslant (1 + \eta_P(\nu_P))\|\mathbf{v}\|_A, \quad \forall\, \mathbf{v}, \tag{3.15}$$

where we choose ν_P for the polynomial degree $2\nu + 1 \simeq (2\nu_P + 1)\frac{H}{h}$ such that $\eta_P(\nu_P) = \sqrt{\hat{\eta}_w}\frac{\sqrt{b}}{\nu_P+1} \to 0$ as $\nu_P \to \infty$.

Proof
From the definition of P, we have

$$\begin{aligned}\|\mathbf{v} - P\mathbf{v}_c\|_A = \|\mathbf{v} - S^2Q\mathbf{v}\|_A &\leqslant \|(I - S^2)\mathbf{v}\|_A + \|S^2(\mathbf{v} - Q\mathbf{v})\|_A \\ &\leqslant \|\mathbf{v}\|_A + \|S(\mathbf{v} - Q\mathbf{v})\|_A \\ &\leqslant \|\mathbf{v}\|_A + \frac{b^{\frac{1}{2}}}{2\nu + 1}\,\|\mathbf{v} - Q\mathbf{v}\|_D.\end{aligned}$$

Next, use estimate (3.7) to bound the term $\|\mathbf{v} - Q\mathbf{v}\|_D$, which gives

$$\|\mathbf{v} - P\mathbf{v}_c\|_A = \|\mathbf{v} - S^2Q\mathbf{v}\|_A \leqslant \left(1 + \sqrt{\hat{\eta}_w}\,\frac{H}{h}\,\frac{b^{\frac{1}{2}}}{2\nu + 1}\right)\|\mathbf{v}\|_A.$$

Thus, letting $2\nu + 1 \simeq (2\nu_P + 1)\frac{H}{h}$ for sufficiently large ν_P, the result (3.15) holds with $\eta_P(\nu_P) := \sqrt{\hat{\eta}_w}\frac{b^{\frac{1}{2}}}{\nu_P+1}$. □

Remark 3.4
Note that the SA-polynomial is defined using $D^{-1}A$ with $b = \mathcal{O}(1)$ as mentioned before. Therefore, for sufficiently large ν_P, we can make the operator $I - P$ as close to an A–orthogonal projection as we want, or equivalently we can have its A-norm as close to unity as we want.

The next lemma shows that the smoothed interpolant P exhibits a weak approximation property, which is somewhat worse than the similar property of the tentative one.

Lemma 3.8
For the smoothed prolongation P defined by (3.11), we have

$$\|\mathbf{v} - P\mathbf{v}_c\|_D^2 \leqslant \eta_w \|\mathbf{v}\|_A^2, \quad \forall\ \mathbf{v}, \tag{3.16}$$

where $\eta_w = \left((2\nu + 1)\eta_w'\right)^2 \simeq \left((2\nu_P + 1)\frac{H}{h}\eta_w'\right)^2$, and η_w' is a uniform constant that is independent of discretization parameters.

Proof
Similarly as before, we have

$$\begin{aligned}
\|\mathbf{v} - P\mathbf{v}_c\|_D = \|\mathbf{v} - S^2 Q\mathbf{v}\|_D &\leqslant \|S^2(\mathbf{v} - Q\mathbf{v})\|_D + \|(I - S^2)\mathbf{v}\|_D \\
&\leqslant \|(I - Q)\mathbf{v}\|_D + \|(I + S)(I - S)\mathbf{v}\|_D \\
&\leqslant \|(I - Q)\mathbf{v}\|_D + 2\|(I - S)\mathbf{v}\|_D \\
&\leqslant \|(I - Q)\mathbf{v}\|_D + 2\left\|D^{\frac{1}{2}}(I - s_\nu(b^{-1}D^{-1}A))A^{-\frac{1}{2}}\right\| \|\mathbf{v}\|_A.
\end{aligned}$$

The last matrix norm can be estimated as follows

$$\left\|D^{\frac{1}{2}}(I - s_\nu(b^{-1}D^{-1}A))A^{-\frac{1}{2}}\right\| = \left\|X^{-\frac{1}{2}}(I - s_\nu(b^{-1}X))\right\|, \quad X = A^{\frac{1}{2}}D^{-1}A^{\frac{1}{2}}.$$

Therefore,

$$\begin{aligned}
\|\mathbf{v} - S^2 Q\mathbf{v}\|_D &\leqslant \|(I - Q)\mathbf{v}\|_D + 2b^{-\frac{1}{2}} \sup_{t\in(0,\,1]} \frac{1 - s_\nu(t)}{\sqrt{t}}\ \|\mathbf{v}\|_A \\
&\leqslant \frac{H}{h}\ \sqrt{\hat{\eta}_w}\ \|\mathbf{v}\|_A + \frac{2}{\sqrt{b}}\ \frac{C_\nu}{2\nu + 1}\ (2\nu + 1)\ \|\mathbf{v}\|_A \\
&\leqslant (2\nu + 1)\ \left[\frac{H}{h}\frac{1}{2\nu + 1}\ \sqrt{\hat{\eta}_w} + \frac{2}{\sqrt{b}}\ \frac{C_\nu}{2\nu + 1}\right] \|\mathbf{v}\|_A \\
&\leqslant (2\nu + 1)\ \eta_w'\ \|\mathbf{v}\|_A.
\end{aligned}$$

In the latter inequality, we used estimate (2.7), which shows that we can choose η_w' to be a uniform constant. Thus, (3.16) holds with

$$\eta_w = \left((2\nu + 1)\ \eta_w'\right)^2,$$

which completes the proof. □

Remark 3.5
We note that $\sqrt{\eta_w}$ deteriorates like $2\nu + 1 \simeq (2\nu_P + 1)\ \frac{H}{h}$ when increasing ν_P. Therefore, for the overall convergence rate of the two-grid method, we need to carefully choose ν_P to balance the deterioration of η_w.

4. MAIN RESULTS: TWO-GRID CONVERGENCE RATE

In this section, based on the construction of the smoother M and the smoothed prolongator P, utilizing their properties, we show that the two-grid method (Algorithm 1) converges uniformly without any regularity assumption. Moreover, its convergence rate improves uniformly when we increase the degrees ν_r and ν_P (or equivalently, ν) of the two polynomials appropriately.

4.1. Simple smoother case

We first consider the case of using simple polynomials to construct M and P. Their properties are summarized in Lemmas 3.1, 3.4, and 3.5. Using the smoothing property (Lemma 3.1) of the polynomial smoother M and the weak approximation property (Lemma 3.5) of P, (recalling that for any $\mathbf{v}$, $\mathbf{v}_c = \hat{P}^T D\mathbf{v}$), we have the estimates

$$\begin{aligned}\|\mathbf{v} - P\mathbf{v}_c\|_{\overline{M}}^2 &\leqslant (1+\tau)\left[\|\mathbf{v} - P\mathbf{v}_c\|_A^2 + \frac{1}{\tau}\eta_S(\nu_r)\ \|\mathbf{v} - P\mathbf{v}_c\|_D^2\right] \\ &\leqslant (1+\tau)\left[\|\mathbf{v} - P\mathbf{v}_c\|_A^2 + \frac{1}{\tau}\eta_S(\nu_r)\ \eta_w\|\mathbf{v}\|_A^2\right]\end{aligned}$$

Therefore, applying the two-grid version of the XZ-identity (Theorem 2.1), we have

$$K_{TG} \leqslant (1+\tau)\ \left[\max_{\mathbf{v}}\ \frac{\|\mathbf{v} - P\mathbf{v}_c\|_A^2}{\|\mathbf{v}\|_A^2} + \frac{1}{\tau}\eta_S(\nu_r)\eta_w\right].$$

By Lemma 3.4, we have

$$K_{TG} \leqslant (1+\tau)\left[(1+\eta_P(\nu_P))^2 + \frac{1}{\tau}\eta_S(\nu_r)\eta_w\right].$$

Then by the definition of η_w in Lemma 3.5, we end up with the estimate

$$K_{TG} \leqslant (1+\tau)\left[(1+\eta_P(\nu_P))^2 + \frac{1}{\tau}\eta_S(\nu_r)(2\nu_P+1)\frac{H^2}{h^2}\left(\eta_w'\right)^2\right].$$

Now we need to carefully choose ν_r such that the term $\frac{1}{\tau}\eta_S(\nu_r)(2\nu_P+1)\frac{H^2}{h^2}\left(\eta_w'\right)^2$ can be made as small as possible. We choose ν_r such that for any given $\tau \in (0,1)$, we have

$$\tau \geqslant \frac{1}{\tau}\eta_S(\nu_r)(2\nu_P+1)\frac{H^2}{h^2}.$$

In particular, we can choose τ as small as we want. Recall that $\eta_S(\nu_r) = \frac{1}{2\nu_r}$; hence, we can choose ν_r such that

$$\tau = \frac{H}{h}\sqrt{\frac{2\nu_P+1}{2\nu_r}} \in (0,1),$$

be as small as needed. As a result, we have

$$K_{TG} \leqslant (1+\tau)\left[\left(1+\sqrt{\hat{\eta}_w}\frac{1}{\sqrt{2\nu_P+1}}\right)^2 + \tau\left(\eta_w'\right)^2\right],$$

which improves with $\nu_P \to \infty$ and $\tau \to 0$.

We summarize the preceding result in the next theorem.

Theorem 4.1

Let the smoother M satisfy the smoothing property (3.3) (or Lemma 3.1) and the smoothed prolongator P be constructed as in (3.8) so that Lemmas 3.4 and 3.5 hold. Then, we have

$$(B_{TG}\mathbf{v},\mathbf{v}) \leqslant K_{TG}(\mathbf{v},\mathbf{v})_A \quad \text{or equivalently} \quad \|E_{TG}\|_A = \rho_{TG} := 1 - \frac{1}{K_{TG}},$$

where

$$K_{TG} \leqslant (1+\tau)\left[\left(1+\frac{\sqrt{\hat{\eta}_w}}{\sqrt{2\nu_P+1}}\right)^2+\tau\left(\eta'_w\right)^2\right],$$

with

$$\tau=\frac{H}{h}\sqrt{\frac{2\nu_P+1}{2\nu_r}}\in(0,1),$$

Moreover, it is clear that $K_{TG}\to 1$ as both, $\nu_r\to\infty$ and $\nu_P\to\infty$, such that $\tau\mapsto 0$.

4.2. *Chebyshev-like polynomial case*

By using the same arguments as in the previous section, we can also derive the two-grid convergence rate when we use Chebyshev-like polynomial to construct both smoother M and prolongation P. We summarize the results in the following theorem.

Theorem 4.2
Let the smoother M satisfy the smoothing property (3.5) (or Lemma 3.2) and the smoothed prolongator P be constructed as in (3.11) so that Lemmas 3.7 and 3.8 hold. Then, we have

$$(B_{TG}\mathbf{v},\mathbf{v})\leqslant K_{TG}(\mathbf{v},\mathbf{v})_A \quad \text{or equivalently} \quad \|E_{TG}\|_A=\rho_{TG}:=1-\frac{1}{K_{TG}},$$

where

$$K_{TG}\leqslant(1+\tau)\left[\left(1+\sqrt{\hat{\eta}_w}\frac{1}{2\nu_P+1}\right)^2+\tau\left(\eta'_w\right)^2\right],$$

with

$$\tau=\frac{H}{h}\,\frac{2\nu_P+1}{2\nu_r+1}\in(0,1).$$

Moreover, it is clear that $K_{TG}\to 1$ as both, $\nu_r\to\infty$ and $\nu_P\to\infty$, such that $\tau\mapsto 0$.

Remark 4.1
We note that, using simple smoother, the constant K_{TG} goes to 1 slower than using Chebyshev-like polynomials although the computational complexity is similar. Moreover, one can also mix these two choices for smoother and smoothed prolongator in practice. Similar two-grid convergence rate results also hold.

5. NUMERICAL EXPERIMENTS

In this section, we illustrate our theoretical results with some numerical experiments. In all tests, we use the weighted $\ell 1$-smoother as D (cf., e.g., [3]). We test the simple polynomial case, further denoted by $z_\nu(t)=(1-t)^\nu$, which results in the following matrix polynomial Z, and respective smoothed interpolant P,

$$Z=z_\nu(D^{-1}A) \text{ and } P=Z\hat{P}.$$

We also test the Chebyshev-like polynomials. Although we analyzed the case $P:=S^2\hat{P}$, for comparison, we also consider the simpler interpolant $P:=S\hat{P}$. In the tables, 'A_c nnz' denotes the number of non-zeros in the coarse-grid matrix A_c, and 'OC' stands for *operator complexity*, that is, $\mathrm{OC}=(A \text{ nnz}+A_c \text{ nnz})/(A \text{ nnz})$.

5.1. 2D finite elements

We consider the diffusion equation on the unit square with checkerboard jump diffusion coefficient as shown in Figure 4(c), where the jump varies up to six orders of magnitude. The mesh is unstructured with 51,681 vertices and 102,400 triangular elements. We construct 200 aggregates by (unstructured) element agglomeration with 350 degrees of freedom for the coarse problem. The coarse degrees of freedom are chosen by selecting a fixed portion of the eigenvectors corresponding to the lower part of the spectrum of the local generalized eigenvalue problems (3.6). More specifically, for a given agglomerate T and corresponding aggregate $\mathcal{A} \subset T$, we choose the smallest $m_{\mathcal{A}}$ eigenvalues and corresponding eigenvectors such that for a tolerance $\theta \in (0, 1)$, we have

$$\lambda_{m_{\mathcal{A}}} \geqslant \theta \lambda_{\max}. \tag{5.1}$$

In the present case, D_T is the diagonal of A_T and $\theta = 0.003$. The results are presented in Table I

Remark 5.1
We note that ν_p in this section stands for the ν used in Section 3. We use a slightly different notation to make the tables more readable. Also, we remind that $\nu_p = \nu$ and ν_P are related as shown in Lemma 3.4 for the simple polynomial, and in Lemma 3.7 for the Chebyshev-like polynomial.

We can compare the rows in the Table I that have the same values 'ν_r'; 'A_c nnz', equivalently 'OC', in which case the total number of prolongation smoothing steps is the same. In such a case, we can see that up to a certain point $P := S\hat{P}$ performs better for this problem than $P := S^2\hat{P}$. When the number of prolongation smoothing steps ν_p is sufficiently large (in this case, also the convergence factor becomes very small), $P := S^2\hat{P}$ results in a better TG convergence. Note that, we compare convergence of the these two methods that require exactly the same computational cost per TG cycle. We can also notice that the simple polynomial, Z, behaves as the theory predicts; however, its convergence factor is of an order of magnitude larger than that of the S-polynomials.

The respective results are also shown graphically in Figures 1, 2, and 3. The aggregates are illustrated in Figure 4.

5.2. 3D finite elements

In this section, we consider the diffusion equation on a unit cube with checkerboard jump coefficient as illustrated in Figure 9, where the jump varies by six orders of magnitude. The mesh is unstructured with 83,497 vertices and 459,776 elements (tetrahedra). We construct 400 (unstructured) aggregates and the spectral SA-AMG methods leads to 439 coarse degrees of freedom corresponding to a tolerance $\theta = 0.003$ (5.1). We use for D_T the diagonal of A_T and $\theta = 0.003$.

The obtained results are summarized in Table II.

Similarly to the 2D case, comparing the entries with equal cost per TG iteration cycle, in Table II, we see that up to a certain point $P := S\hat{P}$ performs better, whereas when the prolongation smoothing steps become sufficiently large, $P := S^2\hat{P}$ starts to perform better. Note the order of magnitude difference in convergence factors between the Z and S polynomials.

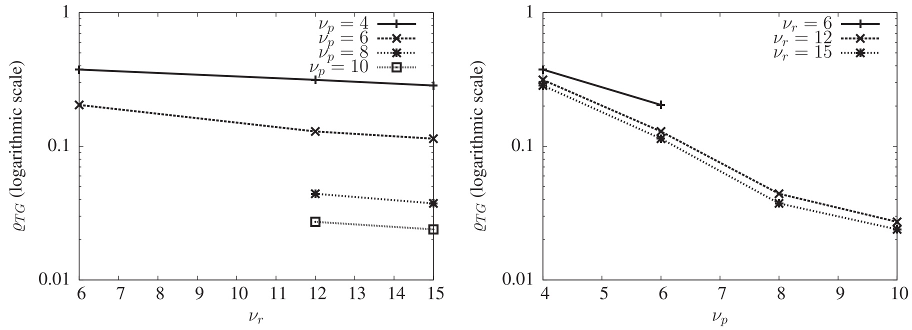
The respective results are also shown graphically in Figures 5, 6, and 7. The mesh, coefficient distribution, and the aggregation are illustrated in Figures 8, 9, 10, and 11.

5.3. Graph Laplacian

Finally, we consider a graph Laplacian problem defined on a graph with 35,638 vertices and 42,827 edges, which is available at http://opte.org (Figure 12). We fix 356 vertices as 'essential boundary'.

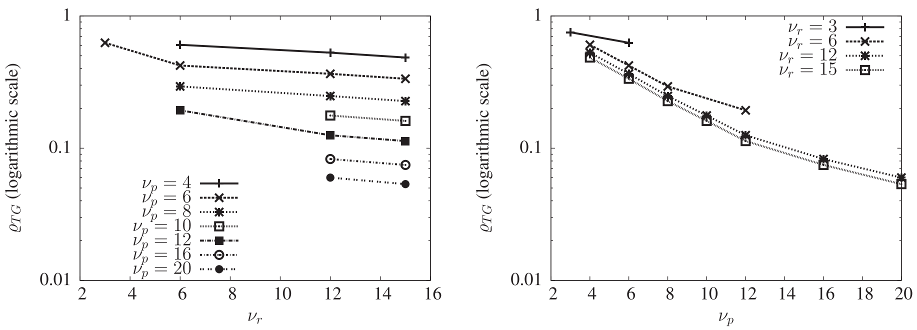
Table I. Results for 2D finite elements.

ν_p	ν_r	A_c nnz	OC	ϱ_{TG}
(a) $P := S^2 \hat{P}$				
3	3	7620	1.02	0.662
4	6	9412	1.03	0.375
4	12	9412	1.03	0.314
4	15	9412	1.03	0.285
6	6	13,660	1.04	0.204
6	12	13,660	1.04	0.1291
6	15	13,660	1.04	0.1140
8	12	18,358	1.05	0.04413
8	15	18,358	1.05	0.03737
10	12	23,490	1.07	0.02721
10	15	23,490	1.07	0.02390
(b) $P := S \hat{P}$				
3	3	5106	1.01	0.753
4	6	5782	1.02	0.605
4	12	5782	1.02	0.528
4	15	5782	1.02	0.485
6	3	7620	1.02	0.627
6	6	7620	1.02	0.422
6	12	7620	1.02	0.365
6	15	7620	1.02	0.335
8	6	9412	1.03	0.293
8	12	9412	1.03	0.248
8	15	9412	1.03	0.227
10	12	11,680	1.03	0.1767
10	15	11,680	1.03	0.1608
12	6	13,660	1.04	0.1934
12	12	13,660	1.04	0.1252
12	15	13,660	1.04	0.1133
16	12	18,358	1.05	0.08281
16	15	18,358	1.05	0.07469
20	12	23,490	1.07	0.05988
20	15	23,490	1.07	0.05351
(c) $P := Z \hat{P}$				
6	3	7620	1.02	0.793
8	6	9412	1.03	0.700
8	12	9412	1.03	0.612
8	15	9412	1.03	0.565
12	6	13,660	1.04	0.615
12	12	13,660	1.04	0.536
12	15	13,660	1.04	0.492
16	12	18,358	1.05	0.471
16	15	18,358	1.05	0.431
20	12	23,490	1.07	0.415
20	15	23,490	1.07	0.379



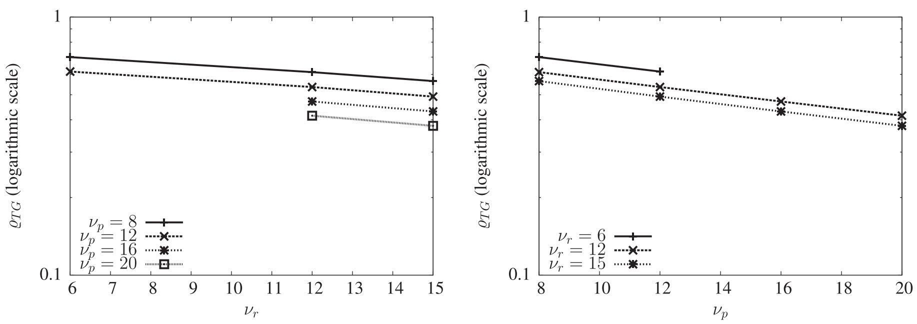
(a) Fixed ν_p and varying ν_r (b) Fixed ν_r and varying ν_p

Figure 1. Results for 2D finite elements with $P := S^2\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.



(a) Fixed ν_p and varying ν_r (b) Fixed ν_r and varying ν_p

Figure 2. Results for 2D finite elements with $P := S\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.

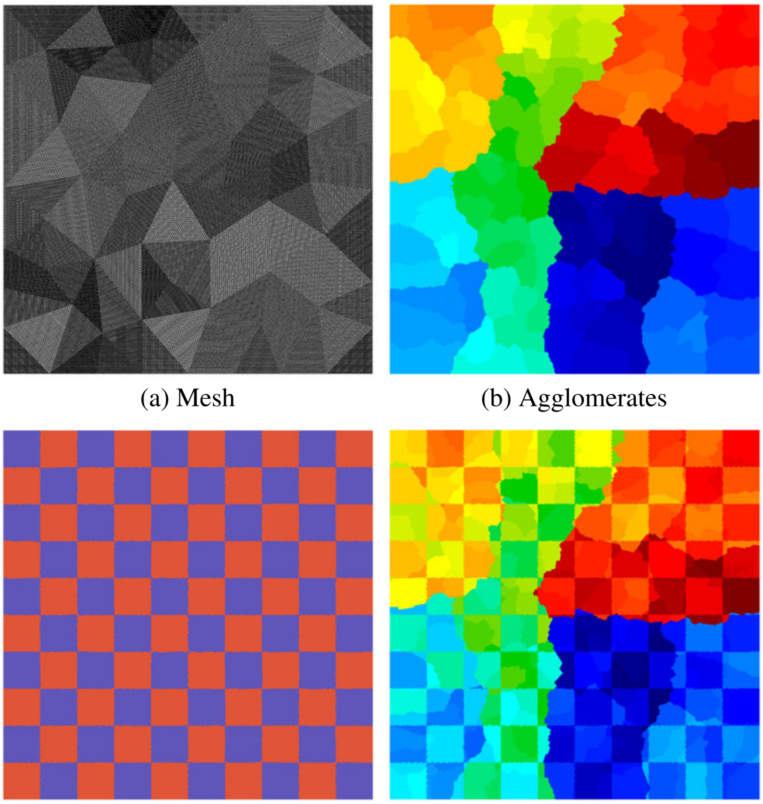


(a) Fixed ν_p and varying ν_r (b) Fixed ν_r and varying ν_p

Figure 3. Results for 2D finite elements with $P := Z\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.

The weight of each edge of the graph is 1. To apply the spectral SA-AMG method, we use the edge matrices A_e, which serve as 'element' matrices (similar to the PDE case), defined as follows:

$$A_e = \begin{bmatrix} 0 & 0 & \dots & & 0 \\ \vdots & 1 & \dots & -1 & 0 \\ 0 & \vdots & \ddots & \vdots & 0 \\ \vdots & -1 & \dots & 1 & 0 \\ 0 & 0 & \dots & & 0 \end{bmatrix}.$$



(a) Mesh (b) Agglomerates

(c) Coefficient (blue $= 1$, red $= 10^6$) (d) Agglomerates and coefficient together

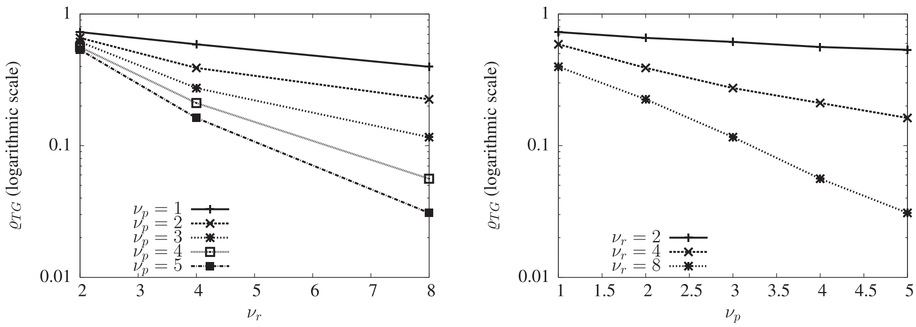
Figure 4. Figures for the 2D finite-element case. (a) mesh; (b) agglomerates; (c) coefficient (blue $= 1$, red $= 10^6$); and (d) agglomerates and coefficient together.

Table II. Results for 3D finite elements.

ν_p	ν_r	A_c nnz	OC	ϱ_{TG}
(a) $P := S^2\hat{P}$				
1	2	26771	1.02	0.729
1	4	26771	1.02	0.589
1	8	26771	1.02	0.398
2	2	62727	1.05	0.659
2	4	62727	1.05	0.390
2	8	62727	1.05	0.225
3	2	108,949	1.09	0.614
3	4	108,949	1.09	0.274
3	8	108,949	1.09	0.1161
4	2	153,545	1.13	0.562
4	4	153,545	1.13	0.211
4	8	153,545	1.13	0.05613
5	2	182,467	1.15	0.535
5	4	182,467	1.15	0.1625
5	8	182,467	1.15	0.03086

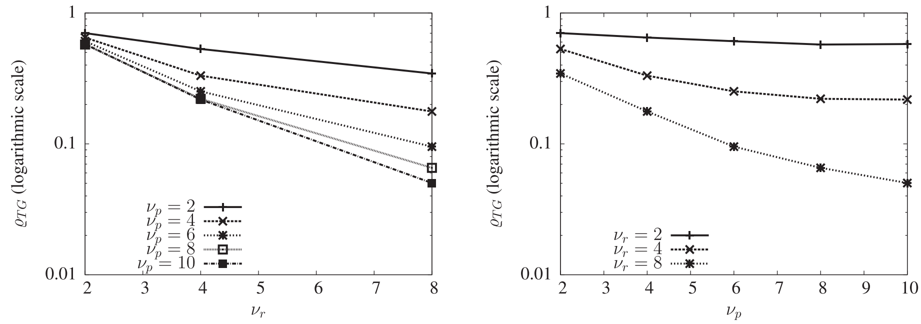
Table II. Continued.

ν_p	ν_r	A_c nnz	OC	ϱ_{TG}
(b) $P := S\hat{P}$				
2	2	26771	1.02	0.702
2	4	26771	1.02	0.531
2	8	26771	1.02	0.345
4	2	62727	1.05	0.649
4	4	62727	1.05	0.332
4	8	62727	1.05	0.1769
6	2	108,949	1.09	0.609
6	4	108,949	1.09	0.252
6	8	108,949	1.09	0.09520
8	2	153,545	1.13	0.574
8	4	153,545	1.13	0.221
8	8	153,545	1.13	0.06559
10	2	182,467	1.15	0.579
10	4	182,467	1.15	0.218
10	8	182,467	1.15	0.05024
(c) $P := Z\hat{P}$				
2	2	26771	1.02	0.759
2	4	26771	1.02	0.635
2	8	26771	1.02	0.441
4	2	62727	1.05	0.703
4	4	62727	1.05	0.535
4	8	62727	1.05	0.349
6	2	108,949	1.09	0.677
6	4	108,949	1.09	0.461
6	8	108,949	1.09	0.285
8	2	153,545	1.13	0.660
8	4	153,545	1.13	0.405
8	8	153,545	1.13	0.238
10	2	182,467	1.15	0.646
10	4	182,467	1.15	0.361
10	8	182,467	1.15	0.201



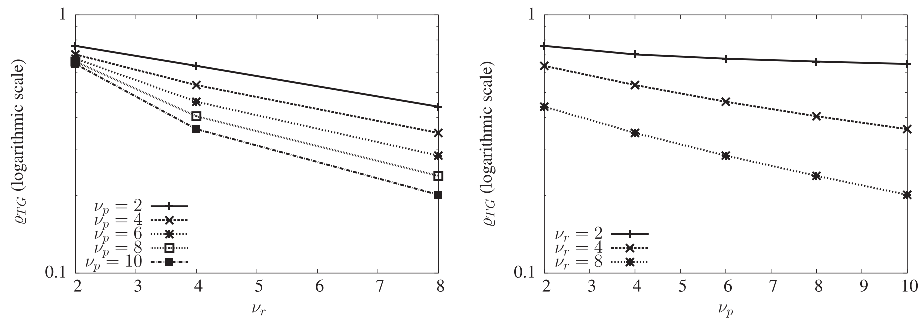
(a) Fixed ν_p and varying ν_r (b) Fixed ν_r and varying ν_p

Figure 5. Results for 3D finite elements with $P := S^2\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.



(a) Fixed ν_p and varying ν_r (b) Fixed ν_r and varying ν_p

Figure 6. Results for 3D finite elements with $P := S\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.



(a) Fixed ν_p and varying ν_r (b) Fixed ν_r and varying ν_p

Figure 7. Results for 3D finite elements with $P := Z\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.

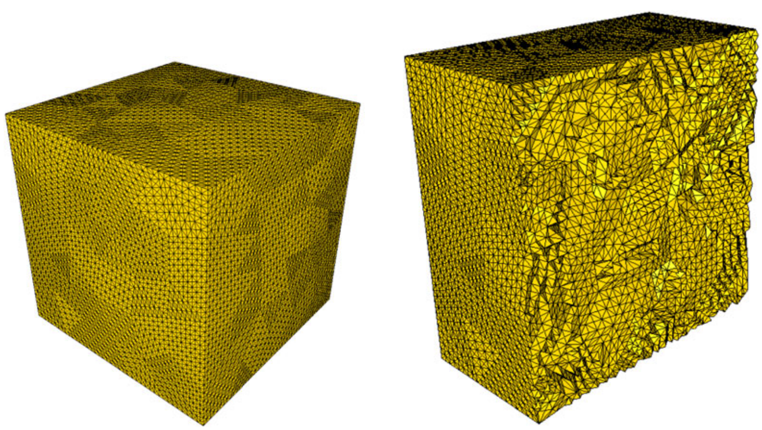


Figure 8. 3D finite-element mesh.

The only non-zero entries in A_e are at positions (i, i), (i, j), (j, i), and (j, j) for any edge $e = (i, j)$ of the given graph. We note that the graph Laplacian is assembled from the preceding symmetric positive semi-definite edge matrices.

We construct 143 aggregates (seen in Figure 13), and the spectral SA-AMG method leads to 502 coarse degrees of freedom, which correspond to a tolerance $\theta = 0.004$ in (5.1). The matrix D_T, as before, is the diagonal of A_T.

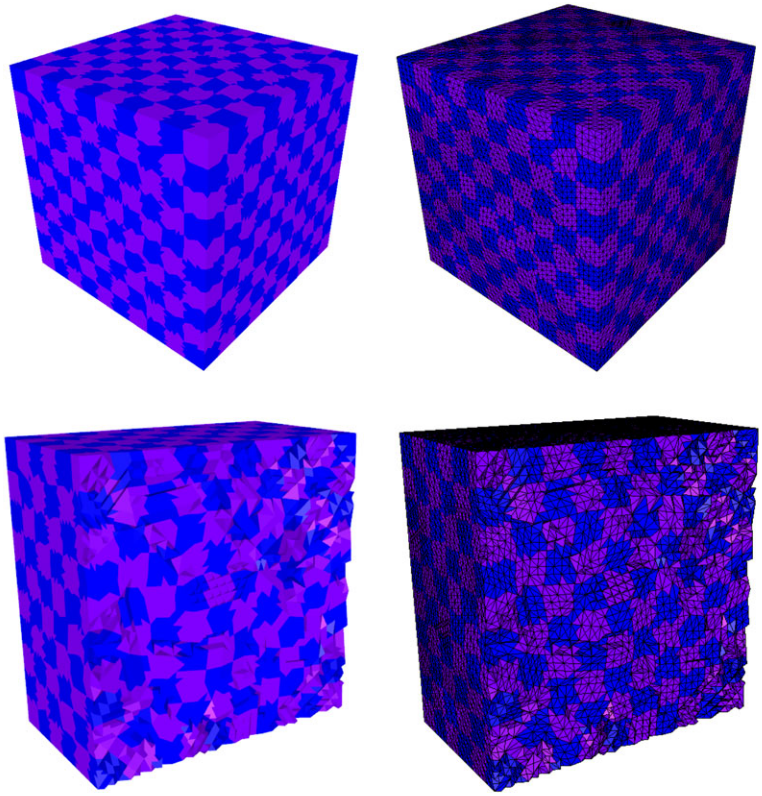


Figure 9. Coefficient variation; 3D finite-element case (dark $= 1$, light $= 10^6$).

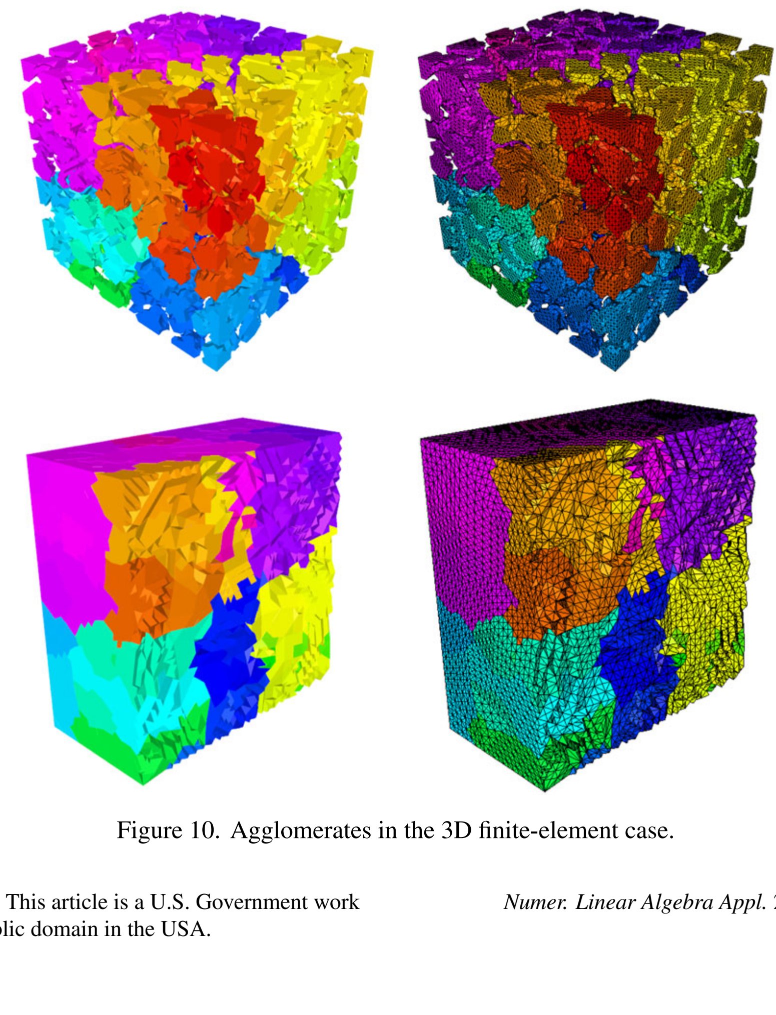


Figure 10. Agglomerates in the 3D finite-element case.

DOI: 10.1002/nla

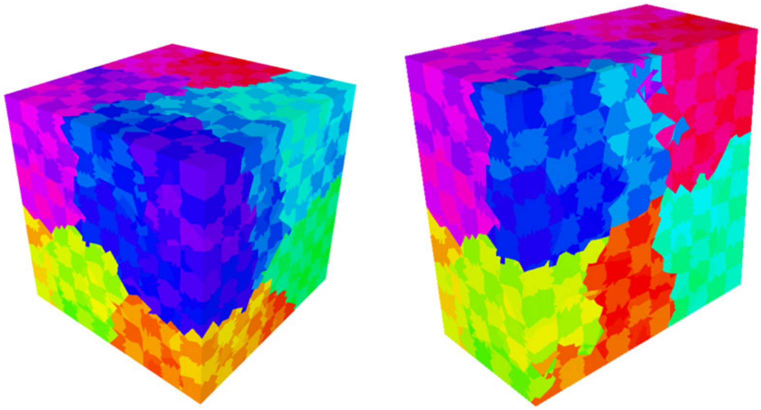


Figure 11. Coefficient and agglomerates together in the 3D finite-element case.

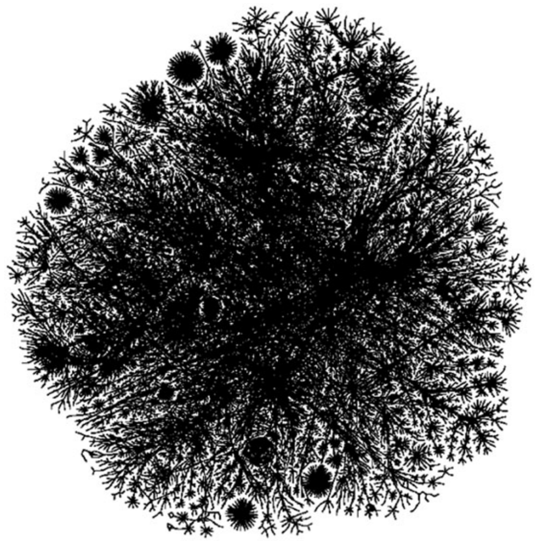


Figure 12. Graph from http://opte.org.

Table III. Results for the graph Laplacian problem.

ν_p	ν_r	A_c nnz	OC	ϱ_{TG}
(a) $P := S^2\hat{P}$				
1	4	134,194	2.11	0.824
1	8	134,194	2.11	0.738
1	16	134,194	2.11	0.540
3	4	247,270	3.04	0.477
3	8	247,270	3.04	0.353
3	16	247,270	3.04	0.257
5	4	251,002	3.07	0.445
5	8	251,002	3.07	0.1416
5	16	251,002	3.07	0.1098
6	4	251,002	3.07	0.443

Table III. Continued.

ν_p	ν_r	A_c nnz	OC	ϱ_{TG}
6	8	251,002	3.07	0.1103
6	16	251,002	3.07	0.08815
7	4	251,002	3.07	0.443
7	8	251,002	3.07	0.08381
7	16	251,002	3.07	0.06935
(b) $P := S\hat{P}$				
2	4	134,194	2.11	0.754
2	8	134,194	2.11	0.671
2	16	134,194	2.11	0.490
6	4	247,270	3.04	0.470
6	8	247,270	3.04	0.352
6	16	247,270	3.04	0.253
10	4	251,002	3.07	0.445
10	8	251,002	3.07	0.1830
10	16	251,002	3.07	0.1333
12	4	251,002	3.07	0.444
12	8	251,002	3.07	0.1261
12	16	251,002	3.07	0.09338
14	4	251,002	3.07	0.443
14	8	251,002	3.07	0.09293
14	16	251,002	3.07	0.06832
(c) $P := Z\hat{P}$				
2	4	134,194	2.11	0.854
2	8	134,194	2.11	0.771
2	16	134,194	2.11	0.567
6	4	247,270	3.04	0.707
6	8	247,270	3.04	0.622
6	16	247,270	3.04	0.452
10	4	251,002	3.07	0.591
10	8	251,002	3.07	0.505
10	16	251,002	3.07	0.365
12	4	251,002	3.07	0.548
12	8	251,002	3.07	0.456
12	16	251,002	3.07	0.330
14	4	251,002	3.07	0.517
14	8	251,002	3.07	0.412
14	16	251,002	3.07	0.299

The results are presented in Table III. The behavior is similar to the finite element cases presented previously. The difference is that the operator complexity ‘OC’ is larger than one and can become unacceptably large, which is typical for the SA method applied to certain, such as scale-free, graph problems.

The respective results are also shown graphically in Figures 14, 15, and 16.

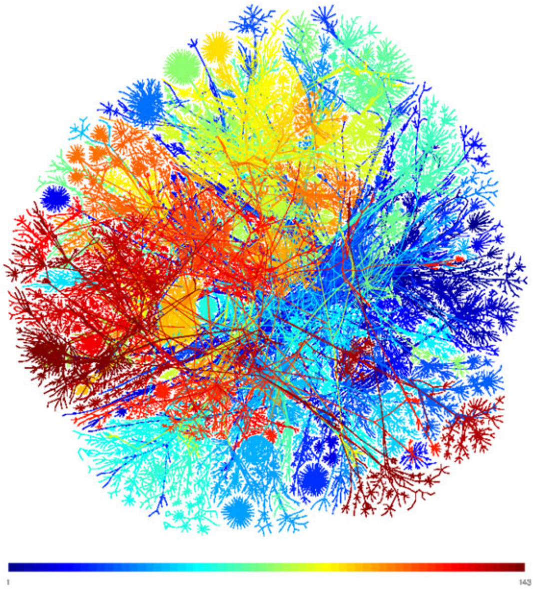
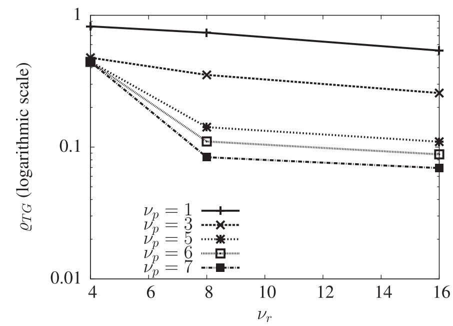
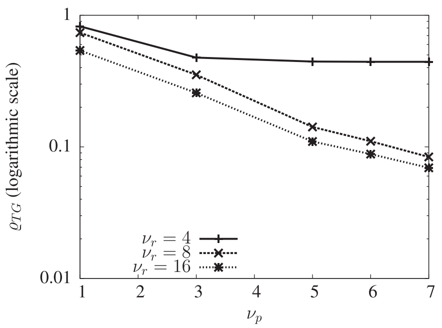


Figure 13. Agglomerates (in different colors) of edges.

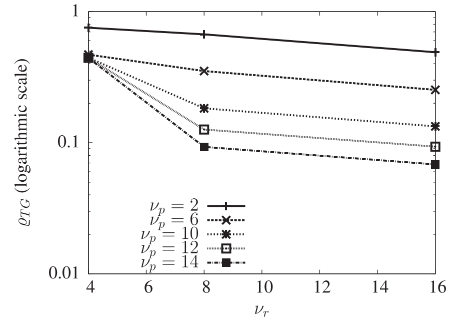


(a) Fixed ν_p and varying ν_r

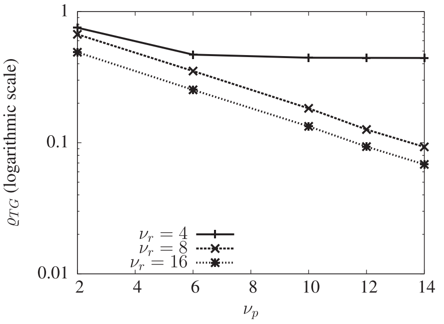


(b) Fixed ν_r and varying ν_p

Figure 14. Results for the graph Laplacian with $P := S^2\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.

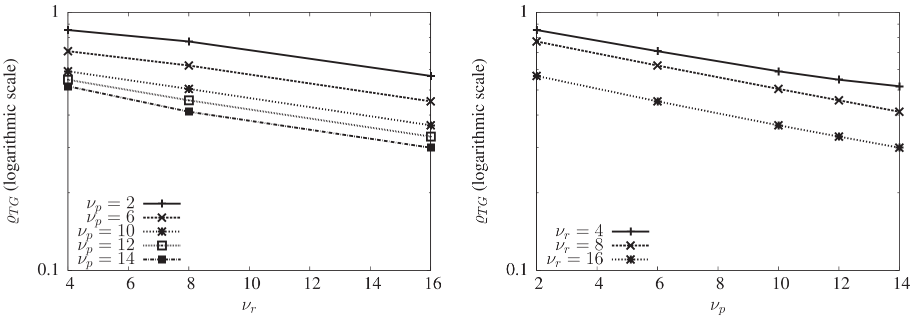


(a) Fixed ν_p and varying ν_r



(b) Fixed ν_r and varying ν_p

Figure 15. Results for the graph Laplacian with $P := S\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.



(a) Fixed ν_p and varying ν_r

(b) Fixed ν_r and varying ν_p

Figure 16. Results for the graph problem with $P := Z\hat{P}$. (a) fixed ν_p and varying ν_r; and (b) fixed ν_r and varying ν_p.

ACKNOWLEDGEMENTS

The numerical tests were run by Delyan Kalchev (presently at CU Boulder) using the SA spectral element agglomeration AMGe code that he developed during his summer internships at LLNL.

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Hu and Xu are partially supported by NSF Grant DMS-1217142.

The valuable and careful comments offered by anonymous referees are also gratefully acknowledged.

REFERENCES

1. Galvis J, Efendiev Y. Domain decomposition preconditioner for multiscale flows in high-contrast media. *Multiscale Modeling and Simulation* 2010; **8**:595–620.
2. Scheichl R, Vassilevski PS, Zikatanov L. Weak approximation properties of elliptic projections with functional constraints. *Multiscale Modeling and Simulation* 2011; **9**:1677–1699.
3. Brezina M, Vassilevski PS. Smoothed aggregation spectral element agglomeration AMG: $SA - \rho$AMGe. In *Large-Scale Scientific Computing, 8th International Conference, LSSC 2011, Sozopol, Bulgaria, June 6–10th, 2011. Revised Selected Papers*, vol. 7116, Lecture Notes in Computer Science. Springer: Berlin Heidelberg, 2012; 3–15.
4. Chartier T, Falgout R, Henson V, Jones J, Manteuffel T, McCormick S, Ruge J, Vassilevski PS. Spectral AMGe (ρAMGe). *SIAM Journal on Scientific Computing* 2003; **25**:1–26.
5. Chartier T, Falgout R, Henson V, Jones J, Manteuffel T, McCormick S, Ruge J, Vassilevski PS. Spectral element agglomerate AMGe. In *Domain Decomposition Methods in Science and Engineering XVI*, vol. 55, Lecture Notes in Computational Science and Engineering. Springer: Heidelberg, 2007; 515–524.
6. Lashuk I, Vassilevski PS. On some versions of the element agglomeration AMGe method. *Numerical Linear Algebra with Applications* 2008; **12**:595–620.
7. Falgout R, Vassilevski PS. On generalizing the algebraic multigrid framework. *SIAM Journal on Numerical Analysis* 2004; **42**:1669–1693.
8. Falgout R, Vassilevski PS, Zikatanov L. On two-grid convergence estimates. *Numerical Linear Algebra with Applications* 2005; **12**:471–494.
9. Zikatanov L. Two-sided bounds on the convergence rate of two-level methods. *Numerical Linear Algebra with Applications* 2008; **15**:439–454.
10. Axelsson O, Gustafsson I. Preconditioning and two-level multigrid methods of arbitrary degree of approximation. *Mathematics of Computation* 1983; **40**:219–242.
11. Vassilevski PS. *Multilevel Block Factorization Preconditioners, Matrix-based Analysis and Algorithms for Solving Finite Element Equations*. Springer: New York, 2008.
12. Hackbusch W. *Multi-Grid Methods and Applications*. Springer-Verlag: Berlin, 1985.
13. Xu J. Theory of multilevel method. *Ph.D. Thesis*, 1989.
14. Xu J, Zikatanov L. The method of alternating projections and the method of subspace corrections in Hilbert space. *Journal of American Mathematical Society* 2002; **15**:573–597.
15. Brezina M, Vaněk P, Vassilevski PS. An improved convergence analysis of smoothed aggregation algebraic multigrid. *Numerical Linear Algebra with Applications* 2012; **19**:441–469.
16. Kraus J, Vassilevski PS, Zikatanov L. Polynomial of best uniform approximation to $1/x$ and smoothing in two-level methods. *Computational Methods in Applied Mathematics* 2012; **12**:448–468.
17. Lashuk I, Vassilevski PS. The construction of Coarse de Rham complexes with improved approximation properties. *Computational Methods in Applied Mathematics* 2014; **14**:257–303.