Applications of Mathematics

Radek Tezaur; Petr Vaněk

Improved convergence estimate for a multiply polynomially smoothed two-level method with an aggressive coarsening

Applications of Mathematics, Vol. 63 (2018), No. 6, 629–641

Persistent URL: http://dml.cz/dmlcz/147561

IMPROVED CONVERGENCE ESTIMATE FOR A MULTIPLY POLYNOMIALLY SMOOTHED TWO-LEVEL METHOD WITH AN AGGRESSIVE COARSENING

Radek Tezaur, Stanford, Petr Vaněk, Plzeň

Received November 13, 2017. Published online November 23, 2018.

Abstract. A variational two-level method in the class of methods with an aggressive coarsening and a massive polynomial smoothing is proposed. The method is a modification of the method of Section 5 of Tezaur, Vaněk (2018). Compared to that method, a significantly sharper estimate is proved while requiring only slightly more computational work.

Keywords: two-level method; aggressive coarsening; smoothed aggregation; polynomial smoother; convergence analysis

MSC 2010: 65F10, 65M55

1. Introduction

We consider a modification of the algorithm defined in Theorem 5.2 of [5] that is used for an iterative solution of linear systems with a positive definite matrix, denoted throughout this paper by A. This algorithm is a smoothed prolongator variational two-level method with a multiple prolongator smoothing using an adequate smoother and an aggressive coarsening. We prove a radically better convergence estimate than in [5].

The presented method is targeted to highly parallel architectures. Unlike many domain decomposition methods [6] that can be, in a multigrid terminology [4], seen as two level multigrid methods with a small coarse space and a massive smoother based on local subdomain solvers, our method uses polynomial smoothers that are a sequence of Richardson iterations, each of which can be performed using up to n processors, where n is the size of A. This fine grain parallelism is not possible with local solver-based domain decomposition methods, which are faced with two conflicting objectives. In order to achieve parallel scalability, it is necessary to use a large number of subdomains and the size of the coarse space, by implication, must be large too. The cost of the coarse-level solver, both in CPU time and memory requirements, then grows quickly, since the parallelism in the coarse-level solver is limited.

By relying on polynomial smoothers instead of subdomain solvers, the method considered here has a lower cost per iteration, and as stated above, it is not limited in the parallelism on the fine level. This allows it to use a number of processors that keeps the cost of the coarse level and the fine level in balance. Furthermore, we will show in Section 3 that its convergence rate is independent of the mesh size and the coarse-space size, which is a property sought in domain decomposition methods for both practical and theoretical reasons.

A key assumption for the method of this paper is a special form of the weak approximation condition (Assumption 2.1 below). In Theorem 5.2 of [5], we have shown that, under this assumption, a radical qualitative acceleration effect is achieved for an approximation constant C_A of Assumption 2.1 that approaches zero. This is relatively reasonable, since the constant C_A of Assumption 2.1 can be made arbitrarily small when the (single) prolongator smoother S is a transformed Chebyshev polynomial in A of a sufficient degree (in applications, the constant is directly proportional to $1/(1+2\deg(S))^2$). Here, for the multiple polynomial prolongator smoother S^{k+1} (instead of S^k in [5]) and the same fine-level smoothing procedure, the radical acceleration effect is achieved for $C_A \approx 1$. The same effect is obtained for the method of Section 6 of [5] but only at a much higher computational cost. In fact, the convergence result of Section 6 is meant to be a mostly theoretical result and the method itself is not very practical. This is not the case here.

An attempt has been made to keep this paper self-contained so that the reader can make his way through it without being acquainted with [5]. Familiarity with [5] is, however, useful since it provides a much wider perspective.

2. The method and its abstract convergence theory

We consider a variational multigrid method with the prolongator $P\colon\ \mathbb{R}^m \to \mathbb{R}^n$, $m < n$ and a post-smoother with the error propagation operator $M\colon\ \mathbb{R}^n \to \mathbb{R}^n$. The error propagation of such two-level multigrid method is given by

$$E = M[I - P(P^{\mathrm{T}}AP)^{+}P^{\mathrm{T}}A].$$

Specifically, we investigate a polynomially accelerated variational multigrid with the multiply smoothed tentative prolongator $P = S^{k+1}p$, where $p\colon\ \mathbb{R}^m \to \mathbb{R}^n$ is a *tenta-*

tive prolongator and S is a polynomial in A of the form

$$S = (I - \alpha_1 A) \ldots (I - \alpha_d A). \tag{2.1}$$

In this case, the error propagation operator E and the error propagation operator of the A-symmetrized method, $E_s = EE^*$, are given by

$$E = S^k S_{A_S}(I - Q_A), \quad E_s = S^k S_{A_S}(I - Q_A) S_{A_S} S^k, \tag{2.2}$$

respectively, where

$$P = S^{k+1} p, \tag{2.3}$$

$$Q_A = P(P^{\mathrm{T}} A P)^+ P^{\mathrm{T}} A, \quad A_S = S^2 A, \quad S_{A_S} = I - \frac{\omega}{\lambda_{A_S}} A_S \tag{2.4}$$

with $\omega \in (0, 2)$ and λ_{A_S} an available upper bound of $\varrho(A_S)$. In other words, we investigate the variational multigrid with prolongator P given by (2.3) and the post-smoother with the error propagation operator $S_{A_S} S^k$, where S_{A_S} is given by (2.4) and S is a polynomial in A in the form (2.1). The operator E_s is the error propagation operator of its A-symmetrization.

The following assumption is crucial to proving rapid convergence of the method above.

Assumption 2.1. Let A be a symmetric, positive definite $n \times n$ matrix, S a polynomial in A of the form (2.1) and $p\colon \mathbb{R}^m \to \mathbb{R}^n$, $m < n$, an injective operator, and λ_{A_S} an available upper bound of $\varrho(S^2 A)$. Assume

$$\exists\, C_A > 0\colon \Big(\forall\, \mathbf{e} \in \mathbb{R}^n\ \exists\, \mathbf{v} \in \mathbb{R}^m\colon\ \|\mathbf{e} - p\mathbf{v}\|^2 \leqslant \frac{C_A}{\lambda_{A_S}} \|\mathbf{e}\|_A^2\Big). \tag{2.5}$$

Remark 2.1. It is desirable to choose the smoother S so that λ_{A_S} is as small as possible. Indeed, the smaller is the upper bound λ_{A_S}, the easier it becomes to satisfy Assumption 2.1 with a good (small) constant C_A. Let λ be an available upper bound of $\varrho(A)$. Then, the linearly transformed Chebyshev polynomial in A given by

$$q(A) = \Big(1 - \frac{1}{r_1} A\Big) \ldots \Big(1 - \frac{1}{r_d} A\Big), \quad r_i = \frac{\lambda}{2}\Big(1 - \cos \frac{2i\pi}{2d+1}\Big) \tag{2.6}$$

minimizes λ_{A_S}, and Lemma 4.4 of [2] shows that, for $S = q(A)$,

$$\lambda_{A_S} = \frac{\lambda}{(1 + 2\deg(q))^2} \geqslant \varrho(S^2 A) \tag{2.7}$$

and $\varrho(S) \leqslant 1$. Assumption 2.1 then becomes

$$\exists\, C_A > 0\colon \Bigl(\forall\, \mathbf{e} \in \mathbb{R}^n\ \exists\, \mathbf{v} \in \mathbb{R}^m\colon\ \|\mathbf{e} - p\mathbf{v}\|^2 \leqslant \frac{C_A(1+2\deg(S))^2}{\lambda}\|\mathbf{e}\|_A^2\Bigr). \tag{2.8}$$

R e m a r k 2.2. For a given prolongator p, the condition (2.8) in the form

$$\forall\, \mathbf{u} \in \mathbb{R}^n\ \exists\, \mathbf{v} \in \mathbb{R}^m\colon\ \|\mathbf{u} - p\mathbf{v}\| \leqslant \frac{C(h,H)}{\lambda}\|\mathbf{u}\|_A, \tag{2.9}$$

is usually verified from properties of finite element spaces, understood as subspaces of Hilbert-Sobolev spaces H^α. The constant $C(h,H)$ cannot be expected to be uniform with respect to the degree of coarsening H/h, where h and H are the mesh sizes characterizing the fine and the coarse space, respectively. For the case of interest,

$$C = C'\Bigl(\frac{H}{h}\Bigr)^2, \tag{2.10}$$

see Section 3, where C' is independent of h and H. The smoother S of a sufficient degree serves as a means to eliminate this non-uniformity: it follows from (2.9) that (2.8) holds with

$$C_A = \frac{C}{(1+2\deg(S))^2}. \tag{2.11}$$

and is uniform with respect to h and H for $\deg(S) \geqslant cH/h$, $c > 0$.

The following theorem states the convergence of the method.

Theorem 2.1. *Let A be a symmetric, positive definite $n\times n$ matrix and p a full-rank $n\times m$ matrix, $m < n$. Furthermore, let S be a polynomial in A of the form (2.1) such that $\varrho(S) \leqslant 1$, P, S_{A_S} and let Q_{A_S} be given by (2.4) and λ_{A_S} be an available upper bound of $\varrho(A_S)$. Under Assumption 2.1, the error propagation operators E and E_s given by (2.2) satisfy $\|E\|_A \leqslant \gamma(C_A)$ and $\|E_s\|_A \leqslant \gamma^2(C_A)$, where the function γ is defined as*

$$\gamma^2(C_A) = \begin{cases} \Bigl(\dfrac{C_A}{\omega(2-\omega)}\dfrac{k}{k+1}\Bigr)^k \dfrac{1}{k+1} & \text{for } C_A \leqslant \dfrac{k+1}{k}\omega(2-\omega),\\[2ex] 1 - \dfrac{\omega(2-\omega)}{C_A} & \text{for } C_A > \dfrac{k+1}{k}\omega(2-\omega). \end{cases} \tag{2.12}$$

The function γ^2 is continuous for $C_A \in [0,\infty)$.

P r o o f. Let $\mathbf{e} \in \mathbb{R}^n$ be the error on the entry of an iteration, $\mathbf{e}_1 = (I - Q_A)\mathbf{e}$, and $\mathbf{e}_2 = S^{k-1}\mathbf{e}_1$. Clearly $E\mathbf{e} = S_{A_S} S\mathbf{e}_2 = S_{A_S} S^k \mathbf{e}_1$. If $\mathbf{e}_1$ or $\mathbf{e}_2 \in \operatorname{Ker}(S)$, $E\mathbf{e} = \mathbf{0}$ and the inequalities underlying the operator norm estimates for $\|E\|_A$ and $\|E_s\|_A$ hold trivially. Therefore, assume the nontrivial case $\mathbf{e}_1, \mathbf{e}_2 \notin \operatorname{Ker}(S)$.

Since $I - Q_A$ is an A-orthogonal projection, we get $\|\mathbf{e}_1\|_A = \|(I - Q_A)\mathbf{e}\|_A \leqslant \|\mathbf{e}\|_A$ and therefore,

$$\frac{\|E\mathbf{e}\|_A}{\|\mathbf{e}\|_A} = \frac{\|S_{A_S} S\mathbf{e}_2\|_A}{\|\mathbf{e}_2\|_A} \frac{\|S^{k-1}\mathbf{e}_1\|_A}{\|\mathbf{e}_1\|_A} \frac{\|\mathbf{e}_1\|_A}{\|\mathbf{e}\|_A} \leqslant \frac{\|S_{A_S} S\mathbf{e}_2\|_A}{\|\mathbf{e}_2\|_A} \frac{\|S^{k-1}\mathbf{e}_1\|_A}{\|\mathbf{e}_1\|_A}. \tag{2.13}$$

Furthermore, the A-symmetry of S and the Cauchy-Schwarz inequality yield that for any $\mathbf{x} \notin \operatorname{Ker}(S)$: $\|S\mathbf{x}\|_A^2 = \langle S\mathbf{x}, S\mathbf{x}\rangle_A = \langle S^2\mathbf{x}, \mathbf{x}\rangle_A \leqslant \|S^2\mathbf{x}\|_A \|\mathbf{x}\|_A$, which can be written equivalently as

$$\frac{\|S\mathbf{x}\|_A}{\|\mathbf{x}\|_A} \leqslant \frac{\|S^2\mathbf{x}\|_A}{\|S\mathbf{x}\|_A}. \tag{2.14}$$

As a consequence,

$$\frac{\|S\mathbf{e}_2\|_A}{\|\mathbf{e}_2\|_A} = \frac{\|S^k\mathbf{e}_1\|_A}{\|S^{k-1}\mathbf{e}_1\|_A} \geqslant \frac{\|S^{k-1}\mathbf{e}_1\|_A}{\|S^{k-2}\mathbf{e}_1\|_A} \geqslant \ldots \geqslant \frac{\|S\mathbf{e}_1\|_A}{\|\mathbf{e}_1\|_A}$$

and therefore,

$$\frac{\|S^{k-1}\mathbf{e}_1\|_A}{\|\mathbf{e}_1\|_A} = \prod_{i=1}^{k-1} \frac{\|S^i\mathbf{e}_1\|_A}{\|S^{i-1}\mathbf{e}_1\|_A} \leqslant \Big(\frac{\|S\mathbf{e}_2\|_A}{\|\mathbf{e}_2\|_A}\Big)^{k-1}.$$

Substituting this bound into (2.13) gives the estimate

$$\frac{\|E\mathbf{e}\|_A}{\|\mathbf{e}\|_A} \leqslant \frac{\|S_{A_S} S\mathbf{e}_2\|_A}{\|\mathbf{e}_2\|_A} \Big(\frac{\|S\mathbf{e}_2\|_A}{\|\mathbf{e}_2\|_A}\Big)^{k-1}. \tag{2.15}$$

Next, we proceed by estimating the first fraction on the right-hand side of (2.15), and use the argument by Achi Brandt [1] based on the orthogonality property known from the proof of Céa's lemma [3]. From the definition (2.4) and by using the inequality $|A_S\mathbf{x}|_{A_S}^2 \leqslant \lambda_{A_S} \|A_S\mathbf{x}\|^2$, ($|\cdot|_{A_S} = \langle\cdot,\cdot\rangle_{A_S}^{1/2} = \langle A_S\cdot,\cdot\rangle^{1/2}$) we obtain

$$\begin{aligned} \|SS_{A_S}\mathbf{e}_2\|_A^2 = |S_{A_S}\mathbf{e}_2|_{A_S}^2 &= |\mathbf{e}_2|_{A_S}^2 - 2\frac{\omega}{\lambda_{A_S}}\|A_S\mathbf{e}_2\|^2 + \Big(\frac{\omega}{\lambda_{A_S}}\Big)^2 |A_S\mathbf{e}_2|_{A_S}^2 \\ &\leqslant |\mathbf{e}_2|_{A_S}^2 - 2\frac{\omega}{\lambda_{A_S}}\|A_S\mathbf{e}_2\|^2 + \frac{\omega^2}{\lambda_{A_S}}\|A_S\mathbf{e}_2\|^2 \\ &\leqslant \Big(1 - \frac{\omega(2-\omega)}{\lambda_{A_S}} \frac{\|A_S\mathbf{e}_2\|^2}{|\mathbf{e}_2|_{A_S}^2}\Big)\Big(\frac{|\mathbf{e}_2|_{A_S}}{\|\mathbf{e}_2\|_A}\Big)^2 \|\mathbf{e}_2\|_A^2. \end{aligned} \tag{2.16}$$

In order to get the lower bound of the fraction $\|A_S\mathbf{e}_2\|^2/|\mathbf{e}_2|^2_{A_S}$, we proceed as follows (using Céa's orthogonality trick): We denote $t = \|S\mathbf{e}_2\|_A/\|\mathbf{e}_2\|_A$. Since $\varrho(S) \leqslant 1$ and $\mathbf{e}_2 \notin \operatorname{Ker}(S)$, we get $t \in (0,1]$. Let $\mathbf{v} = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{e}_2 - p\mathbf{x}\|$. Since Q_A is the A-orthogonal projection onto $\operatorname{Range}(P) = \operatorname{Range}(S^{k+1}p)$, $I - Q_A$ is the A-orthogonal projection onto $\operatorname{Range}(S^{k+1}p)^{\perp_A} = \operatorname{Ker}(p^{\mathrm{T}}S^{k+1}A) = \operatorname{Ker}(p^{\mathrm{T}}AS^{k+1})$, and we get $\mathbf{e}_2 = S^{k-1}\mathbf{e}_1 \in \operatorname{Ker}(p^{\mathrm{T}}AS^2) = \operatorname{Ker}(p^{\mathrm{T}}S^2A) = \operatorname{Range}(p)^{\perp_{A_S}}$. Then, by Assumption 2.1 and the Cauchy-Schwarz inequality $\|S\mathbf{e}_2\|^2_A = \langle A_S\mathbf{e}_2, \mathbf{e}_2\rangle = \langle A_S\mathbf{e}_2, \mathbf{e}_2 - p\mathbf{v}\rangle \leqslant \|A_S\mathbf{e}_2\|\|\mathbf{e}_2 - p\mathbf{v}\| \leqslant \|A_S\mathbf{e}_2\|\sqrt{C_A/\lambda_{A_S}}\|\mathbf{e}_2\|_A = \|A_S\mathbf{e}_2\|\sqrt{C_A/\lambda_{A_S}}\|S\mathbf{e}_2\|_A/t$. Dividing this inequality by $\|S\mathbf{e}_2\|_A$ and squaring the result yields the coercivity bound

$$\Big(\frac{\|A_S\mathbf{e}_2\|}{|\mathbf{e}_2|_{A_S}}\Big)^2 \geqslant t^2\frac{\lambda_{A_S}}{C_A}. \tag{2.17}$$

Substituting this bound into (2.16) and using (2.15) and the trivial identity $\|S\mathbf{e}_2\|_A = |\mathbf{e}_2|_{A_S}$ yields

$$\frac{\|E\mathbf{e}\|^2_A}{\|\mathbf{e}\|^2_A} \leqslant t^{2k}\Big(1 - t^2\frac{\omega(2-\omega)}{C_A}\Big). \tag{2.18}$$

Since $t \in (0,1]$, we have

$$\frac{\|E\mathbf{e}\|^2_A}{\|\mathbf{e}\|^2_A} \leqslant \max_{\xi\in[0,1]} \xi^k\Big(1 - \xi\frac{\omega(2-\omega)}{C_A}\Big). \tag{2.19}$$

By inspecting the function values at the interval ends 0 and 1 and the local maxima on the interval $(0,1)$, we finally obtain

$$\|E\mathbf{e}\|^2_A \leqslant \gamma^2(C_A)\|\mathbf{e}\|^2_A,$$

where γ is given by (2.12), proving the bound for $\|E\|_A$. The bound for $E_s = EE^*$ follows by $\|EE^*\|_A \leqslant \|E\|_A\|E^*\|_A = \|E\|^2_A$. □

Corollary 2.1. *Let A be a symmetric, positive definite $n \times n$ matrix, p a full-rank $n \times m$ matrix, $m < n$. We assume $S = q(A)$ with q given by* (2.6). *We assume P, S_{A_S} and Q_{A_S} are given as in* (2.4) *and λ_{A_S} by* (2.7). *Under assumption* (2.8) *the error propagation operators E and E_s given by* (2.2) *satisfy $\|E\|_A \leqslant \gamma(C_A)$ and $\|E_s\|_A \leqslant \gamma^2(C_A)$ with function γ given by* (2.12).

Proof. By Remark 2.1, for $S = q(A)$ we have $\varrho(S) \leqslant 1$, λ_{A_S} can be chosen as in (2.7) and Assumption 2.1 becomes (2.8). The proof now follows by Theorem 2.1. □

R e m a r k 2.3. We compare the convergence rate of the method of [5], Theorem 5.2 with that of the method analyzed here. Theorem 5.2 of [5] proves that the norm of the error propagation operators of the symmetrized method of Theorem 5.2 of [5] is bounded by

$$\|E_s\|_A \leqslant \gamma_2^2(C_A) = \begin{cases} \dfrac{C_A^k}{(C_A+\omega(2-\omega))^k} & \text{for } C_A \in \Big[0, \dfrac{\omega(2-\omega)}{k-1}\Big), \\ \dfrac{1}{k}\Big(\dfrac{k-1}{\omega(2-\omega)k}C_A\Big)^{k-1} & \text{for } C_A \in \Big[\dfrac{\omega(2-\omega)}{k-1}, \omega(2-\omega)\dfrac{k}{k-1}\Big), \\ 1-\dfrac{\omega(2-\omega)}{C_A} & \text{for } C_A \in \Big[\omega(2-\omega)\dfrac{k}{k-1}, \infty\Big). \end{cases}$$

The kth power acceleration effect can be observed only for $C_A \to 0$. For the method analyzed here, this effect persists for much larger values of C_A. In particular, for the optimal value of $\omega = 1$, it occurs up to $C_A = 1 + 1/k$ (cf. (2.12)). Based on the discussion of Remark 2.2, C_A decreases for the increasing $d = \deg(S)$ as $C_A = C/(1+2d)^2$ assuming (2.9) is satisfied. Figure 2.1 depicts a comparison of the convergence rates of the method of this paper (denoted as Method 2 in the figure) over the method of [5] (denoted as Method 1 in the figure) for $\omega = 1$, $k = 2$ and three different degrees $d = 4, 8, 16$ of the polynomial smoother S. A significant improvement in the practical range of convergence rates is noted.

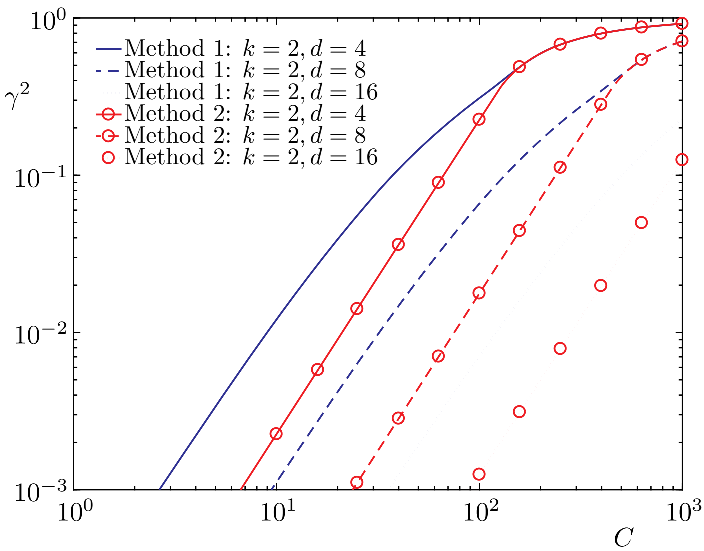

Figure 2.1. Convergence rates of the method of this paper (Method 2) versus the method of [5] (Method 1) for three different degrees d of the polynomial smoother S, where C is a constant of (2.9).

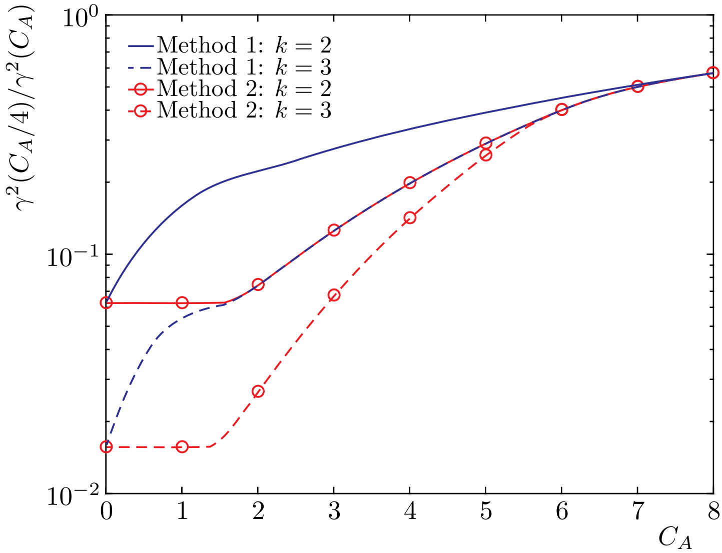

Figure 2.2. Convergence rate improvement achieved by increasing twice the degree of the polynomial smoother S for the method of this paper (Method 2) versus the method of [5] (Method 1).

Let $C_A \equiv C_{A,d} = C/(1+2d)^2$ denote the constant for the smoothing polynomial of degree d and $C_{A,2d} = C/(1+2\cdot 2d)^2$ the constant for the smoothing polynomial of a double degree. Then,

$$C_{A,2d} = \left(\frac{1+2d}{1+4d}\right)^2 C_{A,d} \approx \frac{C_{A,d}}{4} \tag{2.20}$$

for a large enough d. Assuming $C_{A,d} \leqslant \omega(2-\omega)(k+1)/k$, we then obtain by (2.12),

$$\gamma^2(C_{A,2d}) \approx \frac{\gamma^2(C_{A,d})}{4^k}.$$

We recall that $\|E_s\|_A \leqslant \gamma^2(C_A)$. Thus, for a polynomial S of a double degree, the estimate of the rate of convergence decreases almost (and exactly asymptotically) by a factor of $1/4^k$. In [5], this was possible only for $C_A \to 0$. Here it holds, assuming optimal $\omega = 1$, for $C_A \leqslant (k+1)/k > 1$. Figure 2.2 depicts the acceleration effect as a function of $C_{A,d}$ for both methods, $\omega = 1$, and $k = 2, 3$. Again, the improvement of the method of this paper (denoted as Method 2 in the figure) over the method of [5] (denoted as Method 1 in the figure) for $C_{A,d} \approx 1$ is prominently visible.

In light of the discussion in the remark above, we further illustrate the beneficial effect of doubling the degree of the polynomial smoother using the convergence rate estimate (2.12) and the relationship (2.20). Figure 2.3 compares the convergence rate estimate of a method that uses the smoothing polynomial of degree $2d$ and the convergence rate estimate of a method that instead performs two iterations with the the smoothing polynomial of degree d in each iteration (two iterations of the method with the degree of the smoothing polynomial d and a single iteration with the smoothing polynomial of the degree $2d$ are approximately equally expensive, since they use approximately the same number of Richardson smoothing steps and the bandwidth of the coarse-level matrix is, assuming a reasonable numbering, about twice larger, making the backward substitution about twice more expensive). It shows that doubling the degree of the smoothing polynomial *pays off dramatically* (in terms of the time in the iteration stage) unless the method with the polynomial of degree d already exhibits extremely fast convergence (the rate of convergence of 0.0003, approximately).

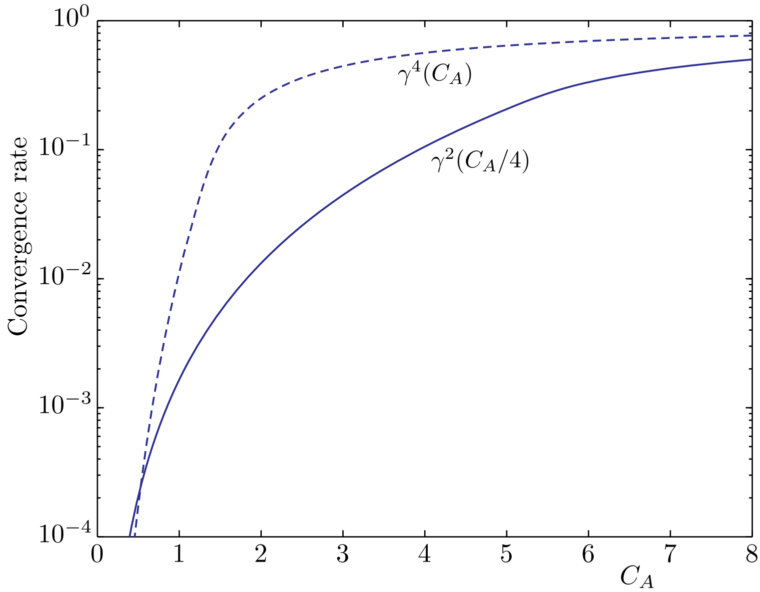

Figure 2.3. Convergence rate comparison of the method with the degree of the smoother $2d$ (solid line) and of two iterations of the method with the degree of the smoother d (dashed line) for $k = 3$.

Next, Figure 2.4 demonstrates how much can be saved in terms of the amount of computational work during the iteration by using the smoothing polynomial of the double degree $2d$ instead of d, and quadruple degree $4d$ instead of d, for $k = 3$. Since the computational work of one iteration is proportional to the degree of the polynomial, the ratios of the work can be estimated using the ratios shown in the figure.

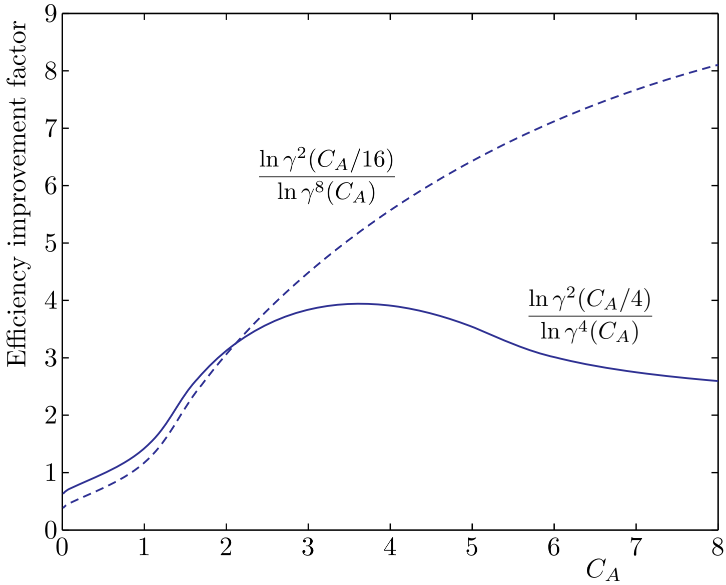

Figure 2.4. Ratio of the amount of work during the iteration needed to satisfy the stopping criterion of the methods with the degree of the smoothing polynomial $2d$ and d (solid line), and the methods with the degree of the smoothing polynomial $4d$ and d (dashed line); $k = 3$.

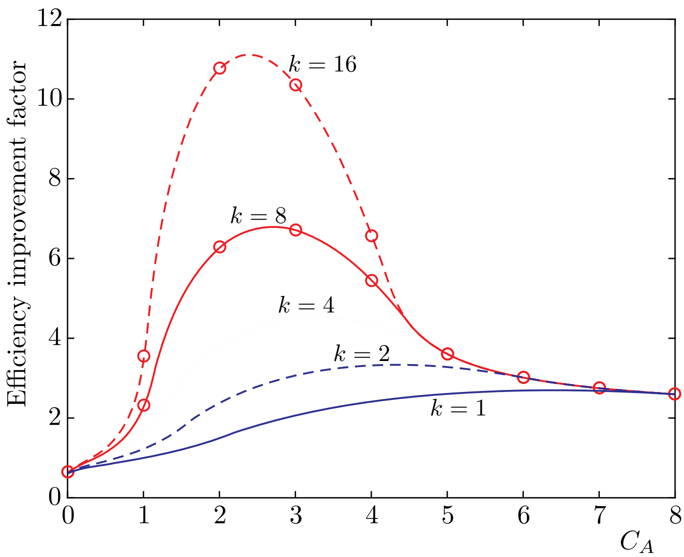

Figure 2.5. Ratio of the amount of work during the iterations needed to satisfy the stopping criterion of the method with the degree of the smoothing polynomial $2d$ and d for $k = 1, 2, 4, 8$ and $k = 16$. The number of Richardson smoothing steps is $1 + (2 + k)d$ and the bandwidth of the coarse-level matrix is proportional to $k + 1$.

Figure 2.5 addresses the computational gain achieved by using the smoothing polynomial of the double degree $2d$ instead of d for $k = 1, 2, 4, 8$ and 16. The beneficial effect increases dramatically with the increasing value of k. However, the high values of k may cause an increase of the computational complexity of the coarse level setup (coarse-level matrix factorization) that may outweigh the gains during the iteration. (When using a polynomial of degree $2d$, the bandwidth of the coarse level matrix increases, for a reasonable numbering, about twice, making backward substitution twice more expensive and the Choleski factorization four times more expensive. Our calculation is therefore valid for the cost of the iteration but not for the cost of the Choleski factorization. This, however, does not matter as far as the coarse-level problem is sufficiently small, see Remark 2.4.) The improvement of the convergence rate is reported in Fig. 2.6.

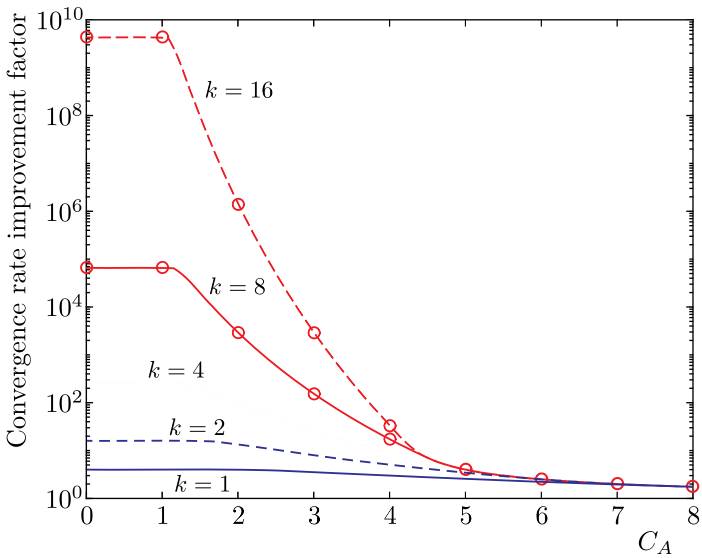

Figure 2.6. Improvement of the rate of convergence when using the double degree of the smoothing polynomial. The value being reported is $\gamma^2(k, C_A)/\gamma^2(k, C_A/4)$.

R e m a r k 2.4. Let us comment on the coarse-space size. This method is targeted to solving problems arising from discretizations of $3D$ linear elasticity. In this case there are 3 degrees of freedom per node on the fine level and 6 degrees of freedom per supernode on the coarse level [8]. For a problem with one million of degrees of freedom and aggregates $10 \times 10 \times 10$ nodes, we obtain a coarse-space of a moderate size of 2000 degrees of freedom. In this case, $d = 4$ is a reasonable initial choice before considering the acceleration (this initial choice has certain "geometrical grounds"; on a model geometry, only the coarse space basis functions that correspond to the immediately adjacent aggregates overlap by their supports, and it is the largest such d). For aggregates of size $15 \times 15 \times 15$ nodes, we get the coarse-space of a negligible size of about 600 degrees of freedom, and $d = 7$ is a reasonable initial choice. For the resulting coarse-space of about 600 degrees of freedom (and within reasonable limits, for a coarse problem of size 2000 also), the increase of the bandwidth of the coarse-level matrix, when performing the Choleski decomposition, can be safely ignored.

3. Application to a method with an aggressive coarsening based on the unknowns aggregation and a massive polynomial smoothing

To make this short paper self-contained, we apply the results of Section 2 to a system arising from a finite element discretization of a second-order elliptic PDE. Up to a different convergence result inherited from Section 2, this section follows closely the final section of [5].

Throughout this section, the prolongator p is assumed to be constructed by a generalized unknowns aggregation method (see [7]). The brief introduction to the unknowns aggregation coarsening can be found in [2].

The resulting method features an aggressive coarsening (with a small coarse-space) based on the unknowns aggregation, balanced by a massive polynomial smoothing (the multiple Richardson iteration). They are optimal in the following sense: for a second order elliptic problem discretized on a mesh with the characteristic mesh size h and a coarse-space characterized by the resolution H, they exhibit a coarse-space size independent rate of convergence for the cost of $O(H/h)$ elementary smoothing steps. The coarse-level matrix is sparse if the aggregates have a reasonably compact shape and approximately the same size.

The theory of the previous section can be readily applied provided that the prolongator satisfies a version of the weak approximation condition required by the method. In order to achieve coarse space size and problem size independent convergence for the considered methods, it is necessary to establish that the prolongator satisfies

$$\exists\, C > 0\colon \Bigl(\forall\, \mathbf{e} \in \mathbb{R}^n\ \exists\, \mathbf{v} \in \mathbb{R}^m\colon\ \|\mathbf{e} - p\mathbf{v}\|^2 \leqslant \frac{C}{\lambda}\Bigl(\frac{H}{h}\Bigr)^2 \|\mathbf{e}\|_A^2\Bigr) \tag{3.1}$$

with a constant C independent of h and H. Here, h is a characteristic element size of the fine-level discretization (assuming the quasi-uniformity of the mesh), H is a characteristic diameter of the aggregates (understood as a set of finite element nodal points) and λ is an available upper bound of $\varrho(A)$. For a scalar elliptic second order problem, (3.1) was proved in [7]. For the case of linear elasticity in 3D, the reader is referred to [8].

We summarize the results in the following uniform theorem:

Theorem 3.1. *Let A be a symmetric, positive definite $n\times n$ matrix, p a full-rank $n\times m$ matrix, $m<n$. Assume the prolongator p satisfies* (3.1), *the smoother S is given by* (2.6), *its degree d satisfies $d\geqslant c_dH/h$ with $c_d>0$ and λ_{A_S} is given by* (2.7). *Then the error propagation operators E and E_s in* (2.2) *satisfy*

$$\|E\|_A\leqslant\gamma\Big(\frac{C_{(3.1)}}{4c_d^2}\Big),\quad \|E_s\|_A\leqslant\gamma^2\Big(\frac{C_{(3.1)}}{4c_d^2}\Big) \tag{3.2}$$

with function γ given by (2.12).

Proof. Due to (3.1) and $d=\deg(S)\geqslant c_dH/h$, (2.8) holds with $C_A=C_{(3.1)}/(4c_d^2)$. The proof now follows by Corollary 2.1. □

References

[1] *A. Brandt*: Algebraic multigrid theory: The symmetric case. Appl. Math. Comput. *19* (1986), 23–56. zbl MR doi

[2] *J. Brousek, P. Franková, M. Hanuš, H. Kopincová, R. Kužel, R. Tezaur, P. Vaněk, Z. Vastl*: An overview of multilevel methods with aggressive coarsening and massive polynomial smoothing. ETNA, Electron. Trans. Numer. Anal. *44* (2015), 401–442. zbl MR

[3] *P. G. Ciarlet*: The Finite Element Method for Elliptic Problems. Studies in Mathematics and Its Applications 4, North-Holland Publishing Company, Amsterdam, 1978. zbl MR doi

[4] *W. Hackbusch*: Multi-Grid Methods and Applications. Springer Series in Computational Mathematics 4, Springer, Berlin, 1985. zbl MR doi

[5] *R. Tezaur, P. Vaněk*: Improved convergence bounds for two-level methods with an aggressive coarsening and massive polynomial smoothing. ETNA, Electron. Trans. Numer. Anal. *48* (2018), 264–285. zbl MR doi

[6] *A. Toselli, O. Widlund*: Domain Decomposition Methods—Algorithms and Theory. Springer Series in Computational Mathematics 34, Springer, Berlin, 2005. zbl MR doi

[7] *P. Vaněk, M. Brezina, J. Mandel*: Convergence of algebraic multigrid based on smoothed aggregation. Numer. Math. *88* (2001), 559–579. zbl MR doi

[8] *P. Vaněk, M. Brezina, R. Tezaur*: Two-grid method for linear elasticity on unstructured meshes. SIAM J. Sci. Comput. *21* (1999), 900–923. zbl MR doi

Authors' addresses: *Radek Tezaur*, Department of Aeronautics and Astronautics, Stanford University, Durand Building, 496 Lomita Mall, Stanford, CA 94305, USA, e-mail: `rtezaur@stanford.edu`; *Petr Vaněk*, Department of Mathematics, University of West Bohemia, Univerzitní 22, 306 14 Plzeň, Czech Republic, e-mail: `petrvanek09@seznam.cz`.