LLNL-JRNL-559551

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Polynomial of Best Uniform Approximation to 1/x and Smoothing in Two-Level Methods

J. Kraus, P. S. Vassilevski, L. T. Zikatanov

June 1, 2012

Computational Methods in Applied Mathematics

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

POLYNOMIAL OF BEST UNIFORM APPROXIMATION TO x^{-1} AND SMOOTHING IN TWO-LEVEL METHODS

JOHANNES K. KRAUS, PANAYOT S. VASSILEVSKI, AND LUDMIL T. ZIKATANOV

Abstract. In this paper, we derive a three-term recurrence relation for computing the polynomial of best approximation in the uniform norm to x^{-1} on a finite interval with positive endpoints. As application, we consider two-level methods for scalar elliptic partial differential equation (PDE), where the relaxation on the fine grid uses the aforementioned polynomial of best approximation. Based on a new smoothing property of this polynomial smoother that we prove, combined with a proper choice of the coarse space, we obtain as a corollary, that the convergence rate of the resulting two-level method is uniform with respect to the mesh parameters, coarsening ratio and PDE coefficient variation.

1. Introduction

The polynomial of best approximation in uniform norm to x^{-1} on a finite interval can be found in different forms in many classical texts on approximation theory, for example, see [1, p. 33, Equation(4.25)], [2, Exercise 1.20]. In fact, the approximating polynomial for $\dfrac{1}{t-a}$, $a > 1$, has already been discovered by Chebyshev in 1887, see [3].

As an application, we study two-level methods with smoothers based on this polynomial of best approximation to x^{-1} on a finite interval $[\lambda_{\min}, \lambda_{\max}]$, $0 < \lambda_{\min} < \lambda_{\max}$, in the $\|\cdot\|_\infty$ (uniform) norm. We derive several results important for applications: a three-term recurrence relation for constructing these polynomials; error estimates; the positivity and monotonicity of the sequence of polynomials of best approximation, and we use these results in designing components of two-level methods. We show a major smoothing property of the polynomial and as a corollary, based on an abstract two-level estimate we derive two-level (TL or TG) convergence estimates in the case of discretized elliptic PDE with heterogeneous coefficients. The estimate explicitly depends on the degree of the polynomial (or on the range of the spectrum which needs to be resolved by the smoother) and we prove that if coarse spaces with stability and approximation properties that are robust with respect coefficient variation are used, then the two-level methods with polynomial smoothers based on the polynomial of best approximation to $1/x$ are robust with respect to the variation in the coefficients of the PDE. Several examples of coarse spaces that provide the required contrast independent approximation property are available in the literature, cf., e.g., [4], [5], and earlier [6] as modified recently in [7]).

Date: May 7, 2012.

The work of the first author has been supported by the Austrian Science Fund, Grant P22989-N18. The work of the second author is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The work of the third author is supported in part by the National Science Foundation, DMS-0810982 and U.S. Department of Energy (Lawrence Livermore National Laboratory, subcontract B595949).

The paper is organized as follows. In Section 2 we derive a three-term recurrence relation for the polynomial of best approximation to x^{-1}. Several properties of the sequence of polynomials of best approximation to $1/x$ are shown in Section 3. In Section 4 we discuss and prove the major smoothing property of the polynomial, which explicitly involves the polynomial degree and we use it an abstract two-level convergence result. As a corollary, we derive an estimate for the convergence rate in case of finite element discretization of scalar elliptic PDE with coarse spaces that provide contrast independent approximation resulting in contrast independent two-grid convergence. This convergence behavior is illustrated also with numerical tests in Section 5.

2. Best polynomial approximation to x^{-1} in uniform norm

We begin with notation and some simple and well known definitions related to Chebyshev polynomials. We consider a finite interval, $[\lambda_{\min}, \lambda_{\max}]$, with $0 < \lambda_{\min} < \lambda_{\max} < \infty$. We denote

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad \sigma = \frac{1}{\lambda_{\max} - \lambda_{\min}}, \quad a = \frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}} = \frac{\kappa + 1}{\kappa - 1}. \tag{2.1}$$

Note that $a > 1$ and $\sigma > 0$. The change of variables

$$t = \frac{2}{\lambda_{\max} - \lambda_{\min}} \left(x - \frac{\lambda_{\max} + \lambda_{\min}}{2} \right) = 2\sigma x - a,$$

maps the interval $[\lambda_{\max}, \lambda_{\min}]$ to $[-1, 1]$. The inverse map is

$$x = \frac{1}{2\sigma}(t + a), \quad \text{and} \quad \frac{1}{x} = \frac{2\sigma}{t + a}.$$

We thus aim to find the polynomial of degree less than or equal to m of best approximation in the norm $\|\cdot\|_{\infty,[-1,1]}$ of $f(t) = \dfrac{1}{t+a}$, $a > 1$. We note that if $Q_m(t)$ is the polynomial of best approximation to $1/(t + a)$ on $[-1, 1]$, and the error of approximation is

$$E_{[-1,1]} = \min_{Q \in \mathcal{P}_m} \left\| \frac{1}{t+a} - Q \right\|_{L_\infty[-1,1]},$$

then

$$q_m(x) := 2\sigma Q_m(2\sigma x - a), \quad \text{and} \quad E = \min_{q \in \mathcal{P}_m} \left\| \frac{1}{x} - q \right\|_{L_\infty[\lambda_{\max}, \lambda_{\min}]} = 2\sigma E_{[-1,1]} \tag{2.2}$$

are the polynomial of best approximation in L^∞-norm on $[\lambda_{\min}, \lambda_{\max}]$ and the error of approximation, respectively.

We denote the (first kind) Chebyshev polynomial of degree k by T_k. For $T_k(\xi) \in \mathcal{P}_k$ we have

$$T_k(\xi) = \frac{1}{2}\left[(\xi + \sqrt{\xi^2 - 1})^k + (\xi + \sqrt{\xi^2 - 1})^{-k}\right] = \frac{1}{2}\left[(\xi + \sqrt{\xi^2 - 1})^k + (\xi - \sqrt{\xi^2 - 1})^k\right].$$

We recall that

$$T_k(t) = \cos k \arccos(t), \quad t \in [-1, 1]$$

and denote

$$\delta := a - \sqrt{a^2 - 1} = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}, \quad \eta = -\delta. \tag{2.3}$$

Evidently, $0 \le \delta < 1$, $\delta^{-1} = a + \sqrt{a^2-1}$, $\eta < 0$ and $\delta = |\eta|$.

With this notation in hand, we have the following identities,

$$a = -\frac{1}{2}(\eta + \eta^{-1}), \quad \frac{1}{t+a} = \frac{2}{2t - \eta - \eta^{-1}}, \tag{2.4}$$

and directly from the expression for $T_k(\xi)$ given above, we also have

$$T_k(a) = \tfrac{1}{2}(-1)^k(\eta^k + \eta^{-k}), \quad T_k(-a) = \tfrac{1}{2}(\eta^k + \eta^{-k}). \tag{2.5}$$

2.1. **Approximation error and three-term recurrence.** Next, in Theorem 2.1 we give a representation of the best polynomial approximation to $\dfrac{1}{t+a}$ in the L^∞-norm on the interval $[-1,1]$. The proof of this theorem is given in the appendix, and amounts to showing that the form given in (2.6) is equivalent to the one given in [1, p. 33, Equation (4.25)].

Theorem 2.1. *Let $m \ge 1$ be a fixed integer. The polynomial $Q_m \in \mathcal{P}_m$, which furnishes the best approximation to $\dfrac{1}{t+a}$ in the L^∞-norm on $[-1,1]$ is*

$$Q_m(t) = \frac{1}{t+a}\left(1 - \frac{2\eta^m}{(\eta-\eta^{-1})^2} R_{m+1}(t)\right), \tag{2.6}$$

where

$$R_{m+1}(t) = \eta^{-1}T_{m+1}(t) - 2T_m(t) + \eta T_{m-1}(t). \tag{2.7}$$

The error of best approximation is

$$E_{[-1,1]} = \min_{Q\in\mathcal{P}_m} \left\| \frac{1}{t+a} - Q \right\|_{L_\infty[-1,1]} = \frac{\delta^m}{a^2-1}.$$

Proof. We prove this theorem in the appendix by showing how one can derive (2.6) from [1, p. 33, Equation (4.25)]. □

The following corollary is immediate and follows after elementary calculations.

Corollary 2.2. *Let $E_{m,[\lambda_{\min},\lambda_{\max}]}$ be the error of approximation with polynomial of degree m on the interval $[\lambda_{\min},\lambda_{\max}]$, $0 < \lambda_{\min} < \lambda_{\max} < \infty$. Then*

$$E_{m,[\lambda_{\min},\lambda_{\max}]} = 2\delta^{m-1}E^2_{0,[\sqrt{\lambda_{\min}},\sqrt{\lambda_{\max}}]}, \tag{2.8}$$

where $E_{0,[\sqrt{\lambda_{\min}},\sqrt{\lambda_{\max}}]}$ is given by the expression

$$E_{0,[\sqrt{\lambda_{\min}},\sqrt{\lambda_{\max}}]} = \frac{1}{2}\left(\frac{1}{\sqrt{\lambda_{\min}}} - \frac{1}{\sqrt{\lambda_{\max}}}\right).$$

Theorem 2.3. *For the polynomials of best approximation to $\dfrac{1}{x}$ given in (2.6), the following three-term recurrence relation holds:*

$$\eta^{-1}Q_{m+2}(t) - 2tQ_{m+1}(t) + \eta Q_m(t) = -2, \quad m = 0,1,\ldots \tag{2.9}$$

with

$$Q_0(t) = \frac{a}{a^2-1}, \qquad Q_1(t) = \frac{1}{\sqrt{a^2-1}} - \frac{t}{a^2-1}.$$

The error of approximation is:

$$E_{[-1,1]} = \min_{Q\in\mathcal{P}_m} \left\| \frac{1}{t+a} - Q \right\|_{L_\infty[-1,1]} = \frac{\delta^m}{a^2-1}.$$

Proof. It is immediate to check that for $m = 0$,

$$Q_0(t) = \frac{1}{2}\left(\frac{1}{a-1} + \frac{1}{a+1}\right) = \frac{1}{2}\left(\frac{2a}{a^2-1}\right) = \frac{2(\eta+\eta^{-1})}{(\eta-\eta^{-1})^2}.$$

Setting

$$r_m(t) = (\eta-\eta^{-1})^2 + q_m(t), \quad \text{with} \quad q_m(t) = 2(-1)^m\eta^{-m}R_{m+1}(t)$$

we have

$$Q_m(t) = \frac{r_m(t)}{(t+a)(\eta-\eta^{-1})^2}.$$

For $m = 1$ we then readily obtain

$$\begin{aligned} r_1(t) &= (\eta-\eta^{-1})^2 - 2\eta^{-1}(\eta(2t^2-1)+2t+\eta^{-1}) \\ &= \eta^2+\eta^{-2}-2-4t^2+2-4\eta^{-1}t-2\eta^{-2} \\ &= \eta^2-4t^2-4\eta^{-1}t-\eta^{-2} = \eta^2-(2t+\eta^{-1})^2 \\ &= (\eta-2t-\eta^{-1})(\eta+2t+\eta^{-1}) = 2(\eta-2t-\eta^{-1})(t+a). \end{aligned}$$

This shows that $Q_1(t)$ has the form given in the statement of the theorem. For $m \geq 2$, using the recurrence relation for $T_m(t)$, it is easy to check that

$$R_{m+2}(t) - 2tR_{m+1}(t) + R_m(t) = 0.$$

We then have

$$\begin{aligned} \eta q_{m+1}(t) + 2tq_m(t) + \eta^{-1}q_{m-1}(t) &= 2\eta(-1)^{m+1}\eta^{-m-1}R_{m+2}(t) \\ &\quad +4t(-1)^m\eta^{-m}R_{m+1}(t) + 2\eta^{-1}(-1)^{m-1}\eta^{-m+1}R_m(t) \\ &= 2(-1)^{m+1}\eta^{-m}(R_{m+2}(t) - 2tR_{m+1}(t) + R_m(t)) = 0. \end{aligned}$$

On the other hand, for any constant K, by the definition of η we have

$$\eta K + 2tK + \eta^{-1}K = 2(t+a)K.$$

Hence, after applying the above identities (with $K = (\eta-\eta^{-1})^2$) we get

$$\eta r_{m+1}(t) + 2tr_m(t) + \eta^{-1}r_{m-1}(t) = 2(t+a)(\eta-\eta^{-1})^2.$$

The proof then is easily completed by using the definition of $Q_m(t)$. □

The next lemma gives an estimate on $|R_{m+1}(t)|$ by a linear polynomial, which is used later to derive a sufficient condition for the positivity of $q_m(\cdot)$.

Lemma 2.4. *The following estimate holds for the polynomial $R_{m+1}(t)$ defined in Theorem 2.1:*

$$-2(t+a) \leq R_{m+1}(t) \leq 2(t+a), \quad t \in [-1,1]. \tag{2.10}$$

Proof. Recall that by the definition of η and δ (see (2.3)), we have that $\eta < 0$, and $|\eta| = \delta$. Let $t = \cos\alpha$, for $\alpha \in [0, \pi]$. Then we find that

$$\begin{aligned} R_{m+1}(t) + 2t - \eta - \eta^{-1} &= \eta^{-1}(T_{m+1}(t) - 1) - 2(T_m(t) - t) + \eta(T_{m-1}(t) + 1) \\ &= -2\eta^{-1}\sin^2\frac{m+1}{2}\alpha + 4\sin\frac{m+1}{2}\alpha\sin\frac{m-1}{2}\alpha \\ &\quad -2\eta\sin^2\frac{m-1}{2}\alpha \\ &= -2\eta^{-1}\left(\sin\frac{m+1}{2}\alpha - \eta\sin\frac{m-1}{2}\alpha\right)^2 \\ &= 2\delta^{-1}\left(\sin\frac{m+1}{2}\alpha + \delta\sin\frac{m-1}{2}\alpha\right)^2 \geq 0. \end{aligned} \tag{2.11}$$

In an analogous fashion we obtain

$$\begin{aligned} R_{m+1}(t) - 2t + \eta + \eta^{-1} &= \eta^{-1}(T_{m+1}(t) + 1) - 2(T_m(t) + t) + \eta(T_{m-1}(t) + 1) \\ &= 2\eta^{-1}\cos^2\frac{m+1}{2}\alpha - 4\cos\frac{m+1}{2}\alpha\cos\frac{m-1}{2}\alpha \\ &\quad +2\eta\cos^2\frac{m-1}{2}\alpha \\ &= 2\eta^{-1}\left(\cos\frac{m+1}{2}\alpha - \eta\cos\frac{m-1}{2}\alpha\right)^2 \\ &= -2\delta^{-1}\left(\cos\frac{m+1}{2}\alpha + \delta\cos\frac{m-1}{2}\alpha\right)^2 \leq 0. \end{aligned} \tag{2.12}$$

Combining (2.11) and (2.12) and using $2t - \eta - \eta^{-1} = 2(t + a)$ yields the desired result. □

2.2. **Algorithm for finding the polynomial of best uniform approximation to x^{-1}.** The result in Theorem 2.3 gives us the polynomial approximation on the interval $[\lambda_{\max}, \lambda_{\min}]$. Indeed, the recurrence relation for $q_{m+1}(x) = 2\sigma Q_{m+1}(2\sigma x - a)$ is:

$$Q_{m+1}(2\sigma x - a) = \eta[-2 + 2(2\sigma x - a)Q_m(2\sigma x - a) - \eta Q_{m-1}(2\sigma x - a)].$$

Multiplying by 2σ then gives

$$q_{m+1}(x) = \eta[-4\sigma + 2\sigma(2\sigma x - a)Q_m(2\sigma x - a) - 2\sigma\eta Q_{m-1}(2\sigma x - a)].$$

Based on this identity, we have the following algorithm in which the formulas are obtained by writing η, σ and a in terms of $\mu_0 = 1/\lambda_{\max}$ and $\mu_1 = 1/\lambda_{\min}$ and δ (defined in (2.3)). The reason for choosing these parameters is because the constants in the algorithm are symmetric with respect to μ_0 and μ_1.

Algorithm 2.5. Set $\mu_0 = 1/\lambda_{\max}$ and $\mu_1 = 1/\lambda_{\min}$.

1. Calculate the 0-th order polynomial q_0 and the first order polynomial q_1:

$$q_0(x) = \frac{1}{2}(\mu_0 + \mu_1), \quad \text{and} \quad q_1(x) = \frac{1}{2}(\sqrt{\mu_0} + \sqrt{\mu_1})^2 - \mu_0\mu_1 x.$$

2. For $k = 1, \ldots, m-1$, q_{k+1} written as a correction to q_k is computed as follows:

$$\begin{aligned} \ell_{k+1}(x) &= \frac{4\mu_0\mu_1}{(\sqrt{\mu_0}+\sqrt{\mu_1})^2}[1 - q_k(x)\,x] + \delta^2[q_k(x) - q_{k-1}(x)] \\ &= \frac{4\mu_0\mu_1}{(\sqrt{\mu_0}+\sqrt{\mu_1})^2}[1 - q_k(x)\,x] + \delta^2\ell_k(x), \\ q_{k+1}(x) &= q_k(x) + \ell_{k+1}(x). \end{aligned}$$

In other words, we have the relation

$$q_{k+1}(x) - q_k(x) = \delta^2(q_k(x) - q_{k-1}(x)) + \frac{4\mu_0\mu_1}{(\sqrt{\mu_0}+\sqrt{\mu_1})^2}\left[1 - xq_k(x)\right]. \tag{2.13}$$

This formula can be used to perform stationary iterations towards solving $A\boldsymbol{u} = \boldsymbol{f}$ for a given symmetric and positive definite matrix A and a given symmetric positive definite preconditioner D to A. A standard stationary iterative method has the form: Given an approximation $\boldsymbol{v}$ to the solution $\boldsymbol{u}$ of the linear system in hand, the next approximation $\boldsymbol{w}$ is defined as

$$\boldsymbol{w} = \boldsymbol{v} + R\,(\boldsymbol{f} - A\boldsymbol{v})\,.$$

A sequence of such approximations, approaching $\boldsymbol{u}$ (when the method is convergent) is obtained by applying this iteration with $\boldsymbol{w} = \boldsymbol{u}_{j+1}$, $\boldsymbol{v} = \boldsymbol{u}_j$, $j = 0, \ldots$, and, with $\boldsymbol{u}_0$, a given initial guess.

We now define

$$R = q_m(D^{-1}A)D^{-1},$$

where q_m is the polynomial of best approximation to x^{-1} on the interval $\left[\frac{\lambda}{\kappa},\ \lambda\right]$ with λ an upper bound for the largest eigenvalue of $D^{-1}A$ and $\kappa > 1$, a parameter controlling the length of the interval.

At every iteration, we need to compute the actions $R\boldsymbol{r}$, where $\boldsymbol{r} = \boldsymbol{f} - A\boldsymbol{v}$ is the current residual. This is accomplished by writing equation (2.13) with a matrix argument, namely:

$$\begin{aligned} \ell_k(D^{-1}A) &= q_k(D^{-1}A) - q_{k-1}(D^{-1}A), \\ \ell_{k+1}(D^{-1}A)D^{-1} &= \delta^2\ell_k(D^{-1}A)D^{-1} \\ &\quad + \frac{4\mu_0\mu_1}{(\sqrt{\mu_0}+\sqrt{\mu_1})^2}D^{-1}\left[I - Aq_k(D^{-1}A)D^{-1}\right]. \end{aligned} \tag{2.14}$$

Algorithm 2.6 (Polynomial Preconditioning with $R = q_m(D^{-1}A)D^{-1}$)**.**

Given $\boldsymbol{r}$, in the following steps the algorithms computes at the end $q_m(D^{-1}A)D^{-1}\boldsymbol{r}$.

(0) Initially, compute $\overline{\boldsymbol{r}} = D^{-1}\boldsymbol{r}$.

(i) Then, compute $\boldsymbol{v}_0 = \frac{1}{2}\,(\mu_0 + \mu_1)\overline{\boldsymbol{r}}$ and $\boldsymbol{v}_1 = \frac{1}{2}\,(\sqrt{\mu_0} + \sqrt{\mu_1})^2\,\overline{\boldsymbol{r}} - \mu_0\mu_1 D^{-1}A\overline{\boldsymbol{r}}$.

(ii) For $k = 1, 2, \ \ldots, \ m-1$, compute the current and preconditioned residuals,

$$\boldsymbol{r}_k = \boldsymbol{r} - A\boldsymbol{v}_k, \qquad \overline{\boldsymbol{r}}_k = D^{-1}\boldsymbol{r}_k.$$

The next $\boldsymbol{v}_{k+1}$ is computed based on the recurrence formula (2.14)

$$\boldsymbol{v}_{k+1} = \boldsymbol{v}_k + \delta^2(\boldsymbol{v}_k - \boldsymbol{v}_{k-1}) + \frac{4\mu_0\mu_1}{(\sqrt{\mu_0}+\sqrt{\mu_1})^2}\overline{\boldsymbol{r}}_k.$$

(iii) At the end, we let $R\boldsymbol{r} = \boldsymbol{v}_m$.

The reason to write q_{k+1} as a correction to q_k is to show that such iterations look like iterations in a defect-correction method: First computing the residual $[1 - q_k(x)\,x]$, and then trying to correct it by adding an additional term. One can also easily see that for any initial q_0 and q_1, if the sequence $q_k(x)$ converges, then it converges to x^{-1}. In other words, choosing q_0 and q_1 different from what they are above, will not generate the sequence of best approximations to x^{-1}, but still this sequence will converge to x^{-1}.

3. Properties of the sequence of polynomials

To simplify the presentation, we now set $\lambda = \lambda_{\max}$ and in this notation we have $\lambda_{\min} = \dfrac{\lambda}{\kappa}$ (recall the definition of κ given in §2). We thus consider the best approximation $q_m(x)$ to $\dfrac{1}{x}$ on the interval $\left[\dfrac{\lambda}{\kappa}, \lambda\right]$. We prove several results on the positivity of the polynomial $q_m(x)$, and the monotonicity of the sequence $\{q_m\}$ for sufficiently large m.

We first note the following identity

$$\begin{aligned} x\,q_m(x) &= 2\sigma x\,Q_m(2\sigma x - a) = (t+a)\,Q_m(t) \\ &= 1 - \frac{2\eta^m}{(\eta-\eta^{-1})^2} R_{m+1}(t) = 1 - \frac{2(-1)^m\delta^m}{(\delta-\delta^{-1})^2} R_{m+1}(t), \quad t \in [-1,1] \end{aligned} \tag{3.1}$$

This gives

$$1 - xq_m(x) = \frac{2(-1)^m\delta^m}{(\delta-\delta^{-1})^2} R_{m+1}(t). \tag{3.2}$$

The next Lemma shows that $(1 - q_m(x)x) > 0$ for all $x \in \left[0, \dfrac{\lambda}{\kappa}\right]$.

Lemma 3.1. *Let $q_m(x)$ be the polynomial of degree less than or equal to m, which furnishes the best approximation to $\dfrac{1}{x}$ in the L^∞-norm on the interval $\left[\dfrac{\lambda}{\kappa}, \lambda\right]$, $\kappa > 1$. Then the following inequality holds:*

$$0 < 1 - xq_m(x), \quad \forall x \in \left(0, \frac{\lambda}{\kappa}\right] \tag{3.3}$$

Proof. Consider the polynomial

$$p(x) = 1 - xq_m(x).$$

Note that $p(x)$ is of degree at most $(m+1)$. Since we have

$$p(x) = x\left(\frac{1}{x} - q_m(x)\right),$$

and $x > 0$ in the intervals of interest, we may conclude that the sign changes in the function $\left(\dfrac{1}{x} - q_m(x)\right)$ are the same as the sign changes in $p(x)$ for any $x > 0$. However, $q_m(x)$ is the polynomial of best uniform approximation to $\dfrac{1}{x}$, and hence there are at least $(m+2)$ points

of Chebyshev alternance in the interval $\left[\frac{\lambda}{\kappa}, \lambda\right]$. Thus, there exist points $\{x_k\}_{k=1}^{m+2}$ such that

$$\frac{\lambda}{\kappa} \leq x_1 < x_2 < \ldots < x_{m+1} < x_{m+2} \leq \lambda,$$

and also such that

$$\left(\frac{1}{x_k} - q_m(x_k)\right) = -\left(\frac{1}{x_{k+1}} - q_m(x_{k+1})\right), \quad k = 1, \ldots, (m+1).$$

We define now $e := \left(\frac{1}{x_1} - q_m(x_1)\right)$, and use the alternation property to get that

$$p(x_k)p(x_{k+1}) = -x_k x_{k+1} e^2 < 0, \quad k = 1, \ldots, (m+1).$$

Hence, we may conclude that all the roots of $p(x)$ are disjoint, and that each of them lies in the *open* interval (x_k, x_{k+1}), $k = 1, \ldots, (m+1)$. We may also conclude that there are no roots of $p(x)$ outside of the open interval $\left(\frac{\lambda}{\kappa}, \lambda\right)$ and there are no roots of its first derivative outside this interval. This is so by the Rolle's theorem: the first derivative $p'(x)$ clearly has m distinct roots, each lying between the roots of $p(x)$. Hence, $p(x)$ is either strictly increasing or strictly decreasing on the interval $\left[0, \frac{\lambda}{\kappa}\right]$ and also it cannot have a zero in this interval. Recall that $0 < \delta = -\eta < 1$ and that $T_k(-1) = (-1)^k$. Using the definition of $R_{m+1}(t)$ from Theorem 2.1, and the relation (3.2) it follows that

$$\begin{aligned} p\left(\frac{\lambda}{\kappa}\right) &= \frac{2(-1)^m \delta^m}{(\delta - \delta^{-1})^2} R_{m+1}(-1) \\ &= \frac{2(-1)^m \delta^m}{(\delta - \delta^{-1})^2} [(-\delta^{-1})(-1)^{m+1} - 2(-1)^m + (-\delta)(-1)^{m-1}] \\ &= \frac{2\delta^m}{(\delta - \delta^{-1})^2} (\delta^{-1} + \delta - 2) = \frac{2\delta^m}{(\delta + \delta^{-1} + 2)} < 1 = p(0). \end{aligned}$$

Here we have used that

$$(\delta - \delta^{-1})^2 = [(\delta^{\frac{1}{2}} + \delta^{-\frac{1}{2}})^2 (\delta^{\frac{1}{2}} + \delta^{-\frac{1}{2}})^2 = (\delta + \delta^{-1} - 2)(\delta + \delta^{-1} + 2). \tag{3.4}$$

We thus conclude that $p(0) > p(\frac{\lambda}{\kappa})$ and therefore $p(x)$ must be decreasing on $\left(0, \frac{\lambda}{\kappa}\right]$, and this leads to

$$0 < \frac{2\delta^m}{(\delta + \delta^{-1} + 2)} = p\left(\frac{\lambda}{\kappa}\right) \leq p(x) \leq 1, \tag{3.5}$$

which concludes the proof. □

The next lemma shows that for $x \in \left[0, \frac{\lambda}{\kappa}\right]$ the sequence of polynomials of best approximation of increasing degree is monotone.

Lemma 3.2. *The following estimate holds:*

$$q_m(x) < q_{m+1}(x), \quad \textit{for all} \quad x \in \left[0, \frac{\lambda}{\kappa}\right], \tag{3.6}$$

where $q_k(x)$, $k = m, (m+1)$ is the best polynomial approximation of degree at most k to $\frac{1}{x}$ in the L^∞-norm on the interval $\left[\frac{\lambda}{\kappa}, \lambda\right]$, $\kappa > 1$.

Proof. The proof amounts to showing that $\ell_{m+1}(x) > 0$ (defined in Step 2. of Algorithm 2.5) for $x \in \left[0, \frac{\lambda}{\kappa}\right]$. With the notation given in Algorithm 2.5 for such values of x we have $x \leq \frac{\lambda}{\kappa} = \mu_1^{-1}$. Therefore,

$$\begin{aligned} \ell_1(x) &= q_1(x) - q_0(x) = \frac{1}{2}(\mu_0 + \mu_1 + 2\sqrt{\mu_0\mu_1}) - \mu_0\mu_1 x - \frac{1}{2}(\mu_0 + \mu_1) \\ &= \sqrt{\mu_0\mu_1}(1 - x\sqrt{\mu_0\mu_1}) \geq \sqrt{\mu_0\mu_1}(1 - \mu_1^{-1}\sqrt{\mu_0\mu_1}) = \frac{\sqrt{\kappa} - 1}{\lambda} > 0. \end{aligned}$$

Further, from Step 2. of Algorithm 2.5 and Lemma 3.1 we have

$$\begin{aligned} \ell_{m+1}(x) &= \frac{4\mu_0\mu_1}{(\sqrt{\mu_0} + \sqrt{\mu_1})^2}[1 - q_m(x)\, x] + \delta^2 \ell_m(x) \\ &= \frac{4\kappa}{\lambda(1 + \sqrt{\kappa})^2}[1 - q_m(x)\, x] + \delta^2 \ell_m(x) \\ &\geq \frac{8\kappa\delta^m}{\lambda(1 + \sqrt{\kappa})^2(\delta + \delta^{-1} + 2)} + \delta^2 \ell_m(x). \end{aligned}$$

Noticing that $(\delta + \delta^{-1} + 2) = \frac{4\kappa}{\kappa - 1}$ then leads to:

$$\ell_{m+1}(x) \geq \frac{2}{\lambda}\delta^{m+1} + \delta^2 \ell_m(x). \tag{3.7}$$

Clearly, $\ell_{m+1} > 0$ if $\ell_m(x) > 0$ and a standard induction argument concludes the proof of the lemma. □

Remark 3.3. *From (3.7) one can have sharper bounds below on $\ell_{m+1}(x)$, but we do not pursue these further here.*

The next lemma is a straightforward corollary of Lemma 3.1.

Lemma 3.4. *Let $q_m(x)$ be the best polynomial approximation of degree at most m to $\frac{1}{x}$ in L^∞-norm on the interval $\left[\frac{\lambda}{\kappa}, \lambda\right]$, $\kappa > 1$. Suppose that $q_m(x)$ is positive on the interval $\left[\frac{\lambda}{\kappa}, \lambda\right]$. Then $q_m(x)$ is positive on the whole interval $x \in (0, \lambda]$.*

Proof. We have already shown in the previous lemma that $q_m(x) > q_0(x) > 0$, for all $m \geq 1$ and $x \in \left[0, \frac{\lambda}{\kappa}\right]$. Since, by assumption $q_m(x)$ is positive on the interval $\left[\frac{\lambda}{\kappa}, \lambda\right]$ the proof is complete. □

In the two-level method convergence estimates in the next section, we will use the following result (which also includes a sufficient condition for the positivity of $q_m(x)$).

Lemma 3.5. *Assume that κ and m are such that the following inequality holds:*

$$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m \leq \frac{\omega}{\kappa-1}, \quad \textit{for some} \quad \omega \in (0,2). \tag{3.8}$$

Then the following inequality holds for for all $x \in (0,\lambda]$:

$$\frac{1}{2}\min\left\{\frac{\kappa+1}{\lambda}, \frac{2-\omega}{x}\right\} \leq q_m(x) \leq \frac{1}{x}\left(1+\frac{\omega}{2}\right). \tag{3.9}$$

Proof. Lower bound: We prove first the lower bound when $x \in \left[\frac{\lambda}{\kappa}, \lambda\right]$. Let $R_{m+1}(t)$ be the polynomial that has been defined in Theorem 2.1. We use the relation (3.1) and Lemma 2.4. Note that $-1 \leq t \leq 1$ for $x \in \left[\frac{\lambda}{\kappa}, \lambda\right]$, and we estimate below $xq_m(x)$ as follows

$$\begin{aligned} xq_m(x) &= 1 - \frac{2(-1)^m\delta^m}{(\delta-\delta^{-1})^2}R_{m+1}(t) \geq 1 - \frac{2\delta^m}{(\delta-\delta^{-1})^2}|R_{m+1}(t)| \\ &\geq 1 - \frac{2\delta^m}{(\delta-\delta^{-1})^2}\left(2t+\delta+\delta^{-1}\right) \\ &\geq 1 - \frac{2\delta^m}{(\delta-\delta^{-1})^2}\left(2+\delta+\delta^{-1}\right) \\ &= 1 - \frac{2\delta^m}{\delta+\delta^{-1}-2} = 1 - \delta^m\frac{\kappa-1}{2} \geq \frac{2-\omega}{2}. \end{aligned}$$

In the last two steps we have used the identity (3.4) and the definition of δ, given in (2.3). We thus have shown that $q_m(x) \geq \frac{2-\omega}{2x}$ for all $x \in \left[\frac{\lambda}{\kappa}, \lambda\right]$. Next, we apply Lemma 3.2 and we have that

$$q_m(x) \geq q_0(x) = \frac{\kappa+1}{2\lambda}, \quad \text{for} \quad x \in \left[0, \frac{\lambda}{\kappa}\right],$$

which concludes the proof of the lower bound.

Upper bound: To prove the upper bound, we need to consider only the case $x \in \left[\frac{\lambda}{\kappa}, \lambda\right]$, because from Lemma 3.1 we already know that $xq_m(x) < 1$ for $x \in \left[0, \frac{\lambda}{\kappa}\right]$. For $x \in \left[\frac{\lambda}{\kappa}, \lambda\right]$, we apply an argument analogous to the one for the lower bound using the relation (3.1) and Lemma 2.4 (just changing "$-$" to "$+$"):

$$xq_m(x) = 1 - \frac{2(-1)^m\delta^m}{(\delta-\delta^{-1})^2}R_{m+1}(t) \leq 1 + \delta^m\frac{\kappa-1}{2} \leq 1 + \frac{\omega}{2}.$$

□

Remark 3.6. *Note that this lemma implies that the polynomial of best approximation is positive on $[0,\lambda]$ as long as (3.8) is satisfied with $\omega \in (0,2)$.*

To conclude this section, we discuss conditions relating κ and the degree of the polynomial m so that (3.8) holds. In what follows, without loss of generality we assume that $\ln((\kappa-1)/\omega) > 1$. In applications (particularly for analysis of convergence of two-level methods)

we are interested in large values of κ (resp. m). Since $\omega \in (0,2)$, such condition is clearly satisfied for $\kappa > 2e+1$.

For fixed and sufficiently large κ, (as we assumed above), let m satisfy

$$\frac{\sqrt{\kappa}+1}{2}\ln[(\kappa-1)/\omega] \le m \le 1+\frac{\sqrt{\kappa}+1}{2}\ln[(\kappa-1)/\omega]. \tag{3.10}$$

We will now show that the lower bound in (3.10) implies (3.8) (and therefore also the conclusion of Lemma 3.5). Since $0<\delta<1$ we have

$$\delta^m = \left(1-\frac{2}{\sqrt{\kappa}+1}\right)^m \le \left[\left(1-\frac{2}{\sqrt{\kappa}+1}\right)^{\sqrt{\kappa}+1}\right]^{\frac{1}{2}\ln[(\kappa-1)/\omega]}$$

On the other hand, the function $(1-2/\xi)^\xi$ is increasing for all $\xi>2$, and hence

$$\delta^m < \left[\lim_{\xi\to\infty}\left(1-\frac{2}{\xi}\right)^\xi\right]^{\frac{1}{2}\ln[(\kappa-1)/\omega]} = \exp\left(-\ln\frac{\kappa-1}{\omega}\right) = \frac{\omega}{\kappa-1}.$$

Thus, if κ is given, the polynomial degree m for which (3.8) holds is bounded below by the right hand side of (3.10).

In addition, it is easy to show that if (3.10) holds, then we also have

$$\frac{1}{\kappa+1} \le c_\omega\left(\frac{\ln m}{m}\right)^2, \quad \text{with} \quad c_\omega = \frac{1}{2}\sup_{\kappa>1;\omega\in(0,2)}\left(\frac{\ln(\kappa/2)}{1+\ln(\kappa/\omega)}\right)^2. \tag{3.11}$$

Note that c_ω is finite. The inequality (3.11) is seen as follows. Since the logarithm is an increasing function on its domain we get

$$\ln m = \ln[\sqrt{\kappa}+1)/2] + \ln\ln[(\kappa-1)/\omega] \ge \frac{1}{2}\ln(\kappa/2).$$

Also, from (3.10), since $\sqrt{\kappa}+1 \ge 2$ we have:

$$\begin{aligned} m^2 &\le \left(1+\frac{\sqrt{\kappa}+1}{2}\ln[(\kappa-1)/\omega]\right)^2 \le \left(\frac{\sqrt{\kappa}+1}{2}(1+\ln[(\kappa-1)/\omega]\right)^2 \\ &\le \frac{1}{2}(\kappa+1)(1+\ln\kappa/\omega)^2. \end{aligned}$$

Combining the last two estimates then gives (3.11).

4. An application to two-level methods

We consider the linear system of equations

$$A\boldsymbol{u} = \boldsymbol{f}, \tag{4.1}$$

where $A \in \mathbb{R}^{N\times N}$ is a symmetric and positive definite matrix, and $\boldsymbol{f} \in \mathbb{R}^N$ is a given right hand side vector. To describe a general two-level multiplicative method, we denote $\boldsymbol{V} = \mathbb{R}^N$, and also introduce a coarse space V_H, $\boldsymbol{V}_H \subset V$, $N_H = \dim V_H$, $N_H < N$. In the following we will always assume that $\boldsymbol{V}_H = \text{range}(P)$, where $P : \mathbb{R}^{N_H} \mapsto \boldsymbol{V}$ and its matrix representation in the canonical basis of $\mathbb{R}^{N_H}$ is given by the coefficients in the expansion of the basis in $\boldsymbol{V}_H$ via the basis in $\boldsymbol{V}$. Clearly, P is a full rank operator and its matrix representation is oftentimes called *prolongation* or *interpolation* matrix. The restriction of A on the coarse space is denoted by $A_H = P^T AP$.

4.1. **Convergence rate estimates.** In this subsection we prove convergence estimates for the classical multiplicative two-level iteration, with polynomial smoother which is used to define a preconditioner $B \approx A^{-1}$. In a recent work [8] the properties of special polynomial smoothers have been exploited in order to conduct an improved convergence analysis of smoothed aggregation algebraic multigrid methods. Here, only for completeness, we include a two-level convergence result presented in [7]. The only difference is that we use a polynomial smoother with polynomial defined via Algorithm 2.6. As in [7] we show explicit dependence of the estimates on the degree of the polynomial.

The results up to and including Theorem 4.3 hold for general SPD A, $\boldsymbol{V}$ and $\boldsymbol{V}_H$, provided that the smoother is constructed using the polynomials of best approximation to $1/x$ on a suitably chosen interval.

In this subsection, by $\rho(X)$ we denote the spectral radius of a matrix X. If, in addition, X is symmetric and positive definite matrix, we denote the X-norm by $\|\boldsymbol{v}\|_X^2 = \boldsymbol{v}^T X \boldsymbol{v}$.

We define the two-grid (or TG) preconditioner using a classical two-level algorithm which reads as follows.

Algorithm 4.1. Given $\boldsymbol{w} \in \boldsymbol{V}$ which approximates the solution of (4.1) we define the next approximation $\boldsymbol{v} \in \boldsymbol{V}$ to $\boldsymbol{u}$ via the following two steps:

1. Coarse grid correction: $\boldsymbol{y}{:=}\boldsymbol{w} + PA_H^{-1}P^T(\boldsymbol{f} - A\boldsymbol{w})$
2. Smoothing: $\boldsymbol{v}{:=}\boldsymbol{y} + R(\boldsymbol{f} - A\boldsymbol{y})$.

We assume that R is symmetric and positive definite and A-norm convergent, namely

$$\|I - RA\|_A^2 < 1. \tag{4.2}$$

The error propagation operator for the two-level iteration above is

$$E_{\mathrm{TL}} = (I - RA)(I - \pi_A), \quad \pi_A = PA_H^{-1}P^T A.$$

We then define the two-level preconditioner as:

$$B = (I - E_{\mathrm{TL}}E_{\mathrm{TL}}^*)A^{-1}.$$

Here E_{TL}^* denotes the adjoint with respect to the inner product defined by A. Introducing $\bar{R}$ such that

$$\left(I - \bar{R}A\right) = (I - RA)^2 \quad \text{and hence} \quad \bar{R} = 2R - RAR. \tag{4.3}$$

it is straightforward then to compute that (see, e.g., [9]):

$$B = \bar{R} + (I - RA)\, PA_H^{-1}P^T\, (I - AR)\,. \tag{4.4}$$

Recall a necessary and sufficient condition for R to be a convergent smoother in A-norm, i.e., (4.2) to hold is that $\bar{R}$ is SPD.

Our goal will be to prove a convergence rate estimate for the two-level method with polynomial smoother. First, let us denote with D the diagonal of A and set

$$R = q_m(D^{-1}A)D^{-1},$$

where $q_m(x)$ is the polynomial of best approximation to $1/x$, generated by the Algorithm 2.5 on a fixed interval $[\lambda/\kappa, \lambda]$. Both λ and κ are to be specified later.

One may also write R in the form

$$R = D^{-1/2}q_m(\widehat{A})D^{-1/2}, \qquad \widehat{A} = D^{-1/2}AD^{-1/2}. \tag{4.5}$$

Using the notation from Section 3, we set $\lambda = \|\widehat{A}\|_{\ell_\infty}$. In what follows, we hold λ fixed and we vary κ and the degree of the polynomial m. However, κ and m do not vary independently and we assume that κ and m satisfy the condition (3.8). With such choice of λ, κ and m, one can easily show that R is a contraction (a convergent smoother) in A-norm and we do so by showing that $\bar{R}$ is SPD, which, as we mentioned earlier, is both necessary and sufficient condition for (4.2) to hold. Clearly, $\bar{R}$ can be written (see (4.3)) as

$$\bar{R} = D^{-1/2}[2q_m(\widehat{A}) - q_m^2(\widehat{A})\widehat{A}]D^{-1/2}.$$

From the upper bound in Lemma 3.5 we immediately get that for all $x \in (0, \lambda]$ we have $xq_m(x) \leq \dfrac{2+\omega}{2}$. Therefore, for all $\boldsymbol{w} \in \boldsymbol{V}$ we get

$$\begin{aligned}\boldsymbol{w}^T(2q_m(\widehat{A}) - [q_m(\widehat{A})]^2\widehat{A})\boldsymbol{w} &\geq (2 - \|xq(x)\|_{\infty,(0,\lambda)})\boldsymbol{w}^T q_m(\widehat{A})\boldsymbol{w}\\ &\geq \frac{2-\omega}{2}\boldsymbol{w}^T q_m(\widehat{A})\boldsymbol{w}.\end{aligned}$$

Applying the inequality above with $\boldsymbol{w} = D^{-1/2}\boldsymbol{y}$ then shows that for all $\boldsymbol{y} \in \boldsymbol{V}$

$$\boldsymbol{y}^T\bar{R}\boldsymbol{y} \geq \frac{2-\omega}{2}\boldsymbol{y}^T R\boldsymbol{y} \geq \frac{2-\omega}{2}\min_{x\in(0,\lambda]} q_m(x)\ \boldsymbol{y}^T D^{-1}\boldsymbol{y}. \tag{4.6}$$

From the lower bound in Lemma 3.5, we conclude that $\bar{R}$ is SPD.

We further note that each of the off-diagonal entries of $(D^{-1/2}AD^{-1/2})$ is less than 1 and the diagonal entry is equal to 1. Therefore, we have that

$$1 \leq \|D^{-1/2}AD^{-1/2}\| = \rho(D^{-1/2}AD^{-1/2}) \leq \|D^{-1/2}AD^{-1/2}\|_{\ell_\infty} = \lambda \leq n_z, \tag{4.7}$$

where n_z is the maximal number of non-zeros in a row of A.

The convergence rate estimates are derived from the following theorem (two-level version of the XZ-identity, cf. [10, 9]).

Theorem 4.2. *Assume that $\bar{R}$ is SPD. Then the following identity holds:*

$$\boldsymbol{v}^T B^{-1}\boldsymbol{v} = \inf_{\boldsymbol{v}_H\in\boldsymbol{V}_H}[\|\boldsymbol{v}_H\|_A^2 + \|\boldsymbol{v} - \boldsymbol{v}_H\|_{\bar{R}^{-1}}^2]. \tag{4.8}$$

Based on Theorem 4.2, we now state and prove a convergence result involving the polynomial smoother.

Theorem 4.3. *Let A be a symmetric positive definite matrix and D be its diagonal. Let $\lambda = \|D^{-1/2}AD^{-1/2}\|_{\ell_\infty}$, and also $\kappa > 1$ and m satisfy (3.8). If $R = q_m(D^{-1}A)D^{-1}$, with $q_m(x)$ the polynomial of best approximation to $1/x$ on the interval $[\lambda/\kappa, \lambda]$, then the following estimate holds for all $\boldsymbol{v} \in \boldsymbol{V}$:*

$$\boldsymbol{v}^T B^{-1}\boldsymbol{v} \leq \frac{4}{(2-\omega)}\inf_{\boldsymbol{v_H}\in\boldsymbol{V_H}}\left[\|\boldsymbol{v}_H\|_A^2 + \frac{\lambda}{(\kappa+1)}\|\boldsymbol{v} - \boldsymbol{v}_H\|_D^2 + \frac{1}{2-\omega}\|\boldsymbol{v} - \boldsymbol{v}_H\|_A^2\right]. \tag{4.9}$$

Proof. First, we see that from (4.6) we have that

$$\boldsymbol{y}^T\bar{R}\boldsymbol{y} \geq \frac{2-\omega}{2}\boldsymbol{y}^T R\boldsymbol{y} \quad \text{and hence} \quad \boldsymbol{y}^T\bar{R}^{-1}\boldsymbol{y} \leq \frac{2}{2-\omega}\boldsymbol{y}^T R^{-1}\boldsymbol{y}. \tag{4.10}$$

Under the assumptions we made in the statement of the theorem we can apply Lemma 3.5, and get that for all $x \in (0, \lambda]$,

$$\frac{1}{q_m(x)} \leq 2\max\left\{\frac{\lambda}{\kappa+1}, \frac{x}{2-\omega}\right\} \leq \left(\frac{2\lambda}{\kappa+1} + \frac{2x}{2-\omega}\right). \tag{4.11}$$

Since $\widehat{A}$ and $q_m(\widehat{A})$ commute, and have the same set of orthonormal eigenvectors, we have that for all $\boldsymbol{w} \in \boldsymbol{V}$ we have

$$\boldsymbol{w}^T[q_m(\widehat{A})]^{-1}\boldsymbol{w} \leq \frac{2\lambda}{\kappa+1}\|\boldsymbol{w}\|^2_{\ell_2} + \frac{2}{2-\omega}\|\boldsymbol{w}\|^2_{\widehat{A}}.$$

Taking $\boldsymbol{y} = D^{1/2}\boldsymbol{w}$ in the inequality above and using the estimate given in (4.10)

$$\boldsymbol{y}^T\bar{R}^{-1}\boldsymbol{y} \leq \frac{2}{2-\omega}\boldsymbol{y}^T R^{-1}\boldsymbol{y} \leq \frac{4}{(2-\omega)}\left[\frac{\lambda}{(\kappa+1)}\|\boldsymbol{y}\|^2_D + \frac{1}{2-\omega}\|\boldsymbol{y}\|^2_A\right]. \tag{4.12}$$

The proof is concluded by taking $\boldsymbol{y} = (\boldsymbol{v} - \boldsymbol{v}_H)$ and applying Theorem 4.2. □

Without loss of generality, we set now $\omega = 1$ and use that in equation (3.11) $c_\omega = c_1 \leq \frac{1}{2}$. The estimate in the Theorem 4.3 takes the form.

Corollary 4.4. *Under the assumptions of Theorem 4.3, with $\omega = 1$ we have*

$$\boldsymbol{v}^T B^{-1}\boldsymbol{v} \leq 4 \inf_{\boldsymbol{v_H}\in\boldsymbol{V_H}}\left[\|\boldsymbol{v}_H\|^2_A + \frac{\lambda}{(\kappa+1)}\|\boldsymbol{v}-\boldsymbol{v}_H\|^2_D + \|\boldsymbol{v}-\boldsymbol{v}_H\|^2_A\right]. \tag{4.13}$$

In addition, if κ and m satisfy (3.10) *we have*

$$\boldsymbol{v}^T B^{-1}\boldsymbol{v} \leq 2 \inf_{\boldsymbol{v_H}\in\boldsymbol{V_H}}\left[\|\boldsymbol{v}_H\|^2_A + \frac{\lambda \ln^2 m}{m^2}\|\boldsymbol{v}-\boldsymbol{v}_H\|^2_D + 2\|\boldsymbol{v}-\boldsymbol{v}_H\|^2_A\right]. \tag{4.14}$$

To stress the fact that estimate (4.12) is purely algebraic, we formulate it separately, as this is our main new result.

Theorem 4.5. *Let A be an s.p.d. matrix and D a given s.p.d. preconditioner for A such that $\|D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\| \leq \lambda$. Consider the polynomial preconditioner*

$$R = q_m(D^{-1}A)D^{-1},$$

where q_m is the polynomial of best approximation of $1/x$ over the interval $\left[\frac{\lambda}{\kappa},\ \lambda\right]$. The parameter κ is chosen depending on m such that (3.8) *holds for a given $\omega \in (0,2)$. Then the following smoothing property holds for R and its symmetrized version $\overline{R}$ (see* (4.3)*):*

$$\frac{2-\omega}{2}\ \boldsymbol{v}^T\overline{R}^{-1}\boldsymbol{v} \leq \boldsymbol{v}^T R^{-1}\boldsymbol{v} \leq \frac{2\lambda}{\kappa+1}\ \boldsymbol{v}^T D\boldsymbol{v} + \frac{2}{2-\omega}\ \boldsymbol{v}^T A\boldsymbol{v}.$$

In addition, if κ and m satisfy (3.10)*, we have*

$$\frac{2-\omega}{2}\ \boldsymbol{v}^T\overline{R}^{-1}\boldsymbol{v} \leq \boldsymbol{v}^T R^{-1}\boldsymbol{v} \leq 2\lambda\ \frac{\ln^2 m}{m^2}\ \boldsymbol{v}^T D\boldsymbol{v} + \frac{2}{2-\omega}\ \boldsymbol{v}^T A\boldsymbol{v}.$$

4.2. **Two-level method for discretized PDE.** In this section we apply the abstract two–level result to the case of a two-level iterative method with large coarsening ratio for the solution of a system of linear algebraic equations arising from a discretization of scalar elliptic equation with heterogeneous coefficients similarly to the presentation in [7], now for the case of a different polynomial smoother from Theorem 4.5. We consider the following variational problem: Find $u \in H^1_D(\Omega)$, for a given polygonal (polyhedral) domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) and a source term $f \in L_2(\Omega)$, such that

$$a(u,\ v) \equiv \int_\Omega \alpha(\boldsymbol{x})\ \nabla u \cdot \nabla v \ \ = \ \ \int_\Omega f(\boldsymbol{x})v(\boldsymbol{x}) = (f,\ v)\ , \quad \text{for all} \quad v \in H^1_D(\Omega). \tag{4.15}$$

Here, $\Omega \subset \mathbb{R}^d$ $d = 2, 3$ is a given domain whose boundary $\Gamma = \partial\Omega$ is partitioned as $\Gamma = \Gamma_D \cup \Gamma_N$. We assume that $\Gamma_D \neq \emptyset$ is closed as a subset of Γ and also has a nonzero $(d-1)$ dimensional measure. We refer to Γ_D as the Dirichlet part of the boundary and Γ_N as the Neumann part of the boundary. In the variational problem (4.15), $H^1_D(\Omega)$ denotes the space of functions in $H^1(\Omega)$ whose traces vanish on Γ_D.

We are interested in the case when the diffusion coefficient $\alpha = \alpha(\boldsymbol{x})$ is a piecewise constant function, that may have large variations within Ω. We thus assume that $\bar{\Omega} = \cup_{l=1}^{m_0} \bar{\mathcal{Y}}_l$, with polygonal (polyhedral) subdomains $\mathcal{Y}_l$, and that $\alpha(\boldsymbol{x}) = \alpha_l$, for all $\boldsymbol{x} \in \mathcal{Y}_l$ and $l = 1, \ldots, m_0$. We introduce the following energy norm

$$\|v\|_a^2 = \int_\Omega \alpha(\boldsymbol{x}) |\nabla v|^2 = \sum_{l=1}^{m_0} \alpha_l \int_{\mathcal{Y}_l} |\nabla v|^2. \tag{4.16}$$

We also need the weighted L_2 norm

$$\|v\|_{0,\alpha}^2 = \int_\Omega \alpha(\boldsymbol{x}) v^2 = \sum_{l=1}^{m_0} \alpha_l \int_{\mathcal{Y}_l} v^2. \tag{4.17}$$

We consider a standard discretization of the variational problem (4.15) with piecewise linear continuous finite elements. To define the finite element spaces and the approximate solution, we assume that we have a locally quasi–uniform, simplicial triangulation $\mathcal{T}_h$ of Ω. We assume that this triangulation also resolves $\mathcal{Y}_l$, namely, for $l = 1, \ldots, m_0$ we have:

$$\bar{\Omega} = \cup_{\tau \in \mathcal{T}_h} \tau, \quad \bar{\mathcal{Y}}_l = \cup_{\tau \in \mathcal{T}_{Y,l}} \tau, \tag{4.18}$$

where $\mathcal{T}_{Y,l} \subset \mathcal{T}_h$, for $l = 1, \ldots, m_0$. The standard space of piecewise linear (w.r.t $\mathcal{T}_h$) and continuous functions vanishing on the boundary of Ω is denoted by V_h.

The discrete problem then reads: Find $u \in V_h$ such that

$$a(u, v) = (f, v), \quad \text{for all} \quad v \in V_h. \tag{4.19}$$

The notation and constructions in the previous section are suitable for the finite element setting as well. Indeed, a coarse space corresponding to $\boldsymbol{V}_H$ (denoted here with V_H) as $V_H = \text{range}(P)$, with the same P as before, but this time representing the coefficients in the expansion of the basis in V_H, $\{\varphi_j^H\}_{j=1}^{N_H}$ via the canonical Lagrange basis $\{\varphi_j\}_{j=1}^N$ in V_h. Evaluating the bilinear form on the basis for V_h and the basis for V_H defines the stiffness matrix A and the matrix A_H:

$$A_{kj} = a(\varphi_j, \varphi_k), \quad (A_H)_{kj} = (P^T A P)_{jk} = a(\varphi_j^H, \varphi_k^H).$$

According to the considerations in the previous section, we use bold face to represent vectors of degrees of freedom and normal font for functions. Thus a function $v \in V_h$ is represented by the vector $\boldsymbol{v} \in \boldsymbol{V}$.

We make the following assumption for the stability and approximation properties of the coarse function space V_H.

- Approximation and stability assumption: For any $v \in V_h$ there exists $v_H \in V_H$ such that

$$H^{-2} \|v - v_H\|_{0,\alpha}^2 + \|v - v_H\|_a^2 \leq c_{as} \|v\|_a^2, \tag{4.20}$$

  where H is the diameter of the support of a typical basis function in V_H, and the constant c_{as} is independent of the variations of the coefficient $\alpha(\boldsymbol{x})$.

Construction of coarse spaces satisfying this assumption is possible as already mentioned, and we refer to [4], [5], and earlier [6] as modified recently in [7] for such constructions.

We next introduce a well-known inequality relating the weighted L^2 norm on the function space V_h and the norm provided by the diagonal of the stiffness matrix on the space of degrees of freedom (nodal values of the piece-wise linear functions). Let $\{\lambda_{j,T}\}_{j=1}^{d+1}$ be the barycentric coordinates in an element $T \in \mathcal{T}_h$ and α_T be the value of the coefficient on T (recall that $\alpha(\boldsymbol{x})$ is piece-wise constant). Let $v \in V_h$ with corresponding vector of degrees of freedom $\boldsymbol{v} \in \boldsymbol{V}$. We have the following simple inequality

$$\begin{aligned} \|\boldsymbol{v}\|_D^2 &= \sum_{T\in\mathcal{T}_h} \alpha_T \sum_{j=1}^{d+1} \boldsymbol{v}_{j,T}^2 |\nabla\lambda_{j,T}|^2 \le \sum_{T\in\mathcal{T}_h} c_T h_T^{-2} \alpha_T \sum_{j=1}^{d+1} \boldsymbol{v}_{j,T}^2 |\lambda_{j,T}|^2 \\ &\le \sum_{T\in\mathcal{T}_h} h_T^{-2} \alpha_T c_T c_{M,T} \|v\|_{L^2(T)}^2. \end{aligned}$$

In the inequalities above, we have used standard inverse inequality, and also that the local mass matrix for an element T is equivalent to its diagonal with a bound $c_{M,T}$ independent of the coefficient variation. Finally,

$$\|\boldsymbol{v}\|_D^2 \le c_M h^{-2} \|v\|_{0,\alpha}^2, \quad \text{with} \quad c_M = \max_{T\in\mathcal{T}_h} c_T c_{M,T}. \tag{4.21}$$

It is also clear that $\|\boldsymbol{v}\|_A = \|v\|_a$ by the definition of the stiffness matrix.

We now formulate the spectral equivalence result for the two-level method when applied to the discretized PDE (4.15).

Theorem 4.6. *Let $\kappa > 1$, and m be such that* (3.8) *holds with $\omega = 1$. Assume that V_H is such that the approximation and stability assumption holds. Then we have the following spectral equivalence (for n_z see* (4.7)*):*

$$\boldsymbol{v}^T A \boldsymbol{v} \le \boldsymbol{v}^T B^{-1} \boldsymbol{v} \le K_{TG}\, \boldsymbol{v}^T A \boldsymbol{v}, \quad K_{TG} = 1 + 4c_{as} \left[\frac{c_M n_z}{(\kappa+1)} \left(\frac{H}{h}\right)^2 + 1 \right]. \tag{4.22}$$

Moreover, if κ (or equivalently the degree of the polynomial m) is sufficiently large the spectral equivalence is uniform with respect to mesh parameters and coefficient variation.

Proof. The proof is the same as the one given in [7] however with a different smoothing property provided by Theorem 4.5.

The lower bound is immediate, since E_{TL} is a contraction in A-norm. The upper bound follows directly from Corollary 4.4 together used in conjunction with the simple inequalities relating the function space V_h and $\boldsymbol{V}$ (see (4.21)). Given $v \in V_h$, let $\boldsymbol{v} \in \boldsymbol{V}$ be the corresponding vector of degrees of freedom. We have

$$\begin{aligned} \boldsymbol{v}^T B^{-1} \boldsymbol{v} &\le 4 \inf_{\boldsymbol{v_H}\in\boldsymbol{V_H}} \left[\|\boldsymbol{v}_H\|_A^2 + \frac{\lambda}{(\kappa+1)} \|\boldsymbol{v}-\boldsymbol{v}_H\|_D^2 + \|\boldsymbol{v}-\boldsymbol{v}_H\|_A^2 \right] \\ &\le 4 \inf_{\boldsymbol{v_H}\in\boldsymbol{V_H}} \left[\|\boldsymbol{v}_H\|_A^2 + \frac{c_M n_z h^{-2}}{(\kappa+1)} \|v-v_H\|_{0,\alpha}^2 + \|v-v_H\|_a^2 \right] \\ &\le \left[1 + \frac{4c_{as} c_M n_z}{(\kappa+1)} \left(\frac{H}{h}\right)^2 + 4c_{as} \right] \|v\|_a^2 = K_{TG}\, \boldsymbol{v}^T A \boldsymbol{v}. \end{aligned}$$

Clearly, for $(\sqrt{\kappa}+1) \geq \dfrac{H}{h}$, and m satisfying (3.10), for example, $m \geq \dfrac{H}{h}\ln(H/h)$, the spectral equivalence is uniform with respect to mesh size and coefficient variation. □

5. Choice of coarse spaces and numerical tests

In this section, we present a number of tests that illustrate the robustness of the two–level methods with the polynomial smoother analyzed in the present paper all in accordance with Theorem 4.6. We consider the second order elliptic equation (4.15) with a mixture of Neumann and Dirichlet boundary conditions. The Dirichlet boundary conditions are imposed on the "east" and "west" vertical boundaries, i.e. $\Gamma_D = \Gamma_E \cup \Gamma_W$ of Ω. As we pointed out, the coefficient $\alpha(\boldsymbol{x})$ is piecewise constant and we assume that the fine triangulation of Ω is aligned with (resolves) all the coefficient discontinuities. In Fig. 1 we show an example of a fine grid $\mathcal{T}_h$, aligned with discontinuities.

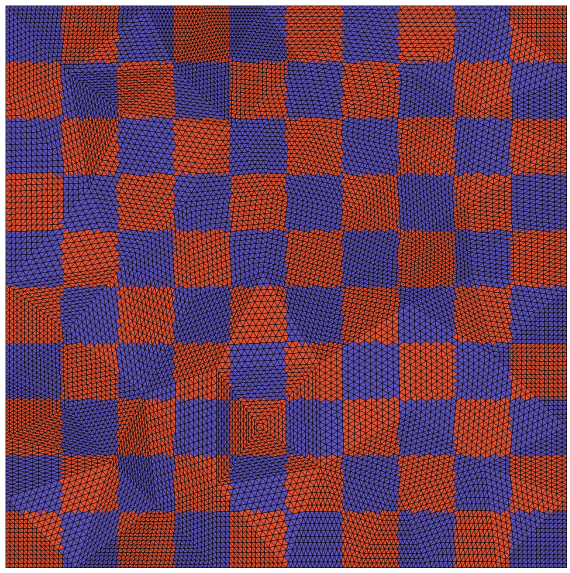
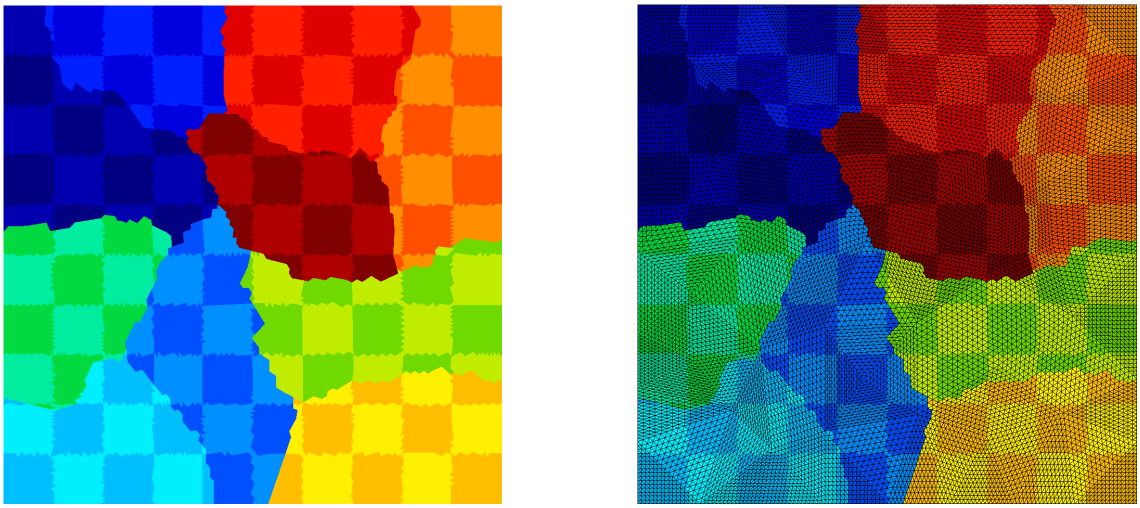


Figure 1: Checkerboard coefficient distribution on a mesh with 25600 elements and 13041 vertices.

5.1. **Coarse spaces.** We use element agglomeration to define "coarse elements" as illustrated in Fig. 2. and a variant of the spectral AMGe method (see, e.g. [9]) in the form presented in [7]. Briefly the main steps in such coarse space construction are:

- Partitioning of the degrees of freedom as a union of non-overlapping sets, $\{\mathcal{A}\}$ called *aggregates*. This is achieved by first partitioning the set of elements into agglomerated elements $\{\tau\}$ (union of fine-grid elements). We use graph partitioner (metis) applied to the graph having vertices the fine-grid elements with edges between two elements if they share a common interface. Then, we form aggregates $\mathcal{A}$, where each aggregate (a set of fine degrees of freedom) corresponds a unique agglomerated element $\tau = \tau_{\mathcal{A}}$ by distributing the shared fine degrees of freedom (fine-grid element vertices belonging to two or more agglomerated elements) to a unique aggregate.



(a) (b)

Figure 2: (a) An example of 10 aggregates; (b) the corresponding element agglomerates (unions of fine grid elements) constructed on a mesh with 6400 elements and 3321 vertices. The distribution of discontinuity of the PDE coefficient is not resolved by the agglomerates.

- Constructing a tentative interpolation matrix $\overline{P}$, defined for an agglomerate τ. Consider the local generalized eigenproblem,

$$A_\tau \boldsymbol{\varphi}_k = \theta_k D_\tau \boldsymbol{\varphi}_k,$$

  where A_τ is the local stiffness matrix corresponding to the agglomerated element τ, D_τ is its diagonal and $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{n_\mathcal{A}}$ with $n_\mathcal{A} = |\mathcal{A}|$ (cardinality of $\mathcal{A}$). Given a *spectral tolerance* θ, we select the eigenvectors $\{\varphi_k\}_{k=1}^{n_\theta}$, where n_θ is the largest integer for which the inequality $\theta_{n_\theta} < \theta$ holds. Extended by zero outside each $\mathcal{A}$, the vectors $\{\varphi_k\}_{k=1}^{n_\theta}$ form n_θ columns of the global tentative interpolation operator $\overline{P}$.
- Constructing the coarse space as the range of the interpolation matrix P, which is defined as

$$P = s_m\left(\lambda^{-1} D^{-1} A\right) \overline{P}.$$

  Here, as in the previous section, D is the diagonal of A, $\lambda \geq \|D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\|$ (e.g. $\lambda = \|D^{-1/2} A D^{-1/2}\|_\infty$), and $s_m(t)$ is the smoothed aggregation (SA) polynomial (cf., e.g., [8])

$$s_m(t) = \frac{(-1)^m}{(2m+1)} \frac{T_{2m+1}(\sqrt{t})}{\sqrt{t}}.$$

5.2. **Numerical tests.** We recall some of the notations and definitions which are used in the tables and figures in this section.

- N is the number of fine grid degrees of freedom;
- N_H is the number of coarse degrees of freedom;
- $nnz(X)$ is the number of the nonzero elements in a matrix X;
- $\widetilde{\varrho}_{TG}$ is the asymptotic convergence factor of the two grid method;

				$\widetilde{\rho}_{TG}$	
θ	N_H	$nnz(A_H)$	$oc(B)$	$a = 0.158$	$a = 0.2$
0.010	774	15,930	1.04	0.995	0.995
0.077	3,629	342,515	1.95	0.879	0.889
0.149	6,557	1,115,207	4.10	0.393	0.492

Table 1: Two-grid convergence, $m = 2$.

				$\widetilde{\rho}_{TG}$	
θ	N_H	$nnz(A_H)$	$oc(B)$	$a = 0.085$	$a = 0.1$
0.010	774	22,092	1.06	0.985	0.986
0.077	3,629	472,907	2.31	0.531	0.538
0.149	6,557	1,538,845	5.28	0.188	0.084

Table 2: Two-grid convergence when $m = 4$.

- $oc(B)$ is the *operator complexity* measure of the two-grid preconditioner B, defined as $oc(B) = \dfrac{nnz(A) + nnz(A_H)}{nnz(A)}$.

The first set of experiments are on a mesh with $102,400$ elements and $N = 51,681$ vertices using 300 agglomerated elements (AEs). We stop the iterations when the relative preconditioned residual norm is reduced by a factor of $\varepsilon = 10^{-8}$. The piecewise constant coefficient $\alpha(\boldsymbol{x})$ is distributed in a checkerboard fashion with values 1 and 10^6 as illustrated in Fig. 1.

The experiments are performed for $m = 2, 4, 6, 8$, $a \equiv \dfrac{1}{\kappa} = 0.158, 0.085, 0.055, 0.04$, and $a = 0.2, 0.1, 0.08, 0.06$ respectively. They are chosen such that the inequality (3.8) (with $\omega = 1$) holds:

$$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m = \left(\frac{1-\sqrt{a}}{1+\sqrt{a}}\right)^m < \frac{a}{1-a} = \frac{1}{\kappa-1}, \quad a = \frac{1}{\kappa}.$$

The same degree m is used for the polynomial smoother in the two-level algorithm and the smoother of the tentative interpolation matrix (it is smoothed out by $s_m\left(D^{-1}A\right)$). The number of non-zero entries of A, is $nnz(A) = 359,841$. We also show how the spectral tolerance θ and the polynomial degree m influences the convergence versus *operator complexity*. The results are presented in Table 1–4. It is evident from the results that the method can become fairly fast (in terms of convergence factors) at the expense of large operator complexity.

In the last experiment shown in Table 5, we illustrate the behavior of the method with respect to varying the contrast 10^c again distributed in a checkerboard fashion. As it is clearly seen, the two-grid method exhibits very good uniform two-grid convergence with operator complexity less than two.

				$\widetilde{\rho}_{TG}$	
θ	N_H	$nnz(A_H)$	$oc(B)$	$a = 0.055$	$a = 0.08$
0.010	774	29,448	1.08	0.965	0.969
0.077	3,629	636,671	2.78	0.205	0.179
0.149	6,557	2,074,291	6.76	0.202	0.026

Table 3: Two-grid convergence when $m = 6$.

				$\widetilde{\rho}_{TG}$	
θ	N_H	$nnz(A_H)$	$oc(B)$	$a = 0.04$	$a = 0.06$
0.010	774	37,618	1.10	0.926	0.933
0.077	3,629	808,357	3.25	0.197	0.111
0.149	6,557	2,632,755	8.32	0.193	0.028

Table 4: Two-grid convergence when $m = 8$.

c	-12	-9	-6	-3	0	3	6	9	12
N_H	2336	2336	2336	2339	2322	2322	2322	2322	2322
$oc(B)$	1.94	1.94	1.94	1.94	1.93	1.93	1.93	1.93	1.93
n_{it}	17	17	17	17	17	16	16	16	16
$\widetilde{\rho}_{TG}$	0.219	0.219	0.219	0.219	0.219	0.200	0.198	0.197	0.198

Table 5: Contrast independent two-grid convergence; coefficient jumps are 10^c. The method corresponds to spectral threshold $\theta = 0.045$, $m = 8$, and $a = 0.04$.

Acknowledgments

The authors thank Delyan Kalchev from Sofia University "St. Kliment Ohridski" for his help with the numerical experiments.

Appendix A.

The proof of Theorem 2.1 presented in this section is based on an equivalent result given in [1, p. 33, Equation (4.25)]. Let us also remark that in this section our considerations are on the interval $[-1, 1]$ and in addition, by *best polynomial approximation* we mean the best polynomial approximation in the norm $\|\cdot\|_\infty$ on $[-1, 1]$.

A.1. **An approximation result equivalent to Theorem 2.1.** We now formulate the result in [1] in the notation introduced earlier and show how the result in Theorem 2.1 can be derived from [1, p. 33, Equation (4.25)].

Theorem A.1 (G. Meinardus, [1])**.** *The polynomial $\widetilde{Q}_m \in \mathcal{P}_m$, of degree less than or equal to m, which furnishes the best approximation to $\dfrac{1}{t-a}$, $a > 1$ on $[-1,1]$ is given by:*

$$\widetilde{Q}_m(t) = \frac{1}{t-a}\left(1 - \frac{(a-\sqrt{a^2-1})^m}{a^2-1}\widetilde{R}_{m+1}(t)\right),$$

where

$$\widetilde{R}_{m+1}(t) = \left[(at-1)T_m(t) + \frac{\sqrt{a^2-1}}{m}(t^2-1)T_m'(t)\right].$$

The result we have just stated is for the best approximation to the function $\dfrac{1}{t-a}$, while to prove Theorem 2.1 we need such result for $\dfrac{1}{t+a}$. It is however easy to show that Theorem A.1 also provides the best polynomial approximation to $\dfrac{1}{t+a}$. Indeed, note that for any polynomial $p(t)$ of degree less than or equal to m, and for all $t \in [-1,1]$, there holds

$$p(-t) - \frac{1}{(-t)-a} = -\left(-p(-t) - \frac{1}{t+a}\right).$$

Further, for a function $g(t)$ continuous on $[-1,1]$ we also have,

$$\max\{g(t) \mid t \in [-1,1]\} = \max\{g(-t) \mid t \in [-1,1]\},$$

These two identities give that

$$\left\|p - \frac{1}{t-a}\right\|_\infty = \max_{t\in[-1,1]}\left|p(-t) - \frac{1}{(-t)-a}\right| = \max_{t\in[-1,1]}\left|(-p(-t)) - \frac{1}{t+a}\right|.$$

Since p was an arbitrary polynomial of degree less than or equal to m, we may take the infimum over all $p \in \mathcal{P}_m$. According to Theorem A.1 the left side is minimized for $p(t) = \widetilde{Q}_m(t)$. Therefore the right side should also be minimized for $p(t) = \widetilde{Q}_m(t)$. More precisely, we have

$$\left\|(-\widetilde{Q}_m(-t)) - \frac{1}{t+a}\right\|_\infty = \inf_{p\in\mathcal{P}_m}\max_{t\in[-1,1]}\left|(-p(-t)) - \frac{1}{t+a}\right| = \inf_{q\in\mathcal{P}_m}\left\|q - \frac{1}{t+a}\right\|_\infty, \tag{A.1}$$

which shows that the best polynomial approximation to $\dfrac{1}{t+a}$ on $[-1,1]$ is $(-\widetilde{Q}_m(-t))$.

A.2. **Proof of Theorem 2.1.** We need to show that for the polynomial $Q_m(t)$ defined as in (2.6) we have $Q_m(t) = (-\widetilde{Q}_m(-t))$. We use properties of Chebyshev polynomials to prove this identity. If we set $\alpha = \arccos t$ we have

$$\begin{aligned}(t^2-1)T_m'(-t) &= (-1)^m m \sin\alpha \sin m\alpha = \frac{(-1)^{m-1}m}{2}(\cos(m+1)\alpha - \cos(m-1)\alpha)\\ &= \frac{(-1)^{m-1}m}{2}(T_{m+1}(t) - T_{m-1}(t)).\end{aligned}$$

Recall that $T_k(-t) = (-1)^k T_k(t)$, $\delta = (a - \sqrt{a^2-1})$, and $2tT_m(t) = (T_{m+1}(t) + T_{m-1}(t))$. Therefore, we have

$$\begin{aligned}
\widetilde{R}_{m+1}(-t) &= -(at+1)T_m(-t) + \frac{\sqrt{a^2-1}}{m}(t^2-1)T_m'(-t) \\
&= (-1)^{m+1}(at+1)T_m(t) + \frac{(-1)^{m-1}\sqrt{a^2-1}}{2}(T_{m+1}(t) - T_{m-1}(t)) \\
&= \frac{(-1)^{m+1}}{2}\left[(a(T_{m+1}(t)+T_{m-1}(t)) + 2T_m(t) + \sqrt{a^2-1}(T_{m+1}(t)-T_{m-1}(t))\right] \\
&= \frac{(-1)^{m+1}}{2}\left[\delta^{-1}T_{m+1}(t) + 2T_m(t) + \delta T_{m-1}(t)\right] \\
&= \frac{(-1)^{m}}{2}\left[\eta^{-1}T_{m+1}(t) - 2T_m(t) + \eta T_{m-1}(t)\right].
\end{aligned}$$

Looking at the definition of $R_{m+1}(t)$, given in Theorem 2.1 (relation (2.7)) it is easily seen that $\widetilde{R}_{m+1}(-t) = \frac{(-1)^m}{2}R_{m+1}(t)$. Since $(\eta^{-1}-\eta)^2 = 4(a^2-1)$, we finally get

$$\begin{aligned}
(-\widetilde{Q}_m(-t)) &= -\frac{1}{-t-a}\left(1 - \frac{4\delta^m}{(\delta+\delta^{-1})^2}\widetilde{R}_{m+1}(-t)\right) = \frac{1}{t+a}\left(1 - \frac{2(-1)^m\delta^m}{(\delta^{-1}-\delta)^2}R_{m+1}(t)\right) \\
&= \frac{1}{t+a}\left(1 - \frac{2\eta^m}{(\eta-\eta^{-1})^2}R_{m+1}(t)\right) = Q_{m+1}(t).
\end{aligned}$$

Thus, $Q_m(t)$ and $(-\widetilde{Q}_m(-t))$ coincide and the proof is complete. □

Remark A.2. *It is also possible to prove directly that the polynomial in* (2.6) *is a polynomial of best approximation to* x^{-1} *by specifying the points of Chebyshev alternance. Such proof is however much more elaborate than the one presented here.*

References

[1] G. Meinardus. *Approximation of functions: Theory and numerical methods.* Expanded translation of the German edition. Translated by Larry L. Schumaker. Springer Tracts in Natural Philosophy, Vol. 13. Springer-Verlag New York, Inc., New York, 1967.

[2] T. J. Rivlin. *An introduction to the approximation of functions.* Dover Publications Inc., New York, 1981. Corrected reprint of the 1969 original, Dover Books on Advanced Mathematics.

[3] P. L. Chebyshev. Sur les polynômes réprésentant le mieux les valeurs des fonctions fractionnaires élémentaires pour les valeurs de la variable contenues entre deu limites données. In *Oeuvres*, volume II, pages 669–678. St Petersburg, 1907.

[4] Juan Galvis and Yalchin Efendiev. Domain decomposition preconditioners for multiscale flows in high-contrast media. *Multiscale Modeling & Simulation*, 8(4):1461–1483, 2010.

[5] Robert Scheichl, Panayot S. Vassilevski, and Ludmil T. Zikatanov. Weak approximation properties of elliptic projections with functional constraints. *Multiscale Modeling & Simulation*, 9(4):1677–1699, 2011.

[6] Marian Brezina, Caroline Heberton, Jan Mandel, and Petr Vaněk. An iterative method with convergence rate chosen a priori. Technical Report 140, University of Colorado

Denver, CCM, University of Colorado Denver, April 1999. An earlier version presented at the 1998 Copper Mountain Conference on Iterative Methods, April 1998.

[7] M. Brezina and P. S. Vassilevski. Smoothed aggregation spectral element agglomeration AMG: SA-ρAMGe. Technical Report LLNL-PROC-490083, Lawrence Livermore National Laboratory, June 29 2011. To appear in Proceedings of the 8th International Conference on "Large-Scale Scientific Computations", Sozopol, Bulgaria, June 6, 2011 through June 10, 2011.

[8] M. Brezina, P. Vaněk, and P. S. Vassilevski. An improved convergence analysis of smoothed aggregation algebraic multigrid. *Numerical Linear Algebra with Applications*, pages n/a–n/a, 2011.

[9] P. S. Vassilevski. *Multilevel block factorization preconditioners*. Springer, New York, 2008. Matrix-based analysis and algorithms for solving finite element equations.

[10] R. D. Falgout, P. S. Vassilevski, and L. T. Zikatanov. On two-grid convergence estimates. *Num. Lin. Alg. Appl.*, 12:471–494, 2005. Also available as LLNL technical report UCRL-JC-150807.

Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences Altenberger Str. 69, 4040 Linz, Austria.
E-mail address: johannes.kraus@oeaw.ac.at

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-560, Livermore, CA 94550, USA.
E-mail address: panayot@llnl.gov

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA.
E-mail address: ltz1@psu.edu