CONVERGENCE THEORY FOR THE EXACT INTERPOLATION SCHEME WITH APPROXIMATION VECTOR USED AS THE FIRST COLUMN OF THE PROLONGATOR: THE PARTIAL EIGENVALUE PROBLEM

PAVLA FRAŇKOVÁ*, MILAN HANUŠ†, HANA KOPINCOVÁ‡, ROMAN KUŽEL§, IVO MAREK¶, IVANA PULTAROVÁ‖, PETR VANĚK**, AND ZBYNĚK VASTL††

Abstract. We analyze our non-linear multigrid method presented in [3], applied to the partial eigenvalue problem $A\mathbf{x} = \lambda\mathbf{x}$ with symmetric, positive definite matrix A. The theory is then extended to the generalized eigenvalue problem $A\mathbf{x} = \lambda B\mathbf{x}$ with symmetric, positive definite matrices A and B. The extension follows from a trivial argument. We prove that a coarse-space of a modest size can be employed, provided its order of approximation is sufficiently high. The demands on the order of approximation of the coarse-space are moderate.

MSC: 65N55, 65N25

1. Introduction. In this paper, we analyze the non-linear multigrid of [3] applied to the partial eigenvalue problem $A\mathbf{x} = \lambda\mathbf{x}$ with symmetric, positive definite matrix A. The theory is then extended to the generalized eigenvalue problem $A\mathbf{x} = \lambda B\mathbf{x}$ with symmetric, positive definite matrices A and B. The convergence proof is local. This means that we prove rapid convergence assuming the input iterate is sufficiently close to the solution. However, the assumption on the input iterate is very weak; the Rayleigh quotient $R(\mathbf{x})$ of the approximation $\mathbf{x}$ only has to be smaller than the second eigenvalue. As a consequence, the convergence estimate has a reasonable global significance. The key result shows that if a small coarse-space (one obtained by aggressive coarsening) with sufficient order of approximation p is used, the rate of convergence $Q(\mathbf{x})$ satisfies

$$\lim_{cond(A)\to\infty} Q(\mathbf{x}) = 0. \tag{1.1}$$

The demand on the order of approximation of the coarse-space is by no means extreme. For A being obtained by a proper discretization of an H^1-equivalent form, we can use a coarse-space exactly approximating linear functions with the resolution $H = h^\alpha$, $\alpha > 1/2$, while still guaranteeing (1.1). Here, h is the resolution on the fine level.

*Department of mathematics, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic (frankova@ntis.zcu.cz)

†Department of mathematics, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic (mhanus@kma.zcu.cz)

‡Department of mathematics, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic (kopincov@kma.zcu.cz)

§Department of mathematics, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic (rkuzel@kma.zcu.cz)

¶Department of Mathematics, Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29, Praha 6, Czech Republic (marek@mbox.ms.mff.cuni.cz)

‖Department of Mathematics, Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29, Praha 6, Czech Republic (ivana@mat.fsv.cvut.cz); Mathematical Institute, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Prague 8, Czech Republic.

**Department of mathematics, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic (ptrvnk@kma.zcu.cz), phone: +420 377 632 697

††Department of mathematics, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic (zvastl@kma.zcu.cz)

Thus, assuming large $cond(A)$ (large problem), we prove extremely rapid convergence for a small coarse-space with sufficient order of approximation p. The convergence theory is fully algebraic; the verification of the assumptions means to prove certain approximation property of the prolongator (coarse-space). The approximation condition we need to be satisfied is a kind of weak approximation condition with A^β norm on the right-hand side and ℓ_2 (Euclidean) norm on the left-hand side. The case of interest is β being integer larger than 1. It is therefore necessary to control the approximation in ℓ_2-norm only.

The theory is developed for the eigenvalue problem $A\mathbf{x} = \lambda\mathbf{v}$ with symmetric, positive definite matrix A. Its extension to generalized eigenvalue problem $A\mathbf{x} = \lambda B\mathbf{x}$, with both matrices A and B being symmetric, positive definite, follows from a trivial argument.

The method of [3] is a special type of Exact Interpolation Scheme (EIS) proposed by Brandt with collaborators in [4, 5] and long before that, by Mandel and Sekerka in [20]. EIS is a non-linear multigrid scheme with the prolongator constructed so that the current approximation $\mathbf{x}$ belongs to its range. While the authors of [4, 5] use a quite complicated way of guaranteeing $\mathbf{x} \in$ Range P, we use a general purpose prolongator and simply add the current approximation $\mathbf{x}$ as its first column. The method was tested with extremely good results on problems of nuclear reactor criticality computations ([3]). We stress that in these experiments, the assumptions of our theory were not satisfied since the matrices A and B of the solved generalized eigenvalue problem $A\mathbf{x} = \lambda B\mathbf{x}$ were non-symmetric and the matrix B singular.

The paper provides theoretical grounds for the improvement of the method of [3]. Namely, it becomes clear that the small coarse-space can be used, provided that it has higher approximation order. This confirms the intuitive guess: in typical applications, good approximation of the first eigenvector can have coarser resolution, but requires high-order approximation, since the typical first eigenvector is very smooth. The approximation itself does not have to be very smooth since our theory requires good approximation only in ℓ_2-sense. We stress again that the demands on the order of approximation of the coarse-space are moderate.

The paper is organized as follows: in Section 2 we describe our algorithm. In the key Section 3, we give the convergence proof. At the end of Section 3 we sketch the verification of the assumptions of the convergence theorem for a problem with H^1-equivalent form. The Section 4 contains the generalization of the algorithm and the theory for the case of generalized eigenvalue problem.

2. Algorithm. Let A be a symmetric, positive definite $n \times n$ matrix with eigenvalues $\lambda_{min} = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_n = \lambda_{max}$. The particular case of interest is A being a finite element stiffness matrix.

We solve the partial eigenvalue problem:

$$\text{find } \lambda_1,\ \mathbf{v}_1 \in \mathbb{R}^n : A\mathbf{v}_1 = \lambda_1\mathbf{v}_1. \tag{2.1}$$

We consider linear injective prolongator $P : \mathbb{R}^m \to \mathbb{R}^n$, $m < n$. We are interested in aggressive coarsening, i.e. $m \ll n$.

Our two-level algorithm with evolving coarse-space for solving (2.1) proceeds as follows:

ALGORITHM 1.

1. *For given input iterate* $\mathbf{x} \in \mathbb{R}^n$*, construct/update the coarse-level matrices*

$$A_2(\mathbf{x}) = [\mathbf{x}|P]^T A[\mathbf{x}|P], \quad B_2(\mathbf{x}) = [\mathbf{x}|P]^T[\mathbf{x}|P], \tag{2.2}$$

see Remark 2.2,

2. *find the eigenvector* $\mathbf{v}^2$ *corresponding to the smallest eigenvalue of the coarse-level problem*

$$A_2(\mathbf{x})\mathbf{v}^2 = \lambda B_2(\mathbf{x})\mathbf{v}^2 \tag{2.3}$$

(if the coarse-level problem (2.3) is to be solved iteratively, natural initial guess for $\mathbf{v}^2$ *is the first canonical basis vector* $\mathbf{e}_1 = (1, 0, \ldots, 0)^T \in \mathbb{R}^{m+1}$*, see Remark 2.1)*

3. *prolongate* $\mathbf{v} \leftarrow [\mathbf{x}|P]\mathbf{v}^2$*,*
4. *post-smooth* $\mathbf{x}^{new} \leftarrow A^{-\nu}\mathbf{v}$*,*
5. *normalize* $\mathbf{x}^{new} \leftarrow 1/\|\mathbf{x}^{new}\|\,\mathbf{x}^{new}$*.*

REMARK 2.1. *Clearly, for the first canonical basis vector* $\mathbf{e}_1 = (1, 0, \ldots, 0)^T \in \mathbb{R}^{m+1}$ *it holds that*

$$\mathbf{x} = [\mathbf{x}|P]\mathbf{e}_1.$$

The vector $\mathbf{e}_1$ *is therefore (assuming* $\mathbf{x} \notin$ Range (P)*) a coarse level isomorphic counterpart of the current approximation* $\mathbf{x}$*; the coarse-level iteration started from* $\mathbf{e}_1$ *is therefore essentially (via the isomorphism* $[\mathbf{x}|P] : \mathbb{R}^{m+1} \to$ Range $([\mathbf{x}|P])$*) started from* $\mathbf{x}$*.*

REMARK 2.2. *Note that only the first column of the prolongator* $[\mathbf{x}|P]$ *changes from the iteration to the next iteration. Therefore only the first row and the first column of matrices* $A_2(\mathbf{x})$ *and* $B_2(\mathbf{x})$ *have to be recalculated in each iteration. If the coarse-level problem (2.3) is to be solved by the inverse power method*

$$\mathbf{v}^2 \leftarrow A_2(\mathbf{x})^{-1}B_2(\mathbf{x})\mathbf{v}^2,$$

the action of the inverse $A_2(\mathbf{x})^{-1}$ *can be performed using pre-calculated Choleski decomposition of the matrix* $A_2(\mathbf{x})$ *with the first column and the first row excluded. This matrix is the same in every iteration.*

The particular form (2.3) of the coarse-level problem will be derived in the next section.

For the sake of convenience, we assume that the input iterate $\mathbf{x}$ is scaled so that $\|\mathbf{x}\| = 1$ and the prolongated solution $\mathbf{v} = [\mathbf{x}|P]\mathbf{v}^2$ is scaled so that $\|\mathbf{v}\| = 1$. This requires only cosmetic changes in the algorithm with no impact on the result. In addition, we assume that the matrix A is scaled so that its lowest eigenvalue is equal to 1. This is possible without loss of generality, because due to the scaling of the vector $\mathbf{x}$ in Step 5, Algorithm 1 is independent on the scaling of A.

3. Local convergence estimate. Let $R(\mathbf{w})$ denote the Rayleigh quotient

$$R(\mathbf{w}) = \frac{\langle A\mathbf{w}, \mathbf{w}\rangle}{\|\mathbf{w}\|^2}, \quad \mathbf{w} \neq \mathbf{0}. \tag{3.1}$$

We will measure the error of approximation $\mathbf{w} \neq \mathbf{0}$ by the expression

$$r(\mathbf{x}) = \frac{\|A\mathbf{w} - R(\mathbf{w})\mathbf{x}\|}{\|\mathbf{w}\|}. \tag{3.2}$$

Clearly, $r(\mathbf{w}) = 0$ if and only if $\mathbf{w}$ is an eigenvector of A.

Let us set

$$T = \text{ Ker } (P^T). \tag{3.3}$$

We start our analysis with a simple observation: for the exact solution $\mathbf{v}^2$ of the coarse-level problem (2.3) it holds that

$$(A - R(\mathbf{v})I)\mathbf{v} \in T, \quad \mathbf{v} = [\mathbf{x}|P]\mathbf{v}^2. \tag{3.4}$$

To prove it, we first notice that the fine-level eigenvalue problem is equivalent to the non-linear equation

$$(A - R(\mathbf{v})I)\mathbf{v} = 0.$$

The coarse-level problem (2.3) is the result (equivalent) of the Galerkin formulation

$$\text{find } \mathbf{v} \in \text{ Range } ([\mathbf{x}|P]) : \langle (A - R(\mathbf{v})I)\mathbf{v}, \mathbf{w}\rangle = 0 \ \forall \mathbf{w} \in \text{ Range } ([\mathbf{x}|P]). \tag{3.5}$$

Indeed, (3.5) is equivalent to the problem

$$\begin{aligned} &\text{find } \mathbf{v}^2 \in \mathbb{R}^{m+1} : \\ &\left\langle \left(A - \frac{\langle A[\mathbf{x}|P]\mathbf{v}^2, [\mathbf{x}|P]\mathbf{v}^2\rangle}{\langle [\mathbf{x}|P]\mathbf{v}^2, [\mathbf{x}|P]\mathbf{v}^2\rangle} I \right) [\mathbf{x}|P]\mathbf{v}^2, [\mathbf{x}|P]\mathbf{w}^2 \right\rangle = 0 \ \forall \mathbf{w^2} \in \mathbb{R}^{m+1}, \end{aligned}$$

that, after transposing prolongators $[\mathbf{x}|P]$ in the right arguments of the inner products, becomes

$$\left\langle \left(A_2(\mathbf{x}) - \frac{\langle A_2(\mathbf{x})\mathbf{v}^2, \mathbf{v}^2\rangle}{\langle B_2(\mathbf{x})\mathbf{v}^2, \mathbf{v}^2\rangle} B_2(\mathbf{x}) \right) \mathbf{v}^2, \mathbf{w}^2 \right\rangle = 0 \ \forall \mathbf{w}^2 \in \mathbb{R}^{m+1}$$

with matrices $A_2(\mathbf{x})$ and $B_2(\mathbf{x})$ given by (2.2). The above identity holds if and only if the left argument of the above inner product is zero, which happens if and only if $\mathbf{v}^2$ is an eigenvector of (2.3). Thus, (3.5) and (2.3) are equivalent. As a consequence, for the solution $\mathbf{v}^2$ of (2.3), the prolongated vector $\mathbf{v} = [\mathbf{x}|P]\mathbf{v}^2$ satisfies (3.5) and therefore

$$(A - R(\mathbf{v})I)\mathbf{v} \in \text{ Range }^{\perp}([\mathbf{x}|P]) = \text{ Ker } ([\mathbf{x}|P]^T) \subset \text{ Ker } (P^T),$$

proving (3.4).

We are interested in estimating $r(\mathbf{x}^{new}) = r(A^{-\nu}\mathbf{v})$, $\mathbf{v} = [\mathbf{x}|P]\mathbf{v}^2$, in terms of $r(\mathbf{x})$. To this end, we first estimate $r(A^{-\nu}\mathbf{v})$ in terms of $r(\mathbf{v})$. We notice that in the expression of $r(\mathbf{w})$ in (3.2), $R(\mathbf{w})\mathbf{w}$ is the projection of $A\mathbf{w}$ onto span$\{\mathbf{w}\}$ orthogonal in Euclidean inner product. Therefore, by the minimizing property of the orthogonal projection,

$$\begin{aligned} \|(A - R(A^{-\nu}\mathbf{v})I)A^{-\nu}\mathbf{v}\| &\leq \|(A - R(\mathbf{v})I)A^{-\nu}\mathbf{v}\| \\ &= \|A^{-\nu}(A - R(\mathbf{v})I)\mathbf{v}\| \\ &\leq \|A^{-\nu}\|_T \|(A - R(\mathbf{v})I)\mathbf{v}\|, \end{aligned} \tag{3.6}$$

since $(A - R(\mathbf{v})I)\mathbf{v} \in T$ by (3.4). Here, $\|A^{-\nu}\|_T$ is the operator norm on the subspace T defined as

$$\|A^{-\nu}\|_T = \sup_{\mathbf{v} \in T \setminus \{\mathbf{0}\}} \frac{\|A^{-\nu}\mathbf{v}\|}{\|\mathbf{v}\|}.$$

Thus, we conclude that

$$r(\mathbf{x}^{new}) = r(A^{-\nu}\mathbf{v}) \leq \|A^{-\nu}\|_T \frac{\|\mathbf{v}\|}{\|A^{-\nu}\mathbf{v}\|} r(\mathbf{v}). \tag{3.7}$$

In the view of (3.6), the inverse power method given by the action of A^{-1} works as a smoother and the entire method resembles the standard linear variational two-level multigrid. The purpose of the coarse-level correction is to guarantee that the residual in (3.4) belongs to space $T =$ Ker (P^T), where the smoother A^{-1} is (hopefully) very efficient in the reduction of the norm of the residual. The efficiency of the smoother A^{-1} in the reduction of the residual that belongs to T is the subject of the following lemma.

Our first key lemma allows us to estimate $\|A^{-\nu}\|_T$ using approximation properties of the prolongator P. The proof is based on the orthogonality trick of Céa ([2]) and is a variation on the theme by Brandt [1].

LEMMA 3.1. *Let A be a symmetric, positive definite $n \times n$ matrix scaled so that $\lambda_{min} = 1$ and $P : \mathbb{R}^m \to \mathbb{R}^n$, $m < n$, a linear injective prolongator. Consider $\alpha \in (0, 1]$ and $\beta > 0$. We assume there is a linear mapping $Q : \mathbb{R}^n \to$ Range (P) such that*

$$\forall \mathbf{w} \in \mathbb{R}^n : \|\mathbf{w} - Q\mathbf{w}\| \leq \frac{C}{\lambda_{max}^{\frac{\alpha\beta}{2}}} \|\mathbf{w}\|_{A^\beta}. \tag{3.8}$$

Then the operator norm $\|A^{-\beta/2}\|_T$ on the subspace T defined in (3.3) satisfies

$$\|A^{-\beta/2}\|_T \leq \frac{C}{\lambda_{max}^{\frac{\alpha\beta}{2}}}. \tag{3.9}$$

REMARK 3.2. *In specific applications, parameter α is related to the mesh resolution H of the coarse level space ($H = h^\alpha$, h being the fine-level resolution). Parameter β corresponds to $p-$approximation quality of the coarse space ($\beta = p+1$).*

REMARK 3.3. *If Q is chosen to be the orthogonal projection onto* Range (P), *it holds that*

$$\|(I - Q)\mathbf{w}\| \leq \|\mathbf{w}\| \leq \lambda_{min}^{\beta/2} \|\mathbf{w}\|_{A^\beta} = \|\mathbf{w}\|_{A^\beta}.$$

The condition (3.8) can be therefore always satisfied with $C/\lambda_{max}^{(\alpha\beta)/2} \leq 1$ and (3.9) gives

$$\|A^{-\beta/2}\|_T \leq 1.$$

Proof. Let $\mathbf{w} \in T$. As $T =$ Ker $(P^T) =$ Range $(P)^\perp$ and Range $(Q) =$ Range (P), it holds that $\mathbf{w} \perp$ Range (Q) and we can estimate using (3.8) and Cauchy-Schwarz inequality,

$$\begin{aligned}
\|A^{-\beta/2}\mathbf{w}\|^2 &= \langle A^{-\beta}\mathbf{w}, \mathbf{w}\rangle \\
&= \langle (I - Q)A^{-\beta}\mathbf{w}, \mathbf{w}\rangle \\
&\leq \|(I - Q)A^{-\beta}\mathbf{w}\|\|\mathbf{w}\| \\
&\leq \|\mathbf{w}\| \frac{C}{\lambda_{max}^{\frac{\alpha\beta}{2}}} \|A^{-\beta}\mathbf{w}\|_{A^\beta} \\
&= \|\mathbf{w}\| \frac{C}{\lambda_{max}^{\frac{\alpha\beta}{2}}} \|A^{-\beta/2}\mathbf{w}\|.
\end{aligned}$$

Dividing both sides of the above estimate by $\|A^{-\beta/2}\mathbf{w}\|$ yields (3.9). □

The fact that the estimate (3.9) improves with growing λ_{max} (that, assuming fixed $\lambda_{min} = 1$, indicates the size of the problem) is, in the context of the final convergence theorem, addressed in Remark 3.7.

EXAMPLE 1. Let us say that we want to guarantee optimal work for a single smoother A^{-1}. Here A is a stiffness matrix obtained by finite difference discretization of the Laplace operator. We choose $\beta = 2$ and the piecewise linear coarse-space (in the discrete sense) with the resolution $H = h^\alpha$, where h is the resolution of the fine-level mesh. The matrix A^2 is then the finite difference discretization of the biharmonic operator and (3.8) holds with uniform C [2]. In this case, we get

$$\|A^{-1}\|_T \leq \frac{C}{\lambda_{max}^\alpha}.$$

If we use the discrete analogue of cubic elements on the coarse-space, (3.8) holds with $\beta = 4$ and we have

$$\|A^{-2}\|_T \leq \frac{C}{\lambda_{max}^{2\alpha}}.$$

LEMMA 3.4. *Let* $\mathbf{x}$ *be the iterate on the entry of Algorithm 1 and* $\mathbf{v} = [\mathbf{x}|P]\mathbf{v}^2$ *the prolongated solution of the coarse-level problem (2.3). Then*

$$R(\mathbf{v}) = \inf_{\mathbf{u} \in \text{ Range } ([\mathbf{x}|P])} R(\mathbf{u}) \leq R(\mathbf{x}). \tag{3.10}$$

Proof. We will derive the coarse-level problem (2.3) by minimizing $R(\mathbf{u})$ on the subspace Range $([\mathbf{x}|P])$.

The minimizer $\mathbf{v}$ of $R(\mathbf{u})$ on Range $([\mathbf{x}|P])$ satisfies the condition

$$\frac{d}{dt}\,|_{t=0} R(\mathbf{v} + t\mathbf{w}) = 0 \;\forall \mathbf{w} \in \text{ Range } ([\mathbf{x}|P]).$$

We have

$$\begin{aligned}
\frac{d}{dt}\,|_{t=0} R(\mathbf{v} + t\mathbf{w}) &= \frac{d}{dt}|_{t=0} \frac{\langle A(\mathbf{v} + t\mathbf{w}), \mathbf{v} + t\mathbf{w}\rangle}{\|\mathbf{v} + t\mathbf{w}\|^2} \\
&= \frac{2\langle A\mathbf{v}, \mathbf{w}\rangle \|\mathbf{v}\|^2 - 2\langle \mathbf{v}, \mathbf{w}\rangle \|\mathbf{v}\|_A^2}{\|\mathbf{v}\|^4} \\
&= \frac{2}{\|\mathbf{v}\|^2} \langle A\mathbf{v} - R(\mathbf{v})\mathbf{v}, \mathbf{w}\rangle .
\end{aligned}$$

Thus the minimizer $\mathbf{v} \in$ Range $([\mathbf{x}|P])$ is the solution of the Galerkin problem (3.5) that leads to the coarse-level problem (2.3). Let $\mathbf{v}^{2,i}$ be the generalized eigenvectors of (2.3) and $\lambda^{2,i}$ the corresponding eigenvalues. We assume the natural numbering $\lambda^{2,i} \leq \lambda^{2,i+1}$. Clearly, $\lambda^{2,i} = R([\mathbf{x}|P]\mathbf{v}^{2,i})$. The eigenvalues $\lambda^{2,i} = R([\mathbf{x}|P]\mathbf{v}^{2,i})$ are (all) extremes and saddle points of $R(\mathbf{v})$ on Range $[\mathbf{x}|P]$. The value $\lambda^{2,1} = R([\mathbf{x}|P]\mathbf{v}^{2,1}) = R([\mathbf{x}|P]\mathbf{v}^2)$ is therefore the global minimum. □

The following lemma puts into relation the expressions $r(\mathbf{x})$ and $r(\mathbf{v})$.

LEMMA 3.5. *Let* A *be an* $n \times n$ *symmetric, positive definite matrix with unique minimal eigenvalue* λ_{min} *and* $\{\mathbf{v}_i\}$ *the orthonormal system of its eigenvectors, with*

λ_i being the eigenvalue corresponding to $\mathbf{v}_i$. Assume A is scaled so that $\lambda_{min} = 1$ and the eigenvalues are numbered so that $\lambda_{min} = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_n = \lambda_{max}$. In addition, for the sake of convenience, we assume that the input iterate $\mathbf{x}$ and the prolongated solution of the coarse-level problem $\mathbf{v} = [\mathbf{x}|P]\mathbf{v}^2$ are normalized such that $\|\mathbf{x}\| = \|\mathbf{v}\| = 1$. Then, under the assumption that $R(\mathbf{x}) < \lambda_2$, it holds that

$$r^2(\mathbf{v}) \leq \frac{\lambda_{max} - R(\mathbf{v})}{\lambda_2 - R(\mathbf{x})} r^2(\mathbf{x}) \leq \frac{\lambda_{max}}{\lambda_2 - R(\mathbf{x})} r^2(\mathbf{x}). \tag{3.11}$$

Proof. The vectors $\mathbf{x}$ and $\mathbf{v}$ can be expressed as a linear combination of the eigenvectors $\{\mathbf{v}_i\}$ as

$$\mathbf{x} = \sum_i c_i(\mathbf{x})\mathbf{v}_i, \qquad \mathbf{v} = \sum_i c_i(\mathbf{v})\mathbf{v}_i$$

with

$$\sum_i c_i^2(\mathbf{x}) = \sum_i c_i^2(\mathbf{v}) = 1. \tag{3.12}$$

Further,

$$\begin{aligned} R(\mathbf{x}) &= \frac{\langle A\mathbf{x}, \mathbf{x}\rangle}{\|\mathbf{x}\|^2} \\ &= \sum_i c_i(\mathbf{x})^2 \lambda_i \\ &= c_1^2(\mathbf{x}) + \sum_{i>1} c_i^2(\mathbf{x})\lambda_i \\ &= 1 - \sum_{i>1} c_i^2(\mathbf{x}) + \sum_{i>1} c_i^2(\mathbf{x})\lambda_i \\ &= 1 + \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1). \end{aligned} \tag{3.13}$$

Similarly, by replacing $\mathbf{x}$ by $\mathbf{v}$, we get

$$R(\mathbf{v}) = 1 + \sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1). \tag{3.14}$$

From Lemma 3.4, it follows that $R(\mathbf{v}) \leq R(\mathbf{x})$ and therefore,

$$\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1) \leq \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1). \tag{3.15}$$

Further, since $R(\mathbf{x})\mathbf{x}$ is the orthogonal projection of $A\mathbf{x}$ onto span$\{\mathbf{x}\}$, we have by Pythagoras theorem and (3.12),

$$\begin{aligned} r^2(\mathbf{x}) &= \frac{\|A\mathbf{x} - R(\mathbf{x})\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \\ &= \frac{\|A\mathbf{x}\|^2 - R^2(\mathbf{x})\|\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2} - R^2(\mathbf{x}) \\
&= c_1^2(\mathbf{x}) + \sum_{i>1} c_i^2(\mathbf{x})\lambda_i^2 - \left(c_1^2(\mathbf{x}) + \sum_{i>1} c_i^2(\mathbf{x})\lambda_i \right)^2 \\
&= 1 - \sum_{i>1} c_i^2(\mathbf{x}) + \sum_{i>1} c_i^2\lambda_i^2(\mathbf{x}) - \left(1 - \sum_{i>1} c_i^2(\mathbf{x}) + \sum_{i>1} c_i^2(\mathbf{x})\lambda_i \right)^2 \\
&= 1 + \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i^2 - 1) - \left(1 + \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2 \\
&= \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i^2 - 1) - 2\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) - \left(\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2 \\
&= \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1)(\lambda_i + 1) - 2\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) - \left(\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2 \\
&= \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1)(\lambda_i + 1 - 2) - \left(\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2 \\
&= \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1)^2 - \left(\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2 .
\end{aligned}
$$

By replacing $\mathbf{x}$ in the previous estimate by $\mathbf{v}$ we get

$$
r^2(\mathbf{v}) = \sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1)^2 - \left(\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1) \right)^2 .
$$

Assume for now that $r(\mathbf{x}) \neq 0$. Using the two identities above, (3.13), (3.14) and the inequality (3.15), we estimate

$$
\begin{aligned}
\frac{r^2(\mathbf{v})}{r^2(\mathbf{x})} &= \frac{\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1)^2 - \left(\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1) \right)^2}{\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1)^2 - \left(\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2} \\
&\leq \frac{(\lambda_n - 1)\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1) - \left(\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1) \right)^2}{(\lambda_2 - 1)\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) - \left(\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right)^2} \\
&= \frac{\left(\lambda_n - 1 - \sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1) \right) \sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1)}{\left(\lambda_2 - 1 - \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1) \right) \sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1)} \\
&= \frac{(\lambda_n - R(\mathbf{v}))\sum_{i>1} c_i^2(\mathbf{v})(\lambda_i - 1)}{(\lambda_2 - R(\mathbf{x}))\sum_{i>1} c_i^2(\mathbf{x})(\lambda_i - 1)} \\
&\leq \frac{\lambda_n - R(\mathbf{v})}{\lambda_2 - R(\mathbf{x})},
\end{aligned}
$$

which completes the proof of (3.11) for $r(\mathbf{x}) \neq 0$. If $r(\mathbf{x}) = 0$, $\mathbf{x}$ is an eigenvector of A, hence a fixed-point of the coarse-level correction step, thus $\mathbf{v}$ is an eigenvector, $r(\mathbf{v}) = 0$ and (3.11) holds trivially. □

Assuming $R(\mathbf{x}) < \lambda_2$, Lemma 3.5 gives the estimate

$$r^2(\mathbf{v}) \leq \frac{\lambda_{max}}{\lambda_2 - R(\mathbf{x})} r^2(\mathbf{x}).$$

From here and (3.7) we get

$$r^2(A^{-\nu}\mathbf{v}) \leq \|A^{-\nu}\|_T^2 \frac{\|\mathbf{v}\|^2}{\|A^{-\nu}\mathbf{v}\|^2} \frac{\lambda_{max}}{\lambda_2 - R(\mathbf{x})} r^2(\mathbf{x}). \tag{3.16}$$

Since $\lambda_1 = 1$ and $\|\mathbf{v}\| = 1$, for $\mathbf{v}$ expressed as $\mathbf{v} = \sum_i c_i(\mathbf{v})\mathbf{v}_i$ we have

$$\frac{\|\mathbf{v}\|^2}{\|A^{-\nu}\mathbf{v}\|^2} = \frac{1}{c_1^2(\mathbf{v}) + \sum_{i>1} c_i^2(\mathbf{v})\lambda_i^{-2\nu}} \leq \frac{1}{c_1^2(\mathbf{v})}. \tag{3.17}$$

By Lemma 3.4 and $\|\mathbf{v}\|^2 = \sum_i c_i^2(\mathbf{v}) = 1$ we get

$$R(\mathbf{x}) \geq R(\mathbf{v}) = c_1^2(\mathbf{v}) + \sum_{i>1} c_i^2(\mathbf{v})\lambda_i \geq c_1^2(\mathbf{v}) + \lambda_2 \sum_{i>1} c_i^2(\mathbf{v}) = c_1^2(\mathbf{v}) + \lambda_2(1 - c_1^2(\mathbf{v})).$$

Thus, $0 > R(\mathbf{x}) - \lambda_2 \geq c_1^2(\mathbf{v})(1 - \lambda_2)$, hence $0 < \lambda_2 - R(\mathbf{x}) \leq c_1^2(\mathbf{v})(\lambda_2 - 1)$ and therefore,

$$\frac{1}{c_1^2(\mathbf{v})} \leq \frac{\lambda_2 - 1}{\lambda_2 - R(\mathbf{x})}.$$

The above estimate, (3.17) and (3.16) give

$$r(A^{-\nu}\mathbf{v}) \leq \sqrt{\lambda_{max}} \|A^{-\nu}\|_T \frac{\sqrt{\lambda_2 - 1}}{\lambda_2 - R(\mathbf{x})} r(\mathbf{x}). \tag{3.18}$$

Now we are ready to formulate our final convergence theorem. In this theorem, we do not assume anymore that the smallest eigenvalue satisfies $\lambda_{min} = 1$.

Define the condition number of the symmetric, positive definite matrix A by

$$cond(A) = \frac{\lambda_{max}}{\lambda_{min}}.$$

THEOREM 3.6. *Let A be a symmetric, positive definite $n \times n$ matrix with unique minimal eigenvalue λ_{min}. We assume the eigenvalues are numbered so that $\lambda_{min} = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n = \lambda_{max}$. Consider $\alpha \in (0, 1]$ and $\beta > 0$ such that $\alpha\beta - 1 > 0$. We assume there is a linear mapping $Q : \mathbb{R}^n \to$ Range (P) such that*

$$\forall \mathbf{u} \in \mathbb{R}^n : \lambda_{min}^{\frac{\beta}{2}} \|\mathbf{u} - Q\mathbf{u}\| \leq \frac{C}{cond(A)^{\frac{\alpha\beta}{2}}} \|\mathbf{u}\|_{A^\beta}.$$

In addition, assume that the input iterate $\mathbf{x}$ is reasonably close to the first eigenvector $\mathbf{v}_1$ so that $R(\mathbf{x}) \in [\lambda_1, \lambda_2)$. Then the result $\mathbf{x}^{new}$ on the exit of Algorithm 1 with $\nu \geq \beta/2$ satisfies the estimate

$$r(\mathbf{x}^{new}) = r(A^{-\nu}\mathbf{v}) \leq Q(\mathbf{x}) r(\mathbf{x}), \quad Q(\mathbf{x}) = \frac{C}{cond(A)^{\frac{\alpha\beta-1}{2}}} \frac{\sqrt{\frac{\lambda_2}{\lambda_{min}} - 1}}{\frac{\lambda_2 - R(\mathbf{x})}{\lambda_{min}}}.$$

The rate of convergence $Q(\mathbf{x})$ satisfies

$$\lim_{cond(A)\to\infty} Q(\mathbf{x}) = 0. \tag{3.19}$$

Proof. For the scaled matrix $1/\lambda_{min}\, A$, the proof follows directly from Lemma 3.1 and (3.18). (We make the substitution $A \leftarrow 1/\lambda_{min}\, A$ in (3.18), (3.8) and (3.9). For the scaled matrix $1/\lambda_{min}\, A$, $\lambda_{max} = \lambda_{max}(A)$ becomes $\lambda_{max}(1/\lambda_{min}\, A) = cond(A)$, λ_2 becomes $\lambda_2(1/\lambda_{min}\, A) = \lambda_2/\lambda_{min}$, $\|\mathbf{w}\|_A$ becomes $\|\mathbf{w}\|_{(1/\lambda_{miin}\, A)} = \lambda_{min}^{-1/2}\|\mathbf{w}\|_A$ and $R(\mathbf{w})$ becomes

$$\frac{\langle (1/\lambda_{min}\, A)\mathbf{w}, \mathbf{w}\rangle}{\|\mathbf{w}\|^2} = 1/\lambda_{min} R(\mathbf{w}), \quad \mathbf{w} \neq \mathbf{0}.)$$

The proof is immediately extended to the case of unscaled matrix A, once we notice that the Algorithm 1 works independently of the scaling of A. □

REMARK 3.7. *On the first sight, the statement of Theorem 3.6 might be perceived as surprising, since the rate of convergence of the analyzed method improves with the increasing condition number of the matrix A. This result ceases to be hard to accept once we acknowledge the following: the crucial element that determines the rate of convergence is the operator norm $\|(1/\lambda_{min}\, A)^{-\nu}\|_T$. Ideally, $T =$ Ker (P^T) is a high-energy space that behaves (with respect to the action of $A^{-\nu}$) similarly as the invariant subspace*

$$T = \text{span}\{\mathbf{v}_i : \lambda_i \geq C\lambda_{max}\},\ C \in (0, 1].$$

Obviously, for this model subspace we get

$$\|(1/\lambda_{min}\, A)^{-\nu}\|_T \leq \left(\frac{1}{C\, cond(A)}\right)^{\nu}.$$

The improvement of the rate of convergence with growing $cond(A)$ is therefore natural; it can be simply explained by the fact that the action of A^{-1}, if understood as a smoother, is extremely powerful.

EXAMPLE 2. Assume that A has been obtained by a finite difference approximation of the Laplace operator and $\lambda_{min} \approx 1$. Then A^2 is the finite difference approximation of the biharmonic operator. Assume we use a coarse-space based on piecewise linear interpolation with the resolution $H = h^{\alpha}$. Then we can prove the weak approximation property

$$\forall \mathbf{u} \in \mathbb{R}^n \exists \mathbf{v} \in \mathbb{R}^n :\ \|\mathbf{u} - P\mathbf{v}\|^2 \leq \frac{C}{\lambda_{max}^{2\alpha}} \|\mathbf{u}\|^2_{A^2}.$$

The assumption of Theorem 3.6 therefore holds with $\beta = 2$. Clearly, assuming $\alpha > 1/2$ we have $\alpha\beta - 1 > 0$ and therefore (3.19) holds for $\mathbf{x}$ such that $R(\mathbf{x}) < \lambda_2$.

If we use the discrete analogue of cubic elements, it holds that

$$\forall \mathbf{u} \in \mathbb{R}^n \exists \mathbf{v} \in \mathbb{R}^n :\ \|\mathbf{u} - P\mathbf{v}\|^2 \leq \frac{C}{\lambda_{max}^{4\alpha}} \|\mathbf{u}\|^2_{A^4}.$$

The weak approximation condition in the assumption of Theorem 3.6 holds therefore with $\beta = 4$. Then $\alpha\beta - 1 > 0$ (and thereby, (3.19)) holds for $\alpha > 1/4$ and $R(\mathbf{x}) < \lambda_2$.

In other words, for a problem that represents a serious challenge in the context of the contemporary (and the near future) scientific computing (say, 10^8 of degrees of freedom), we need a coarse space of the size that ranks in hunderds of degrees of freedom, that is, a coarse-space of a negligible size. This claim is certainly supported by the presented theory, but requires detailed justification both on the theorethical and the computational front.

4. Generalized eigenvalue problem. Consider the generalized eigenvalue problem

$$A\mathbf{x} = \lambda B\mathbf{x}, \text{ or, } B^{-1}A\mathbf{x} = \lambda\mathbf{x}, \tag{4.1}$$

with both A and B being symmetric, positive definite matrices. The matrix $B^{-1}A$ is a symmetric operator in the Hilbert space $(\mathbb{R}^n,\ \langle B\cdot,\cdot\rangle)$. Therefore, both the algorithm and the theory can be, in principle, (that is, regardless of algorithmic feasibility) adapted for the generalized eigenvalue problem simply by replacing systematically $\|\cdot\| \leftarrow \|\cdot\|_B$, $\langle\cdot,\cdot\rangle \leftarrow \langle B\cdot,\cdot\rangle$ and $A \leftarrow B^{-1}A$ in the algorithm, derivation of the coarse-level matrices in (3.5) and in the theory, including the inherent norm $\|\cdot\|_{A^\beta} = \|A^{\beta/2}\cdot\|$ in (3.8). It turns out that the resulting algorithm is the one of [3] with no need to construct B^{-1} and, as such, is indeed algorithmically feasible.

The transpose ^{T} becomes a B-adjoint operator * and the orthogonal complement $^\perp$ a B-orthogonal complement $^{\overset{\perp}{B}}$. As a consequence, the subspace $T = \text{ Ker }(P^T)$ changes to $T = \text{ Ker }(P^*) = \text{ Ker }(P^TB)$. The coarse-level matrices $A_2(\mathbf{x})$ and $B_2(\mathbf{x})$ are then the result of coarsening of $A \leftarrow B^{-1}A$ and I in (3.5) with the replaced inner product $\langle\cdot,\cdot\rangle \leftarrow \langle B\cdot,\cdot\rangle$. The coarse-level matrices are

$$A_2(\mathbf{x}) = [\mathbf{x}|P]^T A_{orig}[\mathbf{x}|P], \quad B_2(\mathbf{x}) = [\mathbf{x}|P]^T B[\mathbf{x}|P]; \tag{4.2}$$

B^{-1} is therefore involved neither in the construction of the coarse-level matrices, nor elsewhere in the algorithm. Here, A_{orig} stands for matrix A before the replacement $A \leftarrow B^{-1}A$. To prove it, we rewrite (3.5) as (denoting A_{orig} as A again)

$$\text{find } \mathbf{v}^2 \in \mathbb{R}^{m+1}:$$
$$\left\langle B\left(B^{-1}A - \frac{\langle BB^{-1}A[\mathbf{x}|P]\mathbf{v}^2, [\mathbf{x}|P]\mathbf{v}^2\rangle}{\langle B[\mathbf{x}|P]\mathbf{v}^2, [\mathbf{x}|P]\mathbf{v}^2\rangle} I\right)[\mathbf{x}|P]\mathbf{v}^2,\ [\mathbf{x}|P]\mathbf{w}^2\right\rangle\ \forall \mathbf{w}^2 \in \mathbb{R}^{m+1}.$$

This is equivalent to

$$\left\langle \left(A_2(\mathbf{x}) - \frac{\langle A_2(\mathbf{x})\mathbf{v}^2, \mathbf{v}^2\rangle}{\langle B_2(\mathbf{x})\mathbf{v}^2, \mathbf{v}^2\rangle} B_2(\mathbf{x})\right)\mathbf{v}^2, \mathbf{w}^2\right\rangle = 0 \quad \forall \mathbf{w}^2 \in \mathbb{R}^{m+1}$$

with matrices $A_2(\mathbf{x})$ and $B_2(\mathbf{x})$ given by (4.2). The above identity holds if and only if the left argument of the above inner product is equal to zero. This happens if and only if $\mathbf{v}^2$ is the generalized eigenvector of the coarse-level generalized eigenvalue problem

$$A_2(\mathbf{x})\mathbf{v}^2 = \lambda B_2(\mathbf{x})\mathbf{v}^2.$$

The inverse power method $\mathbf{x} \leftarrow A^{-\nu}\mathbf{v}$ in Step 4 of Algorithm 1 becomes

$$\mathbf{x} \leftarrow (A^{-1}B)^\nu\mathbf{v},$$

that is

$$
\begin{aligned}
&\mathbf{x} \leftarrow \mathbf{v};\\
&\text{for } i = 1, \ldots, \nu \text{ do}\\
&\quad \{\\
&\qquad \mathbf{b} \leftarrow B\mathbf{x};\\
&\qquad \text{solve } A\mathbf{x} = \mathbf{b}\\
&\quad \}.
\end{aligned}
$$

For the readers convenience, we now summarize the resulting algorithm and the convergence theorem.

ALGORITHM 2.

1. *For given input iterate* $\mathbf{x} \in \mathbb{R}^n$*, construct/update the coarse-level matrices*

$$A_2(\mathbf{x}) = [\mathbf{x}|P]^T A[\mathbf{x}|P], \quad B_2(\mathbf{x}) = [\mathbf{x}|P]^T B[\mathbf{x}|P],$$

   see Remark 2.2,

2. *find the eigenvector* $\mathbf{v}^2$ *corresponding to the smallest eigenvalue of the coarse-level problem*

$$A_2(\mathbf{x})\mathbf{v}^2 = \lambda B_2(\mathbf{x})\mathbf{v}^2, \tag{4.3}$$

   (if the coarse-level problem (2.3) is to be solved iteratively, natural initial guess for $\mathbf{v}^2$ *is the first canonical basis vector* $\mathbf{e}_1 = (1, 0, \ldots, 0)^T \in \mathbb{R}^{m+1}$*, see Remark 2.1),*

3. *prolongate* $\mathbf{v} \leftarrow [\mathbf{x}|P]\mathbf{v}^2$*,*

4. *post-smooth*

$$
\begin{aligned}
&\mathbf{x}^{aux} \leftarrow \mathbf{v};\\
&\textit{for } i = 1, \ldots, \nu \textit{ do}\\
&\quad \{\\
&\qquad \mathbf{b} \leftarrow B\mathbf{x}^{aux};\\
&\qquad \textit{solve } A\mathbf{x}^{aux} = \mathbf{b}\\
&\quad \},
\end{aligned}
$$

5. *normalize* $\mathbf{x}^{new} \leftarrow 1/\|\mathbf{x}^{aux}\|\,\mathbf{x}^{aux}$*.*

Define the mutual condition number

$$cond(A, B) = \frac{\lambda_{max}(B^{-1}A)}{\lambda_{min}(B^{-1}A)},$$

the generalized Rayleigh quotient

$$R_B(\mathbf{w}) = \frac{\langle A\mathbf{w}, \mathbf{w}\rangle}{\langle B\mathbf{w}, \mathbf{w}\rangle}, \quad \mathbf{w} \neq \mathbf{0}$$

and

$$r_B(\mathbf{w}) = \frac{\|(B^{-1}A - R_B(\mathbf{w})I)\mathbf{w}\|_B}{\|\mathbf{w}\|_B}, \quad \mathbf{w} \neq \mathbf{0}.$$

REMARK 4.1. *The definitions of cond(A,B), R_B and r_B are derived from the definitions of cond(A), R and r, respectively, using the rules $A \leftarrow B^{-1}A$, $\|\cdot\| \leftarrow \|\cdot\|_B$ and $\langle\cdot,\cdot\rangle \leftarrow \langle\cdot,\cdot\rangle_B$.*

The mutual condition number cond(A,B) is the condition number of $B^{-1}A$ in the Hilbert space $(\mathbb{R}^n,\ \langle\cdot,\cdot\rangle_B)$, where $B^{-1}A$ is a symmetric operator. The Rayleigh quotient R_B has been obtained from

$$R_B(\mathbf{x}) \equiv \frac{\langle B^{-1}A\mathbf{x},\mathbf{x}\rangle_B}{\langle\mathbf{x},\mathbf{x}\rangle_B}.$$

The origin of r_B in r is obvious.

THEOREM 4.2. *Let A and B be symmetric, positive definite $n\times n$ matrices with unique minimal generalized eigenvalue λ_{min}. We assume the generalized eigenvalues are numbered so that $\lambda_{min} = \lambda_1 < \lambda_2 \le \ldots \le \lambda_n = \lambda_{max}$. Consider $\alpha \in (0,1]$ and $\beta > 0$ such that $\alpha\beta - 1 > 0$. We assume there is a linear mapping $Q : \mathbb{R}^n \to$ Range (P) such that*

$$\forall \mathbf{u} \in \mathbb{R}^n : \lambda_{min}^{\frac{\beta}{2}} \|\mathbf{u} - Q\mathbf{u}\|_B \le \frac{C}{cond(A,B)^{\frac{\alpha\beta}{2}}} \|(B^{-1}A)^{\beta/2}\mathbf{u}\|_B.$$

In addition, assume that the input iterate $\mathbf{x}$ is reasonably close to the first eigenvector $\mathbf{v}_1$ so that $R_B(\mathbf{x}) \in [\lambda_1, \lambda_2)$. Then the result $\mathbf{x}^{new}$ on the exit of Algorithm 2 with $\nu \ge \beta/2$ satisfies the estimate

$$r_B(\mathbf{x}^{new}) \le Q(\mathbf{x}) r_B(\mathbf{x}), \quad Q(\mathbf{x}) = \frac{C}{cond(A,B)^{\frac{\alpha\beta-1}{2}}} \frac{\sqrt{\frac{\lambda_2}{\lambda_{min}} - 1}}{\frac{\lambda_2 - R_B(\mathbf{x})}{\lambda_{min}}}.$$

The rate of convergence $Q(\mathbf{x})$ satisfies

$$\lim_{cond(A,B)\to\infty} Q(\mathbf{x}) = 0.$$

Acknowledgements. This work was sponsored by the TAČR (Technologická Agentura České Republiky) grant TA01020352, ITI (Institut Teoretické Informatiky) grant 1R0545, Department of the Navy Grant N62909-11-1-7032, ERC-CZ project LL1202 and GACR 13-18652S.

REFERENCES

[1] A. BRANDT, *Algebraic Multigrid Theory: The symmetric Case*, Appl. Math. Comput., 19(1986).

[2] P. G. CIARLET, "*The finite element method for elliptic problems,*" Series "Studies in Mathematics and its Applications",North–Holland, Amsterdam, 1978.

[3] R. KUŽEL AND P. VANĚK, *Exact interpolation scheme with approximation vector used as a column of the prolongator*, NLAA, submitted.

[4] D. KUSHNIR, M. GALUN AND A. BRANDT, *Efficient multilevel eigensolvers with application to data analysis tasks*, IEEE transactions on pattern analysis and machine intelligence, 32(2010).

[5] A.BRANDT AND D.RON, *Multigrid Solvers and Multiscale Optimization Strategies* Multilevel Optimization and VLSICAD, Kluwer, 2003.

[6] M. BREZINA, R. FALGOUT, T. MANTEUFFEL, S. MCCORMIC AND J. RUGE, *Adaptive smoothed aggregation (αSA)* SIAM J. of Sci. Comp. 6(2004).

[7] J. Brousek, P. Fraňková, H. Kopincová, R. Kužel, R. Tezaur, P. Vaněk and Z. Vastl, *An overview of multilevel methods with aggressive coarsening and massive polynomial smoothing*, in preparation.

[8] H. Guillard, R. Kužel, P. Vaněk and Z. Vastl, *An alternative to domain decomposition methods based on polynomial smoothing*, Numerical Linear Algebra with Applications, submitted

[9] P. Vaněk, *Smoothed Prolongation Multigrid with Rapid Coarsening and Massive Smoothing*, To appear in Appl. Math.

[10] P. Vaněk, M. Brezina, and R. Tezaur, *Two-grid method for linear elasticity on unstructured meshes*, SIAM J. Sci Comput., 21(1999).

[11] M. Brezina, C. Heberton, J. Mandel and P. Vaněk, *An iterative method with convergence rate chosen a priori*, UCD/CCR Report no. 140, 1999.

[12] P. Vaněk, J. Mandel and M. Brezina, *Algebraic Multigrid by Smoothed Aggregation for Second and Fourth Order Elliptic Problems*, Computing 56(1996).

[13] P. Vaněk, *Acceleration of Convergence of a Two-level Algorithm by Smoothing Transfer Operator*, Appl. Math. 37(1992).

[14] P. Vaněk, *Fast multigrid solver*, Applications of Mathematics 40(1995), no. 1.

[15] P. Vaněk, M. Brezina and J. Mandel, *Convergence of Algebraic Multigrid Based on Smoothed Aggregations*,Numer. Math. 88(2001), no. 3.

[16] P. Vaněk and M. Brezina, *Nearly optimal convergence result for multigrid with aggressive coarsening and polynomial smoothing,* Appl. Math, to appear

[17] M. Brezina, P. Vaněk and P.S. Vassilevski, *An improved convergence analysis of the smoothed aggregation algebraic multigrid,* Numerical Linear Algebra with Applications (to appear).

[18] K. S. Smith, *Nodal method storage reduction by nonlinear it- eration*, Trans. Am. Nucl. Soc., 44, 265 (1983)

[19] R. D. Lawrence, *Progress in nodal methods for the solution of the neutron diffusion and transport equation*, Prog. Nucl. Energy, 17, 271 (1986)

[20] J. Mandel and B. Sekerka, *A local convergence proof of the iterative aggregation method*, Linear Algebra and its Applications, 51(1983).