.

CONVERGENCE ANALYSIS OF GRADIENT ITERATIONS FOR THE SYMMETRIC EIGENVALUE PROBLEM

KLAUS NEYMEYR[α], EVGUENI OVTCHINNIKOV[β] AND MING ZHOU[α] *

Abstract. Gradient iterations for the Rayleigh quotient are simple and robust solvers to determine a few of the smallest eigenvalues together with the associated eigenvectors of (generalized) matrix eigenvalue problems for symmetric matrices. Sharp convergence estimates for the Ritz values and Ritz vectors are derived for various steepest descent/ascent gradient iterations. The analysis shows that poorest convergence of the eigenvalue approximations is attained in a three-dimensional invariant subspace; explicit convergence estimates are then derived by means of a mini-dimensional analysis.

Key words. Gradient iteration, steepest descent/ascent, Rayleigh quotient, preconditioner, elliptic eigenvalue problem.

1. Introduction. Consider the generalized matrix eigenvalue problem

$$Ax_i = \lambda_i M x_i$$

with a symmetric matrix $A \in \mathbb{R}^{n \times n}$ and a symmetric and positive definite matrix $M \in \mathbb{R}^{n \times n}$ to be given. A typical background is an eigenvalue problem for a self-adjoint partial differential operator. Usually the finite element discretization of the operator eigenvalue problem results in large and sparse matrices A and M which are called the discretization matrix and the mass matrix. For an elliptic partial differential operator positive definiteness of A is attainable.

The eigenpairs are denoted by (λ_i, x_i) and are enumerated so that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Here we consider the problems to compute approximations either of the smallest or of the largest eigenvalue together with the associated eigenvectors. In order to solve this (partial) eigenproblem the Rayleigh quotient

$$\rho(x) = (x, Ax)/(x, Mx), \qquad x \neq 0$$

can be minimized/maximized by means of a *gradient iteration*, since $\lambda_1 = \rho(x_1) = \min_{x \neq 0} \rho(x)$ and $\lambda_n = \rho(x_n) = \max_{x \neq 0} \rho(x)$. The *steepest descent* iteration computes a sequence of iterates with decreasing Rayleigh quotients by successive corrections in the negative gradient direction of the current iterate

$$-\nabla \rho(x) = -2(Ax - \rho(x)Mx)/(x, Mx).$$

The *steepest ascent* iteration uses the positive gradient for the maximization of the Rayleigh quotient. This paper deals with steepest descent/ascent gradient iterations for symmetric (and sometimes positive definite) eigenproblems.

Convergence rate estimates for steepest descent/ascent iterations for the Rayleigh quotient have a long history: The classical asymptotic convergence rate estimates go back to Kantorovich [4, 5] as well as to Hestenes and Karush [3]. Non-asymptotic estimates are given by Prikazchikov [15], Zhuk and Bondarenko [20] as well as by Knyazev and Skorokhodov [6, 8]. See Yang [18] for a survey on conjugate gradient iterations for the symmetric eigenvalue problem.

Gradient type iterations can also be considered with respect to general geometries, as pointed out by Samokish [16] and D'yakonov [2], this amounts to *preconditioning*. The importance of such a preconditioning resides in its considerable convergence acceleration; in the context of discretized elliptic operator eigenvalue problems such a preconditioning

---

* April 14, 2011, [α]Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18055 Rostock, Germany; [β] Department of Mathematics, University College London, Gower Street WC1E 6BT, UK.

can be realized, for instance, by multi-grid or multi-level iterations. For a symmetric positive definite matrix B the B-gradient reads

$$(1.1) \qquad\qquad \nabla_B \rho(x) = B^{-1} \nabla \rho(x).$$

By using B-gradients a *preconditioned gradient scheme* can be constructed. The convergence analysis of such B-gradient (preconditioned) eigensolvers still has several open problems. Non-asymptotic estimates are known only for the most simple preconditioned gradient iteration with certain (fixed) step sizes; some of these estimates are sharp with respect to specific parameters, see Samokish [16], Perdon and Gambolati [14], Ovtchinnikov [11] and [7]. For the efficient conjugate-gradient-like iteration LOPCG [1] these (sharp) estimates serve as upper convergence estimates, which do not reflect the cg-like optimal behavior. One exception is Theorem 4.2 from [12], which gives an upper bound for the eigenvalue error reduction over two consecutive iterations of CG algorithms that use the so-called *Jacobi orthogonal correction equation*, see Sleijpen and van der Vorst [17], for conjugation (this estimate applies to LOPCG as well, which implicitly uses such conjugation).

The intention of this paper is to prove convergence estimates for a variety of steepest descent and ascent gradient iterations which reflect limit cases of preconditioned gradient iterations. Most of the available convergence estimates for preconditioned steepest descent iteration give upper bounds on the error reduction which are smallest when the inverse of A is used for preconditioning. One aim of this paper is to demonstrate the ample difference between the two limit cases - the convergence without preconditioning (working with the Euclidean gradient) and preconditioning by the inverse of A. In this way the paper should provide a better theoretical understanding of more general preconditioned gradient iterations.

This paper has a predecessor in the paper of Knyazev and Skorokhodov [8] and overcomes its limitation that the convergence analysis for the eigenvalue approximations is restricted to the extremal eigenvalues; in [8] is assumed that the Rayleigh quotient of the initial iterate is located between the extremal eigenvalue and the next neighboring eigenvalue. The analysis in [8] and the one given here both work with a *mini-dimensional proof technique*. In [8] the steepest ascent scheme (with $M = I$) is analyzed by investigating the convergence behavior of the subspace K of the 3D space H^3 with

$$K = \operatorname{span}\{x, Ax\}, \qquad H^3 = \operatorname{span}\{x_1, x, Ax\},$$

where x_1 is an eigenvector associated with λ_1. The current paper uses, once again, a mini-dimensional analysis. Here a 3D subspace is used, which is an *invariant subspace* of (A, M) since the new Theorem 2.1 below shows that the extremal convergence of the Ritz values (i.e. the fastest or slowest convergence) is attained in an invariant subspace. This invariant subspace approach enables relatively simple (compared to [8]) geometric proofs for the convergence estimates of various cases of steepest descent/ascent both for the Euclidean gradient and for the A-gradient iterations.

The paper is structured as follows: Section 2 shows that the slowest (and also the fastest) convergence of the eigenvalue approximations for certain steepest ascent and steepest descent iterations are attained in three-dimensional invariant subspaces of A. In Section 3 sharp convergence rate estimates are derived by means of a mini-dimensional analysis which is based on a simple geometric construction. In Section 4 the estimates from the low dimensional subspace are embedded into the full space $\mathbb{R}^n$.

2. Extremal convergence of gradient iterations. A basis transformation allows to assume that $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $M = I$. This assumption does not oversimplify the problem as we do not access to the diagonal elements of A. Further let $\lambda_1 < \lambda_2 < \ldots <$

λ_n since multiple eigenvalues do not change the convergence estimates derived below; see the appendix in [10] for the analytical argument.

The steepest descent/ascent B-gradient iteration for the Rayleigh quotient $\rho(\cdot)$ reads

$$(2.1) \qquad x' = x + \alpha_\pm \nabla_B \rho(x)$$

with the optimal step length parameter α_- (α_+) minimizing (maximizing) the Rayleigh quotient in the given affine subspace. In the following analysis $\alpha_\pm$ are always bounded.

Here B-gradients are considered only for $B \in \{I, A\}$. If $B = A$, then regularity of A is required. (Formally an A-gradient is correctly defined only for symmetric positive definite A.) The standard Euclidean gradient iteration works with $B = I$ and the limit case of "exact-inverse preconditioning" uses $B = A$. For these cases the maximization/minimization of $\rho(\cdot)$ amounts to the computation of the Ritz pairs of A with respect to the column space of either $[x, Ax]$ or $[x, A^{-1}x]$.

THEOREM 2.1. *Let $\mathcal{L}(\lambda) := \{x \in \mathbb{R}^n; \rho(x) = (x, Ax)/(x, x) = \lambda\}$ be the level set of the Rayleigh quotient of A with λ being different from the eigenvalues of A and let $n \geq 3$. Consider the gradient iteration*

$$x' = A^\sigma(x + \alpha Ax), \quad \sigma \in \{0, -1\}$$

with optimal step length α maximizing (minimizing) the Rayleigh quotient $\rho(x')$. For $\sigma = -1$ regularity of A is assumed. If $\rho(x')$ takes an extremum in $x \in \mathcal{L}(\lambda)$, then x has at most three non-zero eigenvector components.

Proof. Let $w = A^\sigma(x + \alpha Ax)$ be a Ritz vector of A in $\mathcal{V}^\sigma = \text{span}\{A^\sigma x, A^{\sigma+1}x\}$. By definition of a Ritz vector it holds that $Aw - \rho(w)w \perp \mathcal{V}^\sigma$ and thus $\nabla \rho(w) \perp \mathcal{V}^\sigma$.

Let $x(t)$ be an arbitrary smooth curve in $\mathcal{L}(\lambda)$ and let $w(t) = A^\sigma x(t) + \alpha(t)A^{\sigma+1}x(t)$ be the associated curve of extremum points of the gradient iteration. Then a necessary condition for an extremum of $\rho(w(t))$ in $t = t^*$ with $x^* = x(t^*)$ and $w^* = w(t^*)$ reads

$$0 = \frac{d}{dt}\rho(w(t))\Big|_{t=t^*} = \left(\nabla\rho(w^*), \left(\frac{d}{dt}\left[A^\sigma x(t) + \alpha(t)A^{\sigma+1}x(t)\right]\right)_{t=t^*}\right).$$

For ease of notation the star is always omitted in the sequel. We get that

$$(2.2) \qquad \begin{aligned} 0 &= (\nabla\rho(w), A^\sigma \dot{x} + \dot{\alpha}A^{\sigma+1}x + \alpha A^{\sigma+1}\dot{x}) \\ &= \frac{2}{(w,w)}(Aw - \rho(w)w, A^\sigma(I + \alpha A)\dot{x}) \quad \text{since } \nabla\rho(w) \perp \mathcal{V}^\sigma \\ &= \frac{2}{(w,w)}\underbrace{((A - \rho(w)I)A^{2\sigma}(I + \alpha A)^2 x}_{=:z}, \dot{x}) \end{aligned}$$

where the dot notation is used for derivatives within the extremum point, i.e. $\dot{x} = (d/dt)x(t)|_{t=t^*}$. Furthermore, for any smooth curve $x(t)$ in the level set $\mathcal{L}(\lambda)$ it holds that

$$0 = \frac{d}{dt}\rho(x) = (\nabla\rho(x), \dot{x}) = \frac{2}{(x,x)}(Ax - \rho(x)x, \dot{x}),$$

which means that the gradient $\nabla\rho(x)$ is orthogonal to the tangent plane of $\mathcal{L}(\lambda)$ in x which is spanned by all possible tangent vectors $\dot{x}$. Therefore z in (2.2) and $Ax - \rho(x)x$ are collinear vectors, i.e. $z = \nu(A - \rho(x)I)x$ for a $\nu \in \mathbb{R}$. The last equation for $\sigma = 0$ can be rewritten as $p_a(A)x = 0$ and for $\sigma = -1$ we get after multiplication with A^2 that $p_b(A)x = 0$; $p_a(\cdot)$ and $p_b(\cdot)$ are third order polynomials.

For diagonal $A = \text{diag}(\lambda_1, \ldots, \lambda_n)$ we get from $p_{a,b}(A)x = 0$ for each $i \in \{1, \ldots, n\}$ that $p_{a,b}(\lambda_i) = 0$ or $x_i = 0$. Since $p_{a,b}(\cdot)$ has at most three zeros, the vector $x \in \mathcal{L}(\lambda)$ can have at most three non-zero eigenvector components. $\square$

Thm. 2.1 can easily be generalized to s-step steepest descent/ascent iterations in the Krylov subspaces $\mathcal{K}(x, A, j)$ or in $\mathcal{K}(x, A^{-1}, j)$. Further the Rayleigh-Ritz approximations can also be considered with respect to A^{-1}. In all these cases the fastest/slowest convergence is attained in low-dimensional invariant subspaces.

3. Mini-dimensional analysis. Next we derive a convergence estimate for the smallest Ritz value of A in the space $\text{span}\{x, Ax\}$ and an estimate for the associated Ritz vector. Due to Thm. 2.1 the poorest convergence of the Ritz values is attained in a 3D invariant subspace of A. Let

$$(3.1) \qquad \mathcal{E}_{j,k,l} := \text{span}\{e_j, e_k, e_l\}$$

be such a subspace where e_i is the ith column of the n-by-n identity matrix. For the associated eigenvalues with $\lambda_j < \lambda_k < \lambda_l$ let $\Delta_{p,q}(\theta) := (\theta - \lambda_p)/(\lambda_q - \theta)$. Further $\angle(y, z)$ denotes the acute angle enclosed by $\text{span}\{y\}$ and $\text{span}\{z\}$.

In Sec. 4, we apply these mini-dimensional estimates derived below in the full n-dimensional space, which results in the desired convergence estimates for the gradient iterations.

LEMMA 3.1. *Let $x \in \mathcal{L}(\lambda) \cap \mathcal{E}_{j,k,l}$ with the associated eigenvalues $\lambda_j < \lambda_k < \lambda_l$. If $\lambda_j < \lambda < \lambda_k$, then the smaller Ritz value λ' of A in $\text{span}\{x, Ax\}$ and the associated Ritz vector x' satisfy*

$$\frac{\Delta_{j,k}(\lambda')}{\Delta_{j,k}(\lambda)} \leq \left(\frac{\kappa}{2-\kappa}\right)^2, \qquad \frac{\tan\angle(x', e_j)}{\tan\angle(x, e_j)} \leq \kappa \quad with \quad \kappa = \frac{\lambda_l - \lambda_k}{\lambda_l - \lambda_j}.$$

If A is a positive definite matrix, then an A-angle ratio is bounded, once again, by κ

$$\frac{\tan\angle_A(x', e_j)}{\tan\angle_A(x, e_j)} \leq \kappa.$$

Proof. Next we use the indexes $(1, 2, 3)$ as representatives for (j, k, l) to improve the readability. The associated eigenvalues are $\lambda_1 < \lambda_2 < \lambda_3$. The vectors x and x' in $\mathcal{E}_{1,2,3}$ can be scaled (any non-zero scaling does not change estimates) so that

$$x = e_1 + \alpha_0 e_2 + \beta_0 e_3, \quad x' = e_1 + \alpha_1 e_2 + \beta_1 e_3.$$

(Note that $(x', e_1) \neq 0$. Otherwise it would hold that $\rho(x') \geq \lambda_2$, which contradicts the Courant variational principle.) Then x and x' are elements of the affine space $\mathcal{E}_1 := e_1 + \text{span}\{e_2, e_3\}$. Further x' is an element of the level set (a cone)

$$\mathcal{L}(\lambda') = \{z \in \mathcal{E}_{1,2,3} \; ; \; \rho(z) = \lambda'\}.$$

The intersection of $\mathcal{L}(\lambda')$ and $\mathcal{E}_1$ is an ellipse, see Fig. 3.1. Figure 3.1 also shows a broken line which is the intersection of the two-dimensional space $\text{span}\{x, Ax\}$ and $\mathcal{E}_1$; this line is tangential to the ellipse in x'.

The intersection of this tangent and $e_1 + \text{span}\{e_2\}$ is denoted by y and has the form $y = e_1 + \theta e_2$. Next we compute θ. Therefore in $\text{span}\{x, Ax\}$ the vector $x - Ax/\lambda_3 = (1 - \lambda_1/\lambda_3)e_1 + (1 - \lambda_2/\lambda_3)\alpha_0 e_2$ is formed, which is a multiple of y. The normalization of its first component gives

$$\theta = \frac{1 - \lambda_2/\lambda_3}{1 - \lambda_1/\lambda_3}\alpha_0 = \kappa\alpha_0 \quad with \quad \kappa = \frac{\lambda_3 - \lambda_2}{\lambda_3 - \lambda_1} < 1.$$

Due to $\lambda_2 < \lambda_3$ the larger semi-axis of the ellipse is oriented along the e_2 axis; this implies the first inequality in

$$\tan^2 \angle(x', e_1) \leq \tan^2 \angle(y, e_1) = \kappa^2\alpha_0^2 \leq \kappa^2(\alpha_0^2 + \beta_0^2) = \kappa^2 \tan^2 \angle(x, e_1),$$
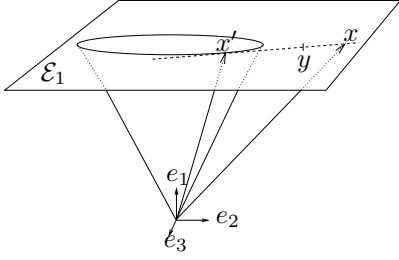
FIG. 3.1. *The cone $\mathcal{L}(\lambda')$ and iterates x, x'.*
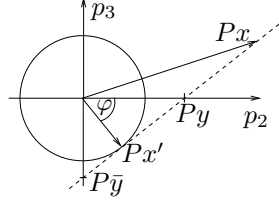
FIG. 3.2. *Geometry in the plane $P\mathcal{E}_1$.*

which proves the tangent estimate.

In the arguments given above one can scale the horizontal and the vertical coordinates (in the plane $\mathcal{E}_1$, see Fig. 3.1) without invalidating the estimate for the tangents as long as the horizontal semi-axis of the ellipse remains not less than the length of the vertical semi-axis. Furthermore, in the particular case of the ellipse becoming a circle, the estimate improves, as we are going to show next, which will deliver the estimate for the Ritz values. First we introduce a coordinatization of the affine plane $\mathcal{E}_1$

$$z = e_1 + \alpha e_2 + \beta e_3 \quad \rightarrow \quad Pz = (\alpha, \beta)^T.$$

Hence the tangent estimate from above can be written as $\|Px'\|^2 \le \kappa^2 \|Px\|^2$ where $\|\cdot\|$ is the Euclidean norm. A rescaling of the axes complies with the introduction of a weighted norm $\|(\alpha, \beta)^T\|_\omega := (\omega_2 \alpha^2 + \omega_3 \beta^2)^{1/2}$ with $\omega_2, \omega_3 > 0$. A horizontal semi-axis which is not smaller than the vertical semi-axis is guaranteed by $(\omega_3/\omega_2) \le (a^2/b^2)$. Therein $a = \big((\lambda' - \lambda_1)/(\lambda_2 - \lambda')\big)^{1/2}$ and $b = \big((\lambda' - \lambda_1)/(\lambda_3 - \lambda')\big)^{1/2}$ are the lengths of the semi-axes of the ellipse $\mathcal{L}(\lambda') \cap \mathcal{E}_1$.

Then $\|Px'\|_\omega^2 \le \kappa^2 \|Px\|_\omega^2$ holds, since

$$\omega_2 \alpha_1^2 + \omega_3 \beta_1^2 \le \omega_2\big(\alpha_1^2 + \tfrac{a^2}{b^2}\beta_1^2\big) = \omega_2 a^2 \le \omega_2 \kappa^2 \alpha_0^2 \le \kappa^2(\omega_2 \alpha_0^2 + \omega_3 \beta_0^2).$$

Next the limit case $\omega_2 = 1$ and $\omega_3 = (a^2/b^2)$, which transforms the ellipse into a circle, is analyzed. We call this limit norm the D-norm with $D = \operatorname{diag}(1, a^2/b^2)$ so that $\|Px'\|_D^2 \le \kappa^2 \|Px\|_D^2$. Next we derive an improved estimate by determining the minimum of $\|Px\|_D^2 / \|Px'\|_D^2$. To this end let $p_2 = (1, 0)^T$, $p_3 = (0, 1)^T \in P\mathcal{E}_1$ and φ be the acute D-angle enclosed by $\operatorname{span}\{p_2\}$ and $\operatorname{span}\{Px'\}$. Further let $P\bar{y} = \big(0, \kappa\beta_0/(\kappa - 1)\big)^T$ be the $\mathcal{E}_1$-representation of the intersection $\bar{y}$ of the tangent and $e_1 + \operatorname{span}\{e_3\}$, whose components can easily be computed like those of y. Then Px' is located on the line segment connecting Py and $P\bar{y}$ and its components can be expressed by $(\alpha_1, \beta_1)^T = (1 - \tau)Py + \tau P\bar{y} = \big((1 - \tau)\kappa\alpha_0,\ \tau\kappa\beta_0/(\kappa - 1)\big)^T$ for some $\tau \in (0, 1)$, so that

$$\operatorname{sgn}(\alpha_0) = \operatorname{sgn}(\kappa\alpha_0) = \operatorname{sgn}(\alpha_1), \quad \operatorname{sgn}(\beta_0) = -\operatorname{sgn}\big(\kappa\beta_0/(\kappa - 1)\big) = -\operatorname{sgn}(\beta_1),$$

since $\kappa \in (0, 1)$. Fig. 3.2 shows the case $\alpha_0 > 0$, $\beta_0 > 0$.

The D-orthogonality $Px' \perp_D Px - Px'$ implies

(3.2)
$$\cos(\varphi) = \frac{\|Px'\|_D}{\|Py\|_D} = \frac{\sqrt{\alpha_1^2 + \beta_1^2 \frac{a^2}{b^2}}}{|\kappa\alpha_0|} = \frac{a}{\kappa|\alpha_0|},$$

$$\sin(\varphi) = \cos(\pi/2 - \varphi) = \frac{\|Px'\|_D}{\|P\bar{y}\|_D} = \frac{\sqrt{\alpha_1^2 + \beta_1^2 \frac{a^2}{b^2}}}{\frac{a}{b}|\kappa\beta_0/(\kappa - 1)|} = \frac{(1 - \kappa)b}{\kappa|\beta_0|}.$$

Therewith the ratio $\|Px\|_D^2/\|Px'\|_D^2$ can be expressed as follows

$$\frac{\|Px\|_D^2}{\|Px'\|_D^2} = \frac{\alpha_0^2 + \beta_0^2 a^2/b^2}{a^2} = (\alpha_0/a)^2 + (\beta_0/b)^2 = \frac{1}{\kappa^2 \cos^2(\varphi)} + \frac{(1-\kappa)^2}{\kappa^2 \sin^2(\varphi)}.$$

Insertion of $\sin^2(\varphi) = t/(1+t)$ and $\cos^2(\varphi) = 1/(1+t)$ with $t = \tan^2(\varphi)$ gives

$$\frac{\|Px\|_D^2}{\|Px'\|_D^2} = \frac{(1+t)(t + (1-\kappa)^2)}{t\kappa^2} =: f(t).$$

By differentiation one sees that the unique minimum of $f(t)$ is taken in $t = 1 - \kappa$ so that

$$\frac{\|Px\|_D^2}{\|Px'\|_D^2} \geq f(1-\kappa) = \left(\frac{2-\kappa}{\kappa}\right)^2.$$

This estimate can be interpreted as a convergence bound for the Ritz value by noting that $\|Px'\|_D^2 = a^2 = (\lambda' - \lambda_1)/(\lambda_2 - \lambda') = \Delta_{1,2}(\lambda')$ and by showing that $\|Px\|_D^2 \leq \Delta_{1,2}(\lambda)$. To see the latter consider the intersection of $\mathcal{L}(\lambda) \cap \mathcal{E}_{1,2,3}$ with $\mathcal{E}_1$, which is an ellipse through x. Thus

$$\frac{\alpha_0^2}{\bar{a}^2} + \frac{\beta_0^2}{\bar{b}^2} = 1 \quad \text{with} \quad \bar{a}^2 = \frac{\lambda - \lambda_1}{\lambda_2 - \lambda}, \quad \bar{b}^2 = \frac{\lambda - \lambda_1}{\lambda_3 - \lambda}.$$

Hence $\lambda' \leq \lambda$ implies $a^2/b^2 \leq \bar{a}^2/\bar{b}^2$ so that

$$\|Px\|_D^2 = \alpha_0^2 + \frac{a^2}{b^2}\beta_0^2 \leq \alpha_0^2 + \frac{\bar{a}^2}{\bar{b}^2}\beta_0^2 = \bar{a}^2 = \Delta_{1,2}(\lambda).$$

For a positive definite matrix A the A-angle estimate is proved as follows

$$\tan^2 \angle_A(x', e_1) = \frac{\lambda_2}{\lambda_1}\alpha_1^2 + \frac{\lambda_3}{\lambda_1}\beta_1^2 = \frac{\lambda_2}{\lambda_1}\left(\alpha_1^2 + \frac{\lambda_3}{\lambda_2}\beta_1^2\right) \leq \frac{\lambda_2}{\lambda_1}\left(\alpha_1^2 + \frac{a^2}{b^2}\beta_1^2\right)$$

$$\leq \frac{\lambda_2}{\lambda_1}\tan^2 \angle(y, e_1) = \frac{\lambda_2}{\lambda_1}\kappa^2\alpha_0^2 \leq \kappa^2\frac{\lambda_2}{\lambda_1}\left(\alpha_0^2 + \frac{\lambda_3}{\lambda_2}\beta_0^2\right) = \kappa^2 \tan^2 \angle_A(x, e_1).$$

□

3.1. An improved estimate for the eigenvector error. The convergence estimate for the Ritz vectors, as given in Lemma 3.1, can be further improved. However, the analysis is lengthy and complicated. Next a short outline is given where the indexes $(1, 2, 3)$ are used as representatives for (j, k, l).

First the D-angle φ can be expressed in terms of α_1 and β_1 as $|\alpha_1| = a\cos(\varphi)$ and $|\beta_1| = b\sin(\varphi)$ since

$$(3.3) \quad \cos(\varphi) = \frac{|(p_2, Px')_D|}{\|p_2\|_D\|Px'\|_D} = \frac{|\alpha_1|}{\sqrt{\alpha_1^2 + \beta_1^2 \frac{a^2}{b^2}}} = \frac{|\alpha_1|}{a}, \quad \sin(\varphi) = \sqrt{1 - \cos^2(\varphi)} = \frac{|\beta_1|}{b},$$

where the ellipse equation $z_2^2/a^2 + z_3^2/b^2 = 1$ is used for the $\sin(\varphi)$ equation.

By using (3.2), (3.3) and

$$\nu := \frac{\lambda_2 - \lambda'}{\lambda_3 - \lambda'} = b^2/a^2, \quad t := \tan^2(\varphi), \quad \xi := \frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1} = 1 - \kappa$$

we get for the ratio $\tan^2 \angle(x', e_1)/\tan^2 \angle(x, e_1)$ that

$$\frac{\tan^2 \angle(x', e_1)}{\tan^2 \angle(x, e_1)} = \frac{\alpha_1^2 + \beta_1^2}{\alpha_0^2 + \beta_0^2} = \frac{(1-\xi)^2 t(1+\nu t)}{(1+t)^2(\xi^2\nu + t)} =: f(\nu, t).$$

Lemma 3.1 proves κ^2 as an upper bound for $f(\nu, t)$; and in fact we find for $\nu \in (0, \xi)$, $t > 0$, that $\lim_{\nu \to 0, t \to 0} f(\nu, t) = (1 - \xi)^2 = \kappa^2$.

However an extremum taken in $(\nu, t) = (0, 0)$ is not admissible. To see this we start with $\rho(x) < \lambda_2$, which implies a bound for β_0^2 as follows

$$\frac{\lambda_1 + \lambda_2 \alpha_0^2 + \lambda_3 \beta_0^2}{1 + \alpha_0^2 + \beta_0^2} = \rho(x) < \lambda_2 \quad \Leftrightarrow \quad \beta_0^2 < \frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_2} = \frac{1 - \kappa}{\kappa}.$$

Combining this bound with (3.2) one gets

$$\frac{b^2}{\sin^2(\varphi)} < \frac{\kappa}{1 - \kappa} \quad \Leftrightarrow \quad \frac{1 - \xi}{\xi} \frac{t}{1 + t} > b^2 = \frac{\xi - \nu}{1 - \xi} \quad \Leftrightarrow \quad \nu > \bar{\nu}(t) := \frac{\xi^2 + 2\xi t - t}{\xi + \xi t}.$$

The insertion of $(\nu, t) = (0, 0)$ in the last inequality gives a contradiction.

The further analysis (by using the constraining inequalities derived above) shows that the maximum of $\tan^2 \angle(x', e_1) / \tan^2 \angle(x, e_1)$ in t for fixed ξ under the given constraints equals the maximum of

$$(3.4) \qquad g(t, \xi) := f(\bar{\nu}(t), \ t) = \frac{(\xi + \xi t + \xi^2 t - t^2 + 2\xi t^2)(1 - \xi)^2 t}{\xi(\xi^3 - \xi t + 2\xi^2 t + t + t^2)(1 + t)^2}$$

with

$$0 < t < \begin{cases} \xi^2/(1 - 2\xi) & \text{if } \xi \in (0, \frac{1}{3}), \\ \xi & \text{if } \xi \in [\frac{1}{3}, 1). \end{cases}$$

One can rewrite (3.4) with $u = t/\xi$ in the more intuitive form

$$g(t, \xi) = \tilde{g}(u, \xi) = (1 - \xi)^2 \frac{u}{(1 + \xi u)^2} \frac{1 + (\xi^2 + \xi)u + (2\xi - 1)\xi u^2}{\xi^2 + (1 + (2\xi - 1)\xi)u + \xi u^2}.$$

The following lemma rests upon an improved estimation of $g(t, \xi)$. For its (technical) proof, see the doctoral thesis of Ming Zhou [19].

LEMMA 3.2. *By using the assumptions and the notation of Lemma 3.1 let*

$$(3.5) \qquad S(\kappa) := \frac{\kappa}{2 - \kappa} \cdot (1 - \kappa) + \kappa \cdot \kappa = \frac{\kappa(1 + \kappa - \kappa^2)}{2 - \kappa} \quad \text{for } \kappa \in (0, 1),$$

which is a convex combination of $\kappa/(2 - \kappa)$ and κ so that $S(\kappa) < \kappa$.

If $\lambda_j < \lambda < \lambda_k$, then the Ritz vector x' associated with the smaller Ritz value λ' of A in $\text{span}\{x, Ax\}$ *satisfies*

$$\frac{\tan \angle(x', e_j)}{\tan \angle(x, e_j)} \leq S(\kappa) \quad \text{with} \quad \kappa = \frac{\lambda_l - \lambda_k}{\lambda_l - \lambda_j}.$$

The numerical computation of the maximum in t of $g(t, \xi) = g(t, 1 - \kappa)$ for $\kappa \in (0, 1)$ illustrates the attainable improvement. Figure 3.3 shows this sharp upper bound and also the bounds κ and $S(\kappa)$. The improved bound of Lemma 3.2 allows to improve the Ritz vector convergence estimates in Thm. 4.1 below.

4. Convergence estimates for gradient iterations. The next theorem, which is the central result of this paper, can be interpreted in two ways. First Thm. 4.1 can be understood as a convergence theorem for Euclidean- and A-gradient steepest descent/ascent iterations for symmetric (positive definite) matrices $A \in \mathbb{R}^{n \times n}$. Second, Thm. 4.1 provides (sharp) estimates for the Ritz values of A and the associated Ritz vectors in the
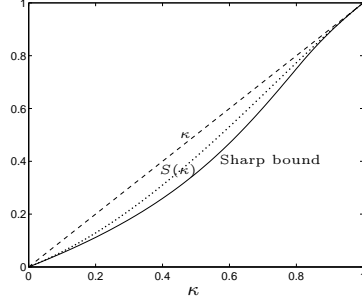
FIG. 3.3. *Bound κ, improved bound $S(\kappa)$ and numerically computed sharp bound.*

two-dimensional Krylov subspaces $\mathrm{span}\{x, Ax\}$ and also $\mathrm{span}\{x, A^{-1}x\}$. The steepest descent (ascent) estimates are estimates on the smallest (largest) Ritz values and the associated Ritz vectors. These two approaches are equivalent; the proposition of Thm. 4.1 reflects both aspects. The theorem applies to non-diagonal symmetric matrices A; however using a diagonalizing basis improves the readabilty of the analysis (see Sec. 2). If one assumes a non-diagonal matrix A in Thm. 4.1, the standard basis vectors e_1 and e_n have to be substituted by the respective eigenvectors x_1 and x_n of A. The proof of the theorem rests upon an application of the mini-dimensional convergence estimates from Sec. 3 in the full vector space $\mathbb{R}^n$ by assigning the index triplets (i, j, k) to the specific indexes between 1 and n.

THEOREM 4.1. *Let $x \in \mathbb{R}^n$ with $\lambda = \rho(x) \in (\lambda_i, \lambda_{i+1})$ for a $i \in \{1, \ldots, n-1\}$. Further let x' the new iterate Euclidean- or A-gradient steepest descent/ascent iteration so that x' is a Ritz vector in $\mathrm{span}\{x, Ax\}$ or $\mathrm{span}\{x, A^{-1}x\}$; $\lambda' := \rho(x')$ is the associated Ritz value.*

1. **Estimates on the Ritz approximations of A in $\mathrm{span}\{\mathbf{x}, \mathbf{Ax}\}$:**
   The **steepest descent iteration** *(2.1) using the gradient $\overline{\nabla}\rho(x)$ satisfies the following estimate: Either $\lambda' := \rho(x') \le \lambda_i$ or it holds that*

   $$0 < \frac{\Delta_{i,i+1}(\lambda')}{\Delta_{i,i+1}(\lambda)} \le \left(\frac{\kappa}{2-\kappa}\right)^2 \quad with \quad \kappa = \frac{\lambda_n - \lambda_{i+1}}{\lambda_n - \lambda_i}.$$

   The bound can be attained for $\lambda \to \lambda_i$ in the 3D invariant subspace $\mathcal{E}_{i,i+1,n}$, see (3.1). If $0 < \tan\angle(x, e_1) < \infty$, then the associated Ritz vector x' satisfies

   $$\frac{\tan\angle(x', e_1)}{\tan\angle(x, e_1)} \le \kappa \quad with \quad \kappa = \frac{\lambda_n - \lambda_2}{\lambda_n - \lambda_1}.$$

   For the **steepest ascent iteration** *is holds either $\lambda' \ge \lambda_{i+1}$ or*

   $$0 < \frac{\Delta_{i+1,i}(\lambda')}{\Delta_{i+1,i}(\lambda)} \le \left(\frac{\kappa}{2-\kappa}\right)^2 \quad with \quad \kappa = \frac{\lambda_i - \lambda_1}{\lambda_{i+1} - \lambda_1}.$$

   The bound can be attained for $\lambda \to \lambda_{i+1}$ in $\mathcal{E}_{1,i,i+1}$. If $0 < \tan\angle(x, e_n) < \infty$, then the Ritz vector x' fulfills

   $$\frac{\tan\angle(x', e_n)}{\tan\angle(x, e_n)} \le \kappa \quad with \quad \kappa = \frac{\lambda_{n-1} - \lambda_1}{\lambda_n - \lambda_1}.$$

2. **Estimates on the Ritz approximations of A in $\mathrm{span}\{\mathbf{x}, \mathbf{A^{-1}x}\}$:**
   Let $A \in \mathbb{R}^{n \times n}$ be a positive definite matrix. The **steepest descent iteration** *(2.1) using the gradient $\nabla_A \rho(x) = A^{-1}\nabla\rho(x)$ satisfies the following estimate:*

Either $\lambda' \leq \lambda_i$ or it holds that

$$0 < \frac{\Delta_{i,i+1}(\lambda')}{\Delta_{i,i+1}(\lambda)} \leq \left(\frac{\kappa}{2-\kappa}\right)^2 \quad with \quad \kappa = \frac{\lambda_i(\lambda_n - \lambda_{i+1})}{\lambda_{i+1}(\lambda_n - \lambda_i)}.$$

The bound can be attained for $\lambda \to \lambda_i$ in $\mathcal{E}_{i,i+1,n}$. If $0 < \tan\angle(x, e_1) < \infty$, then the associated Ritz vector x' satisfies

$$\frac{\tan\angle(x', e_1)}{\tan\angle(x, e_1)} \leq \kappa \quad with \quad \kappa = \frac{\lambda_1(\lambda_n - \lambda_2)}{\lambda_2(\lambda_n - \lambda_1)}.$$

*For the **steepest ascent iteration** is holds either $\lambda' \geq \lambda_{i+1}$ or*

$$0 < \frac{\Delta_{i+1,i}(\lambda')}{\Delta_{i+1,i}(\lambda)} \leq \left(\frac{\kappa}{2-\kappa}\right)^2 \quad with \quad \kappa = \frac{\lambda_{i+1}(\lambda_i - \lambda_1)}{\lambda_i(\lambda_{i+1} - \lambda_1)}.$$

The bound can be attained for $\lambda \to \lambda_{i+1}$ in $\mathcal{E}_{1,i,i+1}$. If $0 < \tan\angle(x, e_n) < \infty$, then the Ritz vector x' fulfills

$$\frac{\tan\angle(x', e_n)}{\tan\angle(x, e_n)} \leq \kappa \quad with \quad \kappa = \frac{\lambda_n(\lambda_{n-1} - \lambda_1)}{\lambda_{n-1}(\lambda_n - \lambda_1)}.$$

Proof. **Case 1 - steepest descent**: Let $\lambda = \rho(x) \in (\lambda_i, \lambda_{i+1})$. Then by Thm. 2.1 poorest convergence is attained in a 3D invariant subspace; Lemma 3.1 proves in $\mathcal{E}_{j,k,l}$ the estimate

$$\frac{\Delta_{j,k}(\lambda')}{\Delta_{j,k}(\lambda)} \leq \left(\frac{\kappa}{2-\kappa}\right)^2 \quad with \quad \kappa = \frac{\lambda_l - \lambda_k}{\lambda_l - \lambda_j}.$$

It holds either that $\lambda_j \leq \lambda_i < \lambda < \lambda_{i+1} \leq \lambda_k \leq \lambda_l$ or alternatively that $\lambda_j < \lambda_k \leq \lambda_i < \lambda < \lambda_{i+1} \leq \lambda_l$. In the latter alternative the smallest Ritz value λ' in $\mathcal{E}_{j,k,l}$ due to the variational principles satisfies that $\lambda' < \lambda_k \leq \lambda_i$, which is the first alternative in the proposition. (However, the estimate also holds for negative $\Delta_{i,i+1}(\lambda')$ trivially.) Next we analyze the first alternative. The upper bound $\left(\kappa/(2-\kappa)\right)^2$ in Lemma 3.1 is a monotone increasing function in $\kappa \in (0,1)$. One has to determine the maximal κ in order to find the convergence estimate. Further, κ is an increasing function in λ_j and λ_l and a decreasing function in λ_k. Hence the maximum is taken in $\lambda_j = \lambda_i$, $\lambda_k = \lambda_{i+1}$ and $\lambda_l = \lambda_n$. This proves that

$$\tilde{\kappa} := \max \kappa = \frac{\lambda_n - \lambda_{i+1}}{\lambda_n - \lambda_i}$$

is the correct bound. For $\lambda \to \lambda_i$ this bound can be attained in $\mathcal{E}_{i,i+1,n}$, which is a result of the analysis in Lemma 3.1.

Next the convergence estimate for the Ritz vector is proved. However, no pendant of Thm. 2.1 is available, i.e. we do not have a subspace of poorest convergence for the Ritz vectors. Instead we use the subspace $\mathcal{U} := \mathrm{span}\{e_1, x, Ax\}$. If x is not an eigenvector, then $\dim(\mathcal{U}) \geq 2$; $\dim(\mathcal{U}) = 2$ is a trivial case in which steepest descent converges in a single step to an eigenvector. So the only relevant case is $\dim(\mathcal{U}) = 3$.

First (e_1, λ_1) is a Ritz pair of A with respect to $\mathcal{U}$. Let θ_2 and θ_3 be the remaining Ritz values so that $\lambda_1 \leq \theta_2 \leq \theta_3$. It is easy to show that $\lambda_1 < \theta_2 < \theta_3$. The associated normalized Ritz vectors are used to form the columns of the orthonormal matrix $U = [e_1, v_2, v_3] \in \mathbb{R}^{n \times 3}$. These Ritz vectors diagonalize A on $\mathcal{U}$ so that

$$U^T A U = \mathrm{diag}(\lambda_1, \theta_2, \theta_3), \quad U^T U = I_3.$$

Next let $\mathcal{V} := \operatorname{span}\{x, Ax\} \subset \mathcal{U}$. By construction the iterate $x' \in \mathcal{V}$ is a Ritz vector of A since $\rho(x') = \min\{\rho(z); \ 0 \neq z \in \mathcal{V}\}$. Moreover, x' is a Ritz vector of $\bar{A} = PAP$ with the orthogonal projection $P = UU^T$ on $\mathcal{U}$. To show this, note that $Px' = x'$ and $Pz = z$ for all $z \in \mathcal{V}$ so that

$$0 = (Ax' - \rho_A(x')x', z) = (APx' - \rho_A(Px')Px', Pz) = (\bar{A}x' - \rho_{\bar{A}}(x')x', z).$$

Hence the eigenvector-residual of x' w.r.t. $\bar{A}$ is orthogonal to $\mathcal{V}$.

This justifies (after a change of the basis) the application of Lemma 3.1 to $U^T \bar{A} U = U^T A U = \operatorname{diag}(\lambda_1, \theta_2, \theta_3)$. So we get (see Lemma 3.1 and note that the tangent estimate also holds under the condition $0 < \tan \angle(x, e_j) < \infty$)

$$\frac{\tan \angle(x', e_1)}{\tan \angle(x, e_1)} \leq \tilde{\kappa} \quad \text{with} \quad \tilde{\kappa} = \frac{\theta_3 - \theta_2}{\theta_3 - \lambda_1}.$$

Further $\lambda_2 \leq \theta_2 < \theta_3 \leq \lambda_n$ shows $0 < \tilde{\kappa} \leq \kappa = (\lambda_n - \lambda_2)/(\lambda_n - \lambda_1) < 1$. Hence we conclude

$$\tan \angle(x', e_1) \leq \tilde{\kappa} \tan \angle(x, e_1) \leq \kappa \tan \angle(x, e_1).$$

Case 1 - steepest ascent: The proof succeeds by applying the result derived above to $-A$. The associated substitution is $(\lambda_1, \lambda_i, \lambda_{i+1}, \lambda_n) \to (-\lambda_n, -\lambda_{i+1}, -\lambda_i, -\lambda_1)$.

Case 2 - steepest ascent: The steepest descent analysis of case 1 is applied to the inverse matrix A^{-1} and to the subspace $\operatorname{span}\{y, A^{-1}y\}$ with $y := A^{1/2}x$. For $\tilde{V} := [x, A^{-1}x]$ and $W := [y, A^{-1}y]$ with $W = A^{1/2}\tilde{V}$ it holds that

$$\tilde{V}^T A \tilde{V} z_i = \theta_i \tilde{V}^T \tilde{V} z_i \quad \Leftrightarrow \quad W^T A^{-1} W z_i = \frac{1}{\theta_i} W^T W z_i.$$

Any Ritz pair $(\theta_i, \tilde{V}z_i)$ of A in $\operatorname{span}\{\tilde{V}\}$ turns into a Ritz pair $(1/\theta_i, Wz_i)$ of A^{-1} in $\operatorname{span}\{W\}$. Further we get $\rho_A(x) = 1/\rho_{A^{-1}}(y)$. Thus the replacement $A \to A^{-1}$ results in the substitution

$$(\lambda_i, \lambda_{i+1}, \lambda_n, \lambda, \lambda') \ \to \ (\lambda_{i+1}^{-1}, \lambda_i^{-1}, \lambda_1^{-1}, \lambda^{-1}, \lambda'^{-1}).$$

This substitution lets the ratio $\Delta_{i,i+1}(\lambda')/\Delta_{i,i+1}(\lambda)$ unchanged and results in

$$\frac{\lambda_n - \lambda_{i+1}}{\lambda_n - \lambda_i} \ \to \ \frac{\lambda_{i+1}(\lambda_i - \lambda_1)}{\lambda_i(\lambda_{i+1} - \lambda_1)}.$$

This proves the Ritz value estimate.

For a positive definite matrix A and $y' := A^{1/2}x'$ Lemma 3.1 proves the A^{-1}-angle estimate

$$\frac{\tan \angle_{A^{-1}}(y', e_n)}{\tan \angle_{A^{-1}}(y, e_n)} \leq \frac{\lambda_1^{-1} - \lambda_{n-1}^{-1}}{\lambda_1^{-1} - \lambda_n^{-1}} = \frac{\lambda_n(\lambda_{n-1} - \lambda_1)}{\lambda_{n-1}(\lambda_n - \lambda_1)}.$$

Direct computation shows that $\angle_{A^{-1}}(y', e_n) = \angle(x', e_n)$. This proves the Ritz vector estimate.

Case 2 - steepest descent follows from case 1 (steepest ascent) by using the arguments used above to prove the first part of case 2. $\square$

5. Conclusion. Gradient iterations for the Rayleigh quotient are basic and (with respect to a proper geometry) potentially fast iterations to compute approximations of the smallest/largest eigenvalue and the corresponding eigenvector. In its most simple form a gradient iteration only needs a computer storage for two vectors so that extremely large

eigenproblems (e.g. those for finite element discretization matrices) can be treated. In this paper convergence estimates for the steepest descent/ascent iterations for the Rayleigh quotient have been derived. The analysis involves the Euclidean gradient $\nabla\rho(x)$ and also the A-gradient $A^{-1}\nabla\rho(x)$, which is the correction direction of an ideally preconditioned gradient iteration.

These two gradient iterations have a very different behavior for discretized operator eigenvalue problems. To compute the smallest eigenvalue of an elliptic and self-adjoint partial differential operator the Euclidean gradient iteration (case 1 in Thm. 4.1) is not to be recommended due to its mesh dependent convergence factor since the largest discrete eigenvalue λ_n typically (e.g. those for the Laplacian) behaves like $\mathcal{O}(h^{-2})$ and $\lim_{\lambda_n\to\infty}\kappa/(2-\kappa)=1$.

In contrast to this, for the A-gradient steepest descent iteration $\lim_{\lambda_n\to\infty}\kappa=\lambda_i/\lambda_{i+1}$ holds. In this limit the $\Delta_{i,i+1}$-ratio for the Ritz values is bounded by

$$(5.1) \qquad \lim_{\lambda_n\to\infty}\left(\frac{\kappa}{2-\kappa}\right)^2=\left(\frac{\lambda_i}{\lambda_{i+1}}\right)^2\left(2-\frac{\lambda_i}{\lambda_{i+1}}\right)^{-2}<1.$$

This allows to point out the potential of (preconditioned) steepest descent iterations compared to the simple *fixed-step-size* preconditioned gradient iteration (also known as the preconditioned inverse iteration, PINVIT) as analyzed in [7]. Its fundamental convergence estimate is

$$\frac{\Delta_{i,i+1}(\lambda')}{\Delta_{i,i+1}(\lambda)}\leq\left(\gamma+(1-\gamma)\frac{\lambda_i}{\lambda_{i+1}}\right)^2,$$

with $\gamma\in[0,1)$ being a measure for the quality of the preconditioner.

For $\gamma=0$ PINVIT reduces to a well-known fixed-step-size A-gradient iteration, namely to the inverse iteration procedure and the convergence estimate reads

$$\frac{\Delta_{i,i+1}(\lambda')}{\Delta_{i,i+1}(\lambda)}\leq\left(\frac{\lambda_i}{\lambda_{i+1}}\right)^2.$$

Comparing this estimate with the bound for the A-gradient steepest descent iteration one gets from (5.1) in the limit $\lambda_n\to\infty$ that

$$\left(\frac{\lambda_{i+1}}{2\lambda_{i+1}-\lambda_i}\right)^2<1$$

is the (grid-independent) *gain-factor* of the optimal step length. In other words the gain-factor describes the benefit of the A-gradient steepest descent iteration. This result is confirmed by the recent analysis of the preconditioned steepest descent iteration in [9], which reproduces case 2 (steepest descent) of Thm. 4.1 in the limit of an (exact-inverse preconditioned) B-gradient iteration.

REFERENCES

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors. *Templates for the solution of algebraic eigenvalue problems: A practical guide.* SIAM, Philadelphia, 2000.
[2] E.G. D'yakonov. *Optimization in solving elliptic problems.* CRC Press, Boca Raton, Florida, 1996.
[3] M.R. Hestenes and W. Karush. A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix. *J. Res. Nat. Bureau Standards*, 47:45–61, 1951.
[4] L. V. Kantorovich. *Functional analysis and applied mathematics.* U. S. Department of Commerce National Bureau of Standards, Los Angeles, Calif., 1952. Translated by C. D. Benster.

[5] L.V. Kantorovich and G.P. Akilov. *Functional analysis in normed spaces*. Pergamon, London, 1964.

[6] A.V. Knyazev. Convergence rate estimates for iterative methods for a mesh symmetric eigenvalue problem. *Russian J. Numer. Anal. Math. Modelling*, 2:371–396, 1987.

[7] A.V. Knyazev and K. Neymeyr. Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers. *SIAM J. Matrix Analysis*, 31:621–628, 2009.

[8] A.V. Knyazev and A.L. Skorokhodov. On exact estimates of the convergence rate of the steepest ascent method in the symmetric eigenvalue problem. *Linear Algebra Appl.*, 154–156:245–257, 1991.

[9] K. Neymeyr. A geometric convergence theory for the preconditioned steepest descent iteration. Technical Report, Universität Rostock, 2010.

[10] K. Neymeyr, E. Ovtchinnikov, and M. Zhou. Convergence analysis of gradient iterations for the symmetric eigenvalue problem. Technical Report, Universität Rostock, 2010.

[11] E. E. Ovtchinnikov. Sharp convergence estimates for the preconditioned steepest descent method for hermitian eigenvalue problems. *SIAM J. Numer. Anal.*, 43(6):2668–2689, 2006.

[12] E.E. Ovtchinnikov. Jacobi correction equation, line search, and conjugate gradients in Hermitian eigenvalue computation. I: Computing an extreme eigenvalue. *SIAM J. Numer. Anal.*, 46(5):2567–2592, 2008.

[13] E.E. Ovtchinnikov. Jacobi correction equation, line search, and conjugate gradients in Hermitian eigenvalue computation. II: Computing several extreme eigenvalues. *SIAM J. Numer. Anal.*, 46(5):2593–2619, 2008.

[14] A. Perdon and G. Gambolati. Extreme eigenvalues of large sparse matrices by Rayleigh quotient and modified conjugate gradients. *Comput. Methods Appl. Mech. Eng.*, 56:251–264, 1986.

[15] V.G. Prikazchikov. Strict estimates of the rate of convergence of an iterative method of computing eigenvalues. *USSR J. Comput. Math. and Math. Physics*, 15:235–239, 1975.

[16] B.A. Samokish. The steepest descent method for an eigenvalue problem with semi-bounded operators. *Izv. Vyssh. Uchebn. Zaved. Mat.*, 5:105–114, 1958. (In Russian).

[17] G.L.G. Sleijpen and H.A. van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 17:401–425, 1996.

[18] X. Yang. A survey of various conjugate gradient algorithms for iterative solution of the largest/smallest eigenvalue and eigenvector of a symmetric matrix, Collection: Application of conjugate gradient method to electromagnetic and signal analysis. *Progress Electromagnetic Res.*, 5:567–588, 1991.

[19] M. Zhou. Doctoral thesis to be completed. Universität Rostock, 2011.

[20] P. F. Zhuk and L. N. Bondarenko. Sharp estimates for the rate of convergence of the *s*-step method of steepest descent in eigenvalue problems. *Ukraïn. Mat. Zh.*, 49(12):1694–1699, 1997.