SIAM J. NUMER. ANAL.
Vol. 50, No. 6, pp. 3188–3207

A GEOMETRIC CONVERGENCE THEORY FOR THE PRECONDITIONED STEEPEST DESCENT ITERATION*

KLAUS NEYMEYR†

Abstract. Preconditioned gradient iterations for very large eigenvalue problems are efficient solvers with growing popularity. However, only for the simplest preconditioned eigensolver, namely the preconditioned gradient iteration (or preconditioned inverse iteration) with fixed step size, are sharp nonasymptotic convergence estimates known. These estimates require a properly scaled preconditioner. In this paper a new sharp convergence estimate is derived for the *preconditioned steepest descent iteration* which combines the preconditioned gradient iteration with the Rayleigh–Ritz procedure for optimal line search convergence acceleration. The new estimate always improves that of the fixed-step-size iteration. The practical importance of this new estimate is that arbitrarily scaled preconditioners can be used. The Rayleigh–Ritz procedure implicitly computes the optimal scaling constant.

Key words. eigenvalue computation, Rayleigh quotient, gradient iteration, steepest descent, preconditioner

AMS subject classifications. 65N12, 65N22, 65N25, 65N30

DOI. 10.1137/11084488X

1. Introduction.

The topic of this paper is a convergence analysis of a *preconditioned gradient iteration with optimal step length* in order to compute the smallest eigenvalue of the generalized eigenvalue problem

$$Ax_i = \lambda_i B x_i \tag{1.1}$$

for symmetric positive definite matrices $A, B \in \mathbb{R}^{n\times n}$. Let T, a preconditioner, approximate the inverse of A. Then the preconditioned gradient iteration with optimal step length ϑ_{opt} reads as

$$x' = x - \vartheta_{\mathrm{opt}} T(Ax - \rho(x)Bx). \tag{1.2}$$

The parameter ϑ_{opt} is determined in such a way that the Rayleigh quotient

$$\rho(x) = \frac{(x, Ax)}{(x, Bx)} \tag{1.3}$$

of x' is minimized. The iterate x' is computed by the Rayleigh–Ritz procedure applied to the two-dimensional space spanned by x and the preconditioned residual $T(Ax - \rho(x)Bx)$. This paper provides a new and sharp estimate for the convergence of $\rho(x')$ toward the smallest eigenvalue of (1.1). The central result is given in Theorem 1.2. To state this theorem some further notation will be introduced, and the assumptions on the preconditioner will be fixed.

A typical source of (1.1) is an eigenproblem for a self-adjoint and elliptic partial differential operator whose weak form reads as

$$a(u, v) = \lambda\,(u, v) \qquad \forall v \in H(\Omega). \tag{1.4}$$

*Received by the editors August 18, 2011; accepted for publication (in revised form) August 14, 2012; published electronically December 4, 2012.
http://www.siam.org/journals/sinum/50-6/84488.html

†Institut für Mathematik, Universität Rostock, 18057 Rostock, Germany (klaus.neymeyr@uni-rostock.de).

The bilinear form $a(\cdot,\cdot)$ is associated with the partial differential operator, and an $L^2(\Omega)$ inner product $(\cdot,\cdot)$ appears on the right side. Further, u is an eigenfunction and λ an eigenvalue if (1.4) is satisfied for all v in an appropriate Hilbert space $H(\Omega)$. A finite element discretization of (1.4) results in (1.1). Then A is called the discretization matrix and B the mass matrix. These matrices are typically sparse and very large.

The eigenvalues of (1.1) are enumerated in increasing order $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. The smallest eigenvalue λ_1 and an associated eigenvector can be computed by means of an iterative minimization of the Rayleigh quotient (1.3). To this end the simplest preconditioned gradient iteration corrects a current iterate x in the direction of the negative preconditioned gradient of the Rayleigh quotient to form the next iterate x':

$$x' = x - T(Ax - \rho(x)Bx). \tag{1.5}$$

This fixed-step-length preconditioned iteration is analyzed in [2, 8, 7, 10]; see also the references in [3]. Embree and Lehoucq demonstrated in [4] a fruitful relationship between simple preconditioned eigensolvers like (1.5) and nonlinear dynamical systems.

Appropriate preconditioners T are available in various ways; especially for the operator eigenproblem (1.4), multigrid or multilevel preconditioners are available. In this context the quality of the preconditioner is typically controlled in terms of a real parameter $\gamma \in [0,1)$ in such a way that

$$(1-\gamma)(z, T^{-1}z) \leq (z, Az) \leq (1+\gamma)(z, T^{-1}z) \quad \forall z \in \mathbb{R}^n, \tag{1.6}$$

or, equivalently, that the spectral radius of the error propagation matrix $I - TA$ is bounded by γ.

The following result for the convergence of (1.5) is known from [8, 10]; the convergence analysis interprets this preconditioned iteration as a preconditioned inverse iteration and makes use of the underlying geometry.

Theorem 1.1. *The iterates of* (1.5) *together with* (1.6) *form a sequence with monotone decreasing Rayleigh quotients. The Rayleigh quotients converge to an eigenvalue, and the iteration vectors converge to an associated eigenvector.*

If the Rayleigh quotient of an iterate x satisfies $\lambda_i \leq \rho(x) < \lambda_{i+1}$, then either the Rayleigh quotient of the next iterate x' fulfills $\rho(x') \leq \lambda_i$ or the following sharp estimate applies:

$$\frac{\rho(x') - \lambda_i}{\lambda_{i+1} - \rho(x')} \leq \sigma^2 \frac{\rho(x) - \lambda_i}{\lambda_{i+1} - \rho(x)}, \qquad \sigma = \gamma + (1-\gamma)\frac{\lambda_i}{\lambda_{i+1}}. \tag{1.7}$$

Theorem 1.1 is a sharp estimate for the fixed-step-length preconditioned gradient iteration and serves as an upper estimate for various improved and faster converging preconditioned gradient-type eigensolvers. The most popular of these improved solvers are the preconditioned steepest descent (PSD) iteration and the locally optimal preconditioned conjugate gradient (LOPCG) iteration (and also their block variants) [7]. All these eigensolvers apply the Rayleigh–Ritz procedure to proper subspaces of iterates for convergence acceleration; see [9]. A systematic hierarchy of these preconditioned gradient iterations and their variants for exact inverse preconditioning (which amounts to certain invert-Lanczos processes [17]) has been suggested in [15]. The aim of this paper is to prove a new sharp convergence estimate for the PSD iteration.

1.1. Assumptions on the preconditioner. A drawback of Theorem 1.1 is its assumption (1.6) on the preconditioner T. The existence of constants $1 \pm \gamma$ with $\gamma < 1$ is not guaranteed for arbitrary (multigrid) preconditioners, but can always be ensured after a proper scaling of the preconditioner. To make this clear, take an arbitrary pair of symmetric positive definite matrices $A, T \in \mathbb{R}^{n\times n}$. Then constants γ_1, $\gamma_2 > 0$ exist, so that the spectral equivalence

$$\gamma_1(z, T^{-1}z) \le (z, Az) \le \gamma_2(z, T^{-1}z) \quad \forall z \in \mathbb{R}^n \tag{1.8}$$

holds. If a preconditioner T satisfies (1.8), then the scaled preconditioner $(2/(\gamma_1 + \gamma_2))T$ fulfills (1.6) with

$$\gamma = \frac{\gamma_2 - \gamma_1}{\gamma_1 + \gamma_2}. \tag{1.9}$$

A clear benefit of the preconditioned steepest descent iteration is that, by computing the optimal-step-length parameter ϑ_{opt} (see (1.2)), the scaling parameter $2/(\gamma_1 + \gamma_2)$ is determined implicitly. Therefore, we can use the assumption (1.8) or alternatively the more convenient form (1.6). This guarantees the practical applicability of the PSD iteration for any preconditioner satisfying (1.8) or in its scaled form satisfying (1.6).

1.2. The optimal-step-length iteration: Preconditioned steepest descent. A disadvantage of the gradient iteration (1.5) is its fixed step length resulting in a nonoptimal new iterate x'. An obvious improvement is to compute x' as the minimizer of the Rayleigh quotient (1.3) in the affine space $\{x - \vartheta T(Ax - \rho(x)Bx);\ \vartheta \in \mathbb{R}\}$. That means we consider the optimally scaled iteration

$$x' = x - \vartheta_{\text{opt}} T(Ax - \rho(x)Bx) \tag{1.10}$$

with the optimal step length

$$\vartheta_{\text{opt}} = \arg\min_{\vartheta \in \mathbb{R}} \rho(x - \vartheta T(Ax - \rho(x)Bx)).$$

This iteration is called the *preconditioned steepest descent iteration* (PSD) [2, 9, 20]. Computationally one gets x' and its Rayleigh quotient $\rho(x')$ by the Rayleigh–Ritz procedure. If $T(Ax - \rho(x)Bx)$ is not an eigenvector, then $(x', \rho(x'))$ is a Ritz pair of (A, B) with respect to the column space of $[x, T(Ax - \rho(x)Bx)]$. As (1.2) aims at a minimization of the Rayleigh quotient, $\rho(x')$ is the smaller Ritz value and x' is an associated Ritz vector. The Rayleigh–Ritz procedure computes the optimal step length implicitly; the step length is determined by the components of the associated eigenvector of Rayleigh–Ritz projection matrices. Consequently the PSD iteration converges faster than the fixed-step-length scheme (1.5) since

$$\rho(x - \vartheta_{\text{opt}} T(Ax - \rho(x)Bx)) \le \rho(x - T(Ax - \rho(x)Bx)). \tag{1.11}$$

Therefore Theorem 1.1 serves as a trivial upper estimate for the accelerated iteration (1.2). The aim of this paper is to prove the following sharp convergence estimate for (1.2).

THEOREM 1.2. *Let $x \in \mathbb{R}^n$ and let x' be the PSD iterate given by* (1.2). *The preconditioner T is assumed to satisfy* (1.8). *If $\lambda_i \le \rho(x) < \lambda_{i+1}$, $i = 1, \dots, n-1$,* *then* $\rho(x') \leq \rho(x)$ *and either* $\rho(x') \leq \lambda_i$ *or*

$$\frac{\rho(x') - \lambda_i}{\lambda_{i+1} - \rho(x')} \leq \sigma^2 \frac{\rho(x) - \lambda_i}{\lambda_{i+1} - \rho(x)}, \tag{1.12}$$
$$\text{with} \quad \sigma = \frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa}, \qquad \kappa = \frac{\lambda_i(\lambda_n - \lambda_{i+1})}{\lambda_{i+1}(\lambda_n - \lambda_i)},$$

and $\gamma := (\gamma_2 - \gamma_1)/(\gamma_1 + \gamma_2)$. *The estimate is sharp and can be attained for* $\rho(x) \to \lambda_i$ *in the three-dimensional invariant subspace associated with the eigenvalues* λ_i, λ_{i+1}, *and* λ_n, $i + 1 \neq n$.

The definition $\gamma := (\gamma_2 - \gamma_1)/(\gamma_1 + \gamma_2)$ in Theorem 1.2 is consistent with γ in (1.6) if $\gamma_1 = 1 - \gamma$ and $\gamma_2 = 1 + \gamma$. The limit case $\gamma = 0$ of Theorem 1.2 is an estimate for the convergence of the steepest descent iteration which minimizes the Rayleigh quotient in the space $\mathrm{span}\{x, A^{-1}Bx\}$. Then the convergence estimate (1.12) reads as

$$\frac{\rho(x') - \lambda_i}{\lambda_{i+1} - \rho(x')} \leq \left(\frac{\kappa}{2-\kappa}\right)^2 \frac{\rho(x) - \lambda_i}{\lambda_{i+1} - \rho(x)}$$

with κ given by (1.12). A proof of this result (in the general setup of steepest ascent and steepest descent for A and A^{-1}) has recently been given in [18]; for the smallest eigenvalue with $i = 1$ the estimate was proved in [11]. This paper generalizes this result on steepest descent for $A^{-1}M$ to the preconditioned variant of this iteration; see also Appendix A for comments on the connection of the preconditioned and the nonpreconditioned iteration. For the following analysis we always assume a properly scaled preconditioner satisfying (1.6). If T fulfills (2.3), we use $(2/(\gamma_1 + \gamma_2))T$ (and call the scaled preconditioner once again T) so that γ is given by (1.9) and (1.6) is fulfilled. This substitution does not restrict the generality of the approach since the scaling constant is implicitly computed with ϑ_{opt} in the Rayleigh–Ritz procedure. We prefer to work with (1.6) since this allows us to set up the proper geometry for the following proof.

Only a few convergence estimates on PSD have been published. Of major importance are the work of Samokish [21], the results of Knyazev given in Theorem 3.3 together with equation (3.3) in [6], and, further, the results of Ovtchinnikov [20]. Knyazev uses similar assumptions and applies Chebyshev polynomials to derive the convergence estimate. Ovtchinnikov in [20] proves nonasymptotic estimates and also an asymptotic convergence factor which represents the average error reduction per iteration. Further, Ovtchinnikov reproduces the result of Samokish in a finite-dimensional nonasymptotic form as Theorem 2.1 in [20] and presents a review and comparison of the asymptotic behavior of various estimates for preconditioned gradient iterations; see section 3 in [20].

Next we compare the estimate given in Theorem 1.2 with the important Corollary 6.4 of Ovtchinnikov [20]. If $\mu_2 < \mu(x) \leq \mu_1$, then the latter result reads in our notation as

$$\mu_1 - \mu(x') \leq \tau^2(\mu_1 - \mu(x)), \tag{1.13}$$
$$\text{with } \tau = \frac{1-\xi}{1+\xi}, \qquad \xi = \frac{1}{\varkappa(TA)} \frac{\mu(x) - \mu_2}{\mu(x) - \mu_n}.$$

Therein $\varkappa(TA)$ is the spectral condition number of TA. The estimate (1.13) is formulated in terms of reciprocal eigenvalues $\mu_i = 1/\lambda_i$ together with the reciprocal

Rayleigh quotient $\mu(x) = 1/\rho(x)$. A reformulation of our Theorem 1.2 in terms of this reciprocal representation is given in Theorem 2.2.

In contrast to (1.13), Theorem 2.2 can be applied to any initial iterate x without assuming $\mu_2 < \mu(x) \le \mu_1$. The restriction of Theorem 2.2 to $i = 1$ reads as

$$\frac{\mu_1 - \mu(x')}{\mu(x') - \mu_2} \le \sigma^2 \frac{\mu_1 - \mu(x)}{\mu(x) - \mu_2}, \tag{1.14}$$

$$\text{with } \sigma = \frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa} \quad \text{and} \quad \kappa = \frac{\mu_2 - \mu_n}{\mu_1 - \mu_n}.$$

In order to convert (1.14) into the form (1.13) we consider the limit $\mu(x), \mu(x') \to \mu_1$ for the denominators on the left and right sides of (1.14). Then the asymptotic behavior for $\mu(x) \to \mu_1$ of the convergence factors τ in (1.13) and σ in (1.14) can easily be compared. With $\xi = (\mu_1 - \mu_2)/(\varkappa(TA)(\mu_1 - \mu_n))$ and $\varkappa(TA) = (1+\gamma)/(1-\gamma)$ from (1.6) we obtain that

$$\tau = \frac{\mu_2 - \mu_n + \gamma(2\mu_1 - \mu_2 - \mu_n)}{2\mu_1 - \mu_2 - \mu_n + \gamma(\mu_2 - \mu_n)} = \sigma.$$

Hence asymptotically (1.13) and (1.14) coincide. An important difference is that Theorem 1.2 can be applied to any iterate x without assuming the Rayleigh quotient of x between the largest eigenvalues (μ-representation) or smallest eigenvalues (λ-representation).

1.3. Overview. This paper is organized as follows. In section 2 the geometry of PSD is introduced. Further, the problem is reformulated in terms of reciprocals of the eigenvalues, which makes the geometry of PSD accessible within the Euclidean space. Section 3 gives a proof that PSD attains its poorest convergence in a three-dimensional invariant subspace of the $\mathbb{R}^n$. Section 4 contains a minidimensional analysis of PSD. Finally, the three-dimensional convergence estimates are embedded into the full $\mathbb{R}^n$, which completes the convergence analysis.

2. The geometry of the PSD iteration.

For the analysis of the PSD iteration it is convenient to work with the linear pencil $B - \mu A$ (instead of $A - \lambda B$). The advantage is that the A-norm by a proper basis transformation turns into the Euclidean norm; see below. A further benefit of this representation is that a generalization to a symmetric positive semidefinite or even only a symmetric B is possible (cf. the analysis of (1.5) in [10]). Hence for the pencil $B - \mu A$ the eigenvalues μ_i are given by

$$Bx_i = \mu_i A x_i \quad \text{with } \mu_i = 1/\lambda_i, \quad i = 1, \dots, n.$$

Therefore the problem is to compute the largest eigenvalue μ_1 by maximizing the inverse of the Rayleigh quotient (1.3):

$$\mu(x) := \frac{(x, Bx)}{(x, Ax)} = \frac{1}{\rho(x)}. \tag{2.1}$$

LEMMA 2.1. *A proper change of the basis allows us to assume for the convergence analysis of* (1.2) *that $A = I$ and that $B = \operatorname{diag}(\mu_1, \dots, \mu_n)$. This allows us to transform* (1.2) *(after multiplication with $\mu(x) = 1/\rho(x)$ and by denoting the transformed preconditioner again by T) in the simplified form*

$$\mu(x)x' = \mu(x)x + \vartheta_{\mathrm{opt}} T(Bx - \mu(x)x) \tag{2.2}$$

with the optimal step length

$$\vartheta_{\text{opt}} = \arg\max_{\vartheta\in\mathbb{R}} \mu(\mu(x)x + \vartheta T(Bx - \mu(x)x)).$$

The quality constraint (1.6) *on the preconditioner* $T \in \mathbb{R}^{n\times n}$ *turns into a bound for the spectral norm* $\|\cdot\|$ *of the symmetric matrix* $I - T$ *which reads as*

$$\|I - T\| \leq \gamma. \tag{2.3}$$

Further, multiple eigenvalues do not change the convergence estimates. This provides the justification to assume that $\mu_1 > \mu_2 > \cdots > \mu_n > 0$.

Proof. The generalized eigenvalue problem (1.1) is first transformed into a standard eigenvalue problem $C^{-1}BC^{-T}y = \mu y$ using the Cholesky factorization $A = CC^T$, $y = C^T x$, and $\mu = 1/\lambda$. The symmetric matrix $C^{-1}BC^{-T}$ can be diagonalized by means of an orthogonal similarity transformation. Then all transformations are applied to (1.2). For convenience, we denote the transformed system matrix by B. Further, the transformed preconditioner is denoted, once again, by T, since (1.6) still holds with $A = I$. All this results in (2.2) and (2.3).

For the fixed-step-size gradient iteration with $\vartheta_{\text{opt}} = 1$ it is known that the multiplicity of all eigenvalues can be assumed to be equal to 1 for the convergence analysis. The proof of this fact either uses a projection argument (see section 3 in [13]) or is based on a continuity argument (see Theorem 2.1 in [10]). These arguments can be transferred to PSD. In particular, the continuity argument can be applied since the additional Rayleigh–Ritz procedure, which leads to ϑ_{opt}, preserves the continuity of the eigenvalue approximations. ∎

Next the reformulation of Theorem 1.2 in terms of the μ-notation is stated.

THEOREM 2.2. *If* $\mu_{i+1} < \mu(x) \leq \mu_i$ *then* $\mu(x') \geq \mu(x)$ *and either* $\mu(x') \geq \mu_i$ *or*

$$\begin{aligned}&\frac{\mu_i - \mu(x')}{\mu(x') - \mu_{i+1}} \leq \sigma^2 \frac{\mu_i - \mu(x)}{\mu(x) - \mu_{i+1}},\\ &\textit{with } \sigma = \frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa} \quad \textit{and} \quad \kappa = \frac{\mu_{i+1} - \mu_n}{\mu_i - \mu_n}.\end{aligned} \tag{2.4}$$

The estimate is sharp and can be attained for $\mu(x) \to \mu_i$ *in the three-dimensional invariant subspace associated with the eigenvalues* μ_i, μ_{i+1}, *and* μ_n, $i + 1 \neq n$.

2.1. The cone of PSD iterates. The starting point of the geometric description of PSD is the *non-scaled* preconditioned gradient iteration (1.5) whose μ-representation reads as

$$\mu(x)x' = \mu(x)x + T(Bx - \mu(x)x) = Bx - (I - T)(Bx - \mu(x)x). \tag{2.5}$$

A central idea of its convergence analysis in [13, 14, 8] is to treat the preconditioners *on the whole*. This means that all admissible preconditioners satisfying the spectral equivalence (2.3) are inserted into (2.5) with x being fixed. This results in a set $\mathcal{B}_\gamma(x)$ of all possible iterates:

$$\mathcal{B}_\gamma(x) := \{Bx - (I - T)(Bx - \mu(x)x);\ T \text{ symmetric and positive definite with } \|I - T\| \leq \gamma\}. \tag{2.6}$$

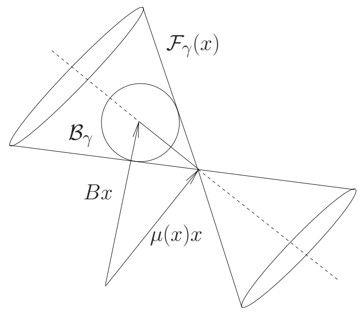

FIG. 2.1. *The circular cone* $\mathcal{F}_\gamma(x)$.

The set $\mathcal{B}_\gamma(x)$ is a full ball with the center Bx and the radius $\gamma\|Bx-\mu(x)x\|$. The subject of the convergence analysis of (2.5) in [13, 14] is to localize a vector of poorest convergence (i.e., with the smallest Rayleigh quotient) in $\mathcal{B}_\gamma(x)$ and to derive an estimate for its Rayleigh quotient.

In contrast to (2.5) the PSD iteration (2.2) works with an optimal-step-length parameter ϑ_{opt} in order to maximize the Rayleigh quotient in the one-dimensional affine space

$$\mu(x)x+\vartheta T(Bx-\mu(x)x) \qquad \vartheta\in\mathbb{R}. \tag{2.7}$$

The union of all these affine spaces for all the preconditioners satisfying (2.3) is the smallest circular cone with its vertex in $\mu(x)x$ which encloses $\mathcal{B}_\gamma(x)$. This cone is denoted by $\mathcal{F}_\gamma(x)$ (see Figure 2.1), and it holds that

$$\begin{aligned}\mathcal{F}_\gamma(x) &:= \{\mu(x)x+\vartheta(y-\mu(x)x);\ y\in\mathcal{B}_\gamma(x);\ \vartheta\in\mathbb{R}\}\\ &= \{\mu(x)x+\vartheta d;\ \|Bx-(\mu(x)\,x+d)\|\le\gamma\|Bx-\mu(x)\,x\|;\ \vartheta\in\mathbb{R}\}.\end{aligned} \tag{2.8}$$

2.2. The geometric convergence analysis as a two-level optimization. The geometric convergence analysis of PSD consists of estimating the poorest convergence behavior. Therefore a two-level optimization problem is to be solved. On the one hand, one has to determine this affine space (2.7) in the cone $\mathcal{F}_\gamma(x)$ in which the maximum of the Rayleigh quotient (i.e., the largest Ritz value in this space) takes its smallest value; this vector is associated with the poorest convergence due to the choice of the preconditioner. On the other hand, the cone $\mathcal{F}_\gamma(x)$ depends on x; hence one can analyze the dependence of this vector of poorest convergence on all vectors in the $\mathbb{R}^n$ having the same Rayleigh quotient as x. This amounts to considering the level set of the Rayleigh quotient of vectors having a fixed Rayleigh quotient μ_0, i.e.,

$$\mathcal{L}(\mu_0):=\{x\in\mathbb{R}^n;\ \mu(x)=\mu_0\}.$$

Let $x^*\in\mathcal{L}(\mu_0)$ be a minimizer representing the poorest convergence and let $d^*\in\mathcal{F}_\gamma(x)-\mu(x)x$ be a search direction of poorest convergence. So the *two-level optimization* is

$$\underline{\mu}:=\min_{x\in\mathcal{L}(\mu_0)}\ \min_{d\in\mathcal{F}_\gamma(x)-\mu_0 x}\mu(\mu_0 x+\vartheta_{\mathrm{opt}}d).$$

Therein $\mu(x)x+\vartheta_{\mathrm{opt}}d$ is a Ritz vector which is associated with the larger Ritz value $\mu(x+\vartheta_{\mathrm{opt}}d)$ in $\mathrm{span}\{x,d\}$. Sometimes we write $\vartheta_{\mathrm{opt}}=\vartheta_{\mathrm{opt}}[x,d]$ to express its dependence on x and d. The minimum $\underline{\mu}$ is now to be estimated from below.

3. The level set optimization—a reduction to three dimensions. The aim of this section is to show that the poorest convergence of PSD with respect to the admissible preconditioners and with respect to all vectors $x \in \mathcal{L}(\mu_0)$ is attained in a three-dimensional B-invariant subspace of the $\mathbb{R}^n$.

The representation (2.7) of the PSD iteration applies the line search to $d \in \mathcal{F}_\gamma(x) - \mu(x)x$. This may result in an unbounded step length. To see this, let $d = e_1 = (1, 0, \ldots, 0)^T$, which is an eigenvector of B. If γ is close to 1, then $e_1 \in \mathcal{F}_\gamma(x) - \mu(x)x$ can be attained since $\lim_{\gamma \to 1} \mathcal{F}_\gamma(x) = \mathbb{R}^n$. The unboundedness is a consequence of $\lim_{\vartheta \to \pm\infty} \mu(\mu(x)x + \vartheta e_1) = \mu_1$. The potential unboundedness of the step length has already been pointed out by Knyazev [12].

Next we want to avoid this singularity. Therefore let $x' = \vartheta x + d$. Due to $\mu(x') > \mu(x)$ (which is guaranteed by Theorem 1.1), ϑ is bounded. So the minimization problem is reformulated as

$$\underline{\mu} := \min_{x \in \mathcal{L}(\mu_0)} \min_{d \in \mathcal{F}_\gamma(x) - \mu_0 x} \mu(\vartheta_{\text{opt}}[x, d]x + d). \tag{3.1}$$

In the next theorem a necessary condition characterizing this minimum is derived by means of the Kuhn–Tucker conditions [19]. The application of the Kuhn–Tucker conditions in the context of the convergence analysis of the fixed-step-size preconditioned gradient iteration has been suggested by Argentati; see [1].

THEOREM 3.1. *The minimum* (3.1) *is attained in a three-dimensional B-invariant subspace of the $\mathbb{R}^n$.*

If PSD does not terminate in an eigenvector, then the associated Ritz vector w of poorest convergence is also contained in the same three-dimensional B-invariant subspace of the $\mathbb{R}^n$, i.e.,

$$(B + a)w = c(B + b)x$$

with $a, b, c \in \mathbb{R}$ and $B + a$ being a regular matrix.

Proof. The minimization problem (3.1) reads as follows:

Minimize

$$\mu(\vartheta_{\text{opt}} x + d)$$

with respect to $x, d \in \mathbb{R}^n$ satisfying the following two constraints:

1. The cone inequality constraint that $d \in \mathcal{F}_\gamma(x) - \mu_0 x$ holds if
$$\begin{aligned} g(x, d) &= \|Bx - (\mu_0 x + d)\|^2 - \gamma^2 \|Bx - \mu_0 x\|^2 \\ &= (1 - \gamma^2)\|Bx - \mu_0 x\|^2 - 2(Bx - \mu_0 x, d) + \|d\|^2 \leq 0. \end{aligned}$$
2. The level set constraint that $x \in \mathcal{L}(\mu_0)$ is satisfied if
$$h(x, d) = (x, Bx) - \mu_0(x, x) = 0.$$

Therein $\vartheta_{\text{opt}} = \vartheta_{\text{opt}}[x, d] \in \mathbb{R}$ is a functional depending on x and d which maximizes the Rayleigh quotient in the two-dimensional subspace span$\{x, d\}$. Equivalently $w := \vartheta_{\text{opt}} x + d$ is a Ritz vector corresponding to the larger Ritz value in just this two-dimensional subspace. The matrix elements of the Rayleigh–Ritz projection of B with respect to span$\{x, d\}$ smoothly depend on x and d and so do the Ritz values and eigenprojections since the number of distinct Ritz values, which equals 2, does not change; see Kato [5, Supplementary Notes on Chapter II]. Thus, ϑ_{opt}, which depends on just the eigenprojection associated with the larger eigenvalue, is a smooth function.

The first constraint guarantees that d is an admissible search direction; i.e., the distance of $\mu_0 x + d$ to the center Bx of the ball $\mathcal{B}_\gamma(x)$ is bounded by its radius $\gamma\|Bx - \mu_0 x\|$. The Karush–Kuhn–Tucker stationarity condition for a local minimizer (x^*, d^*) reads as

$$\nabla_{(x,d)}\mu(\vartheta_{\text{opt}}x^* + d^*) + \alpha\nabla_{(x,d)}g(x^*, d^*) + \beta\nabla_{(x,d)}h(x^*, d^*) = 0$$

with the multipliers α and β. In order to simplify the notation, the asterisks are omitted from now on.

Next we derive the gradients of these functions μ, g, and h with respect to x and d. The chain rule gives (for column vectors)

$$\nabla_x(\mu(\vartheta_{\text{opt}}x + d)) = (D_x(\vartheta_{\text{opt}}x + d))^T\,(\nabla\mu)(\vartheta_{\text{opt}}x + d).$$

It holds that

$$(D_x(\vartheta_{\text{opt}}x + d))_{ij} = (x(\nabla_x\vartheta_{\text{opt}})^T + \vartheta_{\text{opt}}I)_{ij}.$$

With $w := \vartheta_{\text{opt}}x + d$ we get

$$\begin{aligned}\nabla_x(\mu(\vartheta_{\text{opt}}x + d)) &= \vartheta_{\text{opt}}(\nabla\mu)(w) + (\nabla_x\vartheta_{\text{opt}})\,(x, (\nabla\mu)(w))\\ &= \vartheta_{\text{opt}}(\nabla\mu)(w) = \vartheta_{\text{opt}}\frac{2}{(w,w)}(Bw - \mu(w)w).\end{aligned}$$

Therein, $(x, (\nabla\mu)(w)) = 0$ has been used, which holds since $(\nabla\mu)(w)$ is collinear to the residual of the Ritz vector and further, by definition of a Ritz vector, its residual is orthogonal to the approximating subspace $\text{span}\{x, d\}$. For the d-gradient it holds that

$$\nabla_d(\mu(\vartheta_{\text{opt}}x + d)) = (\nabla\mu)(w) = \frac{2}{(w,w)}(Bw - \mu(w)w).$$

The gradients of the constraining functions g and h with $r = Bx - \mu_0 x$ are

$$\begin{aligned}\nabla_x g(x,d) &= (1-\gamma^2)2(B-\mu_0)r - 2(B-\mu_0)d, &\qquad \nabla_x h(x,d) &= 2r,\\ \nabla_d g(x,d) &= -2(B-\mu_0)x + 2d = 2(d-r), &\qquad \nabla_d h(x,d) &= 0.\end{aligned}$$

Hence the x-components of the Karush–Kuhn–Tucker stationarity condition are

(3.2)
$$\frac{\vartheta_{\text{opt}}}{(w,w)}(B-\mu(w))w + \alpha\Big\{(1-\gamma^2)(B-\mu_0)^2x - (B-\mu_0)(w - \vartheta_{\text{opt}}x)\Big\} + \beta r = 0,$$

and the d-components read as $(Bw - \mu(w)w) + \alpha(w,w)(d-r) = 0$. The equation for the d-components can be reformulated as

$$(B+a)w = \alpha(w,w)(B+b)x \tag{3.3}$$

with $a = \alpha(w,w) - \mu(w)$ and $b = \vartheta_{\text{opt}} - \mu_0$. Multiplication of (3.2) with $B + a$ and insertion of (3.3) results in

$$\begin{aligned}&\alpha\Big\{(1-\gamma^2)(B-\mu_0)^2(B+a)x - (B-\mu_0)\left[\alpha(w,w)(B+b)x - \vartheta_{\text{opt}}(B+a)x\right]\Big\}\\ &+ \alpha\vartheta_{\text{opt}}(B-\mu(w))(B+b)x + \beta(B+a)(B-\mu_0)x = 0.\end{aligned}$$

This can be expressed as

$$p_3(B)x = 0 \tag{3.4}$$

with a third order polynomial p_3. Due to the basis assumptions, B is a diagonal matrix and so $p_3(B)$ is diagonal. As p_3 has at most three different zeros, (3.4) can hold only if x has at most three nonzero components, which proves the first assertion.

Hence $x \in \operatorname{span}\{e_j, e_k, e_l\}$ for proper indexes j, k, and l. For this x (3.3) shows that w has no more than four nonzero components; four nonzero components are possible only if $a = -\mu_s$ for $s \neq j, k, l$. Then (3.2) can be written as $p_1(B)w = p_2(B)x \in \operatorname{span}\{e_j, e_k, e_l\}$ with a first order polynomial p_1 and a second order polynomial p_2. The latter equation implies that $p_1(\mu_s) = p_1(-a) = 0$. The sth component of the polynomial identity results in $a = (\alpha\mu_0(w,w) - \mu(w)\vartheta_{\mathrm{opt}})/(\vartheta_{\mathrm{opt}} - \alpha(w,w))$. Together with the known form $a = \alpha(w,w) - \mu(w)$ we get by direct computation that $a = b$. Insertion of this result to (3.3) shows that $w = \alpha(w,w)x + Ce_s$ for a real constant C. Then $x \perp e_s$, and x and e_s are the Ritz vectors. PSD terminates in e_s and w with no more than three nonzero components is the normal case. ∎

4. The cone optimization—a mini-dimensional geometric analysis.

Next the convergence behavior with respect to the cone $\mathcal{F}_\gamma(x)$ is analyzed. Some of the following arguments are valid in the $\mathbb{R}^n$; however, we need these properties only for $n = 3$.

The (half) opening angle φ of the cone $\mathcal{F}_\gamma(x)$ is given by $\sin\varphi = \gamma$, since γ is the ratio of the radius $\gamma\|Bx - \mu(x)x\|$ of the ball $\mathcal{B}_\gamma(x)$ (see (2.6)), and its (maximal) radius $\|Bx - \mu(x)x\|$ for $\gamma \to 1$. With $\cos\varphi = \sqrt{1-\gamma^2}$ the cone $\mathcal{F}_\gamma(x)$ can be written as

$$\mathcal{F}_\gamma(x) := \mu(x)x + \left\{ z \in \mathbb{R}^n;\ \left| \left(\frac{z}{\|z\|}, \frac{Bx - \mu(x)x}{\|Bx - \mu(x)x\|} \right) \right| \geq \sqrt{1-\gamma^2} \right\}.$$

4.1. Restriction to nonnegative vectors.

The analysis of PSD can be restricted to componentwise nonnegative vectors $x \in \mathbb{R}^n$. The justification is as follows. Consider the Householder reflections $H_i = I - 2e_ie_i^T$ for which $x \mapsto H_ix$ changes the sign of the ith component of x. The Rayleigh quotient is invariant under H_i; i.e., $\mu(x) = \mu(H_ix)$. If v is an admissible search direction, i.e., $v \in \mathcal{F}_\gamma(x) - \mu(x)x$, then

$$\begin{aligned}\cos\measuredangle(v, Bx - \mu(x)x) &= \left(\frac{v}{\|v\|}, \frac{Bx - \mu(x)x}{\|Bx - \mu(x)x\|} \right) = \left(\frac{H_iv}{\|H_iv\|}, \frac{BH_ix - \mu(H_ix)H_ix}{\|BH_ix - \mu(H_ix)H_ix\|} \right) \\ &= \cos\measuredangle(H_iv, BH_ix - \mu(H_ix)H_ix),\end{aligned}$$

which means that H_iv encloses the same angle with the residual vector associated with H_ix. As for all $\alpha \in \mathbb{R}$

$$\mu(\mu(H_ix)H_ix + \alpha H_iv) = \mu(H_i(\mu(x)x + \alpha v)) = \mu(\mu(x)x + \alpha v),$$

any Rayleigh quotient in the cone $\mathcal{F}_\gamma(x)$ can be reproduced in the cone $\mathcal{F}_\gamma(H_ix)$ and vice versa. Thus the analysis can be restricted to $x \geq 0$.

4.2. The poorest convergence in the three-dimensional cone $\mathcal{F}_\gamma(x)$.

Any circular cross section S_γ^c (with nonzero radius) of $\mathcal{F}_\gamma(x)$ can serve to represent the admissible search directions; see Figure 4.1. Next we work with the disc

$$S_\gamma^c(x) := \mu(x)x + (1-\gamma^2)r + \{fy;\ y \in \mathbb{R}^3,\ \|y\| \leq 1,\ y \perp r\} \tag{4.1}$$

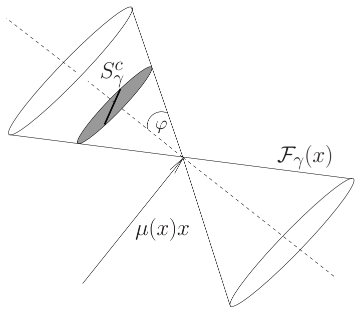

FIG. 4.1. *The cross section S_γ^c and the line segment S_γ (bold line).*

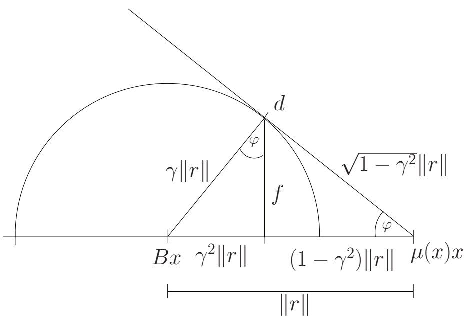

FIG. 4.2. *Three-dimensional-geometry with $r = Bx - \mu(x)x$, $f = \gamma\sqrt{1-\gamma^2}\,\|r\|$.*

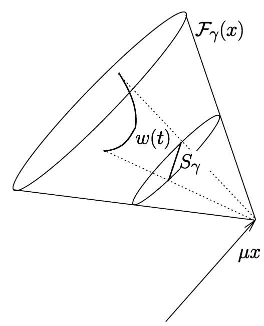

FIG. 4.3. *The line segment $S_\gamma(x)$ and the curve $w(t)$.*

with $r := Bx - \mu(x)x$. Its radius f (see Figure 4.2) is given by

$$f = \gamma\sqrt{1-\gamma^2}\|r\|. \tag{4.2}$$

Further we use only search directions $d \in S_\gamma^c(x) - \mu(x)x$ which are orthogonalized against x; this is justified since the Rayleigh–Ritz approximations (and so the PSD iterate x') depend only on the subspace. So the set of relevant search directions forms a line segment. By using the vector $v = x \times r/\|x \times r\| = x \times r/(\|x\|\,\|r\|)$ one can construct the intersection of this line segment with the surface of the cone. The points of intersection are $d_{1/2}$ with

$$d_1 = \mu(x)x + (1-\gamma^2)r + \gamma\sqrt{1-\gamma^2}\,\|r\|v, \tag{4.3}$$

$$d_2 = \mu(x)x + (1-\gamma^2)r - \gamma\sqrt{1-\gamma^2}\,\|r\|v, \tag{4.4}$$

$$v = \frac{x \times r}{\|x\|\,\|r\|}.$$

Therefore the line segment has the form (see Figure 4.3)

$$S_\gamma(x) := \{d(t) := td_1 + (1-t)d_2;\ t \in [0,1]\}. \tag{4.5}$$

LEMMA 4.1. *The poorest convergence of PSD in three dimensions (aside from the singular cases that PSD terminates in an eigenvector) is attained in d_1 or d_2 as given by* (4.3) *and* (4.4).

Proof. The line segment S_γ has the form $d(t)$ with $t \in [0,1]$ by (4.5). The PSD iteration maps S_γ into a curve $w(t)$, $t \in [0,1]$, where $w(t)$ is the Ritz vector $w(t) = \mu(x)x + \vartheta_{\mathrm{opt}}(t)d(t)$ corresponding to the larger Ritz value in $\mathrm{span}\{x, d(t)\}$. A singularity like that mentioned at the beginning of section 3 is not to be considered since otherwise the first alternative $\mu(x') \geq \mu_i$ in Theorem 2.2 applies and nothing is to be proved. Differentiability of $w(t)$ in t can be assumed as the normalized Ritz vector $w(t) = \mu(x)x + \vartheta_{\mathrm{opt}}(t)d(t)$ smoothly depends on t along the line segment $d(t)$ and also on $\vartheta_{\mathrm{opt}}(t)$, since the two Ritz values in $\mathrm{span}\{x, d(t)\}$ do not coincide; see (4.7) for the explicit form of the Ritz values.

Along $w(t)$ we are looking for a vector $w^* = w(t^*)$ so that

$$\mu(w(t^*)) \leq \mu(w(t)) \quad \forall t \in [0,1].$$

Since $w(t)$ is a Ritz vector, its residual $Bw(t) - \mu(w(t))w(t)$ is orthogonal to the subspace spanned by x and $d(t)$. As the residual is collinear to the gradient vector $\nabla\mu(w(t))$, we get

$$(\nabla\mu(w(t)), \mathrm{span}\{x, d(t)\}) = 0. \tag{4.6}$$

A stationary point of the Rayleigh quotient in a $t \in (0,1)$ is attained if

$$\begin{aligned} 0 = \frac{d}{dt}\mu(w(t)) &= (\nabla\mu(w(t)), w'(t)) \\ &= (\nabla\mu(w(t)), \vartheta'_{\mathrm{opt}}(t)d(t) + \vartheta_{\mathrm{opt}}(t)d'(t)) \\ &= (\nabla\mu(w(t)), \vartheta_{\mathrm{opt}}(t)d'(t)), \end{aligned}$$

where (4.6) has been used for the last identity. As $d'(t)$ is collinear to $x \times r$, we get from $(\nabla\mu(w(t)), d'(t)) = 0$ together with (4.6) that $\nabla\mu(w) = 0$ (since x, d, and d' span the $\mathbb{R}^3$). So any interior stationary point must be an eigenvector and hence $\mu(w(t))$ take the other extrema on the surface for $t = 0$ or $t = 1$ in d_1 or d_2. ∎

Next we apply the Rayleigh–Ritz procedure to the two-dimensional subspaces $[x, d_i - \mu(x)x]$, $i = 1, 2$, in order to determine whether the poorest convergence is attained in d_1 or d_2. First the Euclidean norm of $d_i - \mu(x)x$ is determined:

$$\begin{aligned} \|d_i - \mu(x)x\|^2 &= (1-\gamma^2)^2(r,r) \pm (1-\gamma^2)\gamma\sqrt{1-\gamma^2}(r, x \times r)/\|x\| \\ &\quad + \gamma^2(1-\gamma^2)\|x \times r\|^2/\|x\|^2 \\ &= (1-\gamma^2)^2\|r\|^2 + \gamma^2(1-\gamma^2)\|r\|^2 = (1-\gamma^2)\|r\|^2. \end{aligned}$$

Hence the normalized search directions $(d_i - \mu(x)x)/\|d_i - \mu(x)x\|$ are

$$\bar{d}_{1/2} := \frac{d_{1/2} - \mu(x)x}{\sqrt{1-\gamma^2}\|r\|} = \sqrt{1-\gamma^2}\,\frac{r}{\|r\|} \pm \gamma\,\frac{x \times r}{\|x\|\,\|r\|},$$

and therefore $V_1 = [x, \bar{d}_1]$ and $V_2 = [x, \bar{d}_2] \in \mathbb{R}^{3\times 2}$ are orthonormal matrices. The Ritz values of B in the column space of V_i are the eigenvalues of the projection

$$B_i := V_i^T B V_i = \begin{pmatrix} \mu(x) & (\bar{d}_i, Bx) \\ (\bar{d}_i, Bx) & \mu(\bar{d}_i) \end{pmatrix}.$$

The larger Ritz value (that is, the larger eigenvalue of B_i) reads as

$$\theta_{2,i} = \frac{\mu(x) + \mu(\bar{d}_i)}{2} + \sqrt{\frac{(\mu(x) - \mu(\bar{d}_i))^2}{4} + (\bar{d}_i, Bx)^2}. \tag{4.7}$$

In order to decide whether poorest convergence occurs in d_1 or in d_2, we show that the nondiagonal elements of B_i do not depend on i since

$$(\bar{d}_i, Bx) = (\bar{d}_i, Bx - \mu(x)x) = \|r\| \left(\bar{d}_i, \frac{r}{\|r\|}\right) = \|r\| \cos\measuredangle(\bar{d}_i, r) = \sqrt{1-\gamma^2}\|r\|. \tag{4.8}$$

Hence only the (2,2) element of B_i depends on i. As further

$$\frac{d\theta_{2,i}}{d\mu(\bar{d}_i)} = \frac{1}{2}\left(1 - \frac{1}{1 + \left(\frac{2(\bar{d}_i, Bx)}{\mu(x)-\mu(\bar{d}_i)}\right)^{1/2}}\right) > 0$$

shows that $\theta_{2,i}$ is a monotone increasing function of $\mu(\bar{d}_i)$, we still have to find the d_i with the smaller Rayleigh quotient in order to find the search direction which is associated with the poorer PSD convergence.

Lemma 4.2. *PSD in three dimensions takes its poorest convergence, i.e., the smallest value of* θ_2*, in*

$$d = \mu(x)x + (1-\gamma^2)r + \gamma\sqrt{1-\gamma^2}\,\frac{x \times r}{\|x\|} \tag{4.9}$$

if $x \in \mathbb{R}^n$ *is a componentwise nonnegative vector (cf. section* 4.1*). The associated Ritz value is accessible from* (4.7)*.*

Proof. We show that $\theta_{2,1}$ is the smaller Ritz value by showing (we use the monotonicity of $\theta_{2,i}[\mu(\bar{d}_i)]$) that $\mu(\bar{d}_1) \le \mu(\bar{d}_2)$. This inequality is true if $(r, B(x \times r)) \le 0$. By using $\operatorname{span}\{x, r\} \perp x \times r$ and $r \perp x$, direct computation results in

$$\begin{aligned}(r, B(x\times r)) &= (B(Bx - \mu(x)x), x \times r) = (B^2x, x\times r) - \mu(x)(Bx, x\times r)\\ &= (B^2x, x\times r) - \mu(x)(r + \mu(x)x, x\times r)\\ &= (B^2x, x\times r) = (r, B^2x \times x) = (Bx, B^2x\times x)\\ &= -x_1x_2x_3(\mu_1-\mu_2)(\mu_1-\mu_3)(\mu_2-\mu_3) \le 0.\end{aligned}$$

The last inequality holds since $x \ge 0$ and $\mu_1 > \mu_2 > \mu_3$. ∎

4.3. A mini-dimensional convergence analysis of PSD. Due to Theorem 3.1 the "mini-dimensional" convergence analysis can be restricted to three-dimensional B-invariant subspaces of the $\mathbb{R}^n$. With respect to the basis of eigenvectors these subspaces have the form $\operatorname{span}\{e_j, e_k, e_l\}$, where e_* is the $*$th unit vector. The associated eigenvalues are indexed so that $\mu_j > \mu_k > \mu_l$.

Lemma 4.2 delivers for any $x \in \mathcal{L}(\mu)$ in three dimensions the vector of $\mathcal{B}_\gamma(x)$-poorest PSD convergence. Next we have to analyze the $\mathcal{L}(\mu)$-dependence of the poorest convergence case.

Theorem 4.3. *In the three-dimensional space* $\operatorname{span}\{e_j, e_k, e_l\}$ *the following sharp estimate for PSD holds:*

$$\frac{\Delta_{j,k}(\mu')}{\Delta_{j,k}(\mu)} \le \left(\frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa}\right)^2$$

with

$$\Delta_{j,k}(\xi) = \frac{\mu_j - \xi}{\xi - \mu_k} \quad \textit{and} \quad \kappa = \frac{\mu_k - \mu_l}{\mu_j - \mu_l}.$$

Proof. The starting points of the following analysis are the vectors x and

$$d = \mu(x)x + (1-\gamma^2)r + \gamma\sqrt{1-\gamma^2}\frac{x \times r}{\|x\|}.$$

Without loss of generality, x can be normalized in such a way that

$$x = e_j + \alpha_0 e_k + \beta_0 e_l;$$

hence x is an element of the affine space $\mathcal{E}_j := e_j + \operatorname{span}\{e_k, e_l\}$. The coordinate form of x in three dimensions is then $x = (1, \alpha_0, \beta_0)^T$. Further, let $\tilde{d} = (1, \tilde{\alpha}, \tilde{\beta})^T \in \mathcal{E}_j$ be the corresponding multiple of d. Since $\operatorname{span}\{x, d\}$ is a tangential plane of the ball $\mathcal{B}_\gamma(x)$ in d and $Bx - d$ is a radius vector of the ball, it holds that

$$Bx - d \perp \operatorname{span}\{x, d\} = \operatorname{span}\{x, \tilde{d}\}. \tag{4.10}$$

Hence $Bx - d$ is collinear to

$$x \times \tilde{d} = (\alpha_0\tilde{\beta} - \tilde{\alpha}\beta_0, \beta_0 - \tilde{\beta}, \tilde{\alpha} - \alpha_0)^T.$$

By $S_1 = (1, c_k, 0)^T$ and $S_2 = (1, 0, c_l)^T$ with $S_1, S_2 \in \mathcal{E}_j$ we denote the points of intersection of $\operatorname{span}\{x, \tilde{d}\}$ with $e_j + \operatorname{span}\{e_k\}$ and $e_j + \operatorname{span}\{e_l\}$; see Figure 4.4. Due to (4.10) it holds that $(Bx - d, S_i) = 0$, $i = 1, 2$. Since

$$Bx - d = \gamma^2 r - \gamma\sqrt{1-\gamma^2}\,\frac{x \times r}{\|x\|},$$

we get with

$$r = \begin{pmatrix} \mu_j - \mu \\ (\mu_k - \mu)\alpha_0 \\ (\mu_l - \mu)\beta_0 \end{pmatrix}, \qquad x \times r = \begin{pmatrix} \alpha_0\beta_0(\mu_l - \mu_k) \\ \beta_0(\mu_j - \mu_l) \\ \alpha_0(\mu_k - \mu_j) \end{pmatrix}$$

from $(Bx - d, S_1) = 0$ that

$$c_k = -\frac{(Bx-d)\,|_1}{(Bx-d)\,|_2} = \frac{\|x\|(\mu_j - \mu) + \Gamma\alpha_0\beta_0(\mu_k - \mu_l)}{\|x\|\alpha_0(\mu - \mu_k) + \Gamma\beta_0(\mu_j - \mu_l)}. \tag{4.11}$$

Analogously, $(Bx - d, S_2) = 0$ results in

$$c_l = -\frac{(Bx-d)\,|_1}{(Bx-d)\,|_3} = \frac{\|x\|(\mu_j - \mu) + \Gamma\alpha_0\beta_0(\mu_k - \mu_l)}{\|x\|\beta_0(\mu - \mu_l) + \Gamma\alpha_0(\mu_k - \mu_j)} \tag{4.12}$$

with $\Gamma = \sqrt{1-\gamma^2}/\gamma$.

Any $x \in \mathcal{E}_j \cap \mathcal{L}(\mu)$ is an element of the ellipse $(x_k/a)^2 + (x_l/b)^2 = 1$ with

$$a = \sqrt{\frac{\mu_j - \mu}{\mu - \mu_k}}, \qquad b = \sqrt{\frac{\mu_j - \mu}{\mu - \mu_l}}.$$

As justified in section 4.1 the analysis can be restricted to componentwise nonnegative $x = (1, \alpha_0, \beta_0)^T$ so that its components α_0 and β_0 can be represented in terms of $\psi \in (0, \pi/2)$ and $t = \tan\psi$:

$$\alpha_0 = a\cos(\psi) = a\sqrt{\frac{1}{1+t^2}}, \qquad \beta_0 = b\sin(\psi) = b\sqrt{\frac{t^2}{1+t^2}}. \tag{4.13}$$

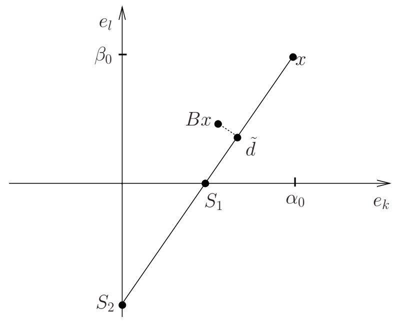

FIG. 4.4. *Geometry in the plane $\mathcal{E}_j$ and $Bx \notin \mathcal{E}_j$.*

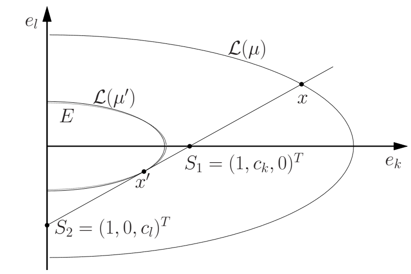

FIG. 4.5. *Ellipses in $\mathcal{E}_j$; E and $\mathcal{L}(\mu') \cap \mathcal{E}_j$ are almost identical.*

Two further ellipses in $\mathcal{E}_j$ are relevant for the subsequent analysis. These ellipses are very similar, each centered in e_j (the origin of $\mathcal{E}_j$) and each tangential to the line through S_1 and S_2. The first ellipse is $\mathcal{E}_j \cap \mathcal{L}(\mu')$ with $\mu' = \mu(x')$ and has the semi-axes

$$a' = \sqrt{\frac{\mu_j - \mu'}{\mu' - \mu_k}}, \qquad b' = \sqrt{\frac{\mu_j - \mu'}{\mu' - \mu_l}}.$$

This ellipse is tangential to the line through S_1 and S_2 since $\mu(x')$ is associated with the poorest convergence on the cone $\mathcal{F}_\gamma(x)$ projected to $\mathcal{E}_j$. Direct computation shows that $a'/b' < a/b$.

The second ellipse E (see Figure 4.5) has the semi-axes $\tilde{a}$ and $\tilde{b}$ so that the ratio of its semi-axes equals that of $\mathcal{E}_j \cap \mathcal{L}(\mu)$. This means that $\tilde{a}/\tilde{b} = a/b$. It holds that $\tilde{a} \geq a'$, since otherwise a contradiction can be derived. Assuming $\tilde{a} < a'$ for any point (α, β) on the ellipse E, it holds that (by using $a'/b' < a/b$)

$$\alpha^2 + \frac{a'^2}{b'^2}\beta^2 < \alpha^2 + \frac{a^2}{b^2}\beta^2 = \alpha^2 + \frac{\tilde{a}^2}{\tilde{b}^2}\beta^2 = \tilde{a}^2 < a'^2$$

so that $\alpha^2/a'^2 + \beta^2/b'^2 < 1$. The latter inequality means that the ellipse E is completely surrounded by the ellipse $\mathcal{L}(\mu') \cap \mathcal{E}_j$, which contradicts its tangentiality to the line through S_1 and S_2. Hence

$$\Delta(\mu') = \frac{\mu_j - \mu'}{\mu' - \mu_k} = a'^2 \leq \tilde{a}^2,$$

and an upper limit for $\tilde{a}^2/\Delta(\mu) = \tilde{a}^2/a^2$ remains to be determined. Next we show that (the case $c_l \to \infty$ is to be treated separately by analyzing the limits of c_k and c_l)

$$\frac{\tilde{a}^2}{a^2} = \frac{c_k^2 c_l^2}{b^2 c_k^2 + a^2 c_l^2}. \tag{4.14}$$

To prove this we determine the point of contact of the line through S_1 and S_2 and the ellipse E. The semi-axes of E are $\tilde{a}$ and $\tilde{b} = b\tilde{a}/a$. By a rescaling of the second semi-axis with the factor a/b the ellipse becomes a circle with the radius $\tilde{a}$, and the point of contact does not change. Further, the line segment connecting S_1 and S_2 is transformed into

$$s(\sigma) = \begin{pmatrix} 0 \\ \frac{a}{b}c_l \end{pmatrix} + \sigma \begin{pmatrix} c_k \\ -\frac{a}{b}c_l \end{pmatrix}, \qquad \sigma \in [0, 1].$$

The point of contact is that point on $s(\sigma)$ with the smallest Euclidean norm. From

$$\|s(\sigma)\|^2 = \sigma^2 c_k^2 + \left(\frac{a}{b} c_l\right)^2 (\sigma - 1)^2$$

direct computation shows that the minimum is attained in $\sigma^* = a^2 c_l^2/(b^2 c_k^2 + a^2 c_l^2)$. The resulting identity $\tilde{a}^2 = \|s(\sigma^*)\|^2$ yields (4.14).

Insertion of (4.11), (4.12), and (4.13) into (4.14) and using the variables $\Gamma := \sqrt{1-\gamma^2}/\gamma \in (0, \infty]$, $\Delta = a^2$, $b^2 = \Delta(1-\kappa)/(1+\kappa\Delta)$ with

$$\kappa = \frac{\mu_k - \mu_l}{\mu_j - \mu_l}$$

results in a representation of $\tilde{a}^2/a^2$ as a function of t, Δ, Γ, and κ. (The limit $\Gamma \to \infty$ needs additional care; however, this limit corresponds to $\gamma = 0$. For $\gamma = 0$ Theorem 2.2 is already proved in [18].) The details are as follows. With

$$\begin{aligned} A &= \sqrt{1+\alpha_0^2+\beta_0^2}\,(\mu_j - \mu) + \Gamma\alpha_0\beta_0(\mu_k - \mu_l), \\ B &= \sqrt{1+\alpha_0^2+\beta_0^2}\,\alpha_0(\mu - \mu_k) + \Gamma\beta_0(\mu_j - \mu_l), \\ C &= \sqrt{1+\alpha_0^2+\beta_0^2}\,\beta_0(\mu - \mu_l) + \Gamma\alpha_0(\mu_k - \mu_j), \end{aligned}$$

it holds that $c_k = A/B$ and $c_l = A/C$. Instead of considering $\tilde{a}^2/a^2$, it is more convenient to estimate its reciprocal from below. From (4.14) one gets

$$\frac{a^2}{\tilde{a}^2} = \frac{\Delta(1-\kappa)}{1+\kappa\Delta}\left(\frac{C}{A}\right)^2 + \Delta\left(\frac{B}{A}\right)^2$$

with

$$\frac{C}{A} = \frac{\sqrt{1+\alpha_0^2+\beta_0^2}\,\beta_0 + \Gamma\alpha_0\frac{\mu_k-\mu_j}{\mu-\mu_l}}{\sqrt{1+\alpha_0^2+\beta_0^2}\,b^2 + \Gamma\alpha_0\beta_0\frac{\mu_k-\mu_l}{\mu-\mu_l}}, \qquad \frac{B}{A} = \frac{\sqrt{1+\alpha_0^2+\beta_0^2}\,\alpha_0 + \Gamma\beta_0\frac{\mu_j-\mu_l}{\mu-\mu_k}}{\sqrt{1+\alpha_0^2+\beta_0^2}\,a^2 + \Gamma\alpha_0\beta_0\frac{\mu_k-\mu_l}{\mu-\mu_k}}.$$

In these formula the ratios of eigenvalue differences are to be expressed in terms of Δ and κ. Therefore let $U := \mu_j - \mu$, $V := \mu - \mu_k$, and $W := \mu - \mu_l$ so that $\mu_k - \mu_l = W - V$, $\mu_j - \mu_l = U + W$, and $\mu_k - \mu_j = -U - V$. Since $\Delta = U/V$ and $\Delta(1-\kappa)/(1+\kappa\Delta) = U/W$, we get that

$$\begin{aligned} \frac{\mu_k - \mu_j}{\mu - \mu_l} &= -\frac{U}{W}\left(1 + \frac{V}{U}\right) = \frac{(\kappa-1)(1+\Delta)}{1+\kappa\Delta}, \\ \frac{\mu_k - \mu_l}{\mu - \mu_l} &= 1 - \frac{V}{U}\frac{U}{W} = \frac{\kappa(1+\Delta)}{1+\kappa\Delta}, \\ \frac{\mu_j - \mu_l}{\mu - \mu_k} &= \frac{U+W}{V} = \frac{U}{V}\left(1 + \frac{W}{U}\right) = \frac{1+\Delta}{1-\kappa}, \\ \frac{\mu_k - \mu_l}{\mu - \mu_k} &= \frac{W-V}{V} = \frac{W}{U}\frac{U}{V} - 1 = \frac{\kappa(1+\Delta)}{1-\kappa}. \end{aligned}$$

Therefore we have

$$\begin{aligned} \frac{a^2}{\tilde{a}^2} =& \frac{\Delta(1-\kappa)}{1+\kappa\Delta}\left(\frac{\sqrt{1+\alpha_0^2+\beta_0^2}\,\beta_0 + \Gamma\alpha_0\frac{(\kappa-1)(1+\Delta)}{1+\kappa\Delta}}{\sqrt{1+\alpha_0^2+\beta_0^2}\,\frac{\Delta(1-\kappa)}{1+\kappa\Delta} + \Gamma\alpha_0\beta_0\frac{\kappa(1+\Delta)}{1+\kappa\Delta}}\right)^2 \\ &+ \Delta\left(\frac{\sqrt{1+\alpha_0^2+\beta_0^2}\,\alpha_0 + \Gamma\beta_0\frac{(1+\Delta)}{1-\kappa}}{\sqrt{1+\alpha_0^2+\beta_0^2}\,\Delta + \Gamma\alpha_0\beta_0\frac{\kappa(1+\Delta)}{1-\kappa}}\right)^2. \end{aligned}$$

Insertion of (4.13) yields $f := f(\Delta, t, \kappa, \Gamma)$ with

$$\begin{aligned} f = \frac{a^2}{\tilde{a}^2} = & \Big((1+\Delta)(\Gamma^2(1-\kappa)^2 + \kappa(1-\kappa) + \Gamma^2 t^2) + (1-\kappa)^2 + t^2(1-\kappa) \\ & + 2\kappa\Gamma t\sqrt{1/(1+t^2)}\sqrt{1+t^2+\kappa\Delta}\sqrt{1-\kappa}\sqrt{1+\Delta}\Big) \Big/ \\ & \Big(\sqrt{1-\kappa}\sqrt{1+t^2+\kappa\Delta} + \kappa\Gamma t\sqrt{1/(1+t^2)}\sqrt{1+\Delta}\Big)^2. \end{aligned}$$

This function is monotone increasing in Δ since $\partial f/\partial\Delta$ equals

$$\frac{\Gamma^2\sqrt{1-\kappa}\Big((1-\kappa)^3 + 3(1-\kappa)^2 t^2 + 3(1-\kappa)t^4 + t^6\Big)}{(1+t^2)\sqrt{1+t^2+\kappa\Delta}\Big(\sqrt{1-\kappa}\sqrt{1+t^2+\kappa\Delta} + \kappa\Gamma t\sqrt{1/(1+t^2)}\sqrt{1+\Delta}\Big)^3} > 0.$$

Therefore $f(0, t, \kappa, \Gamma)$ is a lower bound for $a^2/\tilde{a}^2$ which reads as

$$f(0,t,\kappa,\Gamma) = \frac{(1+t^2)\Big(\Gamma^2(1-\kappa)^2 + (1+t^2)(1-\kappa) + \Gamma^2 t^2 + 2\kappa\Gamma t\sqrt{1-\kappa}\Big)}{(\sqrt{1-\kappa}(1+t^2) + \kappa\Gamma t)^2}.$$

The parameter t determines the choice of x in the level set $\mathcal{L}(\mu)$. The derivative with respect to t reads as

$$\frac{\partial}{\partial t} f(0,t,\kappa,\Gamma) = \frac{2\kappa\Gamma^2(1-\kappa+t^2)\Big(\Gamma t^2 + 2t\sqrt{1-\kappa} - \Gamma(1-\kappa)\Big)}{\Big(\sqrt{1-\kappa}(1+t^2) + \kappa\Gamma t\Big)^3}.$$

The two real zeros of this derivative are

$$t_{1,2} = \frac{\sqrt{1-\kappa}(-1 \pm \sqrt{1+\Gamma^2})}{\Gamma}.$$

The global minimum is taken in

$$0 < t_1 = \frac{\sqrt{1-\kappa}(-1+\sqrt{1+\Gamma^2})}{\Gamma} = \frac{\sqrt{1-\kappa}(1-\gamma)}{\sqrt{1-\gamma^2}}.$$

Therefore the minimum is given by

$$f(0,t_1,\kappa,\Gamma) = \left(\frac{(2-\kappa)+\gamma\kappa}{\kappa+\gamma(2-\kappa)}\right)^2,$$

and its inverse yields the desired convergence estimate

$$\frac{\Delta(\mu')}{\Delta(\mu)} \le \left(\frac{\tilde{a}}{a}\right)^2 \le \left(\frac{\kappa+\gamma(2-\kappa)}{(2-\kappa)+\gamma\kappa}\right)^2.$$

This estimate is sharp since for $\Delta = 0$ the right inequality turns into an identity. Further, $\Delta = 0$ implies $\mu(x) \to \mu_j$ and also $\mu(x') \to \mu_j$ so that $\lim_{\mu(x)\to\mu_j} \tilde{a}/\tilde{b} - a'/b' = 0$ and in this limit $\mathcal{L}(\mu') \cap \mathcal{E}_j$ and E coincide; this implies that the left inequality also turns into an identity. $\square$

Proof of Theorem 2.2 *and Theorem* 1.2. Let $\mu = \mu(x) \in (\mu_{i+1}, \mu_i)$. Theorem 3.1 proves that the poorest convergence is attained in a three-dimensional invariant subspace. Theorem 4.3 proves in $\operatorname{span}\{e_j, e_k, e_l\}$ that

$$\frac{\Delta_{j,k}(\mu')}{\Delta_{j,k}(\mu)} \leq \left(\frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa}\right)^2.$$

It holds that either $\mu_l \leq \mu_{i+1} \leq \mu(x) < \mu_i \leq \mu_k < \mu_j$ or $\mu_l < \mu_k \leq \mu_{i+1} < \mu(x) < \mu_i \leq \mu_j$. In the first case the Ritz value $\mu(x')$ in $\operatorname{span}\{e_j, e_k, e_l\}$ satisfies that $\mu_k \leq \mu(x')$, which is the first alternative in Theorem 2.2. To analyze the second case, we get that the convergence factor is a monotone increasing function in $\kappa \in (0,1)$ since

$$\frac{\partial}{\partial\kappa}\frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa} = \frac{2(1-\gamma^2)}{(2-\kappa) + \gamma\kappa} \geq 0.$$

Further, $\kappa = (\mu_k - \mu_l)/(\mu_j - \mu_l)$ is a monotone decreasing function in μ_j and μ_l and a monotone increasing function in μ_k. Hence the poorest convergence with the maximal convergence factor is attained in $j = i$, $k = i+1$, and $l = n$, which proves Theorem 2.2:

$$\frac{\Delta_{i,i+1}(\mu')}{\Delta_{i,i+1}(\mu)} \leq \left(\frac{\kappa + \gamma(2-\kappa)}{(2-\kappa) + \gamma\kappa}\right)^2 \quad \text{with } \kappa = \frac{\mu_{i+1} - \mu_n}{\mu_i - \mu_n}.$$

Then Theorem 1.2 follows by inserting the reciprocals of the eigenvalues and Ritz values. ∎

Conclusions. The new convergence bound given in Theorem 1.2 completes the efforts to find sharp convergence estimates within the hierarchy of preconditioned PINVIT(k) and nonpreconditioned INVIT(k) eigensolvers for the index $k = 2$; a hierarchy of these solvers has been suggested in [15]. Next the results are summarized. All these convergence estimates have the common form

$$\Delta_{i,i+1}(\rho(x')) \leq \sigma^2 \Delta_{i,i+1}(\rho(x))$$

with $\Delta_{i,i+1}(\xi) = (\xi - \lambda_i)/(\lambda_{i+1} - \xi)$.

The convergence factor for the nonpreconditioned inverse iteration INVIT(1) procedure is (see [16])

$$\sigma(\text{INVIT}(1)) = \frac{\lambda_i}{\lambda_{i+1}}.$$

The associated preconditioned scheme, i.e., the preconditioned inverse iteration PINVIT(1) or preconditioned gradient iteration, has the convergence factor (see [8])

$$\sigma(\text{PINVIT}(1)) = \gamma + (1-\gamma)\frac{\lambda_i}{\lambda_{i+1}}.$$

Further, the convergence factor of the nonpreconditioned steepest descent iteration INVIT(2) reads as (see [18])

$$\sigma(\text{INVIT}(2)) = \frac{\kappa}{2-\kappa} \quad \text{with } \kappa = \frac{\lambda_i(\lambda_n - \lambda_{i+1})}{\lambda_{i+1}(\lambda_n - \lambda_i)}.$$

The new result on PINVIT(2), which is the PSD iteration, is now

$$\sigma(\text{PINVIT}(2)) = \frac{\kappa + \gamma(2-\kappa)}{(2-\kappa)+\gamma\kappa} \quad \text{with } \kappa = \frac{\lambda_i(\lambda_n - \lambda_{i+1})}{\lambda_{i+1}(\lambda_n - \lambda_i)}.$$

All these convergence factors are sharp.

Further progress in deriving convergence estimates for the hierarchy of nonpreconditioned and preconditioned iteration is a matter for future work. Especially for the practically important LOPCG iteration [7], sharp convergence estimates are highly desired.

Appendix A. Preconditioning for the eigenvalue problem. Preconditioning for linear systems $Au = f$ can be introduced by a left-multiplication with a preconditioner $T \approx A^{-1}$ so that $(TA)u = (Tf)$. If the spectral condition number is decreased, i.e., $\kappa(TA) < \kappa(A)$, then preconditioning accelerates the convergence of an iterative solver.

The left-multiplication with T does not work for the eigenvalue problem in the same way. From (1.1) we get with a regular matrix T that

$$(TA)x_i = \lambda_i(TB)x_i.$$

The eigenvalues of (A, B) and (TA, TB) coincide. Similarly, the application of the linear transformation $y = T^{-1/2}x$ to (1.2) results in

$$y' = y - \vartheta_{\text{opt}}(\tilde{A}y - \tilde{\rho}(y)\tilde{B}y), \quad \tilde{\rho}(y) = \frac{(y, \tilde{A}y)}{(y, \tilde{B}y)} \tag{A.1}$$

with $\tilde{A} = T^{1/2}AT^{1/2} \sim TA$ and $\tilde{B} = T^{1/2}BT^{1/2} \sim TB$. Thus (A.1) has the form of the steepest descent iteration for $(\tilde{A}, \tilde{B})$ as treated in [18]. However, this transformation does not simplify the convergence analysis of the PSD iteration because of the key role of the level set $\mathcal{L}(\rho_0) = \{x \in \mathbb{R}^n;\ \rho(x) = \rho_0\}$ in the analysis of the preconditioned iteration (see section 2.2) and also in the analysis of the steepest descent iteration in [18]. The basis transformation by $T^{-1/2}$ maps the level set $\mathcal{L}(\rho_0)$ to $T^{-1/2}\mathcal{L}(\rho_0)$, which is not a level set. Hence the proof techniques from [18] cannot be applied. Further, we have to analyze the convergence of (A.1) not only for a single preconditioner T but for all preconditioners satisfying (1.6). In this paper all these admissible preconditioners are taken into account by the set of possible iterates $\mathcal{F}_\gamma(x)$ given in (2.8). The transformation $y = T^{-1/2}x$ applied to $\mathcal{F}_\gamma(x)$ destroys its circular cone geometry. To summarize, the simple transformation which maps (1.2) to (A.1) appears to be promising, but it is not clear how this basis transformation can successfully be used to simplify the convergence analysis of preconditioned gradient-type iterations for the eigenvalue problem.

Acknowledgment. The author is very grateful to Ming Zhou, University of Rostock, for his help with the introduction of the ellipse E in section 4.3, which was a valuable input in finalizing the convergence proof.

REFERENCES

[1] R. Argentati, A. Knyazev, K. Neymeyr, and E. Ovtchinnikov, *Preconditioned Eigensolver Convergence Theory in a Nutshell*, Technical report, in preparation.

[2] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.

[3] J. Bramble, J. Pasciak, and A. Knyazev, *A subspace preconditioning algorithm for eigenvector/eigenvalue computation*, Adv. Comput. Math., 6 (1996), pp. 159–189.

[4] M. Embree and R. Lehoucq, *Dynamical systems and non-Hermitian iterative eigensolvers*, SIAM J. Numer. Anal., 47 (2009), pp. 1445–1473.

[5] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1976.

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

[7] A. Knyazev, *Preconditioned eigensolvers—an oxymoron?*, Electron. Trans. Numer. Anal., 7 (1998), pp. 104–123.

[8] A. Knyazev and K. Neymeyr, *A geometric theory for preconditioned inverse iteration.* III: *A short and sharp convergence estimate for generalized eigenvalue problems*, Linear Algebra Appl., 358 (2003), pp. 95–114.

[9] A. Knyazev and K. Neymeyr, *Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method*, Electron. Trans. Numer. Anal., 15 (2003), pp. 38–55.

[10] A. V. Knyazev and K. Neymeyr, *Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 621–628.

[11] A. Knyazev and A. Skorokhodov, *On exact estimates of the convergence rate of the steepest ascent method in the symmetric eigenvalue problem*, Linear Algebra Appl., 154/156 (1991), pp. 245–257.

[12] A. V. Knyazev, *Modified gradient methods for spectral problems*, Differ. Uravn., 23 (1987), pp. 715–717 (in Russian).

[13] K. Neymeyr, *A geometric theory for preconditioned inverse iteration.* I: *Extrema of the Rayleigh quotient*, Linear Algebra Appl., 322 (2001), pp. 61–85.

[14] K. Neymeyr, *A geometric theory for preconditioned inverse iteration.* II: *Convergence estimates*, Linear Algebra Appl., 322 (2001), pp. 87–104.

[15] K. Neymeyr, *A Hierarchy of Preconditioned Eigensolvers for Elliptic Differential Operators*, Habilitationsschrift an der Mathematischen Fakultät, Universität Tübingen, Tübingen, Germany, 2001.

[16] K. Neymeyr, *A note on inverse iteration*, Numer. Linear Algebra Appl., 12 (2005), pp. 1–8.

[17] K. Neymeyr, *On preconditioned eigensolvers and invert-Lanczos processes*, Linear Algebra Appl., 430 (2009), pp. 1039–1056.

[18] K. Neymeyr, E. Ovtchinnikov, and M. Zhou, *Convergence analysis of gradient iterations for the symmetric eigenvalue problem*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 443–456.

[19] J. Nocedal and S. Wright, *Numerical optimization*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.

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

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