

CONVERGENCE THEORY FOR PRECONDITIONED EIGENVALUE SOLVERS IN A NUTSHELL *

MERICO E. ARGENTATI[†], ANDREW V. KNYAZEVA[‡], KLAUS NEYMEYR[§],
EVGUENI E. OVTCHINNIKOV[¶], AND MING ZHOU[§]

Abstract. Preconditioned iterative methods for numerical solution of large matrix eigenvalue problems are increasingly gaining importance in various application areas, ranging from material sciences to data mining. Some of them, e.g., those using multilevel preconditioning for elliptic differential operators or graph Laplacian eigenvalue problems, exhibit almost optimal complexity in practice, i.e., their computational costs to calculate a fixed number of eigenvalues and eigenvectors grow linearly with the matrix problem size. Theoretical justification of their optimality requires convergence rate bounds that do not deteriorate with the increase of the problem size. Such bounds were pioneered by E. D'yakonov over three decades ago, but to date only a handful have been derived, mostly for symmetric eigenvalue problems. Just a few of known bounds are sharp. One of them is proved in [doi:10.1016/S0024-3795(01)00461-X] for the simplest preconditioned eigensolver with a fixed step size. The original proof has been greatly simplified and shortened in [doi:10.1137/080727567] by using a gradient flow integration approach. In the present work, we give an even more succinct proof, using novel ideas based on Karush-Kuhn-Tucker theory and nonlinear programming.

Key words. symmetric; preconditioner; eigenvalue; eigenvector; Rayleigh quotient; gradient; iterative method; Karush-Kuhn-Tucker theory.

AMS subject classifications. 65F15 65K10 65N25

Dedicated to the memory of Evgenii G. D'yakonov, Moscow, Russia, 1935–2006.

1. Introduction. Preconditioning is a technique developed originally for the iterative solution of linear systems that aims at the acceleration of convergence of the iterations. In its simplest form, the system $Ax = b$ is multiplied by a matrix T such that the spectral condition number of TA , the ratio of the largest to the smallest singular value thereof, is considerably smaller than that of A , which generally leads to faster convergence.

Iterative methods for solving linear systems normally do not require A and T to be explicitly formed as matrices: it is sufficient that matrix-vector multiplications are implemented and performed via user-defined procedures. The same is true for iterative methods that compute eigenvalues and eigenvectors of a very large matrix, as, e.g., in [24], calculating one eigenvector of a 100-billion size matrix, or in [4].

A classical application area for preconditioned solvers is the discretized boundary value problems for elliptic partial differential operators; see, e.g., [5]. With multigrid preconditioning, preconditioned solvers may achieve linear complexity on problems from this area; see, e.g., [12] and references there for symmetric eigenvalue problems. D'yakonov seminal work, summarized in [5], proposes “spectrally equivalent”

* Communicated by Nicholas Higham. A preliminary version is posted at <http://arxiv.org>.

[†] Department of Mathematical and Statistical Sciences, University Colorado Denver, P.O. Box 173364, Campus Box 170, Denver, CO 80217-3364. (merico.argentati at ucdenver.edu).

[‡] Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139-1955. (andrew.knyazev at merl.com).

[§] Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18055 Rostock, Germany. (klaus.neymeyr at mathematik.uni-rostock.de and ming.zhou at mathematik.uni-rostock.de).

[¶] Numerical Analysis Group, Building R18, STFC Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxfordshire, OX11 0QX, UK. (evgueni.ovtchinnikov at stfc.ac.uk).

preconditioning for elliptic operator eigenvalue problems in order to guarantee convergence that does not deteriorate with the increasing dimension of the discretized problem. Owing to this, for large enough problems such preconditioners outperform direct solvers, which factorize the original sparse matrix A . Inevitable matrix fill-ins, especially prominent in discretized differential problems in more than two spatial dimensions, destroy the matrix sparsity, resulting in computer memory overuse and non-optimal performance.

Preconditioning has also long since been a key technique in *ab initio* calculations in material sciences; see, e.g., [2] and references therein. In the last decade, preconditioning for graphs is attracting growing attention as a tool for achieving an optimal complexity for large data mining problems, e.g., for graph bisection and image segmentation using graph Laplacian and Fiedler vectors since [10]; for recent work see, e.g., [22].

Preconditioned iterative methods for the original linear system $Ax - b = 0$ are in many cases mathematically equivalent to standard iterative methods applied for the preconditioned system $T(Ax - b) = 0$. For example, the classical Richardson iteration step applied to the preconditioned system becomes

$$(1.1) \quad x_{n+1} = x_n - \tau_n T(Ax_n - b),$$

where τ_n is a suitably chosen scalar.

Turning now to eigenvalue problems, let us consider the computation of an eigenvector of a real symmetric positive definite matrix A corresponding to its smallest eigenvalue. Borrowing an argument from [9], suppose that the targeted eigenvalue λ_* , or a sufficiently good approximation thereof, is known. Then the corresponding eigenvector can be computed by solving a homogeneous linear system $(A - \lambda_* I)x = 0$, or, equivalently, the system $T(A - \lambda_* I)x = 0$, where I is an identity. The Richardson iteration step now becomes

$$(1.2) \quad x_{n+1} = x_n - \tau_n T(A - \lambda_* I)x_n.$$

Theoretically, the best preconditioners for $Ax - b = 0$ and $(A - \lambda_* I)x = 0$ are, correspondingly, $T = A^{-1}$ and $T = (A - \lambda_* I)^\dagger$, where \dagger denotes a pseudo-inverse, making both Richardson iteration schemes, (1.1) and (1.2), converge in a single step with $\tau_n = 1$. Under the standard assumption $T \approx A^{-1}$, both in (1.1) and (1.2), convergence theory is straightforward, e.g., in terms of the spectral radius $\rho(I - TA)$ of $I - TA$. Sharp explicit convergence bounds, not relying on generic constants, can be derived in the form of inequalities that allow one to determine whether the convergence deteriorates with the increasing problem size by analyzing every term in the bound.

For some classes of eigenvalue problems, the efficiency of choosing $T \approx A^{-1}$ has been demonstrated, both numerically and theoretically, in [1, 11]. This choice allows the easy adaptation of a vast variety of preconditioners already developed for linear systems to the eigensolvers.

In practice, the theoretical value λ_* in the Richardson iteration above has to be replaced with its approximation. A standard choice for λ_* is a Rayleigh quotient function $\lambda(x) = x^T Ax / x^T x$, leading to

$$(1.3) \quad x_{n+1} = x_n - \tau_n T(A - \lambda(x_n)I)x_n.$$

It is well known that the Rayleigh quotient $\lambda(x_n)$ gives a high quality (quadratic) approximation of the eigenvalue λ_* , if the sequence x_n converges to the corresponding

eigenvector. Thus, asymptotically as $\lambda(x_n) \rightarrow \lambda_*$, where $n \rightarrow \infty$, methods (1.2) and (1.3) are equivalent, and so may be their asymptotic convergence rate bounds. However, asymptotic convergence rate bounds naturally contain generic constants, which are independent of $n \rightarrow \infty$, but may depend on the problem size.

Due to the changing value $\lambda(x_n)$, a non-asymptotic theoretical convergence analysis is much more difficult, compared to the case for linear systems, even for the simplest methods, such as the Richardson iteration (1.3). D'yakonov pioneering work from the eighties, summarized in [5, Chapter 9], includes the first non-asymptotic convergence bounds for preconditioned eigensolvers proving their linear convergence with a rate, which can be bounded above independently of the problem size.

Just a few of the known bounds are sharp. One of them is proved for the simplest preconditioned eigensolver with a fixed step size (1.3) in a series of papers by Neymeyr over a decade ago; see [11] and references therein. The original proof has been greatly simplified and shortened in [13] by using a gradient flow integration approach.

In this paper we present a new self-contained proof of a sharp convergence rate bound from [11] for the preconditioned eigensolver (1.3), Theorem 2.1. Following the geometrical approach of [11, 13], we reformulate the problem of finding the convergence bound for (1.3) as a constrained optimization problem for the Rayleigh quotient. The main novelty of the proof is that here we use inequality constraints, which brings to the scene the Karush-Kuhn-Tucker (KKT) theory; see, e.g. [6, 18]. KKT conditions allow us to reduce our convergence analysis to the simplest scenario in two dimensions, which is the key step in the proof. We have also found several simplifications in the two dimensional convergence analysis, compared to that of [11, 13]. We believe that the new proof will greatly enhance the understanding of the convergence behavior of increasingly popular preconditioned eigensolvers, whose application area is quickly expanding; see, e.g., [14, 15, 16, 17, 20, 21, 22].

2. Convergence rate bound. We consider a real generalized eigenvalue problem $Ax = \lambda Bx$ with real symmetric positive definite matrices A and B . The objective is to approximate iteratively the smallest eigenvalue λ_1 by minimizing the Rayleigh quotient $\lambda(x) = x^T Ax / x^T Bx$. A direct formulation of the convergence analysis with respect to this form of the eigenvalue problem has some disadvantages. Instead, the inverted form $Bx = \mu Ax$ with $\mu = 1/\lambda$ results in more compact representation of the problem and the proof (many inverses like A^{-1} and $1/\lambda$ can be avoided), cf. [11, 13]. For this inverted form the objective is to approximate the largest eigenvalue μ_1 of $Bx = \mu Ax$ by maximizing the Rayleigh quotient $\mu(x) = x^T Bx / x^T Ax$. We denote the eigenvalues by $\mu_1 > \dots > \mu_m > 0$, which can have arbitrary multiplicity. Corresponding eigenspaces are denoted by $\mathcal{V}_1, \dots, \mathcal{V}_m$.

The increase of $\mu(x)$ can be achieved by correcting the current iterate x along the preconditioned gradient of the Rayleigh quotient, i.e.

$$(2.1) \quad x' = x + \frac{1}{\mu(x)} T(Bx - \mu(x)Ax);$$

see [11, 19, 13] and references therein. If $B = I$, then $\mu(x) = 1/\lambda(x)$ and method (2.1) turns into (1.3) with $\tau_n = 1$, discussed in the Introduction.

In all our prior work on preconditioned eigensolvers for symmetric eigenvalue problems, including [11, 13], we have always assumed that the preconditioner T is a symmetric and positive definite matrix, typically satisfying conditions

$$(2.2) \quad (1 - \gamma)z^T T^{-1}z \leq z^T Az \leq (1 + \gamma)z^T T^{-1}z, \quad \forall z, \text{ for a given } \gamma \in [0, 1),$$

or equivalent, up to the scaling of T . Recently, the authors of [3] have noticed and demonstrated that T does not have to be symmetric positive definite, and a less restrictive assumption

$$(2.3) \quad s_{\max} \left(I - A^{1/2} T A^{1/2} \right) \leq \gamma < 1,$$

can be used instead, where s_{\max} denotes the matrix largest singular value, and $A^{1/2}$ is the symmetric positive definite square root of A . It is verified in [3] that (2.2) and (2.3) are equivalent if T is symmetric and positive definite.

In what follows, we give a complete and concise proof of the following convergence rate bound, first proved in [11, 13],

THEOREM 2.1. *If $\mu_{i+1} < \mu(x) < \mu_i$ and T satisfies (2.3), then for x' given by (2.1) it holds that either $\mu(x') \geq \mu_i$ or*

$$(2.4) \quad 0 < \frac{\mu_i - \mu(x')}{\mu(x') - \mu_{i+1}} \leq \sigma^2 \frac{\mu_i - \mu(x)}{\mu(x) - \mu_{i+1}}, \quad \sigma = \gamma + (1 - \gamma) \frac{\mu_{i+1}}{\mu_i}.$$

The first step, Lemma 2.2, of the proof of Theorem 2.1 is the same as that in [11, 13], where we characterize a set of possible next step iterates x' in (2.1) varying the preconditioner T constrained by assumption (2.3), aiming at eliminating the preconditioner T from consideration. The only difference is that in [11, 13] we start with changing an original coordinate basis to an A -orthogonal basis, which transforms A into the identity I , resulting in a one-line proof of Lemma 2.2. Here, we choose to present a detailed proof of Lemma 2.2, for a general A , demonstrating that the transformation of A into the identity I , made after Lemma 2.2, is well justified.

LEMMA 2.2. *Let us denote $\kappa = \mu(x)$ and*

$$x'_A = A^{1/2} x', \quad x_A = A^{1/2} x, \quad B_A = A^{-1/2} B A^{-1/2}, \quad T_A = A^{1/2} T A^{1/2}, \quad r_A = B_A x_A - \kappa x_A,$$

and define a closed ball

$$\mathcal{B}_A = \{y : (B_A x_A - y)^T (B_A x_A - y) \leq \gamma^2 (r_A)^T r_A\}$$

centered at $B_A x_A$. Let T satisfy (2.3), then for x' given by (2.1) it holds that

$$\kappa x'_A = B_A x_A - (I - T_A)(B_A x_A - \kappa x_A) \in \mathcal{B}_A.$$

Proof. Left-multiplying (2.1) by $\mu(x) A^{1/2}$ gives

$$\mu(x) A^{1/2} x' = \mu(x) A^{1/2} x + A^{1/2} T A^{1/2} A^{-1/2} B A^{-1/2} A^{1/2} x - \mu(x) A^{1/2} T A^{1/2} A^{1/2} x,$$

or, in our new notation,

$$\kappa x'_A = \kappa x_A + T_A B_A x_A - \kappa T_A x_A = B_A x_A - (I - T_A)(B_A x_A - \kappa x_A),$$

resulting $B_A x_A - \kappa x'_A = (I - T_A) r_A$. Since $s_{\max}(I - T_A) \leq \gamma$ by (2.3), we get $(B_A x_A - \kappa x'_A)^T (B_A x_A - \kappa x'_A) = ((I - T_A) r_A)^T ((I - T_A) r_A) \leq \gamma^2 (r_A)^T r_A$. \square

The second step of the proof is traditional—reducing the generalized symmetric eigenvalue problem $Bx = \mu Ax$ to the standard eigenvalue problem for the symmetric

positive definite matrix $B_A = A^{-1/2}BA^{-1/2}$ by making the change of variables as hinted by Lemma 2.2. We use the standard inner product in \cdot_A variables, i.e.

$$(y_A, z_A) = y_A^T z_A = \left(A^{1/2}y\right)^T \left(A^{1/2}z\right) = y^T Az,$$

and the corresponding vector norm $\|y_A\| = (y_A, y_A)^{1/2}$, so, e.g.,

$$\kappa = \mu(x) = \frac{x^T Bx}{x^T Ax} = \frac{x_A^T B_A x_A}{x_A^T x_A} = \frac{(x_A, B_A x_A)}{(x_A, x_A)} = \mu_A(x_A).$$

We later use B_A and B_A^{-1} -based scalar products and norms defined as follows, e.g.,

$$(y_A, z_A)_{B_A} = y_A^T B_A z_A = \left(A^{1/2}y\right)^T A^{-1/2}BA^{-1/2} \left(A^{1/2}z\right) = y^T Bz.$$

For brevity we drop the subscript \cdot_A in the rest of the paper. In the following T refers to $T_A = A^{1/2}TA^{1/2}$, B refers to $B_A = A^{-1/2}BA^{-1/2}$, x refers to $x_A = A^{1/2}x$, and so on, cf. Lemma 2.2. Furthermore, $\mu(x) = (x, Bx)/(x, x)$, and method (2.1) is $\mu(x)x' = Bx - (I - T)(Bx - \mu(x)x)$. The new form of condition (2.3) is $\|I - T\| \leq \gamma$. This means that $A^{1/2}TA^{1/2}$ approximates the identity matrix with respect to the notation used in Lemma 2.2. The closed ball has the form $\mathcal{B} = \{y : \|Bx - y\| \leq \gamma\|r\|\}$ with the radius $\gamma\|r\|$ centered at Bx . Since $\mu(x)x' \in \mathcal{B}$ and $\mu(x') = \mu(\mu(x)x')$, we can estimate $\mu(x')$ by using a minimizer of $\mu(\cdot)$ in \mathcal{B} (i.e. by considering the worst case). We observe that, effectively, we set $A = I$ without loss of generality.

3. The special case with $\gamma = 0$: the power method. The main idea of the geometrical approach of [11, 13], which we also employ in this paper, is that the convergence rate of iterations (2.1) is slowest, in terms of the Rayleigh quotient, if x is a linear combination of two eigenvectors, which makes the further convergence analysis trivial. A new proof of this fact actually occupies a major part of our paper. In order to illustrate how such a dramatic reduction in dimension becomes possible, in this section we apply our technique to a simplified case $T = I$ corresponding to $\gamma = 0$. It is not difficult to see that under this assumption (2.1) turns into one iteration $\mu(x)x' = Bx$ of the power method, and bound (2.4) holds with $\gamma = 0$ and thus $\sigma = \mu_{i+1}/\mu_i$. Let us make a historic note that exactly this result has apparently first appeared in [7, 8].

The left-hand side of bound (2.4) is monotone in $\mu(x') = \mu(Bx)$. One way to find out at which x the behavior of (2.1) is the worst is to minimize $f(x) = \mu(Bx)$ for all x that satisfy $\mu(x) = \kappa$ for some fixed $\kappa \in (\mu_{i+1}, \mu_i)$. Slightly abusing the notation in the proof, we keep denoting by x both the initial approximation in (2.1) and the vector in the minimization problem.

We notice that $\mu(x) = \kappa$ is equivalent to $h(x) = \kappa(x, x) - (x, Bx) = 0$. Therefore, at a stationary point we have, using Lagrangian multipliers, that

$$(3.1) \quad \nabla f(x) + a\nabla h(x) = 0,$$

where a is some constant. This yields

$$\frac{2B(B - \mu(Bx)I)Bx}{\|Bx\|^2} + 2a(\kappa x - Bx) = 0,$$

which can be rewritten as

$$(3.2) \quad B^3x - \mu(Bx)B^2x - cBx + c\kappa x = 0,$$

where $c = \|Bx\|^2 a$. Since

$$\mu(Bx) = \frac{(Bx, B(Bx))}{(Bx, Bx)}$$

implies $(B^3x - \mu(Bx)B^2x, x) = 0$, we obtain

$$\begin{aligned} c\|Bx - \kappa x\|^2 &= (B^3x - \mu(Bx)B^2x, Bx - \kappa x) \\ &= (B^3x - \mu(Bx)B^2x, Bx - \mu(Bx)x) = \|B^2x - \mu(Bx)Bx\|^2, \end{aligned}$$

which shows that $c > 0$. Thus, equation (3.2) can be viewed as a polynomial equation $p_3(B)x = 0$, where $p_3(t)$ is a third degree polynomial with positive first and last coefficients, specifically 1 and $c\kappa$, correspondingly.

Inserting $x = \sum_{i=1}^m v_i$, where v_i are the projections of x onto the eigenspaces \mathcal{V}_i , leads to $\sum_{i=1}^m p_3(\mu_i)v_i = 0$. Since the eigenspaces are orthogonal to each other, the products $p_3(\mu_i)v_i$ must be zero for each i . Owing to the positiveness of the first and last coefficients, the polynomial p_3 must have a non-positive root, and thus at most two positive roots, i.e. $p_3(\mu_i)$ can be zero for some two indexes $i = k$ and $i = l$ at most, allowing the only possibly nonzero v_k and v_l from all projections v_i . We conclude that x is a linear combination of at most two normalized eigenvectors x_k and x_l , corresponding to distinct eigenvalues μ_k and μ_l of the matrix B .

We assume without loss of generality that $x = x_k + \alpha x_l$, then

$$\alpha^2 = \frac{\mu_k - \mu(x)}{\mu(x) - \mu_l} = \tan^2 \angle(x, x_k).$$

Similarly, since $Bx = \mu_k x_k + \alpha \mu_l x_l$, we obtain

$$(3.3) \quad \tan^2 \angle(Bx, x_k) = \frac{\mu_k - \mu(Bx)}{\mu(Bx) - \mu_l} = \frac{\mu_l^2}{\mu_k^2} \alpha^2 = \sigma^2 \frac{\mu_k - \mu(x)}{\mu(x) - \mu_l}, \quad \text{with } \sigma = \frac{\mu_l}{\mu_k}.$$

Let $\mu_k > \mu_l$, then $\kappa \in (\mu_l, \mu_k)$ implies $\mu_l \leq \mu_{i+1} < \mu(x) = \kappa < \mu_i \leq \mu_k$. By using monotonicity of the ratio of the quotients in μ_k and μ_l and the fact that the vector x here corresponds to the worst-case scenario, i.e. minimizing $\mu(x') = \mu(Bx)$ over all x with the fixed value $\mu(x) = \kappa$, we obtain (2.4) with $\gamma = 0$. Since (3.3) is an equality, we also prove that the upper bound in (2.4) with $\gamma = 0$ is sharp, turning into an equality if the initial approximation in (2.1) satisfies $x \in \text{span}\{x_i, x_{i+1}\}$.

In the next section, we apply the described dimensionality reduction technique to the general case. We formulate the conditions that “the worst case” x must satisfy, which yield the generalization of equation (3.1), and rewrite this equation as a cubic equation $p_3(B)x = 0$. We show that the first and last coefficients of this equation are positive, which, as we have just seen, implies that x is a linear combination of two eigenvectors. A simple two-dimensional analysis completes the proof of Theorem 2.1.

4. The general case with $\gamma \in [0, 1)$: the preconditioned eigensolver (2.1).

Next the proof of Theorem 2.1 is given: Let us denote $r = Bx - \mu(x)x$ and define $\mathcal{B} = \{y : \|Bx - y\| \leq \gamma \|r\|\}$, a closed ball with the radius $\gamma \|r\|$ centered at Bx . On the one hand, it holds that $\mu(y) > \mu(x)$ for any vector $y \in \mathcal{B}$ since x is not an eigenvector and $r \neq 0$. Indeed, taking into account $\|Bx - y\|^2 < \|r\|^2 = \|Bx - \mu(x)x\|^2$, we have

$$\begin{aligned} \|y\|^2 &< 2(x, y)_B - 2\mu(x)\|x\|_B^2 + \mu(x)^2\|x\|^2 = 2(x, y)_B - \mu(x)\|x\|_B^2 \\ &= (\|y\|_B^2 - \|y - \mu(x)x\|_B^2)/\mu(x) \leq \|y\|_B^2/\mu(x) \end{aligned}$$

so that $\mu(x) < \|y\|_B^2 / \|y\|^2 = \mu(y)$. On the other hand, $\mu(x)x' = Bx - (I - T)r \in \mathcal{B}$, since $\|(I - T)r\| \leq \gamma\|r\|$, see Lemma 2.2. This proves $\mu(x') > \mu(x) > \mu_{i+1}$ and, thus, the left inequality in (2.4), provided that $\mu(x') < \mu_i$.

In the previous proof with $\gamma = 0$, the ball \mathcal{B} shrinks to a single point Bx and the only choice of $y = Bx$ is possible. The present case $\gamma > 0$ is significantly more difficult for the worst-case scenario analysis, involving a minimization problem with two variables, x and y . In our previous work, see [11, 13] and references therein, we first vary $y \in \mathcal{B}$ intending to minimize $\mu(y)$ for a given x , and then vary x fixing $\mu(x) = \kappa$. The first minimization problem defines y as an implicit function of x , and then Lagrangian multipliers are used, as in Section 3, to analyze the second minimization problem, in x . It turns out that the proof is much simpler if we vary both x and y at the same time and attack the required two-parameter minimization problem in x and y directly by using the KKT arguments as provided below.

LEMMA 4.1. *For $\gamma \in [0, 1)$ and a fixed value κ that is not an eigenvalue of B , let a pair of vectors $\{x^*, y^*\}$ denote a solution of the following constrained minimization problem:*

$$\text{minimize } \mu(y) \quad \text{subject to } \|Bx - y\| \leq \gamma\|Bx - \kappa x\| \quad \text{and } \mu(x) = \kappa.$$

If x^ is not an eigenvector of B , then both x^* and y^* belong to a two-dimensional invariant subspace of B corresponding to two distinct eigenvalues, and*

$$(4.1) \quad \sin \angle(Bx^*, y^*) = \gamma \sin \angle(Bx^*, x^*),$$

where $\angle(\cdot, \cdot)$ denotes an angle between two vectors defined by $\angle(u, v) := \arccos \left(\frac{(u, v)}{\|u\|\|v\|} \right)$.

Proof. We consider the equivalent problem

$$\begin{aligned} &\text{minimize } f(x, y) = \mu(y), \quad x \neq 0, \\ &\text{subject to } g(x, y) = \|Bx - y\|^2 - \gamma^2\|Bx - \kappa x\|^2 \leq 0, \quad \text{and} \\ &\quad h(x, y) = \kappa(x, x) - (x, Bx) = 0. \end{aligned}$$

We first notice that the assumption $x \neq 0$ implies $y \neq 0$ because of the first constraint and $\gamma < 1$. Thus, $\mu(y)$ is correctly defined. Next, let us temporarily consider a stricter constraint $\|x\| = 1$, instead of $x \neq 0$. Combined with the other constraints, this results in minimization of the smooth function $f(x, y)$ on a compact set, so there exists a solution $\{x^*, y^*\}$. Finally, let us remove the artificial constraint $\|x\| = 1$ and notice that any nonzero multiple of $\{x^*, y^*\}$ is also a solution. Thus we can consider the Karush-Kuhn-Tucker (KKT) conditions, e.g., [6, Theorem 9.1.1], [6, 18], in any neighborhood of $\{x^*, y^*\}$, which does not include the origin.

Next we show that the gradients of g and h are linearly independent. For the gradient of h , it holds that $\partial h / \partial x = -2r$ with $r \neq 0$, since x is not an eigenvector of B , and it holds that $\partial h / \partial y = 0$, since h does not depend on y . Assuming the linear dependence of the gradients of g and h implies that $\partial g / \partial y = 0$, so that $2(y - Bx) = 0$ and $y = Bx$. By using $y = Bx$, it holds that

$$\frac{\partial g}{\partial x} = 2(B^2x - By - \gamma^2(B - \kappa I)r) = -2\gamma^2(B - \kappa I)r,$$

while $\partial h / \partial x = -2r$, i.e. (using again the assumed linear dependence) the vector $r = Bx - \kappa x$ is an eigenvector of $B - \kappa I$, and, hence x is an eigenvector of B , contradicting the lemma assumption.

Therefore, the gradients of g and h are linearly independent. All functions involved in our constrained minimization are smooth. We conclude that the stationary point $\{x^*, y^*\}$ is regular, i.e., the KKT conditions are valid. The KKT stationarity condition states that there exist constants a and b such that

$$\nabla f(x^*, y^*) + a\nabla g(x^*, y^*) + b\nabla h(x^*, y^*) = 0$$

at the critical point $\{x^*, y^*\}$. The independent variables $\{x, y\}$ no longer appear, so to simplify the notation, in the rest of the proof we drop the superscript $*$ and substitute $\{x, y\}$ for $\{x^*, y^*\}$. We separately write the partial derivatives with respect to x ,

$$(4.2) \quad 2a(B^2x - By - \gamma^2(B - \kappa I)r) - 2br = 0,$$

and with respect to y ,

$$(4.3) \quad 2\frac{By - \mu(y)y}{(y, y)} + 2a(y - Bx) = 0.$$

The KKT complementary slackness condition $ag(x, y) = 0$ must be satisfied, implying

$$(4.4) \quad \|Bx - y\| = \gamma\|r\| \text{ if } a \neq 0.$$

If y is an eigenvector then $By - \mu(y)y = 0$ in condition (4.3), leading to $y = Bx$, i.e. vector x is also an eigenvector of B , thus we are done. Now we consider a nontrivial case, where neither x nor y is an eigenvector. Condition (4.3) then implies $a \neq 0$, so identity (4.4) holds unconditionally, condition (4.2) turns into

$$(4.5) \quad B(Bx - \gamma^2r - y) = cr \quad \text{with } c = \frac{b}{a} - \gamma^2\kappa,$$

and taking the inner product of (4.3) with y gives

$$(4.6) \quad (Bx - y, y) = 0.$$

Taking the inner products of both sides of (4.5) with $B^{-1}r$ results in

$$c\|r\|_{B^{-1}}^2 = (Bx - y, r) - \gamma^2\|r\|^2 = (Bx - y, r) - \|Bx - y\|^2 = -\kappa(Bx - y, x).$$

Therein we use (4.4) and (4.6).

Denoting $d = a\|y\|^2 - \mu(y)$, we rewrite (4.3) as

$$(4.7) \quad By - \mu(y)Bx = d(Bx - y).$$

Taking the inner products of both sides of (4.7) with $y - \mu(y)x$ yields

$$0 \leq \|y - \mu(y)x\|_B^2 = d(Bx - y, y - \mu(y)x) = -d\mu(y)(Bx - y, x),$$

where the orthogonality $(Bx - y, y) = 0$ has been used again. Therefore, we obtain $cd\|r\|_{B^{-1}}^2 = -d\kappa(Bx - y, x) \geq 0$, which implies $cd \geq 0$.

Substituting $r = Bx - \kappa x$ and multiplying through by B in (4.5) results in

$$(1 - \gamma^2)B^3x + (\kappa\gamma^2 - c)B^2x - B^2y + c\kappa Bx = 0.$$

Multiplying through by $d + \mu(y)$ and substituting $(d + \mu(y))Bx = (B + dI)y$, which follows from (4.7), we obtain $p_3(B)y = (c_3B^3 + c_2B^2 + c_1B + c_0)y = 0$, where $p_3(\cdot)$

is a third degree polynomial with $c_3 = 1 - \gamma^2 > 0$ and $c_0 = cd\kappa \geq 0$, which cannot have more than two positive roots. Thus, y is a linear combination of two normalized eigenvectors x_k and x_l corresponding to two distinct eigenvalues (cf. Section 3), i.e. $y \in \mathcal{Z} := \text{span}\{x_k, x_l\}$. Since $d + \mu(y) = a\|y\|^2 \neq 0$, by (4.7) so is x .

Furthermore, the orthogonality $(Bx - y, y) = 0$ from (4.6) shows that

$$\cos^2 \angle(Bx, y) = \frac{(Bx, y)^2}{\|Bx\|^2 \|y\|^2} = \frac{(y, y)^2}{\|Bx\|^2 \|y\|^2} = \frac{\|y\|^2}{\|Bx\|^2},$$

and $\sin^2 \angle(Bx, y) = 1 - \cos^2 \angle(Bx, y) = (\|Bx\|^2 - \|y\|^2) / \|Bx\|^2 = \|Bx - y\|^2 / \|Bx\|^2$. This leads to $\sin \angle(Bx, y) = \|Bx - y\| / \|Bx\|$, since the angles between vectors have the range $[0, \pi]$ (due to arccos). Similarly, $(Bx - \kappa x, x) = 0$ together with $\kappa > 0$ implies $\sin \angle(Bx, x) = \sin \angle(Bx, \kappa x) = \|Bx - \kappa x\| / \|Bx\|$. Then we have $\sin \angle(Bx, y) = \gamma \sin \angle(Bx, x)$ by using (4.4). \square

We now derive bound (2.4) in a two-dimensional B -invariant subspace.

LEMMA 4.2. *Let x^* and y^* belong to a two-dimensional invariant subspace of B corresponding to the eigenvalues $\mu_k > \mu_l$ and satisfy (4.1), where x^* is not an eigenvector. It holds that*

$$(4.8) \quad \frac{\mu_k - \mu(y^*)}{\mu(y^*) - \mu_l} \frac{\mu(x^*) - \mu_l}{\mu_k - \mu(x^*)} \leq \left(\gamma + (1 - \gamma) \frac{\mu_l}{\mu_k} \right)^2.$$

Proof. In this proof we drop the superscript $*$ upon x and y . The vectors x , y and Bx can be represented by the coefficient vectors

$$u := c_1(1, \alpha)^T, \quad v := c_2(1, \beta)^T \quad \text{and} \quad w := c_1(\mu_k, \alpha\mu_l)^T$$

with respect to an orthonormal basis $\{x_k, x_l\}$, where x_k and x_l are eigenvectors associated with μ_k and μ_l . Evidently, it holds that $(Bx, y) = (w, v)$, $\|Bx\| = \|w\|$, $\|y\| = \|v\|$ by using the orthonormal basis. Therefore, $\angle(Bx, y) = \angle(w, v)$, and similarly $\angle(Bx, x) = \angle(w, u)$. This allows us to rewrite (4.1) in the form $\sin \angle(w, v) = \gamma \sin \angle(w, u)$. Using the geometric property of the cross products

$$\tilde{v} := \begin{bmatrix} w \\ 0 \end{bmatrix} \times \begin{bmatrix} v \\ 0 \end{bmatrix} \quad \text{and} \quad \tilde{u} := \begin{bmatrix} w \\ 0 \end{bmatrix} \times \begin{bmatrix} u \\ 0 \end{bmatrix},$$

we have

$$\frac{\|\tilde{v}\|}{\|w\| \|v\|} = \gamma \frac{\|\tilde{u}\|}{\|w\| \|u\|},$$

which yields

$$\gamma^2 = \frac{\|\tilde{v}\|^2 \|u\|^2}{\|\tilde{u}\|^2 \|v\|^2} = \frac{(\beta\mu_k - \alpha\mu_l)^2 (1 + \alpha^2)}{(\alpha\mu_k - \alpha\mu_l)^2 (1 + \beta^2)}.$$

Further, we use the equalities

$$\alpha^2 = \tan^2 \angle(x, x_k) = \frac{\mu_k - \mu(x)}{\mu(x) - \mu_l}, \quad \beta^2 = \tan^2 \angle(y, x_k) = \frac{\mu_k - \mu(y)}{\mu(y) - \mu_l},$$

which can be derived in a similar way to Section 3. Then $\frac{1 + \alpha^2}{1 + \beta^2} = \frac{\mu(y) - \mu_l}{\mu(x) - \mu_l} \geq 1$, so that

$$\gamma^2 \geq \frac{(\beta\mu_k - \alpha\mu_l)^2}{(\alpha\mu_k - \alpha\mu_l)^2}, \quad \text{and} \quad \left| \frac{\beta}{\alpha} - \frac{\mu_l}{\mu_k} \right| \leq \gamma \left| 1 - \frac{\mu_l}{\mu_k} \right|.$$

Since $0 < \mu_l/\mu_k < 1$, we have

$$\left| \frac{\beta}{\alpha} \right| \leq \left| \frac{\beta}{\alpha} - \frac{\mu_l}{\mu_k} \right| + \left| \frac{\mu_l}{\mu_k} \right| \leq \gamma \left(1 - \frac{\mu_l}{\mu_k} \right) + \frac{\mu_l}{\mu_k} = \gamma + (1 - \gamma) \frac{\mu_l}{\mu_k},$$

which proves (4.8) by using $\frac{\mu_k - \mu(y)}{\mu(y) - \mu_l} \frac{\mu(x) - \mu_l}{\mu_k - \mu(x)} = \left| \frac{\beta}{\alpha} \right|^2$. \square

The proof of Theorem 2.1 is completed by deriving convergence bound (2.4) from its two-dimensional version. We restate the assumption $\mu_{i+1} < \mu(x) < \mu(x') < \mu_i$. According to Lemma 4.1, there exists a minimizer y with $\mu(y) \leq \mu(x')$, which satisfies (4.8) with $\mu_l < \mu(x) < \mu(y) < \mu_k$, and the interval (μ_{i+1}, μ_i) is a subset of (μ_l, μ_k) . Using monotonicity arguments, (4.8), and the same arguments as Section 3, we obtain

$$\begin{aligned} \frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}} \frac{\mu(x) - \mu_{i+1}}{\mu_i - \mu(x)} &= \frac{\mu_i - \mu(y)}{\mu_i - \mu(x)} \frac{\mu(x) - \mu_{i+1}}{\mu(y) - \mu_{i+1}} \leq \frac{\mu_k - \mu(y)}{\mu_k - \mu(x)} \frac{\mu(x) - \mu_l}{\mu(y) - \mu_l} \\ &= \frac{\mu_k - \mu(y)}{\mu(y) - \mu_l} \frac{\mu(x) - \mu_l}{\mu_k - \mu(x)} \leq \left(\gamma + (1 - \gamma) \frac{\mu_l}{\mu_k} \right)^2 \leq \left(\gamma + (1 - \gamma) \frac{\mu_{i+1}}{\mu_i} \right)^2. \end{aligned}$$

This proves (2.4) since $(\mu_i - \mu(x'))/(\mu(x') - \mu_{i+1}) \leq (\mu_i - \mu(y))/(\mu(y) - \mu_{i+1})$.

Conclusions. We have presented a succinct proof of the standard sharp convergence rate bound for the simplest fixed step size preconditioned eigensolver. The key argument of the new proof is the characterization of the case of poorest convergence as a constrained optimization problem for the Rayleigh quotient. Employing the Karush-Kuhn-Tucker conditions and some elementary matrix algebra, we dramatically simplify the convergence analysis by reducing it to a subspace spanned by two eigenvectors. We expect the analytical framework developed in this paper to be a valuable tool in the convergence analysis of a variety of preconditioned eigensolvers; see, e.g., the analysis of the preconditioned steepest descent method in [16].

Appendix. I. An alternative estimate for the left-hand side of (4.8), which is sharp with respect to all variables, can be derived as follows: With $\delta := \beta/\alpha$ and $\varepsilon := \mu_l/\mu_k$, a new representation of γ^2 is given by

$$\gamma^2 = \frac{(\delta - \varepsilon)^2(1 + \alpha^2)}{(1 - \varepsilon)^2(1 + \alpha^2\delta^2)}.$$

This results in a quadratic equation for δ with the roots

$$\delta_{\pm} = \frac{\varepsilon(1 + \alpha^2) \pm \gamma(1 - \varepsilon)\sqrt{(1 + \alpha^2)(1 + \alpha^2\varepsilon^2) - \alpha^2\gamma^2(1 - \varepsilon)^2}}{(1 + \alpha^2) - \alpha^2\gamma^2(1 - \varepsilon)^2}.$$

Since $\delta^2 = \frac{\beta^2}{\alpha^2} = \frac{\mu_k - \mu(y)}{\mu(y) - \mu_l} \frac{\mu(x) - \mu_l}{\mu_k - \mu(x)}$, a strictly sharp bound for the estimate in (4.8) is given by $\max\{\delta_+^2, \delta_-^2\} = \delta_+^2$. We note that in the limit case $\mu(x) \rightarrow \mu_k$ it holds that $\alpha \rightarrow 0$, and δ_+^2 turns into $(\varepsilon + \gamma(1 - \varepsilon))^2$. This coincides with the known bound in (4.8).

II. The bound in (4.8) contains a convex combination of 1 and μ_l/μ_k . Interestingly, this bound can also be derived by using a convex function as follows: Without loss of generality, we assume that x has a positive x_k coordinate. Then $\angle(x, x_k)$ is an acute angle. Since $B > 0$, $\angle(Bx, x)$ and $\angle(Bx, x_k)$ are also acute angles. The equality (4.1)

together with $\gamma < 1$ shows further $\angle(Bx, y) < \angle(Bx, x) < \pi/2$. Since $\angle(Bx, x_k)$ and $\angle(Bx, y)$ are acute angles, the vectors x_k and y are located in a half-plane whose boundary line is orthogonal to Bx . A simple case differentiation shows that $\angle(y, x_k)$ is either equal to $|\angle(Bx, x_k) - \angle(Bx, y)|$ or equal to $\angle(Bx, x_k) + \angle(Bx, y)$. Further, we use the equalities

$$\tan^2 \angle(y, x_k) = \frac{\mu_k - \mu(y)}{\mu(y) - \mu_l}, \quad \tan^2 \angle(x, x_k) = \frac{\mu_k - \mu(x)}{\mu(x) - \mu_l}, \quad \frac{\tan^2 \angle(Bx, x_k)}{\tan^2 \angle(x, x_k)} = \frac{\mu_l^2}{\mu_k^2},$$

which can be derived in a similar way to Section 3. The last equality proves $\angle(Bx, x_k) < \angle(x, x_k)$, since the tangent is an increasing function for acute angles, and $\mu_l < \mu_k$. This leads to $\angle(x, x_k) = \angle(Bx, x_k) + \angle(Bx, x)$, since x, Bx, x_k are all in the same quadrant. In summary, it holds that

$$(4.9) \quad \angle(y, x_k) \leq \angle(Bx, x_k) + \angle(Bx, y) < \angle(Bx, x_k) + \angle(Bx, x) = \angle(x, x_k) < \pi/2,$$

i.e., $\angle(y, x_k)$ is a further acute angle. Using these acute angles, we write (4.8) equivalently as

$$(4.10) \quad \tan \angle(y, x_k) \leq \gamma \tan \angle(x, x_k) + (1 - \gamma) \tan \angle(Bx, x_k).$$

In order to prove (4.10), we use (4.1) again, together with $\varphi := \angle(Bx, x)$, $\vartheta := \angle(Bx, x_k)$ and the first inequality in (4.9). It holds that

$$\tan \angle(y, x_k) \leq \tan[\vartheta + \arcsin(\gamma \sin(\varphi))] =: f(\gamma).$$

Because of

$$f'(\gamma) = \frac{(1 + f(\gamma)^2) \sin(\varphi)}{\sqrt{1 - (\gamma \sin(\varphi))^2}} \geq 0 \quad \text{for } \gamma \in [0, 1],$$

$f(\gamma)$ is a monotonically increasing function in $[0, 1]$. The numerator of $f'(\gamma)$ is also a monotonically increasing function and its denominator is monotonically decreasing in $\gamma \in [0, 1]$. These two functions are positive so that $f'(\gamma)$ is also a monotonically increasing function. Thus $f(\gamma)$ is a convex function in $[0, 1]$, and

$$\tan \angle(y, x_k) \leq f(\gamma) \leq (1 - \gamma)f(0) + \gamma f(1) = (1 - \gamma) \tan(\vartheta) + \gamma \tan(\vartheta + \varphi),$$

which proves (4.10) and hence (4.8).

REFERENCES

1. 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.
2. F. BOTTIN, S. LEROUX, A. KNYAZEV, G. ZERAH, Large-scale ab initio calculations based on three levels of parallelization, *Computational Materials Science*, 42(2008), 2, pp. 329–336. doi:10.1016/j.commatsci.2007.07.019
3. H. BOUWMEESTER, A. DOUGHERTY, A. V. KNYAZEV, Nonsymmetric Preconditioning for Conjugate Gradient and Steepest Descent Methods, *Procedia Computer Science*, 51 (2015), pp. 276–285. doi:10.1016/j.procs.2015.05.241. A preliminary version available at arXiv:1212.6680 [cs.NA], 2012. <http://arxiv.org/abs/1212.6680>
4. S. BRIN AND L. PAGE, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, 30 (1998), 17, pp. 107–117. doi:10.1016/S0169-7552(98)00110-X

5. E. G. D'YAKONOV, *Optimization in Solving Elliptic Problems*, CRC Press, Boca Raton, Florida, 1996. ISBN: 978-0849328725
6. R. FLETCHER, *Practical Methods of Optimization*, John Wiley & Sons, Second Edition, 1987.
7. A. V. KNYAZEV, *Computation of eigenvalues and eigenvectors for mesh problems: algorithms and error estimates*, (In Russian), Dept. Num. Math., USSR Ac. Sci., Moscow, 1986.
8. A. V. KNYAZEV, Convergence rate estimates for iterative methods for a mesh symmetric eigenvalue problem, *Russian J. Numer. Anal. Math. Modelling*, 2 (1987), pp. 371–396. doi:10.1515/rnam.1987.2.5.371
9. A. V. KNYAZEV, Preconditioned eigensolvers - an oxymoron?, *Electronic Transactions on Numerical Analysis*, 7(1998), pp. 104–123. <http://etna.mcs.kent.edu/vol.7.1998/pp104-123.dir/pp104-123.pdf>
10. A. V. KNYAZEV, Modern Preconditioned Eigensolvers for Spectral Image Segmentation and Graph Bisection, Workshop on Clustering Large Data Sets Third IEEE International Conference on Data Mining (ICDM 2003), 2003. <http://math.ucdenver.edu/~aknyazev/research/conf/ICDM03.pdf>
11. A. V. 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. doi:10.1016/S0024-3795(01)00461-X
12. A. V. KNYAZEV AND K. NEYMEYR, Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method, *Electronic Transactions on Numerical Analysis*, 15 (2003), pp. 38–55. <http://etna.mcs.kent.edu/vol.15.2003/pp38-55.dir/pp38-55.pdf>
13. 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. doi:10.1137/080727567
14. D. KRESSNER, M. STEINLECHNER, AND A. USCHMAJEV, Low-rank tensor methods with subspace correction for symmetric eigenvalue problems, *SIAM J. Sci. Comput.*, 36(2014), 5, pp. A2346–A2368. <http://sma.epfl.ch/~anchpcommon/publications/EVAMEN.pdf>
15. D. KRESSNER, M. M. PANDUR, M. SHAO, An indefinite variant of LOBPCG for definite matrix pencils, *J Numerical Algorithms*, 66(2014), 4, pp. 681–703. doi:10.1007/s11075-013-9754-3
16. K. NEYMEYR, A geometric convergence theory for the preconditioned steepest descent iteration, *SIAM J. Numer. Anal.*, 50 (2012), pp. 3188–3207.
17. 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.
18. J. NOCEDAL AND S.J. WRIGHT, *Numerical Optimization*, Springer, 2006.
19. 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. doi:10.1137/040620643
20. D. B. SZYLD AND F. XUE, Preconditioned eigensolvers for large-scale nonlinear Hermitian eigenproblems with variational characterizations. I. Conjugate gradient methods, Research Report 14-08-26, Department of Mathematics, Temple University, August 2014. Revised April 2015. To appear in *Mathematics of Computation*. <https://www.math.temple.edu/~szyld/reports/NLPCG.report.rev.pdf>
21. D. B. SZYLD, E. VECHARYNSKI AND F. XUE, Preconditioned eigensolvers for large-scale nonlinear Hermitian eigenproblems with variational characterizations. II. Interior eigenvalues, Research Report 15-04-10, Department of Mathematics, Temple University, April 2015. To appear in *SIAM Journal on Scientific Computing*. <http://arxiv.org/abs/1504.02811>
22. E. VECHARYNSKI, Y. SAAD, AND M. SOSONKINA Graph partitioning using matrix values for preconditioning symmetric positive definite systems, *SIAM J. Sci. Comput.*, 36(2014), 1, pp. A63–A87. doi:10.1137/120898760
23. E. VECHARYNSKI, C. YANG, AND J. E. PASK, A projected preconditioned conjugate gradient algorithm for computing a large invariant subspace of a Hermitian matrix, *Journal of Computational Physics*, Vol. 290, pp. 73–89, 2015.
24. S. YAMADA, T. IMAMURA, T. KANO, AND M. MACHIDA, High-performance computing for exact numerical approaches to quantum many-body problems on the earth simulator, In Proceedings of the 2006 ACM/IEEE conference on Supercomputing (SC '06). ACM, New York, NY, USA, article 47, 2006. doi:10.1145/1188455.1188504