A geometric theory for preconditioned inverse iteration. III: A short and sharp convergence estimate for generalized eigenvalue problems.

Andrew V. Knyazev

Department of Mathematics, University of Colorado at Denver, P.O. Box 173364, Campus Box 170, Denver, CO 80217-3364 [1]

Klaus Neymeyr

Mathematisches Institut der Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany [2]

Abstract

In two previous papers by Neymeyr: A geometric theory for preconditioned inverse iteration I: Extrema of the Rayleigh quotient, LAA 322: (1-3), 61-85, 2001, and A geometric theory for preconditioned inverse iteration II: Convergence estimates, LAA 322: (1-3), 87-104, 2001, a sharp, but cumbersome, convergence rate estimate was proved for a simple preconditioned eigensolver, which computes the smallest eigenvalue together with the corresponding eigenvector of a symmetric positive definite matrix, using a preconditioned gradient minimization of the Rayleigh quotient. In the present paper, we discover and prove a much shorter and more elegant, but still sharp in decisive quantities, convergence rate estimate of the same method that also holds for a generalized symmetric definite eigenvalue problem. The new estimate is simple enough to stimulate a search for a more straightforward proof technique that could be helpful to investigate such practically important method as the locally optimal block preconditioned conjugate gradient eigensolver. We demonstrate practical effectiveness of the latter for a model problem, where it compares favorably with two well-known Jacobi-Davidson type methods, JDQR and JDCG.

Key words: Symmetric generalized eigenvalue problem, preconditioning, preconditioned eigensolver, gradient, steepest descent, conjugate gradient, matrix-free, Rayleigh, Ritz, Davidson, eigenvector, iterative method
1991 MSC: 65F15

[1] *E-mail*: `andrew.knyazev@cudenver.edu`
WWW URL: `http://www-math.cudenver.edu/~aknyazev`

[2] *E-mail*: `neymeyr@na.uni-tuebingen.de`

1 Introduction

Let A and T be real symmetric positive definite n-by-n matrices. We consider the problem of computing the smallest eigenvalue λ_1 and the corresponding eigenvector u_1 of matrix A by preconditioned iterative methods, where T will play the role of the preconditioner, e.g., Knyazev (2000). Such eigensolvers are matrix–free, i.e. no A, neither the preconditioner T need to be available as matrices, and are designed to solve efficiently and accurately extremely large and ill–conditioned eigenvalue problems.

The trivial choice $T = I$, see Kantorovich (1952) and Hestenes and Karush (1951), suffers from poor convergence for ill-conditioned matrices, cf. Bradbury and Fletcher (1966); Feng and Owen (1996); Rodrigue (1973); Yang (1991); Knyazev and Skorokhodov (1991). Preconditioned gradient methods with a general preconditioner T for symmetric eigenvalue problem have been studied, e.g., by Samokish (1958), Petryshyn (1968), Godunov et al. (1976), D'yakonov and Orekhov (1980); D'yakonov (1983), Knyazev (1987, 1998) as well as in the monograph D'yakonov (1996) and in a recent survey Knyazev (2000), which include extensive bibliography. Such preconditioned eigensolvers have been used in practice, e.g., for band structure calculations Dobson (1999); Dobson et al. (2000), thin elastic structures Ovtchinnikov and Xanthis (2000), and a real-space *ab initio* method for electronic structure calculations in terms of nonorthogonal orbitals defined on a grid Fattebert and Bernholc (2000). In the latter paper, a multigrid preconditioner is employed to improve the steepest descent directions used in the iterative minimization of the energy functional.

Let us also mention here briefly a number of very recent articles on preconditioned eigensolvers, even though they are not as closely related to the subject of the present paper as the papers cited in the previous paragraph. Oliveira (1999) obtains asymptotic convergence rate estimate of the generalized Davidson method similar to that by Samokish (1958) for the preconditioned steepest descent. Sadkane and Sidje (1999) discuss the block Davidson method with deflation. Smit and Paardekooper (1999) study inexact inverse and Rayleigh quotient iterations, using a perturbation technique somewhat comparable with that used in Neymeyr (2001a,b), but explicitly based on the error reduction rate of the inner iterations. Basermann (2000) applies a block incomplete LU decomposition for preconditioning in the Jacobi-Davidson method Sleijpen and Van der Vorst (1996); Bai et al. (2000). Ng (2000) uses for Toeplitz matrices the preconditioned Lanczos method suggested and analyzed in Scott (1981); Knyazev (1987); Morgan and Scott (1993), see also Bai et al. (2000).

Let $\|\cdot\|_A$ denote the A-based vector norm $\|\cdot\|_A = (\cdot, A\cdot)$ as well as the corresponding induced operator norm. For our theoretical estimates, we assume that the preconditioner T approximates the matrix A, such that

$$||I - T^{-1}A||_A \leq \gamma, \; 0 \leq \gamma < 1. \tag{1}$$

In general, as both matrices A and T are symmetric positive definite, the following always holds:

$$\delta_0(u, Tu) \leq (u, Au) \leq \delta_1(u, Tu), \; 0 < \delta_0 \leq \delta_1. \tag{2}$$

The ratio δ_1/δ_0 can be viewed as the spectral condition number $\kappa(T^{-1}A)$ of the preconditioned matrix $T^{-1}A$ and measures how well the preconditioner T approximates, up to a scaling, the matrix A. A smaller ratio δ_1/δ_0 typically ensures faster convergence. For mesh problems, matrices A and T are called *spectrally equivalent* if the ratio is bounded from above uniformly in the mesh size parameter, see D'yakonov (1996).

Assumption (1) leads to (2) with $\delta_0 = 1 - \gamma$ and $\delta_1 = 1 + \gamma$. *Vice versa*, assumption (2) leads to (1), but only if T is properly scaled. Namely, if T satisfies (2) then optimally scaled $2T/(\delta_0 + \delta_1)$ substituting T satisfies (1) with

$$\gamma = \frac{\kappa(T^{-1}A) - 1}{\kappa(T^{-1}A) + 1}. \tag{3}$$

Our convergence estimates in the present paper for methods with optimal scaling will be based on assumption (2) and will use γ given by (3). We note that some preconditioned eigensolvers, e.g., the steepest descent method we will discuss later, implicitly provide the optimal scaling of the preconditioner. In the rest of the paper, we will assume (1), unless explicitly stated otherwise, in order to be consistent with the previous papers Neymeyr (2001a,b).

It is well-known that the minimum of the Rayleigh quotient

$$\lambda(u) = \frac{(u, Au)}{(u, u)}, \text{ where } u \in \mathbb{R}^n, u \neq 0, \tag{4}$$

is λ_1 and the corresponding stationary point is the eigenvector u_1 of A. Gradient preconditioned eigensolvers generate a sequence of nonzero vectors, which minimizes the Rayleigh quotient, using its gradient, computed in the T-based scalar product $(\cdot, \cdot)_T = (\cdot, T\cdot)$, see, e.g., D'yakonov (1996):

$$\nabla_T \lambda(u) = \frac{2}{(u, u)_T} T^{-1}(Au - \lambda(u)u). \tag{5}$$

The simplest method in this class, a two–term gradient minimization, can be written as

$$u^{(i+1)} = u^{(i)} - \omega^{(i)} T^{-1} \left(Au^{(i)} - \lambda(u^{(i)}) u^{(i)} \right), \tag{6}$$

where $\omega^{(i)}$ is a scalar step size. We will analyze the error reduction of one step of the method,

$$u' = u - \omega T^{-1}(Au - \lambda u), \tag{7}$$

where we discard upper indexes and denote $u' = u^{(i+1)}$, $u = u^{(i)}$, $\omega = \omega^{(i)}$, and $\lambda = \lambda(u^{(i)})$.

We will consider two choices of ω here. The first case is an *a priori* fixed choice $\omega = 1$. This choice is evidently affected by a preconditioner scaling.

The second choice corresponds to the well-known, e.g., D'yakonov (1996); Knyazev (2000), preconditioned steepest descent for the Rayleigh quotient, where ω is chosen to minimize the Rayleigh quotient on the two-dimensional subspace span$\{u, T^{-1}(Au - \lambda u)\}$ by means of the Rayleigh–Ritz method. This leads to a 2-by-2 generalized eigenvalue problem that can be solved explicitly by using formulas for roots of the corresponding characteristic equation, which is in this case quadratic. We emphasize again that such choice of ω implicitly determines the optimal preconditioner scaling constant; thus, (3) can be used in convergence rate estimates in this case.

Preconditioned steepest descent is an obvious way to accelerate the convergence of the basic preconditioned eigensolver (7) with $\omega = 1$. There are several practically more efficient algorithms, e.g., the recent successive eigenvalue relaxation method of Ovtchinnikov and Xanthis (2001), and preconditioned conjugate gradient algorithms for minimizing the Rayleigh quotient, using an approximate inverse preconditioner, see a recent paper Bergamaschi et al. (2000) and references there.

The most promising preconditioned eigensolver is the locally optimal block preconditioned conjugate gradient (LOBPCG) method suggested and analyzed in Knyazev (1991, 1998, 2000, 2001). In LOBPCG for computing the first eigenpair, the new iterate is determined by the Rayleigh–Ritz method on a three-dimensional subspace, which includes the previous iterate in addition to the current iterate and the preconditioned residual of the two-dimensional trial subspace of the steepest descent method. The LOBPCG converges many times faster than the steepest descent in numerical tests, and is argued in Knyazev (2001) to be practically the optimal method on the whole class of preconditioned eigensolvers. However, no simple comprehensive convergence theory of the LOBPCG, explaining its apparent optimality, is yet known. The reason is that deriving sharp convergence estimates is challenging even for simplest preconditioned eigensolvers, such as that described by (7).

While an apparently sharp asymptotic convergence rate estimate for the preconditioned steepest descent method appeared in the very first paper Samokish (1958), a sharp non-asymptotic convergence rate estimate is not yet known despite of major efforts over the decades; see Knyazev (1998) for the review and references. For a simpler method, namely, (7) with $\omega = 1$, a sharp non-asymptotic convergence rate estimate was proved only recently, in Neymeyr (2001a,b). There, Neymeyr interpreted a preconditioned gradient method with a fixed step size as a *perturbation of a well known inverse iteration method*, in such a way that the associated system of linear equations was solved approximately by using a *preconditioner*. To highlight this, the method (7) with $\omega = 1$ was called the *Preconditioned INVerse ITeration* (PINVIT). A simple geometric interpretation of the method was discovered that provided a basis for derivation of sharp convergence estimates Neymeyr (2001b).

The estimate of Neymeyr (2001a,b) is sharp, but too cumbersome for a human being. In the present paper, we discover and prove a much shorter and more elegant, but still sharp, convergence rate estimate for the same method. The new estimate also holds for a generalized symmetric definite eigenvalue problem. It is simple enough to stimulate a search for a more straightforward proof technique that might finally lead to considerable progress in theory of practically important methods, such as LOBPCG Knyazev (2001).

There are several preconditioned eigensolvers, similar to classical subspace iterations, for computing an invariant subspace spanned by a group of eigenvectors corresponding to several smallest eigenvalues of A; see; e.g., McCormick and Noe (1977); Longsine and McCormick (1980); Bramble et al. (1996); Knyazev (2000); Zhang et al. (1999) and, for trace minimization methods, see Bai et al. (2000); Sameh and Tong (2000) and references there.

In Neymeyr (2000), the sharp convergence rate estimate of Neymeyr (2001a,b) for single-vector preconditioned solver is generalized to cover similar subspace iterations. A sharp simplification of the estimate of Neymeyr (2000) is suggested in Knyazev (2001), but the proof is sketchy and not complete. In the present paper, we fill these gaps in the arguments of Knyazev (2001).

The paper is organized as follows. In Section 2, we derive a new simple and sharp convergence estimate for the PINVIT. Furthermore, we derive an upper estimate for the convergence of preconditioned steepest descent. We extend these results to generalized symmetric definite eigenproblems in Section 3. In Section 4, we present similar convergence estimates for preconditioned subspace iterations. Numerical results are given in Section 5.

2 Preconditioned inverse iteration

According to formula (1.5) of Theorem 1.1 in Neymeyr (2001b), the sharp estimate from above for the Rayleigh quotient of u', computed by (7) with $\omega = 1$ is the following lengthy and, therefore, somewhat unreadable result: if $\lambda = \lambda(u) \in [\lambda_k, \lambda_{k+1}[$ then

$$\lambda' = \lambda(u') \leq \lambda_{k,k+1}(\lambda, \gamma), \tag{8}$$

$$\begin{aligned}
&\lambda_{k,k+1}(\lambda,\gamma) = \\
&\lambda\lambda_k\lambda_{k+1}(\lambda_k+\lambda_{k+1}-\lambda)^2 \\
&\qquad \Big(\gamma^2(\lambda_{k+1}-\lambda)(\lambda-\lambda_k)(\lambda\lambda_{k+1}+\lambda\lambda_k-\lambda_k^2-\lambda_{k+1}^2) \\
&\qquad\quad -2\gamma\sqrt{\lambda_k\lambda_{k+1}}(\lambda-\lambda_k)(\lambda_{k+1}-\lambda) \\
&\qquad\qquad\quad \sqrt{\lambda_k\lambda_{k+1}+(1-\gamma^2)(\lambda-\lambda_k)(\lambda_{k+1}-\lambda)} \\
&\qquad\quad -\lambda(\lambda_k+\lambda_{k+1}-\lambda)(\lambda\lambda_{k+1}+\lambda\lambda_k-\lambda_k^2-\lambda_k\lambda_{k+1}-\lambda_{k+1}^2)\Big)^{-1},
\end{aligned} \tag{9}$$

see the theorem below for the exact meaning of notations.

The estimate (8) is sharp in a sense that a preconditioner T and a vector u can be found such that the bound for the Rayleigh quotient is attained. Here, we present a concise convergence rate estimate for PINVIT, written in different terms, which is also sharp, but in a different somewhat weaker sense; see Remark 2 below.

Theorem 1 *Let $u \in \mathbb{R}^n$ and let $\lambda = \lambda(u) \in [\lambda_1, \lambda_n[$ be its Rayleigh quotient, where $\lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of A. The preconditioner is assumed to satisfy (1) for some $\gamma \in [0, 1[$. If $\lambda = \lambda(u) \in [\lambda_k, \lambda_{k+1}[$ then it holds for the Rayleigh quotient $\lambda' = \lambda(u')$ with u' computed by (7) with $\omega = 1$ that either $\lambda' < \lambda_k$ (unless $k = 1$), or $\lambda' \in [\lambda_k, \lambda[$. In the latter case,*

$$\frac{\lambda' - \lambda_k}{\lambda_{k+1} - \lambda'} \leq \left(q\left(\gamma, \lambda_k, \lambda_{k+1}\right)\right)^2 \frac{\lambda - \lambda_k}{\lambda_{k+1} - \lambda}, \tag{10}$$

where

$$q\left(\gamma, \lambda_k, \lambda_{k+1}\right) = \gamma + (1-\gamma)\frac{\lambda_k}{\lambda_{k+1}} = 1 - (1-\gamma)\left(1 - \frac{\lambda_k}{\lambda_{k+1}}\right) \tag{11}$$

is the convergence factor.

PROOF. Evidently, having the estimate (8), we only need to show that the maximum for all $\lambda \in [\lambda_k, \lambda_{k+1}[$ of the function

$$\frac{\lambda_{k,k+1}(\lambda,\gamma) - \lambda_k}{\lambda_{k+1} - \lambda_{k,k+1}(\lambda,\gamma)} \frac{\lambda_{k+1} - \lambda}{\lambda - \lambda_k}, \tag{12}$$

where $\lambda_{k,k+1}(\lambda,\gamma)$ is explicitly given in (9), is exactly $\left(q\left(\gamma, \lambda_k, \lambda_{k+1}\right)\right)^2$. It is easy to check that the function takes this value, when $\lambda = \lambda_k$, however we are not able to find a simple proof that it is the maximal value, using the expression for $\lambda_{k,k+1}(\lambda,\gamma)$ from (9). Instead, we will use a different, though equivalent, representation of $\lambda_{k,k+1}(\lambda,\gamma)$ from the Theorem 1.1 in Neymeyr (2001b), which provides the "mini–dimensional analysis" in $S = \text{span}\{u_k, u_{k+1}\}$, see also Theorem 5.1 in Neymeyr (2001a). We adopt the notations of the latter theorem and set for convenience $k = 1$ and $k + 1 = 2$, without a loss of generality.

It is shown in Neymeyr (2001a) that the set of all iterates E_γ, when one fixes the vector u and chooses all preconditioners T satisfying (1), is a ball, in the A–based scalar product. In the two-dimensional subspace S, the intersection $S \cap E_\gamma$ is a disk. The quantity r will denote the radius of the disk, and y and x will be Cartesian coordinates of its center with respect to a Cartesian system of coordinates, given by the A–orthonormal eigenvectors u_1 and u_2 of A, which span S, correspondingly. Neymeyr (2001a) obtains the following formulas:

$$x = \sqrt{\frac{\lambda(\lambda - \lambda_1)}{\lambda_2(\lambda_2 - \lambda_1)}},\ y = \sqrt{\frac{\lambda(\lambda_2 - \lambda)}{\lambda_1(\lambda_2 - \lambda_1)}},\ r = \gamma\sqrt{\frac{(\lambda - \lambda_1)(\lambda_2 - \lambda)}{\lambda_1\lambda_2}}.$$

According to Neymeyr (2001a), the unique maximum of the Rayleigh quotient on the whole E_γ is actually attained on the disc $S \cap E_\gamma$ and is given by $\lambda_{1,2}(\lambda,\gamma)$ defined by formula (5.6) of Neymeyr (2001a), reproduced here:

$$\lambda_{1,2}(\lambda,\gamma) = \frac{\eta^2 + \xi^2}{\eta^2/\lambda_1 + \xi^2/\lambda_2}, \tag{13}$$

where

$$(\eta, \xi) = (\sqrt{l^2 - \xi^2},\ \frac{xl^2 + ryl}{x^2 + y^2})$$

are the coordinates of the point of the maximum and l is its Euclidean norm; moreover,

$$l = \sqrt{x^2 + y^2 - r^2}.$$

Formula (9) is then derived from (13) in Neymeyr (2001a).

For our present proof, the geometric meaning of quantities is not at all important. The only important fact is that (13) provides a formula for $\lambda_{1,2}(\lambda,\gamma)$

for known x, y and r, which, in their turn, are explicitly given as functions of $\gamma, \lambda, \lambda_1$, and λ_2 only. The rest of the proof is nothing but simple, though somewhat tedious, algebraic manipulations.

Directly from (13), we have

$$\frac{\lambda_{12}-\lambda_1}{\lambda_2-\lambda_{12}} = \frac{\xi^2\lambda_1}{\eta^2\lambda_2} = \frac{\lambda_1}{\lambda_2}\,\frac{(xl+ry)^2}{(x^2+y^2)^2-(xl+ry)^2},$$

where in denominator

$$(x^2+y^2)^2-(xl+ry)^2 = (yl-xr)^2.$$

Here, $yl-xr$ is positive because of $y>r$.

Explicit expressions for x and y give

$$\frac{\lambda_2-\lambda}{\lambda-\lambda_1} = \frac{y^2\lambda_1}{x^2\lambda_2}.$$

Therefore, the convergence factor q, defined by

$$\frac{\lambda_{12}-\lambda_1}{\lambda_2-\lambda_{12}}\,\frac{\lambda_2-\lambda}{\lambda-\lambda_1} = \frac{\lambda_1^2 y^2}{\lambda_2^2 x^2}\,\frac{(xl+ry)^2}{(yl-rx)^2} =: q^2,$$

is equal to

$$q = \frac{\lambda_1 y(xl+ry)}{\lambda_2 x(yl-rx)} = \frac{\lambda_1}{\lambda_2}\,\frac{1+\frac{yr}{xl}}{1-\frac{xr}{yl}} > 0. \qquad (14)$$

Direct computation shows that

$$\frac{yr}{xl} = \gamma(\lambda_2-\lambda)\left(\frac{\lambda_2}{\lambda_1}\right)^{1/2} z^{-1/2}$$

and

$$\frac{xr}{yl} = \gamma(\lambda-\lambda_1)\left(\frac{\lambda_1}{\lambda_2}\right)^{1/2} z^{-1/2}$$

with $z = \gamma^2(\lambda_1-\lambda)(\lambda_2-\lambda)+\lambda(\lambda_1+\lambda_2-\lambda) > 0$. Hence,

$$q[\lambda] = \frac{\sqrt{\frac{\lambda_1}{\lambda_2}}z^{1/2}+\gamma(\lambda_2-\lambda)}{\sqrt{\frac{\lambda_2}{\lambda_1}}z^{1/2}-\gamma(\lambda-\lambda_1)}. \qquad (15)$$

We note again that value of $q[\lambda]$ squared in (15) must be the same as that of the expression (12) with $\lambda_{1,2}(\lambda,\gamma)$ given by (9) — it is just written in a more civilized way.

We now want to eliminate dependence of the convergence factor q on λ, by finding a sharp upper bound, independent of λ. For that, let us show

$$q'(\lambda) < 0,$$

which is equivalent to

$$\gamma\sqrt{\lambda_1\lambda_2}(\lambda_2-\lambda_1) < (\lambda_2-\lambda_1)z^{1/2} + (\frac{d}{d\lambda}z^{1/2})\left\{\lambda_2(\lambda_2-\lambda)+\lambda_1(\lambda-\lambda_1)\right\}.$$

Taking the square of both sides and inserting z and $\frac{d}{d\lambda}z^{1/2}$, we observe after factorization that the last inequality holds provided that the following quantity

$$(1-\gamma^2)(\lambda_2-\lambda_1)^2(\lambda_1+\lambda_2-\lambda)^2\left[(1+\gamma)\lambda_1+(1-\gamma)\lambda_2\right]\left[(1-\gamma)\lambda_1+(1+\gamma)\lambda_2\right]$$

is positive, which it trivially is under our assumptions $0 \leq \gamma < 1$ and $0 < \lambda_1 \leq \lambda < \lambda_2$. Thus, $q[\lambda]$ takes its largest value, when $\lambda = \lambda_1$:

$$q[\lambda_1] = \gamma + (1-\gamma)\frac{\lambda_1}{\lambda_2} = \frac{\lambda_1}{\lambda_2} + \gamma\left(1-\frac{\lambda_1}{\lambda_2}\right) = 1-(1-\gamma)\left(1-\frac{\lambda_1}{\lambda_2}\right).$$

□

Remark 2 *It follows directly from the proof of the theorem above that the true convergence factor in the estimate (10) may depend on λ, but this dependence is not decisive. We eliminate λ to make the estimate much shorter.*

Thus, our upper bound (11) of the convergence factor does not depend on λ and is sharp, as a function of the decisive quantities γ, λ_k, λ_{k+1} only. The estimate (10) is also asymptotically sharp, when $\lambda \to \lambda_k$, as it then turns into the sharp estimate (8).

Remark 3 *The preconditioned steepest descent for the Rayleigh quotient when ω is computed to minimize the Rayleigh quotient on the two-dimensional subspace* $\operatorname{span}\{u, T^{-1}(Au-\lambda u)\}$*, evidently produces a smaller value λ' compared to that when ω is chosen a priori. Thus, the convergence rate estimate (10) with the convergence factor (11) holds for the preconditioned steepest descent method, too. Moreover, we can now assume (2) instead of (1) and use (3) as we already discussed in the Introduction, which leads to*

$$1-\gamma = \frac{2}{\kappa(T^{-1}A)+1}, \tag{16}$$

This estimate for the preconditioned steepest descent is not apparently sharp as can be seen by comparing it with the asymptotic estimate by Samokish (1958).

3 Generalized symmetric definite eigenvalue problems

We now consider a generalized *symmetric definite* eigenvalue problem of the form $(A-\lambda B)u = 0$ with real symmetric n-by-n matrices A and B, assuming that A is positive definite. That describes a regular matrix pencil $A-\lambda B$ with a discrete spectrum (set of eigenvalues λ). It is well known that such generalized eigenvalue problem has all real eigenvalues λ_i and corresponding (right) eigenvectors u_i, satisfying $(A-\lambda_i B)u_i = 0$, can be chosen orthogonal in the following sense: $(u_i, Au_j) = (u_i, Bu_j) = 0,\ i \neq j$. In some applications, the matrix B is simply the identity, $B = I$, and then we have the standard symmetric eigenvalue problem with matrix A, which has n real positive eigenvalues

$$0 < \lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n = \lambda_{\max}.$$

We already discussed the case $B = I$ in the previous section.

In general, when $B \neq I$, the pencil $A - \lambda B$ has n real, some possibly infinite, eigenvalues. If B is nonsingular, all eigenvalues are finite. If B is positive semidefinite, some eigenvalues are infinite, all other eigenvalues are positive, and we consider the problem of computing the smallest m eigenvalues of the pencil $A - \lambda B$. When B is indefinite, it is convenient to consider the pencil $\mu A - B$ with eigenvalues

$$\mu = \frac{1}{\lambda},\ \mu_{\min} = \mu_n \leq \cdots \leq \mu_1 = \mu_{\max},$$

where we want to compute the largest m eigenvalues, $\mu_1, \ldots \mu_m$, and corresponding eigenvectors.

We first consider the case $B > 0$, when we may still use λ's. We naturally redefine the Rayleigh quotient (4) to

$$\lambda(u) = \frac{(u, Au)}{(u, Bu)},\ \text{where } u \in \mathbb{R}^n, u \neq 0, \tag{17}$$

and replace method (7) with the following:

$$u' = u - \omega T^{-1}(Au - \lambda(u)Bu), \tag{18}$$

still assuming that the preconditioner T approximates A according to (1).

A different popular approach to deal with a generalized eigenvalue problem, e.g., utilized in the ARPACK based MATLAB code EIGS.m, relies on explicit factorizations of the matrix B, A, or their linear combination. It cannot, of course, be used in a matrix-free environment, when all matrices are only available as matrix-vector-multiply (MVM) functions.

The method (18) is not new. It was previously studied, e.g., by D'yakonov and Orekhov (1980); D'yakonov (1983, 1996). Here, we easily derive a new sharp convergence estimate for it, using our previous result for $B = I$.

Theorem 4 *Let $B > 0$. Let $u \in \mathbb{R}^n$ and let $\lambda = \lambda(u) \in [\lambda_1, \lambda_n[$ be its Rayleigh quotient, where $\lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of $B^{-1}A$. The preconditioner is assumed to satisfy (1) for some $\gamma \in [0, 1[$. If $\lambda = \lambda(u) \in [\lambda_k, \lambda_{k+1}[$, then it holds for the Rayleigh quotient $\lambda' = \lambda(u')$ with u' computed by (18) with $\omega = 1$ that either $\lambda' < \lambda_k$ (unless $k = 1$), or $\lambda' \in [\lambda_k, \lambda[$. In the latter case, the convergence estimate (10) holds with the convergence factor (11).*

PROOF. As $B > 0$, the bilinear form $(\cdot, \cdot)_B = (\cdot, B\cdot)$ describes a scalar product, in which matrices $B^{-1}T$ and $B^{-1}A$ are symmetric positive definite. Let us make all the following substitutions at once:

$$(\cdot, \cdot)_B \Rightarrow (\cdot, \cdot),\ B^{-1}A \Rightarrow A,\ B^{-1}T \Rightarrow T.$$

Then, the formula (18) turns into (7) and the generalized eigenvalue problem for the pencil $A - \lambda B$ becomes a standard eigenvalue problem for the matrix $B^{-1}A$. Thus, we can use Theorem 4 that gives us the present theorem after the back substitution to the original terms of the present section. □

Remarks 2 and 3 hold with evident modifications for $B > 0$.

To cover the general case, when B may not be definite, we replace λ's with μ's by switching from the pencil $A - \lambda B$ to the pencil $B - \mu A$. We redefine the Rayleigh quotient (17) to

$$\mu(u) = \frac{(u, Bu)}{(u, Au)}, \text{ where } u \in \mathbb{R}^n, u \neq 0, \tag{19}$$

and replace method (18) with the following:

$$u' = u + \omega T^{-1}(Bu - \mu(u)Au), \tag{20}$$

still assuming that the preconditioner T approximates A according to (1). We are now interested in the largest eigenvalue μ_1 of the matrix $A^{-1}B$.

The method (18) was previously suggested, e.g., in Knyazev (1986) and reproduced in D'yakonov (1996). Now, we obtain a new sharp convergence estimate for it, using our previous theorem.

Theorem 5 *Let $u \in \mathbb{R}^n$ and let $\mu = \mu(u) \in]\mu_n, \mu_1]$ be its Rayleigh quotient, where $\mu_1 \geq \ldots \geq \mu_n = \mu_{\min}$ are the eigenvalues of $A^{-1}B$. The preconditioner is assumed to satisfy (1) for some $\gamma \in [0, 1[$. If $\mu = \mu(u) \in]\mu_{k+1}, \mu_k]$ then it holds for the Rayleigh quotient $\mu' = \mu(u')$ with u' computed by (20) with*

$$\omega = \frac{1}{\mu - \mu_{\min}}$$

that either $\mu' > \mu_k$ (unless $k = 1$), or $\mu' \in]\mu, \mu_k]$. In the latter case, the convergence estimate

$$\frac{\mu_k - \mu'}{\mu' - \mu_{k+1}} \leq q^2 \frac{\mu_k - \mu}{\mu - \mu_{k+1}}, \tag{21}$$

holds with the convergence factor

$$q = 1 - (1 - \gamma)\frac{\mu_j - \mu_{j+1}}{\mu_j - \mu_{\min}}. \tag{22}$$

PROOF. We first rewrite the estimate of the previous theorem for $B > 0$ in terms of μ's:

$$\frac{\mu_k - \mu'}{\mu' - \mu_{k+1}} \leq q^2 \frac{\mu_k - \mu}{\mu - \mu_{k+1}}, \; q = 1 - (1 - \gamma)\frac{\mu_j - \mu_{j+1}}{\mu_j}. \tag{23}$$

Here we use the fact that

$$\frac{\mu_k - \mu'}{\mu' - \mu_{k+1}} \frac{\mu - \mu_{k+1}}{\mu_k - \mu} = \frac{\lambda' - \lambda_k}{\lambda_{k+1} - \lambda'} \frac{\lambda_{k+1} - \lambda}{\lambda - \lambda_k}$$

and that

$$q = 1 - (1 - \gamma)\left(1 - \frac{\lambda_k}{\lambda_{k+1}}\right) = 1 - (1 - \gamma)\frac{\mu_k - \mu_{k+1}}{\mu_k}. \tag{24}$$

Now, we are ready to deal with a general symmetric B. We use a trick, suggested in Knyazev (1986) and reproduced in D'yakonov (1996). Namely, we substitute our actual matrix B, which is not necessarily positive definite with positive definite matrix $B_\alpha = B - \alpha A > 0$, where a scalar $\alpha < \mu_{\min}$, and apply the previous estimate (23) to the pencil $B_\alpha - \mu_\alpha A$ with eigenvalues $\mu_\alpha = \mu - \alpha$. This gives (23), but with

$$q = 1 - (1 - \gamma)\frac{\mu_k - \mu_{k+1}}{\mu_k - \alpha}.$$

Finally, we realize that the method itself is invariant with respect to α, except for the scalar shift that must be now chosen as

$$\omega = \frac{1}{\mu - \alpha}.$$

Moreover, everything depends continuously on $\alpha < \mu_{\min}$, so we can take the limit $\alpha = \mu_{\min}$ as well. This proves estimate (21) with q given by (22) $\square$

Remarks 2 and 3 for general B turn into the following.

Remark 6 *The convergence factor (22) is sharp, as a function of the decisive quantities γ, $\mu_k - \mu_{k+1}$, $\mu_k - \mu_{\min}$ only. The estimate (21) is also asymptotically sharp, when $\mu \to \mu_k$, as it then turns into a sharp estimate.*

Remark 7 *The preconditioned steepest ascent for the Rayleigh quotient (19) when ω in (20) is computed to maximize the Rayleigh quotient on the two-dimensional subspace* $\mathrm{span}\{u, T^{-1}(Bu - \mu Au)\}$*, evidently produces a larger value μ' compare to that when ω is chosen a priori. Thus, the convergence rate estimate (21) with the convergence factor (22) holds for the preconditioned steepest ascent method, too. Moreover, we can now assume (2) instead of (1) and use (16).*

Remark 8 *In the locally optimal preconditioned conjugate gradient method (4.2) of Knyazev (2001), the trial subspace is enlarged compare to that of the preconditioned steepest ascent method of Remark 7. Thus, the convergence rate estimate (21) with q given by (22) holds for the former method, too, assuming (2) and taking (16). Our preconditioner T was denoted as T^{-1} in Knyazev (2001).*

4 Preconditioned subspace iterations

In this section, we will present a generalization of results of the previous two sections to the case, where m extreme eigenpairs are computed simultaneously in so-called *subspace*, or *block* iteration methods.

We need to return to the case $B = I$ again and consider first the following block version of method (6).

Let the current iterate $U^{(i)}$ be an n-by-m matrix with columns, approximating m eigenvectors of A, corresponding to m smallest eigenvalues. We assume that

$$\left(U^{(i)}\right)^T U^{(i)} = I,\ \left(U^{(i)}\right)^T AU^{(i)} = \mathrm{diag}(\lambda_1^{(i)}, \ldots, \lambda_m^{(i)}) = \Lambda^{(i)}.$$

We perform one step of iterations

$$\hat{U}^{(i+1)} = U^{(i)} - T^{-1}\left(AU^{(i)} - U^{(i)}\Lambda^{(i)}\right)\Omega^{(i)}, \tag{25}$$

where $\Omega^{(i)}$ is an m-by-m matrix, a generalization of the scalar step size. Finally, we compute the next iterate $U^{(i+1)}$ by the Rayleigh-Ritz procedure for the pencil $A - \lambda I$ on the trial subspace given by the column-space of $\hat{U}^{(i+1)}$ such that

$$\left(U^{(i+1)}\right)^T U^{(i+1)} = I,\ \left(U^{(i+1)}\right)^T AU^{(i+1)} = \operatorname{diag}(\lambda_1^{(i+1)}, \ldots, \lambda_m^{(i+1)}) = \Lambda^{(i+1)}.$$

The preconditioned iterative method (25) with $\Omega^{(i)} = I$ is analyzed in Bramble et al. (1996), where a survey on various attempts to analyze this and simplified preconditioned subspace schemes is also given. In this analysis, restrictive conditions on the initial subspace are assumed to be satisfied.

An alternative theory for method (25) with $\Omega^{(i)} = I$ is developed in Neymeyr (2000), based on the sharp convergence rate estimate (8) of Neymeyr (2001a,b) for single-vector preconditioned solver that we use in the previous two sections. The advantages of the approach of Neymeyr (2000) are that:

- it is applicable to any initial subspaces,
- the convergence rate estimate can be used recursively, while the estimate of Bramble et al. (1996) cannot,
- the estimates for the convergence of the Ritz values are individually sharp in a sense that an initial subspace and a preconditioner can be constructed so that the convergence rate estimate for a fixed index $j \in [1, m]$ is attained,
- the convergence rate estimate for a fixed index j is exactly the same as (8) for the single-vector method (6) with $\omega^{(i)} = 1$.

The only serious disadvantage of the estimates of Neymeyr (2000) is that they deteriorate when eigenvalues of interest $\lambda_1, \ldots, \lambda_m$ include a cluster. The actual convergence of method (25) in numerical tests is known not to be sensitive to clustering of eigenvalues, and estimates of Bramble et al. (1996) do capture this property, essential for subspace iterations.

A sharp simplification of the estimate of Neymeyr (2000) is suggested in Theorem 5.1 of Knyazev (2001), but the proof is sketchy and not complete. In this section, we fill these gaps in the arguments of Knyazev (2001).

First, we reproduce here the result of Theorem 3.3 of Neymeyr (2000): for a fixed index $j \in [1, m]$, if $\lambda_j^{(i)} \in [\lambda_{k_j}, \lambda_{k_j+1}[$ and the method (25) with $\Omega^{(i)} = I$ is used, then

$$\lambda_j^{(i+1)} \leq \lambda_{k_j,k_j+1}(\lambda_j^{(i)}, \gamma), \tag{26}$$

where the latter quantity is given by (9). Now, using the fact that the estimate (26) is identical to (8) and that our proof of Theorem 1 provides an equivalent representation of expression (9), we immediately derive the following generalization of Theorem 1 to the block method

Theorem 9 *The preconditioner is assumed to satisfy (1) for some* $\gamma \in [0, 1[$. *For a fixed index* $j \in [1, m]$, *if* $\lambda_j^{(i)} \in [\lambda_{k_j}, \lambda_{k_j+1}[$ *then it holds for the Ritz value* $\lambda_j^{(i+1)}$ *computed by (25) with* $\Omega^{(i)} = I$ *that either* $\lambda_j^{(i+1)} < \lambda_{k_j}$ *(unless* $k_j = j$*), or* $\lambda_j^{(i+1)} \in [\lambda_{k_j}, \lambda_j^{(i)}[$.

In the latter case,

$$\frac{\lambda_j^{(i+1)} - \lambda_{k_j}}{\lambda_{k_j+1} - \lambda_j^{(i+1)}} \leq \left(q\left(\gamma, \lambda_{k_j}, \lambda_{k_j+1}\right)\right)^2 \frac{\lambda_j^{(i)} - \lambda_{k_j}}{\lambda_{k_j+1} - \lambda_j^{(i)}}, \tag{27}$$

where

$$q\left(\gamma, \lambda_{k_j}, \lambda_{k_j+1}\right) = \gamma + (1-\gamma)\frac{\lambda_{k_j}}{\lambda_{k_j+1}} = 1 - (1-\gamma)\left(1 - \frac{\lambda_{k_j}}{\lambda_{k_j+1}}\right) \tag{28}$$

is the convergence factor.

By analogy with Remarks 2 and 3, we have the following.

Remark 10 *For a fixed index* j, *the convergence factor* $q\left(\gamma, \lambda_{k_j}, \lambda_{k_j+1}\right)$ *given by (28) is sharp, as a function of the decisive quantities* γ, λ_{k_j}, λ_{k_j+1} *only. The estimate (27) is also asymptotically sharp, when* $\lambda_j^{(i)} \to \lambda_{k_j}$, *as it then turns into the sharp estimate (26).*

Let us highlight again that, while the convergence factors (28) are sharp individually, when we fix the index j, *they are not sharp collectively, for all* $j = 1, \ldots, m$, *neither asymptotically, when the initial subspace is already close to the seeking subspace spanned by the first* m *eigenvectors. In the latter case, the estimates of Bramble et al. (1996) are better.*

Remark 11 *There are several different versions of the preconditioned block steepest descent; see, e.g., Knyazev (2000). In one of them,* $U^{(i+1)}$ *is computed by the Rayleigh-Ritz method of the* $2m$*-dimensional trial subspaces, spanned by columns of* $U^{(i)}$ *and* $T^{-1}\left(AU^{(i)} - U^{(i)}\Lambda^{(i)}\right)$. *This leads to Ritz values* $\lambda_j^{(i+1)}$, *which are not larger than those produced by (25) with any* $\Omega^{(i)}$, *in particular, with* $\Omega^{(i)} = I$. *Thus, the convergence rate estimate (27) with the convergence factor (28) holds for this version of the preconditioned block steepest descent method, too. Moreover, we can now assume (2) instead of (1) and use (16).*

Let now $B \neq I$, $B > 0$. Then we assume that

$$\left(U^{(i)}\right)^T BU^{(i)} = I,\ \left(U^{(i)}\right)^T AU^{(i)} = \mathrm{diag}(\lambda_1^{(i)}, \ldots, \lambda_m^{(i)}) = \Lambda^{(i)}.$$

We perform one step of iterations

$$\hat{U}^{(i+1)} = U^{(i)} - T^{-1}\left(AU^{(i)} - BU^{(i)}\Lambda^{(i)}\right)\Omega^{(i)}, \tag{29}$$

and compute the next iterate $U^{(i+1)}$ by the Rayleigh-Ritz procedure for the pencil $A - \lambda B$ on the trial subspace given by the column-space of $\hat{U}^{(i+1)}$ such that

$$\left(U^{(i+1)}\right)^T BU^{(i+1)} = I,\ \left(U^{(i+1)}\right)^T AU^{(i+1)} = \mathrm{diag}(\lambda_1^{(i+1)}, \ldots, \lambda_m^{(i+1)}) = \Lambda^{(i+1)}.$$

Repeating the same arguments as those in the proof of Theorem 4, we conclude that Theorem 9 also trivially holds for the method (29) with $\Omega^{(i)} = I$ for solving an generalized eigenvalue problem for pencil $A - \lambda B$, when $B > 0$.

Finally, in the general case, when B may not be definite, we modify the method (29) for the pencil $B - \mu A$ the following way: assuming that

$$\left(U^{(i)}\right)^T AU^{(i)} = I,\ \left(U^{(i)}\right)^T BU^{(i)} = \mathrm{diag}(\mu_1^{(i)}, \ldots, \mu_m^{(i)}) = M^{(i)},$$

we perform one step of iterations

$$\hat{U}^{(i+1)} = U^{(i)} - T^{-1}\left(BU^{(i)} - AU^{(i)}M^{(i)}\right)\Omega^{(i)}, \tag{30}$$

and compute the next iterate $U^{(i+1)}$ by the Rayleigh-Ritz procedure for the pencil $B - \mu A$ on the trial subspace given by the column-space of $\hat{U}^{(i+1)}$ such that

$$\left(U^{(i+1)}\right)^T AU^{(i+1)} = I,\ \left(U^{(i+1)}\right)^T BU^{(i+1)} = \mathrm{diag}(\mu_1^{(i+1)}, \ldots, \mu_m^{(i+1)}) = M^{(i+1)}.$$

By analogy with the proof of Theorem 5, we derive

Theorem 12 *The preconditioner is assumed to satisfy (1) for some $\gamma \in [0, 1[$. For a fixed index $j \in [1, m]$, if $\mu_j^{(i)} \in]\mu_{k_j+1}, \mu_{k_j}]$ then it holds for the Ritz value $\mu_j^{(i+1)}$ computed by (30) with*

$$\Omega^{(i)} = \left(M^{(i)} - \mu_{\min} I\right)^{-1}$$

that either $\mu_j^{(i+1)} > \mu_{k_j}$ (unless $k_j = j$), or $\mu_j^{(i+1)} \in]\mu_j^{(i)}, \mu_{k_j}]$. In the latter case,

$$\frac{\mu_{k_j+1} - \mu_j^{(i+1)}}{\mu_j^{(i+1)} - \mu_{k_j}} \leq \left(q\left(\gamma, \mu_{k_j}, \mu_{k_j+1}\right)\right)^2 \frac{\mu_{k_j+1} - \mu_j^{(i)}}{\mu_j^{(i)} - \mu_{k_j}}, \tag{31}$$

where

$$q\left(\gamma, \mu_{k_j}, \mu_{k_j+1}\right) = 1 - (1-\gamma)\left(\frac{\mu_{k_j} - \mu_{k_j+1}}{\mu_{k_j} - \mu_{\min}}\right) \tag{32}$$

is the convergence factor.

Remark 13 *If columns of $U^{(i+1)}$ are computed by the Rayleigh-Ritz method for the pencil $B - \mu A$, as m Ritz vectors corresponding to the m largest Ritz values, on the $2m$-dimensional trial subspace spanned by columns of $U^{(i)}$ and $T^{-1}\left(BU^{(i)} - U^{(i)}M^{(i)}\right)$, the convergence rate estimate (31) with the convergence factor (32) holds for this version of the preconditioned block steepest ascent method, too. Moreover, we can now assume (2) instead of (1) and use (16).*

Remark 14 *In the locally optimal block preconditioned conjugate gradient (LOBPCG) method of Knyazev (2001), $U^{(i+1)}$ is computed by the Rayleigh-Ritz method on the $3m$-dimensional trial subspaces, spanned by columns of $U^{(i-1)}$, $U^{(i)}$ and $T^{-1}\left(BU^{(i)} - U^{(i)}M^{(i)}\right)$. Thus, in LOBPCG the trial subspace is enlarged compare to that of the preconditioned block steepest ascent method, described in the previous remark. Therefore, evidently, the convergence rate estimate (31) with the convergence factor given by (32) with (16), assuming (2), holds for the LOBPCG method, too; see Theorem 5.1 of Knyazev (2001).*

Remark 14 provides us with the only presently known theoretical convergence rate estimate of the LOBPCG. However, this estimates is, by construction, the same as that for the preconditioned block steepest ascent method, which, in its turns, is the same as that of the PINVIT with the optimal scaling. Numerical comparison of these methods according to Knyazev (1998, 2000, 2001) demonstrates, however, that the LOBPCG method is in practice much faster. Therefore, firstly, our theoretical convergence estimates of the LOBPCG of the present paper are not sharp enough yet to explain excellent convergence properties of the LOBPCG in numerical simulations, which we illustrate next. Secondly, we can only present here numerical tests for the LOBPCG, and we do not need to return back to numerical simulations of Knyazev (1998, 2000, 2001), which already showed a superiority of the LOBPCG with respect to the steepest ascent method and PINVIT.

5 A numerical example

In this final section, we demonstrate practical effectiveness of the LOBPCG method for a model problem by comparing it with JDCG and JDQR methods, see Notay (2001); Sleijpen and Van der Vorst (1996), using a test program written by Notay, which is publicly available at
http://homepages.ulb.ac.be/~ynotay/.
We refer to a recent paper Morgan (2000) for numerical comparisons of JDQR with the generalized Davidson method, the preconditioned Lanczos methods, and the inexact Rayleigh quotient iterations.

We consider, as in Notay (2001), the eigenproblem for the Laplacian on the L-shaped domain embedded in the unit square with the homogeneous Dirichlet boundary conditions. We compute several smallest eigenvalues and corresponding eigenfunctions of its finite difference discretization, using the standard five point pencil on a uniform mesh with the mesh size $h = 1/180$ and 23941 inner nodes. The discretized eigenproblem is, therefore, a matrix eigenvalue problem for the pencil $A - \lambda I$, where A is the stiffness matrix for the Laplacian and the mass matrix is simply the identity. The matrix A is of the size 23941 and has 118989 nonzero elements. Thus, the matrix vector multiplication (MVM) Au has approximately the same costs as five vector operations (VO).

To choose a preconditioner, we follow Notay (2001) and use an incomplete Cholesky factorization of the stiffness matrix A with a drop tolerance (DT). Two DT values are tested: 10^{-3} and 10^{-4}. A smaller DT improves the quality of the preconditioner, but at the same time increases the costs of constructing it and the costs of applying it on every iteration. The latter is called the costs of the preconditioner solve (PrecS) and is approximately equivalent to 30 VO and 65 VO for DT 10^{-3} and 10^{-4}, correspondingly, for the preconditioner chosen.

For each preconditioner, we run two tests: one to compute only the single smallest eigenpair and another to compute ten smallest eigenpairs. The accuracy of the output results is checked by computing the Euclidean norm of the residual $\|AU^{(i)} - U^{(i)}\Lambda^{(i)}\| < \epsilon$, where $U^{(i)}$ is the matrix whose columns are computed orthonormal approximations to eigenvectors and $\Lambda^{(i)}$ is the diagonal matrix with the corresponding computed approximations to the eigenvalues. Two accuracy levels are tested, $\epsilon = 10^{-5}$ and $\epsilon = 10^{-5}$.

Table 1 provides numerical comparison of the latest, as of April 12, 2001, revision 3.3 of the LOBPCG code with the data from Notay (2001) for the previous revision 3.2 of the LOBPCG code, JDCG code by Yvan NOTAY, and JDQR code by Gerard Sleijpen, for DT=10^{-3}. Here, MVM and PrecS lines

	$Ax = b$	One eigenpair computed by				Ten eigenpairs computed by			
	by PCG	3.3	3.2	JDCG	JDQR	3.3	3.2	JDCG	JDQR
	Accuracy $\epsilon = 10^{-5}$:								
MVM	33	15	15	21	27	140	350	165	144
PrecS	16	13	13	21	28	120	200	164	174
CPU	6	7	NA	9	13	70	NA	80	110
	Accuracy $\epsilon = 10^{-10}$:								
MVM	57	35	46	45	50	260	903	375	280
PrecS	28	33	33	45	55	240	380	375	350
CPU	12	17	NA	20	27	120	NA	190	210

Table 1
Comparison of PCG, LOBPCG, JDCG, and JDQR codes for DT=10^{-3}.

show how many times the matrix vector product Au and the preconditioner solve $Tu = f$, respectively, are performed in different methods. We also add results for the preconditioned conjugate gradient (PCG) linear solver for the system $Ax = b$ using the same preconditioner, where b is a pseudo random vector and the initial guess is simply the zero vector.

All tests are performed on a PIII 500Mhz computer with 400Mb of PC100 RAM, running MATLAB release 12 under MS Windows NT SP6. As flops count, used in Notay (2001), is no longer available in MATLAB release 12, it is replaced with CPU timing, in seconds, obtained by MATLAB's profiler. For the PCG tests, the built-in MATLAB PCG code is used. The incomplete Cholesky factorization is computed by the built-in MATLAB CHOLINC code.

Table 2 provides similar comparison of the revision 3.3 of the LOBPCG code with PCG, JDCG and JDQR for DT=10^{-4}.

Before we start discussing main results demonstrated on Tables 1 and 2, let us highlight the main improvements made in LOBPCG revision 3.3:

- In revision 3.2, some vectors in the basis of the trial subspace for the Rayleigh-Ritz method were not normalized, which resulted, when high accuracy was required, in instabilities due to badly scaled Gram matrices. This forced the method to perform extra orthogonalization with respect to the A-based scalar product, thus, increasing the MVM number, cf. the corresponding data for one eigenpair in Table 1. Without extra orthogonalization, each step of LOBPCG requires one MVM and one PrecS, see Knyazev (2001). In revision 3.3, an implicit normalization of all vectors in the basis of the trial subspace for the Rayleigh-Ritz method is implemented,

	$Ax = b$	One eigenpair computed by			Ten eigenpairs computed by		
	by PCG	3.3	JDCG	JDQR	3.3	JDCG	JDQR
	Accuracy $\epsilon = 10^{-5}$:						
MVM	15	10	13	19	100	115	100
PrecS	7	8	13	19	80	115	120
CPU	7	7	10	15	60	90	120
	Accuracy $\epsilon = 10^{-10}$:						
MVM	27	20	26	31	170	246	200
PrecS	13	18	26	34	150	246	260
CPU	10	15	20	26	120	200	250

Table 2
Comparison of PCG, LOBPCG, JDCG, and JDQR codes for DT=10^{-4}.

which increased stability and eliminated any need for extra orthogonalization in the problem tested, cf. 3.3 and 3.2 columns at the bottom of Table 1.

- LOBPCG iterates vectors simultaneously, similar to classical subspace iterations. In some cases different eigenpairs converge with different speed, see Figure 1 for the problems tested. In revision 3.2, the stopping criteria was such that simultaneous iterations were continually performed on *all* eigenpairs until the one with the worst speed was converged. This resulted in some eigenpairs computed much more accurately than the others in the final output, e.g., in the 10 vectors case with $\epsilon = 10^{-5}$ presented in the right upper corner of Table 1, some eigenpairs were in reality computed with accuracy 10^{-10}! Revision 3.3 freezes already converged vectors and excludes them from further iterations, which reduces significantly the total number of MVM and PrecS. This behavior is well illustrated on Figure 1 that presents convergence history in LOBPCG revision 3.3 for different eigenpairs. Here, the smallest eigenpairs converge much faster and get frozen when they reach the required accuracy level. We note, however, that the frozen eigenpairs in revision 3.3 still participate actively in the Rayleigh-Ritz procedure, thus, they do get improved in the process of iterations, e.g., in the test with ten eigenpairs and $\epsilon = 10^{-10}$, presented at the bottom of Table 2 and on the right picture on Figure 1 the singular value decomposition of $AU^{(i)} - U^{(i)}\Lambda^{(i)}$ of the final output reveals singular values ranging from 10^{-11} to 10^{-13}.
- Finally, more attention is paid in revision 3.3 to eliminate redundant algebraic computations in the code, which somewhat decreases the costs of every iteration outside of the MVM and PrecS.

We first notice that the LOBPCG code revision 3.3 converges with essentially the same speed as the linear solver PCG, especially for $\epsilon = 10^{-5}$, in both

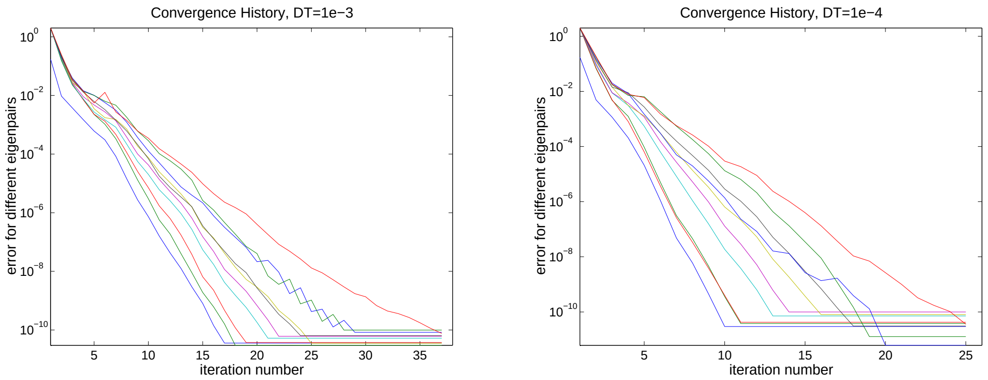

Fig. 1. LOBPCG convergence history.

tables. PCG does not involve computing as many scalar products and linear combinations as in LOBPCG, which leads to a better CPU time for PCG. We note that the number of MVM is artificially doubled in PCG, because the second MVM is performed on every step in the code only to compute the actual residual.

However, comparing an eigensolver, which finds the smallest eigenpair of the matrix A, to a linear solver, which solves the system $Ax = b$, cannot be possibly accurate, simply because the convergence of the eigensolver depends on the gap between the smallest eigenvalue λ_1 and the next one, while the convergence of the linear solver does not. A more precise comparison of the eigensolver, according to Knyazev (2001), is with an iterative solver, which finds a nontrivial solution of the homogeneous equation $(A - \lambda_1 I)u = 0$.

We provide such comparison for both choices of the preconditioner on Figure 2, using a code PCGNULL, described in Knyazev (2001), that is a trivial modification of the MATLAB built-in PCG code. We take the value of λ_1 from the LOBPCG run. The same initial guess, simply a vector with all components equal to one, is used in LOBPCG and PCGNULL.

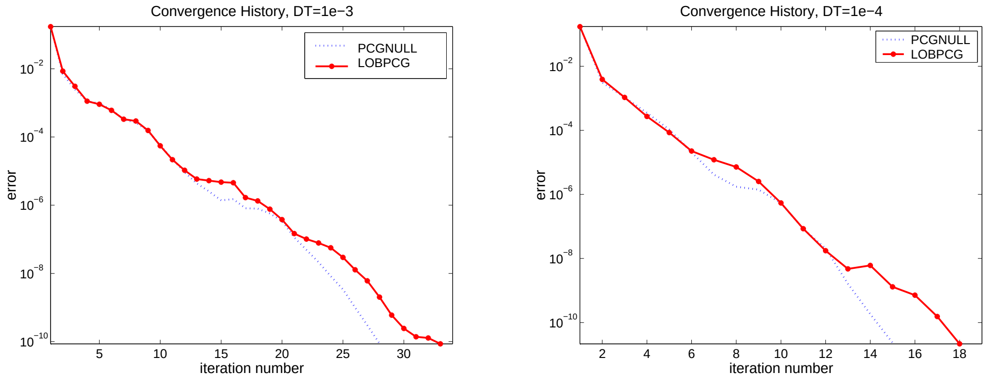

Fig. 2. LOBPCG vs. PCGNULL

We observe on Figure 2 not just a similar convergence speed but a striking correspondence of the error history lines. There is no an adequate explanation of such a correspondence and it remains a subject of the current research.

We are now prepared to compare the LOBPCG code revision 3.3 with JDCG and JDQR. Most importantly, LOBPCG is always faster, in numbers of MVM and PrecS, and in a raw CPU time. A faster convergence of the LOBPCG evidently turns into even bigger advantage in terms of the CPU time when a preconditioner solve gets more expensive, as we observe by comparing Table 1 with Table 2.

This is no big surprise as far as JDQR is concerned, because JDQR is a general code that works in the nonsymmetric case, too. The JDCG is, however, a specially tuned, for the symmetric case, version of the JDQR. The JDCG has much fewer, compare to the LOBPCG, algebraic overheads, according to numerical results of Notay (2001), as it does not include the Rayleigh-Ritz procedure and orthogonalization is performed in the standard Euclidean geometry. The problem tested is especially beneficial for the JDCG, because MVM is so inexpensive and the mass matrix is identity.

Let us remind the reader that the LOBPCG code is written for generalized eigenproblems, thus, even when the mass matrix is identity, such a code will be more expensive compare to a code for $Au = \lambda u$. No JDCG, or JDQR code is publicly available for generalized eigenproblems.

The fact that JDCG is slower in our tests than the LOBPCG could be attributed to a common devil of all outer-inner iterative solvers, like JDCG: no matter how smart a strategy is used to determine the number of inner iterations, one cannot match the performance of analogous methods without inner steps, like LOBPCG.

Despite of the fact that the revision 3.3 of LOBPCG computes eigenpairs simultaneously, dissimilar to JDCG and JDQR, which compute eigenpairs one by one, they all scale well with respect to the number of eigenpairs seeking. The number of PrecS and the CPU time for ten eigenpairs grow, compare to that for one eigenpair, no more than 10 times for all methods in all tests. We note that the chosen test problem does not have big clusters of eigenvalues. It might be expected that JDCG and JDQR would not perform as well in situations where many eigenvalues are close to each other, simply because JDCG and JDQR compute eigenvectors separately, while LOBPCG is specifically designed for clusters.

Availability of Software for the Preconditioned Eigensolvers

The Internet page
`http:// www-math.cudenver.edu/~aknyazev/software/CG/`
is maintained by the first author. It contains, in particular, the MATLAB code LOBPCG used for numerical experiments of the present paper.

Conclusion

We derive a short and sharp convergence rate estimate for basic preconditioned eigensolvers. The analysis presented here should increase understanding and provide tools for investigation of more efficient preconditioned eigensolvers, such as LOBPCG Knyazev (2001), under development. Our numerical tests support the main thesis of Knyazev (2001) that LOBPCG is, perhaps, a practically optimal preconditioned eigensolver for symmetric eigenproblems.

References

Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H. (Eds.), 2000. Templates for the solution of algebraic eigenvalue problems. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Basermann, A., 2000. Parallel block ILUT/ILDLT preconditioning for sparse eigenproblems and sparse linear systems. Numer. Linear Algebra Appl. 7 (7-8), 635–648, Preconditioning techniques for large sparse matrix problems in industrial applications (Minneapolis, MN, 1999).

Bergamaschi, L., Pini, G., Sartoretto, F., 2000. Approximate inverse preconditioning in the parallel solution of sparse eigenproblems. Numer. Linear Algebra Appl. 7 (3), 99–116.

Bradbury, W., Fletcher, R., 1966. New iterative methods for solution of the eigenproblem. Numer. Math. 9, 259–267.

Bramble, J., Pasciak, J., Knyazev, A., 1996. A subspace preconditioning algorithm for eigenvector/eigenvalue computation. Adv. Comput. Math. 6, 159–189.

Dobson, D. C., 1999. An efficient method for band structure calculations in 2D photonic crystals. J. Comput. Phys. 149 (2), 363–376.

Dobson, D. C., Gopalakrishnan, J., Pasciak, J. E., 2000. An efficient method for band structure calculations in 3D photonic crystals. J. Comput. Phys. 161 (2), 668–679.

D'yakonov, E., 1983. Iteration methods in eigenvalue problems. Math. Notes 34, 945–953.

D'yakonov, E., 1996. Optimization in solving elliptic problems. CRC Press, Boca Raton, Florida.

D'yakonov, E., Orekhov, M., 1980. Minimization of the computational labor in determining the first eigenvalues of differential operators. Math. Notes 27, 382–391.

Fattebert, J., Bernholc, J., 2000. Towards grid-based O(N) density-functional theory methods: Optimi zed nonorthogonal orbitals and multigrid acceleration. PHYSICAL REVIEW B (Condensed Matter and Materials Physics) 62 (3), 1713–1722.

Feng, Y., Owen, D., 1996. Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems. Int. J. Numer. Meth. Engrg. 39, 2209–2229.

Godunov, S., Ogneva, V., Prokopov, G., 1976. On the convergence of the modified method of steepest descent in the calculation of eigenvalues. Amer. Math. Soc. Transl. Ser. 2 105, 111–116.

Hestenes, M., Karush, W., 1951. A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix. J. Res. Nat. Bureau Standards 47, 45–61.

Kantorovich, L. V., 1952. Functional analysis and applied mathematics. Transl. of Uspehi Mat. Nauk 3 (1949), no.6 (28), 89–185, National Bureau of Standards Report, Washington.

Knyazev, A. V., 1986. Computation of eigenvalues and eigenvectors for mesh problems: algorithms and error estimates. Dept. Numerical Math. USSR Academy of Sciences, Moscow, (In Russian).

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

Knyazev, A. V., 1991. A preconditioned conjugate gradient method for eigenvalue problems and its implementation in a subspace. In: International Ser. Numerical Mathematics, 96, Eigenwertaufgaben in Natur- und Ingenieurwissenschaften und ihre numerische Behandlung, Oberwolfach, 1990. Birkhäuser, Basel.

Knyazev, A. V., 1998. Preconditioned eigensolvers -an oxymoron? . Electron. Trans. Numer. Anal. 7, 104–123.

Knyazev, A. V., 2000. Preconditioned eigensolvers: practical algorithms. In: Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H. (Eds.), Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide. SIAM, Philadelphia, pp. 352–368, section 11.3. An extended version published as a technical report UCD-CCM 143, 1999, at the Center for Computational Mathematics, University of Colorado at Denver.

Knyazev, A. V., 2001. Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method. SIAM J. Sci. Comp. To appear.

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

Longsine, D., McCormick, S., 1980. Simultaneous Rayleigh–quotient minimization methods for $Ax = \lambda Bx$. Linear Algebra Appl. 34, 195–234.

McCormick, S., Noe, T., 1977. Simultaneous iteration for the matrix eigenvalue problem. Linear Algebra Appl. 16, 43–56.

Morgan, R. B., 2000. Preconditioning eigenvalues and some comparison of solvers. J. Comput. Appl. Math. 123 (1-2), 101–115, Numerical analysis 2000, Vol. III. Linear algebra.

Morgan, R. B., Scott, D. S., 1993. Preconditioning the Lanczos algorithm for of sparse symmetric eigenvalue problems. SIAM J. Sci. Comput. 14 (3), 585–593.

Neymeyr, K., 2000. A geometric theory for preconditioned inverse iteration applied to a subspace. Accepted for publication in Math. Comput.

Neymeyr, K., 2001a. A geometric theory for preconditioned inverse iteration. I: Extrema of the Rayleigh quotient. Linear Algebra Appl. 322, 61–85.

Neymeyr, K., 2001b. A geometric theory for preconditioned inverse iteration. II: Convergence estimates. Linear Algebra Appl. 322, 87–104.

Ng, M. K., 2000. Preconditioned Lanczos methods for the minimum eigenvalue of a symmetric positive definite Toeplitz matrix. SIAM J. Sci. Comput. 21 (6), 1973–1986 (electronic).

Notay, Y., 2001. Combination of Jacobi-Davidson and conjugate gradients for the pa rtial symmetric eigenproblem. Numer. Lin. Alg. Appl. To appear.

Oliveira, S., 1999. On the convergence rate of a preconditioned subspace eigensolver. Computing 63 (3), 219–231.

Ovtchinnikov, E., Xanthis, L., 2001. Successive eigenvalue relaxation: a new method for generalized eigenvalue problems and convergence estimates. Proc. R. Soc. Lond. A 457, 441–451.

Ovtchinnikov, E. E., Xanthis, L. S., 2000. Effective dimensional reduction algorithm for eigenvalue problems for thin elastic structures: A paradigm in three dimensions. Proc. Natl. Acad. Sci. USA 97 (3), 967–971.

Petryshyn, W., 1968. On the eigenvalue problem $Tu - \lambda Su = 0$ with unbounded and non–symmetric operators T and S. Philos. Trans. Roy. Soc. Math. Phys. Sci. 262, 413–458.

Rodrigue, G., 1973. A gradient method for the matrix eigenvalue problem $Ax = \lambda Bx$. Numer. Math. 22, 1–16.

Sadkane, M., Sidje, R. B., 1999. Implementation of a variable block Davidson method with deflation for solving large sparse eigenproblems. Numer. Algorithms 20 (2-3), 217–240.

Sameh, A., Tong, Z., 2000. The trace minimization method for the symmetric generalized eigenvalue problem. J. Comput. Appl. Math. 123 (1-2), 155–175, numerical analysis 2000, Vol. III. Linear algebra.

Samokish, B., 1958. The steepest descent method for an eigenvalue problem with semi–bounded operators. Izv. Vyssh. Uchebn. Zaved. Mat. 5, 105–114.

Scott, D. S., 1981. Solving sparse symmetric generalized eigenvalue problems without factorization. SIAM J. Numer. Analysis 18, 102–110.

Sleijpen, G. L. G., Van der Vorst, H. A., 1996. A Jacobi-Davidson iteration method for linear eigenvalue problems. SIAM J. Matrix Anal. Appl. 17 (2), 401–425.

Smit, P., Paardekooper, M. H. C., 1999. The effects of inexact solvers in algorithms for symmetric eigenvalue problems. Linear Algebra Appl. 287 (1-3), 337–357, Special issue celebrating the 60th birthday of Ludwig Elsner.

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

Zhang, T., Golub, G. H., Law, K. H., 1999. Subspace iterative methods for eigenvalue problems. Linear Algebra Appl. 294 (1-3), 239–258.