

## A GEOMETRIC THEORY FOR PRECONDITIONED INVERSE ITERATION APPLIED TO A SUBSPACE

KLAUS NEYMEYR

**ABSTRACT.** The aim of this paper is to provide a convergence analysis for a preconditioned subspace iteration, which is designated to determine a modest number of the smallest eigenvalues and its corresponding invariant subspace of eigenvectors of a large, symmetric positive definite matrix. The algorithm is built upon a subspace implementation of preconditioned inverse iteration, i.e., the well-known inverse iteration procedure, where the associated system of linear equations is solved approximately by using a preconditioner. This step is followed by a Rayleigh–Ritz projection so that preconditioned inverse iteration is always applied to the Ritz vectors of the actual subspace of approximate eigenvectors. The given theory provides sharp convergence estimates for the Ritz values and is mainly built on arguments exploiting the geometry underlying preconditioned inverse iteration.

### 1. INTRODUCTION

Consider the problem to determine a modest number of the smallest eigenvalues together with its invariant subspace of eigenvectors of a large, symmetric positive definite matrix  $A$ . The eigenvalues of  $A \in \mathbf{R}^{n \times n}$  may have arbitrary multiplicity and are denoted in a way that  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ . To be more precise, we are interested in the first  $s$  eigenvalues and the corresponding eigenvectors. Hence, we assume that  $\lambda_s < \lambda_{s+1}$  in order to make the corresponding invariant subspace unique.

There are numerous techniques to solve the given problem (see [10]). In this work we focus on a subspace implementation of the method of preconditioned inverse iteration. Hence our method is attractive for those operators which are mesh analogs of differential operators with possibly multiple eigenvalues and for which also reliable (multigrid) preconditioners are known. In the following we summarize the idea of inverse iteration and show how to apply a preconditioner.

Inverse iteration [2, 11, 12, 17] maps a given vector  $x$  to  $x'$  by

$$(1.1) \quad A\hat{x} = \kappa x, \quad x' = \frac{\hat{x}}{|\hat{x}|},$$

where an additional nonzero constant  $\kappa$  is introduced which is usually considered to be equal to 1 ( $|\cdot|$  denotes the Euclidean norm). It is a well-known fact that inverse

---

Received by the editor July 27, 1999.

2000 *Mathematics Subject Classification.* Primary 65N30, 65N25; Secondary 65F10, 65F15.

*Key words and phrases.* Symmetric eigenvalue problem, subspace iteration, preconditioning, multigrid, inverse iteration.

iteration converges to an eigenvector  $z$  corresponding to the smallest eigenvalue if the acute angle between the initial vector and  $z$  is unequal to  $\frac{\pi}{2}$  (see [17]). The new iterate  $x'$  in (1.1) does not depend on the choice of the scaling constant  $\kappa$  provided that  $\kappa \neq 0$ . Thus we make the choice  $\kappa = \lambda(x)$ , where

$$(1.2) \quad \lambda(x) = \frac{(x, Ax)}{(x, x)}$$

denotes the Rayleigh quotient of  $x$  and  $(\cdot, \cdot)$  is the Euclidean inner product. This choice of  $\kappa$  has the effect that  $\hat{x} - x$  converges to the null vector if  $\lambda(\hat{x})$  converges to  $\lambda_1$  and hence provides the basis for the application of a preconditioner to solve approximately the system of linear equations (1.1).

A preconditioner  $B^{-1}$  for  $A$  is a matrix which approximates the inverse of  $A$  in a way that the spectral radius of the error propagation matrix  $I - B^{-1}A$  is bounded. For our purposes we assume that there is a real constant  $\gamma \in [0, 1[$  so that

$$(1.3) \quad \|I - B^{-1}A\|_A \leq \gamma,$$

where  $\|\cdot\|_A$  denotes the operator norm induced by  $A$ .

The determination of  $\hat{x}$  in (1.1) requires the solution of a system of linear equations in  $A$ . Approximate solution of this system by using a preconditioner  $B^{-1}$  for  $A$  leads to the error propagation equation (with  $\kappa = \lambda(x)$ )

$$(1.4) \quad \tilde{x} - \lambda(x)A^{-1}x = (I - B^{-1}A)(x - \lambda(x)A^{-1}x),$$

where  $\tilde{x}$  is an approximation for  $\hat{x}$ . We define for the rest of the paper  $\lambda = \lambda(x)$ . Hence, the new iterate  $\tilde{x}$  is given by

$$(1.5) \quad \tilde{x} = \lambda A^{-1}x + (I - B^{-1}A)(x - \lambda A^{-1}x)$$

or by its equivalent representation (containing no inverse of  $A$ )

$$(1.6) \quad \tilde{x} = x - B^{-1}(Ax - \lambda x).$$

Since the iterative scheme (1.5) directly derives from inverse iteration (INVIT), it is referred to as preconditioned inverse iteration (PINVIT), while its second representation (1.6) is usually associated with preconditioned gradient methods for the eigenvalue problem. The latter naming relies on the fact that the gradient of the Rayleigh quotient (1.2) is collinear to the residual  $Ax - \lambda x$ . Thus the actual iterate  $x$  in (1.6) is corrected in the direction of the negative preconditioned gradient of the Rayleigh quotient. Preconditioned gradient methods for the eigenvalue problem were first studied by Samokish [19] and later by Petryshyn [18]. Estimates on the convergence rate were given by Godunov et. al. [9] and D'yakonov et. al. [4, 8]. See Knyazev for a recent survey on preconditioned eigensolvers [13].

The viewpoint of preconditioned gradient methods for the convergence analysis of PINVIT seems to be less than optimal, since the convergence estimates are not sharp and some assumptions on the Rayleigh quotient of the initial vector have to be made [4, 8].

The exploitation of the structure of equation (1.5), which represents  $\tilde{x}$  as the result of (scaled) inverse iteration plus a perturbation by the product of the error propagation matrix and the vector  $x - \lambda A^{-1}x$ , results in sharp convergence estimates for the Rayleigh quotients of the iterates and in a clear description of the geometry of preconditioned inverse iteration [15, 16]. Moreover, it becomes to apparent that PINVIT essentially behaves like inverse iteration. This means that the convergence estimates for PINVIT take their extremal values in the same

vectors as those of INVIT and that convergence of both methods is shown under similar assumptions on the initial vectors. The theory in [15, 16] is mainly based on an analysis of extremal properties of certain quantities defining the geometry of PINVIT.

We come now to the question of how to generalize the vector iterations of INVIT and PINVIT to subspace algorithms. One generalization of INVIT is the so-called subspace iteration (also known as block inverse power method) with Rayleigh–Ritz projections (see for instance Section 1.4 in Parlett [17]). We next describe this algorithm and then discuss its modification in order to apply preconditioning.

Therefore, let  $\mathcal{V} = \text{span}\{v_1, \dots, v_s\}$  be a given  $s$ -dimensional subspace of the  $\mathbf{R}^n$  spanned by the vectors  $v_i$ ,  $i = 1, \dots, s$ , which are assumed to be the Ritz vectors of  $A$ . Hence, for  $V = [v_1, \dots, v_s] \in \mathbf{R}^{n \times s}$  holds

$$(1.7) \quad V^T A V = \Theta = \text{diag}(\theta_1, \dots, \theta_s) \quad \text{and} \quad V^T V = I,$$

where  $I \in \mathbf{R}^{s \times s}$  is the identity matrix. The Ritz vectors are in an order that the Ritz values  $\theta_i$  increase with  $i$ .

Subspace iteration to determine an invariant subspace of  $A$  corresponding to some of the smallest eigenvalues is a straightforward generalization of inverse iteration. The new subspace  $\hat{\mathcal{V}}$  results from applying  $A^{-1}$  to  $\mathcal{V}$ , i.e.,  $\hat{\mathcal{V}} = A^{-1}\mathcal{V}$ . Using the matrix notation, the column space of  $\hat{V}$ , defined by

$$(1.8) \quad A\hat{V} = VD,$$

is equal to  $\hat{\mathcal{V}}$ , where  $D \in \mathbf{R}^{s \times s}$  in (1.8) denotes a diagonal matrix with nonzero diagonal elements which gives rise to a scaling of the columns of  $V$ . The column space  $\hat{\mathcal{V}} = \text{span}(\hat{V})$  does not depend on  $D$  so that  $D$  is usually considered to be equal to the identity matrix. Convergence of subspace iteration to the  $A$ -invariant subspace spanned by the  $s$  eigenvectors to the eigenvalues  $\lambda_1, \dots, \lambda_s$  is guaranteed [17] if the initial subspace is not orthogonal to any of the  $s$  eigenvectors of  $A$  to the  $s$  smallest eigenvalues and  $\lambda_s < \lambda_{s+1}$  (to make the subspace unique).

Now we describe how to apply a preconditioner to solve the equation (1.8) approximately. First we make the choice  $D = \Theta$  in (1.8), where  $\Theta$  is the diagonal matrix of the Ritz values, expecting that  $\hat{V} - V$  converges to the null matrix if the subspace algorithm converges to a subspace of eigenvectors of  $A$ . Thus in analogy to (1.5) and (1.6) the subspace implementation of PINVIT is given by

$$(1.9) \quad \tilde{V} = A^{-1}V\Theta + (I - B^{-1}A)(V - A^{-1}V\Theta).$$

For the preconditioner we assume (1.3) again. The simplified form of equation (1.9) (containing no inverse of  $A$ ) reads

$$(1.10) \quad \tilde{V} = V - B^{-1}(AV - V\Theta).$$

To determine the approximate eigenvectors and eigenvalues, we now apply the Rayleigh–Ritz procedure. The Ritz vectors  $v'_i$ ,  $i = 1, \dots, s$ , of the column space  $\text{span}(\tilde{V})$  define the columns of  $V'$  and the corresponding Ritz values  $\theta'_i$ ,  $i = 1, \dots, s$ , are the diagonal elements of the matrix  $\Theta'$ . The preconditioned subspace algorithm iterates the transformation  $V, \Theta \rightarrow V', \Theta'$ .

The preconditioned iterative scheme in the form (1.10) was recently analyzed by Bramble et al. [1], where a survey on various attempts to analyze this and alternative (simplified) preconditioned subspace schemes is also given. One such simplification is that in equation (1.10): instead of the matrix  $\Theta$ , a constant diagonal matrix is

considered in order to make the convergence analysis feasible (see [5, 6, 7]). Usually the main difficulty for the convergence analysis of (1.10) is seen in the dependence of the iteration operator (which acts in (1.10) on the columns of  $V$ ) on the Ritz values  $\theta_i$ . Here we do not consider these simplified schemes. We further note that the subspace implementation of PINVIT can be understood as a modified version of a block Davidson method [3], in a way that the dimension of the iteration subspace is not modified in the course of the iteration.

In the analysis of Bramble et al. [1] very restrictive conditions on the initial subspace are assumed to be satisfied, which are far from being fulfilled in practice. The analysis presented here is applicable to initial subspaces generated from scratch. Moreover, our estimates for the convergence of the Ritz values are sharp in a sense that an initial subspace and a preconditioner can be constructed so that the convergence estimates are attained. The given analysis shows that the convergence estimates for each Ritz value are the same as those which are derived for the Rayleigh quotient of the iterates of PINVIT. Hence the subspace implementation of PINVIT, in the following referred to as SPINVIT, essentially behaves like the classical subspace iteration (block inverse power method).

The rest of this paper is organized as follows. In the next section we give a more detailed description of PINVIT, state its central convergence theorem and prove a monotonicity lemma. In Section 3 the convergence analysis of SPINVIT is given. Lemma 3.1 proves that SPINVIT preserves the dimension of the subspace in the course of the iteration. Theorem 3.2 provides an estimate from above for the largest Ritz value. Finally, the central Theorem 3.3 contains sharp convergence estimates for all Ritz values.

## 2. PRECONDITIONED INVERSE ITERATION (PINVIT)

In this section we recapitulate those facts and results from the convergence analysis of the preconditioned inverse iteration method (see [15, 16]), which are needed here for the analysis of its subspace implementation. Furthermore, we prove some monotonicity of the convergence estimates.

The iterative scheme of PINVIT has the two representations (1.5) and (1.6). The first one points out the relation to the method of inverse iteration and turns out as a valuable tool for the analysis. Obviously,  $\tilde{x}$  is always computed by evaluation of (1.6). In practice the new iterate  $\tilde{x}$  is normed to 1, but theoretically the convergence does not depend on the scaling of  $x$ , since preconditioned inverse iteration is homogeneous with respect to scaling of  $x$ .

The convergence theory of PINVIT in [15, 16] is mainly based on an analysis of the geometry of PINVIT with respect to the  $A$ -orthonormal basis of eigenvectors of  $A$ . To introduce this basis, let  $X \in \mathbf{R}^{n \times n}$  be the orthogonal matrix containing in its columns the eigenvectors of  $A$ , so that  $X^T A X = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  and  $X^T X = I$ . Furthermore, let  $c$  ( $\tilde{c}$ ) be the coefficient vectors of  $x$  ( $\tilde{x}$ ) with respect to this basis so that  $x = X \Lambda^{-1/2} c$  and  $\tilde{x} = X \Lambda^{-1/2} \tilde{c}$ . Then for any iterate  $\tilde{x}$  (1.5) there is a  $\tilde{\gamma}$  (with  $0 \leq \tilde{\gamma} \leq \gamma$  where  $\gamma$  is determined by (1.3)) and a Householder reflection  $H = I - 2vv^T$  (with  $v \in \mathbf{R}^n$  and  $|v|^2 = v^T v = 1$ ), so that

$$(2.1) \quad \tilde{c} = \lambda_\Lambda(c) \Lambda^{-1} c - \tilde{\gamma} H (I - \lambda_\Lambda(c) \Lambda^{-1}) c$$

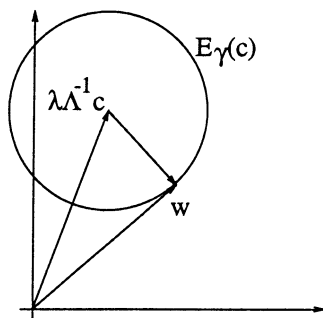


FIGURE 1. The set  $E_\gamma(c)$  with the center  $\lambda\Lambda^{-1}c$  and the radius  $|\lambda\Lambda^{-1}c - w|$ .

(see Lemma 2.4 in [15]). Therein  $\lambda_\Lambda(\cdot)$  denotes the Rayleigh quotient within this basis which for a nonzero  $d \in \mathbf{R}^n$  reads

$$(2.2) \quad \lambda_\Lambda(d) := \frac{(d, d)}{(d, \Lambda^{-1}d)}.$$

Then we have  $\lambda = \lambda(x) = \lambda_\Lambda(c)$ . For given  $c$  and  $\gamma$  the set of all iterates  $\tilde{c}$  resulting from (2.1) is a ball denoted by  $E_\gamma(c)$

$$(2.3) \quad E_\gamma(c) := \{\lambda\Lambda^{-1}c + d; d \in \mathbf{R}^n, |d| \leq \gamma|(I - \lambda\Lambda^{-1})c|\},$$

with the radius  $\gamma|(I - \lambda\Lambda^{-1})c|$  and the center  $\lambda\Lambda^{-1}c$ ;  $|\cdot|$  denotes the Euclidean norm. Equivalently,  $E_\gamma(c)$  results by applying PINVIT, as given by equation (1.6) for all preconditioners  $B^{-1}$  satisfying (1.3), to the vector  $x$  and subsequent transformation of all iterates to the  $A$ -orthonormal basis of eigenvectors of  $A$  (see [15]).

Therefore, the problem of deriving convergence estimates for the Rayleigh quotient of the iterates of PINVIT is reduced to the problem of analyzing the extremal behavior of the Rayleigh quotient (2.2) on the ball  $E_\gamma(c) \subset \mathbf{R}^n$ . Here, we are particularly interested in suprema of the Rayleigh quotient on  $E_\gamma(c)$  in order to analyze the case of poorest convergence. In [15] it is shown that this supremum is attained in a point  $w$  of the form

$$(2.4) \quad w = \beta(\alpha + \Lambda)^{-1}c,$$

and that  $w$  is an element of the boundary of  $E_\gamma(c)$ . The real constants  $\alpha \geq 0$  and  $\beta$  in (2.4) are unique. Figure 1 illustrates the position of  $w$  in  $E_\gamma(c)$ . It is a somewhat surprising result that the point of a supremum on  $E_\gamma(c)$  can be represented in the form (2.4), i.e., by inverse iteration with a positive shift  $\alpha$  if applied to  $c$ . Furthermore, similiar properties hold for the infimum of the Rayleigh quotient on  $E_\gamma(c)$ , which describes the best possible convergence of PINVIT.

The central convergence theorem for PINVIT reads as follows (see [16]).

**Theorem 2.1.** *Let a vector  $c \in \mathbf{R}^n$  with  $|c| = 1$  and  $\gamma \in [0, 1[$  be given. Let  $\lambda = \lambda_\Lambda(c)$  be the Rayleigh quotient of  $c$ . Then the Rayleigh quotient  $\tilde{\lambda} = \lambda_\Lambda(\tilde{c})$  of the new iterate  $\tilde{c}$  of PINVIT (2.1) can be estimated from above, in order to describe the case of poorest convergence, in the following way:*

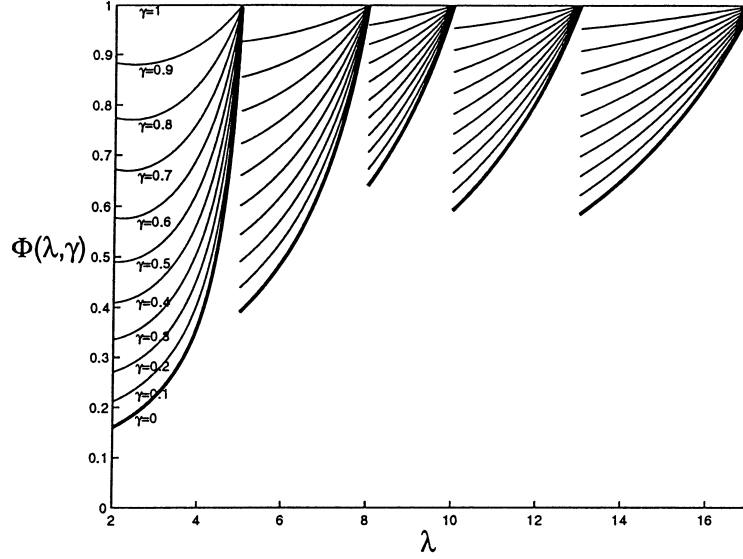


FIGURE 2. Convergence estimates  $\Phi_{i,i+1}(\lambda, \gamma)$  in the interval  $[\lambda_i, \lambda_{i+1}[$  for  $i = 1, \dots, 5$ . The curves are shown for 11 values of  $\gamma = 0, 0.1, \dots, 1.0$  against  $\lambda \in [2, 17]$  for the example system with the eigenvalues 2, 5, 8, 10, 13 and 17.

- (a) If  $\lambda = \lambda_i$ ,  $i = 1, \dots, n$ , then  $\tilde{\lambda}$  takes its maximum  $\tilde{\lambda} = \lambda_i$  if  $c$  is collinear to the  $i$ th unit vector  $e_i$ . (In this case  $\text{PINVIT}$  is stationary or equivalently  $X\Lambda^{-1/2}e_i$  is collinear to the  $i$ th eigenvector of  $A$ .)
- (b) If  $\lambda_i < \lambda < \lambda_{i+1}$ , then the maximal Rayleigh quotient on the set  $E_\gamma(c)$  takes its maximal value under variation of  $c$ , with  $\lambda_\Lambda(c) = \lambda$  and  $|c| = 1$ , in a vector of the form

$$c_{i,i+1} = (0, \dots, 0, c_i, c_{i+1}, 0, \dots, 0)^T,$$

with  $\lambda(c_{i,i+1}) = \lambda$ . Furthermore, we explicitly have  $\tilde{\lambda} = \lambda_{i,i+1}(\lambda, \gamma)$  with

$$(2.5) \quad \begin{aligned} \lambda_{i,j}(\lambda, \gamma) = & \lambda\lambda_i\lambda_j(\lambda_i + \lambda_j - \lambda)^2 / (\gamma^2(\lambda_j - \lambda)(\lambda - \lambda_i)(\lambda\lambda_j + \lambda\lambda_i - \lambda_i^2 - \lambda_j^2) \\ & - 2\gamma\sqrt{\lambda_i\lambda_j}(\lambda - \lambda_i)(\lambda_j - \lambda)\sqrt{\lambda_i\lambda_j + (1 - \gamma^2)(\lambda - \lambda_i)(\lambda_j - \lambda)} \\ & - \lambda(\lambda_i + \lambda_j - \lambda)(\lambda\lambda_j + \lambda\lambda_i - \lambda_i^2 - \lambda_j^2)). \end{aligned}$$

*Proof.* See Theorem 1.1 in [16]. □

It is not easy to see that  $\lambda_{i,i+1}(\lambda, \gamma) \leq \lambda$ . Lemma A.1 in Appendix A gives a crude estimate from above:

$$(2.6) \quad \lambda_{i,i+1}(\lambda, \gamma) \leq \lambda - (1 - \gamma)^2 \frac{(\lambda - \lambda_i)(\lambda_{i+1} - \lambda)}{\lambda_{i+1}}.$$

Figure 2 illustrates the convergence estimates  $\Phi_{i,i+1}(\lambda, \gamma)$  (describing the relative decrease to the nearest eigenvalue  $\lambda_i$  less than  $\lambda$ )

$$(2.7) \quad \Phi_{i,i+1}(\lambda, \gamma) := \frac{\lambda_{i,i+1}(\lambda, \gamma) - \lambda_i}{\lambda - \lambda_i}.$$

for a matrix having the eigenvalues 2, 5, 8, 10, 13 and 17. In each interval  $[\lambda_i, \lambda_{i+1}[$  the estimate  $\Phi_{i,i+1}(\lambda, \gamma)$  is shown for 11 values of  $\gamma$ , i.e.,  $\gamma = 0, 0.1, \dots, 1.0$ . The estimate for  $\gamma = 0$ , which describes the poorest convergence of inverse iteration, corresponds to the bold curves. To derive this estimate explicitly, we insert  $\gamma = 0$  and  $j = i + 1$  in (2.5) and obtain

$$(2.8) \quad \lambda_{i,i+1}(\lambda, 0) = (\lambda_i^{-1} + \lambda_{i+1}^{-1} - (\lambda_i + \lambda_{i+1} - \lambda)^{-1})^{-1}.$$

Hence  $\Phi_{i,i+1}(\lambda, 0)$  in the interval  $[\lambda_i, \lambda_{i+1}[$  reads

$$\Phi_{i,i+1}(\lambda, 0) = \frac{\lambda_i^2}{\lambda_i^2 + (\lambda_{i+1} - \lambda)(\lambda_i + \lambda_{i+1})},$$

which is a convergence estimate from above for inverse iteration if applied to a given vector whose Rayleigh quotient is equal to  $\lambda$  (see Section 1 in [16]). For  $\lambda = \lambda_{i+1}$  we have  $\Phi_{i,i+1}(\lambda_{i+1}, \gamma) = 1$ , which expresses the fact that PINVIT is stationary in the eigenvectors of  $A$ . The curves in Figure 2 for  $\gamma > 0$  describe the case of poorest convergence of PINVIT. In each interval  $[\lambda_i, \lambda_{i+1}[$  poorest convergence of INVIT is attained in those vectors which are spanned by the eigenvectors corresponding to  $\lambda_i$  and  $\lambda_{i+1}$ . In the same case the poorest convergence of PINVIT is attained if additionally the preconditioner is chosen appropriately [15]. For a random initial vector with a Rayleigh quotient  $\lambda > \lambda_2$ , it cannot be guaranteed that PINVIT (and in the same way INVIT) converges to an eigenvector corresponding to the smallest eigenvalue of  $A$ , since both methods may reach stationarity in the orthogonal complement of the eigenvectors to the smallest eigenvalue. In practice this situation is very unlikely due to rounding errors.

It is worth noting that PINVIT, depending on the eigenvector expansion of the actual iterate and on the choice of the preconditioner, may converge much more rapidly as suggested by the worst case estimate (2.5).

In preparation of our central convergence Theorem 3.3, the next lemma shows that for fixed  $\gamma$  the function  $\lambda_{i,j}(\lambda, \gamma)$  is strictly monotone increasing in  $\lambda$ .

**Lemma 2.2.** *Let  $\tilde{\lambda}, \lambda \in ]\lambda_1, \lambda_n[$  with  $\tilde{\lambda} < \lambda$  be given for which indexes  $p$  and  $q$  can be determined so that  $\tilde{\lambda} \in [\lambda_p, \lambda_{p+1}[$  and  $\lambda \in [\lambda_q, \lambda_{q+1}[$ . Then it holds that*

$$\lambda_{p,p+1}(\tilde{\lambda}, \gamma) < \lambda_{q,q+1}(\lambda, \gamma).$$

*Proof.* Adopting the notation of Theorem 2.1 we have

$$\lambda_{i,j}(\lambda, \gamma) = \sup\{\lambda_\Lambda(z); z \in E_\gamma(c_{i,j})\},$$

where the vector  $c_{i,j} =: c$  has exactly two nonzero components  $c_i$  and  $c_j$  whose absolute values are uniquely determined by  $|c| = 1$  and  $\lambda_\Lambda(c) = \lambda$ . Moreover, for  $0 \leq \gamma_1 \leq \gamma_2 \leq 1$  we have  $E_{\gamma_1}(c) \subseteq E_{\gamma_2}(c)$ . Thus we conclude that the function  $\lambda_{i,j}(\lambda, \gamma)$  is monotone increasing in  $\gamma$ . Inserting  $\gamma = 1$  in (2.5) leads to  $\lambda_{i,j}(\lambda, 1) = \lambda$  and for  $\gamma = 0$ , one obtains

$$\lambda_i \leq \lambda_{i,j}(\lambda, 0) = (\lambda_i^{-1} + \lambda_j^{-1} - (\lambda_i + \lambda_j - \lambda)^{-1})^{-1}.$$

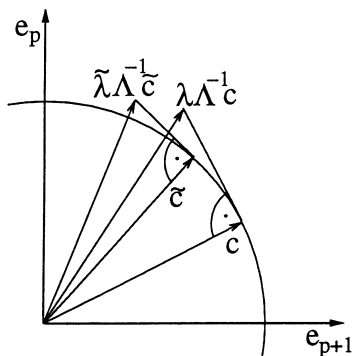


FIGURE 3. Two vectors  $c, \tilde{c} \in \text{span}\{e_p, e_{p+1}\}$  with  $\lambda_\Lambda(\tilde{c}) < \lambda_\Lambda(c)$  and the result of inverse iteration.

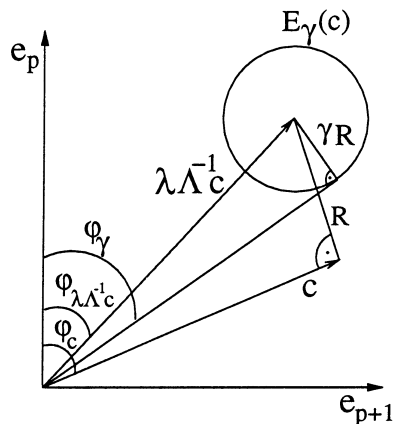


FIGURE 4. The angle  $\varphi_\gamma$  which is the largest angle enclosed by any vector of  $E_\gamma(c)$  and the axis  $e_p$ .

We first treat the case  $\lambda_p < \lambda_q$  so that  $p < q$ . Using the preceding arguments we conclude

$$\lambda_{p,p+1}(\tilde{\lambda}, \gamma) \leq \lambda_{p,p+1}(\tilde{\lambda}, 1) = \tilde{\lambda} < \lambda_{p+1} \leq \lambda_q \leq \lambda_{q,q+1}(\lambda, 0) \leq \lambda_{q,q+1}(\lambda, \gamma).$$

Next we treat  $\lambda_p = \lambda_q$  or  $p = q$ . Thus we have to show the monotonicity of  $\lambda_{p,p+1}(\lambda, \gamma)$  in  $\lambda$  for  $\lambda \in [\lambda_p, \lambda_{p+1}[$ . This can be done by analyzing the geometry of PINVIT within the plane, which is spanned by the unit vectors  $e_p$  and  $e_{p+1}$ . This “mini-dimensional” analysis is justified by Theorem 1.1 in [16].

We first observe that for  $z \in \text{span}\{e_p, e_{p+1}\}$  the Rayleigh quotient  $\lambda(z)$  is a monotone increasing function in the acute angle  $\varphi_z$  enclosed by the axis  $e_p$  and  $z$ . Inserting  $z$  in (2.2) one obtains

$$\lambda_\Lambda(z) = \frac{1 + z_p^2/z_{p+1}^2}{\lambda_p^{-1} + \lambda_{p+1}^{-1} z_p^2/z_{p+1}^2}.$$

Differentiation with respect to  $\tan(\varphi_z) = \frac{z_{p+1}}{z_p}$  provides the required result.

Now let  $c, \tilde{c} \in \text{span}\{e_p, e_{p+1}\}$  with  $|c| = |\tilde{c}| = 1$  and  $\lambda_\Lambda(\tilde{c}) = \tilde{\lambda} < \lambda_\Lambda(c) = \lambda$ . The given situation is shown in Figure 4. Then for the acute angles enclosed by these vectors and the axis  $e_p$  holds that

$$\varphi_{\tilde{c}} < \varphi_c.$$

Furthermore, application of inverse iteration to  $c$  and  $\tilde{c}$  and subsequent evaluation of the Rayleigh quotients leads to

$$\varphi_{\tilde{\lambda}\Lambda^{-1}\tilde{c}} < \varphi_{\lambda\Lambda^{-1}c},$$

since by (2.8) the Rayleigh quotient  $\lambda_{p,p+1}(\lambda, 0)$  is monotone increasing in  $\lambda$  (see Figure 3). We conclude that  $\varphi_c$  and  $\varphi_{\lambda\Lambda^{-1}c}$  are monotone increasing in  $\lambda$ .

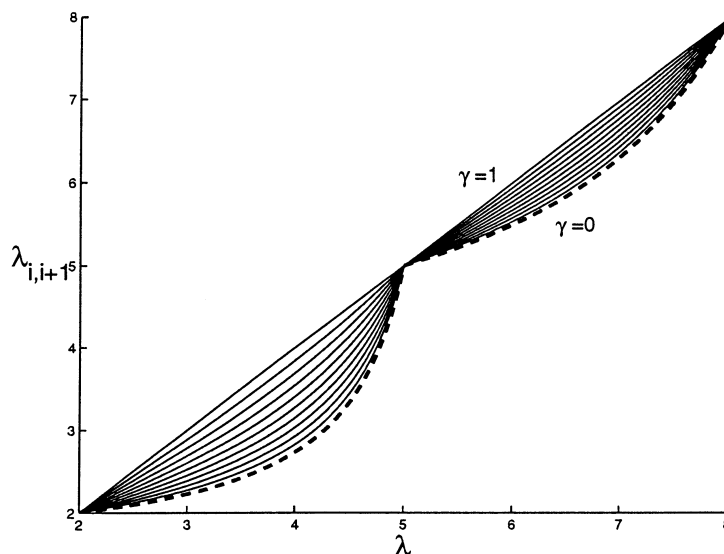


FIGURE 5. The Rayleigh quotient  $\lambda_{i,i+1}(\lambda, \gamma)$ ,  $i = 1, \dots, 3$ , as given by (2.5) against  $\lambda \in [2, 8]$  for the example system  $\Lambda = \text{diag}(2, 5, 8)$ . The 11 curves correspond to  $\gamma = 0, 0.1, \dots, 1.0$ . The dashed curve stands for  $\gamma = 0$ . For  $\gamma = 1$  we have  $\lambda = \lambda_{i,i+1}(\lambda, \gamma)$  or stationarity of PINVIT.

Since

$$(2.9) \quad w = \arg \max_{w \in E_\gamma(c)} \angle(e_p, w)$$

maximizes the Rayleigh quotient on  $E_\gamma(c)$  we only have to show that the acute angle enclosed by  $e_p$  and  $w$

$$\varphi_\gamma := \max_{w \in E_\gamma(c)} \angle(e_p, w)$$

is a monotone increasing function of  $\lambda$ . Therefore let  $\varphi_c = \angle(e_p, c)$  and  $\varphi_{\lambda\Lambda^{-1}c} = \angle(e_p, \lambda\Lambda^{-1}c)$  so that  $\varphi_c - \varphi_{\lambda\Lambda^{-1}c}$  for  $\gamma = 1$  is the opening angle of the cone

$$C_\gamma(c) := \{\zeta d; d \in E_\gamma(c), \zeta > 0\},$$

of all positive multiples of the vectors in  $E_\gamma(c)$ . The geometric setup is shown in Figure 4.

Applying the orthogonal decomposition

$$|c|^2 + |c - \lambda\Lambda^{-1}c|^2 = |\lambda\Lambda^{-1}c|^2,$$

we have  $\tan(\varphi_c - \varphi_{\lambda\Lambda^{-1}c}) = R$  with  $R^2 = |\lambda\Lambda^{-1}c|^2 - |c|^2$ . Hence

$$\varphi_\gamma = \varphi_{\lambda\Lambda^{-1}c} + \arcsin\left(\frac{\gamma R}{\sqrt{1 + R^2}}\right) = \varphi_{\lambda\Lambda^{-1}c} + \arcsin(\gamma \sin(\varphi_c - \varphi_{\lambda\Lambda^{-1}c})).$$

Applying Lemma A.2 from Appendix A for  $\alpha = \varphi_c$ ,  $\hat{\alpha} = \varphi_{\tilde{c}}$ ,  $\beta = \varphi_{\lambda\Lambda^{-1}c}$  and  $\hat{\beta} = \varphi_{\tilde{\lambda}\Lambda^{-1}\tilde{c}}$  completes the proof, since now  $\varphi_\gamma$ ,  $\varphi_c$  and  $\varphi_{\lambda\Lambda^{-1}c}$  are all monotone increasing in  $\lambda$  so that  $\lambda(w)$  ( $w$  by (2.9)) increases in  $\lambda$  too.  $\square$

Figure 5 illustrates the content of Lemma 2.2 by using a subsystem of the example shown in Figure 2, i.e.,  $\Lambda = \text{diag}(2, 5, 8)$ . For  $\gamma = 0, 0.1, \dots, 1.0$ , the strictly monotone increasing function  $\lambda_{i,i+1}(\lambda, \gamma)$  is plotted against  $\lambda \in [2, 8]$ . The dashed curve corresponds to  $\gamma = 0$ .

### 3. SUBSPACE IMPLEMENTATION OF PINVIT (SPINVIT)

We start the presentation of the convergence theory of SPINVIT by proving a certain  $A$ -orthogonal decomposition which immediately implies that the dimension of the iteration subspace of SPINVIT is preserved, i.e.,  $\text{rank}(\tilde{V}) = s$ . Then in Theorem 3.2 an estimate for the largest Ritz value is established. Finally, Theorem 3.3 contains sharp convergence estimates for all Ritz values.

For the rest of the paper PINVIT and SPINVIT are represented with respect to the initial basis, i.e., the  $A$ -orthogonal basis of eigenvectors of  $A$ , as used in Section 2, is not considered furthermore.

**Lemma 3.1.** *Let  $V \in \mathbf{R}^{n \times s}$  be the matrix of the Ritz vectors of  $A$  with  $s = \text{rank}(V)$ , i.e., the dimension of the linear space spanned by the columns of  $V$ , and let  $\Theta = V^T A V$  be the diagonal matrix of the Ritz values. Then it holds that*

$$(3.1) \quad (a) \quad V^T A(V - A^{-1}V\Theta) = 0 \in \mathbf{R}^{s \times s},$$

$$(3.2) \quad (b) \quad (A^{-1}V\Theta)^T A(A^{-1}V\Theta) = \\ V^T A V + (V - A^{-1}V\Theta)^T A(V - A^{-1}V\Theta),$$

$$(3.3) \quad (c) \quad \text{rank}(\tilde{V}) = s, \quad \text{for } \tilde{V} \text{ by (1.9)}.$$

*Proof.* Property (a) follows from the definition of the Ritz vectors and Ritz values (see equation (1.7)). Furthermore, by applying (a) one obtains

$$\begin{aligned} & V^T A V + (V - A^{-1}V\Theta)^T A(V - A^{-1}V\Theta) \\ &= \Theta - (A^{-1}V\Theta)^T A(V - A^{-1}V\Theta) = (A^{-1}V\Theta)^T A(A^{-1}V\Theta). \end{aligned}$$

Finally, we show  $\|\tilde{V}y\|_A > 0$ ,  $\tilde{V}$  by (1.9), for all nonzero vectors  $y \in \mathbf{R}^s$ :

$$\begin{aligned} \|\tilde{V}y\|_A &= \|A^{-1}V\Theta y + (I - B^{-1}A)(V - A^{-1}V\Theta)y\|_A \\ &\geq \|A^{-1}V\Theta y\|_A - \|(I - B^{-1}A)(V - A^{-1}V\Theta)y\|_A \\ &\geq \|A^{-1}V\Theta y\|_A - \|(V - A^{-1}V\Theta)y\|_A \\ &= \frac{\|A^{-1}V\Theta y\|_A^2 - \|(V - A^{-1}V\Theta)y\|_A^2}{\|A^{-1}V\Theta y\|_A + \|(V - A^{-1}V\Theta)y\|_A} \\ &= \frac{\|Vy\|_A^2}{\|A^{-1}V\Theta y\|_A + \|(V - A^{-1}V\Theta)y\|_A} > 0. \end{aligned}$$

The last inequality holds, since by  $\text{rank}(V) = s$  we have  $\|Vy\|_A > 0$  and by regularity of  $A$  also  $\|A^{-1}V\Theta y\|_A > 0$ . Hence,  $\text{rank}(\tilde{V}) = s$ .  $\square$

Figure 6 illustrates the  $A$ -orthogonal decomposition within the  $\mathbf{R}^{n \times s}$  as presented in Lemma 3.1.

The next theorem provides an estimate for the maximal Ritz value  $\theta'_s$  of  $A$  with respect to  $\tilde{V}$ . The theorem simply says that the worst case estimate for the largest Ritz value is the same as that of PINVIT, i.e., the estimate for the decrease of the largest Ritz value coincides with that which results from Theorem 2.1 if PINVIT is applied to an initial vector with the Rayleigh quotient  $\theta_s$ . The estimate (3.4) in

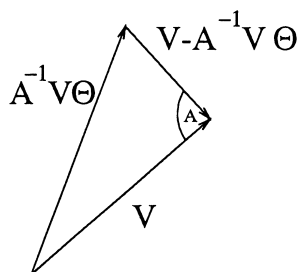


FIGURE 6.  $A$ -orthogonal decomposition in the  $\mathbf{R}^{n \times s}$ .

Theorem 3.2 is sharp (the proof is given later in Theorem 3.3) in a sense that an initial subspace  $V$  and a preconditioner can be constructed so that the estimate is attained.

**Theorem 3.2.** *Let  $\theta_s$  be the largest Ritz value of  $A$  with respect to  $V$  and let  $p$  be the index so that  $\theta_s \in [\lambda_p, \lambda_{p+1}[$ . For the maximal Ritz value  $\theta'_s$  of  $A$  with respect to the new subspace  $\tilde{V}$  generated by SPINVIT, as defined by equation (1.9) or (1.10),*

$$(3.4) \quad \theta'_s \leq \lambda_{p,p+1}(\theta_s, \gamma)$$

holds. Therein the function  $\lambda_{p,p+1}$  is given by (2.5).

*Proof.* We use the minmax characterization of the Ritz values (see [17]), which says

$$(3.5) \quad \theta'_s = \max_{y \neq 0} \lambda(\tilde{V}y).$$

From (1.9) for any  $y$  the vector  $\tilde{V}y$  is given by

$$(3.6) \quad \begin{aligned} \tilde{V}y &= A^{-1}V\Theta y + (I - B^{-1}A)(V - A^{-1}V\Theta)y \\ &= \lambda(z)A^{-1}z + (I - B^{-1}A)(Vy - \lambda(z)A^{-1}z), \end{aligned}$$

where in the last equation  $z = \lambda(V\Theta y)^{-1}V\Theta y$  is introduced for which we have  $\lambda(z) = \lambda(V\Theta y)$ .

To find an estimate from above for  $\theta'_s$  we define the set  $\hat{F}_\gamma(y)$

$$\hat{F}_\gamma(y) := \{\lambda(z)A^{-1}z + (I - B^{-1}A)(Vy - \lambda(z)A^{-1}z); \|I - B^{-1}A\|_A \leq \gamma\},$$

of all obtainable iterates which result from application of all admissible preconditioners satisfying (1.3). Since  $\tilde{V}y \in \hat{F}_\gamma(y)$ , we can apply the Rayleigh quotient (1.2), so that  $\lambda(\tilde{V}y)$  is less or equal than the maximum on the set  $\lambda(\hat{F}_\gamma(y))$ , i.e.,

$$\lambda(\tilde{V}y) \leq \max \lambda \left( \hat{F}_\gamma(y) \right).$$

The set's maximum  $\max \lambda(\hat{F}_\gamma(y))$  depends on  $y$ . Taking a further maximum with respect to  $y$  and applying (3.5) leads to

$$(3.7) \quad \theta'_s = \max_{y \neq 0} \lambda(\tilde{V}y) \leq \max_{y \neq 0} \max \lambda \left( \hat{F}_\gamma(y) \right).$$

What we have analyzed so far is the convergence of PINVIT, as given by equation (1.5), if applied to  $z$

$$\tilde{z} = \lambda(z)A^{-1}z + (I - B^{-1}A)(z - \lambda(z)A^{-1}z),$$

whose iterates, for all admissible preconditioners, define the set  $F_\gamma(z)$

$$F_\gamma(z) := \{\lambda(z)A^{-1}z + (I - B^{-1}A)(z - \lambda(z)A^{-1}z); \|I - B^{-1}A\|_A \leq \gamma\}.$$

Next we show that  $\hat{F}_\gamma(y) \subseteq F_\gamma(z[y])$ , where  $z[y] = \lambda(V\Theta y)^{-1}V\Theta y$ . Now let  $w \in \hat{F}_\gamma(y)$ . From Lemma 2.2 in [15] a preconditioner  $B_1$  with  $\|I - B_1^{-1}A\|_A \leq \gamma$  in the form

$$(3.8) \quad B_1^{-1} = A^{-1} + \gamma_1 A^{-1/2} H_1 A^{-1/2}$$

can be constructed so that

$$w = \lambda(z)A^{-1}z + (I - B_1^{-1}A)(Vy - \lambda(z)A^{-1}z)$$

( $H_1$  is a Householder matrix and  $0 \leq \gamma_1 \leq \gamma$ ). To show  $w \in F_\gamma(z[y])$  we take a second preconditioner  $B_2 = A^{-1} + \gamma_2 A^{-1/2} H_2 A^{-1/2}$ , having the same form as (3.8) but with different  $H_2$  and  $0 < \gamma_2 < \gamma$ , which has to satisfy

$$(3.9) \quad (I - B_1^{-1}A)t_1 = (I - B_2^{-1}A)t_2,$$

where  $t_1 = Vy - \lambda(z)A^{-1}z$  and  $t_2 = z - \lambda(z)A^{-1}z$ . From (3.9) we obtain

$$(3.10) \quad \gamma_1 A^{-1/2} H_1 A^{1/2} t_1 = \gamma_2 A^{-1/2} H_2 A^{1/2} t_2.$$

Multiplication with  $A^{1/2}$  and application of the Euclidean norm results in

$$\gamma_1 \|t_1\|_A = \gamma_2 \|t_2\|_A$$

for the norm induced by  $A$ . For given  $t_1$ ,  $H_1$  and  $\gamma_1 \leq \gamma$  we can easily determine a Householder matrix  $H_2$  so that the vectors  $H_1 A^{1/2} t_1$  and  $H_2 A^{1/2} t_2$  are collinear. Furthermore, we can also determine a constant  $\gamma_2$ , with  $0 < \gamma_2 < \gamma$ , so that (3.10) is satisfied, since  $\|t_2\|_A > \|t_1\|_A$  as shown in the following.

By Cauchy's inequality we have for any  $y \in \mathbf{R}^s$

$$(y, \Theta^2 y)^2 \leq (y, \Theta y)(y, \Theta^3 y),$$

since the Ritz values  $\theta_i$ ,  $i = 1, \dots, s$  are positive. Hence

$$(y, \Theta y) + \frac{(y, \Theta^2 y)^2}{(y, \Theta^3 y)^2} (y, \Theta^3 y) \leq 2(y, \Theta y),$$

from which with  $\vartheta = \lambda(V\Theta y) = \frac{(y, \Theta^3 y)}{(y, \Theta^2 y)}$  the estimate

$$(y, \Theta y) + \vartheta^{-2} (\Theta y, V^T A V \Theta y) \leq 2(y, V^T V \Theta y)$$

follows. With  $z = \vartheta^{-1} V \Theta y$  one obtains

$$(Vy, AVy) + (z, Az) \leq 2(Vy, \lambda(z)A^{-1}z)_A,$$

or equivalently

$$\|Vy\|_A^2 - 2(Vy, \lambda(z)A^{-1}z)_A \leq \|z\|_A^2 - 2(z, \lambda(z)A^{-1}z)_A.$$

From this we conclude

$$\|t_1\|_A^2 = \|Vy - \lambda(z)A^{-1}z\|_A^2 \leq \|z - \lambda(z)A^{-1}z\|_A^2 = \|t_2\|_A^2.$$

Hence we have  $\hat{F}_\gamma(y) \subseteq F_\gamma(z[y])$  and thus for the maximal Rayleigh quotient

$$(3.11) \quad \max \lambda \left( \hat{F}_\gamma(y) \right) \leq \max \lambda \left( F_\gamma(z[y]) \right).$$

If one takes the maximum with respect to  $y$  on the right hand side of (3.11), then on the left hand side this will lead to

$$(3.12) \quad \max_{y \neq 0} \max \lambda \left( \hat{F}_\gamma(y) \right) \leq \max_{y \neq 0} \max \lambda (F_\gamma(z[y])).$$

The monotonicity lemma (Lemma 2.2) says that the convergence estimate for any  $V\Theta y$  with  $y \neq 0$ , because of  $\lambda(V\Theta y) \leq \theta_s$ , is dominated by  $\lambda_{p,p+1}(\theta_s, \gamma)$ , i.e.,

$$(3.13) \quad \max_{y \neq 0} \lambda_{k,k+1}(\lambda(V\Theta y), \gamma) \leq \lambda_{p,p+1}(\theta_s, \gamma).$$

Therein, the index  $k$  depends on  $y$  in a way that  $\lambda(V\Theta y) \in [\lambda_k, \lambda_{k+1}[$ . We combine the estimates (3.7), (3.12), Theorem 2.1, and (3.13) to obtain

$$\theta'_s \leq \max_{y \neq 0} \max \lambda (F_\gamma(z[y])) \leq \max_{y \neq 0} \lambda_{k,k+1}(\lambda(V\Theta y), \gamma) \leq \lambda_{p,p+1}(\theta_s, \gamma).$$

□

In the next theorem, which is the central convergence theorem for SPINVIT, sharp estimates from above for each Ritz value are given. Theorem 3.3 says that for any of the  $s$  Ritz values an estimate for the decrease of the Rayleigh quotient, like inequality (3.4), holds (see Theorem 3.2). In other words, each Ritz value behaves like the Rayleigh quotient in the preconditioned vector scheme PINVIT: convergence of PINVIT to the smallest eigenvalue  $\lambda_1$  is only guaranteed if the Rayleigh quotient of the initial vector is less than  $\lambda_2$ ; if the last condition is not fulfilled, the very unlikely situation may occur that PINVIT converges to an eigenvalue larger or equal to  $\lambda_2$  so that the found eigenvector is located in the orthogonal complement of the eigenspace corresponding to  $\lambda_1$ . But in practice, as a result of rounding errors, PINVIT always converges from scratch to the smallest eigenvalue and a corresponding eigenvector.

For SPINVIT we have a comparable situation. If the column space of  $V_0 \in \mathbf{R}^{n \times s}$  describes a given initial space, then an exact arithmetic convergence of the smallest Ritz value  $\theta_1$  to the eigenvalue  $\lambda_1$  can only be guaranteed if the Ritz value  $\theta_1(V_0)$  with respect to  $V_0$  is less than  $\lambda_2$ ; convergence of the second Ritz value to  $\lambda_2$  is guaranteed if  $\theta_2(V_0) < \lambda_3$  and so on. These assumptions seem to be very restrictive, but similar assumptions have to be made to prove convergence of the subspace implementation of INVIT. If the smallest Ritz value of  $V_0$  is equal or less than  $\lambda_2$ , then the column space  $\text{span}(V_0)$  is possibly orthogonal to the eigenspace to the smallest eigenvalue  $\lambda_1$ . Hence in exact arithmetic, inverse iteration does not converge to the smallest eigenvalue and its invariant subspace of  $A$ . It is important to note that it is well believed that in practice due to rounding errors the subspace implementation of INVIT converges from scratch to the eigenspace associated with the  $s$  smallest eigenvalues. In the same way SPINVIT will converge from scratch, as already observed in [1].

**Theorem 3.3.** *Let  $\mathcal{V} = \text{span}\{v_1, \dots, v_s\}$  be a given  $s$ -dimensional subspace of the  $\mathbf{R}^n$ , let  $V \in \mathbf{R}^{n \times s}$  be the matrix which contains in its columns the Ritz vectors of  $A$  with respect to  $\mathcal{V}$ , and let  $\Theta$  be the diagonal matrix of the Ritz values (see equation (1.7)). Let indexes  $k_i$  be given in a way that  $\theta_i \in [\lambda_{k_i}, \lambda_{k_i+1}[$ . Moreover, let  $\theta'_1 \leq \dots \leq \theta'_s$  be the Ritz values of  $\tilde{V}$ , which results from application of SPINVIT, by (1.9) or (1.10), to  $V$ . Then for  $i = 1, \dots, s$ ,*

$$(3.14) \quad \theta'_i \leq \lambda_{k_i, k_i+1}(\theta_i, \gamma)$$

holds. If the subspace dimension  $s$  is less than  $n$  (otherwise the Ritz values coincide with the eigenvalues), then the estimate (3.14) is sharp in a sense that for each Ritz value  $\theta_i$  an initial space  $V$  and a preconditioner  $B^{-1}$  (depending on  $\theta_i$ ) can be constructed so that (3.14) is attained. (The estimate is not always sharp for all Ritz values at the same time.)

*Proof.* The proof is given by induction on  $s$ . For a one-dimensional subspace, or  $s = 1$ , SPINVIT is the same as PINVIT so that by Theorem 2.1

$$\theta'_1 \leq \lambda_{p,p+1}(\theta_1, \gamma),$$

where  $\theta_1 \in [\lambda_p, \lambda_{p+1}[$ .

For  $V = [v_1, \dots, v_s] \in \mathbf{R}^{n \times s}$ , consisting of  $s$  Ritz vectors of  $A$ , by deleting its last column we define

$$V^{(s-1)} := [v_1, \dots, v_{s-1}].$$

The columns of  $V^{(s-1)}$  remain to be Ritz vectors of  $A$ . The corresponding diagonal matrix containing the Ritz values reads  $\Theta^{(s-1)} = \text{diag}(\theta_1, \dots, \theta_{s-1})$ .

Application of SPINVIT to  $V^{(s-1)}$  and  $\Theta^{(s-1)}$  results in

$$\tilde{V}^{(s-1)} = V^{(s-1)} - B^{-1} \left( AV^{(s-1)} - V^{(s-1)}\Theta^{(s-1)} \right).$$

For the Ritz values  $\theta'_i(\tilde{V}^{(s-1)})$  of  $A$  with respect to  $\tilde{V}^{(s-1)}$  by the induction hypothesis,

$$\theta'_i(\tilde{V}^{(s-1)}) \leq \lambda_{k_i, k_i+1}(\theta_i, \gamma), \quad i = 1, \dots, s-1,$$

holds.

Applying SPINVIT to  $V = [v_1, \dots, v_s] = [V^{(s-1)}, v_s]$  we obtain

$$\tilde{V} = [\tilde{V}^{(s-1)}, \tilde{v}_s],$$

since by equation (1.9) the last column  $\tilde{v}_s$  has no influence on the previous columns. Let  $\tilde{\mathcal{V}}^{(s-1)}$  the column space of  $\tilde{V}^{(s-1)}$  and  $\tilde{\mathcal{V}}$  be the column space of the enlarged matrix  $\tilde{V}$ . By the minmax characterization of the Ritz values the first  $s-1$  Ritz values decrease while expanding  $\tilde{\mathcal{V}}^{(s-1)}$  to  $\tilde{\mathcal{V}}$ . Hence for  $i = 1, \dots, s-1$ ,

$$\theta'_i(\tilde{V}^{(s-1)}) = \min_{\mathcal{V}_i \leq \tilde{\mathcal{V}}^{(s-1)}} \max_{x \in \mathcal{V}_i \setminus \{0\}} \lambda(x) \geq \min_{\mathcal{V}_i \leq \tilde{\mathcal{V}}} \max_{x \in \mathcal{V}_i \setminus \{0\}} \lambda(x) = \theta'_i(\tilde{V}),$$

where the minimum is taken over all  $i$ -dimensional subspaces denoted by  $\mathcal{V}_i$ . For the remaining Ritz value  $\theta_s$ , Theorem 3.2 provides the required result.

To show that the estimate (3.14) is sharp, let a particular  $i$ ,  $1 \leq i \leq s$ , with  $\theta_i \in [\lambda_p, \lambda_{p+1}[$  be given for which we now construct an initial matrix  $V \in \mathbf{R}^{n \times s}$  and a preconditioner  $B$  so that

$$\theta'_i = \lambda_{p,p+1}(\theta_i, \gamma).$$

Let  $x_j$ ,  $j = 1, \dots, n$ , be the eigenvectors of  $A$  with  $|x_j| = 1$  and  $\lambda_j = (x_j, Ax_j)$ . Let the first column  $v_1$  of  $V$  be given by

$$v_1 = \sqrt{\frac{\lambda_{p+1} - \theta}{\lambda_{p+1} - \lambda_p}} x_p + \sqrt{\frac{\theta - \lambda_p}{\lambda_{p+1} - \lambda_p}} x_{p+1}.$$

Then  $|v_1| = 1$  and  $\lambda(v_1) = \theta$ . Since by Lemma B.1 in Appendix B,  $\lambda_i \leq \theta_i \leq \lambda_{i+(n-s)}$ , we can fill up the remaining columns of  $V$  with those  $s-1$  eigenvectors of  $A$ , that are orthogonal to  $x_p$  and  $x_{p+1}$ , in a way that we take exactly  $i-1$  of the eigenvectors of  $A$  to eigenvalues less or equal to  $\lambda_p$  and  $s-i$  eigenvectors to

eigenvalues larger or equal to  $\lambda_{p+1}$ . Then  $V$  is an orthogonal matrix and  $\theta_i$  is the  $i$ th Ritz vector of  $A$  with respect to  $V$ .

Furthermore, for this choice of  $V$  the residual matrix  $AV - V\Theta$  is the zero matrix with exception of the first column. The preconditioner  $B$  is taken in the form

$$B^{-1} = A^{-1} + \gamma A^{-1/2} H A^{-1/2},$$

where  $H = I - 2xx^T$  is a Householder matrix with  $x \in \mathbf{R}^n$ . Then  $\|I - B^{-1}A\|_A = \gamma$  and a vector  $x \in \text{span}\{x_p, x_{p+1}\}$  can be determined so that the Rayleigh quotient  $\lambda(\tilde{v}_1)$  with

$$\tilde{v}_1 = \theta_1 A^{-1} v_1 + (I - B^{-1}A)(v_1 - \theta_1 A^{-1} v_1)$$

takes its maximum

$$\lambda(\tilde{v}_1) = \lambda_{p,p+1}(\theta, \gamma)$$

with respect to any  $B^{-1}$  satisfying (1.3) (see Section 5 of [15]). It is also shown that  $\tilde{v}_1 \in \text{span}\{x_p, x_{p+1}\}$ , so that  $\tilde{v}_1$  is a Ritz vector of  $A$  with respect to  $\tilde{V}$ . Therefore SPINVIT, if applied to  $V$ , collapses to the vector iteration of PINVIT, which is applied to the single vector  $v_1$ , since all other columns of  $V$  are not modified. Hence from  $\lambda(\tilde{v}_1) = \lambda_{p,p+1}(\theta, \gamma) \in [\lambda_p, \lambda_{p+1}[$  the  $i$ th Ritz value with respect to  $\tilde{V}$  is guaranteed to be equal to  $\lambda(\tilde{v}_1)$ , since the other Ritz values remain stationary in the eigenvalues of  $A$ .  $\square$

**3.1. Convergence of the Ritz vectors.** So far we have not given any estimates on the convergence of the Ritz vectors generated by SPINVIT to the eigenvectors of  $A$ . The reason for this lack is to be seen in the fact that the acute angle between the  $i$ th Ritz vector and the  $i$ th eigenvector is not necessarily a monotone decreasing sequence (see Section 3.2 in [16] discussing the case  $s = 1$ , for which SPINVIT is the same as PINVIT).

But what can be done to prove convergence of the Ritz vectors toward the eigenvectors of  $A$ ? From Corollary 3.1 in [16] we obtain

$$(3.15) \quad \|Av_1 - \theta_1 v_1\|_{A^{-1}} \leq \left( \frac{\theta_1}{\lambda_1} - 1 \right)^{1/2}$$

as an estimate from above for the residual associated with the Ritz vector  $v_1$  and their corresponding Ritz value  $\theta_1$ . The simplest way to derive error estimates for Ritz vectors  $v_i$ ,  $i \geq 2$ , is to determine first the invariant subspace to the smallest eigenvalue and then to switch over to the orthogonal complement. Now one can apply the estimate (3.15) in the orthogonal complement and so on, moving toward the interior of the spectrum. We further note that it is not possible to bound the residual of the Ritz vectors  $v_i$ ,  $i > 1$ , by the difference  $\lambda_i - \theta_i$  without previous knowledge of the convergence of the Ritz vectors  $v_j$ ,  $j < i$ . It is easy to construct appropriate counterexamples.

#### 4. AN EXAMPLE

To demonstrate that the convergence rates for the Ritz values of SPINVIT are of comparable magnitude with that of multigrid methods for boundary value problems, we consider now the five-point finite difference discretization of the eigenproblem

TABLE 1. Convergence rate estimates for the 6 Ritz values  $\theta_i$  in the case of block inverse iteration, i.e.,  $\gamma = 0$ , and for  $\gamma = 0.2$  and  $\gamma = 0.8$ . The 6 smallest (exact) eigenvalues of the finite difference discretization with  $h = \pi/50$  are given by  $\lambda_i^h$ .

$i$	$\lambda_i^h$	$\theta_i$	$\Theta_{i,i+1}(\theta_i, 0)$	$\Theta_{i,i+1}(\theta_i, 0.2)$	$\Theta_{i,i+1}(\theta_i, 0.8)$
1	1.99934	3.5	0.277	0.376	0.804
2	4.99441	5.5	0.436	0.530	0.868
3	4.99441	6.5	0.563	0.642	0.905
4	7.98948	7.2	0.680	0.740	0.931
5	9.97305	10.3	0.619	0.688	0.917
6	9.97305	11.0	0.688	0.747	0.934

for the Laplacian on the square  $[0, \pi]^2$  with homogeneous Dirichlet boundary conditions. The eigenvalues of the continuous problem  $\lambda_{k,l}$  and of the finite difference discretization  $\lambda_{k,l}^h$ , for the mesh size  $h$ , are given by

$$\lambda_{k,l} = k^2 + l^2, \quad \lambda_{k,l}^h = \frac{4}{h^2} \left( \sin^2\left(\frac{kh}{2}\right) + \sin^2\left(\frac{lh}{2}\right) \right).$$

For  $h = \pi/50$  the 10 smallest eigenvalues (with multiplicity) read explicitly

$$\begin{aligned} \lambda_{1,\dots,10} &= (2, 5, 5, 8, 10, 10, 13, 13, 17, 17), \\ \lambda_{1,\dots,10}^h &= (1.99934, 4.99441, 4.99441, 7.98948, 9.97305, \\ &\quad 9.97305, 12.96812, 12.96812, 16.91563, 16.91563). \end{aligned}$$

Hence these eigenvalues  $\lambda_i$  and  $\lambda_i^h$  coincide within the 1 percent range. Figure 2 shows the convergence estimates  $\Phi_{i,i+1}(\lambda, \gamma)$  for the eigenvalues  $\lambda_i$ . Note that the estimates are valid independently of the multiplicity of the eigenvalues.

To illustrate the convergence rates  $\Theta_{i,i+1}(\lambda, \gamma)$ , which describe the convergence of  $\lambda$  to  $\lambda_i$  for some Ritz values of SPINVIT, we define

$$\Theta_{i,i+1}(\lambda, \gamma) := \max_{\lambda_i \leq \tilde{\lambda} \leq \lambda} \Phi_{i,i+1}(\tilde{\lambda}, \gamma)$$

(refer to Figure 2 to see that  $\Theta_{i,i+1}(\lambda, \gamma)$  only slightly differs from  $\Phi_{i,i+1}(\lambda, \gamma)$ ; for the given example both quantities coincide). The Ritz values  $\theta_i$  in the third column of Table 1 are given in a way that, if  $m$  is the multiplicity of the eigenvalue  $\lambda_i^h$ , then  $m$  Ritz values are located in the interval between  $\lambda_i^h$  and the nearest larger eigenvalue. In this situation SPINVIT is guaranteed to converge to the eigenvectors corresponding to the 6 smallest eigenvalues. The convergence rate estimates for these Ritz values are given in the last three columns of Table 1. The column  $\gamma = 0$  describes the case of block inverse iteration while the columns  $\gamma = 0.2$  and  $\gamma = 0.8$  correspond to SPINVIT.

It is worth noting that the convergence rate estimates do not depend on the mesh size  $h$  and hence on the number of the variables. To derive a crude estimate from above we insert equation (2.6) in (2.7) and obtain

$$\Phi_{i,i+1}(\lambda, \gamma) \leq 1 - (1 - \gamma)^2 \left(1 - \frac{\lambda}{\lambda_{i+1}}\right) = \frac{\lambda}{\lambda_{i+1}} + \gamma(2 - \gamma) \left(1 - \frac{\lambda}{\lambda_{i+1}}\right).$$

Since the right hand side is strictly monotone increasing in  $\lambda \in [\lambda_i, \lambda_{i+1}]$ , it is also an estimate for  $\Theta_{i,i+1}(\lambda, \gamma)$ . Therein the first term  $\frac{\lambda}{\lambda_{i+1}}$  describes the behavior of

inverse iteration, which converges to 1 for  $\lambda \rightarrow \lambda_{i+1}$  (this is also shown by the bold curves in Figure 2). The further term of the order  $O(\gamma)$  estimates the influence of the preconditioner and thus describes the influence of PINVIT.

Hence, depending on the quality of the preconditioner, eigenvector/eigenvalue computation can be done reliably with the SPINVIT algorithm. So SPINVIT can be viewed as the eigenproblem counterpart of multigrid algorithms for the solution of boundary value problems, for which preconditioners satisfying an estimate of the form (1.3) are known and which have optimal convergence properties.

### 5. CONCLUSION

A new theoretical framework for the method of preconditioned inverse iteration in a subspace (SPINVIT) has been presented. For the most part the convergence analysis is built on an analysis of the geometry underlying preconditioned inverse iteration. Sharp convergence estimates for each Ritz value, which are independent of the number of unknowns, have been given. These estimates coincide with those derived for the Rayleigh quotient in the vector iteration of PINVIT. Furthermore, SPINVIT turns out to behave comparably to the subspace implementation of inverse iteration; for  $\gamma = 0$  the methods coincide. Hence, SPINVIT can be viewed as a reliable method for determining some of the smallest eigenvalues and its corresponding eigenvectors of a large, symmetric positive definite matrix from scratch (as already observed by Bramble et al. [1]).

SPINVIT can also be embedded in an adaptive multigrid algorithm to solve eigenproblems for elliptic operators. Such a method and a posteriori error estimation for SPINVIT is the topic of [14].

### APPENDIX A. SOME AUXILIARY LEMMATA

The next lemma provides a crude estimate from above for the sharp convergence estimate (2.5) of PINVIT (see [8]).

**Lemma A.1.** *Let  $\lambda \in ]\lambda_i, \lambda_{i+1}[$  and  $\gamma \in [0, 1]$ . Then*

$$(A.1) \quad \lambda_{i,i+1}(\lambda, \gamma) \leq \lambda - (1 - \gamma)^2 \frac{(\lambda - \lambda_i)(\lambda_{i+1} - \lambda)}{\lambda_{i+1}}.$$

*Proof.* Inserting (1.5) in (1.2) one obtains by direct computation (with  $r = Ax - \lambda x$ ,  $d = B^{-1}r$  and  $|x| = 1$ )

$$\lambda - \lambda(\tilde{x}) = \frac{[2 - \|d\|_A^2 \|d\|_B^{-2}] \|d\|_B^2 + \lambda(d, d)}{1 - 2(x, d) + (d, d)}.$$

Estimating the nominator from below using (1.3) ( $\|\cdot\|_B$  is the norm induced by  $B$ )

$$[2 - \|d\|_A^2 \|d\|_B^{-2}] \|d\|_B^2 + \lambda(d, d) \geq (1 - \gamma) \|d\|_B^2 + \lambda(d, d),$$

and the denominator from above for  $\epsilon > 0$

$$1 - 2(x, d) + (d, d) \leq (1 + \epsilon) + (1 + \epsilon^{-1})(d, d),$$

leads to

$$\lambda - \lambda(\tilde{x}) \geq \frac{(1 - \gamma) \|d\|_B^2 + \lambda |d|^2}{(1 + \epsilon) + (1 + \epsilon^{-1}) |d|^2} \geq \min \left\{ \frac{1 - \gamma}{1 + \epsilon} \|d\|_B^2, \frac{\lambda}{1 + \epsilon^{-1}} \right\}.$$

The minimum takes its maximal value in  $\epsilon = \frac{1-\gamma}{\lambda} \|d\|_B^2$ . Hence,

$$(A.2) \quad \lambda - \lambda(\tilde{x}) \geq (1-\gamma)^2 \|r\|_{A^{-1}}^2 \frac{\lambda(\tilde{x})}{\lambda}.$$

The term  $\|r\|_{A^{-1}}^2$  can be estimated from below by Temple's inequality in the form

$$(A.3) \quad \|r\|_{A^{-1}}^2 \geq \frac{\lambda(\lambda - \lambda_i)(\lambda_{i+1} - \lambda)}{\lambda_i \lambda_{i+1}}$$

provided that  $(x, x) = 1$  and  $\lambda(x) \in [\lambda_i, \lambda_{i+1}[$ . Inserting (A.3) in (A.2) and estimating  $\lambda(\tilde{x}) > \lambda_i$ , one derives

$$\lambda - \lambda(\tilde{x}) \geq (1-\gamma)^2 \frac{(\lambda - \lambda_i)(\lambda_{i+1} - \lambda)}{\lambda_{i+1}}.$$

We finally note that  $\lambda_{i,i+1}(\lambda, \gamma)$  is dominated by  $\lambda(\tilde{x})$  so that (A.1) follows.  $\square$

For the proof of Lemma 2.2 the next lemma is required, which proves the monotonicity of some trigonometric function.

**Lemma A.2.** *Let  $\alpha, \tilde{\alpha}, \beta, \tilde{\beta} \in \mathbf{R}$  with*

$$(A.4) \quad 0 < \tilde{\alpha} < \alpha < \frac{\pi}{2}, \quad 0 < \tilde{\beta} < \beta < \frac{\pi}{2}, \quad \tilde{\beta} < \tilde{\alpha} \quad \text{and} \quad \beta < \alpha$$

*be given. Furthermore, define  $\phi(\alpha, \beta) := \beta + \arcsin(\gamma \sin(\alpha - \beta))$ . Then for all  $\gamma \in [0, 1[$*

$$\phi(\alpha, \beta) > \phi(\tilde{\alpha}, \tilde{\beta})$$

*holds.*

*Proof.* It suffices to show  $\frac{\partial}{\partial \alpha} \phi(\alpha, \beta) > 0$  and  $\frac{\partial}{\partial \beta} \phi(\alpha, \beta) > 0$  for  $\alpha, \beta$  satisfying (A.4). The first derivative reads

$$\frac{\partial}{\partial \alpha} \phi(\alpha, \beta) = \frac{\gamma \cos(\alpha - \beta)}{(1 - \gamma^2 \sin^2(\alpha - \beta))^{1/2}} > 0,$$

while the second derivative is given by

$$\frac{\partial}{\partial \beta} \phi(\alpha, \beta) = 1 - \left( \frac{\gamma^2 \cos^2(\alpha - \beta)}{1 - \gamma^2 \sin^2(\alpha - \beta)} \right)^{1/2}.$$

Since  $1 > \gamma^2$ , we obtain the required result.  $\square$

## APPENDIX B. AN INTERLACE LEMMA ON RITZ VALUES

The next lemma, which generalizes Cauchy's interlace theorem [17] to Ritz values, is required for the proof of Theorem 3.3.

**Lemma B.1.** *Let  $A \in \mathbf{R}^{n \times n}$  and orthogonal  $V \in \mathbf{R}^{n \times s}$  with  $V^T V = I \in \mathbf{R}^{s \times s}$  be given. Let the eigenvalues of  $A$  be given by  $\lambda_1 \leq \dots \leq \lambda_n$  and the Ritz values of  $A$  with respect to  $V$  by  $\theta_1 \leq \dots \leq \theta_s$ . Then it holds that*

$$\lambda_i \leq \theta_i \leq \lambda_{i+(n-s)}, \quad i = 1, \dots, s.$$

*Proof.* By the minmax characterization of the eigenvalues and of the Ritz values we have

$$\lambda_i = \min_{\mathcal{V}_i \leq \mathbf{R}^n} \max_{x \in \mathcal{V}_i \setminus \{0\}} \lambda(x) \leq \min_{\mathcal{V}_i \leq \text{span}(V)} \max_{x \in \mathcal{V}_i \setminus \{0\}} \lambda(x) = \theta_i,$$

where  $\mathcal{V}_i$  denotes an arbitrary subspace of dimension  $i$ . It remains to show  $\theta_{s-j} \leq \lambda_{n-j}$  for  $j = 0, \dots, s-1$ . Hence we have to show that the column space  $\text{span}(V)$  of  $V$  has an  $s-j$  dimensional subspace  $\mathcal{V}_{s-j}$  so that

$$(B.1) \quad \max_{x \in \mathcal{V}_{s-j} \setminus \{0\}} \lambda(x) \leq \lambda_{n-j}.$$

Therefore let  $x_{n-s+1}, \dots, x_n$  be the eigenvectors of  $A$  corresponding to the  $s$  eigenvalues  $\lambda_{n-s+1}, \dots, \lambda_n$ . We define  $j$  vectors  $y^{(n-l+1)} \in \mathbf{R}^s$ ,  $l = 1, \dots, j$ ,

$$y^{(n-l+1)} := ((v_1, x_{n-l+1}), \dots, (v_s, x_{n-l+1}))^T.$$

There are  $s-j$  orthogonal vectors  $a^{(1)}, \dots, a^{(s-j)} \in \mathbf{R}^s$  which are orthogonal to each of the vectors  $y^{(n-l+1)} \in \mathbf{R}^s$ ,  $l = 1, \dots, j$ . We now construct  $s-j$  orthogonal vectors in  $\text{span}(V)$

$$\hat{v}^{(k)} := Va^{(k)} \in \text{span}(V), \quad k = 1, \dots, s-j.$$

These vectors are elements of  $\text{span}\{x_1, \dots, x_{n-j}\}$ , since for  $l = n-j+1, \dots, n$  we have

$$(\hat{v}^{(k)}, x_l) = \sum_{i=1}^s a_i^{(k)}(v_i, x_l) = (a^{(k)}, y^{(l)}) = 0.$$

With the choice  $\mathcal{V}_{s-j} = \text{span}\{\hat{v}^{(1)}, \dots, \hat{v}^{(s-j)}\}$ , equation (B.1) is obviously satisfied.  $\square$

## REFERENCES

- [1] J.H. Bramble, J.E. Pasciak, and A.V. Knyazev. A subspace preconditioning algorithm for eigenvector/eigenvalue computation. *Adv. Comput. Math.*, 6:159–189, 1996. MR **98c**:65057
- [2] F. Chatelin. *Eigenvalues of matrices*. Wiley, Chichester, 1993. MR **94d**:65002
- [3] E.R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.*, 17:87–94, 1975. MR **52**:2168
- [4] E.G. D'yakonov. Iteration methods in eigenvalue problems. *Math. Notes*, 34:945–953, 1983. MR **86d**:65048
- [5] E.G. D'yakonov. *Optimization in solving elliptic problems*. CRC Press, Boca Raton, Florida, 1996. MR **97e**:65004
- [6] E.G. D'yakonov and A.V. Knyazev. Group iterative method for finding lower-order eigenvalues. *Moscow Univ. Comput. Math. Cybern.*, 2:32–40, 1982. MR **84a**:65030
- [7] E.G. D'yakonov and A.V. Knyazev. On an iterative method for finding lower eigenvalues. *Russian J. Numer. Anal. Math. Modelling*, 7(6):473–486, 1992. MR **94f**:65035
- [8] E.G. D'yakonov and M.Y. Orekhov. Minimization of the computational labor in determining the first eigenvalues of differential operators. *Math. Notes*, 27:382–391, 1980. MR **82a**:65086
- [9] S.K. Godunov, V.V. Ogneva, and G.P. Prokopov. On the convergence of the modified method of steepest descent in the calculation of eigenvalues. *Amer. Math. Soc. Transl. Ser. 2*, 105:111–116, 1976. MR **48**:3235
- [10] G.H. Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, MD, 3rd edition, 1996. MR **97g**:65006
- [11] I. Ipsen. *A history of inverse iteration*, volume in Helmut Wielandt, *Mathematische Werke*, Mathematical Works, Vol. 2: Linear Algebra and Analysis, pages 464–472. Walter de Gruyter, Berlin, 1996.
- [12] I. Ipsen. Computing an eigenvector with inverse iteration. *SIAM Rev.*, 39:254–291, 1997. MR **98f**:65041

- [13] A.V. Knyazev. Preconditioned eigensolvers—an oxymoron? *Electron. Trans. Numer. Anal.*, 7:104–123, 1998. MR **99h**:65068
- [14] K. Neymeyr. A posteriori error estimation for elliptic eigenproblems. Sonderforschungsbereich 382, Universität Tübingen, Report 132, 1999.
- [15] K. Neymeyr. A geometric theory for preconditioned inverse iteration. I: Extrema of the Rayleigh quotient. *Linear Algebra Appl.* 322(1–3):61–85, 2001. CMP 2001:06
- [16] K. Neymeyr. A geometric theory for preconditioned inverse iteration. II: Convergence estimates. *Linear Algebra Appl.* 322(1–3):87–104, 2001. CMP 2001:06
- [17] B.N. Parlett. *The symmetric eigenvalue problem*. Prentice Hall, Englewood Cliffs New Jersey, 1980. MR **81j**:65063
- [18] W.V. Petryshyn. 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, 1968. MR **36**:5747
- [19] B.A. Samokish. The steepest descent method for an eigenvalue problem with semi-bounded operators. *Izv. Vyssh. Uchebn. Zaved. Mat.*, 5:105–114, 1958.

MATHEMATISCHES INSTITUT DER UNIVERSITÄT TÜBINGEN, AUF DER MORGENSTELLE 10, 72076 TÜBINGEN, GERMANY

*E-mail address:* `neymeyr@na.uni-tuebingen.de`