ON RITZ APPROXIMATIONS FOR POSITIVE DEFINITE OPERATORS I (THEORY)

LUKA GRUBIŠIĆ AND KREŠIMIR VESELIĆ

ABSTRACT. We give new lower bounds of the Rayleigh–Ritz approximations of a part of the spectrum of an elliptic operator. Furthermore, we present bounds for the accompanying Ritz vectors. The bounds include a form of a relative gap between the Ritz values and the rest of the spectrum of the operator. A model example shows that the obtained bounds may be very sharp.

1. INTRODUCTION

In this paper we present a new estimate of the quality of the Rayleigh–Ritz approximations to the part of the spectrum of a positive definite selfadjoint operator in a Hilbert space.

Our method is based on two recently published works for finite matrices [5] and [10]. Sharp estimates obtained in these works are based on the maximal angle Θ between the subspaces spanned by RX and $R^{-*}X$ where X are (orthonormal) test vectors and $H = R^*R$ is the given matrix. The Ritz values are the eigenvalues of the matrix $M = X^*HX$. The maximal angle Θ appears to be equal to the norm of

$$R^{-1}(H - H')R^{-*},$$

where

$$\begin{aligned} H' &= PHP + (\mathbf{I} - P)H(\mathbf{I} - P) \\ &= XMX^* + (\mathbf{I} - P)H(\mathbf{I} - P) \end{aligned}$$

is the "block diagonal part" of H with respect to the projections $P = XX^*$ and $\mathbf{I} - P$.

This setting appears to be particularly suitable in the infinite–dimensional case since all crucial quantities may be expressed not in operators themselves but in the corresponding quadratic forms. This brings two important advantages

- quadratic forms simplify the calculations and, which is most important, they allow the test vectors to belong to the form

Date: May 26, 2003.

domain of H. This naturally includes linear finite elements for the second order elliptic differential operators.

- The obtained estimates are of the "relative type" i.e. the eigenvalue bound looks like

$$\frac{|\lambda-\mu|}{\lambda}\leq\frac{\sin\Theta}{1-\sin\Theta},$$

whereas the subspace bound works with the "relative gap" between the relevant groups of eigenvalues. The relative gap separates well two close eigenvalues that are themselves small and is therefore particularly suitable for dealing with the lower part of the spectrum of a positive definite operator (see also the diagram on Figure 1).

Our results can be seen as a natural generalization of the known result of Davis, Kahan and Weinberger [4] for the eigenvalues and Davis–Kahan [3] for the eigenvectors.

In the finite element literature one usually finds estimates of the type: *Let u, $\|u\|=1$ be a test vector, let $\mu=\big(H^{1/2}u,H^{1/2}u\big)$ be the Ritz value then if μ approximates an eigenvalue λ we have*

$$\frac{|\lambda-\mu|}{\lambda}\leq c\,\|D\|_{\tau}\,.$$

Here, c is a constant of moderate size, τ a finite–element approximation and $\|D\|_{\tau}$ is a measure of the residual $D=Hu-\mu u$. In fact, (cf. [9])

$$\|Hu-\mu u\|_{E}\leq c\,\|D\|_{\tau}\ ,$$

where E denotes the dual energy norm. The Ritz value bound is accompanied by the corresponding subspace error estimate. The subspace error estimate is a function of $c\,\|D\|_{\tau}$ and a subspace stability factor (in the terminology of the paper [9]). The subspace stability factor implicitly contains some measure of a spectral gap and a constant depending on the geometry of the domain.

Compared to these results our bound is more general (we are not limited to differential operators) and more explicit; all of its constants are directly calculated from the input data.

An important feature of our technique is a constructive proof of the existence and uniqueness of the solution of a weakly formulated Sylvester equation. We feel that through a combination of a residual theorem, that establishes a matching, and a subspace theorem, that reveals a nature of that matching, we get bounds that accurately reflect underlying properties of the Rayleigh–Ritz approximation.

In this paper we bring only one application of our bounds to a simple model problem, the Sturm–Liouville operator with coupled boundaries.

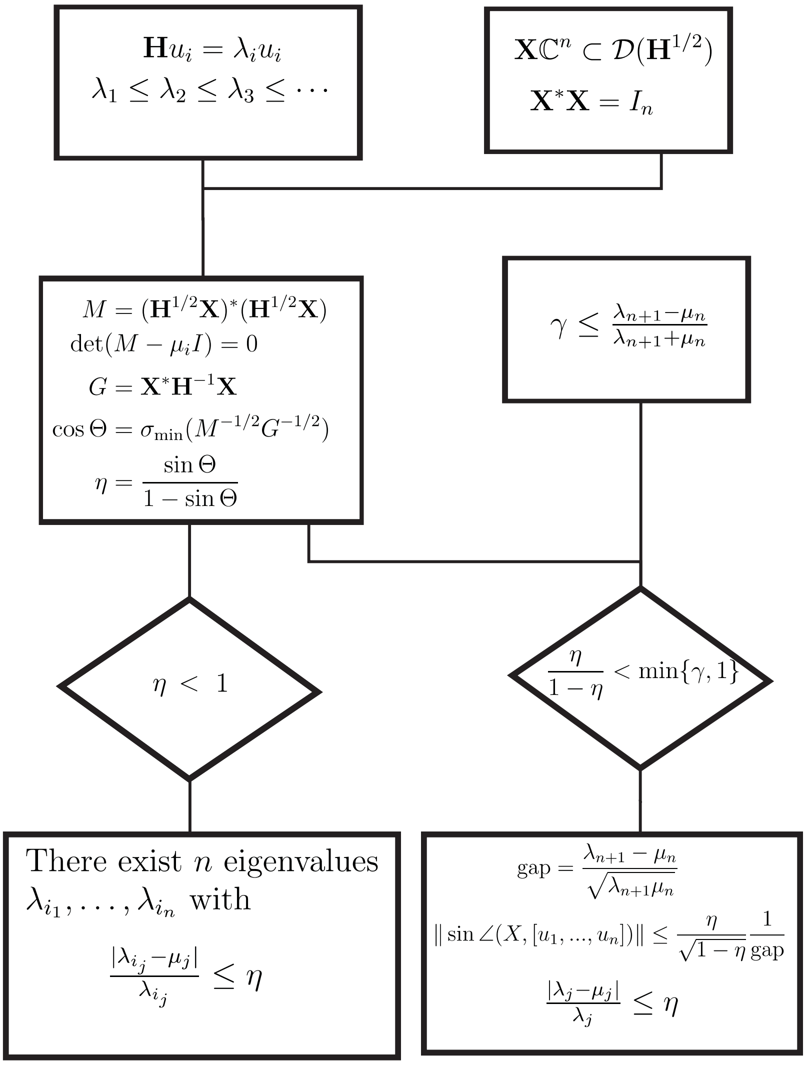

FIGURE 1. A diagrammatic overview of the results of this article.

The obtained bounds are, in fact, very sharp. Application to less trivial boundary value problems of Mathematical Physics will be brought in a subsequent paper.

2. A bound for the Ritz values and vectors

In what follows we shall use the standard notion of the generalized inverse of a degenerate operator $A = BC$, where

$$\begin{aligned} B &: \mathcal{H}'' \to \mathcal{H}', \quad \text{injective}, \\ C &: \mathcal{H} \to \mathcal{H}'', \quad \text{surjective}, \end{aligned}$$

and $\mathcal{H}, \mathcal{H}', \mathcal{H}''$ are any Hilbert spaces. The generalized inverse A^+ is given (cf. [7, Gantmacher, Ch I, §5]) by

$$A^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*. \tag{1}$$

Let $U, V : \mathbf{C}^n \to \mathcal{H}$ be isometries. *The angles* $\theta_1, \ldots, \theta_n$ between the spaces $U\mathbf{C}^n$ and $V\mathbf{C}^n$ are defined as

$$\theta_i = \arccos \sigma_i, \quad i = 1, \ldots, n, \tag{2}$$

where $\sigma_1, \ldots, \sigma_n$ are the singular values of any of the operators

$$V^*U,\ U^*V,\ P_V U,\ P_U V,$$

(P_Y generically denotes the orthogonal projection on the space $Y\mathbf{C}^n$, for $Y : \mathbf{C}^n \to \mathcal{H}$). Furthermore, it should be noted that the non–vanishing singular values of any of the degenerate operators

$$P_{V^\perp} U,\ P_{U^\perp} V$$

are the sines of the canonical angles that are not zero. We say that $U\mathbf{C}^n$, $V\mathbf{C}^n$ are in the acute position if

$$(U\mathbf{C}^n)^\perp \cap V\mathbf{C}^n = U\mathbf{C}^n \cap (V\mathbf{C}^n)^\perp = \{0\}$$

or, equivalently, if

$$\Theta = \max_i \theta_i < \frac{\pi}{2}. \tag{3}$$

The following lemma is a generalization of a result by Drmač [5] for finite matrices.

Lemma 2.1. *Let $\mathcal{H}, \mathcal{H}'$ be Hilbert spaces and $R : \mathcal{H} \to \mathcal{H}'$ a closed, densely defined linear operator satisfying*

$$\|Rx\| \geq \delta \|x\|, \quad \delta > 0.$$

Let $X : \mathbf{C}^n \to \mathcal{H}$ be a degenerate isometry with $X\mathbf{C}^n \subset \mathcal{D}(R)$. Let

$$Y = RX, \quad Z = RH^{-1}X, \quad H = R^*R$$

and suppose

$$Y\mathbf{C}^n \cap (Z\mathbf{C}^n)^\perp = (Y\mathbf{C}^n)^\perp \cap Z\mathbf{C}^n = \{0\}. \tag{4}$$

then there is an orthonormal basis in $\mathcal{K} = Y\mathbf{C}^n \dotplus Z\mathbf{C}^n$ such that in $\mathcal{K} \oplus \mathcal{K}^\perp$ the operator YZ^* is represented by

$$YZ^* = \left[\begin{array}{ccc|c} \mathbf{I}_k & & & \\ & \bigoplus_{i=1}^{l} \begin{bmatrix} 1 & \tan\theta_i \\ 0 & 0 \end{bmatrix} & & 0 \\ & & 0 & \\ \hline & 0 & & 0 \end{array}\right], \tag{5}$$

where $\theta_1, \dots, \theta_l$ are those angles between $Y\mathbf{C}^n$ and $Z\mathbf{C}^n$, which are different from 0 and $\frac{\pi}{2}$.

Proof. Note that RH^{-1} is everywhere defined and bounded. We will use the natural notation

$$R^+ = \overline{H^{-1}R^*} = (RH^{-1})^*,$$

moreover, [8, Ch. V], we conclude $\mathcal{R}(R^{+*}) \subseteq \mathcal{D}(R^*)$. Thus, $Y, Z : \mathbf{C}^n \to \mathcal{H}'$ are bounded and injective. We will prove the identities

$$Z^*Y = \mathbf{I} \qquad (\text{on } \mathbf{C}^n) \tag{6}$$

$$YZ^* = (P_Z P_Y)^+, \tag{7}$$

where P_Z, P_Y are the orthogonal projections onto $Z\mathbf{C}^n$, $Y\mathbf{C}^n$, respectively. For any x, y we have $R^{+*}Xy \in \mathcal{D}(R^*)$,

$$\begin{aligned}
(Yx, Zy) &= \big(RXx, R^{+*}Xy\big) \\
&= \big(Xx, R^*R^{+*}Xy\big) = \big(Xx, R^*RH^{-1}Xy\big) \\
&= (Xx, Xy) = (x, y)
\end{aligned}$$

and (6) follows. Furthermore, since $Y\mathbf{C}^n$ and $(Z\mathbf{C}^n)^\perp$ have the zero intersection the operator

$$P_Z Y = ZZ^+Y,$$

is injective, whereas Y^+ is surjective. Thus, we compute

$$\begin{aligned}
(P_Z P_Y)^+ &= (ZZ^+YY^+)^+ \\
&= (Y^+)^+[(ZZ^+Y)^*(ZZ^+Y)]^{-1}(ZZ^+Y)^* \\
&= Y[(Z^+Y)^*Z^*Z(Z^+Y)]^{-1}(Z^+Y)^*Z^* \\
&= Y(Z^+Y)^{-1}(Z^*Z)^{-1}(Z^+Y)^{-*}(Z^+Y)^*Z^* \\
&= Y(Z^+Y)^{-1}(Z^*Z)^{-1}Z^* \\
&= Y(Z^*ZZ^+Y)^{-1}Z^* \\
&= Y(Z^*Y)^{-1}Z^*
\end{aligned}$$

Since both P_Z and P_Y are degenerate — the finite dimensional subspace $\mathcal{K} = X\mathbf{C}^n \dotplus Y\mathbf{C}^n$ reduces both of them — the Wedin's theorem [12] guarantees the existence of an orthonormal basis in $\mathcal{K}$ such that in

$\mathcal{K} \oplus \mathcal{K}^\perp$ projections P_Z, P_Y are represented as

$$P_{\mathcal{X}} = \left[\begin{array}{ccc|c} \mathbf{I}_k & & & \\ & \bigoplus_{i=1}^l \Phi & & \\ & & \Delta_{\mathcal{X}} & \\ \hline & & & 0 \end{array}\right],$$

$$P_{\mathcal{Y}} = \left[\begin{array}{ccc|c} \mathbf{I}_k & & & \\ & \bigoplus_{i=1}^l \Psi(\theta_i) & & \\ & & \Delta_{\mathcal{Y}} & \\ \hline & & & 0 \end{array}\right],$$

with

$$\Phi = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \Psi(\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix}.$$

By (1) we have

$$\begin{aligned} &\left(\begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix} \begin{bmatrix} \cos\theta_i & 0 \\ 0 & 0 \end{bmatrix} \right)^+ \\ &= \left(\begin{bmatrix} \cos\theta_i \\ \sin\theta_i \end{bmatrix} \begin{bmatrix} \cos\theta_i & 0 \end{bmatrix} \right)^+ \\ &= \begin{bmatrix} \cos\theta_i \\ 0 \end{bmatrix} \frac{1}{\cos^2\theta} \cdot 1 \cdot \begin{bmatrix} \cos\theta_i & \sin\theta_i \end{bmatrix} = \begin{bmatrix} 1 & \tan\theta_i \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Now by (7) the conclusion (5) follows. □

The following theorem is again a generalization of a corresponding result by Drmač [5] for finite matrices.

Theorem 2.2. *Let R, H, X, Y, Z be as in the Lemma 2.1 an let H^{-1} be compact. Let M be the Hermitian matrix, defined by*

$$(Mx, y) = (RXx, RXy), \quad x, y \in \mathbf{C}^n. \tag{8}$$

and denote its eigenvalues (with multiplicities) by

$$\mu_1 \leq \cdots \leq \mu_n.$$

Then there exist n eigenvalues (with multiplicities)

$$\lambda_{i_1}, \ldots, \lambda_{i_n}$$

of H such that

$$\frac{|\lambda_{i_j} - \mu_j|}{\lambda_{i_j}} \leq \eta \equiv \frac{\sin\Theta}{1 - \sin\Theta}, \tag{9}$$

where Θ is the maximal acute angle, defined by (3), between $Y\mathbf{C}^n$ and $Z\mathbf{C}^n$ and η is supposed to be less then 1.

Proof. Let $h[x,y] = (Rx, Ry) = \big(H^{1/2}x, H^{1/2}y\big)$. This is a positive definite, closed, sesquilinear form defined on $\mathcal{D}(h) = \mathcal{D}(R) = \mathcal{D}(H^{1/2})$. Also, by [8, Ch. VI.7]

$$R = UH^{1/2}, \quad R^* = H^{1/2}U^*, \tag{10}$$

where U is the isometry from $\mathcal{H}'$ onto $\mathcal{R}(R)$. By $R^+ = H^{-1/2}U^*$ we have

$$R^+Rx = x, \quad x \in \mathcal{D}(H^{1/2}), \tag{11}$$

whereas $RR^+ = UU^*$ is the orthogonal projection onto $\mathcal{R}(R)$. We will now show that the form

$$\begin{aligned} \delta h[x,y] \;&=\; (R(\mathbf{I} - XX^*)x, RXX^*y) + \\ &\quad\;\, (RXX^*x, R(\mathbf{I} - XX^*)y)\,, \quad x, y \in \mathcal{D}(h). \end{aligned}$$

is h–bounded. For, $x, y \in \mathcal{H}'$ we have $R^+x, R^+y \in \mathcal{D}(h)$ and

$$\begin{aligned} \delta h[R^+x, R^+y] \;&=\; \big(R(\mathbf{I} - XX^*)R^+x, RXX^*R^+y\big) \\ &\quad + \big(RXX^*R^+x, R(\mathbf{I} - XX^*)R^+y\big) \\ &=\; \big((RR^+ - YZ^*)x, YZ^*y\big) \\ &\quad + \big(YZ^*x, (RR^+ - YZ^*)y\big)\,. \end{aligned}$$

Obviously $\delta h[R^+x, R^+y] = 0$ whenever $y \perp \mathcal{R}(R)$ or $x \perp \mathcal{R}(R)$. Hence, it is enough to take $x, y \in \mathcal{R}(R)$. By the Lemma 2.1 we obtain

$$\delta h[R^+x, R^+y] = \left(\left[\begin{array}{c|c} \begin{matrix} 0 & & \\ & \bigoplus_{i=1}^{l} \begin{bmatrix} 0 & -\tan\theta_i \\ -\tan\theta_i & -2\tan^2\theta_i \end{bmatrix} & \\ & & 0 \end{matrix} & 0 \\ \hline 0 & 0 \end{array} \right] x, y \right)$$

Thus, for any x, y we have

$$\begin{aligned} \big|\delta h[R^+x, R^+y]\big| \;&\leq\; \max_i \left\| \begin{bmatrix} 0 & -\tan\theta_i \\ -\tan\theta_i & -2\tan^2\theta_i \end{bmatrix} \right\|_2 \|x\|\|y\| \\ &=\; \frac{\sin\Theta}{1 - \sin\Theta}\, \|x\|\,\|y\| \end{aligned}$$

and for $x, y \in \mathcal{D}(h)$

$$|\delta h[x,y]| = \big|\delta h[R^+Rx, R^+Ry]\big| \leq \eta\sqrt{h[x]h[y]}. \tag{12}$$

Since the relative bound is less then one, the form

$$h' = h - \delta h, \quad \mathcal{D}(h') = \mathcal{D}(h) \tag{13}$$

is closed and positive definite and

$$(1-\eta)h \leq h' \leq (1+\eta)h, \quad \eta < 1. \tag{14}$$

The operator H' defined by the form h' is again positive definite and has a compact inverse [8]. We will show that the subspace $X\mathbf{C}^n$ reduces H'. Indeed, for $y \in \mathcal{D}(h)$, $x \in \mathbf{C}^n$ we have

$$\begin{aligned} h'[y, Xx] &= (Ry, RXx) - (RXX^*y, R(\mathbf{I} - XX^*)Xx) \\ &\quad - (R(\mathbf{I} - XX^*)y, RXX^*Xx) \\ &= (Ry, RXx) - (R(\mathbf{I} - XX^*)y, RXx) \\ &= (RXX^*y, RXx) \\ &= (MX^*y, x)\,. \end{aligned}$$

This is equivalent to

$$\left(H'^{1/2}y, H'^{1/2}Xx\right) = (y, XMx)\,, \quad y \in \mathcal{D}(h'),\ x \in \mathbf{C}^n.$$

This implies $X\mathbf{C}^n \subset \mathcal{D}(H')$ and

$$(y, H'Xx - XMx) = 0,$$

for all $y \in \mathcal{H}$. Hence,

$$H'X = XM. \tag{15}$$

We now use the standard min–max formula, cf. [2],

$$\lambda_k(H) = \min_{\dim(\mathcal{M}_k)=k} \max \left\{ \frac{h[f]}{\|f\|^2} : 0 \neq f \in \mathcal{M}_k \right\},$$

and similarly for the eigenvalues λ' of H' (both λ'_k and λ_k are counted together with their multiplicities). The min–max formulae, together with (14), immediately imply

$$(1-\eta)\lambda_i \leq \lambda'_i \leq (1+\eta)\lambda_i.$$

By (15) the restriction of H' onto $X\mathbf{C}^n$ is represented by the matrix M and thus $\mu_j = \lambda'_{i_j}$. □

Remark 2.1. The eigenvalues

$$\mu_1 \leq \cdots \leq \mu_n$$

of the matrix M, defined in (8), are called the Ritz values of the operator H associated with the subspace $X\mathbf{C}^n$.

Remark 2.2. Obviously, in order for Theorem 2.2 to hold it is not necessary that H^{-1} be compact. It is sufficient to assume that

(1) the essential spectrum of H lies right from some λ_e and
(2) the whole spectrum of M lies left from that λ_e.

Remark 2.3. The estimate (9) can be used to obtain a bound of $\frac{|\mu_j - \lambda_{i_j}|}{\mu_j}$, namely

$$\frac{|\mu_j - \lambda_{i_j}|}{\mu_j} = \left| \frac{\frac{\mu_j - \lambda_{i_j}}{\lambda_{i_j}}}{1 + \frac{\mu_j - \lambda_{i_j}}{\lambda_{i_j}}} \right| \leq \frac{\eta}{1-\eta}. \tag{16}$$

The following Lemma will be a key technical result for the rest of this section.

Lemma 2.3. *Let $h, \delta h, h', H, H', \eta$ be as in the proof of the preceding theorem. Set*

$$s[x,y] = \delta h[H^{-1/2}x, H'^{-1/2}y],$$

then

$$|s[x,y]| \leq \frac{\eta}{\sqrt{1-\eta}} \|x\| \, \|y\|, \quad x, y \in \mathcal{H}. \tag{17}$$

Proof. By (12) we have

$$|\delta h[H^{-1/2}x, H'^{-1/2}y]| \leq \eta \, \|x\| \sqrt{h[H'^{-1/2}]}$$

and also

$$\|H^{1/2}H'^{-1/2}\| \leq \frac{1}{\sqrt{1-\eta}}.$$

Altogether, the estimate (17) follows. □

In fact, by (12) we have

$$\begin{aligned} s[x,y] &= (Sx,y), && (18)\\ S &= H^{1/2}H'^{-1/2} - \overline{H^{-1/2}H'^{1/2}}, && (19)\\ \|S\| &\leq \frac{\eta}{\sqrt{1-\eta}}. && (20) \end{aligned}$$

The foregoing Lemma is in fact, implicitly contained in [10] for finite matrices. The following Corollary is explicitly obtained in [10] — again for the finite dimensional case.

Corollary 2.4. *Denote by $v_1, \ldots, v_n$ the orthogonal eigenvectors of H, belonging to the eigenvalues $\lambda_{i_1}, \ldots \lambda_{i_n}$, numbered as in (9) and similarly by $u_1, \ldots, u_n$ for the eigenvalues $\mu_1, \ldots, \mu_n$ of H'. Then,*

$$\|v_j - u_j\| \leq \frac{\sqrt{2}\,\eta}{\sqrt{1-\eta}} \max_{k \neq j} \frac{\sqrt{\mu_j \lambda_{i_k}}}{|\lambda_{i_k} - \mu_j|}. \tag{21}$$

PROOF. We have

$$\begin{aligned} s[v_k, u_j] &= \delta h[H^{-1/2}v_k, H'^{-1/2}u_j] \\ &= \Big(v_k, H^{1/2}H'^{-1/2}u_j\Big) - \Big(H'^{1/2}H^{-1/2}v_k, u_j\Big) \\ &= (\lambda_{i_k}^{1/2}\mu_j^{-1/2} - \lambda_{i_k}^{-1/2}\mu_j^{1/2})\,(v_k, u_j)\,, \end{aligned}$$

and

$$\begin{aligned} \sum_{k\neq j} |\,(v_k, u_j)\,|^2 &\leq \max_{k\neq j} \frac{\lambda_{i_k}\mu_j}{(\lambda_{i_k} - \mu_j)^2} \sum_{k\neq j} |s[v_k, u_j]|^2 \\ &\leq \max_{k\neq j} \frac{\lambda_{i_k}\mu_j}{(\lambda_{i_k} - \mu_j)^2} \frac{\eta^2}{1-\eta}. \end{aligned}$$

Choosing v_j, u_j so that $(v_j, u_j) \geq 0$ we obtain

$$\begin{aligned} \|v_j - u_j\| &= \sqrt{2}\sqrt{1-(v_j,u_j)} = \sqrt{2}\sqrt{1 - \sqrt{1 - \sum_{k\neq j} |\,(v_k, u_j)\,|^2}} \\ &\leq \sqrt{2}\sqrt{1 - \sqrt{1 - \max_{k\neq j} \frac{\lambda_{i_k}\mu_j}{(\lambda_{i_k} - \mu_j)^2} \frac{\eta^2}{1-\eta}}} \\ &\leq \frac{\sqrt{2}\,\eta}{\sqrt{1-\eta}} \max_{k\neq j} \frac{\sqrt{\mu_j\lambda_{i_k}}}{|\lambda_{i_k} - \mu_j|}. \qquad \square \end{aligned}$$

3. A bound for the invariant subspace

Let $V : \mathbf{C}^n \to \mathcal{H}$ be the isometry defined by

$$Ve_i = v_i, \quad i = 1, \ldots, k,$$

where (e_i) is the canonical orthonormal basis of $\mathbf{C}^n$ and v_i are as in the proof of the Corollary 2.4. The orthogonal projection

$$P_\perp = \mathbf{I} - VV^* \tag{22}$$

commutes with H and $H^{1/2}$. By $\Lambda_\perp$ we denote the restriction of H to its invariant subspace $P_\perp\mathcal{H}$. By (19) and (20) for $x \in \mathbf{C}^n$, $v \in P_\perp\mathcal{D}(H^{1/2})$ we have

$$\begin{aligned} |\,(v, SXx)\,| &= |\,\Big(v, (H^{1/2}H'^{-1/2} - H^{-1/2}H'^{1/2})Xx\Big)\,| \\ &= |(\Big(\Lambda_\perp^{1/2}v, XM^{-1/2}x\Big) - \Big(\Lambda_\perp^{-1/2}v, XM^{1/2}x\Big))| \\ &\leq \frac{\eta}{\sqrt{1-\eta}}\,\|v\|\,\|x\|\,. \end{aligned} \tag{23}$$

This relation is valid for any $x \in \mathbf{C}^n$, $v \in P_\perp \mathcal{D}(H^{1/2})$ and then also for any $v \in P_\perp \mathcal{H}$. The above bound should make it possible to estimate

$$T = P_\perp X : \mathbf{C}^n \to P_\perp \mathcal{H}, \tag{24}$$

which is the solution of the formal Sylvester equation

$$\Lambda_\perp T - TM = \Lambda_\perp^{1/2} S X M^{1/2}. \tag{25}$$

for the unknown T. Here by "formal" we mean that the right hand side of (25) need not represent a bona fide operator. This equation in its weak form reads

$$\left(\Lambda_\perp^{1/2} v, TM^{-1/2} x\right) - \left(v, \Lambda_\perp^{-1/2} TM^{1/2} x\right) = (v, SXx)\,, \tag{26}$$

for every $v \in P_\perp \mathcal{D}(h)$ and $x \in \mathbf{C}^n$.

Theorem 3.1. *Let $\Lambda_\perp$ be the positive definite operator in the Hilbert space $\mathcal{H}$, let $M \in \mathbf{C}^{n\times n}$ be the positive definite matrix and let $Q : \mathbf{C}^n \to \mathcal{H}$ be bounded. Provided*

$$\|M\| < \frac{1}{\|\Lambda_\perp^{-1}\|}\,, \tag{27}$$

the weakly formulated Sylvester equation

$$\left(\Lambda_\perp^{1/2} v, TM^{-1/2} x\right) - \left(v, \Lambda_\perp^{-1/2} TM^{1/2} x\right) = (v, Qx)\,, \tag{28}$$

has a unique solution given by the norm–convergent integral

$$T = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \Lambda_\perp^{1/2} (\Lambda_\perp - \mathrm{i}\zeta - d)^{-1} Q (M - \mathrm{i}\zeta - d)^{-1} M^{1/2} d\zeta, \tag{29}$$

where d is any number satisfying

$$\|M\| < d < \frac{1}{\|\Lambda_\perp^{-1}\|}\,. \tag{30}$$

Proof. The uniqueness means that

$$\left(\Lambda_\perp^{1/2} v, WM^{-1/2} x\right) - \left(v, \Lambda_\perp^{-1/2} WM^{1/2} x\right) = 0, \tag{31}$$

for $x \in \mathbf{C}^n$, $v \in P_\perp \mathcal{D}(H^{1/2})$, has the only solution $W = 0$. Indeed, (31) implies $WM^{-1/2} x \in \mathcal{D}(\Lambda_\perp^{1/2})$ and

$$\Lambda_\perp^{1/2} WM^{-1/2} = \Lambda_\perp^{-1/2} WM^{1/2} = 0$$

or, equivalently

$$WM^{-1} y = \Lambda_\perp^{-1} W y, \qquad y \in \mathbf{C}^n.$$

Now, let $y = u_1, \dots, u_n$, then

$$(\mu_j \Lambda_\perp^{-1} - \mathbf{I}) W u_j = 0.$$

By the relation (30) the operator $(\mu_j \Lambda_\perp^{-1} - \mathbf{I})$ is boundedly invertible and we obtain $W u_1 = \cdots = W u_n = 0$ and hence $W = 0$. Thus the equation (28) has not more than one solution.

Now for the existence. If $\Lambda_\perp$ is bounded then (28) is equivalent to (25) and the formula (29) is known to yield a unique solution to (28). Define a cut–off function

$$f_n(x) = \begin{cases} x, & D \le x \le n \\ n, & n \le x \end{cases}$$

with $D = 1/\|\Lambda_\perp^{-1}\|$. The operator $f_n(\Lambda_\perp)$ is bicontinuous and

$$T_n = -\frac{1}{2\pi} \int_{-\infty}^{\infty} f_n(\Lambda_\perp)^{1/2} (f_n(\Lambda_\perp) - \mathrm{i}\zeta - d)^{-1} Q (M - \mathrm{i}\zeta - d)^{-1} M^{1/2} d\zeta$$

satisfies

$$\left(f_n(\Lambda_\perp)^{1/2} v, T_n M^{-1/2} x\right) - \left(v, f_n(\Lambda_\perp)^{-1/2} T_n M^{1/2} x\right) = (v, Qx)\,. \tag{32}$$

Now, for $v \in \mathcal{D}(\Lambda_\perp^{1/2})$ we have

$$f_n(\Lambda_\perp)^{1/2} v \xrightarrow[n\to\infty]{} \Lambda_\perp^{1/2} v. \tag{33}$$

and similarly $f_n(\Lambda_\perp)^{-1/2} \to \Lambda_\perp^{-1/2}$ strongly. As a consequence of the elementary spectral theory,

$$f_n(\Lambda_\perp)^{1/2} (f_n(\Lambda_\perp) - \mathrm{i}\zeta - d)^{-1} \xrightarrow[n\to\infty]{} \Lambda_\perp^{1/2} (\Lambda_\perp - \mathrm{i}\zeta - d)^{-1}$$

in the strong sense. Since the operator X is degenerate the convergence

$$\Phi_n(\zeta, d) \xrightarrow[n\to\infty]{} \Phi(\zeta, d), \tag{34}$$

for

$$\Phi_n(\zeta, d) = f_n(\Lambda_\perp)^{1/2} (f_n(\Lambda_\perp) - \mathrm{i}\zeta - d)^{-1} Q (M - \mathrm{i}\zeta - d)^{-1} M^{1/2}$$

and

$$\Phi(\zeta, d) = \Lambda_\perp^{1/2} (\Lambda_\perp - \mathrm{i}\zeta - d)^{-1} Q (M - \mathrm{i}\zeta - d)^{-1},$$

takes place even in the norm sense. To set the scene for the Lebesgue dominated convergence theorem we estimate

$$\begin{aligned} |\,(\Phi_n(\zeta, d)x, y)\,|^2 \;&\le\; \left(f_n(\Lambda_\perp)((f_n(\Lambda_\perp) - d)^2 + \zeta^2)^{-1} y, y\right) \\ &\qquad \cdot \left((M(M-d)^2 + \zeta^2)^{-1} x, x\right) \frac{\eta^2}{1-\eta}. \end{aligned}$$

For $\lambda \geq D$ the functions

$$\frac{f_n(\lambda)}{(f_n(\lambda)-d)^2+\zeta^2} = \begin{cases} \dfrac{\lambda}{(\lambda-d)^2+\zeta^2}, & D \leq \lambda \leq n \\ \dfrac{n}{(n-d)^2+\zeta^2}, & n \leq \lambda \end{cases}$$

are decreasing with $\lambda \geq D$. This is so, because $D > d$ and

$$\frac{\lambda}{(\lambda-d)^2+\zeta^2}$$

is decreasing with $\lambda > D$. In particular

$$\frac{f_n(\lambda)}{(f_n(\lambda)-d)^2+\zeta^2} \leq \frac{D}{(D-d)^2+\zeta^2}.$$

An analogous estimate is obtained with M. Thus we have obtained

$$\|\Phi_n(\zeta, d)\| \leq \sqrt{\frac{D}{(D-d)^2+\zeta^2}}\sqrt{\frac{\|M\|}{(\|M\|-d)^2+\zeta^2}},$$

an integrable uniform bound. By the Lebesgue dominated convergence theorem we have $T_n \to T$ in norm. Letting $n \to \infty$ in (32) and (33) we conclude that the operator T satisfies (28). $\square$

Theorem 3.2. *Let $X, M, \Lambda_\perp, P_\perp, D, \eta$ be as defined above. Then*

$$\|P_\perp X\| \leq \frac{\eta}{\sqrt{1-\eta}}\frac{\sqrt{D\,\|M\|}}{D-\|M\|}. \tag{35}$$

Proof. $P_\perp X = T$ satisfies the equation (28) where $Q = SX$. By (23) we can estimate the norm of T from (29):

$$\begin{aligned} |\,(Tx,y)\,|^2 &\leq \frac{1}{4\pi^2}\beta(y)\alpha(x)\,\|SX\|^2, \\ \beta(y) &= \int_{-\infty}^{\infty} \left(\Lambda_\perp((\Lambda_\perp-d)^2+\zeta^2)^{-1}y, y\right) d\zeta, \\ \alpha(x) &= \int_{-\infty}^{\infty} \left((M(M-d)^2+\zeta^2)^{-1}x, x\right) d\zeta, \end{aligned}$$

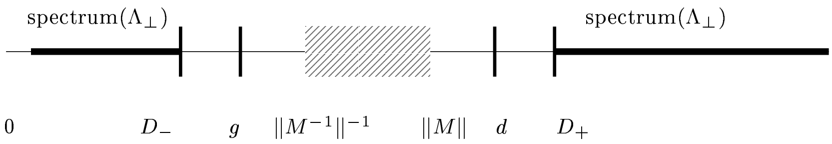

FIGURE 2. The spectral gaps

Using the spectral integral $\Lambda_\perp = \int \lambda \, dE(\lambda)$ we obtain

$$\begin{aligned}
\beta(y) &= \int_{-\infty}^{\infty} d\zeta \int_D^{\infty} \frac{\lambda \, d\,(E(\lambda)y, y)}{(\lambda - d)^2 + \zeta^2} \\
&= \int_D^{\infty} \lambda \, d\,(E(\lambda)y, y) \int_{-\infty}^{\infty} \frac{d\zeta}{(\lambda - d)^2 + \zeta^2} \\
&= \int_D^{\infty} \frac{\pi \lambda \, d\,(E(\lambda)y, y)}{\lambda - d} = \pi \left(\Lambda_\perp (\Lambda_\perp - d)^{-1} y, y \right)
\end{aligned}$$

and similarly

$$\alpha(x) = \pi \left(M(d - M)^{-1} x, x \right).$$

Together with (23) this gives

$$\begin{aligned}
|\,(Tx, y)\,| &\le \frac{\eta}{\sqrt{1-\eta}} \frac{1}{2} \sqrt{(\Lambda_\perp(\Lambda_\perp - d)^{-1} y, y)} \sqrt{(M(d-M)^{-1}x, x)} \\
&\le \frac{\eta}{2\sqrt{1-\eta}} \sqrt{\frac{D\|M\|}{(D-d)(d-\|M\|)}} \, \|x\| \, \|y\| \qquad (36)
\end{aligned}$$

for any $d \in \langle \|M\|, D \rangle$. The optimal d equals $\frac{(D + \|M\|)}{2}$ and

$$\|P_\perp V\| \le \frac{\eta}{\sqrt{1-\eta}} \frac{\sqrt{D\|M\|}}{D - \|M\|}$$

□

Remark 3.1. An analogue of Theorem 3.2 holds, if the assumption (30) is replaced by a more general one, namely that the interval

$$\left[\|M^{-1}\|^{-1}, \|M\| \right]$$

be contained in the resolvent set of the operator $\Lambda_\perp$. In this way we approximate not merely the lower part of the spectrum but any desired spectral interval. In this case, instead of (29) we have

$$T = -\frac{1}{2\pi}\int_{-\infty}^{\infty}\Lambda_\perp^{1/2}(\Lambda_\perp - \mathrm{i}\zeta - d)^{-1}SX(M-\mathrm{i}\zeta-d)^{-1}M^{1/2}d\zeta$$
$$+\frac{1}{2\pi}\int_{-\infty}^{\infty}\Lambda_\perp^{1/2}(\Lambda_\perp - \mathrm{i}\zeta - g)^{-1}SX(M-\mathrm{i}\zeta-g)^{-1}M^{1/2}d\zeta,$$

where d, g are chosen from the right and left spectral gap, see Fig. 2.

The same proof as in the Theorem 3.2 yields the estimate

$$\|P_\perp X\| \le \frac{\eta}{\sqrt{1-\eta}}\left(\frac{\sqrt{D_+\|M\|}}{D_+-\|M\|}+\frac{\sqrt{\|M^{-1}\|^{-1}D_-}}{\|M^{-1}\|^{-1}-D_-}\right). \tag{37}$$

Remark 3.2. We can interpret the form δh as the operator

$$\delta H : \mathcal{D}^1 \to \mathcal{D}^{-1},$$

where $\mathcal{D}^1$ is the Hilbert space $\mathcal{D}(h)$ with the scalar product

$$(f, g)_1 = h[f, g]$$

and $\mathcal{D}^{-1}$ is its dual, see [6]. The operator δH is defined by

$$(\delta H f, g) = \delta h[f, g].$$

The estimate

$$|\,(\delta H f, g)\,| = |\delta h[f, g]| \le \eta\sqrt{h[f]h[g]},$$

together with the relation (12) implies

$$\eta = \max_{f,g\neq 0}\frac{|\delta h[f,g]|}{\sqrt{h[f]h[g]}}. \tag{38}$$

This is equivalent to the statement

$$\|\delta H\|_{\mp 1} = \eta,$$

where $\|\cdot\|_{\mp 1}$ is the operator norm on $\mathcal{L}(\mathcal{D}^1, \mathcal{D}^{-1})$. We conclude that η does not depend on the choice of the factor R. The same conclusion can be obtained by a geometric reasoning (unitary invariance of an angle implies the conclusion, cf. [5]).

4. A numerical example

In this section we demonstrate the use of Theorems 2.2 and 3.2 on a model problem. We compare our bounds with other explicit bounds (i.e. bounds that are free from unknown quantities) found in the literature. Our subspace theorem is compared with the Davis–Kahan SinΘ theorem and the Temple–Kato bound, cf. [1, 3]. On the other hand, the Ritz value bound is compared with the Temple–Lehmann and the

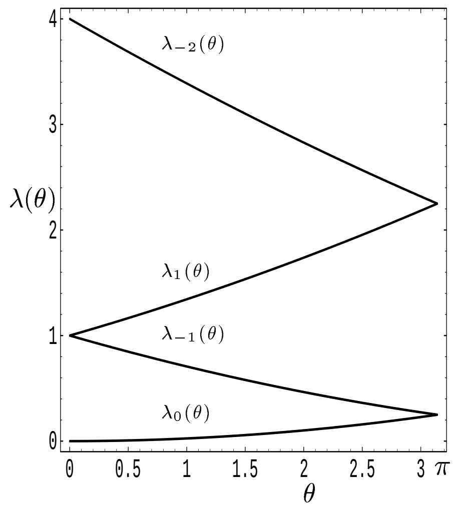

FIGURE 3. Increasingly ordered eigenvalues of the family of problems (39) as functions of θ.

Temple–Kato inequalities, cf. [1, 2, 11]. The model problem is (cf. [11])

$$\begin{aligned} -z'' - \alpha\, z &= \lambda z, \\ e^{i\theta} z(0) &= z(2\pi) \\ e^{i\theta} z'(0) &= z'(2\pi), \end{aligned} \tag{39}$$

where $\theta \in [0, \pi]$ and $\alpha \in \mathbf{R}$ is a constant fixed so that all the eigenvalues are positive. The solution to the problem (39) is given by the pairs

$$\begin{aligned} \lambda_{\pm k} &= \left(\pm k + \frac{\theta}{2\pi}\right)^2 - \alpha, \quad z_{\pm k}(t) = e^{-i\left(\pm k + \frac{\theta}{2\pi}\right)t}, \quad k \in \mathbf{N} \\ \lambda_0 &= \left(\frac{\theta}{2\pi}\right)^2 - \alpha, \qquad z_0(t) = e^{-i\frac{\theta}{2\pi}t}. \end{aligned}$$

On Figure 3 we see increasingly ordered eigenvalues of the family of problems (39) displayed as functions of θ. For $\theta = \pi$ we have an eigenvalue problem that has all the eigenvalues of multiplicity two. By varying the parameter θ in a "neighborhood" of π we construct eigenvalue problems that have as tightly clustered eigenvalues as we desire. A "relative" distance function is not shift invariant. For every $\theta \in [0, \pi]$ we can choose a shift α so as to make the two lowermost eigenvalues well separated in a "relative" sense.

For the parameters θ and α we choose

$$\theta = \frac{9999\,\pi}{10000}, \qquad \alpha = 0.2499\ . \tag{40}$$

The two smallest eigenvalues of the problem (39) are λ_0 and λ_{-1}. We compute

$$\min_{\substack{i=0,-1\\ p\neq 0,-1}} \frac{|\lambda_i - \lambda_p|}{\sqrt{\lambda_i \lambda_p}} = 115.459, \qquad \frac{|\lambda_0 - \lambda_{-1}|}{\sqrt{\lambda_0 \lambda_{-1}}} = 1.15466$$

and

$$\min_{\substack{i=0,-1\\ p\neq 0,-1}} |\lambda_i - \lambda_p| = 1.9998, \quad |\lambda_0 - \lambda_{-1}| = 10^{-4}.$$

The eigenvalue problem (39) with this choice of parameters α and θ will demonstrate why in some situations it is preferable to have bounds that are functions of the "relative" gaps rather then "absolute" ones.

Let $Z : \mathbf{C}^n \to \mathcal{H}$ be an isometry. By $\mathcal{E}_Z$ we denote a subspace associated, in the sense of Theorem 2.2, with the test space $Z\mathbf{C}^n$. In this case study we compute bounds of the approximation of the invariant spaces

$$\mathcal{E}^{\mathrm{X}} = \operatorname{span}\{z_0\}, \qquad \mathcal{E}^{\mathrm{Y}} = \operatorname{span}\{z_0, z_{-1}\}.$$

In a weak formulation (39) reads

$$h[z,v] = \int_0^{2\pi} z'\overline{v'} - \alpha \int_0^{2\pi} z\overline{v} = \lambda \int_0^{2\pi} z\overline{v} = \lambda\,(z,v) \tag{41}$$

$$z, v \in \{f : f, f' \in L^2[0,2\pi], \mathrm{e}^{\mathrm{i}\theta} f(0) = f(2\pi)\} = \mathcal{D}(h).$$

The assumptions on the constant α guarantee that there exists $\delta > 0$ such that

$$\delta\,\|v\| \le h[v,v], \qquad v \in \mathcal{D}(h),$$

so the form h satisfies the assumptions of Theorem 2.2. For the equidistant subdivision

$$0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 2\pi, \quad x_{i+1} - x_i = \frac{1}{n+1},$$

we introduce the finite dimensional function spaces

$$\begin{aligned}
\mathcal{V}_n^1 &= \{f : f \in C[0,2\pi], \mathrm{e}^{\mathrm{i}\theta} f(0) = f(2\pi), \\
&\qquad f \text{ is linear on } \langle x_{j-1}, x_j\rangle, j = 1,\ldots,n+1\} \\
\mathcal{V}_n^3 &= \{f : f \in C^1[0,2\pi], \mathrm{e}^{\mathrm{i}\theta} f(0) = f(2\pi), \\
&\qquad f \text{ is cubic on } \langle x_{j-1}, x_j\rangle, j = 1,\ldots,n+1\}\,.
\end{aligned}$$

Let Π_n^1 and Π_n^3 denote the interpolation operators

$$\begin{aligned}\Pi_n^1 &: C[0,2\pi] \to \mathcal{V}_n^1,\\ \Pi_n^3 &: C^1[0,2\pi] \to \mathcal{V}_n^3.\end{aligned}$$

For a given $n \in \mathbf{N}$, as approximate eigenvectors we choose

$$\begin{aligned}u^1 = \Pi_n^1 z_0 \quad &, \quad v^1 = \Pi_n^1 z_{-1}\\ u^3 = \Pi_n^3 z_0 \quad &, \quad v^3 = \Pi_n^3 z_{-1}.\end{aligned}$$

Let the matrices

$$\begin{aligned}X^1 = [u^1] \quad &, \quad X^3 = [u^3]\\ Y^1 = [u^1\, v^1] \quad &, \quad Y^3 = [u^3\, v^3],\end{aligned}$$

be understood as operators. For the test subspaces we take $X^1\mathbf{C}$, $X^3\mathbf{C}$, $Y^1\mathbf{C}^2$ and $Y^3\mathbf{C}^2$. Obviously,

$$\lim_n (\Pi_n^i z_j) = z_j, \quad i = 1,3, \quad j = 0,-1\,,$$

so we compute the error bounds as functions of n.

Remark 4.1. Note that for the test spaces $X^1\mathbf{C}$ and $Y^1\mathbf{C}^2$ of all compared bounds only our bound is meaningful. Our theory can handle the test vectors that are in the form domain whereas other bounds require test vectors from the operator domain!

We illustrate some practical issues concerning the computation of η by computing

$$\sin\Theta(H^{1/2}Y, H^{-1/2}Y).$$

Here $Y : \mathbf{C}^2 \to \mathcal{H}$ is an arbitrary isometry.

We define

$$\begin{aligned}Q_{-1} &= (H^{-1/2}Y)((H^{-1/2}Y)^*(H^{-1/2}Y))^{-1/2}\\ Q_1 &= (H^{1/2}Y)((H^{1/2}Y)^*(H^{1/2}Y))^{-1/2}.\end{aligned}$$

The operators Q_1 and Q_{-1} are obviously isometries. The singular values of the matrix

$$Q_{-1}^* Q_1 \in \mathbf{C}^{2\times 2}$$

are $\cos\theta_1$ and $\cos\theta_2$ as defined in the relation (2).

By $\sigma_{\min}(\cdot)$ we denote the smallest singular value of a matrix. One computes

$$\cos\Theta(H^{1/2}Y, H^{-1/2}Y) = \sigma_{\min}(Q_{-1}^* Q_1) = \sigma_{\min}(K_{-1}^{-1/2} Y^* Y K_1^{-1/2}),$$

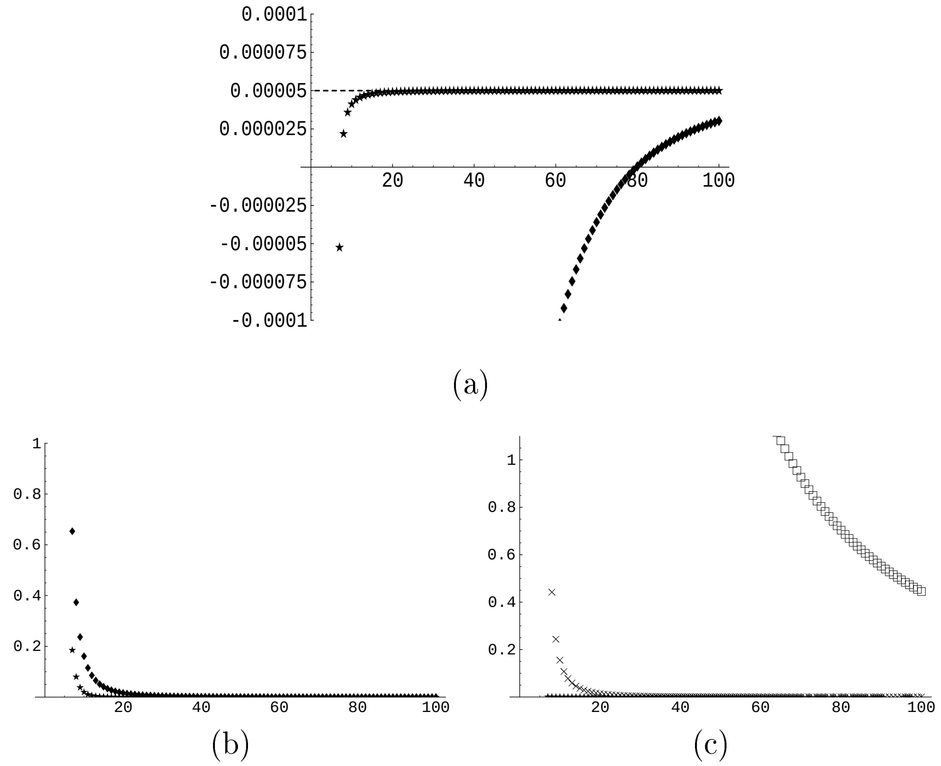

FIGURE 4. The eigenvalue and the eigenvector estimates from the space $X^3\mathbf{C} \subset \mathcal{V}_n^3$ as functions of n.

where

$$\begin{aligned}
K_1 = (H^{1/2}Y)^*(H^{1/2}Y) &= \begin{bmatrix} \left(H^{1/2}u, H^{1/2}u\right) & \left(H^{1/2}v, H^{1/2}u\right) \\ \left(H^{1/2}u, H^{1/2}v\right) & \left(H^{1/2}v, H^{1/2}v\right) \end{bmatrix} \\
K_{-1} = (H^{-1/2}Y)^*(H^{-1/2}Y) &= \begin{bmatrix} (H^{-1}u, u) & (H^{-1}v, u) \\ (H^{-1}u, v) & (H^{-1}v, v) \end{bmatrix} \\
Y^*Y &= \begin{bmatrix} (u, u) & (v, u) \\ (u, v) & (v, v) \end{bmatrix}.
\end{aligned}$$

The Green function of the operator $Hu = -u'' - \alpha u$, with the boundary conditions (39), is

$$G(t_1, t_2) = \frac{\mathrm{i}}{2\sqrt{\alpha}} \left(e^{\mathrm{i}\sqrt{\alpha}\,|t_1 - t_2|} + \frac{e^{\mathrm{i}\,(t_1 - t_2)\sqrt{\alpha}}}{-1 + e^{-2\,\mathrm{i}\,\pi\sqrt{\alpha} - \mathrm{i}\,\theta}} + \frac{e^{\mathrm{i}\,(t_2 - t_1)\sqrt{\alpha}}}{-1 + e^{-2\,\mathrm{i}\,\pi\sqrt{\alpha} + \mathrm{i}\,\theta}} \right),$$

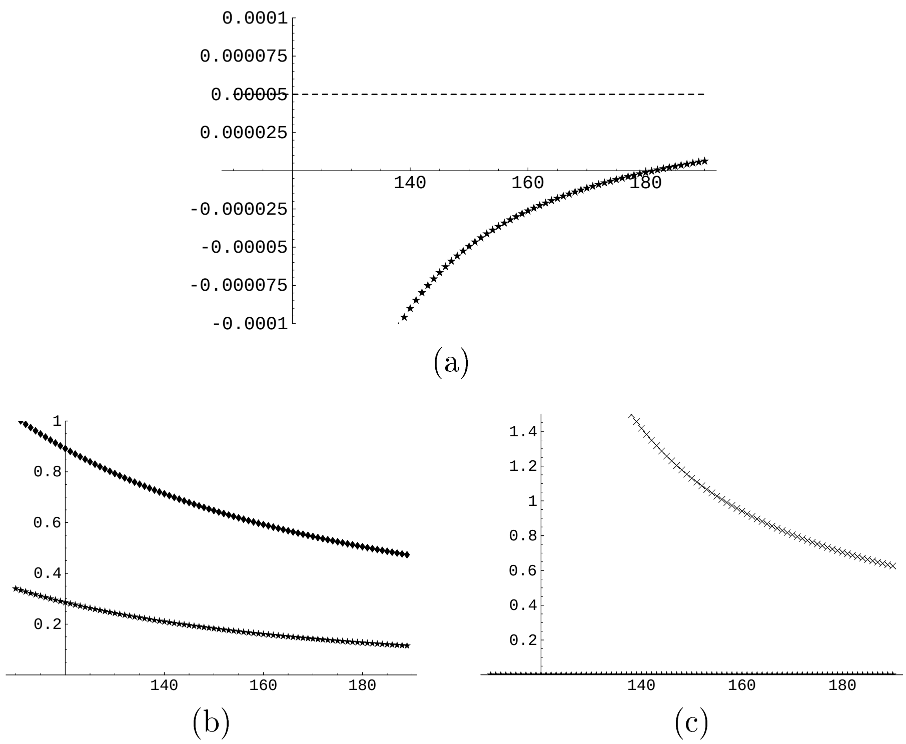

FIGURE 5. The eigenvalue and the eigenvector estimates from the space $X^1\mathbf{C} \subset \mathcal{V}_n^1$ as functions of n.

so in this case we use the formula

$$\left(H^{-1}u, v\right) = \int_0^{2\pi} dt_1 \int_0^{2\pi} G(t_1, t_2)u(t_2)\overline{v(t_1)}\, dt_2.$$

Finally, we have

$$\eta = \frac{\sin\Theta(H^{1/2}Y, H^{-1/2}Y)}{1 - \sin\Theta(H^{1/2}Y, H^{-1/2}Y)}.$$

The results of the numerical experiments are presented on Figures 4 and 5 . The figures are organized so that

Graph (a): (★) denotes our lower bound for the λ_0 obtained from the subspace X using Theorem 2.2

$$\lambda_0 \geq \left(1 - \frac{\eta}{1-\eta}\right)\left(H^{1/2}u, H^{1/2}u\right)$$

and (♦) denotes a lower bound for the λ_0 obtained from the Temple–Kato inequality

$$\lambda_0 \geq \left(H^{1/2}u, H^{1/2}u\right) - \frac{(Hu, Hu) - \left(H^{1/2}u, H^{1/2}u\right)^2}{\lambda_{-1} - (H^{1/2}u, H^{1/2}u)},$$

The dashed line represents the value of λ_0.

Graph (b): (★) denotes the true error and (♦) denotes our bound from Theorem 2.2 ,

Graph (c): ($\times$) denotes the subspace bound from Theorem 3.2

$$\sin\Theta(\mathcal{E}^X, X) \leq \frac{\eta}{\sqrt{1-\eta}} \frac{\sqrt{(H^{1/2}u, H^{1/2}u)\,\lambda_{-1}}}{\lambda_{-1} - (H^{1/2}u, H^{1/2}u)},$$

($\square$) denotes the Davis–Kahan bound

$$\sin\Theta(\mathcal{E}^X, X) \leq \frac{\|D\|}{\lambda_{-1} - (H^{1/2}u, H^{1/2}u)}$$

whereas the Temple–Kato bound

$$\sin\Theta(\mathcal{E}^X, X) \leq \frac{2}{\lambda_{-1} - 0} \times\sqrt{\left((H^{1/2}u, H^{1/2}u) - \frac{\lambda_{-1} - 0}{2}\right)^2 + \|D\|^2}.$$

turns out to be larger than 1 and so is not graphically represented. The residual vector D is defined as

$$D = Hu - \left(H^{1/2}u, H^{1/2}u\right)u.$$

The vertical line (|) marks the true error

$$\sin\Theta(\mathcal{E}^X, X\mathbf{C}^2).$$

Remark 4.2. Here we have assumed that the eigenvalues that match the computed Ritz values are indeed the eigenvalues λ_{-1} and λ_0. This is equivalent to the assumption $\mathcal{E}^X = \mathcal{E}_X$, in our subspace terminology. This assumption will be justified in Theorem 5.1.

In both the Temple–Kato and the Davis–Kahan bounds

$$\lambda_{-1} - \left(H^{1/2}u, H^{1/2}u\right)$$

should be understood as the best possible estimate of the spectral gap

$$\min_{\lambda\in\sigma(H)\setminus\lambda_0} \left|\left(H^{1/2}u, H^{1/2}u\right) - \lambda\right|.$$

All measures of the spectral gap, that appear in the subspace theorems, are estimated in the same fashion.

We stress again that on Figure 5 only our bound is presented since

$$X^1\mathbf{C} \subset \mathcal{D}(h) \setminus \mathcal{D}(H)$$

and other competing bounds are not applicable.

Remark 4.3. Let $\alpha = 0$, then for the problem (39) we may choose $R \in \mathcal{C}(\mathcal{H})$ defined by

$$Ru = \mathrm{i}u', \quad u \in \mathcal{D}(R) = \mathcal{D}(h).$$

The operator R is selfadjoint but not positive definite in

$$\mathcal{H} = L^2[0, 2\pi],$$

see [11]. In this particular case we know how to compute the formula for the R^+. Such a formula may be used as an alternative to the Green function in the computation of the bound η.

For $\xi \in \mathcal{H}$ the vector $R^+\xi = v$ is computed as the solution of the first order problem

$$\begin{aligned} \mathrm{i}v' &= \xi \\ \mathrm{e}^{\mathrm{i}\theta}v(0) &= v(2\pi). \end{aligned} \tag{42}$$

The formula for the function v is

$$v(x) = \frac{-\mathrm{i}}{-1 + e^{\mathrm{i}\,\theta}} \int_0^{2\,\pi} \xi(t)\, dt - \mathrm{i} \int_0^x \xi(t)\, dt.$$

Now, given functions $\xi, \zeta \in \mathcal{H}$ one computes

$$\big(H^{-1}\xi, \zeta\big) = \big(R^+\xi, R^+\zeta\big) = \int_0^{2\pi} v(t)\overline{g(t)}\, dt.$$

The functions v and g are the solutions of the appropriate first order systems of the form (42).

5. The right matching

In this section we give a theorem that establishes which eigenvalues of the operator H, given by a sesquilinear form h, are approximated by the Ritz values associated with the test subspace $X\mathbf{C}^n \subset \mathcal{D}(h) = \mathcal{D}(H^{1/2})$. We illustrate the result on the model problem. The theory of this section justifies the identification of the subspaces $\mathcal{E}^X$ and $\mathcal{E}_X$, tacitly assumed to produce Figures 4 and 5. Before we proceed with the formulation of the theorem we state, once again, that given $0 < \lambda, \mu$ and $\eta < 1$ the relation

$$\frac{|\lambda - \mu|}{\lambda} \leq \eta < 1$$

implies the relation

$$\frac{|\lambda - \mu|}{\mu} \le \frac{\eta}{1-\eta},$$

see Remark 2.3 .

Theorem 5.1. *Let positive definite sesquilinear form h and the subspace $X\mathbf{C}^n \subset \mathcal{D}(h)$ be given. By h' we denote the form constructed in the proof of Theorem 2.2 . By H and H' we denote the selfadjoint operators defined by the forms h and h'. Let*

$$\mu_1 \le \cdots \le \mu_n$$

be the Ritz values of the operator H associated with the subspace $X\mathbf{C}^n$. If

$$\frac{\eta}{1-\eta} < \min\{\gamma, 1\}, \tag{43}$$

where η, assumed to be less than 1, is as in (9) and (38) and

$$\gamma \;=\; \min_{\substack{k=1,\dots,n\\ p=n+1,\dots,\infty}} \frac{\lambda_p - \mu_k}{\lambda_p + \mu_k}$$

then[1]

$$\frac{|\lambda_i - \mu_i|}{\lambda_i} \le \eta, \quad i = 1, ..., n.$$

Proof. We note that under the assumption (43) we conclude that the operator H' is positive definite. Hence, we denote its eigenvalues by

$$0 < \widetilde{\mu}_1 \le \widetilde{\mu}_2 \le \cdots \le \widetilde{\mu}_n \le \cdots \infty$$

The estimate

$$1 - \eta \le \frac{\widetilde{\mu}_i}{\lambda_i} \le 1 + \eta, \quad i \in \mathbf{N}.$$

is a consequence of Minimax theorem. Alternatively we write this assertion as

$$\frac{|\lambda_i - \widetilde{\mu}_i|}{\lambda_i} \le \eta, \quad i \in \mathbf{N}. \tag{44}$$

By the construction of the operator H' we know that the there exists a 1-1 mapping $i_{(\cdot)}$ of $\mathbf{N}$ onto itself such that

$$\mu_k = \widetilde{\mu}_{i_k}, \quad k = 1, ..., n.$$

[1]Note that (43) implies $\gamma > \eta \ge 0$. The supposition $\gamma > 0$ is equivalent to the supposition (27).

Note that $k < d$ implies $i_k < i_d$ for $k, d \leq n$. Now the relation (44) implies

$$\frac{|\lambda_{i_k} - \mu_k|}{\lambda_{i_k}} \leq \eta, \quad k = 1, ..., n.$$

To prove the theorem we show a slightly stronger assertion, namely,

$$\mu_k < \widetilde{\mu}_p, \quad p = n+1, ..., \infty, k = 1, ..., n \tag{45}$$

$$\lambda_{i_k} < \lambda_p, \quad p = n+1, ..., \infty, k = 1, ..., n. \tag{46}$$

In the other words we show that $i_k \leq n, k = 1, ..., n$ which together with (44) implies the assertion

$$\frac{|\lambda_i - \mu_i|}{\lambda_i} \leq \eta, \quad i = 1, ..., n.$$

Let us prove the first statement $\mu_k \neq \widetilde{\mu}_p$, $p = n+1, ..., \infty$. Choosing $k \in \{1, ..., n\}$ we have

$$\begin{aligned}
\frac{\widetilde{\mu}_p - \mu_k}{\mu_k} &\geq \frac{\lambda_p - \mu_k}{\mu_k} - \frac{|\widetilde{\mu}_p - \lambda_p|}{\mu_k} \\
&\geq \frac{\lambda_p - \mu_k}{\mu_k + \lambda_p}\frac{\mu_k + \lambda_p}{\mu_k} - \frac{|\widetilde{\mu}_p - \lambda_p|}{\lambda_p}\frac{\lambda_p}{\mu_k} \\
&> \gamma(1 + \frac{\lambda_p}{\mu_k}) - \gamma\frac{\lambda_p}{\mu_k} = \gamma > 0
\end{aligned}$$

which proves (45), whereas (46) follows from

$$\begin{aligned}
\frac{\lambda_p - \lambda_{i_k}}{\mu_k} &\geq \frac{\lambda_p - \mu_k}{\mu_k} - \frac{|\mu_k - \lambda_{i_k}|}{\mu_k} \\
&\geq \frac{\lambda_p - \mu_k}{\mu_k + \lambda_p}\frac{\mu_k + \lambda_p}{\mu_k} - \frac{\eta}{1 - \eta} \\
&> \gamma(1 + \frac{\lambda_p}{\mu_k}) - \gamma = \gamma\frac{\lambda_p}{\mu_k} > 0
\end{aligned}$$

□

Remark 5.1. Provided

$$\frac{\eta}{\sqrt{1-\eta}}\frac{\sqrt{D\,\|M\|}}{D - \|M\|} = \frac{\eta}{\sqrt{1-\eta}}\frac{1}{\min\limits_{\substack{k=1,...,n \\ p=n+1,...,\infty}} \frac{\lambda_p - \mu_k}{\sqrt{\lambda_p \mu_k}}} < 1 \tag{47}$$

Theorem 3.2 guarantees that we have a good approximation of the desired eigenspace. Relation (43) is of the same type as relation (47). The only difference is the use of a different measure for the spectral

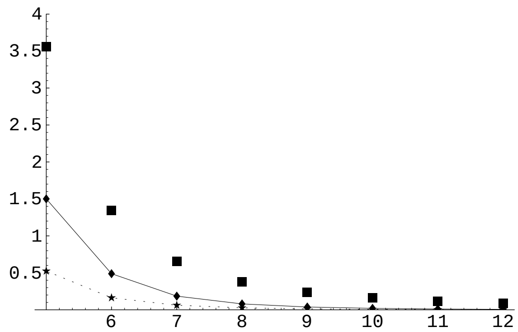

FIGURE 6. The true error and the Ritz value estimate for the approximation from the subspace $Y^3\mathbf{C}^2 \subset \mathcal{V}_n^3$ as a function of n.

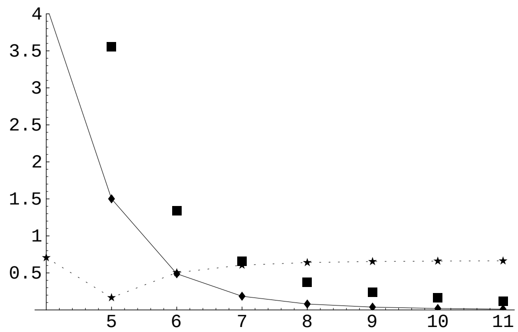

FIGURE 7. Right and wrong matching

gap. We note that the measure of the spectral gap γ is representable by a suboptimal d in formula (36). Note, that for $0 < \eta < 1$

$$\frac{\eta}{1-\eta} > \frac{\eta}{\sqrt{1-\eta}}$$

holds and that we subsequently have

$$\frac{\eta}{\sqrt{1-\eta}} \frac{\sqrt{D\,\|M\|}}{D - \|M\|} \leq \frac{\eta}{1-\eta} \frac{D + \|M\|}{D - \|M\|}.$$

Remark 5.2. An analogue to Remark 3.1 is appropriate here. If we are provided with the information that

$$\frac{\eta}{1-\eta} < \min\{\gamma_l, \gamma_r, 1\},$$

where

$$\gamma_r = \min_{\substack{k=1,\dots,n \\ p=r+n+1,\dots,\infty}} \frac{\lambda_p - \mu_k}{\lambda_p + \mu_k} \qquad \text{and} \qquad \gamma_l = \min_{\substack{k=1,\dots,n \\ p=1,\dots,r}} \frac{\mu_k - \lambda_p}{\lambda_p + \mu_k},$$

then

$$\mu_1 \leq \cdots \leq \mu_n$$

match the eigenvalues

$$\lambda_{r+1} < \lambda_{r+2} \cdots < \lambda_{r+n}.$$

Note that the assumption $0 < \min\{\gamma_l, \gamma_r, 1\}$ is equivalent to the assumption that the interval

$$\big[\|M^{-1}\|^{-1}, \|M\|\big]$$

is contained in the resolvent set of the operator $\Lambda_\perp$.

On Figure 6 we have displayed the comparison of the true error in the Rayleigh–Ritz approximation from the subspace $Y^3\mathbf{C}^2 \subset \mathcal{V}_n^3$ with our bound. The error in approximation of λ_{-1} is denoted as $(\star)$ while the error in the approximation of λ_0 is denoted as $(\blacklozenge)$. The bound, denoted as $(\blacksquare)$, follows the error in λ_0 since Theorem 2.2 guarantees the existence of the matching between the Ritz values and the part of the spectrum of the same multiplicity.

Figure 7 is even more instructive, it illustrates the real strength of our bounds. For the same example it displays

- $\blacklozenge$**:** the error between $\mu_{\min} = \min\, \sigma(M)$ and λ_0 (the expected matching),
- $\bigstar$**:** the error between $\mu_{\min}$ and λ_{-1} (the wrong matching),
- $\blacksquare$**:** our bound.

We see that as long as our bound is larger than one ($n < 7$) the Ritz value “fails to match”, i.e. it is closer to a “wrong” eigenvalue of the operator H. Based on the information

$$\min_{\substack{i=1,2\\ p\neq 0,-1}} \frac{\lambda_p - \mu_i}{\mu_i + \lambda_p} > 0.99$$

and on Theorem 5.1, we can guarantee that for $n \geq 8$ the Ritz values from the subspace $Y^3\mathbf{C}^2$ match λ_0 and λ_{-1}. This shows that — at least on the model example — our bound is quite sharp.

Remark 5.3. Theorem 5.1 is the reason we have opted for Temple–Kato inequality rather than the Residual theorem of Davis, Kahan and Weinberger [4]. The residual theorem of Davis, Kahan and Weinberger can also be employed for the subspaces X^3 and Y^3. However, we would need to compare Davis, Kahan and Weinberger's residual to the "absolute" gap if we were to guarantee that the Ritz values match the lowermost eigenvalues. This conclusion is necessary in order to produce the plots analogous to Figures 4 and 5. On the other hand, Temple–Kato inequality is a quadratic estimate which utilizes the same residual as the result of Davis, Kahan and Weinberger (cf. [4]). Furthermore, using the additional information contained in the "absolute" gap it provides the bound of the lowermost eigenvalue.

6. Conclusion

A method to compute an estimate of the accuracy of the subspace approximation method is presented. It can also be used to obtain accurate lower estimates of the desired group of eigenvalues. The bounds have to be viewed as a combination of the Ritz value bound, which gives an existence of the matching of the Ritz values and eigenvalues, and the subspace bound, which describes a nature of that matching. The main features of our theory are

- We allow any subspace $X\mathbf{C}^n \subset \mathcal{D}(h)$ to be taken as a test space.
- Our bounds contain computable quantities only.
- Our estimate of a subspace error is a function the computable quantity η and a relative gap between the Ritz values and the "unwanted" part of the spectrum.
- The bounds are computed at the price of the solution of one second order system or, if possible, at the price of the solution of two first order systems.
- Our theory is general enough not to be limited to the second order eigenvalue problems only.
- Positive definite eigenvalue problems in more then one dimension are within the scope of our theory.

The practical issues that arise in connection to these conclusions will be presented in the second part of our article.

Acknowledgment. The authors are indebted to I. Veselić, Bochum for a helpful discussion.

References

[1] F. Chatelin. *Spectral approximation of linear operators.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983. With a foreword by P. Henrici, With solutions to exercises by Mario Ahués.

[2] E. B. Davies. ICMS lecture notes on computational spectral theory. In *Spectral theory and geometry (Edinburgh, 1998)*, pages 76–94. Cambridge Univ. Press, Cambridge, 1999.

[3] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.

[4] C. Davis, W. M. Kahan, and H. F. Weinberger. Norm-preserving dilations and their applications to optimal error bounds. *SIAM J. Numer. Anal.*, 19(3):445–469, 1982.

[5] Z. Drmač. On relative residual bounds for the eigenvalues of a Hermitian matrix. *Linear Algebra Appl.*, 244:155–163, 1996.
[6] W. G. Faris. *Self-adjoint operators.* Springer-Verlag, Berlin, 1975. Lecture Notes in Mathematics, Vol. 433.
[7] F. R. Gantmacher. *The theory of matrices. Vols. 1, 2.* Chelsea Publishing Co., New York, 1959.
[8] T. Kato. *Perturbation theory for linear operators.* Springer-Verlag, Berlin, second edition, 1976. Grundlehren der Mathematischen Wissenschaften, Band 132.
[9] M. G. Larson. A posteriori and a priori error analysis for finite element approximations of self-adjoint elliptic eigenvalue problems. *SIAM J. Numer. Anal.*, 38(2):608–625 (electronic), 2000.
[10] R. Mathias and K. Veselić. A relative perturbation bound for positive definite matrices. *Linear Algebra Appl.*, 270:315–321, 1998.
[11] M. Reed and B. Simon. *Methods of modern mathematical physics. I–IV.* Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.
[12] P.-Å. Wedin. On angles between subspaces of a finite dimensional inner product space. In *Proceedings of a Conference on Matrix Pencils.* Springer Verlag, 1982.

FernUniversität Hagen, LG Mathematische Physik, Feithstr. 140, D-58084 Hagen

E-mail address: luka.grubisic@fernuni-hagen.de

E-mail address: kresimir.veselic@fernuni-hagen.de