Topics in Validated Computations
J. Herzberger (Editor)

Result verification for eigenvectors and eigenvalues

G. Mayer

Fachbereich Mathematik, Universität Rostock, D–18051 Rostock, Germany

1. INTRODUCTION

In this contribution we address to the matrix eigenproblem and to related topics such as the generalized eigenproblem, the singular value problem and the inverse eigenvalue problem.

As is well-known, the *matrix eigenproblem* for an $n \times n$ matrix A consists in finding numbers λ^* and associated vectors $x^* \neq 0$ such that the equation

$$Ax = \lambda x \tag{1}$$

holds for $\lambda = \lambda^*$ and $x = x^*$. Often A is specialized to be a Hessenberg matrix or a band matrix, or to be symmetric, Hermitian, positive definite, or tridiagonal. There are many methods to solve (1) iteratively partly depending on these properties. Among them can be found the von Mises iteration, Wielandt's inverse iteration, the QR method, the Jacobi method, Hyman's method and the bisection method based on Sturm sequences. For details we refer to [44], [99], [124], [139], [140] or any other textbook in numerical analysis. Based on these methods software packages such as EISPACK [40], [120], LAPACK [16] or the NAG–Library [101] provide algorithms to compute approximations $\tilde{\lambda}$, $\tilde{x}$ of eigenvalues λ^* and eigenvectors x^*, respectively. Shifting and balancing (permutations, scalings) is incorporated in these algorithms in order to increase their efficiency.

Although the matrix eigenproblem (1) has to be handled on the field of complex numbers for full generality (as is seen, e.g., by $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$) most applications deal with real matrices A, real eigenvalues λ^* and real eigenvectors x^*. Therefore, we restrict ourselves in the main sections of our paper to such matrices and eigenvalues adding, however, references and remarks if A or λ^* are complex. We want to consider the following problems:

1. Find an interval $[\lambda]$ which contains at least one (or exactly one) real eigenvalue λ^* of A. In particular, find $[\lambda]$ such that the bounds $\underline{\lambda}$, $\overline{\lambda}$ of $[\lambda]$ differ only in the

last few digits. In this case their common leading digits coincide with those of λ^*. This aspect is particularly important if one wants to control rounding errors when approximating λ^* on a computer.

2. Find interval enclosures $[\lambda]_i$ simultaneously for all real eigenvalues λ_i of A.

3. Find an interval vector $[x]$ which contains at least one (or exactly one) eigenvector x^* associated with an eigenvalue λ^*. Again tight bounds $\underline{x}$, $\overline{x}$ are required in order to guarantee digits of x^*.

4. Find the interval quantities in 1.–3. when A is replaced there by an interval matrix $[A] \in M_{nn}(I(\boldsymbol{R}))$. In this case $[\lambda]$ means an interval which contains for each $A \in [A]$ at least one (or exactly one) eigenvalue of A; $[x]$ means an interval vector with the following property:
For any $\lambda^* \in [\lambda]$ and any $A \in [A]$ for which λ^* is an eigenvalue, the vector $[x]$ contains at least one (or exactly one) eigenvector of A associated with λ^*.

Most of these items can be handled (more or less well) by means of classical existence theorems and error estimates. Therefore, after a section with preliminaries and a section on quadratic systems, we repeat some of these classical results in Section 4. Some of the results are also used for the interval methods to be described later on. For the matrix eigenproblem we have divided these methods in two major parts: In Section 5 we present a method for general matrices with simple, double or nearly double eigenvalues; in Section 6 we specialize A to be symmetric, allowing now multiple eigenvalues and clusters of eigenvalues. In Section 7 we consider the generalized eigenproblem and in Section 8 we derive a method to verify and enclose singular values and singular vectors. Section 9 describes an inverse eigenvalue problem for which the methods of the Sections 5 and 6 apply. In Section 10 we briefly outline additional topics in result verification for eigenpairs, and we make brief bibliographical remarks on the results of the previous sections in which we mostly dispensed with references.

All our numerical examples are programmed in PASCAL–XSC (or its predecessor PASCAL–SC). Cf. [67] for a description of this language.

2. PRELIMINARIES

In this section we list some notations, and we recall some auxiliary results used later on.

For the standard notations such as for vectors, matrices, interval quantities and their entries we refer to the contribution [55] in this book.

By $\mathrm{spec}(A)$ and $\rho(A)$ we denote the *spectrum* and the *spectral radius*, respectively, of a matrix $A \in M_{nn}(\boldsymbol{C})$, by $p_A(\lambda)$ we abbreviate its *characteristic polynomial* $\det(A - \lambda I)$. The *algebraic multiplicity* of an eigenvalue λ^* of A is its multiplicity as a zero of $p_A(\lambda)$, the *geometric multiplicity* is the dimension of the corresponding eigenspace.

By a *cluster* of eigenvalues we mean a set of eigenvalues which lie close together. Multiple eigenvalues are allowed.

Instead of I we mostly use I_n for the $n \times n$ identity matrix. We write e for the vector $e = (1, 1, \ldots, 1)^T \in V_n(\boldsymbol{R})$ and $e^{(i)}$ for the i-th column of I_n with n depending on the context.

We define $A \in M_{nn}(\boldsymbol{C})$ to be *irreducible* if for each index pair (i, j) with $i,\ j \in \{1, \ldots, n\}$ there is a finite sequence $a_{ii_1},\ a_{i_1i_2},\ a_{i_2i_3}, \ldots,\ a_{i_lj}$ of non-vanishing entries of A, where l depends on (i, j).

It is easily seen that A is irreducible if and only if there is no permutation matrix $P \in M_{nn}(\boldsymbol{R})$ such that PAP^T has the block form $\begin{pmatrix} C_{11} & O \\ C_{21} & C_{22} \end{pmatrix}$ with quadratic diagonal blocks $C_{11},\ C_{22}$.

We call $A \in M_{nn}(\boldsymbol{R})$ *nonnegative* writing $A \geq 0$, if this inequality holds entrywise. Analogously we introduce nonnegative vectors. We call $x \in V_n(\boldsymbol{R})$ *positive* using the notation $x > 0$, if $x_i > 0,\ i = 1, \ldots, n$. The partial orderings '$\leq$', '$\geq$' on $M_{nn}(\boldsymbol{R})$ are defined by $A \leq B :\Longleftrightarrow B \geq A :\Longleftrightarrow B - A \geq 0$.

By $\|x\|_2,\ \|x\|_\infty, \|A\|_2$, and $\|A\|_\infty$ we mean the *Euclidean norm* of $x \in V_n(\boldsymbol{C})$, its *maximum norm*, the *spectral norm* of $A \in M_{nn}(\boldsymbol{C})$, and its *row sum norm*, respectively.

Our first theorem considers the smoothness of eigenvalues and eigenvectors with respect to changes in the matrix entries. In order to make the eigenvector x^* unique we assume here that x^* is normalized by

$$x^*_{i_0} = 1 \tag{2}$$

where i_0 denotes a fixed component.

Theorem 1 *Let $A \in M_{nn}(\boldsymbol{C})$.*

a) The eigenvalues of A depend continuously on the entries a_{ij}.

b) All algebraic simple eigenvalues of A and the corresponding eigenvectors (normalized by (2)) are infinitely often continuously differentiable with respect to the entries a_{ij}.

Proof.

a) Since, by construction, the coefficients of $p_A(\lambda)$ depend analytically on a_{ij} it suffices to prove that the zeros $\lambda_i(a_0, \ldots, a_n),\ i = 1, \ldots, n$, of a polynomial

$$p(\lambda; a_0, \ldots, a_n) := \sum_{j=0}^{n} a_j \lambda^j, \qquad a_n \neq 0,\ n \geq 1, \tag{3}$$

depend continuously on its coefficients a_j. To this end assume that the zeros of p do not behave continuously at some point $(a^*_0, \ldots, a^*_n)$, and denote the zeros of $p(\lambda; a^*_0, \ldots, a^*_n)$ by $\lambda^*_i,\ i = 1, \ldots, n$. Then there are a real number $\varepsilon > 0$, a zero λ^*_α of $p(\lambda; a^*_0, \ldots, a^*_n)$ and a sequence $\{(a^{(k)}_0, \ldots, a^{(k)}_n)\}_{k=1}^\infty$ such that for all $k \in \boldsymbol{N}$ the inequalities

$$|a^{(k)}_j - a^*_j| \leq \frac{1}{k},\ j = 0, 1, \ldots, n,$$

and

$$|\lambda_i(a_0^{(k)}, \ldots, a_n^{(k)}) - \lambda_\alpha^*| \geq \varepsilon, \; i = 1, \ldots, n \; ,$$

hold. Letting k tend to infinity in the inequality

$$|p(\lambda_\alpha^*; a_0^{(k)}, \ldots, a_n^{(k)})| = |a_n^{(k)}| \prod_{\substack{i=1 \\ i \neq \alpha}}^{n} |\lambda_\alpha^* - \lambda_i(a_0^{(k)}, \ldots, a_n^{(k)})| \geq |a_n^{(k)}| \varepsilon^n$$

yields the contradiction

$$0 = |p(\lambda_\alpha^*; a_0^*, \ldots, a_n^*)| \geq |a_n^*| \varepsilon^n > 0.$$

b) The smoothness of algebraic simple eigenvalues λ^* follows immediately from the implicit function theorem [135], applied to the polynomial p in (3), since

$$\frac{\partial p(\lambda^*; a_0, \ldots, a_n)}{\partial \lambda} = p'(\lambda^*; a_0, \ldots, a_n) \neq 0 \tag{4}$$

for simple zeros λ^* of p.
To prove the smoothness of the corresponding normalized eigenvectors x^* we assume w.l.o.g. $i_0 = n$ in (2). Then trivially $z_1 = x_1^*, \; \ldots, \; z_{n-1} = x_{n-1}^*$ is a solution of the linear system

$$\sum_{j=1}^{n-1} (a_{ij} - \delta_{ij}\lambda^*) z_j = -(a_{in} - \delta_{in}\lambda^*), \; i = 1, \ldots, n, \tag{5}$$

where δ_{ij} denotes the Kronecker symbol. Since λ^* is a simple eigenvalue, the solution of (5) is unique, hence the $n \times (n-1)$ coefficient matrix of (5) has rank $n-1$. Therefore, one can delete some equation k of (5) such that the modified system remains uniquely solvable. To facilitate the notation we assume $k = n$. Consider now the matrices $A + E$ with $E \in M_{nn}(C)$ and let $\lambda(E)$ be an eigenvalue of $A + E$ such that $\lim_{E \to O} \lambda(E) = \lambda^*$ holds. For $\|E\|_\infty$ being sufficiently small, $\lambda(E)$ is a simple eigenvalue (which follows from (4) and from the continuity of $p'(\lambda; a_0, \ldots, a_n)$), and the system

$$\sum_{j=1}^{n-1} (a_{ij} + e_{ij} - \delta_{ij}\lambda(E)) z_j = -(a_{in} + e_{in} - \delta_{in}\lambda(E)), \; i = 1, \ldots, n-1, \tag{6}$$

has a nonsingular $(n-1) \times (n-1)$ coefficient matrix $B(E)$. Let $x(E)$ be an eigenvalue of $A + E$ associated with $\lambda(E)$. Then $x_n(E) \neq 0$ since otherwise $B(E)(x_1(E), \ldots, x_{n-1}(E))^T = 0$ which implies $x(E) = 0$. Therefore, we can assume that $x(E)$ is normalized according to (2) with $i_0 = n$. In this case $z_1 = x_1(E), \ldots, z_{n-1} = x_{n-1}(E)$ is the unique solution of (6). By Cramer's rule, this solution is a rational function of $a_{ij} + e_{ij}$ and $\lambda(E)$, hence the theorem is proved by the smoothness property of such functions and of $\lambda(E)$.

□

Note that for multiple eigenvalues the corresponding eigenvectors need not behave continuously, as is seen in [98], Ex. 3.1.5, p. 47.

Our next theorem characterizes algebraic simple eigenvalues. It is used as an auxiliary result in Section 4.

Theorem 2 *Let (x^*, λ^*) be an eigenpair of $A \in M_{nn}(\boldsymbol{C})$, let $e^{(n)}$ denote the n-th column of the $n \times n$ identity matrix I, and write $(A - \lambda^* I)_{*,i}$ for the i-th column of $A - \lambda^* I$.*

a) If $x_n^* \neq 0$ then the follwing assertions are equivalent.

(i) λ^* is an algebraic simple eigenvalue.

(ii) $B^* := \begin{pmatrix} A - \lambda^* I & -x^* \\ (e^{(n)})^T & 0 \end{pmatrix}$ is nonsingular.

(iii) $(B^*)' := ((A - \lambda^* I)_{*,1}, \ldots, (A - \lambda^* I)_{*,n-1}, -x^*)$ is nonsingular.

b) If $(x^*)^H x^* = 1$ then λ^* is an algebraic simple eigenvalue if and only if $(B^*)'' := \begin{pmatrix} A - \lambda^* I & -x^* \\ 2(x^*)^H & 0 \end{pmatrix}$ is nonsingular.

Proof.
a) (i) $\Rightarrow$ (ii)
Assume B^* to be singular with

$$B^* z = 0 \tag{7}$$

for some $z \in V_{n+1}(\boldsymbol{C}) \backslash \{0\}$. With z we construct the vector $\hat{z} := (z_1, \ldots, z_n)^T \in V_n(\boldsymbol{C})$. From (7) we get

$$(A - \lambda^* I)\hat{z} = z_{n+1}\, x^* \tag{8}$$

and

$$z_n = 0. \tag{9}$$

By (8), $\hat{z} \neq 0$. Otherwise (8) implies the contradiction $z = 0$. If $z_{n+1} = 0$ then (8), (9) and $x_n^* \neq 0$ show that $\hat{z}$ and x^* are linearly independent eigenvectors of A. If $z_{n+1} \neq 0$ then (8) yields

$$(A - \lambda^* I)^2 \frac{\hat{z}}{z_{n+1}} = 0$$

hence $\dfrac{\hat{z}}{z_{n+1}}$ is a principal vector of A. Thus in both cases λ^* is an algebraic multiple eigenvalue. This contradicts the assumption.

(ii) $\Rightarrow$ (iii)
follows by evaluating $\det(B^*)$ along the $(n+1)$-st row which gives $-\det(B^*)'$.

(iii) $\Rightarrow$ (i)
Assume that λ^* is an algebraic multiple eigenvalue. Then A has a left eigenvector y^* which is associated with this eigenvalue and which satisfies $(y^*)^H x^* = 0$. This can easily be seen by going back to the Jordan normal form of A. Since $(y^*)^H (B^*)' = 0$, the matrix $(B^*)'$ is singular contradicting the assumption.

b) " $\Rightarrow$ "
follows nearly literally as '(i) $\Rightarrow$ (ii)' in a). Replace (9) by $(x^*)^H \hat{z} = 0$ only.

" $\Leftarrow$ "
follows as '(iii) $\Rightarrow$ (i)' in a) with $(y^*)^H(B^*)' = 0$ being replaced by $((y^*)^H, 0)(B^*)'' = 0$.

□

In Section 4 we will deal with *normal* matrices A which are defined by the property

$AA^H = A^H A$.

We recall one of the standard results for such matrices.

Theorem 3 *$A \in M_{nn}(\boldsymbol{C})$ is unitarily similar to a diagonal matrix if and only if A is normal. In particular, normal matrices are exactly those that possess an orthonormal basis of eigenvectors.*

Proof. See [125], p. 311 .

□

3. QUADRATIC SYSTEMS

In this section, we consider functions f of the form

$$f : \begin{cases} V_n(\boldsymbol{R}) & \rightarrow & V_n(\boldsymbol{R}) \\ x & \mapsto & f(x) := r + Sx + Tx^2 \end{cases} \tag{10}$$

with $r \in V_n(\boldsymbol{R})$, $S \in M_{nn}(\boldsymbol{R})$,

$$T : \begin{cases} V_n(\boldsymbol{R}) \times V_n(\boldsymbol{R}) & \rightarrow & V_n(\boldsymbol{R}) \\ (x, y) & \mapsto & \left(\sum_{j=1}^{n} \sum_{k=1}^{n} t_{ijk} x_k y_j \right) \end{cases} ,$$

and $Tx^2 := T(x,x)$. The matrix S represents a linear mapping whereas T is a bilinear one. We identify T with the three dimensional array $(t_{ijk}) \in M_{nnn}(\boldsymbol{R})$. By means of the given quantities r_i, s_{ij} and t_{ijk} we want to construct an interval vector which, under weak conditions, contains a fixed point of f. Since for the eigenproblem as well as for the singular value problem there are functions of the form (10) for which a fixed point is a pair of an eigenvalue/eigenvector, and of a singular value/singular vector, respectively, one then has automatically a verification method for each of these problems.

Note that the system $x = f(x)$ is equivalent to the quadratic system $x - f(x) = 0$ which justifies the title of this section.

Theorem 4 *Let f be defined as in (10), and let*

$$\rho := \|r\|_\infty, \ \sigma := \|S\|_\infty, \ \tau := \|T\|_\infty := \max_{1\le i\le n}\left\{\sum_{j=1}^{n}\sum_{k=1}^{n}|t_{ijk}|\right\}, \tag{11}$$

$$\sigma \le 1, \ \Delta := (1-\sigma)^2 - 4\rho\tau \ge 0, \tag{12}$$

$$\left.\begin{array}{lcl} \beta^- & := & (1-\sigma-\sqrt{\Delta})/(2\tau) = 2\rho/(1-\sigma+\sqrt{\Delta}) \ , \\ \beta^+ & := & (1-\sigma+\sqrt{\Delta})/(2\tau) \ . \end{array}\right\} \tag{13}$$

a) *For any $\beta \in [\beta^-, \beta^+]$ the function f has at least one fixed point x^* in $[x]^0 := [-\beta, \beta]e$, and the iteration*

$$[x]^{(k+1)} := f([x]^{(k)}), \ k = 0, 1, \ldots, \tag{14}$$

converges to some interval vector $[x]^*$ with

$$x^* \in [x]^* \subseteq [x]^{(k)} \subseteq [x]^{(k-1)} \subseteq \ldots \subseteq [x]^{(0)}, \ k \in \boldsymbol{N}. \tag{15}$$

b) *For any $\beta \in [\,\beta^-, (\beta^- + \beta^+)/2\,)$ the function f has a unique fixed point x^* in $[x]^{(0)} := [-\beta, \beta]e$, and (15) holds with $[x]^* = [x^*, x^*]$, i. e., (14) converges to x^*.*

c) *For any $\beta \in (\beta^-, \beta^+)$ and $[x] := [-\beta, \beta]e$ the function f satisfies*

$$f([x]) \subseteq \mathrm{int}([x]). \tag{16}$$

Proof.

a) Let

$$T([x],[y]) := \sum_{j=1}^{n}\sum_{k=1}^{n} t_{ijk}[x]_k[y]_j \ .$$

We show that $[x] := [-\beta, \beta]e$ satisfies

$$f([x]) = r + S[x] + T[x]^2 \subseteq [x] \tag{17}$$

whenever $\beta \in [\beta^-, \beta^+]$. Then Brouwer's fixed point theorem yields the first part of the assertion, since trivially $f(x) \in f([x])$ holds for all $x \in [x]$.
The inclusion (17) is equivalent to

$$\begin{array}{rcl} [-\beta,\beta]e & \supseteq & r + \left(\sum_{j=1}^{n} s_{ij}[-\beta,\beta]\right) + \left(\sum_{j=1}^{n}\sum_{k=1}^{n} t_{ijk}[-\beta,\beta][-\beta,\beta]\right) \\ & = & r + [-\beta,\beta]|S|e + [-\beta^2,\beta^2]|T|e^2 \end{array} \tag{18}$$

with the bilinear operator $|T| := (|t_{ijk}|)$. Since for two intervals $[a]$, $[b]$ the inclusion $[a] \subseteq [b]$ holds if and only if $|\text{mid}([a]) - \text{mid}([b])| + \frac{1}{2}w([a]) \leq \frac{1}{2}w([b])$ is valid, (18) is equivalent to

$$|r| + \beta|S|e + \beta^2|T|e^2 \leq \beta e \; . \tag{19}$$

The inequality (19) certainly holds if

$$\rho + \beta\sigma + \beta^2\tau \leq \beta$$

whence $\beta \in [\beta^-, \beta^+]$.
The second part of a) follows easily from (17).

b) Let $[x]^*$ be the limit of (14). Then we get for the widths

$$\begin{aligned} w([x]^*) &= w(f([x]^*)) \\ &= |S|w([x]^*) + \left(\sum_{j=1}^{n} \sum_{k=1}^{n} |t_{ijk}| w([x]_k^*[x]_j^*) \right) \; , \end{aligned}$$

and with

$$w_\infty := \|w([x]^*)\|_\infty$$

and

$$\begin{aligned} w([x]_k^*[x]_j^*) &\leq w([x]_k^*)|[x]_j^0| + |[x]_k^0|w([x]_j^*) \\ &\leq \beta\{w([x]_k^*) + w([x]_j^*)\} \end{aligned}$$

we obtain

$$w_\infty \leq \sigma w_\infty + 2\beta\tau w_\infty \; . \tag{20}$$

If $w_\infty > 0$ then (20) implies $1 \leq \sigma + 2\beta\tau$, hence

$$\beta \geq \frac{1-\sigma}{2\tau} = \frac{\beta_1 + \beta_2}{2}$$

which contradicts our assumption. Thus, $w_\infty = 0$, and $x^* \in [x]^*$ yields $[x]^* = [x^*, x^*]$.

Since each fixed point of f which is contained in $[x]^{(0)}$, remains in $[x]^{(k)}$, $k \in \boldsymbol{N}$, the uniqueness of x^* follows from $w_\infty = 0$.

c) Since $[a] \subseteq \mathrm{int}([b])$ holds for two intervals $[a]$, $[b]$ if and only if $|\mathrm{mid}([a]) - \mathrm{mid}([b])| + \frac{1}{2}w([a]) < \frac{1}{2}w([b])$, one can proceed as in a) ending up with

$$\rho + \beta\sigma + \beta^2\tau < \beta$$

which means $\beta \in (\beta^-, \beta^+)$.

□

The assumptions (12) are necessary for the existence of β^-, β^+, and for $\beta^- \geq 0$. In our applications we shall have $\rho \approx 0$, $\sigma \approx 0$ so that these assumptions are fulfilled.

Instead of one quadratic system one sometimes has to consider the interval arithmetic evaluation of f with respect to x, r, S and T, i. e., interval functions defined by

$$f([x],[r],[S],[T]) := [r] + [S][x] + [T][x]^2$$

with $[r] \in V_n(I(\boldsymbol{R}))$, $[S] \in M_{nn}(I(\boldsymbol{R}))$, and $[T] = ([t]_{ijk}) \in M_{nnn}(I(\boldsymbol{R}))$. It is clear that an interval vector $[x]$, which satisfies

$$f([x],[r],[S],[T]) \subseteq [x]$$

contains for *each* function $f(x,r,S,T) := r + Sx + Tx^2$ with $r \in [r]$, $S \in [S]$, $T \in [T]$ at least one fixed point. We leave it to the reader as an easy problem to reformulate Theorem 4 for interval data. We only mention that ρ, σ, τ have to be replaced by

$$\rho := \|\, |[r]|\, \|_\infty, \quad \sigma := \|\, |[S]|\, \|_\infty, \quad \tau := \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^{n} \sum_{k=1}^{n} |[t_{ijk}]| \right\}.$$

4. ESTIMATES FOR EIGENVALUES AND EIGENVECTORS

We recall here some existence theorems for eigenvalues λ^* and corresponding eigenvectors x^*, respectively, which also yield bounds for λ^* and x^*.

Probably the most famous estimate is contained in the following theorem.

Theorem 5 *Let $A \in M_{nn}(\boldsymbol{C})$, $\lambda^* \in spec(A)$. Then $|\lambda^*| \leq \|A\|$, where $\|\ .\ \|$ is allowed to be any (submultiplicative) matrix norm. In particular,*

$$\rho(A) \leq \|A\|\,. \tag{21}$$

Proof.

For any $x \in \boldsymbol{C}^n$ let $(x, 0, \ldots, 0) \in \boldsymbol{C}^{n \times n}$ be the matrix of which the columns two up to n are zero. It is easily seen that $\|\ .\ \|_v$ defined by $\|x\|_v := \|(x, 0, \ldots, 0)\|$ is a vector norm. Applying this norm on the equation $Ax^* = \lambda^* x^*$, $x^* \neq 0$, and using the submultiplicativity of the matrix norm $\|\ .\ \|$ yields

$$|\lambda^*| \|x\|_v = \|Ax^*\|_v \leq \|A\|\, \|(x, 0, \ldots, 0)\| = \|A\|\, \|x^*\|_v$$

whence the assertion follows.

□

We remark that Theorem 5 holds, in particular, for any operator norm since such norms are always submultiplicative; cf. [124].

Let SAS^{-1} be the Jordan normal form of $A \in M_{nn}(\boldsymbol{C})$ and let $D_\varepsilon := \mathrm{diag}(\varepsilon, \varepsilon^2, \ldots, \varepsilon^n) \in M_{nn}(\boldsymbol{R})$ for $\varepsilon > 0$. Then

$$\|A\| := \|D_\varepsilon^{-1} SAS^{-1} D_\varepsilon\|_\infty \tag{22}$$

is an operator norm which is generated by the vector norm $\|x\| := \|D_\varepsilon^{-1} Sx\|_\infty$, $x \in \boldsymbol{C}^n$. The particular form of the entries of $D_\varepsilon^{-1} SAS^{-1} D_\varepsilon$ yields at once the inequality $\|A\| \le \rho(A) + \varepsilon$, i. e., $\rho(A)$ can be approximated arbitrarily well by some appropriate matrix norm. Nevertheless this does not mean that there is always a matrix norm for which equality holds in (21). As a counter-example consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{23}$$

for which $\rho(A) = 0$, but $\|A\| > 0$ for each matrix norm $\|\cdot\|$ because of $A \ne 0$.

For particular norms, equality holds in (21) if A is diagonalizable (symmetric or normal, e. g.). This is easily proved using (22) with $\varepsilon = 1$.

Normally, $\rho(A)$ is not an eigenvalue of A. In some cases, however, $\rho(A) \in \mathrm{spec}(A)$ can be guaranteed. For nonnegative matrices this is valid by the subsequent Theorem of Perron and Frobenius.

Theorem 6 *Let $0 \le A \in M_{nn}(\boldsymbol{R})$ be irreducible. Then the following assertions hold.*

a) $\rho(A)$ is an eigenvalue of A.

b) There is a positive eigenvector x_P^* associated with $\rho(A)$. (Frequently, x_P^* is called a Perron vector *of A.)*

c) If $A \le B$ for some $B \in M_{nn}(\boldsymbol{R}) \backslash \{A\}$, then $\rho(A) < \rho(B)$.

Proof. See [133], pp. 28–31, e. g.

□

If one drops the irreducibility of A then $x > 0$ in b) must be weakened to $x \ge 0$, and the strict inequality in c) must be replaced be '$\le$' as is seen by the matrices A and $2A$ with A from (23). Cf. [133] for details. With Theorem 6 one can prove the following comparison result for the spectral radii.

Theorem 7 *[133], p. 47.*
Let $0 \le |B| \le A \in M_{nn}(\boldsymbol{R})$. Then $\rho(B) \le \rho(A)$.

□

For nonnegative matrices (21) can be specialized in the following way.

Theorem 8 *Let* $0 \leq A \in M_{nn}(\boldsymbol{R}),\ 0 < x \in V_n(\boldsymbol{R}),\ 0 \leq \alpha \in \boldsymbol{R}$. *If*

$$Ax \leq \alpha x \tag{24}$$

holds, then $\rho(A) \leq \alpha$. *If strict inequality holds in (24) then* $\rho(A) < \alpha$.

Proof.
The inequality (24) implies

$$\max_{1\leq i\leq n} \frac{(Ax)_i}{x_i} \leq \alpha \, . \tag{25}$$

Since the lefthand side of (25) equals the weighted row sum norm $\|D_x^{-1}AD_x\|_\infty$ with $D_x := \mathrm{diag}(x_1, \ldots, x_n)$ the assertion follows from Theorem 5.

□

If one restricts A to be Hermitian or normal, the estimate in Theorem 5 can be considerably sharpened if one only looks for bounds of single eigenvalues. In this case we are going to prove several classical estimates in a unified way following the lines given in [94]. The crucial tool is the subsequent lemma.

Lemma 9 *Let* $A = A^H \in M_{nn}(\boldsymbol{C})$ *and let* h *be a rational function with real coefficients, for which* $h(A)$ *exists. If* $x \in V_n(\boldsymbol{C})\backslash\{0\}$ *is a vector such that*

$$x^H h(A) x \leq 0 \tag{26}$$

holds, then there is at least one eigenvalue λ^* *of* A *which satisfies* $h(\lambda^*) \leq 0$.

Proof.
The matrix $h(A)$ is Hermitian but not positive definite by (26). Therefore, $h(A)$ has at least one eigenvalue $\mu^* \leq 0$. Since the eigenvalues of $h(A)$ are given by $h(\lambda)$, $\lambda \in \mathrm{spec}(A)$, the assertion follows.

□

To formulate the announced collection of estimates, we need the following definitions.

Definition 10 *Let* $\alpha \in \boldsymbol{C},\ x \in V_n(\boldsymbol{C})\backslash\{0\},\ A \in M_{nn}(\boldsymbol{C})$. *Define*

$$\begin{aligned}
m_i &:= x^H A^i x \, , \\
m_{ij} &:= (A^i x)^H (A^j x) \, , \\
R_x &:= \frac{m_{01}}{m_{00}} = \frac{m_1}{m_0} \, , && \textit{(Rayleigh quotient)} \\
T_x(\alpha) &:= \frac{m_{11} - \bar{\alpha} m_{01}}{m_{10} - \bar{\alpha} m_{00}} \, , && \textit{(Temple quotient)} \\
\varepsilon_x^2 &:= \frac{m_{11}}{m_{00}} - \frac{m_{10} m_{01}}{m_{00}^2} = \frac{m_{11}}{m_{00}} - |R_x|^2 \in \boldsymbol{R} \, .
\end{aligned}$$

□

Note, that for Hermitian matrices A, the Rayleigh quotient is real and $T_x(\alpha) = \dfrac{m_2 - \alpha m_1}{m_1 - \alpha m_0} \in \boldsymbol{R}$, provided that $\alpha \in \boldsymbol{R}$.

Theorem 11 *Let $A = A^H \in M_{nn}(\boldsymbol{C})$, $\alpha \in \boldsymbol{R}$, $x \in V_n(\boldsymbol{C})\backslash\{0\}$. Furthermore, let f be a rational function with real coefficients, for which $f(A)$ exists. Then the following assertions hold.*

a) $\exists\ \lambda^* \in \operatorname{spec}(A): \quad |\lambda^* - \alpha| \le \dfrac{\|Ax - \alpha x\|_2}{\|x\|_2}$ *(Walker/Weston [134])*

b) $\exists\ \lambda^* \in \operatorname{spec}(A): \quad \dfrac{|\lambda^* - \alpha|}{|\lambda^*|} \le \dfrac{\|Ax - \alpha x\|_2}{\|Ax\|_2}$ *, provided $Ax \neq 0$*
(Stummel/Hainer [126], p. 222)

c) $\exists\ \lambda^* \in \operatorname{spec}(A): \quad |f(\lambda^*)| \le \dfrac{\|f(A)x\|_2}{\|x\|_2}$ *(Stoer/Bulirsch [124], p. 392)*

d) $\exists\ \lambda^* \in \operatorname{spec}(A): \quad |\lambda^* - R_x| \le \varepsilon_x$ *(Krylov/Bogolyubov [74], Weinstein [136])*

e) $\exists\ \lambda^* \in \operatorname{spec}(A):$ *λ^* lies between α and $T_x(\alpha)$* *(Temple [128], Wielandt [137])*

f) $\exists\ \lambda^* \in \operatorname{spec}(A):$ *λ^* lies between α and* $\tau_x(\alpha) := \alpha + \dfrac{x^H x}{x^H(A - \alpha I)^{-1}x}$
(Löwdin [78])

g) $\exists\ \lambda^* \in \operatorname{spec}(A): \quad \underline{\mu} := \min\limits_{1\le l\le n} \dfrac{(Ax)_l}{x_l} \le \lambda^* \le \bar{\mu} := \max\limits_{1\le l\le n} \dfrac{(Ax)_l}{x_l}$ *, provided $A \in M_{nn}(\boldsymbol{R})$ and $x_l \in \boldsymbol{R}\backslash\{0\}$, $l = 1, \ldots, n$.* *(Collatz [28])*

Proof. [94]
a) – f) Apply Lemma 9 with $h(t)$ as listed below with b denoting the bound of the respective inequality which is to be shown. Then $x^H h(A)x = 0$ by straight forward computations, and Lemma 9 proves the assertion.

a) $h(t) := (t - \alpha)^2 - b^2$;

b) $h(t) := (t - \alpha)^2 - b^2 t^2$;

c) $h(t) := f(t)^2 - b^2$;

d) $h(t) := (t - R_x)^2 - \varepsilon_x^2$;

e) $h(t) := (t - b)(t - \alpha)$, $\quad (b := T_x(\alpha))$;

f) $h(t) := (t - b)/(t - \alpha)$, $\quad (b := \tau_x(\alpha))$.

g) Choose $h(t) := (t-\underline{\mu})(t-\overline{\mu})$ in Lemma 9. Then

$$\begin{aligned} x^H h(A)x &= \sum_{l=1}^{n}\left\{\left(\frac{(Ax)_l}{x_l}\right)^2 - \overline{\mu}\frac{(Ax)_l}{x_l} - \underline{\mu}\frac{(Ax)_l}{x_l} + \underline{\mu}\overline{\mu}\right\} x_l^2 \\ &= \sum_{l=1}^{n}\left(\frac{(Ax)_l}{x_l} - \underline{\mu}\right)\left(\frac{(Ax)_l}{x_l} - \overline{\mu}\right) x_l^2 \le 0 \, . \end{aligned}$$

Hence there is an eigenvalue λ^* of A satisfying $h(\lambda^*) = (\lambda^* - \underline{\mu})(\lambda^* - \overline{\mu}) \le 0$ which proves the assertion.

□

Note that Theorem 11g) does no longer hold if $x_l = 0$ is permitted and $\underline{\mu}$, $\overline{\mu}$ are built only with the non–vanishing components of x. This is shown by the example $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ and $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, since then $\mathrm{spec}(A) = \{0, 2\}$ and $\underline{\mu} = \overline{\mu} = 1$.

We now generalize Theorem 11 to normal matrices and $\alpha \in \boldsymbol{C}$. To this end we need the following analogue of Lemma 9.

Lemma 12 *Let $A \in M_{nn}(\boldsymbol{C})$ be normal, and let h be a rational function in the variables s, t, for which $h(A^H, A)$ exists. If $x \in V_n(\boldsymbol{C})\backslash\{0\}$ is a vector such that*

$$Re\{x^H h(A^H, A)x\} \le 0 \tag{27}$$

then there is at least one eigenvalue λ^* of A which satisfies $Re\{h(\overline{\lambda}, \lambda)\} \le 0$.

Proof.
By Theorem 3 there is a basis of orthonormal eigenvectors $(x^*)^1, \ldots, (x^*)^n$ of A corresponding to the eigenvalues $\lambda_1^*, \ldots, \lambda_n^*$, with which x can be represented as

$$x = \sum_{i=1}^{n} \alpha_i (x^*)^i \, .$$

Taking into account that A is normal, we get

$$\begin{aligned} x^H h(A^H, A)x &= \sum_{i=1}^{n} \left((x^*)^i\right)^H h(\overline{\lambda_i^*}, \lambda_i^*)(x^*)^i |\alpha_i|^2 \\ &= \sum_{i=1}^{n} h(\overline{\lambda_i^*}, \lambda_i^*)|\alpha_i|^2 \, . \end{aligned}$$

Then (27) yields

$$\sum_{i=1}^{n} \mathrm{Re}\{h(\overline{\lambda_i^*}, \lambda_i^*)\}|\alpha_i|^2 \le 0 \, ,$$

which proves the assertion.

□

Theorem 13 *Let $A \in M_{nn}(\boldsymbol{C})$ be normal, $\alpha \in \boldsymbol{C}$, $x \in V_n(\boldsymbol{C})\backslash\{0\}$. Furthermore, let f be a rational function, for which $f(A)$ exists. Then the assertions a) – d) of Theorem 11 remain valid. The assertions e) – g) change as follows:*

If K_T, K_τ and K_μ, respectively, are closed discs in the complex plane which contain α, $T_x(\alpha)$, and α, $\tau_x(\alpha)$, and all $\mu_l := \frac{(Ax)_l}{x_l}$, respectively, then each of these discs contains at least one eigenvalue of A. In g) only the restriction $x_l \in C\backslash\{0\}$, $l = 1, \ldots, n$, is required.

Proof.
a) – d) are proved by means of Lemma 12 analogously to Theorem 11 using the following functions h. Again, b denotes the bound of the respective inequality.

a) $h(s,t) := (s-\overline{\alpha})(t-\alpha) - b^2$;

b) $h(s,t) := (s-\overline{\alpha})(t-\alpha) - b^2 st$;

c) $h(s,t) := \overline{f(s)}f(t) - b^2$;

d) $h(s,t) := (s-\overline{R_x})(t-R_x) - \varepsilon_x^2$.

e) Let $h(s,t) := \kappa(s - \overline{T_x(\alpha)})(t-\alpha)$ with $\kappa \in C\backslash\{0\}$ at the moment being arbitrary. Then a simple computation yields $x^H h(A^H, A)x = 0$, hence by Lemma 12 there is an eigenvalue λ^* of A such that

$$\mathrm{Re}\left\{\kappa\left(\overline{\lambda^*} - \overline{T_x(\alpha)}\right)(\lambda^* - \alpha)\right\} \le 0 \tag{28}$$

holds. W.l.o.g. we can assume that α and $T_x(\alpha)$ do not belong to spec(A). Then (28) is equivalent to

$$\mathrm{Re}\{\kappa\frac{\lambda^*-\alpha}{\lambda^* - T_x(\alpha)}\} \le 0 \; , \tag{29}$$

with α differing from $T_x(\alpha)$, since $\alpha = T_x(\alpha)$ yields the contradiction

$$0 = m_{11} - \overline{\alpha}m_{01} - \alpha m_{10} + \alpha\overline{\alpha}m_{00} = \|Ax - \alpha x\|_2^2 \; ,$$

i.e., $\alpha \in$ spec(A). Let $K \subseteq K_T$ be a closed disc containing α, $T_x(\alpha)$ on its boundary ∂K. Choose $\gamma \in \partial K\backslash\{\alpha, T_x(\alpha)\}$ such that α, γ, $T_x(\alpha)$ are met in this order when running through ∂K in a positively oriented sense. Fix now κ by $\kappa := i\, e^{-i \arg((\gamma-\alpha)/(\gamma - T_x(\alpha)))}$ Then the linear fractional transformation

$$w = g(z) := \kappa\,\frac{z-\alpha}{z - T_x(\alpha)}$$

maps α, γ, $T_x(\alpha)$ into the points 0, $i\left|\frac{\gamma-\alpha}{\gamma - T_x(\alpha)}\right|$, ∞ of the imaginary axis. Hence, by the usual properties of linear fractional transformations (cf. [2], e.g.), K is mapped onto the closed left half plane, which, by (29), also contains $g(\lambda^*)$. Since g is bijective, λ^* is contained in K and therefore in K_T.

f) follows analogously to e) with $h(s,t) := \kappa\,\frac{t-\tau(\alpha)}{t-\alpha}$, and an appropriate κ. Here, h

does not depend explicitly on s.

g) Let β, γ, δ be three different points of ∂K_μ which are met in this order when running through ∂K in a positively oriented sense. Let β, γ, δ differ from μ_l, $l = 1, \ldots, n$, and assume w.l.o.g. $\delta \notin \operatorname{spec}(A)$. Then, analogously to e), the transformation

$$w = g(z) := \kappa \frac{z-\beta}{z-\delta} \quad \text{with} \quad \kappa := i\, e^{-i \arg((\gamma-\beta)/(\gamma-\delta))}$$

maps K_μ 1–1 onto the closed left half plane. Defining $h(s,t) := \kappa(s-\overline{\delta})(t-\beta)$ yields

$$\begin{aligned} x^H h(A^H, A)x &= \kappa \sum_{l=1}^{n} (\overline{\mu}_l \mu_l - \overline{\delta}\mu_l - \beta\overline{\mu}_l + \overline{\delta}\beta)|x_l|^2 \\ &= \sum_{l=1}^{n} \kappa(\overline{\mu}_l - \overline{\delta})(\mu_l - \beta)|x_l|^2 \\ &= \sum_{l=1}^{n} g(\mu_l)|\mu_l - \delta|^2 |x_l|^2 . \end{aligned}$$

Since $\mu_l \in K_\mu$ implies $\operatorname{Re}\{g(\mu_l)\} \le 0$, we obtain $\operatorname{Re}\{x^H h(A^H, A)x\} \le 0$, hence by Lemma 12 there is an eigenvalue λ^* of A which satisfies

$$0 \ge \operatorname{Re}\{h(\overline{\lambda^*}, \lambda^*)\} = \operatorname{Re}\{\kappa(\overline{\lambda^*} - \overline{\delta^*})(\lambda^* - \beta)\} ,$$

whence $\operatorname{Re}\{g(\lambda^*)\} \le 0$ and $\lambda^* \in K_\mu$.

□

In our next theorem we put together several variants of the classical theorem of Gershgorin. To formulate them we recall the definition of a Gershgorin disc.

Definition 14 *Let* $A \in M_{nn}(\boldsymbol{C})$, $r_i := \sum_{\substack{j=1 \\ j\ne i}}^{n} |a_{ij}|$.

Then $K_{r_i}(a_{ii}) := \{ z \in \boldsymbol{C} |\ |z - a_{ii}| \le r_i \}$ *is called* Gershgorin disc, $K_{r_i}(a_{ii}) \cap \boldsymbol{R}$ *is called* Gershgorin interval. *The set* $K(A)$ *is defined by*

$$K(A) := \bigcup_{i=1}^{n} K_{r_i}(a_{ii}) .$$

Theorem 15 *Let* $A \in M_{nn}(\boldsymbol{C})$. *Then the following assertions hold.*

a) $K(A)$ *contains all eigenvalues of* A.

b) If the union $M_1 := \bigcup_{k=1}^{m} K_{r_{i_k}}(a_{i_k i_k})$ *and the union* $M_2 = K(A)\backslash M_1$ *of the remaining Gershgorin discs are disjoint, then* M_1 *contains exactly* m *eigenvalues and* M_2 *contains* $n - m$ *eigenvalues.*

c) If A *is irreducible and if* λ^* *is an eigenvalue of* A *which lies on the boundary* $\partial K(A)$ *of* $K(A)$, *then* λ^* *lies on the boundary of each Gershgorin disc.*

d) If A *is normal then each Gershgorin disc contains at least one eigenvalue of* A.

Proof.
a) Let

$$Ax^* = \lambda^* x^* \quad \text{with} \quad |x_s^*| = \max_{1 \le i \le n} |x_i^*| = 1 \ . \tag{30}$$

Then by (30) we get

$$|a_{ss} - \lambda^*| = |a_{ss} x_s^* - \lambda^* x_s^*| \le \sum_{\substack{j=1 \\ j \ne s}}^{n} |a_{sj}| \; |x_j^*| \le r_s \ , \tag{31}$$

hence $\lambda^* \in K_{r_s}(a_{ss})$.

b) W.l.o.g. assume $i_k = k$, i. e., $M_1 = \bigcup_{k=1}^{m} K_{r_k}(a_{kk})$. Let $D := \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, $C := A - D$, $A_\varepsilon := D + \varepsilon C$. Then

$$K(A_\varepsilon) \subseteq K(A), \; 0 \le \varepsilon \le 1 \ . \tag{32}$$

M_1 contains exactly the m eigenvalues $a_{11}, a_{22}, \ldots, a_{mm}$ of A_0. Letting ε tend from 0 to 1 the assertion follows from (32) and Theorem 1a).

c) If $\lambda^* \in \partial K(A)$ then $\lambda^* \in \partial K_{r_s}(a_{ss})$ with s from (30). Hence equality holds in (31) implying

$$|x_j| = 1 \ \text{ for each } j \text{ with } a_{sj} \ne 0. \tag{33}$$

Choose now $k \in \{1, \ldots, n\}$. By the irreducibility of A there is a sequence $a_{si_1}, a_{i_1 i_2}, \ldots, a_{i_l k}$ of non–vanishing entries of A. In particular, $a_{si_1} \ne 0$, hence $|x_{i_1}| = 1$ by (33), and the steps of this proof can be repeated first with s being replaced by i_1, then with i_2 replacing i_1, etc., yielding $|x_k| = 1$ and $\lambda^* \in \partial K_{r_k}(a_{kk})$ at the end. Since k was arbitrary, the assertion is proved.

d) follows from Theorem 13d) with x being the i–th column of the identity matrix. Then $R_x = a_{ii}$ and $\varepsilon_x^2 = \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ij}|^2 \le r_i^2$.

□

Theorem 15d) need not be true for non–normal matrices, as is shown by the example $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ where both eigenvalues -1 and 3 are not contained in $K_1(1)$.

We conclude this section by a result of Wilkinson [138], III, 49., which yields bounds for eigenpairs.

Theorem 16 *Let $A = A^T \in M_{nn}(\boldsymbol{R})$, $\tilde{x} \in V_n(\boldsymbol{R}) \backslash \{0\}$, $\tilde{\lambda} \in \boldsymbol{R}$, $r := \|A\tilde{x} - \tilde{\lambda}\tilde{x}\|_2$. Then the following assertions hold.*

a) The interval $\left[\tilde{\lambda} - \dfrac{r}{\|\tilde{x}\|_2}, \tilde{\lambda} + \dfrac{r}{\|\tilde{x}\|_2}\right]$ *contains at least one eigenvalue of A.*

b) If $a \ge r$, $\|\tilde{x}\|_2 = 1$, and if $[\tilde{\lambda} - a, \tilde{\lambda} + a]$ contains a simple eigenvalue λ^* but no more eigenvalues of A, then there is an eigenvector x^* associated with λ^*, satisfying $\|x^*\|_2 = 1$ and

$$\|x^* - \tilde{x}\|_2^2 \le \frac{r^2}{a^2} + \left\{1 - \left(1 - \frac{r^2}{a^2}\right)^{1/2}\right\}^2 =: \varepsilon^2. \tag{34}$$

In particular, one gets $x^* \in \tilde{x} + [-\varepsilon, \varepsilon]\ (1, \ldots, 1)^T$.

Proof.

a) W.l.o.g. we assume $\tilde{\lambda} \notin \mathrm{spec}(A) = \{\lambda_1^*, \ldots, \lambda_n^*\}$. Let $U \in M_{nn}(\boldsymbol{R})$ be an orthogonal matrix which transforms A to diagonal form, i. e., $UAU^T = D := \mathrm{diag}(\lambda_1^*, \ldots, \lambda_n^*)$. Then, with $z := A\tilde{x} - \tilde{\lambda}\tilde{x}$, we get

$$Uz = UAU^TU\tilde{x} - \tilde{\lambda}U\tilde{x} = (D - \tilde{\lambda}I)U\tilde{x}\ ,$$

whence

$$\tilde{x} = U^T(D - \tilde{\lambda}I)^{-1}Uz$$

and

$$\begin{aligned}\|\tilde{x}\|_2 &\le \|U^T(D - \tilde{\lambda}I)^{-1}U\|_2\ r = \|(D - \tilde{\lambda}I)^{-1}\|_2\, r \\ &= \max_{1 \le i \le n} \frac{1}{|\lambda_i^* - \tilde{\lambda}|} \cdot r\ .\end{aligned}$$

Thus

$$\min_{1 \le i \le n} |\lambda_i^* - \tilde{\lambda}| \le \frac{r}{\|\tilde{x}\|_2}\ ,$$

which proves a).

b) We adopt the notation of a), denoting by u^i the columns of U and assuming w.l.o.g. $\lambda^* = \lambda_1$. Let $\tilde{x}$ be represented as

$$\tilde{x} = \sum_{i=1}^n \alpha_i u^i$$

with $\alpha_1 \ge 0$. (If $\alpha_1 < 0$, replace u^i by $-u^i$.) Then

$$z = \sum_{i=1}^n \alpha_i(\lambda_i^* - \tilde{\lambda})u^i$$

and

$$r^2 = z^Tz = \alpha_1^2(\lambda_1^* - \tilde{\lambda})^2 + \sum_{i=2}^n \alpha_i^2(\lambda_i^* - \tilde{\lambda})^2 \ge \alpha_1^2(\lambda_1^* - \tilde{\lambda})^2 + a^2\sum_{i=2}^n \alpha_i^2 \ge a^2\sum_{i=2}^n \alpha_i^2$$

hence

$$\sum_{i=2}^{n} \alpha_i^2 \leq \frac{r^2}{a^2} \,. \tag{35}$$

This yields

$$1 = \|\tilde{x}\|_2^2 = \sum_{i=1}^{n} \alpha_i^2 \leq \alpha_1^2 + \frac{r^2}{a^2}$$

and

$$\alpha_1 \geq \left(1 - \frac{r^2}{a^2}\right)^{1/2} \,, \tag{36}$$

whence

$$1 - \alpha_1 \leq 1 - \left(1 - \frac{r^2}{a^2}\right)^{1/2} \,. \tag{37}$$

By (35) and (37) we finally get

$$\|\tilde{x} - u^1\|_2^2 = (\alpha_1 - 1)^2 + \sum_{i=2}^{n} \alpha_i^2 \leq \varepsilon^2 \,. \tag{38}$$

□

Note that $\|\tilde{x}\|_2 = 1$ enters only in the last four inequalities of the proof. If $\|\tilde{x}\|_2 \neq 1$ in Theorem 16b) then (34) can be generalized to

$$\|x^* - \tilde{x}\|_2^2 \leq \frac{r^2}{a^2} + \max\left\{(1 - \|\tilde{x}\|_2)^2, \left\{1 - \left(\|\tilde{x}\|_2^2 - \frac{r^2}{a^2}\right)^{1/2}\right\}^2\right\}$$

provided the square root exists. This follows from

$$\|\tilde{x} - u^1\|_2^2 \leq (\alpha_1 - 1)^2 + \frac{r^2}{a^2} \tag{39}$$

which is independent from the value of $\|\tilde{x}\|_2$. For $\alpha_1 \geq 1$ we have $\|\tilde{x}\|_2 \geq \alpha_1 \geq 1$, hence (39) yields

$$\|\tilde{x} - u^1\|_2^2 \leq (\|\tilde{x}\|_2 - 1)^2 + \frac{r^2}{a^2} \,.$$

For $0 \leq \alpha_1 < 1$ and $\|\tilde{x}\|_2^2 - \frac{r^2}{a^2} \geq 0$, we get $\alpha_1 \geq \left(\|\tilde{x}\|_2^2 - \frac{r^2}{a^2}\right)^{1/2}$ analogously to (36). A similar conclusion as for (38) proves then

$$\|\tilde{x} - u^1\|_2^2 \leq \frac{r^2}{a^2} + \left\{1 - \left(\|\tilde{x}\|_2^2 - \frac{r^2}{a^2}\right)^{1/2}\right\}^2 \,.$$

5. GENERAL MATRICES

To verify simple *real* eigenvalues and corresponding *real* eigenvectors of general *real* matrices (cf. Section 5.1) one starts with the function

$$f(x,\lambda) := \begin{pmatrix} Ax - \lambda x \\ x_n - \alpha \end{pmatrix} , \ \alpha \neq 0, \tag{40}$$

the zeros (x^*, λ^*) of which are eigenpairs of A, with the eigenvector x^* being normalized to

$$x_n^* = \alpha \ . \tag{41}$$

In this way, the eigenvector becomes unique, provided λ^* is (at least geometric) simple. The reason why the n–th component of x^* is normalized has purely notational character. Replacing $x_n - \alpha$ in (40) by $x_s - \alpha$ yields a normalization for the s–th component. If (41) is changed to

$$(x^*)^T x^* = \alpha \tag{42}$$

then (40) has to be modified accordingly. In this case (and for simple eigenvalues) the eigenvector is still ambiguous; it can be made unique by prescribing, e. g., a sign of one of its non–zero components. We will return to (42) later on in Section 5.1.

There is no need in (40) to restrict A a priori to be real and λ to be real and simple. In fact, the method to be derived can be applied for $A \in M_{nn}(\boldsymbol{C})$ and for $\lambda \in \boldsymbol{C}$. For multiple eigenvalues, however, it has to be modified in order to guarantee success in the verification process. We shall present such a modification for double or nearly double eigenvalues in Section 5.2. In the sequel we shall also touch the eigenproblem for interval data which occurs, e. g., in practical computation when A is not representable in the given floating point system.

5.1. Simple eigenvalues

Let $A \in M_{nn}(\boldsymbol{R})$, and let $(\tilde{x}, \tilde{\lambda}) \in V_n(\boldsymbol{R}) \times \boldsymbol{R}$ be any approximation of an unknown eigenpair $(x^*, \lambda^*) \in V_n(\boldsymbol{R}) \times \boldsymbol{R}$ satisfying

$$\tilde{x}_n = \alpha \ . \tag{43}$$

Imagine, e. g., that $(\tilde{x}, \tilde{\lambda})$ is computed by some standard software package. At this moment it is not yet relevant that λ^* is simple. It will turn out, however, that λ^* must be simple if the crucial assumption (49) of our first theorem is to be fulfilled. To derive the method, expand f from (40) into its Taylor series at $(\tilde{x}, \tilde{\lambda})$, and introduce the differences

$$\Delta x := x - \tilde{x} \ , \quad \Delta\lambda := \lambda - \tilde{\lambda} \ . \tag{44}$$

Then (40) reads

$$f(x,\lambda) = f(\tilde{x},\tilde{\lambda}) + \begin{pmatrix} A - \tilde{\lambda} I_n & -\tilde{x} \\ (e^{(n)})^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} - \begin{pmatrix} \Delta\lambda\, \Delta x \\ 0 \end{pmatrix} \ . \tag{45}$$

Note that $f_{n+1}(\tilde{x},\tilde{\lambda}) = 0$ and that the Taylor expansion is finite due to the particular nonlinearity λx of f. Multiplying (45) by a preconditioning matrix $-C \in M_{n+1,n+1}(\boldsymbol{R})$, and adding $\begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix}$ on both sides, yields the righthand side function $g : V_{n+1}(\boldsymbol{R}) \to V_{n+1}(\boldsymbol{R})$ which is given by

$$g(\Delta x, \Delta\lambda) := -Cf(\tilde{x},\tilde{\lambda}) + \left\{ I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda} I_n & -\tilde{x} - \Delta x \\ (e^{(n)})^T & 0 \end{pmatrix} \right\} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} , \tag{46}$$

and which has the error

$$(\Delta x^*, \Delta\lambda^*) := (x - x^*, \lambda - \lambda^*)$$

as a fixed point. It is easily seen that (46) can be interpreted as the Taylor expansion at $(\tilde{x},\tilde{\lambda})$ of the function s defined by

$$s(x,\lambda) := \begin{pmatrix} x - \tilde{x} \\ \lambda - \tilde{\lambda} \end{pmatrix} - Cf(x,\lambda) = \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} - Cf(x,\lambda) . \tag{47}$$

Therefore, trivially,

$$g(\Delta x, \Delta\lambda) = s(x,\lambda) \tag{48}$$

holds.

With $[x] \in V_n(I(\boldsymbol{R}))$, $[\lambda] \in I(\boldsymbol{R})$ we define

$$[\Delta x] := [x] - \tilde{x},\ [\Delta\lambda] := [\lambda] - \tilde{\lambda} .$$

For the interval arithmetic evaluation of g we obtain one of the main results of this article.

Theorem 17 *Let $A \in M_{nn}(\boldsymbol{R})$, $\tilde{\lambda} \in \boldsymbol{R}$, $\tilde{x} \in V_n(\boldsymbol{R})$, $C \in M_{n+1,n+1}(\boldsymbol{R})$, and define g by (46). Let $\tilde{x}$ be normalized by (43). If g fulfills the inclusion*

$$g([\Delta x],[\Delta\lambda]) \subseteq \operatorname{int}([\Delta x]^T,[\Delta\lambda])^T \tag{49}$$

then the following assertions hold.

a) C is nonsingular.

b) There exists exactly one eigenvector $x^* \in \tilde{x} + [\Delta x]$ which is normalized by (41).

c) There exists exactly one eigenvalue $\lambda^* \in \tilde{\lambda} + [\Delta\lambda]$.

d) $Ax^* = \lambda^* x^*$ (with x^* from b) and λ^* from c)).

e) λ^* is a geometric simple eigenvalue.

f) If $(\tilde{x},\tilde{\lambda})$ are sufficiently good approximations of (x^*,λ^*) then it can be guaranteed that λ^* is algebraic simple.

g) If one starts the iteration

$$\begin{pmatrix} [\Delta x]^{(k+1)} \\ [\Delta\lambda]^{(k+1)} \end{pmatrix} := g([\Delta x]^{(k)}, [\Delta\lambda]^{(k)}),\ k = 0, 1, \ldots, \tag{50}$$

with

$$([\Delta x]^{(0)}, [\Delta\lambda]^{(0)}) := ([\Delta x], [\Delta\lambda])$$

from (49) then the iterates converge satisfying

$$([\Delta x]^{(k+1)}, [\Delta\lambda]^{(k+1)}) \subseteq ([\Delta x]^{(k)}, [\Delta\lambda]^{(k)}),\ k = 0, 1, \ldots,$$

and

$$(x^*, \lambda^*) \in (\tilde{x}, \tilde{\lambda}) + ([\Delta x]^{(k)}, [\Delta\lambda]^{(k)}),\ k = 0, 1, \ldots .$$

□

To prove Theorem 17 we need several auxiliary results which we are now going to state.

Lemma 18 *With the notations of Theorem 17 the assumption (49) implies that C and $\begin{pmatrix} A - \tilde{\lambda}\, I_n & -\tilde{x} - \Delta x \\ (e^{(n)})^T & 0 \end{pmatrix}$ are nonsingular for all $\Delta x \in [\Delta x]$.*

Proof.
By (49) we get for the widths

$$\begin{aligned} w\begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} &> w\,(g([\Delta x], [\Delta\lambda])) \\ &\geq \left| I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda}\, I_n & -\tilde{x} - [\Delta x] \\ (e^{(n)})^T & 0 \end{pmatrix} \right| w \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} \\ &\geq \left| I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda}\, I_n & -\tilde{x} - \Delta x \\ (e^{(n)})^T & 0 \end{pmatrix} \right| w \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} \end{aligned}$$

for all $\Delta x \in [\Delta x]$. Hence the Theorems 7 and 8 imply

$$\rho\left(I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda}\, I_n & -\tilde{x} - \Delta x \\ (e^{(n)})^T & 0 \end{pmatrix} \right) < 1$$

whence the assertion follows.

□

Lemma 19 *Let (x^*, λ^*) be an eigenpair of $A \in M_{nn}(\boldsymbol{R})$ with $\Delta x^* \in [\Delta x] \in V_n(I(\boldsymbol{R}))$. Then (49) implies $x_n^* \neq 0$.*

Proof.
Assume that $x_n^* = 0$ holds. Then

$$\begin{pmatrix} A - \tilde{\lambda} I_n & -x^* \\ \left(e^{(n)}\right)^T & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \Delta\lambda^* \end{pmatrix} = \begin{pmatrix} Ax^* - \lambda^* x^* \\ 0 \end{pmatrix} = 0 ,$$

and Lemma 18 implies the contradiction $x^* = 0$.

□

Lemma 20 *Let (x^*, λ^*) be an eigenpair of $A \in M_{nn}(\boldsymbol{R})$ with $\Delta x^* \in [\Delta x] \in V_n(I(\boldsymbol{R}))$, $x_n^* = \alpha$. Then (49) implies $\Delta\lambda^* \in [\Delta\lambda]$.*

Proof.
With s, g from (46), (47) define p as the projection

$$p(\Delta\lambda) := g_{n+1}(\Delta x^*, \Delta\lambda) = s_{n+1}(x^*, \lambda) = \left\{ \begin{pmatrix} \Delta x^* \\ \Delta\lambda \end{pmatrix} - Cf(x^*, \lambda) \right\}_{n+1} .$$

Then, by (49), we have $p(\Delta\lambda) \in \operatorname{int}([\Delta\lambda])$ for all $\Delta\lambda \in [\Delta\lambda]$. Hence Brouwer's fixed point theorem guarantees a fixed point $\Delta\hat{\lambda} \in \operatorname{int}([\Delta\lambda])$ of p. We will show that $\Delta\hat{\lambda} = \Delta\lambda^*$ holds. Let $\hat{\lambda} := \tilde{\lambda} + \Delta\hat{\lambda}$, define $\Delta\hat{y} \in V_n(\boldsymbol{R})$ by the equation

$$g(\Delta x^*, \Delta\hat{\lambda}) = \begin{pmatrix} \Delta\hat{y} \\ \Delta\hat{\lambda} \end{pmatrix}$$

and assume for the moment $\Delta\hat{\lambda} \neq \Delta\lambda^*$, or, equivalently, $\hat{\lambda} \neq \lambda^*$. With (46) we get

$$\begin{aligned} \begin{pmatrix} \Delta\hat{y} \\ \Delta\hat{\lambda} \end{pmatrix} &= g(\Delta x^*, \Delta\hat{\lambda}) = s(x^*, \hat{\lambda}) = \begin{pmatrix} \Delta x^* \\ \Delta\hat{\lambda} \end{pmatrix} - C \begin{pmatrix} Ax^* - \hat{\lambda} x^* \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \Delta x^* \\ \Delta\hat{\lambda} \end{pmatrix} - (\lambda^* - \hat{\lambda}) C \begin{pmatrix} x^* \\ 0 \end{pmatrix} . \end{aligned}$$

Therefore,

$$C \begin{pmatrix} x^* \\ 0 \end{pmatrix} = \frac{1}{\lambda^* - \hat{\lambda}} \begin{pmatrix} x^* - \hat{y} \\ 0 \end{pmatrix} , \tag{51}$$

and with $\underline{\lambda} := \tilde{\lambda} + \underline{\Delta\lambda}$ we have

$$\begin{aligned} g(\Delta x^*, \underline{\Delta\lambda}) = s(x^*, \underline{\lambda}) &= \begin{pmatrix} \Delta x^* \\ \underline{\Delta\lambda} \end{pmatrix} - C \begin{pmatrix} Ax^* - \underline{\lambda} x^* \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \Delta x^* \\ \underline{\Delta\lambda} \end{pmatrix} - \frac{\lambda^* - \underline{\lambda}}{\lambda^* - \hat{\lambda}} \begin{pmatrix} x^* - \hat{y} \\ 0 \end{pmatrix} . \end{aligned}$$

This implies the contradiction

$$p(\underline{\Delta\lambda}) = \underline{\Delta\lambda} = g_{n+1}(\Delta x^*, \underline{\Delta\lambda}) \in \operatorname{int}([\Delta\lambda]) ,$$

whence $\Delta\lambda^* = \Delta\hat{\lambda} \in [\Delta\lambda]$.

□

Lemma 21 *Let (x^*, λ^*) be an eigenpair of $A \in M_{nn}(\boldsymbol{R})$ with $\Delta\lambda^* \in [\Delta\lambda]$. Then (49) implies the existence of some vector $\Delta y^* \in [\Delta x]$ such that $(y^*, \lambda^*) = (\tilde{x} + \Delta y^*, \lambda^*)$ is an eigenpair satisfying $\Delta y^* := y^* - \tilde{x} \in [\Delta x]$ and $(\Delta y^*)_n = 0$.*

Proof.
Similarly to the previous proof start with a projection p which this time is defined by

$$p(\Delta x) := \begin{pmatrix} g_1(\Delta x, \Delta\lambda^*) \\ \vdots \\ g_n(\Delta x, \Delta\lambda^*) \end{pmatrix} = \begin{pmatrix} s_1(x, \lambda^*) \\ \vdots \\ s_n(x, \lambda^*) \end{pmatrix} .$$

For all $\Delta x \in [\Delta x]$ the inclusion (49) implies

$$p(\Delta x) \in \operatorname{int}([\Delta x]) , \tag{52}$$

hence Brouwer's fixed point theorem guarantees a fixed point $\Delta\hat{y} \in \operatorname{int}([\Delta x])$ of p. Let $\hat{y} := \tilde{x} + \Delta\hat{y}$, define $\Delta\hat{\lambda}$ by the equation

$$s(\hat{y}, \lambda^*) = g(\Delta\hat{y}, \Delta\lambda^*) = \begin{pmatrix} \Delta\hat{y} \\ \Delta\hat{\lambda} \end{pmatrix} , \tag{53}$$

and assume for the moment

$$\Delta\hat{\lambda} \neq \Delta\lambda^* , \tag{54}$$

or, equivalently, $\hat{\lambda} \neq \lambda^*$. Let

$$u(t) := (1 - x_n^* t)\hat{y} + \alpha t x^* = \hat{y} + t(\alpha x^* - x_n^* \hat{y})$$

and $\Delta u(t) := u(t) - \tilde{x}$.

If $\alpha x^* - x_n^* \hat{y} = 0$ then $x_n^* \neq 0$ because of $x^* \neq 0$. Hence $\hat{y} = \dfrac{\alpha}{x_n^*} x^*$. This implies $\hat{y}_n = \alpha$ and, by (53),

$$\begin{pmatrix} \Delta\hat{y} \\ \Delta\lambda^* \end{pmatrix} = s(\hat{y}, \lambda^*) = g(\Delta\hat{y}, \Delta\lambda^*) = \begin{pmatrix} \Delta\hat{y} \\ \Delta\hat{\lambda} \end{pmatrix} ,$$

contradicting (54).

Let now $\alpha x^* - x_n^* \hat{y} \neq 0$. Then $\lim_{t\to\infty} \|\Delta u(t)\|_\infty = \infty$ and $\Delta u(0) = \Delta\hat{y} \in [\Delta x]$. Hence there exists a real number $\tilde{t}$ such that $\Delta u(\tilde{t})$ lies on the boundary $\partial[\Delta x]$ of $[\Delta x]$. By (52) we get

$$p(\Delta u(\tilde{t})) \in \operatorname{int}([\Delta x]) . \tag{55}$$

On the other hand we obtain

$$\begin{aligned} s(u(\tilde{t}), \lambda^*) &= \begin{pmatrix} \Delta u(\tilde{t}) \\ \Delta\lambda^* \end{pmatrix} - C \begin{pmatrix} (1 - x_n^* \tilde{t})(A\hat{y} - \lambda^* \hat{y}) \\ (1 - x_n^* \tilde{t})\hat{y}_n + \alpha \tilde{t} x_n^* - \alpha \end{pmatrix} \\ &= \begin{pmatrix} \Delta u(\tilde{t}) \\ \Delta\lambda^* \end{pmatrix} + (1 - x_n^* \tilde{t}) \left\{ -C \begin{pmatrix} A\hat{y} - \lambda^* \hat{y} \\ \hat{y}_n - \alpha \end{pmatrix} + \begin{pmatrix} \Delta\hat{y} \\ \Delta\lambda^* \end{pmatrix} - \begin{pmatrix} \Delta\hat{y} \\ \Delta\lambda^* \end{pmatrix} \right\} \end{aligned}$$

$$= \begin{pmatrix} \Delta u(\tilde{t}) \\ \Delta\lambda^* \end{pmatrix} + (1 - x_n^*\tilde{t}) \left\{ s(\hat{y}, \lambda^*) - \begin{pmatrix} \Delta\hat{y} \\ \Delta\lambda^* \end{pmatrix} \right\}$$

$$= \begin{pmatrix} \Delta u(\tilde{t}) \\ \Delta\lambda^* \end{pmatrix} + (1 - x_n^*\tilde{t}) \left\{ \begin{pmatrix} \Delta\hat{y} \\ \Delta\hat{\lambda} \end{pmatrix} - \begin{pmatrix} \Delta\hat{y} \\ \Delta\lambda^* \end{pmatrix} \right\}$$

$$= \begin{pmatrix} \Delta u(\tilde{t}) \\ \Delta\lambda^* \end{pmatrix} + \begin{pmatrix} 0 \\ (1 - x_n^*\tilde{t})(\hat{\lambda} - \lambda^*) \end{pmatrix} .$$

This implies

$$p(\Delta u(\tilde{t})) = \Delta u(\tilde{t}) \in \partial[\Delta x]$$

contradicting (55).

□

Lemma 22 *If (49) is valid then there is exactly one eigenpair* $(x^*, \lambda^*) \in (\tilde{x}, \tilde{\lambda}) + ([\Delta x], [\Delta\lambda])$ *of* A *with* x^* *being normalized by* $x_n^* = \alpha$.

Proof.
The existence of (x^*, λ^*) follows at once from Lemma 18 and from Brouwer's fixed point theorem applied to g or s from (46) and (47), respectively.

To prove the uniqueness assume that there is a second eigenpair (y^*, μ^*) of A satisfying $y_n^* = \alpha$, $y^* \neq x^*$ and

$$(\Delta y^*, \Delta\mu^*) := (y^* - \tilde{x}, \mu^* - \tilde{\lambda}) \in ([\Delta x], [\Delta\lambda]) .$$

Case 1 : $\lambda^* \neq \mu^*$.

First we will show that λ^* and μ^* differ both from $\tilde{\lambda}$. To this end assume $\lambda^* = \tilde{\lambda}$. Then

$$\begin{pmatrix} A - \tilde{\lambda} I_n & -y^* \\ (e^{(n)})^T & 0 \end{pmatrix} \begin{pmatrix} y^* - x^* \\ \mu^* - \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} (\mu^* - \tilde{\lambda})\, y^* - y^*(\mu^* - \tilde{\lambda}) \\ 0 \end{pmatrix} = 0 ,$$

and Lemma 18 implies $\mu^* = \tilde{\lambda}$. Our assumption then yields the contradiction $\mu^* = \tilde{\lambda} = \lambda^*$. Therefore, $\lambda^* \neq \tilde{\lambda}$. Analogously one shows $\mu^* \neq \tilde{\lambda}$.

In the sequel we will use the function h which is defined by the expression

$$h(t) := (1 - t)(\mu^* - \tilde{\lambda}) + t(\lambda^* - \tilde{\lambda}) = \mu^* - \tilde{\lambda} + t(\lambda^* - \mu^*) , \qquad t \in \boldsymbol{R} .$$

We note that

$$h(t) = 0 \qquad \text{is equivalent to} \qquad t = \frac{\mu^* - \tilde{\lambda}}{\mu^* - \lambda^*} \tag{56}$$

with t differing from zero by the previous conclusions. In addition, we define for $z \in \boldsymbol{R}$ and $t \in \boldsymbol{R}$ the expressions

$$g_z(\Delta x, \Delta, \lambda) \ := \ s(\tilde{x}, \tilde{\lambda}) + \left\{ I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda} I_n & -z \\ \left(e^{(n)}\right)^T & 0 \end{pmatrix} \right\} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} ,$$

$$
\begin{aligned}
u(t) &:= x^* + t(y^* - x^*)\ , \\
v(t) &:= x^* + t\,\frac{\lambda^* - \tilde{\lambda}}{h(t)}(y^* - x^*)\ , \\
\sigma(t) &:= \tilde{\lambda} + \frac{\lambda^* - \tilde{\lambda}}{h(t)}(\mu^* - \tilde{\lambda})\ , \\
\Delta u(t) &:= u(t) - \tilde{x}\ , \\
\Delta v(t) &:= v(t) - \tilde{x}\ , \\
\Delta \sigma(t) &:= \sigma(t) - \tilde{\lambda}\ .
\end{aligned}
$$

Then we obtain

$$
\begin{aligned}
&Av(t) - \tilde{\lambda}v(t) - (\sigma - \tilde{\lambda})u(t) \\
&= \lambda^* x^* + t\frac{\lambda^* - \tilde{\lambda}}{h(t)}(\mu^* y^* - \lambda^* x^*) - \tilde{\lambda}x^* - t\frac{\lambda^* - \tilde{\lambda}}{h(t)}\tilde{\lambda}(y^* - x^*) \\
&\quad - \frac{\lambda^* - \tilde{\lambda}}{h(t)}(\mu^* - \tilde{\lambda})\{x^* + t(y^* - x^*)\} \\
&= \frac{\lambda^* - \tilde{\lambda}}{h(t)}\{x^*\left(h(t) - t\lambda^* + t\tilde{\lambda} - \mu^* + \tilde{\lambda} + t(\mu^* - \tilde{\lambda})\right) + y^*\left(t\mu^* - t\tilde{\lambda} - t(\mu^* - \tilde{\lambda})\right)\} \\
&= 0
\end{aligned}
$$

and

$$v_n(t) - \alpha = 0,$$

which implies

$$
\begin{aligned}
g_{u(t)}\left(\Delta v(t), \Delta\sigma(t)\right) &= s(\tilde{x}, \tilde{\lambda}) + \begin{pmatrix} \Delta v(t) \\ \Delta\sigma(t) \end{pmatrix} \\
&\quad - C\begin{pmatrix} Av(t) - \tilde{\lambda}v(t) - (A\tilde{x} - \tilde{\lambda}\tilde{x}) - (\sigma - \tilde{\lambda})u(t) \\ v_n(t) - \tilde{x}_n \end{pmatrix} \\
&= -C\begin{pmatrix} A\tilde{x} - \tilde{\lambda}\tilde{x} \\ \tilde{x}_n - \alpha \end{pmatrix} + \begin{pmatrix} \Delta v(t) \\ \Delta\sigma(t) \end{pmatrix} + C\begin{pmatrix} A\tilde{x} - \tilde{\lambda}\tilde{x} \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \Delta v(t) \\ \Delta\sigma(t) \end{pmatrix}\ .
\end{aligned}
$$

By (49) we get

$$
\begin{pmatrix} \Delta x^* \\ \Delta\lambda^* \end{pmatrix} = s(x^*, \lambda^*) = g(\Delta x^*, \Delta\lambda^*) \in \text{ int } \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} \tag{57}
$$

whence $\Delta u(0) = \Delta x^* \in \text{int }([\Delta x])$.

Analogously, $\Delta u(1) = \Delta y^* \in \text{int }([\Delta x])$.

Therefore, there are real numbers $\underline{t} < 0$, $\overline{t} > 1$ such that $\Delta u(\underline{t})$, $\Delta u(\overline{t})$ lie on the boundary $\partial[\Delta x]$ of $[\Delta x]$ with $\Delta u(t) \in \text{int }(\Delta x)$ for $\underline{t} < t < \overline{t}$.

We will show now, that

$$
\begin{pmatrix} \Delta v(t) \\ \Delta\sigma(t) \end{pmatrix} \in \text{ int} \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} \qquad \text{for all } t \in [\underline{t}, \overline{t}]\ . \tag{58}
$$

This relation is certainly true for $t = 0$ because of $(\Delta v(0), \Delta\sigma(0)) = (\Delta x^*, \Delta\lambda^*)$ and (57). Assuming

$$\begin{pmatrix} \Delta v(\tilde{t}) \\ \Delta\sigma(\tilde{t}) \end{pmatrix} \in \partial \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} \quad \text{for some } \tilde{t} \in [\underline{t}, \overline{t}] \tag{59}$$

yields the contradiction

$$\begin{pmatrix} \Delta v(\tilde{t}) \\ \Delta\sigma(\tilde{t}) \end{pmatrix} = g_{u(\tilde{t})}\left(\Delta v(\tilde{t}),\ \Delta\sigma(\tilde{t})\right) \in g([\Delta x], [\Delta\lambda]) \subseteq \ \text{int} \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix}.$$

Thus, (58) is valid.

We remark that

$$\frac{\mu^* - \tilde{\lambda}}{\mu^* - \lambda^*} \notin [\underline{t}, \overline{t}] \tag{60}$$

holds. Otherwise let t tend to $\dfrac{\mu^* - \tilde{\lambda}}{\mu^* - \lambda^*}$ which is different from zero since $\mu^* \neq \tilde{\lambda}$ by the arguments at the beginning of this proof. Then $h(t)$ tends to zero whence $\|\Delta v(t)\|_\infty$ tends to infinity contradicting (58). Therefore, (60) is valid. Together with (56) it implies $h(t) \neq 0$ for $\underline{t} \leq t \leq \overline{t}$. W.l.o.g. let $h(t) > 0$ for $\underline{t} \leq t \leq \overline{t}$.

We will next prove the inequality

$$\underline{t} < t\frac{\lambda^* - \tilde{\lambda}}{h(t)} < \overline{t} \ \text{ for } \ \underline{t} \leq t \leq \overline{t}. \tag{61}$$

By the definition of $\underline{t}$, $\overline{t}$, the inequality (61) is certainly fulfilled for $t = 0$. Assume that there is a real number $t_1 \in [\underline{t}, \overline{t}]$ such that $t_1\dfrac{\lambda^* - \tilde{\lambda}}{h(t_1)} \leq \underline{t}$ holds. Since (61) holds for $t = 0$, there is some $\tilde{t} \in [\underline{t}, \overline{t}]$ satisfying

$$\tilde{t}\frac{\lambda^* - \tilde{\lambda}}{h(\tilde{t})} = \underline{t}.$$

This yields

$$\Delta v(\tilde{t}) = \Delta x^* + \underline{t}(y^* - x^*) = \Delta u(\underline{t}) \in \partial([\Delta x])$$

contradicting (58). Thus the left inequality of (61) holds. The right inequality can be proven analogously.

Choose now $t := \underline{t}$ in (61). Taking into account $\underline{t} < 0,\ h(\underline{t}) > 0$ this implies

$$\begin{aligned} \underline{t} < \underline{t}\frac{\lambda^* - \tilde{\lambda}}{h(\underline{t})} \quad &\Longleftrightarrow \quad (1 - \underline{t})(\mu^* - \tilde{\lambda}) + \underline{t}(\lambda^* - \tilde{\lambda}) > \lambda^* - \tilde{\lambda} \\ &\Longleftrightarrow \quad (1 - \underline{t})(\mu^* - \tilde{\lambda}) > (1 - \underline{t})(\lambda^* - \tilde{\lambda}) \\ &\Longleftrightarrow \quad \mu^* > \lambda^*. \end{aligned} \tag{62}$$

Choosing $t := \bar{t}$ in (61) yields analogously

$$\begin{aligned}
\bar{t}\frac{\lambda^* - \tilde{\lambda}}{h(\bar{t})} < \bar{t} \quad &\Longleftrightarrow \quad \lambda^* - \tilde{\lambda} < (1-\bar{t})(\mu^* - \tilde{\lambda}) + \bar{t}(\lambda^* - \tilde{\lambda}) \\
&\Longleftrightarrow \quad (1-\bar{t})(\lambda^* - \tilde{\lambda}) < (1-\bar{t})(\mu^* - \tilde{\lambda}) \\
&\Longleftrightarrow \quad \lambda^* - \tilde{\lambda} > \mu^* - \tilde{\lambda} \\
&\Longleftrightarrow \quad \lambda^* > \mu^*
\end{aligned}$$

contradicting (62).

Case 2 : $\lambda^* = \mu^*$

Analogously to case 1 one proves $\tilde{\lambda} \neq \lambda^* = \mu^*$, and with

$$\begin{aligned}
h(t) &:= \lambda^* - \tilde{\lambda} \equiv \text{ constant } \neq 0, \\
u(t) &:= v(t) := x^* - t(y^* - x^*), \\
\sigma(t) &:= \lambda^* \equiv \text{ constant}, \\
\Delta u(t) &:= \Delta v(t) := v(t) - \tilde{x}, \\
\Delta\sigma(t) &:= \sigma(t) - \tilde{\lambda}
\end{aligned}$$

one can repeat all the steps of the previous case up to (59) which yields

$\Delta u(\tilde{t}) \in \partial([\Delta x])$.

contradicting

$$\begin{pmatrix} \Delta u(\tilde{t}) \\ \Delta\sigma(\tilde{t}) \end{pmatrix} = \begin{pmatrix} \Delta v(\tilde{t}) \\ \Delta\sigma(\tilde{t}) \end{pmatrix} = g_{u(\tilde{t})}\left(\Delta v(\tilde{t}), \Delta\sigma(\tilde{t})\right) \in g([\Delta x], [\Delta\lambda]) \subseteq \text{ int } \begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} .$$

□

We are now ready to prove Theorem 17.

Proof of Theorem 17.

a) is contained in Lemma 18.

b) The existence of $x^* \in \tilde{x} + [\Delta x]$, $x_n^* = \alpha$, follows from Lemma 22.

In order to prove the uniqueness, assume first, that there are two independent eigenvectors x^*, $y^* \in \tilde{x} + [\Delta x]$ which are normalized according to (41) and which are associated with two eigenvalues λ^* and μ^* of A. By Lemma 20, both eigenvalues necessarily are contained in $\tilde{\lambda} + [\Delta\lambda]$, hence Lemma 22 implies $\lambda^* = \mu^*$ and $x^* = y^*$ contradicting the linear independency of x^*, y^*.

c) The existence of $\lambda^* \in \tilde{\lambda} + [\Delta\lambda]$ follows again from Lemma 22. If there is a second eigenvalue μ^* which differs from λ^* and which is also contained in $\tilde{\lambda} + [\Delta\lambda]$, then the Lemmas 21 and 22 yield the contradiction $\lambda^* = \mu^*$.

d) follows from Lemma 22 and b), c).

e) If λ^* is not geometric simple, then there are two linearly independent eigenvectors x^*, y^* associated with λ^*, with x^* being normalized by (40) and $x^* \in \tilde{x} + [\Delta x]$ taking into account Lemma 22. For $\varepsilon > 0$ sufficiently small, the vector

$$z^* := \begin{cases} x^* + \varepsilon \left(\alpha \dfrac{y^*}{y_n^*} - x^* \right) & , \text{ if } y_n^* \neq 0 \\ x^* + \varepsilon\, y^* & , \text{ if } y_n^* = 0 \end{cases}$$

is an eigenvector of A associated with λ^* which is linearly independent of x^* and which satisfies $z^* \in \tilde{x} + [\Delta x]$ and $z_n^* = \alpha$. This contradicts b).

f) By Lemma 18, the matrix

$$\begin{pmatrix} A - \tilde{\lambda} I_n & -x^* \\ (e^{(n)})^T & 0 \end{pmatrix} \qquad \text{is nonsingular.}$$

Hence for sufficiently good approximations $\tilde{\lambda} \approx \lambda^*$ the same holds for $\begin{pmatrix} A - \lambda^* I_n & -x^* \\ (e^{(n)})^T & 0 \end{pmatrix}$ because of Theorem 1a). Then Theorem 2 proves the assertion.

g) follows from (49) and from the inclusion monotonocity of the interval arithmetic.

□

Note that Theorem 17 neither gives a hint how to construct $([\Delta x], [\Delta\lambda])$ such that (49) holds nor does it guarantee that the iterates from (50) converge to the error $(\Delta x^*, \Delta\lambda^*)$. Both items are handled in the following theorem which is a direct consequence of Theorem 4.

Theorem 23 *With the notations of Theorem 17 define*

$$\rho := \left\| C \begin{pmatrix} A\tilde{x} - \tilde{\lambda}\tilde{x} \\ 0 \end{pmatrix} \right\|_\infty, \quad \sigma := \left\| I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda} I_n & -\tilde{x} \\ (e^{(n)})^T & 0 \end{pmatrix} \right\|_\infty, \quad \tau := \|C\|_\infty \tag{63}$$

and assume

$$\sigma < 1, \qquad \Delta := (1-\sigma)^2 - 4\rho\tau \geq 0. \tag{64}$$

Then the numbers

$$\begin{aligned} \beta^- &:= (1 - \sigma - \sqrt{\Delta})/(2\tau) = \frac{2\rho}{1 - \sigma + \sqrt{\Delta}}, \\ \beta^+ &:= (1 - \sigma + \sqrt{\Delta})/(2\tau) \end{aligned}$$

are nonnegative, and the condition (49) of Theorem 17 is fulfilled for $([\Delta x]^T, [\Delta\lambda])^T := [-\beta, \beta] e \in V_{n+1}(I(\boldsymbol{R}))$ *with arbitrary* $\beta \in (\beta^-, \beta^+)$. *In particular, all the assertions of that theorem hold.*

If β *is restricted to* $[\beta^-, (\beta^- + \beta^+)/2)$ *then the iterates of (49) converge to the error* $\begin{pmatrix} \Delta x^* \\ \Delta\lambda^* \end{pmatrix}$.

Proof.
Note that $\sigma < 1$ implies that C is nonsingular. Now apply Theorem 4 with

$$r := -C\begin{pmatrix} A\tilde{x} - \lambda\tilde{x} \\ 0 \end{pmatrix}, \; S := I_{n+1} - C\begin{pmatrix} A - \tilde{\lambda} I_n & -\tilde{x} \\ (e^{(n)})^T & 0 \end{pmatrix},$$

$$t_{ijk} := \begin{cases} c_{ij}, & \text{if } j \in \{1, \dots, n\} \text{ and } k = n+1 \\ 0 & \text{otherwise} \end{cases},$$

and note that

$$g([\Delta x], [\Delta\lambda]) \subseteq r + S\begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} + C\left\{\begin{pmatrix} 0 & [\Delta x] \\ 0 & 0 \end{pmatrix}\begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix}\right\}$$

holds because of the subdistributivity of the interval arithmetic and because

$$(P[Q])[z] \subseteq P([Q][z])$$

holds for matrices $P \in M_{n+1,n+1}(\boldsymbol{R})$, $[Q] \in M_{n+1,n+1}(I(\boldsymbol{R}))$ and vectors $[z] \in V_{n+1}(I(\boldsymbol{R}))$. Denote the quantities in (11) and (13) by $\hat{\rho}, \hat{\sigma}, \hat{\tau}, \hat{\beta}^-, \hat{\beta}^+$ to distinguish them from those in Theorem 23. Then $\rho = \hat{\rho}$, $\sigma = \hat{\sigma}$, $\tau \geq \hat{\tau}$, hence $\hat{\beta}^- \leq \beta^-$, $\beta^+ \leq \hat{\beta}^+$, and Theorem 4 proves the assertion.

□

Another possibility to achieve (60) consists in applying epsilon–inflation starting with $\begin{pmatrix} [\Delta x] \\ [\Delta\lambda] \end{pmatrix} := 0$. See [48], [86], [90], [116], e. g.

As preconditioning matrix C one normally chooses $C := \tilde{B}^{-1}$ with

$$\tilde{B} := \begin{pmatrix} A - \tilde{\lambda} I_n & -\tilde{x} \\ (e^{(n)})^T & 0 \end{pmatrix}, \tag{65}$$

or at least an approximation to $\tilde{B}^{-1}$. Then σ of Theorem 23 is zero or will at least be small, and so is ρ, provided $(\tilde{x}, \tilde{\lambda})$ approximates (x^*, λ^*) sufficiently well. Hence, in this case the assumptions (64) of Theorem 23 will certainly hold.

By the Theorems 1 and 2 the inverse $\tilde{B}^{-1}$ exists by continuity if, again, the approximations $\tilde{x}$, $\tilde{\lambda}$ are sufficiently good. However, due to rounding errors, $\tilde{B}^{-1}$ normally cannot be computed exactly. Defining C nevertheless by $C := \tilde{B}^{-1}$ means that $\sigma = 0$ and that C in the expressions for ρ and τ in (63) has to be replaced by a computable enclosure of $\tilde{B}^{-1}$. This can be done by enclosing the solutions $z = z^{(i)}$ of the n linear systems

$$\tilde{B}z = e^{(i)}, \; i = 1, \dots, n \,.$$

Such supersets can be obtained by using, e. g., the interval Gaussian algorithm or Krawczyk's method. See [12] or [85] for details.

We will add now another remark unifying two verification methods which, at a first glance, seem to be quite different. To this end, let again $C := \tilde{B}^{-1}$ hold. Then C has the block form

$$C = \begin{pmatrix} C_{11} & C_{12} \\ 0 & 1 \\ C_{21} & C_{22} \end{pmatrix} \tag{66}$$

with $C_{11} \in M_{n-1,n}(\boldsymbol{R})$, $C_{12} \in \boldsymbol{R}^{n-1}$, $C_{21} \in M_{1,n}(\boldsymbol{R})$, $C_{22} \in \boldsymbol{R}$. This can be seen, e. g., by inspecting the adjoints. Any preconditioning matrix C of the form (66) effects

$$g_n([\Delta x],[\Delta\lambda]) = 0 \ , \tag{67}$$

if $\tilde{x}$ is normalized by $\tilde{x}_n = \alpha$. In particular, this means

$$[\Delta x]_n^{(k+1)} = 0 \ , \quad k = 0, 1, \ldots ,$$

in (50). Therefore it seems reasonable to allow only starting vectors $[\Delta x]^{(0)}$ with $[\Delta x]_n^{(0)} = 0$. In this case, the blocks C_{12}, C_{22} do not influence the values of g since they are multiplied by zero. Therefore, one can shrink the whole problem to an n–dimensional one, deleting the n–th component of g and of $[\Delta x]$ and using $\hat{C} = \begin{pmatrix} C_{11} \\ C_{22} \end{pmatrix} \in M_{nn}(\boldsymbol{R})$ instead of C.

This results in a new function $\hat{g}$, defined by

$$\hat{g}(\Delta\hat{x},\Delta\lambda) := \hat{C}(\tilde{\lambda}\tilde{x} - A\tilde{x}) + \left\{ I - \hat{C}\left(\hat{B} + \begin{pmatrix} 0 & -\Delta\hat{x} \\ 0 & 0 \end{pmatrix}\right)\right\}\begin{pmatrix} \Delta\hat{x} \\ \Delta\lambda \end{pmatrix}$$

with

$$\Delta\hat{x} := (\Delta x_1, \ldots, \Delta x_{n-1})^T$$

and

$$\hat{B} := ((A - \tilde{\lambda}I)_{*,1}, \ldots, (A - \tilde{\lambda}I)_{*,n-1}, -\tilde{x}) \tag{68}$$

where $(A - \tilde{\lambda}I)_{*,j}$ denotes the j–th column of $A - \tilde{\lambda}I$. Because of $\det(\hat{C}) = -\det(C)$ it is clear that $\hat{C}$ is nonsingular if and only if C has this property.

Theorem 17 remains valid for the new situation, the proof can be reduced to the former one: Reconstruct g from $\hat{g}$ via (66), (67) with $C_{12} = C_{22} := 0$; let

$$\hat{g}([\Delta x]',[\Delta\lambda]) \subseteq \operatorname{int}\begin{pmatrix} [\Delta x]' \\ [\Delta\lambda] \end{pmatrix}$$

hold and set $[\Delta x] := \begin{pmatrix} [\Delta x]' \\ [-\varepsilon,\varepsilon] \end{pmatrix}$, $\varepsilon > 0$. For sufficiently small ε we get

$$\begin{aligned} &g_i([\Delta x],[\Delta\lambda]) \subseteq \operatorname{int}([\Delta x]_i) \ , \quad i = 1, \ldots, n-1 \ , \\ &g_{n+1}([\Delta x],[\Delta\lambda]) \subseteq \operatorname{int}([\Delta\lambda]) \ , \end{aligned}$$

and, by (67),

$$0 = g_n([\Delta x],[\Delta\lambda]) \subseteq \operatorname{int}([\Delta x]_n) = (-\varepsilon,\varepsilon) \ .$$

Hence (49) is fulfilled for $([\Delta x],[\Delta\lambda])$. Taking into account $x_n^* = \tilde{x}_n = \alpha$ and $\Delta x_n^* = 0$ one can replace $[\Delta x]_n$ by $[\Delta x]_n = 0$ in Theorem 17b). With these remarks the analogue of Theorem 17 is easily seen.

The assertions of Theorem 23 remain also true. One only has to replace ρ, σ, τ by

$$\rho := \|\hat{C}(\tilde{\lambda}\tilde{x} - A\tilde{x})\|_\infty \ , \quad \sigma := \|I - \hat{C}\hat{B}\|_\infty \ , \quad \tau := \|\hat{C}\|_\infty \ , \tag{69}$$

and $[\Delta x]$ by $[\Delta \hat{x}] \in V_{n-1}(I(\boldsymbol{R}))$. Cf. [4] or [88] for details.

The complex case $A \in M_{nn}(\boldsymbol{C})$ is considered in [88] and [117]. Here, the complex interval arithmetic is needed as described in [55].

We again refer to [88] when replacing A by an interval matrix $[A]$ dealing thus with topic 4 of Section 1. In this case $\tilde{x}$, $\tilde{\lambda}$ can be thought to be approximations for an eigenpair of the midpoint matrix $\check{A} := \text{mid}([A])$, e. g., C can be chosen as an approximate inverse of $\check{A}^{-1}$, and A has to be replaced by $[A]$ in (46). Then Theorem 17 remains valid, where the assertions b) – g) hold for each matrix $A \in [A]$ with the eigenpair (x^*, λ^*) depending on A, of course.

In a similar way, Theorem 23 remains true. Here, ρ has to be replaced by

$$\rho := \| \, |C \begin{pmatrix} [A]\tilde{x} - \tilde{\lambda}\tilde{x} \\ 0 \end{pmatrix} | \, \|_\infty$$

and σ has to be changed to

$$\sigma := \| \, |I_{n+1} - C \begin{pmatrix} [A]\tilde{x} - \tilde{\lambda}\tilde{x} & -\tilde{x} \\ (e^{(n)})^T & 0 \end{pmatrix} | \, \|_\infty \, .$$

If the normalization (41) is replaced by (42) then some of the previous results can again be shown. The function f then has to be redefined by

$$f(x, \lambda) := \begin{pmatrix} Ax - \lambda x \\ x^T x - \alpha \end{pmatrix} , \tag{70}$$

whence g from (49) now reads

$$g(\Delta x, \Delta\lambda) := -C\, f(\tilde{x}, \tilde{\lambda}) + \left\{ I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda} I & -\tilde{x} - \Delta x \\ 2\tilde{x}^T + (\Delta x)^T & 0 \end{pmatrix} \right\} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} . \tag{71}$$

with $\tilde{x}^T\tilde{x} = \alpha$. Note that $f_{n+1}(x, \lambda)$ now depends quadratically on x while it depended only linearly on x in (40). This complicates matters when trying to transfer the proof of Theorem 17 to the new situation. Lemmas 18 and 20 can be restated and proved (cf. [88]), while the analogue of Lemma 21 and of Lemma 22 are still unproved. Therefore, up to now, only the following statements can be guaranteed.

Theorem 24 *Let A, C, $\tilde{x}$, $\tilde{\lambda}$ be defined as in Theorem 17 with $\tilde{x}$ being normalized by $\tilde{x}^T\tilde{x} = \alpha \neq 0$. If*

$$g([\Delta x], [\Delta\lambda]) \subseteq \text{int}([\Delta x]^T, [\Delta\lambda])^T \tag{72}$$

holds for g from (71), then the following assertions are true:

a) C is nonsingular.

b) There is an eigenpair $(x^*, \lambda^*) \in (\tilde{x} + [\Delta x], \tilde{\lambda} + [\Delta\lambda])$ satisfying $(x^*)^T x^* = \alpha$.

c) For sufficiently good approximations $(\tilde{x}, \tilde{\lambda})$ of (x^*, λ^*) from b), the eigenvalue λ^* is algebraic simple.

d) *If* $(\hat{x},\hat{\lambda})$ *is any eigenpair of* A *satisfying* $\hat{x}^T\hat{x}=\alpha$ *and* $\hat{x}\in\tilde{x}+[\Delta x]$, *then* $\hat{\lambda}\in\tilde{\lambda}+[\Delta\lambda]$.

e) *If one starts the iteration*

$$\begin{pmatrix} [\Delta x]^{(k+1)} \\ [\Delta\lambda]^{(k+1)} \end{pmatrix} := g([\Delta x]^{(k)},[\Delta\lambda]^{(k)}) \ , \quad k=0,1,\ldots \tag{73}$$

with

$$([\Delta x]^{(0)},[\Delta\lambda]^{(0)}) := ([\Delta x],[\Delta\lambda])$$

from (72) then the iterates converge satisfying

$$\left([\Delta x]^{(k+1)},[\Delta\lambda]^{(k+1)}\right) \subseteq \left([\Delta x]^{(k)},[\Delta\lambda]^{(k)}\right) \ , \quad k=0,1,\ldots$$

and

$$(x^*,\lambda^*) \in (\tilde{x},\tilde{\lambda}) + ([\Delta x]^{(k)},[\Delta\lambda]^{(k)}) \ , \quad k=0,1,\ldots \ .$$

Proof.

a) is proved analogously to Lemma 18.

b) follows from (72) and from Brouwer's fixed point theorem.

c) is a consequence of Theorem 2a), and of the analogue of Lemma 18.

d) is proved by applying the analogue of Lemma 20.

e) follows from (72) and from the inclusion monotonicity of the interval arithmetic.

□

Theorem 25 *With the notations of Theorem 17 and with* $\tilde{x}^T\tilde{x}=\alpha\neq 0$ *define*

$$B'' := \begin{pmatrix} A-\tilde{\lambda}I_n & -\tilde{x} \\ 2\tilde{x}^T & 0 \end{pmatrix} ,$$

$$\rho := \|C\,f(\tilde{x},\tilde{\lambda})\|_\infty \ , \quad \sigma := \|I-CB''\|_\infty \ , \quad \tau := \|\,|C|\cdot(1,\ldots,1,n)^T\|_\infty \ . \tag{74}$$

n being the number of the components of Δx. Assume

$$\sigma<1 \quad \text{and} \quad \Delta := (1-\sigma)^2-4\rho\tau \geq 0 \ . \tag{75}$$

Then the numbers

$$\begin{aligned} \beta^- &:= (1-\sigma-\sqrt{\Delta})/(2\tau) \ , \\ \beta^+ &:= (1-\sigma+\sqrt{\Delta})/(2\tau) \end{aligned}$$

are nonnegative, and the condition (72) of Theorem 24 is fulfilled for $([\Delta x]^T,[\Delta\lambda])^T := [-\beta,\beta]e\in V_{n+1}(I(\boldsymbol{R}))$ *with arbitrary* $\beta\in(\beta^-,\beta^+)$. *In particular, all the assertions of this theorem hold.*

If β *is restricted to* $[\beta^-,(\beta^-+\beta^+)/2)$ *then* $(\tilde{x}+[\Delta x],\tilde{\lambda}+[\Delta\lambda])$ *contains exactly one eigenpair* (x^*,λ^*) *of* A *satisfying* $(x^*)^Tx^*=\alpha$, *and the iterates of (73) converge to the error* $\begin{pmatrix} \Delta x^* \\ \Delta\lambda^* \end{pmatrix}$.

Proof.
Theorem 25 is proved analogously to Theorem 23 using $S := I - CB''$ and

$$t_{ijk} := \begin{cases} c_{ij}, & \text{if } j \in \{1,\ldots,n\} \text{ and if } k = n+1 \\ -c_{i,n+1}, & \text{if } j = k < n+1 \\ 0 & \text{otherwise} \end{cases}$$

hence

$$\tau = \|T\|_\infty = \max_{1\le i\le n+1} \left(\sum_{j=1}^{n} |c_{ij}| + \sum_{j=1}^{n} |c_{i,n+1}| \right) = \| \, |C| \cdot (1,\ldots,1,n)^T \|_\infty .$$

□

We close this subsection with two examples, starting with the Wilkinson matrix W_{2n+1}^+.

Example 26 *(Cf. [139])*

Let $W_{2n+1}^+ = (w_{ij}^{(2n+1)})$ be defined by

$$W_{2n+1}^+ = (w_{ij}^{(2n+1)}) := \begin{pmatrix} n & 1 & & & & & & & \\ 1 & n-1 & 1 & & & & & O & \\ & 1 & n-2 & 1 & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & 1 & 0 & 1 & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & 1 & n-2 & 1 & \\ & O & & & & & 1 & n-1 & 1 \\ & & & & & & & 1 & n \end{pmatrix}$$

with ones in the first superdiagonal and in the first subdiagonal, and with the diagonal entries

$$w_{ii}^{(2n+1)} := |\, n+1-i \,| \,, \quad i = 1, \ldots, 2n+1 \,.$$

We want to verify the two largest eigenvalues λ_{20}^* and λ_{21}^* which are known to differ only in the sixteenth significant digit. To this end we compute the approximations for the eigenvalues and for the corresponding eigenvectors from EISPACK and $C \approx B^{-1}$ from the software package distributed with PASCAL–XSC. The verification process (50) stopped with

$$\lambda_{20}^* \in [10.74619418290333_2^3]$$

and

$$\lambda_{21}^* \in [10.746194182903_{39}^{40}]$$

Instead of listing the components of the enclosures $[x]^{20}$ and $[x]^{21}$ for the corresponding eigenvectors $(x^*)^{20}$, $(x^*)^{21}$ we mention only bounds of the relative width $q([x])$ which we defined for $[x] \in V_n(I(\boldsymbol{R}))$ by

$$q([x]) := \max_{1\le i\le n} \frac{w([x]_i)}{|[x]_i|} \,. \tag{76}$$

In both cases $[x] = [x]^{20}$ and $[x] = [x]^{21}$ we obtained $q([x]) \leq 3.4E-016$. This shows how standard software packages and verification numerics can collaborate to verify eigenpairs and to enclose them tightly.

$\square$

Our second example illustrates the verification process for an unsymmetric matrix which occurs in the connection with the Riemann hypothesis (cf. [104]).

Example 27

Let

$$A_n = (a_{ij}) := \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & \dots & \vdots \\ & 2 & -1 & -1 & 2 & -1 & -1 & \dots & \vdots \\ & & 3 & -1 & -1 & -1 & 3 & \dots & \vdots \\ & & & 4 & -1 & -1 & -1 & \dots & \vdots \\ & & & & 5 & -1 & -1 & \dots & \vdots \\ & & & & & 6 & -1 & \dots & \vdots \\ & & -1 & & & & 7 & \dots & \vdots \\ & & & & & & & \ddots & \vdots \\ & & & & & & & & n-1 \end{pmatrix}$$

with

$$a_{ij} := (i+1)\delta_{(i+1)|(j+1)} - 1$$

and

$$\delta_{i|j} := \begin{cases} 1 & , \text{ if } i \text{ is a divisor of } j \\ 0 & \text{ otherwise} \end{cases} .$$

Roesler showed in [104] that the Riemann hypothesis is valid if and only if the determinant $\det(A_n)$ of A_n increases for any $\varepsilon > 0$ at most like

$$\det(A_n) = O\left(n!\; n^{-1/2\;+\;\varepsilon}\right) \;, \; n \to \infty$$

where $O(\cdot)$ denotes the usual Landau symbol. Since $\det(A_n)$ can be represented as the product of the eigenvalues of A_n, we get in this way a relationship to the subject of our article. For $n = 11$, which means $A_n \in M_{10,10}(\boldsymbol{R})$, we computed the following enclosures $[\lambda]_i$ for the eigenvalues λ_i^*. Instead of listing the enclosures for the corresponding eigenvectors we again show only bounds q_i for the relative widths.

n = 11	i	$[\lambda]_i$	q_i
	1	$-1.970\,214\,329\,754\,72^{0}_{1}E-002$	$2.1E-016$
	2	$3.758\,517\,054\,844\,65^{6}_{4}E-001$	$2.1E-016$
	3	$2.714\,315\,143\,311\,93^{9}_{7}E+000$	$2.0E-016$
	4	$4.0+[-1,1]\cdot 10^{-15}$	$2.6E-016$
	5	$5.0+[-1,1]\cdot 10^{-15}$	$4.4E-016$
	6	$6.534\,132\,065\,892\,6^{40}_{38}E+000$	$2.2E-016$
	7	$7.314\,390\,058\,013\,41^{8}_{6}E+000$	$2.1E-016$
	8	$8.655\,903\,539\,939\,00^{6}_{4}E+000$	$2.0E-016$
	9	$9.588\,680\,211\,084\,14^{8}_{5}E+000$	$2.1E-016$
	10	$1.083\,642\,941\,957\,19^{4}_{3}E+001$	$2.1E-016$

As the theory shows, the exact values of λ_4^* and λ_5^* are 4 and 5, respectively.

□

5.2. Double or nearly double eigenvalues

While the method in the previous subsection is taylored to simple eigenvalues, we address now to a double eigenvalue or to two different eigenvalues with clustering being allowed. I.e., we consider the following three cases

(i) $\lambda_i^* \neq \lambda_j^*$ are two algebraic simple eigenvalues of A;

(ii) $\lambda_i^* = \lambda_j^*$ is a geometric and algebraic double eigenvalue of A;

(iii) $\lambda_i^* = \lambda_j^*$ is an algebraic double but geometric simple eigenvalue of A.

All the three cases share the same property: The eigenvalues of A associated with λ_i^*, λ_j^*, and (in the case (iii)) the corresponding principal vectors span a two–dimensional subspace V of $V_n(\boldsymbol{C})$ which is invariant with respect to the mapping represented by A; i. e., $w \in V$ implies $Aw \in V$. Any pair u^*, v^* of linearly independent vectors from V are called *generators* of V since each element w of V can be represented as linear combination $w = \mu u^* + \nu v^*, \mu, \nu \in \boldsymbol{C}$, of u^* and v^*. For each pair of generators u^*, v^* there is a 2×2 matrix M such that

$$A\,U^* = U^*\,M \tag{77}$$

holds with $U^* := (u^*, v^*) \in M_{n,2}(\boldsymbol{C})$.

In the cases (i) and (ii), u^*, v^* can be chosen to be eigenvectors x^*, y^*, in the case (iii) it is possible to choose u^* as eigenvector x^* and v^* as an associated principal vector y^* of degree two. Then M has the form

$$M = J := \begin{pmatrix} \lambda_i^* & \kappa \\ 0 & \lambda_j^* \end{pmatrix} \tag{78}$$

with $\kappa = 0$ in the cases (i), (ii) and $\kappa = 1$ for (iii). If u^*, v^* are linear combinations of x^*, y^* one can obtain the underlying vectors x^*, y^* by essentially finding the eigenvalues of the 2×2 matrix M and by reducing it to the Jordan normal form (78). If $M = S^{-1}JS$ holds with some 2×2 matrix S and with J from (78) then (77) is equivalent to

$$A(U^*S) = (U^*S)(S^{-1}MS) = (U^*S)J$$

where the columns of U^*S are x^* and y^*, respectively.

Although our aim consists in computing enclosures of eigenvectors and principal vectors, we will start with the general case (77) where u^*, v^* are any generators of V. First we remark that it is always possible to find two components i_1, i_2 and two generators u^*, v^* of V for which

$$u^*_{i_1} = \alpha \ , \quad u^*_{i_2} = \beta \ , \quad v^*_{i_1} = \gamma \ , \quad v^*_{i_2} = \delta \tag{79}$$

holds with prescribed values $\alpha, \beta, \gamma, \delta \in \boldsymbol{C}$ provided $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, $\begin{pmatrix}\gamma\\ \delta\end{pmatrix}$ are linearly independent. This makes the generators unique. If u, v are generators which do not yet satisfy the normalizations (79), we make the ansatz

$$u^* = d_{11}u + d_{12}v \ , \qquad v^* = d_{21}u + d_{22}v$$

requiring

$$\left.\begin{aligned} u^*_{i_1} &= d_{11}u_{i_1} + d_{12}v_{i_1} = \alpha \\ u^*_{i_2} &= d_{11}u_{i_2} + d_{12}v_{i_2} = \beta \end{aligned}\right\} \tag{80}$$

and

$$\left.\begin{aligned} v^*_{i_1} &= d_{21}u_{i_1} + d_{22}v_{i_1} = \gamma \\ v^*_{i_2} &= d_{21}u_{i_2} + d_{22}v_{i_2} = \delta \end{aligned}\right\} \tag{81}$$

The two 2×2 systems (80), (81) have the same coefficient matrix $K := \begin{pmatrix} u_{i_1} & v_{i_1} \\ u_{i_2} & v_{i_2} \end{pmatrix}$. Therefore, both are uniquely solvable if and only if $\det K \neq 0$, i. e., $u_{i_1}v_{i_2} - u_{i_2}v_{i_1} \neq 0$. This is possible by using the same strategy as in [15] and [35]: First define i_1 by

$$u_{i_1} := \max_{1 \le i \le n} |u_i| \ . \tag{82}$$

The value u_{i_1} certainly differs from zero because u, v are generators of V. In particular they are linearly independent. With i_1 define i_2 by

$$u_{i_1}v_{i_2} - u_{i_2}v_{i_1} = \max_{1 \le j \le n} |u_{i_1}v_j - u_jv_{i_1}| \ . \tag{83}$$

If the lefthand side of (83) were zero, then $v = \dfrac{v_{i_1}}{u_{i_1}}u$, hence u, v are linearly dependent contradicting the generator property of u, v. Therefore (79) is possible. Since $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, $\begin{pmatrix}\gamma\\ \delta\end{pmatrix}$ are assumed to be linearly independent, the same holds for u^*, v^*. Hence they are generators of V. For ease of notation we assume $i_1 = n-1$ and $i_2 = n$ in the sequel.

We restrict now to real matrices A, real eigenvalues λ^*, and real generators u^*, v^*.

Let $f : \left\{ \begin{array}{lcl} V_{2n+4}(\boldsymbol{R}) & \to & V_{2n+4}(\boldsymbol{R}) \\ z & \mapsto & f(z) \end{array} \right.$ be defined by

$$z := (u^T, m_{11}, m_{21}, v^T, m_{12}, m_{22})^T \, , \qquad f(z) := \begin{pmatrix} Au - m_{11}u - m_{21}v \\ u_{n-1} - \alpha \\ u_n - \beta \\ Av - m_{12}u - m_{22}v \\ v_{n-1} - \gamma \\ v_n - \delta \end{pmatrix} .$$

It is obvious that the vectors u^*, v^* in the representation $z^* = \left((u^*)^T, m_{11}^*, m_{21}^*, (v^*)^T, m_{12}^*, m_{22}^*\right)^T$ for a zero z^* of f are the generators of V satisfying (79).

Assume now that we are given approximations $\tilde{u}, \tilde{v}, \tilde{M}$ of u^*, v^*, M, which are gathered in the vector $\tilde{z} := \left(\tilde{u}^T, \tilde{m}_{11}, \tilde{m}_{21}, \tilde{v}^T, \tilde{m}_{12}, \tilde{m}_{22}\right)^T$. Let $\tilde{u}, \tilde{v}$ be linearly independent. It is described in [35] how to obtain $\tilde{u}, \tilde{v}, \tilde{M}$ using the QR-algorithm. We proceed now as in Section 5.1, expanding the function $z - \tilde{z} - Cf(z)$ in a Taylor series at $\tilde{z}$, where $C \in M_{2n+4,2n+4}(\boldsymbol{R})$ is some nonsingular preconditioning matrix. Introducing

$$\Delta z = \left(\Delta u^T, \Delta m_{11}, \Delta m_{21}, \Delta v^T, \Delta m_{12}, \Delta m_{22}\right)^T := z - \tilde{z}$$

we get

$$f(z) = 0 \iff \Delta z = g(\Delta z) := -C\,f(\tilde{z}) + (I_{2n+4} - CB)\Delta z + \tilde{T}\Delta z \tag{84}$$

with the $(2n+4) \times (2n+4)$ matrices

$$B := \begin{pmatrix} A - \tilde{m}_{11}I_n & -\tilde{u} & -\tilde{v} & -\tilde{m}_{21}I_n & 0 & 0 \\ (e^{(n-1)})^T & 0 & 0 & 0 & 0 & 0 \\ (e^{(n)})^T & 0 & 0 & 0 & 0 & 0 \\ -\tilde{m}_{12}I_n & 0 & 0 & A - \tilde{m}_{22}I_n & -\tilde{u} & -\tilde{v} \\ 0 & 0 & 0 & (e^{(n-1)})^T & 0 & 0 \\ 0 & 0 & 0 & (e^{(n)})^T & 0 & 0 \end{pmatrix} \tag{85}$$

and

$$\tilde{T} := C \begin{pmatrix} 0 & \Delta u & \Delta v & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta u & \Delta v \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{86}$$

which have the same block partitioning. Thus we again arrived at a quadratic function g, for which Theorem 4 applies with

$$r := -C\,f(\tilde{z}) \, , \quad S := I - CB^{\cdot}, \tag{87}$$

and

$$t_{ijk} := \begin{cases} c_{ij} & , \text{ if } k \in \{n+1, n+2\} \text{ and } j \in \{1, \dots, n\}, \\ c_{ij} & , \text{ if } k \in \{2n+3, 2n+4\} \text{ and } j \in \{n+3, \dots, 2n+2\}, \\ 0 & \text{otherwise,} \end{cases} \tag{88}$$
$$i = 1, \dots, 2n+4 \; .$$

Hence

$$\left.\begin{array}{l} \rho := \|C\, f(\tilde{z})\|_\infty \; , \quad \sigma := \|I - CB\|_\infty \; , \\ \tau := \|T\|_\infty = \left\| 2|C|(\hat{e}^T, 0, 0, \hat{e}^T, 0, 0)^T \right\|_\infty \leq 2\|C\|_\infty \end{array}\right\} \tag{89}$$

where $\hat{e} := (1, \dots, 1)^T \in V_n(\boldsymbol{R})$.

We then get the following result in which we require the normalization (79) with

$$\alpha := \tilde{u}_{n-1} \; , \quad \beta := \tilde{u}_n \; , \quad \gamma := \tilde{v}_{n-1} \; , \quad \delta := \tilde{v}_n \; . \tag{90}$$

Theorem 28 *Let g, ρ, σ, τ be defined as in (84), (89), and let (90) hold. Assume*

$$\sigma < 1 \qquad \textit{and} \qquad \Delta := (1-\sigma)^2 - 4\rho\tau \geq 0 \; , \tag{91}$$

and let

$$\beta^- := (1 - \sigma - \sqrt{\Delta})/(2\tau) \; ,$$
$$\beta^+ := (1 - \sigma + \sqrt{\Delta})/(2\tau) \; .$$

a) If $\beta \in [\beta^-, \beta^+]$ then g has in $[-\beta, \beta]e \in V_{2n+4}(\boldsymbol{R})$ at least one fixed point Δz^*. With $U^* = (u^*, v^*)$ and M from $\left((u^*)^T, m_{11}^*, m_{21}^*, (v^*)^T, m_{12}^*, m_{22}^*\right)^T := \tilde{z} + \Delta z^*$, the equation (77) holds. In particular, u^*, v^* are linearly independent vectors which are normalized by (90) and which are generators of an invariant two dimensional subspace of $V_n(\boldsymbol{R})$.

The iteration

$$[\Delta z]^{(k+1)} := g([\Delta z]^{(k)}) \; , \quad k = 0, 1, \dots \tag{92}$$

converges to some interval vector $[\Delta z]^*$ with

$$\Delta z^* \in [\Delta z]^* \subseteq [\Delta z]^{(k)} \subseteq \cdots \subseteq [\Delta z]^{(0)} \; , \quad k = 0, 1, \dots \; . \tag{93}$$

b) If $\beta \in [\beta^-, (\beta^- + \beta^+)/2)$ then g has in $[\Delta z]^{(0)} := [-\beta, \beta]e \in V_{2n+4}(\boldsymbol{R})$ a unique fixed point Δz^*, and (93) holds with $[\Delta z]^* = \Delta z^*$, i. e., (92) converges by contracting to Δz^*.

□

We remark that C^{-1}, B^{-1} exist since $\sigma < 1$. Choosing $C := B^{-1}$ leads to the block form

$$\begin{array}{cl} \left(\begin{array}{cccc} C_{11} & C_{12} & C_{13} & C_{14} \\ 0\ldots 1\,0 & 0 & 0 & 0 \\ 0\ldots 0\,1 & 0 & 0 & 0 \\ C_{31} & C_{32} & C_{33} & C_{34} \\ 0 & 0 & 0\ldots 1\,0 & 0 \\ 0 & 0 & 0\ldots 0\,1 & 0 \\ C_{41} & C_{42} & C_{43} & C_{44} \end{array}\right) & \begin{array}{l} \\ \leftarrow n-1 \\ \leftarrow n \\ \\ \leftarrow 2n+1 \\ \leftarrow 2n+2 \\ \\ \end{array} \\ \begin{array}{cccc} \underbrace{\qquad}_{n} & \underbrace{\qquad}_{2} & \underbrace{\qquad}_{n} & \underbrace{\qquad}_{2} \end{array} & \end{array}$$

whence $g_i(\Delta z) = 0$ for $i \in \{n-1, n, 2n+1, 2n+2\}$ and for any $\Delta z \in V_{2n+4}(\boldsymbol{R})$. Thus, as in the previous subsection, it makes sense to start with interval vectors $[\Delta z]$ for which $[\Delta z]_i = 0$, $i \in \{n-1, n, 2n+1, 2n+2\}$ holds, shrinking the matrices and vectors in g to get a new function $\hat{g} : V_{2n}(\boldsymbol{R}) \to V_{2n}(\boldsymbol{R})$ with modifications analogously to those in Section 5.1. Essentially in this form the method (92) has been presented in [15].

We want to consider another specialization of (92). To this end let $\tilde{m}_{21} = 0$ and $\tilde{m}_{12} \in \{0,1\}$. Then $\tilde{M}$ has Jordan normal form, hence $\tilde{m}_{11}, \tilde{m}_{22}$ can be thought to approximate eigenvalues λ^*, μ^*, and $\tilde{u}, \tilde{v}$ can be considered to approximate corresponding eigenvectors and/or principal vectors according to the cases (i) – (iii). The matrix B has the block form

$$B = \begin{pmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{pmatrix} \qquad \text{with} \qquad B_{ii} = \begin{pmatrix} A - \tilde{m}_{ii} I_n & -\tilde{u} & -\tilde{v} \\ (e^{(n-1)})^T & 0 & 0 \\ (e^{(n)})^T & 0 & 0 \end{pmatrix}, \quad i = 1,2.$$

Hence B is nonsingular if and only if B_{ii}^{-1} exists for $i = 1,2$. This holds certainly if one of the cases (i) – (iii) are handled and if the approximation $\tilde{z}$ is sufficiently good, as can be immediately seen from the subsequent result.

Theorem 29 *Let*

$$B_i^* := \begin{pmatrix} A - \lambda_i^* I_n & -u^* & -v^* \\ (e^{(n-1)})^T & 0 & 0 \\ (e^{(n)})^T & 0 & 0 \end{pmatrix} \in M_{n+2,n+2}(\boldsymbol{R}), \quad i = 1,2,$$

with λ_1^*, λ_2^* being two eigenvalues of A ($\lambda_1^* = \lambda_2^*$ being allowed) and with u^*, v^* being two linearly independent vectors from the largest invariant subspace V belonging to λ_1^*, λ_2^* which satisfy

$$u_{n-1}^* \neq 0\ , \quad u_{n-1}^* v_n^* - u_n^* v_{n-1}^* \neq 0\ . \tag{94}$$

Let (77) hold with $U^* = (u^*, v^*)$. Then the following statements are equivalent.

a) Either λ_1^*, λ_2^* are two different algebraic simple eigenvalues of A or $\lambda_1^* = \lambda_2^*$ is an algebraic double eigenvalue of A.

b) B_1 and B_2 are both nonsingular.

Proof.
Let a) hold and let u^*, v^* be generators of the invariant two–dimensional subspace V associated with λ_1^*, λ_2^*. Assume that u^*, v^* satisfy (94) and that B_1^* is singular. (If B_2^* is singular the proof proceeds analogously.) Then there is a vector $w \in V_{n+2}(\boldsymbol{R})\backslash\{0\}$ such that

$$B_1^* w = 0\ . \tag{95}$$

Decomposing w into $w = \left((w^*)^T, w_{n+1}, w_{n+2}\right)^T$ with $w^* \in V_n(\boldsymbol{R})$ yields

$$\begin{aligned} (A - \lambda_1^* I_n) w^* &= w_{n+1} u^* + w_{n+2} v^*\ , & (96)\\ w_{n-1}^* &= 0\ , & (97)\\ w_n^* &= 0\ . & (98) \end{aligned}$$

If $w^* = 0$ then $w_{n+1}u^* + w_{n+2}v^* = 0$. Since u^*, v^* are linearly independent we get $w_{n+1} = w_{n+2} = 0$ contradicting $w \neq 0$. Therefore, $w^* \neq 0$ with $w_{n-2}^* \neq 0$ w.l.o.g., whence by (94), (97), (98)

$$\det \begin{pmatrix} w_{n-2}^* & u_{n-2}^* & v_{n-2}^* \\ w_{n-1}^* & u_{n-1}^* & v_{n-1}^* \\ w_n^* & u_n^* & v_n^* \end{pmatrix} = w_{n-2}^*(u_{n-1}^* v_n^* - u_n^* v_{n-1}^*) \neq 0\ .$$

Hence u^*, v^*, w^* are linearly independent.

If $w_{n+1} = w_{n+2} = 0$ then w^* is an eigenvector associated with λ_1^*. Hence the dimension $\dim V$ of the invariant subspace V exceeds two, contradicting a). If $w_{n+1} \neq 0$ or $w_{n+2} \neq 0$, represent u^*, v^* as linear combinations of corresponding eigenvectors / principal vectors contained in V. If $\lambda_1^* \neq \lambda_2^*$, multiply (96) by $(A - \lambda_2^* I_n)$ and commute both matrix factors to see that $(A - \lambda_2^* I_n)w^*$ is a principal vector associated with λ_1^*, although λ_1^*, λ_2^* are algebraic simple eigenvalues. If $\lambda_1^* = \lambda_2^*$ is a geometric double eigenvalue then w^* is a corresponding principal vector of degree two. If $\lambda_1^* = \lambda_2^*$ is a geometric simple but algebraic double eigenvalue then w^* is a corresponding principal vector of degree three. Hence in all three cases we get $\dim V \geq 3$ contradicting a). Therefore, B_1^* cannot be singular, and b) is valid.

To prove the converse, let b) hold and assume a) to be false. Since (77) holds for U^* the subspace $\hat{V}$ spanned by u^*, v^* is invariant with respect to A. Due to this fact, u^*, v^* can be written as linear combinations of two eigenvectors x^*, y^* of A or of an eigenvector x^* and a corresponding principal vector y^* of degree 2. Since we assumed $\dim V \geq 3$, the Jordan normal form of A shows that there is a left eigenvector w^* of A associated with λ_1^* or λ_2^* which is orthogonal to x^* and y^*. Therefore $(w^*)^T u^* = (w^*)^T v^* = 0$, whence $\left((w^*)^T, 0, 0\right) B_i^* = 0$ for $i = 1$ or $i = 2$. This contradicts b).

□

We remark that an analogous theorem also holds in the modified case dealing with $\hat{g}$. This can be seen as in the proof of Theorem 2. For this modification, and with $\tilde{m}_{12} = 0$, B has the form

$$B = \hat{B} = \begin{pmatrix} \hat{B}_{11} & 0 \\ \tilde{m}_{21} I' & \hat{B}_{22} \end{pmatrix}\ .$$

where $\hat{B}_{ii}$, $i = 1, 2$, is $A - \tilde{m}_{ii} I_n$ with the columns $n-1$ and n being replaced by $-\tilde{u}, -\tilde{v}$, respectively, and where $I' := I_n - e^{(n-1)}\,(e^{(n-1)})^T - e^{(n)}\,(e^{(n)})^T$. Its inverse reads

$$\hat{B}^{-1} = \begin{pmatrix} \hat{B}_{11}^{-1} & 0 \\ -\hat{B}_{22}^{-1} \tilde{m}_{21} I' \hat{B}_{11}^{-1} & \hat{B}_{22}^{-1} \end{pmatrix}.$$

Applying Theorem 28 with C having the block form

$$C = \begin{pmatrix} C_1 & 0 \\ -C_2 \tilde{m}_{21} I' C_1 & C_2 \end{pmatrix} \tag{99}$$

yields expressions for ρ, σ, τ which are essentially identical with the corresponding quantities in [15], (2.11) – (2.13).

For numerical examples we refer to [15], where among others the following one was presented.

Example 30 *[15]*

Consider the 7×7 matrix

$$A = \begin{pmatrix} -6 & 0 & 0 & -1 & -4 & -4 & 0 \\ 0 & 4 & 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 4 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -6 & -4 & -4 & 0 \\ -4 & 0 & 0 & -4 & -6 & -1 & 0 \\ -4 & 0 & 0 & -4 & -1 & -6 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 4 \end{pmatrix}$$

with the eigenvalues

$$\lambda_1 = 6,\ \lambda_2 = \lambda_3 = 3,\ \lambda_4 = 1,\ \lambda_5 = \lambda_6 = -5,\ \lambda_7 = -15$$

and with the corresponding eigenvectors / principal vectors

$$x^1 = \frac{1}{4}\begin{pmatrix} 0 \\ 4 \\ 2 \\ 0 \\ 0 \\ 0 \\ 3 \end{pmatrix},\ x^2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},\ x^3 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},\ x^4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ -1 \\ 0 \end{pmatrix},\ x^5 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 1 \\ 0 \end{pmatrix},$$

$$x^6 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 1 \\ -1 \\ 0 \end{pmatrix},\ x^7 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Here x^3 is a principal vector associated with $\lambda_2 = \lambda_3 = 3$ and x^5, x^6 are two linearly independent eigenvectors belonging to $\lambda_5 = \lambda_6 = -5$. The approximations $\tilde{M}, \tilde{u}, \tilde{v}$ were chosen to be

$$\tilde{M} = \begin{pmatrix} -4.999\,999\,99 & 1\,E-8 \\ 0 & -5.000\,000\,01 \end{pmatrix},$$

$$\tilde{u} = \begin{pmatrix} 9.999\,999\,9\,E-1 \\ 1\,E-8 \\ -1\,E-8 \\ -1 \\ -9.999\,999\,9\,E-1 \\ 9.999\,999\,9\,E-1 \\ 1\,E-8 \end{pmatrix} \approx x^5, \quad \tilde{v} = \begin{pmatrix} 1 \\ 1\,E-8 \\ -1\,E-8 \\ -9.999\,999\,9\,E-1 \\ 9.999\,999\,9\,E-1 \\ -9.999\,999\,9\,E-1 \\ 1\,E-8 \end{pmatrix} \approx x^6 .$$

The modified method with $i_1 = 4$, $i_2 = 5$ (instead of $i_1 = n-1 = 6$, $i_2 = n = 7$) and with $C \approx B^{-1}$ from (99) yields

$$\begin{pmatrix} -5 & 0 \\ 0 & -5 \end{pmatrix} + \begin{pmatrix} [-1,1]\cdot 10^{-12} & [-1,0.3]\cdot 10^{-19} \\ [-2.6,2.6]\cdot 10^{-20} & [-1,1]\cdot 10^{-12} \end{pmatrix},$$

$$x^5 + \begin{pmatrix} [-0.1,1]\cdot 10^{-11} \\ [-1,0.2]\cdot 10^{-19} \\ [-0.2,1]\cdot 10^{-19} \\ 0 \\ 1.0\cdot 10^{-8} \\ [-1.000\,1 - 0.009\,9]\cdot 10^{-8} \\ [-1,0.2]\cdot 10^{-19} \end{pmatrix}, \qquad x^6 + \begin{pmatrix} [-1.000\,1 - 0.999\,9]\cdot 10^{-8} \\ [-2,1.2]\cdot 10^{-19} \\ [-0.3,1]\cdot 10^{-19} \\ -0.999\,999\,99 \\ 0.999\,999\,99 \\ [0.999\,9, 1.000\,1]\cdot 10^{-8} \\ [-1,0.3]\cdot 10^{-19} \end{pmatrix}$$

as verified enclosures for $M = \begin{pmatrix} -5 & 0 \\ 0 & -5 \end{pmatrix}$ from (77) and for the generators $u^* = (1 - 0.5\cdot 10^{-8})\,x^5 + 0.5\cdot 10^{-8}\,x^6$ and $v^* = 0.999\,999\,99 \cdot x^6$, respectively.

□

6. SYMMETRIC MATRICES

To enclose eigenvalues λ^* and corresponding eigenvectors x^* (at least if λ^* is simple) for symmetric matrices A we present a method described in [79]. The basic idea is to transform A by a finite number of Jacobi rotations $J_{pq}^T A J_{pq} =: A'$ with

$$J_{pq} := \begin{array}{cl} \left(\begin{array}{ccccccccccc} 1 & & & & & & & & & & \\ & \ddots & & & & & & O & & & \\ & & 1 & & & & & & & & \\ & & & c & 0 & \cdots & 0 & -s & & & \\ & & & 0 & 1 & & & 0 & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & 0 & & & 1 & 0 & & & \\ & & & s & 0 & \cdots & 0 & c & & & \\ & & & & & & & & 1 & & \\ O & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{array}\right) & \begin{array}{l} \\ \\ \\ \leftarrow p \\ \\ \\ \\ \leftarrow q \\ \\ \\ \\ \end{array} \\ \begin{array}{ccccccccccc} & & & \uparrow & & & & \uparrow & & & \\ & & & p & & & & q & & & \end{array} & \end{array} , \quad c := \cos\varphi, \quad s := \sin\varphi \qquad (100)$$

nearly to diagonal form. Then an interval version of Gershgorin's Theorem 15 is used to enclose the eigenvalues of A, and Wilkinson's results in Theorem 16 are applied to enclose eigenvectors associated with simple eigenvalues. The angle φ in (100) is determined so that $a'_{pq} = 0$ holds where (p, q) are chosen in a row cyclic way, i. e., (p, q) runs through the scheme

$$\begin{array}{ccccccc} (1,2) & \rightarrow & (1,3) & \rightarrow & \dots & \rightarrow & (1,n) \\ & & (2,3) & \rightarrow & \dots & \rightarrow & (2,n) \\ & & & & \ddots & & \vdots \\ & & & & & & (n-1,n) \end{array}$$

rowwise from left to right starting with (1,2) and restarting with the same index pair if one arrives at $(n-1, n)$. The method is known as the *row cyclic Jacobi method.* As in [124], the entries of A' are computed via the following procedure, provided $a_{pq} \neq 0$.

Algorithm 31 *(Jacobi method; one step)*

$$\theta := \frac{a_{pp} - a_{qq}}{2a_{pq}} \qquad (= \cot(2\varphi))$$

$$t := \tan\varphi := \begin{cases} \dfrac{1}{\theta + (\mathit{sign}\,\theta)\sqrt{\theta^2+1}} & \mathit{if}\ \theta \neq 0 \\ 1 & \mathit{if}\ \theta = 0 \end{cases}$$

$$c := \frac{1}{\sqrt{1+t^2}} \qquad s := ct \ ; \qquad \tau := \tan\left(\frac{\varphi}{2}\right) = \frac{s}{1+c}$$

$$\begin{aligned} a'_{pp} &:= a_{pp} + t a_{pq} \\ a'_{qq} &:= a_{qq} - t a_{pq} \\ a'_{pq} &:= a'_{qp} := 0 \\ a'_{pj} &:= a'_{jp} := a_{pj} + s(a_{pj} - \tau a_{pj}) \ , \qquad j \neq q \\ a'_{qj} &:= a'_{jq} := a_{qj} - s(a_{pj} + \tau a_{qj}) \ , \qquad j \neq p \\ a'_{ij} &:= a_{ij} \ , \qquad i, j \notin \{p, q\} \end{aligned}$$

□

As is well-known, the explicit computation of φ is not necessary. Instead of assuming $a_{pq} \neq 0$ for Algorithm 31, one often uses a threshold variant of the Jacobi method, skipping the computational process for an index pair (p, q) if $|a_{pq}| \leq \varepsilon_k$ with $\varepsilon_k > 0$ being some given small number acting as threshold. If $|a_{ij}| < \varepsilon_k$ for all $i < j$, then ε_k is replaced by some $\varepsilon_{k+1} \in (0, \varepsilon_k)$ such that the sequence $\{\varepsilon_k\}_{k=1}^{\infty}$ decreases strictly to zero. Taking into account $J_{pq}^T = J_{pq}^{-1}$ it is known ([37], [50]) that the (infinite) row cyclic Jacobi method converges to the Jordan normal form of A, which, by the symmetry of A, is a diagonal matrix with the eigenvalues of A as diagonal entries.

The method in [79] now transforms A by means of (100) into a diagonally dominant matrix $\hat{A}$ with

$$\varepsilon' |\hat{a}_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |\hat{a}_{ij}| , \qquad i = 1, \ldots, n , \qquad \varepsilon' \text{ sufficiently small.} \tag{101}$$

The transforming matrices J_{pq} are multiplied together to form a single transformation matrix S. Unfortunately, due to rounding errors, one only gets an approximation $\tilde{S}$ of S satisfying $\tilde{S}^{-1} \approx \tilde{S}^T$ because of $S^{-1} = S^T$. To get an enclosure for $\tilde{S}^{-1} A \tilde{S}$ one first forms $A\tilde{S}$ using interval arithmetic. Then one has essentially two possibilities.

(i) Solve the n interval systems

$$[A\tilde{S}]\,[z]^{(i)} = e^{(i)} \ .$$ [1]

Then

$$\tilde{S}^{-1} \in [S]^{-1} := \left([z]^{(1)}, \ldots, [z]^{(n)}\right) \in M_{nn}(I(\boldsymbol{R})) \ . \tag{102}$$ [2]

(ii) Compute $\alpha := \|\tilde{S}^T\|_\infty \dfrac{\|I - \tilde{S}^T\tilde{S}\|_\infty}{1 - \|I - \tilde{S}^T\tilde{S}\|_\infty}$. Then

$$\tilde{S}^1 \in [S]^{-1} := \tilde{S} + [-\alpha, \alpha]\, ee^T \ , \ \text{provided} \ \|I - \tilde{S}^T\tilde{S}\|_\infty < 1 \ . \tag{103}$$

(102) is trivial, (103) follows from the next lemma.

Lemma 32 *Let*

$$\|I - CS\|_\infty < 1 \quad \textit{for} \quad C, S \in M_{nn}(\boldsymbol{R}) \ . \tag{104}$$

Then C, S are nonsingular, and

$$S^{-1} \in C + [-\alpha, \alpha] ee^T \quad \textit{with} \quad \alpha := \|C\|_\infty \frac{\|I - CS\|_\infty}{1 - \|I - CS\|_\infty} \ .$$

[1] This equation has to be interpreted in the usual way; see [95], e. g.
[2] $[S]^{-1}$ is used only symbolically; it is not an inverse in the algebraic sense.

Proof.
The existence of C^{-1}, S^{-1} follows immediately from (104). By Neumann's series, we obtain

$$\begin{aligned} S^{-1} - C &= (I - CS)\{I - (I - CS)\}^{-1}C \\ &= (I - CS) \sum_{k=0}^{\infty} (I - CS)^k C , \end{aligned}$$

hence

$$\|S^{-1} - C\|_\infty \leq \|I - CS\|_\infty \sum_{k=0}^{\infty} \|I - CS\|_\infty^k \|C\|_\infty = \alpha .$$

□

Having computed $[B] := [S]^{-1}(A\tilde{S})$, Theorem 15 shows by a worst case analysis that the eigenvalues of A are contained in the set

$$[g] := \bigcup_{i=1}^{n} [g]_i$$

with the Gershgorin intervals

$$[g]_i := [\underline{b}_{ii} - \sum_{\substack{j=1 \\ j\neq i}}^{n} |[b]_{ij}| \ , \ \overline{b}_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{n} |[b]_{ij}|] \qquad i = 1, \ldots, n .$$

If

$$[g]_i \cap \bigcup_{\substack{j=1 \\ j\neq i}}^{n} [g]_j = \emptyset \tag{105}$$

then we call a Gershgorin interval $[g]_i$ *isolated.*

Such an isolated interval $[g]_i$ contains exactly one eigenvalue λ^* of A by Theorem 15b). It is algebraic simple. This remark holds trivially for *all* eigenvalues of A if the intervals $[g]_i$, $i = 1, \ldots, n$, turn out to be pairwise disjoint. Such pairwise disjoint Gershgorin intervals are expected when dealing with the inverse eigenvalue problem described in Section 9. Clustering can also be verified by the method above by modifying Theorem 15b) in a straightforward way.

The bounds $\underline{g}_i, \overline{g}_i$ can be improved if one intersects $[g]_i$ with additional enclosures based on estimates like those of Theorem 11, or if one increases the precision which can be done using particular features of modern programming languages like PASCAL – XSC, combined with computations in staggered correction format described in [19], [81], [123]. In addition, the Gram–Schmidt procedure can be applied to re–orthogonalize $\tilde{S}$. For a description of such computational details we refer to [79].

Let now λ^* be a simple eigenvalue which is enclosed by an isolated Gershgorin interval $[g]_i$. In order to obtain an enclosure for a corresponding eigenvector x^* which is normalized by $(x^*)^T x^* = 1$, apply Theorem 16b) with

$$\begin{aligned} &\tilde{x} := \frac{\tilde{S}_{*,i}}{\|\tilde{S}_{*,i}\|_2} \\ &\tilde{\lambda} := \mathrm{mid}\,([g]_i) \ , \quad r := \|A\tilde{x} - \tilde{\lambda}\tilde{x}\|_2 \ , \\ &a := \min_{\substack{1\leq j\leq n \\ j\neq i}} \left\{\min\left\{|c| \mid c \in [g]_j - \tilde{\lambda}\right\}\right\} - \eta \qquad (\eta \text{ any small number}) . \end{aligned}$$

Define ε according to (34). If $a \geq r$ holds then $x^* \in \tilde{x} + [-\varepsilon, \varepsilon]e$.

Before we illustrate the foregoing results by an example we point out that in the contribution [25] of this book another enclosure method is described for symmetric matrices. Cf. also [21] - [23], [86], [89].

Example 33

We want to verify the eigenvalues of the modified 18×18 Hilbert matrix [1]

$$A = (a_{ij}) \quad \text{with} \quad a_{ij} := 144\,403\,552\,893\,600/(i + j - 1) \; .$$

This matrix is symmetric and positive definite having integer entries. Using the method described above and improving the Gershgorin intervals by the Temple estimate from Theorem 11 yields the following enclosures $[\lambda]_i$ for selected eigenvalues λ_i^*.

i	$[\lambda]_i$
1	$1.225\,612\,625\,222\,07\,^{4}_{3}E - 011$
2	$4.765\,338\,711\,530\,30\,^{7}_{5}E - 009$
5	$8.722\,494\,855\,391\,5\,^{21}_{18}E - 003$
10	$1.020\,114\,001\,053\,09\,^{2}_{1}E + 006$
18	$2.722\,318\,555\,347\,66\,^{1}_{0}E + 014$

Note that the smallest eigenvalue λ_1 is particularly difficult to verify.

□

7. THE GENERALIZED EIGENPROBLEM

As is well-known, the generalized eigenproblem

$$Ax = \lambda Bx \; , \qquad A,\, B \in M_{nn}(C) \; , \quad B \text{ nonsingular,} \tag{106}$$

can be reduced at once to the simple eigenproblem (1) by multiplying (106) with B^{-1}. In practice, because of rounding errors, $B^{-1}A$ normally cannot be computed exactly, but it can be enclosed 'solving', e. g., the n linear systems

$$Bz^j = A_{*,j} \; , \quad j = 1, \ldots, n \; , \quad (A_{*,j} \; j\text{–th column of } A)$$

by the verification methods mentioned in Section 5.1. The resulting inclusions $[z]^j$ of z^j then yield an enclosure for the columns of B^{-1}, and the eigenvalue / eigenvector methods of Section 5 or 6 can be applied to the interval matrix $([z]^1, \ldots, [z]^n) \cdot A$. There are also other possibilities to attack (106). The first one can be used for simple eigenvalues $\lambda = \lambda^*$ of (106), i. e., for simple eigenvalues of

$$B^{-1}A\,x = \lambda x \; , \tag{107}$$

although there is no need to require simplicity for λ^* from the beginning. But it will turn out as in Section 5.1 that some matrix being involved is nonsingular if λ^* is simple. The

[1] Cf. [45] for eigenpairs of smaller Hilbert matrices.

method proceeds analogously to that in Section 5 for the standard eigenproblem. Start with

$$f(x,\lambda) := \begin{pmatrix} Ax - \lambda Bx \\ x_n - \alpha \end{pmatrix}, \quad \alpha \neq 0,$$

precondition with $-C \in M_{n+1,n+1}(\boldsymbol{C})$, choose an approximation $(\tilde{x}, \tilde{\lambda})$ for an eigenpair (x^*, λ^*) of (106) and expand

$$s(x,\lambda) := \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} - C\, f(x,\lambda)$$

in a Taylor series at $\begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \end{pmatrix}$. Then the exact error $\begin{pmatrix} \Delta x^* \\ \Delta\lambda^* \end{pmatrix} := \begin{pmatrix} x^* - \tilde{x} \\ \lambda^* - \tilde{\lambda} \end{pmatrix}$ is a fixed point of the equation

$$\begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = g(\Delta x, \Delta\lambda),$$

where g is defined by

$$\begin{aligned} g(\Delta x, \Delta\lambda) &:= -C\, f(\tilde{x},\tilde{\lambda}) + \left\{ I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda} B I_n & -B(\tilde{x} + \Delta x) \\ (e^{(n)})^T & 0 \end{pmatrix} \right\} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} \\ &= s(x,\lambda). \end{aligned}$$

Let now A, B, C be real and let $\tilde{x}_n = x_n^* = \alpha$.

Checking the proofs for the Lemmas and Theorems in Section 5.1 shows that with minor modifications the analogues of the Theorems 23 and 24 hold with

$$\begin{aligned} \rho &:= \left\| C \begin{pmatrix} A\tilde{x} - \tilde{\lambda} B\tilde{x} \\ 0 \end{pmatrix} \right\|_\infty, \quad \sigma := \left\| I_{n+1} - C \begin{pmatrix} A - \tilde{\lambda} B & -B\tilde{x} \\ (e^{(n)})^T & 0 \end{pmatrix} \right\|_\infty, \\ \tau &:= \| |C| \cdot |B| \|_\infty, \text{ and} \\ [t]_{ijk} &:= \begin{cases} (|C| \cdot |B|)_{ij} \cdot [-1,1] & \text{if } j \in \{1,\dots,n\} \text{ and if } k = n+1 \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

The absolute values arise because of

$$\begin{aligned} \left(-C \begin{pmatrix} -B[\Delta x] \\ 0 \end{pmatrix} \right) [\Delta\lambda] &= \left(\sum_{s=1}^{n} \sum_{j=1}^{n} c_{is} b_{sj} [\Delta x]_j \right) [\Delta\lambda] \\ &\subseteq \sum_{j=1}^{n} \sum_{s=1}^{n} c_{is} b_{sj} [-1,1] [\Delta x]_j [\Delta\lambda] \\ &= \sum_{j=1}^{n} \left(\sum_{s=1}^{n} |c_{is}| \cdot |b_{sj}| [-1,1] \right) [\Delta x]_j [\Delta\lambda]. \end{aligned}$$

In practicular, the interval version of Theorem 4 has to be applied.

Note that Theorem 2, which is needed for the proofs, still holds if its matrices are adapted according to (106). The two modifications of g in Section 5.1 also hold; replace $\hat{B}$ in (68) by

$$\hat{B} := \left((A - \tilde{\lambda}B)_{*,1}, \ldots, (A - \tilde{\lambda}B)_{*,n-1}, -B\tilde{x}\right) \in M_{nn}(\boldsymbol{R})$$

and ρ, τ in (69) by

$$\rho := \left\|\hat{C}(\tilde{\lambda}B\tilde{x} - A\tilde{x})\right\|_\infty , \quad \tau := \left\||\hat{C}| \cdot |B|\right\|_\infty ,$$

respectively. The function f in (70) reads now

$$f(x, \lambda) := \begin{pmatrix} Ax - \lambda Bx \\ x^T x - \alpha \end{pmatrix} .$$

We also remark that A and B can be replaced by interval matrices $[A]$, $[B]$, and that A, B as well as $[A]$, $[B]$ can be complex.

Another procedure for (106) consists in generalizations of the results in Section 6 provided A and B are symmetric and B is positive definite.

Example 34 *[15]*

Let

$$A = \begin{pmatrix} 10 & 2 & 3 & 1 & 1 \\ 2 & 12 & 1 & 2 & 1 \\ 3 & 1 & 11 & 1 & -1 \\ 1 & 2 & 1 & 9 & 1 \\ 1 & 1 & -1 & 1 & 15 \end{pmatrix}, \qquad B = \begin{pmatrix} 12 & 1 & -1 & 2 & 1 \\ 1 & 14 & 1 & -1 & 1 \\ -1 & 1 & 16 & -1 & 1 \\ 2 & -1 & -1 & 12 & -1 \\ 1 & 1 & 1 & -1 & 11 \end{pmatrix}$$

be the matrices from [140], p. 312. To enclose the smallest eigenvalue λ^* of the generalized eigenproblem $Ax = \lambda Bx$ we chose

$$\tilde{\lambda} = 0.432\,787 \qquad \text{and} \qquad \tilde{x} = \begin{pmatrix} 0.134\,591 \\ -0.612\,947\,E-1 \\ -0.157\,902\,562\,211 \\ 0.109\,466 \\ -0.414\,730\,\;E-1 \end{pmatrix} .$$

With the exception of $\tilde{x}_3$ these values coincide with the first six significant digits of Table 3 in [140], p. 313. The digits of $\tilde{x}_3$ are the same as in that table. Prescribing the third component of x^* by $x_3^* = \alpha := \tilde{x}_3$ instead of the n–th one, and changing $\hat{B}$ accordingly one gets, with $\hat{C} := \hat{B}^{-1}$, for the modified method described above the verified enclosures

$$\lambda^* \in [\,0.432\,787\,211\,01_7^6\,] \qquad \text{and} \qquad x^* \in \begin{pmatrix} [\;\; 0.134\,590\,573\,96_2^1 &] \\ [\; -0.612\,947\,224\,71_4^5\;E-1 &] \\ [\; -0.157\,902\,562\,21_1^1 &] \\ [\;\; 0.109\,465\,787\,72_4^3 &] \\ [\; -0.414\,730\,117\,96_5^6\;E-1 &] \end{pmatrix} .$$

Another example and more details can be found in [7].

8. THE SINGULAR VALUE PROBLEM

The singular value problem for rectangular matrices $A \in M_{mn}(\boldsymbol{C})$ consists in finding two unitary matrices $U \in M_{nn}(\boldsymbol{C})$, $V \in M_{mm}(\boldsymbol{C})$ and a rectangular matrix $\Sigma \in M_{mn}(\boldsymbol{C})$ such that

$$A = V\Sigma U^H \tag{108}$$

holds where Σ has the form

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & O & \\ & & \sigma_r & & \\ & & & \ddots & \\ & & & & \sigma_n \\ & & & & 0 \\ & O & & & \vdots \\ & & & & 0 \end{pmatrix}$$

in the case $m \geq n$, and

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & & \\ & \ddots & & & O & & \\ & & \sigma_r & & & & \\ & O & & \ddots & & & \\ & & & & \sigma_m & 0 & \dots & 0 \end{pmatrix}$$

in the case $m < n$, with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = 0 = \dots = \sigma_{\min\{m,n\}}$ and $r = \operatorname{rank}(A)$. The numbers σ_i, $i = 1, \dots, \min\{m,n\}$ are called *singular values* of A, the corresponding columns $u^{(i)}$, $v^{(i)}$ of U and V are called *right singular vectors* and *left singular vectors*, respectively. The task of computing the singular values σ_i is called *singular value problem.* From theory one knows that the decomposition in (108) exists. Since (108) yields

$$\left.\begin{array}{rcl} Au^{(i)} & = & \sigma_i v^{(i)} \\ A^H v^{(i)} & = & \sigma_i u^{(i)} \end{array}\right\} \qquad i = 1, \dots, \min\{m,n\} \tag{109}$$

the vectors $u^{(i)}$, $v^{(i)}$ are eigenvectors of the matrices $A^H A \in M_{nn}(\boldsymbol{C})$ and $AA^H \in M_{mm}(\boldsymbol{C})$, respectively, associated with the eigenvalues σ_i^2. In this respect one could use the methods in the Sections 5 and 6 to enclose $u^{(i)}$, $v^{(i)}$ and σ_i^2. Another way consists of considering the function

$$f : V_{n+m+1}(\boldsymbol{R}) \to V_{n+m+1}(\boldsymbol{R})$$

defined by

$$f(u, v, \sigma) := \begin{pmatrix} Au - \sigma v \\ A^H v - \sigma u \\ u^H u - 1 \end{pmatrix} \tag{110}$$

of which the zeros (u^*, v^*, σ^*), $\sigma^* \neq 0$, contain a pair of singular vectors u^*, v^* associated with the singular value σ^*. The singular vectors are normalized according to

$$(u^*)^H u^* = 1, \qquad (v^*)^H v^* = 1 \tag{111}$$

which follows from the definition of f and from

$$Au^* = \sigma^* v^*, \quad A^H v^* = \sigma^* u^*$$

whence

$$\begin{aligned} (v^*)^H v^* &= (v^*)^H \frac{1}{\sigma^*} A u^* = \frac{1}{\sigma^*} \left(A^H v^*\right)^H u^* \\ &= (u^*)^H u^* = 1. \end{aligned}$$

For $\sigma^* = 0$ the zeros (u^*, v^*, σ^*) of f do not necessarily fulfill $(v^*)^H v^* = 1$, since obviously $f(u^*, 0, 0) = 0$ for each right singular vector u^* of A which is associated with $\sigma^* = 0$. To force v^* to be normalized by (111) even in this case, one can start with the function

$$f(u, v, \sigma, \sigma') := \begin{pmatrix} Au - \sigma v \\ A^H v - \sigma' u \\ u^H u - 1 \\ v^H v - 1 \end{pmatrix} \tag{112}$$

repeating the steps to follow. The zeros of (112) trivially fulfill (111), and they also satisfy $\sigma = \sigma'$. This latter equality can be seen from (111) and from

$$\begin{aligned} Au = \sigma v &\;\Rightarrow\; v^H A u = \sigma v^H v = \sigma; \\ A^H v = \sigma' u &\;\Rightarrow\; u^H A^H v = \sigma' u^H u = \sigma'; \\ &\;\Rightarrow\; \sigma = v^H A u = \left(u^H A^H v\right)^H = \sigma'. \end{aligned}$$

Cf. [5] for details.

We restrict ourselves now to the real case, i.e., we assume $A \in M_{mn}(\boldsymbol{R})$. Then Σ, U, V are real matrices. Introducing approximations $\tilde{u}, \tilde{v}, \tilde{\sigma}$ and the differences

$$\Delta u := u - \tilde{u}, \quad \Delta v := v - \tilde{v}, \quad \Delta\sigma := \sigma - \tilde{\sigma}$$

yields the fixed point problem

$$\begin{pmatrix} \Delta u \\ \Delta v \\ \Delta\sigma \end{pmatrix} = g(\Delta u, \Delta v, \Delta\sigma) := -Cf(\tilde{u}, \tilde{v}, \tilde{\sigma}) + (I - CB) \begin{pmatrix} \Delta u \\ \Delta v \\ \Delta\sigma \end{pmatrix} + \tilde{T} \begin{pmatrix} \Delta u \\ \Delta v \\ \Delta\sigma \end{pmatrix} \tag{113}$$

with the $(m+n+1) \times (n+m+1)$ matrices

$$B := \begin{pmatrix} A & -\tilde{\sigma} I_m & -\tilde{v} \\ -\tilde{\sigma} I_n & A^T & -\tilde{u} \\ 2\tilde{u}^T & 0 & 0 \end{pmatrix} \tag{114}$$

and

$$\tilde{T} := C \begin{pmatrix} 0 & 0 & \Delta v \\ 0 & 0 & \Delta u \\ (\Delta u)^T & 0 & 0 \end{pmatrix} .$$

This follows analogously to Section 5.1 by preconditioning f with $-C$, adding $\begin{pmatrix} \Delta u \\ \Delta v \\ \Delta\sigma \end{pmatrix}$ on both sides of the equation

$$0 = -Cf(u, v, \sigma)$$

and evaluating the righthand side into a Taylor series at the approximation $\begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \\ \tilde{\sigma} \end{pmatrix}$. The series terminates after the third summand of the expansion, hence g is a quadratic function as defined in (10) with

$$\begin{aligned} r &:= -Cf(\tilde{u}, \tilde{v}, \tilde{\sigma}) \ , \\ S &:= I - CB \ , \end{aligned}$$

and with $T = (t_{ijk})$ given by

$$t_{ijk} := \begin{cases} c_{ij} & \text{if } j \in \{1, \dots, m\} \text{ and } k = m + n + 1 \\ c_{i,m+j} & \text{if } j \in \{1, \dots, n\} \text{ and } k = m + n + 1 \\ c_{i,m+n+1} & \text{if } j = k \in \{1, \dots, n\} \\ 0 & \text{otherwise} \end{cases} .$$

Therefore, Theorem 4 yields a method and at the same time a criterion for verifying solutions (u^*, v^*, σ^*) of the singular value problem. With

$$\rho := \|Cf(\tilde{u}, \tilde{v}, \tilde{\sigma})\|_\infty \, , \ \hat{\sigma} := \|I - CB\|_\infty \, ,^1 \ \text{ and } \ \tau := \| |C| \cdot (1, \dots, 1, n)^T \|_\infty$$

one gets at once the following result.

Theorem 35 *With the notations above let*

$$\begin{aligned} \hat{\sigma} &< 1, \ \Delta := (1 - \hat{\sigma})^2 - 4\rho\tau \geq 0, \\ \beta^- &:= (1 - \hat{\sigma} - \sqrt{\Delta})/(2\tau), \\ \beta^+ &:= (1 - \hat{\sigma} + \sqrt{\Delta})/(2\tau), \end{aligned} \tag{115}$$

a) If $\beta \in [\beta^-, \beta^+]$ then g has in $\begin{pmatrix} [\Delta u]^{(0)} \\ [\Delta v]^{(0)} \\ [\Delta\sigma]^{(0)} \end{pmatrix} := [-\beta, \beta]e \in V_{n+m+1}(I(\boldsymbol{R}))$ at least one fixed point $\begin{pmatrix} \Delta u^* \\ \Delta v^* \\ \Delta\sigma^* \end{pmatrix} \in V_{n+m+1}(\boldsymbol{R})$ with $u^* := \tilde{u} + \Delta u^* \in V_n(\boldsymbol{R})$, $v^* := \tilde{v} + \Delta v^* \in$

[1] To distinguish the quantity σ of Theorem 4 from a singular value, we replaced the notation σ in Theorem 4 by $\hat{\sigma}$.

$V_m(\boldsymbol{R})$, $\sigma^* := \tilde{\sigma} + \Delta\sigma^* \in \boldsymbol{R}$ forming for $\sigma^* \neq 0$ a triple of a right singular vector, a left singular vector and a corresponding singular value such that the normalizations (111) are satisfied. The iteration

$$\begin{pmatrix} [\Delta u]^{(k+1)} \\ [\Delta v]^{(k+1)} \\ [\Delta\sigma]^{(k+1)} \end{pmatrix} := g\left([\Delta u]^{(k)}, [\Delta v]^{(k)}, [\Delta\sigma]^{(k)}\right) \ , \quad k = 0, 1, \dots \ , \tag{116}$$

converges to some interval vector $\begin{pmatrix} [\Delta u]^* \\ [\Delta v]^* \\ [\Delta\sigma]^* \end{pmatrix}$ *with*

$$\begin{aligned} \begin{pmatrix} \Delta u^* \\ \Delta v^* \\ \Delta\sigma^* \end{pmatrix} \in \begin{pmatrix} [\Delta u]^* \\ [\Delta v]^* \\ [\Delta\sigma]^* \end{pmatrix} \subseteq \begin{pmatrix} [\Delta u]^{(k)} \\ [\Delta v]^{(k)} \\ [\Delta\sigma]^{(k)} \end{pmatrix} \subseteq \begin{pmatrix} [\Delta u]^{(k-1)} \\ [\Delta v]^{(k-1)} \\ [\Delta\sigma]^{(k-1)} \end{pmatrix} \\ \subseteq \dots \subseteq \begin{pmatrix} [\Delta u]^{(0)} \\ [\Delta v]^{(0)} \\ [\Delta\sigma]^{(0)} \end{pmatrix} , \qquad k \in \boldsymbol{N} \ . \end{aligned} \tag{117}$$

b) If $\beta \in [\beta^-, (\beta^- + \beta^+)/2)$ then g has in $\begin{pmatrix} [\Delta u]^{(0)} \\ [\Delta v]^{(0)} \\ [\Delta\sigma]^{(0)} \end{pmatrix} := [-\beta, \beta]e \in V_{n+m+1}(I(\boldsymbol{R}))$ *a unique fixed point* $\begin{pmatrix} \Delta u^* \\ \Delta v^* \\ \Delta\lambda^* \end{pmatrix}$*, and (117) holds with* $\begin{pmatrix} [\Delta u]^* \\ [\Delta v]^* \\ [\Delta\sigma]^* \end{pmatrix} = \begin{pmatrix} \Delta u^* \\ \Delta v^* \\ \Delta\sigma^* \end{pmatrix}$*, i. e., the iteration (116) converges to* $\begin{pmatrix} \Delta u^* \\ \Delta v^* \\ \Delta\sigma^* \end{pmatrix}$.

□

As in Section 5.1 one normally chooses $C \approx B^{-1}$ so that $\hat{\sigma} \approx 0$. If the approximations $\tilde{u}, \tilde{v}, \tilde{\sigma}$ are sufficiently close to a solution $u^*, v^*, \sigma^* \neq 0$, the weighted residual r will be small, hence the assumptions (115) of Theorem 35 will be fulfilled. Note that $\sigma \neq 0$ certainly holds if $0 \notin \tilde{\sigma} + [\Delta\sigma]^{(0)}$ and that a similar Theorem holds if g is based on (112).

As we are going to see by the subsequent Theorem 36 and by continuity arguments, the inverse B^{-1} certainly exists if the approximations $\tilde{u}$, $\tilde{v}$ of the singular values u^*, v^* are not too bad and if the corresponding singular value σ^* is simple, i. e., if it is a simple eigenvalue of the matrix A^TA. Since this matrix is symmetric σ^* is then automatically algebraic simple. We already remarked that the columns $u^{(i)}$, $i = 1, \dots, \min\{m, n\}$, of the matrix U from (108) fulfill

$$A^T A\, u^{(i)} = (\sigma_i^*)^2\, u^{(i)}, \tag{118}$$

with σ^* being the i–th singular value. Therefore, because U is nonsingular, a singular value $\sigma^* \neq 0$ is simple if and only if it occurs only once in the matrix Σ whence, by

$$AA^T v^{(i)} = (\sigma_i^*)^2 v^{(i)}, \quad v^{(i)} \ i\text{–th column of } U,$$

a simple non-zero singular value is also an algebraic simple eigenvalue of AA^T and vice versa.

We are now ready to prove the following theorem on the simplicity of singular values.

Theorem 36 *Let $\sigma^* \neq 0$ be a singular value of $A \in M_{mn}(\boldsymbol{R})$. Then σ^* is simple if and only if the matrix*

$$B^* := \begin{pmatrix} A & -\sigma^* I_m & -v^* \\ -\sigma^* I_n & A^T & -u^* \\ 2(u^*)^T & 0 & 0 \end{pmatrix} \in M_{m+n+1,m+n+1}(\boldsymbol{R})$$

is nonsingular with u^*, v^* denoting a right and a left singular vector of B^*, respectively, associated with $\sigma^* \neq 0$.

Proof.
The idea of the proof is already contained in the proof of Theorem 2. Let $\sigma^* \neq 0$ be simple and assume B^* to be singular. Then there is a vector $z = \begin{pmatrix} z^1 \\ z^2 \\ z' \end{pmatrix}$ with $z^1 \in V_n(\boldsymbol{R})$, $z^2 \in V_m(\boldsymbol{R})$ and $z' \in \boldsymbol{R}$ such that $z \neq 0$ and $B^* z = 0$. Hence

$$Az^1 - \sigma^* z^2 = z'v^*, \tag{119}$$
$$A^T z^2 - \sigma^* z^1 = z'u^*, \tag{120}$$
$$(u^*)^T z^1 = 0\,. \tag{121}$$

If $z' = 0$ then

$$A^T A z^1 = \sigma^* A^T z^2 = (\sigma^*)^2 z^1$$

whence, by the simplicity of σ^*, we have $z^1 = \alpha u^*$. Therefore, (121) implies $z^1 = 0$, and (119) together with $\sigma^* \neq 0$ yields $z^2 = 0$ contradicting $z \neq 0$.

If $z' \neq 0$ then we get from (119), (120)

$$A^T A z^1 - (\sigma^*)^2 z^1 - z'\sigma^* u^* = z' A^T v^* = z'\sigma^* u^*$$

hence

$$(A^T A - (\sigma^*)^2 I_n) z^1 = 2z'\sigma^* u^*.$$

Multiplying this equation with $A^T A - (\sigma^*)^2 I_n$ and taking (118) into account shows that z^1 is a principal vector of degree two for the symmetric matrix $A^T A$ which is impossible since $A^T A$ is diagonalizable. Therefore, B^* is nonsingular.

Assume now B^* to be nonsingular and $\sigma^* \neq 0$ to be a multiple singular value. According to the discussion preceding Theorem 36, σ^* occurs multiply in Σ, whence there is a second pair $\hat{u} \in V_n(\boldsymbol{R})$, $\hat{v} \in V_m(\boldsymbol{R})$ of singular vectors, associated with σ^*. With (109) and by the orthonormality of U, V this yields the contradiction

$$\left((\hat{v}^*)^T, (\hat{u}^*)^T, 0\right) B^* = \left((\hat{v}^*)^T A - \sigma^*(\hat{u}^*)^T, (\hat{u}^*)^T A^T - \sigma^*(\hat{v}^*)^T, -(\hat{u}^*)^T u^* - (\hat{v}^*)^T v^*\right) = 0\,.$$

□

We finally remark that A can again be replaced by interval matrices $[A] \in M_{mn}(I(\boldsymbol{R}))$ and that the complex case can be handled analogously.

The subsequent example which we borrowed from [5] illustrates the efficiency of the method described in Theorem 35 .

Example 37 *[5]*

We want to enclose the largest singular value σ_1 of the 5×3 matrix

$$A = \begin{pmatrix} 1 & 6 & 11 \\ 2 & 7 & 12 \\ 3 & 8 & 13 \\ 4 & 9 & 14 \\ 5 & 10 & 15 \end{pmatrix}$$

with the approximations $\tilde{\sigma}_1 = 0.351\,272\,233E+2$,

$$\tilde{u} = \begin{pmatrix} 0.201\,664\,911 \\ 0.516\,830\,501 \\ 0.831\,996\,092 \end{pmatrix}, \quad \tilde{v} = \begin{pmatrix} 0.354\,557\,057 \\ 0.398\,696\,370 \\ 0.442\,835\,683 \\ 0.486\,974\,996 \\ 0.531\,114\,309 \end{pmatrix}.$$

With $C := B^{-1}$ (cf. the paragraphs following the proof of Theorem 17 for a discussion of this choice) we got

$$\sigma_1 \in [\,0.351\,272\,233\,33^{5}_{6}\ E+2\,]\ ,$$

$$u^* \in \begin{pmatrix} [\,0.201\,664\,911\,19^{2}_{3}\,] \\ [\,0.516\,830\,501\,39^{2}_{3}\,] \\ [\,0.831\,996\,091\,59^{1}_{2}\,] \end{pmatrix}, \qquad \text{and } v^* \in \begin{pmatrix} [\,0.354\,557\,057\,03^{7}_{8}\,] \\ [\,0.398\,696\,369\,99^{8}_{9}\,] \\ [\,0.442\,835\,682\,9^{59}_{60}\,] \\ [\,0.486\,974\,995\,92^{1}_{2}\,] \\ [\,0.531\,114\,308\,88^{2}_{3}\,] \end{pmatrix}.$$

9. AN INVERSE EIGENVALUE PROBLEM

The inverse eigenvalue problem which we want to consider in this section, consists in finding n real numbers c_i^*, $i = 1, \ldots, n$, such that the matrix

$$A(c) := A_0 + \sum_{i=1}^{n} c_i A_i, \quad c := (c_i) \in V_n(\boldsymbol{R})\ , \tag{122}$$

has for $c = c^* = (c_i^*)$ prescribed eigenvalues

$$\lambda_1^* < \lambda_2^* < \ldots < \lambda_n^*\ . \tag{123}$$

Here, A_i, $i = 0, \ldots, n$, are given symmetric $n \times n$ matrices, so that $A(c) = A(c)^T$ holds. The problem arises, e. g., if one wants to find the spring constants in a spring–mass–wall device sketched in Fig. 1, where the eigenfrequencies of the system and one single spring constant are given. The situation is made clearer in the following example.

Example 38

Let n masses m_i be coupled by springs on a straight line and by two rigid walls standing in parallel as is indicated in Fig. 1.

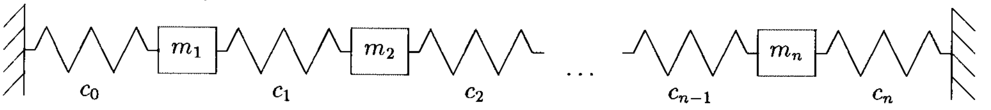

Fig. 1

By Hooke's law and by Newton's second law, the system is described by

$$\left.\begin{aligned} B\ddot{x}(t) &= -Ax(t), \quad x(t) = (x_i(t)) \in V_n(\boldsymbol{R}) \\ x(t_0) &= x_0 \\ \dot{x}(t_0) &= \dot{x}_0 \end{aligned}\right\} \tag{124}$$

with $x_i(t)$ denoting the distance of the i–th mass from its equilibrium position at the time t, with the mass matrix

$B := \operatorname{diag}(m_1, \ldots, m_n) \in M_{nn}(\boldsymbol{R})$

and with the symmetric $n \times n$ matrix

$$A := \begin{pmatrix} c_0 + c_1 & -c_1 & & & & \\ -c_1 & c_1 + c_2 & -c_2 & & O & \\ & -c_2 & c_2 + c_3 & -c_3 & & \\ & & \ddots & \ddots & \ddots & \\ & O & & -c_{n-2} & c_{n-2} + c_{n-1} & -c_{n-1} \\ & & & & -c_{n-1} & c_{n-1} + c_n \end{pmatrix}$$

in which the c_i mean the spring constants. The vectors x_0 and $\dot{x}_0$ are the initial values for the displacement x and for the velocity $\dot{x}$. A fundamental system of (124) is given by

$\{\, x(t) \,|\, x(t) = v^i \cos(\sqrt{\lambda_i} t) \text{ or } x(t) = v^i \sin(\sqrt{\lambda_i} t)\,,\ i = 1, \ldots, n \,\}$

with

$$\lambda_i B v^i = A v^i. \tag{125}$$

As is well–known, $\sqrt{\lambda_i}$ are the eigenfrequencies of the oscillating system. It is easy to see that a matrix of the form (122) can be introduced in (125). To this end let $B^{\frac{1}{2}} := \operatorname{diag}(\sqrt{m_1}, \ldots, \sqrt{m_n})$, $w^i := B^{\frac{1}{2}} v^i$, and define the symmetric $n \times n$ matrices $H^{(l)}$, $l = 1, \ldots, n$, by

$$h_{ij}^{(0)} := \begin{cases} c_0 & \text{if } i = j = 1 \\ 0 & \text{otherwise} \end{cases} ,$$

$$h_{ij}^{(l)} := \left\{ \begin{array}{rl} 1 & \text{if } i=j \in \{l, l+1\} \\ -1 & \text{if } (i,j) \in \{(l,l+1),\ (l+1,l)\} \\ 0 & \text{otherwise} \end{array} \right\} \quad l = 1, \ldots, n-1,$$

$$h_{ij}^{(n)} := \begin{cases} 1 & \text{if } i=j=n \\ 0 & \text{otherwise} \end{cases} .$$

With $B^{-\frac{1}{2}} := (B^{\frac{1}{2}})^{-1}$, equation (125) can be written equivalently as

$$\lambda_i w^i = B^{-\frac{1}{2}} A B^{-\frac{1}{2}} w^i = \left(B^{-\frac{1}{2}} H^{(0)} B^{-\frac{1}{2}} + \sum_{l=1}^{n} c_l B^{-\frac{1}{2}} H^{(l)} B^{-\frac{1}{2}} \right) w^i$$

which is (122) with the symmetric matrices $A_i := B^{-\frac{1}{2}} H^{(i)} B^{-\frac{1}{2}}$.

□

To find a verified solution $c^* \in V_n(\boldsymbol{R})$ of the inverse eigenvalue problem we start with the function

$$f : \begin{cases} U^* & \rightarrow & V_n(\boldsymbol{R}) \\ c & \mapsto & \lambda(c) - \lambda^* \end{cases} \tag{126}$$

where $\lambda(c) := (\lambda_i(c))$ is the vector of which the components are the eigenvalues of $A(c)$, ordered increasingly like the components of $\lambda^* = (\lambda_i^*) = \lambda(c^*)$; cf. (123). If we assume for the moment that c^* exists then, by Theorem 1b), there is certainly a neighbourhood U^* of c^* such that for $c \in U^*$ the eigenvalues of $A(c)$ are simple and can therefore be thought to be ordered as required. This neighbourhood has been chosen in (126) to define f. Trivially, the zeros of f are just the solutions of the inverse eigenvalue problem.

We want to apply Newton's method to obtain an approximation of c^*. To this end we express the Jacobian f' of f by means of the given matrices A_i and by the eigenvectors $(x^*)^i(c)$ of $A(c)$, associated with $\lambda_i(c)$ and normalized by

$$\left((x^*)^i(c)\right)^T (x^*)^i(c) = 1 . \tag{127}$$

To avoid the ambiguity which we mentioned in Section 5, we think of $sign((x^*)^i_{i_0}(c))$ being fixed for some component $(x^*)^i_{i_0}(c)$ of $(x^*)^i(c)$. Although Theorem 1b) was proved assuming the normalization (2), it is easy to see that its assertion also holds in the present situation. One only has to consider $\dfrac{x^*}{\sqrt{(x^*)^T x^*}}$ with $x^*_{i_0}$. Multiplying

$$A(c)(x^*)^i(c) = \lambda_i^*(c)(x^*)^i(c)$$

by $((x^*)^i(c))^T$ from the left and taking into account (127) yields

$$\left((x^*)^i(c)\right)^T A(c)(x^*)^i(c) = \lambda_i^*(c).$$

Differentiating both sides with respect to c_j results in

$$\begin{aligned} \frac{\partial \lambda_i^*(c)}{\partial c_j} &= \frac{\partial \left((x^*)^i(c)\right)^T}{\partial c_j} A(c) \left((x^*)^i(c)\right) + \left((x^*)^i(c)\right)^T A_j \left((x^*)^i(c)\right) \\ &\quad + \left((x^*)^i(c)\right)^T A(c) \frac{\partial \left((x^*)^i(c)\right)}{\partial c_j} \\ &= \left((x^*)^i(c)\right)^T A_j \left((x^*)^i(c)\right) + 2\lambda_i^*(c) \left((x^*)^i(c)\right)^T \frac{\partial \left((x^*)^i(c)\right)}{\partial c_j} . \end{aligned}$$

Differentiating (127) with respect to c_j yields

$$2\left((x^*)^i(c)\right)^T \frac{\partial\left((x^*)^i(c)\right)}{\partial c_j} = 0 ,$$

thus

$$\frac{\lambda_i^*(c)}{\partial c_j} = \left((x^*)^i(c)\right)^T A_j\,(x^*)^i(c) . \tag{128}$$

Given c, the vectors $\lambda^*(c)$ and $(x^*)^i(c)$, $i = 1, \ldots, n$, can be computed approximately using any of the software packages mentioned in the beginning of this article. Therefore the matrix and the righthand side of the Newton equation

$$\left(\left((x^*)^i(c)\right)^T A_j\,(x^*)^i(c)\right)(c^{k+1} - c^k) = -(\lambda(c^k) - \lambda^*) \tag{129}$$

are known at least approximately; hence c^{k+1} can be computed from (129). Stopping the iteration and verifying c^* can now be done following the lines of [8]. In the verification step, $f'(c)$ has to be enclosed by a tight interval vector, i. e., the eigenvalues of $A(c^k)$ have to be enclosed by one of the methods described in Section 5.1 or 6. In addition, an enclosure of $f'(c)$ is needed with c varying in an interval vector $[c]$. This means that the normalized eigenvectors $(x^*)^i(c)$, $c \in [c]$ have to be enclosed. Again the methods in the Sections 5.1 or 6 can be applied, this time on the interval matrix $A([c]) = A_0 + \sum_{i=1}^{n}[c]_i A_i$ showing that enclosure methods are also needed for a whole set of point problems and not only for a single one.

The numerical examples which follow are taken from [9]. They prove that the inverse eigenvalue problem can have several solutions, a phenomenon which has already been remarked in [38].

Example 39

The matrices $A_i \in M_{5,5}(\boldsymbol{R})$, $i = 0, \ldots, 5$, are given by

$$A_0 = \begin{pmatrix} 6 & 1 & 3 & -2 & 0 \\ 1 & 2 & 2 & 0 & 4 \\ 3 & 2 & 1 & 2 & 0 \\ -2 & 0 & 2 & -2 & 0 \\ 0 & 4 & 0 & 0 & -3 \end{pmatrix} , \quad A_1 = \begin{pmatrix} 2 & 1 & 0 & -1 & 1 \\ 1 & 0 & -4 & -1 & 0 \\ 0 & -4 & -2 & 1 & 3 \\ -1 & -1 & 1 & 0 & 5 \\ 1 & 0 & 3 & 5 & -1 \end{pmatrix} ,$$

$$A_2 = \begin{pmatrix} 1 & 2 & -3 & 0 & -1 \\ 2 & -1 & -3 & 1 & 0 \\ -3 & -3 & 0 & -2 & 2 \\ 0 & 1 & -2 & 0 & 6 \\ -1 & 0 & 2 & 6 & 1 \end{pmatrix} , \quad A_3 = \begin{pmatrix} 2 & -1 & 0 & 2 & 1 \\ -1 & 2 & 1 & 0 & -6 \\ 0 & 1 & -3 & 8 & -3 \\ 2 & 0 & 8 & 6 & -3 \\ 1 & -6 & -3 & -3 & 4 \end{pmatrix} ,$$

$$A_4 = \begin{pmatrix} -3 & -2 & 2 & 0 & 4 \\ -2 & 1 & 2 & -4 & 0 \\ 2 & 2 & -2 & -1 & 2 \\ 0 & -4 & -1 & -5 & 0 \\ 4 & 0 & 2 & 0 & 1 \end{pmatrix} , \quad A_5 = \begin{pmatrix} -3 & -1 & -5 & 3 & 2 \\ -1 & 2 & 7 & -1 & -2 \\ -5 & 7 & 5 & -3 & 0 \\ 3 & -1 & -3 & 0 & -2 \\ 2 & -2 & 0 & -2 & 4 \end{pmatrix} .$$

The eigenvalues λ_i^* are prescribed by

$\lambda^* = (-10, -5, -1, 4, 10)^T$.

As can be seen by a direct computation, the vector $c^* = (-3, 4, 1, 2, -1)^T$ is a solution of the problem. It is verified when starting the Newton iteration with $c^0 = (-2.9, 4.1, 0.9, 2.01, -1.01)^T$. With this iterative process we got the enclosure

$$[c] = \begin{pmatrix} -3 \\ 4 \\ 1 \\ 2 \\ -1 \end{pmatrix} + 10^{-15} \cdot \begin{pmatrix} [-2, 1] \\ [-1, 1] \\ [-1, 1] \\ [-1, 1] \\ [-1, 1] \end{pmatrix} .$$

Starting with $c^0 = (10, 10, 10, 10, 10)^T$ yields

$$[c] = \begin{pmatrix} [& -3.879\,049\,564\,183\,73^{0}_{6} &] \\ [& 4.305\,375\,937\,429\,0^{95}_{89} &] \\ [& 0.729\,062\,953\,735\,3^{612}_{577} &] \\ [& 1.682\,982\,632\,583\,79^{7}_{3} &] \\ [& -1.092\,532\,116\,503\,92^{6}_{9} &] \end{pmatrix}$$

which means that

$$c^{**} = \begin{pmatrix} -3.879\,049\,564\,183\,7\ldots \\ 4.305\,375\,937\,429\,0\ldots \\ 0.729\,062\,953\,735\,3\ldots \\ 1.682\,982\,632\,583\,7\ldots \\ -1.092\,532\,116\,503\,9\ldots \end{pmatrix}$$

is another solution of the problem.

$\square$

Our second example originates from [39], where an approximation of c^* has been derived.

Example 40

$$A_0 = \begin{pmatrix} 0 & 4 & -1 & 1 & 1 & 5 & -1 & 1 \\ 4 & 0 & -1 & 2 & 1 & 4 & -1 & 2 \\ -1 & -1 & 0 & 3 & 1 & 3 & -1 & 3 \\ 1 & 2 & 3 & 0 & 1 & 2 & -1 & 4 \\ 1 & 1 & 1 & 1 & 0 & 1 & -1 & 5 \\ 5 & 4 & 3 & 2 & 1 & 0 & -1 & 6 \\ -1 & -1 & -1 & -1 & -1 & -1 & 0 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 0 \end{pmatrix} , \qquad A_i = e^{(i)}(e^{(i)})^T, \quad i = 1, \ldots, 8 ,$$

$\lambda^* = (10, 20, 30, 40, 50, 60, 70, 80)^T$.

Starting with $c^0 = (10, 20, 30, 40, 50, 60, 70, 80)^T$ yields

$$[c] = \begin{pmatrix} [& 1.190\,787\,610\,247\,27^{2}_{0}E+001 &] \\ [& 1.970\,552\,150\,808\,69^{9}_{8}E+001 &] \\ [& 3.054\,549\,818\,697\,70^{5}_{4}E+001 &] \\ [& 4.006\,265\,748\,844\,80^{5}_{3}E+001 &] \\ [& 5.158\,714\,029\,072\,5^{51}_{48}E+001 &] \\ [& 6.470\,213\,143\,217\,9^{53}_{48}E+001 &] \\ [& 7.017\,067\,582\,089\,11^{8}_{3}E+001 &] \\ [& 7.131\,849\,917\,021\,90^{9}_{4}E+001 &] \end{pmatrix}, \text{ hence } c^* = \begin{pmatrix} 11.907\,876\,102\,472\ldots \\ 19.705\,521\,508\,086\ldots \\ 30.545\,498\,186\,977\ldots \\ 40.062\,657\,488\,448\ldots \\ 51.587\,140\,290\,725\ldots \\ 64.702\,131\,432\,179\ldots \\ 70.170\,675\,820\,891\ldots \\ 71.318\,499\,170\,219\ldots \end{pmatrix}.$$

Starting with $c^0 = (-10, -10, -30, -30, -50, -50, -70, -70)^T$ results in

$$[c] = \begin{pmatrix} [& 1.146\,135\,429\,773\,86^{4}_{3}E+001 &] \\ [& 7.888\,082\,936\,085\,4^{30}_{27}E+001 &] \\ [& 6.835\,339\,960\,285\,13^{6}_{2}E+001 &] \\ [& 4.987\,833\,041\,174\,66^{7}_{4}E+001 &] \\ [& 5.916\,891\,783\,339\,2^{30}_{28}E+001 &] \\ [& 3.041\,047\,014\,754\,0^{40}_{38}E+001 &] \\ [& 2.483\,432\,401\,438\,6^{21}_{18}E+001 &] \\ [& 3.701\,237\,433\,149\,02^{3}_{0}E+001 &] \end{pmatrix},$$

guaranteeing a different solution

$$c^{**} = \begin{pmatrix} 11.461\,354\,297\,738\ldots \\ 78.880\,829\,360\,854\ldots \\ 68.353\,399\,602\,851\ldots \\ 49.878\,330\,411\,746\ldots \\ 59.168\,917\,833\,392\ldots \\ 30.410\,470\,147\,540\ldots \\ 24.834\,324\,014\,386\ldots \\ 37.012\,374\,331\,490\ldots \end{pmatrix}.$$

□

10. ADDITIONAL REMARKS

In this section we want to supplement our article by bibliographical notes and additional topics appearing in verification numerics for eigenproblems.

The results of Section 3 on quadratic systems are due to Alefeld [6]. The systematic application of Theorem 4 in the Sections 5 – 9 is new.

The contents of Section 4 are well-known, the unified treatment of the inequalities in the Theorems 11 and 13 can be found in [94]. Unifying results on eigenvalues are also derived in [3].

One of the first papers concerning enclosures for eigenproblems with interval methods is certainly that of Krawczyk [70], in which he already uses the method of Section 5.1 to enclose eigenpairs. (Cf. also [71], [72].) This procedure of expanding the function f from (40) in a Taylor series seems to go back to Unger [132] in the non–interval case. Later on, Rump used in [114], [115], [116] the same access to state and prove Theorem 17. In his computations he already used C in the form (66), and he exploited there (67); cf. also Section 4 in [48]. Similarly to [127] Alefeld started in [4] from the beginning with the modification of f which yields $\hat{g}$ of Section 5.1, and proved Theorem 23 for this case. In [88] it is shown, among others, that the procedures of Alefeld and Rump are equivalent and that therefore Theorem 17 holds for $\hat{g}$ and Theorem 23 is valid for g, too. (Cf. also [87] and [89].) Note that the convergence results in [70] – [72] also fit in this theory and that the Taylor expansion above was also used in [93] and [111]. The normalization $(x^*)^T x^* = 1$ has been considered in [70].

In [72] the case of *complex* eigenpairs for *real* matrices was reduced to a real system with $n+2$ unknowns by splitting $x = x_{re} + ix_{im} \in V_n(\boldsymbol{C})$, $\lambda = \lambda_{re} + i\lambda_{im} \in \boldsymbol{C}$ into real and imaginary part. This yields

$$\begin{aligned}&\{(A-\lambda_{re}I_n)^2 + (\lambda_{im})^2 I_n\}x_{re} = 0\\ &(x_{re})_{i_0} = 1, \quad (Ax_{re})_{i_0} - \lambda_{re} = 0\end{aligned}$$

which is equivalent to

$$Ax = \lambda x\ , \quad x_{i_0} = 1$$

via $x_{im} = -\dfrac{1}{\lambda_{im}}(A - \lambda_{re}I_n)x_{re}$, provided $\lambda_{im} \neq 0$. Again Taylor's expansion is used to end up with a method analogously to that in Section 5.1 .

Complex matrices in combination with a Taylor expansion are also considered in [48], [86], [117], again with similar results as above.

Methods with higher order of convergence are derived in [13] using higher order divided differences and the interval Gaussian algorithm.

The results in Section 5.2 for double or nearly double eigenvalues are due to Alefeld and Spreuer [15] based on an idea of [35]. They already used the modified function $\hat{g}$. Enclosures for multiple eigenvalues and for a basis of the associated subspaces are also considered in [73].

The particular case of symmetric matrices seems to be handled first by Alefeld and Herzberger in [11], restricting themselves there to tridiagonal matrices A. The method starts with the characteristic equation

$$p_n(\lambda) := \det(\lambda I_n - A) = \prod_{j=1}^{n}(\lambda - \lambda_j^*) = 0 \tag{130}$$

with λ_j^*, $j = 1, \ldots, n$, being the real eigenvalues of A; $\lambda_1^*, \ldots, \lambda_{n-m}^*$ are assumed to be simple whereas the other eigenvalues λ_j^* can be simple or multiple. Solving (130) for the i–th eigenvalue yields

$$\lambda_i^* = \lambda - \frac{p_n(\lambda)}{\prod\limits_{\substack{j=1\\ j\neq i}}^{n}(\lambda - \lambda_j^*)}\ .$$

This forms the basis for the two iterative processes

$$[\lambda]_i^{(k+1)} = \left\{ \tilde{\lambda}_i^{(k)} - \frac{p_n(\tilde{\lambda}_i^{(k)})}{\prod\limits_{\substack{j=1\\ j\neq i}}^{n-m} (\tilde{\lambda}_i^{(k)} - [\lambda]_j^{(k)}) \prod\limits_{j=n-m+1}^{n} (\tilde{\lambda}_i^{(k)} - [\lambda]_j^{(0)})} \right\} \cap [\lambda]_i^{(k)} , \qquad i = 1, \ldots n-m , \tag{131}$$

and

$$[\lambda]_i^{(k+1)} = \left\{ \tilde{\lambda}_i^{(k)} - \frac{p_n(\tilde{\lambda}_i^{(k)})}{\prod\limits_{j=1}^{i-1} (\tilde{\lambda}_i^{(k)} - [\lambda]_j^{(k+1)}) \prod\limits_{j=i+1}^{n-m} (\tilde{\lambda}_i^{(k)} - [\lambda]_j^{(k)}) \prod\limits_{j=n-m+1}^{n} (\tilde{\lambda}_i^{(k)} - [\lambda]_j^{(0)})} \right\} \cap [\lambda]_i^{(k)} , \qquad i = 1, \ldots n-m , \tag{132}$$

respectively, with $\lambda_j^* \in [\lambda]_j^{(0)}$, $j = 1, \ldots, n$, $\tilde{\lambda}_i^{(k)} \in [\lambda]_i^{(k)}$, $i = 1, \ldots, n-m$, $[\lambda]_i^{(0)} \cap [\lambda]_j^{(0)} = \emptyset$, $i \leq n-m$, $j = 1, \ldots, n$, $j \neq i$. The characteristic polynomial is evaluated via the usual recursive formulae for symmetric tridiagonal matrices (cf. [124], e. g.). For convergence results and modifications of (131), (132) we refer to [11] and [12].

The enclosure method described in Section 6 is due to Lohner. He gives a very detailed description of the method in [79] using Gershgorin's theorem. Applying the Temple quotient was proposed in [86]. We also mention a paper of Yamamoto [143] for more sophisticated bounds, and we want to address the attention to [31].

A similar approach as in [79] was used by Klug in [68] based on a result of Heindl [49] to improve orthonormality of nearly othonormal sets of vectors.

A different method for verifying eigenvalues of symmetric matrices was derived by Behnke [21] using results of [43]. Here, a smaller generalized eigenvalue problem is associated with the original one. From the eigenvalues of this smaller problem one can compute bounds for those of the given one. We refer to the contribution [25] of this book for more details, and also to [22], [23].

We finally mention the paper [36] in which guaranteed bounds for the eigenvalues of Hermitian matrices were computed.

The generalized eigenproblem (100) was investigated in [7], [117] by the methods outlined in Section 7. There, the theoretical results can be found which were only indicated in this section. The problem can also be handled by the method in [21]. An overview can be found in [86].

The verification method in Section 8 for the singular value problem is again due to Alefeld. Based on a paper of Dongarra [34], he uses in [5] the function f from (112), and he already states and proves Theorem 35. In [80] the singular value problem (108) was reduced to the eigenvalue problem

$$B \begin{pmatrix} u \\ v \end{pmatrix} = \lambda \begin{pmatrix} u \\ v \end{pmatrix} , \tag{133}$$

with $B := \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix} \in M_{n+m,m+n}(\boldsymbol{R})$, and with the normalization

$$\left\| \begin{pmatrix} u \\ v \end{pmatrix} \right\|_2^2 = \|v\|_2^2 + \|u\|_2^2 = 2 \ . \tag{134}$$

Each eigenvalue of (133) satisfies

$$Au = \lambda v \ , \tag{135}$$
$$A^T v = \lambda u \ , \tag{136}$$

hence

$$\begin{aligned} A^T A u &= \lambda A^T v = \lambda^2 u \ , \\ A A^T v &= \lambda A u = \lambda^2 v \ . \end{aligned}$$

Thus, each positive eigenvalue of B is a non–zero singular value of A . By (136), (136) the corresponding eigenvectors $\begin{pmatrix} u \\ v \end{pmatrix}$, normalized by (134), satisfy

$$\|u\|_2^2 = \left(\frac{1}{\lambda} A^T v\right)^T \cdot u = \frac{1}{\lambda} \lambda v^T v = \|v\|_2^2 \ ,$$

hence by (134) we obtain $\|u\|_2^2 = \|v\|_2^2 = 1$. Thus u, v are normalized right and left singular vectors, respectively. In this way the procedure of Section 6 can be applied to B to enclose all singular values and the corresponding singular vectors – at least theoretically. For practical computations we mention the difficulties in deciding whether an eigenvalue is zero, if it is enclosed in a tight zero containing interval, and we also merely point out that it is nearly impossible in practice to determine the order of the eigenvalues in a cluster and therefore to guarantee the dimension of the corresponding eigenspaces. As in Section 5.2 one can only enclose a basis of the eigenspace which contains all eigenvectors of the cluster.

We also mention the paper [33] of Deif in which singular values of interval matrices are considered, and the thesis [56] of Hoffmann in which error bounds and verification algorithms are presented for the generalized singular value decomposition.

Even when dealing with point matrices, topic 4 of Section 1 has to be considered when executing the algorithms on a computer, since the verification steps always have to be done in interval arithmetic. For the theoretical treatment of this topic we mention the papers of [32], [86], [88] pointing out that enclosures for *interval* matrices $[A]$ are certainly not handled there for the first time. In this connection we also cite the paper [118] of Rump, in which he shows a way how to obtain inner and outer enclosures for the eigenvalues of all matrices contained in $[A]$. (Cf. [117] and [118] for the same question concerning the generalized eigenvalue problem.) An *inner enclosure* $[\lambda]_i^{int}$ of the i–th eigenvalue of $[A]$ is an interval such that any $\lambda \in [\lambda]_i^{int}$ is the i–th eigenvalue of some matrix $A \in [A]$. In contrast to this an *outer enclosure* $[\lambda]_i^{out}$ is an enclosure $[\lambda]_i$ as described in topic 4.

Enclosures for simple zeros of real λ–matrices, i. e. of matrix polynomials

$$P(\lambda) := A_0 + \lambda A_1 + \ldots + \lambda^m A_m, \quad \lambda \in \boldsymbol{R},\ A_i \in M_{nn}(\boldsymbol{R}),\ i = 0, \ldots, m,$$

are considered in [112]. A *zero* of P is a value λ^* for which there exists a vector $x^* \in V_n(\boldsymbol{R})\backslash\{0\}$ such that

$$P(\lambda^*)x^* = 0 \tag{137}$$

holds. Note that (137) is equivalent to the eigenproblem $Bx = \lambda x$ with

$$B := \begin{pmatrix} O & I_n & & O \\ \vdots & & \ddots & \\ O & & & I_n \\ -A_m^{-1}A_0 & -A_m^{-1}A_1 & \cdots & -A_m^{-1}A_{m-1} \end{pmatrix} \in M_{mn,mn}(\boldsymbol{R}) ,$$

provided A_m is non–singular. In [112] Taylor's expansion is again applied. This time the starting function reads

$$\begin{pmatrix} P(\lambda)x \\ x_n - 1 \end{pmatrix} .$$

We also mention [113] for enclosures of zeros of non–real λ–matrices and [24] for another way to enclose eigenvalues of parameter dependent matrices.

In [105] enclosures for real eigenpairs of matrices

$$A \in [A] := A_0 + [-1,1]\,qp^T,\ 0 \le q \in V_n(\boldsymbol{R}),\ 0 < p \in V_n(\boldsymbol{R})$$

are considered. Theoretical results are given.

Stability results used in control theory are derived in [106], [107], [109]. There, an interval matrix $[A] \in M_{nn}(I(\boldsymbol{R}))$ is called *Hurwitz stable* if each matrix $A \in [A]$ has only eigenvalues λ^* with $Re\{\lambda^*\} < 0$. Similarly, $[A]$ is called *Schur stable* if $\rho(A) < 1$ for each matrix $A \in [A]$. (In [109] only symmetric matrices $A \in [A] = [A]^T$ are required to have this latter property.) Stability is also studied in [17], [20], [26], [29], [30], [46], [47], [51] – [54], [57], [59], [60], [62], [63], [69], [82] – [84], [92], [100], [102], [103], [119], [121], [122], [141], [142], [144]. An overview is given in [41].

In [108] it is remarked that checking stability of $[A] = [A]^T$ is NP–hard.

In [107] the problem of describing and enclosing the set

$$\Lambda := \{\lambda \in \boldsymbol{R} \mid Ax = \lambda x \text{ for some } A \in [A],\ x \neq 0\}$$

is considered. A condition for $x \in V_n(\boldsymbol{R})$ to be an eigenvector of some $A \in [A]$ was derived there. In [110], the set

$$\Lambda_S := \{\lambda \in V_n(\boldsymbol{R}) \mid Ax = \lambda x \text{ for some } A \in [A] \text{ and an } x \text{ with } Sx > 0\}$$

is described with

$$S := \mathrm{diag}(\sigma_1, \ldots, \sigma_n), \quad \sigma_i \in \{-1, +1\},\ i = 1, \ldots, n,$$

being a signature matrix.

We conclude our contribution by listing the papers [58], [61], [77], [97], which also deal with enclosures of eigenvalues.

Acknowledgement

The author thanks the referees for carefully reading the manuscript and for making valuable suggestions which improved the paper.

REFERENCES

1. E. Adams, U. Kulisch (eds.), Scientific Computing with Automatic Result Verification, Academic Press, New York, 1993.
2. L. V. Ahlfors, Complex Analysis, McGraw–Hill, New York, 1966.
3. J. Albrecht, L. Collatz and W. Velte, Numerical Treatment of Eigenvalue Problems Vol. 3, ISNM 69 (1984).
4. G. Alefeld, Berechenbare Fehlerschranken für ein Eigenpaar unter Einschluß von Rundungsfehlern bei Verwendung des genauen Skalarprodukts, Z. angew. Math. Mech., 67 (1987) 145–152.
5. G. Alefeld, Rigorous Error Bounds for Singular Values of a Matrix Using the Precise Scalar Product, in [64], 1987, 9–30.
6. G. Alefeld, Errorbounds for Quadratic Systems of Nonlinear Equations Using the Precise Scalar Product, in [76], 1988, 59–68.
7. G. Alefeld, Berechenbare Fehlerschranken für ein Eigenpaar beim verallgemeinerten Eigenwertproblem, Z. angew. Math. Mech., 68 (1988) 181–184.
8. G. Alefeld, Inclusion Methods for Systems of Nonlinear Equations – The Interval Newton Method and Modifications. This volume.
9. G. Alefeld, A. Gienger and G. Mayer, Numerical Validation for an Inverse Matrix Eigenvalue Problem, Computing, to appear.
10. G. Alefeld, A. Gienger, F. Potra, Efficient Numerical Validation of Solutions of Nonlinear Systems, SIAM J. Numer. Anal., 31 (1994) 252–260.
11. G. Alefeld and J. Herzberger, Über die Verbesserung von Schranken für die Eigenwerte symmetrischer Tridiagonalmatrizen, Angew. Informatik, 16 (1974) 27–35.
12. G. Alefeld and J. Herzberger, Introduction to Interval Computations, Academic Press, New York, 1983.
13. G. Alefeld, B. Illg, F. Potra, An Enclosure Method with Higher Order of Convergence – Applications to the Algebraic Eigenvalue Problem, Int. Ser. Numer. Math., 96 (1991) 9–22.
14. G. Alefeld and G. Mayer, A Computer Aided Existence and Uniqueness Proof for an Inverse Matrix Eigenvalue Problem, J. Interval Computations, to appear. Preprint: Report 93/2, Institut für Wissenschaftliches Rechnen, Universität Karlsruhe, Karlsruhe, 1993.
15. G. Alefeld and H. Spreuer, Iterative Improvement of Componentwise Errorbounds for Invariant Subspaces Belonging to a Double or Nearly Double Eigenvalue, Computing, 36 (1986) 321–334.
16. E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov and D. Sorensen, LAPACK Users' Guide, SIAM, Philadelphia, 1992.
17. M. B. Argoun, On Sufficient Conditions for the Stability of Interval Matrices, Int. J. Control, 44 (1986) 1245–1250.
18. L. Atanassova and J. Herzberger (eds.), Computer Arithmetic and Enclosure Methods, Elsevier (North–Holland), Amsterdam, 1992.
19. W. Auzinger and H. J. Stetter, Accurate Arithmetic Results for Decimal Data and Non-Decimal Computers, Computing, 35 (1985) 141–151.
20. B. R. Barmish and C. V. Hollot, Counter–example to a Recent Result on the Stability of Interval Matrices by S. Białas, Int. J. Control, 39 (1984) 1103–1104.
21. H. Behnke, Inclusion of Eigenvalues of General Eigenvalue Problems for Matrices, in [76], 1988, 69–78.
22. H. Behnke, The Determination of Guaranteed Bounds to Eigenvalues with the Use of Variational Methods II, in [129], 1990, 155–170.
23. H. Behnke, The Calculation of Guaranteed Bounds for Eigenvalues Using Complementary Variational Principles, Computing, 47 (1991) 11–27.
24. H. Behnke, Bounds for Eigenvalues of Parameter–Dependent Matrices, Computing, 49 (1992) 159–167.
25. H. Behnke and F. Goerisch, Inclusions for Eigenvalues of Selfadjoint Problems. This volume.
26. S. Białas, A Necessary and Sufficient Condition for the Stability of Interval Matrices, Int. J. Control, 37 (1983) 717–722.
27. S. D. Chatterji, B. Fuchssteiner, U. Kulisch, R. Liedl (eds.), Jahrbuch Überblicke Mathematik 1994,

Vieweg, Braunschweig, 1994.
28. L. Collatz, Einschließungssatz für charakteristische Zahlen von Matrizen, Math. Z., 48 (1942) 221–226.
29. D. Cordes, Verifizierter Stabilitätsnachweis für Lösungen von Systemen periodischer Differentialgleichungen auf dem Rechner mit Anwendungen, Thesis, Universität Karlsruhe, 1987.
30. Xu Daoyi, Simple Criteria for Stability of Interval Matrices, Int. J. Control, 38 (1985) 289–295.
31. A. S. Deif, Realistic A Priori and A Posteriori Error Bounds for Computed Eigenvalues, IMA J. Numer. Anal., 9 (1990) 323–329.
32. A. S. Deif, The Interval Eigenvalue Problem, Z. angew. Math. Mech., 71 (1991) 61–64.
33. A. S. Deif, Singular Values of an Interval Matrix, Linear Algebra Appl., 151 (1991) 125–133.
34. J. J. Dongarra, Improving the Accuracy of Computed Singular Values, SIAM J. Sci. Stat. Comput., 4 (1983) 712–719.
35. J. J. Dongarra, C. B. Moler and J. H. Wilkinson, Improving the Accuracy of Computed Eigenvalues and Eigenvectors, SIAM J. Numer. Anal., 20 (1983) 23–45.
36. K. V. Fernando and M. W. Pont, Computing Accurate Eigenvalues of a Hermitian Matrix, in [131], 1989, 104–115.
37. G. E. Forsythe and P. Henrici, The Cyclic Jacobi Method for Computing the Principal Values of a Complex Matrix, Trans. Amer. Math. Soc., 94 (1960) 1–23.
38. S. Friedland, Inverse Eigenvalue Problems, Linear Algebra Appl., 17 (1977) 15–51.
39. S. Friedland, J. Nocedal, M. L. Overton, The Formulation and Analysis of Numerical Methods for Inverse Eigenvalue Problems, SIAM J. Numer. Anal., 24 (1987) 634–667.
40. B. S. Garbow, J. M. Boyle, J. J. Dongarra and C. B. Moler, Matrix Eigensystem Routines – EISPACK Guide Extension, Lecture Notes in Computer Science 51, Springer, Berlin, 1977.
41. J. Garloff and N. K. Bose, Boundary Implications for Stability Properties: Present Status, in [M88], 391–402.
42. F. Goerisch and J. Albrecht, Eine einheitliche Herleitung von Einschließungssätzen für Eigenwerte, in [3], 1984, 58–88.
43. F. Goerisch and H. Haunhorst, Eigenwertschranken für Eigenwertaufgaben mit partiellen Differentialgleichungen, Z. angew. Math. Mech., 65 (1985) 129–135.
44. G. H. Golub and C. F. van Loan, Matrix Computations, Second Edition, Johns Hopkins University Press, Baltimore, Maryland, 1989.
45. R. T. Gregory and D. L. Karney, A Collection of Matrices for Testing Computational Algorithms, Wiley, New York, 1969.
46. W. M. Grimm and U. Piechottka, Comments on 'Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices', Int. J. Control, 49 (1989) 1099–1102.
47. B. Groß, Verification of Asymptotic Stability for Interval Matrices and Applications in Control Theory, in [1], 1993, 357–395.
48. K. Grüner, Solving the Complex Algebraic Eigenvalue Problem with Verified High Accuracy, in [131], 1989, 59–78.
49. G. Heindl, An Efficient Method for Improving Orthonormality of Nearly Orthonormal Sets of Vectors, in [130], 1990, 83–91.
50. P. Henrici, On the Speed of Convergence of Cyclic and Quasi–Cyclic Jacobi Methods for Computing Eigenvalues of Hermitian Matrices, SIAM J. Appl. Math., 6 (1958) 144–162.
51. D. Hertz, Root Clustering of Interval Matrices, Birkhäuser Symp. on Fundamentals of Discrete–Time Systems, 1992.
52. D. Hertz, The Extreme Eigenvalues and Stability of Real Symmetric Interval Matrices, IEEE Trans. Autom. Contr., 37 (1992) 532–535.
53. D. Hertz, The Extreme Eigenvalues and Stability of Hermitian Interval Matrices, IEEE Trans. Circuits Syst., I, Fundam. Theory Appl., 39 (1992) 463–466.
54. D. Hertz, The Maximal Eigenvalue and Stability of a Class of Real Symmetric Interval Matrices, IEEE Trans. Circuits Syst., I, Fundam. Theory Appl., 40 (1993) 56–57.
55. J. Herzberger, Basic Definitions and Properties of Interval Arithmetic. This volume.
56. R. Hoffmann, Konstruktion von Fehlerschranken bei der verallgemeinerten Singulärwertzerlegung und ihre iterative Verbesserung, Thesis, Universität Karlsruhe, 1993.

57. C. V. Hollot and A. C. Bartlett, On the Eigenvalues of Interval Matrices, Proc. 26th IEEE Conference on Decision and Control, Los Angeles, CA, 1987, 794–799.
58. D. Hudak, Ein Eigenwertproblem für Intervall–Matrizen, Z. angew. Math. Mech., 64 (1984) 503–505.
59. C.–L. Jiang, Another Sufficient Condition for the Stability of Interval Matrices, Int. J. Control, 47 (1988) 181–186.
60. C.–L. Jiang, Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices, Int. J. Control, 47 (1988) 1563–1565.
61. Z. Jin, An Interval Method for Bounding the Eigenvalue and Eigenvectors of a Symmetric Matrix, Freiburger Intervall–Berichte 87/7, Universität Freiburg, 1987, 17–21.
62. W. C. Karl, J. P. Greschak and G. C. Verghese, Comments on 'A Necessary and Sufficient Condition for the Stability of Interval Matrices', Int. J. Control, 39 (1984) 849–851.
63. W. C. Karl and G. C. Verghese, Comments on 'Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices', Int. J. Control, 48 (1988) 2159–2160.
64. E. Kaucher, U. Kulisch and Ch. Ullrich (eds.), Computerarithmetic, Teubner, Stuttgart.
65. E. Kaucher, S. M. Markov and G. Mayer (eds.), Computer Arithmetic, Scientific Computation and Mathematical Modelling, IMACS Annals on Computing and Applied Mathematics, 12, Baltzer, Basel, 1991.
66. R. Klatte, U. Kulisch, C. Lawo, M. Rauch, A. Wiethoff, C–XSC, A C++ Class Library for Extended Scientific Computing, Springer, Berlin, 1993.
67. R. Klatte, U. Kulisch, M. Neaga, D. Ratz and Ch. Ullrich, PASCAL–XSC, Sprachbeschreibung mit Beispielen, Springer, Berlin, 1991.
68. U. Klug, Verified Inclusions for Eigenvalues and Eigenvectors of Real Symmetric Matrices, in [130], 1990, 111–125.
69. S. R. Kolla and J. B. Farison, Counter–example to 'Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices', Int. J. Control, 48 (1988) 1751–1752.
70. R. Krawczyk, Iterative Verbesserung von Schranken für Eigenwerte und Eigenvektoren reeller Matrizen, Z. angew. Math. Mech., 48 (1968) T80–T83.
71. R. Krawczyk, Fehlerabschätzung reeller Eigenwerte und Eigenvektoren von Matrizen, Computing, 4 (1969) 281–293.
72. R. Krawczyk, Verbesserung von Schranken für Eigenwerte und Eigenvektoren von Matrizen, Computing, 5 (1970) 200–206.
73. O. Kreß, Über Fehlerabschätzungen für Eigenwerte und Eigenvektoren reellsymmetrischer Matrizen, Thesis, Interner Bericht 73/4, Institut für Praktische Mathematik, Universität Karlsruhe, 1973.
74. N. Krylov and N. Bogolyubov, Bull. Acad. Sci. USSR, Classe Phys. Math. Leningrad, 471 (1929).
75. U. W. Kulisch and W. L. Miranker (eds.), A New Approach to Scientific Computation, Academic Press, New York, 1983.
76. U. Kulisch and H. J. Stetter (eds.), Scientific Computation with Automatic Result Verification, Computing, Suppl. 6 (1988).
77. Q. Lian and Z. Jin, A Method for Solving the Eigenproblem of a Symmetric Interval Matrix, Freiburger Intervall–Berichte 87/7, Universität Freiburg, 1987, 9–14.
78. P.–O. Löwdin, Studies in Perturbation Theory, IV, Solution of Eigenvalue Problem by Projection Operator Formalism, J. Math. Phys., 3 (1962) 969–982.
79. R. Lohner, Enclosing all Eigenvalues of Symmetric Matrices, in [131], 1989, 87–103.
80. R. Lohner, Einschließung der singulären Werte und Vektoren reeller Matrizen, GAMM–Vortrag, Karlsruhe, 1989.
81. R. J. Lohner, Interval Arithmetic in Staggered Correction Format, in [1], 1993, 301–321.
82. M. Mansour, Robust Stability of Interval Matrices, Proc. 28th Conference on Decision and Control, Tampa, FL, 1989.
83. G. Mayer, On the Convergence of Powers of Interval Matrices, Linear Algebra Appl., 58 (1984) 201–216.
84. G. Mayer, On the Convergence of Powers of Interval Matrices (2), Numer. Math., 46 (1985) 69–83.
85. G. Mayer, Old and New Aspects on the Interval Gaussian Algorithm, in [65], 1991, 329–349.
86. G. Mayer, Enclosures for Eigenvalues and Eigenvectors, in [18], 1992, 49–68.
87. G. Mayer, Taylor-Verfahren für das algebraische Eigenwertproblem, Z. angew. Math. Mech., 73

(1993) T857–T860.
88. G. Mayer, A Unified Approach to Enclosure Methods for Eigenpairs, Z. angew. Math. Mech., 74 (1994) 115-128.
89. G. Mayer, Ergebnisverifikation beim algebraischen Eigenwertproblem, in [27], 1994, 111–130.
90. G. Mayer, Epsilon–Inflation in Verification Algorithms. J. Comp. Appl. Math., to appear.
91. R. E. Moore (ed.), Reliability in Computing, The Role of Interval Methods in Scientific Computing, Academic Press, Boston, 1988.
92. T. Mori and H. Kokame, Convergence Property of Interval Matrices and Interval Polynomials, Int. J. Control, 45 (1987) 481–484.
93. V. Moynihan, Techniques for Generating Accurate Eigensolutions in ADA, in [131], 1989, 79-86.
94. A. Neumaier, Eigenwertabschätzungen für Matrizen, J. Reine und Angew. Math., 354 (1984) 206–212.
95. A. Neumaier, Interval Methods for Systems of Equations, Cambridge University Press, Cambridge, 1990.
96. K. L. E. Nickel, Interval Mathematics 1980, Academic Press, New York, 1980.
97. K. Nickel, Bounding Eigenvectors of Symmetric Matrices, in [96], 1980, 433–447.
98. J. M. Ortega, Numerical Analysis, A Second Course, Classics in Applied Mathematics, SIAM, Philadelphia, 1990.
99. B. N. Parlett, The Symmetric Matrix Eigenvalue Problem, Prentice Hall, Englewood Cliffs, N. J., 1980.
100. D. B. Petkovsky, Stability Analysis of Interval Matrices: Improved Bounds, Int. J. Control, 48 (1988) 2265–2273.
101. J. Phillips, The NAG Library: A Beginner's Guide, Clarendon, Oxford, 1986.
102. A. Rachid, Comments on 'Another Sufficient Condition for the Stability of Interval Matrices', Int. J. Control, 49 (1989) 1103.
103. A. Rachid, Comments on 'Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices', Int. J. Control, 49 (1989) 1105.
104. F. Roesler, Riemann's Hypothesis as an Eigenvalue Problem, Linear Algebra Appl., 81 (1986) 153–198.
105. J. Rohn, Real Eigenvalues of an Interval Matrix with Rank One Radius, Z. angew. Math. Mech., 70 (1990) T562–T563.
106. J. Rohn, Stability of Interval Matrices: The Real Eigenvalue Case, IEEE Trans. Autom. Control, 37 (1992), 1604–1605.
107. J. Rohn, Interval Matrices: Singularity and Real Eigenvalues, SIAM J. Matrix Anal. Appl., 14 (1993) 82–91.
108. J. Rohn, Checking Stability of Symmetric Interval Matrices is NP–hard, Preprint, 1993.
109. J. Rohn, Positive Definiteness and Stability of Interval Matrices, SIAM J. Matrix Anal. Appl., 15 (1994) 175–184.
110. J. Rohn and A. Deif, On the Range of Eigenvalues of an Interval Matrix, Computing, 47 (1992) 373–377.
111. J. Rokne, Fehlererfassung bei Eigenwerten von Matrizen, Computing, 7 (1971) 145–152.
112. J. Rokne, Errorbounds for Simple Zeros of λ–Matrices, Computing, 16 (1976) 17–27.
113. J. Rokne, Including Iterations for the Lambda–Matrix Eigenproblem, Computing, 35 (1985) 207–218.
114. S. M. Rump, Kleine Fehlerschranken bei Matrixproblemen, Thesis, Universität Karlsruhe, 1980.
115. S. M. Rump, Exakte Fehlerschranken für Eigenwerte und Eigenvektoren, Z. angew. Math. Mech., 61 (1981) T311–T313.
116. S. M. Rump, Solving Algebraic Problems with High Accuracy, in [75], 1983, 53–120.
117. S. M. Rump, Guaranteed Inclusions for the Complex Generalized Eigenproblem, Computing, 42 (1989) 225–238.
118. S. M. Rump, Zur Außen– und Inneneinschließung von Eigenwerten bei toleranzbehafteten Matrizen, Z. angew. Math. Mech., 73 (1993) T861–T863.
119. Y. Q. Shi and H. Zhang, Counter–example to 'Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices', Int. J. Control, 49 (1989) 1095–1097.

120. B. T. Smith, J. M. Boyle, J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema and C. B. Moler, Matrix Eigensystem Routines – EISPACK Guide, 2nd ed., Lecture Notes in Computer Science 6, Springer, Berlin, 1976.
121. C. B. Soh, Counter–example to 'Sufficient and Necessary Condition for the Asymptotic Stability of Discrete Linear Interval Matrices', Int. J. Control, 49 (1989) 1107–1108.
122. C. B. Soh, Necessary and Sufficient Conditions for Stability of Symmetric Interval Matrices, Int. J. Control, 51 (1990) 243–248.
123. H. J. Stetter, Sequential Defect Correction for High–Accuracy Floating–Point Algorithms, Lecture Notes in Mathematics, Vol. 1066 (1984) 186–202.
124. J. Stoer and R. Bulirsch, Introduction to Numerical Analysis, Springer, New York, 1980.
125. G. Strang, Linear Algebra and its Applications, Harcourt Brace Jovanovich, 3rd ed., San Diego, 1988.
126. F. Stummel and K. Hainer, Praktische Mathematik, Teubner, Stuttgart, 1971.
127. H. J. Symm and J. H. Wilkinson, Realistic Error Bounds for a Simple Eigenvalue and its Associated Eigenvector, Numer. Math., 35 (1980) 113–126.
128. G. Temple, The Computation of Characteristic Numbers and Characteristic Functions, Proc. London Math. Soc. (2), 29 (1929) 257–280.
129. Ch. Ullrich (ed.), Computer Arithmetic and Self-Validating Numerical Methods, Academic Press, New York, 1990.
130. Ch. Ullrich (ed.), Contributions to Computer Arithmetic and Self–Validating Numerical Methods, Baltzer, IMACS, Basel, 1990.
131. Ch. Ullrich and J. Wolff von Gudenberg (eds.), Accurate Numerical Algorithms, A Collection of Research Papers, Research Reports ESPRIT, Project 1072, DIAMOND, Vol. 1, Springer, Berlin, 1989.
132. H. Unger, Nichtlineare Behandlung von Eigenwertaufgaben, Z. angew. Math. Mech., 30 (1950) 281–282.
133. R. S. Varga, Matrix Iterative Analysis, Prentice–Hall, Englewood Cliffs, N. J., 1962.
134. A. G. Walker and J. D. Weston, Inclusion Theorems for the Eigenvalues of a Normal Matrix, J. London Math. Soc. 24 (1949) 28–31.
135. W. Walter, Analysis 2, Springer, Berlin, 1992.
136. D. H. Weinstein, Modified Ritz Methods, Proc. Nat. Acad. Sc. USA, 20 (1934) 529–532.
137. H. Wielandt, Ein Einschließungssatz für charakteristische Wurzeln normaler Matrizen, Arch. Math., 1 (1948) 348–352.
138. J. H. Wilkinson, Rounding Errors in Algebraic Processes, Notes on Applied Science 32. National Physical Laboratory, London, 1963.
139. J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon, Oxford, 1988.
140. J. H. Wilkinson and C. Reinsch, Handbook for Automatic Computation, Volume II, Linear Algebra, Springer, Berlin, 1971.
141. L. X. Xin, Necessary and Sufficient Conditions for Stability of a Class of Interval Matrices, Int. J. Control, 45 (1987) 211–214.
142. S.–J. Xu and C.–X. Shao, Comments on 'Sufficient Conditions for the Stability of Interval Matrices', Int. J. Control, 49 (1989) 379–381.
143. T. Yamamoto, Error Bounds for Computed Eigenvalues and Eigenvectors, Numer. Math., 34 (1980) 189–199.
144. R. K. Yedavalli, Stability Analysis of Interval Matrices: Another Sufficient Condition, Int. J. Control, 43 (1986) 767–772.